

**UNIVERSIDAD AUTÓNOMA DE NUEVO LEÓN**

**FACULTAD DE CIENCIAS BIOLÓGICAS**



**ANÁLISIS TRANSCRIPCIONAL DE LA VÍA DE SECRECIÓN Y  
METABOLISMO DEL METANOL DURANTE LA  
PRODUCCIÓN DE UNA PROTEÍNA  
HETERÓLOGA EN *Pichia pastoris***

**Por**

**KARLA BEATRÍZ FERNÁNDEZ CANO**

Como requisito parcial para obtener el Grado de  
**MAESTRO EN CIENCIAS CON ACENTUACIÓN EN MICROBIOLOGÍA**

Julio, 2015

**ANÁLISIS TRANSCRIPCIONAL DE LA VÍA DE SECRECIÓN Y  
METABOLISMO DEL METANOL DURANTE LA  
PRODUCCIÓN DE UNA PROTEÍNA  
HETERÓLOGA EN *Pichia pastoris***

**Comité de Tesis**

---

**Director/Presidente**

Dra. Martha Guerrero Olazarán

---

**Secretario**

Dr. José M. Viader Salvadó

---

**Vocal:**

Dra. Laura María Trejo Ávila

---

**Vocal:**

M.C. Eddy Luz Cab Barrera

---

**Vocal:**

Dr. Juan Antonio Gallegos López

"Un verdadero espíritu de rebeldía es aquel que busca la felicidad en esta vida".  
-Henrik Johan Ibsen

"El individuo ha luchado siempre para no ser absorbido por la tribu. Si lo intentas, a menudo estarás solo, y a veces asustado. Pero ningún precio es demasiado alto por el privilegio de ser uno mismo."  
-Friedrich Nietzsche

"Por muy dura que nos parezca la vida, mientras haya vida hay esperanza."  
-Stephen Hawking

"El payaso no soy yo, sino esa sociedad tan monstruosamente cínica e inconscientemente ingenua que interpreta un papel de serio para disfrazar su locura"  
-Salvador Dalí

"La magia es un puente que te permite ir del mundo visible hacia el invisible. Y aprender las lecciones de ambos mundos."  
-Paulo Coelho

## LUGAR DE TRABAJO

Este trabajo se realizó en el laboratorio 5 del instituto de Biotecnología de la Facultad de Ciencias Biológicas de la Universidad Autónoma de Nuevo León bajo la dirección de la Dra. Martha Guerrero Olazarán

### **Este trabajo se presentó en los siguientes congresos:**

- 1.- M. Guerrero-Olazarán, K.B. Fernández-Cano, J.M. Viader-Salvadó. Transcriptional analysis of *Pichia pastoris* under processes of high and low levels of recombinant protein production. 8th Conference on Recombinant Protein Production, Palma de Mallorca (España), 22-24 de Abril de 2015. (Modalidad cartel).
- 2.- K.B. Fernández-Cano, J.M. Viader-Salvadó, M. Castillo-Galván, M. Guerrero-Olazarán. Análisis transcripcional durante la producción de una proteína heteróloga en *Pichia pastoris*. Congreso Nacional de Biotecnología y Bioingeniería, Sociedad Mexicana de Biotecnología y Bioingeniería, A.C. Guadalajara, Jalisco, (México). 21-26 de Junio del 2015. (Modalidad oral).

## **AGRADECIMIENTOS**

Agradezco a CONACYT por haberme otorgado la beca para realizar mis estudios de maestría

Gracias a mis tutores, la Doctora Martha Guerrero Olazarán y el Doctor José María Viader Salvadó, por haberme permitido ser parte de éste proyecto, gracias por sus consejos, criterio, paciencia y dedicación.

Gracias a la Dra. Laura M. Trejo Ávila y la MC. Eddy Luz Cab Barrera por haber sido mis tutoras durante los seminarios, por sus comentarios positivos y palabras de aliento.

Gracias a mis compañeros del L5 por haberme ayudado a integrarme en el laboratorio y algunos de ellos por enseñarme técnicas indispensables para éste trabajo, gracias Mónica, Ana, Antonio, Paco, Mariana, Mauricio, Jonathan y David.

Gracias a mis compañeros de posgrado por sus palabras de aliento, consejos, convivencia, trabajo de equipo y de más, gracias Yolanda, Alberto, Adrián, Diego, Ramiro, Luis Fernando, Fabiola, Lety y todas esas personas que estuvieron en los salones de clase.

Gracias a mi familia por el apoyo incondicional que siempre me han dado, por tener confianza en mí, y apoyarme en mis decisiones, en especial a mis padres. Gracias a mis tíos y primo por la estancia larga en su hogar que hasta el momento se ha dado.

Gracias a Karlenne por su apoyo y comprensión durante todo el tiempo en que realicé mis estudios.

## **DEDICATORIA**

Esta tesis está dedicada a mis padres, porque siempre me han dado la oportunidad de elegir mis propios caminos, a mi hermano y a mi familia en general, que aunque nuestras convivencias son cortas, siempre han estado ahí para mí cuando lo necesito.

# ÍNDICE DE CONTENIDO

<b>AGRADECIMIENTOS</b>	<b>I</b>
<b>DEDICATORIA</b>	<b>I</b>
<b>ÍNDICE DE CONTENIDO</b>	<b>II</b>
<b>ÍNDICE DE TABLAS</b>	<b>VI</b>
<b>ÍNDICE DE FIGURAS</b>	<b>IX</b>
<b>LISTA DE SÍMBOLOS Y ABREVIATURAS</b>	<b>XIV</b>
<b>RESUMEN</b>	<b>XVI</b>
<b>1. INTRODUCCIÓN</b>	<b>1</b>
<b>2. ANTECEDENTES</b>	<b>3</b>
<b>2.1 Microorganismos metilotróficos</b>	<b>3</b>
2.1.1 Importancia de las especies de levaduras metilotróficas	3
2.1.1.1 Vía de utilización del metanol (vía MUT) en <i>P. pastoris</i>	7
2.1.2 <i>P. pastoris</i> como hospedero para la expresión de proteínas heterólogas y genotipo de la cepa KM71	9
2.1.2.1 <i>Pichia pastoris</i> , productora de proteínas recombinantes	11
2.1.2.2 Factores que influyen en la producción y secreción de proteínas heterólogas	12
2.1.2.2.1 Bioproceso optimizado para la sobreproducción de una proteína heteróloga en <i>P. pastoris</i>	13
2.1.3 Vía de secreción de proteínas en <i>P. pastoris</i>	15
2.1.3.1 Respuesta a proteínas mal plegadas (UPR)	20
2.1.3.2 Vía de degradación de proteínas asociadas al Retículo endoplásmico (ERAD)	22
2.1.3.3 Autofagia	23
2.1.3.4 Factor sensitivo N-etilmaleimida de fijación a receptores de proteína (SNAREs)	24
2.1.4 Estudio de genes que afectan la producción de proteínas heterólogas en <i>P. pastoris</i>	25
<b>2.2 Estudio de Transcriptomas</b>	<b>26</b>
2.2.1 Técnicas utilizadas para obtener perfiles de expresión génica	26
2.2.1.1 Técnica de RNA-Seq	27
2.2.1.1.1 Plataformas utilizadas para RNA-Seq	28

2.2.1.1.2 Preparación de bibliotecas de DNAC y secuenciación	29
2.2.1.1.4 Análisis de datos obtenidos por RNA-Seq	34
2.2.1.1.5 Análisis de calidad	35
2.2.1.1.5 Alineamientos y ensamblaje	35
2.2.1.1.5 Experimentos de RNA-seq con Tophat y Cufflinks	36
2.2.1.1.6 Análisis de enriquecimiento de genes mediante GO	37
2.2.1.2 Validación del análisis transcripcional obtenido por la técnica de RNA-seq	38
2.2.1.2.1 Técnica de RT-qPCR para el análisis de expresión diferencial de genes	39
2.2.1.2.2 Análisis de los datos obtenidos por RT-qPCR	40
<b>3. JUSTIFICACIÓN</b>	<b>45</b>
<b>4. HIPÓTESIS</b>	<b>45</b>
<b>5. OBJETIVO GENERAL</b>	<b>46</b>
<b>6. OBJETIVOS ESPECÍFICOS</b>	<b>46</b>
<b>7. MATERIALES Y MÉTODOS</b>	<b>47</b>
<b>7.1 Área de trabajo</b>	<b>47</b>
<b>7.2 Material biológico</b>	<b>47</b>
<b>7.3 Materiales y Equipos Empleados</b>	<b>48</b>
<b>7.3.1 Extracción de RNA</b>	<b>48</b>
7.3.1.1 Preparación de reactivos extras al kit de extracción	48
7.3.2 Tratamiento DNasa	48
7.3.3 Cuantificación e integridad de las muestras de RNA	48
7.3.4 Síntesis de DNAC	49
7.3.5 Ensayos de qPCR	49
<b>7.4 Diseño experimental</b>	<b>49</b>
<b>7.5 Análisis transcripcional</b>	<b>52</b>
7.5.1 Protocolo de Extracción de RNA	52
7.5.2 Obtención de las librerías de DNAC y secuenciación de los transcritos	53
7.5.3 Análisis Bioinformáticos	54
7.5.3.1 Análisis de calidad de los archivos FASTQ	54
7.5.3.2 Mapeo y ensamblaje de los transcritos	54
7.5.3.3 Analisis diferencial con Cuffdiff	56
7.5.4 Análisis global de la expresión diferencial de genes	59

