

UNIVERSIDAD NACIONAL DEL LITORAL

Facultad de Bioquímica y Ciencias Biológicas



Tesis para la obtención del Grado Académico de Doctor en Ciencias Biológicas

Aplicación de tecnologías de secuenciación masiva para el análisis transcripcional de células CHO-K1 y utilización de nuevas regiones promotoras endógenas para la expresión de proteínas recombinantes en cultivos celulares

Lic. Ileana del Rosario Tossolini

Director de Tesis: Dr. Claudio C. Prieto

Co-Director de Tesis: Dr. Ricardo Kratje

Lugar de realización:

Laboratorio de Cultivos Celulares

Facultad de Bioquímica y Ciencias Biológicas

Universidad Nacional del Litoral

AGRADECIMIENTOS

“La vida es como una cámara de fotos. Sólo enfócate en lo que es importante, captura los buenos momentos, saca de lo negativo un aprendizaje revelado, y si las cosas no salen como deseabas, intenta una nueva toma”.

Anónimo.



Durante el transcurso de estos años de formación doctoral, muchas son las “fotografías” que he tomado a lo largo del camino. La gran mayoría han sido retratos de buenos e inolvidables momentos, mientras que otras tantas me han permitido revelar un aprendizaje que forjó mi desarrollo y crecimiento, no sólo a nivel académico y profesional, sino también en un plano personal.

Estoy profundamente agradecida por las diversas oportunidades y experiencias que fueron enriqueciendo este recorrido. Afortunadamente he podido cruzarme con hermosas personas y visitar lugares que nunca pensé que conocería gracias a esta profesión. Asimismo, me genera una enorme satisfacción ver cómo el esfuerzo, la dedicación y la perseverancia, volcados a este trabajo de tesis durante este tiempo, hicieron posible que hoy llegue al final del camino, pero (como dice José Saramago) “resulta que era sólo una curva abierta a otro paisaje y a nuevas curiosidades”.

Sin embargo, no he transitado sola este camino, sino que son muchas las personas que me han acompañado en este andar y a las cuales debo un enorme agradecimiento, ya que “los grandes logros de cualquier persona generalmente dependen de muchas manos, corazones y mentes” (Walter Disney). Por ello, quiero agradecer inmensamente:

A la Facultad de Bioquímica y Ciencias Biológicas (FBCB) de la Universidad Nacional del Litoral (UNL) y al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por haberme dado la posibilidad de crecer profesionalmente, brindándome las herramientas necesarias para el desarrollo de este trabajo de tesis.

A Ricardo y Marina, por abrirme las puertas del Laboratorio de Cultivos Celulares, cuando en aquel momento yo era una “forastera” en la institución, dándome la oportunidad de formar parte del grupo de trabajo. Por las oportunidades que me brindaron, por su calidad humana y acompañamiento tanto en los buenos momentos como en los más difíciles que me ha tocado vivir.

A Claudio, por la confianza que depositó en mí, por creer en mis capacidades y conocimientos para comenzar un nuevo tema de trabajo. Por su entusiasmo y optimismo, por felicitarme frente a los buenos resultados, pero más aún por ayudarme

y alentarme frente a la presencia de dificultades (tanto académicas como personales). Por todas las oportunidades brindadas y los momentos compartidos ¡Muchas gracias!

A todo el personal del flamante CBL y Zelltek, y a aquellos que han pasado y han dejado su huella en estos años. Gracias por ser excelentes compañeros de trabajo y un cálido grupo humano, por tantos momentos divertidos compartidos dentro y fuera del Laboratorio.

A mis compañeros y amigos del Laboratorio, por la ayuda que me ofrecieron cada vez que fue necesario, por sus consejos, por transmitirme sus conocimientos y experiencias, por las anécdotas y risas, las rondas de mate, las charlas, y también por las salidas, los congresos y los viajes compartidos, que hicieron que cada día de trabajo sea más ameno y agradable. Gracias a mis “compañeros de oficina”, a los que siguen firme, los que estuvieron antes, los que se fueron y los que pasan de a ratos. Un agradecimiento afectuoso y especial a: Belu, Seba, Diego, Regi, Mili B., Agus, Soni, Caro, Edu, Ale Raimondi, Gustavo “el Doc”, Batta, Erne, Vero, Lupe, Mile, Jesu, Javi, Juani, Fran, Carlitos, Marian, Marcos, Nati, Gise y Andre.

A Seba A., por transmitirme generosamente sus conocimientos de biología molecular y cultivos. El día que pisé el Laboratorio, me enseñó lo que es una pipeta, cómo se usa, hasta me explicó cómo abrir y cerrar tubos eppendorf (a pesar de que volqué todo a la primera, no me olvido más de ese día). ¡Gracias por toda tu ayuda!

A Fabri, mi maestro de la Biotecnología... Gracias por ser tan buena persona y compañero de trabajo, por tu generosidad, por enseñarme con dedicación, por las anécdotas divertidas y las charlas interminables en el “gabinete psicológico” del *sorter*. Durante el tiempo que trabajamos juntos nos complementamos tanto, que ahora te dedicas a programar y yo a hacer experimentos en el laboratorio!

A Agus, por la inmensa ayuda que me brindó en la última etapa de esta tesis, por enseñarme con paciencia y esmero. Por entusiasmarse (tanto o más que yo) en la búsqueda de “los promotores” y acompañarme en esta investigación tan interesante ¡Gracias por tu predisposición y esfuerzo!

A los grandes amigos y amigas que me regaló esta tesis: Fabri, Luisi, Evange, Cele, Diane, Blondie, Sofi, Mili S., Ani y Fer. Agradezco infinitamente su amistad, por estar siempre, por escucharme, aconsejarme y acompañarme en los buenos y no tan buenos momentos, y alegrarse por cada uno de mis logros. Desde aquel día que entré al Laboratorio, me “adoptaron” y me hicieron sentir una más del grupo. A veces dudo si estudié Bioinformática o Biotecnología, porque gracias a ellos conozco todas las anécdotas de la FBCB. En el transcurso de estos años hicieron que Santa Fe sea mi segunda casa (literalmente) ¡Gracias por estar a mi lado!

A Fernando, por abrirme no solamente las puertas del Laboratorio en EE.UU., sino también las puertas de su casa. Gracias por tu extrema generosidad, por tu dedicación, por ser un gran maestro, empujándome a que cada día de un poquito más de mí, y por renovarme las ganas y el entusiasmo de seguir en esta área tan apasionante que es la investigación. Gracias también por ser un estupendo guía turístico, por las clases de *surf* y por los hermosos momentos compartidos con tu familia y amigos. Todo esto hizo de mi viaje una experiencia más que enriquecedora que nunca olvidaré.

A Ariel A., por su apoyo incondicional, por sus consejos, por su humildad y generosidad al transmitir sus conocimientos. Desde la época en que fue mi docente de grado, hasta hoy que somos compañeros de trabajo en el mundo apasionante de la educación, ha sido mi ejemplo profesional a seguir ¡Gracias por estar siempre!

A mis seres queridos, mi familia y amigos, por el apoyo incondicional recibido durante estos años, al interesarse por mi profesión y preocuparse por mis actividades. También, gracias a todas aquellas personas que se cruzaron en mi camino, que me tendieron una mano, me dieron una palabra de aliento o me regalaron una sonrisa.

A Renzo, hoy mi marido, mi compañero de viaje en este camino que es la vida. Hemos crecido juntos y allí siempre estás para acompañarme, sostenerme y alegrarte conmigo cada vez que llego a una meta. Simplemente gracias por estar a mi lado, por escucharme, por entenderme y por apoyarme en todas mis decisiones. Gracias por tus palabras de aliento, tu escucha serena y por darme siempre energías para seguir y hacer lo que me gusta. Gracias por creer en mí y hacerme mejor persona.

A mis padres, Carmen y Andrés, ya que si no fuera por ellos hoy no habría llegado hasta este punto. Sin duda han sido los mejores maestros de mi vida, porque me han enseñado con el ejemplo, con las buenas acciones, con la dedicación diaria, con el trabajo arduo y la superación. Gracias por darme con gran esfuerzo mi educación, por todos los valores que me han inculcado, por el amor que me han brindado, por los consejos y las palabras de aliento.

En especial, gracias a mi papá, mi ángel de la guarda. Gracias por ser mi gran maestro (espero haber sido una gran aprendiz). Por guiarme y aconsejarme, por ser mi ejemplo de superación y de que nada es imposible si uno se lo propone y lucha por ello. Nunca voy a olvidar lo feliz que te pusiste cuando nos enteramos que me habían otorgado la beca para hacer este Doctorado. Desde aquel momento transitaste estos cinco años a mi lado, aunque no físicamente pero sí espiritualmente; siempre sentí tu acompañamiento en todos los momentos que he vivido, desde aquellos más simples hasta los más importantes. A pesar de nuestra distancia física, siento que estás conmigo, y sé que este momento hubiera sido tan especial para vos como lo es hoy para mí. Esta tesis va dedicada especialmente a vos.

PUBLICACIONES

Parte de los resultados de la presente tesis fueron publicados en:

I Tossolini, FJ López-Díaz, R Kratje, and CC Prieto (2018) Characterization of cellular states of CHO-K1 suspension cell culture through cell cycle and RNA-sequencing profiling. *J Biotechnol* (286):56–67.

ÍNDICE

| | |
|---|------|
| ABREVIATURAS Y SÍMBOLOS | VIII |
| UNIDADES | XIV |
| RESUMEN | XV |
| ABSTRACT | XIX |
| | |
| INTRODUCCIÓN GENERAL | 1 |
| 1. Proteínas recombinantes terapéuticas | 2 |
| 2. Expresión de proteínas recombinantes terapéuticas en células eucariotas | 6 |
| 3. Células CHO | 7 |
| 4. Proceso de producción de una proteína recombinante de uso terapéutico en cultivo de células de mamífero | 8 |
| 5. Estrategias para la generación de líneas celulares productoras de proteínas recombinantes | 11 |
| 5.1. Expresión transitoria del transgén de interés | 12 |
| 5.2. Expresión estable del transgén de interés | 13 |
| 5.2.1. Amplificación del transgén de interés | 14 |
| 5.2.2. Transferencia de genes mediante el empleo de vectores lentivirales | 14 |
| 5.2.3. Ingeniería de genomas: modificaciones sitio dirigidas | 17 |
| 6. Tecnologías de secuenciación masiva de ADN | 21 |
| 7. Tecnologías “ómicas” aplicadas al estudio de líneas celulares productoras de proteínas recombinantes | 26 |
| | |
| OBJETIVOS | 30 |
| GENERAL | 31 |
| ESPECÍFICOS | 32 |
| | |
| CAPÍTULO 1: | |
| <i>Estudio del ciclo celular de células CHO-K1 cultivadas en biorreactor</i> | 33 |
| INTRODUCCIÓN | 34 |
| 1. Análisis del ciclo celular en células de mamíferos | 34 |
| 1.1. Principios básicos del ciclo celular | 34 |
| 1.2. Análisis del ciclo celular a través de citometría de flujo | 36 |
| 2. Sincronización de cultivos de células de mamífero en la fase G0/G1 para incrementar la productividad de proteínas recombinantes terapéuticas | 37 |
| | |
| OBJETIVOS | 40 |
| GENERAL | 40 |
| ESPECÍFICOS | 40 |

| | |
|--|----|
| MATERIALES Y MÉTODOS | 41 |
| 1. Técnicas de biología molecular | 41 |
| 2. Técnicas de cultivo celular | 41 |
| 2.1. Líneas celulares | 41 |
| 2.1.1. CHO-K1 (ATCC) | 41 |
| 2.2. Medios de cultivo y suplementos para células animales..... | 41 |
| 2.2.1. CHO-K1s..... | 41 |
| 2.3. Sistemas de cultivos de células..... | 41 |
| 2.3.1. Cultivos en suspensión en pequeña escala..... | 41 |
| 2.3.2. Cultivos en biorreactor tanque agitado con volumen de trabajo de un litro | 42 |
| 2.4. Determinaciones analíticas en cultivos celulares..... | 42 |
| 2.4.1. Determinación de la concentración y viabilidad de células | 42 |
| 2.4.2. Monitoreo de la concentración de glucosa, lactato y amonio en sobrenadantes de cultivo | 43 |
| 2.5. Revitalización de líneas celulares de mamífero | 44 |
| 3. Metodologías para el análisis de líneas y clones celulares..... | 45 |
| 3.1. Citometría de flujo | 45 |
| 3.1.1. Análisis del ciclo celular | 45 |
| 3.1.1.1. Fijación celular | 45 |
| 3.1.1.2. Protocolo para la tinción con yoduro de propidio | 45 |
| 3.1.1.3. Settings para citometría | 46 |
| RESULTADOS Y CONCLUSIONES | 47 |
| 1. Cultivo de células CHO-K1 en alta densidad..... | 47 |
| 2. Caracterización del ciclo celular | 51 |
| CAPÍTULO 2: <i>Análisis del transcriptoma de células CHO-K1 cultivadas en alta densidad, mediante RNA-seq</i> | 59 |
| INTRODUCCIÓN | 60 |
| 1. Secuenciación de ADN mediante tecnologías de secuenciación masiva de segunda generación (NGS)..... | 60 |
| 1.1. Tecnología <i>Illumina</i> | 60 |
| 1.1.1. Preparación de la biblioteca | 60 |
| 1.1.2. Generación de clusters | 62 |
| 1.1.3. Secuenciación..... | 62 |
| 1.1.4. Análisis de datos | 62 |
| 1.2. Orientación de las lecturas | 63 |
| 1.3. Longitud de las lecturas | 64 |

| | |
|---|----|
| 1.4. Profundidad de cobertura (<i>depth of coverage</i>) | 65 |
| 1.5. <i>Multiplexing</i> | 66 |
| 1.6. Valor de calidad de las bases asignadas a cada secuencia | 66 |
| 2. Transcriptómica..... | 68 |
| 2.1. Protocolo general para la secuenciación del transcriptoma mediante RNA-seq | 69 |
| 2.1.1. Diseño del experimento..... | 69 |
| 2.1.2. Preparación del ARN..... | 69 |
| 2.1.3. Construcción de las bibliotecas | 70 |
| 2.1.4. Secuenciación..... | 70 |
| 2.1.5. Análisis de los datos..... | 70 |
| OBJETIVOS | 73 |
| GENERAL | 73 |
| ESPECÍFICOS..... | 73 |
| MATERIALES Y MÉTODOS | 74 |
| 1. Técnicas de biología molecular | 74 |
| 1.1. Purificación de ARN total a partir de cultivos de células CHO-K1 en suspensión empleando el <i>kit RNeasy mini</i> | 74 |
| 1.2. Determinación de la concentración y pureza del ARN..... | 74 |
| 2. Técnicas de cultivo celular | 75 |
| 3. Metodologías para el análisis de líneas y clones celulares..... | 75 |
| 4. Experimentos y técnicas específicas para la secuenciación del transcriptoma (RNA-Seq) de células CHO-K1 adaptadas al crecimiento en suspensión y a medio libre de SFB..... | 76 |
| 4.1. Extracción de ARN total | 76 |
| 4.2. Control de la integridad de las muestras | 76 |
| 4.2.1. <i>RNA ScreenTape</i> (para el control de las muestras iniciales de ARN total) | 76 |
| 4.2.2. <i>High Sensitivity RNA ScreenTape</i> (para el análisis del ARNm purificado)..... | 76 |
| 4.2.3. <i>D1000 ScreenTape</i> (para el análisis de las bibliotecas de ADNc) | 77 |
| 4.3. Cuantificación de las muestras de ARN | 77 |
| 4.4. Digestión de ADN contaminante..... | 77 |
| 4.5. Purificación del ARNm y construcción de las bibliotecas de ADNc para secuenciación mediante <i>Illumina</i> | 77 |
| 4.6. Secuenciación de las bibliotecas de ADNc..... | 84 |
| 4.7. Procesamiento de los datos | 85 |
| 4.8. Validación de la expresión de los genes seleccionados por RT-qPCR..... | 85 |
| RESULTADOS Y CONCLUSIONES | 88 |
| 1. Cultivos de células CHO-K1 en alta densidad en forma de réplicas biológicas para RNA-Seq..... | 88 |

| | | |
|--|--|-----|
| 2. | Secuenciación del transcriptoma (RNA-seq) a partir de cultivos de células CHO-K1 en biorreactor..... | 90 |
| 2.1. | Extracción de las muestras de ARN total, control de calidad de las mismas y generación de las bibliotecas | 90 |
| 2.2. | Procesamiento de los datos | 95 |
| 2.2.1. | Control de calidad y pre-procesamiento de las secuencias | 95 |
| 2.2.2. | Alineamiento de las lecturas y cuantificación de los transcritos | 96 |
| 2.2.3. | Validación de la expresión de los genes seleccionados por RT-qPCR..... | 100 |
| 2.2.4. | Análisis de expresión diferencial entre las fases exponencial y estacionaria de cultivos CHO-K1 en suspensión | 102 |
| 2.2.4.1. | Análisis de la expresión de genes que responden a cambios en la temperatura | 103 |
| 2.2.5. | Análisis funcionales de los datos transcriptómicos | 104 |
| 2.2.5.1. | Enriquecimiento en términos <i>Gene Ontology</i> (GO) de los genes expresados de manera diferencial..... | 104 |
| 2.2.5.2. | Estudio de las vías metabólicas (<i>pathways</i>) en las cuales están implicados los genes diferencialmente expresados | 106 |
| 2.2.5.2.1. | Ciclo celular | 106 |
| 2.2.5.2.2. | Apoptosis | 109 |
| 2.2.5.2.3. | Metabolismo energético | 109 |
| 2.2.5.2.4. | Procesamiento del ADN | 110 |
| 2.2.5.2.5. | Procesamiento de proteínas en el retículo endoplasmático y modificaciones postraduccionales | 111 |
| 2.2.5.2.6. | Otras vías metabólicas..... | 112 |
| CAPÍTULO 3: | | |
| <i>Utilización de promotores endógenos identificados en células CHO-K1 para la producción de proteínas recombinantes terapéuticas</i> | | |
| 114 | | |
| INTRODUCCIÓN | | |
| 115 | | |
| 1. | Análisis de promotores..... | 115 |
| 1.1. | Proceso de transcripción eucariota | 115 |
| 1.2. | Uso de promotores en la industria biotecnológica para la producción de proteínas recombinantes terapéuticas..... | 117 |
| 2. | Identificación de sitios <i>hotspots</i> y edición genética en líneas celulares de mamífero..... | 121 |
| OBJETIVOS | | |
| 123 | | |
| GENERAL | | |
| 123 | | |
| ESPECÍFICOS..... | | |
| 123 | | |
| MATERIALES Y MÉTODOS | | |
| 124 | | |
| 1. | Técnicas de biología molecular | 124 |
| 1.1. | Líneas celulares procariontas..... | 124 |

| | |
|---|-----|
| 1.2. Medios de cultivo para células procariotas | 124 |
| 1.2.1. Medio LB (Luria Bertani) | 124 |
| 1.2.2. Medio LB-Antibiótico | 124 |
| 1.2.3. Agar LB..... | 124 |
| 1.2.4. Agar LB-Antibiótico | 124 |
| 1.3. Purificación de ADN plasmídico | 125 |
| 1.3.1. Minipreparación de ADN plasmídico | 125 |
| 1.3.2. Maxipreparación de ADN plasmídico | 125 |
| 1.4. Determinación de la concentración y pureza del ADN..... | 127 |
| 1.5. Digestiones enzimáticas..... | 127 |
| 1.6. Electroforesis de ADN en geles de agarosa..... | 127 |
| 1.7. Precipitación de ADN | 128 |
| 1.8. Purificación de ADN a partir de geles de agarosa | 128 |
| 1.9. Reacciones de ligación de ADN | 129 |
| 1.10. Preparación de células <i>E. coli</i> TOP 10 competentes, por el método químico.. | 129 |
| 1.11. Transformación de células competentes..... | 130 |
| 1.12. Criopreservación de cepas bacterianas | 130 |
| 1.13. Revitalización de cepas bacterianas | 130 |
| 1.14. Extracción de ADN genómico | 131 |
| 1.15. Purificación de ARN total a partir de cultivos de células CHO-K1 en suspensión | 131 |
| 1.16. Reacciones de PCR..... | 132 |
| 1.16.1. Amplificación de los promotores endógenos y de las secuencias de los brazos necesarios para recombinación homóloga (CRISPR), a partir de ADN genómico de células CHO-K1 en suspensión | 132 |
| 1.16.2. PCR de colonias (<i>Colony PCR</i>) | 134 |
| 1.16.3. Reacciones de RT-PCR..... | 136 |
| 1.16.3.1. Reacción de retrotranscripción (RT) | 136 |
| 1.16.3.2. Reacción de qPCR a partir del ADNc..... | 136 |
| 1.17. Estrategias de clonado para la generación de los vectores de expresión..... | 138 |
| 1.17.1. Esquema y descripción del plásmido de expresión eucariota pZsGreen1-1 | 138 |
| 1.17.2. Obtención del plásmido pZsGreen1-1-PromotorEndógeno | 139 |
| 1.18. Ensayos con CRISPR/Cas9 | 139 |
| 1.18.1. Amplificación y purificación del vector lentiCRISPRv2 | 139 |
| 1.18.2. Selección de las regiones <i>target</i> y diseño de los sgRNAs..... | 141 |
| 1.18.3. Clonado de los oligos en el vector lentiCRISPRv2 | 143 |
| 2. Técnicas de cultivo celular | 144 |
| 2.1. Líneas celulares | 144 |

| | |
|--|------------|
| 2.1.1. CHO-K1 (ATCC) | 144 |
| 2.1.2. HEK293T/17 (ATCC, CRL-11268)..... | 144 |
| 2.2. Medios de cultivo y suplementos para células animales..... | 144 |
| 2.2.1. CHO-K1s..... | 145 |
| 2.2.2. CHO-K1adh..... | 145 |
| 2.2.3. HEK293T/17 adherentes | 145 |
| 2.2.4. Suero fetal bovino | 145 |
| 2.2.5. Tripsina | 145 |
| 2.3. Sistemas de cultivos de células..... | 145 |
| 2.3.1. Cultivos en suspensión en pequeña escala..... | 145 |
| 2.3.2. Cultivos en adherencia..... | 146 |
| 2.4. Determinaciones analíticas en cultivos celulares..... | 146 |
| 2.4.1. Determinación de la concentración y viabilidad de células | 146 |
| 2.5. Revitalización de líneas celulares de mamífero..... | 146 |
| 2.6. Criopreservación de líneas celulares de mamífero | 147 |
| 2.7. Desarrollo de líneas celulares recombinantes | 147 |
| 2.7.1. Transfección mediante lipofección | 147 |
| 2.7.2. Transducción..... | 148 |
| 2.7.2.1. Obtención de partículas lentivirales a partir del plásmido de transferencia lentiCRISPR, codificante de Cas9-sgRNA-EGFP | 148 |
| 2.7.2.2. Concentración de LVs mediante ultracentrifugación..... | 149 |
| 2.7.2.3. Transducción lentiviral..... | 149 |
| 2.8. Experimentos con el sistema CRISPR/Cas9 en células CHO-K1 | 149 |
| 2.8.1. Experimento de knock-out con el sistema CRISPR/Cas9 en células CHO-K1 adherentes productoras de EGFP..... | 149 |
| 2.8.2. Experimento de <i>knock-in</i> mediante CRISPR/Cas9 en células CHO-K1 | 150 |
| 3. Metodologías para el análisis de líneas y clones celulares..... | 151 |
| 3.1. Citometría de flujo | 151 |
| 3.2. Microscopía de fluorescencia | 152 |
| RESULTADOS Y CONCLUSIONES | 153 |
| 1. Estudio comparativo de la expresión de cuatro genes con mayor expresión con la de un transgén previamente caracterizado | 153 |
| 2. Generación de nuevos vectores de expresión para la producción de proteínas recombinantes terapéuticas | 154 |
| 2.1. Aislamiento de promotores endógenos derivados de genes previamente identificados mediante RNA-seq con elevada actividad transcripcional en cultivos CHO-K1 en biorreactor..... | 154 |
| 2.2. Construcción de nuevos vectores de expresión, conteniendo los promotores endógenos aislados, destinados a la producción de proteínas recombinantes en células de mamífero | 156 |

| | |
|---|-----|
| 2.3. Evaluación de la actividad transcripcional de los nuevos vectores de expresión de forma transiente en células de mamífero..... | 158 |
| 3. Perspectivas futuras..... | 166 |
| 3.1. Experimentos con el sistema CRISPR/Cas9 en células CHO-K1 | 167 |
| 3.1.1. Experimento de <i>knock-out</i> con el sistema CRISPR/Cas9 en células CHO-K1 adherentes productoras de EGFP | 169 |
| 3.1.2. Experimento de <i>knock-in</i> mediante CRISPR/Cas9 en células CHO-K1 | 172 |
| CONCLUSIONES | 178 |
| BIBLIOGRAFÍA | 188 |

ABREVIATURAS Y SÍMBOLOS

| | |
|----------------|---|
| α | alpha |
| β | beta |
| γ | gamma |
| κ | kappa |
| λ | lambda |
| μ | micro |
| ψ | psi |
| Abs | <i>absorbancia</i> |
| ACSS2 | <i>acyl-CoA synthetase short chain family member 2</i> |
| Actb | <i>actin beta</i> |
| ADN | ácido desoxirribonucleico |
| ADNc | ADN copia |
| Akt | <i>AKT serine/threonine kinase</i> |
| AMPK | <i>AMP-activated protein kinase</i> |
| ARN | ácido ribonucleico |
| ARNasa – RNasa | ribonucleasa |
| ARNm | ARN mensajero |
| ARNr | ARN ribosomal |
| ARNt | ARN de transferencia |
| Asn | asparagina |
| ATCC | <i>American Type Culture Collection</i> |
| ATF | <i>alternating tangential flow</i> – flujo tangencial alternado |
| ATF4 | <i>activating transcription factor 4</i> |
| ATP | adenosín trifosfato |
| BARD1 | <i>BRCA1 associated RING domain 1</i> |
| BGH | <i>bovine growth hormone</i> |
| BHK | <i>baby hamster kidney</i> – células de riñón de hámster bebé |
| BIM | <i>BCL2 like 11</i> |
| BLM | <i>bloom syndrome RecQ like helicase</i> |
| BRCA1 | <i>breast cancer gene 1</i> |
| BRCA2 | <i>breast cancer gene 2</i> |
| BRIP1 | <i>BRCA1 interacting protein C-terminal helicase 1</i> |
| Calr | <i>calreticulin</i> |
| Cas9 | <i>CRISPR associated protein 9</i> |
| Cd63 | <i>cd63 molecule</i> |
| Cdc45 | <i>cell division cycle protein 45</i> |
| Cdc6 | <i>cell division cycle protein 6</i> |
| CDK1 | <i>cyclin dependent kinase 1</i> |
| ChIP-seq | <i>chromatin-immunoprecipitation sequencing</i> |
| CHO | <i>chinese hamster ovary</i> – células de ovario de hámster chino |
| CHO-K1adh | células CHO-K1 adherentes |

| | |
|-----------------|---|
| CHO-K1s | células CHO-K1 en suspensión |
| CHOP | <i>DDIT3, DNA damage inducible transcript 3</i> |
| Cip1 – p21 | <i>cyclin dependent kinase inhibitor 1a</i> |
| CIRP | <i>cold-inducible RNA binding protein</i> |
| CMV | citomegalovirus |
| CO ₂ | dióxido de carbon |
| col. | colaboradores |
| cPPT | <i>central polypurine tract</i> |
| CRISPR | <i>clustered regularly interspaced short palindromic repeats</i> |
| CRT | <i>cyclic reversible termination</i> |
| csp | cantidad suficiente para |
| Ctsz | <i>cathepsin z</i> |
| CV | coeficiente de variación |
| DE | diferencialmente expresados |
| DHFR | enzima dihidrofolato reductasa |
| DMEM | medio de Eagle modificado por Dulbecco |
| DMSO | dimetilsulfóxido |
| DNasa | desoxirribonucleasa |
| dNTP | desoxirribonucleótido trifosfato |
| DO | densidad óptica |
| DSB | <i>double strand break</i> |
| DsRed | <i>red fluorescent protein</i> – proteína roja fluorescente derivada de Discosoma |
| DTT | ditiotreitól |
| <i>E. coli</i> | <i>Escherichia coli</i> |
| EDTA | ácido etilendiaminotetraacético |
| EF1 α | <i>elongation factor 1α</i> - Factor de elongación 1 α |
| EGFP | <i>enhanced green fluorescent protein</i> |
| Eif3i | <i>eukaryotic translation initiation factor 3 subunit I</i> |
| EM | esclerosis múltiple |
| EPD | <i>Eukaryotic Promoter Database</i> |
| EPO | eritropoyetina |
| ERLEC2 | <i>OS9, endoplasmic reticulum lectin</i> |
| ESC | <i>embryonic stem cells</i> – células madres embrionarias |
| EST | <i>expressed sequence tags</i> |
| etc. | etcétera |
| FACS | <i>fluorescence-activated cell sorting</i> - clasificación celular activada por fluorescencia |
| FDA | <i>Food and Drug Administration</i> – Administración de Alimentos y Medicamentos |
| FDR | <i>false discovery rate</i> |
| FEN1 | <i>flap structure-specific endonuclease 1</i> |
| FGF13 | <i>fibroblast growth factor 13</i> |
| FGFR2 | <i>fibroblast growth factor receptor 2</i> |
| FPKM | <i>fragments per kilobase million</i> |
| FSC | <i>forward scatter</i> |
| FT | <i>flowthrough</i> |
| Fth1 | <i>ferritin</i> |

| | |
|-------------------------------|--|
| FVIII | factor VIII de la cascada de coagulación sanguínea |
| Fw | <i>forward</i> – en sentido directo |
| <i>g</i> | fuerza centrífuga relativa |
| GADD45 | <i>growth arrest and DNA damage inducible alpha</i> |
| <i>gag</i> | proteína de la cápside lentiviral |
| GalNAc-T | acetil-galactosamina-transferasas |
| GALNT6 | <i>polypeptide N-acetylgalactosaminyltransferase 6</i> |
| GAPDH | gliceraldehído-3-fosfato deshidrogenasa |
| Gb | giga pares de bases |
| GFP | <i>green fluorescent protein</i> – proteína verde fluorescente |
| GO | <i>Gene Ontology</i> |
| GOD | glucosa oxidasa |
| GOI | <i>gene of interest</i> – gen de interés |
| GS | glutamina sintetasa |
| <i>H. influenzae</i> | <i>Haemophilus influenzae</i> |
| H ₂ O ₂ | peróxido de hidrógeno |
| HA | <i>homology arm</i> – brazo homólogo |
| HC | <i>hierarchical clustering</i> |
| HCP | <i>host cell proteins</i> |
| HDR | <i>homology-directed repair</i> |
| HEK | <i>human embryonic kidney</i> – células embrionarias de riñón humano |
| HF | <i>hollow fiber filter</i> – filtro de fibra hueca |
| Hmox1 | <i>heme oxygenase 1</i> |
| HR | <i>homologous recombination</i> |
| Hspa5 | <i>heat shock protein family a (hsp70) member 5</i> |
| IgG1 | inmunoglobulina G |
| IGV | <i>Integrative Genomics Viewer</i> |
| Inr | <i>initiator element</i> – región iniciadora |
| IQGAP3 | <i>IQ motif containing GTPase activating protein 3</i> |
| IRES | <i>internal ribosome entry site</i> - sitio interno de entrada al ribosoma |
| ITGA5 | <i>integrin subunit alpha 5</i> |
| ITGA8 | <i>integrin subunit alpha 8</i> |
| JNK | <i>c-Jun N-terminal kinase</i> |
| Kb | kilo pares de bases |
| KEGG | <i>Kyoto Encyclopedia of Genes and Genomes</i> |
| λ | longitud de onda |
| Lamp1 | <i>lysosomal associated membrane protein 1</i> |
| LB | Luria Bertani |
| LDHA | lactato deshidrogenasa A |
| Lgals1 | <i>galectin 1</i> |
| LiCl | cloruro de litio |
| LTR | <i>long terminal repeat</i> |
| LVs | <i>lentiviral vectors</i> |
| MCB | <i>master cell bank</i> – banco celular maestro |
| MCM | complejo proteico de mantenimiento de minicromosomas |
| MCS | sitio de múltiple clonado |

| | |
|---------------------------|---|
| Met | metionina |
| MgCl ₂ | cloruro de magnesio |
| MMEJ | <i>microhomology-mediated end joining</i> |
| MP | <i>mate-pairs</i> |
| MPS1 | <i>monopolar spindle 1 kinase</i> |
| MSX | metionina sulfoximina |
| MTX | metotrexato |
| N° | número |
| NaAc | acetato de sodio |
| NaCl | cloruro de sodio |
| NaOH | hidróxido de sodio |
| NGF | <i>nerve growth factor</i> |
| NGS | <i>next generation sequencing</i> |
| NHEJ | <i>non-homologous end-joining</i> |
| NS0 | células derivadas de mieloma murino |
| nt | nucleótido |
| OD | oxígeno disuelto |
| ON | <i>overnight</i> |
| ORC | complejo de reconocimiento de origen |
| Ori | origen de replicación bacteriano |
| <i>P. pastoris</i> | <i>Pichia pastoris</i> |
| p/ | para |
| P53 | <i>tumor protein p53</i> |
| PacBio | Pacific Biosciences |
| PALB2 | <i>partner and localizer of BRCA2</i> |
| PAM | <i>protospacer adjacent motif</i> – motivo adyacente al protoespaciador |
| pb | par de bases |
| PBS | solución salina de fosfatos |
| PCA | <i>principal component analysis</i> |
| PCK1 | <i>phosphoenolpyruvate carboxykinase 1</i> |
| PCR | <i>polymerase chain reaction</i> – reacción en cadena de la polimerasa |
| PDI | <i>protein disulphide isomerase</i> |
| PE | <i>paired-end</i> |
| PEG | polietilenglicol |
| PEI | polietilenimina |
| PGK1 | <i>phosphoglycerate kinase 1</i> |
| PGM | <i>Personal Genome Machine</i> |
| PI | <i>propidium iodide</i> – yoduro de propidio |
| PI3K | <i>phosphatidylinositol-3-kinase</i> |
| POD | peroxidasa |
| <i>pol</i> | proteína polimerasa lentiviral |
| POLD1 | <i>DNA polymerase delta 1, catalytic subunit</i> |
| POLE | <i>DNA polymerase epsilon, catalytic subunit</i> |
| polyA – poly(A) – poli(A) | poliadenilación |
| Rad50 | <i>Rad50 double strand break repair protein</i> |
| Rad51B | <i>Rad51 paralog B</i> |

| | |
|----------------------|---|
| Rad51C | <i>Rad51</i> paralog C |
| Rad54L | <i>DNA repair and recombination protein Rad54-like</i> |
| RBM3 | <i>RNA binding motif protein 3</i> |
| RCLs | partículas lentivirales competentes replicativas |
| RE | retículo endoplasmático |
| Rev | proteína lentiviral Rev |
| RIN | <i>RNA integrity number</i> |
| RMCE | <i>recombinase mediated cassette exchange</i> |
| RNA | ARN de interferencia |
| RNA-seq | <i>RNA-sequencing</i> – secuenciación del ARN |
| RPKM | <i>reads per kilobase million</i> |
| Rps18 | <i>ribosomal protein S18</i> |
| RRE | <i>Rev response element</i> |
| RT-qPCR | <i>real-time quantitative reverse transcription PCR</i> |
| Rv | reverse – en sentido reverso |
| <i>S. cerevisiae</i> | <i>Saccharomyces cerevisiae</i> |
| <i>S. pyogenes</i> | <i>Streptococcus pyogenes</i> |
| s/ | sin |
| S/MARs | <i>scaffold/matrix attachment regions</i> |
| S100a6 | <i>s100 calcium binding protein a6</i> |
| SAGE | <i>serial analysis of gene expression</i> |
| SBE | Sociedad de Ingenieros Biológicos |
| SBS | <i>sequencing by synthesis</i> |
| SDS | dodecilsulfato de sodio |
| SE | <i>single-end</i> |
| SEC21B | <i>coatomer subunit gamma</i> |
| SFB | suero fetal bovino |
| sgRNA | <i>single guide RNA</i> – ARN guía |
| SIN | <i>self-inactivated</i> |
| SMRT | <i>single-molecule real-time</i> |
| SOLiD | <i>sequencing by oligo ligation detection</i> |
| SSC | <i>side scatter</i> |
| SV40 | <i>simian vacuolating virus 40</i> – virus vacuolado del simio 40 |
| TA | temperatura ambiente |
| TALENs | <i>transcription activator like effectors nucleases</i> |
| TE | solución de Tris-EDTA |
| TFF | <i>tangencial flow filtration</i> – filtración por flujo tangencial |
| TG | transgén |
| TNF- α | receptor del factor de necrosis tumoral-alfa |
| TOOS | <i>N-etil-N-(2 hidroxil-3-sulfopropil) m-toluidina</i> |
| tPA | activador tisular del plasminógeno humano |
| TPI1 | <i>triosephosphate isomerase 1</i> |
| TRAIL | <i>TNF superfamily member 10</i> |
| Tris-HCl | <i>Tris(hidroximetil)aminometano hydrochloride</i> |
| TSS | <i>transcription start site</i> – sitio de inicio de la transcripción |
| UCOEs | <i>ubiquitous chromatin opening element</i> |
| UGGT2 | <i>UDP-glucose glycoprotein glucosyltransferase 2</i> |

| | |
|---------------|--|
| USB | <i>Universal Serial Bus</i> |
| USER | <i>uracil-specific excision reagent</i> |
| UTR | <i>untranslated region</i> |
| UV | ultravioleta |
| V+I | vector + inserto |
| VIH | virus de la inmunodeficiencia humana |
| Vim | <i>vimetin</i> |
| Vol. | volumen |
| VR | vector religado |
| vs | <i>versus</i> |
| VSV-G | glicoproteína del virus de la estomatitis vesicular |
| WCB | <i>working cell bank</i> – banco celular de trabajo |
| WPRE | <i>woodchuck hepatitis virus postranscriptional regulatory element</i> |
| wt | <i>wild type</i> |
| <i>x-mean</i> | intensidad media de fluorescencia |
| Ybx1 | <i>y-box binding protein 1</i> |
| ZFNs | <i>zinc finger nucleases</i> |
| ZsGreen | proteína verde fluorescente derivada de <i>Zoanthus sp.</i> |

UNIDADES

| | |
|---------------------------------|--|
| cm | centímetro |
| mm | milímetro |
| μm | micrómetro |
| nm | nanómetro |
| cm^2 | centímetro cuadrado |
| cél./ml o cél. ml^{-1} | células por mililitro |
| g | gramo |
| mg | miligramo |
| μg | microgramo |
| ng | nanogramo |
| pg | picogramo |
| l | litro |
| dl | decilitro |
| ml | mililitro |
| μl | microlitro |
| fl | femtolitro |
| d | día |
| h | hora |
| min | minuto |
| s | segundo |
| M | molar |
| mM | milimolar |
| μM | micromolar |
| nmol | nanomol |
| pmol | picomol |
| U | unidades |
| UI | unidades internacionales |
| mU | miliunidades |
| kDa | kilodalton |
| rpm | revoluciones por minuto |
| V | voltio |
| % | por ciento |
| % (v/v) | mililitros de soluto en 100 ml de solución |
| % (p/v) | gramos de soluto en 100 ml de solución |
| $^{\circ}\text{C}$ | grado Celsius |

RESUMEN

La producción de proteínas recombinantes terapéuticas es una de las áreas de mayor crecimiento dentro de la industria farmacéutica. Actualmente, los cultivos de células de mamífero son el sistema de producción dominante para dichas proteínas, principalmente porque pueden llevar a cabo las modificaciones postraduccionales complejas (por ejemplo la glicosilación) que son necesarias para una correcta secreción, eficacia y estabilidad del producto terapéutico. Las células de ovario de hámster chino (*chinese hamster ovary*, CHO) constituyen el sistema huésped de elección para la producción industrial de las proteínas terapéuticas. Aproximadamente el 70% de los biofármacos son producidos en esta línea celular. Sin embargo, a pesar de su creciente empleo biotecnológico e industrial y su importancia –tanto científica como económica– todavía no se encuentra disponible una completa información del genoma y transcriptoma de las células CHO.

Para la producción de proteínas recombinantes a escala industrial se emplean frecuentemente los cultivos bifásicos, que consisten en una fase inicial de crecimiento exponencial a 37°C seguida de una fase de arresto celular, que es inducida por la reducción de la temperatura del cultivo (28–33°C). Esta inhibición del crecimiento está asociada con el arresto de las células en la fase G1 del ciclo celular, y está relacionada con diversas características positivas, como –p. ej.– una mejora tanto en la productividad (al prolongarse la fase estacionaria de producción) como en la viabilidad celular. Por ello, se considera a dicha fase del ciclo celular como el período ideal para incrementar la producción de proteínas recombinantes. Sin embargo, es escasa la información disponible a nivel molecular acerca los mecanismos que regulan estos fenotipos a temperaturas reducidas en cultivos de células de mamífero.

La expresión óptima de las proteínas heterólogas en células de mamífero requiere de la optimización de diversos parámetros, entre los cuales es crucial un correcto diseño del vector de expresión. Actualmente, debido a la escasa información genómica referida a las células CHO, la alternativa más usada es el empleo de promotores virales (como CMV o SV40), ya que en la mayoría de las condiciones estos proporcionan niveles altos de expresión. Pero en muchos casos estos imponen a las células un estrés que interrumpe el crecimiento celular y afecta diversos procesos celulares. Además, idealmente la función promotora debe coordinarse con el proceso productivo. Por estas razones, el uso de promotores endógenos constituye una alternativa interesante.

En este contexto, se planteó la hipótesis del presente trabajo de tesis, cuyo objetivo general es la aplicación de las nuevas tecnologías de secuenciación masiva (NGS) a la línea celular CHO-K1, con el fin de desarrollar estrategias de ingeniería

genética mediante el análisis bioinformático de los datos generados. Específicamente, se propuso llevar a cabo la secuenciación del transcriptoma de células CHO-K1 cultivadas en alta densidad, con el propósito de comparar perfiles de expresión génica en diferentes etapas del cultivo en biorreactor, y seleccionar ciertos genes con elevada expresión, con el objetivo de usarlos para la construcción de vectores de expresión de proteínas recombinantes. Los mismos son una alternativa al empleo de los promotores habitualmente usados, principalmente virales, para lograr incrementar la estabilidad genética y mejorar la eficiencia del proceso de obtención de las proteínas recombinantes.

En primer lugar se caracterizó el ciclo celular de células CHO-K1 cultivadas en un biorreactor, analizando diariamente la distribución de las mismas en las diferentes fases del ciclo celular. Se llevaron a cabo dos cultivos independientes de células CHO-K1 en suspensión. En ambas experiencias fue posible evidenciar el enriquecimiento del cultivo en células sincronizadas en fase G0/G1 en las etapas tanto del crecimiento exponencial tardío como hacia los últimos días de la fase estacionaria. En particular, durante el segundo cultivo, más del 80% de las células permanecieron en fase G0/G1 durante toda la fase estacionaria. Esto puede deberse al marcado gradiente de temperatura aplicado, ya que no se usó ningún método químico o físico para lograr el arresto celular.

Estos análisis arrojan una tendencia evidente: cuanto más temprano se aplica el gradiente de temperatura, más rápidamente las células se arrestan en la fase G0/G1. Ello genera un mayor control del bioproceso, asociado con los siguientes beneficios: menor tiempo de cultivo, disminución de la proliferación celular y mayor viabilidad celular. En conjunto, estos factores pueden generar mayores tasas específicas de producción de proteínas recombinantes durante el proceso productivo.

Considerando que en la siguiente etapa de trabajo se propuso la secuenciación del transcriptoma de células CHO-K1 cultivadas en alta densidad y sincronizadas en fase G0/G1 y –como se determinó que los cultivos de células CHO-K1 en alta densidad tienden a arrestarse en la fase G0/G1–, se decidió realizar las extracciones de ARN total directamente a partir del cultivo celular en cada biorreactor. Se secuenció el transcriptoma a partir de tres réplicas biológicas de células CHO-K1 cultivadas en medio libre de suero fetal bovino y empleando la perfusión como modo de operación del biorreactor (que es un cultivo continuo con retención celular). Se compararon los perfiles de expresión génica entre las diferentes etapas del cultivo y se observó que los mayores cambios regulatorios ocurrieron durante la fase estacionaria, predominando marcadamente la *up*-regulación génica. Además, se determinó que la exposición de las células CHO-K1 a temperaturas reducidas de cultivo desencadena una respuesta coordinada que comprende la modulación del ciclo celular, del metabolismo energético,

de la proliferación celular y de procesos como la replicación y reparación del ADN, transcripción, metabolismo proteico y apoptosis.

Además, para cada etapa del cultivo se identificaron genes con funciones clave en el arresto de las células en la fase G0/G1, correlacionando así los análisis transcriptómicos con los de ciclo celular. Cabe destacar que este trabajo es el primero en cual se reporta una combinación entre los resultados del estudio del ciclo celular y el análisis del transcriptoma de la línea celular CHO-K1 huésped durante un bioproceso en condiciones típicas a escala industrial.

Por otro lado, se seleccionaron los primeros 200 transcriptos con mayor expresión, tanto en la etapa exponencial como estacionaria de cultivo. De ese grupo se eligieron los siguientes genes: S100a6, Fth1, Lgals1, Vim, Actb, Rps18, Cd63, Calr, Ybx1, Hspa5, Hmox1, Ctsz y Lamp1, considerándolos como “candidatos” para el empleo de sus regiones promotoras. Así se planteó utilizar las secuencias promotoras de dichos genes para la construcción de nuevos vectores de expresión destinados a la producción de proteínas recombinantes, logrando que dichos promotores respondan a la dinámica de los bioprocesos de nuestro laboratorio. Esto constituye una alternativa al empleo de promotores virales, aprovechando la maquinaria transcripcional regulatoria endógena disponible en las condiciones de un cultivo en alta densidad celular con un medio de cultivo definido.

Se pudo llevar a cabo el clonado de los promotores Rps18, Hspa5 y Hmox1 en el vector pZsGreen1-1 (para Hmox1 se evaluaron dos clones). El *screening* de la actividad transcripcional se realizó únicamente de manera transiente, en células HEK y CHO-K1 (para esta última en adherencia y en suspensión), analizando por citometría de flujo la expresión de la proteína verde fluorescente ZsGreen1 bajo el control del promotor CMV y de los promotores endógenos aislados. El resultado más promisorio se obtuvo con el promotor endógeno Hspa5, el cual mostró un comportamiento equivalente al promotor CMV en células HEK y CHO-K1 (para esta última línea, tanto en adherencia como en suspensión). En estos estudios se obtuvieron resultados muy prometedores, ya que se pudo constituir un esquema racional de trabajo para la identificación de nuevos promotores a partir de los análisis transcriptómicos, el aislamiento de sus secuencias y generación de nuevos vectores de expresión.

Dado que para los cuatro genes “candidatos” que exhibieron mayor expresión (S100a6, Fth1, Lgals1 y Vim) no se pudo evaluar la actividad transcripcional en forma transiente, se decidió como alternativa proyectar e iniciar experimentos de *knock-in* utilizando el sistema CRISPR/Cas9.

En resumen, los resultados alcanzados permiten establecer una novedosa plataforma de trabajo aplicada a cultivos celulares, que resulta muy promisoriosa. En

efecto, se cuenta ahora con promotores endógenos de células CHO-K1 para la producción de cualquier proteína recombinante de interés bioterapéutico –como alternativa al uso de promotores virales comerciales– y, además, que los mismos responden a las condiciones de cultivo habituales implementadas en la escala productiva (alta densidad en cultivos en perfusión), sumado a la posibilidad de aplicar la novedosa tecnología de edición genética CRISPR a líneas celulares de mamífero. Por otro lado, esta optimización de los tiempos y costos de producción de proteínas recombinantes se puede aplicar tanto en investigación básica como aplicada en la producción a nivel industrial, ya sea en medicina humana como veterinaria.

ABSTRACT

The production of therapeutic recombinant proteins is one of the fastest growing areas within the pharmaceutical industry. At present, mammalian cells are the leading production system for these proteins since they can produce complex post-translational modifications (e.g. glycosylation) that are necessary for proper secretion, drug efficacy and stability. Chinese hamster ovary (CHO) cells are the main platform commonly used for the industrial production of therapeutic proteins. Nearly 70% of biologics are produced in these cells. However, despite its wide use in biotechnology and the industry, as well as its scientific and economic importance, there is no in-depth information about the genome and transcriptome of CHO cells available yet.

Biphasic cultures are frequently used for the production of recombinant proteins at industrial scale. In these cultures, the initial phase of exponential cell growth at 37°C is followed by a growth arrest phase, which is induced by the reduction of the culture temperature (28–33°C). This inhibition is associated with G1-phase cell cycle arrest and it is connected with many positive properties, such as improved productivity (by prolonging the stationary/production phase) and desirable cell viability. For this reason, the G1-phase is considered the appropriate stage for increasing the production of recombinant proteins. Nevertheless, the mechanisms that regulate these phenotypes during mild hypothermia are still not completely understood.

Optimal heterologous protein expression from mammalian hosts requires tuning up many parameters among which the expression vector design plays a pivotal role. Today, due to the limited genomic information of CHO cells, promoter sequences with viral heritage (CMV, SV40) are a typical choice, since they deliver very high expression levels under most conditions. However, there are significant drawbacks linked with the usage of viral promoters because they initiate a stress response that culminates in cell growth interruption and they affect several cellular processes. Besides, the promoter function should be connected with the bioprocess dynamics. A solution to this problem would be the use of endogenous promoters which, in general, should not introduce any side effects due to their optimization for the host organism.

In this context, we proposed the application of next-generation sequencing technologies to CHO-K1 host cell line, in order to develop genetic engineering strategies from the bioinformatic data analysis. Specifically, we proposed sequencing the transcriptome of CHO-K1 cells cultured in high density, with the purpose of comparing gene expression profiles between growth stages during a bioreactor culture, and selecting genes with high levels of expression, so as to use them for the construction of expression vectors for recombinant protein production. They are an

alternative to viral promoters and they are meant to achieve better genetic stability and improve production process efficiency.

First, we analyzed cell cycle behavior during a typical industrial bioprocess of suspension-adapted CHO-K1 cells and we carried out two independent bioreactor cultures. We found a cell culture state characterized by G0/G1 phase synchronization, mainly during the late exponential growth phase and towards the last days of the stationary phase. In particular, more than 80% of the cells were in G0/G1 phase during the whole stationary phase in the second fermentation. This is probably due to the early marked temperature gradient applied during the bioprocess, considering no chemical or physical agent was used for cell arrest.

An evident trend in all these results is that the earlier the temperature gradient is applied, the sooner cells arrest in G0/G1 phase, resulting in a major bioprocess performance control by shortening culture time, diminishing cell proliferation and improving culture longevity. Altogether, these factors may render higher specific rates of recombinant protein production.

Since the next stage of this thesis project aims at sequencing the transcriptome of CHO-K1 cells cultured in high density and synchronized in G0/G1 phase and —as we saw that CHO-K1 cultures tend to be arrested in G0/G1 phase— we decided to purify total RNA directly from cells samples from each bioreactor. Transcriptome sequencing was carried out from three biological replicates of CHO-K1 cells, cultured in perfusion mode and with serum-free media. The number of differentially expressed (DE) genes between exponential phase vs. inoculation compared to the number of DE genes between stationary phase vs. inoculation day reveals that the most important regulatory changes happened throughout the stationary phase when gene up-regulation markedly prevailed. Our results indicate that the exposure of CHO-K1 cells to sub-optimal temperatures triggers a coordinated response, comprising the modulation of the cell cycle, energy metabolism and cell proliferation, and biological processes, such as DNA replication and repair, transcription, protein metabolism and apoptosis.

Also, we successfully identified key genes and pathways connected with the particular culture states, for each culture stage, especially with the arrest of the cells in G0/G1 phase. Thus, there was a correlation between transcriptomic analysis and cell cycle study. This is the first report combining cell cycle and transcriptome analysis of CHO-K1 host cells during typical culture bioprocesses at industrial scale.

In addition, we selected the first 200 transcripts with the highest levels of expression, both in the exponential and stationary culture stages. From this set of genes, the following ones were chosen: S100a6, Fth1, Lgals1, Vim, Actb, Rps18, Cd63,

Calr, Ybx1, Hspa5, Hmox1, Ctsz y Lamp1. Besides, they were designated as “candidates” for using their promoter regions. By doing so, we aimed at using the promoter sequences of those genes for the construction of new expression vectors for the production of recombinant proteins. Consequently, the endogenous promoters will respond to the dynamics of the bioprocesses carried out in our laboratory. This constitutes a promising alternative to the use of viral promoters and leverages the endogenous transcriptional regulatory machinery available during suspension cell cultures using a defined culture medium.

The promoters of Rps18, Hspa5 and Hmox1 were inserted into pZsGreen1-1 expression vector. For the last promoter, two clones were analyzed. The screening of transcriptional activity was only carried out in a transient manner, namely, transfecting HEK and CHO-K1 cells (for the last one, adherent and suspension cultures were analyzed). The results obtained were highly promising since it was possible to establish a rational workflow for the identification of new promoters from transcriptomic analysis, the isolation of their sequences and the generation of new expression vectors.

On the other hand, we could not evaluate the transcriptional activity of the rest of the genes that were “candidates”, especially, the four genes that showed the highest expression (S100a6, Fth1, Lgals1 and Vim). Therefore, we decided to design knock-in experiments using the CRISPR/Cas9 technology as an alternative way of using the endogenous promoters identified.

In summary, our results enable the development of a novel and promising biotechnological platform applied to cell cultures. We now have endogenous promoters for the production of any recombinant protein of biotherapeutic interest—as an alternative to the use of commercial viral promoters—and they respond to the typical culture conditions at large scale (high cell density in perfusion cultures). Besides, we have now the possibility of applying the novel editing technology CRISPR to mammalian cell lines. Moreover, this time and cost optimization for the production of recombinant proteins can be applied both at laboratory and industrial scale in the fields of human and veterinary medicine.

INTRODUCCIÓN GENERAL



INTRODUCCIÓN GENERAL

“Creo que las mayores innovaciones del siglo XXI estarán en la intersección de la biología y la tecnología. Una nueva era está comenzando”.

Steve Jobs

El mundo ha entrado en lo que ya es considerada la “revolución más grande de la historia de la humanidad”: la Cuarta Revolución Industrial, según lo planteado en el *World Economic Forum* en 2016 [1]. Esta revolución se caracteriza por una fusión de tecnologías que borra progresivamente los límites entre los dominios de lo físico, lo biológico y lo digital. Durante las últimas cinco décadas, los avances tecnológicos en el campo digital han cambiado completamente la forma en que nos comunicamos, cómo comerciamos y cómo vemos al mundo. La Cuarta Revolución Industrial promete basarse en dichos avances para transformar aún más los modelos económicos y sociales, las industrias existentes y el campo de la salud.

Como ejemplo, se puede mencionar la forma en que una combinación de saberes digitales y nuevas técnicas de edición de genes está comenzando a transformar la salud humana y animal, la industria de alimentos y la producción química. El método CRISPR/Cas9 (*clustered regularly interspaced short palindromic repeats*) es el avance más reciente en las técnicas de edición de genes. El impacto de CRISPR en los sistemas se ve multiplicado por el hecho de que el costo de secuenciar un genoma humano completo ha pasado de US\$ 1000 millones (en 2003) a menos de US\$ 1000 (en 2016), y se espera que en poco tiempo sea de sólo US\$ 100. Debido a esta reducción sustancial en los costos, muchos laboratorios alrededor del mundo están adoptando nuevos enfoques y sistemas de producción.

1. Proteínas recombinantes terapéuticas

Las proteínas son sintetizadas por todos los seres vivos como parte de su metabolismo natural. Algunas proteínas, como las enzimas, actúan como biocatalizadores acelerando la tasa de reacción al disminuir la energía de activación, mientras que otras tienen roles importantes en la respuesta inmune, señalización, adhesión y ciclo celular. A su vez, son producidas en industrias biotecnológicas con la ayuda de la ingeniería genética y la ingeniería de proteínas. Tanto las proteínas obtenidas a partir de fuentes naturales como aquellas obtenidas de modo recombinante han generado un incremento productivo en áreas como medicina, diagnóstico, alimentos, nutrición, industria química, entre otras [2].

La producción de proteínas recombinantes terapéuticas ha revolucionado la medicina moderna en las últimas décadas. La Figura 1 muestra diversos hitos en el

campo de la biotecnología que tienen o probablemente tendrán un gran impacto en la producción de proteínas terapéuticas.

La industria de la fermentación microbiana surgió a principios de 1900, cuando comenzaron las primeras fermentaciones anaeróbicas a mayor escala para la producción de químicos como la acetona y el butanol, seguido de la producción de ácido cítrico. La penicilina se descubrió en 1927 pero su desarrollo no ocurrió sino hasta comienzos de 1940. En 1922, Banting y Bens produjeron la primera proteína de uso farmacéutico: la insulina. La era de la biotecnología moderna comenzó en 1971 con el establecimiento de Cetus Corporation en California, alrededor de uno a dos años previamente al descubrimiento de la tecnología de ADN recombinante. El empleo de esta técnica ha revolucionado la biología, la medicina y la biotecnología. Su impacto no sólo ha afectado al desarrollo de prácticamente todo lo “bio”, sino que también se ha convertido en una herramienta fundamental en otros ámbitos como la industria y la justicia.

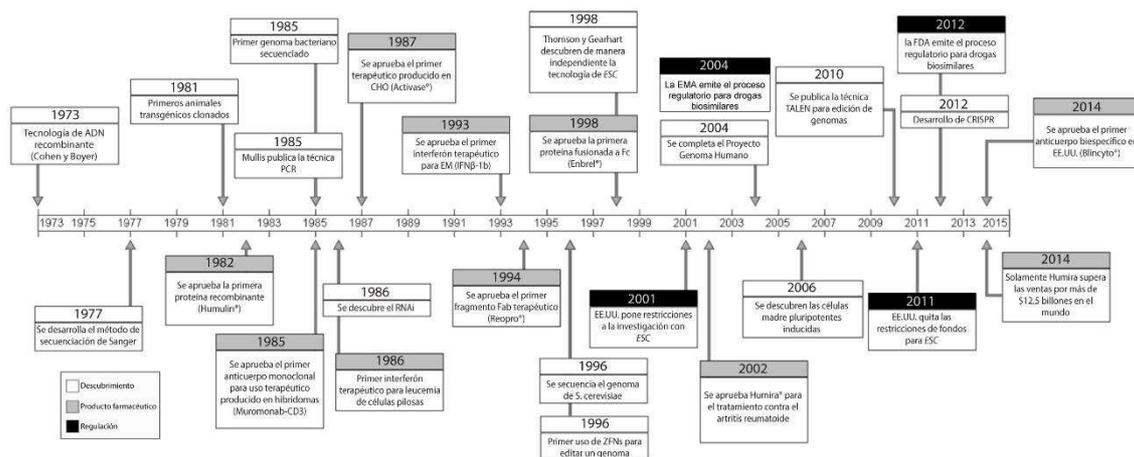


Figura 1. Cronología de grandes descubrimientos y aprobaciones de medicamentos importantes.

Figura modificada de Wells y col., 2016 [3].

El primer paso en la producción de una proteína en forma recombinante es la obtención de la secuencia codificante, luego dicha proteína se expresa en el huésped heterólogo seleccionado. Existen diversos sistemas de expresión, tales como cultivos en bacterias, levaduras, hongos filamentosos, células de mamíferos, plantas, insectos o a través del empleo de plantas o animales transgénicos. La calidad, funcionalidad, velocidad de producción y rendimiento son los factores más importantes a controlar. Habitualmente, las proteínas no glicosiladas son producidas en *Escherichia coli* (*E. coli*) o levaduras, constituyendo un 40% de las proteínas que existen en el mercado. Las glicoproteínas, por su parte, son producidas en cultivos de células de mamíferos.

Levaduras, células de insectos y hongos usualmente no son capaces de dar lugar a proteínas con el perfil de glicosilación deseado [2].

Dentro de las proteínas recombinantes de uso farmacéutico se pueden mencionar a la insulina, la albúmina, la hormona de crecimiento humano, Factor VIII, entre otras. Entre 2011 y 2016, la Administración de Alimentos y Medicamentos (FDA) aprobó 48 nuevas proteínas terapéuticas, de un total de 197 autorizaciones de medicamentos (Tabla 1). De esos 48 productos bioterapéuticos, 29 corresponden a anticuerpos monoclonales producidos en células CHO (*chinese hamster ovary*) y en células de mieloma murino. Este tipo de biofármacos superó los cuarenta mil millones de dólares en ventas anuales en 2014.

Tabla 1. Nuevos productos bioterapéuticos y sus organismos de producción aprobados por la FDA en el período 2011-2016.

Tabla modificada de Wells y col., 2016 [3].

| Año | Marca comercial | Principio activo | Organismo de producción |
|------|-----------------|----------------------------|--|
| 2016 | Zinplava | Bezlotoxumab | - |
| 2016 | Lartruvo | Olaratumab | Células NS0, mieloma murino |
| 2016 | Adlyxin | Lixisenatide | - |
| 2016 | Zinbryta | Daclizumab | - |
| 2016 | Tecentriq | Atezolizumab | - |
| 2016 | Cinqair | Reslizumab | Células NS0, mieloma murino |
| 2016 | Taltz | Ixekizumab | - |
| 2016 | Anthim | Obiltoxaximab | - |
| 2015 | Kanuma | Sebelipase alfa | Huevos blancos de gallinas modificadas |
| 2015 | Empliciti | Elotuzumab | Células NS0, mieloma murino |
| 2015 | Portrazza | Necitumumab | Células NS0, mieloma murino |
| 2015 | Darzalex | Daratumumab | CHO |
| 2015 | Nucala | Mepolizumab | CHO |
| 2015 | Strensiq | Asfotase alfa | CHO |
| 2015 | Praxbind | Idarucizumab | CHO |
| 2015 | Tresiba | Insulin degludec injection | <i>S. cerevisiae</i> |
| 2015 | Repatha | Evolocumab | CHO |
| 2015 | Praluent | Alirocumab | CHO |
| 2015 | Unituxin | Dinutuximab | Células SP2/0, mieloma murino |
| 2015 | Natpara | Parathyroid hormone | <i>E. coli</i> |
| 2015 | Cosentyx | Secukinumab | CHO |
| 2014 | Opdivo | Nivolumab | - |
| 2014 | Blinicyto | Blinatumomab | CHO |
| 2014 | Trulicity | Dulaglutide | Células de mamífero |
| 2014 | Keytruda | Pembrolizumab | - |
| 2014 | Plegridy | Peginterferon beta-1-a | CHO |
| 2014 | Entyvio | Vedolizumab | CHO |
| 2014 | Sylvant | Siltuximab | CHO |
| 2014 | Cyramza | Ramucirumab | Células NS0, mieloma murino |
| 2014 | Tanzeum | Albiglutide | <i>S. cerevisiae</i> |

| | | | |
|------|-------------|------------------------------|-----------------------------|
| 2014 | Vimizim | Elosulfase alfa | CHO |
| 2014 | Myalept | Metreleptin | <i>E. coli</i> |
| 2013 | Gazyva | Obinutuzumab | CHO |
| 2013 | Kadcyla | Ado-trastuzumab emtansine | CHO |
| 2012 | Gattex | Teduglutide | <i>E. coli</i> |
| 2012 | Raxibacumab | Raxibacumab | Células murinas |
| 2012 | Jetrea | Ocriplasmin | <i>P. pastoris</i> |
| 2012 | Neutroval | Tbo-filgrastim | <i>E. coli</i> |
| 2012 | Zaltrap | Ziv-aflibercept | CHO |
| 2012 | Perjeta | Pertuzumab | CHO |
| 2012 | ElELYso | Taliglucerasa alfa | Planta |
| 2012 | Voraxaze | Glucarpidase | <i>E. coli</i> |
| 2011 | Eylea | Aflibercept | CHO |
| 2011 | Erwinaze | Asparaginase | <i>Erwinia chrysanthemi</i> |
| 2011 | Adcetris | Brentuximab vedotin | CHO |
| 2011 | Nulojix | Belatacept | Células de mamífero |
| 2011 | Yervoy | Ipilimumab | CHO |
| 2011 | Benlysta | Belimumab | Células de mamífero |

En 2017, 10 de los 15 medicamentos más vendidos alrededor del mundo fueron biofármacos, recaudando aproximadamente US\$ 79,3 mil millones (Tabla 2).

Tabla 2. Lista de los diez productos bioterapéuticos más vendidos en 2017 [4].

| | Biomolécula | Fabricante | Sistema de expresión | Ganancia |
|-----------|--|-------------------|-----------------------------|-------------------------|
| 1 | Humira® (adalimumab) | AbbVie | Células de mamífero | US\$ 18,43 mil millones |
| 2 | Rituxan® (rituximab) | Roche | Células de mamífero | US\$ 9,24 mil millones |
| 3 | Enbrel® (etanercept) | Amgen/Pfizer | Células de mamífero | US\$ 7,88 mil millones |
| 4 | Herceptin® (trastuzumab) | Roche/Genentech | Células de mamífero | US\$ 7,44 mil millones |
| 5 | Remicade® (infliximab) | Janssen/Merck | Células de mamífero | US\$ 7,15 mil millones |
| 6 | Avastin® (bevacizumab) | Roche/Genentech | Células de mamífero | US\$ 7,10 mil millones |
| 7 | Eylea (aflibercept) | Bayer/Regeneron | Células de mamífero | € 603 mil millones |
| 8 | Lantus® (insulin glargine) | Sanofi | <i>E. coli</i> | US\$ 5,73 mil millones |
| 9 | Prevnar13® (pneumococcal conjugate vaccine) | Pfizer | Bacteria | US\$ 5,60 mil millones |
| 10 | Neulasta / Peglasta | Amgen | <i>E. coli</i> | US\$ 4,72 mil millones |

2. Expresión de proteínas recombinantes terapéuticas en células eucariotas

La elección del sistema de expresión para la producción de proteínas recombinantes terapéuticas depende de varios factores, entre los cuales se destacan: las características de crecimiento del cultivo, los niveles de expresión buscados, la conformidad de que la proteína se exprese de manera intra o extracelular, las modificaciones postraduccionales esperadas, la actividad biológica del producto, así como también de cuestiones relacionadas a los aspectos económicos y a las pautas establecidas por las entidades regulatorias.

En 1987 se aprobó la comercialización de la primera proteína recombinante terapéutica producida en células de mamíferos. En la actualidad, los cultivos de células eucariotas dominan el campo de producción de proteínas recombinantes empleadas como agentes terapéuticos para el diagnóstico y la prevención de enfermedades [5]. Aproximadamente un 60-70% de las proteínas recombinantes bioterapéuticas son producidas en cultivos de células animales [6], principalmente debido a que la maquinaria de procesamiento proteico, incluyendo el proceso de glicosilación y plegamiento, se asemeja a la humana [7].

Dentro de las desventajas del empleo del cultivo de células animales se pueden mencionar los bajos niveles de productividad en comparación a otros sistemas, el elevado costo o la posible contaminación con virus [2,8]. Además, la utilización de sistemas procariontes se caracteriza por ser más rápida y económica. Sin embargo, al momento de producir productos bioterapéuticos para humanos, la balanza costo-beneficio se inclina a favor del sistema eucariota [9], ya que como se mencionó previamente, estas células permiten llevar a cabo el correcto procesamiento y plegamiento de la proteína de interés y además introducen las modificaciones postraduccionales apropiadas.

Numerosos sistemas de expresión eucariotas han sido estudiados y evaluados con el objetivo de lograr niveles elevados de expresión de la proteína recombinante en estudio. En las últimas dos décadas se han llevado a cabo grandes mejoras no sólo relacionadas al huésped elegido para la producción, sino también en lo que respecta a la construcción y diseño de vectores, optimización de codones, amplificación de genes, métodos de transfección, entre otros. Sin embargo, es imperativo llevar a cabo la optimización del bioproceso a los fines de obtener el producto de interés en grandes cantidades y con las características buscadas [10,11].

3. Células CHO

Una gran variedad de proteínas recombinantes terapéuticas, aprobadas para el tratamiento de diversas patologías, son producidas a gran escala empleando células CHO (derivadas de ovario de hámster chino), células BHK (derivadas de riñón de hámster bebé) y células NS0 (derivadas de mieloma murino). Todas ellas han demostrado ser seguras para la producción de una variedad de proteínas recombinantes de uso terapéutico [12,13].

El primer producto producido a partir de células animales, específicamente células CHO, fue el activador tisular del plasminógeno humano (tPA), aprobado para su uso clínico en 1987. Desde entonces, las células CHO han sido empleadas para la producción de numerosas proteínas recombinantes de uso terapéutico, tales como: eritropoyetina (EPO), etanercept (proteína de fusión entre la región extracelular del receptor del factor de necrosis tumoral-alfa (TNF- α) y la porción cristalizable de la IgG1 humana Fc γ 1), factor VIII (FVIII), adalimumab (anticuerpo monoclonal anti-TNF- α), etc. En la actualidad, más del 70% de las proteínas terapéuticas aprobadas son producidas en células CHO, mientras que sólo un 8% se obtienen a partir de células humanas [9,11,14,15].

La línea celular CHO fue aislada por Theodore Puck en 1957 [16] a partir de la inmortalización de cultivos primarios de células ováricas de hámster chino (*Cricetulus griseus*). Una de las líneas celulares ancestrales más utilizada es la línea CHO-K1, constituyendo una plataforma de producción robusta y confiable. Se caracteriza por su fácil adaptación al crecimiento en suspensión y en condiciones libre de suero fetal bovino (SFB), lo que permite llevar a cabo su cultivo en altas densidades celulares. De este modo, el incremento de la concentración de la proteína recombinante de interés en el sobrenadante de cultivo y la ausencia de proteínas contaminantes características del SFB (así como también virus y priones) permiten simplificar los procesos posteriores de recuperación, purificación y refinamiento del producto final [17]. Además, desde el punto de vista regulatorio, las células CHO-K1 constituyen un huésped de producción adecuado, ya que son incapaces de replicar patógenos humanos como el VIH (virus de la inmunodeficiencia humana), *H. influenzae*, poliovirus y herpesvirus, entre otros [18].

La variabilidad genómica de las células CHO ha permitido aislar líneas celulares mutantes deficientes en enzimas metabólicas. La supervivencia de estas células depende de la presencia de ciertos nutrientes, y es por ello que dicha dependencia puede ser usada como marcador de selección para la generación de líneas celulares productoras de proteínas recombinantes. Entre las diversas variantes generadas, las más comúnmente usadas son las líneas celulares deficientes en la enzima dihidrofolato

reductasa (DHFR) [17,19], como las líneas CHODUKX y CHO-DG44 que requieren glicina, hipoxantina y timidina para su crecimiento [20].

A diferencia de otros organismos relacionados (como rata y ratón), la generación de información genómica para las células CHO ha sido bastante limitada. Técnicas como la genómica, la transcriptómica y la proteómica contribuyen a mejorar la comprensión y a optimizar los procesos productivos de proteínas recombinantes, lo que conduce a una mejora significativa en la velocidad con la que estas terapias pasan de moléculas de laboratorio a su aplicación clínica como medicamentos.

En 2002 se inició un proyecto de colaboración entre investigadores de la Universidad de Minnesota y del Instituto de Ingeniería de Bioprocesos de Singapur para aplicar las tecnologías “ómicas” a las células CHO. En 2004 se creó un microarreglo de ADN copia (ADNc) con más de 4000 secuencias diferentes de células CHO [18]. Estos resultados promovieron la fundación del consorcio para el estudio del genoma de células CHO (*Consortium on Chinese Hamster Ovary Cells Genomics*) en 2006, junto con la Sociedad de Ingenieros Biológicos (SBE). El principal objetivo de este consorcio, que involucró también a numerosas empresas de la industria biotecnológica y farmacéutica, fue promover y facilitar el empleo de células CHO para el descubrimiento de nuevas drogas mediante el incremento de la calidad y eficiencia de producción de medicinas [17, 18].

4. Proceso de producción de una proteína recombinante de uso terapéutico en cultivo de células de mamífero

El desarrollo de un proceso de producción de proteínas recombinantes en cultivo de células animales sigue un esquema bien establecido (Figura 2).

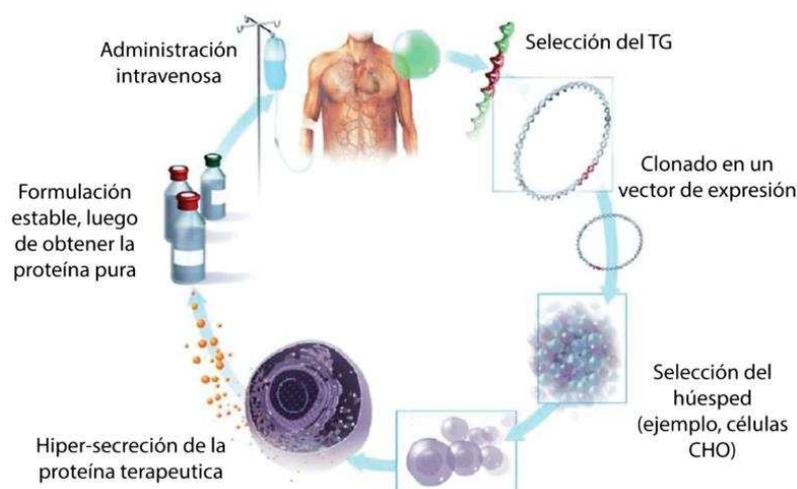


Figura 2. Representación esquemática del proceso de producción de una proteína recombinante para uso terapéutico.

Figura modificada de Dingermann, 2008 [21].

En la Figura 3 se encuentran representadas las principales etapas implicadas en el desarrollo de líneas celulares para la producción de proteínas recombinantes; en este caso empleando células CHO con un sistema de selección de DHFR/MTX. Para cada nuevo candidato terapéutico, este proceso puede durar más de seis meses, incluso antes de que ingrese en la fase de evaluación cuando se prueba la eficacia y seguridad de la molécula en animales o seres humanos.

Inicialmente, el transgén (TG) de interés (junto con los elementos regulatorios) es insertado dentro del genoma celular. Asimismo, se suele emplear un segundo gen como marcador de selección, de manera que en presencia de un agente de selección (que es aplicado días posteriores a realizar la inserción del TG), sólo sobrevivan aquellas células que incorporaron dicho marcador.

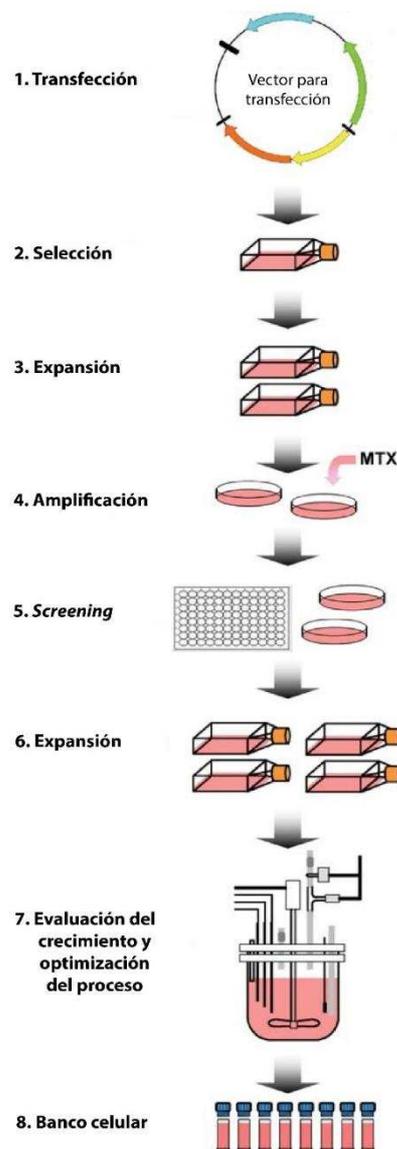


Figura 3. Representación esquemática de las etapas implicadas en el desarrollo de líneas celulares para la producción de proteínas recombinantes utilizando células CHO. Figura modificada de Jayapal y col., 2007 [18].

Luego del proceso de selección, las células que sobreviven (línea celular heterogénea) se transfieren como células individuales a un segundo cultivo. Posteriormente, estos se expanden para producir poblaciones clonales que luego son caracterizadas de manera de evaluar el nivel de expresión de la proteína recombinante. Sólo aquellos clones celulares que presentan productividades aceptables son resguardados para análisis posteriores, seleccionando únicamente un candidato clonal para la producción de la proteína recombinante [6]. A partir de este clon se obtiene el banco celular maestro (MCB, *master cell bank*) y de este el banco de trabajo (WCB, *working cell bank*), de manera de garantizar la consistencia lote a lote en la producción. También, se debe definir la identidad, pureza, número de células, viabilidad, estabilidad génica, números de copia, identidad e integridad del TG y ausencia de agentes adventicios (virus, bacterias, hongos) y micoplasma.

Por otro lado, es necesario establecer un sistema de cultivo en función de las necesidades de producción. En general, la proteína recombinante es naturalmente secretada al sobrenadante de cultivo o bien, se utilizan construcciones que garanticen la secreción de la proteína al medio de cultivo.

El protocolo de cultivo en el biorreactor seleccionado afecta al perfil de impurezas y, por lo tanto, a la eficacia del *downstream processing*. De igual manera, puede alterar la solubilidad y estabilidad de la proteína de interés, afectando con ello la calidad y rendimiento del producto purificado. Asimismo, una sobreexpresión de la proteína puede llegar a disminuir ciertos factores necesarios para un adecuado plegamiento proteico. Por ello, el cultivo en el biorreactor se desarrolla en concordancia con el protocolo de purificación. Este cultivo se puede realizar en modo *batch*, *fed-batch* o continuo. Los procesos en modo *batch* son simples y robustos, pero no permiten alcanzar densidades celulares elevadas, a diferencia de los cultivos en modo *fed-batch*, en los cuales el metabolismo celular puede ser controlado [22].

Por otro lado, los cultivos continuos (en perfusión) proveen un ambiente favorable para el cultivo de células, removiendo los productos formados y suplementando nutrientes de manera continua. En comparación con los sistemas *batch* o *fed-batch*, este sistema ofrece cultivos prolongados, alcanzando densidades celulares superiores y menor tiempo de residencia del producto en el biorreactor. Esto resulta más favorable para la estabilidad del producto y es mandatorio en el caso de glicoproteínas inestables. Otra ventaja del modo continuo es el uso de reactores más pequeños en comparación con los procesos *fed-batch*.

Existen distintos sistemas que permiten la separación celular continua asegurando condiciones de esterilidad adecuadas, basados en diferentes propiedades físicas, tales como la filtración, la centrifugación o el ultrasonido. Todos los sistemas de filtración

corresponden a los de filtración tangencial (*tangential flow filtration*, TFF), tales como el sistema de filtro giratorio (*spin-filter*), filtración por flujo tangencial en cartuchos y flujo tangencial alternado (*alternating tangential flow*, ATF) [23].