7.5.4.1	Análisis de enriquecimiento de genes	59
7.5.4.1.1	Diagrama de Venn y enriquecimiento de genes	61
7.5.4.2	Análisis de expresión diferencial del gen heterólogo FTEII	61
7.5.4.3	Análisis de expresión diferencial de rutas metabólicas específicas	62
7.5.4.3.1	Análisis de expresión diferencial de los genes en la vía de asimilación del metanol (vía MUT)	62
7.5.4.3.2	Análisis de expresión diferencial de los genes en la vía de secreción	62
7.5.4.3.3	Análisis de expresión diferencial de genes involucrados en el estrés en RE durante la secreción	63
<b>7.6</b>	<b>Validación del análisis transcripcional</b>	<b>63</b>
7.6.1	Tratamiento con DNasa	63
7.6.2	Protocolo de Síntesis de DNAc	63
7.6.3	Selección y Diseño de oligonucleótidos	64
7.6.4	Descripción de las mezclas de trabajo utilizadas para la estandarización de los ensayos de qPCR mediante SYBR Green	66
7.6.4.1	Curvas estándar para la corrección de la eficiencia en los ensayos de qPCR para cada uno de los genes seleccionados	69
7.6.4.2	Ensayos de RT-qPCR para la determinación de los niveles de expresión de cada gen	70
7.6.4.3	Calculo de expresión relativa para cada uno de los genes	71
7.6.5	Comparación de las técnicas de análisis de expresión diferencial de genes (RNA-seq vs RT-qPCR)	72
<b>8.</b>	<b>RESULTADOS</b>	<b>73</b>
<b>8.1</b>	<b>Obtención de preparaciones de RNA a partir de células de <i>P. pastoris</i> KM71FTEII cultivadas en biorreactor para la secuenciación de transcritos</b>	<b>73</b>
8.1.1	Obtención de las librerías de DNAc y secuenciación de transcritos	75
<b>8.2</b>	<b>Análisis Bioinformático de los transcritos secuenciados</b>	<b>76</b>
8.2.1	Análisis de calidad de las secuencias	76
8.2.2	Modificación del cromosoma antes del mapeo con el genoma de referencia	78
8.2.3	Mapeo de los transcritos con un genoma de referencia	78
8.2.4	Análisis diferencial con cufflinks	80



8.2.5 Análisis global de expresión diferencial de genes	80
8.2.5.1 Análisis global de enriquecimiento de genes	83
8.2.5.2 Genes sobre- y sub-expresados comunes en las tres comparaciones determinados mediante el diagrama de Venn	92
8.2.5.2 Análisis de expresión diferencial del gen Heterólogo	101
8.2.5.3 Análisis de expresión diferencial de las Rutas metabólicas específicas	102
8.2.5.3.1 Análisis de expresión diferencial de genes de la vía del metabolismo del metanol (vía MUT)	104
8.2.5.3.2 Análisis de expresión diferencial de genes involucrados en la vía de secreción	109
8.2.5.3.3 Análisis de expresión diferencial de genes involucrados en el estrés en RE durante la secreción	128
<b>8.3 Validación del análisis transcripcional mediante RT-qPCR de los genes seleccionados</b>	<b>141</b>
8.3.1 Selección y Diseño de oligonucleótidos específicos para los genes involucrados en la vía de secreción y metabolismo del metanol.	141
8.3.3 Curva Estándar para la obtención de los intervalos de linealidad y eficiencia de cada uno de los genes seleccionados para los ensayos de qPCR.	143
8.3.4 Ensayos de qPCR mediante SYBR Green para cada uno de los genes evaluados	144
8.3.4.1 Calculo de los niveles de expresión relativa	146
8.3.4.2 Comparación de las técnicas de RT-qPCR y RNA-seq	147
<b>9. DISCUSIÓN</b>	<b>150</b>
<b>10. CONCLUSIONES</b>	<b>170</b>
<b>11. PERSPECTIVAS</b>	<b>171</b>
<b>12. BIBLIOGRAFÍA</b>	<b>172</b>
<b>ANEXO 1. OLIGONUCLEÓTIDOS DISEÑADOS</b>	<b>181</b>
<b>ANEXO 2. CARACTERÍSTICAS DE OLIGONUCLEÓTIDOS DISEÑADOS</b>	<b>182</b>
<b>ANEXO 3. ALINEAMIENTOS Y FORMACIÓN DE ESTRUCTURAS SECUNDARIAS</b>	<b>185</b>
<b>ANEXO 4. CURVAS DE DISOCIACIÓN, AMPLIFICACIÓN Y ESTÁNDAR</b>	<b>188</b>
<b>RESUMEN AUTOBIOGRÁFICO</b>	<b>194</b>

## ÍNDICE DE TABLAS

<b>Tabla 1.</b> Genes que codifican para las enzimas clave localizadas en los peroxisomas de algunas especies de levadura que comparten alguna variante conservada (Van der Klei <i>et al.</i> , 2006).	5
<b>Tabla 2.</b> Los promotores más usados y más recientemente establecidos para la expresión de proteínas heterólogas (Ahmad <i>et al.</i> , 2014).	6
<b>Tabla 3.</b> Genes de <i>P. pastoris</i> involucrados en la ruta de asimilación del metanol (De Shutter <i>et al.</i> , 2009).	9
<b>Tabla 4.</b> Cepas disponibles para expresar proteínas heterólogas en <i>P. pastoris</i> (Nett & Gerngross, 2003).	10
<b>Tabla 5.</b> Secuencias señal usadas para secretar la proteína al espacio extracelular (Ahmad <i>et al.</i> , 2014).	16
<b>Tabla 6.</b> Procesamiento de Archivos FASTQ que se generan después de la secuenciación.	37
<b>Tabla 7.</b> Tecnologías disponibles para RT-qPCR (VanGuilder <i>et al.</i> , 2008).	40
<b>Tabla 8.</b> Diferencias entre las dos diferentes estrategias de cuantificación y sus ejemplos.	42
<b>Tabla 9.</b> Condiciones de cultivo y rendimiento en el paso de inducción con MeOH en CM y CP (Viader-Salvadó <i>et al.</i> , 2013).	47
<b>Tabla 10.</b> Mezcla de trabajo utilizada en el ensayo de qPCR con SYBR Green de todos los DNAc (dilución 1:25).	66
<b>Tabla 11.</b> Mezcla de trabajo utilizada en el ensayo de qPCR con SYBR Green de los controles negativos de la qPCR.	67
<b>Tabla 12.</b> Mezcla de trabajo para el ensayo de qPCR con SYBR Green de los controles negativos de la qPCR.	67
<b>Tabla 13.</b> Condiciones empleadas en el termociclador para los ensayos de qPCR de los genes seleccionados.	68
<b>Tabla 14.</b> Concentración de RNA y relación A260/A280 en las preparaciones de RNA que provienen de células de <i>P. pastoris</i> KM71FTEII cultivadas en biorreactor.	73
<b>Tabla 15.</b> Número de integridad de RNA (RIN) de la preparaciones de RNA que provienen de las células de <i>P. pastoris</i> KM71FTEII cultivadas en biorreactor.	74
<b>Tabla 16.</b> Resumen de las lecturas obtenidas por el secuenciador y la calidad de las cinco librerías de DNAc obtenidas.	76
<b>Tabla 17.</b> Número de lecturas finales y removidas después del recorte de secuencias con el criterio de calidad de un valor de puntaje Phred $\geq 20$ .	77
<b>Tabla 18.</b> Anotaciones que se añadieron al archivo gtf que contiene las anotaciones del genoma de la cepa GS115 para agregar las regiones incorporadas del plásmido pPIC9fteII.	78
<b>Tabla 19.</b> Número y porcentaje de lecturas mapeadas en base al genoma modificado de referencia de la cepa GS115 de <i>P. pastoris</i> .	79
<b>Tabla 20.</b> Nombre y función de las proteínas codificadas por los genes involucrados en el inicio del metabolismo del metanol.	107
<b>Tabla 21.</b> Nombre y función de las proteínas codificadas por los genes involucrados en la vía de asimilación del metanol.	107
<b>Tabla 22.</b> Nombre y función de las proteínas codificadas por los genes involucrados en la vía de desasimilación del metanol.	108