Otro aspecto clave a considerar es el empleo de un medio de cultivo adecuado para el crecimiento celular. En el pasado, el uso de suero fetal bovino era necesario para el cultivo de células animales. Sin embargo, actualmente un gran porcentaje de los cultivos a gran escala se realizan en medios libres de suero, empleando péptidos pequeños, factores de crecimiento, lípidos, carbohidratos y moléculas pequeñas. Se evita el uso de suero por su composición indefinida, variación lote a lote, el riesgo de transmitir agentes adventicios y su elevado costo [6]. También se emplean medios definidos libre de proteínas (sin la suplementación de proteínas).

Una vez establecido el protocolo de purificación se realiza una caracterización bioquímica adecuada de la proteína pura. Se deben desarrollar diferentes técnicas analíticas que permitan definir distintos atributos, y en paralelo, aplicar técnicas analíticas ortogonales que conduzcan a determinar una propiedad específica. Finalmente, una vez definida cada etapa del proceso de producción, se procede al registro del medicamento, el cual es un procedimiento que implica la evaluación y autorización del terapéutico, siendo condición necesaria para la comercialización o puesta en el mercado del producto.

5. Estrategias para la generación de líneas celulares productoras de proteínas recombinantes

Para que un proceso de producción sea rentable es de primordial importancia seleccionar clones con elevados niveles de productividad. A continuación se describen distintas estrategias implementadas para lograr dicho objetivo.

La capacidad de la línea celular para ofrecer elevadas productividades volumétricas de la proteína en estudio dependerá de muchos factores, incluyendo: elementos de control transcripcional y traduccional, procesamiento del ARN, número de copias, estabilidad del ARNm, sitio de integración del TG en el genoma, potencial toxicidad de la proteína recombinante en la célula huésped, así como también de las propiedades genéticas de la célula productora [10].

Para favorecer el proceso de generación de la línea celular productora de la proteína de interés es fundamental llevar a cabo un correcto diseño del vector de expresión. Dichos vectores comúnmente utilizados presentan un promotor viral fuerte (como el promotor constitutivo del citomegalovirus, CMV, por ejemplo) o un promotor (o potenciador, *enhancer*) celular (como por ejemplo el *elongation factor 1 α* , EF1 α). También deben contar con marcadores de selección que permitan generar líneas

celulares estables y amplificar la expresión, un origen de replicación procariota y marcadores de selección para bacterias [24].

En la mayoría de los casos, el TG de interés se encuentra como ADNc sin intrones. Sin embargo, se ha reportado que los eventos de *splicing* producen un aumento en la eficiencia del transporte hacia el citoplasma y del proceso de traducción. Es por ello que diversos vectores de expresión contienen secuencias de intrones, los cuales se encuentran normalmente localizados entre el promotor y la secuencia del TG de interés [25]. A su vez, es deseable que la proteína pueda ser recuperada a partir del sobrenadante del cultivo celular para facilitar los procesos posteriores de purificación y refinamiento del producto final. Por ello se debe incluir la señal de secreción adecuada para cada proteína [12].

Por otra parte, la accesibilidad y demás propiedades específicas de cada metodología de transferencia del material genético ejercen cierta influencia sobre el sitio del genoma en el cual se integrará el TG de interés. La integración del mismo puede ocurrir en sitios transcripcionalmente activos, regiones promotoras, secuencias codificantes o regiones intergénicas; conduciendo, en algunos casos, a la desregulación o a la interrupción accidental de genes endógenos [26].

La mayoría de los procesos de producción de biofarmacéuticos utilizan, principalmente, sistemas de expresión en los cuales el TG de interés se integra en el genoma celular de manera aleatoria. Actualmente, existen distintos sistemas para la transferencia del material genético: precipitación por fosfato de calcio, electroporación, lipofección y transferencia mediada por polímeros (PEI, polietilenimina). Dentro de este tipo de procedimientos también se puede mencionar la “infección” con una partícula viral, proceso conocido como transducción. Asimismo, el TG puede expresarse en modo transitorio o de manera estable [17,27].

5.1. Expresión transitoria del transgén de interés

La expresión transitoria del TG se define como la producción de la proteína recombinante en un período corto de tiempo (10 a 14 días), lo cual permite llevar a cabo rápidamente su caracterización inicial. A este procedimiento se lo conoce como “transfección” y ha sido desarrollado principalmente con células HEK293 y CHO. Permite evaluar la funcionalidad del vector construido, así como también optimizar diferentes combinaciones de promotores y otros elementos útiles para la expresión de la proteína recombinante. Regularmente, el TG de interés se clona en un vector no viral y las células se transfectan mediante un método químico, empleando por ejemplo fosfato de calcio, lípidos catiónicos o PEI. A diferencia de lo que ocurre en un método de

expresión estable del TG, las células no atraviesan un proceso de selección durante la fase de producción.

La implementación de esta tecnología presenta ciertas desventajas asociadas a la densidad de células a transfectar, la productividad específica de la proteína en estudio y la heterogeneidad del producto. Sin embargo, recientemente se ha propuesto el empleo del método de expresión transitoria en gran escala, utilizando reactores tipo tanque agitado. Pero hasta el momento no se ha reportado la aprobación de la producción de una proteína terapéutica mediante este método, debido a que no resulta sencillo demostrar la reproducibilidad y consistencia del producto proteico obtenido [28].

5.2. Expresión estable del transgén de interés

La generación de líneas celulares estables, por su parte, depende de la integración permanente del plásmido en el cromosoma celular. La principal desventaja asociada al empleo de líneas celulares estables radica en que se requiere un tiempo prolongado para su establecimiento.

Tradicionalmente, la generación de líneas celulares recombinantes estables se ha llevado a cabo mediante el empleo de la metodología de transfección que resulta en una baja frecuencia de integración del TG de interés en el genoma celular. Es por ello que la etapa de selección es sumamente importante a los fines de que sobrevivan sólo aquellas células que hayan incorporado el TG de interés al genoma celular. Aunque exitoso, este método consume mucho tiempo, entre 5 y 12 meses, desde la transfección del material genético hasta la obtención de una línea productora [28], debido a que el aislamiento y la caracterización de un gran número de clones son los pasos que más tiempo insumen.

El sitio de integración tiene un efecto importante en la tasa de transcripción del TG (efecto posición). Se ha descrito que la integración en regiones de heterocromatina resulta en una baja o nula tasa de expresión, mientras que la integración en regiones de eucromatina frecuentemente permite la expresión génica. Sin embargo, la inserción del TG en dichas zonas a veces no resulta suficiente para asegurar su expresión a través del tiempo ya que puede ser silenciado, llevando a la pérdida de productividad de la línea celular. Se han reportado diferentes estrategias para contrarrestar el efecto posición, empleando elementos regulatorios en *cis* como *insulators*, elementos de apertura de la cromatina ubicuos (UCOEs, *ubiquitous chromatin opening element*), regiones S/MARs y elementos conservados anti-represores [17,27,29,30].

Asimismo, recientemente se han desarrollado técnicas que permiten amplificar el TG de interés, o bien, insertarlo en sitios específicos del genoma celular. Afortunadamente, esta metodología ha permitido mejorar notablemente los procesos de

generación de líneas celulares y selección de clones productores, reduciendo también los tiempos requeridos [10,31]. A continuación, se detallan algunas metodologías que han posibilitado alcanzar dichas mejoras.

5.2.1. Amplificación del transgén de interés

Existe una amplia variedad de sistemas para la selección de las células transfectadas, dentro de las que se puede mencionar la resistencia a antibióticos como neomicina y puomicina. Sin embargo, la selección por la presencia del gen DHFR en las células CHO deficientes en DHFR (CHO dhfr-), enzima esencial para la síntesis de purina y pirimidina, es una de las metodologías más ampliamente utilizadas [6], ya que permite lograr la amplificación del número de copias del TG hasta 1000 veces.

Esta enzima es inhibida en presencia del análogo de ácido fólico metotrexato (MTX). En este sistema de expresión, las células CHO dhfr- deben ser transfectadas con un vector que contenga el TG de interés y el gen DHFR. El empleo de MTX en concentraciones crecientes durante sucesivas rondas permitirá seleccionar aquellas células que demuestren una elevada capacidad de producción de la enzima DHFR y, por lo tanto, de la proteína de interés. La principal desventaja de esta metodología radica en que requiere de aproximadamente cinco rondas de selección por lo que la generación de la línea celular puede llevar más de un año [10,32].

Otro sistema de amplificación consiste en emplear células CHO y cultivarlas en presencia de un inhibidor de la glutamina sintetasa (GS), como la metionina sulfoximina (MSX). Dicha enzima es fundamental para la síntesis de glutamina a partir de glutamato y amonio y, en consecuencia, para la supervivencia de la línea celular en momentos de privación de dicho nutriente en el medio de crecimiento. Mediante el proceso de transfección se puede introducir en las células un vector que codifique para la GS y la selección se realiza empleando MSX, de modo que las células que no hayan incorporado el vector de interés mueren durante el proceso de selección. En este caso no se requieren rondas de amplificación [6,33].

5.2.2. Transferencia de genes mediante el empleo de vectores lentivirales

Otra metodología para la transferencia del material genético es la utilización de vectores virales recombinantes. Los vectores lentivirales (LVs, lentiviral vectors), miembros de la familia Retroviridae (retrovirus) se utilizan ampliamente en investigación y como vectores en terapia génica, debido a su capacidad de transducir tanto células en división como quiescentes, integrándose así en regiones de la eucromatina [34,35]. Además, presentan un elevado tropismo hacia distintos tipos celulares [36]. Una gran

ventaja de este sistema radica en que las células transducidas mantienen el TG en su genoma luego de sucesivas etapas de división.

Los LVs han sido contruidos a partir de diversos tipos de lentivirus, siendo el virus de la inmunodeficiencia humana tipo 1 (VIH-1) el más estudiado y utilizado en la actualidad. Inicialmente, el empleo de LVs requería de la implementación de diversas medidas de seguridad debido a la posible generación de partículas lentivirales competentes replicativas (RCLs). A los fines de minimizar el riesgo de formación de dichas RCLs y mejorar su eficiencia se han desarrollado varias “generaciones” de vectores lentivirales [37]. Desde la primera generación de LVs utilizados en estudios de terapia génica hasta la actual tercera generación, se han ido removiendo genes accesorios y regulatorios dispensables para el ensamblado de la partícula, como así también se han ido separando las secuencias estructurales en distintos plásmidos, buscando disminuir al mínimo la probabilidad de recombinación y posible generación de un nuevo microorganismo [38].

El sistema de LVs de tercera generación (Figura 4) está conformado por dos partes principales: el vector de transferencia que posee el TG de interés, y el sistema de empaquetamiento.

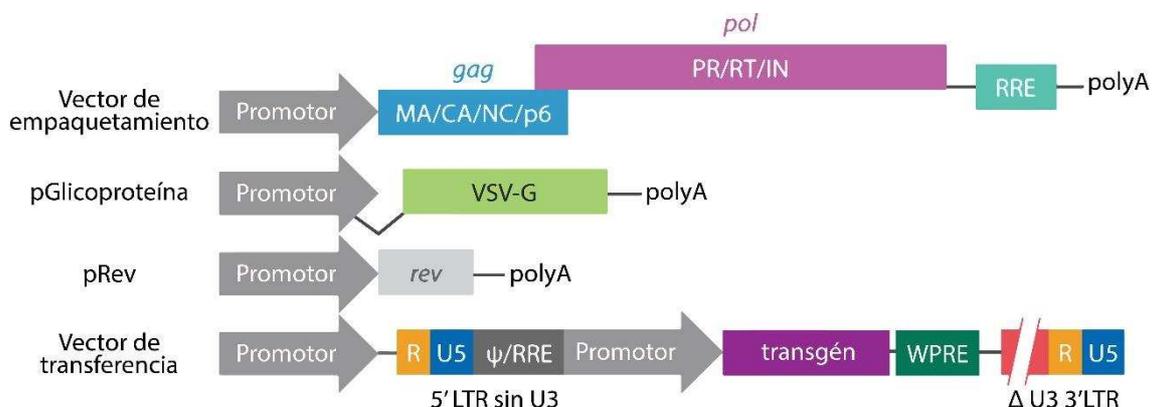


Figura 4. Tercera generación de vectores lentivirales.

En esta generación se desarrollaron los vectores SIN (*self-inactivated*), incrementando aún más la seguridad del sistema, el cual consta de cuatro plásmidos conteniendo los genes *gag* y *pol*, VSV-G, el elemento de respuesta a Rev y el transgén de interés.

En el vector de transferencia (pLV-GOI) se encuentran todas las secuencias necesarias para la expresión del TG de interés, como así también las secuencias *cis*-actuantes requeridas para el correcto ensamblado del vector lentiviral [35], [39]. A continuación se detallan las regiones más importantes:

- **LTR (*long terminal repeat*):** son regiones ubicadas en cada uno de los extremos del ADN proviral, encargadas de actuar como promotores y/o *enhancers* de la transcripción del ARN lentiviral. La región 5'LTR fue

modificada, intercambiando el promotor dependiente de *tat* original del VIH-1 por un promotor constitutivo del virus del sarcoma de Rous. Una región del 3'LTR fue delecionada para generar de esta forma un vector auto-inactivante (SIN, *self-inactivating*).

- **Psi (ψ):** secuencia de empaquetamiento, requerida para la incorporación del ARN genómico a las partículas virales.
- **RRE (*Rev response element*):** región que posee la secuencia necesaria para la unión de la proteína Rev, la cual promueve la exportación del ARN del núcleo al citoplasma.
- **cPPT (*central polypurine tract*):** región rica en purinas que promueve la retrotranscripción viral cuando las partículas lentivirales transducen una célula.
- **Promotor:** el promotor mayormente utilizado es el promotor constitutivo del citomegalovirus (CMV), pero puede ser reemplazado por otros promotores para células eucariotas.
- **WPRE (*woodchuck hepatitis virus posttranscriptional regulatory element*):** región estabilizadora del ARN mensajero.

Por otro lado, se encuentran los vectores plasmídicos que portan las secuencias codificantes de las proteínas que serán expresadas en *trans* y son necesarias para que el ARN lentiviral sea ensamblado. En el sistema lentiviral de tercera generación los vectores utilizados son los siguientes:

- **Vector de empaquetamiento pMDLg/pRRE:** posee la secuencia codificante de la poliproteína de cápside (*gag*) y la secuencia de la retrotranscriptasa viral (*pol*).
- **pRSV-Rev:** contiene la secuencia de la proteína Rev, la cual tiene capacidad de unirse a las secuencias RRE, presentes tanto en pMDL/pRRE como en pLV, y promover su exportación desde el núcleo al citoplasma de la célula transducida.
- **pGlicoproteína:** este vector posee la secuencia codificante para la glicoproteína de envoltura necesaria para que la partícula sea capaz de transducir las células blanco. La proteína más utilizada es la glicoproteína del virus de la estomatitis vesicular (VSV-G) codificada en el vector pMD.G, debido a su amplio tropismo celular y a su elevada estabilidad.

La transfección de células HEK293 o HEK293T/17 con estos cuatro plásmidos y su correcta estequiometría permite llevar a cabo el ensamblado de LVs (Figura 5) [35], [40].

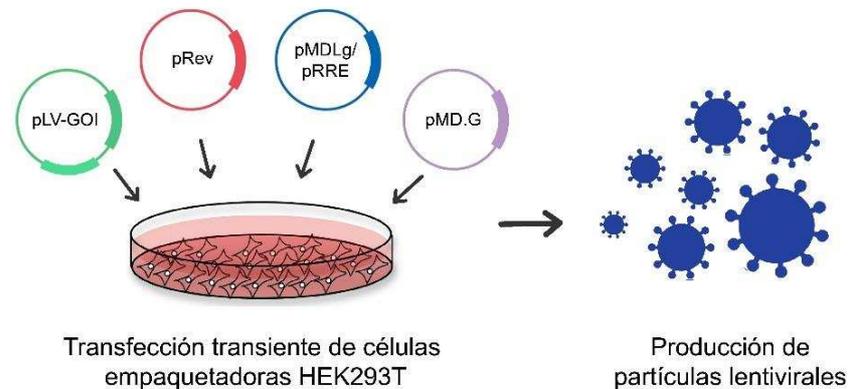


Figura 5. Ensamblado de partículas lentivirales mediante transfección de células empaquetadoras HEK293T.

Se ha demostrado que los vectores lentivirales presentan una elevada frecuencia de integración en regiones codificantes, permitiendo así la elevada expresión de los transgenes que transfieren [26,41,42]. Este resultado impulsó su empleo para la generación de líneas celulares estables productoras de proteínas recombinantes, obteniéndose como resultado líneas celulares (CHO y HEK293) con elevadas productividades específicas de las proteínas complejas como el FVIII, la EPO y diversos anticuerpos [25,36,43,44].

5.2.3. Ingeniería de genomas: modificaciones sitio dirigidas

Como se mencionó anteriormente, la integración del TG en el genoma de una célula es un evento generalmente aleatorio y el sitio de integración resulta ser un factor crucial que afectará su estabilidad y su nivel de expresión, siendo la principal causa de la heterogeneidad de expresión entre clones individuales.

Numerosos esfuerzos se han realizado con el objetivo de desarrollar nuevas estrategias que permitan llevar a cabo la manipulación de células eucariotas de manera efectiva, confiable y predecible. En los últimos años han surgido nuevas metodologías que permiten direccionar el gen de interés hacia sitios del genoma previamente caracterizados, en una región activa y estable (generalmente sitios conocidos como “hotspots”). Estas estrategias permiten sortear los obstáculos asociados al “efecto de posición” y así alcanzar niveles elevados de expresión de la proteína codificante.

El método de intercambio de *cassette* mediado por una recombinasa o *recombinase mediated cassette exchange* (RMCE) permite llevar a cabo el intercambio de dos secuencias de ADN entre el genoma celular y un plásmido que contiene el TG

de interés. Como su nombre lo indica, este procedimiento es catalizado por una recombinasa que reconoce secuencias heteroespecíficas de ADN. De esta manera, la reacción entre cada una de las secuencias mencionadas será independiente, evitando reacciones cruzadas y permitiendo la integración de una secuencia de ADN en un sitio específico del genoma [45-47]. La principal desventaja de este método es que deben introducirse previamente al genoma las secuencias heteroespecíficas de ADN que median el intercambio de *cassette*.

Las metodologías de edición de genomas desarrolladas hasta el momento se basan en la generación de un corte en las dos hebras de la molécula del ADN (*double strand break*, DSB) realizado en forma precisa y dirigida en la región a editar [48]. Este corte es luego reparado por la célula que dispone, para esto, de dos mecanismos alternativos (Figura 6).

La vía de reparación preferencial es la recombinación no homóloga o unión de extremos no homólogos (*non-homologous end-joining*, NHEJ) que consiste en la simple unión de los extremos generados y típicamente introduce mutaciones adicionales, al generar inserciones o deleciones en la zona de la unión (lo que se conoce como *knock-out*). La otra vía de reparación es la recombinación homóloga (*homologous recombination*, HR o *homology-directed repair*, HDR) que puede utilizar como molde a una molécula exógena de ADN suministrada para reparar el corte generado (conocido como *knock-in*). El cromosoma así editado es luego heredado durante la división celular.

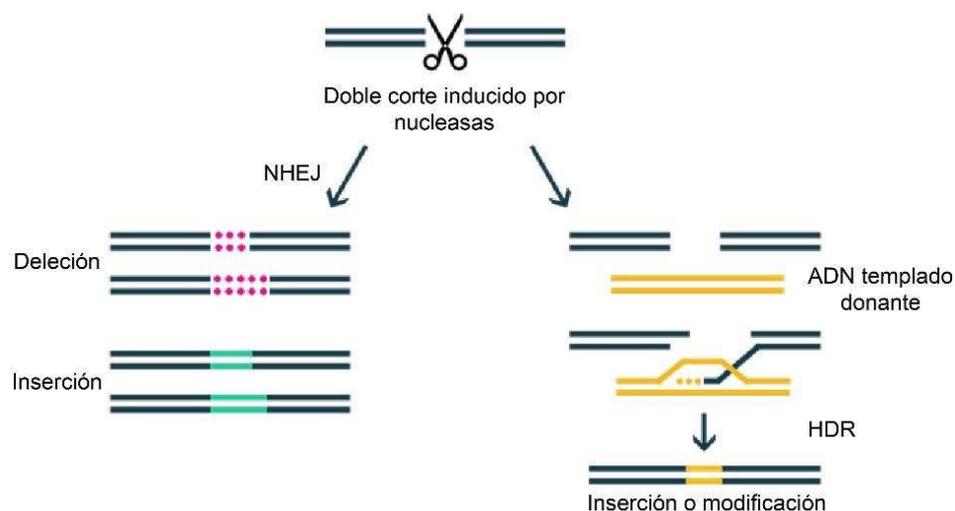


Figura 6. Mecanismos celulares de reparación del ADN: NHEJ y HDR, desencadenados por un corte en ambas hebras de la molécula de ADN.

La tecnología denominada nucleasas con dominios de dedos de zinc (*zinc finger nucleases*, ZFNs) fue usada por primera vez para modificar células humanas en 2005 [48]. Las mismas son enzimas de restricción creadas a partir de la fusión del

dominio dedo de zinc de unión al ADN, que reconoce trinucleótidos específicos, con un dominio de ruptura de ADN (Figura 7.A). La más comúnmente utilizada es la nucleasa *FokI*. Una de las ventajas que presenta este sistema es su especificidad de unión, ya que para que *FokI* produzca la escisión del ADN debe dimerizarse; por lo tanto, es necesario un par de ZFNs para que se reconozcan sitios de ADN. Sin embargo, una de las principales desventajas es que las ZFNs son difíciles de diseñar y modificar [49-52].

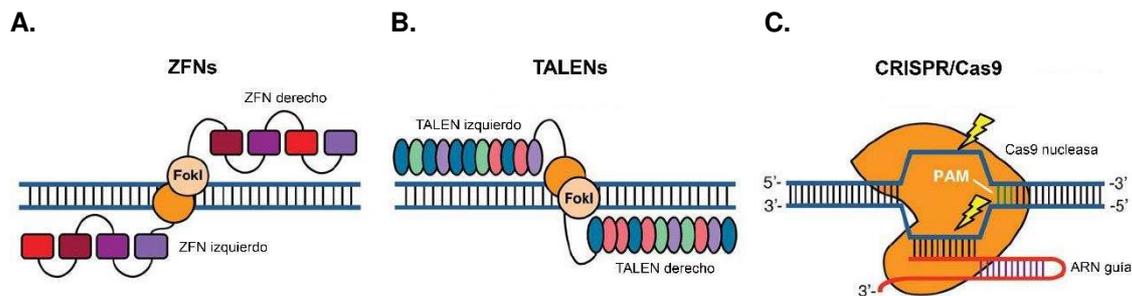


Figura 7. Tecnologías de edición genética.

- A.** Nucleasas con dominios de dedos de zinc (ZFNs), compuestas por dos dominios de dedos de zinc de unión al ADN y un dominio endonucleasa (*FokI*).
 - B.** Nucleasas efectoras tipo activadores de la transcripción (TALENs), formadas por dos dominios de unión al ADN personalizables y un dominio nucleasa (*FokI*).
 - C.** Sistema CRISPR/Cas9, compuesto por la nucleasa Cas9 y un ARN guía, que dirige la Cas9 al sitio de corte en el genoma.
- Figura modificada de Gersbach y col., 2014 [49].

Otro sistema de edición genética lo constituyen las nucleasas efectoras tipo activadores de la transcripción (*transcription activator like effectors nucleases*, TALENs) (Figura 7.B). Originalmente fueron caracterizadas en *Xanthomonas* en 2009 [53], en donde las proteínas TALENs son secretadas cuando infectan una gran variedad de plantas, activando en ellas genes que ayudan a la patogénesis. Se pueden generar secuencias personalizadas de TALENs para reconocer sitios específicos del genoma, a las que se fusiona la nucleasa *FokI* para generar el dímero funcional de corte. Son más sencillas de modificar que las proteínas ZFNs [49,50,53]. Sin embargo, dado que son de mayor tamaño, su integración génica es menos eficiente y, a su vez, al ser de origen bacteriano pueden desencadenar una respuesta inmune.

Un tercer sistema que actualmente está reemplazando ampliamente el uso de los anteriores es la tecnología CRISPR/Cas9 (Figura 7.C). Las secuencias CRISPR se encuentran en aproximadamente el 40% de los genomas bacterianos y en el 90% de los genomas secuenciados de las arqueas [54]. Con frecuencia se hallan asociadas con los genes Cas, que codifican para proteínas nucleasas. El sistema CRISPR/Cas9 es un sistema inmunitario procariótico que confiere resistencia a agentes externos como plásmidos y fagos, proveyendo así una forma de inmunidad adquirida [55].

El sistema se compone de dos partes principales, un componente proteico Cas9 con actividad nucleasa y un ARN guía (*single guide RNA*, sgRNA) que brinda especificidad al sistema [48,55-59]. La secuencia blanco o *target*, que es cortada por Cas9 en el genoma, está determinada por 20 nucleótidos presentes en la secuencia del ARN guía, por lo que basta con sintetizar una nueva molécula de ARN para redireccionar a la nucleasa hacia otro *locus*.

En este aspecto, CRISPR se distingue de las nucleasas mencionadas anteriormente para las que, por cada secuencia que se quiere editar, se requiere diseñar una nueva enzima. Esto convierte al sistema CRISPR/Cas9 en una técnica de edición extremadamente simple y accesible (en términos económicos y metodológicos), comparada con los métodos disponibles hasta su descubrimiento.

En resumen, CRISPR fue adaptado y rediseñado para un amplio abanico de aplicaciones: el encendido o apagado de genes específicos para el análisis de su función, el desarrollo de estudios en gran escala (empleando numerosos sgRNAs diferentes al mismo tiempo) para la identificación de genes involucrados en un proceso celular determinado, la creación de modelos animales para el estudio de enfermedades humanas, el direccionamiento de proteínas a secuencias específicas del genoma, la modificación genética de plantas y embriones, el tratamiento de enfermedades genéticas en seres humanos, entre muchas otras.

En la Tabla 3 se muestra una comparación de diversos aspectos técnicos para el diseño y el empleo de los sistemas de edición de genomas ZFNs, TALENs y CRISPR/Cas9.

Tabla 3. Comparación de diferentes aspectos técnicos para el diseño y uso de las tecnologías de edición genética ZFNs, TALENs y CRISPR/Cas9.

Tabla modificada de Feng y col., 2015 [60].

| Factores | ZFNs | TALENs | CRISPR/Cas9 |
|----------------------------------|--|------------------------------|--------------------------|
| Construcción | Significativo | Significativo | Simple |
| Evaluación <i>in vitro</i> | Significativo | Significativo | Simple |
| Eficiencia (<i>target</i>) | Factor limitante | Regular | Bueno |
| Efectos <i>off-targets</i> | Alto | Bajo | Alto |
| Elección del sitio <i>target</i> | Limitado | Limitado | Prácticamente ilimitado |
| Mutaciones en múltiples genes | Limitado | Limitado | Ilimitado |
| Componente a diseñar | Proteína | Proteína | ARN |
| Componentes esenciales | Proteínas dedos de zinc y nucleasa <i>FokI</i> | TALEs y nucleasa <i>FokI</i> | Proteína Cas9 y ARN guía |
| Tiempo necesario | Mucho (7-15 días) | Mucho (7-15 días) | Poco (1-3 días) |
| Costo | Alto | Alto | Bajo |

6. Tecnologías de secuenciación masiva de ADN

Desde la aparición en 1977 del método de Sanger, considerado de primera generación, las técnicas de secuenciación de ADN han evolucionado enormemente, dando lugar en los últimos años a las nuevas tecnologías de secuenciación masiva o de nueva generación (*next generation sequencing*, NGS). Si bien la técnica desarrollada por Sanger se fue automatizando y perfeccionando, permitiendo hoy obtener lecturas relativamente largas (desde 600 a 900 pb aproximadamente) con una tasa de error entre 0,01 y 0,10%; los costos (en relación a tiempo y dinero) de producir suficientes lecturas para secuenciar incluso un genoma pequeño siguen siendo muy elevados.

A diferencia de los sistemas de secuenciación tradicionales, las tecnologías NGS son capaces de secuenciar paralelamente, y de forma masiva, millones de fragmentos de ADN en un único proceso de secuenciación, en un tiempo récord y con costos cada vez más reducidos [61] (Figura 8).

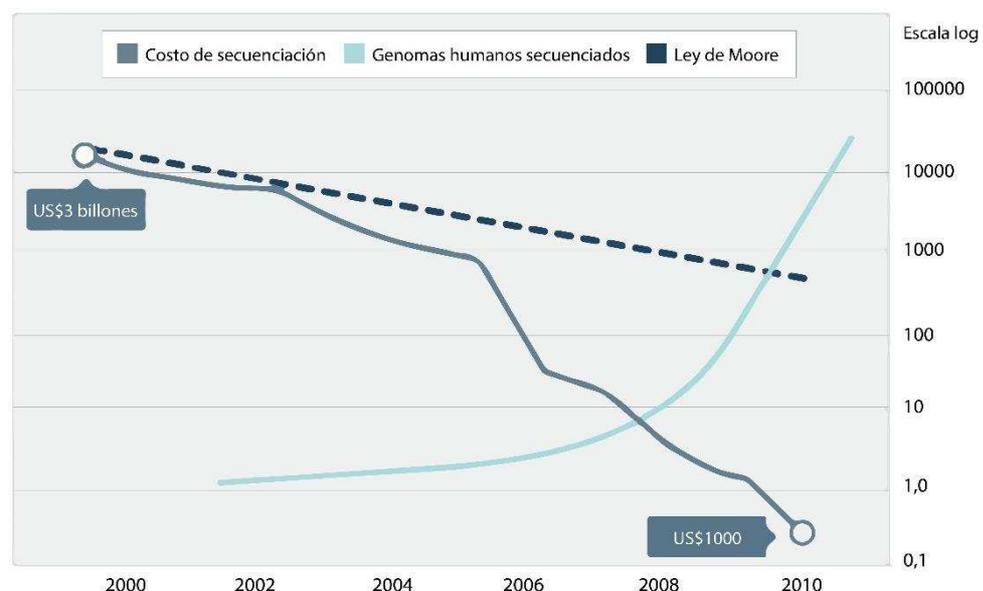


Figura 8. Caída en los costos de secuenciación y aumento en la cantidad de datos genómicos.

El Proyecto Genoma Humano (que empleó la técnica de Sanger) comenzó en 1990 y culminó en 2001, y costó 3 mil millones de dólares. Actualmente, con el uso de las tecnologías de NGS y las de secuenciación de tercera generación, un genoma puede secuenciarse por aproximadamente US\$ 1000 en un día. Desde 2000 hasta 2010, la cantidad de datos de genomas aumentó exponencialmente, mientras que el costo de secuenciación se redujo sustancialmente.

Figura modificada de Mallawaarachchi, 2018 [62].

La segunda generación de plataformas de secuenciación inicia su actividad comercial en 2005, provocando una auténtica revolución en la investigación biológica. Si bien es cierto que la secuenciación tradicional, mediante la tecnología Sanger, ha dominado el mercado durante al menos dos décadas, la acelerada evolución de estos

nuevos secuenciadores, tanto en precisión como en rendimiento, así como el abaratamiento del costo por base leída han permitido la rápida expansión de su uso en la comunidad científica, ofreciendo nuevas alternativas para la secuenciación de genomas completos, resecuenciación dirigida de zonas concretas del genoma, secuenciación de transcriptomas (*RNA-sequencing*, RNA-seq), identificación de microARNs, estudios de interacción proteína-ADN (*chromatin-immunoprecipitation sequencing*, ChIP-seq), estudios de metilación, entre otros.

Cada plataforma de secuenciación se basa en principios químicos distintos que generan diferencias cualitativas y cuantitativas [63], [64]. A pesar de ello, presentan varias particularidades comunes:

- El ADN se fragmenta al azar y se unen adaptadores específicos en ambos extremos de cada fragmento, directamente sin necesidad de clonarlo en bacterias.
- La amplificación de la biblioteca se produce mediante el anclaje del fragmento de ADN, a través de sus adaptadores, a una superficie sólida como son las microesferas o directamente a la placa de secuenciación.
- La secuenciación y detección de las bases ocurren al mismo tiempo en todas las moléculas de ADN (secuenciación masiva y paralela).
- Las plataformas NGS permiten realizar un tipo de secuenciación de lecturas pareadas (*paired-end*, PE), mediante la cual es posible leer ambos extremos del mismo fragmento de ADN. Esta estrategia facilita no sólo el posicionamiento de aquellas lecturas que pueden alinear en múltiples sitios del genoma, sino que también posibilita la identificación de variantes estructurales.

A pesar de sus innegables ventajas, la secuenciación masiva también presenta una serie de limitaciones que van superándose conforme se desarrollan nuevas mejoras tecnológicas. Precisamente, una de las mayores dificultades es el manejo del elevado volumen de datos generado en cada experimento [65], que requiere un cierto manejo de técnicas bioinformáticas avanzadas y de una adecuada infraestructura computacional, no siempre al alcance de los grupos de investigación [66]. Además, generalmente hay una ausencia de bases de datos y recursos “ómicos” para organismos que no son considerados modelos, lo cual dificulta ciertos análisis.

Además de las diferencias metodológicas para el diseño de bibliotecas genómicas a secuenciar y de los fundamentos de cada técnica de secuenciación, existe una división entre las plataformas de secuenciación masiva en función de la longitud de las lecturas que generan (Figura 9), de manera que se habla de plataformas de lecturas largas

(>300 pb) y de lecturas cortas (<300 pb) [67]. A lo largo de los últimos años, las plataformas de lecturas cortas han ido incrementando progresivamente la longitud de lectura alcanzada, manteniendo una tasa de error baja (0,01%), mientras que las plataformas de lecturas largas han conseguido reducir el error (10-40%), logrando secuenciar moléculas completas, como por ejemplo de ARNm [68]. Estos avances metodológicos abren la puerta a la obtención de resultados más robustos, incluso con el uso combinado de distintas plataformas.

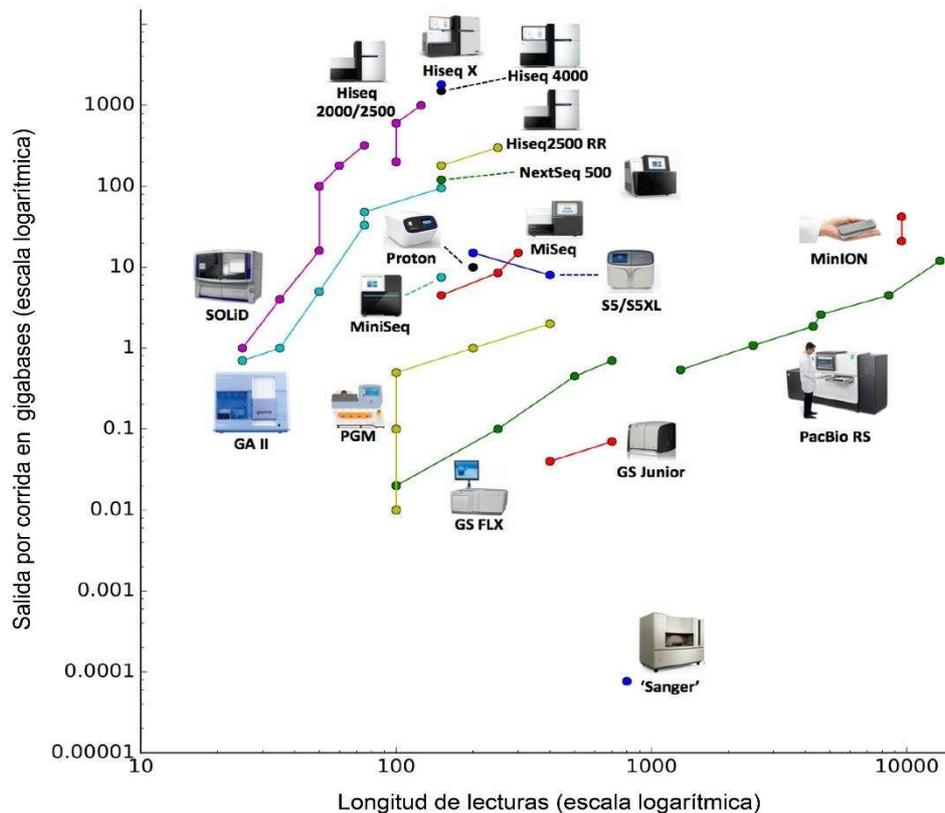


Figura 9. Representación gráfica del rendimiento de cada equipo de secuenciación (Gb) en función de la longitud de lectura (pb), mostrando el desarrollo de las diferentes plataformas de secuenciación.

Los datos se basan en las mediciones del rendimiento de cada tecnología, desde la aparición de la primera versión del equipo. Figura modificada de Nederbragt, 2016 [69].

La primera plataforma de secuenciación masiva que se empezó a comercializar, en 2005, fue el equipo 454 *Genome Sequencer* de Life Sciences (adquirido por Roche en 2007). Esta plataforma se basa en la técnica de pirosecuenciación [70], la cual detecta la señal luminosa generada por una reacción quimioluminiscente, donde se utiliza como sustrato el pirofosfato liberado al incorporarse un nuevo nucleótido a la cadena de ADN naciente. Tras sucesivos perfeccionamientos, esta plataforma consiguió alcanzar longitudes de lectura de 700 pb en promedio, con mayor precisión y un volumen total de secuencia por corrida de 0,7 Gb en 24 h. Aunque en sus inicios la tecnología

454 constituía una alternativa rápida y eficiente para conseguir lecturas largas con una precisión media [71], los principales inconvenientes de esta tecnología son el alto costo de los reactivos y las elevadas tasas de error en las regiones de homopolímeros. Por ello, Roche anunció en octubre de 2013 que para mediados de 2016 no seguiría dando soporte a esta plataforma, por su incapacidad de competir con las nuevas tecnologías que se habían desarrollado.

Al poco tiempo de la aparición de la tecnología 454 surgieron dos plataformas caracterizadas por la identificación de las bases a través de fluorescencia y por la generación de un número muy superior de lecturas más cortas, pero más precisas.

Por un lado, Applied Biosystems (posteriormente Life Technologies y actualmente Thermo Fisher Scientific) desarrolló en 2006 el sistema SOLiD (*sequencing by oligo ligation detection*), basado en la tecnología de secuenciación por ligación y codificación por dos nucleótidos [72]. Esta tecnología producía inicialmente 3 Gb de lecturas de 35 pb, pero los modelos actuales consiguen generar salidas de hasta 300 Gb y lecturas pareadas de 2 x 75 pb, con una precisión del 99,9%. La ventaja principal de este método es la precisión, ya que cada base es interrogada dos veces. Las principales desventajas radican en las lecturas cortas (50-75 pb), los tiempos de ejecución muy largos (de 7 a 14 días) y la necesidad de contar con una infraestructura computacional avanzada y personal experto para el análisis de los datos crudos.

Por otro lado, la compañía *Illumina* Inc. ha desarrollado distintos modelos de secuenciadores, tras adquirir el primer *Genome Analyzer* comercializado por Solexa en 2006. Los diferentes secuenciadores de *Illumina* (HiSeq, MiSeq, NextSeq, etc.) utilizan la tecnología de secuenciación por síntesis [73]. La misma se caracteriza por el uso de nucleótidos marcados con fluoróforos que bloquean de forma reversible la elongación de la cadena. De este modo, tras la detección de la incorporación del nucleótido marcado, y la eliminación del mismo, es posible continuar con un nuevo ciclo de adición de otro nucleótido. La amplificación clonal previa a la secuenciación se realiza por la técnica de PCR de puente (*bridge PCR*). Los secuenciadores de *Illumina* pueden generar una salida de 15 Gb de lecturas pareadas 2 x 300 pb (modelo MiSeq) o bien lecturas de 2 x 150 pb hasta 1800 Gb de salida (modelo HiSeqX). Actualmente, es la plataforma de NGS más exitosa con un dominio mayor al 70% del mercado mundial, particularmente con los modelos mencionados.

En 2010, Ion Torrent (actualmente Thermo Fisher Scientific) comercializó su sistema PGM (*Personal Genome Machine*), el cual incorpora la tecnología de semiconductores y no depende del uso de fluorescencia, sino de los cambios en el pH que se producen cuando se libera un protón al incorporarse un nucleótido a la molécula de ADN. El proceso se lleva a cabo en un *chip* con micro-pocillos cuya base actúa como

un pH-metro microscópico para detectar dichos cambios. El sistema PGM es capaz de generar lecturas largas de hasta 400 pb y 2 Gb como salida. La evolución de la tecnología de semiconductores ha confluído en los secuenciadores Ion Proton que son capaces de generar lecturas de 200 pb con salidas de hasta 15 Gb. La principal desventaja del sistema radica en la lectura de regiones con repeticiones de homopolímeros. Dentro de las ventajas de esta plataforma se pueden nombrar las longitudes de lecturas relativamente largas, un flujo de trabajo flexible y tiempos y costos reducidos.

Más recientemente se han desarrollado los secuenciadores denominados de tercera generación. Las principales diferencias de estas plataformas con las de segunda generación radican en que no es necesario realizar una PCR antes de la secuenciación (aunque existen protocolos para la construcción de bibliotecas genómicas sin PCR para su secuenciación en *Illumina*), reduciendo así el tiempo de análisis y los errores derivados de dicha técnica y, además, en que la detección de nucleótidos se realiza en tiempo real.

La plataforma PacBio (Pacific Biosciences) se basa en fluorescencia para identificar los nucleótidos mediante un procedimiento de secuenciación de una única molécula en tiempo real (*single-molecule real-time*, SMRT). Dado que los cuatro nucleótidos se añaden simultáneamente y se detectan en tiempo real, la velocidad de secuenciación aumenta mucho en comparación con las tecnologías previamente mencionadas, reduciendo considerablemente los tiempos. Además, se consiguen longitudes de secuencia, en promedio, desde 4200 a 8500 pb, llegando hasta 30000 pb.

Oxford Nanopore Technologies proporciona el último sistema de secuenciación de una única molécula. Esta plataforma de secuenciación utiliza un nanoporo (similar a un canal iónico de una membrana plasmática) por el que transitan las moléculas de ADN, produciendo una interrupción en el voltaje del canal [74] y cuyo registro permite identificar la secuencia de nucleótidos. La idea principal detrás de la secuenciación de ADN y ARN con nanoporos es que la conductividad de las corrientes iónicas en el poro cambia cuando la cadena de ácido nucleico pasa a través del mismo. El flujo de la corriente iónica depende de la forma de la molécula que pasa a través del poro y, como los nucleótidos tienen diferentes estructuras, cada uno es reconocido por el efecto que induce sobre la corriente iónica [74]. Oxford Nanopore Technologies ha comercializado una serie de secuenciadores que utilizan esta tecnología (MinION, GridION y PromethION). El secuenciador MinION pesa menos de 100 g y se conecta directamente a una computadora mediante un puerto USB 3.0. Puede generar en promedio 10-20 Gb de salida, siendo las lecturas de más de 10000 pb. El GridION X5 es un sistema de

mesada compacto, diseñado para correr y analizar hasta cinco placas de secuenciación (*flow-cell*) de tipo MinION. PromethION se utiliza para secuenciación a mayor escala.

Existen otros sistemas de secuenciación de tercera generación como el de la plataforma Helicos (Helicos Bioscience), que no se comercializan actualmente, mientras que otros sistemas se encuentran en fase incipiente de desarrollo.

7. Tecnologías “ómicas” aplicadas al estudio de líneas celulares productoras de proteínas recombinantes

Los desarrollos en el campo de la bioinformática, junto al perfeccionamiento de las plataformas de alto rendimiento, han contribuido al avance de las técnicas denominadas "ómicas", permitiendo desarrollos extraordinarios en los campos de la salud y la biotecnología.

El término “ómica” hace referencia a la cuantificación y caracterización de diferentes tipos de moléculas presentes en una condición particular en un momento específico [75]. Bajo este concepto se engloban diferentes disciplinas como la genómica (estudio del genoma y la función de los genes), la proteómica (estudio de la estructura, función, localización e interacción de proteínas), la transcriptómica (estudio del transcriptoma y la expresión génica), la metabolómica (estudio de metabolitos y productos metabólicos), la glicómica (estudio de la identidad de los carbohidratos), entre otras.

A pesar de la rapidez con la cual estas técnicas han avanzado en los últimos años, la aplicación de las mismas para el estudio y optimización de los cultivos celulares de mamífero ha sido mucho más lenta que en otras áreas, particularmente en células CHO. Esto puede deberse, en parte, a los desafíos generados por la complejidad de dichas células (genomas más grandes, compartimentalización, modificaciones proteicas complejas) y a un financiamiento limitado [18]. No obstante, el uso de las “ómicas” está ganando prevalencia en la investigación académica y en la industria de producción de proteínas recombinantes terapéuticas, lo que provoca un gran progreso y maduración de la cantidad y calidad de los datos disponibles en esta área.

En la actualidad, los procesos de mejoramiento de los cultivos celulares dependen en gran medida de una optimización empírica, que insume tiempo y recursos considerables [18]. Si bien algunas estrategias, como los diseños experimentales multifactoriales, han logrado optimizar ciertos aspectos de los procesos de cultivo, estos enfoques empíricos deben aplicarse a cada línea celular, lo que demanda bastante tiempo y esfuerzos [32].

Además, poco se puede comprender sobre cómo o por qué algunas condiciones específicas mejoran dichos procedimientos [76,77], lo cual a su vez resulta en

comportamientos impredecibles cuando los procesos celulares son escalados o transferidos a otras instalaciones [18]. Por estas razones, las tecnologías “ómicas” permiten capturar diversas características intracelulares, mejorando y transformando el desarrollo de las líneas celulares con aplicación industrial [18].

El uso de las “ómicas” en el área de producción de proteínas recombinantes terapéuticas representa una oportunidad enorme para caracterizar y comprender el estado del cultivo celular y optimizar racionalmente los procesos en biorreactor. Además, la información obtenida mediante estos estudios posibilita elegir ciertos blancos genómicos para favorecer fenotipos deseados y aportar información para la formulación de los medios de cultivo y de los diferentes diseños de biorreactores [78].

Las líneas celulares empleadas en aplicaciones industriales son modificadas genéticamente con el fin de lograr ciertos atributos fenotípicos requeridos, que incluyen la producción y la secreción de proteínas exógenas. Además, diversas modificaciones no deseadas, como reordenamientos cromosómicos, duplicaciones genéticas y cambios epigenéticos pueden desencadenarse en respuesta señales de estrés, tales como la presión de selección, las condiciones de cultivo y la adaptación a los diferentes medios [79]. Como resultado, cada línea celular industrial tiene un perfil genómico y epigenético único.

Por ello, dada la variedad de modificaciones que pueden ocurrir, se dificultan las comparaciones entre diferentes líneas celulares recombinantes. Sin embargo, contrastar con líneas celulares parentales, que están modificadas en menor medida, suele ser un enfoque más común, ya que se puede utilizar la caracterización de la línea salvaje (*wild type*, *wt*) como estado de referencia contra el que pueden compararse simultáneamente otras líneas celulares modificadas [75]. Por estas razones, las líneas celulares CHO parentales, incluyendo el hámster chino, DG44, CHO-S y CHO-K1, se están analizando cada vez más en diferentes estudios “ómicos” a gran escala, especialmente en el área académica. Los estudios más relevantes hasta el momento se resumen en la Tabla 4.

Mediante el empleo de las tecnologías de NGS para la secuenciación masiva de ADNc, lo que se conoce como RNA-Seq, se obtiene un panorama global del transcriptoma de diversos organismos a un nivel sin precedentes de sensibilidad y precisión. Este método permite medir niveles de expresión génica y obtener información que, previamente a esta técnica, era inaccesible, como la expresión alelo específica y el descubrimiento de nuevos promotores e isoformas [80].

Sin embargo, los volúmenes de datos producidos son grandes y complejos, por lo cual su comprensión no es una tarea sencilla. La etapa de procesamiento de datos y su

posterior análisis es crucial, y es aquí donde la disciplina Bioinformática adquiere vital importancia.

Tabla 4. Principales estudios “ómicos” en células CHO, considerados de referencia.
Tabla modificada de Lewis y col., 2016 [75].

| Referencia | Técnica | Resumen |
|---|-----------------|--|
| Xu y col., 2011 [81] | Genómica | Primera publicación disponible del genoma (borrador) de CHO-K1. |
| Lewis y col., 2013 [82] | Genómica | Primera publicación disponible del genoma (borrador) de Hámster chino (<i>Chinese hamster</i>). Seis genomas borradores de líneas celulares CHO derivadas de los linajes CHO-K1, DG44 y CHO-S. |
| Becker y col., 2011 [83]; Rupp y col., 2014 [84] | Transcriptómica | Publicación de bibliotecas de ADNc de células CHO. Particular énfasis en el metabolismo central de azúcares y de la vía de N-glicosilación. |
| Bort y col., 2012 [85] | Transcriptómica | Análisis de expresión de ARNm y de microARN durante el transcurso de cultivos en forma <i>batch</i> : fases <i>lag</i> , exponencial y estacionaria. |
| Baycin-Hizal y col., 2012 [86] | Proteómica | Primera publicación disponible del proteoma de CHO, se identificaron más de 6000 proteínas expresadas. |
| Slade y col., 2012 [87] | Proteómica | Se identificaron 352 proteínas secretadas en líneas celulares CHO-S y DG44. |
| Lim y col., 2013 [88] | Proteómica | Se identificaron proteínas secretadas en cultivo <i>fed-batch</i> de células CHO-S. |
| Levy y col., 2014 [89]; Valente y col., 2014 [90] | Proteómica | Se caracterizaron y cuantificaron HCP (<i>host cell proteins</i>) de células CHO. |
| North y col., 2010 [91] | Proteómica | Se caracterizaron patrones de glicosilación de proteínas expresadas en nueve líneas celulares CHO resistentes a lectina. |
| Tep y col., 2012 [92] | Proteómica | Se desarrolló un método MALDI-TOF MS para cuantificar cambios glicómicos en células CHO en biorreactor. |

En los últimos años se ha fomentado el análisis transcripcional de los cultivos celulares y se han realizado diversas investigaciones con el fin de producir herramientas e información genómica, particularmente sobre las células CHO. En 2006 se fundó el

consorcio para la genómica de células CHO, con el objetivo de acercar a la comunidad que trabaja con dicha línea celular a la era genómica. Sin embargo, a pesar de su creciente empleo biotecnológico e industrial y su importancia tanto científica como económica, todavía no se encuentra disponible una completa información del genoma y transcriptoma de las células CHO.

La mayoría de las investigaciones transcriptómicas en células CHO desarrolladas hasta el momento han empleado microarreglos de ADN para explorar las diferentes características causantes de la alta productividad en células de mamíferos. Estas investigaciones han demostrado que aquellas células altamente productoras presentan perfiles de expresión únicos para aquellos genes relacionados con el crecimiento y muerte celular [93-95]. Además, se ha estudiado el impacto de los diferentes parámetros de cultivo en el transcriptoma celular, incluyendo cultivos a temperatura reducida [96,97], cultivos en forma *batch* y *fed-batch* [98] y estudios de fenotipos relevantes [78,99]. Todas estas investigaciones han generado estrategias valiosas para la ingeniería celular de estas células, con el fin de aumentar la productividad y el crecimiento de los cultivos.

Sin embargo, los análisis empleando microarreglos de ADN se reducen a un número de genes previamente identificados y esto limita las posibilidades de obtener nuevos resultados. Las tecnologías de NGS han permitido superar estas desventajas y, en diversos trabajos, se ha demostrado la factibilidad de realizar análisis confiables de perfiles de expresión en organismos sin un genoma de referencia apropiado [100,101], proporcionando además nuevas herramientas y flujos de trabajo adecuados para el procesamiento de los datos originados mediante NGS.

El mayor avance logrado hasta el momento ha sido la publicación del genoma de la línea ancestral CHO-K1 [81]. Gracias a la publicación de este genoma, se comenzaron a diseñar y aplicar estrategias para la manipulación genética de la línea celular parental. A su vez, se potenció el desarrollo y el empleo de las técnicas “ómicas” en células CHO-K1 y en líneas celulares relacionadas. Sin embargo, este genoma aún se encuentra en estado borrador, con regiones aún sin haber sido secuenciadas y numerosos genes sin anotación, por lo cual tampoco debe considerarse completamente representativo de todas las líneas celulares CHO [81,101]. Ello se debe a que si bien las mismas derivan de la línea CHO-K1 ancestral, como han sido manipuladas con el fin de optimizarlas, esto puede provocar diversos cambios genómicos, como se mencionaba anteriormente. Por lo cual, la estrategia más adecuada consiste en analizar las líneas celulares utilizadas en cada laboratorio.

OBJETIVOS



OBJETIVOS

GENERAL

Se plantea como objetivo general la aplicación de las nuevas tecnologías de secuenciación masiva (NGS) a la línea celular CHO-K1, con el fin de desarrollar estrategias de ingeniería genética mediante el análisis bioinformático de los datos generados.

En el presente trabajo se pretende aplicar RNA-Seq a células CHO-K1 adaptadas al crecimiento en suspensión con el propósito de comparar perfiles de expresión génica en diferentes etapas de cultivo en biorreactor. Los resultados obtenidos en la secuenciación del transcriptoma estarán destinados a diseñar estrategias de ingeniería genética con un fundamento más sólido, indagar características que permitan mejorar la productividad, conocer en forma global las células con las que se está trabajando y, además, poder articular la información obtenida con otras áreas de investigación que se llevan a cabo en nuestro laboratorio. El uso de nuevos promotores endógenos constituirá una de las principales estrategias a ser aplicada a partir de los análisis transcriptómicos.

Como se mencionó en la introducción, resulta más conveniente secuenciar la línea celular con la que se trabaja en nuestro laboratorio, debido a las variaciones que presenta con respecto a la línea CHO-K1 ancestral. Además, teniendo en cuenta que los clones productores desarrollados en este grupo de trabajo provienen del mismo banco celular, se plantea el análisis de la línea *wild type* (wt). Esta denominación se refiere a aquellas células que no han sido sometidas a eventos de modificación genética y que, por lo tanto, no expresan ninguna proteína recombinante. De esta manera, los resultados serán, en principio, extrapolables a cualquier clon productor proveniente de CHO-K1 wt.

A través del análisis de los perfiles de expresión obtenidos por RNA-Seq se seleccionarán ciertos genes con elevada expresión durante el cultivo de células CHO-K1 en biorreactor, con el objetivo de usarlos para la construcción de vectores de expresión de proteínas recombinantes. Cabe esperar que los mismos puedan convertirse en una alternativa al empleo de los promotores habitualmente usados – principalmente promotores virales– con el fin de incrementar la estabilidad genética y mejorar la eficiencia del proceso de obtención de las proteínas recombinantes desarrolladas.

ESPECÍFICOS

Los objetivos particulares perseguidos son los siguientes:

- 1.** Analizar a distintos tiempos la distribución en las distintas fases del ciclo celular de una suspensión de células CHO-K1, cultivadas en un medio libre de suero en alta densidad en un biorreactor de tanque agitado.
- 2.** Obtener poblaciones celulares enriquecidas en células G0/G1 a partir de las cuales se propone extraer el ARN total para los análisis transcriptómicos.
- 3.** Secuenciar el transcriptoma de células CHO-K1 en las condiciones de cultivo descritas y, posteriormente, procesar y reconstruir los datos generados mediante diversas estrategias bioinformáticas.
- 4.** Llevar a cabo el análisis de expresión diferencial entre las etapas de cultivo en biorreactor, respecto al inicio del cultivo. Caracterizar los procesos biológicos y vías metabólicas involucradas en el bioproceso.
- 5.** Examinar particularmente aquellos genes que se encuentren sobre-expresados en condiciones de cultivo en alta densidad celular.
- 6.** Utilizar las secuencias promotoras de un grupo de genes con elevada expresión para la construcción de nuevos vectores de expresión destinados a la producción de proteínas recombinantes.
- 7.** Evaluar y contrastar la actividad transcripcional de los vectores de expresión creados, empleando un gen reportero como la proteína verde fluorescente ZsGreen1, con la actividad de aquellos promotores usados habitualmente para producir proteínas recombinantes terapéuticas (por ejemplo, CMV).



CAPÍTULO 1

*Estudio del ciclo celular de
células CHO-K1 cultivadas
en biorreactor*

CAPÍTULO 1

Estudio del ciclo celular de células CHO-K1 cultivadas en biorreactor

INTRODUCCIÓN

“La mente que se abre a una nueva idea jamás volverá a su tamaño original”.

Albert Einstein.

1. Análisis del ciclo celular en células de mamíferos

1.1. Principios básicos del ciclo celular

Tal como lo expresa la teoría celular, todas las células se forman a partir de células preexistentes. La reproducción celular requiere de la división celular, la que produce dos células hijas. Cuando una célula se divide, la información genética contenida en su ADN se duplica de manera precisa y luego las copias se transmiten a cada célula hija, conteniendo cada una la misma cantidad de ADN [102]. En los procariontes, este proceso de división es sencillo y recibe el nombre de fisión binaria. En los eucariotes, el ADN está organizado en más de un cromosoma, siendo más complejo dicho proceso de división celular. Este ciclo, coordinado y regulado, de crecimiento y división, se denomina ciclo celular. El ciclo de la célula eucariota se compone de cinco etapas, o fases, como se muestra en la Figura 10.

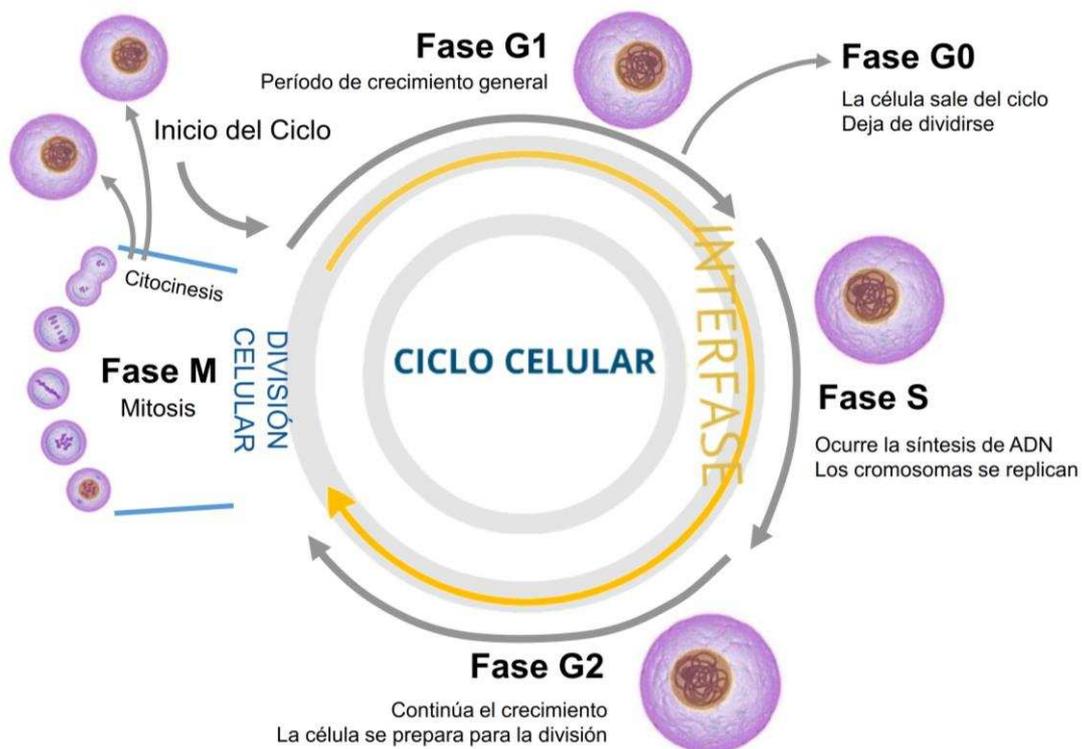


Figura 10. Representación esquemática de las fases que componen el ciclo celular eucariota.

La célula puede encontrarse en dos estados muy diferenciados: el estado de no división o interfase, en el cual la célula realiza sus funciones específicas y, si está destinada a progresar hacia la división celular, comienza realizando la duplicación de su ADN; y el estado de división, conocido como mitosis o fase M.

Dentro de la interfase, se encuentran tres grandes etapas [103], que abarcan el 90% de la totalidad del ciclo celular:

- **Fase G1** (*growth* o *Gap1*): primera fase de crecimiento, se "inicia" con una célula que proviene de una división previa. Durante esta fase existe crecimiento celular –con síntesis de proteínas y ARN– duplicando la célula su tamaño y masa debido a la continua síntesis de todos sus componentes. Tiene una duración de aproximadamente 12 h [102].
- **Fase S** (síntesis): en esta etapa se produce la replicación o síntesis del ADN; como resultado cada cromosoma se duplica y queda formado por dos cromátidas idénticas. Con esta duplicación del ADN, el núcleo celular presenta el doble de proteínas nucleares y de ADN que al inicio del ciclo. Tiene una duración aproximada de 6 h [102].
- **Fase G2** (*growth* o *Gap 2*): es la tercera fase de crecimiento del ciclo celular en la que continúa la síntesis de proteínas y ARN. Se incrementan las proteínas citoplasmáticas y organelas, por lo que la célula aumenta de tamaño y se observan cambios visibles en la estructura celular que indican el principio de la mitosis. Tiene una duración estimada de 4 h [102].

Luego comienza la **Fase M** (mitosis y citocinesis), en la cual la célula progenitora (célula somática) se divide en dos células hijas idénticas. Esta etapa incluye la mitosis, que a su vez se divide en: profase, metafase, anafase, telofase y la citocinesis, que se inicia ya en la telofase mitótica. Esta etapa dura aproximadamente 0,5 h [102].

Las células de mamífero proliferan sólo cuando son estimuladas a través de señales intracelulares (como factores de transcripción) y extracelulares (factores de crecimiento, hormonas o mitógenos). Si tales señales no existen, el ciclo celular se detendrá en un punto de control G1 y la célula entrará en la **Fase G0**, permaneciendo en este estado por días, semanas o incluso años antes de volver al ciclo para dividirse.

Mientras tanto, la célula continúa metabólicamente activa, a pesar de haber cesado su crecimiento. Si recibe nuevamente señales, en particular extracelulares, la célula es estimulada para salir de G0 y entrar a G1, de forma tal de iniciar un nuevo ciclo de división.

1.2. Análisis del ciclo celular a través de citometría de flujo

Una de las primeras aplicaciones de la citometría de flujo fue la medición del contenido de ADN en las células, logrando una identificación rápida de las fases del ciclo celular. Este análisis se basa en la capacidad de tinción del ADN celular de manera estequiométrica (la cantidad de colorante es directamente proporcional a la cantidad de ADN dentro de la célula). De este modo, se produce un patrón característico que refleja las fases del ciclo celular dentro de la población de células en estudio [104].

Para ello, existen diversos marcadores fluorescentes, los cuales tienen una alta afinidad por el ADN. En el análisis de ácidos nucleicos es posible utilizar diversos tipos de fluorocromos, dependiendo de su espectro de emisión y fundamentalmente del tipo de láser con que esté equipado el citómetro de flujo a usar [105]. Generalmente, se usa yoduro de propidio (*propidium iodide*, PI), excitándolo con un láser de argón que emite luz a 488 nm. Este tipo de láser es el que tiene incorporado la mayoría de los citómetros de flujo disponibles.

Las células diploides ($2N$) en las fases G₀ y G₁ tienen una única copia de ADN; por lo tanto, la intensidad de fluorescencia será de $1X$ (lo que en el campo de la citometría de flujo se conoce como índice de ADN). En cambio, las células en las fases G₂ y M tienen dos copias de ADN; por lo tanto, tendrán el doble de intensidad de fluorescencia (índice de ADN igual a 2). Debido a que las células en la fase S están sintetizando el ADN, tendrán una fluorescencia con valores variables entre las poblaciones de $1X$ y $2X$. De esta manera, el ciclo celular se puede graficar como una curva que muestra las variaciones del contenido de ADN en función del número de células analizadas (Figura 11) [102,104].

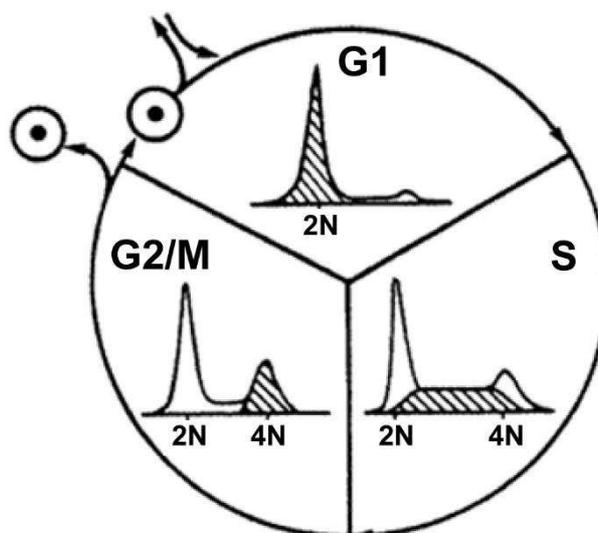


Figura 11. Representación esquemática del ciclo celular junto con los histogramas típicos de intensidad de contenido de ADN, para cada fase.

Figura modificada de Rabinovitch, 2017 [102].

La fluorescencia medida en las células en fase G0 y en fase G1 produce una curva gaussiana. De la misma forma, las células en G2 –que tienen duplicado su ADN con respecto a las células en G1– también producen un pico distribuido como una curva normal. Los mismos factores que afectan la amplitud de los picos G0/G1 y G2/M también afectan la amplitud de la fase S. Por ello, resulta que en un histograma de este tipo, las células que inician la fase S se superponen con las células en G1 así como también lo hacen las células al final de la fase S con las células en G2 [106,107]. Por esta razón, los programas informáticos usados para modelar ciclo celular deben realizar una deconvolución de las poblaciones y asignar los porcentajes de células correspondientes a cada población.

Existen diversos modelos matemáticos para estimar la fase S y los demás componentes de un histograma de ADN (como *debris* y agregados), siendo los modelos rectangular, trapezoidal y polinomial los más utilizados. Además, para un correcto modelado de las poblaciones celulares en cada fase del ciclo, se deben discriminar los dobletes de células G0/G1, ya que serán interpretados en el pico correspondiente a G2/M, provocando a una sobreestimación de esta población [106].

2. Sincronización de cultivos de células de mamífero en la fase G0/G1 para incrementar la productividad de proteínas recombinantes terapéuticas

La producción de proteínas terapéuticas es un proceso largo y complejo, desde el clonado inicial del TG en un vector de expresión apropiado hasta el procesamiento final como producto farmacéutico en el mercado, requiriéndose una gran cantidad de recursos para controlar, monitorear y asegurar la calidad en cada etapa. Además, la producción de cada proteína terapéutica específica debe optimizarse, principalmente a escala industrial [3]. En particular, para lograr una mayor productividad se debe conocer el efecto que tienen diversos parámetros sobre el cultivo, como la temperatura, la capacidad del biorreactor, la velocidad de agitación, la composición del medio empleado, la saturación de oxígeno, la concentración de los productos de desecho, entre otros.

El crecimiento de los cultivos celulares en procesos productivos puede dividirse en cuatro etapas: *lag* (latencia), *log* (exponencial), estacionaria y muerte. No es deseable que la fase de latencia sea muy prolongada, por lo cual es necesario comenzar un bioproceso con concentraciones elevadas, para lograr así suficiente cantidad de células de forma tal que la producción sea máxima [108].

Asimismo, tener una proliferación descontrolada, más allá de una densidad celular deseada, también es algo poco deseable, ya que se agotan los nutrientes y el oxígeno

disuelto, se acumulan metabolitos tóxicos, se degrada el producto y esto conlleva a la muerte celular [109-111].

Existen diversas estrategias para el control de la proliferación excesiva, que se implementan típicamente en la mitad de la fase exponencial o hacia el final de la misma y que optimizan tanto la productividad como la duración de tales procesos productivos [108]. Dicha regulación de la proliferación celular generalmente se puede lograr a través del arresto del ciclo celular o mediante el control de los nutrientes disponibles en el medio de cultivo.

Se sabe que la producción proteica es dependiente de la fase del ciclo celular y, particularmente, diversos genes implicados en la biogénesis del ribosoma y la traducción proteica se encuentran altamente expresados en la fase G1 [112-121]. Las células arrestadas en el final de la fase G1 (G0/G1) del ciclo celular son metabólicamente más activas y de mayor tamaño que aquellas no arrestadas [116,120]. Por estas razones, se considera a dicha fase como la etapa ideal para incrementar la producción de proteínas recombinantes. Esta estrategia se ha implementado en numerosas investigaciones que han logrado aumentar la productividad de diversas líneas celulares comercialmente relevantes, como hibridomas y células CHO [112-115,117-119,122-124].

Existen diferentes enfoques para arrestar los cultivos en la fase G1 del ciclo celular, entre ellos la reducción de la temperatura, la ingeniería celular y la utilización de compuestos químicos [108]. En particular, se ha visto que la disminución de la temperatura reduce la proliferación celular, incrementando la longevidad del cultivo y la productividad específica en células CHO productoras de diversas proteínas recombinantes [118,119,124-131]. Además, la calidad del producto recombinante, en relación al patrón de isoformas, contenido de ácido siálico y actividad biológica *in vivo*, se mantiene en condiciones de baja temperatura [118,131,132]. Dicha reducción en el metabolismo implica la disminución del consumo de glucosa (por ende, la del medio de cultivo), del oxígeno disuelto y de la generación de productos tóxicos, como lactato y amonio. Esto conlleva a que las células entren en apoptosis de manera tardía, favoreciendo así que la fase estacionaria o de producción se prolongue por más tiempo [114,118,126,127,133-135]. Asimismo, diferentes estudios han comprobado que el uso de una menor temperatura de cultivo incrementa los niveles de ARNm recombinante, debido a que aumentan los niveles de transcripción del gen de interés o bien, mejora la estabilidad del mismo [118,130].

Por estas razones, en muchos cultivos celulares se lleva a cabo un proceso bifásico para la producción de proteínas recombinantes. El mismo consta de dos fases: una etapa de proliferación inicial a 37°C, en la cual se trata de maximizar la biomasa; seguida de una fase de producción extendida, a una temperatura reducida (en el rango

de 28°C a 33°C), durante la cual las células permanecen arrestadas mientras la producción de proteínas heterólogas se mantiene en un nivel alto y estable.

Sin embargo, actualmente se conoce muy poco a nivel molecular sobre la respuesta de los cultivos de células de mamífero a bajas temperaturas, o el mecanismo por el cual las células mejoran la productividad en estas condiciones [108]. Se ha visto que la respuesta celular es activa y coordinada y, aunque el metabolismo proteico global se reduce, se desencadena la síntesis de proteínas inducibles a bajas temperaturas, como CIRP (*cold-inducible RNA binding protein*), RBM3 (*RNA binding motif protein 3*), Vim (*vimetin*), PDI (*protein disulphide isomerase*), entre otras [96,136-142].

Además, producir el arresto celular en la fase G1 mediante la reducción de la temperatura, también tiene la ventaja de que todas las células en cultivo se verán afectadas de forma similar. En cambio, la aplicación de estrategias de ingeniería genética para este fin, conlleva a que ciertos clones puedan no incorporar los cambios deseados e, incluso, la mutación o pérdida de función de genes necesarios para el control de la proliferación puede desencadenar una situación desfavorable que produzca una pérdida en la productividad esperada [108]. De igual manera, el empleo de agentes químicos para producir el arresto celular puede tener efectos secundarios no deseados sobre el comportamiento y la fisiología celular. Por lo tanto, el uso de un gradiente de temperatura para la producción de diversas proteínas recombinantes es el método de elección empleado en los bioprocesos industriales.

OBJETIVOS

GENERAL

Como objetivo general para esta primera etapa nos planteamos caracterizar el ciclo celular de las células CHO-K1 cultivadas en biorreactor, con el fin de dilucidar si es posible establecer una correlación entre las fases de crecimiento celular y la variación de la distribución relativa de las células en las diferentes fases del ciclo celular.

Por otro lado, sincronizar dichos cultivos para obtener mayor cantidad de células en la fase G0/G1 del ciclo celular, ya que dicha etapa es considerada el período ideal para incrementar la producción de proteínas recombinantes.

ESPECÍFICOS

- 1.** Analizar a distintos tiempos la distribución en las distintas fases del ciclo celular de una suspensión de células CHO-K1, cultivadas en un medio libre de suero en alta densidad en un biorreactor de tanque agitado.
- 2.** Determinar si el cultivo tiende a estacionarse en una fase particular del ciclo, en respuesta a las condiciones típicas de un bioproceso de células CHO-K1 adaptadas al crecimiento en suspensión.
- 3.** Obtener poblaciones celulares enriquecidas en células G0/G1 a partir de las cuales se propone extraer el ARN total para los análisis transcriptómicos subsiguientes.

MATERIALES Y MÉTODOS

1. Técnicas de biología molecular

No se aplicaron técnicas de biología molecular para las actividades desarrolladas en este Capítulo.

2. Técnicas de cultivo celular

2.1. Líneas celulares

2.1.1. CHO-K1 (ATCC)

La línea CHO-K1 fue obtenida de ATCC (*American Type Culture Collection*®, EE.UU.) (ATCC CCL-61). Es un clon de la línea celular CHO derivada de ovario de hámster chino adulto.

2.2. Medios de cultivo y suplementos para células animales

Los medios de cultivos utilizados para el cultivo de células animales fueron preparados utilizando agua Milli Q y esterilizados por filtración, utilizando un cartucho Sartopore 2 (Sartorius, Alemania) con membranas de 0,22 µm de tamaño de poro. Cada medio preparado queda en cuarentena durante 48 h mientras se llevan a cabo los controles microbiológicos y se conservan a 4°C al resguardo de la luz.

A continuación se detallan los medios de cultivo empleados en este Capítulo:

2.2.1. CHO-K1s

Se utilizó el medio comercial Ex-Cell 302 (Sigma-Aldrich, EE.UU.), suplementado con glutamina 2 mM.

2.3. Sistemas de cultivos de células

Todas las líneas celulares eucariotas se cultivaron en una incubadora a 37°C, con atmósfera de CO₂ al 5% saturada de humedad. En esta etapa de trabajo, se utilizaron cultivos en suspensión.

2.3.1. Cultivos en suspensión en pequeña escala

Para los cultivos adaptados a crecer en suspensión, utilizando una incubadora orbital (Certomat CTplus, Sartorius, Alemania) con control de humedad y atmósfera de CO₂ al 5%, se emplearon los siguientes sistemas de cultivo:

- a) Tubos CellSTAR® CellReactor™ (Greiner Bio-One, Austria).
- b) Frascos Schott de 100, 250 y 500 ml.

- c) Frascos Erlenmeyer para cultivo celular (Corning, EE.UU.) de 125, 250 y 500 ml.

2.3.2. Cultivos en biorreactor tanque agitado con volumen de trabajo de un litro

Este es un sistema de multivasos con un volumen de trabajo de un litro (BIOSTAT Qplus, Sartorius, Alemania), compuesto de 6 biorreactores en forma paralela. En nuestro caso se llevaron a cabo cultivos continuos en perfusión, utilizando un filtro tipo *spin-filter* con una malla metálica con un tamaño de poro aproximado de 7 a 8 μm (*Twilled Dutch Weave*, BOPP, Países Bajos), como dispositivo de retención celular.

El biorreactor cuenta con todos los sensores y periferia necesaria para su funcionamiento (electrodo de oxígeno, electrodo redox, electrodo de pH, sensor de temperatura, agitador tipo *impeller*, sistema de aireación con burbujeador y/o membrana de aireación, botellas de alimentación, siembra y cosecha).

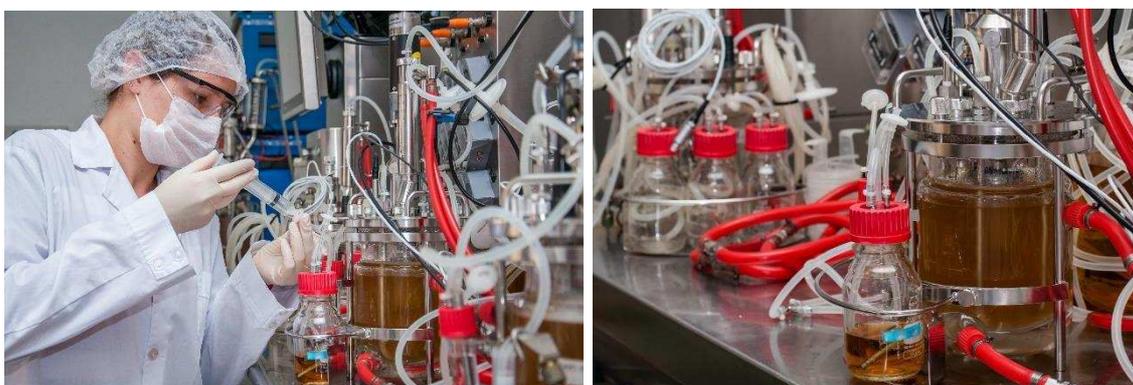


Figura 12. Fotografías del sistema multivasos (BIOSTAT Qplus) compuesto de 6 biorreactores con cultivos de células de mamífero.
Imágenes tomadas en el Laboratorio de Cultivos Celulares (FBCB-UNL).

2.4. Determinaciones analíticas en cultivos celulares

2.4.1. Determinación de la concentración y viabilidad de células

El recuento celular se realizó en cámara de Neubauer, estimando la proporción de células muertas por el método de exclusión con azul de tripán. La viabilidad del cultivo se determinó calculando el porcentaje de células viables con respecto a las totales, teniendo en cuenta que la concentración de células totales es la suma de la concentración de células viables y no viables.

2.4.2. Monitoreo de la concentración de glucosa, lactato y amonio en sobrenadantes de cultivo

a) Glucosa

La determinación de la concentración de glucosa se llevó a cabo por el método enzimático de la glucosa oxidasa/peroxidasa (GOD/POD), utilizando el *kit* comercial Glucosa Enzimática (Sociedad de Bioquímicos, Argentina). El mismo se basa en la oxidación específica de D-glucosa por parte de la enzima GOD, con formación en forma estequiométrica de H₂O₂. Este último, en presencia de 4-aminofenazona y fenol, forma una quinoneimina roja mediante una reacción catalizada por la enzima POD. De esta manera, el dosaje espectrofotométrico del producto coloreado permite la cuantificación indirecta de glucosa en la muestra.

El protocolo utilizado fue el siguiente:

1. Preparar el reactivo de trabajo de acuerdo a las instrucciones del fabricante.
2. Encender baño termostático (37°C).
3. En tubos tipo eppendorf, preparar un blanco de reactivos (B_R), un blanco de muestra (B_M), un testigo (T) y las muestras (M) de la siguiente manera:

| | B _R | B _M | T | M |
|---|----------------|----------------|-------|-------|
| Reactivo de trabajo | 1 ml | - | 1 ml | 1 ml |
| Solución testigo de glucosa 1 g.l⁻¹ | - | - | 10 µl | - |
| Muestra | - | 10 µl | - | 10 µl |
| Agua destilada | 10 µl | 1 ml | - | - |

4. Mezclar por inversión e incubar los tubos a 37°C durante 10 min.
5. Leer en espectrofotómetro a λ= 505 nm. El color es estable 1 h a temperatura ambiente.
6. Calcular la concentración de glucosa en la muestra de acuerdo a la expresión:

$$C_{glucosa} (g.l^{-1}) = [Abs_M - (Abs_{B_M} + Abs_{B_R})] * f * (D^{-1}) \quad (1)$$

Donde:

$$f = \frac{1}{(Abs_T - Abs_{B_R})} g.l^{-1}$$

La reacción de color sigue un comportamiento lineal hasta 4,5 g.l⁻¹ y las diluciones de las muestras deben efectuarse en solución tampón PBS.

b) Lactato

La determinación de lactato se llevó a cabo por el método enzimático, utilizando el *kit* comercial *Lactate* (Wiener Lab, Argentina). El mismo se basa en la oxidación específica del lactato por parte de la enzima lactato-oxidasa, con formación en forma estequiométrica de H₂O₂. Este último, en presencia de TOOS (*N*-etil-*N*-(2 hidroxil-3-sulfopropil) *m*-toluidina) y 4-aminoantipirina, forma un cromógeno mediante una reacción catalizada por la enzima POD. De esta manera, el dosaje espectrofotométrico del producto coloreado ($\lambda = 540$ nm) permite la cuantificación indirecta de lactato en la muestra.

El protocolo llevado a cabo fue el siguiente:

1. Preparar en placas de 96 pozos el estándar de lactato (por duplicado) y las muestras.
2. Realizar las diluciones de las muestras en solución tampón PBS, teniendo en cuenta que la reacción de color sigue es lineal hasta 1,3 g.l⁻¹. Se recomienda trabajar en un rango de concentraciones de 0,1 a 0,2 g.l⁻¹.
3. Sembrar 2 μ l por pozo de estándar y de las diluciones de las muestras. Procesar también un blanco de reactivos (B_R), sembrando 2 μ l de diluyente. Agregar 175 μ l por pozo de Reactivo A e incubar en estufa a 37°C durante 2 min.
4. Leer a $\lambda = 540$ nm. La lectura corresponde a los blancos de muestra (B_M).
5. Agregar 35 μ l por pozo de Reactivo B e incubar en estufa a 37°C durante 5 min (en este paso, es muy importante respetar el tiempo).
6. Leer inmediatamente a $\lambda = 540$ nm (lectura final).

La concentración de las muestras se calcula como:

$$C_{lactato} (g.l^{-1}) = (Abs_M - Abs_{B_M} - Abs_{B_R}) * f * (D^{-1}) \quad (2)$$

Donde:

$$f = \frac{Conc. \text{ estándar}}{(Abs_T - Abs_{B_M} - Abs_{B_R})} g.l^{-1}$$

2.5. Revitalización de líneas celulares de mamífero

La descongelación rápida de la suspensión celular es esencial para su óptima recuperación. Para la revitalización de las líneas celulares criopreservadas se retiró el criotubo del termo de nitrógeno líquido y se lo sumergió rápidamente en un baño de agua a 37°C.

La suspensión celular descongelada se transfirió a un tubo cónico conteniendo 40 ml de medio de cultivo y se centrifugó a 200 *g* durante 5 min a temperatura ambiente. Esta etapa se realizó con el objeto de diluir rápidamente el dimetilsulfóxido (DMSO) y luego eliminarlo, ya que es un reactivo tóxico para las células en cultivo. Se descartó el

sobrenadante y el sedimento de células se resuspendió en un volumen de medio de cultivo adecuado. Finalmente, se realizó el recuento celular con el objeto de evaluar la recuperación y la viabilidad celular.

3. Metodologías para el análisis de líneas y clones celulares

3.1. Citometría de flujo

3.1.1. Análisis del ciclo celular

Para los estudios de ciclo celular se han reportado diferentes protocolos basados en el uso de diversos agentes intercalantes de ácidos nucleicos [105,144]. Con el objetivo de realizar una primera caracterización del ciclo celular de las células CHO-K1 cultivadas en biorreactor, se recurrió a un método frecuentemente reportado en la bibliografía, el cual se basa en la determinación del contenido de ADN mediante coloración de núcleos con PI [145,146].

Las muestras se fijaron en el momento de la extracción e inmediatamente se conservaron a 4°C ON. Luego, la tinción se realizó el mismo día en que fue llevada a cabo la citometría celular de forma tal de procesar todas las muestras en el mismo experimento.

3.1.1.1. Fijación celular

Se llevó a cabo siguiendo el protocolo detallado a continuación:

1. Realizar el recuento celular de la muestra y tomar un volumen correspondiente a 14×10^6 de células viables totales. Si la densidad celular o el volumen de muestra no son suficientes, se puede tomar un volumen menor teniendo en cuenta la proporción 1,5 ml de solución de fijación por cada 1×10^6 de células totales.
2. Centrifugar las células durante 5 min a 1000 rpm. Realizar un lavado con 10 ml de PBS-EDTA 2 mM (distribuir el volumen total en dos tubos cónicos de 15 ml). Centrifugar nuevamente.
3. Descartar el sobrenadante por volcado y en el pequeño volumen remanente de PBS-EDTA, resuspender el *pellet* celular (volumen aproximado 100 ul). Realizar la homogeneización con vórtex.
4. Mantener las células en agitación con vórtex a baja velocidad y mientras tanto agregar lentamente por goteo 10,5 ml/tubo de solución de fijación fría (etanol 70% (v/v) en agua Milli Q). Homogeneizar varias veces por inversión.
5. Rotular y conservar inmediatamente a 4 °C.

3.1.1.2. Protocolo para la tinción con yoduro de propidio

1. Cosechar las células fijadas contenidas en uno de los tubos cónicos. Centrifugar durante 5 min a 1000 rpm y resuspender el *pellet* en 1 ml de PBS-EDTA. Volver a hacer un recuento celular y realizar una dilución

apropiada en PBS-EDTA para tener 1×10^6 cél./ml en un volumen mínimo de 1 ml.

2. En un tubo cónico nuevo, tomar 810 μ l de la suspensión anterior y agregar 45 μ l de RNAsa A 1 mg/ml (concentración final 50 μ g/ml). Incubar durante 15 min a 37°C.
3. Cubrir el tubo cónico con papel de aluminio, rotular y adicionar 45 μ l de solución de PI 1 mg/ml (concentración final 50 μ g/ml). Incubar durante 30 min a TA.

Luego se determinó el contenido de ADN analizando los niveles de fluorescencia por citometría de flujo, utilizando el equipo FACS ARIA III (BD, EE.UU.).

3.1.1.3. Settings para citometría

| Láser | Filtro | Espejo | Nozzle | Velocidad de flujo | Nº de eventos a grabar |
|------------|--------|--------|------------|--------------------|------------------------|
| 488 (Blue) | 695/40 | 655 LP | 70 μ m | 1,0 | 30.000 |

- Parámetros FSC, SSC, PI: amplificación lineal.
- Ajustar *Area Scaling* para FSC.
- Definir *Threshold* para FSC (aproximadamente 50.000) y para IP (aproximadamente 5.000).
- Generar gráficos *dot-plots* SSC-A vs. FSC-H, PI-A vs. PI-W; histograma PI-A.
- Ajustar voltaje PMT para FSC, SSC, PI (el pico G0/G1 debe estar centrado en el canal 50.000 del histograma).

Los datos generados fueron procesados utilizando los *software* FACSDiva v6.1.2 (BD), FlowJo X v10 (Tree Star Inc., EE.UU.) y ModFit LT v5.0 (Verity, Inglaterra).

Para garantizar la confiabilidad de los resultados, previamente se llevó a cabo el control de calidad del equipo empleando el *kit* comercial *DNA QC Particles* (BD) que permite comprobar la linealidad y resolución del instrumento necesarios para los ensayos de ciclo celular. Se siguieron las instrucciones del fabricante.

RESULTADOS Y CONCLUSIONES

1. Cultivo de células CHO-K1 en alta densidad

En primer lugar, se realizó el armado y calibración de un biorreactor multivaso *Biostat Q-plus* con un volumen de trabajo de un litro, operado en modo continuo con retención celular (modo denominado como perfusión), utilizando un dispositivo tipo *spin-filter* como sistema de retención celular.



Figura 13. Cultivo de células CHO-K1 en biorreactor con volumen de trabajo de un litro. El cultivo fue operado de modo continuo, utilizando un filtro de tipo *spin-filter* para llevar a cabo la retención celular. El equipo posee los sensores y la periferia necesarios para realizar el control *online* de diferentes variables de cultivo.

Para el estudio del ciclo celular correspondiente a esta etapa de trabajo, se llevaron a cabo dos cultivos independientes de células CHO-K1 en suspensión, empleando el medio comercial Ex-Cell 302 libre de SFB (Sigma-Aldrich).

Para la preparación del inóculo, se descongelaron células CHO-K1 con un bajo número de pasaje, adaptadas al crecimiento en suspensión. Una vez revitalizado el cultivo, este se amplificó gradualmente hasta obtener medio litro de suspensión celular con una densidad de $3,1 \times 10^5$ cél./ml, con la cual se inoculó el biorreactor. Pasadas las primeras 24 h, se realizó el recuento celular y se agregaron 250 ml de medio nuevo, alcanzando una densidad celular de $5,2 \times 10^5$ cél./ml en 750 ml de volumen final.

Se consideró dicho momento como el inicio (día cero) del primer cultivo en el biorreactor. Al día siguiente, se aumentó el volumen de trabajo a un litro, y a partir del día 4 se comenzó con un régimen de perfusión utilizando un filtro rotatorio (*spin-filter*) como dispositivo de retención celular. El abastecimiento de oxígeno se llevó a cabo mediante aireación en la cámara superior y el burbujeo a través de difusor o *sparger*, utilizando mezclas de diferentes proporciones de aire y oxígeno puro.

El primer cultivo en el biorreactor tuvo una duración de 22 días (Figura 14.A). Durante este período, se realizó el monitoreo y control de las variables operativas que se describen a continuación:

- Densidad celular y viabilidad: diariamente se tomó una muestra y se determinaron las concentraciones de células totales, células viables y viabilidad celular mediante el método de exclusión de colorantes empleando el azul de tripán y realizando los recuentos en cámara de Neubauer.
- Concentración de glucosa y lactato: una alícuota de la muestra diaria se centrifugó y el sobrenadante clarificado se utilizó para determinar las concentraciones de glucosa y lactato. Estos valores fueron tenidos en cuenta para establecer, cuando fuera necesario, el caudal de perfusión aplicado en el biorreactor.
- Concentración de oxígeno disuelto (OD): se midió *online* y se controló en forma automática, fijando como valor de *set point* un 60% de saturación de aire.
- Temperatura: se monitoreó *online* y se varió progresivamente desde 37°C a 32°C, hacia el final del cultivo.
- Agitación: se empleó un agitador de tipo hélice a una velocidad controlada que osciló entre 130 y 160 rpm.
- pH: se midió *online* y, modificando el volumen de perfusión, su valor se mantuvo en el rango 6,80 – 7,30.

Los resultados de los recuentos celulares permitieron elaborar la curva de crecimiento de las células CHO-K1 cultivadas en el biorreactor, en la que se graficó la concentración de células viables, la variación de la viabilidad celular, las determinaciones diarias de concentraciones de glucosa y de lactato y los valores de la temperatura y del volumen de perfusión, en función del tiempo de cultivo (Figura 14).

El biorreactor comenzó a perfundirse en el día 4 y el caudal varió entre 0,25 y 1,50 l/día, con el objetivo de mantener los valores de OD, pH y los metabolitos glucosa y lactato en rangos apropiados. Asimismo, los valores de temperatura, agitación y presión parcial de oxígeno se manipularon para tratar de evitar la proliferación celular excesiva, controlando así el consumo de oxígeno y la formación de espuma.

Como puede observarse en la Figura 14.A., el cultivo presentó una fase de latencia de aproximadamente cinco días, lo cual es esperable teniendo en cuenta que el cultivo se inició con una densidad celular relativamente baja. A esta fase *lag* le siguió una marcada fase de crecimiento exponencial, llegando rápidamente a una concentración celular de aproximadamente $3,0 \times 10^7$ cél./ml.

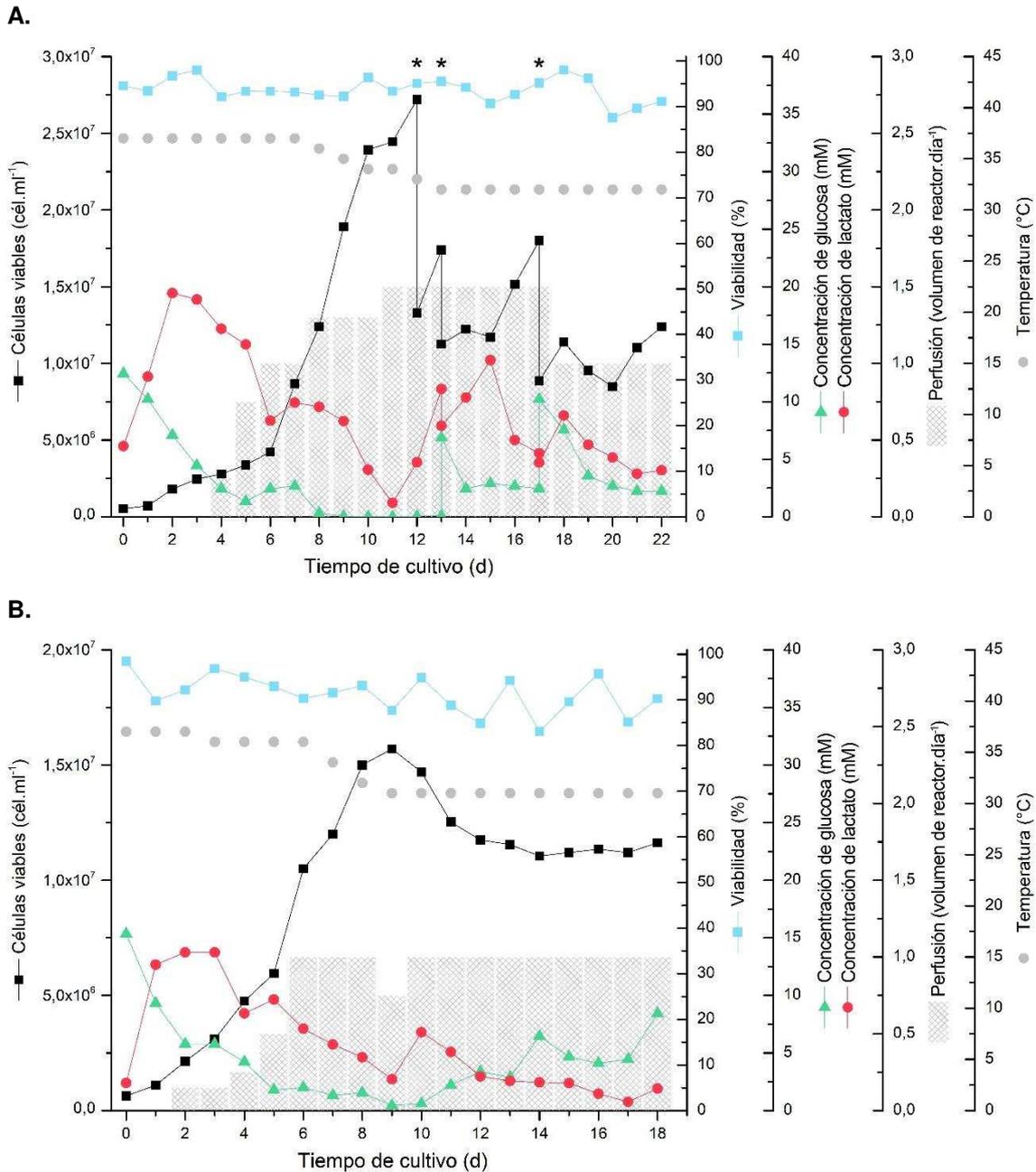


Figura 14. Cultivos de células CHO-K1 en un biorreactor con volumen de trabajo de un litro operado en modo de perfusión destinados al análisis del ciclo celular. Para ambos cultivos, se muestran las curvas correspondientes a la concentración de células viables, de glucosa y de lactato así como la variación de la viabilidad, de la velocidad de perfusión y de la temperatura en función del tiempo de cultivo. Los asteriscos indican las diluciones del cultivo llevadas a cabo con el fin de controlar las variables del proceso.

Debido a que el crecimiento acelerado de las células excedió la capacidad de control de las variables del proceso (fundamentalmente OD, disponibilidad de nutrientes y generación no deseada de espuma), se realizaron operaciones de eliminación de fracciones de suspensión celular (diluciones del cultivo) los días 12 ($1,5 \times 10^7$ cél./ml), 13 y 17 ($1,0 \times 10^7$ cél./ml).

De esta manera, es posible asumir que, a partir del día 18, se lograron estabilizar las variables operativas y el cultivo alcanzó la fase estacionaria. El día 22 se decidió finalizar el bioproceso.

En relación a la variación de la viabilidad celular, se observa un comportamiento estable a lo largo de todo el proceso, presentando un valor medio que supera el 90%. Respecto al consumo de glucosa, se aprecia una tendencia siempre decreciente hasta el día 13, que se corresponde con la elevada tasa de crecimiento celular alcanzada el día 12. Asimismo, a pesar de que –en general– el lactato tiende a acumularse debido a la oxidación poco eficiente de la glucosa, este mostró un decaimiento rápido. Además del efecto de dilución de este metabolito debido al modo de operación del biorreactor, este fenómeno también podría deberse a una re-captación del lactato para su utilización como fuente de energía por parte de las células.

Durante cinco días los valores de glucosa fueron prácticamente indetectables y esta situación se logró revertir cuando se diluyó la suspensión a la mitad, alcanzándose el máximo volumen de perfusión. A su vez, la disminución de la temperatura permitió reducir progresivamente la actividad metabólica para contribuir a la estabilización del proceso.

Considerando todo lo expuesto anteriormente, es posible concluir que el cultivo de células CHO-K1 en alta densidad se llevó a cabo satisfactoriamente. Los resultados generados a partir del primer cultivo de células CHO-K1 permitieron obtener información relevante acerca del comportamiento de dichas células en cultivos de alta densidad, logrando caracterizar el ciclo celular durante el proceso (Capítulo 1 - Sección 2. Caracterización del ciclo celular). En particular, se verificó la capacidad del equipo FACS ARIA III para efectuar los análisis de ciclo celular y, a su vez, también demostró ser apropiado el protocolo de coloración con PI.

Se decidió llevar a cabo un segundo cultivo (Figura 14.B) con el fin de lograr una mejor regulación de las variables operativas, para evitar que la proliferación celular exceda la capacidad de control del sistema. En este caso, el biorreactor fue inoculado con una mayor densidad celular ($6,0 \times 10^6$ cél./ml). La agitación osciló entre 120 y 150 rpm y los valores de pH se mantuvieron en el rango 6,60 – 7,20. El biorreactor comenzó a perfundirse en el día 2 y el caudal varió entre 0,15 y 1,00 l/día.

Como se observa en la Figura 14.B, el cultivo presentó una fase de latencia de muy corta duración, seguida de una fase exponencial. Asimismo, con el objetivo de evitar una proliferación celular excesiva, se aplicó un gradiente de temperatura decreciente, de 37°C a 31°C, en forma más temprana que en el cultivo anterior (desde el día 3) y tanto el abastecimiento de oxígeno (régimen de aireación) como de nutrientes

(volumen de perfusión) se realizó de manera de mantener dichas variables en los valores límites inferiores para disminuir la tasa de crecimiento celular.

El cultivo alcanzó rápidamente el estado estacionario, ya que se controlaron antes los parámetros temperatura y abastecimiento de oxígeno, reduciendo así el volumen de perfusión necesario, respecto al empleado en el primer cultivo. De esta manera se evitó realizar diluciones de la suspensión celular, estabilizándose el cultivo en torno a una densidad de $1,1 \times 10^7$ cél./ml. Durante todo el proceso, la viabilidad celular se mantuvo alrededor del 90%.

También se observa que los perfiles de glucosa y lactato fueron adecuados. En relación a la variación de la concentración de glucosa, se evidenció una disminución paulatina, con valores superiores al límite de detección. Por otro lado, los niveles de lactato se mantuvieron siempre bajos. Ambos metabolitos se controlaron apropiadamente modificando el régimen de perfusión.

El cultivo se interrumpió en el día 18, ya que se consiguió estabilizar el sistema en torno a una densidad celular deseada y de una manera relativamente rápida, gracias a la manipulación adecuada de variables operativas sin necesidad de diluir el cultivo. Se minimizó la fase de latencia y se alcanzó una densidad celular apropiada para el sistema de trabajo empleado. A su vez, se trabajó con concentraciones detectables de glucosa, niveles aceptables de lactato, alcanzando concentraciones constantes hacia el final del cultivo.

2. Caracterización del ciclo celular

Para los dos cultivos se tomó una muestra diaria del biorreactor y se siguió el protocolo de coloración con PI. A su vez, se efectuó la calibración del citómetro y la adquisición de los datos a través de la plataforma FACSDiva. Se adquirieron diariamente 30000 eventos, para mejorar la confiabilidad en los análisis posteriores.

Con respecto a la caracterización del ciclo celular, primeramente se realizó el control de calidad del citómetro en relación a parámetros específicos. Para este objetivo, se utilizó el *kit DNA QC Particles* (BD) según las instrucciones del fabricante.

En primer lugar, se verificó la linealidad; es decir, la capacidad del equipo de establecer una relación directamente proporcional entre el contenido de ADN de las muestras y la señal de fluorescencia generada por el colorante. Para ello, se aplicó la fórmula $\frac{\text{Media dobles}}{\text{Media singles}}$, obteniéndose un valor de 1,99 (comprendido dentro del rango estipulado 1,95 – 2,00). A su vez, se localizó en el histograma de PI el coeficiente de variación (CV) correspondiente a los singles, siendo este igual a 3% (valor aceptable de acuerdo con lo establecido en el protocolo).

Por otra parte, se analizaron la resolución y la eficiencia del citómetro respecto a los picos de fluorescencia en un histograma típico para la caracterización del ciclo celular. Asimismo, analizando el *dot-plot* PI-A vs. PI-W se evidenció que el equipo permite una correcta separación entre la proporción de G2/M (correspondiente a la población de singletes) y la proporción de dobles en G0/G1. Esto es de gran importancia ya que, al no distinguirse dicha separación, se estaría sobreestimando la población en G2/M, por presentar la misma intensidad de fluorescencia que la población de dobles en G0/G1.

La adquisición de datos y calibración de los voltajes de los detectores se realizó a través de la plataforma FACSDiva. El procesamiento de los resultados para estimar la distribución relativa de las células en las diferentes fases del ciclo celular (G0/G1, S y G2/M) se efectuó con FlowJo X y ModiFit LT.

A partir de los datos generados, se graficaron los diagramas de dispersión de puntos (*dot-plot*): SSC vs. FSC, para la selección de la población de interés (Figura 15.A); y PI-W vs. PI-A, para la identificación de la subpoblación de singletes celulares (Figura 15.B).

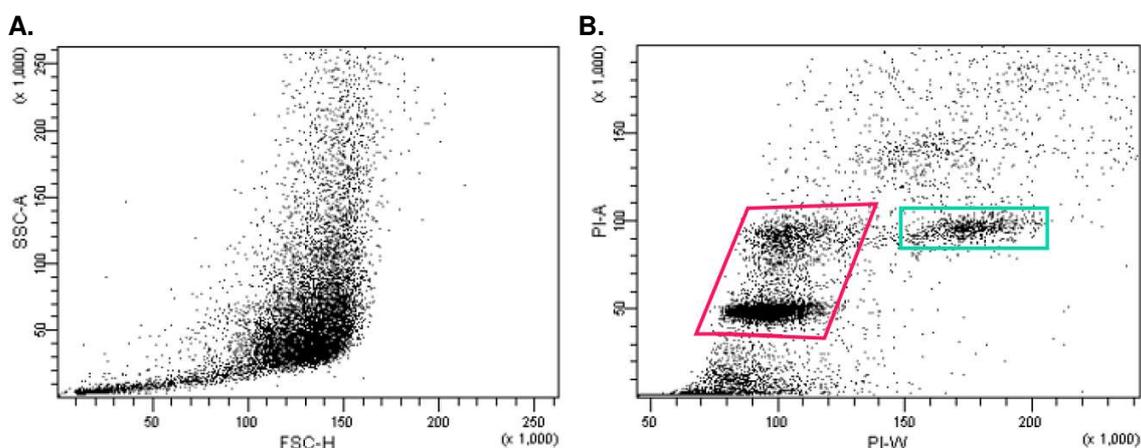


Figura 15. Análisis del ciclo celular por citometría de flujo con el equipo FACS ARIA III, empleando la plataforma FACSDiva. A modo de ejemplo se muestran los gráficos elaborados para el día 11 del primer cultivo.

A. Identificación de la población celular.

B. Discriminación entre singletes (rojo) y dobles (verde).

Para el primer cultivo, los archivos generados por FACSDiva se analizaron en forma automática con ModFit LT, con el objetivo de modelar el ciclo celular mediante la deconvolución de los histogramas PI-A. Sin embargo, esta estrategia no resultó satisfactoria, ya que la presencia de *debris* y agregados celulares interfiere en el modelado de las fases del ciclo, ocasionando pérdida de información. Por estos motivos, se decidió modificar la forma de analizar los datos generados tanto en el primer como segundo cultivo.

Los archivos crudos provenientes de la citometría se procesaron en FlowJo X, seleccionando la población de interés (singletes celulares) por inspección visual, debido a que este programa presenta una mayor resolución gráfica.

En la Figura 16 se muestra, a modo de ejemplo, el archivo generado para los datos del día de cultivo 6 del segundo cultivo. En este caso, se guardaron únicamente los eventos delimitados por el trapecio rojo. Seguidamente, con el programa ModFit LT se efectuó una detección automática del pico correspondiente a la población G0/G1, mientras que la media del pico de G2/M se ajustó por inspección visual, evitando de esta manera aplicar la auto-linealidad. Finalmente, la fase S fue modelada según un modelo rectangular de tres compartimentos [102,143].

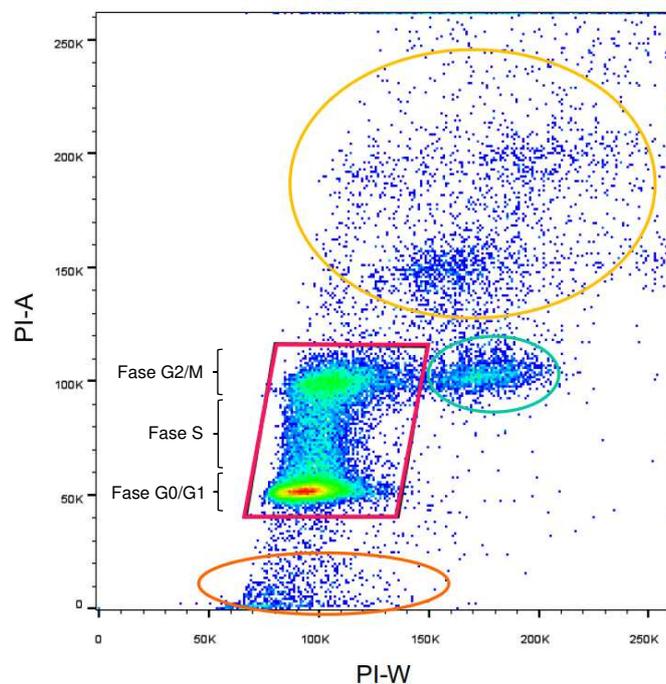


Figura 16. Análisis del ciclo celular: diagrama de dispersión de puntos PI-A vs. PI-W generado por FlowJo X para la discriminación de poblaciones celulares. Identificación de singletes (trapezoido rojo), dobletes (óvalo verde), *debris* (naranja) y agregados (amarillo). Dentro de la población de singletes se encuentran indicados, de manera aproximada, los eventos correspondientes a cada fase.

En la Figura 17 se muestran los histogramas obtenidos a partir de los datos de las muestras diarias del segundo cultivo, empleando ModFit. Se observan las poblaciones típicas que representan la distribución relativa de las células entre las diferentes etapas del ciclo celular para una misma muestra; visualizándose los picos de G0/G1 (en amarillo), S (en verde) y G2/M (en rojo).

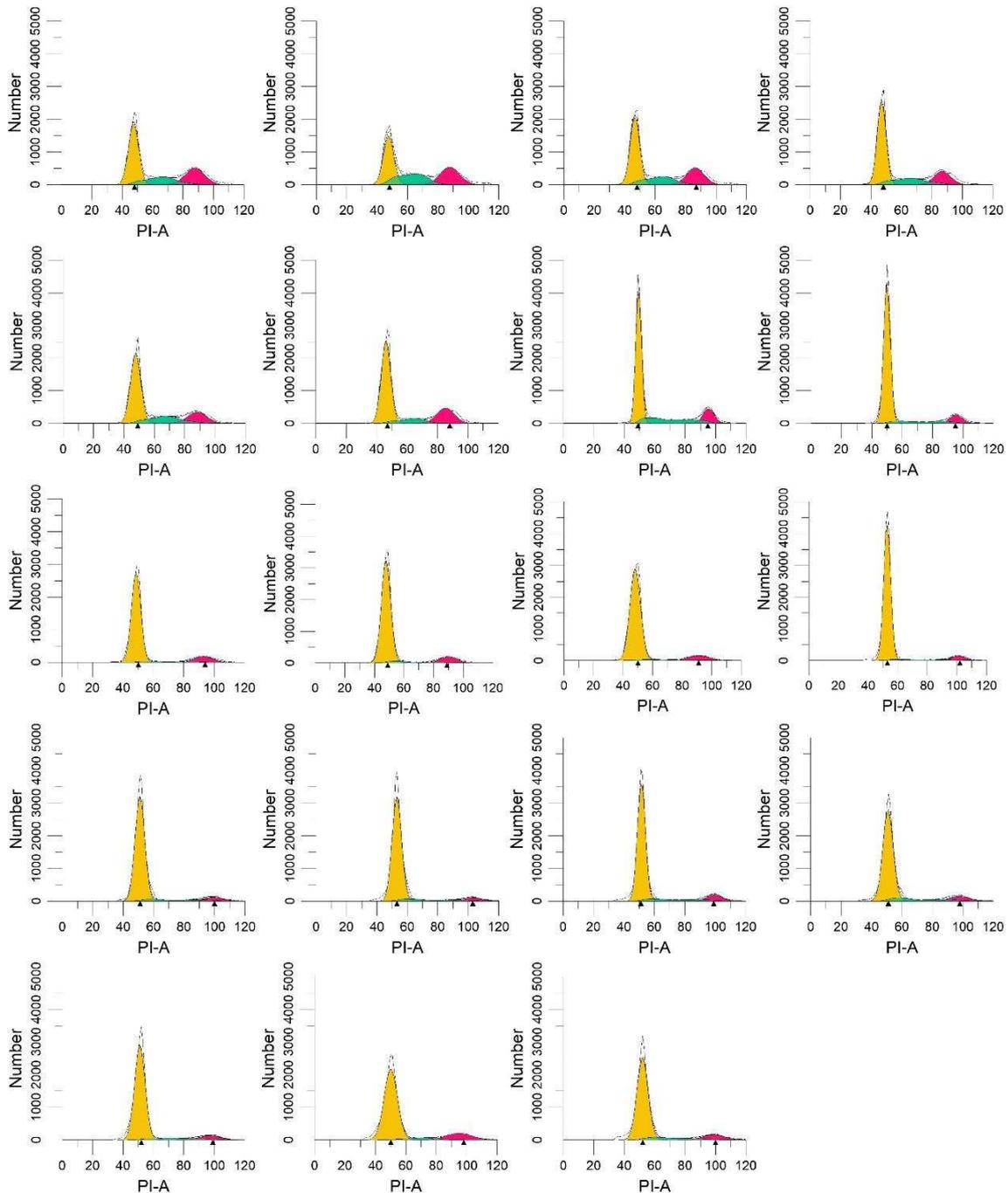


Figura 17. Análisis del ciclo celular: gráficos generados por ModFit LT para el modelado de la distribución relativa de fases del ciclo celular, para cada día del segundo cultivo. Se observan los histogramas típicos de ciclo celular, con las poblaciones correspondientes a las fases G0/G1 (amarillo), S (verde) y G2/M (rojo). El día 1 corresponde al histograma superior izquierdo, incrementándose los días de cultivo de forma sucesiva por fila.

Con los valores obtenidos (porcentaje de células en cada fase) se construyeron los gráficos de barra que muestran la distribución de las células en las fases del ciclo celular, por cada día de cultivo, junto con las curvas de crecimiento (Figura 18).

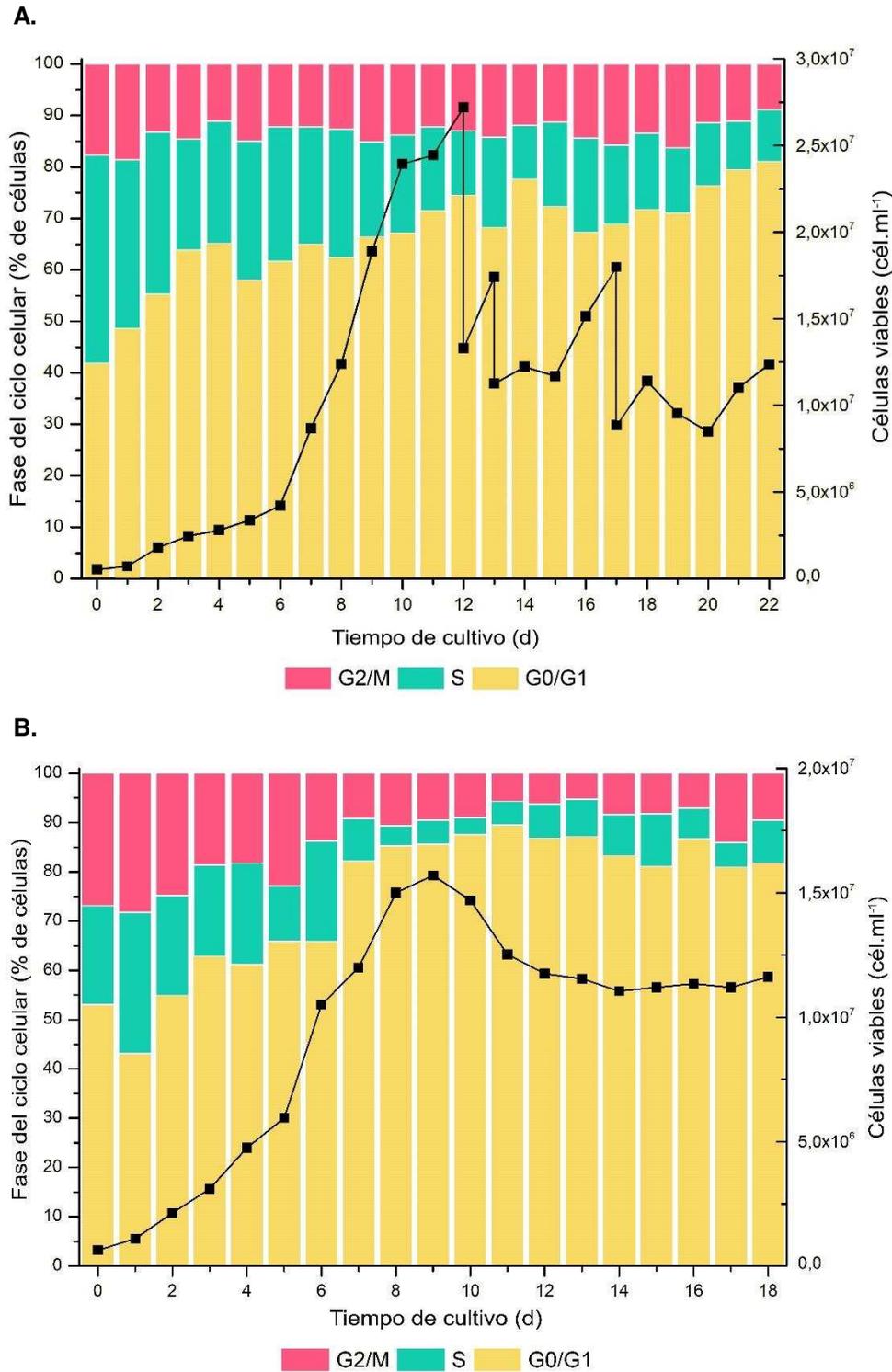


Figura 18. Análisis del ciclo celular: curvas de crecimiento y porcentajes diarios de células en cada una de las fases del ciclo celular, para el primer (A) y segundo (B) cultivo. Valores obtenidos con ModFit LT: G0/G1 (amarillo), S (verde) y G2/M (rojo).

Las características de ambos bioprocesos, descritas previamente en la Sección 1 de este Capítulo, se ven reflejadas en los análisis del ciclo celular. Durante el primer cultivo, el porcentaje máximo de células en la fase G0/G1 (81%) fue alcanzado

el último día del cultivo, cuando la temperatura había permanecido en 32°C desde los nueve días previos. Asimismo, durante el segundo cultivo se logró obtener antes una mayor proporción de células en G0/G1 (89%) para el día 11 del cultivo (al inicio de la fase estacionaria), luego de dos días de mantener la temperatura a 31°C.

Para el segundo cultivo, se observa que la proporción de células en fase S se reduce paulatinamente al progresar el cultivo, llegando a una cantidad mínima de 3% en el día 10; identificándose una clara tendencia de las células a estacionarse en las fases G0/G1, particularmente en la fase exponencial tardía y en el transcurso de la fase estacionaria. Más del 80% de las células permanecieron en G0/G1 durante toda la fase estacionaria. Esto puede deberse al marcado gradiente de temperatura que fue aplicado, comenzando en los primeros días del cultivo, para controlar adecuadamente las concentraciones de metabolitos (glucosa y lactato); en consecuencia se utilizó un menor volumen de perfusión y se evitó la separación de suspensión celular.

Además, puede distinguirse que, luego de cada una de las sucesivas diluciones celulares durante el primer cultivo (días 12, 13 y 17), los porcentajes de células en fase S se incrementan levemente, para luego disminuir nuevamente. Esto puede deberse a que el crecimiento celular se ve favorecido por el aumento de nutrientes disponibles luego de incorporarse medio no metabolizado. Por otro lado, el porcentaje de células en la fase G2/M se mantuvo relativamente constante durante el primer cultivo, mientras que en el segundo se observa una pequeña disminución de esta población a partir de la fase exponencial tardía.

Por otro lado, analizamos la distribución de la población celular en cada día del segundo cultivo con el objetivo de identificar cambios en el tamaño y la complejidad de las células. Como se mencionó en la Introducción de este Capítulo, las células arrestadas en G0/G1 son metabólicamente más activas y de mayor tamaño que aquellas no arrestadas. En la Figura 19 se muestran los gráficos *dot-plot* SSC vs. FSC, en los cuales se visualiza la población celular de singletes para cada día del cultivo. Con estos gráficos se obtiene información general respecto a las características intrínsecas de las células, en relación a los cambios de tamaño y complejidad de las mismas, conforme pasan los días de cultivo en alta densidad.

Puede verse claramente que, a medida que el cultivo progresa, la forma que adopta la población celular, en los gráficos de SSC vs. FSC, varía en cuanto a su localización y dispersión. En particular, se observa un cambio muy marcado a partir del día 8, que corresponde al inicio de la fase estacionaria. La población se va desplazando hacia la derecha del eje FSC, lo que indica un incremento en el tamaño de las células, y hacia el extremo superior del eje SSC, evidenciando también una mayor granulosidad

o complejidad celular, lo cual sugiere una composición celular más compleja, dada por un aumento de los componentes celulares, como proteínas y otras biomoléculas.

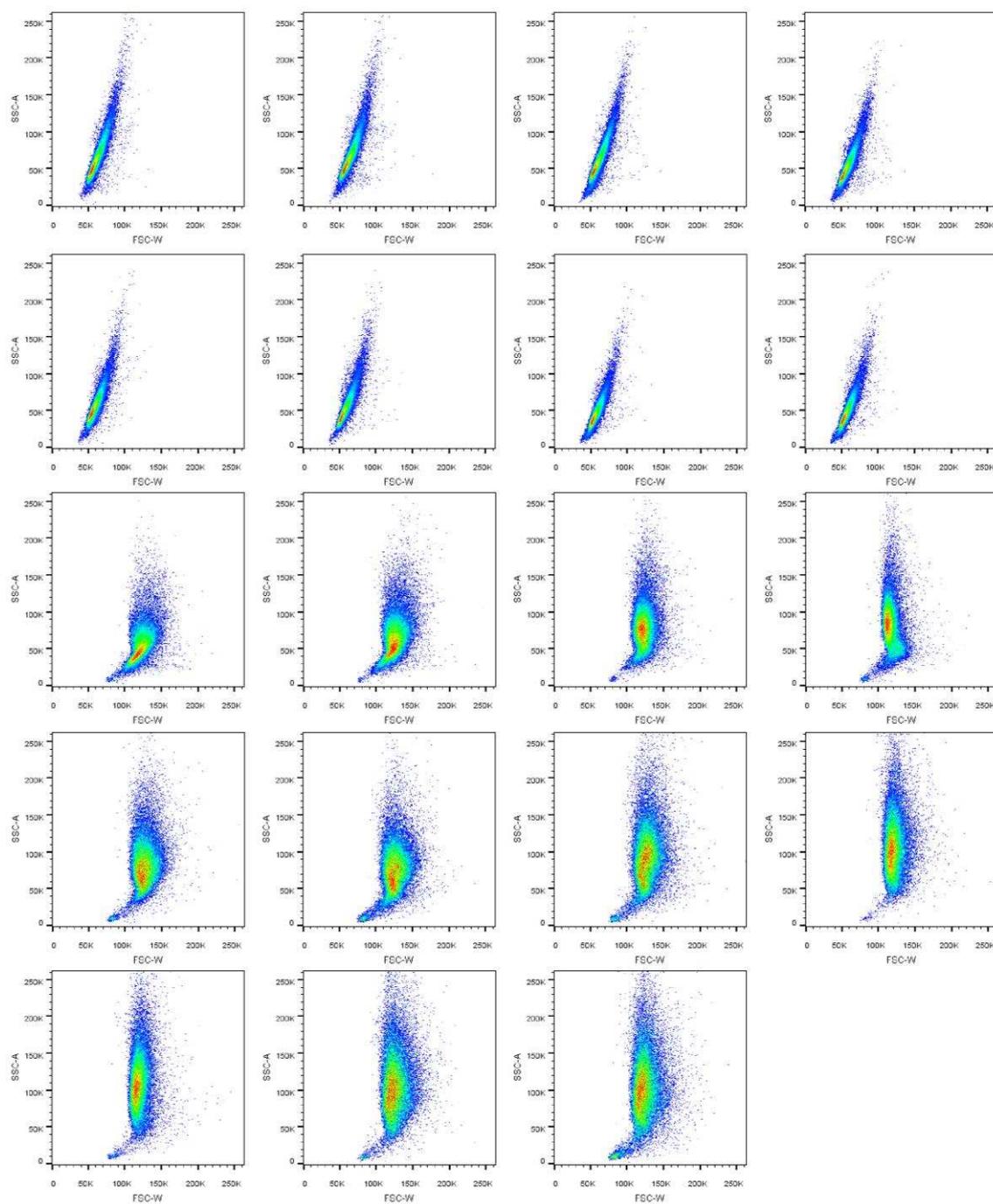


Figura 19. Análisis del ciclo celular: gráficos de tipo *dot-plot* SSC vs. FSC, en los cuales se visualiza la población celular de singletes para cada día del segundo cultivo. El día 1 corresponde al gráfico superior izquierdo, incrementándose los días de cultivo de forma sucesiva por fila.

En síntesis, en ambas experiencias fue posible evidenciar un enriquecimiento del cultivo en células sincronizadas en fase G₀/G₁, en la etapa de crecimiento exponencial tardío y hacia los últimos días de la fase estacionaria. Estos análisis arrojan una

tendencia evidente: cuanto más temprano se aplica el gradiente de temperatura, más rápidamente las células son arrestadas en la fase G0/G1, generando un mayor control del bioproceso, lo cual deriva en un menor tiempo de cultivo, una disminución de la proliferación celular y una mayor viabilidad celular. En conjunto, estos factores pueden generar mayores tasas específicas de producción de proteínas recombinantes durante un proceso productivo.

Teniendo en cuenta que el objetivo siguiente de este trabajo es secuenciar el transcriptoma de células CHO-K1 cultivadas en alta densidad y sincronizadas en fase G0/G1, ya que como se mencionó, es considerada la mejor etapa para la producción de proteínas recombinantes; y en virtud de los resultados obtenidos de la caracterización del ciclo celular, es posible evidenciar que los cultivos de CHO-K1 tienden a estacionarse espontáneamente en G0/G1, con un alto porcentaje de células en dicha fase. Ciertas estrategias de control del proceso, como el descenso gradual de la temperatura, favorecerían esta situación [108].

Por estas razones, se decidió realizar las extracciones de ARN total de las células muestreadas directamente a partir del biorreactor; es decir, sin recurrir a ningún método de sincronización (métodos físicos, químicos o separación por *Fluorescence Activated Cell Sorting, FACS*), preservando totalmente la integridad celular y, por lo tanto, la del transcriptoma en las condiciones propias de un bioproceso a escala industrial.



CAPÍTULO 2

*Análisis del transcriptoma de
células CHO-K1 cultivadas en
alta densidad, mediante RNA-seq*

CAPÍTULO 2

Análisis del transcriptoma de células CHO-K1 cultivadas en alta densidad, mediante RNA-seq

INTRODUCCIÓN

“La esperanza reside en los sueños, en la imaginación y en el coraje de aquellos que se atreven a convertir sus sueños en realidad”.

Jonas Salk

1. Secuenciación de ADN mediante tecnologías de secuenciación masiva de segunda generación (NGS)

1.1. Tecnología *Illumina*

En principio, el concepto detrás de esta tecnología de NGS es similar a la secuenciación por electroforesis capilar. La ADN polimerasa cataliza la incorporación de desoxirribonucleótidos trifosfatos (dNTPs) marcados fluorescentemente en una cadena molde durante los ciclos sucesivos de síntesis de ADN. Con esta metodología, dichos nucleótidos bloquean de forma reversible la elongación de la cadena. Así, luego de la detección del nucleótido marcado, y la eliminación del mismo, es posible continuar con un nuevo ciclo de adición de otro nucleótido.

La diferencia fundamental con la tecnología de primera generación es que *Illumina* extiende este proceso a través de millones de fragmentos en forma paralela y masiva. Las secuencias cortas producidas por el instrumento se denominan lecturas o *reads*.

Más del 90% de los datos de secuenciación (NGS) del mundo son generados mediante la secuenciación por síntesis química de *Illumina* [147]. Esta empresa ofrece equipos y productos que tienen una precisión muy alta, un gran rendimiento de lecturas sin errores y un elevado porcentaje de bases con calidad por encima de Q30 [148,149].

El flujo de trabajo (*workflow*) es bastante sencillo, y comprende cuatro etapas principales [147], detalladas a continuación (Figura 20).

1.1.1. Preparación de la biblioteca

Comienza con la fragmentación aleatoria de la muestra de ADN o ADNc, luego se ligan los adaptadores en los extremos 5' y 3' de los fragmentos generados (Figura 20.A).

De manera alternativa, el método “*tagmentation*” (*Illumina*) realiza en un único paso tanto la fragmentación como las reacciones de ligación de los adaptadores, lo cual incrementa la eficiencia del proceso de preparación de la biblioteca. Por último, los fragmentos se amplifican mediante PCR.

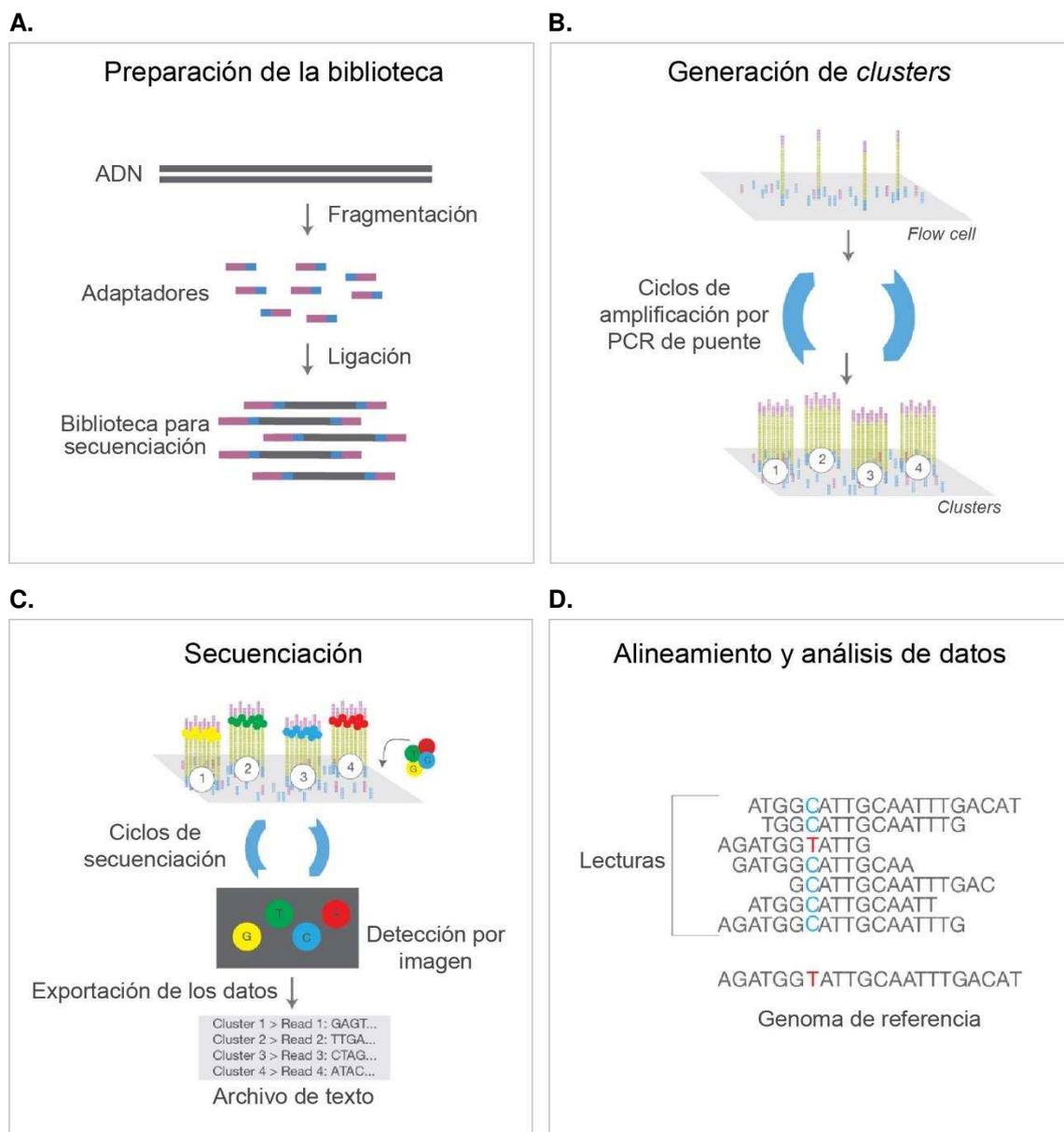


Figura 20. Representación esquemática de las cuatro etapas principales en el proceso de secuenciación mediante la tecnología *Illumina*.

- A.** Preparación de la biblioteca: se fragmenta el ADN (o ADNc) y se ligan los adaptadores en ambos extremos.
- B.** Amplificación mediante la generación de *clusters*: los fragmentos de la biblioteca se unen a la placa de secuenciación por complementariedad de los adaptadores y cada fragmento se amplifica mediante “PCR de puente”, formando *clusters* de una misma secuencia.
- C.** Secuenciación: se añaden todos los reactivos necesarios para la secuenciación, incluidos dNTPs marcados con fluorescencia. Se incorpora la primera base, se toma una imagen de la placa y así se detecta la fluorescencia emitida por cada *cluster*. Este ciclo se repite “n” veces, generando lecturas de “n” bases.
- D.** Alineamiento y análisis de los datos: las secuencias generadas se alinean contra un genoma de referencia empleando *software* bioinformático específico. Luego se pueden llevar a cabo diversos análisis.

Figura modificada de *Illumina* Inc. [147].

1.1.2. Generación de clusters

La biblioteca se coloca en una placa de secuenciación (*flow-cell*), compuesta generalmente por ocho carriles (*lanes*), donde los fragmentos desnaturalizados se unen por complementariedad de sus adaptadores a los oligonucleótidos que se encuentran inmovilizados en la superficie de la *flow-cell*. Así, cada fragmento de ADN monocatenario se unirá por uno de sus extremos a uno de los oligonucleótidos complementarios presentes en la placa.

Tras la acción de la polimerasa –que utiliza estos adaptadores como cebador (*primer*)– se sintetiza la cadena complementaria, quedando inmovilizada. Luego, tendrá lugar un nuevo ciclo de desnaturalización, provocando que las hebras de ADN formen puentes gracias a la unión de su extremo libre con otro de los adaptadores en la placa. Este proceso se repite varias veces, hasta formar lo que se conoce como *clusters*, una agrupación de secuencias idénticas inmovilizadas sobre una superficie sólida (Figura 20.B). Este mecanismo es propio de *Illumina* y se denomina “PCR de puente”.

El inconveniente de esta etapa es la posible incorporación de errores debidos a fallas en la polimerasa durante el proceso de síntesis de nuevas copias por PCR.

1.1.3. Secuenciación

La tecnología *Illumina* utiliza el método propio de “terminador reversible” (*cyclic reversible termination, CRT*) en el cual se utilizan cuatro nucleótidos con terminadores reversibles que están presentes simultáneamente en cada ciclo (Figura 20.C).

Los cuatro terminadores reversibles bloquean la polimerización, por lo que la enzima polimerasa sólo puede añadir una base por vez a cada cadena de ADN que se está sintetizando. Al añadirse los nucleótidos en forma simultánea y no secuencialmente, se produce una competencia natural, por lo que se minimizan los errores de incorporación. Como las bases individuales se unen a todos los fragmentos de ADN de manera uniforme, de a un ciclo por vez, el proceso de secuenciación produce secuencias (lecturas) de igual longitud.

Después de cada ciclo de incorporación se eliminan los terminadores y se determina la identidad de la base incorporada mediante una “fotografía” de la placa (proceso denominado en inglés *imaging*), ya que cada dNTP está marcado con un fluoróforo diferente. Este método de secuenciación, propio de *Illumina*, se conoce como secuenciación por síntesis (*sequencing by synthesis, SBS*).

1.1.4. Análisis de datos

Luego de la asignación de bases y de los valores de calidad correspondientes a cada una, se producen millones de secuencias que son almacenadas en archivos de

texto, comúnmente archivos en formato “fastq” (Figura 20.C). Con estos se procede primero realizando un control de calidad de las secuencias generadas, para luego alinearlas contra un genoma de referencia (Figura 20.D).

Posteriormente, se pueden realizar diversos análisis como, p.ej., tratar de identificar polimorfismos, inserciones/eliminaciones, realizar un recuento de lecturas para RNA-seq, etc.

1.2. Orientación de las lecturas

La secuenciación puede ser de tipo “*single-end*” (SE), en la cual la secuencia se lee a partir de un solo extremo del fragmento de ADN, siendo un proceso más rápido, económico y sencillo.

En general, los nuevos secuenciadores generan lecturas a partir de cada uno de los extremos de un fragmento de ADN [150], dando lugar a lecturas apareadas, y lo hacen usando dos estrategias diferentes (Figura 21):

- *Mate-pairs* (MP): se crean a partir de fragmentos de ADN de tamaño conocido, generando bibliotecas con tamaños superiores a los 600 pb, llegando a varios kb de longitud. Estos fragmentos son biotinilados en sus extremos, se circularizan y se ligan empleando un adaptador interno o *linker*. Luego, se fragmentan al azar, se purifican aquellos fragmentos que contienen el adaptador y se secuencian (Figura 21.A).
- *Paired-end* (PE): se generan lecturas a partir de bibliotecas de un tamaño aproximado de 300 pb o menos. Se secuencian ambos extremos de los fragmentos; se comienza como una secuenciación típica SE, y luego de producida la lectura en dicha dirección, inicia otro ciclo desde el extremo opuesto (Figura 21.B).

La secuenciación de tipo PE mejora la capacidad de identificar las posiciones relativas que las lecturas tienen en el genoma, optimizando el alineamiento de las secuencias. De esta manera, este tipo de secuenciación es mucho más efectiva que la secuenciación de tipo SE, especialmente para resolver diversos reordenamientos estructurales, como inserciones, eliminaciones o inversiones de genes. Además, mejora el ensamblado de regiones repetitivas y permite eliminar los duplicados de PCR que se originan durante la preparación de la biblioteca [147]. Sin embargo, este elevado nivel de precisión no es necesario para todos los experimentos, por ejemplo para cuantificar los niveles de expresión génica por RNA-seq.

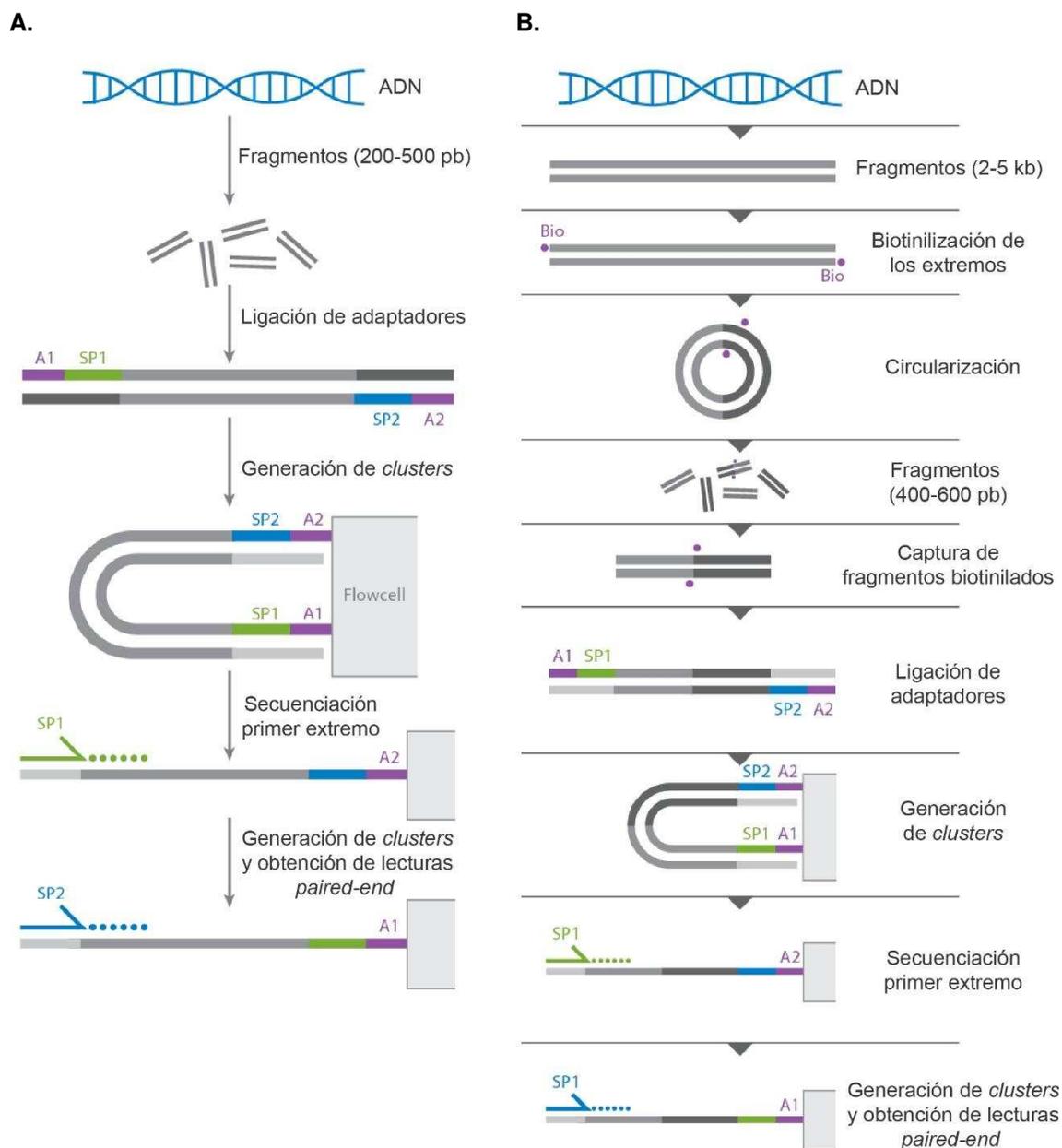


Figura 21. Esquema del proceso de obtención de lecturas pareadas mediante la tecnología *Illumina*.

A. Secuenciación de tipo *paired-end*.

B. Secuenciación de tipo *mate-pair*. SP1 y SP2 indican los *primers* de secuenciación; A1 y A2 hacen referencia a los adaptadores unidos a la superficie de la placa de secuenciación.

Figura modificada de Mardis, 2013 [151].

1.3. Longitud de las lecturas

El concepto de longitud de lectura hace referencia al número de pares de bases que se leen en una ronda de secuenciación. Por ejemplo, una lectura puede consistir en 50 pb, 100 pb o más. Para una longitud de secuencia de 50 pb, las lecturas de un solo extremo (SE) tienen 50 pb de cada fragmento, mientras que las secuencias pareadas (PE) consistirán en dos lecturas de 50 pb, cubriendo hasta 100 pb en el mismo

fragmento. Las secuencias de mayor tamaño brindan una información más confiable sobre las ubicaciones de las lecturas, particularmente si corresponden a regiones repetitivas. No obstante, suele ser más costoso generar lecturas más largas.

Como se mencionó en la Sección 6 de la Introducción General existe una segmentación entre las plataformas de secuenciación masiva en función de la longitud de las lecturas que generan; por ello se hace referencia a plataformas de lecturas largas (>300 pb) y de lecturas cortas (<300 pb) [67]. Para el caso de *Illumina*, los secuenciadores pueden generar tanto lecturas SE como PE que van desde los 50 pb hasta los 300 pb.

La longitud de lectura es una de las principales variables y limitaciones que se deben tener en cuenta a la hora de diseñar los experimentos. Lecturas más largas generalmente no son requeridas o recomendadas para casos donde la secuencia específica no es importante (por ejemplo, experimentos de RNA-seq, para cuantificación de expresión), ya que aumenta innecesariamente el costo y el tiempo de la secuenciación.

1.4. Profundidad de cobertura (*depth of coverage*)

Un aspecto importante en un proceso de secuenciación es el número de veces que cada nucleótido del genoma es secuenciado en una corrida. Este valor se denomina profundidad de cobertura (*depth of coverage*, o simplemente *coverage*) y es uno de los factores determinantes para evaluar la confiabilidad del nucleótido asignado a esa posición del genoma. Se calcula de la siguiente manera:

$$C = Y / G; \text{ donde: } Y = N \times L \quad (3)$$

siendo:

C= cobertura; Y= rendimiento secuenciación (*yield*); G= tamaño del genoma; N= número de lecturas; L= longitud de las lecturas.

Sin embargo, el valor de cobertura estimado es un valor promedio, ya que no todas las regiones se van a secuenciar el mismo número de veces; así, algunos fragmentos se van a leer 100 veces o más, mientras que otros sólo se secuenciarán una o dos veces, o ninguna. A mayor número de cobertura se obtienen datos de mejor calidad y confiables.

Para aplicaciones como RNA-seq, generalmente se recomienda un mínimo de 20 millones de lecturas por muestra. Para el caso de estudios que necesitan de una mayor precisión, como estudios de *splicing* alternativo, la obtención de 40-60 millones de lecturas pareadas por muestra generará mejores resultados [152].

1.5. Multiplexing

Al mismo tiempo que ha aumentado progresivamente la cantidad de datos producidos en una corrida de secuenciación, también se ha incrementado el número de muestras que se pueden secuenciar en una misma placa.

El *multiplexing* (Figura 22) permite mezclar gran cantidad de bibliotecas (*pool*) y secuenciarlas simultáneamente durante una misma corrida. En la preparación de las bibliotecas, se añaden secuencias únicas, llamadas índice (*index*), a cada fragmento de ADN, por cada muestra; que permiten luego identificar, en la etapa de procesamiento de los datos, a qué biblioteca pertenece cada secuencia, proceso denominado *demultiplexing* [147].

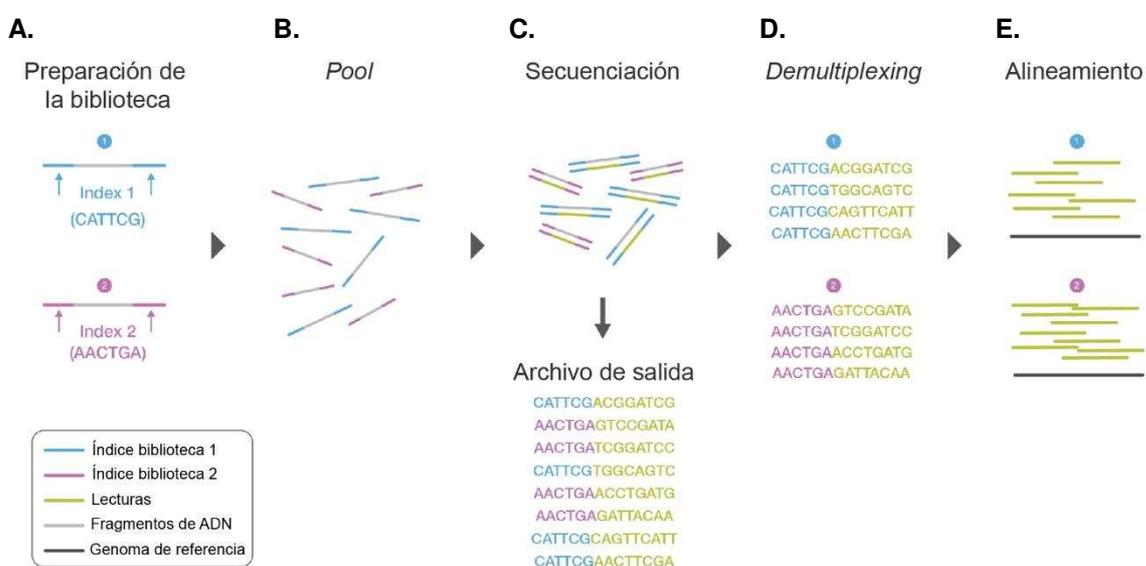


Figura 22. Proceso de preparación de bibliotecas con diferentes índices (*multiplexing*).

- Se añaden dos secuencias (índices) diferentes a los fragmentos de ADN provenientes de dos muestras, durante la preparación de las bibliotecas.
 - Se realiza el *pool* de las bibliotecas y se ponen en la misma *flow-cell*.
 - Se secuencian dichas bibliotecas durante la misma corrida de secuenciación. Todas las lecturas generadas se exportan a un único archivo de salida.
 - Se realiza el *demultiplexing* de las secuencias, agrupando las lecturas acorde a sus índices, creando así dos archivos diferentes.
 - Cada grupo de secuencias se alinea contra el genoma de referencia apropiado.
- Figura modificada de *Illumina Inc.* [147].

1.6. Valor de calidad de las bases asignadas a cada secuencia

Todos los sistemas de secuenciación estiman la probabilidad de que cada uno de los nucleótidos secuenciados sea erróneo; a este parámetro se lo conoce generalmente como calidad. Esta estimación del error es específica de cada tecnología y la calcula el *software* de procesamiento de secuencias del equipo.

Para facilitar el análisis y la interpretación de los resultados, estos valores suelen convertirse a una escala normalizada que utilizan todas las tecnologías de secuenciación, la escala de *Phred*. La misma se define como:

$$Q = -10 \log_{10} p \quad (4)$$

siendo p la probabilidad de error de que ese nucleótido fue asignado incorrectamente. Por ejemplo, si el valor de calidad Q es igual a 30, las chances de que esa base haya sido asignada incorrectamente son de 1 en 1000 (Tabla 5).

En el análisis de secuencias producidas por NGS, el formato estándar para los archivos que almacenan las millones de lecturas generadas, es el formato “fastq”. Este es un formato de texto, una simple extensión del formato “fasta”, usado para almacenar tanto la secuencia de nucleótidos como sus correspondientes valores de calidad, codificados mediante un único carácter del código ASCII (haciendo que los archivos sean concisos y compactos).

Tabla 5. Interpretación de los valores de calidad Q .

| Valor de calidad <i>Phred</i> | Probabilidad de error | Precisión |
|-------------------------------|-----------------------|-----------|
| 10 | 1 en 10 | 90% |
| 20 | 1 en 100 | 99% |
| 30 | 1 en 1000 | 99,9% |
| 40 | 1 en 10000 | 99,99% |
| 50 | 1 en 100000 | 99,999% |
| 60 | 1 en 1000000 | 99,9999% |

En dicho formato, cada secuencia se representa por cuatro líneas:

1. Identificador de la secuencia (comienza con el carácter @ y contiene información del secuenciador, las coordenadas del *cluster* en la *flow*-cell en el cual se generó la secuencia, la secuencia del adaptador, del índice, etc.).
2. Secuencia.
3. + (es un espaciador, puede seguirle una descripción opcional).
4. Valores de calidad alfanuméricos (valor de calidad codificado)

Luego, con estos archivos en formato “fastq” se procede a realizar un control de calidad y una revisión de las estadísticas de las lecturas producidas en la secuenciación.

En función de ello, se lleva a cabo lo que se conoce como el pre-procesamiento de las secuencias, en el cual se utilizan herramientas informáticas específicas para filtrar aquellas lecturas de baja calidad, remover los adaptadores que se hayan secuenciado, eliminar los extremos de baja calidad en las secuencias, entre otras.

Con el nuevo *set* de datos generados se llevan a cabo los análisis bioinformáticos específicos, según las preguntas biológicas que se quieran responder.

2. Transcriptómica

El transcriptoma es el conjunto completo de transcritos en una célula, y la proporción de los mismos, durante un momento específico o condición dada. Entender el transcriptoma es esencial para interpretar los elementos funcionales que componen el genoma y para dilucidar los constituyentes moleculares de células y tejidos, y además para comprender los diversos estadios de desarrollo y enfermedad [153].

Los principales objetivos de la transcriptómica son: catalogar las especies de transcritos presentes en la muestra, incluyendo ARNm, ARN no codificante y pequeños ARNs; determinar la estructura transcripcional de los genes, en cuanto a sus sitios de inicio de la transcripción y extremos 5' y 3', patrones de *splicing* y otras modificaciones postranscripcionales; y cuantificar los cambios en los niveles de expresión de cada transcritos durante las diferentes condiciones estudiadas [153].

Diversas tecnologías se han desarrollado en las últimas décadas para caracterizar y cuantificar el transcriptoma. Entre ellas se encuentran los microarreglos de ADN (*DNA microarrays*, también llamados *DNA chip* o *gene chip*), los cuales consisten en múltiples fragmentos de ADNc (cada uno de los cuales representa un gen diferente) adheridos a un soporte sólido en una disposición regular y prefijada [154]. A grandes rasgos, el procedimiento consiste en aislar el ARNm de las muestras y obtener su correspondiente ADNc, que luego se marca con un compuesto fluorescente. A continuación, se incuban las muestras de ADNc en el microarreglo, de manera tal que hibriden con los fragmentos de ADN inmovilizados en el *chip*. Luego, se realiza la excitación y la detección de la emisión de fluorescencia; y básicamente cuanto mayor sea la intensidad, mayor será la expresión de un gen dado.

En contraste con los microarreglos, los métodos basados en secuenciación directamente determinan la secuencia del ADNc. Inicialmente, se secuenciaban bibliotecas de EST (*expressed sequence tags*, marcador de secuencia expresada), luego surgieron otros métodos como el “análisis en serie de la expresión génica” (*serial analysis of gene expression*, SAGE); sin embargo presentan diversas limitaciones, principalmente por realizar la secuenciación empleando el método de Sanger [153].

Recientemente, se desarrolló el método denominado RNA-seq, el cual presenta notables ventajas sobre las tecnologías existentes y ha revolucionado la manera en que los transcriptomas son analizados (especialmente para los eucariotas), ya que se producen millones de secuencias provenientes de las muestras de ARN mediante NGS.

Empleando esta técnica, en contraste con los microarreglos, se puede no sólo cuantificar la expresión génica, sino también descubrir y anotar nuevos transcritos, detectar fusiones génicas y pequeñas inserciones y eliminaciones, estudiar las diversas especies de transcritos, determinar la estructura transcripcional de los genes, identificar eventos de *splicing* alternativo y patrones de poliadenilación, y detectar aquellos genes con muy baja expresión. Además, se puede realizar la secuenciación de muestras que no dispongan de un genoma de referencia, pudiendo cuantificarse la expresión de las mismas aún cuando no existan anotaciones disponibles, mediante el ensamblado *de novo* de las lecturas producidas.

2.1. Protocolo general para la secuenciación del transcriptoma mediante RNA-seq

A continuación se detalla un esquema general de trabajo para la secuenciación de muestras de ARN mediante la técnica RNA-seq (Figura 23).

2.1.1. Diseño del experimento

Se debe planificar el experimento de acuerdo con las preguntas biológicas que se quieran responder. En particular, es crítico identificar el objetivo principal del ensayo, que puede ser obtener información de expresión diferencial entre dos o más grupos o tratamientos, ensamblar el transcriptoma *de novo*, encontrar nuevas isoformas, etc.

También se debe considerar la cantidad de réplicas biológicas y/o técnicas que se llevarán a cabo, el valor de cobertura deseado de acuerdo con la aplicación específica (por lo tanto, la profundidad de secuenciación requerida), la orientación de las lecturas, entre otros. En esta etapa es importante que participen tanto el personal que realizará el trabajo de laboratorio, como los bioinformáticos que analizarán los datos.

2.1.2. Preparación del ARN

Se procede a purificar el ARN total de las muestras a ser analizadas. El éxito de los experimentos de RNA-Seq es altamente dependiente de la extracción del ARN, si se recupera puro e intacto. El valor RIN (*RNA integrity number*) –arrojado por el equipo Agilent Bioanalyzer (Agilent Technologies)– brinda información sobre la integridad del ARN purificado. Debido a que el ARN es más inestable que el ADN y las ARNasas son

enzimas muy estables, se debe tener un cuidado adicional al purificar y trabajar con las muestras.

Luego, se realiza un enriquecimiento de la población de interés a partir de las muestras de ARN total, generalmente en ARNm. Más del 90% del ARN total es ARN ribosomal (ARNr), así como cantidades considerables de ARN de transferencia (ARNt), que normalmente no son motivo de análisis [155].

Se puede capturar directamente el ARNm o bien eliminar el ARNr. La primera estrategia se basa en que la mayoría de los ARNm maduros, transcritos por la ARN polimerasa II, tienen una cola de poli(A) en su extremo 3'; por lo cual el ARNm se captura mediante hibridación a unas perlas magnéticas (*magnetic beads*) compuestas por oligo(dT), dejando de lado también a las moléculas de ARNm inmaduras [155]. Para el segundo caso, el ARNr se elimina a través de hibridación a unas perlas que tienen secuencias complementarias a los ARNr [155]. Seguidamente, se fragmentan las muestras.

2.1.3. Construcción de las bibliotecas

En esta etapa se obtiene el ADNc a partir de las muestras de ARN y se ligan los adaptadores (junto a los índices en el caso de realizar *multiplexing*). Existen diversos protocolos para la construcción de las bibliotecas, en función de la especie de ARN que se quiere secuenciar y si se desea obtener información específica de cada hebra (*strand-specific library*) [156-158]. Luego, es necesario cuantificar las bibliotecas y validar la calidad de las mismas. Es crítico que la distribución de los tamaños de los fragmentos sea unimodal; es decir, que se observe un único pico en el resultado arrojado, por ejemplo por el equipo Agilent Bioanalyzer [155].

2.1.4. Secuenciación

Actualmente, la tecnología dominante para esta aplicación es *Illumina*, ya que brinda una elevada profundidad de secuenciación, lecturas de tamaño adecuado y con un bajo porcentaje de errores.

2.1.5. Análisis de los datos

En general, se realiza el control de calidad y el pre-procesamiento de las secuencias, las lecturas se ensamblan en transcritos o se alinean contra un genoma de referencia, se anotan los posibles transcritos, se cuantifica el número de lecturas por cada uno de ellos, se estima la abundancia de los transcritos, se normalizan los valores (esto asegura que las medidas de expresión puedan compararse entre transcritos de una misma muestra, así como también entre muestras) y se realiza un

análisis estadístico para identificar la expresión diferencial de los mismos en las muestras secuenciadas [159].

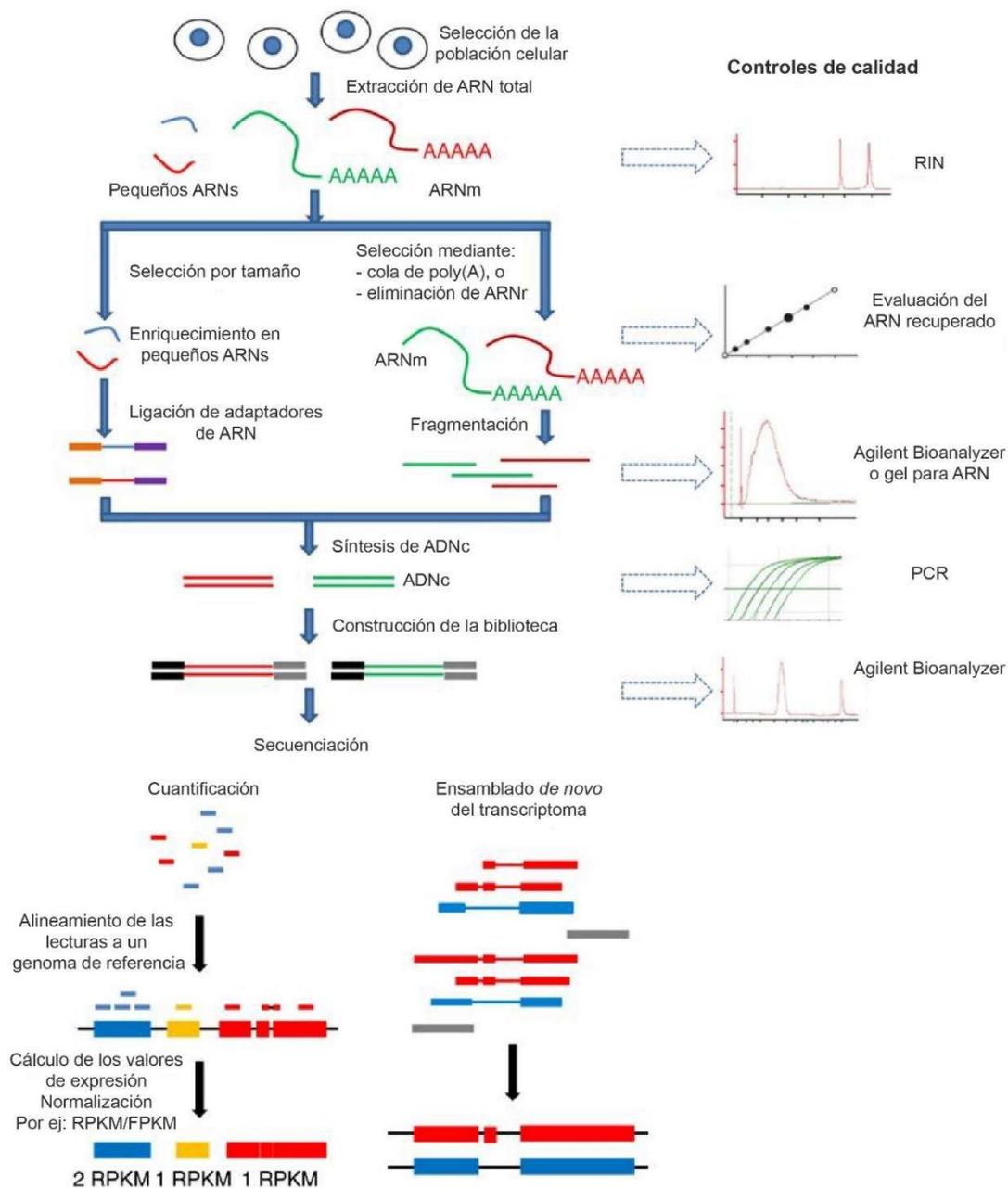


Figura 23. Esquema general de los pasos a seguir al realizar RNA-seq y controles de calidad asociados a etapas específicas.
Figura modificada de Zeng, 2012 [155].