<b>Tabla 23.</b> Nombre y función de proteínas codificadas por los genes involucrados en la translocación y transporte de proteínas.	114
<b>Tabla 24.</b> Nombre y función de genes que codifican chaperonas involucradas en el plegamiento de proteínas en el RE	118
<b>Tabla 25.</b> Nombre y función de las proteínas codificadas por los genes involucrados en la síntesis de glicoproteínas en RE.	121
<b>Tabla 26.</b> Nombre y función de proteínas codificadas por genes involucrados en el procesamiento y tráfico de glicoproteínas en aparato de Golgi.	124
<b>Tabla 27.</b> Nombre y función de proteínas codificadas por genes involucrados en la homeostásis intracelular y degradación o recambio proteolítico en sistemas vacuolares.	127
<b>Tabla 28.</b> Nombre y función de las proteínas codificadas por los genes involucrados en UPR.	133
<b>Tabla 29.</b> Nombre y función de las proteínas codificadas por los genes involucrados en el complejo proteico del proteosoma mediado por ERAD.	135
<b>Tabla 30.</b> Nombre y función de las proteínas codificadas por los genes involucrados en autofagia.	138
<b>Tabla 31.</b> Nombre y función de las proteínas codificadas por los genes involucrados en SNARE.	140
<b>Tabla 32.</b> Genes utilizados en la validación transcripcional involucrados en la vía de secreción seleccionados a partir de los resultados obtenidos en la comparación CM/CP.	141
<b>Tabla 33.</b> Genes utilizados para la validación transcripcional involucrados en la vía del metabolismo del metanol seleccionados a partir de los resultados obtenidos en la comparación CM/CP.	142
<b>Tabla 34.</b> Eficiencia, linealidad, intervalos de linealidad y umbral de fluorescencia de los amplicones evaluados.	143
<b>Tabla 35.</b> Valores de ciclo umbral ( $C_T$ ), media y CV (%) de cada uno de los genes evaluados en el cultivo C0 antes de la inducción con metanol.	144
<b>Tabla 36.</b> Valores de ciclo umbral ( $C_T$ ), media y CV(%) de cada uno de los genes evaluados en el cultivo CM47-1 a las 47 horas de inducción con metanol.	145
<b>Tabla 37.</b> Valores de ciclo umbral ( $C_T$ ), media y CV(%) de cada uno de los genes evaluados en el cultivo CM47-2 a las 47 horas de inducción con metanol.	145
<b>Tabla 38.</b> Valores de ciclo umbral ( $C_T$ ), media y CV(%) de cada uno de los genes evaluados en el cultivo CP47-1 a las 47 horas de inducción con metanol.	146
<b>Tabla 39.</b> Valores de Expresión relativa de cada uno de los genes en las condiciones de cultivo evaluadas.	147
<b>Tabla 40.</b> Comparación del $\log_2$ (Fold_change) obtenido en el análisis transcripcional y el $\log_2$ calculado a partir de los niveles de expresión relativa correspondiente a cada uno de los genes.	148
<b>Tabla 41.</b> Oligonucleótidos específicos utilizados para la amplificación de los genes seleccionados de la vía de secreción.	181
<b>Tabla 42.</b> Oligonucleótidos específicos utilizados para la amplificación de los genes seleccionados de la vía del metabolismo del metanol.	181
<b>Tabla 43.</b> Formación de estructuras secundarias de tipo horquilla , homodímero y heterodímero de las secuencias oligonucleotídicas que amplifican los genes involucrados en la vía de secreción.	186

**Tabla 44.** Formación de estructuras secundarias de tipo horquilla , homodímero y heterodímero de las secuencias oligonucleotídicas que amplifican los genes involucrados en la vía del metabolismo del metanol.

187

## ÍNDICE DE FIGURAS

<b>Figura 1.</b> Metabolismo del metanol en levaduras metiltróficas (Hartner & Glieder, 2006).	4
<b>Figura 2.</b> Ruta de asimilación del metanol de <i>P. pastoris</i> (De Shutter <i>et al.</i> , 2009).	7
<b>Figura 3.</b> Genes involucrados en los complejos SEC y OST, importantes para la glicosilación temprana (Delic <i>et al.</i> , 2013).	18
<b>Figura 4.</b> Representación esquemática de la vía de secreción de proteínas en <i>P. pastoris</i> (De Shutter <i>et al.</i> , 2009).	20
<b>Figura 5.</b> Plataformas de Illumina (Illumina, 2011).	30
<b>Figura 6.</b> Purificación del RNA con el Kit Tru-seq de Illumina.	31
<b>Figura 7.</b> Construcción de la biblioteca de DNAC.	32
<b>Figura 8.</b> Secuenciación de cDNA mediante PCR puente (Illumina, 2011).	33
<b>Figura 9.</b> Detección de fluorescencia y captación de la primera base correspondiente a cada cluster (Illumina, 2011).	34
<b>Figura 10.</b> Resumen de las estrategias de cuantificación.	44
<b>Figura 11.</b> Diseño experimental.	50
<b>Figura 12.</b> Parámetros que se modificaron en el programa bioinformático TopHat2 localizado en la plataforma Galaxy.	55
<b>Figura 13.</b> Parámetros que se modificaron en el programa bioinformático Cufflinks localizado en la plataforma Galaxy.	57
<b>Figura 14.</b> Parámetros que se modificaron en el programa bioinformático Cuffcompare localizado en la plataforma Galaxy.	57
<b>Figura 15.</b> Parámetros que se modificaron en el programa bioinformático Cuffdiff localizado en la plataforma Galaxy.	58
<b>Figura 16.</b> Estrategia para la preparación del mix para 10 genes por duplicado de una sola muestra utilizado en los ensayos de qPCR.	68
<b>Figura 17.</b> Estrategia para la generación de la curva estándar para la corrección de la eficiencia en el análisis de expresión relativa de cada gen en el cultivo CM48h.	70
<b>Figura 18.</b> Estrategia para el cálculo de expresión relativa de cada gen.	71
<b>Figura 19.</b> Gel de agarosa al 0.8% de las preparaciones de RNA.	74
<b>Figura 20.</b> Gel de electroforesis digital de las preparaciones de RNA de las muestras obtenido por el Bioanalizador Agilent 2100.	74
<b>Figura 21.</b> Electroferogramas de las preparaciones de RNA.	75
<b>Figura 22.</b> Electroferogramas de las librerías de DNAC obtenidas a partir de las cinco preparaciones de RNA.	76
<b>Figura 23.</b> Ejemplo de los gráficos obtenidos con la herramienta “FASTQ Summary Statistics” de las lecturas correspondientes a la muestra CM47-1 R2 (3’), antes y después del recorte de las secuencias.	77
<b>Figura 24.</b> Secuencia del plásmido que contiene el gen heterólogo de interés insertado " <i>in silico</i> " al final del cromosoma 1.	78
<b>Figura 25.</b> Visualización con el programa IGV de los transcritos mapeados correspondientes a la muestra C0 del cromosoma 4 respecto al genoma de <i>P. pastoris</i> .	79
<b>Figura 26.</b> Datos generados con la herramienta Cuffdiff que corresponden a la cuantificación y normalización de los transcritos durante la prueba de expresión diferencial de CM respecto a CP visualizada en Excel.	80

<b>Figura 27.</b> Expresión global diferencial de genes encontrados con Cuffdiff, considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	81
<b>Figura 28.</b> Expresión global diferencial de genes considerando un $\log_2(\text{Fold\_change}) \leq -1.0$ y $\geq 1.0$ .	81
<b>Figura 29.</b> Análisis global de la expresión diferencial de genes en la comparación CM/C0 considerando un $\log_2(\text{Fold\_change}) \geq 1.0$ y $\leq -1.0$ .	82
<b>Figura 30.</b> Análisis global de la expresión diferencial de genes en la comparación CP/C0 considerando un $\log_2(\text{Fold\_change}) \geq 1.0$ y $\leq -1.0$ .	82
<b>Figura 31.</b> Análisis Global de la expresión diferencial de genes en la comparación CM/CP considerando un $\log_2(\text{Fold\_change}) \geq 1.0$ y $\leq -1.0$ .	83
<b>Figura 32.</b> Términos GO sobre-representados de los genes sobre-expresados en la comparación CM/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	84
<b>Figura 33.</b> Porcentaje de genes asociados a los términos GO sobre-representados de los genes sobre-expresados más importantes en la comparación CM/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	85
<b>Figura 34.</b> Términos GO sobre-representados de los genes sub-expresados en la comparación CM/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	85
<b>Figura 35.</b> Porcentaje de genes asociados a los términos GO sobre-representados de los genes sub-expresados más importantes en la comparación CM/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	86
<b>Figura 36.</b> Términos GO sobre-representados de los genes sobre-expresados en la comparación CP/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	87
<b>Figura 37.</b> Porcentaje de genes asociados a los términos GO sobre-representados de los genes sobre-expresados más importantes en la comparación CP/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	87
<b>Figura 38.</b> Términos GO sobre-representados de los genes sub-expresados en la comparación CP/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	88
<b>Figura 39.</b> Porcentaje de genes asociados a los términos GO sobre-representados de los genes sub-expresados en la comparación CP/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	88
<b>Figura 40.</b> Términos GO sobre-representados de los genes sobre-expresados en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	89
<b>Figura 41.</b> Porcentaje de genes asociados a los términos GO sobre-representados de los genes sobre-expresados en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	90
<b>Figura 42.</b> Términos GO sobre-representados de los genes sub-expresados en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	90
<b>Figura 43.</b> Porcentaje de genes asociados a los términos GO sobre-representados de los genes sub-expresados en la comparación CM/CP utilizando la	

ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	91
<b>Figura 44.</b> Diagrama de Venn que muestra los genes sobre-expresados en las tres comparaciones considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	93
<b>Figura 45.</b> Diagrama de Venn que muestra los genes sub-expresados en las tres comparaciones considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	93
<b>Figura 46.</b> Porcentaje de términos agrupados sobre-representados de los genes sobre-expresados debido al cambio en la fuente de carbono en CM/C0 y CP/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	95
<b>Figura 47.</b> Porcentaje de genes asociados a los Términos GO más importantes sobre-representados de los genes sobre-expresados debido al cambio en la fuente de carbono en CM/C0 y CP/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	95
<b>Figura 48.</b> Porcentaje de términos agrupados sobre-representados de los genes sobre-expresados en la comparación CM/CP que excluye a los genes sobre-expresados tanto en CM como en CP debido al cambio en la fuente de carbono utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	97
<b>Figura 49.</b> Porcentaje de genes asociados a los términos GO más importantes sobre-representados de los genes sobre-expresados en la comparación CM/CP que excluye a los genes sobre-expresados tanto en CM como en CP debido al cambio en la fuente de carbono utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	98
<b>Figura 50.</b> Porcentaje de términos agrupados sobre-representados de los genes sub-expresados debido al cambio en la fuente de carbono en las comparaciones CM/C0 y CP/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	99
<b>Figura 51.</b> Porcentaje de genes asociados a los términos GO mas importantes sobre-representados de los genes sub-expresados debido al cambio en la fuente de carbono en las comparaciones CM/C0 y CP/C0 utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	99
<b>Figura 52.</b> Porcentaje de términos agrupados sobre-representados de los genes sub-expresados que excluye a los genes sub-expresados tanto en CM como en CP debido al cambio en la fuente de carbono en las comparaciones CM/CP utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	100
<b>Figura 53.</b> Porcentaje de genes asociados a los términos GO sobre-representados de los genes sub-expresados que excluye a los genes sub-expresados tanto en CM como en CP debido al cambio en la fuente de carbono en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un $\log_2(\text{Fold\_change}) \leq -0.7$ y $\geq 0.7$ .	101
<b>Figura 54.</b> Expresión diferencial del gen heterólogo <i>FTEII</i> en las tres comparaciones (CM/C0, CP/C0 y CM/CP).	102
<b>Figura 55.</b> Distribución del número de genes sobre-expresados considerando un $\log_2(\text{Fold\_change}) \leq 1.0$ ó $\geq 1.0$ en las rutas metabólicas analizadas.	103
<b>Figura 56.</b> Distribución del número de genes sub-expresados considerando un $\log_2(\text{Fold\_change}) \leq 1.0$ ó $\geq 1.0$ en las rutas metabólicas analizadas.	104

<b>Figura 57.</b> Expresión diferencial considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ de los genes importantes en el inicio del metabolismo del metanol ( <i>AOX2</i> y <i>CAT</i> ) en las tres comparaciones.	105
<b>Figura 58.</b> Genes diferencialmente expresados considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ involucrados en la vía de asimilación del metanol en las tres comparaciones.	105
<b>Figura 59.</b> Expresión diferencial de genes considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ involucrados en la vía de desasimilación del metanol en las tres comparaciones.	106
<b>Figura 60.</b> Vía del metabolismo del metanol para la generación de biomasa y energía.	109
<b>Figura 61.</b> Expresión Global Diferencial de genes de la vía de secreción considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	110
<b>Figura 62.</b> Distribución del número de genes sobre-expresados correspondientes a la vía de secreción considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ en las tres comparaciones.	111
<b>Figura 63.</b> Distribución del número de genes sub-expresados correspondientes a la vía de secreción considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ en las tres comparaciones.	111
<b>Figura 64.</b> Genes diferencialmente expresados considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ involucrados en la translocación y transporte de proteínas.	112
<b>Figura 65.</b> Genes diferencialmente expresados que codifican chaperonas involucradas en el plegamiento de proteínas, con un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	115
<b>Figura 66.</b> Genes diferencialmente expresados involucrados en la síntesis de glicoproteínas en RE, considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	119
<b>Figura 67.</b> Genes diferencialmente expresados involucrados en el procesamiento y tráfico de glicoproteínas en aparato de Golgi, considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	123
<b>Figura 68.</b> Genes diferencialmente expresados que codifican para proteínas implicadas mantenimiento de la homeostasis intracelular y en la degradación o recambio proteolítico en sistemas vacuolares, considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	125
<b>Figura 69.</b> Genes sobre-expresados con un log en la vía de secreción de proteínas.	128
<b>Figura 70.</b> Expresión Global diferencial de otros genes indirectamente relacionados con la secreción considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	129
<b>Figura 71.</b> Distribución del número de genes sobre-expresados correspondientes a otros genes indirectamente relacionados con la secreción considerando un $\log_2$ $\leq -1.0$ ó $\geq 1.0$ en las tres comparaciones.	130
<b>Figura 72.</b> Distribución del número de genes sub-expresados correspondientes a otros genes indirectamente relacionados con la secreción considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ en las tres comparaciones.	130
<b>Figura 73.</b> Genes diferencialmente expresados involucrados en la respuesta a proteínas mal plegadas (UPR), considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	131
<b>Figura 74.</b> Genes diferencialmente expresados involucrados en el complejo proteico del proteosoma mediado por ERAD, considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	134



<b>Figura 75.</b> Genes diferencialmente expresados involucrados en autofagia, considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	136
<b>Figura 76.</b> Genes diferencialmente expresados involucrados en los receptores de ataque a proteínas solubles NSF (SNARE), considerando un $\log_2$ (Fold_change) $\leq -1.0$ ó $\geq 1.0$ .	139
<b>Figura 77.</b> Correlación entre las dos técnicas utilizadas en el análisis transcripcional.	148
<b>Figura 78.</b> Comportamiento de los genes evaluados en las dos técnicas utilizadas RNA-seq vs RT-qPCR.	149
<b>Figura 79.</b> Características de los oligonucleótidos que amplifican los genes seleccionados por su sobre-expresión de la vía de secreción diseñados con Primer3.	182
<b>Figura 80.</b> Características de los oligonucleótidos que amplifican los genes seleccionados por su sub-expresión de la vía de secreción diseñados con Primer3.	183
<b>Figura 81.</b> Características de los oligonucleótidos que amplifican los genes seleccionados por su sobre-expresión de la vía del metabolismo del metanol con Primer3.	184
<b>Figura 82.</b> Alineamientos de los oligonucleótidos 5' y 3' que amplifican los genes seleccionados por su sobre-expresión de la vía de secreción.	185
<b>Figura 83.</b> Alineamientos de los oligonucleótidos 5' y 3' que amplifican los genes seleccionados por su sub-expresión de la vía de secreción.	185
<b>Figura 84.</b> Alineamientos de los oligonucleótidos 5' y 3' que amplifican los genes seleccionados por su sobre-expresión de la vía del metabolismo del metanol.	186
<b>Figura 85.</b> Curvas de disociación de todos los productos amplificados que corresponden a los genes seleccionados.	188
<b>Figura 86.</b> Curvas de amplificación correspondientes a los factores de dilución empleados para el cálculo de la eficiencia, linealidad e intervalos de linealidad correspondientes a los genes seleccionados.	189
<b>Figura 87.</b> Curva estándar de cada uno de los productos de amplificación correspondientes a cada uno de los genes evaluados.	190
<b>Figura 88.</b> Curvas de amplificación de los genes sobre-expresados de la vía de secreción en las condiciones CM, CP y C0.	191
<b>Figura 89.</b> Curvas de amplificación de los genes seleccionados por su sub-expresión en la vía de secreción en las condiciones CM,CP y C0.	192
<b>Figura 90.</b> Curvas de amplificación de los genes seleccionados por su sobre-expresión en la vía del metanol en las condiciones CM, CP y C0.	193
<b>Figura 91.</b> Curvas de amplificación de los normalizadores empleados para el calculo de la expresión relativa en las condiciones CM, CP y C0.	193