El procesamiento de datos obtenidos por RNA-seq es computacionalmente más complejo que para otros tipos de análisis de datos de NGS, ya que para genomas eucariotas, una gran proporción de lecturas alinean más allá de los sitios de *splicing* (lo que se conoce como *splice junctions*) y dicha proporción aumenta a medida que se incrementa la longitud de lectura usada [155]. Esto, a su vez, vuelve más complejos los

análisis posteriores, como la cuantificación de las diversas isoformas. Existen varios programas que tienen en cuenta los sitios de *splicing* al momento de alinear las lecturas contra el genoma de referencia, los cuales implementan diferentes estrategias computacionales [160]. Adicionalmente, con subgrupos de genes expresados diferencialmente de manera significativa, se pueden realizar análisis de vías metabólicas (*pathways*) y enriquecimiento de términos de ontología genética (lo que se conoce como *GO enrichment analysis*); con el objetivo de atribuirle un significado biológico a los hallazgos por RNA-seq.

OBJETIVOS

GENERAL

Se plantea como objetivo general para esta segunda etapa del proyecto, secuenciar el transcriptoma de células CHO-K1 –cultivadas en alta densidad, en modo de perfusión y con un medio de cultivo libre de suero fetal bovino– con el propósito de comparar los perfiles de expresión génica entre las diferentes etapas del cultivo (exponencial y estacionaria).

Los resultados obtenidos en la secuenciación del transcriptoma permiten conocer en forma global el comportamiento de la línea celular CHO-K1 en condiciones experimentales análogas a las de un bioproceso típico a escala industrial, diseñar estrategias de ingeniería genética e indagar características que permitan mejorar la producción de proteínas recombinantes.

ESPECÍFICOS

1. Secuenciar el transcriptoma de células CHO-K1 en condiciones de cultivo en alta densidad celular y posteriormente, procesar y reconstruir los datos generados mediante diversas estrategias bioinformáticas.
2. Llevar a cabo un análisis de expresión diferencial entre las etapas de cultivos en biorreactor, respecto al inicio del cultivo.
3. Explorar los cambios globales en los perfiles de expresión génica de cultivos bifásicos, luego de una reducción de la temperatura del cultivo desde 37°C a 31°C.
4. Caracterizar los procesos biológicos y las vías metabólicas involucradas en dichos cambios.
5. Correlacionar los resultados obtenidos con los análisis de ciclo celular llevados a cabo previamente.

MATERIALES Y MÉTODOS

1. Técnicas de biología molecular

1.1. Purificación de ARN total a partir de cultivos de células CHO-K1 en suspensión empleando el kit RNeasy mini (QIAGEN, Alemania)

Se siguieron las instrucciones del fabricante, como se explica a continuación:

1. Cosechar 4-5 x 10⁶ células totales y centrifugar 5 min a 300 g a TA.
2. Remover el sobrenadante cuidadosamente con pipeta. Asegurarse de que quede la menor cantidad posible de medio de cultivo.
3. Lisar las células mediante el agregado de buffer RLT. Agregar 600 µl del reactivo por cada 5-10 x 10⁶ células totales. Pipetear generosamente para realizar una buena mezcla.
4. Homogeneizar el lisado pasándolo al menos cinco veces por una aguja de 0,9 mm colocada en una jeringa estéril.
5. Añadir un volumen de etanol 70% (v/v) y homogeneizar con pipeta.
6. Transferir 700 µl a la columna colocada en un tubo de recolección de 2 ml.
7. Centrifugar 15 s a ≥8000 g y descartar el FT.
8. Añadir 700 µl del buffer RW1 a la columna y centrifugar 15 s a ≥8000 g para lavar la membrana de la columna. Descartar el FT.
9. Remover cuidadosamente la columna del tubo de recolección para evitar que entre en contacto con el FT, y colocar en un tubo nuevo.
10. Añadir 500 µl del buffer RPE a la columna. Centrifugar 15 s a ≥8000 g para lavar la membrana de la columna. Descartar el FT y conservar el tubo de recolección.
11. Repetir el paso 10, centrifugando 2 min a ≥8000 g.
12. Remover cuidadosamente la columna y colocarla en un tubo de recolección de 1,5 ml nuevo.
13. Añadir 30-50 µl de agua libre de RNasa a la columna y centrifugar 1 min a ≥8000 g para eluir el ARN.
14. Determinar la concentración y pureza del ARN obtenido.
15. Conservar a -70°C.

1.2. Determinación de la concentración y pureza del ARN

Mediante espectrofotometría se puede determinar la concentración y la pureza de una muestra de ARN (o ADN) basándose en la capacidad de absorbancia de un compuesto presente en una solución a una longitud de onda determinada. De este modo la concentración de la muestra se calcula teniendo en cuenta el valor de absorbancia obtenido a una longitud de onda de 260 nm. Mientras que la relación de absorbancias A₂₆₀/A₂₈₀ y A₂₆₀/A₂₃₀ se utilizan para evaluar la pureza de las muestras.

Una unidad de absorbancia a λ= 260 nm equivale a una concentración de 40 µg.ml⁻¹ (y 50 µg.ml⁻¹ de ADN doble cadena). De esta manera, la concentración y

pureza de las muestras se determinó por medida de su absorbancia a $\lambda = 260$ nm, $\lambda = 280$ nm y $\lambda = 230$ nm en espectrofotómetro UV-visible Ultrospec 2000 (LabWrench, Canadá), según las siguientes ecuaciones:

$$\text{Concentración de ADN} (\mu\text{g. ml}^{-1}) = DO(260 \text{ nm}) \times 50 \frac{\mu\text{g. ml}^{-1} \text{ ADN}}{DO} \times \frac{1}{\text{Dilución}} \quad (5)$$

$$\text{Pureza de la muestra} = \frac{DO(260 \text{ nm})}{DO(280 \text{ nm})} \quad (6)$$

Un valor de pureza comprendido en el rango entre 1,8 y 2,0 indica que la absorción es fundamentalmente debida a los ácidos nucleicos. Un valor menor a 1,8 indica que puede haber contaminación de proteínas (cuyo máximo de absorción está en torno a 280 nm).

En el caso del ARN la relación de absorbancias A260/280 con un valor entre 2,0 y 2,2 se considera indicativa de un ARN de pureza óptima. Valores A260/280 $>1,7$ se corresponden a una muestra de ARN con una pureza aceptable. En cambio, un ratio A260/280 $<1,7$ es indicativo de contaminación por la presencia de compuestos aromáticos.

A 230 nm absorben contaminantes como sales caotrópicas, fenoles o carbohidratos. Generalmente, la relación A260/230 en una muestra de ARN se corresponde a un valor de 2 o ligeramente superior. Sin embargo, no existe una opinión consensuada sobre el valor mínimo en la relación A260/230 que una muestra debe tener para ser considerada funcional. En general se considera que el ARN es aceptable si A260/230 es $>1,5$; ya que valores menores indicarían la presencia de contaminantes en la muestra.

2. Técnicas de cultivo celular

Se empleó la línea CHO-K1 (ATCC) adaptada al crecimiento en suspensión y se cultivaron tres réplicas en biorreactor para un correcto análisis estadístico por RNA-seq, según lo detallado metodológicamente en las secciones previas: Capítulo 1 / Materiales y Métodos / Sección 2. Técnicas de cultivo celular / Subsecciones 2.1.1., 2.2., 2.2.1., 2.3., 2.3.1., 2.3.2., 2.4., 2.4.2. y 2.5.

3. Metodologías para el análisis de líneas y clones celulares

No se aplicaron estas metodologías para las actividades desarrolladas en este Capítulo.

4. Experimentos y técnicas específicas para la secuenciación del transcriptoma (RNA-Seq) de células CHO-K1 adaptadas al crecimiento en suspensión y a medio libre de SFB

4.1. Extracción de ARN total

El seguimiento de los parámetros de cultivo durante los bioprocesos, en conjunto con los ensayos de caracterización del ciclo celular, permitió decidir acerca de la forma más conveniente de llevar a cabo las extracciones de ARN.

Para realizar las extracciones finales de ARN se empleó el *kit RNEasy minikit* (subsección 1.1 de este apartado), el cual se basa en el uso de mini-columnas de sílice y evita la utilización de reactivos orgánicos.

4.2. Control de la integridad de las muestras

Se realizó el control de integridad de las muestras empleando el sistema Agilent2200 TapeStation, según recomendaciones del fabricante:

Muestras de ARN:

4.2.1. RNA ScreenTape (para el control de las muestras iniciales de ARN total):

- a. Mezclar 1 μ l de cada muestra de ARN con 5 μ l del *RNA Sample Buffer*.
- b. En caso de correr un marcador de peso molecular, preparar 5 μ l del *RNA Sample Buffer* con 1 μ l del *RNA Ladder*.
- c. Homogeneizar con vórtex. Centrifugar por 1 min a 2000 rpm.
- d. Realizar la desnaturalización de las muestras:
 - i. Calentar el marcador y las muestras a 72°C durante 3 min.
 - ii. Colocar los tubos en hielo por 2 min.
- e. Realizar un *spin* de todas las muestras.
- f. Cargar los tubos en el equipo, colocar la *RNA screen tape* y configurar el programa correspondiente.

4.2.2. High Sensitivity RNA ScreenTape (para el análisis del ARNm purificado):

- a. Mezclar 2 μ l de cada muestra de ARN con 1 μ l del *HS RNA Sample Buffer*.
- b. En caso de correr un marcador de peso molecular, preparar 1 μ l del *HS RNA Sample Buffer* con 2 μ l del *High Sensitivity RNA Ladder* diluido con 10 μ l de agua libre de ARNasa.
- c. Homogeneizar con vórtex. Centrifugar por 1 min a 2000 rpm.
- d. Realizar la desnaturalización de las muestras:
 - i. Calentar el marcador y las muestras a 72°C durante 3 min.
 - ii. Colocar los tubos en hielo por 2 min.
- e. Realizar un *spin* de todas las muestras.
- f. Cargar los tubos en el equipo, colocar la *HS RNA screen tape* y configurar el programa correspondiente.

Muestras de ADN:

4.2.3. D1000 ScreenTape (para el análisis de las bibliotecas de ADNc):

- a. Mezclar 1 μ l de cada muestra de ADN con 3 μ l del *D1000 Sample Buffer*.
- b. En caso de correr un marcador de peso molecular, preparar 3 μ l del *D1000 Sample Buffer* con 1 μ l del *D1000 Ladder*.
- c. Homogeneizar con vórtex. Centrifugar por 1 min a 2000 rpm.
- d. Realizar un *spin* de todas las muestras.
- e. Cargar los tubos en el equipo, colocar la *D1000 screen tape* y configurar el programa correspondiente.

4.3. Cuantificación de las muestras de ARN

Para la cuantificación de las mismas se usó el espectrofotómetro NanoDrop2000 (Thermo Fisher Scientific, EE.UU.). El protocolo simplemente consiste en seleccionar en la computadora conectada al equipo el tipo de muestra a analizar (ácidos nucleicos en este caso), medir el blanco pipeteando 1 μ l de agua Milli Q en la parte inferior del pedestal, bajar el brazo e indicar la medición en el programa. Limpiar el pedestal, pipetear 1 μ l de la muestra y medir. Repetir este último paso para cada muestra.

4.4. Digestión de ADN contaminante

Posteriormente, se tomaron 2 μ g de cada muestra de ARN y se llevó a cabo la digestión de ADN contaminante mediante el *kit DNA-free* (Ambion, Thermo Fisher Scientific), según el siguiente protocolo:

1. Agregar a cada muestra de ARN: 2 μ l del *DNase I Buffer* 10X, 1 μ l de la enzima *rDNase I* y Agua Milli Q autoclavada csp 20 μ l.
2. Incubar por 20 min a 37°C.
3. Adicionar 2 μ l de *DNase inactivation reagent*. Homogeneizar bien.
4. Incubar durante 2 min a TA, mezclando ocasionalmente.
5. Centrifugar a 10000 *g* durante 1,5 min.
6. Transferir aproximadamente 16 μ l a un nuevo tubo (sin tocar el *pellet* blanco que corresponde al gel que forma el reactivo incorporado en el paso 3).
7. Cuantificar las muestras nuevamente.
8. Almacenar a -70°C.

4.5. Purificación del ARNm y construcción de las bibliotecas de ADNc para secuenciación mediante Illumina

La purificación de ARNm se hizo según protocolo del módulo *NEBNext Poly(A) mRNA Magnetic Isolation* (NEB, EE.UU.) y la construcción de las bibliotecas de ADNc empleando el *kit NEBNext Ultra Directional RNA Library Prep* y *NEBNext Multiplex Oligos dual Index Primers Set 1* (NEB) para *Illumina*.

Se llevó a cabo el siguiente protocolo:

1. Preparar la siguiente *mix* con *First Strand Reaction Buffer* y *random primers*.

| Reactivos | Volumen (por muestra) |
|---|-----------------------|
| <i>First Strand Synthesis Reaction Buffer (5X)</i> | 8 μ l |
| <i>Random Primers</i> | 2 μ l |
| Agua libre de nucleasas | 10 μ l |

Mantener la *mix* en hielo.

2. Purificación, fragmentación y *priming* a partir de ARN total:
 1. Diluir el ARN total en agua libre de nucleasas en un volumen final de 50 μ l en tubo de 0,2 ml y mantener en hielo.
 2. Alicuotar 20 μ l de *NEBNext Oligo d(T)₂₅ beads* (perlas magnéticas) en un nuevo tubo de 0,2 ml de PCR.
 3. Lavar las perlas agregando al tubo anterior 100 μ l de *RNA Binding Buffer (2X)*. Homogeneizar *up and down* 6 veces para mezclar bien.
 4. Colocar los tubos con las perlas (un tubo por muestra de ARN) en un *rack* magnético a TA por 2 min.
 5. Remover y descartar el sobrenadante cuidadosamente sin tocar las perlas.
 6. Retirar los tubos del *rack* magnético.
 7. Repetir pasos 2.3-2.6.
 8. Resuspender las perlas en 50 μ l de *RNA Binding Buffer (2X)* y adicionar 50 μ l de la muestra de ARN total del paso 1.
 9. Calentar a 65°C durante 5 min y mantener a 4°C para desnaturalizar el ARN y facilitar la unión de la poli(A) del ARNm a las perlas.
 10. Resuspender las perlas. Homogeneizar *up and down* 6 veces para mezclar bien.
 11. Incubar a TA por 5 min para permitir la unión del ARNm a las perlas.
 12. Resuspender las perlas. Homogeneizar *up and down* 6 veces para mezclar bien.
 13. Incubar a TA por 5 min.
 14. Colocar los tubos en el *rack* magnético a TA durante 2 min para separar de la solución el ARNm unido a las perlas.
 15. Remover y descartar el sobrenadante cuidadosamente sin tocar las perlas.
 16. Retirar los tubos del *rack* magnético.
 17. Lavar las perlas agregando 200 μ l del *Buffer* de lavado (*Wash Buffer*) para quitar el ARN que no se logró unir. Homogeneizar *up and down* 6 veces para mezclar bien.
 18. Colocar los tubos en el *rack* magnético a TA durante 2 min para separar de la solución el ARNm unido a las perlas.
 19. Remover y descartar el sobrenadante cuidadosamente sin tocar las perlas.
 20. Retirar los tubos del *rack* magnético.
 21. Repetir pasos 2.17-2.20.
 22. Añadir 50 μ l de *Buffer* Tris a cada tubo. Homogeneizar suavemente *up and down* 6 veces para mezclar bien.
 23. Calentar las muestras a 80°C por 2 min, luego mantener a 25°C para eluir el ARNm poli(A) de las perlas.

24. Agregar 50 μ l de *RNA Binding Buffer* (2X) a las muestras, para que el ARNm se vuelva a unir a las perlas. Homogeneizar suavemente *up and down* 6 veces para mezclar bien.
25. Incubar a TA por 5 min.
26. Resuspender las perlas. Homogeneizar *up and down* 6 veces para mezclar bien.
27. Incubar a TA por 5 min para que el ARN se una a las perlas.
28. Colocar los tubos en el *rack* magnético a TA durante 2 min.
29. Remover y descartar el sobrenadante cuidadosamente sin tocar las perlas.
30. Retirar los tubos del *rack* magnético.
31. Lavar las perlas agregando 200 μ l del *Wash Buffer*. Homogeneizar *up and down* 6 veces para mezclar bien.
32. Colocar los tubos en el *rack* magnético a TA durante 2 min.
33. Remover y descartar el sobrenadante cuidadosamente sin tocar las perlas. Asegurarse de retirar todo el volumen para asegurar el éxito de los pasos siguientes.
34. Retirar los tubos del *rack* magnético.
35. Eluir el ARNm agregando a cada muestra 15,5 μ l de la *mix* preparada en el paso 1.
36. Calentar las muestras a 80°C por 2 min, luego mantener a 25°C.
37. Colocar las muestras en el *rack* magnético. Transferir 0,5 μ l del sobrenadante a un nuevo tubo de PCR y agregar 2 μ l de agua libre de nucleasas. Realizar una medición de la integridad de las muestras empleando el sistema Agilent2200 TapeStation (*High Sensitivity RNA ScreenTape*) (según Sección 4.2.2) y comprobar que no exista ARNr.
38. Incubar el resto de las muestras a 94°C por 15 min.
39. Colocar los tubos en el *rack magnético* a TA y transferir 13,5 μ l del sobrenadante a un nuevo tubo de PCR.
40. Colocar las muestras en hielo y proceder inmediatamente a la síntesis de la primera hebra de ADNc.

3. Síntesis de la primera hebra de ADNc:

1. A las muestras anteriores (13,5 μ l) agregar los siguientes componentes y mezclar cuidadosamente:
 - i. *Murine RNase Inhibitor*: 0,5 μ l.
 - ii. Actinomicina D (0,1 μ g/ μ l): 5 μ l.
 - iii. *ProtoScript II Reverse Transcriptase*: 1 μ l.
2. Incubar las muestras en un termociclador precalentado (con la tapa a 105°C) según los siguientes pasos:
 - i. 25°C durante 10 min.
 - ii. 42°C durante 15 min.
 - iii. 70°C durante 15 min.
 - iv. Mantener a 4°C.
3. Inmediatamente realizar la síntesis de la segunda hebra de ADNc.

4. Síntesis de la segunda hebra de ADNc:

1. Adicionar los siguientes reactivos a las muestras anteriores y homogeneizar cuidadosamente:
 - i. Agua libre de nucleasas: 48 μ l.
 - ii. *Second Strand Synthesis Reaction Buffer* (10X): 8 μ l.
 - iii. *Second Strand Synthesis Enzyme Mix*: 4 μ l.
2. Incubar en un termociclador durante 1 hora a 16°C, con la tapa a 40°C.

5. Purificación del ADNc usando las *Agencourt AMPure XP Beads* 1,8X:

1. Resuspender las perlas AMPure XP mediante vórtex.
2. Agregar 144 μ l de las perlas resuspendidas a las muestras de ADNc (\approx 80 μ l). Mezclar bien con vórtex o con pipeta haciendo *up and down* al menos unas 10 veces.
3. Incubar a TA durante 5 min.
4. Realizar un *spin* para recolectar cualquier volumen de muestra que haya quedado en las paredes de cada tubo. Colocar los tubos en un *rack* magnético para separar las perlas del sobrenadante. Luego de 5 min remover y descartar el sobrenadante con cuidado sin tocar las perlas que contienen el ADNc.
5. Agregar a cada tubo, aún en el *rack* magnético, 200 μ l de etanol 80% (v/v) recién preparado. Incubar a TA durante 30 s, remover y descartar el sobrenadante cuidadosamente.
6. Repetir el paso 5.5 para un total de dos pasos de lavado.
7. Secar las perlas durante 5 min mientras los tubos están en el *rack* magnético con la tapa abierta.
8. Sacar los tubos del *rack* magnético. Eluir el ADN *target* de las perlas colocando 60 μ l de *Buffer* TE 0,1X o Tris-HCl 10 mM. Homogeneizar bien mediante vórtex. Realizar un *spin* e incubar por 2 min a TA. Colocar los tubos en el *rack* magnético hasta que la solución se vea clara.
9. Transferir 55,5 μ l del sobrenadante a un nuevo tubo de PCR.

6. Reparación y dA-tailing de los extremos:

1. Agregar a las muestras del último paso (volumen final 65 μ l):
 - i. *NEBNext End Repair Reaction Buffer* (10X): 6,5 μ l.
 - ii. *NEBNext End Prep Enzyme Mix*: 3 μ l.
2. Incubar las muestras en un termociclador (con la tapa a 75°C) de la siguiente manera:
 - i. 20°C durante 30 min.
 - ii. 65°C durante 30 min.
 - iii. Mantener a 4°C.
3. Proceder inmediatamente con la ligación de los adaptadores.

7. Ligación de los adaptadores:

1. Diluir el adaptador *NEBNext* para *Illumina** (15 μ M) a 1,5 μ M con Tris-HCl 10 mM pH 7,5 o Tris-HCl 10 mM pH 7,5 con NaCl 10 mM, para su uso inmediato.
2. Adicionar a las muestras previas (65 μ l) los siguientes componentes:

- i. *Blunt/TA Ligase Master Mix*: 15 μ l.
 - ii. Adaptador *NEBNext* diluido en el paso 7.1: 1 μ l.
 - iii. Agua libre de nucleasas: 2,5 μ l.
3. Homogeneizar bien con pipeta. Realizar un *spin*.
 4. Incubar a 20°C por 15 min.
- [†]El adaptador se encuentra en el *kit* de oligos para *Illumina NEBNext Multiplex*.

8. Purificación de la reacción de ligación empleando *AMPure XP Beads*:

1. Adicionar agua libre de nucleasas a las muestras anteriores para llegar a un volumen final de 100 μ l.
2. Agregar 100 μ l de las perlas *AMPure XP* resuspendidas. Mezclar bien con vórtex o con pipeta haciendo *up and down* al menos 10 veces.
3. Incubar a TA durante 5 min.
4. Colocar los tubos en un *rack* magnético para separar las perlas del sobrenadante. Luego de 5 min, descartar el sobrenadante ya que contiene fragmentos no deseados (no descartar las perlas).
5. Agregar a cada tubo, aún en el *rack* magnético, 200 μ l de etanol 80% (v/v) recién preparado. Incubar a TA durante 30 s, remover y descartar el sobrenadante cuidadosamente.
6. Repetir el paso 8.5 para un total de dos pasos de lavado.
7. Realizar un breve *spin* de los tubos y volver a colocar en el *rack* magnético.
8. Remover completamente los restos de etanol y dejar secar las perlas durante 5 min mientras los tubos están en el *rack* magnético con la tapa abierta.
9. Sacar los tubos del *rack* magnético. Eluir el ADN *target* de las perlas colocando 52 μ l de *Buffer TE 0,1X* o *Tris-HCl 10 mM*. Homogeneizar bien mediante vórtex. Realizar un *spin* e incubar por 2 min a TA. Colocar los tubos en el *rack* magnético hasta que la solución se vea clara.
10. Transferir 50 μ l del sobrenadante a un nuevo tubo de PCR. Descartar las perlas.
11. Agregar 50 μ l de las perlas *AMPure XP* resuspendidas. Mezclar bien con vórtex o con pipeta haciendo *up and down* al menos 10 veces.
12. Incubar a TA durante 5 min.
13. Realizar un breve *spin* de los tubos y volver a colocar en el *rack* magnético. Luego de 5 min remover y descartar el sobrenadante (no descartar las perlas).
14. Agregar a cada tubo, aún en el *rack* magnético, 200 μ l de etanol 80% (v/v) recién preparado. Incubar a TA durante 30 s, remover y descartar el sobrenadante cuidadosamente.
15. Repetir el paso 8.14 para un total de dos pasos de lavado.
16. Realizar un *spin* y colocar los tubos en el *rack* magnético.
17. Remover completamente los restos de etanol y dejar secar las perlas durante 5 min mientras los tubos están en el *rack* magnético con la tapa abierta.
18. Sacar los tubos del *rack* magnético. Eluir el ADN *target* colocando 19 μ l de *TE 0,1X* o *Tris-HCl 10 mM*. Homogeneizar bien mediante vórtex. Incubar por 2 min a TA. Colocar los tubos en el *rack* magnético hasta que la solución se vea clara. Transferir 17 μ l del sobrenadante (cuidadosamente sin tocar las perlas) a un nuevo tubo de PCR y proceder con la amplificación por PCR.

9. Amplificación mediante PCR del ADN unido a los adaptadores:

1. A los 17 µl de cada muestra de ADNc agregar los siguientes componentes y homogeneizar bien mediante pipeta (volumen final 50 µl):
 - i. *NEBNext USER Enzyme*: 3 µl.
 - ii. *NEBNext Q5 Hot Start HiFi PCR Master Mix 2X*: 25 µl.
 - iii. *Universal PCR primer/ i5 Primer*: 2,5 µl.
 - iv. *Index primer/ i7 Primer*: 2,5 µl.

Se empleó el *kit NEBNext Multiplex Oligos para Illumina (Dual Index Primers Set 1, NEB #E7600S)*.

2. Llevar a cabo la reacción de PCR según las siguientes condiciones:

| Paso | Temperatura | Tiempo | Nº de ciclos |
|-------------------------------------|-------------|--------|--------------|
| Digestión con la enzima USER | 37°C | 15 min | 1 |
| Desnaturalización inicial | 98°C | 30 s | 1 |
| Desnaturalización | 98°C | 10 s | 13 |
| Annealing / Extensión | 65°C | 75 s | |
| Extensión final | 65°C | 5 min | 1 |
| Mantener | 4°C | ∞ | |

El *multiplexing* de las muestras se llevó a cabo según la siguiente tabla:

| Multiplex primers | | | | |
|--------------------------|------------------------|------------------------|--------------------------|--------------------------|
| Nº de muestra | <i>Index 1 Forward</i> | <i>Index 2 Reverse</i> | <i>Index 1 Secuencia</i> | <i>Index 2 Secuencia</i> |
| 1 | i701 | i501 | ATTACTCG | TATAGCCT |
| 2 | i702 | i501 | TCCGGAGA | TATAGCCT |
| 3 | i703 | i501 | CGCTCATT | TATAGCCT |
| 4 | i704 | i501 | GAGATTCC | TATAGCCT |
| 5 | i705 | i501 | ATTCAGAA | TATAGCCT |
| 6 | i706 | i501 | GAATTCTG | TATAGCCT |
| 7 | i707 | i501 | CTGAAGCT | TATAGCCT |
| 8 | i708 | i502 | TAATGCGC | ATAGAGGC |
| 9 | i709 | i502 | CGGCTATG | ATAGAGGC |
| 10 | i710 | i502 | TCCGCGAA | ATAGAGGC |
| 11 | i711 | i502 | TCTCGCGC | ATAGAGGC |
| 12 | i712 | i502 | AGCGATAG | ATAGAGGC |
| 13 | i701 | i502 | ATTACTCG | ATAGAGGC |

10. Purificación de la reacción de PCR empleando *Agencourt AMPure XP Beads*:

1. Realizar un vórtex de las perlas *Agencourt AMPure XP* para resuspenderlas.
2. Agregar 50 µl de agua libre de nucleasas a las muestras anteriores (volumen final 100 µl).
3. Agregar 90 µl (0,9X) de las perlas resuspendidas. Mezclar bien mediante vórtex o pipeta al menos 10 veces.
4. Incubar a TA por 5 min.

5. Realizar un breve *spin* a los tubos y colocarlos en un *rack* magnético para separar las perlas del sobrenadante. Luego de 5 min, cuando la solución se vea clara, descartar el sobrenadante (no descartar las perlas que contienen el ADN *target*).
6. Agregar a cada tubo, aún en el *rack* magnético, 200 µl de etanol 80% (v/v). Incubar a TA durante 30 s, remover y descartar el sobrenadante cuidadosamente.
7. Repetir el paso 10.6 para un total de dos pasos de lavado.
8. Secar las perlas durante 5 min mientras los tubos están en el *rack* magnético con la tapa abierta.
9. Sacar los tubos del *rack* magnético. Eluir el ADN *target* de las perlas colocando 23 µl de TE 0,1X o Tris-HCl 10 mM. Homogeneizar bien con vórtex o pipeta. Realizar un *spin* e incubar por 2 min a TA. Colocar los tubos en el *rack* magnético hasta que la solución se vea clara.
10. Transferir 20 µl del sobrenadante a un nuevo tubo de PCR y almacenar a -20°C.

11. Análisis de la calidad de las bibliotecas y cuantificación de las mismas:

La calidad de las bibliotecas fue analizada empleando equipo Agilent2200 TapeStation con un *chip* para ADN (*D1000 ScreenTape*), siguiendo las instrucciones del fabricante (Sección 4.2.3).

Para la cuantificación de las bibliotecas se usó el fluorómetro Qubit 2.0 (Thermo Fisher Scientific) y el *kit dsDNA HS* de la siguiente manera:

1. Preparar la solución de trabajo colocando 1 µl de *Qubit Reagent* y 199 µl de *Qubit Buffer*.
2. Preparar dos tubos para los *standards* y un tubo por muestra a analizar de la siguiente forma:

| | Tubo <i>standard</i> | Tubo muestra |
|----------------------------|-----------------------------|---------------------|
| Solución de trabajo | 190 µl | 199 µl |
| <i>Standard</i> | 10 µl | - |
| Muestra de ADN | - | 1 µl |
| Volumen total | 200 µl | 200 µl |

3. Homogeneizar todos los tubos con vórtex unos segundos.
4. Incubar a TA por 2 min.
5. Colocar los tubos en el fluorómetro *Qubit* y cuantificar.

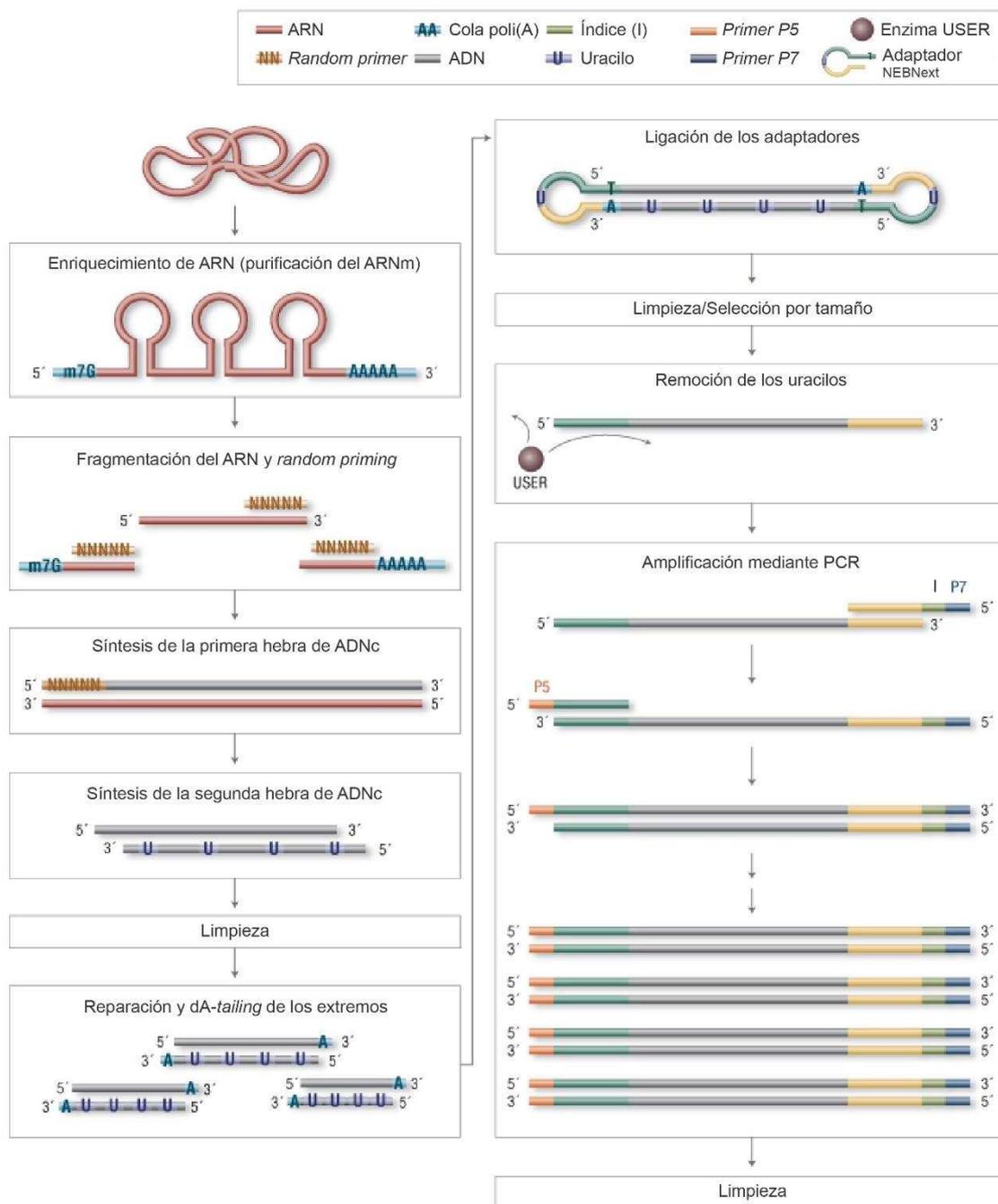


Figura 24. Protocolo general para la preparación de las bibliotecas de ADNc para su secuenciación en la plataforma *Illumina*, a partir de las muestras de ARN total de células CHO-K1 cultivadas en biorreactor. Dicho protocolo corresponde al NEBNext® Ultra™ Directional RNA Library Prep Kit para *Illumina* [178].

4.6. Secuenciación de las bibliotecas de ADNc

El *pooling* de las bibliotecas se generó agrupando las mismas acorde a la profundidad de secuenciación requerida, luego de evaluar el rendimiento de cada biblioteca de ADNc. Las mismas se secuenciaron empleando un equipo HiSeq2500 (*Illumina*), generando lecturas de una longitud de 50 pb de tipo *single-end*.

4.7. Procesamiento de los datos

Se realizó el control de calidad de los datos empleando FastQC v0.11.3 [179] y se efectuó el pre-procesamiento de los mismos utilizando Trimmomatic v0.33 [180]. Parte de los análisis de los datos se llevaron a cabo en el servidor del *Integrative Genomics and Bioinformatics Core* (Salk Institute, La Jolla, CA, EE.UU.).

Se creó un índice de la actual versión del genoma de referencia de CHO-K1 (CriGri_1.0 *RefSeq assembly accession* GCF_000223135.1) con el programa Bowtie v.1.1.2 [181], y las lecturas filtradas anteriormente se alinearon contra dicho genoma empleando TopHat v2.1.1 [182] y STAR v2.5 [183]. Para los análisis de cuantificación génica (valores FPKM normalizados) y de expresión diferencial entre las fases exponencial y estacionaria de cultivo se utilizó el paquete Cufflinks v2.2.1 [182] y HOMER v4.8 [184], aplicando el método DESeq2.

Los gráficos *scatter plots* de los valores de expresión, entre las réplicas biológicas, se crearon en el entorno R v3.3.1 [185], y también se emplearon paquetes como *cummeRbund* [186] para visualizar los resultados de los análisis de expresión. Además, se usaron los *software* Gene Cluster v3.0 [187] y Java TreeView v1.1.6 [188] para llevar a cabo los análisis de agrupamiento de datos (*hierarchical clustering*). Para el análisis de componentes principales (*principal component analysis*, PCA), con los valores de expresión normalizados, también se utilizó el paquete *cummeRbund*.

Los análisis de enriquecimiento en términos de *Gene Ontology* (GO), con los subgrupos de genes expresados diferencialmente, se realizaron utilizando PANTHER [189]. Asimismo, los análisis de vías metabólicas (empleando los mismos subgrupos de genes) se llevaron a cabo con el paquete (entorno R) Pathview [190], que permite “mapear” los genes a los *pathways* de la base de datos KEGG [191]. Para estos estudios se utilizaron los *pathways* del organismo *Mus musculus*. Para los análisis de GO y *pathways* se tuvieron en cuenta los genes diferencialmente expresados, tanto en la fase exponencial como en la fase estacionaria, en ambos casos comparado al día de inoculación (considerando valores de $\log_2\text{-fold-change} \geq 1$ y *FDR-adjusted p-value* (Benjamini Hochberg, [192]) ≤ 0.05).

Los programas mencionados se ejecutaron en Linux v16.04.

4.8. Validación de la expresión de los genes seleccionados por RT-qPCR

Con el objetivo de validar la expresión génica obtenida por RNA-seq se llevó a cabo una reacción de PCR cuantitativa (qPCR). La síntesis de ADNc se realizó a partir de las muestras de ARN de 10 días diferentes de cultivo, usando el *kit SuperScript III Reverse Transcriptase* (Invitrogen, EE.UU.), de acuerdo con las instrucciones del fabricante.

Se diseñaron pares de *primers* para 19 genes (incluido un gen *housekeeping*), para la mayoría dos pares por gen, empleando primer-BLAST [193] y Primer3 [194], teniendo en cuenta que expandan una unión exón-intrón, para brindar mayor especificidad a la reacción.

En la siguiente tabla se detallan las secuencias de los *primers* diseñados:

| Gen | Secuencia <i>primer forward</i> (5'-3') | Secuencia <i>primer reverse</i> (5'-3') |
|----------|---|---|
| S100a6_1 | GAAGGAGCTCACCATTGGCT | CGTCCATCAGCCTTGCAATC |
| S100a6_2 | CTGATCCAGAAGGAGCTCACC | TCAGCCTTGCAATCTCAGCA |
| Fth1_1 | ACTGTACGCTCCTACGTCT | GCCACATCATCCCGGTCAAA |
| Fth1_2 | GAGGTGGCCGAATCTTCCTG | CCACTCTCCCAGTCATCACG |
| Rpl34_1 | GGGAAAGCACCTAAATCCGC | CACGAACCCCTCGAAGTCTG |
| Rpl34_2 | AAGAAGCATGTCAGCAGGGC | TCCTGTCACGGACACACTTG |
| Rps18_1 | GGACGTGAAGGACGGAAAGT | CTCGCAGCTTGTTGTCCAGA |
| Rps18_2 | GAAGATCAGAGCCCACAGGG | CTGACCTCGGACACGAAGAC |
| Psap_1 | TCTGCACTACCCAAACAGCC | CCACCGGTCTTCTTCTGAGG |
| Psap_2 | ACGGCTAACAAAGACGCTGT | GTGTGCATGCTTTGCTCACA |
| Lgmn_1 | GGGCTCCAATGGCTGGTATAA | TGATCTGGTAAGCGTGGCAT |
| Lgmn_2 | ACGACGACATTGCCAACTCT | GGTCGGTTGATCACAATGCC |
| Ctsz_1 | TGCAACAACACTACCAGGCCAA | GCAGGTCCCACACTGGTTAAA |
| Ctsz_2 | GGTCGGGAGAAGATGATGGC | ACCATTATCCCGCAGCTGAT |
| Hmox1_1 | TTTCAGAAGGGCCAGGTGAC | GTACAAGGAGGCCATCACCA |
| Hmox1_2 | CAGAGGAGGCTAAGACTGCAT | TCCTGGAGCTCCTCAAACAG |
| Anxa2_1 | CGCCTATCAGAGAAGGACCAA | TGGCCAGATAAGGCTGACTTC |
| Anxa2_2 | GCCCAATACGACGCTTCTGA | GTCTTCATCAGTCCCCAGGC |
| Lamp1_1 | AACATGACTCTTCCCGATGCC | CAGTGATTGGTTGGAGGCCA |
| Lamp1_2 | TGACAGTGGAGAGCAAGAACAG | GCTAGAACTGCCATTCATCCCA |
| Actb_1 | CCAAGGCCAACCGTGAAAAG | ACATGGCTGGGGTGTGTAAG |
| Actb_2 | TTGCCCTGAGGCTCTTTTCC | TGGTGGATGCCACAGGATTC |
| Lgals1_1 | GTCTCAAAGTTCCGGGGCGA | CTTTCCCAGGTTCCAGCACA |
| Lgals1_2 | CTGGAAGCACTGTGGAGGT | GGCAGCTTGATGGTCAGGT |
| Calr_1 | CCAGAAGGACATGCATGGAGA | CAGATGTCCGGACCAAACATGA |
| Calr_2 | GGGAACCACCGTAATTCAGA | AATCTGACGTGGCTTCCACT |
| Cd63_1 | GACGGCCTTAGTTCTGGACG | GTTGTTAGCTCCACAGCACTC |
| Cd63_2 | CACTGTGGGCTGTGGAATTG | CCACGCAGCCCTGAGTATG |
| Hspa5_1 | TGGGAAAGAAGGTTACCCATGC | GGGCATCATTGAAGTAAGCCG |
| Hspa5_2 | TGGACTGAATGTCATGCGGA | CCAGGCCATACGCAATAGCA |
| Pkm_1 | GGACTGGACTCATCAAAGGCA | GTGGCTCCCTTCTTCCAGCTC |
| Pkm_2 | TGGATGACGGGCTCATTTC | GTCACCAGGTAGTCAGCACC |
| Ybx1_1 | GCGTCGACCACAGTATTCCA | TTGTCAGCACCCCTCCATCAC |
| Ybx1_2 | GTATCGGGTTACAGACCGC | TCTAGGCTGTCTTTGGCGAG |
| Vim_1 | GTGCAGTCACTTACCTGCGA | GGCGTTCCAGGGACTCATT |
| Vim_2 | GAGGTGGAGCGTGACAATCT | CCTGCAATTTCTCTCGCAGC |
| Eif3i | GCAGGACATGAGAGTGGTGA | CCAACACCTCCCAGACTTG |

Para cada uno de estos pares, utilizando el equipo *Stratagene MX3005P* (Agilent Technologies), se determinó primero el ciclo umbral y la eficiencia para la PCR cuantitativa en tiempo real mediante una curva de calibración, empleando diluciones seriadas a 1/5 de un *pool* de ADNc y agua como control sin templado. Luego, se evaluaron los resultados y se eligió un *set* de *primers* por gen.

Posteriormente, se realizó la cuantificación de la expresión génica (por triplicado) usando el colorante *SYBR green* (*SYBR green Master Mix*, Thermo Fisher Scientific) en el equipo *Stratagene MX3005 P* (Agilent), de acuerdo a los siguientes pasos:

Paso 1: 50°C por 2 min.

Paso 2: 95°C por 10 min.

Paso 3: 95°C por 15 s.

Paso 4: 60°C por 1 min.

Volver al paso 3: 40 veces.

La fluorescencia emitida se detectó luego de finalizado cada ciclo a 60°C.

El gen *GAPDH* –comúnmente usado como control interno (gen *housekeeping*)– se observó significativamente regulado en los datos de RNA-seq. Por lo tanto, se eligió el gen *Eif3i* como referencia interna.

Los análisis de los datos se llevaron a cabo en aplicaciones de hojas de cálculo, aplicando el método 2 delta-delta Ct [162].

RESULTADOS Y CONCLUSIONES

Con los resultados de ambas experiencias de cultivo se logró no sólo obtener información valiosa acerca del comportamiento de la línea celular en términos de procesos, sino que –además– caracterizar el ciclo celular de esta línea en las condiciones análogas a las de un bioproceso típico a escala industrial. Para esta etapa del trabajo, se procedió a realizar tres cultivos de células CHO-K1 adaptadas al crecimiento en suspensión en forma de réplicas, de manera tal de abarcar la variabilidad biológica inherente a los cultivos celulares y contar así con un mayor poder estadístico para los análisis posteriores. De estos tres cultivos se extrajo ARN total para su subsiguiente secuenciación con la tecnología *Illumina*.

1. Cultivos de células CHO-K1 en alta densidad en forma de réplicas biológicas para RNA-Seq

Se llevaron a cabo, en forma simultánea, tres cultivos de células CHO-K1 adaptadas al crecimiento en suspensión, de manera de contar con réplicas biológicas para realizar RNA-Seq. Se buscó imitar el comportamiento de la segunda experiencia llevada a cabo para el estudio del ciclo celular, para poder correlacionar también los análisis del ciclo celular con los transcriptómicos.

Se utilizaron biorreactores Biostat Q-plus (Sartorius), de un litro de volumen de trabajo y con régimen de perfusión, empleando medio Ex-Cell 302 libre de SFB (Sigma-Aldrich). Uno de estos cultivos tuvo una duración de 12 días, mientras que para los dos restantes fue de 17 días. Diariamente se monitorearon y controlaron las diferentes variables operativas (oxígeno disuelto, pH, temperatura, agitación y espuma) y se tomaron muestras para realizar el seguimiento de los parámetros de cultivo, de acuerdo con lo descrito en la Sección 1 del apartado Resultados y Conclusiones del Capítulo 1.

Los tres biorreactores fueron inoculados con una densidad celular de $6,0 \times 10^5$ cél./ml, buscando minimizar la fase de latencia. Se fijaron las mismas condiciones operativas en los tres cultivos (régimen de aireación, agitación, control de la espuma). Con el objetivo de evitar una proliferación celular excesiva, se aplicó un gradiente de temperatura decreciente desde el día 3 (de 37°C a 31°C).

La agitación osciló entre 120 y 150 rpm, y los valores de pH se mantuvieron en el rango 6,70 – 7,30. Los cultivos en los biorreactores comenzaron a perfundirse en el día 2 y el caudal varió entre 0,15 y 1,00 l/día. Se trabajó con concentraciones detectables de glucosa, niveles adecuados de lactato, observándose un equilibrio de ambos metabolitos (Figura 25).

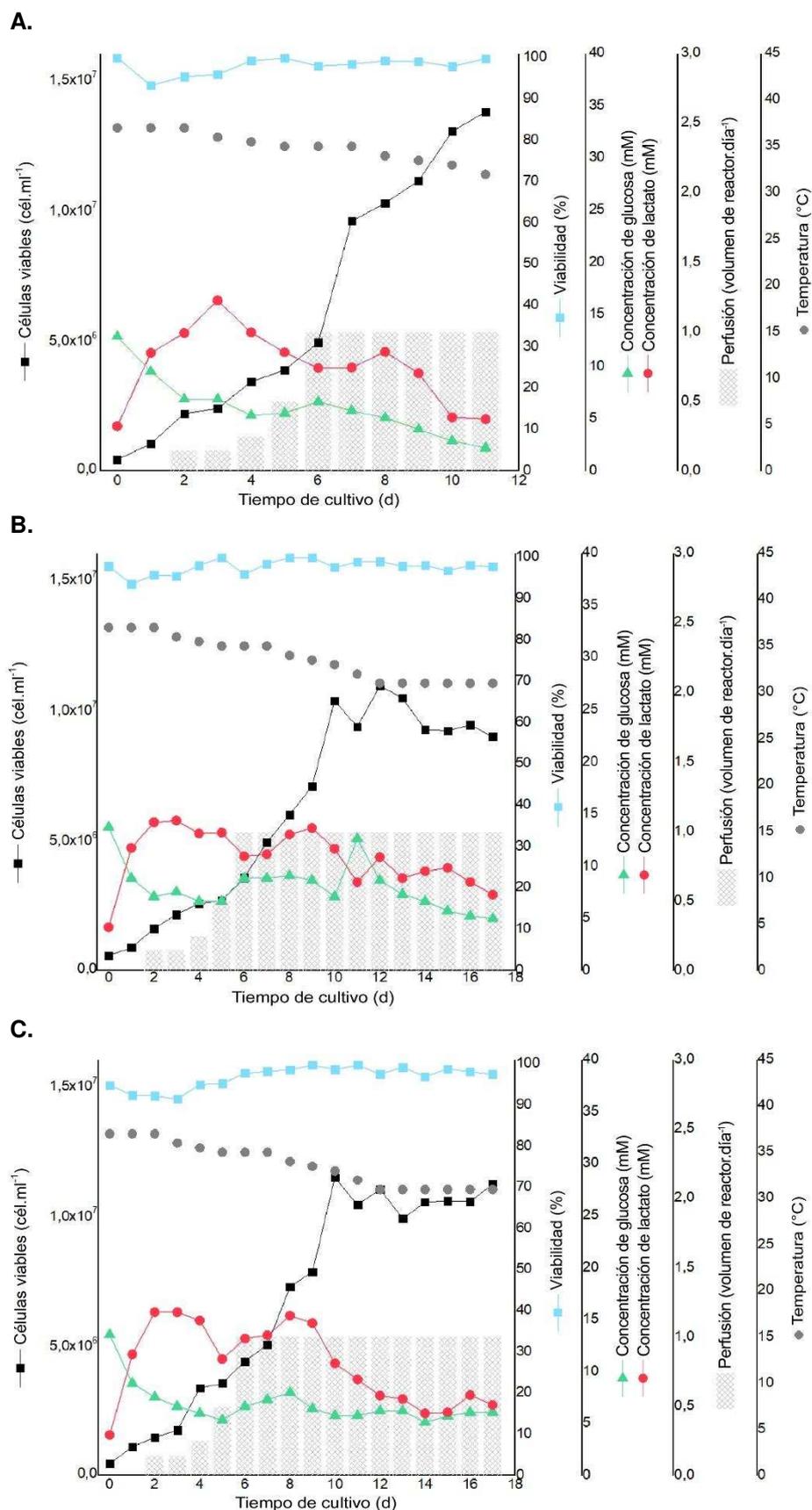


Figura 25. Réplicas de cultivos de células CHO-K1 en biorreactor para RNA-seq. Para los tres cultivos, se muestran las curvas correspondientes a la concentración de células viables, de glucosa y de lactato así como la variación de la viabilidad, de la velocidad de perfusión y de la temperatura en función del tiempo de cultivo.

Como se aprecia en la Figura 25, los cultivos exhibieron una fase de latencia de muy corta duración, seguida de una fase exponencial de rápido crecimiento. Las experiencias de las réplicas 2 y 3 alcanzaron rápidamente la fase estacionaria, ya que se controlaron en forma temprana los parámetros temperatura y abastecimiento de oxígeno, necesitando un adecuado volumen de perfusión. No fue necesario llevar a cabo operaciones de separación de suspensión celular del biorreactor, alcanzando una densidad de $1,2 \times 10^7$ cél./ml. Para las tres réplicas, el porcentaje de viabilidad fue superior al 91% durante todo el cultivo. En general, los cultivos mostraron un comportamiento muy similar al de la segunda experiencia del Capítulo 1.

Por otro lado, se distingue que los perfiles de consumo de glucosa y producción de lactato fueron aceptables y ambos se correlacionan con la dinámica de los cultivos. En cuanto a la variación de la concentración de glucosa, se observa un consumo paulatino, siempre con valores por encima del límite de detección. Además, los niveles de lactato se mantuvieron bajos. Ambos metabolitos se lograron controlar adecuadamente mediante la modificación del régimen de perfusión y los valores de temperatura de los cultivos.

Los cultivos de las réplicas 2 y 3 se interrumpieron en el día 17. Como puede observarse en la Figura 25.A, el cultivo correspondiente a la réplica 1 se interrumpió en el día 12 debido a un problema con el controlador de los biorreactores. De esta manera, se contó con tres réplicas biológicas para la fase exponencial y dos réplicas para la fase estacionaria.

2. Secuenciación del transcriptoma (RNA-seq) a partir de cultivos de células CHO-K1 en biorreactor

2.1. Extracción de las muestras de ARN total, control de calidad de las mismas y generación de las bibliotecas

Se extrajo ARN de cada cultivo, por duplicado o triplicado en algunos casos, los días: 1 (para utilizar como control), 4, 6, 7 y 8 (para la fase exponencial) y 12, 13, 14, 15 y 16 (para la fase estacionaria), empleando el *kit RNEasy minikit* (QIAGEN). Para estos extractos, se evaluó la integridad, pureza y concentración de ARN utilizando los equipos Agilent2200 TapeStation (Agilent) y NanoDrop2000 (Thermo Fisher Scientific), respectivamente.

En cuanto al análisis de integridad de estas muestras, los valores RIN (*RNA integrity number*) variaron entre 8 y 10 (Figura 26); siendo 7 el valor mínimo necesario para proceder con la secuenciación. A su vez, el análisis de pureza resultó satisfactorio, obteniéndose valores deseables para la proporción de ARNr (*rRNA ratio*).

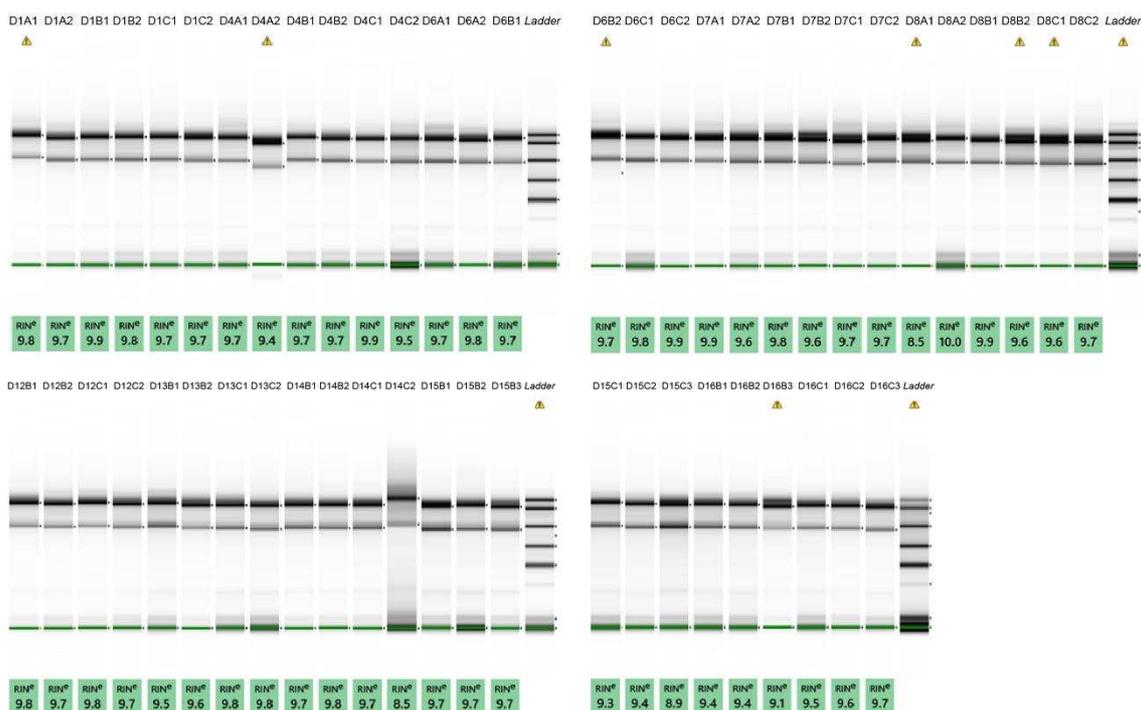


Figura 26. Resultado del análisis de integridad de las muestras de ARN.

En la parte superior de cada carril se encuentra codificada la muestra, siendo la letra “D” indicativa del día de cultivo, “A”, “B” y “C” indican las réplicas biológicas 1, 2 y 3, respectivamente; y el número final indica la réplica técnica correspondiente a la extracción de ARN. En la parte inferior se aprecia el valor RIN, indicativo de la integridad de los extractos.

Por otro lado, las concentraciones de ARN también resultaron apropiadas, según se puede ver en la Tabla 6.

Tabla 6. Concentraciones de las muestras de ARN total, valores de absorbancia a 260 nm y 280 nm y proporciones indicativas de la pureza de los extractos.

| Muestra* | Concentración (ng/μl) | Abs 260 | Abs 280 | Abs 260 / Abs 280 | Abs 260 / Abs 230 | Vol. para tratamiento con DNasa (μl)† |
|----------|-----------------------|---------|---------|-------------------|-------------------|---------------------------------------|
| D1 A1 | 395,80 | 9,90 | 4,77 | 2,08 | 1,24 | 5,10 |
| D1 A2 | 398,20 | 9,90 | 4,82 | 2,06 | 1,16 | 5,00 |
| D1 B1 | 364,00 | 9,10 | 4,41 | 2,06 | 1,83 | 5,50 |
| D1 B2 | 300,50 | 7,51 | 3,65 | 2,06 | 1,18 | 6,70 |
| D1 C1 | 317,00 | 7,92 | 3,84 | 2,06 | 1,74 | 6,30 |
| D1 C2 | 281,70 | 7,04 | 3,44 | 2,05 | 1,82 | 7,10 |
| D4 A1 | 294,80 | 7,37 | 3,58 | 2,06 | 1,17 | 6,80 |
| D4 A2 | 381,60 | 9,54 | 4,63 | 2,06 | 1,79 | 5,20 |
| D4 B1 | 341,40 | 8,54 | 4,14 | 2,06 | 1,38 | 5,90 |
| D4 B2 | 313,60 | 7,84 | 3,80 | 2,06 | 1,71 | 6,40 |
| D4 C1 | 245,20 | 6,13 | 2,98 | 2,06 | 1,19 | 8,20 |
| D4 C2 | 367,80 | 9,19 | 4,44 | 2,07 | 1,58 | 5,40 |
| D6 A1 | 444,90 | 11,12 | 5,37 | 2,07 | 1,58 | 4,50 |
| D6 A2 | 481,80 | 12,04 | 5,81 | 2,07 | 1,94 | 4,20 |

| | | | | | | |
|---------------|--------|-------|------|------|------|------|
| D6 B1 | 345,60 | 8,64 | 4,18 | 2,07 | 1,86 | 5,80 |
| D6 B2 | 365,70 | 9,14 | 4,38 | 2,09 | 0,86 | 5,50 |
| D6 C1 | 210,50 | 5,26 | 2,53 | 2,08 | 0,68 | 9,50 |
| D6 C2 | 300,90 | 7,523 | 3,66 | 2,06 | 1,16 | 6,60 |
| D7 A1 | 329,70 | 8,24 | 3,97 | 2,08 | 1,23 | 6,10 |
| D7 A2 | 385,40 | 9,64 | 4,64 | 2,08 | 1,09 | 5,20 |
| D7 B1 | 396,10 | 9,90 | 4,77 | 2,08 | 1,35 | 5,00 |
| D7 B2 | 432,40 | 10,81 | 5,19 | 2,08 | 1,02 | 4,60 |
| D7 C1 | 380,60 | 9,52 | 4,60 | 2,07 | 1,10 | 5,30 |
| D7 C2 | 370,60 | 9,27 | 4,46 | 2,08 | 1,19 | 5,40 |
| D8 A1 | 517,60 | 12,94 | 6,22 | 2,08 | 1,46 | 3,90 |
| D8 A2 | 564,00 | 14,10 | 6,80 | 2,07 | 1,61 | 3,50 |
| D8 B1 | 536,10 | 13,40 | 6,46 | 2,08 | 1,53 | 3,70 |
| D8 B2 | 469,60 | 11,74 | 5,66 | 2,07 | 1,64 | 4,30 |
| D8 C1 | 525,40 | 13,14 | 6,32 | 2,08 | 1,33 | 3,80 |
| D8 C2 | 533,30 | 13,33 | 6,43 | 2,07 | 1,45 | 3,80 |
| D12 B1 | 368,10 | 9,20 | 4,46 | 2,06 | 1,42 | 5,40 |
| D12 B2 | 314,60 | 7,87 | 3,82 | 2,06 | 1,2 | 6,40 |
| D12 C1 | 252,00 | 6,30 | 3,06 | 2,06 | 1,16 | 7,90 |
| D12 C2 | 355,90 | 8,90 | 4,34 | 2,05 | 1,45 | 5,60 |
| D13 B1 | 440,00 | 11,00 | 5,35 | 2,06 | 1,94 | 4,50 |
| D13 B2 | 380,50 | 9,51 | 4,65 | 2,05 | 1,79 | 5,30 |
| D13 C1 | 359,30 | 8,98 | 4,35 | 2,06 | 1,34 | 5,60 |
| D13 C2 | 264,30 | 6,61 | 3,23 | 2,05 | 1,01 | 7,60 |
| D14 B1 | 395,50 | 9,89 | 4,81 | 2,06 | 1,07 | 5,10 |
| D14 B2 | 330,30 | 8,26 | 4,03 | 2,05 | 1,20 | 6,10 |
| D14 C1 | 368,90 | 9,22 | 4,50 | 2,05 | 1,60 | 5,40 |
| D14 C2 | 287,90 | 7,20 | 3,54 | 2,04 | 1,41 | 6,90 |
| D15 B1 | 498,90 | 12,48 | 6,14 | 2,03 | 2,15 | 4,00 |
| D15 B2 | 242,40 | 6,06 | 3,06 | 1,98 | 2,0 | 8,30 |
| D15 B3 | 507,30 | 12,68 | 6,22 | 2,04 | 1,51 | 3,90 |
| D15 C1 | 271,40 | 6,79 | 3,31 | 2,05 | 1,85 | 7,40 |
| D15 C2 | 319,40 | 7,98 | 3,86 | 2,07 | 1,17 | 6,30 |
| D15 C3 | 450,30 | 11,26 | 5,46 | 2,06 | 1,62 | 4,40 |
| D16 B1 | 417,20 | 10,43 | 5,04 | 2,07 | 1,84 | 4,80 |
| D16 B2 | 303,20 | 7,58 | 3,69 | 2,05 | 1,73 | 6,60 |
| D16 B3 | 650,60 | 16,27 | 7,92 | 2,05 | 2,14 | 3,10 |
| D16 C1 | 234,40 | 5,86 | 2,85 | 2,06 | 1,51 | 8,50 |
| D16 C2 | 322,90 | 8,07 | 3,90 | 2,07 | 1,68 | 6,20 |
| D16 C3 | 343,30 | 8,58 | 4,20 | 2,04 | 1,23 | 5,80 |

* El nombre de cada muestra se encuentra codificado, siendo la letra "D" indicativa del día de cultivo, "A", "B" y "C" indican las réplicas biológicas 1, 2 y 3, respectivamente; y el número final indica la réplica técnica correspondiente a la extracción de ARN.

† Se muestra el volumen necesario para el posterior tratamiento con *DNasa* para la digestión del ADN contaminante.

Según los resultados obtenidos del análisis de integridad y pureza de los extractos de ARN total, se procedió a seleccionar las muestras que se detallan en la Tabla 7 para su posterior secuenciación.

Tabla 7. Muestras de ARN elegidas para su secuenciación mediante *Illumina*.

| N° de muestra | Nombre* | Descripción† |
|---------------|---------|---|
| 1 | D1_r1 | D1A1 Día 1 - Cultivo A (réplica 1) - Extracción ARN 1 |
| 2 | D1_r2 | D1B1 Día 1 - Cultivo B (réplica 2) - Extracción ARN 1 |
| 3 | D1_r3 | D1C2 Día 1 - Cultivo C (réplica 3) - Extracción ARN 2 |
| 4 | D7_r1 | D7A1 Día 7 - Cultivo A (réplica 1) - Extracción ARN 1 |
| 5 | D7_r2 | D7B1 Día 7 - Cultivo B (réplica 2) - Extracción ARN 1 |
| 6 | D7_r3 | D7C1 Día 7 - Cultivo C (réplica 3) - Extracción ARN 1 |
| 7 | D8_r1 | D8A2 Día 8 - Cultivo A (réplica 1) - Extracción ARN 2 |
| 8 | D8_r2 | D8B1 Día 8 - Cultivo B (réplica 2) - Extracción ARN 1 |
| 9 | D8_r3 | D8C2 Día 8 - Cultivo C (réplica 3) - Extracción ARN 2 |
| 10 | D14_r2 | D14B2 Día 14 - Cultivo B (réplica 2) - Extracción ARN 2 |
| 11 | D14_r3 | D14C1 Día 14 - Cultivo C (réplica 3) - Extracción ARN 1 |
| 12 | D15_r2 | D15B1 Día 15 - Cultivo B (réplica 2) - Extracción ARN 1 |
| 13 | D15_r3 | D15C2 Día 15 - Cultivo C (réplica 3) - Extracción ARN 2 |

* Se indica el nuevo nombre codificado para cada muestra, empleando para la secuenciación.

† Se detalla a qué cultivo celular (réplica biológica) pertenecen y a qué réplica técnica de extracción de ARN.

Se continuó con la construcción de las bibliotecas de ADNc de las muestras seleccionadas, previo tratamiento con *DNasa* para eliminar restos de ADN que pueda contaminar e interferir en la secuenciación.

Seguidamente, se amplificaron mediante PCR los fragmentos conteniendo los adaptadores y se realizó un control de calidad y cuantificación de las bibliotecas generadas (Figura 27).

Se puede ver que el promedio de tamaño de inserto de las bibliotecas es de 300 pb (Figura 27.A). Al aumentar el contraste (Figura 27.B) se aprecian, para ciertas muestras, bandas de bajo peso molecular, alrededor de los 150 pb. Esto también se observó en los electroferogramas generados por el equipo, como un pequeño pico en 150 pb. Dichas bandas pueden deberse a *primer dimers*. Por lo cual, decidimos llevar a cabo una segunda ronda de purificación, realizando así una nueva selección por tamaño de fragmento. Si bien la proporción de *primer dimers* que se observa es mínima respecto al total, aun así esto puede amplificarse unas cinco o más veces y quitar profundidad a la biblioteca durante la secuenciación.

De igual manera, volvimos a analizar la integridad de las bibliotecas (Figura 27.C) y no se identificaron los fragmentos de bajo peso molecular, eliminándose los artefactos producidos durante la PCR para la amplificación de las bibliotecas. Tampoco se observan en la misma imagen del gel con mayor contraste (Figura 27.D).

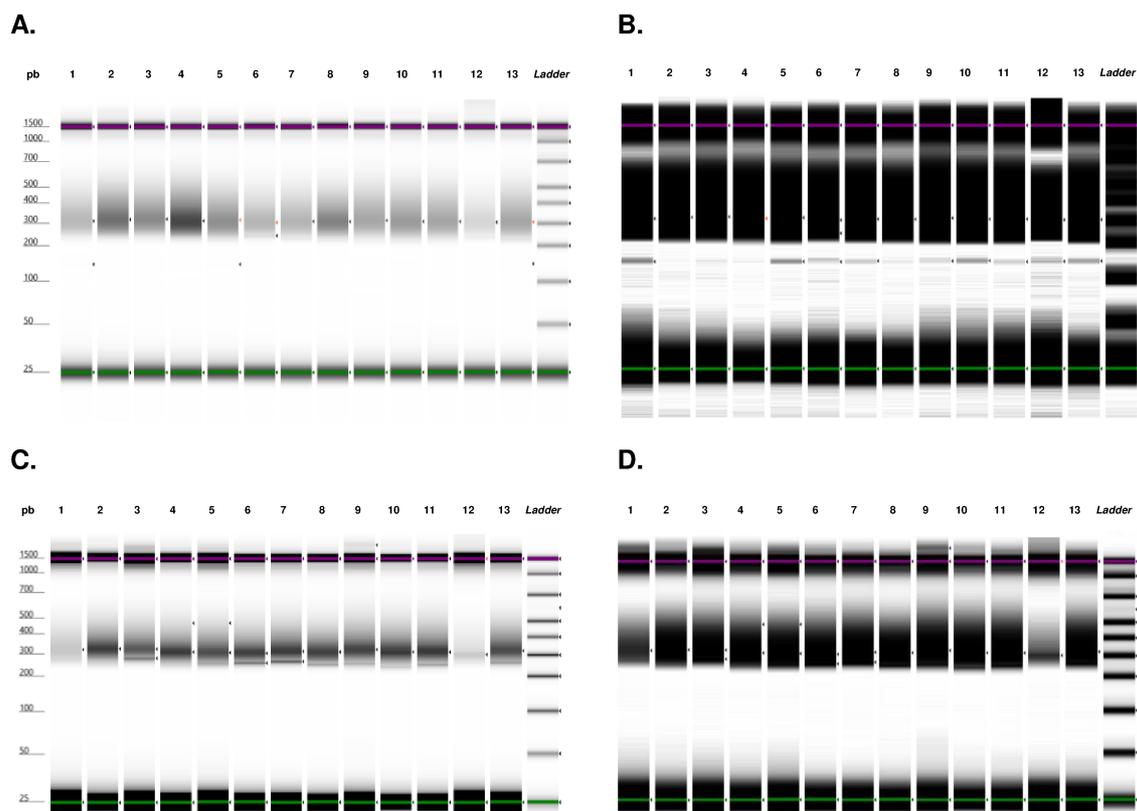


Figura 27. Control de la integridad de las bibliotecas empleando el equipo Agilent2200 TapeStation (D1000 Screen Tape).

- A.** Imagen de la electroforesis en gel automatizada, generada luego de la amplificación de las bibliotecas.
- B.** Misma imagen, pero con mayor contraste para apreciar las bandas de bajo peso molecular (aproximadamente a los 150 pb), debidas a *primer dimers*.
- C.** Imagen de la electroforesis en gel automatizada, generada luego de la amplificación de las bibliotecas seguida de una nueva ronda de selección por tamaño de fragmento, para eliminar los *primer dimers*.
- D.** Misma imagen, pero con mayor contraste.

Seguidamente, se cuantificaron las bibliotecas utilizando el fluorómetro Qubit. Los valores de concentración obtenidos se muestran en la Tabla 8. Si bien se obtuvieron rendimientos menores, luego de la segunda ronda de selección por tamaño de fragmento, las concentraciones resultaron adecuadas para proseguir con la secuenciación.

Luego, se continuó con la generación del *pooling* de las bibliotecas para secuenciar todas en una misma línea de la placa de secuenciación, del equipo HiSeq2500 de *Illumina*, de tal manera que el rendimiento esperado sea similar para todas las muestras. Se realizó un cálculo matemático, recomendado por el personal del área de secuenciación del *Salk Institute* (EE.UU.).

Se generaron finalmente 13 archivos de secuencias en formato *fastq*, obteniéndose en total un promedio de 520 millones de lecturas de 50 pb de tipo *single-end*; es decir, aproximadamente 40 millones de lecturas por biblioteca.

Tabla 8. Concentración (pre y pos-selección por tamaño de fragmento) de las bibliotecas generadas a partir de las muestras de ARN de cultivos CHO-K1 en biorreactor.

| Muestra | Concentración (ng/μl) | |
|---------|--------------------------|--------------------------|
| | Pre-selección por tamaño | Pos-selección por tamaño |
| D1_r1 | 8,04 | 1,16 |
| D1_r2 | 22,80 | 3,24 |
| D1_r3 | 8,90 | 1,92 |
| D7_r1 | 21,60 | 3,40 |
| D7_r2 | 11,90 | 2,10 |
| D7_r3 | 12,70 | 2,80 |
| D8_r1 | 14,20 | 2,74 |
| D8_r2 | 12,60 | 3,48 |
| D8_r3 | 9,74 | 2,85 |
| D14_r2 | 13,30 | 3,78 |
| D14_r3 | 10,30 | 3,34 |
| D15_r2 | 7,46 | 1,47 |
| D15_r3 | 12,10 | 2,46 |

2.2. Procesamiento de los datos

2.2.1. Control de calidad y pre-procesamiento de las secuencias

Con los archivos de secuencias generados, se realizó en primer lugar un control de calidad de los datos. En la Figura 28 se aprecian algunos de los gráficos obtenidos con el programa FASTQC; de manera representativa se indican sólo las lecturas obtenidas para el día 1 de la réplica 1.

En la gráfica de la Figura 28.A, se muestran –en forma de diagrama de cajas– los valores de calidad de todas las bases en cada posición. Se aprecian valores relativamente altos, siendo 36 el valor promedio de calidad. Esto puede verse claramente en la Figura 28.B, donde el mayor porcentaje de lecturas posee valores altos de calidad. Por otro lado, en la Figura 28.C, se encuentran graficadas las proporciones de cada una de las bases de ADN en las distintas posiciones de las secuencias. En una biblioteca aleatoria se espera que prácticamente no existan diferencias entre las distintas bases de una corrida de secuenciación. No obstante, en este análisis se observa un sesgo en la composición de las bases para aproximadamente los primeros 12 nucleótidos. Sin embargo, Hansen y col. (2010) [161] señalan que es normal y esperado observar este fenómeno debido a los *primers* hexámeros usados al momento de construir las bibliotecas. Por último, en la Figura 28.D, se muestra la gráfica para la distribución de contenido de GC a lo largo de todas las secuencias y su comparación con la distribución teórica. No se observan diferencias significativas entre ambas, lo cual es un buen indicativo de que la muestra no presenta contaminación con ADN genómico o proveniente de otro organismo.

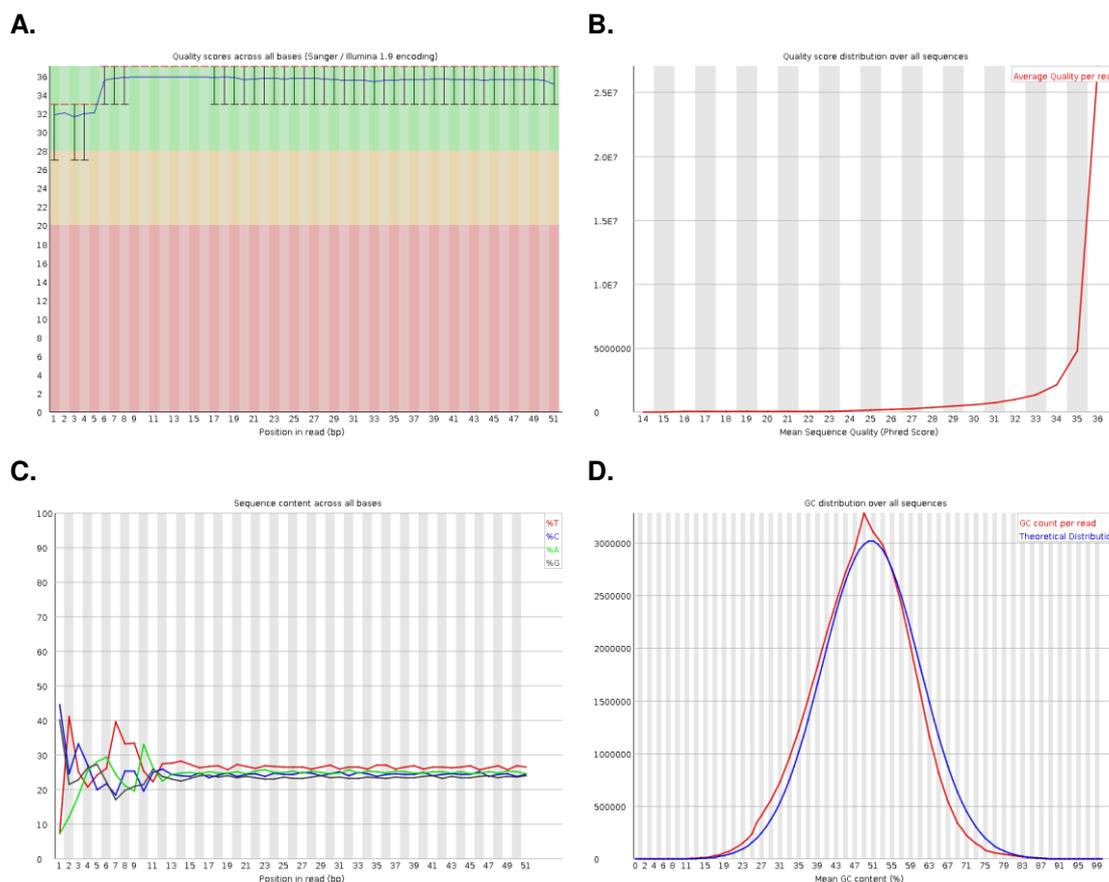


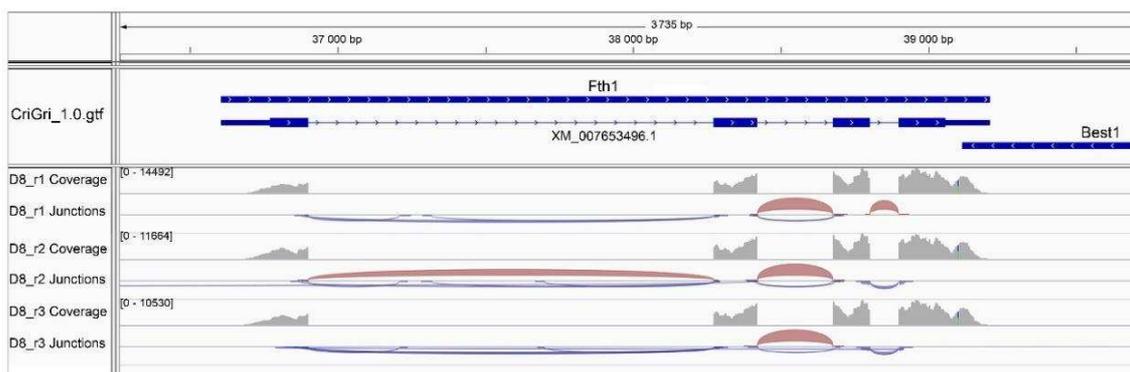
Figura 28. Control de calidad, mediante FASTQC, de los datos producidos por RNA-Seq, para las lecturas del día 1 réplica 1.
A. Gráfico *Per base sequence quality*.
B. Gráfico *Per sequence quality score*.
C. Gráfico *Per base sequence content*.
D. Gráfico *Per sequence GC content*.

Los controles de calidad de los archivos de secuencias resultaron adecuados, siendo los gráficos obtenidos con FASTQC, para todas las muestras, similares a los mostrados en la Figura 28. Ello permitió continuar con el pre-procesamiento de las lecturas, realizando la remoción de los adaptadores del proceso de secuenciación que pudieran estar presentes y de bases con una baja calidad, descartándose aproximadamente un 3% de secuencias por cada archivo.

2.2.2. Alineamiento de las lecturas y cuantificación de los transcriptos

Se compararon los resultados de los alineamientos obtenidos, contra el genoma de referencia, utilizando tanto STAR como TopHat. Se vio que el porcentaje de lecturas alineadas adecuadamente fue mayor (90% *uniquely mapped reads*) empleando STAR. Considerando que el genoma de CHO-K1 se encuentra aún en estado borrador, este valor indica un buen alineamiento. Además, se evaluó el alineamiento gráficamente, detectando que –en general– las lecturas mapean contra los exones, observándose pocas secuencias alineadas en regiones intrónicas o intergénicas (Figura 29.A).

A.



B.

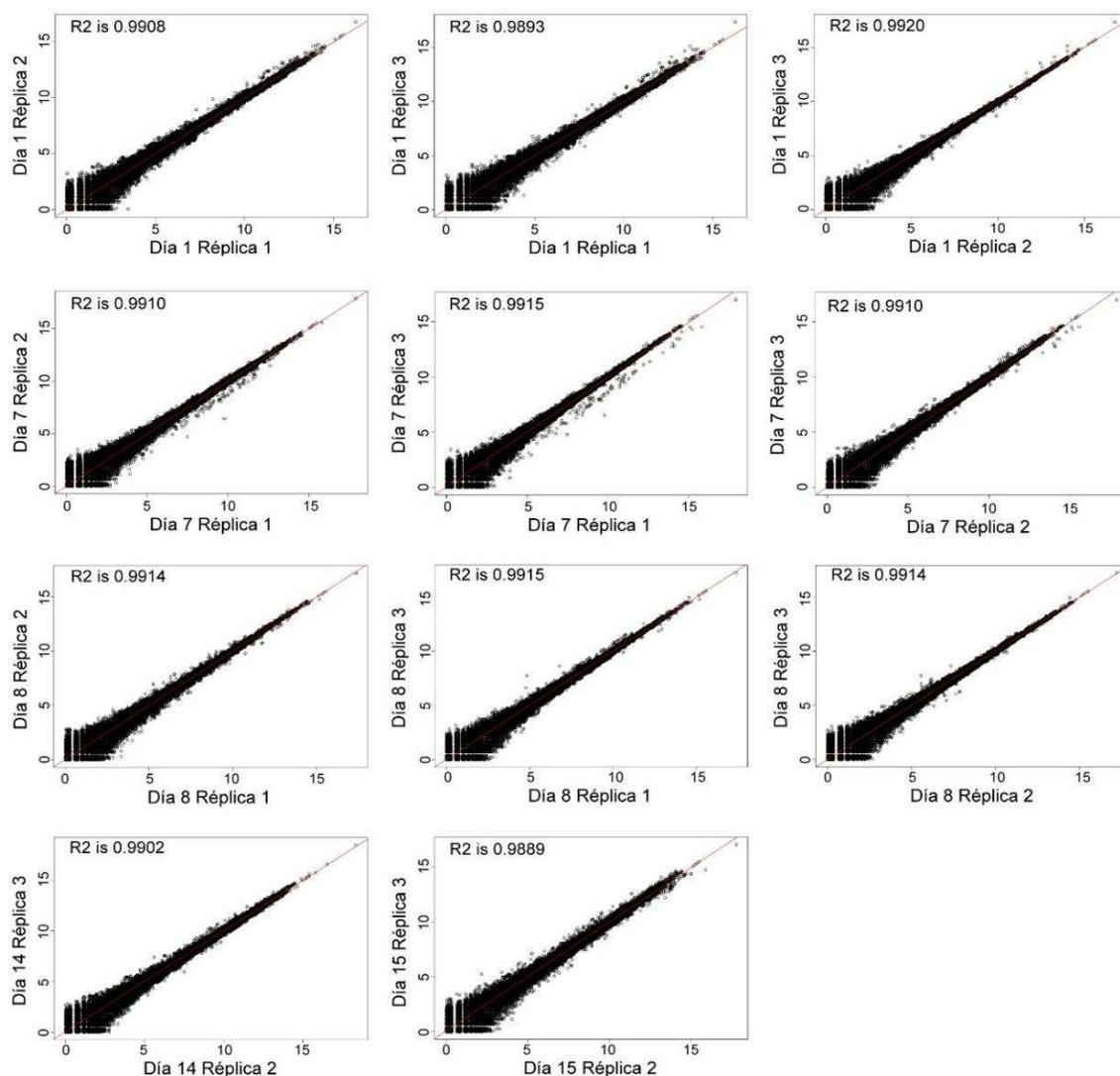


Figura 29. Análisis de células CHO-K1 mediante RNA-seq.

- A. Ejemplo de visualización con IGV de los datos de secuencia para a las réplicas biológicas del día 8, alineadas al gen *Fth1*. Se muestran los *tracks* para la cobertura (*coverage*) y las lecturas que expanden uniones exón-intrón (*splice junctions*).
- B. Gráficos de dispersión (*scatter plots*) de los valores de expresión entre réplicas para un mismo día.

Se realizó el ensamblado de los transcritos y la estimación de la abundancia de los mismos con Cufflinks. Consecutivamente, se ejecutó Cuffmerge que unifica los ensamblados, facilitando así el cálculo de la expresión en cada día. Luego se utilizó Cuffquant y Cuffnorm, para producir finalmente una serie de archivos que contienen los valores de expresión de los transcritos y genes, debidamente normalizados por el tamaño de cada biblioteca.

Para el control de calidad, luego de la cuantificación, se realizaron gráficos de dispersión de los valores de expresión entre réplicas para un mismo día (Figura 29.B). Se puede observar una correlación alta entre las réplicas de un mismo día, siendo para todos los casos los valores de R^2 superiores a 0,98. Esto indica que –como cabe esperar– no existen grandes variabilidades entre las réplicas biológicas.

A su vez, se efectuaron análisis de agrupamiento de datos (*hierarchical clustering*) y de componentes principales (*principal component analysis*, PCA) (Figura 30).

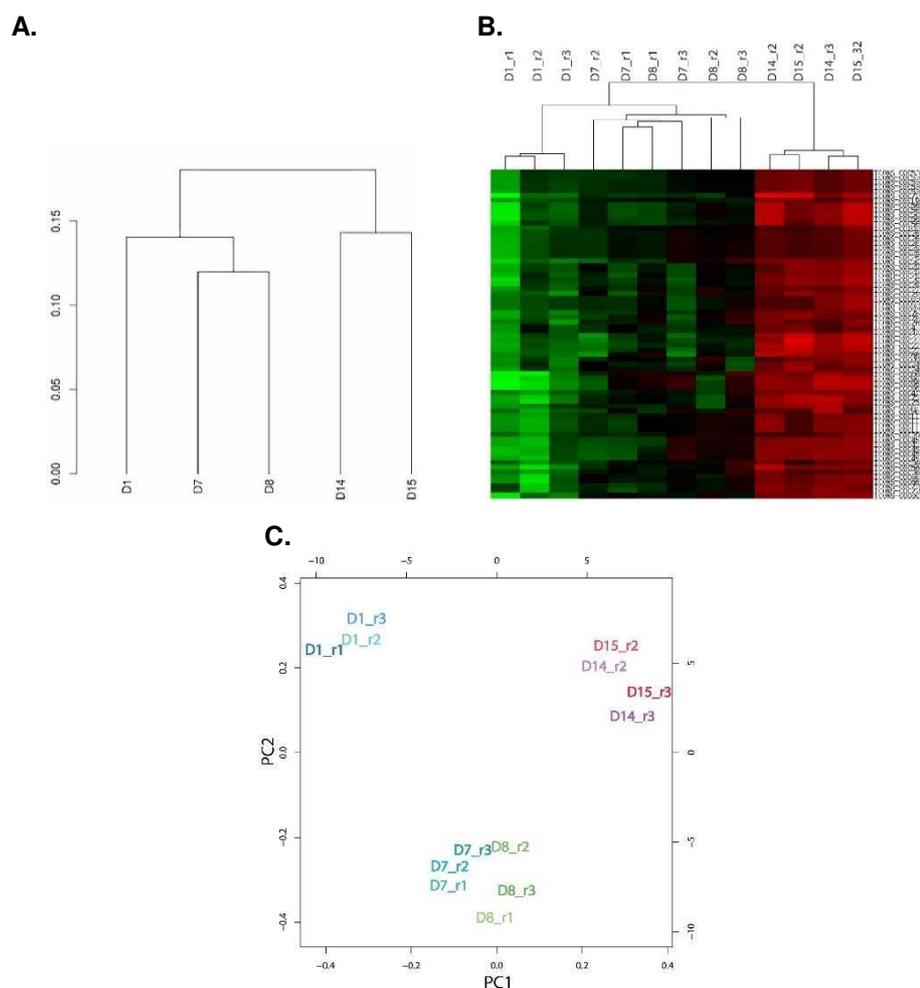


Figura 30. Agrupamiento de los datos producidos por RNA-seq.
A-B. Dendogramas obtenidos por análisis de *unsupervised hierarchical clustering*.
A. Las réplicas biológicas están agrupadas.
B. Se muestran las réplicas y la expresión de un subgrupo de genes.
C. Análisis de componentes principales (PCA) de los datos.

En particular, se observó que las réplicas de un mismo día formaron un único grupo o *cluster*. En la Figura 30.A-B se visualizan los dendogramas obtenidos. Como puede verse, los días 14 y 15 formar un *cluster*, los días 7 y 8 otro, junto a uno de mayor nivel con el día 1, evidenciando que los mayores cambios de expresión se presentan en la fase estacionaria, respecto al día de inoculación. Asimismo, en la Figura 30.C, se aprecian tres grupos diferentes, comparando PC1 y PC2, lo cual también indica una correlación alta entre las réplicas biológicas y valores de expresión similares entre los días de una misma fase de cultivo celular.

Por otro lado, también se realizaron gráficos de dispersión de los valores de expresión entre muestras de distintas etapas de cultivo. En la Figura 31 se observan los *scatter plots* generados para los valores del día 1 respecto al día 8, día 1 vs. día 15 y día 8 comparado con día 15. Puede verse que la mayor dispersión de datos ocurre al comparar los valores de expresión del día 15 de la fase estacionaria respecto al día 1 de cultivo, predominando la *up*-regulación génica.

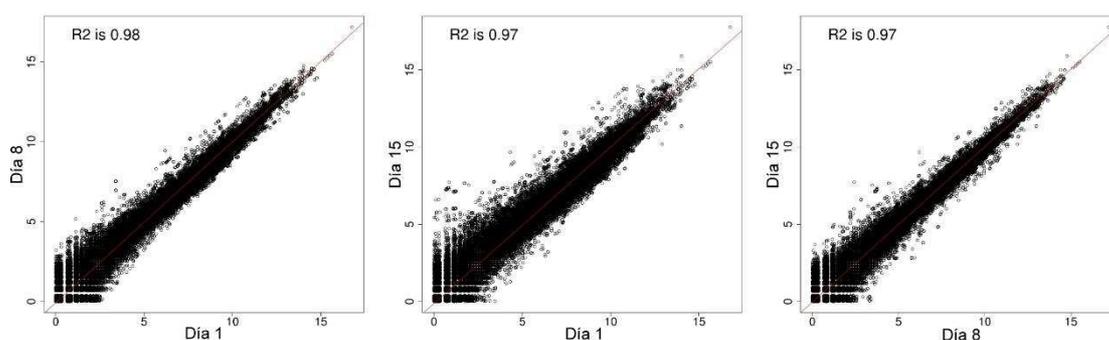


Figura 31. Gráficos de dispersión (*scatter plots*) entre los datos de las distintas etapas de cultivo en biorreactor.

Seguidamente a la cuantificación, se seleccionaron los primeros 200 transcritos mayormente expresados, con nombre y función conocida. Se descartaron así aquellos con un tamaño menor a 300 pb y sin anotación funcional, por tratarse en su mayoría de artefactos de PCR. Se indagó sobre la función biológica de los mismos, la longitud, se estudió visualmente el mapeo de los datos en cada uno, al igual que la estructura exón-intrón, y la cantidad de isoformas de cada gen. Así, se eligieron los siguientes genes: **S100a6** (*s100 calcium binding protein a6*), **Fth1** (*ferritin*), **Lgals1** (*galectin 1*), **Vim** (*vimentin*), **Actb** (*actin beta*), **Rps18** (*ribosomal protein s18*), **Cd63** (*cd63 molecule*), **Calr** (*calreticulin*), **Ybx1** (*y-box binding protein 1*), **Hspa5** (*heat shock protein family a (Hsp70) member 5*), **Hmox1** (*heme oxygenase 1*), **Ctsz** (*cathepsin z*) y **Lamp1** (*lysosomal associated membrane protein 1*). Se consideraron entonces estos genes como “candidatos” para la próxima etapa del proyecto (Capítulo 3), que implica como principal estrategia, la utilización de las regiones promotoras de dichos genes para la

construcción de vectores de expresión de proteínas recombinantes en células de mamífero.

2.2.3. Validación de la expresión de los genes seleccionados por RT-qPCR

Luego de determinar para cada par de *primers* diseñado, el ciclo umbral y la eficiencia para la PCR cuantitativa en tiempo real mediante una curva de calibración, se realizó la cuantificación relativa de la expresión génica (por el método conocido como $\Delta\Delta Ct$ [162]) de 18 genes seleccionados a partir de los análisis de RNA-Seq para 10 días de cultivo (Figura 32). El gen endógeno de referencia elegido fue Eif3i para la normalización, y se utilizaron los valores del día 1 como control.

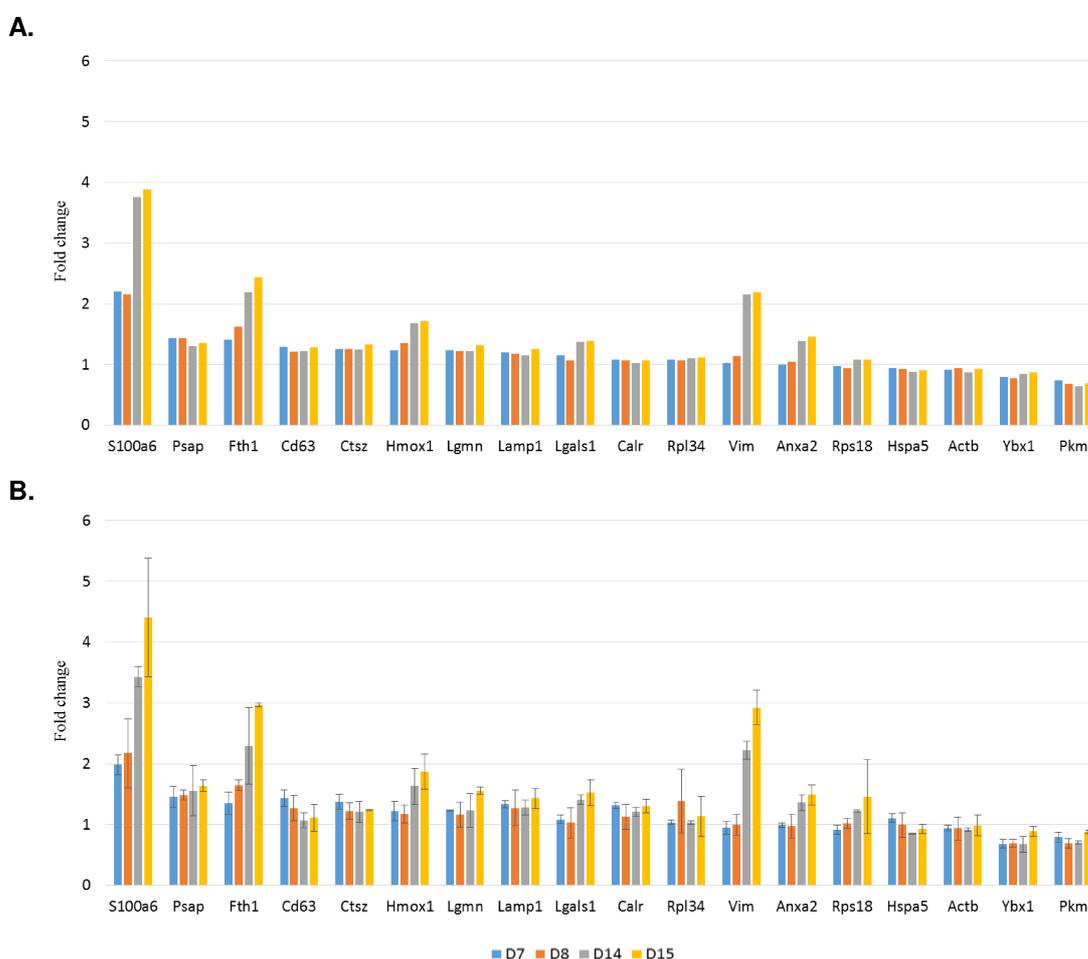


Figura 32. Cambio en la expresión de los genes seleccionados, respecto al inicio del cultivo.

A. Valores de *fold change* obtenidos mediante Cuffdiff al analizar los datos producidos por RNA-Seq.

B. Valores de *fold change* producidos por RT-qPCR.

En la Figura 32 se pueden observar los valores de *fold change* obtenidos para los genes seleccionados, ya sea mediante el análisis bioinformático de los datos (Figura 32.A) como a través de RT-qPCR (Figura 32.B). En la Figura 33 se muestra la

comparación de los valores de *fold change* (por día) de 10 genes representativos y sus correspondientes gráficos de regresión lineal.

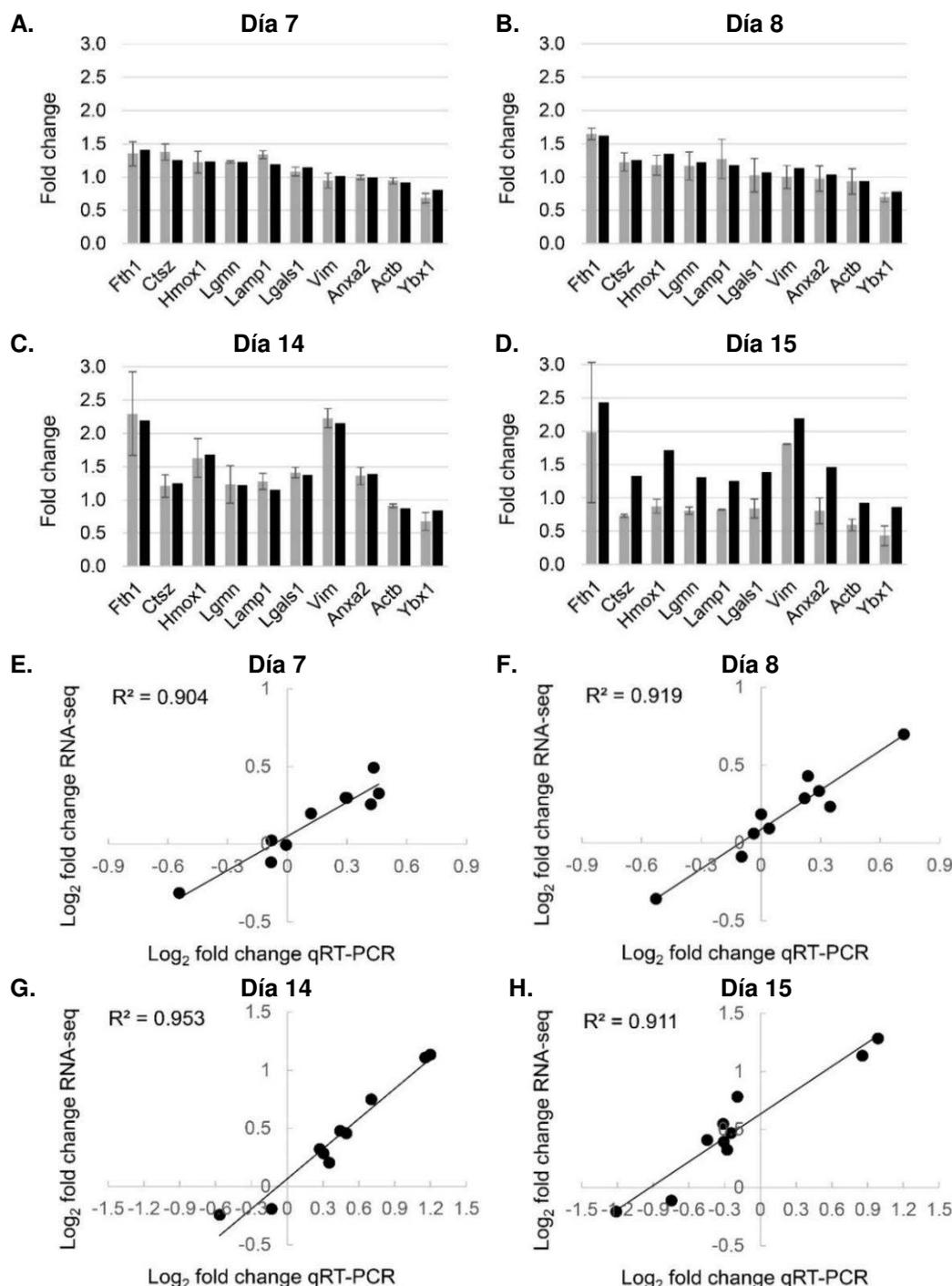


Figura 33. Validación de los datos de expresión obtenidos mediante RNA-seq a través de RT-qPCR.

Se muestran los resultados para diez genes.

A-D. Las barras de color negro representan los datos de RNA-seq y las barras grises los valores obtenidos por RT-qPCR, por día comparado con el día 1. Los valores de *fold change* de RNA-seq fueron primeramente reconstituídos a 2^x .

E-H. Análisis de regresión lineal de los valores de \log_2 *fold change* entre datos de expresión de RNA-seq y RT-qPCR, por día comparado con el día 1.

Se observa una gran consistencia en los valores obtenidos por ambos métodos. Se identificaron valores de correlación altos entre los datos de RNA-seq y RT-qPCR ($R^2 > 0,90$), verificándose así los valores de expresión derivados de los análisis *in silico*.

2.2.4. Análisis de expresión diferencial entre las fases exponencial y estacionaria de cultivos CHO-K1 en suspensión

Se llevó a cabo una caracterización general del transcriptoma con el objetivo de comprender los mecanismos celulares que regulan los fenotipos observados en los cultivos, mediante el análisis de expresión diferencial entre las fases exponencial y estacionaria, respecto al día de inicio del cultivo. Solamente se tuvieron en cuenta aquellos genes con un valor de \log_2 *fold change* ≥ 1 (es decir *up* o *down* regulados dos veces) y un p-valor ajustado (Benjamini Hochberg) $\leq 0,05$. En la Figura 34 se observan dos gráficos que resumen la variabilidad en la expresión génica entre las fases de crecimiento de los cultivos de células CHO-K1 en suspensión.

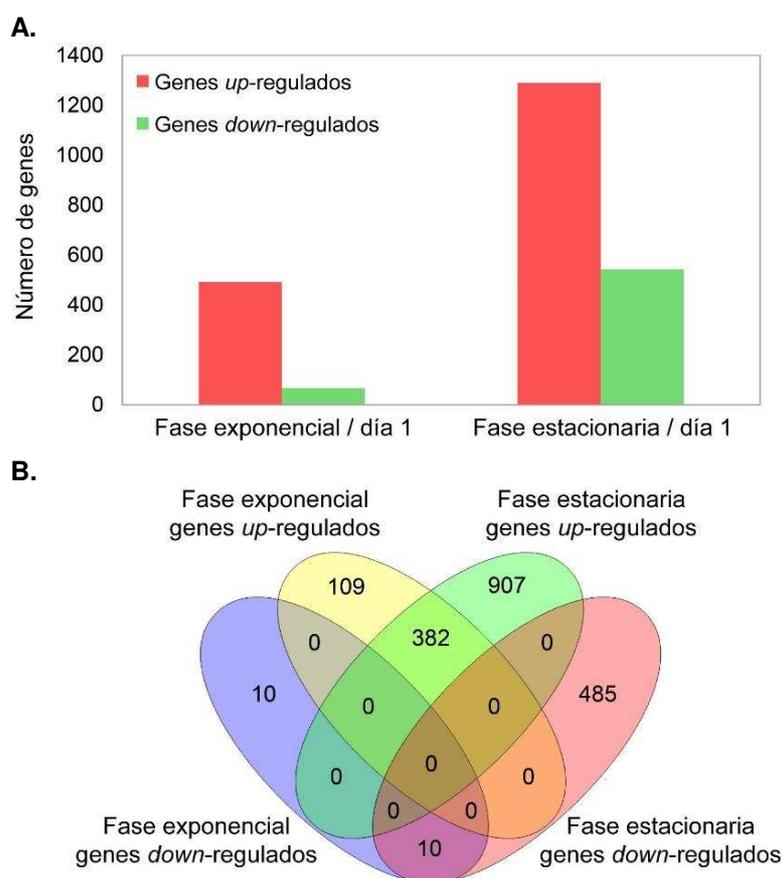


Figura 34. Análisis de expresión diferencial de las fases exponencial y estacionaria vs. el día de inoculación.

Se consideran los genes con un valor de \log_2 *fold change* ≥ 1 y un p-valor ajustado $\leq 0,05$.

- A.** Las barras de color rojo representan la cantidad de genes *up*-regulados y las barras verdes indican aquellos genes *down*-regulados en cada fase de cultivo.
- B.** Diagrama de Venn indicativo de la proporción de genes *up*-regulados y *down*-regulados durante las etapas de cultivo en biorreactor.

Como se ve en la Figura 34.A, el número de genes diferencialmente expresados (DE) entre la fase exponencial y el día 1, comparado con la cantidad de genes DE en la fase estacionaria respecto al día 1, indica que los mayores cambios regulatorios ocurren durante la fase estacionaria, predominando marcadamente la *up*-regulación génica. Esto también se pudo observar al realizar los análisis de agrupamiento y dispersión de datos, descritos en la subsección anterior. Que exista menor cantidad de genes DE durante la fase exponencial (respecto al día 1) puede estar indicando un crecimiento celular exponencial preexistente al momento de la inoculación, ya que puede apreciarse una fase *lag* de muy corta duración para los cultivos de células CHO-K1 mostrados en la Figura 25. Por otro lado, tales proporciones de genes DE muestran que las células CHO-K1 cultivadas en suspensión necesitan adecuar sus patrones de expresión génica hacia los últimos días del cultivo, de una forma más notoria que las células durante la fase exponencial. En consecuencia, las células activan una gran cantidad de genes para continuar proliferando en respuesta a las condiciones de cultivo.

Con el objetivo de identificar genes exclusivamente *up/down* regulados en cada fase de cultivo celular, se realizó un diagrama de Venn con los genes DE. Como se observa en la Figura 34.B, de los 1289 genes *up*-regulados en la fase estacionaria, 907 de ellos se encontraron específicamente DE en dicha fase, mientras que 382 genes ya se habían activado durante la fase exponencial. De igual manera, de los 495 genes *down*-regulados durante la fase estacionaria, solamente 10 ya se encontraban *down*-regulados en la fase exponencial.

Estos subgrupos de genes DE –tanto en la fase exponencial como en la fase estacionaria– representan el 2,0% y 6,4% del total de genes con actividad transcripcional en cada fase, respectivamente. Por lo cual, dichos genes DE constituyen un subgrupo altamente específico con cambios regulatorios significativos, volviéndolos *targets* (blancos moleculares) muy interesantes para experimentos de ingeniería genética, con el fin de mejorar y optimizar los bioprocesos de células CHO-K1.

2.2.4.1. Análisis de la expresión de genes que responden a cambios en la temperatura

En la industria farmacéutica, la incorporación de un gradiente de temperatura es un paso comúnmente empleado durante bioprocesos productivos, para inducir simultáneamente el arresto de las células en la fase G0/G1 del ciclo celular e incrementar la duración y la viabilidad de los cultivos (aumentando así la productividad y el rendimiento de la proteína recombinante producida).

Por este motivo, se examinó la expresión de genes que responden a cambios de la temperatura del cultivo. Ciertos genes que exhiben una *up*-regulación a bajas

temperaturas han sido ampliamente estudiados; entre ellos, CIRP (*cold-inducible RNA-binding protein*) y RBM3 (*RNA-binding motif protein 3*) [139,141,163-165]. Ambos genes están implicados en la modulación de la transcripción y de la traducción, actuando como ARN chaperonas. En particular, en condiciones de bajas temperaturas, RBM3 facilita la síntesis proteica por el mecanismo cap-independiente; y en estos estudios transcriptómicos se observó *up*-regulado más de 3,0 veces en la fase exponencial (a 35°C) y más de 5,5 veces durante la fase estacionaria (a 31°C). En cuanto al gen CIRP, se lo ha reportado con elevada expresión en condiciones de temperaturas reducidas y sin actividad transcripcional a 37°C. En coincidencia con tales hallazgos, en este estudio CIRP se vio 2,7 veces *up*-regulado durante la fase estacionaria y sin cambios en su expresión en la fase exponencial (respecto al día de inoculación). Dicho gen incrementa la productividad de las proteínas recombinantes y provoca el arresto del crecimiento de manera específica para cada línea celular. Tan y col. (2008) [164] demostraron que, al sobre-expresar CIRP en cultivos de células CHO a 37°C, el título de proteína recombinante producida aumenta un 25% en cultivos adherentes y hasta un 40% en cultivos en suspensión. Por lo tanto, estos genes que responden a temperaturas reducidas de cultivo son *targets* promisorios para la ingeniería de células CHO.

Otros genes que se han identificado diferencialmente expresados luego de cambios en la temperatura son Vim y GAPDH. En nuestros análisis de expresión, Vim se encontró 2,1 veces *up*-regulado en la fase estacionaria. Este resultado coincide con lo observado por Kumar y col. (2008) [166] al realizar el análisis proteómico de cultivos de células CHO llevados a cabo a temperaturas de cultivo reducidas (menores a 37°C). Para este mismo estudio, se informó que el gen GAPDH –que tiene un papel fundamental en las vías metabólicas de la glicólisis y la gluconeogénesis– está *up*-regulado; en cambio, en nuestros análisis encontramos que GAPDH está 1,9 veces *down*-regulado en la fase estacionaria. Esto puede deberse a una reducción global del metabolismo energético percibido durante dicha fase de los cultivos CHO-K1 en biorreactor.

2.2.5. Análisis funcionales de los datos transcriptómicos

2.2.5.1. Enriquecimiento en términos *Gene Ontology* (GO) de los genes expresados de manera diferencial

Los análisis de términos de ontología genética (*Gene Ontology*, GO) se llevaron a cabo con los genes DE durante la fase estacionaria (1289 genes *up*-regulados y 495 genes *down*-regulados), debido a que en esta fase se presentan los mayores cambios regulatorios; para explorar así los procesos funcionales en los que están implicados dichos genes. Sólo se examinaron los términos asociados a procesos

biológicos con un p-valor $\leq 0,05$, dada que esta es la categoría más informativa para nuestros propósitos.

En el subgrupo de genes *down*-regulados se identifica una clara prevalencia de términos asociados a los mecanismos de reparación y replicación del ADN, y a los procesos de ciclo celular. Mientras que en los términos enriquecidos para los genes *up*-regulados se observa una variación mayor. Las Tablas 9 y 10 muestran el resultado del *overrepresentation test* aplicado, obtenido con PANTHER. Se detallan los términos enriquecidos en cada subgrupo de genes, sus números de acceso GO, las veces que dicho término se encuentra enriquecido (*fold enrichment*) respecto a la base de datos de referencia del organismo *Mus musculus*, y el p-valor correspondiente.

Tabla 9. Análisis de enriquecimiento en términos GO relacionados a procesos biológicos, con los genes *down*-regulados durante la fase estacionaria.

| GO Procesos Biológicos | Referencia (<i>Mus musculus</i>) | | Genes <i>down</i> -regulados* (datos de la fase estacionaria) | | |
|--|---------------------------------------|-----------------|--|------------------------|----------|
| | Número de genes | Número de genes | Valor esperado | <i>Fold enrichment</i> | p-valor† |
| DNA replication initiation (GO:0006270) | 25 | 10 | 0,3 | 31,8 | 1,4E-08 |
| Mitotic chromosome condensation (GO:0007076) | 13 | 7 | 0,2 | 42,8 | 4,3E-06 |
| Positive regulation of cytokinesis (GO:0032467) | 31 | 8 | 0,4 | 20,5 | 7,3E-05 |
| Centromere complex assembly (GO:0034508) | 20 | 7 | 0,3 | 27,8 | 8,1E-05 |
| Chromosome separation (GO:0051304) | 27 | 7 | 0,3 | 20,6 | 6,1E-04 |
| Nuclear DNA replication (GO:0033260) | 17 | 6 | 0,2 | 28,1 | 8,9E-04 |
| Regulation of double-strand break repair via homologous recombination (GO:0010569) | 29 | 7 | 0,4 | 19,2 | 9,9E-04 |
| Regulation of sister chromatid cohesion (GO:0007063) | 21 | 6 | 0,3 | 22,7 | 3,1E-03 |
| DNA integrity checkpoint (GO:0031570) | 96 | 10 | 1,2 | 8,3 | 4,6E-03 |
| Positive regulation of chromosome segregation (GO:0051984) | 24 | 6 | 0,3 | 19,9 | 6,6E-03 |
| Mitotic G2/M transition checkpoint (GO:0044818) | 27 | 6 | 0,3 | 17,7 | 1,3E-02 |
| Microtubule depolymerization (GO:0007019) | 15 | 5 | 0,2 | 26,5 | 1,4E-02 |
| Centromeric sister chromatid cohesion (GO:0070601) | 7 | 4 | 0,1 | 45,4 | 1,9E-02 |
| Chromatin remodeling at centromere (GO:0031055) | 8 | 4 | 0,1 | 39,8 | 3,3E-02 |
| Positive regulation of cell cycle phase transition (GO:1901989) | 73 | 8 | 0,9 | 8,7 | 4,4E-02 |

* Se considera un valor de \log_2 *fold change* ≥ 1 y un p-valor ajustado $\leq 0,05$.

† Se muestran sólo los términos obtenidos mediante el *overrepresentation test* de PANTHER con un p-valor $\leq 0,05$.

Tabla 10. Análisis de enriquecimiento en términos GO relacionados a procesos biológicos, con los genes *up*-regulados durante la fase estacionaria.

| GO Procesos Biológicos | Referencia (<i>Mus musculus</i>) | | Genes <i>up</i> -regulados* (datos de la fase estacionaria) | | |
|---|---------------------------------------|-----------------|--|-----------------|----------|
| | Número de genes | Número de genes | Valor esperado | Fold enrichment | p-valor† |
| Cell adhesion (GO:0007155) | 756 | 47 | 17,0 | 2,8 | 5,4E-06 |
| Negative regulation of cellular protein metabolic process (GO:0032269) | 950 | 51 | 21,3 | 2,4 | 1,2E-04 |
| Cellular response to oxygen-containing compound (GO:1901701) | 713 | 41 | 16,0 | 2,6 | 5,8E-04 |
| Cell surface receptor signaling pathway (GO:0007166) | 1698 | 73 | 38,1 | 1,9 | 7,6E-04 |
| Negative regulation of cell proliferation (GO:0008285) | 645 | 38 | 14,5 | 2,6 | 9,8E-04 |
| Regulation of cytokine secretion (GO:0050707) | 176 | 18 | 4,0 | 4,6 | 1,5E-03 |
| Regulation of peptidase activity (GO:0052547) | 369 | 26 | 8,3 | 3,1 | 4,2E-03 |
| Regulation of protein modification process (GO:0031399) | 1645 | 68 | 37,0 | 1,8 | 8,9E-03 |
| Negative regulation of intracellular signal transduction (GO:1902532) | 503 | 30 | 11,3 | 2,7 | 1,7E-02 |
| Negative regulation of catalytic activity (GO:0043086) | 678 | 36 | 15,2 | 2,4 | 2,3E-02 |
| Negative regulation of extrinsic apoptotic signaling pathway (GO:2001237) | 97 | 12 | 2,2 | 5,5 | 2,5E-02 |
| Positive regulation of secretion by cell (GO:1903532) | 422 | 26 | 9,5 | 2,7 | 4,7E-02 |

* Se considera un valor de \log_2 fold change ≥ 1 y un p-valor ajustado $\leq 0,05$.

† Se muestran sólo los términos obtenidos mediante el *overrepresentation test* de PANTHER con un p-valor $\leq 0,05$.

2.2.5.2. Estudio de las vías metabólicas (*pathways*) en las cuales están implicados los genes diferencialmente expresados

Del mismo modo, se llevaron a cabo análisis de enriquecimiento en *pathways* o vías metabólicas de KEGG, teniendo en cuenta los genes DE *up* y *down*-regulados en las fases exponencial y estacionaria (genes DE con un valor de \log_2 fold change ≥ 1 y un p-valor ajustado $\leq 0,05$).

2.2.5.2.1. Ciclo celular

En primer lugar, se examinó el *pathway* correspondiente a ciclo celular. Como puede verse en la Figura 35.A, durante la fase exponencial se encuentran *down*-regulados genes implicados en la replicación del ADN. Sin embargo, en la fase estacionaria (Figura 35.B) se observa una mayor cantidad de genes *down*-regulados y

solamente tres *up*-regulados. Específicamente, el 26% de los genes de dicho *pathway* están *down*-regulados en la fase estacionaria, en contraste con sólo el 4% durante la fase exponencial (que en promedio estos genes están 1,4 veces más *down*-regulados en la fase estacionaria).

Puede verse una *down*-regulación de las ciclinas A y E en la fase estacionaria, las mismas conducen a las células hacia el arresto en la fase G1 del ciclo celular, lo cual concuerda con nuestros estudios previos. La ciclina E es esencial para que la célula progrese en el ciclo durante la fase G1 y para la iniciación de la replicación del ADN; a su vez estimula la expresión de la ciclina A, posibilitando que la célula entre y avance en la fase S.

También se identificaron genes *down*-regulados más de 2 veces en dicho *pathway*, que son esenciales para la iniciación de la replicación del ADN, como Cdc6 y Cdc45, junto con genes que forman el complejo proteico de mantenimiento de minicromosomas (MCM) y el complejo de reconocimiento de origen (ORC). Todos estos son reguladores claves de los primeros pasos del proceso de replicación del ADN, antes y durante la fase S del ciclo celular.

Además, los genes Cip1 (p21) y GADD45 se encontraron *up*-regulados durante la fase estacionaria. Estudios previos, empleando microarreglos de ADN, han reportado que la expresión de p21 se correlaciona positivamente con la supresión de genes que son importantes para el progreso del ciclo celular y con la inducción de genes asociados con la senescencia [167]. Asimismo, Cip1 tiene un papel clave en el arresto de las células en G1, lo cual concuerda con nuestros hallazgos de una proporción mayor de células en G0/G1 durante la fase estacionaria (en promedio, un 80% en G0/G1 y un 10% en fase S). Una cuestión que resulta interesante es que este efecto del gen Cip1 ha sido asociado con un incremento en la productividad de proteína recombinante en células CHO [120].

También, la ciclina B, necesaria para avanzar hacia y fuera de la fase M del ciclo celular, se vio significativamente *down*-regulada, al igual que la quinasa CDK1. Al mismo tiempo, otros genes que cumplen funciones críticas en la fase M se encontraron *down*-regulados, por ejemplo MPS1, un regulador clave que actúa como punto de control (*checkpoint*) para que se lleve a cabo la correcta segregación de los cromosomas durante la mitosis.

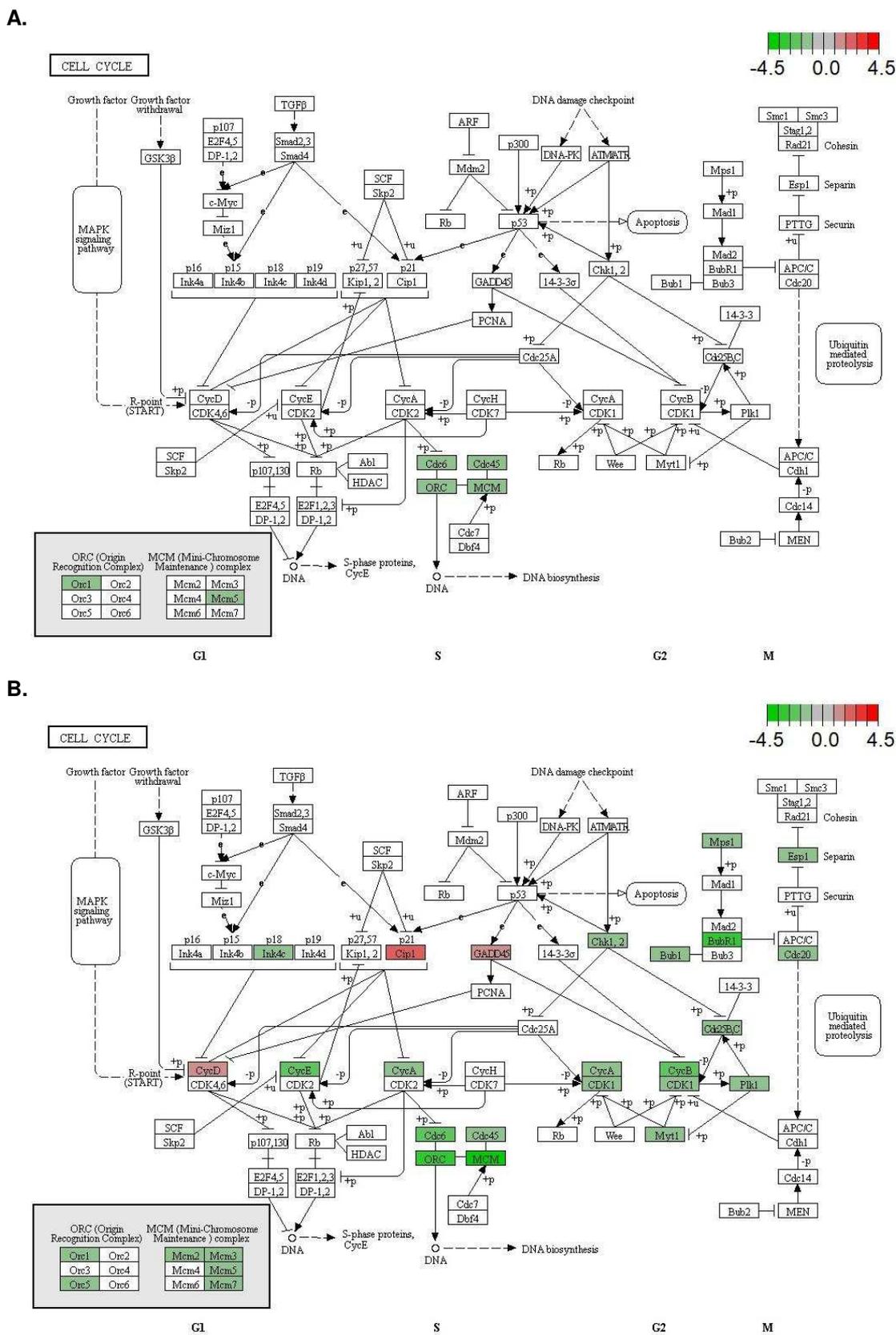


Figura 35. Genes DE mapeados al *pathway* de ciclo celular de KEGG.
Los rectángulos rojos indican los genes *up*-regulados y los verdes los *down*-regulados. Se muestra en el extremo superior derecho la escala de colores indicativa de los valores de expresión (\log_2 fold change).
A. Genes DE durante la fase exponencial.
B. Genes DE en la fase estacionaria.

2.2.5.2.2. Apoptosis

Un aspecto crítico en los cultivos celulares de mamíferos es el problema de la muerte celular, la cual puede desencadenarse por falta de nutrientes, acumulación de producto recombinante y otras señales de estrés; las células inician así la apoptosis o muerte celular programada. El empleo de bajas temperaturas de cultivo reduce el metabolismo celular y, en consecuencia, las células no entran en apoptosis [168]. En nuestros cultivos de células CHO-K1 se observaron porcentajes de viabilidad celular altos y relativamente constantes, principalmente durante la fase estacionaria.

Al analizar el *pathway* de KEGG vinculado a la apoptosis, con los genes DE durante la fase estacionaria, se encontraron significativamente regulados diferentes genes con funciones en las dos vías de señalización que dan inicio a la apoptosis. Por ejemplo, el gen TRAIL –un receptor que detecta la presencia de señales extracelulares de muerte y promueve la vía extrínseca de apoptosis– se vio 4,6 veces *up*-regulado. Asimismo, el gen BIM –que está involucrado en la vía intrínseca de apoptosis desencadenada por diversos eventos mitocondriales– se encontró 2,0 veces *up*-regulado; ambos durante la fase estacionaria respecto al inicio del cultivo.

Asimismo, genes implicados en procesos que evaden la apoptosis se vieron también sobre-expresados, favoreciendo de esta manera la supervivencia celular. Por ejemplo, NGF se observó casi 9,0 veces *up*-regulado en la fase estacionaria (y 3,0 veces *up*-regulado durante la fase exponencial). En ausencia de dicho gen, las células entran en apoptosis y, en consecuencia, se activa el gen PI3K (*phosphatidylinositol-3-kinase*), que tiene un rol fundamental en la vía metabólica PI3K-Akt. Este *pathway* se observó globalmente *up*-regulado en la fase estacionaria (alrededor del 15% de los genes de dicha vía se encontraron sobre-expresados y solamente un 3% se encontró *down*-regulado). El *pathway* de PI3K es crucial en diversos aspectos de crecimiento y supervivencia celular. De hecho, ningún gen implicado en la apoptosis se vio activado en esta vía y aquellos genes relacionados con la regulación del ciclo celular y la proliferación estaban *up*-regulados.

2.2.5.2.3. Metabolismo energético

En un estudio realizado por Moore y col. (1997) [168] observaron que, al cultivar células CHO en condiciones de bajas temperaturas, se reducía el metabolismo energético global. Al analizar los *pathways* de glicólisis y gluconeogénesis, se observaron cinco genes significativamente *down*-regulados más de 2,0 veces en la fase estacionaria: TPI1, GAPDH, PGK1, PCK1 y ACSS2. Esta reducción general en dichas vías metabólicas puede resultar suficiente para aumentar la viabilidad del cultivo, por disminuir la acumulación de metabolitos tóxicos para las células.

Un aspecto bien conocido de las células CHO es el uso ineficiente que dichas células hacen de la glucosa para producir ATP [169], las cuales convierten la mayor cantidad de glucosa disponible en lactato, produciendo ATP a través de la glicólisis aeróbica. Se sabe que la acumulación del ácido láctico en el medio de cultivo inhibe el crecimiento celular y la producción de proteína recombinante [170,171].

La enzima lactato deshidrogenasa A (LDHA) cataliza la conversión reversible de piruvato a lactato. La misma presenta elevados niveles de expresión en las líneas celulares CHO. Sin embargo, en este análisis transcriptómico no se observó ningún cambio en el perfil de expresión de dicho gen a lo largo de todo el cultivo; de hecho presentó niveles constantes y relativamente bajos de expresión durante el bioproceso. Por otro lado, ciertos estudios han demostrado que en las fases tardías de crecimiento de cultivos en modo *fed-batch* o continuo, las células de mamífero alteran su metabolismo y pasan de producir lactato a consumirlo [169,172]. En contraste, las células quiescentes metabolizan la glucosa a una tasa menor y la mayoría de la glucosa consumida se convierte en biomasa. Por esto, las células en cultivo pueden o no alterar su metabolismo de forma tal de consumir lactato de acuerdo con la disponibilidad de glucosa y a la tasa de proliferación celular.

Al no detectar cambios significativos en la expresión del gen LDHA, pero sí observar una reducción de la concentración de ácido láctico en el medio de cultivo, cuando las células alcanzan la fase estacionaria, no queda claro si en nuestros bioprocesos de células CHO-K1 el lactato está siendo consumido o directamente las células no lo están generando.

2.2.5.2.4. Procesamiento del ADN

Al analizar los *pathways* relacionados con el procesamiento y reparación del ADN encontramos una *down*-regulación general tanto en la vía de recombinación homóloga (HR) como en la de replicación del ADN, especialmente durante la fase estacionaria.

Se ha reportado que la ausencia de ciertos genes involucrados en el *pathway* de HR, como Rad51B, Rad51C, BRCA1 y Rad50, llevan al arresto de las células en la fase G0/G1 del ciclo celular, aún en ausencia de ADN dañado. En los análisis de expresión diferencial se observó al gen Rad51C (esencial para dicha vía) 2,2 veces *down*-regulado en la fase estacionaria. Además, BARD1 se vio 3,9 veces *down*-regulado; dicho gen interactúa con el extremo N-terminal de BRCA1, el cual se encontró también 2,3 veces *down*-regulado. El heterodímero BRCA1-BARD1 coordina una gran cantidad de *pathways* celulares, como la reparación del ADN dañado, ubiquitinación y regulación transcripcional, para mantener la estabilidad genómica. Asimismo, otros genes esenciales en la vía de recombinación homóloga se vieron *down*-regulados en la fase

estacionaria, como BRIP1 (2,3 veces), PALB2 (2,5 veces), BRCA2 (2,4 veces), Rad54L (3,5 veces), POLD1 (2,6 veces) y BLM (2,0 veces).

Del mismo modo, genes importantes para la replicación del ADN se observaron *down*-regulados durante la fase estacionaria de los cultivos CHO-K1. Los genes que codifican las subunidades catalíticas de la ADN polimerasa delta y epsilon (POLD1 y POLE) presentaron una *down*-regulación de 2,6 y 2,0 veces, respectivamente. También, varios genes que conforman el complejo MCM exhibieron una *down*-regulación: MCM2 (2,0 veces), MCM3 (2,8 veces), MCM5 (3,1 veces) y MCM7 (2,4 veces). Dicho complejo tiene una función helicasa de ADN y actúa en las etapas de inicio y de elongación de la replicación del ADN eucariota, precisamente en la formación y elongación de la horquilla de replicación. Además, el gen FEN1 –que codifica una endonucleasa esencial para la reparación de cortes en ambas hebras del ADN en la vía de recombinación mediada por microhomología (MMEJ)– se vio 2,3 veces *down*-regulado.

Tales variaciones en los perfiles de expresión génica de dichos *pathways* se advirtieron de manera notoria a medida que el cultivo progresaba en la fase estacionaria, acorde a lo que observamos en los análisis de ciclo celular. Es decir, la maquinaria del ADN se *down*-reguló; en consecuencia, las células no fueron capaces de progresar por el ciclo celular y así aumentó la proporción de células arrestadas en G0/G1 durante la fase estacionaria. Esta dinámica puede deberse, en parte, a la activación de vías que responden a señales de estrés, como los *pathways* de señalización p53 y AMPK; ambos globalmente *up*-regulados durante la fase estacionaria. Por un lado, está reportado que p53 induce el arresto en el ciclo celular, la senescencia o la apoptosis. Por otro lado, una vez que se activa AMPK, este desencadena una inhibición de las vías metabólicas que consumen energía.

2.2.5.2.5. Procesamiento de proteínas en el retículo endoplasmático y modificaciones postraduccionales

En relación al procesamiento de las proteínas en el retículo endoplasmático (RE), durante la fase estacionaria se pudo ver una *up*-regulación en las vías de plegamiento normal, de control de calidad y de secreción, así como también en las vías de respuesta a estrés y mal plegamiento proteico. Específicamente, los genes UGGT2, ERLEC2, SEC21B y JNK se encontraron 2,7; 2,0; 3,4 y 2,0 veces *up*-regulados, respectivamente.

Harreither y col. (2015) [173] reportaron que, en líneas celulares con niveles altos de productividad, ciertas etapas de la vía secretora estaban *up*-reguladas en respuesta al elevado contenido de proteína recombinante en las células. En nuestros estudios de expresión diferencial de células CHO-K1 no productoras, se obtuvieron resultados similares. Esto sugiere que la regulación del procesamiento proteico está relacionada

no sólo con la producción de proteína recombinante, sino también con las condiciones del bioproceso, por ejemplo, con el uso de temperaturas reducidas de cultivo. En dicho estudio también observaron que genes como CHOP y ATF4, con funciones regulatorias importantes dentro del *pathway*, estaban *up*-regulados. En los análisis por RNA-seq aquí descritos, ATF4 solamente presentó expresión diferencial significativa durante la fase estacionaria, al comparar con el día de inoculación (2,0 veces *up*-regulado), y el gen CHOP se expresó en niveles muy bajos durante todo el cultivo. Esto indica que el retículo endoplasmático presentó bajos niveles de estrés, y, en consecuencia, las células tampoco desencadenaban la apoptosis.

Otro aspecto que es crítico en los cultivos celulares de mamíferos es que la glicosilación de la proteína recombinante en cuestión se lleve a cabo de manera correcta; siendo problemas típicos en los procesos de producción de biofarmacéuticos, la obtención de valores bajos de productividad específica y una incorrecta glicosilación proteica. Estudios previos describen que las modificaciones postraduccionales, en particular la glicosilación, se ven favorecidas en condiciones de bajas temperaturas de cultivo [118,174].

Al analizar la expresión de genes involucrados en el *pathway* de N-glicosilación no se observaron cambios significativos de expresión, tanto en la fase exponencial como en la fase estacionaria. El único gen, relacionado a un proceso de glicosilación, que exhibió un cambio significativo en sus valores de expresión fue GALNT6, el cual codifica para una enzima de la familia de acetil-galactosamina-transferasas (GalNAc-T). Dicha enzima inicia la O-glicosilación en el aparato de Golgi. El gen GALNT6 se vio 3,0 veces *up*-regulado en la fase exponencial y 10,5 veces *up*-regulado durante la fase estacionaria. Estos resultados indican que la vía de O-glicosilación se ve favorecida en cultivos de células CHO bajo las condiciones descritas previamente en este trabajo.

2.2.5.2.6. Otras vías metabólicas

Ciertos genes pertenecientes al *pathway* del citoesqueleto de actina mostraron una *up*-regulación en la fase estacionaria. Por ejemplo, el factor de crecimiento FGF13 y el receptor FGFR2 se vieron 3,0 veces y 4,3 veces *up*-regulados, respectivamente; encontrándose este último *up*-regulado de manera similar durante la fase exponencial.

Además, diversos genes que codifican para integrinas se encontraron *up*-regulados: ITGA8 se observó 2,0 veces *up* en la fase exponencial y 2,9 veces en la estacionaria; y el gen ITGA5 mostró una *up*-regulación de 2,0 veces solamente en la fase estacionaria de cultivo. Por otro lado, el gen IQGAP3, que controla la adhesión celular, se vio 2,8 veces *down*-regulado durante el transcurso de la fase estacionaria. Walthers y col. (2016) [175] reportaron que la expresión de genes codificantes de

integrinas tienen una función fundamental en los procesos de señalización y proliferación celular, en diferentes cultivos de células CHO adaptadas al crecimiento en suspensión. Esto demuestra la importancia de las integrinas para la supervivencia celular, incluso en un entorno sin ligando.

Por otra parte, diferentes genes relacionados a vías metabólicas implicadas en el desarrollo del cáncer se encontraron diferencialmente expresados, predominando la *up*-regulación génica. Lee y col. (2016) [176] identificaron genes relacionados al cáncer que estaban significativamente *up*-regulados. Estos genes fueron utilizados como *targets* mediante estrategias de ingeniería genética para acelerar la adaptación de las células CHO al crecimiento en suspensión, basados en que las células CHO-K1 adaptadas al crecimiento en suspensión exhiben un comportamiento similar al de las células que se encuentran en procesos de metástasis del cáncer. Asimismo, Pourhassan-Moghaddam y col. (2012) [177] plantearon el uso de regiones promotoras de ciertos genes implicados en el desarrollo del cáncer en el diseño de nuevos vectores de expresión, basándose en el hecho de que dichos genes presentan elevados niveles de transcripción.

En resumen, estos hallazgos sugieren que la exposición de las células CHO-K1 a temperaturas reducidas de cultivo desencadena una respuesta coordinada que comprende la modulación del ciclo celular, del metabolismo energético, de la proliferación celular, la reorganización del citoesqueleto y de procesos como la replicación y reparación del ADN, transcripción, metabolismo proteico y apoptosis.

Se pudieron identificar genes con funciones clave en el arresto de las células en la fase G0/G1, para cada etapa del cultivo; correlacionando así los análisis transcriptómicos con los de ciclo celular (Capítulo 1).

Todos estos genes identificados, que exhibieron cambios significativos de expresión, pueden resultar potenciales *targets* para la ingeniería celular y/o metabólica de las células CHO, en particular de las células huésped usadas para producir proteínas recombinantes terapéuticas. De esta manera, los bioprocesos de células CHO-K1 se pueden optimizar, tanto a escala de laboratorio como industrial.



CAPÍTULO 3

*Utilización de promotores
endógenos identificados en células
CHO-K1 para la producción de
proteínas recombinantes
terapéuticas*

CAPÍTULO 3

Utilización de promotores endógenos identificados en células CHO-K1 para la producción de proteínas recombinantes terapéuticas

INTRODUCCIÓN

“(...) parecía que habíamos llegado al final del camino y resulta que era sólo una curva abierta a otro paisaje y a nuevas curiosidades”.

José Saramago

1. Análisis de promotores

Los promotores son uno de los componentes clave en la regulación de la expresión génica. En el contexto de producción de proteínas recombinantes, una etapa muy importante corresponde a la selección del promotor. El conocimiento pleno de la secuencia, los elementos constitutivos y la actividad de los promotores determina en gran parte el éxito del proceso productivo.

El gran interés en el estudio de los promotores radica en el potencial que ofrecen para regular y controlar la expresión del TG, ya que no sólo permiten incrementar la actividad transcripcional sino también proporcionar niveles adicionales de control, por ejemplo, en tiempos o estadios específicos [195].

Por ello, constituyen un área interesante de investigación mediante la aplicación de tecnologías de secuenciación, particularmente RNA-seq. Recientemente, se ha comenzado a estudiar en células CHO diversas regiones promotoras, sus componentes e interacciones, así como la búsqueda de nuevos promotores que dirijan niveles altos de expresión transgénica de forma estable.

1.1. Proceso de transcripción eucariota

La transcripción en eucariotas se realiza en el núcleo y, aunque el proceso básico es similar al de procariontes, la maquinaria transcripcional, la regulación y el destino de los ARNm es mucho más complejo. La primera etapa del proceso de transcripción en eucariotas es llevada a cabo por los complejos enzimáticos formados por las ARN polimerasas, que transcriben el ADN genómico en ARN. La estructura de la ARN polimerasa eucariota es más complicada que en bacterias; además, se requiere una gran cantidad extra de elementos proteicos y secuencias en el ADN para señalar y regular los niveles de síntesis del ARNm. Existen tres ARN polimerasas diferentes, cada una con una función específica. La ARN polimerasa I produce el pre-ARNr 45S involucrado en la formación de los ribosomas; la ARN polimerasa II sintetiza ARNm, microARNs y ARN pequeños nucleares; la polimerasa III produce los ARNs de

transferencia, el ARNr 5S y los pequeños ARNs. Probablemente, la enzima más estudiada sea la polimerasa II, debido a que es la responsable de la síntesis de las secuencias codificantes para las proteínas.

En general, para comenzar el proceso de transcripción, la polimerasa se posiciona en la vecindad del sitio de inicio de la transcripción, localizado en el extremo 5' del gen (Figura 36). La región promotora está formada por diversas subregiones con secuencias específicas que son capaces de unir a los factores de transcripción. Dichos factores posibilitan la posterior unión del complejo compuesto por la ARN polimerasa y sus proteínas auxiliares. Una vez formada la burbuja de transcripción, la polimerasa puede comenzar a transcribir.

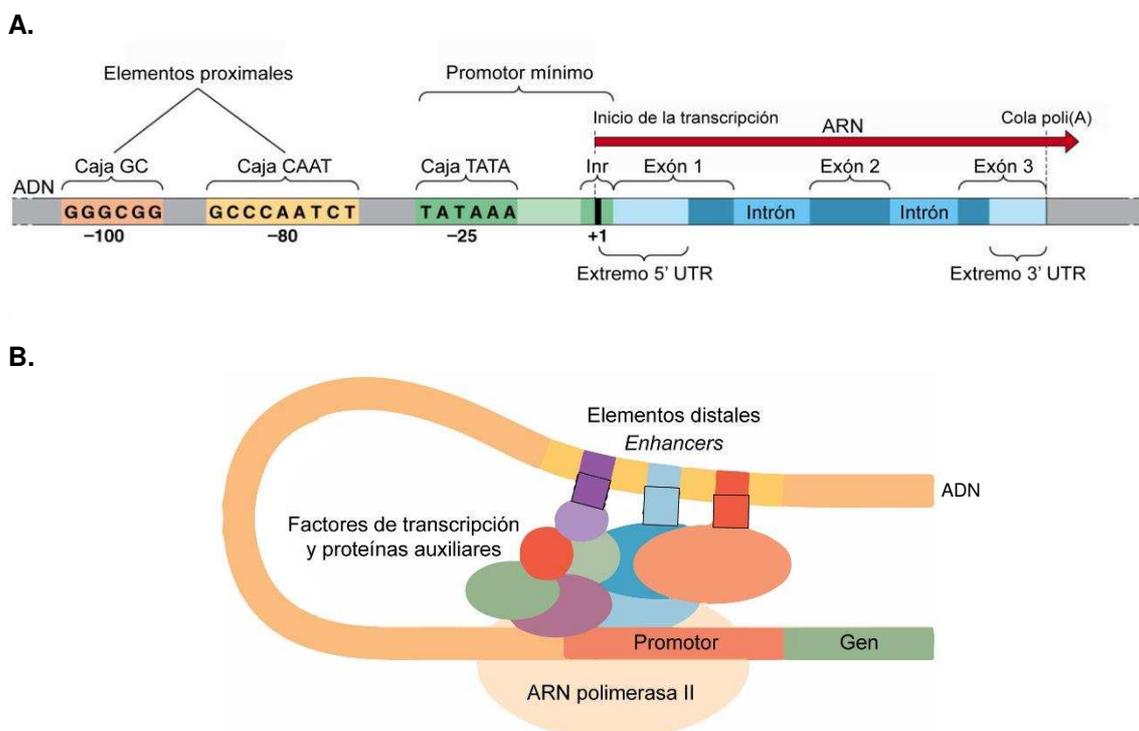


Figura 36. Proceso de transcripción en células de mamíferos.

A. Configuración general de un promotor eucariota. Se muestran sus regiones regulatorias, el sitio de inicio de la transcripción y la estructura de la región codificante formada por exones e intrones, que son procesados posteriormente.

B. Mecanismo de regulación transcripcional a través de elementos distales. Se muestra la acción de potenciadores o *enhancers*, ubicados a mayor distancia *upstream* del promotor mínimo.

Figura modificada de Romanova y col. (2018) [196].

Un ejemplo de región promotora se muestra en la Figura 36.A, la cual se divide en el promotor mínimo (*core promoter*) y los elementos proximales (*proximal promoter*). Estas regiones tienen un papel fundamental en la interacción y el procesamiento de las señales transcripcionales. En mamíferos, el promotor mínimo abarca aproximadamente

unas 80 pb, que puede ir desde los -40 a +40 pb, respecto al sitio de inicio de la transcripción (*transcription start site*, TSS). Se lo define como la secuencia mínima de ADN en dirección 5' de un gen que es suficiente para iniciar la transcripción del mismo. Sus elementos constitutivos más conocidos son la caja TATA y la región iniciadora (Inr). Además, el promotor mínimo está formado por diversos sitios de unión a factores de transcripción, que son secuencias bastante conservadas de unos 3-20 pb de longitud. Por otro lado, la región integrada por elementos proximales se encuentra *upstream* del promotor mínimo (aproximadamente a 250 pb), a los que se unen proteínas y factores de transcripción que facilitan el ensamblaje del complejo de iniciación.

Posteriormente, los ARNs transcritos son modificados. El pre-ARNm sufre un proceso de maduración de "corte y empalme" (*splicing*), en el cual se eliminan los intrones. Durante esta etapa se pueden originar diferentes moléculas de ARNm (*splicing alternativo*), derivadas de distintas combinaciones de exones. Por lo cual, una misma secuencia codificante, puede dar lugar a distintos ARNm y, por lo tanto, producir diferentes proteínas en respuesta a estímulos diversos.

Por otro lado, la actividad de los promotores también está influenciada por otras regiones regulatorias, como los potenciadores (*enhancers*) o silenciadores, que activan o reprimen la expresión génica y están ubicados a mayor distancia del sitio de inicio de la transcripción, constituyendo otro nivel de regulación transcripcional (Figura 36.B).

1.2. Uso de promotores en la industria biotecnológica para la producción de proteínas recombinantes terapéuticas

La generación de clones celulares recombinantes constituye un procedimiento extenso, compuesto de varias etapas. Entre ellas podemos mencionar la elección de la línea celular, el diseño del *cassette* de expresión, la generación de líneas celulares recombinantes y la obtención de clones productores, los cuales deberán ser minuciosamente evaluados a los fines de seleccionar el candidato que ofrezca las mejores características productivas.

Tradicionalmente, la generación de líneas celulares recombinantes estables se ha llevado a cabo mediante transfección, que resulta en una baja frecuencia de integración del TG de interés en el genoma celular. Por ello, la etapa de selección es sumamente importante a los fines de que sobrevivan sólo aquellas células que hayan incorporado el TG a su genoma. Este método consume mucho tiempo –entre 5 y 12 meses– desde la transfección del material genético hasta la obtención de una línea productora [28], ya que la etapa más demandante consiste en aislar y caracterizar un gran número de clones.

En los últimos años, se han desarrollado diversas técnicas para la optimización de la producción de proteínas heterólogas. Estas incluyen la ingeniería metabólica de la línea productora, el estudio y perfeccionamiento de la composición de los medios de crecimiento, la incorporación de elementos regulatorios, como secuencias modificadoras de la cromatina, entre otros. Sin embargo, no se le ha otorgado demasiada importancia al estudio del promotor que se usará, el cual constituye un mecanismo de control fundamental [196].

En todo este contexto, una etapa crucial dentro del proceso productivo corresponde a la elección de un promotor que permita la adecuada expresión del TG de interés [197]. Para poder alcanzar niveles óptimos en la producción de proteínas recombinantes en células de mamíferos, se necesitan niveles altos de expresión.

Podemos clasificar a los promotores en diferentes tipos: constitutivos, inducibles y sintéticos [195]. Los promotores constitutivos naturalmente inducen la expresión en la mayoría de los tejidos del organismo, independientemente de los factores ambientales o del estadio de desarrollo. La principal ventaja de los mismos es que, generalmente, tienen altos niveles de expresión transgénica, por lo que han sido ampliamente utilizados en la industria biotecnológica. Sin embargo, a veces tal sobreexpresión es contraproducente para las células; o bien, no se mantiene en el tiempo, algo que resulta desfavorable para los bioprocesos productivos. Los promotores inducibles, en cambio, pueden ser activados en respuesta a diferentes estímulos, incluyendo señales endógenas y factores físicos y/o químicos externos, bajo condiciones experimentales controladas [195]. La ventaja de estos promotores es en que permiten un control más preciso de la expresión del TG de interés. Pero, la principal desventaja radica en el uso de agentes que pueden resultar tóxicos para las células, como los antibióticos, lo cual además no es deseable a escala industrial.

Actualmente, debido a la escasa información genómica referida a las células CHO, la alternativa más usada es el empleo de promotores constitutivos, generalmente virales, ya que proporcionan niveles elevados necesarios en un proceso productivo [24,198]. Dos de los promotores más empleados en la industria biofarmacéutica son el promotor temprano del virus vacuolado del simio 40 (*Simian vacuolating virus 40*, SV40), aislado en 1960 de células de riñón de *Macaca mulatta* [199]; y el promotor del citomegalovirus (CMV), originario de la familia de patógenos humanos *Herpesviridae*.

No obstante, existen diversas desventajas asociadas al uso de promotores virales, en lugar de promotores endógenos eucariotas. En muchos casos estos imponen a las células un estrés que interrumpe el crecimiento celular, afectan los procesos de

traducción, plegamiento y ensamblaje dentro del RE y, además, se ha demostrado que son propensos al silenciamiento epigenético [200]. Otro problema se relaciona con la dependencia del ciclo celular de estos promotores, siendo muy variable los niveles de expresión de un mismo TG entre diferentes células [201]. Es más, se ha demostrado que la expresión dirigida por CMV aumenta en la fase S del ciclo celular, disminuyendo considerablemente cuando las células están en la fase G0/G1 [202].

Una solución a dichos problemas la constituye el uso de promotores endógenos, que generalmente no producen ninguno de estos efectos no deseados. El primer trabajo llevado a cabo con promotores endógenos humanos se realizó en 1990 [203], donde el promotor del factor de elongación 1 alfa (*elongation factor 1 alpha*, EF1 α) resultó ser apropiado para la expresión de proteínas recombinantes en células de mamíferos. Más tarde, se comenzó a utilizar el EF1 α de hámster chino [204]. Sin embargo, la búsqueda de nuevos promotores endógenos para la expresión de proteínas recombinantes en células CHO, de manera estable y con alta productividad, recién ha comenzado recientemente.

La actividad de un promotor es altamente dependiente del contexto y de los cambios dinámicos en el ambiente extracelular (disponibilidad de nutrientes, pH, temperatura, oxígeno disuelto, etc.), de la maquinaria transcripcional celular (abundancia y estado de activación de los factores de transcripción, disponibilidad de co-reguladores, entre otros) y de la estructura de la cromatina en el sitio de integración del transgén [200]. Estos tres tipos de regulación transcripcional deben tenerse en cuenta al momento de elegir y/o diseñar promotores para la producción de biofármacos.

Principalmente, la función promotora debe coordinarse con el proceso productivo, asegurando que los promotores sean compatibles con las características celulares deseadas, tales como la proliferación y la supervivencia, y con el tiempo de cultivo (por ejemplo con procesos bifásicos en biorreactor). Por estas razones, promotores endógenos han sido investigados para su utilización en procesos productivos. La mayoría de ellos derivan de genes *housekeeping*, ya que presentan una elevada actividad transcripcional. Otros promotores de interés han sido identificados en células CHO [201], entre ellos un promotor que exhibe actividad inducible a temperaturas reducidas. Sin embargo, en comparación con los promotores virales, es muy reducido el conocimiento que se tiene acerca de los promotores celulares de mamíferos, debido a que –como se mencionó anteriormente– la expresión génica eucariota es más compleja y requiere una coordinación celular muy precisa.

La identificación de genes (y de sus promotores) con el nivel y la dinámica de expresión deseada no es una tarea sencilla y se necesita del uso de herramientas como

la transcriptómica, para identificar tales patrones únicos de expresión en condiciones determinadas. El siguiente desafío lo constituye el aislamiento de la secuencia del promotor correspondiente, ya que las regiones regulatorias y los límites de los promotores no suelen ser fácilmente identificables. Además, recientemente se han publicado los genomas de líneas celulares de interés industrial (por ejemplo de células CHO y BHK). Como se detalló previamente, la mayoría de los promotores de eucariotas tienen una región principal (*core promoter*), mientras que otros elementos regulatorios pueden extenderse desde unos cientos a varios miles de pares de bases y comprenden múltiples sitios de unión a factores de transcripción, muchos de los cuales no se conocen aún. Por ello, debido a que no existen límites definidos para los promotores, se emplean fragmentos relativamente largos, de al menos 1 kb [201,205,206].

Algunas investigaciones incluyen el uso de regiones endógenas, como el trabajo de Kang y col. (2016) [207], en el cual emplearon un fragmento genómico de aproximadamente 2900 pb, compuesto por dos secuencias aisladas de dos *loci* diferentes del genoma, las cuales están formadas por diferentes sitios de unión a factores de transcripción. Luego, clonaron este nuevo elemento regulatorio *upstream* del promotor CMV y formaron así nuevos vectores de expresión que mejoraron la expresión de proteínas recombinantes en células CHO [207].

Una opción relativamente nueva para evitar el uso de secuencias virales radica en la construcción de promotores artificiales o sintéticos. Para construirlos se utilizan secuencias conservadas llamadas *cis*-reguladoras o motivos (elementos que contienen sitios de unión a factores de transcripción) derivados de promotores presentes en la naturaleza [208,209]. La bioinformática brinda herramientas que facilitan la identificación de dichos motivos para crear promotores sintéticos específicamente activos en determinadas condiciones, explorando las secuencias reguladoras en las regiones *upstream* de los genes de interés. Diversos trabajos han logrado construir este tipo de promotores empleando diferentes estrategias [197,210-212]. En la mayoría, la actividad transcripcional resultó ser equivalente o superior a la de los promotores comúnmente usados, demostrando el alto impacto de la aplicación de estos nuevos promotores en la industria biotecnológica.

En resumen, la manera más conveniente de estudiar nuevos promotores (endógenos, regiones reguladoras o sintéticos) es utilizar un vector que contenga una proteína fluorescente, sin promotor, y clonar la secuencia aislada en dicho plásmido de manera tal de monitorear la actividad transcripcional del promotor acorde a la dinámica del cultivo. Sin embargo, este enfoque implica la inserción del TG de manera aleatoria en el genoma (en caso de ser expresión estable). Una alternativa sería insertar el TG

de interés en un *locus* específico, bajo la actividad del promotor endógeno identificado. Esta inserción sitio dirigida es posible actualmente gracias a las tecnologías de edición genética como ZFN, TALEN y CRISPR/Cas9.

2. Identificación de sitios *hotspots* y edición genética en líneas celulares de mamífero

La integración del TG en el genoma de una célula es un evento aleatorio y el sitio de integración resulta ser un factor crucial que afectará su estabilidad y su nivel de expresión, siendo la principal causa de la heterogeneidad de expresión entre clones individuales. Numerosos esfuerzos se han realizado con el objetivo de desarrollar nuevas estrategias que permitan llevar a cabo la manipulación de células eucariotas de manera efectiva, confiable y predecible.

En los últimos años han surgido nuevas metodologías que permiten direccionar el gen de interés hacia sitios del genoma previamente caracterizados, en una región activa y estable (generalmente sitios conocidos como “*hotspots*”). Estas estrategias permiten sortear los obstáculos asociados al “efecto de posición” y así alcanzar niveles elevados de expresión de la proteína codificante.

La escasa presencia de dichos sitios en el genoma celular se basa en que sólo el 0,1% del ADN genómico posee sitios transcripcionalmente activos [47,213,214]. Por ello es fundamental desarrollar metodologías que permitan encontrar y reutilizar dichos *hotspots* a través de la introducción de diferentes transgenes que codifiquen para diversas proteínas de interés terapéutico.

La aplicación de estudios transcriptómicos en bioprocesos de cultivos celulares facilita la identificación de potenciales *targets* para la ingeniería celular, con el propósito de optimizar el bioproceso y las características del producto. Para esto, la tecnología CRISPR/Cas9 abre un abanico de posibilidades mediante la edición precisa de células CHO huésped destinadas a la producción de proteína recombinante, mejorando así la productividad y la calidad del producto y reduciendo también los costos de producción.

En los últimos años, ciertas investigaciones se han focalizado en realizar el *knock-out* de uno o varios genes en simultáneo en células CHO con el propósito de mejorar diversas características de los cultivos celulares; por ejemplo, optimizar el procesamiento proteico en el RE [215], reducir los procesos no deseados de apoptosis y fucosilación [216], disminuir la metilación del ADN [217] e incluso acelerar el proceso de adaptación de células CHO-K1 al crecimiento en suspensión [176].

Asimismo, se ha logrado exitosamente insertar genes de interés en *loci* específicos de células CHO [218-220], aunque estos trabajos en los cuales se realizó el

knock-in de una secuencia específica, aún son escasos y recién están comenzando a llevarse a cabo en células CHO. Estas últimas investigaciones demuestran que la integración sitio específica de transgenes, en diversos *hotspots* de células CHO, es una estrategia ideal para superar los desafíos que implica la integración aleatoria convencional del gen de interés.

OBJETIVOS

GENERAL

En esta última etapa de trabajo se propone el empleo de un *set* de promotores endógenos con elevada expresión, previamente identificados mediante RNA-seq en células CHO-K1, cultivadas en alta densidad, en modo de perfusión y con un medio de cultivo libre de suero fetal bovino.

Se plantea utilizar dichos promotores para la construcción de nuevos vectores de expresión destinados a la producción de proteínas recombinantes de uso en salud humana y animal. Con esta estrategia se pretende aprovechar la maquinaria transcripcional regulatoria endógena disponible en las condiciones de un bioproceso de alta densidad celular con un medio de cultivo definido. Así, se busca lograr una expresión óptima de transgenes codificantes de proteínas terapéuticas en un contexto normal de producción industrial.

Los mismos serían una alternativa al empleo de los promotores habitualmente usados, principalmente promotores virales, con el fin de incrementar la estabilidad genética y mejorar la eficiencia del proceso de obtención de las proteínas recombinantes desarrolladas.

ESPECÍFICOS

1. Examinar particularmente aquellos genes que exhiban elevada actividad transcripcional en condiciones de cultivo en alta densidad celular.
2. Utilizar las secuencias promotoras de ese grupo de genes con una alta expresión, para la construcción de nuevos vectores de expresión destinados a la producción de proteínas recombinantes de uso en salud humana y animal.
3. Evaluar y contrastar la actividad transcripcional de los vectores de expresión creados, empleando un gen reportero como la proteína verde fluorescente ZsGreen1, con la actividad de aquellos promotores usados habitualmente para producir proteínas recombinantes terapéuticas (por ejemplo, CMV).

MATERIALES Y MÉTODOS

1. Técnicas de biología molecular

1.1. Líneas celulares procariotas

Escherichia coli TOP 10 (Invitrogen). Se empleó para el aislamiento y la amplificación de diferentes plásmidos.

1.2. Medios de cultivo para células procariotas

La esterilización de los medios de cultivo para bacterias se realizó en autoclave durante 15 min a 1 atmósfera y 120°C. A continuación se detallan los medios de cultivo empleados:

1.2.1. Medio LB (*Luria Bertani*)

Se empleó para el crecimiento bacteriano en medio líquido (pH 7,2):

| | |
|----------------------|--------------------------------------|
| Triptona | 10 g.l ⁻¹ (Sigma, Japón). |
| Extracto de levadura | 5 g.l ⁻¹ (BD, EE.UU.). |
| NaCl | 10 g.l ⁻¹ (Sigma). |

1.2.2. Medio LB-Antibiótico

Se empleó para el crecimiento de clones bacterianos resistentes a antibiótico:

Medio LB.

Antibióticos:

- Ampicilina 100 µg.ml⁻¹ (Bagó, Argentina).
- Kanamicina 100 µg.ml⁻¹ (Ivax, Argentina).

1.2.3. Agar LB

Medio LB.

Agar al 2% (p/v) (Sigma).

1.2.4. Agar LB-Antibiótico

Se empleó para la selección de clones bacterianos:

Agar LB

Antibiótico:

- Ampicilina 100 µg.ml⁻¹ (Bagó).
- Kanamicina 100 µg.ml⁻¹ (Ivax, Argentina).

El antibiótico se suplementó al medio estéril fundido una vez que la temperatura alcanzó aproximadamente los 50°C, ya que el mismo es termolábil. Luego de

homogeneizar, se procedió a adicionar 10 ml del mismo en placas de Petri estériles de 100 mm de diámetro (Greiner Bio-One, Austria).