## LISTA DE SÍMBOLOS Y ABREVIATURAS

KM71FTEII	Cepa recombinante de <i>P. pastoris</i> productora y secretora de una fitasa termoestable
Mut <sup>s</sup>	Mutante que carece del gen <i>AOXI</i> funcional
DEG	Expresión diferencial de genes
RNA-seq	Secuenciación de RNA
pAOX	Promotor del gen que codifica para la enzima alcohol oxidasa
CM/C0	Comparación de la condición sobre-productora de la proteína heteróloga respecto al cambio de la fuente de carbono
CP/C0	Comparación de la condición de baja producción de la proteína heteróloga respecto al cambio de la fuente de carbono
CM/CP	Comparación de la condición sobre-productora de la proteína heteróloga respecto a la de baja producción
RE	Retículo endoplásmico
UPR	Respuesta a proteínas mal plegadas o no plegadas
ERAD	Vía de degradación de proteínas asociadas al retículo endoplásmico
RT-qPCR	Transcripción reversa cuantitativa de la reacción en cadena de la polimerasa
NGS	Secuenciación de nueva generación
DNAc	Ácido desoxirribonucleico complementario
RNA <sub>m</sub>	Ácido ribonucleico mensajero
RNA <sub>t</sub>	Ácido ribonucleico de transferencia
RNA <sub>r</sub>	Ácido ribonucleico ribosomal
MUT	Utilización de metanol
α-MF	Secuencia señal pre-pro del factor α de <i>S. cerevisiae</i>
GS115	Cepa de <i>P. pastoris</i> GS115
OST	Complejo oligosacariltransferasa
SEC	Complejo de secreción, homólogo al translocón del retículo endoplasmático
BiP	Proteína de unión a proteínas codificada por el gen <i>KAR2</i>
NEF	Factor de intercambio de nucleótidos
s-UPR	Super respuesta a proteínas mal plegadas o no plegadas
t.-SNARE	Factor sensitivo N-etilmaleimida de fijación a receptores de proteína localizada en la membrana
v.-SNARE	Factor sensitivo N-etilmaleimida de fijación a receptores de proteína vesicular
miRNAs	Micro ácidos ribonucleicos
dNTP	Deoxinucleótido
Phred	Programa que calcula el puntaje de calidad de secuencias nucleotídicas para obtener la eficiencia de la secuenciación
contigs	segmentos de DNA superpuestos que en conjunto

	representan una región consenso de DNA
FPKM	número de lecturas mapeadas por kilobase de regiones de exones por millón de lecturas mapeadas en el genoma de referencia
log <sub>2</sub> (Fold_change)	Logaritmo en base 2 utilizado para representar el nivel de expresión
IGV	Visor integrativo de genomas
BAM	Formato binario de salida de datos después de alineamientos de secuencias biológicas
BAI	Formato que permite indexar los datos obtenidos en formato BAM
FASTQ	Formato basado en texto utilizado para la salida de secuencias biológicas, usualmente nucleotídicas, después de la secuenciación
FASTA	Formato basado en texto utilizado para contener secuencias biológicas nucleotídicas y aminoacídicas
RIN	Número de integridad de RNA
GO	Ontología de genes
BOGAS	Sistema de Bioinformática Online de Anotaciones de genomas
MIQE	Mínima información para la Publicación de Experimentos de PCR Cuantitativo
CT	Umbral de fluorescencia
MeOH	Metanol
DEPC	Dietilpirocarbonato
EDTA	Ácido etilendiaminotetraacético
dT	Nucleótidos deoxitimina

## RESUMEN

**Introducción:** La levadura metilotrófica *Pichia pastoris* es de gran importancia industrial principalmente en la producción de proteínas heterólogas. En un estudio reciente se emplearon cinco factores ambientales para definir condiciones de cultivo a nivel de bioreactor que condujeron a altos (CM) y bajos (CP) niveles de la producción extracelular de una fitasa recombinante en una cepa Mut<sup>s</sup> de *P. pastoris*. Los resultados de este estudio mostraron que bajo las condiciones CM, la demanda y consumo de O<sub>2</sub> y de metanol fueron más altos y condujeron a valores más altos en la velocidad específica de crecimiento ( $\mu$ ), biomasa (2.7 veces), niveles de producción de fitasa extracelular (5.5 veces) y rendimientos ( $Y_{p/x}$ ) que en CP. Con el fin de comprender los mecanismos de regulación transcripcional que afectan a la fisiología de *P. pastoris* por la sobre-producción de la proteína recombinante y las condiciones de cultivo, en este trabajo se realizó un análisis de expresión diferencial de genes (DGE) empleando la tecnología de secuenciación masiva de mRNA (RNAseq) de la cepa Mut<sup>s</sup> de *P. pastoris* crecida bajo las condiciones CM y CP reportadas previamente. Además se validaron los resultados del estudio de DGE mediante RT-qPCR. **Resultados:** La expresión de 4,950 genes, el 93% de los genes totales anotados, fueron detectados. Se sub- y sobre-expresaron 350 y 413 genes respectivamente en CM respecto a CP. En CM vs CP se sobre-expresaron significativamente términos relacionados con la biosíntesis de aminoácidos, biosíntesis de nucleósidos de purina, regulación de la traducción, glicosilación de proteínas y mitosis, indicando una mayor actividad anabólica en CM. La transcripción del gen heterólogo y de los genes de la ruta de desasimilación del metanol no mostraron diferencias entre ambas condiciones de cultivo y fue inducida en metanol. Sin embargo las enzimas claves (DAS1 y DAS2) de la ruta de asimilación del metanol se sobre-expresaron significativamente en CM vs CP, indicando que CM está favorecida la producción de biomasa y la generación de energía a través de esta vía, explicando los valores más altos para la  $\mu$  y biomasa obtenidos en CM respecto a CP. De 110 genes analizados involucrados en la vía de secreción, 20 se sobre-expresaron en CM vs CP, la sobre-expresión de estos genes indicaron que bajo las condiciones de CM, se presenta una mayor actividad transcripcional de los genes implicados en el transporte y translocación hacia el RE (15%), genes implicados en el plegamiento de proteínas en RE (25%), así como genes relacionados en el procesamiento de las proteínas a través del RE (30%) y Golgi (35%) que permitieron un estado fisiológico favorable para la secreción de la proteína heteróloga. De los 44 genes relacionados con el estrés en RE durante la secreción, en CM vs CP se sobre-expresaron genes UPR indicando, que bajo condiciones de CM, se promueve la expresión de genes relacionados con el plegamiento de proteínas y probablemente se evita el acumulamiento de proteínas mal plegadas. La sub-expresión de todos los genes relacionados con autofagia, es uno de los factores que podría explicar la menor actividad proteolítica observada en CM. Finalmente se observó una correlación entre los métodos de RNA-seq y RTqPCR ( $r^2=0.7$ ). **Conclusiones:** El análisis de la DGE señala que los factores ambientales en CM condujeron a la regulación de la expresión de genes del proceso de secreción y genes relacionados al estrés en RE durante la secreción que condujeron a valores de  $Y_{p/x}$ , más altos en CM que en CP y no se atribuyen a una expresión diferencial del gen heterólogo. La regulación de la ruta del metanol hacia la asimilación y una mejor respuesta de adaptación al estrés en CM condujeron a un mayor crecimiento y producción de biomasa en CM que en CP.

## ABSTRACT

**Background:** The methylotrophic yeast *Pichia pastoris* has great industrial interest due to production of heterologous proteins. In a recent study, five environmental factors were used to define culture conditions, at bioreactor level, which led to high (CM) and low (CP) levels of extracellular phytase production in a Mut<sup>s</sup> recombinant strain of *P. pastoris*. The results of this study showed that under CM conditions, demand and consumption of O<sub>2</sub> and methanol were higher and led to higher values in the specific growth rate ( $\mu$ ), biomass (2.7 times), production levels of extracellular phytase (5.5 times) and yield ( $Y_{p/x}$ ) than in CP. In order to understand the mechanisms of transcriptional regulation affecting the *P. pastoris* physiology as a consequence of the recombinant protein overproduction and culture conditions, in the present work, a differential gene expression (DGE) analysis, employing mRNA massive sequencing (RNAseq) technology, of *P. pastoris* Mut<sup>s</sup> strain grown under the CM and CP conditions described previously was performed. Also, validation of RNA-seq results was performed with RT-qPCR.

**Results:** The expression of 4,950 genes, 93% of total annotated genes, was identified. CM led to 350 and 413 genes up- and down-regulated, respectively, compared to CP. In CM vs CP, the significant up-regulated terms were related with amino acids and nucleotides biosynthesis, protein glycosylation, regulation of translation and mitosis; indicating a greater anabolic activity in CM vs CP. The transcription of heterologous gene and genes of methanol dissimilative pathway was induced in methanol and no differences between the two culture conditions was showed. However key enzymes (DAS1 and DAS2) of methanol assimilation pathway were significantly up-regulated in CM vs CP, indicating that biomass production and energy generation was favored through this pathway, and so thus explaining the higher values of  $\mu$  and biomass obtained in CM vs CP. From 110 analyzed genes involved in the secretion pathway, 20 were up-regulated in CM vs CP, the up-regulation of this genes indicated that under CM conditions, it was observed a high transcriptional activity of genes related to transport and translocation through the ER (15%), genes related to protein folding in ER (25%), as well as genes related to protein processing in ER (30%) and Golgi apparatus (35%) that allowed a favorable physiological condition for heterologous protein secretion. From the 44 genes related to stress in RE during secretion, UPR genes were over-expressed in CM vs CP, indicating that under CM, the expression of genes related with the correct protein folding was promoted and the accumulation of misfolded proteins was probably also avoided. Down-regulation of whole autophagy-related genes in CM vs CP, is one of the factors that could explain the lower proteolytic activity that was observed in CM. Finally, a correlation between RNA-seq and RT-qPCR methods ( $r^2 = 0.7$ ) was observed.