1.3. Purificación de ADN plasmídico

Los plásmidos usados se purificaron a partir de cultivos de bacterias transformadas en medio líquido (Medio LB-Antibiótico). Todos los métodos de purificación se basan en el tamaño relativamente pequeño de los plásmidos y su naturaleza circular cerrada covalentemente.

1.3.1. Minipreparación de ADN plasmídico

Se utilizó el *kit* comercial *Wizard Plus SV Minipreps DNA purification system* (Promega, EE.UU.) siguiendo las recomendaciones del fabricante, como se describe a continuación:

1. A partir de colonias individuales de cultivos en medio Agar-LB-Antibiótico inocular frascos conteniendo 10 ml de medio LB-Antibiótico y cultivar durante 16-20 h con agitación a 180 rpm y a 37°C.
2. Centrifugar a 2500 *g* durante 10 min 1,5 ml del cultivo en un tubo Eppendorf. Descartar el sobrenadante y repetir este procedimiento con otros 1,5 ml de cultivo.
3. Resuspender el *pellet* celular con 250 μ l de solución de resuspensión y adicionar 250 μ l de solución de lisis celular. Invertir el tubo 4 veces.
4. Inmediatamente, adicionar 10 μ l de solución de proteasa alcalina, invirtiendo el tubo 4 veces e incubando a temperatura ambiente 5 min.
5. Agregar 350 μ l de solución de neutralización e invertir el tubo varias veces. Centrifugar a 5000 *g* por 10 min.
6. Tomar el sobrenadante (aproximadamente 850 μ l) y transferir a una columna de purificación, colocada dentro de un tubo colector. Centrifugar a 5000 *g* durante 1 min y descartar el FT.
7. Agregar a la columna 750 μ l de solución de lavado, centrifugar 1 min a 5000 *g* y repetir el lavado con 250 μ l, centrifugando 2 min.
8. Posteriormente, transferir la columna a un tubo Eppendorf nuevo y eluir el ADN plasmídico agregando sobre la resina de la columna 100 μ l de agua Milli Q autoclavada. Dejar reposar un momento y centrifugar 1 min a 5000 *g*. Determinar la concentración y pureza del ADN plasmídico obtenido.

Por último, se evaluó la integridad de los plásmidos purificados mediante electroforesis en gel de agarosa, como se explica en el apartado 1.6 de esta Sección.

1.3.2. Maxipreparación de ADN plasmídico

Con el propósito de obtener cantidades suficientes de los vectores de expresión para su empleo en, por ejemplo, el ensamblado de partículas lentivirales, se realizaron

preparaciones de ADN plasmídico en gran escala (maxiprep), tal como se describe a continuación:

1. Inocular 10 ml de medio de cultivo LB-Antibiótico con una colonia proveniente de medio LB-Agar-Antibiótico e incubar a 37°C o 30°C (para vectores lentivirales) a 180 rpm durante 16 h.
2. Adicionar a este precultivo 90 ml de LB-Antibiótico y cultivar durante otras 8 h en las mismas condiciones.
3. Inocular con este precultivo 200 ml de LB-Antibiótico y cultivar durante 16-20 h adicionales en las mismas condiciones.
4. Centrifugar los cultivos a 2500 *g* durante 10 min a 4°C.
5. Resuspender el *pellet* en 10 ml de TEG estéril (Tris-HCl 25 mM, EDTA 10 mM; glucosa 1% (p/v), pH 8,3), transferir la suspensión celular a un tubo cónico de 50 ml y mantener en hielo durante 5 min.
6. Agregar 20 ml de NaOH 0,2 M, SDS 1% (p/v), y homogeneizar adecuadamente para lograr la lisis celular en hielo durante 5 min.
7. Adicionar 15 ml de acetato de potasio 5 M (pH 5,2) a 4°C (solución de neutralización), homogeneizar, agregar 5 ml de CaCl₂ 6,6 M; homogeneizar e incubar en hielo durante 10 min.
8. Centrifugar 10 min a 4°C a 2500 *g*.
9. Filtrar el sobrenadante con papel de filtro, en tubos cónicos de 50 ml. Precipitar con 0,6 volúmenes de isopropanol (Baker, EE.UU.), incubando 20 min a -70°C.
10. Centrifugar 20 min a 2500 *g*, a 4°C y eliminar el sobrenadante.
11. Resuspender el *pellet* en 1 ml de agua Milli Q estéril, y transferir de a 500 µl en dos tubos Eppendorf nuevos.
12. Agregar 250 µl de fenol y luego 250 µl de cloroformo a cada tubo, agitar y centrifugar la mezcla 2 min a 13000 *g*. Tomar la fase acuosa superior de ambos tubos y juntar en un único tubo.
13. Precipitar con 0,6 volumen de isopropanol, agitar y mantener 20 min a -70°C. Centrifugar 15 min a 4°C a 13000 *g* y lavar el *pellet* con 300 µl de etanol 80% (v/v) en agua Milli Q estéril, centrifugando 5 min.
14. Dejar secar a 37°C en baño seco termostatzado.
15. Disolver el *pellet* en 500 µl de TE (Tris-HCl 25 mM, NaCl 100 mM, pH 8,3) 0,5 X conteniendo 100 µg.ml⁻¹ de RNasa A (Promega) e incubar 60 min a 37°C, para eliminar restos de ARN.
16. Agregar 250 µl de fenol y luego 250 µl de cloroformo a cada tubo, agitar y centrifugar la mezcla 2 min a 13000 *g*. Tomar la fase acuosa superior y transferir a un nuevo tubo.
17. Agregar a la fase acuosa 50 µl de NaAc 3 M (pH 7,0) y 0,6 volumen de isopropanol. Agitar e incubar 15 min a -70°C, de modo de precipitar el ADN plasmídico. Centrifugar 15 min a 13000 *g* a 4°C. Descartar sobrenadante.
18. Lavar el *pellet* generado con 300 µl de etanol 80% (v/v) en agua Milli Q estéril. Centrifugar 5 min a 13000 *g* a 4°C. Descartar sobrenadante.
19. Dejar secar a 37°C en baño seco termostatzado.
20. Disolver el *pellet* en 300 µl de agua Milli Q estéril. Determinar la concentración y pureza del ADN plasmídico obtenido.

Por último, se evaluó la integridad de los plásmidos purificados mediante electroforesis en gel de agarosa, como se explica en el apartado 1.6 de esta Sección.

1.4. Determinación de la concentración y pureza del ADN

Se realizó según lo detallado previamente en: Capítulo 2 / Materiales y Métodos / Sección 1.2.

1.5. Digestiones enzimáticas

Las digestiones con endonucleasas de restricción (Promega) se realizaron respetando los medios y condiciones recomendadas por el fabricante. En todos los casos, la reacción se llevó a cabo durante 2 h a 37°C, empleándose de 1-5 U de enzima por cada 1 µg de ADN, en un volumen final de 20 µl.

Las digestiones con la enzima de restricción *BsmBI* (o *Esp3I*) (Thermo Fisher Scientific) para digerir el vector lentiCRISPRv2 (Addgene #52961) se llevaron a cabo según las condiciones indicadas por el fabricante:

| | |
|------------------------------|--------|
| Agua Milli Q estéril | 11 µl. |
| 10X Buffer Tango | 2 µl |
| ADN (1 µg.µl ⁻¹) | 5 µl |
| DTT 10 mM | 1 µl |
| <i>BsmBI</i> | 1 µl |

La reacción se llevó a cabo durante 2 h a 37°C, en un volumen final de 20 µl.

Luego de las digestiones enzimáticas, los fragmentos se separaron mediante electroforesis en geles de agarosa para su posterior análisis y/o purificación.

1.6. Electroforesis de ADN en geles de agarosa

Las separaciones de ADN en geles de agarosa se realizaron empleando el sistema submarino. Para la preparación de los geles, se pesó la cantidad adecuada de agarosa (Biodynamics SRL, Argentina) y se disolvió en solución tampón TAE (Tris-acetato 40 mM; EDTA 1mM; pH 8) en horno microondas hasta su completa solubilización.

La solución se enfrió a 60°C y se adicionó el colorante *SYBR Safe DNA gel stain* (Invitrogen) en una dilución 1/10.000. Esta solución se colocó en un soporte adecuado y se le introdujo un peine de 1 mm de espesor para formar los carriles de siembra, el cual se retiró una vez solidificado. El gel se colocó en una cuba electroforética y se cubrió completamente con solución TAE.

Las muestras de ADN se mezclaron con solución de siembra 10X [Glicerol (Cicarelli) 30% (v/v); azul de bromofenol (Sigma) 0,25% (p/v)] y se sembraron en cada una de las calles del gel. Se tapó la cuba y se aplicó un voltaje entre 80-100 V durante 30 min o hasta que el frente de corrida se desplace la distancia requerida.

Finalmente, el gel fue examinado utilizando un transiluminador Safe Imager (Invitrogen) y fotografiado. El fundamento de la detección de los fragmentos de ADN se basa en que el colorante SYBR es un agente intercalante que se introduce en la estructura secundaria de la doble hélice del ADN y se acopla energéticamente a los ácidos nucleicos que lo forman, de manera que su tasa de emisión fluorescente se incrementa notablemente. El complejo resultante ADN/SYBR presenta un pico de excitación a $\lambda = 498$ nm y un pico de emisión a $\lambda = 522$ nm.

Para estimar la longitud de los fragmentos de ADN separados se emplearon los marcadores de masa molecular (según los requerimientos de cada experimento): 1Kb *DNA Ladder*, *Trans2K[®] Plus II DNA Marker* y 100bp Plus *DNA Ladder* (TransGen Biotech, China).

1.7. Precipitación de ADN

Los fragmentos de ADN obtenidos luego de la digestión con enzimas de restricción fueron aislados y concentrados, en determinados casos, mediante precipitación con etanol absoluto. Esto se efectuó con el fin de eliminar los fragmentos cortos de ADN liberados, las sales y las enzimas de restricción.

Para esto, la muestra se llevó a volumen final de 200 μ l con agua autoclavada adicionando un volumen de NaAc 3 M equivalente al 10% del volumen de la muestra y se agitó. Se agregaron entre 2 y 3 volúmenes de etanol absoluto y la mezcla se incubó a -70°C durante 20 min. Luego, se centrifugó a 13000 *g* durante 15 min a 4°C y se lavó con 200 μ l de etanol 70% (v/v) frío. Las muestras se centrifugaron nuevamente a 13000 *g* durante 7 min a 4°C y se secaron en baño seco termostatzado a 37°C durante 10 min. El *pellet* de ADN se resuspendió en un volumen final de 50 μ l en agua Milli Q autoclavada.

1.8. Purificación de ADN a partir de geles de agarosa

Los fragmentos de ADN separados por electroforesis en geles de agarosa fueron purificados utilizando el *kit* comercial *QIAquick Gel Extration Kit* (QIAGEN), siguiendo las recomendaciones del fabricante, como se describe a continuación:

1. Separar con bisturí la fracción de gel conteniendo la banda de interés y colocar dentro de un tubo Eppendorf. Agregar 3 volúmenes de solución QG por cada volumen de gel (100 mg de gel equivalen aproximadamente a 100 µl).
2. Incubar durante 10 min en baño de agua a 50°C, mezclando por inversión cada 2-3 min.
3. Transferir la muestra a una columna de purificación (*QIAquick spin column*) dentro de su respectivo tubo colector y centrifugar durante 1 min a 13000 *g*.
4. Descartar el sobrenadante y agregar 0,5 ml de solución QG a la columna, volviendo a centrifugar de la misma manera.
5. Para lavar, agregar 0,75 ml de solución PE centrifugando durante 1 min a 13000 *g*. Descartar el sobrenadante y centrifugar de la misma manera, para descartar toda la solución PE residual.
6. Transferir la columna a otro tubo Eppendorf y eluir el ADN con 50 µl de agua Milli Q autoclavada.

Finalmente, se analizó una alícuota mediante electroforesis en gel de agarosa y se determinó la pureza y la concentración de los fragmentos mediante lectura espectrofotométrica.

1.9. Reacciones de ligación de ADN

Para las reacciones de ligación se utilizó una determinada relación inserto/plásmido en un volumen de reacción de 20 µl, utilizando la enzima T4 ADN Ligasa (Invitrogen), según el protocolo indicado por el proveedor. La mezcla de reacción se incubó ON (*overnight*, toda la noche) a 4°C. El control de plásmido religado se llevó a cabo reemplazando el volumen correspondiente al inserto por agua Milli Q autoclavada.

Luego de la reacción de ligación, se utilizó la mezcla de reacción obtenida para transformar células *E. coli* competentes.

1.10. Preparación de células *E. coli* TOP 10 competentes, por el método químico

A partir de un banco celular de *E. coli* TOP 10 conservado a -70°C:

1. Realizar una estría sobre una placa de Agar LB e incubar durante 16-20 h a 37°C.
2. Escoger una colonia e inocular con ella 5 ml de caldo LB (cultivo *starter*), incubando a 37°C durante 18 h con agitación a 180 rpm.
3. Realizar un subcultivo tomando 500 µl del cultivo *starter* a 50 ml de medio LB y cultivar durante 3 h a 37°C y 180 rpm o hasta lograr una DO de 0,6.
4. Incubar el cultivo en baño de hielo durante 10 min y luego recuperar por centrifugación a 2500 *g* por 10 min a 4°C. A partir de este punto todos los procedimientos se efectúan en baño de hielo.
5. Descartar el sobrenadante y resuspender el *pellet* celular en 500 µl de medio LB frío.

6. Agregar 2,5 ml de *storage solution* [medio LB conteniendo DMSO 5% (v/v); PEG 10% (p/v) (MW 8000, Sigma); MgCl₂ 1% (p/v), esterilizado por filtración con membrana de 0,22 µm] y mezclar completamente mediante agitación suave.
7. Finalmente, dispersar alícuotas de 100 µl en criotubos (Greiner Bio-One) fríos, y conservarlos inmediatamente a -70°C.
8. Realizar un control de las células competentes preparadas, para asegurar la continuidad en la esterilidad del cultivo y estimar la eficiencia de transformación. Para esto, descongelar 2 criotubos y transformarlos con 100 ng de un vector plasmídico conocido (control positivo) y sin ADN (control negativo). Las células competentes serán utilizadas sólo cuando la placa del control negativo no muestre desarrollo de colonias y cuando se obtenga buena eficiencia de transformación en el control positivo.

1.11. Transformación de células competentes

En el momento de la transformación se agregó el ADN a una alícuota de células competentes recién descongeladas y se incubó 30 min en baño de hielo. Luego, las células fueron colocadas 1,5 min a 42°C, a fin de aumentar la eficiencia de transformación, y se incubó inmediatamente en baño de hielo durante 2 min. A cada tubo se le agregaron 700 µl de medio LB y se incubó 60 min a 37°C. Transcurrida la hora de recuperación, se centrifugó a 2500 g durante 5 min, se descartaron 600 µl de sobrenadante y el *pellet* fue resuspendido en el volumen remanente de medio de cultivo. La totalidad de la suspensión celular se distribuyó homogéneamente sobre una placa conteniendo agar LB-Antibiótico. La placa se incubó ON a 37°C.

1.12. Criopreservación de cepas bacterianas

Se cultivaron durante 18 h a 37°C en medio LB-Antibiótico (en el caso de bacterias recombinantes). Posteriormente, se tomaron 800 µl del cultivo fresco y se trasvasaron a un criovial (Greiner Bio-One). Se adicionaron 200 µl de glicerol 50% (v/v) estéril (Cicarelli), homogeneizando para asegurar la distribución del glicerol.

Los criotubos se conservaron a -70°C.

1.13. Revitalización de cepas bacterianas

Se tomó una pequeña porción de material criopreservado con un ansa o *tip* desde la superficie del cultivo celular congelado y se inoculó en medio LB, en caso de bacterias no transformadas, o en medio LB antibiótico, para bacterias recombinantes.

El criotubo se conservó nuevamente a -70°C y el tubo inoculado se incubó a 37°C durante 16 h.

1.14. Extracción de ADN genómico

La purificación de ADN genómico se realizó a partir de cultivos de células CHO-K1 en suspensión, siguiendo el protocolo que se detalla a continuación:

1. Cosechar un mínimo de 10×10^7 células totales, a partir de cultivos CHO-K1 en suspensión.
2. Resuspender el *pellet* en PBS frío. Centrifugar 2 min a 166 *g*.
3. Resuspender las células en PBS frío a razón de una concentración aproximada de 5×10^7 cél.ml⁻¹.
4. Tomar un volumen de la suspensión celular y agregar 0,5 volumen de *Buffer* de Lisis (EDTA 20 mM, Tris 10 Mm pH 8, NaCl 200 mM, Triton X-100 0,2% (v/v), Proteinasa K 100 µg/ml). Incubar 40 min a 4°C.
5. Centrifugar a 13000 *g* durante 5 min a 4°C.
6. Tomar el sobrenadante, dividirlo en tres tubos Eppendorf y agregar en cada uno un volumen de isopropanol y NaAc a una concentración final de 0,3 M. Incubar 20 min a -70°C.
7. Centrifugar 15 min a 13700 *g* a 4°C. Resuspender el *pellet* en 100 µl de LiCl 2M y juntar las mismas muestras en un tubo. Mantener a 4°C durante 30 min. Centrifugar 15 min a 13000 *g* a 4°C.
8. Tomar el sobrenadante y realizar una extracción fenol/cloroformo (0,5/0,5 v/v). Centrifugar durante 2 min a 13000 *g*.
9. Tomar la fase acuosa y agregar un volumen de isopropanol y NaAc a una concentración final de 0,3 M. Incubar durante 20 min a -70°C.
10. Centrifugar 15 min a 13000 *g* a 4°C. Lavar el *pellet* generado con 300 µl de etanol 80% (v/v) en agua Milli Q estéril. Centrifugar 2 min a 13000 *g* a 4°C. Descartar sobrenadante. Dejar secar a 37°C en baño seco termostatizado.
11. Disolver el *pellet* en 300 µl de agua Milli Q estéril y *RNAsa* a una concentración final de 100 µg/ml. Incubar a 37°C durante 1 hora.
12. Agregar un volumen de isopropanol y NaAc a una concentración final de 0,3 M. Mantener a -70°C durante 20 min.
13. Repetir el paso 10.
14. Resuspender en 100 µl de agua Milli Q estéril.

Se determinó la concentración y pureza del ADN genómico purificado y se evaluó la integridad del mismo mediante electroforesis en gel de agarosa, como se detalló en previamente.

1.15. Purificación de ARN total a partir de cultivos de células CHO-K1 en suspensión

Se realizó según lo detallado previamente en: Capítulo 2 / Materiales y Métodos / Sección 1.1.

1.16. Reacciones de PCR

1.16.1. Amplificación de los promotores endógenos y de las secuencias de los brazos necesarios para recombinación homóloga (CRISPR), a partir de ADN genómico de células CHO-K1 en suspensión

Para estas reacciones de PCR se utilizó la enzima *TransTaq DNA Polymerase High Fidelity* - HiFi (TransGen Biotech), cuya fidelidad es 18 veces superior a la polimerasa *EasyTaq*. La mezcla de reacción se preparó siguiendo las instrucciones del fabricante, como se detalla a continuación:

| Reactivo | Concentración final | Volumen (por tubo de reacción) |
|---|---------------------|--------------------------------|
| <i>TransTaq® HiFi Buffer I 10X</i> | 1 X | 5,0 µl |
| dNTPs 2,5 mM | 0,2 mM | 4,0 µl |
| <i>Primer Forward</i> 10 µM | 0,2 µM | 1,0 µl |
| <i>Primer Reverse</i> 10 µM | 0,2 µM | 1,0 µl |
| Templado | 150 ng | Variable |
| <i>TransTaq® HiFi DNA Polymerase</i> | 1,25-2,5 U | 1,0 µl |
| Agua | - | csp 50 µl |

Las reacciones se llevaron a cabo en un termociclador (Biometra, Analytik Jena, Alemania) de acuerdo con el siguiente programa:

Paso 1: Desnaturalización inicial: 94°C por 3 min

Paso 2: Desnaturalización: 94°C por 0,5 min

Paso 3: *Annealing* (según la temperatura indicada más abajo) por 0,5 min

Paso 4: Extensión 72°C por 1 min

Volver al paso 2: 30 veces

Paso 5: Extensión 72°C por 10 min

Paso 6: Conservación a 4°C

La temperatura de *annealing* se modificó según el fragmento de ADN a amplificar:

| Secuencia amplificada* | Primers [†] | Secuencia (sentido 5'-3') | Temp. de annealing |
|------------------------|----------------------------------|--|--------------------|
| Promotor S100a6 | Prom_S100a6_Fw Prom_S100a6_Rv | CCCGAGCTCGATCTTCCCTGGAGCTGACCTTG CCCGGTACCTGAGCACCCCAGAAAAGCAGACT | 67°C |
| Promotor Fth1 | Prom_Fth1_Fw Prom_Fth1_Rv | CCCGAGCTCGATGGCCTAGCTAGATGCAGTGG CCCGGTACCTGAGGTTCCGTTCAAGCACTGTT | 67°C |
| Promotor Lgals1 | Prom_Lgals1_Fw Prom_Lgals1_Rv | CCCGAGCTCGATCCCGGTCATTCAACACCCT CCCGGTACCTGAGATGTCAGGAACGGGGTGG | 68°C |

| | | | |
|----------------------------|--------------------|-----------------------------------|------|
| Promotor | Prom_Vim_Fw | CCCGAGCTCGATATGTTTGCGGTCTGGGAGG | 68°C |
| Vim | Prom_Vim_Rv | CCCGGTACCTGACAAGAGTGGCAGAGGACCG | |
| Promotor | Prom_Actb_Fw | CCCGAGCTCGATATCGGGCAGTGTCCACAAG | 68°C |
| Actb | Prom_Actb_Rv | CCCGGTACCTGACCGCGCTCCAAGTCTCAAT | |
| Promotor | Prom_Rps18_Fw | CCCGAGCTCGATGGGCAGACATAGGGAGTGC | 68°C |
| Rps18 | Prom_Rps18_Rv | CCCGGTACCTGACATATCTCCGGCCCACACC | |
| Promotor | Prom_Cd63_Fw | CCCGAGCTCGATGCTTCCTCCGTTTGCTCA | 68°C |
| Cd63 | Prom_Cd63_Rv | CCCGGTACCTGACGGAAAGCACGGGGATGAG | |
| Promotor | Prom_Calr_Fw | CCCGAGCTCGATACCGTACCCCTAGTCCTCTG | 68°C |
| Calr | Prom_Calr_Rv | CCCGGTACCTGAGCACGGGCCCTTTAAAACG | |
| Promotor | Prom_Ybx1_Fw | CCCGAGCTCGATAAAGCCTGAGCCATTAAGGAG | 64°C |
| Ybx1 | Prom_Ybx1_Rv | CCCGGTACCTGAAAACCTTCGTTGCTGGAAAA | |
| Promotor | Prom_Hspa5_Fw | CCCGAGCTCGATCGGGAACATTATGGGGCGA | 68°C |
| Hspa5 | Prom_Hspa5_Rv | CCCGGTACCTGAGGTCTCCGGTCACTGTTG | |
| Promotor | Prom_Hmox1_Fw | CCCGAGCTCGATTAGAAGGGAGGCCTGGCAT | 68°C |
| Hmox1 | Prom_Hmox1_Rv | CCCGGTACCTGACCGAGCCAGCCCTTTAAGT | |
| Promotor | Prom_Ctsz_Fw | CCCGAGCTCGATCCCCAGATTTGCCTCCGAG | 69°C |
| Ctsz | Prom_Ctsz_Rv | CCCGGTACCTGACGACCCTTTGACCTGACCC | |
| Promotor | Prom_Lamp1_Fw | CCCGAGCTCGATCAGCATGCCACCTATTGCT | 67°C |
| Lamp1 | Prom_Lamp1_Rv | CCCGGTACCTGAAAACCTGGATCCCACAGAA | |
| S100a6 HA izquierdo | S100a6_HA_Left_Fw | CCCATCGATGGATGGAGCAGCCTCCATCTC | 67°C |
| | S100a6_HA_Left_Rv | CCCCATATGGCTGCCTCCAAAGGGCATCTC | |
| S100a6 HA derecho | S100a6_HA_Right_Fw | CCCGATCCTGGTGGCTCCCCACTGATTTT | 65°C |
| | S100a6_HA_Right_Rv | CCCGAATTCTGCTCCGTGACACCCAGACAA | |
| Fth1 HA izquierdo | Fth1_HA_Left_Fw | CCCATCGATGGACGGGATGATGTGGCTCTG | 67°C |
| | Fth1_HA_Left_Rv | CCCCATATGGCTGACATCACTGGTGGCTTCG | |
| Fth1 HA derecho | Fth1_HA_Right_Fw | CCCGATCCTGGTGGCTGCCGTTACCTTTTC | 65°C |
| | Fth1_HA_Right_Rv | CCCGAATTCTGCCAGGAATTGGCCATGAG | |
| Lgals1 HA izquierdo | Lgals1_HA_Left_Fw | CCCATCGATGGAGGGAGAGAGATCCCTTCAGC | 67°C |
| | Lgals1_HA_Left_Rv | CCCCATATGGCTGGGGCTTGTGACTTCACTC | |
| Lgals1 HA derecho | Lgals1_HA_Right_Fw | CCCGATCCTGGCTGCCTCTGTTCCCCACA | 63°C |
| | Lgals1_HA_Right_Rv | CCCGAATTCTGCCCCATGCACAAATGAAAA | |
| Vim HA izquierdo | Vim_HA_Left_Fw | CCCATCGATGGACCAGAGATGGACAGGTTGGT | 67°C |
| | Vim_HA_Left_Rv | CCCGCTAGCACGCACGGAGTGTGTGCAATTTT | |
| Vim HA derecho | Vim_HA_Right_Fw | CCCGATCCTGGCAACGCAGTACCAGCAAGAA | 67°C |
| | Vim_HA_Right_Rv | CCCGAATTCTGCCCTCCCTGTGCTGAAATC | |

* HA significa brazo homólogo (*homology arm*).

† *Left y Right*, hacen referencia a los *primers* para amplificar los brazos homólogos izquierdo y derecho, respectivamente.

Cada reacción se llevó a cabo por duplicado y luego de amplificar los fragmentos, se sembró cada producto de reacción en un gel de agarosa (Sección 1.6) y se purificó del gel (apartado 1.8).

En el diseño de los *primers* (realizado con Primer3 [194] y primer-BLAST [193]) para la amplificación de los promotores endógenos se incorporaron sitios de reconocimiento para enzimas de restricción (*SacI* en el *primer forward* y *KpnI* en el *primer reverse*), lo cual permitió clonar las secuencias en el vector pZsGreen1-1 (Clontech, EE.UU.).

De igual manera, los *primers* para amplificar los brazos homólogos fueron diseñados de tal manera que contienen sitios de corte para las enzimas de restricción: *Clal* y *NdeI* para los HA izquierdos de S100a6, Fth1 y Lgals1, y *Clal* y *NheI* para Vim, en los extremos 5' y 3' respectivamente; y *BamHI* y *EcoRI* para los HA derechos, en los extremos 5' y 3' respectivamente. De esta manera se pueden clonar dentro del *cassette* de *knock-in* diseñado para los experimentos con CRISPR.

1.16.2. PCR de colonias (Colony PCR)

Se utilizó esta técnica para determinar si los diferentes clonados que se llevaron a cabo resultaron correctos, determinando (con *primers* específicos para cada inserto) la presencia o ausencia del inserto.

Para el caso de los clonados de los promotores endógenos en el vector pZsGreen1-1, el tamaño de los fragmentos amplificados fue de aproximadamente 1000 pb, ya que se usaron los mismos *primers* empleados para amplificar cada inserto (promotor). Se aislaron, para cada clonado, un número variable de colonias transformadas y se resuspendieron en 10 µl de Agua Milli Q autoclavada.

Para los clonados de los sgRNA en el vector lentiCRISPRv2, el tamaño de cada fragmento que se amplificó fue de 1500 pb. Se aislaron entre 2 y 7 colonias transformadas con cada construcción plasmídica y se resuspendieron en 10 µl de Agua Milli Q autoclavada.

Para las reacciones de PCR se utilizó la enzima *EasyTaq DNA Polymerase* (TransGen Biotech), preparando la mezcla de reacción de acuerdo con las instrucciones del fabricante:

| Reactivo | Concentración final | Volumen (por tubo de reacción) |
|-------------------------------|---------------------|-----------------------------------|
| EasyTaq Buffer 10X | 1 X | 5,0 µl |
| dNTPs 2,5 mM | 0,2 mM | 4,0 µl |
| Primer Forward 10 µM | 0,2 µM | 1,0 µl |
| Primer Reverse 10 µM | 0,2 µM | 1,0 µl |
| Templado | X | 3,0 µl |
| EasyTaq DNA Polymerase | 2,5-5 U | 0,5 µl |
| Agua | - | csp 50 µl |

Además, se realizó un control negativo sin el agregado de muestra y un control positivo (para el caso de los promotores con ADN genómico como templado). Las reacciones se llevaron a cabo en un termociclador (Biometra, Analytik Jena) de acuerdo con el siguiente programa:

Paso 1: Desnaturalización inicial: 94°C por 3 min

Paso 2: Desnaturalización: 94°C por 0,5 min

Paso 3: *Annealing* 57°C por 0,5 min

Paso 4: Extensión 72°C por 1,5 min

Volver al paso 2: 30 veces

Paso 5: Extensión 72°C por 10 min

Paso 6: Conservación a 4°C

Siendo las secuencias de cada oligonucleótido (*primer*), las siguientes:

| Primer | Secuencia (sentido 5'-3') |
|-------------------------|----------------------------------|
| Primer Forward | ACAGCAAGCGCCGCTGATCT |
| Primers Reverse: | |
| S100a6_1_Rv | AAACGTGGTGGGTAATTGTACAATC |
| Fth1_1_Rv | AAACACATGCATACACTTCCTCAGC |
| Lgals1_Rv | AAACCTAGGCCTCAATAAAGGCAGC |
| Vim_Rv | AAACTGCAACAACGCAGTACCAGC |
| EGFP2_Rv | AAACCGGTGAACAGCTCCTCGCCC |

Los productos amplificados se evaluaron mediante electroforesis en gel de agarosa.

1.16.3. Reacciones de RT-PCR

Para la detección de ARN mensajero en las células CHO-K1 adaptadas al crecimiento en suspensión se llevó a cabo un protocolo de extracción de ARN total, para luego realizar una reacción de retrotranscripción, obteniendo así el ADN copia que se utilizó luego como templado para las reacciones de PCR específicas.

1.16.3.1. Reacción de retrotranscripción (RT)

Para obtener el ADN copia (ADNc) a partir del ARN purificado se utilizó la enzima *M-MLV Reverse Transcriptase* (Promega) y un oligonucleótido específico para la cola de poli(A) presentes en los ARN mensajeros.

1. Agregar 2 µg de ARN purificado en un volumen final de 15 µl de agua autoclavada.
2. Desnaturalizar la muestra, calentando 5 min a 65°C y llevando rápidamente a baño de hielo.
3. Preparar la muestra de reacción agregando los siguientes componentes:

| Componente | Volumen (por tubo de reacción) |
|------------------------------------|-----------------------------------|
| <i>M-MLV Buffer 5X</i> | 6,0 µl |
| dNTPs mix 10mM | 2,0 µl |
| Oligo dT (10 µM) | 1,5 µl |
| Templado ARN | 15 µl (del paso 2) |
| <i>M-MLV Reverse Transcriptase</i> | 1,0 µl |
| Agua | csp 30 µl |

4. Programa de ciclado:
Paso 1: 22°C por 15 min
Paso 2: 30°C por 15 min
Paso 3: 37°C por 15 min
Paso 4: 42°C por 90 min
Paso 5: 80°C por 5 min
Paso 6: 94°C por 30 s
5. Conservar a -20°C hasta su utilización.

1.16.3.2. Reacción de qPCR a partir del ADNc

Se contrastó la expresión génica de 4 genes identificados en la etapa anterior mediante RNA-seq (son los 4 genes con los mayores niveles de expresión), contra la expresión de un transgén (α -Galactosidasa A) incorporado en un clon productor de

células CHO-K1 adaptado al crecimiento en suspensión, generado en nuestro laboratorio.

La síntesis de ADNc se llevó a cabo a partir de las muestras de ARN purificadas de un día de cultivo en biorreactor de células CHO-K1 en fase estacionaria (empleado en los análisis de RNA-seq - Capítulo 2) y un día de cultivo en alta densidad del clon productor antes mencionado, según el protocolo detallado en la Sección 3.3 de este apartado.

Se utilizaron los *primers* diseñados para los genes con mayor expresión: S100a6, Fth1, Lgals1 y Vim; los *primers* para el gen de referencia Eif3i (Capítulo 2 – Materiales y Métodos – Sección 4.8); y un *set* de *primers* para una región contenida dentro de la secuencia WPRE, la cual forma parte del vector pLV-EF1-rhGal-F, con el que se generó el clon productor CHO-K1 α -Galactosidasa A:

| Primers | Secuencia (sentido 5' - 3') |
|----------------|------------------------------------|
| Fw WPRE | CGCTGCTTTAATGCCTTTGT |
| Rv WPRE | GGGCCACAACCTCCTCATAAA |

Para las reacciones de PCR se utilizó la *master mix 5X HOT FIRE Pol EvaGreen qPCR Mix Plus (Rox)* (Solis BioDyne, Estonia), preparando la mezcla de reacción por triplicado, de acuerdo con las instrucciones del fabricante:

| Reactivo | Concentración final | Volumen (por tubo de reacción) |
|---|----------------------------|---------------------------------------|
| HOT FIRE Pol EvaGreen qPCR Mix Plus 5X | 1 X | 4,0 μ l |
| Primers | | |
| Forward 10 μM | 0,25 μ M | 0,5 μ l |
| Reverse 10 μM | 0,25 μ M | 0,5 μ l |
| Templado | X | 10 μ l |
| Agua | - | csp 20 μ l |

Las reacciones se llevaron a cabo en el equipo StepOne (Applied Biosystems, EE.UU.) de acuerdo con el siguiente programa de ciclado:

Paso 1: Desnaturalización inicial: 95°C por 15 min

Paso 2: Desnaturalización: 95°C por 15 s

Paso 3: *Annealing* 60°C por 20 s

Paso 4: Extensión 72°C por 20 s

Volver al paso 2: 40 veces

Los análisis de los datos se llevaron a cabo en aplicaciones de hojas de cálculo, aplicando el método 2 delta-delta Ct [162].

1.17. Estrategias de clonado para la generación de los vectores de expresión

1.17.1. Esquema y descripción del plásmido de expresión eucariota pZsGreen1-1

El plásmido empleado para evaluar la actividad transcripcional en células eucariotas, dirigida por los promotores endógenos identificados por RNA-seq, fue el vector pZsGreen1-1 (Clontech) (Figura 37).

Este vector contiene la secuencia codificante de la proteína verde fluorescente ZsGreen1, derivada de *Zoanthus sp.*, a la cual se le han optimizado sus codones para una expresión óptima en células de mamíferos. Asimismo, a dicha secuencia wt se le ha realizado una sustitución única de aminoácidos (Asn-66 por Met), para mejorar las características de emisión de la proteína ZsGreen1 (máximo de excitación = 493 nm; máximo de emisión = 505 nm). En la región *upstream* a ZsGreen1 se incluyó la secuencia consenso de inicio de la traducción Kozak, con el fin de incrementar la eficiencia de traducción en células eucariotas

El vector pZsGreen1-1 no posee una secuencia promotora que dirija la expresión de ZsGreen1, por lo cual es ideal para monitorear la actividad de diferentes promotores y/o *enhancers*, los cuales se insertan en el sitio de múltiple clonado (MCS).

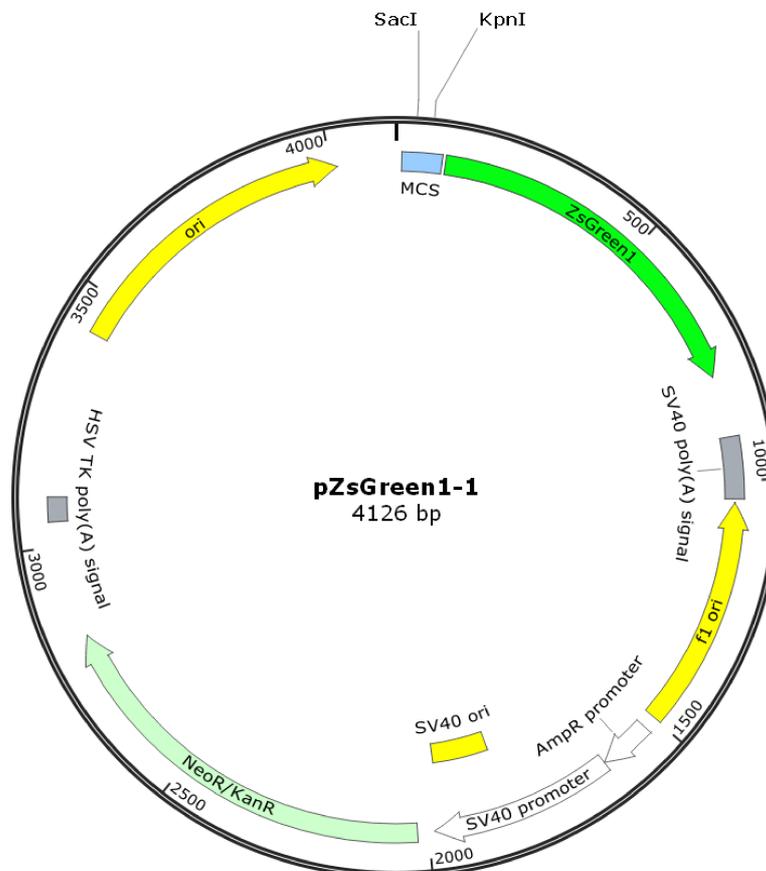


Figura 37. Representación esquemática del vector de expresión eucariota pZsGreen1-1.
Figura realizada con SnapGene® Viewer 4.1.9.

Además, el plásmido posee la señal de poliadenilación de SV40 *downstream* del gen ZsGreen1 (SV40 *polyA signal*), para un correcto procesamiento del ARNm; un origen de replicación en células de mamíferos (SV40 ori), expresando el antígeno T de SV40; un origen de replicación bacteriano (ori); un origen de replicación f1 para la producción de ADN de hebra simple (f1 ori).

Conjuntamente, el vector contiene un *cassette* de selección, compuesto por el promotor temprano SV40, el gen de resistencia a neomicina/kanamicina y una señal de poliadenilación del gen timidina quinasa (TK) del virus herpes simple (HSV TK *polyA signal*). Un promotor bacteriano *upstream* de dicho *cassette* expresa la resistencia a kanamicina en bacterias.

1.17.2. Obtención del plásmido pZsGreen1-1-PromotorEndógeno

Para generar los diversos vectores de expresión conteniendo los diferentes promotores endógenos aislados de células CHO-K1, y la secuencia codificante para la proteína verde fluorescente ZsGreen1, se realizó la digestión del plásmido pZsGreen1-1 con las enzimas *SacI* y *KpnI* (Promega), como se indica en la Sección 1.5 de este apartado. Al mismo tiempo, se digirieron las secuencias de los promotores endógenos amplificadas por PCR (Sección 1.16.1) con las mismas enzimas de restricción, ya que los sitios de reconocimiento para tales enzimas se incorporaron en los *primers* diseñados para PCR.

Luego, se realizó la ligación del vector con cada promotor endógeno utilizando la enzima T4 ADN Ligasa. Las ligaciones se incubaron a 4°C ON (Sección 1.9). La proporción de vector más inserto (V+I) utilizada en las ligaciones fue de 1:5, respectivamente. También se realizó un control de vector religado (VR). Transcurrido dicho tiempo, se utilizó la totalidad de la mezcla de ligación para transformar células competentes *E. coli* TOP 10, como se describe en la Sección 1.11.

A continuación se tomaron las colonias obtenidas en la placa de V+I y se analizó la presencia del inserto en los mismos mediante reacción de PCR de colonias (apartado 1.16.2). La presencia del inserto en los clones que resultaron positivos en la PCR se corroboró mediante minipreparación de ADN plasmídico seguida por digestión con las enzimas de restricción *SacI* y *KpnI*.

1.18. Ensayos con CRISPR/Cas9

1.18.1. Amplificación y purificación del vector lentiCRISPRv2

Para la expresión de la endonucleasa Cas9 y el sgRNA se utilizó el vector lentiCRISPRv2 (*Addgene plasmid #52961*), donado por el *Salk Institute* (La Jolla, CA,

EE.UU.). El mismo es un plásmido lentiviral de tercera generación de casi 15 kb, producido por Feng Zhang (Sanjana y col., 2014 [223]). Además, es un plásmido del tipo “todo en uno” (*all-in-one*), ya que expresa la endonucleasa Cas9 bajo el promotor EF-1 α y el sgRNA bajo el control del promotor U6.

En la Figura 38 se puede observar un esquema de dicho plásmido. El mismo es un vector complejo, que posee diversos elementos, entre ellos se destacan:

- un origen de replicación bacteriano (*ori*);
- un gen que otorga resistencia a ampicilina;
- el promotor U6;
- una región denominada *filler* de aproximadamente 2 kb, que se utiliza para distinguir las bandas originadas producto de la digestión con la enzima *BsmBI* mediante electroforesis en gel de agarosa, para aislar así el vector digerido y clonar luego el sgRNAs en el lugar de esta región *filler*;
- la secuencia *scaffold* del sgRNA;
- el promotor EF1 α ;
- el gen de la endonucleasa Cas9 fusionado a un gen de resistencia a puromicina, mediante la secuencia p2A que codifica para el péptido de auto clivaje 2A;
- una señal de localización nuclear (NLS), que dirige a la proteína Cas9 al núcleo.
- el promotor CMV (CMV);
- la presencia de una región RRE, altamente estructurada, que actúa como sitio de unión para la proteína Rev, la cual permite el transporte de los ARNs sin procesar o parcialmente procesados desde el núcleo al citoplasma;
- la secuencia cPPT (segmento polipurina central), que actúa como cebador de la hebra (+) en la síntesis del ADN viral;
- WPRE, un elemento regulatorio postranscripcional del virus de la hepatitis de *Woodchuck*;
- una señal de empaquetamiento (Psi), una secuencia corta y altamente estructurada requerida para la incorporación del ARN genómico en el interior de las partículas virales.

Todas las regiones necesarias para la generación de los vectores lentivirales se encuentran comprendidas entre las repeticiones terminales largas (5' LTR y 3' LTR).

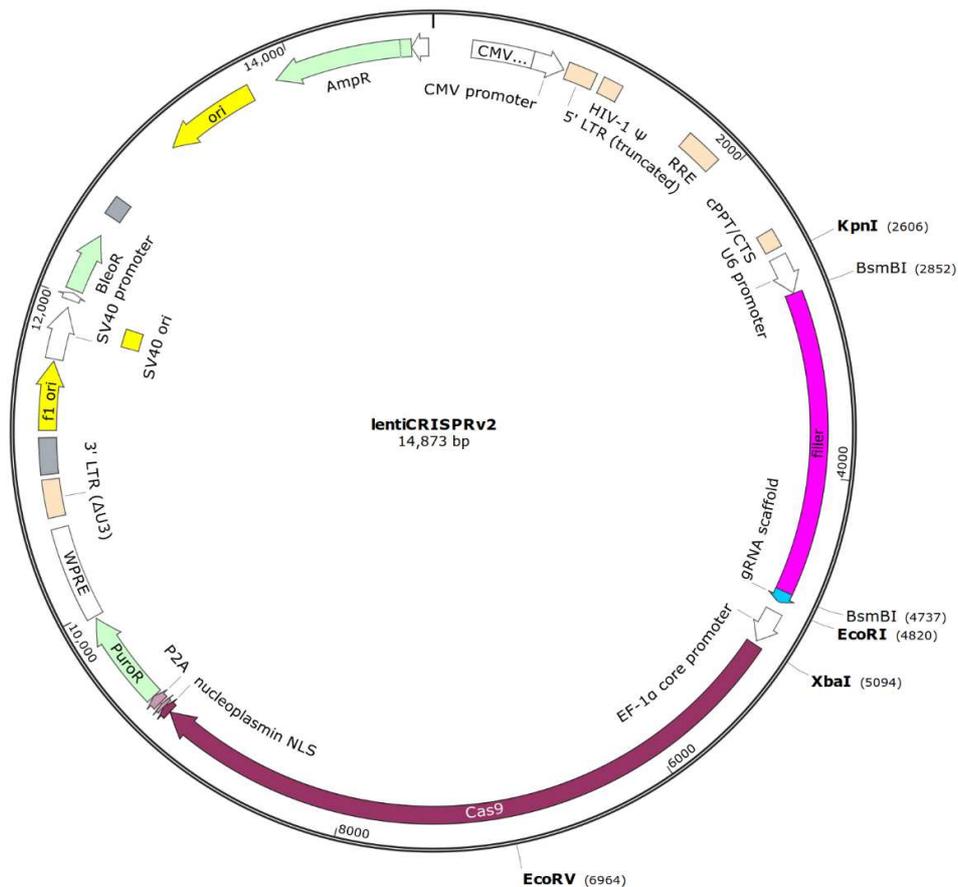


Figura 38. Representación esquemática del vector lentiCRISPRv2.

Figura realizada con SnapGene® Viewer 4.1.9.

Se realizó la purificación del ADN plasmídico por maxipreparación (Sección 1.3.2), a partir de células competentes *E. Coli* TOP10 transformadas (apartado 1.11), con el fin de obtener un plásmido con alto grado de pureza y en cantidad suficiente para realizar los experimentos posteriores.

Para verificar la identidad del plásmido obtenido, se realizó un chequeo mediante digestión con las enzimas de restricción *EcoRV* y *XbaI* (Promega) (Sección 1.5) y el producto de la misma se sometió a una electroforesis en gel de agarosa (Sección 1.6).

1.18.2. Selección de las regiones target y diseño de los sgRNAs

Se analizaron las regiones genómicas de cuatro genes identificados a partir de los análisis transcriptómicos (S100a6, Fth1, Lgals1 y Vim). En particular, se buscaron sitios *targets* para el diseño de los sgRNAs en las regiones 5' y 3' intergénicas (es decir, hacia el extremo 5' antes del promotor y hacia el extremo 3' luego de la señal de poliadenilación) y en las regiones 5' UTR y 3' UTR de cada gen (Figura 39). También, se diseñó un sgRNA para la proteína verde fluorescente EGFP, empleado en el ensayo de *knock-out* (Sección 2.8.1).



Figura 39. Esquema de las regiones genómicas *target* para el diseño de los sgRNAs para el experimento de *knock-in* mediante el sistema CRISPR/Cas9.
Las flechas indican los sitios seleccionados para el diseño.

Para cada región *target* se consideraron aproximadamente entre 100-150 pb de la secuencia del transcriptoma secuenciado, y se diseñaron los sgRNAs utilizando los *software* CCTop [224] y CHOPCHOP [225]. Dichos oligos se detallan a continuación.

| Oligo* | Secuencia (5'→3') |
|------------------------|----------------------------------|
| EGFP_gRNA1_fw | CACCGAGCTGGACGGCGACGTA |
| EGFP_gRNA1_rv | AACTTTACGTCGCCGTCAGCTC |
| EGFP_gRNA2_fw | CACCGGGCGAGGAGCTGTTACCG |
| EGFP_gRNA2_rv | AAACCGGTGAACAGCTCCTCGCC |
| EGFP_gRNA3_fw | CACCGTCTTTGCTCAGGGCGGACT |
| EGFP_gRNA3_rv | AAACAGTCCGCCCTGAGCAAAGAC |
| S100a6_3UTR1_fw | CACCGATTGTACAATTACCCACCAC |
| S100a6_3UTR1_rv | AAACGTGGTGGTAATTGTACAATC |
| S100a6_3UTR2_fw | CACCGCAATTACCCACCACTGGATT |
| S100a6_3UTR2_rv | AAACAATCCAGTGGTGGTAATTGC |
| S100a6_5UTRinter_fw | CACCGCCAATCTAGTAGACCACTA |
| S100a6_5UTRinter_rv | AAACTAGTGGTCTACTAGATTGGC |
| S100a6_3UTRinter_fw | CACCGCGGAAGGCCTAATCCAAGTT |
| S100a6_3UTRinter_rv | AAACAACCTGGATTAGGCCTTCCGC |
| Fth1_3UTR_fw | CACCGCTGAGGAAGTGTATGCATGT |
| Fth1_3UTR_rv | AAACACATGCATACACTTCCTCAGC |
| Fth1_3UTRinter_fw | CACCGATTTATTCAACTCCTTAATT |
| Fth1_3UTRinter_rv | AAACAATTAAGGAGTTGAATAAATC |
| Fth1_5UTRinter_fw | CACCGAAGGCATGCCTTTCAAATTA |
| Fth1_5UTRinter_rv | AAACTAATTTGAAAGGCATGCCTTC |
| Lgals1_3UTR_fw | CACCGCTGCCTTTATTGAGGCCTAG |
| Lgals1_3UTR_rv | AAACCTAGGCCTCAATAAAGGCAGC |
| Lgals1_3UTRinter_fw | CACCGAGCCTAGCACATAGAGGGTT |
| Lgals1_3UTRinter_rv | AAACAACCCTCTATGTGCTAGGCTC |
| Lgals1_5UTR_fw | CACCGATTGGCCTAGCCTACCGGAG |
| Lgals1_5UTR_rv | AAACCTCCGGTAGGCTAGGCCAATC |
| Lgals1_5UTRinter_fw | CACCGAGTGCTGGTATGGTATCAAA |
| Lgals1_5UTRinter_rv | AACTTTGATACCATAACAGCACTC |
| Vim_3UTR1_fw | CACCGCTGGTACTGCGTTGTTGCA |
| Vim_3UTR1_rv | AAACTGCAACAACGCAGTACCAGC |
| Vim_3UTR2_fw | CACCGTTATTGAAGCAGTAACAAGT |
| Vim_3UTR2_rv | AAACACTTGTTACTGCTTCAATAAC |
| Vim_3UTRinter_fw | CACCGCATAAATACTATATAGTGCT |
| Vim_3UTRinter_rv | AAACAGCACTATATAGTATTTATGC |
| Vim_5UTRinter_fw | CACCGCTCTTAATTGCCAGGTAAGC |
| Vim_5UTRinter_rv | AAACGCTTACCTGGCAATTAAGAGC |

*Los oligos resaltados en negrita fueron los utilizados en los experimentos posteriores.

Las consideraciones generales aplicadas al diseño fueron: sitio 5'-PAM: "NGG"; longitud del sgRNA (sin la secuencia PAM) de 20 nucleótidos; transcripción *in vitro* mediante promotor U6; predicción de *off-targets* utilizando el genoma de *Cricetulus griseus*, con un máximo total de *mismatches* de 4 nucleótidos. Se trató de diseñar más de un sgRNA por región, en algunos casos se obtuvieron hasta tres sgRNAs, en otros casos solamente uno, por diversas razones como la falta de secuencia o la presencia de muchos *off-targets*.

Una vez definida la secuencia *target*, se diseñaron y sintetizaron dos oligos por cada sgARN (uno correspondiente a la hebra *forward* y el otro a la *reverse*). Para cada secuencia *target* el diseño quedó de la siguiente manera:



Donde las "N" representan los 20 nucleótidos de la secuencia *target*, los extremos "CACCG" y "AAAC" (en sentido 5'→3') son necesarios para un correcto clonado de cada par de oligos en el vector lentiCRISPRv2, y la base "G", inmediatamente antes de la secuencia *target*, es esencial para la correcta expresión de los sgRNAs una vez dentro de la célula bajo el promotor U6.

1.18.3. Clonado de los oligos en el vector lentiCRISPRv2

El plásmido lentiCRISPRv2 fue digerido con la enzima de restricción *BsmBI* (Sección 1.5). La misma corta fuera del sitio de reconocimiento y crea extremos no palindrómicos, en este caso se generan extremos complementarios a los anteriormente descritos, incorporados al diseño de los oligos. El resultado de la digestión se separó por electroforesis en gel de agarosa (apartado 1.6), y el fragmento del tamaño esperado (aproximadamente 12 kb) se escindió del gel y se purificó (Sección 1.8), dejando de lado el fragmento correspondiente a la región *filler*.

Por otro lado, se realizó el *annealing* de cada par de oligos. Para el estudio de los promotores endógenos, solamente se emplearon los sgRNAs para los sitios *targets* del extremo 3' UTR de cada gen (en un futuro se plantea analizar el resto de las regiones).

Se resuspendieron los oligos a 100 µM en agua Milli Q doble estéril. Luego, se tomaron 2 µl de cada oligo (fw y rv) y se llevó a un volumen final de 2 µl con agua Milli Q estéril. La reacción de *annealing* se realizó en termociclador durante 5 min a 95°C, luego se dejaron los tubos a temperatura ambiente por 15 min. Posteriormente, cada reacción se diluyó a razón de 1:200 en agua Milli Q estéril.

Luego, se realizó la ligación del vector con cada sgRNA utilizando la enzima T4 ADN Ligasa. Las ligaciones se incubaron a 4°C ON (Sección 1.9). La proporción de vector más inserto (V+I) utilizada en las ligaciones fue de 1:5, respectivamente. También se realizó un control de vector religado (VR). Transcurrido dicho tiempo, se utilizó la totalidad de la mezcla de ligación para transformar células competentes *E. coli* TOP 10, como se describe en la Sección 1.11.

A continuación se tomaron las colonias obtenidas en la placa de V+I y se analizó la presencia del inserto en los mismos mediante reacción de PCR de colonias (apartado 1.16.2). Todos los clones elegidos (por cada par de oligos clonados) fueron positivos, se eligió un clon de cada uno, se realizó una maxipreparación de ADN plasmídico (Sección 1.3.2). Posteriormente, se verificó la identidad de los vectores generados mediante digestión con las enzimas de restricción *EcoRV* y *XbaI* (Promega) (Sección 1.5) y el producto de la misma se sometió a una electroforesis en gel de agarosa (Sección 1.6).

2. Técnicas de cultivo celular

2.1. Líneas celulares

2.1.1. CHO-K1 (ATCC)

La línea CHO-K1 fue obtenida de ATCC (ATCC CCL-61). Es un clon de la línea celular CHO derivada de ovario de hámster chino adulto.

2.1.2. HEK293T/17 (ATCC, CRL-11268)

Es un clon que deriva de la línea celular HEK293, establecida a partir de riñón de embrión humano por Graham en 1977. El clon 293T/17 fue modificado para generar la expresión del antígeno T del virus SV40. Se utilizó para el ensamblado de partículas lentivirales y como huésped para la expresión de proteínas recombinantes. En el resto del manuscrito se hace referencia a esta línea celular como HEK.

2.2. Medios de cultivo y suplementos para células animales

Todos los medios de cultivos utilizados para el cultivo de células animales fueron preparados utilizando agua Milli Q y esterilizados por filtración, utilizando dispositivos tipo *Steritop* (Merck Millipore, EE.UU.) con membranas de 0,22 µm de tamaño de poro. Cada medio preparado queda en cuarentena durante 48 h mientras se llevan a cabo los controles microbiológicos y se conservan a 4°C al resguardo de la luz. A continuación se detallan los medios de cultivo y suplementos empleados en este capítulo:

2.2.1. CHO-K1s

Se utilizó el medio comercial Ex-Cell 302 (Sigma-Aldrich), suplementado con glutamina 2 mM.

2.2.2. CHO-K1adh

Se utilizó el medio de cultivo DMEM/Ham's F12 10% SFB, el cual consiste en una mezcla 1:1 (v/v) de los medios DMEM (Dulbecco's Modified Eagle's Medium, Gibco) y Ham's-F12 (Gibco) suplementados con NaHCO₃ (Gibco) 2,441 g.l⁻¹, glutamina (Sigma-Aldrich) 200 mM, sulfato de gentamicina (Gibco) 0,05 mg.ml⁻¹ y SFB 10% (v/v).

2.2.3. HEK293T/17 adherentes

Se utilizó un medio compuesto por medio DMEM (Gibco) 13,37 g.l⁻¹, NaHCO₃ 1,5 g.l⁻¹, piruvato de sodio 0,11 g.l⁻¹, gentamicina 0,05 g.l⁻¹ y SFB 10% (v/v).

2.2.4. Suero fetal bovino

Para suplementar los medios de cultivo utilizados en cultivos en adherencia, se adicionó a cada medio basal la concentración de SFB deseada. Los SFB utilizados (PAA Laboratories, Austria) son de calidad certificada y previo a su uso son descomplementados por incubación en baño de agua termostaticado a 56°C durante 30 min.

2.2.5. Tripsina

Tripsina (Gibco) 0,05% (p/v).

Na₂EDTA (Merck) 0,02% (p/v).

Preparado en buffer PBS [KH₂PO₄ (Cicarelli) 0,6 g/l; NaHPO₄ (Cicarelli) 0,78 g/l; NaCl (Cicarelli) 8,8 g/l; pH 7,4]. Se respetó una relación de aproximadamente 50 µl de solución de tripsina-EDTA por cada cm² de superficie.

2.3. Sistemas de cultivos de células

Todas las líneas celulares eucariotas se cultivaron en una incubadora a 37°C, con atmósfera de CO₂ al 5% saturada de humedad. Según la etapa de trabajo, se utilizaron cultivos en adherencia o suspensión.

2.3.1. Cultivos en suspensión en pequeña escala

Según lo detallado en el Capítulo 1 / Materiales y Métodos / Sección 2. Técnicas de cultivo celular / subsección 2.3.1.

2.3.2. Cultivos en adherencia

Para cultivos en adherencia, se emplearon los siguientes sistemas de cultivo:

- Placas multipozos de poliestireno (Greiner Bio-One o Jet Biofil, China) de 24, 12 y 6 pozos con una superficie de 2, 5 y 9,6 cm² por pozo, respectivamente.
- Frascos T (Greiner Bio-One o Jet Biofil) de 25, 75 y 175 cm² de superficie.

En todos los casos se mantuvo una relación constante de 0,4 ml de medio de cultivo por cada cm² de superficie.

Las células que presentan un crecimiento en adherencia a un sustrato, esto es, las líneas CHO-K1adh y HEK293T/17, forman una monocapa. Para su propagación es necesario liberarlas del sustrato por medios mecánicos o enzimáticos. Cuando los cultivos alcanzaron la confluencia, es decir, crecieron formando una monocapa en la totalidad de la superficie de los frascos o placas sobre los cuales estaban multiplicándose, se procedió a retirar el sobrenadante y lavar la monocapa celular con una solución de tripsina-EDTA, compuesta por tripsina 0,05% (p/v) y EDTA 0,02% (p/v) en PBS durante 15 s a TA con el objetivo de neutralizar los factores antitripsina presentes en el SFB. Se respetó una relación de aproximadamente 50 µl de solución de tripsina-EDTA por cada cm² de superficie. Luego de eliminar la solución de lavado, se agregó nuevamente la misma proporción de la solución de tripsina-EDTA, dejándola actuar el tiempo necesario para cada línea celular empleada, que permita el desprendimiento completo de las células adheridas a la superficie. A continuación, se agregó el medio de cultivo correspondiente, suplementado con SFB, para neutralizar la tripsina y homogeneizar la suspensión celular de tal forma de obtener las células en suspensión.

Por otro lado, las células capaces de crecer en suspensión, las células CHO-K1s, fueron propagadas mediante dilución de la suspensión original en un mayor volumen del medio de cultivo correspondiente.

2.4. Determinaciones analíticas en cultivos celulares

2.4.1. Determinación de la concentración y viabilidad de células

Según lo detallado en el Capítulo 1 / Materiales y Métodos / Sección 2. Técnicas de cultivo celular / subsección 2.4.1.

2.5. Revitalización de líneas celulares de mamífero

Según lo detallado en el Capítulo 1 / Materiales y Métodos / Sección 2. Técnicas de cultivo celular / subsección 2.5.

2.6. Criopreservación de líneas celulares de mamífero

La conservación de las líneas y clones celulares se realizó en nitrógeno líquido, ya que a temperaturas de -196°C se observan deterioros celulares mínimos. Se empleó una técnica de enfriamiento gradual desde temperatura ambiente hasta -196°C .

Para ello, una vez realizado el recuento, suspensiones celulares en fase de crecimiento exponencial se centrifugaron a 200 *g* durante 5 min a temperatura ambiente en una centrífuga Eppendorf 5403, rotor 16A4-44 (Alemania). Se descartó el sobrenadante y el sedimento celular se resuspendió en una solución de SFB y dimetilsulfóxido (DMSO, Sigma-Aldrich) en una relación 90:10 (v/v), de manera de obtener una concentración de $1-2 \times 10^6$ cél.ml⁻¹. La suspensión celular se distribuyó a razón de 1 ml por criotubo (Greiner Bio-One), los que luego fueron guardados en un contenedor apropiado, *MrFrosty* (Nalgene), y llevados a -70°C por 24 h. Este dispositivo contiene alcohol isopropílico, permitiendo que en el interior del mismo la temperatura disminuya a razón de 1°C por minuto, de manera de lograr un enfriamiento gradual de las células, que luego permita una correcta revitalización de las mismas. Finalmente, los criotubos se almacenaron en termos de nitrógeno líquido a -196°C .

2.7. Desarrollo de líneas celulares recombinantes

2.7.1. Transfección mediante lipofección

El día previo a la transfección, células CHO-K1adh y HEK fueron desprendidas enzimáticamente utilizando una solución de tripsina-EDTA. Luego se cuantificaron y cultivaron en placas multipozos de poliestireno de 12 pocillos en una concentración de $2,0 \times 10^5$ cél.ml⁻¹. Veinticuatro horas después, la monocapa del cultivo celular fue lavada con medio basal en cada pocillo con la finalidad de eliminar el SFB, para favorecer así la eficiencia de transfección. Luego se continuó con los siguientes pasos:

1. Preparar un tubo Eppendorf para cada vector plasmídico conteniendo la cantidad necesaria para dos transfecciones (una en células CHO-K1 y otra en células HEK). Para ello, colocar en cada tubo 4 μg del vector de interés (pZsGreen1-1, pCMV-ZsGreen1-1, pRps18-ZsGreen1-1, pHspa5-ZsGreen1-1, pHmox-C1-ZsGreen1-1 y pHmox-C2-ZsGreen1-1) junto con 400 μl de medio basal.
2. Por otro lado, poner 2 μl de *Lipofectamine™ 2000 Reagent* (Invitrogen) en un tubo Eppendorf junto con 200 μl de medio basal, para cada transfección. Es decir, preparar en un cónico 24 μl de *Lipofectamine™ 2000 Reagent* con 2,4 ml de medio basal. Fraccionar la mezcla en seis tubos Eppendorf. Dejar reposar durante 5 min.
3. Mezclar las dos soluciones y dejar reposar 20-25 min a temperatura ambiente para la formación de los complejos ADN-lípido.
4. En el tiempo de incubación, remover el medio de cultivo de las células y adicionar 400 μl de medio basal a cada pozo.

5. Una vez transcurrido el tiempo de incubación, agregar lentamente la mezcla ADN-lípidos a cada pocillo conteniendo células CHO-K1adh y HEK.
6. Incubar el cultivo celular en estufa gaseada a 37°C durante 4-5 h.
7. Una vez transcurrida la incubación, agregar medio DMEM/Ham's F12 10% SFB (v/v) a las células CHO-K1 y medio DMEM p/HEK 10% SFB (v/v), de forma tal de completar a 2 ml el volumen de cada pozo.
8. Transcurridas 72 h, evaluar la fluorescencia de las células por citometría de flujo.

2.7.2. Transducción

2.7.2.1. Obtención de partículas lentivirales a partir del plásmido de transferencia lentiCRISPR, codificante de Cas9-sgRNA-EGFP

Para el ensamblado de LVs se llevó a cabo una transfección transiente mediante lipofección de células HEK empaquetadoras con cuatro plásmidos distintos: pMDLg/pRRE, pMD.G, pRSV-Rev [226] y el vector de transferencia correspondiente (lentiCRISPRv2-sgRNA-EGFP).

Luego de 48 h postransfección, se cosecharon los LVs a partir del sobrenadante de cultivo. A continuación se detalla el procedimiento llevado a cabo:

1. Cultivar células HEK a una densidad de 4×10^5 céls.ml⁻¹ en placas de 100 mm de diámetro (Greiner Bio-One), empleando medio DMEM p/HEK suplementado con SFB 10% (v/v).
2. Luego de 24 h de incubación a 37°C, CO₂ 5%, llevar adelante el procedimiento de lipofección. Para ello, preparar 2,5 ml de medio DMEM p/HEK sin SFB conteniendo 2,5 µg pRSV-REV; 3,6 µg pMD.G; 6,5 µg pMDLg/pRRE y 10 µg del plásmido de transferencia. Por otro lado, en otro tubo cónico de 15 ml, mezclar 2,5 ml de medio DMEM p/HEK sin SFB con 50 µl de *Lipofectamine 2000* (1 mg/ml). Agitar vigorosamente e incubar esta última mezcla por 5 min a TA.
3. Mezclar el contenido de ambos tubos cónicos gota a gota para favorecer la formación homogénea de los complejos lípido-ADN e incubar a TA durante 20 min.
4. Retirar el sobrenadante de las placas de cultivo conteniendo las células HEK y adicionar el volumen de 5 ml de la mezcla lípido-ADN. Incubar en las mismas condiciones de cultivo durante al menos 4 h.
5. Al cabo de este tiempo, suplementar el cultivo con 5 ml de medio DMEM p/HEK SFB 20% (v/v), de modo que la concentración final de suero sea 10% (v/v).
6. Cosechar el sobrenadante del cultivo conteniendo los vectores lentivirales 48 h después de la transfección, centrifugando a 200 g durante 8 min a TA para eliminar las células que pudieran haberse desprendido de la monocapa al retirar el medio.
7. Filtrar el sobrenadante utilizando un filtro para jeringa con membrana de 0,45 µm de tamaño de poro.
8. Almacenar el sobrenadante a -70°C hasta efectuar el paso de concentración de las partículas por centrifugación.

2.7.2.2. Concentración de LVs mediante ultracentrifugación

Una vez cosechados los LVs del sobrenadante de células HEK empaquetadoras, se realizó la ultracentrifugación como se describe a continuación:

1. Colocar el sobrenadante filtrado en dos tubos para ultracentrífuga Beckman Coulter (EE.UU.), completando con medio DMEM y controlando que el peso de ambos tubos con sobrenadante quede equilibrado.
2. Centrifugar los tubos a 65.000 *g* durante 3 h a 4°C con desaceleración mínima, (rotor JA 30.50).
3. Una vez finalizado el tiempo de centrifugación, eliminar el sobrenadante por volcado, dejando el tubo invertido para retirar completamente el mismo.
4. Resuspender completamente el *pellet* de partículas lentivirales, incubar ON el tubo con 300 μ l de medio de cultivo Ex-Cell 302 a 4°C.
5. Al día siguiente, alicuotar las partículas lentivirales en tubos Eppendorf conteniendo 100 μ l cada uno y almacenarlos a -70°C.

2.7.2.3. Transducción lentiviral

Células CHO-K1adh (clon celular que expresa EGFP) se transdujeron con los LVs concentrados. Se procedió de la siguiente manera:

1. Cultivar células CHO-K1adh en placas multipozos de 6 pozos a una densidad de 3×10^4 cél. ml^{-1} . Luego de incubar a 37°C, CO₂ 5% durante 24 h, remover el sobrenadante de cultivo y adicionar en cada caso los LVs concentrados.
2. Incubar las células en estas condiciones durante otras 24 h, luego remover el medio conteniendo los LVs y reemplazarlo por 3 ml del medio de cultivo correspondiente.
3. Veinticuatro horas luego del cambio de medio de cultivo (48 h postransducción), tratar las células con tripsina, y cultivarlas nuevamente en placas a una densidad de 3×10^4 cél. ml^{-1} para efectuar dos transducciones sucesivas en idénticas condiciones.

2.8. Experimentos con el sistema CRISPR/Cas9 en células CHO-K1

2.8.1. Experimento de knock-out con el sistema CRISPR/Cas9 en células CHO-K1 adherentes productoras de EGFP

Con el objetivo de poner a punto e implementar en células CHO el sistema CRISPR/Cas9, se realizó un ensayo de *knock-out* de células CHO-K1 reporteras productoras de EGFP. Se diseñó el sgRNA (considerando como secuencia *target* la de EGFP) de la manera descrita en la Sección 1.18.2, y se clonó en el plásmido lentiviral de la forma mencionada en el apartado 1.18.3, generándose el vector lentiCRISPRv2-sgRNA-EGFP.

Se procedió a ensamblar partículas lentivirales a partir de los vectores estructurales pMDLg/pRRE, pMD.G, pRSV-Rev, y el vector de transferencia lentiCRISPRv2-sgRNA-EGFP. Para ello, se realizó una transfección mediada por lípidos

catiónicos en células empaquetadoras HEK y 48 h posttransfección se cosecharon los sobrenadantes de cultivo conteniendo las partículas lentivirales, se concentraron y se almacenaron a -70°C (Sección 2.7).

Luego, se transdujeron células de un clon CHO-K1 productor de EGFP (apartado 2.7.2.3) y se mantuvieron en cultivo durante un mes utilizando medio DMEM/Ham's F12 (suplementado con 10% SFB). Aproximadamente cada tres días se realizó citometría de flujo (Guava easyCyte) para evaluar la expresión de la proteína fluorescente EGFP (Sección 3.1). Posteriormente, se criopreservaron las células transducidas (Sección 2.6) trascurrido un mes de cultivo y seis meses después se descongelaron (apartado 2.5) para evaluar la estabilidad del *knock-out* generado.

A continuación, se realizó la selección de la población celular que había perdido la expresión de la proteína fluorescente mediante FACS (*fluorescence-activated cell sorting*), empleando el equipo FACS ARIA III (BD), con el objetivo de enriquecernos en dicha población con el cambio producido a través del sistema CRISPR/Cas9. Esta población celular se mantuvo en cultivo durante 48 h (*pos-sorting*) para evaluar la estabilidad de la modificación. Este ensayo sirvió como una forma de evaluación del proceso para futuros experimentos con esta técnica de edición genética.

2.8.2. Experimento de *knock-in* mediante CRISPR/Cas9 en células CHO-K1

Los brazos homólogos necesarios para el *knock-in* en cada gen elegido (Sección 1.18.2) se amplificaron mediante PCR empleando la enzima *TransTaq DNA polymerase High Fidelity* (TransGen Biotech) (apartado 1.16.1), utilizando como ADN molde el genoma de células CHO-K1 adaptadas al crecimiento en suspensión (Sección 1.14).

En la secuencia de los *primers* diseñados para tal fin se incluyeron sitios de reconocimiento para enzimas de restricción, necesarias para el clonado en el plásmido donante. El tamaño de cada uno de los brazos homólogos es de 800-850 pb. Los productos de PCR se separaron por electroforesis en gel de agarosa (Sección 1.6) y los fragmentos con tamaño esperado se escindieron del gel y se purificaron (apartado 1.8).

Por otro lado, se diseñó la secuencia de un *cassette* de expresión para el ensayo de *knock-in* (Figura 40).

Esta secuencia contiene dos sitios de múltiple clonado en donde se insertarán las secuencias necesarias para la recombinación homóloga en cada sitio *target* (brazos homólogos izquierdo y derecho). Entre dichos sitios se encuentra una secuencia IRES y la proteína reportera ZsGreen1, flanqueada por sitios FRT del sistema de recombinación FLP/FRT, la secuencia regulatoria WPRE, y por fuera del sitio de inserción del brazo homólogo derecho se incluyó el promotor SV40, la proteína fluorescente DsRed y la señal de poliadenilación BGH.

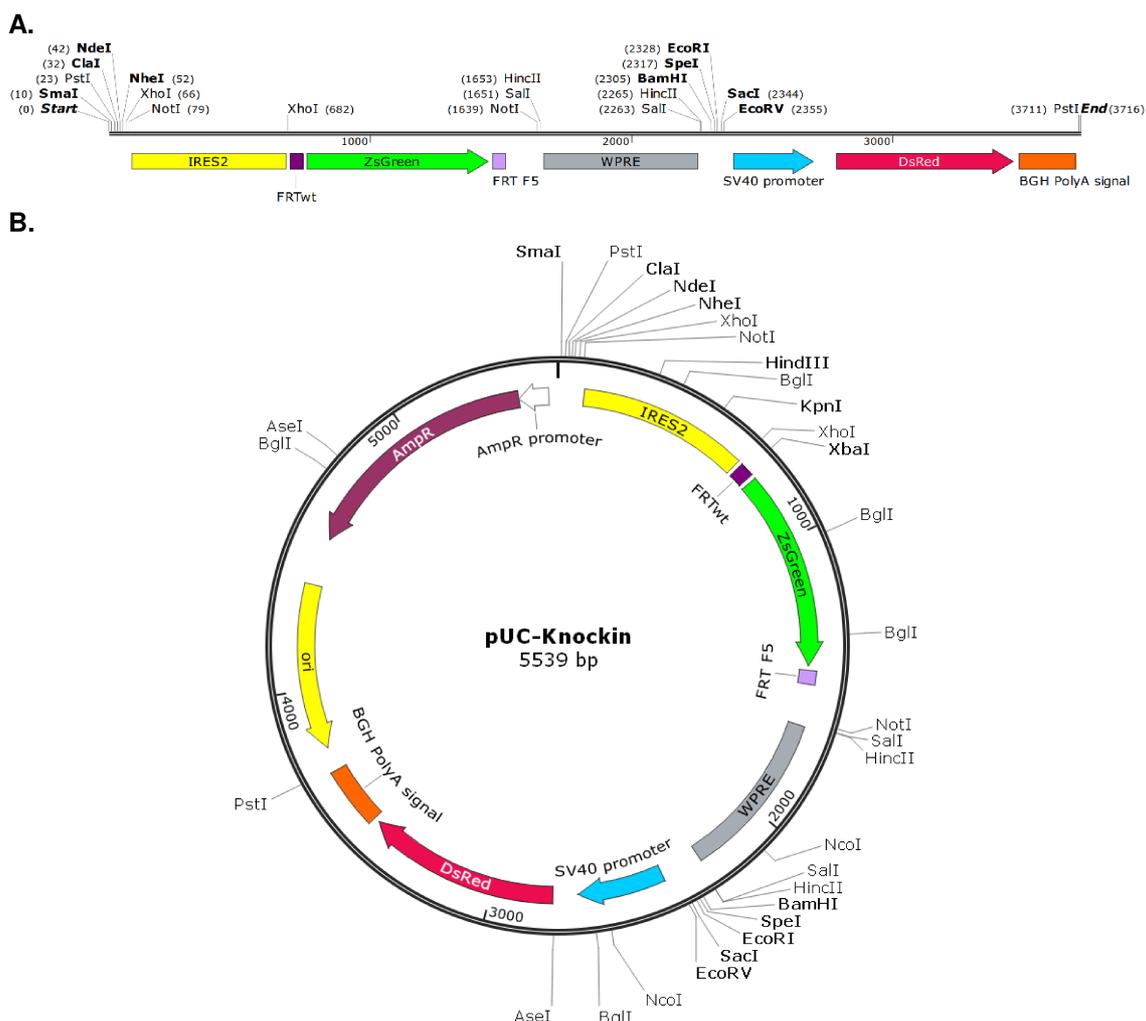


Figura 40. Representación esquemática de la secuencia diseñada para insertarse en células CHO-K1 mediante *knock-in* por recombinación homóloga directa.
A. *Cassette* diseñado *in silico* y sintetizado.
B. Plásmido donante con dicho *cassette* clonado.
 Figura realizada con SnapGene® Viewer 4.1.9

El tamaño total de la secuencia diseñada es de 3716 pb (Figura 40.A). Esta secuencia fue sintetizada por la empresa Gene Universal (EE.UU.) y clonada en el vector pUC57 (1,8 kb), originándose el vector pUC-Knockin (Figura 40.B).

3. Metodologías para el análisis de líneas y clones celulares

3.1. Citometría de flujo

Tanto para la evaluación de la actividad transcripcional de los promotores endógenos (intensidad de fluorescencia de ZsGreen1) –seleccionados a partir de RNA-seq– en células de mamífero como para evaluar la intensidad de fluorescencia de EGFP en los ensayos de *knock-out* mediante el sistema CRISPS/Cas9 en células CHO-K1, se utilizó un citómetro de flujo Guava (EasyCyte, Merck Millipore). Este

citómetro tiene la posibilidad de adquirir muestras presentes en una placa de 96 pozos. Los datos fueron analizados con el *software* GUAVA ExpressPlus (Merck Millipore).

Para el estudio de los promotores endógenos, se tripsinaron los pozos con células CHO-K1 adaptadas al crecimiento en suspensión (a las cuales se les incorporó SFB para los ensayos de transfección y *screening* de la actividad transcripcional), y los pozos con células HEK transfectadas también con las diferentes construcciones pZsGreen1-1-PromotorEndógeno. Se realizó una dilución de cada muestra con PBS en una placa de 96 pozos y se analizó la intensidad de fluorescencia en dicho citómetro. Como control se emplearon células CHO-K1 y HEK transfectadas con el vector pZsGreen1-1 (sin promotor) y el vector pCMV-ZsGreen1-1 (generado previamente en nuestro laboratorio).

Para los experimentos de *knock-out* con CRISPR se tripsinaron los pozos con células CHO-K1adh Clon 52 que expresan EGFP (copia única), en ambos casos transducidas con los LVs que expresan los componentes del sistema CRISPR/Cas9 necesarios para el ensayo de *knock-out*. Se utilizaron como controles células CHO-K1 *wild type* y células del clon CHO-K1-EGFP sin transducir. Se realizó una dilución de cada muestra con PBS en una placa de 96 pozos y se analizó la intensidad de fluorescencia en el citómetro.

Posteriormente, se seleccionó la población celular con el *knock-out* mediante FACS, empleando el equipo FACS ARIA III (BD). Los datos fueron analizados con el *software* FACSDiva v6.1.2 (BD).

3.2. Microscopía de fluorescencia

Paralelamente a los ensayos de citometría, en ambos experimentos, las células fueron evaluadas diariamente en el microscopio de fluorescencia invertido *Eclipse Ti-S* (Nikon Instruments Inc).

RESULTADOS Y CONCLUSIONES

1. Estudio comparativo de la expresión de cuatro genes con mayor expresión con la de un transgén previamente caracterizado

Como se detalló en el Capítulo 2, hemos caracterizado el transcriptoma de células CHO-K1 huésped adaptadas al crecimiento en suspensión mediante RNA-seq, identificando perfiles de expresión génica característicos para cada etapa de cultivo en biorreactor, así como también las principales vías metabólicas reguladas. Asimismo, se ha identificado un grupo de genes que exhibe elevada actividad transcripcional durante dicho bioproceso.

Luego de llevar a cabo la cuantificación génica mediante RNA-seq, se seleccionaron los primeros 200 transcriptos con mayor expresión, tanto en la etapa exponencial como estacionaria de cultivo. De ese grupo, se eligieron los siguientes genes: S100a6, Fth1, Lgals1, Vim, Actb, Rps18, Cd63, Calr, Ybx1, Hspa5, Hmox1, Ctsz y Lamp1. Se consideraron entonces dichos genes como “candidatos” para el empleo de sus regiones promotoras con el objetivo de construir nuevos vectores de expresión para la producción de proteínas recombinantes en células de mamífero.

Posteriormente, se contrastó la expresión de los cuatro genes con mayor expresión (S100a6, Fth1, Lgals1 y Vim), contra la expresión de un transgén (α -Galactosidasa A) incorporado en un clon productor de células CHO-K1, generado en nuestro laboratorio, mediante RT-qPCR (Figura 41). Esto se realizó con el fin de comparar la expresión de dichos genes contra la de este TG previamente caracterizado, y que tiene elevada actividad transcripcional, posee únicamente dos copias en el genoma y, además, su expresión está potenciada por secuencias reguladoras incorporadas en el vector de expresión lentiviral con el cual se generó este clon. Para ello, se utilizaron –por un lado– los extractos de ARN disponibles de los cultivos en biorreactores detallados en el Capítulo 2 y –por otro lado– nuevas muestras de ARN a partir del cultivo en biorreactor del clon productor de α -Galactosidasa A, durante la fase estacionaria de crecimiento celular.

Como se aprecia en la Figura 41, el gen α -Galactosidasa A está 4,2 veces más expresado que el gen S100a6; 8,2 veces más que Fth1; 18,2 veces más que Lgals1 y 18,5 veces más que Vim. Además, se distinguió que estos genes “candidatos”, en general, presentan una expresión más elevada en el clon CHO-K1 α -Galactosidasa A que en las células CHO-K1 wt (aproximadamente entre 0,4 y 0,6 veces más).

Este análisis arrojó resultados muy prometedores, particularmente con respecto al gen S100a6, como potencial candidato mediante el empleo de su secuencia

promotora para la producción de proteínas recombinantes terapéuticas, como alternativa al uso de promotores virales.

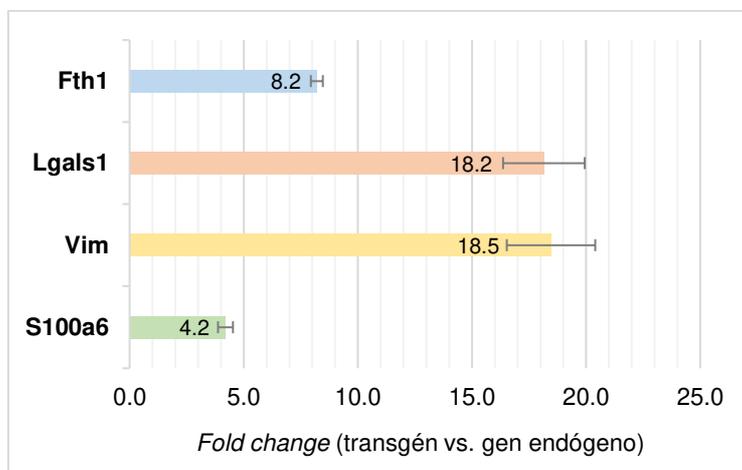


Figura 41. Cambio en la expresión del transgén α -Galactosidasa A, incorporado en un clon de células CHO-K1, respecto a los cuatro genes seleccionados con mayor expresión.

Se muestran los valores de *fold change* obtenidos por RT-qPCR.

2. Generación de nuevos vectores de expresión para la producción de proteínas recombinantes terapéuticas

Como objetivo principal de esta etapa de trabajo, se planteó utilizar las secuencias promotoras de los genes identificados con alta actividad transcripcional, para la construcción de nuevos vectores de expresión destinados a la producción de proteínas recombinantes, logrando así que dichos promotores respondan a las condiciones celulares y a la dinámica de los procesos productivo en escala industrial.

Esta estrategia de trabajo constituye una alternativa al empleo de promotores virales usados habitualmente, aprovechando la maquinaria transcripcional regulatoria endógena disponible en las condiciones de un bioproceso de alta densidad celular con un medio de cultivo definido. Así, se busca lograr una expresión óptima del TG de interés en un contexto normal de producción industrial.

2.1. Aislamiento de promotores endógenos derivados de genes previamente identificados mediante RNA-seq con elevada actividad transcripcional en cultivos CHO-K1 en biorreactor

Los *primers* necesarios para la amplificación de los promotores endógenos seleccionados (S100a6, Fth1, Lgals1, Vim, Actb, Rps18, Cd63, Calr, Ybx1, Hspa5, Hmox1, Ctsz y Lamp1) se diseñaron de manera tal de amplificar entre 900 pb y 1100 pb de ADN genómico de células CHO-K1 adaptadas al crecimiento en suspensión, inmediatamente *upstream* del sitio de inicio de la transcripción de cada gen.

Además, se compararon las regiones seleccionadas con las secuencias genómicas de los mismos genes en los organismos *Mus musculus* y *Homo sapiens* (ratón y humano). Para esto se empleó la base de datos EPD (*Eukaryotic Promoter Database*) [221], de la cual se extrajeron las secuencias de los promotores en dichos organismos; y el *software* Meme Suite [222] con el que se compararon las secuencias en búsqueda de motivos conservados, particularmente en las regiones del promotor mínimo y de elementos proximales. Esto se realizó con el fin de asegurar que la secuencia aislada de CHO-K1 pertenezca al promotor del gen correspondiente en cada caso, y que la misma contenga diversos elementos regulatorios de la actividad transcripcional.

Por otro lado, se incorporaron a las secuencias de los *primers* diseñados los sitios de reconocimiento para las enzimas de restricción *SacI* (en el *primer forward*) y *KpnI* (en el *primer reverse*), lo cual permitió clonar posteriormente los promotores en el vector pZsGreen1-1 (Clontech), como se detalla en la siguiente Sección.

Posteriormente, los promotores endógenos se amplificaron por duplicado mediante PCR y el resultado se sometió a una electroforesis en gel de agarosa, como se observa en la Figura 42.

Solamente se obtuvieron productos de PCR esperados para los promotores de los genes: *Hspa5*, *Hmox1*, *Rps18*, *Lgals1*, *Vim*, *Calr* y *Fth1*. Los fragmentos correspondientes se escindieron del gel, se purificaron y cuantificaron. Para los demás promotores, se probaron diferentes condiciones de reacción de PCR, obteniéndose resultados negativos en todos los casos. Por ello, se diseñaron nuevos pares de *primers* para tales secuencias y se realizaron las correspondientes reacciones de PCR; sin embargo, tampoco se pudieron amplificar dichos fragmentos.

Una de las razones principales, particularmente para los genes *S100a6*, *Ybx1* y *Cd63*, radica en la falta de secuencia genómica en la región del promotor, ya que como se ha mencionado, el genoma de células CHO-K1 se encuentra aún en estado borrador; esto implica que ciertas regiones del genoma no estén secuenciadas o bien ensambladas, lo cual dificulta este tipo de experimentos. Por lo tanto, se decidió proceder con la generación de nuevos vectores de expresión únicamente a partir de las secuencias promotoras amplificadas correctamente.

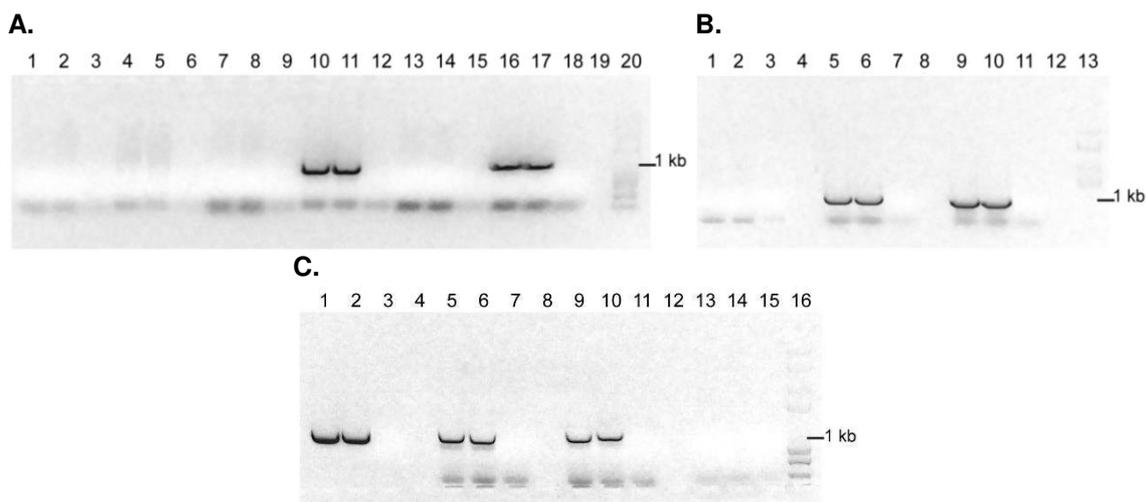


Figura 42. Electroforesis en geles de agarosa 1% (p/v) de los productos amplificados por PCR para el aislamiento de los promotores endógenos seleccionados.

- A.** Carriles 1-3: Actb (duplicados) y su control negativo. Carriles 4-6: Cd63 (duplicados) y su control negativo. Carriles 7-9: Ybx1 (duplicados) y su control negativo. Carriles 10-12: Rps18 (duplicados) y su control negativo. Carriles 13-15: S100a6 (duplicados) y su control negativo. Carriles 16-18: Fth1 (duplicados) y su control negativo. Carril 20: marcador de peso molecular *Trans2K Plus II DNA Marker* (TransGen Biotech).
- B.** Carriles 1-3: Lamp1 (duplicados) y su control negativo. Carriles 5-7: Carl (duplicados) y su control negativo. Carriles 9-11: Hmox1 (duplicados) y su control negativo. Carril 13: marcador de peso molecular *Trans2K Plus II DNA Marker* (TransGen Biotech).
- C.** Carriles 1-3: Hspa5 (duplicados) y su control negativo. Carriles 5-7: Lgals1 (duplicados) y su control negativo. Carriles 9-11: Vim (duplicados) y su control negativo. Carriles 13-15: Ctsz (duplicados) y su control negativo. Carril 16: marcador de peso molecular *Trans2K Plus II DNA Marker* (TransGen Biotech).

2.2. Construcción de nuevos vectores de expresión, conteniendo los promotores endógenos aislados, destinados a la producción de proteínas recombinantes en células de mamífero

En esta etapa de trabajo se usó la proteína verde fluorescente (ZsGreen1) para evaluar en células de mamíferos la actividad transcripcional de los promotores seleccionados. El empleo de esta proteína es muy útil ya que permite monitorear los niveles y estabilidad de expresión de manera rápida y sencilla, empleando citometría de flujo. Para ello, se utilizó el vector pZsGreen1-1 (Clontech). También, se usó el plásmido pCMV-ZsGreen1-1 (generado con anterioridad en nuestro laboratorio), que contiene la secuencia del promotor CMV, para contrastar la actividad transcripcional de los promotores endógenos.

Se amplificó y purificó el ADN de ambos vectores mediante una mini y maxipreparación de ADN plasmídico, lo que permitió obtener cantidad suficiente del ácido nucleico, necesario para los pasos siguientes. A continuación, se realizó la digestión simultánea del vector pZsGreen1-1 y los promotores endógenos purificados, con las enzimas de restricción *SacI* y *KpnI*. Los productos de las reacciones de digestión

se cuantificaron por lectura espectrofotométrica a $\lambda=260$ nm para efectuar, posteriormente, la ligación y transformación de células competentes. Se obtuvieron entre 4 y 10 colonias en la placa de V+I para cada clonado y se analizó la presencia del inserto (promotor) mediante reacción de PCR de colonias (Figura 43).

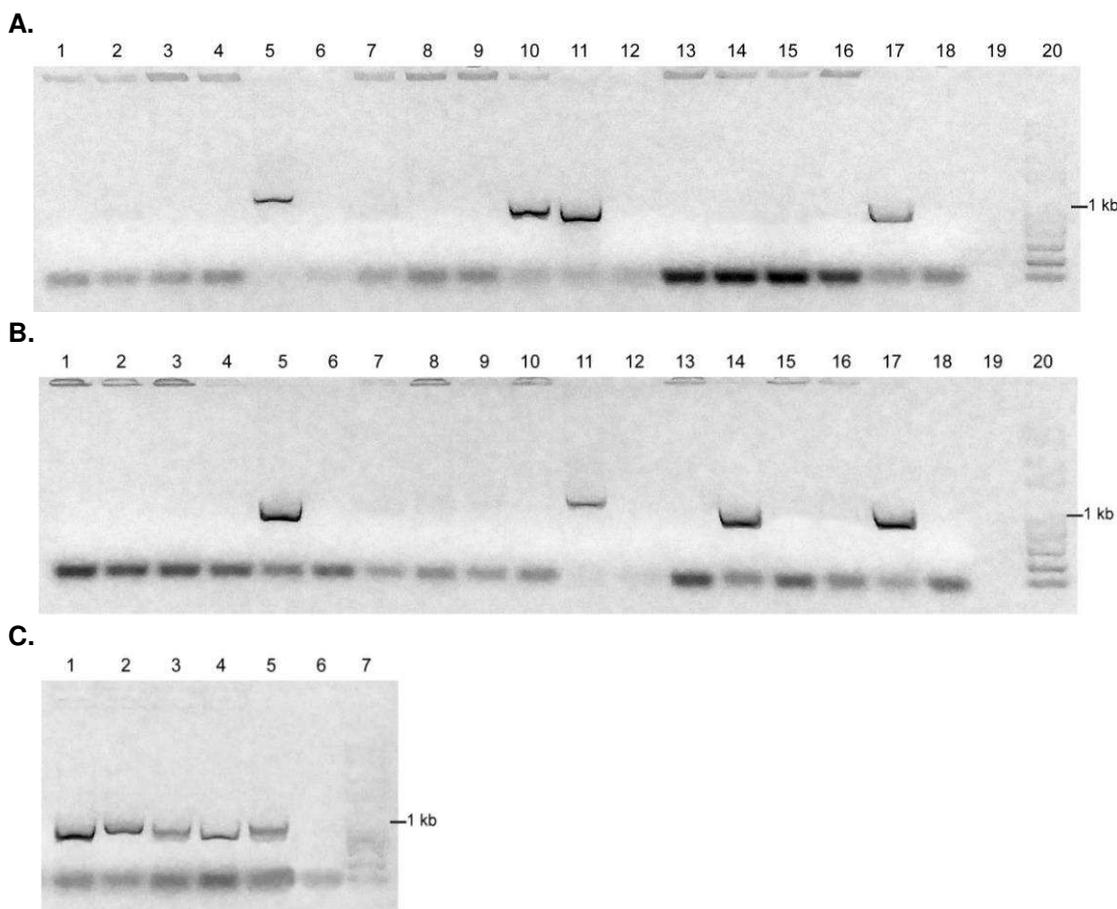


Figura 43. Electroforesis en geles de agarosa 1% (p/v) de los productos amplificados por PCR de colonias.

- A.** Carriles 1-4: productos obtenidos partir de las colonias recombinantes para Fth1. Carril 5: control positivo de reacción. Carril 6: control de reactivos de PCR. Carriles 7-10: productos obtenidos partir de las colonias recombinantes para Hspa5. Carril 11: control positivo de reacción. Carril 12: control de reactivos de PCR. Carriles 13-16: productos obtenidos partir de las colonias recombinantes para Vim. Carril 17: control positivo de reacción. Carril 18: control de reactivos de PCR. Carril 20: marcador de peso molecular *Trans2K Plus II DNA Marker* (TransGen Biotech).
- B.** Carriles 1-4: productos obtenidos partir de las colonias recombinantes para Lgals1. Carril 5: control positivo de reacción. Carril 6: control de reactivos de PCR. Carriles 7-10: productos obtenidos partir de las colonias recombinantes para Calr. Carril 11: control positivo de reacción. Carril 12: control de reactivos de PCR. Carriles 13-16: productos obtenidos partir de las colonias recombinantes para Rps18. Carril 17: control positivo de reacción. Carril 18: control de reactivos de PCR. Carril 20: marcador de peso molecular *Trans2K Plus II DNA Marker* (TransGen Biotech).
- C.** Carriles 1-4: productos obtenidos partir de las colonias recombinantes para Hmox1. Carril 5: control positivo de reacción. Carril 6: control de reactivos de PCR. Carril 7: marcador de peso molecular *Trans2K Plus II DNA Marker* (TransGen Biotech).

Como se observa en la Figura 43, solamente se consiguieron colonias positivas para los clonados de los promotores de los genes Rps18, Hspa5 y Hmox1, obteniéndose una colonia positiva para los dos primeros, y cuatro positivas para Hmox1.

Para aquellos clonados que no resultaron positivos, se repitió la reacción de ligación (probando diferentes relaciones inserto/vector) y posterior control por PCR de colonias, no obteniéndose resultados satisfactorios. Por este motivo, se continuó solamente con los clonados positivos mencionados. Se realizó la amplificación (minipreparación) de Rps18, Hspa5 y Hmox clones 1 y 2; y posterior digestión con las enzimas *SacI* y *KpnI* que liberaron el inserto (promotor de aproximadamente 1 kb), tal como se observa en la Figura 44. Estos clones, se amplificaron y criopreservaron a -70°C.

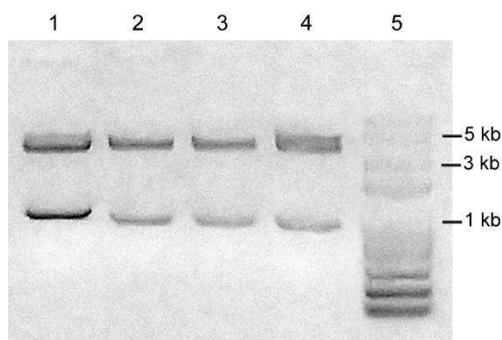


Figura 44. Electroforesis en geles de agarosa 1% (p/v) de las reacciones de digestión con las enzimas de restricción *SacI* y *KpnI* de los clonados de los promotores endógenos en pZsGreen1-1.

Producto de las digestiones de los clonados: Carril 1: Hspa5. Carriles 2-3: Hmox1 clones 1 y 2. Carril 4: Rps18. Carril 5: marcador de peso molecular *Trans2K Plus II DNA Marker*.

2.3. Evaluación de la actividad transcripcional de los nuevos vectores de expresión de forma transiente en células de mamífero

La línea celular CHO-K1 constituye uno de los sistemas de producción más empleados en la industria biofarmacéutica para la generación de proteínas recombinantes de uso terapéutico a gran escala. Por otro lado, la línea celular HEK es muy fácil de transfectar y, en consecuencia, permite generar líneas celulares con elevados niveles de producción de la proteína de interés. Por estos motivos, en esta etapa de trabajo no solamente se utilizó la línea celular CHO-K1, a partir de la cual se identificaron los promotores endógenos, sino que también se emplearon células HEK para el *screening* de la actividad transcripcional de los nuevos vectores de expresión generados.

Se llevó a cabo la transfección transitoria de células CHO-K1 (adaptadas al crecimiento en suspensión) y de células HEK (adherentes), con los plásmidos

construidos previamente, empleando lípidos catiónicos (Lipofectamina 2000, Invitrogen). A su vez, se transfectaron dichas células con el plásmido pZsGreen1-1 (sin promotor), el cual se usó como control basal; y con el vector pCMV-ZsGreen1-1, para comparar la actividad transcripcional de los promotores endógenos con la de un promotor usado habitualmente.

Para las células CHO-K1, se utilizaron placas de 12 pozos y medio de crecimiento correspondiente para células CHO adherentes, conteniendo 10% (v/v) de SFB. A las 72 h postransfección, con un volumen de las células CHO-K1 transfectadas (con cada plásmido) se inició el cultivo en suspensión en pequeña escala, empleando medio de cultivo Ex-Cell 302 libre de SFB. Las células HEK se cultivaron en placas de 12 pozos con medio de crecimiento para células HEK, conteniendo 10% (v/v) de SFB.

Por citometría de flujo se analizó la expresión transiente –durante 13 días (312 h) de cultivo– de la proteína verde fluorescente ZsGreen1 bajo el control del promotor viral CMV y de los promotores endógenos aislados: Hspa5, Rps18 y Hmox1 (considerando dos clones para este último). Asimismo, las líneas celulares cultivadas en adherencia fueron monitoreadas mediante microscopía de fluorescencia durante ese período.

En la Figura 45 se observan los diagramas de dispersión de puntos (*dot-plot*): SSC vs. FSC, para la selección de las poblaciones celulares de interés, generados a partir de las muestras de células 72 h postransfección. Los *gates* se mantuvieron iguales para todas las muestras analizadas, para cada línea celular.

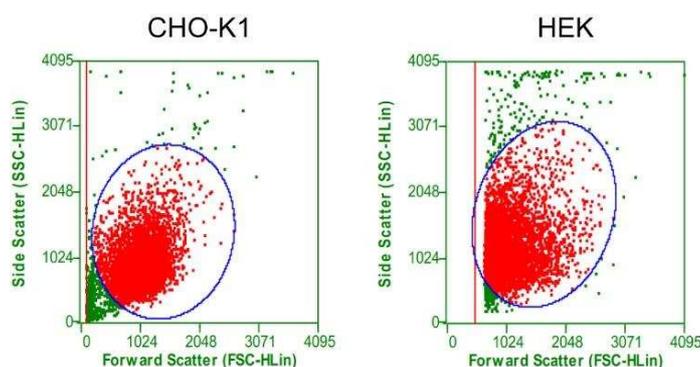


Figura 45. Gráficos SSC vs. FSC realizados con el software GUAVA ExpressPlus.

A modo de ejemplo, se muestran los datos obtenidos por citometría de flujo de las muestras 72 h postransfección con pZsGreen1-1 (sin promotor), de células CHO-K1 y HEK.

Los resultados de los análisis de la expresión transiente de la proteína ZsGreen1 (por microscopía de fluorescencia y citometría de flujo) se exponen en las Figuras 46-48 y en las Tablas 11-13, para células HEK, CHO-K1adh y CHO-K1s, respectivamente. No se muestran las imágenes de microscopía para las células CHO-K1 y HEK a las 312 h postransfección, ya que no se evidenció fluorescencia en dichas muestras.

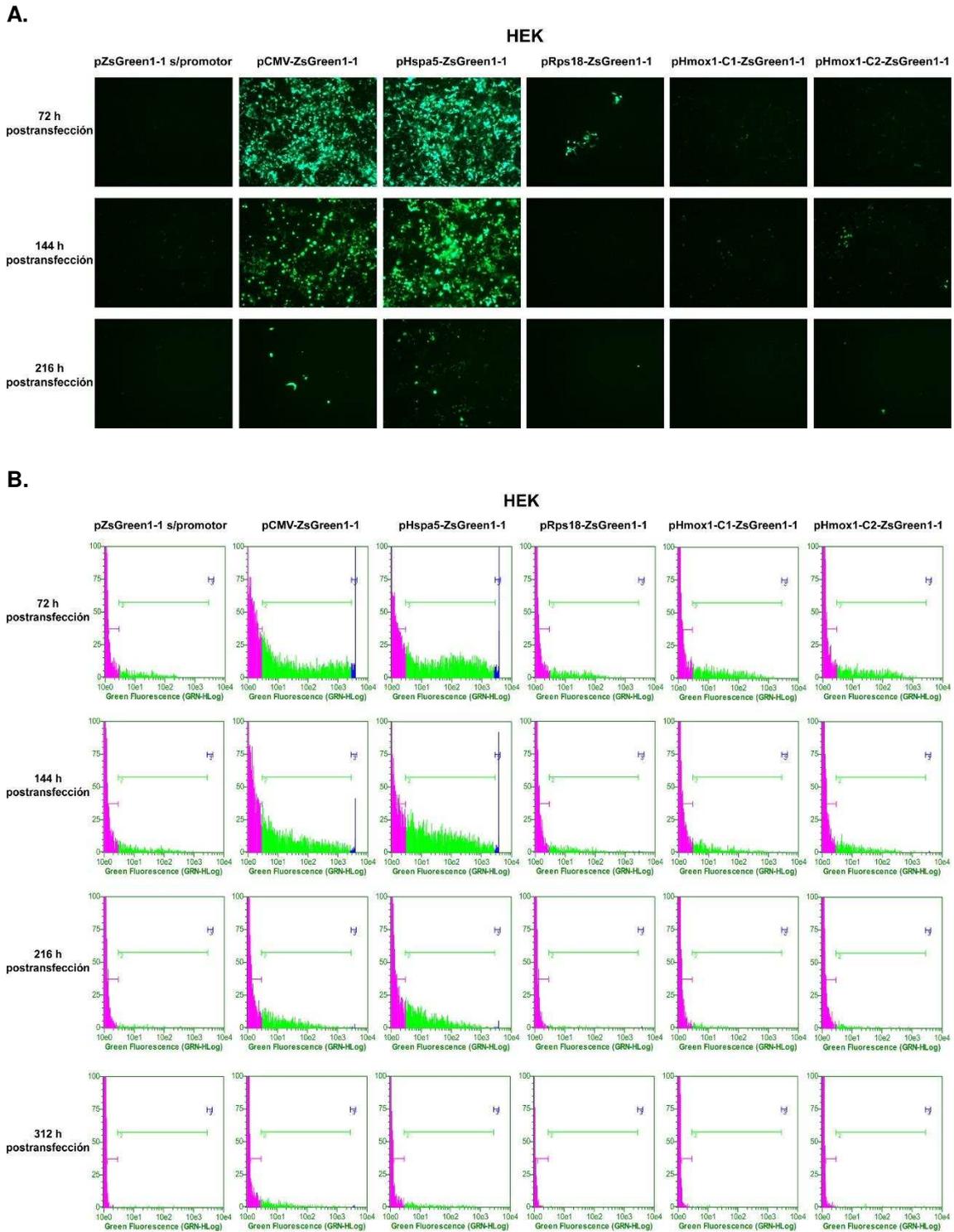


Figura 46. Evaluación de la actividad transcripcional de los vectores de expresión generados, de forma transiente en células HEK (adherentes).
A. Imágenes representativas de cada transfección, obtenidas por microscopía de fluorescencia (aumento 100X y tiempo de exposición 200 ms).
B. Histogramas (número de eventos en la población seleccionada vs. intensidad de fluorescencia de ZsGreen1) obtenidos por citometría de flujo. El Marcador 1 (M1 - fucsia) indica células negativas; el Marcador 2 (M2 - verde), células positivas con intensidad media de fluorescencia, y el Marcador 3 (M3 - azul), células positivas con alta intensidad de fluorescencia.

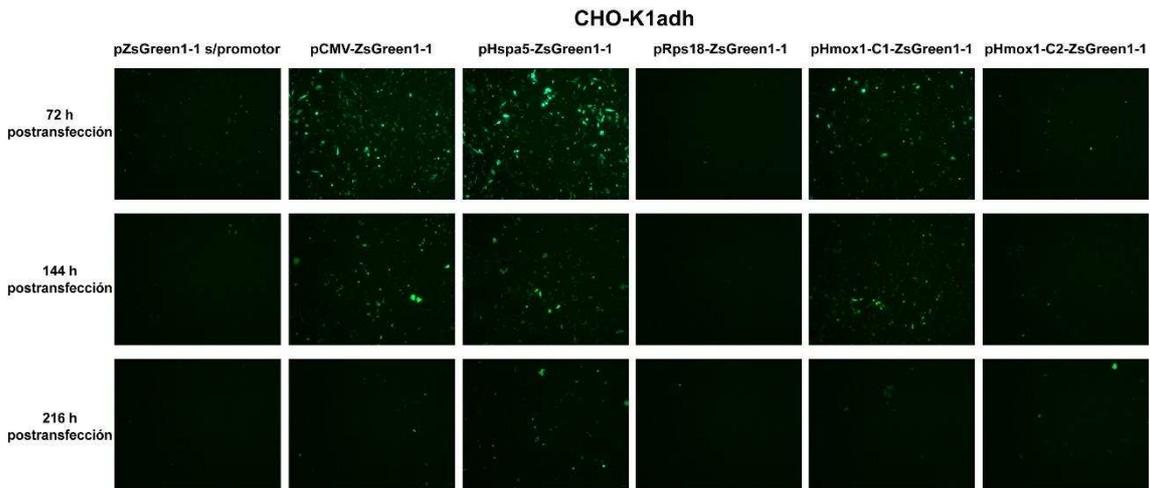
Tabla 11. Evaluación de la actividad transcripcional de los vectores de expresión generados, de forma transiente en células HEK.
Valores obtenidos por citometría de flujo.

| HEK | | | | | |
|------------------------------|----------|----------|---------------|----------|---------------|
| 72 h postransfección | | | | | |
| | M1 %* | M2 %* | M2 x-mean† | M3 %* | M3 x-mean† |
| pZsGreen1-1 s/promotor | 94,50 | 5,50 | 34 | 0 | 0 |
| pCMV-ZsGreen1-1 | 44,82 | 45,33 | 313 | 9,84 | 3763 |
| pHspa5-ZsGreen1-1 | 40,71 | 51,26 | 316 | 8,03 | 3780 |
| pRps18-ZsGreen1-1 | 93,31 | 6,69 | 39 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 86,06 | 13,94 | 86 | 0 | 0 |
| pHmox1-C2-ZsGreen1-1 | 86,93 | 13,07 | 79 | 0 | 0 |
| 144 h postransfección | | | | | |
| pZsGreen1-1 s/promotor | 93,26 | 6,74 | 195 | 0 | 0 |
| pCMV-ZsGreen1-1 | 58,22 | 40,35 | 180 | 1,44 | 3722 |
| pHspa5-ZsGreen1-1 | 44,48 | 52,72 | 210 | 2,80 | 3748 |
| pRps18-ZsGreen1-1 | 95,22 | 4,76 | 54 | 0,02 | 2985 |
| pHmox1-C1-ZsGreen1-1 | 92,17 | 7,83 | 63 | 0 | 0 |
| pHmox1-C2-ZsGreen1-1 | 91,54 | 8,42 | 54 | 0,04 | 3867 |
| 216 h postransfección | | | | | |
| pZsGreen1-1 s/promotor | 98,86 | 1,14 | 33 | 0 | 0 |
| pCMV-ZsGreen1-1 | 85,25 | 14,67 | 60 | 0,08 | 3740 |
| pHspa5-ZsGreen1-1 | 79,91 | 22,94 | 64 | 0,15 | 3667 |
| pRps18-ZsGreen1-1 | 99,27 | 0,70 | 32 | 0,02 | 3909 |
| pHmox1-C1-ZsGreen1-1 | 98,82 | 1,18 | 23 | 0 | 0 |
| pHmox1-C2-ZsGreen1-1 | 98,47 | 1,53 | 33 | 0 | 0 |
| 312 h postransfección | | | | | |
| pZsGreen1-1 s/promotor | 99,79 | 0,21 | 71 | 0 | 0 |
| pCMV-ZsGreen1-1 | 95,29 | 4,62 | 62 | 0,08 | 3778 |
| pHspa5-ZsGreen1-1 | 94,82 | 5,18 | 65 | 0 | 0 |
| pRps18-ZsGreen1-1 | 99,91 | 0,09 | 4 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 99,85 | 0,15 | 11 | 0 | 0 |
| pHmox1-C2-ZsGreen1-1 | 99,75 | 0,25 | 348 | 0 | 0 |

* M1 representa el porcentaje de células negativas (que no expresan ZsGreen1); M2, el porcentaje de células positivas con intensidad media de fluorescencia, y M3, el porcentaje de células positivas con alta intensidad de fluorescencia.

† M2 x-mean indica la media de los valores de fluorescencia dentro de M2, y M3 x-mean, la media de los valores de fluorescencia dentro de M3.

A.



B.

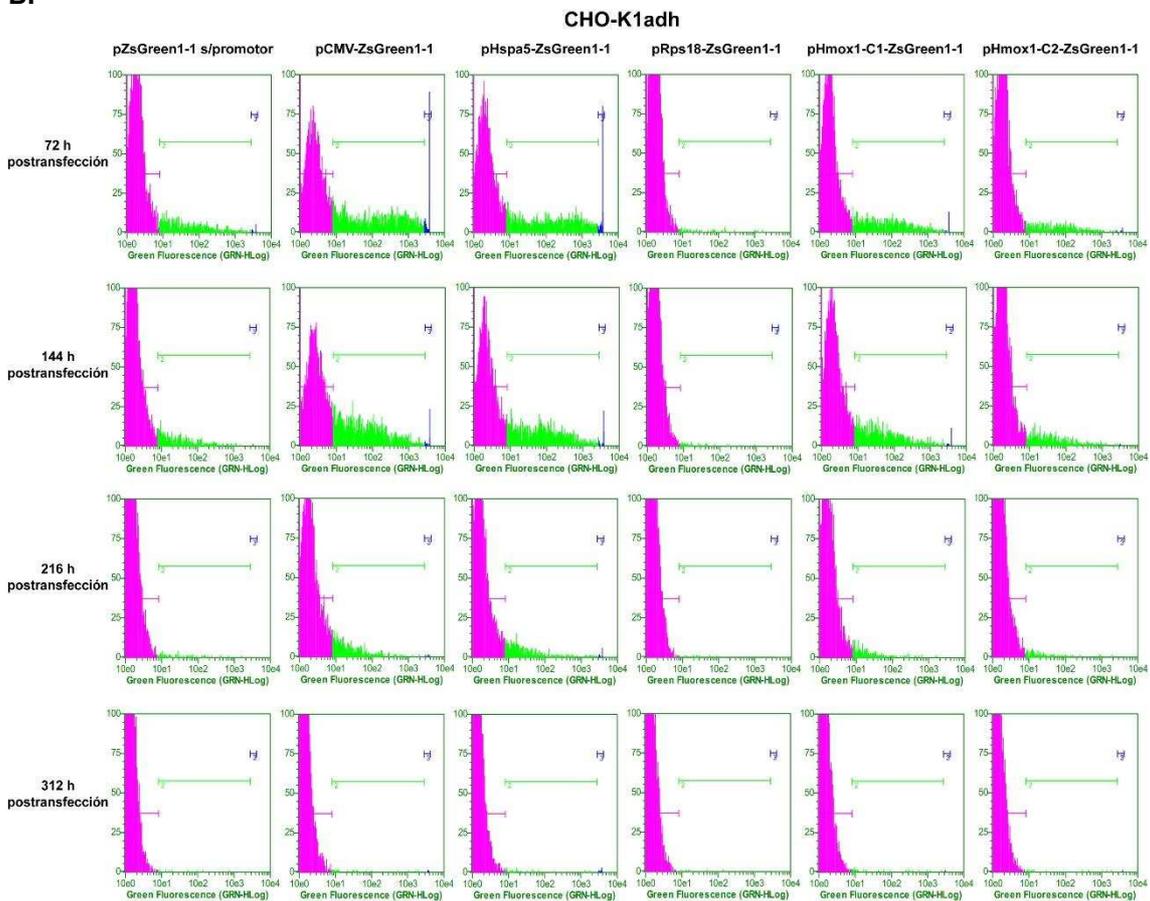


Figura 47. Evaluación de la actividad transcripcional de los vectores de expresión generados, de forma transiente en células CHO-K1adh.

A. Imágenes representativas de cada transfección, obtenidas por microscopía de fluorescencia (aumento 100X y tiempo de exposición 300 ms).

B. Histogramas (número de eventos en la población seleccionada vs. intensidad de fluorescencia de ZsGreen1) obtenidos por citometría de flujo. El Marcador 1 (M1 - fucsia) indica células negativas; el Marcador 2 (M2 - verde), células positivas con intensidad media de fluorescencia, y el Marcador 3 (M3 - azul), células positivas con alta intensidad de fluorescencia.

Tabla 12. Evaluación de la actividad transcripcional de los vectores de expresión generados, de forma transiente en células CHO-K1adh.
Valores obtenidos por citometría de flujo.

| CHO-K1adh | | | | | |
|------------------------------|----------|----------|---------------|----------|---------------|
| 72 h postransfección | | | | | |
| | M1 %* | M2 %* | M2 x-mean† | M3 %* | M3 x-mean† |
| pZsGreen1-1 s/promotor | 86,05 | 13,73 | 155 | 0,22 | 3524 |
| pCMV-ZsGreen1-1 | 62,32 | 34,43 | 384 | 3,25 | 3641 |
| pHspa5-ZsGreen1-1 | 69,06 | 28,12 | 395 | 2,82 | 3684 |
| pRps18-ZsGreen1-1 | 98,82 | 1,18 | 134 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 80,80 | 18,8 | 202 | 0,39 | 3794 |
| pHmox1-C2-ZsGreen1-1 | 90,71 | 9,16 | 145 | 0,13 | 3625 |
| 144 h postransfección | | | | | |
| pZsGreen1-1 s/promotor | 90,39 | 9,59 | 79 | 0,02 | 3379 |
| pCMV-ZsGreen1-1 | 63,57 | 35,67 | 177 | 0,76 | 3693 |
| pHspa5-ZsGreen1-1 | 68,90 | 30,28 | 212 | 0,82 | 3745 |
| pRps18-ZsGreen1-1 | 99,12 | 0,88 | 55 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 77,13 | 22,55 | 141 | 0,32 | 3735 |
| pHmox1-C2-ZsGreen1-1 | 89,64 | 10,34 | 107 | 0,02 | 3158 |
| 216 h postransfección | | | | | |
| pZsGreen1-1 s/promotor | 98,61 | 1,39 | 85 | 0 | 0 |
| pCMV-ZsGreen1-1 | 88,69 | 11,24 | 53 | 0,06 | 3503 |
| pHspa5-ZsGreen1-1 | 90,03 | 9,76 | 88 | 0,22 | 3675 |
| pRps18-ZsGreen1-1 | 99,79 | 0,21 | 30 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 94,85 | 5,15 | 58 | 0 | 0 |
| pHmox1-C2-ZsGreen1-1 | 97,51 | 2,50 | 67 | 0 | 0 |
| 312 h postransfección | | | | | |
| pZsGreen1-1 s/promotor | 99,79 | 0,21 | 110 | 0 | 0 |
| pCMV-ZsGreen1-1 | 99,41 | 0,53 | 229 | 0,06 | 3778 |
| pHspa5-ZsGreen1-1 | 99,35 | 0,51 | 158 | 0,15 | 3600 |
| pRps18-ZsGreen1-1 | 99,87 | 0,13 | 40 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 99,28 | 0,70 | 228 | 0,02 | 3156 |
| pHmox1-C2-ZsGreen1-1 | 99,69 | 0,29 | 379 | 0,02 | 3922 |

* M1 representa el porcentaje de células negativas (que no expresan ZsGreen1); M2, el porcentaje de células positivas con intensidad media de fluorescencia, y M3, el porcentaje de células positivas con alta intensidad de fluorescencia.

† M2 x-mean indica la media de los valores de fluorescencia dentro de M2, y M3 x-mean, la media de los valores de fluorescencia dentro de M3.

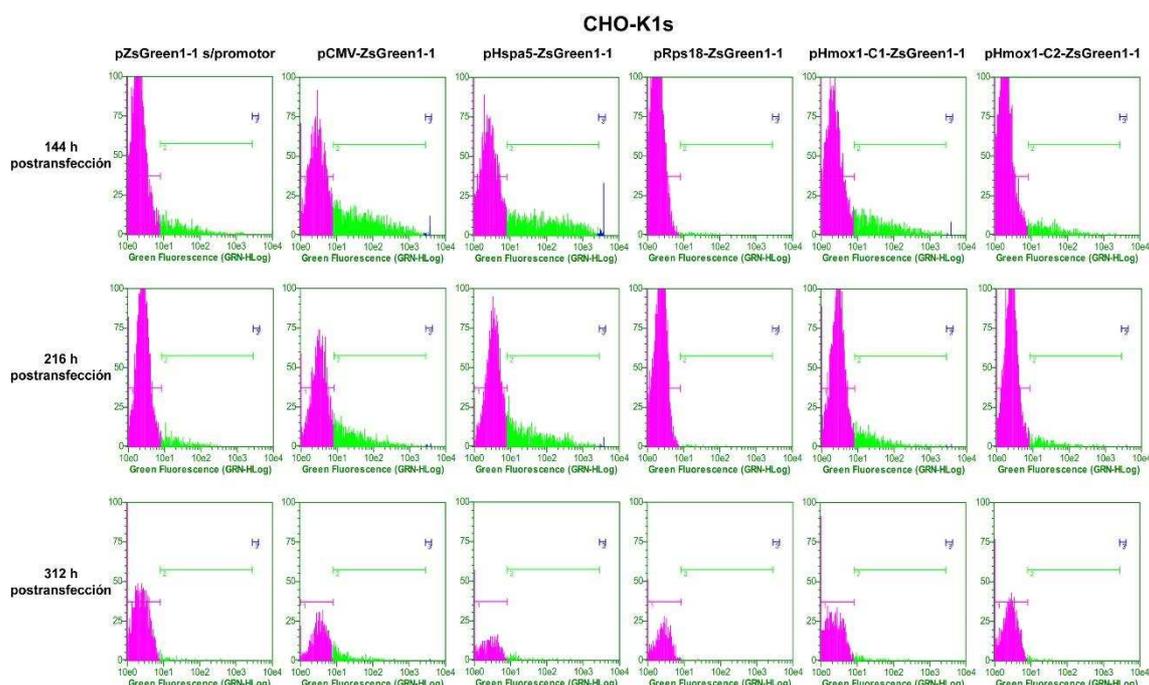


Figura 48. Evaluación de la actividad transcripcional de los vectores de expresión generados, de forma transiente en células CHO-K1s (suspensión).

Histogramas (número de eventos en la población seleccionada vs. intensidad de fluorescencia de ZsGreen1) obtenidos por citometría de flujo. El Marcador 1 (M1 - fucsia) indica células negativas; el Marcador 2 (M2 - verde), células positivas con intensidad media de fluorescencia, y el Marcador 3 (M3 - azul), células positivas con alta intensidad de fluorescencia.

Tabla 13. Evaluación de la actividad transcripcional de los vectores de expresión generados, de forma transiente en células CHO-K1s.

Valores obtenidos por citometría de flujo.

| CHO-K1s | | | | | |
|------------------------|----------|----------|---------------|----------|---------------|
| 144 h postransfección | | | | | |
| | M1 %* | M2 %* | M2 x-mean† | M3 %* | M3 x-mean† |
| pZsGreen1-1 s/promotor | 88,98 | 11,02 | 78 | 0 | 0 |
| pCMV-ZsGreen1-1 | 64,81 | 34,63 | 166 | 0,56 | 3659 |
| pHspa5-ZsGreen1-1 | 67,09 | 31,74 | 259 | 1,17 | 3681 |
| pRps18-ZsGreen1-1 | 99,11 | 0,89 | 30 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 79,45 | 20,32 | 130 | 0,23 | 3812 |
| pHmox1-C2-ZsGreen1-1 | 90,55 | 9,45 | 86 | 0 | 0 |
| 216 h postransfección | | | | | |
| pZsGreen1-1 s/promotor | 93,43 | 6,57 | 37 | 0 | 0 |
| pCMV-ZsGreen1-1 | 75,42 | 25,45 | 83 | 0,12 | 3447 |
| pHspa5-ZsGreen1-1 | 72,87 | 26,83 | 130 | 0,29 | 3640 |

| | | | | | |
|------------------------------|-------|-------|----|------|------|
| pRps18-ZsGreen1-1 | 99,58 | 0,42 | 33 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 87,14 | 12,78 | 74 | 0,08 | 3544 |
| pHmox1-C2-ZsGreen1-1 | 94,17 | 5,81 | 69 | 0,03 | 3922 |
| 312 h postransfección | | | | | |
| pZsGreen1-1 s/promotor | 97,80 | 2,20 | 28 | 0 | 0 |
| pCMV-ZsGreen1-1 | 82,00 | 17,86 | 34 | 0,09 | 3860 |
| pHspa5-ZsGreen1-1 | 90,20 | 9,80 | 39 | 0 | 0 |
| pRps18-ZsGreen1-1 | 99,62 | 0,38 | 14 | 0 | 0 |
| pHmox1-C1-ZsGreen1-1 | 95,55 | 4,45 | 31 | 0 | 0 |
| pHmox1-C2-ZsGreen1-1 | 97,93 | 2,07 | 14 | 0 | 0 |

* M1 representa el porcentaje de células negativas (que no expresan ZsGreen1); M2, el porcentaje de células positivas con intensidad media de fluorescencia, y M3, el porcentaje de células positivas con alta intensidad de fluorescencia.

† M2 *x-mean* indica la media de los valores de fluorescencia dentro de M2, y M3, *x-mean* la media de los valores de fluorescencia dentro de M3.

De estos resultados puede evidenciarse que el vector pZsGreen1-1 (sin promotor) presenta actividad transcripcional basal, probablemente debido a la presencia del promotor fuerte SV40 en dicho plásmido, actuando como promotor del gen ZsGreen1 aún a una distancia significativa.

En todos los casos, la actividad transcripcional máxima se advirtió entre las 72 h y 144 h postransfección, para luego ir disminuyendo con el transcurso de los días de cultivo, lo cual es un resultado esperado, por ser un ensayo en forma transiente.

El promotor endógeno Rps18 exhibió muy baja actividad transcripcional, incluso llamativamente menor, en algunos casos, al plásmido sin promotor. En células HEK, el promotor Hmox1 mostró expresión similar en sus dos clones; mientras que en las células CHO-K1, el Clon 1 presentó aproximadamente el doble de actividad transcripcional que el Clon 2, pero siempre menor al promotor viral CMV empleado como control.

El resultado más promisorio se obtuvo con el promotor endógeno Hspa5, el cual mostró un comportamiento equivalente al promotor CMV, en células HEK y CHO-K1 (tanto en crecimiento en adherencia como en suspensión). En las células HEK se lograron valores superiores al 50% de células positivas, que expresaron la proteína verde fluorescente analizada; mientras que en las células CHO este porcentaje fue menor al 40%, probablemente debido a que la transfección en células HEK es más eficiente.

A su vez, al analizar los valores de las medias de intensidad de fluorescencia (*x-mean*) para las células en M2 y M3, en general dichos valores fueron superiores para

Hspa5, al contrastar con CMV; es decir, que las células que expresan la proteína ZsGreen1, lo hacen con mayor intensidad bajo el control del promotor endógeno Hspa5.

Por otro lado, al evaluar la actividad transcripcional de los promotores en las células CHO-K1 transfectadas cultivadas en suspensión y con medio libre de SFB, se observaron valores ligeramente superiores de la actividad transcripcional de todos los promotores durante los 13 días que duró el ensayo.

En conclusión, se lograron identificar y aislar promotores endógenos a partir de los análisis transcripcionales de células CHO-K1 adaptadas al crecimiento en suspensión. Por diversas razones ya expuestas, sólo se pudo evaluar la actividad transcripcional de los promotores correspondientes a los genes Rps18, Hspa5 y Hmox1; de un total de 13 promotores seleccionados.

El *screening* de la actividad transcripcional de los mismos se realizó únicamente de manera transiente, con el objetivo de obtener de forma rápida un panorama global del comportamiento en células de mamífero de las secuencias promotoras aisladas. Los estudios realizados arrojaron resultados muy prometedores, ya que se pudo establecer un esquema racional de trabajo para la identificación de nuevos promotores, a partir del análisis de los perfiles de expresión génica de células CHO-K1 cultivadas en biorreactor, el aislamiento de las secuencias y generación de nuevos vectores de expresión. Esto se llevó a cabo con el objetivo de reemplazar el uso de promotores virales para la producción de proteínas recombinantes en células de mamíferos; el cual es un campo que está comenzando a estudiarse en la actualidad.

En particular, el promotor Hspa5 mostró ser una gran alternativa al empleo del promotor CMV, tanto en células CHO-K1 como en células HEK. El hecho de que se obtuvieran resultados similares en ambas líneas celulares, a pesar de que las secuencias fueron aisladas del genoma de células CHO-K1, abre un abanico de posibilidades para el estudio de nuevos promotores en diversas líneas celulares y en diferentes condiciones de cultivo (adherencia/suspensión, *batch*/continuo).

3. Perspectivas futuras

Los ensayos descritos anteriormente demostraron que las secuencias aisladas efectivamente tienen actividad promotora en células de mamífero. Por lo cual, experimentos a futuro implicarían caracterizar dichas secuencias, mediante secuenciación de ADN y búsqueda de motivos funcionales y dominios conservados; y también evaluar la actividad transcripcional de los promotores endógenos de manera estable.

De igual manera, se podrían emplear otras líneas celulares, y comparar además con otros promotores ampliamente usados en el área de producción de proteínas recombinantes terapéuticas, como por ejemplo SV40 y EF1 α . A su vez, en los próximos análisis se propone la co-transfección con otra proteína fluorescente (como DsRed), con el objetivo de poder normalizar dichos ensayos por eficiencia de transfección.

Asimismo, estos promotores se deberían evaluar en las mismas condiciones en las cuales se llevaron a cabo los experimentos para RNA-seq, cuyos análisis permitieron obtener los diferentes perfiles de expresión génica, en las etapas de crecimiento celular exponencial y estacionaria, a partir de los cuales se identificaron los genes “candidatos” para esta etapa de trabajo.

Por otro lado, particularmente para aquellos genes a partir de los cuales no se pudieron amplificar sus promotores, principalmente debido a falta de secuencia genómica o anotación insuficiente del genoma de la línea celular CHO-K1, se decidió aplicar la reciente tecnología de edición genética CRISPR/Cas9.

3.1. Experimentos con el sistema CRISPR/Cas9 en células CHO-K1

Se decidió plantear experimentos de *knock-in* utilizando el sistema CRISPR/Cas9, debido principalmente al hecho de que para los cuatro genes “candidatos” que exhibieron mayor expresión (S100a6, Fth1, Lgals1 y Vim) no se pudieron llevar a cabo los experimentos de evaluación de la actividad transcripcional, ya sea porque para algunos no se logró aislar sus secuencias promotoras o porque para otros no resultaron positivos los clonados en el vector pZsGreen1-1. De esta manera, se pretende insertar el TG de interés en el genoma de la célula huésped, bajo el control de tales promotores endógenos, evitando de esta forma el empleo de secuencias virales y otras exógenas.

Como se mencionó en la Introducción de este Capítulo es fundamental desarrollar metodologías que permitan encontrar y reutilizar los sitios denominados *hotspots* a través de la introducción de diferentes transgenes que codifiquen para diversas proteínas de interés terapéutico. Para esto, la tecnología CRISPR/Cas9 abre un abanico de posibilidades para la edición precisa de células CHO huésped destinadas a la producción de proteínas recombinantes, mejorando así la productividad y la calidad del producto y reduciendo también los costos de producción. Diversas investigaciones han logrado exitosamente insertar genes de interés en *loci* específicos de células CHO, demostrando que la integración sitio específica de transgenes en diversos *hotspots* de células CHO es una estrategia ideal para superar los desafíos que implica la integración aleatoria convencional de un gen de interés.

En síntesis, se propone insertar un *cassette* de *knock-in* en el extremo 3' UTR de cada gen seleccionado, con el objetivo de utilizar el promotor y las regiones reguladoras del gen endógeno para expresar un gen reportero, que luego se podrá intercambiar mediante tecnología de recombinación de ADN por un TG de interés terapéutico. Para ello se utilizará CRISPR/Cas9 sin alterar la expresión del gen endógeno y sin introducir regiones virales ni antibióticos de selección, optimizando así aspectos regulatorios de la producción de proteínas recombinantes terapéuticas.

Para estos experimentos, se utilizó el vector lentiCRISPRv2, que fue donado a nuestro laboratorio por el *Salk Institute* (La Jolla, CA, EE.UU.). El mismo es un plásmido lentiviral de tercera generación de aproximadamente 15 kb [223]. Además, es un vector del tipo “todo en uno”, ya que expresa la endonucleasa Cas9 bajo el control del promotor EF1 α y también el sgRNA bajo el promotor U6.

Inicialmente, se amplificó y purificó el ADN plasmídico de dicho vector y luego se realizó una digestión con las enzimas de restricción *EcoRV* y *XbaI* para corroborar su identidad. El producto de la digestión se sometió a una electroforesis en gel de agarosa (Figura 49.A).

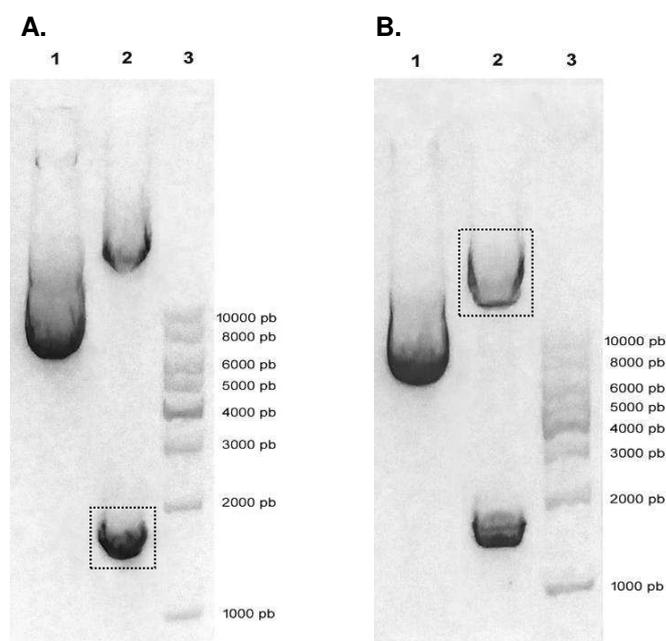


Figura 49. Electroforesis en gel de agarosa 1% (p/v)

- A.** Digestión del plásmido lentiCRISPRv2 con las enzimas *EcoRV* y *XbaI*. Carril 1: lentiCRISPRv2 sin digerir. Carril 2: producto de la digestión. Se encuentra señalada la banda de tamaño esperado de 1870 pb. Carril 3: marcador de peso molecular 1KB *DNA Ladder* (TransGen Biotech).
- B.** Digestión del plásmido lentiCRISPRv2 con la enzima *BsmBI*. Carril 1: lentiCRISPRv2 sin digerir. Carril 2: producto de la digestión. Se encuentra señalada la banda de 12988 pb que se purificó del gel. Carril 3: marcador de peso molecular 1KB *DNA Ladder* (TransGen Biotech).

Se observó la liberación del fragmento esperado (1870 pb). Posteriormente, el plásmido lentiCRISPRv2 fue digerido con la enzima de restricción *BsmBI*. El resultado se separó por electroforesis en gel de agarosa (Figura 49.B) y el fragmento del tamaño esperado (12988 pb) se escindió del gel y se purificó.

3.1.1. Experimento de *knock-out* con el sistema CRISPR/Cas9 en células CHO-K1 adherentes productoras de EGFP

En primer lugar, se llevó a cabo un ensayo de *knock-out* de un clon CHO-K1 productor de EGFP, a través del mecanismo celular NHEJ, con la finalidad de implementar y optimizar el sistema CRISPR/Cas9 en células CHO. En la Figura 50 se muestra el esquema de trabajo aplicado.

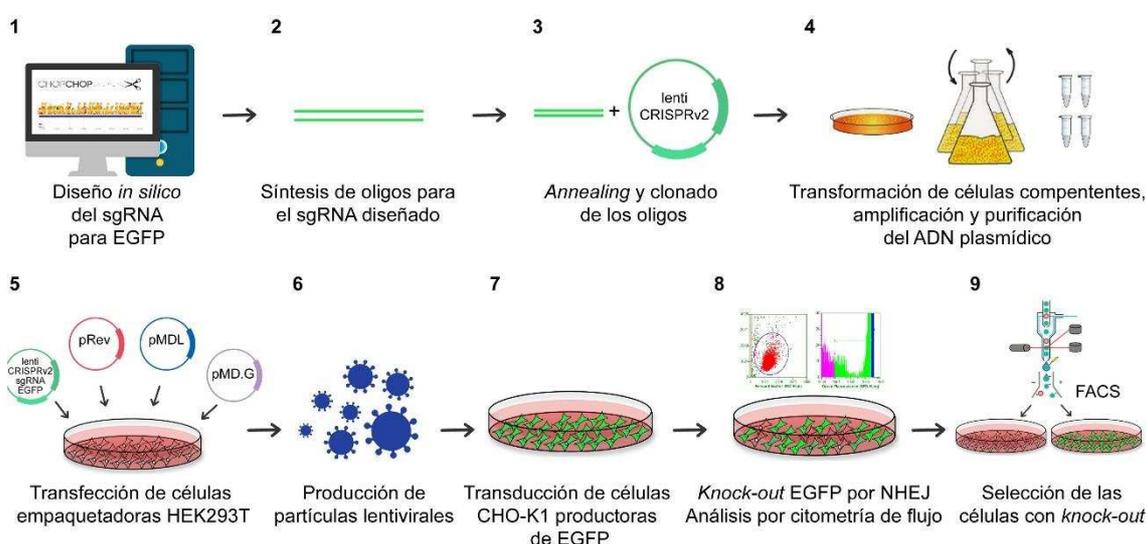


Figura 50. Esquema de trabajo llevado a cabo para el experimento de *knock-out* de EGFP en células CHO-K1adh.

Se diseñó y sintetizó el sgRNA, considerando como secuencia *target* la de EGFP, y se clonó al plásmido lentiviral, generándose el vector lentiCRISPRv2-sgRNA-EGFP. Se detallan los pasos y técnicas empleadas en la Sección de Materiales y Métodos de este Capítulo.

Se procedió a ensamblar partículas lentivirales a partir de los vectores estructurales pMDLg/pRRE, pMD.G, pRSV-Rev y el vector de transferencia generado lentiCRISPRv2-sgRNA-EGFP, mediante transfección mediada por lípidos catiónicos en células empaquetadoras HEK. Transcurridas 48 h postransfección se cosecharon los sobrenadantes de cultivo conteniendo las partículas lentivirales, se concentraron y se almacenaron a -70°C.

Luego, se transdujeron células de un clon CHO-K1adh, que contiene una sola copia de EGFP, y se mantuvieron en cultivo durante un mes. Aproximadamente cada tres días se evaluó mediante citometría de flujo (Guava easyCyte) la intensidad de expresión de la proteína verde fluorescente EGFP (Figura 51). Posteriormente, se criopreservaron las células transducidas luego de transcurrido un mes de cultivo, y seis meses después se descongelaron para evaluar la estabilidad del *knock-out* generado. Se realizó un ensayo de FACS (*fluorescence-activated cell sorting*) de la población celular que había perdido la expresión de la proteína fluorescente, para evaluar la eficiencia del procedimiento.

En la Figura 51 se muestran los histogramas de la intensidad de fluorescencia de EGFP, para cuatro días representativos del ensayo. Se utilizó como control negativo la línea celular CHO-K1 wt (adherente) y como control positivo células del clon empleado para el *knock-out*. Se muestra también el histograma de fluorescencia resultante de la evaluación posFACS de la población que presenta el *knock-out* para EGFP.

Al transcurrir los días de cultivo se aprecia claramente que la población celular que expresa EGFP va disminuyendo. La población presente en el Marcador 3 (M3) se compone de células que expresan EGFP con alta intensidad de fluorescencia. Se observa un corrimiento de las células desde el M3 al M2 (intensidad media de fluorescencia) y, finalmente, hacia M1 (población que no expresa EGFP).

En la Tabla 14 se muestran los porcentajes de células en cada población para cada día analizado. Cabe aclarar que no se logró el 100% de células *knock-out*. Esto puede deberse a muchos factores entre los cuales podemos mencionar: la eficiencia del sgRNA diseñado o la eficacia de la maquinaria de reparación celular por recombinación no homóloga (NHEJ) que no siempre genera *in-deles* o mutaciones al producirse un corte en la hebra de ADN que provoquen el *knock-out*.

De todas maneras, al llevar a cabo un proceso de FACS de la población de interés, se produjo un enriquecimiento eficiente de células con el *knock-out* deseado. Las células que fueron seleccionadas mediante FACS se mantuvieron en cultivo durante 48 h y se les realizó citometría de flujo nuevamente, revelando resultados similares al primer análisis posFACS mostrado. Se demostró así que el cambio generado en el ADN es permanente y estable en el tiempo.

Con este experimento pudimos comprobar que el vector funciona correctamente al expresar Cas9 y el sgRNA clonado. Además, el resultado del análisis mediante FACS es útil para aplicar en el ensayo de *knock-in* planteado (Sección 3.1.2), para el posterior enriquecimiento de las células que hayan incorporado eficientemente el *cassette* reportero diseñado.

Cabe destacar que si bien en este ensayo, el *knock-out* generado no tiene otro propósito funcional más allá del *screening* y evaluación de la técnica, resulta que con este análisis se sentó precedente en nuestro laboratorio para futuros experimentos de *knock-out* de genes de interés funcional con el sistema CRISPR/Cas9.

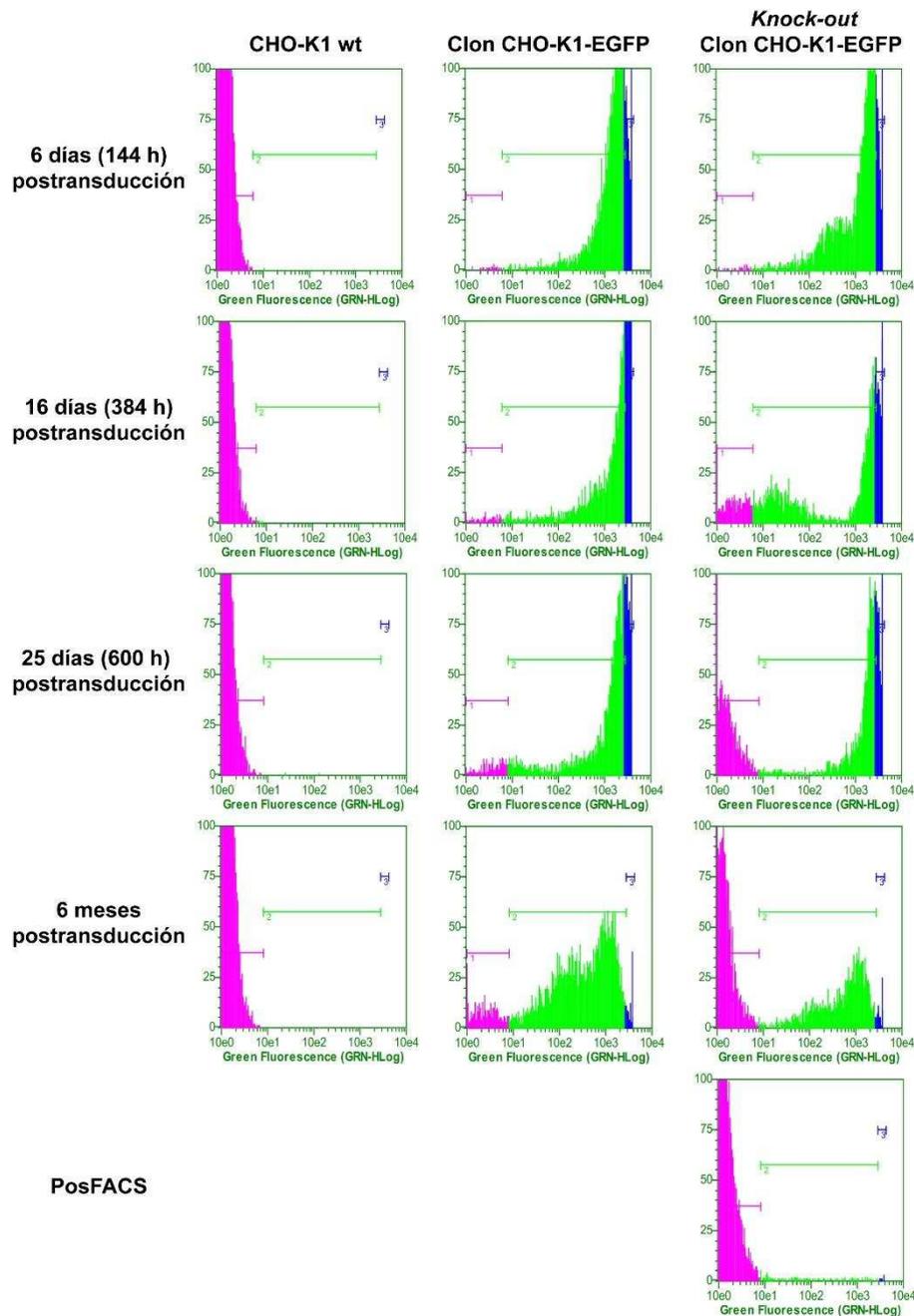


Figura 51. Histogramas de los niveles de fluorescencia de EGFP obtenidos por citometría de flujo.

Muestras correspondientes a cuatro días representativos del experimento de *knock-out* en células CHO-K1adh. Se muestran los controles negativo CHO-K1 *wild type* y positivo Clon CHO-K1 productor de EGFP. El Marcador 1 (M1 - fucsia) corresponde a la población celular que no expresa EGFP; el Marcador 2 (M2 - verde), población que expresa EGFP, y el Marcador 3 (M3 - azul), población celular que expresa EGFP en intensidades altas.

Tabla 14. Resultados de la citometría de flujo para el experimento de *knock-out* de EGFP en células CHO-K1adh.

Porcentajes de células en cada población por cada muestra, por día analizado.

| | CHO-K1 <i>wild type</i> | | | Clon CHO-K1 EGFP | | | Clon CHO-K1 EGFP <i>knock-out</i> | | |
|---------------------------------|-------------------------|------|------|------------------|-------|-------|-----------------------------------|-------|-------|
| | M1 % | M2 % | M3 % | M1 % | M2 % | M3 % | M1 % | M2 % | M3 % |
| 6 días pos transducción | 99,98 | 0,02 | 0,00 | 0,53 | 78,16 | 21,31 | 0,58 | 79,84 | 19,58 |
| 16 días pos transducción | 99,85 | 0,15 | 0,00 | 1,74 | 47,89 | 50,37 | 13,07 | 58,22 | 28,71 |
| 25 días pos transducción | 99,96 | 0,04 | 0,00 | 4,63 | 60,05 | 35,32 | 32,20 | 42,31 | 25,49 |
| 6 meses pos transducción | 100,00 | 0,00 | 0,00 | 9,61 | 88,02 | 2,30 | 57,89 | 40,40 | 1,71 |
| PosFACS | | | | | | | 98,00 | 1,87 | 0,13 |

3.1.2. Experimento de *knock-in* mediante CRISPR/Cas9 en células CHO-K1

El protocolo general de trabajo planteado para el ensayo de *knock-in* en células CHO-K1 se muestra en la Figura 52.

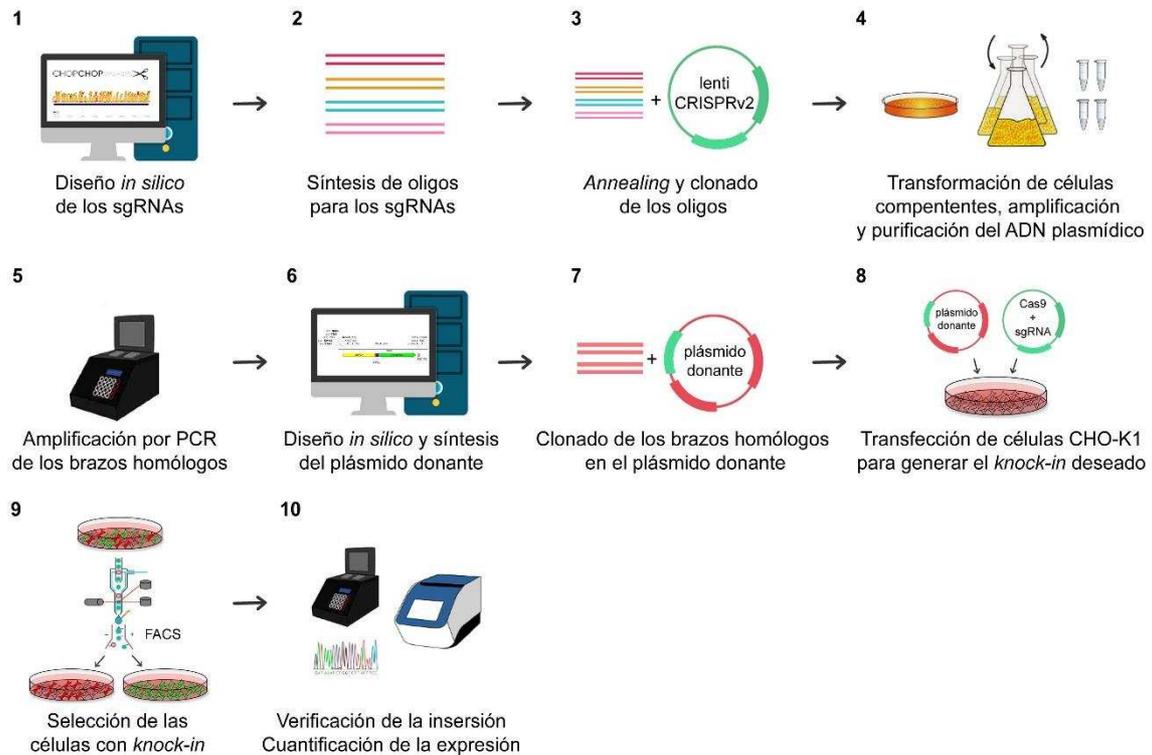


Figura 52. Esquema de trabajo planteado para el experimento de *knock-in* en células CHO-K1 adaptadas al crecimiento en suspensión y a medio libre de SFB.

El primer paso de dicho protocolo es el diseño *in silico* de los sgRNAs, a partir de la secuencia de la región específica donde se quiere realizar la edición genética. Para ello se utilizaron herramientas bioinformáticas desarrolladas para tal fin, que disponían del genoma de *Cricetulus griseus* para el análisis de *off-targets*.

La especificidad de la nucleasa Cas9 está determinada por la secuencia guía de 20 nucleótidos presente en el sgRNA. Para el sistema de *S. pyogenes* usado, la secuencia *target* debe preceder a la secuencia 5'-PAM "NGG"; así la secuencia ARN guía se empareja con la hebra complementaria y Cas9 produce el corte en ambas hebras del ADN a 3 pb *upstream* de la secuencia PAM.

Al momento de diseñar los sgRNAs se consideraron diferentes regiones genómicas (Figura 53). Para cada gen elegido se buscaron sitios *target* en las regiones 5' UTR y 3' UTR de cada gen y en las regiones 5' y 3' intergénicas. Las consideraciones específicas respectivas al diseño se detallan en la Sección 1.18.2 de Materiales y Métodos de este Capítulo.



Figura 53. Representaciones esquemáticas de las regiones genómicas *target* consideradas en el diseño de los sgRNAs, para el experimento de *knock-in* con el sistema CRISPR/Cas9.

Las flechas indican los sitios seleccionados para el diseño.

El principal objetivo de este trabajo es utilizar los sgRNAs de la región 3' UTR con el fin de introducir un *cassette* reportero (sin promotor) y así evaluar la expresión tanto del gen endógeno como del gen reportero, bajo el mismo promotor celular. La región 5' UTR tiene la misma finalidad, pero no se pudieron diseñar sgRNAs para todos los genes seleccionados de los estudios transcriptómicos por falta de secuencia. Las regiones intergénicas se evaluarán a futuro, introduciendo en esos *loci* un promotor y un gen reportero, de forma tal de evaluar la actividad transcripcional mediada por ese contexto cromosomal.

Se eligieron aquellos sgRNAs con menos cantidad de *off-targets* y/o que los mismos estén en regiones intergénicas o intrónicas. Para algunas regiones *target* consideradas no se pudieron diseñar por dos grandes razones: gran cantidad de *off-targets* predichos y por falta de secuencia genómica; es decir, ADN no secuenciado de células CHO-K1. En los casos posibles se diseñaron dos pares de oligonucleótidos por cada región genómica *target*. Se realizó el *annealing* de un par de oligos diseñado para la región 3' UTR de cada gen y se clonaron en el vector lentiCRISPRv2, previamente digerido con la enzima *BsmBI*, purificado y cuantificado. Mediante PCR de

colonias recombinantes se verificó la correcta inserción de los mismos (Figura 54.A), ya que –en todos los casos– se amplificó el fragmento del tamaño esperado (1500 pb).

Para llevar a cabo la inserción de ADN en un *locus* particular se requiere el uso de plásmidos donantes que contengan regiones homólogas flanqueantes al sitio de modificación. Los brazos homólogos necesarios para el *knock-in* en el extremo 3' UTR de cada gen elegido se amplificaron mediante PCR y el resultado se sometió a una electroforesis en gel de agarosa (Figura 54.B). Todos los fragmentos resultaron del tamaño esperado (aproximadamente 800-850 pb), luego se escindieron del gel y se purificaron.

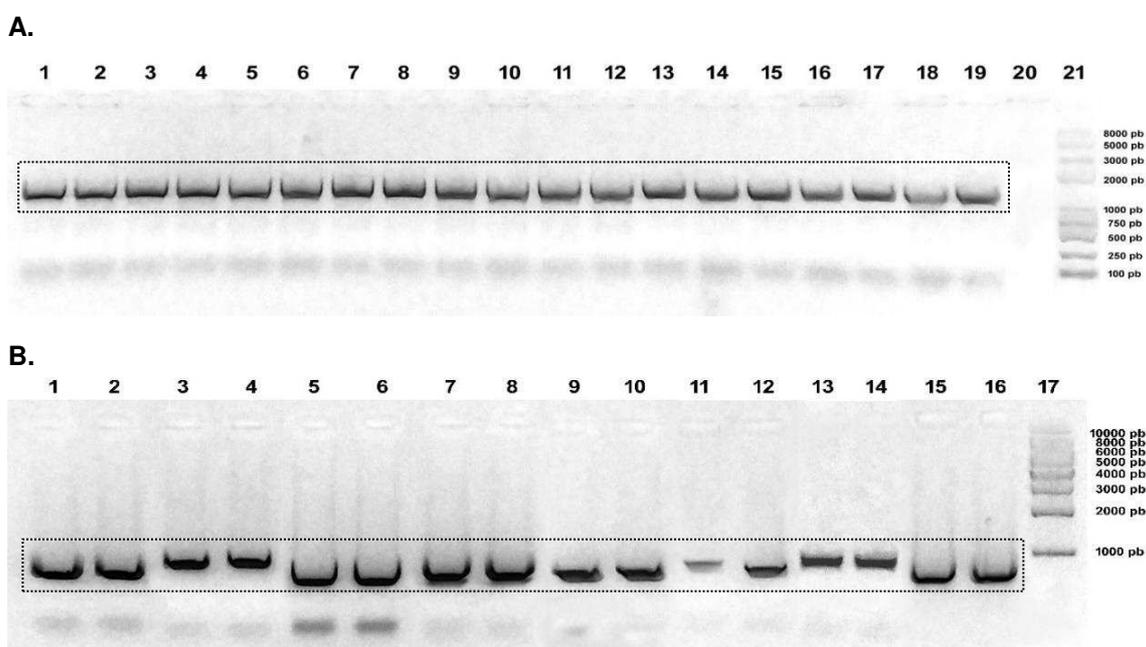


Figura 54. Electroforesis en gel de agarosa 1% (p/v).

- A.** Amplicones de 1500 pb resultantes de la PCR de colonias recombinantes. Carriles 1-7: clones 1 a 7 con sgRNA para S100a6. Carriles 8-12: clones 1 a 5 con sgRNA para Fth1. Carriles 13-17: clones 1 a 5 con sgRNA para Vim. Carriles 18-19: clones 1 a 2 con sgRNA para Lgals1. Carril 20: control de los reactivos de reacción. Carril 21: marcador de peso molecular *Trans2K Plus II DNA Marker* (TransGen Biotech). Todos los clones resultaron positivos, ya que en sus carriles se observa la banda del tamaño esperado señalada.
- B.** Brazos homólogos (HA) para *knock-in* amplificados. Carriles 1-2: HA derecho S100a6. Carriles 3-4: HA izquierdo S100a6. Carriles 5-6: HA derecho Fth1. Carriles 7-8: HA izquierdo Fth1. Carriles 9-10: HA derecho Vim. Carriles 11-12: HA izquierdo Vim. Carriles 13-14: HA izquierdo Lgals1. Carriles 15-16: HA derecho Lgals1. Carril 17: marcador de peso molecular *1KB DNA Ladder* (TransGen Biotech).

En la Figura 55 se muestra una representación esquemática del *cassette*, diseñado, para llevar a cabo el experimento de *knock-in*.

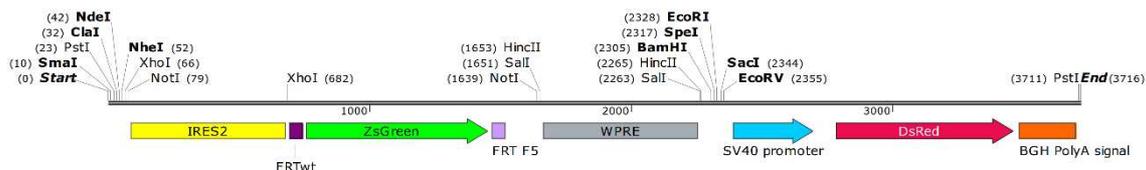


Figura 55. Representación esquemática de la secuencia diseñada para insertarse en células CHO-K1 con el plásmido donante mediante *knock-in* por el mecanismo celular de recombinación homóloga directa (HDR).
Figura realizada con SnapGene® Viewer 4.1.9

Esta secuencia tiene un tamaño de 3716 pb y posee dos sitios de múltiple clonado para la inserción de los brazos izquierdo y derecho necesarios para la recombinación homóloga directa (HDR), una vez dentro de la célula. Entre dichos sitios se encuentra una secuencia IRES (sitio interno de entrada al ribosoma) que permitirá la traducción del gen reportero; el gen de la proteína verde fluorescente ZsGreen1, flanqueada por sitios FRT del sistema de recombinación FLP/FRT necesarios para intercambiar dicho gen por otro de interés (proteína terapéutica) una vez en la célula y la secuencia regulatoria WPRE para estabilizar y mejorar la expresión del TG.

Luego del sitio de inserción del brazo homólogo derecho se incluyó el promotor SV40 que dirige la expresión de la proteína fluorescente DsRed y la señal de poliadenilación BGH. Esto último para futuros ensayos de FACS de manera tal de seleccionar únicamente la población que exprese ZsGreen1, ya que aquellas células que expresen DsRed indicarán eventos de recombinación no deseados, debido a que dicho promotor y la secuencia codificante para la proteína roja fluorescente se encuentran por fuera de los brazos homólogos.

Además, se agregaron sitios de reconocimiento para enzimas de restricción flanqueando las secuencias IRES y WPRE, con el fin de removerlas y de ser necesario clonar IRES en el lugar de WPRE para usar este *cassette* en el sitio 5' UTR del gen en futuros ensayos. De igual manera, se podrá incorporar un promotor en el sitio de la secuencia IRES, en caso de llevar a cabo el *knock-in* en las regiones 5' y 3' intergénicas.

La secuencia del *cassette* diseñado fue sintetizada por la empresa Gene Universal Inc. (EE.UU.), y posterior a su síntesis fue clonada en el plásmido pUC57 (1800 pb), generándose el vector pUC-Knockin (con un tamaño de 5539 pb). El mismo fue luego resuspendido en agua Milli Q estéril y amplificado mediante maxipreparación de ADN plasmídico.