**Conclusions:** The DGE analysis indicates that environmental factors in CM resulted in the regulation of gene expression involved in secretion and genes related to stress in ER during secretion, leading to higher values of  $Y_{p/x}$  in CM than in CP and was not attributed to a differential expression of heterologous gene. The regulation towards the methanol assimilative pathway and a better adaptation to stress in CM, lead to a higher growth and biomass production in CM vs CP.

# 1. INTRODUCCIÓN

*Pichia pastoris* es una levadura metilotrófica capaz de usar metanol como única fuente de carbono para la obtención de energía y crecimiento celular (Van der Klei *et al.*, 2006). Además el desarrollo de vectores, cepas y métodos para la manipulación genética de *P. pastoris* ha contribuido al desarrollo de sistemas de expresión de proteínas heterólogas que pueden controlar la secreción de la proteína al medio extracelular (Schutter *et al.*, 2009). Este sistema de expresión generalmente emplea al promotor del gen *AOX1*, el cual es regulado por metanol para la expresión de genes heterólogos (Gasser *et al.*, 2007) y debido a un mejor control del bioproceso se han utilizado cepas (KM71, Mut<sup>s</sup>) solo funcionales en el gen *AOX2*, habiendo deletado el gen *AOX1*, capaces de producir altos niveles de proteínas heterólogas (Chiruvolu *et al.*, 1997).

*P. pastoris* ha sido utilizada por más de 1500 laboratorios académicos e industriales alrededor del mundo para la producción de proteínas con aplicaciones médicas y en la investigación; hasta el 2013 se registran más de 20,000 publicaciones acerca del uso de *Pichia pastoris* para la producción de más de 5,000 proteínas heterólogas, con más de 70 productos que se pueden encontrar en el mercado, los cuáles incluyen la insulina humana, vacuna del hepatitis B, anticuerpos monoclonales, etc. (Gehlsen; 2013).

Algunos estudios fisiológicos han demostrado que muchos procesos, incluyendo respuestas de estrés a factores ambientales y el plegamiento de las proteínas, influyen en los rendimientos de la producción de las proteínas heterólogas en *P. pastoris* (Gasser *et al.*, 2007).

El análisis transcripcional de genes que influyen en la síntesis y secreción de proteínas heterólogas es un factor clave para poder entender las respuestas de los organismos hospederos durante estos procesos, así como comprender la influencia de las condiciones de cultivo específicas (Resina *et al.*, 2007).

El perfil de un transcriptoma se realiza tradicionalmente mediante RT-qPCR para analizar unos cuantos genes, o microarreglos (DNAc) para analizar la actividad transcripcional del genoma completo (Fang *et al.*, 2012).

La plataforma de secuenciación masiva de RNAm (RNA-seq), basada en tecnologías de nueva generación (NGS), ha sido usada para mapear y cuantificar perfiles de transcriptomas completos que reflejan ventajas importantes comparadas con otras herramientas de análisis transcripcional como los microarreglos (Liang *et al.*, 2012).

Existen tres importantes plataformas de NGS, de las cuáles Illumina se caracteriza por requerir una mínima cantidad de RNA y producir bibliotecas específicas DNAC altamente sensibles mediante un kit (Tru-seq) que permite seleccionar específicamente la molécula de RNA que tiene un grupo monofosfato en su extremo 5' y un grupo hidroxilo en el 3' (paired-end), disminuyendo el intervalo de error y aumentando la robustez (Chu-Corey, 2012).

Recientemente Viader-Salvadó y colaboradores (2013) optimizaron cinco factores ambientales que permitieron a una cepa Mut<sup>s</sup> de *P. pastoris* (KM71FTEII), incrementar la producción de una fitasa recombinante a nivel de biorreactor. A partir de estos resultados se emplearon los cinco factores ambientales para definir condiciones de cultivo a nivel de bioreactor que condujeron a altos (CM) y bajos (CP) niveles de la producción extracelular de la proteína heteróloga en la cepa KM71FTEII. Los resultados de este estudio mostraron que bajo las condiciones CM, la demanda y consumo de O<sub>2</sub> y de metanol fueron más altos y condujeron a valores más altos en la velocidad específica de crecimiento ( $\mu$ ), biomasa (2.7 veces), niveles de producción de fitasa extracelular (5.5 veces) y rendimientos ( $Y_{p/x}$ ) que en CP. Con el fin de comprender los mecanismos de regulación transcripcional que afectan a la fisiología de *P. pastoris* por la sobreproducción de la proteína recombinante y las condiciones de cultivo, en este trabajo se realizó un análisis de expresión diferencial de genes (DGE) empleando la tecnología de secuenciación masiva de mRNA (RNAseq) de la cepa Mut<sup>s</sup> de *P. pastoris* crecida bajo las condiciones CM y CP reportadas previamente. Además se validaron los resultados del estudio de DGE mediante RT-qPCR.

## 2. ANTECEDENTES

### 2.1 Microorganismos metilotróficos

Los microorganismos metilotróficos procariotas y eucariotas tienen la capacidad de utilizar compuestos C1 reducidos como única fuente de energía (Yurimoto *et al.*, 2011).

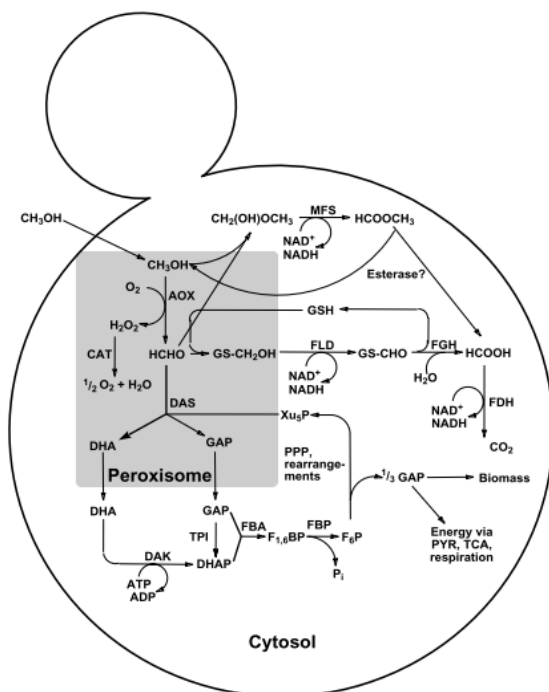
Estos microorganismos utilizan un amplio rango de vías metabólicas para asimilar y desasimilar compuestos C1 reducidos. La forma general de estos microorganismos para asimilar el carbono es mediante la conversión de tres moléculas de C1 en una molécula C3 mediante una ruta cíclica. Entre los compuestos que los microorganismos procariotas pueden asimilar se encuentran el metanol, el metano y la metilamina, mientras que los microorganismos eucariotas únicamente asimilan el metanol (Van der Klei *et al.*, 2006; Yurimoto *et al.*, 2011).

#### 2.1.1 Importancia de las especies de levaduras metilotróficas

Todas las especies metilotróficas pertenecen a los géneros de *Pichia*, *Hansenula*, *Candida* y *Torulopsis*; algunas de las especies más estudiadas son *P. pastoris*, *P. methanolica*, *C. boidinni* y *H. polymorpha* (Van der Klei *et al.*, 2006).

La especies de levaduras metilotróficas en general, utilizan una vía de asimilación y desasimilación de metanol en común, donde enzimas como la alcohol oxidasa, dihidroxiacetona sintasa, catalasa, formaldehído deshidrogenasa, S-formilglutatión hidrolasa, formato deshidrogenasa, alcohol deshidrogenasa y Pmp20 están asociadas con el metabolismo del metanol. Esta vía genera dos compuestos muy reactivos, el formaldehído y el peróxido de hidrógeno (Yurimoto *et al.*, 2011). En la figura 1 se representa el metabolismo del metanol de levaduras metilotróficas.





**Figura 1.** Metabolismo del metanol en levaduras metiltróficas (Hartner & Glieder, 2006).

La enzima que inicia la ruta es la alcohol oxidasa (AOX) que cataliza la conversión del metanol a formaldehído formando peróxido de hidrógeno como subproducto, el cual es convertido en oxígeno y agua por la acción de la catalasa previniendo la difusión de estas especies reactivas hacia el citoplasma, ya que ambas reacciones son compartimentalizadas en los peroxisomas. Además existe una tercera enzima del metabolismo del metanol localizada en los peroxisomas, la dihidroxiacetona sintasa (DHAS), la cual es un componente del ciclo de la xilulosa-5-fosfato, y cataliza el primer paso de la asimilación de formaldehído con la consecuente formación de las dos moléculas C3 (dihidroxiacetona y gliceraldehído-3-fosfato) provenientes de una molécula C1 (formaldehído) y una C5 (xilulosa-5-fosfato). Estas tres enzimas localizadas en los peroxisomas son esenciales para permitir a las levaduras crecer en metanol (Van der Klei *et al.*, 2006).

A continuación se muestran los genes que codifican a las enzimas clave en los peroxisomas en la tabla 1.

**Tabla 1.** Genes que codifican para las enzimas clave localizadas en los peroxisomas de algunas especies de levadura que comparten alguna variante conservada (Van der Klei *et al.*, 2006).

<b>Especie</b>	<b>Gen</b>	<b>Enzima</b>	<b>Residuo C-terminal</b>
<i>C. boidinni</i>	<i>AOD1</i>	AOX	-ARY
<i>P. methanolica</i>	<i>MOD1</i>	AOX	-ARF
<i>P. methanolica</i>	<i>MOD2</i>	AOX	-GRF
<i>H. polymorpha</i>	<i>AOX</i>	AOX	-AKL
<i>P. pastoris</i>	<i>AOX1</i>	AOX	-ARF
<i>P. pastoris</i>	<i>AOX2</i>	AOX	-ARF
<i>C. boidinni</i>	<i>CTA1</i>	CAT	-NKF
<i>H. polymorpha</i>	<i>CAT</i>	CAT	-SKI
<i>C. boidinni</i>	<i>DAS1</i>	DHAS	-NHL
<i>H. polymorpha</i>	<i>DAS</i>	DHAS	-NKL

La proliferación masiva de los peroxisomas ha vuelto a las levaduras muy atractivas en diversas aplicaciones industriales, especialmente como hospederas en la producción de proteínas heterólogas, además poseen promotores inducibles, por ejemplo para la expresión de los genes que codifican para AOX y DHAS, que regulan la expresión de los genes responsables del metabolismo del metanol; donde el promotor del gen de la alcohol oxidasa (*pAOX*) es el más fuerte conocido en la naturaleza del genoma de estos microorganismos (Van der Klei *et al.*, 2006).

Además existen otros promotores utilizados en los sistemas de expresión de proteínas heterólogas mostrados en la tabla 2 que son muy utilizados en los sistemas de expresión de proteínas heterólogas (Ahmad *et al.*, 2014).

Los sistemas de expresión de genes heterólogos están guiados por un gran número de promotores fuertemente inducibles con el metanol que han sido desarrollados en cepas de levaduras metilotróficas que incluyen a *P. pastoris*, *H. polymorpha*, *P. methanolica* y *C. boidinni*. El aumento del uso académico e industrial de estos sistemas de expresión han llevado a la producción de un gran número de proteínas heterólogas incluyendo enzimas, anticuerpos, citocinas, proteínas plasmáticas y hormonas. Algunas de las ventajas de estos sistemas incluyen:

1. Medio de cultivo económico basado en sales para el crecimiento de las levaduras.
2. Regulación fuerte y prolongada de promotores inducidos por el metanol y reprimidos por glucosa o etanol.

3. El hecho de que el proceso de plegamiento de proteínas, secreción y otras funciones en estas levaduras son similares en muchos procesos de eucariotas superiores (Yurimoto *et al.*, 2011).

**Tabla 2.** Los promotores más usados y más recientemente establecidos para la expresión de proteínas heterólogas (Ahmad *et al.*, 2014).

<b>Promotores Inducibles</b>	<b>Gen correspondiente</b>	<b>Regulación</b>	<b>Referencia</b>
<i>AOX1</i>	Alcohol oxidasa 1	Inducible con MeOH	Tschopp <i>et al.</i> 1987a
<i>DAS</i>	Dihidroxiacetona sintasa	Inducible con MeOH	Ellis <i>et al.</i> 1985; Tschopp <i>et al.</i> 1987a)
<i>FLD1</i>	Formaldehído dehidrogenasa 1	Inducible con MeOH o metilamina	Shen <i>et al.</i> 1998
<i>ICL1</i>	Isocitrato liasa	Reprimido con glucosa, inducido en ausencia de glucosa/ por adición de metanol	Menendez <i>et al.</i> 2003
<i>PHO89</i>	Na <sup>+</sup> / simportador de fosfato	Inducción a través del ayuno de fosfato	Ahn <i>et al.</i> 2009
<i>THI1</i>	Gen de la biosíntesis de la tiamina	Reprimido por tiamina	Stadlmayr <i>et al.</i> 2010
<i>ADH1</i>	Alcohol deshidrogenasa	Reprimido en glucosa y metanol, inducido en glicerol o etanol	Stadlmayr <i>et al.</i> 2010
<i>ENO1</i>	Enolasa	Reprimido en glucosa, metanol y etanol, inducido en glicerol	Stadlmayr <i>et al.</i> 2010
<i>GUT1</i>	Glicerol Kinasa	Reprimido en metanol, inducido en glucosa, glicerol y etanol	Stadlmayr <i>et al.</i> 2010
<b>Promotores Constitutivos</b>	<b>Gen correspondiente</b>	<b>Regulación</b>	<b>Referencia</b>
<i>GAP</i>	Gliceraldehído 3-P deshidrogenasa	Expresión constitutiva en glucosa, en menor medida en glicerol y metanol	Waterham <i>et al.</i> 1997
<i>TEF1</i>	Factor de elongación de la traducción 1	Expresión constitutiva en glicerol y glucosa	Ahn <i>et al.</i> 2007
<i>PGK1</i>	3-fosfoglicerato kinasa	Expresión constitutiva en glucosa, en menor medida en glicerol y metanol	De Almeida <i>et al.</i> 2005
<i>GCW14</i>	Potencial Glicosil Fosfatidil Inocitol (GPI)- proteína de anclaje	Expresión constitutiva en glicerol, glucosa y metanol	Liang <i>et al.</i> 2013b
<i>G1</i>	Transportador de alta afinidad de glucosa	Reprimido en glicerol, inducido sobre limitación de glucosa	Prielhofer <i>et al.</i> 2013)
<i>G6</i>	Aldehído deshidrogenasa	Reprimido en glicerol, inducido sobre limitación de glucosa	Prielhofer <i>et al.</i> 2013

### 2.1.1.1 Vía de utilización del metanol (vía MUT) en *P. pastoris*

AOX cataliza la primera reacción en la vía MUT, oxidando el metanol a formaldehído mientras se reduce el  $O_2$  a  $H_2O_2$  por acción de la catalasa (CAT); ambas enzimas son secuestradas y utilizadas en los peroxisomas para evitar la toxicidad ocasionada por el peróxido de hidrógeno. En la figura 2 se representa el metabolismo del metanol (Krainer *et al.*, 2012).

AOX es un homo-octámero donde cada subunidad contienen un FAD unido de manera no covalente. En *P. pastoris* se encuentran dos genes que codifican para AOX (*AOX1* y *AOX2*), siendo el gen *AOX1* el responsable de la mayor actividad de la alcohol oxidasa en la célula. La expresión del gen *AOX1* es regulada e inducida mediante metanol (Cregg *et al.*, 1989). Los promotores inducibles comúnmente usados en la vía MUT, pAOX1 y pFDH llevan a la producción de las enzimas necesarias para la asimilación del metanol (De Shutter *et al.*, 2009).

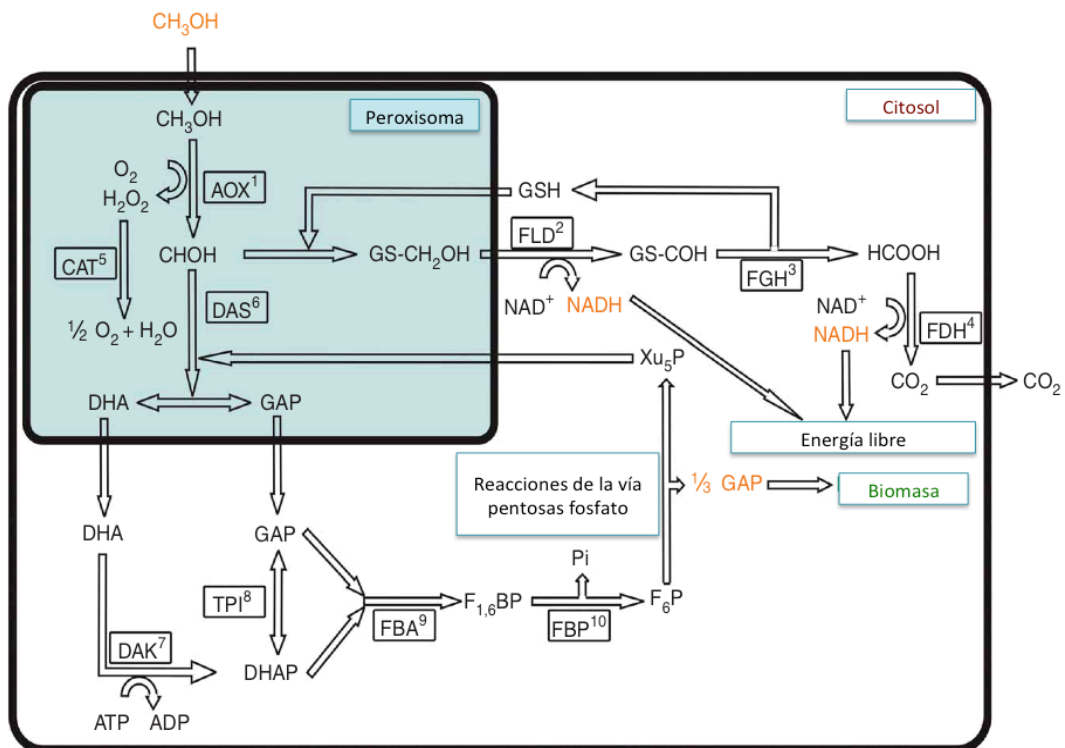


Figura 2. Ruta de asimilación del metanol de *P. pastoris* (De Shutter *et al.*, 2009).