Para corroborar su identidad, se realizó una reacción de digestión con la enzima de restricción *NotI*, presente en dos sitios de la secuencia de dicho plásmido. Se observó la liberación de la banda de tamaño esperado (1560 pb) (Figura 56).

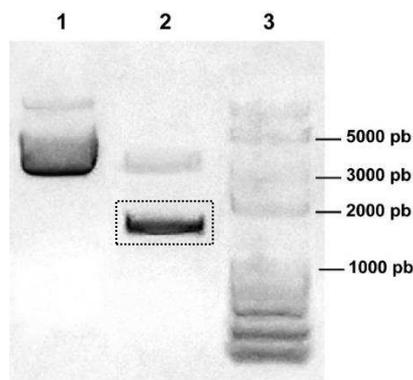


Figura 56. Electroforesis en gel de agarosa 1% (p/v) del vector pUC-Knockin.

Carril 1: Plásmido pUC-Knockin sin digerir enzimáticamente. Carril 2: producto de la digestión de dicho vector con la enzima *NotI*. Se encuentra señalada la banda liberada de tamaño esperado (1560 pb). Carril 3: marcador de peso molecular *Trans2K Plus II DNA Marker* (TransGen Biotech).

Debido a una demora considerable en el envío a Argentina de la secuencia sintetizada, las actividades desarrolladas en relación al *knock-in* en células CHO-K1 mediante la tecnología CRISPR/Cas9 hasta el momento en nuestro laboratorio fueron las descritas previamente en esta Sección. Los experimentos propuestos a futuro, respecto a este ensayo, incluyen la digestión enzimática secuencial tanto del plásmido pUC-Knockin como de los brazos homólogos para S100a6, Fth1, Lgals1 y Vim; lo cual permitirá el clonado de los mismos en dicho vector.

Posteriormente este plásmido donante se linealizará y co-transfectará con cada vector lentiCRISPRv2-sgRNA generado, en células CHO-K1 adaptadas al crecimiento en suspensión. Luego, se evaluará la eficiencia de cada *knock-in* por FACS, seleccionando las poblaciones celulares con la modificación deseada; es decir, aquellas células que expresen ZsGreen y no DsRed. Por último, mediante secuenciación de ADN de las regiones flanqueantes a cada sitio de inserción, se analizará la secuencia tanto del gen endógeno como del *cassette* introducido, esperando que no se produzcan mutaciones o re-arreglos genómicos; es decir, que la inserción haya sido completa y correcta. Asimismo, se analizará por RT-qPCR la expresión del gen reportero y de los genes endógenos seleccionados para estos experimentos de *knock-in*.

Las actividades propuestas a más largo plazo incluyen el intercambio, mediante la tecnología de recombinación FLP/FRT, de la secuencia codificante para ZsGreen1 (una vez incorporada en las células) por la de otra secuencia de interés, como puede ser el gen de una proteína terapéutica. Así como también llevar a cabo el *knock-in* en los otros sitios *target* para cada gen (5' UTR, y regiones 5' y 3' intergénicas). Por otro lado, una vez establecido el protocolo de trabajo con el sistema CRISPR/Cas9 se podrá

aplicar a otras líneas celulares, utilizando el plásmido donante diseñado, en el cual se clonarán los brazos de recombinación propios del genoma de la línea celular empleada.

En resumen, contar con promotores endógenos para la producción de cualquier proteína recombinante de interés bioterapéutico, sumado a la posibilidad de aplicar la novedosa tecnología de edición genética CRISPR a líneas celulares de mamífero, establece una plataforma de trabajo promisoría e inagotable para nuestro laboratorio. La obtención de esta plataforma de producción representará una herramienta biotecnológica de gran utilidad, ya que permitirá obtener clones –en principio– de células CHO-K1 candidatos para la producción de proteína recombinantes con fines biofarmacéuticos. Además, posibilitará optimizar los tiempos y costos de producción de cualquier proteína recombinante tanto a escala industrial como en investigación básica, y con aplicabilidad en medicina humana y veterinaria.

CONCLUSIONES



CONCLUSIONES

“Me lo contaron y lo olvidé; lo vi y lo entendí; lo hice y lo aprendí”.

Confucio

La producción de proteínas recombinantes terapéuticas ha revolucionado la medicina moderna en las últimas décadas. Esta área se ha desarrollado enormemente desde 1987, año en el que se aprobó la comercialización de la primera proteína recombinante terapéutica producida en células de mamíferos. Los cultivos de células eucariotas dominan el campo de producción de productos bioterapéuticos para el diagnóstico y la prevención de enfermedades [5]. Aproximadamente un 60-70% de estas proteínas son producidas en cultivos de células animales [6], principalmente porque la maquinaria de procesamiento proteico, que incluye los procesos de glicosilación y plegamiento, es similar a la humana [7].

Una de las líneas celulares de mamífero más utilizada es la línea ancestral CHO-K1, la cual constituye una plataforma de producción robusta y confiable. Se caracteriza por su fácil adaptación al crecimiento en suspensión y en condiciones libre de SFB, lo que permite llevar a cabo su cultivo en altas densidades celulares. En la actualidad, más del 70% de las proteínas terapéuticas aprobadas son producidas en células CHO, mientras que sólo un 8% en células humanas [9,11,14,15].

El proceso de producción de proteínas terapéuticas se caracteriza por ser complejo y demandar mucho tiempo; desde el clonado inicial del transgén en un vector de expresión apropiado, hasta el procesamiento final como producto farmacéutico en el mercado, requiriéndose una gran cantidad de recursos para controlar, monitorear y asegurar la calidad en cada etapa. Además, la producción de cada proteína terapéutica específica debe optimizarse, principalmente a escala industrial [3]. En particular, para alcanzar mayores productividades se debe conocer el efecto que tienen ciertos parámetros sobre el cultivo, como la temperatura, la capacidad del biorreactor, la velocidad de agitación, la composición del medio empleado, entre otros.

Se ha reportado que la producción de proteínas recombinantes es dependiente de la fase del ciclo celular. Especialmente, diversos genes implicados en la biogénesis del ribosoma y la traducción proteica exhiben una elevada expresión en la fase G1 [112-121]. Además, las células sincronizadas en G0/G1 son metabólicamente más activas y de mayor tamaño que aquellas no arrestadas [116,120]. Por ello, se considera a dicha fase del ciclo celular como el período ideal para incrementar la producción de proteínas recombinantes.

Existen diferentes métodos para sincronizar los cultivos en la fase G0/G1 del ciclo celular. Particularmente, se ha visto que la disminución de la temperatura reduce la

proliferación celular y así se incrementa la duración del cultivo, la viabilidad celular y la productividad específica en células CHO [118,119,124-131]. Esta reducción en el metabolismo implica una disminución del consumo de glucosa (por lo tanto del medio de crecimiento empleado), del oxígeno disuelto y de la generación de productos tóxicos. Esto conlleva a que las células entren en apoptosis de manera tardía, favoreciendo así que la fase estacionaria o de producción se prolongue por más tiempo [114,118,126,127,133-135]. De igual manera, diferentes estudios han comprobado que el uso de una menor temperatura de cultivo aumenta los niveles de ARNm recombinante [118,130].

Por estas razones, en muchos cultivos celulares se lleva a cabo un proceso bifásico para la producción de proteínas recombinantes. El mismo consta de dos fases: una etapa de proliferación inicial a 37°C, seguida de una fase de producción extendida a una temperatura reducida (28°C-33°C). Sin embargo, actualmente es escasa la información a nivel molecular sobre la respuesta de los cultivos de células de mamífero a bajas temperaturas, o sobre el mecanismo por el cual las células mejoran la productividad en estas condiciones [108].

La aplicación de las técnicas “ómicas” en el área de producción de bioterapéuticos representa una gran oportunidad para caracterizar y comprender el estado de los cultivos celulares y optimizar así los procesos de producción a escala industrial. La aplicación de estudios transcriptómicos en bioprocesos facilita la identificación de potenciales *targets* para la ingeniería celular. Para ello, la tecnología CRISPR/Cas9 abre un abanico de posibilidades a través de la edición precisa de células CHO huésped destinadas a la producción de proteínas recombinantes.

La mayoría de los procesos de producción de biofarmacéuticos utilizan, principalmente, sistemas de expresión en los cuales el transgén de interés se integra en el genoma celular de manera aleatoria. En los últimos años se han desarrollado nuevas técnicas que permiten incorporar el gen de interés en sitios del genoma ya caracterizados, en regiones transcripcionalmente activas y estables (“*hotspots*”). Estas estrategias permiten evadir las dificultades asociadas al “efecto de posición”, logrando niveles elevados de expresión del transgén.

En la generación de una línea celular productora de una proteína de interés terapéutico es fundamental el diseño del vector de expresión. En todo este contexto, una etapa crucial dentro del proceso productivo corresponde a la elección de un promotor que permita la adecuada expresión del transgén [197], ya que para alcanzar niveles óptimos en la producción de proteínas recombinantes en células de mamíferos, se necesitan niveles elevados de expresión. Este interés en el análisis de los promotores reside en el gran potencial que tienen para regular y controlar la expresión del transgén,

ya que no sólo pueden incrementar la actividad transcripcional sino que también proporcionar otros niveles de control, por ejemplo, en tiempos o estadios concretos [195]. Por esto, constituyen un área interesante de investigación mediante la aplicación de tecnologías de secuenciación, particularmente RNA-seq, para la identificación de los mismos.

En la actualidad, debido a la escasa información genómica disponible para las células CHO, la alternativa más usada es el empleo de promotores constitutivos, generalmente virales, ya que proporcionan dichos niveles elevados de expresión, necesarios en un proceso productivo [24,198]. No obstante, el uso de promotores virales presenta ciertas desventajas. En muchos casos estos imponen a las células un estrés que interrumpe el crecimiento celular, afectan los procesos de traducción, plegamiento y ensamblaje dentro del retículo endoplásmico y, además, se ha visto que son susceptibles al silenciamiento epigenético [200]. Otro problema se relaciona con la dependencia de estos promotores respecto a las fases del ciclo celular [201]. Se ha demostrado que la expresión dirigida por CMV aumenta en la fase S, disminuyendo considerablemente cuando las células están en G0/G1 [202].

Principalmente, la función promotora debe coordinarse con el bioproceso, asegurando que los promotores sean compatibles con las características celulares deseadas, tales como la proliferación y la viabilidad, y con la duración del cultivo. Por estas razones, la utilización de promotores endógenos constituye una alternativa a los virales para su uso en procesos productivos.

En este contexto, se planteó la hipótesis del presente trabajo de tesis, cuyo objetivo general es la aplicación de las nuevas tecnologías de secuenciación masiva (NGS) a la línea celular CHO-K1, con el fin de desarrollar estrategias de ingeniería genética mediante el análisis bioinformático de los datos generados. Específicamente, se propuso llevar a cabo la secuenciación del transcriptoma (RNA-seq) de células CHO-K1 cultivadas en alta densidad, con el propósito de comparar perfiles de expresión génica en diferentes etapas del cultivo en biorreactor, y seleccionar ciertos genes con elevada expresión, con el objetivo de usarlos para la construcción de vectores de expresión de proteínas recombinantes. Los mismos son una alternativa al empleo de los promotores habitualmente usados, principalmente promotores virales, para lograr incrementar la estabilidad genética y mejorar la eficiencia del proceso de obtención de las proteínas recombinantes desarrolladas.

Para tal fin, en primer lugar nos planteamos caracterizar el ciclo celular de células CHO-K1 cultivadas en biorreactor, analizando diariamente la distribución de las mismas en las diferentes fases del ciclo celular. Esto con el objetivo de determinar si es posible establecer una correlación entre las fases de crecimiento celular y la variación de la

distribución de las células en las diferentes fases del ciclo celular; y, además, si el cultivo tiende a estacionarse en una fase particular del ciclo, en respuesta a las condiciones típicas de un bioproceso de células CHO-K1 adaptadas al crecimiento en suspensión. A su vez, sincronizar dichos cultivos para lograr mayor cantidad de células en la fase G0/G1, ya que como se mencionó previamente, dicha etapa es considerada el período ideal para incrementar la producción de proteínas recombinantes, y obtener así poblaciones celulares enriquecidas en células G0/G1 de las cuales extraer el ARN total para los análisis transcriptómicos subsiguientes.

En esta etapa de trabajo, se llevaron a cabo dos cultivos independientes de células CHO-K1 en biorreactor para el estudio del ciclo celular y en ambos procesos se utilizó el medio comercial Ex-Cell 302 libre de SFB (Sigma-Aldrich). Diariamente se extrajeron muestras para estos análisis y se llevó a cabo el modelado del ciclo celular con el programa ModFit LT. Con los valores obtenidos (porcentaje de células en cada fase) se construyeron los gráficos de barra de la Figura 18, los cuales muestran la distribución de las células en las fases del ciclo celular, por cada día de cultivo, junto con las curvas de crecimiento para cada una de las dos experiencias.

Las características de ambos bioprocesos, detalladas en el Capítulo 1 - Sección 1, se ven reflejadas en los análisis del ciclo celular. Durante el primer cultivo el porcentaje máximo de células en la fase G0/G1 (81%) fue alcanzado el último día del cultivo, cuando la temperatura había permanecido en 32°C desde los nueve días previos. En el segundo cultivo se logró obtener tempranamente una mayor proporción de células en G0/G1 (89%), en el día 11 del cultivo (al inicio de la fase estacionaria), luego de dos días con una temperatura de 31°C. En particular, durante este segundo cultivo, más del 80% de las células permanecieron en fase G0/G1 durante toda la fase estacionaria. Esto puede deberse al marcado gradiente de temperatura aplicado, ya que no se usó ningún método químico o físico para lograr el arresto celular. Dicho gradiente, que comenzó en los primeros días del cultivo, se llevó a cabo para controlar adecuadamente las concentraciones de los metabolitos (glucosa y lactato); y en consecuencia se utilizó un menor volumen de perfusión y se evitó la separación de suspensión celular del biorreactor.

Por otro lado, analizamos la distribución de la población celular para cada día de la segunda experiencia con el objetivo de identificar cambios en el tamaño y la complejidad de las células. Como se mencionó anteriormente, las células arrestadas en fase G0/G1 son metabólicamente más activas y de mayor tamaño que aquellas no arrestadas. En la Figura 19 puede verse claramente que, a medida que el cultivo progresa, la forma que adopta la población celular, en los gráficos de SSC vs. FSC, varía en cuanto a su localización y dispersión. En particular, se observa un cambio muy

marcado a partir del día 8, que corresponde con el inicio de la fase estacionaria. La población se va desplazando hacia la derecha del eje FSC, lo que indica un incremento en el tamaño de las células, y hacia el extremo superior del eje SSC, evidenciando también una mayor granulosidad o complejidad celular, lo cual sugiere una composición celular más compleja, dada por un aumento de los componentes celulares, como proteínas y otras biomoléculas.

En síntesis, en ambos cultivos fue posible evidenciar un enriquecimiento del cultivo en células sincronizadas en fase G0/G1, en la etapa de crecimiento exponencial tardío y hacia los últimos días de la fase estacionaria. Estos análisis arrojan una tendencia evidente: cuanto más temprano se aplica el gradiente de temperatura, más rápidamente las células son arrestadas en la fase G0/G1, generando un mayor control del bioproceso, lo cual deriva en un menor tiempo de cultivo, una disminución de la proliferación celular y una mayor viabilidad celular. En conjunto, estos factores pueden generar mayores tasas específicas de producción de proteínas recombinantes durante un proceso productivo.

Considerando que en la siguiente etapa de trabajo se propuso la secuenciación del transcriptoma de células CHO-K1 cultivadas en alta densidad y sincronizadas en fase G0/G1, y en función de los resultados obtenidos de la caracterización del ciclo celular, es posible evidenciar que los cultivos de CHO-K1 tienden a estacionarse espontáneamente en G0/G1, con un alto porcentaje de células en dicha fase. Por estas razones, se decidió realizar las extracciones de ARN total de las células muestreadas directamente a partir del biorreactor; es decir, sin recurrir a ningún método de sincronización, preservando totalmente la integridad celular y, por lo tanto, la del transcriptoma en las condiciones propias de un bioproceso a escala industrial.

Con los resultados de los dos primeros cultivos en biorreactores se logró no sólo obtener información valiosa acerca del comportamiento de la línea celular en términos de procesos, sino que además se consiguió caracterizar el ciclo celular de esta línea en las condiciones habituales de trabajo de nuestro laboratorio. Por ello se procedió a realizar otros tres cultivos análogos de células CHO-K1 adaptadas al crecimiento en suspensión, controlando las variables y parámetros operacionales para imitar el comportamiento del segundo cultivo. Dichos cultivos se llevaron a cabo en forma de réplicas biológicas, de manera tal de abarcar la variabilidad inherente a los cultivos celulares y contar así con un mayor poder estadístico para los análisis posteriores. De estos cultivos se extrajo ARN total para su subsiguiente secuenciación con la tecnología Illumina.

Se secuenció y analizó el transcriptoma de células CHO-K1, cultivadas en alta densidad, en modo de perfusión y con medio de cultivo libre de SFB. Se compararon los

perfiles de expresión génica entre las diferentes etapas del cultivo (exponencial y estacionaria, respecto al inicio del cultivo), particularmente luego de una reducción de la temperatura desde 37°C a 31°C. Solamente se tuvieron en cuenta aquellos genes con un valor de $\log_2 \text{fold change} \geq 1$ (es decir *up* o *down* regulados dos veces) y un p-valor ajustado $\leq 0,05$. En la Figura 34 se observan dos gráficos que resumen la variabilidad en la expresión génica entre las fases de crecimiento de los cultivos de CHO-K1 en suspensión. Se vio que los mayores cambios regulatorios ocurrieron durante la fase estacionaria, predominando marcadamente la *up*-regulación génica. Estos subgrupos de genes diferencialmente expresados, tanto en la fase exponencial como en la fase estacionaria, representan un 2,0% y un 6,4% del total de genes con actividad transcripcional en cada fase, respectivamente. También, se examinó la expresión de genes que responden a cambios de la temperatura del cultivo, como CIRP y RBM3.

Todos los genes identificados, que exhibieron cambios significativos de expresión y aquellos que responden a temperaturas reducidas de cultivo, constituyen un subgrupo altamente específico con cambios regulatorios significativos, los cuales pueden resultar potenciales *targets* para la ingeniería celular y/o metabólica de las células CHO, en particular de las células huésped usadas para producir proteínas recombinantes terapéuticas. De esta manera, los bioprocesos de células CHO-K1 se pueden optimizar, tanto a escala de laboratorio como industrial.

Los resultados obtenidos permitieron obtener un panorama global del comportamiento de esta línea celular en respuesta a un bioproceso típico a escala industrial. Se lograron caracterizar los procesos biológicos y las vías metabólicas involucradas en dichos cambios. Además, se pudieron identificar genes con funciones clave en el arresto de las células en la fase G0/G1, para cada etapa del cultivo; correlacionando así los análisis transcriptómicos con los de ciclo celular. Cabe destacar que este trabajo es el primero en cual se reporta una combinación entre los resultados del estudio del ciclo celular y el análisis del transcriptoma de la línea celular CHO-K1 huésped durante un bioproceso en condiciones típicas a escala industrial.

En resumen, se pudo verificar que la exposición de las células CHO-K1 a temperaturas reducidas de cultivo desencadena una respuesta coordinada que comprende la modulación del ciclo celular, del metabolismo energético, de la proliferación celular, la reorganización del citoesqueleto y de procesos como la replicación y reparación del ADN, transcripción, metabolismo proteico y apoptosis.

Además de los análisis mencionados para la caracterización del transcriptoma de cultivos CHO-K1 en alta densidad, se seleccionaron los primeros 200 transcritos con mayor expresión, tanto en la etapa exponencial como estacionaria de cultivo. De ese grupo se eligieron los siguientes genes: S100a6, Fth1, Lgals1, Vim, Actb, Rps18, Cd63,

Calr, Ybx1, Hspa5, Hmox1, Ctsz y Lamp1, y se los consideró “candidatos” para el empleo de sus regiones promotoras.

Como objetivo principal de la última etapa de trabajo, se planteó utilizar las secuencias promotoras de los genes identificados con alta actividad transcripcional mediante RNA-seq, para la construcción de nuevos vectores de expresión destinados a la producción de proteínas recombinantes, logrando así que dichos promotores respondan a las condiciones celulares y a la dinámica de los procesos productivos de nuestro laboratorio. Esta estrategia de trabajo constituye una alternativa al empleo de promotores virales usados habitualmente, aprovechando la maquinaria transcripcional regulatoria endógena disponible en las condiciones de un bioproceso de alta densidad celular con un medio de cultivo definido. Así, se busca lograr una expresión óptima del transgén de interés en un contexto normal de producción industrial.

Se pudo llevar a cabo el clonado de los promotores Rps18, Hspa5 y Hmox1 en el vector pZsGreen1-1 (para Hmox1 se evaluaron dos clones). Luego, con cada construcción plasmídica se transfectaron células CHO-K1 adaptadas al crecimiento en suspensión y HEK, mediante el empleo de lípidos catiónicos en condiciones de crecimiento en adherencia. A las 72 h postransfección, un volumen de las células CHO-K1 transfectadas (con cada vector) se llevó al crecimiento en suspensión en pequeña escala, empleando medio de cultivo Ex-Cell 302 libre de SFB.

El *screening* de la actividad transcripcional se realizó únicamente de manera transiente, con el objetivo de obtener de forma rápida un panorama global del comportamiento de las secuencias promotoras aisladas, en dos de las líneas celulares de mamífero más usadas. Por citometría de flujo se analizó la expresión de la proteína verde fluorescente ZsGreen1 bajo el control del promotor viral CMV y de los promotores endógenos aislados durante 13 días (312 h). Asimismo, las líneas celulares cultivadas en adherencia fueron monitoreadas mediante microscopía de fluorescencia durante ese período. Los resultados de los análisis de la expresión transiente de la proteína ZsGreen1 se muestran en las Figuras 46-48 y en las Tablas 11-13, para células HEK, CHO-K1adh y CHO-K1s, respectivamente.

En todos los casos, la actividad transcripcional máxima se observó entre las 72 h y 144 h postransfección, para luego ir disminuyendo con el transcurso de los días de cultivo. El promotor endógeno Rps18 mostró muy baja actividad transcripcional, incluso llamativamente menor, en algunos casos, al plásmido sin promotor usado como control basal. En células HEK, el promotor Hmox1 exhibió una expresión similar en sus dos clones; mientras que en las células CHO-K1, el Clon 1 presentó aproximadamente el doble de actividad transcripcional que el Clon 2, pero siempre en una proporción menor al promotor viral CMV.

El resultado más promisorio se obtuvo con el promotor endógeno Hspa5, el cual mostró un comportamiento equivalente al promotor CMV, en células HEK y CHO-K1 (tanto en crecimiento en adherencia como en suspensión). En las células HEK se lograron valores superiores al 50% de células positivas, que expresaron ZsGreen1; mientras que en las células CHO este porcentaje fue menor al 40%, probablemente debido a que las células HEK son más fáciles de transfectar y presentan una mayor eficiencia de este proceso. Por otro lado, al evaluar la actividad transcripcional de los promotores en las células CHO-K1 llevadas a un crecimiento en suspensión y con medio libre de SFB, se observaron valores ligeramente superiores de la actividad transcripcional de todos los promotores durante los 13 días que duró el ensayo.

Con estos estudios se obtuvieron resultados muy prometedores, ya que se pudo constituir un esquema racional de trabajo para la identificación de nuevos promotores, a partir de los análisis de expresión génica de células CHO-K1 cultivadas en biorreactor, el aislamiento de sus secuencias y generación de nuevos vectores de expresión. El hecho de que se obtuvieran resultados similares tanto en CHO-K1 como en HEK, a pesar de que las secuencias fueron aisladas del genoma de la primera línea celular, impulsa el estudio de los promotores endógenos en otras líneas celulares y en diferentes condiciones de cultivo.

Por otro lado, particularmente para aquellos genes a partir de los cuales no se pudieron amplificar sus promotores, principalmente debido a falta de secuencia genómica o anotación insuficiente del genoma de la línea celular CHO-K1, se decidió aplicar la reciente tecnología de edición genética CRISPR/Cas9.

Como se ha mencionado, es fundamental desarrollar metodologías que permitan encontrar y reutilizar los sitios denominados *hotspots* a través de la introducción de diferentes transgenes que codifiquen para diversas proteínas de interés terapéutico. Para esto, la tecnología CRISPR/Cas9 promueve un nuevo horizonte en la edición precisa de células CHO huésped destinadas a la producción de proteína recombinante. Diversas investigaciones describen la inserción exitosa de genes de interés en *loci* específicos de células CHO, demostrando que la integración sitio específica de transgenes en diversos *hotspots* de células CHO es una estrategia prometedora para superar los desafíos que implica la integración aleatoria convencional de un transgén de interés.

Se decidió plantear experimentos de *knock-in* utilizando el sistema CRISPR/Cas9 debido a que para los cuatro genes “candidatos” que exhibieron mayor expresión (S100a6, Fth1, Lgals1 y Vim) no se pudieron llevar a cabo los experimentos de evaluación de la actividad transcripcional en forma transiente, ya que para algunos no

se logró aislar sus secuencias promotoras y para otros no resultaron positivos los clonados en el vector pZsGreen1-1.

El objetivo que se propuso para esta nueva etapa de trabajo fue insertar un *cassette* de *knock-in* en el extremo 3' UTR de cada gen seleccionado, con el fin de utilizar el promotor y las regiones reguladoras del gen endógeno para expresar un gen reportero, que luego se podrá intercambiar mediante tecnología de recombinación de ADN por un TG de interés terapéutico. Para ello se utilizará CRISPR/Cas9 sin alterar la expresión del gen endógeno y sin introducir regiones virales ni antibióticos de selección, optimizando así aspectos regulatorios de la producción de proteínas recombinantes terapéuticas.

En primer lugar, se llevó a cabo un ensayo de *knock-out* de un clon CHO-K1 productor de EGFP, a través del mecanismo celular NHEJ, con la finalidad de implementar y optimizar el sistema CRISPR/Cas9 en células CHO. Luego, se realizaron las etapas 1-6 del esquema de trabajo planteado en la Figura 52, para el ensayo de *knock-in*. En resumen, se logró exitosamente el diseño de los sgRNAs para cada *loci* propuesto en cada uno de los cuatro genes “candidatos” con mayor expresión, se llevó a cabo la síntesis de los oligonucleótidos (para los sgRNAs diseñados en el extremo 3' UTR), y el *annealing* y clonado de los mismos en el vector lentiCRISPRv2. También, se logró la correcta amplificación mediante PCR de los brazos homólogos, requeridos para el *knock-in* mediante el mecanismo celular de HDR, y el diseño *in silico* y síntesis del *cassette* de expresión que se co-transfectará con el vector generado para cada sgRNA, que contiene la endonucleasa Cas9. Luego, se evaluará la eficiencia de cada *knock-in* por FACS, secuenciación de ADN de las regiones flanqueantes a cada sitio de inserción y por RT-qPCR se analizará la expresión del gen reportero y de los genes endógenos seleccionados para estos ensayos.

En resumen, contar con promotores endógenos para la producción de cualquier proteína recombinante de interés bioterapéutico, como alternativa al uso de promotores virales y que, además, responden a las condiciones de cultivo habituales en escala industrial, sumado a la posibilidad de aplicar la novedosa tecnología de edición genética CRISPR a líneas celulares de mamífero, establece una plataforma de trabajo promisorio e inagotable para nuestro laboratorio.

La obtención de esta plataforma de producción representa una herramienta biotecnológica de gran utilidad ya que permitirá obtener clones, en principio, de células CHO-K1 candidatos para la producción de proteína recombinantes con fines biofarmacéuticos. Además, posibilitará optimizar los tiempos y costos de producción de cualquier proteína recombinante tanto a escala industrial como en investigación básica, y con aplicabilidad en medicina humana y veterinaria.

BIBLIOGRAFÍA



BIBLIOGRAFÍA

- [1] K Schwab (2016) The Fourth Industrial Revolution: what it means, how to respond. *World Economic Forum*. Disponible en: <https://www.weforum.org/agenda/2016/01/the-fourth-industrial-revolution-what-it-means-and-how-to-respond/>. Último acceso: 02/04/18.
- [2] AL Demain, P Vaishnav (2009) Production of recombinant proteins by microbes and higher organisms. *Biotechnol Adv* 3 (27):297-306.
- [3] E Wells, AS Robinson (2017) Cellular engineering for therapeutic protein production: product quality, host modification, and process improvement. *Biotechnol J* 1 (12):1-12.
- [4] B Adda (2018) Top 10 Best Selling Biotherapeutics 2017. Disponible en: <https://www.biochemadda.com/top-10-best-selling-biotherapeutics/>. Último acceso: 11/02/18.
- [5] J Zhu (2012) Mammalian cell protein expression for biopharmaceutical production. *Biotechnol Adv* 5 (30):1158-1170.
- [6] FM Wurm (2004) Production of recombinant protein therapeutics in cultivated mammalian cells. *Nat Biotechnol* 11 (22):1393-1398.
- [7] P Hossler, SF Khattak, ZJ Li (2009) Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology* 9 (19):936-949.
- [8] D Clark, N Pazdernik (2016) Recombinant Proteins, En: *Biotechnology*, Elsevier, p. 335-363.
- [9] SC Almo, JD Love (2014) Better and faster: improvements and optimization for mammalian recombinant protein production. *Curr Opin Struct Biol* (26):39-43.
- [10] J Zhang (2010) Mammalian cell culture for biopharmaceutical production. En: *Manual of Industrial Microbiology and Biotechnology*, 3rd Edition, (Ed: R Baltz, A Demain, J Davies, A Bull, B Junker, L Katz, L Lynd, P Masurekar, C Reeves, H Zhao), Washington DC, ASM Press, p. 157-178.
- [11] S Xiao, J Shiloach, MJ Betenbaugh (2014) Engineering cells to improve protein expression. *Curr Opin Struct Biol* (26):32-38.
- [12] Y Durocher, M Butler (2009) Expression systems for therapeutic glycoprotein production. *Curr Opin Biotechnol* 6 (20):700-707.
- [13] S Estes, M Melville (2014) Mammalian Cell Line Developments in Speed and Efficiency, *Adv Biochem Eng Biotechnol* (139):11-33.
- [14] A Kantardjieff, W Zhou (2013) Mammalian Cell Cultures for Biologics Manufacturing. En: *Advances in biochemical engineering/biotechnology*, p. 1-9.
- [15] J Dumont, D Ewart, B Mei, S Estes, R Kshirsagar (2016) Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Crit Rev Biotechnol* 6 (36):1110-1122.
- [16] TT Puck, SJ Cieciura, A Robinson (1958) Genetics of somatic mammalian cells. III. Long-term cultivation of euploid cells from human and animal subjects. *J Exp Med* 6 (108):945-56.
- [17] AD Bandaranayake, SC Almo (2014) Recent advances in mammalian protein production. *FEBS Lett* 2 (588):253-260.

- [18] KP Jayapal, KF Wlaschin, W-S Hu, MGS Yap (2007) Recombinant Protein Therapeutics from CHO Cells — 20 Years and Counting. CHO Consortium, SBE *Spec Sect*, p. 40-47.
- [19] G Urlaub, LA Chasin (1980) Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity. *Proc Natl Acad Sci USA* 7 (77):4216-20.
- [20] FM Wurm (2013) CHO Quasispecies—Implications for Manufacturing Processes. *Processes* 3 (1):296-311.
- [21] T Dingermann (2008) Recombinant therapeutic proteins: Production platforms and challenges. *Biotechnol J* 1 (3):90-97.
- [22] M Thiry, D Cingolani (2002) Optimizing scale-up fermentation processes. *Trends Biotechnol* 3 (20):103-5.
- [23] M-F Clincke, C Mölleryd, Y Zhang, E Lindskog, K Walsh, V Chotteau (2013) Very high density of CHO cells in perfusion by ATF or TFF in WAVE bioreactor™. Part I. Effect of the cell density on the process. *Biotechnol Prog* 3 (29):754-67.
- [24] SC Makrides (1999) Components of Vectors for Gene Transfer and Expression in Mammalian Cells. *Protein Expr Purif* 2 (17):183-202.
- [25] EF Mufarrege, S Antuña, M Etcheverrigaray, R Kratje, C Prieto (2014) Development of lentiviral vectors for transient and stable protein overexpression in mammalian cells. A new strategy for recombinant human FVIII (rhFVIII) production. *Protein Expr Purif* (95):50-56.
- [26] N Kruse, S Spencer, D Wirth (2014) Rational approaches for transgene expression: targeted integration and episomal maintenance, En: *Animal Cell Biotechnology in Biologics Production*, (Ed: H Hauser, R Wagner), Braunschweig, Alemania, De Gruyter, p. 173-215.
- [27] DL Hacker, S Balasubramanian (2016) Recombinant protein production from stable mammalian cell lines and pools. *Curr Opin Struct Biol* (38):129-136.
- [28] K Ding, L Han, H Zong, J Chen, B Zhang, J Zhu (2017) Production process reproducibility and product quality consistency of transient gene expression in HEK293 cells with anti-PD1 antibody as the model protein. *Appl Microbiol Biotechnol* 5 (101):1889-1898.
- [29] M-E Lalonde, Y Durocher (2017) Therapeutic glycoprotein production in mammalian cells. *J Biotechnol* (251):128-140.
- [30] N Harraghy, D Calabrese, I Fisch, P-A Girod, V LeFourn, A Regamey, N Mermoud (2015) Epigenetic regulatory elements: Recent advances in understanding their mode of action and use for recombinant protein production in mammalian cells. *Biotechnol J* 7 (10):967-978.
- [31] K Büssow (2015) Stable mammalian producer cell lines for structural biology. *Curr Opin Struct Biol* (32):81-90.
- [32] M Matasci, DL Hacker, L Baldi, FM Wurm (2008) Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects. *Drug Discov Today Technol* 2–3 (5):e37-e42.
- [33] M De Jesus, FM Wurm (2011) Manufacturing recombinant proteins in kg-ton quantities using animal cells in bioreactors. *Eur J Pharm Biopharm* 2 (78):184-188.

- [34] B Dropulić (2011) Lentiviral Vectors: Their Molecular Design, Safety, and Use in Laboratory and Preclinical Research. *Hum Gene Ther* 6 (22):649-657.
- [35] T Sakuma, MA Barry, Y Ikeda (2012) Lentiviral vectors: basic to translational. *Biochem J* 3 (443):603-618.
- [36] L Baranyi, A Roy, HD Embree, B Dropulic (2010) Lentiviral vector-mediated genetic modification of cell substrates for the manufacture of proteins and other biologics. *PDA J Pharm Sci Technol* 5 (64):379-85.
- [37] V Picanco-Castro, EM de Sousa Russo-Carbolante, D Tadeu Covas (2012) Advances in lentiviral vectors: a patent review. *Recent Pat DNA Gene Seq* 2 (6):82-90.
- [38] C Delenda (2004) Lentiviral vectors: optimization of packaging, transduction and gene expression. *J Gene Med* S1 (6):S125-S138.
- [39] O-W Merten, M Hebben, C Bovolenta (2016) Production of lentiviral vectors. *Mol Ther Methods Clin Dev* (3):16017.
- [40] HT Spencer, G Denning, RE Gautney, B Dropulic, AJ Roy, L Baranyi, B Gangadharan, ET Parker, P Lollar, CB Doering (2011) Lentiviral vector platform for production of bioengineered recombinant coagulation factor VIII. *Mol Ther* 2 (19):302-9.
- [41] X Wu, Y Li, B Crise, SM Burgess (2003) Transcription Start Regions in the Human Genome Are Favored Targets for MLV Integration. *Science* 300 (5626):1749-1751.
- [42] RS Mitchell, BF Beitzel, ARW Schroder, P Shinn, H Chen, CC Berry, JR Ecker, FD Bushman (2004) Retroviral DNA Integration: ASLV, HIV, and MLV Show Distinct Target Site Preferences. *PLoS Biol* 8 (2):e234.
- [43] A Oberbek, M Matasci, DL Hacker, FM Wurm (2011) Generation of stable, high-producing cho cell lines by lentiviral vector-mediated gene transfer in serum-free suspension culture. *Biotechnol Bioeng* 3 (108):600-610.
- [44] MC Rodríguez, N Ceaglio, S Antuña, MB Tardivo, M Etcheverrigaray, C Prieto (2017) High yield process for the production of active human α -galactosidase a in CHO-K1 cells through lentivirus transgenesis. *Biotechnol Prog* 5 (33):1334-1345.
- [45] S Turan, C Zehe, J Kuehle, J Qiao, J Bode (2013) Recombinase-mediated cassette exchange (RMCE) — A rapidly-expanding toolbox for targeted genomic modifications. *Gene* 1 (515):1-27.
- [46] D Wirth, L Gama-Norton, P Riemer, U Sandhu, R Schucht, H Hauser (2007) Road to precision: recombinase-based targeting technologies for genome engineering. *Curr Opin Biotechnol* 5 (18):411-419.
- [47] D Wirth, H Hauser (2004) Flp-Mediated Integration of Expression Cassettes into FRT-Tagged Chromosomal Loci in Mammalian Cells, En: *Recombinant Gene Expression*, New Jersey, Humana Press, p. 467-476.
- [48] S Chandrasegaran, D Carroll (2016) Origins of Programmable Nucleases for Genome Engineering. *J Mol Biol* 5 (428):963-989.
- [49] C Gersbach, T Gaj, C Barbas (2014) Comparing Genome Editing Technologies. *Genet Eng Biotechnol News* 5 (34).

- [50] AJ Wood, T-W Lo, B Zeitler, CS Pickle, EJ Ralston, AH Lee, R Amora, JC Miller, E Leung, X Meng, L Zhang, EJ Rebar, PD Gregory, FD Urnov, BJ Meyer (2011) Targeted genome editing across species using ZFNs and TALENs. *Science* 333 (6040):307.
- [51] A Xiao, Y-Y Hu, W-Y Wang, Z-P Yang, Z-X Wang, P Huang, X-J Tong, Z Bo, S LIN, X An, H Ying-Ying, W Wei-Ye, Y Zhi-Peng, W Zhan-Xiang, H Feng, T Xiang-Jun, Z Bo, L Shuo (2011) Progress in zinc finger nuclease engineering for targeted genome modification. *Hered* 7 (33):665-683.
- [52] F Le Provost, S Lillico, B Passet, R Young, B Whitelaw, J-L Vilotte (2010) Zinc finger nuclease technology heralds a new era in mammalian transgenesis. *Trends Biotechnol* 3 (28):134-141.
- [53] JK Joung, JD Sander (2013) TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 1 (14):49-55.
- [54] LA Marraffini, EJ Sontheimer (2010) CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* 3 (11):181-90.
- [55] R Barrangou, P Horvath (2017) A decade of discovery: CRISPR functions and applications. *Nat Microbiol* 7 (2):17092.
- [56] JA Doudna, E Charpentier (2014) Genome editing. The new frontier of genome engineering with CRISPR-Cas9. *Science* 346 (6213):1258096.
- [57] RE Haurwitz, M Jinek, B Wiedenheft, K Zhou, JA Doudna, AP Snijders, MJ Dickman, KS Makarova, E V. Koonin, J van der Oost (2010) Sequence- and Structure-Specific RNA Processing by a CRISPR Endonuclease. *Science* 329 (5997):1355-1358.
- [58] SQ Tsai, JK Joung (2016) Defining and improving the genome-wide specificities of CRISPR-Cas9 nucleases. *Nat Rev Genet* 5 (17):300-312.
- [59] H Wang, M La Russa, LS Qi (2016) CRISPR/Cas9 in Genome Editing and Beyond. *Annu Rev Biochem* 1 (85):227-264.
- [60] W Feng, Y Dai, L Mou, D Cooper, D Shi, Z Cai (2015) The Potential of the Combination of CRISPR/Cas9 and Pluripotent Stem Cells to Provide Human Organs from Chimaeric Pigs. *Int J Mol Sci* 12 (16):6545-6556.
- [61] L Liu, Y Li, S Li, N Hu, Y He, R Pong, D Lin, L Lu, M Law (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol* (2012):251364.
- [62] V Mallawaarachchi (2018) Latest Advances in DNA Sequencing Bioinformatics and Precision Medicine. Disponible en: <https://thearachchistyle.com/2018/04/22/latest-advances-in-dna-sequencing-bioinformatics-and-precision-medicine/>. Último acceso: 25/05/18.
- [63] L Liu, Y Li, S Li, N Hu, Y He, R Pong, D Lin, L Lu, M Law (2012) Comparison of next-generation sequencing systems. *J Biomed Biotechnol* (2012):1-11.
- [64] ML Metzker (2010) Sequencing technologies - the next generation. *Nat Rev Genet* 1 (11):31-46.
- [65] S Zhao, K Prenger, L Smith (2013) Stormbow: A Cloud-Based Tool for Reads Mapping and Expression Quantification in Large-Scale RNA-Seq Studies. *ISRN Bioinforma* (2013):1-8.

- [66] U López de Heredia, JL Vázquez-Poletti (2016) RNA-seq analysis in forest tree species: bioinformatic problems and solutions. *Tree Genet Genomes* 2 (12):30.
- [67] EL van Dijk, H Auger, Y Jaszczyszyn, C Thermes (2014) Ten years of next-generation sequencing technology. *Trends Genet* 9 (30):418-426.
- [68] U López De Heredia (2016) Massive sequencing techniques in the study of biological diversity. *Munibe, Cienc nat* (64):7-31.
- [69] L Nederbragt (2016) Developments in high throughput sequencing. Disponible en: <https://flxlexblog.wordpress.com/2016/07/08/developments-in-high-throughput-sequencing-july-2016-edition/>. Último acceso: 20/05/18.
- [70] M Margulies, M Egholm, WE Altman, S Attiya, JS Bader, LA Bemben, J Berka, MS Braverman, Y-J Chen, Z Chen, SB Dewell, L Du, JM Fierro, XV Gomes, BC Godwin, W He, S Helgesen, CH Ho, GP Irzyk, SC Jando, MLI Alenquer, TP Jarvie, KB Jirage, J-B Kim, JR Knight, JR Lanza, JH Leamon, SM Lefkowitz, M Lei, J Li, KL Lohman, H Lu, VB Makhijani, KE McDade, MP McKenna, EW Myers, E Nickerson, JR Nobile, R Plant, BP Puc, MT Ronan, GT Roth, GJ Sarkis, JF Simons, JW Simpson, M Srinivasan, KR Tartaro, A Tomasz, KA Vogt, GA Volkmer, SH Wang, Y Wang, MP Weiner, P Yu, RF Begley, JM Rothberg (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 7057 (437):376-380.
- [71] D Pushkarev, NF Neff, SR Quake (2009) Single-molecule sequencing of an individual human genome. *Nat Biotechnol* 9 (27):847-850.
- [72] ER Mardis (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet* 3 (24):133-41.
- [73] J Ju, D Hyun Kim, L Bi, Q Meng, X Bai, Z Li, X Li, M Sano Marma, S Shi, J Wu, JR Edwards, A Romu, NJ Turro (2006) Four-color DNA sequencing by synthesis using cleavable fluorescent nucleotide reversible terminators. *PNAS* 103 (52):19635-19640.
- [74] J Clarke, H-C Wu, L Jayasinghe, A Patel, S Reid, H Bayley (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat Nanotechnol* 4:265-270.
- [75] AM Lewis, NR Abu-Absi, MC Borys, ZJ Li (2016) The use of 'Omics technology to rationally improve industrial mammalian cell line performance. *Biotechnol Bioeng* 1 (113):26-38.
- [76] FC Courtes, J Lin, HL Lim, SW Ng, NSC Wong, G Koh, L Vardy, MGS Yap, B Loo, D-Y Lee (2013) Translatome analysis of CHO cells to identify key growth genes. *J Biotechnol* 3 (167):215-224.
- [77] S Selvarasu, YS Ho, WPK Chong, NSC Wong, FNK Yusufi, YY Lee, MGS Yap, D-Y Lee (2012) Combined in silico modeling and metabolomics analysis to characterize fed-batch CHO cell culture. *Biotechnol Bioeng* 6 (109):1415-1429.
- [78] C Clarke, P Doolan, N Barron, P Meleady, F O'Sullivan, P Gammell, M Melville, M Leonard, M Clynes (2011) Large scale microarray profiling and coexpression network analysis of CHO cells identifies transcriptional modules associated with growth and productivity. *J Biotechnol* 3 (155):350-359.
- [79] DM Wuest, SW Harcum, KH Lee (2012) Genomics in mammalian cell culture bioprocessing. *Biotechnol Adv* 3 (30):629-638.

- [80] A Oshlack, MD Robinson, MD Young (2010) From RNA-seq reads to differential expression results. *Genome Biol* 12 (11):220.
- [81] X Xu, H Nagarajan, NE Lewis, S Pan, Z Cai, X Liu, W Chen, M Xie, W Wang, S Hammond, MR Andersen, N Neff, B Passarelli, W Koh, HC Fan, J Wang, Y Gui, KH Lee, MJ Betenbaugh, SR Quake, I Famili, BO Palsson, J Wang (2011) The genomic sequence of the Chinese hamster ovary (CHO)-K1 cell line. *Nat Biotechnol* 8 (29):735-41.
- [82] NE Lewis, X Liu, Y Li, H Nagarajan, G Yerganian, E O'Brien, A Bordbar, AM Roth, J Rosenbloom, C Bian, M Xie, W Chen, N Li, D Baycin-Hizal, H Latif, J Forster, MJ Betenbaugh, I Famili, X Xu, J Wang, BO Palsson (2013) Genomic landscapes of Chinese hamster ovary cell lines as revealed by the *Cricetulus griseus* draft genome. *Nat Biotechnol* 8 (31):759-65.
- [83] J Becker, C Timmermann, T Jakobi, O Rupp, R Szczepanowski, M Hackl, A Goesmann, A Tauch, N Borth, J Grillari, A Pühler, T Noll, K Brinkrolf (2011) Next-generation sequencing of the CHO cell transcriptome. *BMC Proc Suppl* 8 (5 Suppl 8):P6.
- [84] O Rupp, J Becker, K Brinkrolf, C Timmermann, N Borth, A Pühler, T Noll, A Goesmann (2014) Construction of a public CHO cell line transcript database using versatile bioinformatics analysis pipelines. *PLoS One* 1 (9):e85568.
- [85] JAH Bort, M Hackl, H Höflmayer, V Jadhav, E Harreither, N Kumar, W Ernst, J Grillari, N Borth (2012) Dynamic mRNA and miRNA profiling of CHO-K1 suspension cell cultures. *Biotechnol J* 4 (7):500-515.
- [86] D Baycin-Hizal, DL Tabb, R Chaerkady, L Chen, NE Lewis, H Nagarajan, V Sarkaria, A Kumar, D Wolozny, J Colao, E Jacobson, Y Tian, RN O'Meally, SS Krag, RN Cole, BO Palsson, H Zhang, M Betenbaugh (2012) Proteomic Analysis of Chinese Hamster Ovary Cells. *J Proteome Res* 11 (11):5265-5276.
- [87] PG Slade, M Hajivandi, CM Bartel, SF Gorfien (2012) Identifying the CHO Secretome using Mucin-type O-Linked Glycosylation and Click-chemistry. *J Proteome Res* 12 (11):6175-6186.
- [88] UM Lim, MGS Yap, YP Lim, L-T Goh, SK Ng (2013) Identification of Autocrine Growth Factors Secreted by CHO Cells for Applications in Single-Cell Cloning Media. *J Proteome Res* 7 (12):3496-3510.
- [89] NE Levy, KN Valente, LH Choe, KH Lee, AM Lenhoff (2014) Identification and characterization of host cell protein product-associated impurities in monoclonal antibody bioprocessing. *Biotechnol Bioeng* 5 (111):904-912.
- [90] KN Valente, AK Schaefer, HR Kempton, AM Lenhoff, KH Lee (2014) Recovery of Chinese hamster ovary host cell proteins for proteomic analysis. *Biotechnol J* 1 (9):87-99.
- [91] SJ North, H-H Huang, S Sundaram, J Jang-Lee, AT Etienne, A Trollope, S Chalabi, A Dell, P Stanley, SM Haslam (2010) Glycomics Profiling of Chinese Hamster Ovary Cell Glycosylation Mutants Reveals N -Glycans of a Novel Size and Complexity. *J Biol Chem* 8 (285):5759-5775.
- [92] S Tep, M Hincapie, WS Hancock (2012) The characterization and quantitation of glycomic changes in CHO cells during a bioreactor campaign. *Biotechnol Bioeng* 12 (109):3007-3017.

- [93] P Doolan, M Melville, P Gammell, M Sinacore, P Meleady, K McCarthy, L Francullo, M Leonard, T Charlebois, M Clynes (2008) Transcriptional profiling of gene expression changes in a PACE-transfected CHO DUKX cell line secreting high levels of rhBMP-2. *Mol Biotechnol* 3 (39):187-99.
- [94] PM Nissom, A Sanny, YJ Kok, YT Hiang, SH Chuah, TK Shing, YY Lee, KT Wong, WS Hu, MY Sim, R Philp (2006) Transcriptome and proteome profiling to understanding the biology of high productivity CHO cells. *Mol Biotechnol* 2 (34):125-40.
- [95] E Trummer, W Ernst, F Hesse, K Schriebl, C Lattenmayer, R Kunert, K Vorauer-Uhl, H Katinger, D Müller (2008) Transcriptional profiling of phenotypically different Epo-Fc expressing CHO clones by cross-species microarray analysis. *Biotechnol J* 7 (3):924-37.
- [96] JY Baik, MS Lee, SR An, SK Yoon, EJ Joo, YH Kim, HW Park, GM Lee (2006) Initial transcriptome and proteome analyses of low culture temperature-induced expression in CHO cells producing erythropoietin. *Biotechnol Bioeng* 2 (93):361-71.
- [97] H Swiderek, M Al-Rubeai (2007) Functional genome-wide analysis of antibody producing NS0 cell line cultivated at different temperatures. *Biotechnol Bioeng* 3 (98):616-30.
- [98] DC Wong, KT Wong, YY Lee, PN Morin, CK Heng, MG Yap (2006) Transcriptional profiling of apoptotic pathways in batch and fed-batch CHO cell cultures. *Biotechnol Bioeng* 2 (94):373-82.
- [99] A Kantardjieff, PM Nissom, SH Chuah, F Yusufi, NM Jacob, BC Mulukutla, M Yap, W-S Hu (2009) Developing genomic platforms for Chinese hamster ovary cells. *Biotechnol Adv* 6 (27):1028-35.
- [100] F Birzele, J Schaub, W Rust, C Clemens, P Baum, H Kaufmann, A Weith, TW Schulz, T Hildebrandt (2010) Into the unknown: expression profiling without genome sequence information in CHO by next generation sequencing. *Nucleic Acids Res* 12 (38):3999-4010.
- [101] S Hammond, JC Swanberg, M Kaplarevic, KH Lee (2011) Genomic sequencing and analysis of a Chinese hamster ovary cell line using Illumina sequencing technology. *BMC Genomics* 67 (12):1471-2164.
- [102] P Rabinovitch Introduction to cell cycle analysis. Disponible en: <http://www.phnxflow.com/Introduction%20to%20Cell%20Cycle%20Analysis.pdf>. Último acceso: 11/10/17.
- [103] KA Schafer (1998) The Cell Cycle: A Review. *Vet Pathol* 6 (35):461-478.
- [104] M Bono, V Simon (1998) Citometría de flujo: Principios básicos y aplicaciones, En: *Fundamentos de Inmunología*, (Ed: I Palomo, A Ferreira, C Sepúlveda, M Roseblatt, U Vergara), Universidad de Talca, p. 647-665.
- [105] Z Darzynkiewicz, G Juan (1997) Overview of Nuclei Acid Analysis, En: *Current Protocols in Cytometry* (Ed: J Robinson, Z Darzynkiewicz, P Dean, L Dressler, P Rabinovitch).
- [106] JL Castillo N, F Kawaguchi P, J Madariaga B, O Venegas R, E Lecannelier F, S Ocampo R, M Castillo N (1999) Aspectos que afectan el análisis de contenido de ADN por citometría de flujo. *Rev Med Chil* 11 (127):1385-1397.
- [107] TV Shankey, PS Rabinovitch, B Bagwell, KD Bauer, RE Duque, DW Hedley, BH Mayall, L Wheelless, C Cox (1993) Guidelines for implementation of clinical DNA cytometry. *Cytometry* 5 (14):472-477.

- [108] N Kumar, P Gammell, M Clynes (2007) Proliferation control strategies to improve productivity and survival during CHO based production culture. *Cytotechnology* 1–3 (53):33-46.
- [109] M Al-Rubeai, RP Singht (1998) Apoptosis in cell culture. *Curr Opin Biotechnol* (9):152-156.
- [110] A-P Zeng, W-D Deckwer, W-S Hu (1998) Determinants and rate laws of growth and death of hybridoma cells in continuous culture. *Biotechnol Bioeng* 6 (57):642-654.
- [111] A-P Zeng, W-D Deckwer (1999) Model Simulation and Analysis of Perfusion Culture of Mammalian Cells at High Cell Density. *Biotechnol Prog* 3 (15):373-382.
- [112] M Al-Rubeai, AN Emery (1990) Mechanisms and kinetics of monoclonal antibody synthesis and secretion in synchronous and asynchronous hybridoma cell cultures. *J Biotechnol* 1-2 (16):67-85.
- [113] M Al-Rubeai, AN Emery, S Chalder, DC Jan (1992) Specific monoclonal antibody productivity and the cell cycle-comparisons of batch, continuous and perfusion cultures. *Cytotechnology* 1-3 (9):85-97.
- [114] A Moore, J Mercer, G Dutina, CJ Donahue, KD Bauer, JP Mather, T Etcheverry, T Ryll (1997) Effects of temperature shift on cell cycle, apoptosis and nucleotide pools in CHO cell batch cultures. *Cytotechnology* (23):47-54.
- [115] M Fussenegger, S Schlatter, D Dätwyler, X Mazur, JE Bailey (1998) Controlled proliferation by multigene metabolic engineering enhances the productivity of Chinese hamster ovary cells. *Nat Biotechnol* 5 (16):468-472.
- [116] A V. Carvalhal, I Marcelino, MJT Carrondo (2003) Metabolic changes during cell growth inhibition by p27 overexpression. *Appl Microbiol Biotechnol* 2 (63):164-173.
- [117] N Ibarra, S Watanabe, J-X Bi, J Shuttleworth, M Al-Rubeai (2003) Modulation of Cell Cycle for Enhancement of Antibody Productivity in Perfusion Culture of NS0 Cells. *Biotechnol Prog* 1 (19):224-228.
- [118] SK Yoon, JY Song, GM Lee (2003) Effect of low culture temperature on specific productivity, transcription level, and heterogeneity of erythropoietin in Chinese hamster ovary cells. *Biotechnol Bioeng* 3 (82):289-98.
- [119] MB Fogolín, R Wagner, M Etcheverrigaray, R Kratje (2004) Impact of temperature reduction and expression of yeast pyruvate carboxylase on hGM-CSF-producing CHO cells. *J Biotechnol* 1-2 (109):179-191.
- [120] J-X Bi, J Shuttleworth, M Al-Rubeai (2004) Uncoupling of cell growth and proliferation results in enhancement of productivity in p21CIP1-arrested CHO cells. *Biotechnol Bioeng* 7 (85):741-749.
- [121] E Trummer, K Fauland, S Seidinger, K Schriebl, C Lattenmayer, R Kunert, K Vorauer-Uhl, R Weik, N Borth, H Katinger, D Mu (2006) Process Parameter Shifting: Part II . Biphase Cultivation — A Tool for Enhancing the Volumetric Productivity of Batch Processes Using Epo-Fc Expressing CHO Cells. *Biotechnol Bioeng* (94):1045-1052.
- [122] M Fussenegger, D Fassnacht, R Schwartz, JA Zanghi, M Graf, JE Bailey, R Pörtner (2000) Regulated overexpression of the survival factor bcl-2 in CHO cells increases viable cell density in batch culture and decreases DNA release in extended fixed-bed cultivation. *Cytotechnology* 1 (32):45-61.

- [123] H Kaufmann, X Mazur, R Marone, JE Bailey, M Fussenegger (2001) Comparative analysis of two controlled proliferation strategies regarding product quality, influence on tetracycline-regulated gene expression, and productivity. *Biotechnol Bioeng* 6 (72):592-602.
- [124] E Trummer, K Fauland, S Seidinger, K Schriebl, C Lattenmayer, R Kunert, K Vorauer-Uhl, R Weik, N Borth, H Katinger, D Müller (2006) Process parameter shifting: Part II. Biphasic cultivation—A tool for enhancing the volumetric productivity of batch processes using Epo-Fc expressing CHO cells. *Biotechnol Bioeng* 6 (94):1045-1052.
- [125] M Torres, R Zúñiga, M Gutierrez, M Vergara, N Collazo, J Reyes, J Berrios, JC Aguillon, MC Molina, C Altamirano (2018) Mild hypothermia upregulates myc and xbp1s expression and improves anti-TNF α production in CHO cells. *PLoS One* 3 (13):e0194510.
- [126] K Furukawa, K Ohsuye (1998) Effect of culture temperature on a recombinant CHO cell line producing a C-terminal α -amidating enzyme. *Cytotechnology* 2 (26):153.
- [127] K Furukawa, K Ohsuye (1999) Enhancement of productivity of recombinant α -amidating enzyme by low temperature culture. *Cytotechnology* 1/2 (31):85-94.
- [128] SM Schatz, RJ Kerschbaumer, G Gerstenbauer, M Kral, F Dorner, F Scheiflinger (2003) Higher expression of fab antibody fragments in a CHO cell line at reduced temperature. *Biotechnol Bioeng* 4 (84):433-438.
- [129] SK Yoon, JK Hong, SH Choo, JY Song, HW Park, GM Lee (2006) Adaptation of Chinese hamster ovary cells to low culture temperature: Cell growth and recombinant protein production. *J Biotechnol* 4 (122):463-472.
- [130] SR Fox, UA Patel, MGS Yap, DIC Wang (2004) Maximizing interferon-gamma production by chinese hamster ovary cells through temperature shift optimization: Experimental and modeling. *Biotechnol Bioeng* 2 (85):177-184.
- [131] M Bollati-Fogolín, G Forno, M Nimtz, HS Conradt, M Etcheverrigaray, R Kratje (2005) Temperature Reduction in Cultures of hGM-CSF-expressing CHO Cells: Effect on Productivity and Product Quality. *Biotechnol Prog* 1 (21):17-21.
- [132] KJR Clark, FWR Chaplin, SW Harcum (2004) Temperature Effects on Product-Quality-Related Enzymes in Batch CHO Cell Cultures Producing Recombinant tPA. *Biotechnol Prog* 6 (20):1888-1892.
- [133] S Chuppa, Y-S Tsai, S Yoon, S Shackelford, C Rozales, R Bhat, G Tsay, C Matanguihan, K Konstantinov, D Naveh (1997) Fermentor temperature as a tool for control of high-density perfusion cultures of mammalian cells. *Biotechnol Bioeng* 2 (55):328-338.
- [134] P Jorjani, SS Ozturk (1999) Effects of Cell Density and Temperature on Oxygen Consumption Rate for Different Mammalian Cell Lines. *Biotechnol Bioeng* 3 (64):349-356.
- [135] DCF Wong, KTK Wong, YY Lee, PN Morin, CK Heng, MGS Yap (2006) Transcriptional profiling of apoptotic pathways in batch and fed-batch CHO cell cultures. *Biotechnol Bioeng* 2 (94):373-382.

- [136] H Wassmann, C Greiner, S Hülsmann, D Moskopp, EJ Speckmann, J Meyer, H Van Aken (1998) Hypothermia as cerebroprotective measure. Experimental hypoxic exposure of brain slices and clinical application in critically reduced cerebral perfusion pressure. *Neurol Res* (20 Suppl 1):S61-5.
- [137] S Phadtare, J Alsina, M Inouye (1999) Cold-shock response and cold-shock proteins. *Curr Opin Microbiol* 2 (2):175-180.
- [138] MB Al-Fageeh, RJ Marchant, MJ Carden, CM Smales (2006) The cold-shock response in cultured mammalian cells: Harnessing the response for the improvement of recombinant protein production. *Biotechnol Bioeng* 5 (93):829-835.
- [139] H Nishiyama, H Higashitsuji, H Yokoi, K Itoh, S Danno, T Matsuda, J Fujita (1997) Cloning and characterization of human CIRP (cold-inducible RNA-binding protein) cDNA and chromosomal assignment of the gene. *Gene* 1-2 (204):115-20.
- [140] MS Sheikh, F Carrier, MA Papathanasiou, MC Hollander, Q Zhan, K Yu, AJ Fornace (1997) Identification of several human homologs of hamster DNA damage-inducible transcripts. Cloning and characterization of a novel UV-inducible cDNA that codes for a putative RNA-binding protein. *J Biol Chem* 42 (272):26720-6.
- [141] S Danno, K Itoh, T Matsuda, J Fujita (2000) Decreased Expression of Mouse Rbm3, a Cold-Shock Protein, in Sertoli Cells of Cryptorchid Testis. *Am J Pathol* 5 (156):1685-1692.
- [142] LA Sonna, J Fujita, SL Gaffin, CM Lilly (2002) Invited Review: Effects of heat and cold stress on mammalian gene expression. *J Appl Physiol* 4 (92):1725-1742.
- [143] C Jayat, MH Ratinaud (1993) Cell cycle analysis by flow cytometry: principles and applications. *Biol cell* 1-2 (78):15-25.
- [144] Z Darzynkiewicz (2011) Chapter 7: Critical Aspects in Analysis of Cellular DNA Content. En: *Curr Protoc Cytom* 1 (56):7.2.1-7.2.8.
- [145] AA Hurley (1997) Quality Control in Nucleic Acid Analysis. *Curr Protoc Cytom* 1 (00):7.2.1-7.2.5.
- [146] Z Darzynkiewicz, G Juan, E Bedner (2001) Chapter 8: Determining Cell Cycle Stages by Flow Cytometry, En *Current Protocols in Cell Biology*, Hoboken, NJ, EE.UU., John Wiley & Sons, Inc., Unit 8.4.
- [147] Illumina An introduction to Next-Generation Sequencing Technology. Disponible en: http://blumberg.bio.uci.edu/biod145-w2018/papers%20to%20consider%20adding/illumina_sequencing_introduction.pdf. Último acceso: 16/16/18.
- [148] MG Ross, C Russ, M Costello, A Hollinger, NJ Lennon, R Hegarty, C Nusbaum, DB Jaffe (2013) Characterizing and measuring bias in sequence data. *Genome Biol* 5 (14):R51.
- [149] T Nakazato, T Ohta, H Bono (2013) Experimental Design-Based Functional Mining and Characterization of High-Throughput Sequencing Data in the Sequence Read Archive. *PLoS One* 10 (8):e77910.
- [150] B Rodríguez-Santiago, L Armengol (2012) Tecnologías de secuenciación de nueva generación en diagnóstico genético pre- y postnatal. *Diagnóstico Prenat* 2 (23):56-66.

- [151] ER Mardis (2008) Next-Generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9):387-402.
- [152] Columbia Genome Center Genome Sequencing: Defining Your Experiment, Columbia University Department of Systems Biology. Disponible en: <https://systemsbiology.columbia.edu/genome-sequencing-defining-your-experiment>. Último acceso: 23/07/18.
- [153] Z Wang, M Gerstein, M Snyder (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet* 1 (10):57-63.
- [154] C Vallin Plous (2007) Microarreglos de ADN y sus aplicaciones en investigaciones biomédicas. *Revista CENIC* 28 (2).
- [155] W Zeng, A Mortazavi (2012) Technical considerations for functional sequencing assays. *Nat Immunol* 9 (13):802-807.
- [156] J Podnar, H Deiderick, G Huerta, S Hunicke-Smith (2014) Next-Generation Sequencing RNA-Seq Library Construction, En: *Current Protocols in Molecular Biology*, Hoboken, NJ, EE.UU., John Wiley & Sons, Inc., p. 4.21.1-4.21.19.
- [157] S Zhong, JG Joung, Y Zheng, YR. Chen, B Liu, Y Shao, JZ Xiang, Z Fei, JJ Giovannoni (2011) High-Throughput Illumina Strand-Specific RNA Sequencing Library Preparation. *Cold Spring Harb Protoc* 2011 (8):940-949.
- [158] SR Head, HK Komori, SA LaMere, T Whisenant, F Van Nieuwerburgh, DR Salomon, P Ordoukhanian (2014) Library construction for next-generation sequencing: Overviews and challenges. *Biotechniques* 2 (56).
- [159] C Monger, PS Kelly, C Gallagher, M Clynes, N Barron, C Clarke (2015) Towards next generation CHO cell biology: Bioinformatics methods for RNA-Seq-based expression profiling. *Biotechnol J* 7 (10):950-966.
- [160] R Lindner, CC Friedel (2012) A Comprehensive Evaluation of Alignment Algorithms in the Context of RNA-Seq. *PLoS One* 12 (7):e52403.
- [161] KD Hansen, SE Brenner, S Dudoit (2010) Biases in Illumina transcriptome sequencing caused by random hexamer priming. *Nucleic Acids Res* 12 (38):e131-e131.
- [162] KJ Livak, TD Schmittgen (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the 2- $\Delta\Delta$ CT Method. *Methods* 4 (25):402-408.
- [163] SA Chappell, GC Owens, VP Mauro (2001) A 5' Leader of Rbm3, a Cold Stress-induced mRNA, Mediates Internal Initiation of Translation with Increased Efficiency under Conditions of Mild Hypothermia. *J Biol Chem* 40 (276):36917-22.
- [164] HK Tan, MM Lee, MGS Yap, DIC Wang (2008) Overexpression of cold-inducible RNA-binding protein increases interferon- γ production in Chinese-hamster ovary cells. *Biotechnol Appl Biochem* 4 (49):247.
- [165] J Dresios, A Aschrafi, GC Owens, PW Vanderklish, GM Edelman, VP Mauro (2005) Cold stress-induced protein Rbm3 binds 60S ribosomal subunits, alters microRNA levels, and enhances global protein synthesis. *Proc Natl Acad Sci* 6 (102):1865-1870.
- [166] N Kumar, P Gammell, P Meleady, M Henry, M Clynes (2008) Differential protein expression following low temperature culture of suspension CHO-K1 cells. *BMC Biotechnol* 1 (8):42.

- [167] BD Chang, K Watanabe, E V Broude, J Fang, JC Poole, T V Kalinichenko, IB Roninson (2000) Effects of p21Waf1/Cip1/Sdi1 on cellular gene expression: implications for carcinogenesis, senescence, and age-related diseases. *Proc Natl Acad Sci U S A* 8 (97):4291-6.
- [168] A Moore, J Mercer, G Dutina, CJ Donahue, KD Bauer, JP Mather, T Etcheverry, T Ryll (1997) Effects of temperature shift on cell cycle, apoptosis and nucleotide pools in CHO cell batch cultures. *Cytotechnology* 1/3 (23):47-54.
- [169] BC Mulukutla, M Gramer, W-S Hu (2012) On metabolic shift to lactate consumption in fed-batch culture of mammalian cells. *Metab Eng* 2 (14):138-149.
- [170] W Zhou, J Rehm, W-S Hu (1995) High viable cell concentration fed-batch cultures of hybridoma cells through on-line nutrient feeding. *Biotechnol Bioeng* 6 (46):579-587.
- [171] W Zhou, W Zhou, J Rehm, A Europa, W-S Hu (1997) Alteration of mammalian cell metabolism by dynamic nutrient feeding. *Cytotechnology* 2 (24):99-108.
- [172] AF Europa, A Gambhir, PC Fu, WS Hu (2000) Multiple steady states with distinct cellular metabolism in continuous culture of mammalian cells. *Biotechnol Bioeng* 1 (67):25-34.
- [173] E Harreither, M Hackl, J Pichler, S Shridhar, N Auer, PP Łabaj, M Scheideler, M Karbiener, J Grillari, DP Kreil, N Borth (2015) Microarray profiling of preselected CHO host cell subclones identifies gene expression patterns associated with in-creased production capacity. *Biotechnol J* 10 (10):1625-1638.
- [174] M Bollati-Fogolín, G Forno, M Nimtz, HS Conradt, M Etcheverrigaray, R Kratje (2008) Temperature Reduction in Cultures of hGM-CSF-expressing CHO Cells: Effect on Productivity and Product Quality. *Biotechnol Prog* 1 (21):17-21.
- [175] CG Walther, R Whitfield, DC James (2016) Importance of Interaction between Integrin and Actin Cytoskeleton in Suspension Adaptation of CHO cells. *Appl Biochem Biotechnol* 7 (178):1286-1302.
- [176] N Lee, J Shin, JH Park, GM Lee, S Cho, BK Cho (2016) Targeted Gene Deletion Using DNA-Free RNA-Guided Cas9 Nuclease Accelerates Adaptation of CHO Cells to Suspension Culture. *ACS Synth Biol* 11 (5):1211-1219.
- [177] M Pourhassan-Moghaddam, B Farhadi, K Nejati-Koshki (2012) Promoters of Cancer Genes for Recombinant Protein Expression in Human Cancer Cell Lines. *BiolImpacts* 1 (2):55-60.
- [178] New England Biolabs Inc. NEBNext Ultra Directional RNA Library Prep Kit for Illumina® | NEB.
Disponibile en: <https://international.neb.com/products/e7420-nebnext-ultra-directional-rna-library-prep-kit-for-illumina#ProductInformation>.
Último acceso: 06/09/18.
- [179] S Andrews (2010) FastQC: a quality control tool for high throughput sequence data.
Disponibile en: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
Último acceso: 09/08/18.

- [180] AM Bolger, M Lohse, B Usadel (2014) Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 15 (30):2114-2120.
- [181] B Langmead, C Trapnell, M Pop, SL Salzberg (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 3 (10):R25.
- [182] C Trapnell, A Roberts, L Goff, G Pertea, D Kim, DR Kelley, H Pimentel, SL Salzberg, JL Rinn, L Pachter (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* 3 (7):562-78.
- [183] A Dobin, CA Davis, F Schlesinger, J Drenkow, C Zaleski, S Jha, P Batut, M Chaisson, TR Gingeras (2013) STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 1 (29):15-21.
- [184] S Heinz, C Benner, N Spann, E Bertolino, YC Lin, P Laslo, JX Cheng, C Murre, H Singh, CK Glass (2010) Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Mol Cell* 4 (38):576-589.
- [185] B Ripley (2001) The R Project in Statistical Computing. Disponible en: <https://www.r-project.org/>. Último acceso: 13/08/18.
- [186] L Goff, C Trapnell, D Kelley (2013) CummeRbund: Analysis, exploration, manipulation, and visualization of Cufflinks high-throughput sequencing data. Disponible en: <https://rdrr.io/bioc/cummeRbund/>. Último acceso: 29/03/18.
- [187] MJL de Hoon, S Imoto, J Nolan, S Miyano (2004) Open source clustering software. *Bioinformatics* 9 (20):1453-1454.
- [188] AJ Saldanha (2004) Java Treeview--extensible visualization of microarray data. *Bioinformatics* 17 (20):3246-3248.
- [189] PD Thomas, MJ Campbell, A Kejariwal, H Mi, B Karlak, R Daverman, K Diemer, A Muruganujan, A Narechania (2003) PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res* 9 (13):2129-2141.
- [190] W Luo, C Brouwer (2013) Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics* 14 (29):1830-1831.
- [191] M Kanehisa, S Goto (2000) KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 1 (28):27-30.
- [192] Y Benjamini, Y Hochberg (1995) Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1 (57):289-300.
- [193] J Ye, G Coulouris, I Zaretskaya, I Cutcutache, S Rozen, TL Madden (2012) Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 1 (13):134.
- [194] A Untergasser, I Cutcutache, T Koressaar, J Ye, BC Faircloth, M Remm, SG Rozen (2012) Primer3--new capabilities and interfaces. *Nucleic Acids Res* 15 (40):e115.
- [195] ZM De Guglielmo, R Fernandez, D Silva, P Ciencias (2016) Principales promotores utilizados en la transformación genética de plantas Main promoters used in plant genetic transformation. *Rev Colomb Biotecnol* (2):119-128.

- [196] N Romanova, T Noll (2018) Engineered and Natural Promoters and Chromatin-Modifying Elements for Recombinant Protein Expression in CHO Cells. *Biotechnol J* 3 (13):1700232.
- [197] JK Cheng, HS Alper (2016) Transcriptomics-Guided Design of Synthetic Promoters for a Mammalian System. *ACS Synth Biol* 12 (5):1455–1465.
- [198] JY Qin, L Zhang, KL Clift, I Hular, AP Xiang, B-Z Ren, BT Lahn (2010) Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLoS One* 5 (5):e10611.
- [199] BE Eddy, GS Borman, WH Berkeley, RD Young (1961) Tumors Induced in Hamsters by Injection of Rhesus Monkey Kidney Cell Extracts. *Exp Biol Med* 1 (107):191-197.
- [200] AJ Brown, DC James (2016) Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnol Adv* 5 (34):492-503.
- [201] H Thaisuchat, M Baumann, J Pontiller, F Hesse, W Ernst (2011) Identification of a novel temperature sensitive promoter in CHO cells. *BMC Biotechnol* 1 (11):51.
- [202] G Brightwell, V Poirier, E Cole, S Ivins, KW Brown (1997) Serum-dependent and cell cycle-dependent expression from a cytomegalovirus-based mammalian expression vector. *Gene* 1 (194):115-23.
- [203] DW Kim, T Uetsuki, Y Kaziro, N Yamaguchi, S Sugano (1990) Use of the human elongation factor 1 alpha promoter as a versatile and efficient expression system. *Gene* 2 (91):217-23.
- [204] J Running Deer, DS Allison (2004) High-level expression of proteins in mammalian cells using transcription regulatory sequences from the Chinese hamster EF-1alpha gene. *Biotechnol Prog* 3 (20):880-9.
- [205] AH Minn, C Hafele, A Shalev (2005) Thioredoxin-Interacting Protein Is Stimulated by Glucose through a Carbohydrate Response Element and Induces β -Cell Apoptosis. *Endocrinology* 5 (146):2397-2405.
- [206] Z Wang, YP Rong, MH Malone, MC Davis, F Zhong, CW Distelhorst (2006) Thioredoxin-interacting protein (txnip) is a glucocorticoid-regulated primary response gene involved in mediating glucocorticoid-induced apoptosis. *Oncogene* 13 (25):1903-1913.
- [207] S-Y Kang, Y-G Kim, S Kang, HW Lee, EG Lee (2016) A novel regulatory element (E77) isolated from CHO-K1 genomic DNA enhances stable gene expression in Chinese hamster ovary cells. *Biotechnol J* 5 (11):633-41.
- [208] R Mehrotra, G Gupta, R Sethi, P Bhalothia, N Kumar, S Mehrotra (2011) Designer promoter: an artwork of cis engineering. *Plant Mol Biol* 6 (75):527-36.
- [209] ML Roberts (2011) The Use of Functional Genomics in Synthetic Promoter Design, En: *Computational Biology and Applied Bioinformatics* (Ed: H Silverio Lopes, L Magalhães Cruz), p. 375-396.
- [210] R Ogawa, G Kagiya, T Kodaki, S Fukuda, K Yamamoto (2007) Construction of strong mammalian promoters by random cis-acting element elongation. *Biotechniques* 5 (42):628-633.
- [211] GM Edelman, R Meech, GC Owens, FS Jones (2000) Synthetic promoter elements obtained by nucleotide sequence variation and selection for activity. *Proc Natl Acad Sci U S A* 7 (97):3038-43.

- [212] MG Grabherr, J Pontiller, E Mauceli, W Ernst, M Baumann, T Biagi, R Swofford, P Russell, MC Zody, F Di Palma, K Lindblad-Toh, RM Grabherr (2011) Exploiting nucleotide composition to engineer promoters. *PLoS One* 5 (6):e20136.
- [213] H Zhou, Z Liu, Z Sun, Y Huang, W Yu (2010) Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. *J Biotechnol* 2 (147):122-129.
- [214] S Turan, M Galla, E Ernst, J Qiao, C Voelkel, B Schiedlmeier, C Zehe, J Bode (2011) Recombinase-Mediated Cassette Exchange (RMCE): Traditional Concepts and Current Challenges. *J Mol Biol* 2 (407):193-221.
- [215] W Wang, W Zheng, F Hu, X He, D Wu, W Zhang, H Liu, X Ma (2018) Enhanced Biosynthesis Performance of Heterologous Proteins in CHO-K1 Cells Using CRISPR-Cas9. *ACS Synth Biol* 5 (7):1259-1268.
- [216] LM Grav, JS Lee, S Gerling, TB Kallehauge, AH Hansen, S Kol, GM Lee, LE Pedersen, HF Kildegaard (2015) One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol J* 9 (10):1446-1456.
- [217] YL Jia, X Guo, JT Lu, XY Wang, L Le Qiu, TY Wang (2018) CRISPR/Cas9-mediated gene knockout for DNA methyltransferase Dnmt3a in CHO cells displays enhanced transgenic expression and long-term stability. *J Cell Mol Med* 22 (9):4106-4116.
- [218] JS Lee, TB Kallehauge, LE Pedersen, HF Kildegaard (2015) Site-specific integration in CHO cells mediated by CRISPR/Cas9 and homology-directed DNA repair pathway. *Sci Rep* (5):1-11.
- [219] R Bachu, I Bergareche, LA Chasin (2015) CRISPR-Cas targeted plasmid integration into mammalian cells via non-homologous end joining. *Biotechnol Bioeng* 10 (112):2154-2162.
- [220] M Zhao, J Wang, M Luo, H Luo, M Zhao, L Han, M Zhang, H Yang, Y Xie, H Jiang, L Feng, H Lu, J Zhu (2018) Rapid development of stable transgene CHO cell lines by CRISPR/Cas9-mediated site-specific integration into C12orf35. *Appl Microbiol Biotechnol* 102 (14):6105-6117.
- [221] R Dreos, G Ambrosini, R Groux, R Cavin Périer, P Bucher (2017) The eukaryotic promoter database in its 30th year: focus on non-vertebrate organisms. *Nucleic Acids Res* 45 (D1):D51–D55.
- [222] TL Bailey, M Boden, F a Buske, M Frith, CE Grant, L Clementi, J Ren, WW Li, WS Noble (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* (37):W202-8.
- [223] NE Sanjana, O Shalem, F Zhang (2014) Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* 8 (11):783-784.
- [224] M Stemmer, T Thumberger, M del Sol Keyer, J Wittbrodt, JL Mateo (2015) CCTop: An Intuitive, Flexible and Reliable CRISPR/Cas9 Target Prediction Tool. *PLoS One* 4 (10):e0124633.
- [225] K Labun, TG Montague, JA Gagnon, SB Thyme, E Valen (2016) CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. *Nucleic Acids Res* 44 (W1):W272-6.
- [226] T Dull, R Zufferey, M Kelly, RJ Mandel, M Nguyen, D Trono, L Naldini (1998) A third-generation lentivirus vector with a conditional packaging system. *J Virol* 11 (72):8463-71.