Continuando con ésta vía, el formaldehído generado por AOX se oxida a CO<sub>2</sub> en la vía de desasimilación del metanol, donde aporta 2 moléculas NADH por molécula de formaldehído, o es condensado con xilulosa-5-fosfato y subsecuentemente convertido a dihidroxiacetona y gliceraldehído-3-fosfato en la vía de asimilación del metanol (Krainer *et al.*, 2012).

La enzima clave en la vía de desasimilación es la FLD (formaldehído deshidrogenasa), que cataliza la oxidación dependiente de NAD<sup>+</sup> de S-hidroxiacetilglutathion a S-formilglutathion. Otras enzimas importantes en ésta vía son la enzima FDH (formato deshidrogenasa dependiente de NAD<sup>+</sup>) y FGH (S-formilglutathion hidrolasa). Además la vía de desasimilación del metanol juega un papel importante en el metabolismo de fuentes de nitrógeno metiladas como metilamina, donde el formaldehído es un subproducto. La detoxificación de formaldehído es la principal tarea de la vía de desasimilación (Hartner and Glieder, 2006).

La enzima clave en la vía de asimilación es la dihidroxiacetona sintasa (isoformas DAS1 y DAS2), que cataliza la conversión de la xilulosa 5-fosfato y formaldehído en dihidroxiacetona y gliceraldehído-3-fosfato. Posteriormente la DHA es fosforilada a través de la enzima dihidroxiacetona quinasa (DAK), resultando en la dihidroxiacetona fosfato. A continuación, ocurre la condensación del DHAP con GAP formando una molécula de fructosa 1,6 bifosfato (FBP), esta condensación es realizada a través de la enzima fructosa 1,6 bifosfato aldolasa (FBA2). Posteriormente, a través de la catálisis realizada por la enzima fructosa 1,6 bifosfatasa (FBP) ocurre la formación de una molécula de fructosa 6-fosfato. Dos moléculas de F6P y una molécula de DHAP sufren un reordenamiento para formar tres moléculas de XU5P a través de la catálisis realizada por las enzimas transaldolasa (TAL); transquetolasa (TKL1), que cataliza la reacción entre la xilulosa-5-fosfato y la ribosa-5-fosfato para formar gliceraldehído-3-fosfato y sedoheptulosa-7-fosfato y viceversa; y la triosa fosfato isomerasa (TPI), el resultado de este ciclo es la formación de un mol de triosa fosfato para cada tres moles de formaldehído y un gasto de tres moles de ATP. La triosa fosfato formada a partir de metanol puede ser utilizada tanto anabólicamente (síntesis de componentes celulares) como catabólicamente (obtención de energía vía ciclo de los TCA) (Krainer *et al.*, 2012).

La secuenciación del genoma ha logrado identificar a todos los genes que codifican a las enzimas involucradas en la asimilación del metanol en *P. pastoris* (Tabla 3, De Shutter *et al.*, 2009).

**Tabla 3.** Genes de *P. pastoris* involucrados en la ruta de asimilación del metanol (De Shutter *et al.*, 2009).

<b>Gen</b>	<b>Código EC</b>	<b>Id Locus</b>
<i>AOX</i>	1.1.3.13	Chr4_0152 Chr4_0821
<i>FLD</i>	1.2.1.1	Chr3_1028
<i>FGH</i>	3.1.2.12	Chr3_0867
<i>FDH</i>	1.2.1.2	Chr3_0932
<i>CAT</i>	1.11.1.6	Chr2-2_0131
<i>DAS</i>	2.2.1.3	Chr3_0834 Chr3_0832
<i>DAK</i>	2.7.1.29	Chr3_0841
<i>TPI</i>	5.3.1.1	Chr3_0951
<i>FBA</i>	4.1.21.13	Chr1-1_0319 Chr1-1_0072
<i>FBP</i>	3.1.3.11	Chr3_0868

### **2.1.2 *P. pastoris* como hospedero para la expresión de proteínas heterólogas y genotipo de la cepa KM71**

Todas las cepas de *P. pastoris* son derivados de la cepa de tipo silvestre Y-11430 de Northern Regional Research Laboratories, (NRRL; Peoria, IL). Un listado de cepas disponibles para la expresión en *P. pastoris* se muestra en la tabla 4. Las cepas de tipo salvaje, las cepa X-33 y Y-11430, y las cepa GS115 (*HIS4*) y KM71 transformadas con vectores conteniendo el gen *HIS4* de *P. pastoris* no requieren de algún suplemento para el crecimiento en medios mínimos.

Muchas cepas de expresión se derivan de la cepa auxotrófica a histidina (*HIS4*) GS115 para permitir la transformación con vectores basados en *HIS4*. Muchas otras cepas están disponibles con diferentes combinaciones con una o más mutaciones auxotróficas en los genes biosintéticos *ARG4*, *URA3*, *ADE1*, *HIS4* y *URA5* de *P. pastoris*. Estas cepas permiten la selección de vectores de expresión que contiene el gen marcador de selección correspondiente a la transformación. Aunque la integración de más de un cassette de expresión requiere transformaciones secuenciales, la disponibilidad de marcadores múltiples permite la expresión de varios genes heterólogos relacionados (Nett & Gerngross, 2003).

La mayoría de las cepas de *P. pastoris* crecen en metanol a la misma velocidad que las de tipo silvestre (Mut<sup>+</sup>, fenotipo que metaboliza metanol).

**Tabla 4.** Cepas disponibles para expresar proteínas heterólogas en *P. pastoris* (Nett & Gerngross, 2003).

Cepa	Genotipo
Y-11430	Tipo salvaje
X-33	Tipo salvaje
<b>Cepas auxotróficas</b>	
GS115	<i>HIS4</i>
GS190	<i>ARG4</i>
JC220	<i>ADE1</i>
JC254	<i>URA3</i>
GS200	<i>ARG4 HIS4</i>
JC227	<i>ADE1 ARG4</i>
JC304	<i>ADE1 HIS4</i>
JC305	<i>ADE1 URA3</i>
JC306	<i>ARG4 URA3</i>
JC307	<i>HIS4 URA3</i>
JC300	<i>ADE1 ARG4 HIS4</i>
JC301	<i>ADE1 HIS4 URA3</i>
JC302	<i>ADE1 ARG4 URA3</i>
JC303	<i>ARG4 HIS4 URA3</i>
JC308	<i>ADE1 ARG4 URA3</i>
YJN165	<i>URA5</i>
KM71	<i>HIS4 ARG4 AOX1Δ::SCARG4</i>
MC100-3	<i>HIS4 ARG4 AOX1Δ::SCARG4 AOX2Δ::PPHIS4</i>
<b>Cepas deficientes de proteasa</b>	
SMD1168	<i>HIS4 ΔPEP4::URA3 URA3</i>
SMD1165	<i>PRB1 HIS4</i>
SMD1163	<i>PEP4 PRB1 HIS4</i>
SMD1168 <i>kex1::SUC2</i>	<i>ΔPEP4:: URA3 ΔKEX1::SUC2 HIS4 URA3</i>
<b>Otras cepas</b>	
GS241	<i>FLD1</i>
MS105	<i>HIS4 FLD1</i>

Sin embargo, las cepas con mutaciones en el gen *AOX* son a veces capaces de producir altos niveles de proteínas heterólogas, por lo que se tienen cepas con deleciones en uno o ambos genes *AOX*. Un beneficio adicional de usar este tipo de cepas es que no se requieren grandes cantidades de metanol en escala de fermentación como las Mut<sup>+</sup> convencionales. La cepa KM71 (*HIS4 ARG4 AOX1Δ::SCARG4*) tiene una deleción parcial del gen *AOX1* que ha sido reemplazado por el gen *ARG4* de *S. cerevisiae* (Cregg

& Madden, 1987). Esta cepa (Mut<sup>s</sup>, fenotipo de metabolismo lento del metanol) crece de manera lenta en metanol porque depende de la expresión del gen *AOX2*, el cual es menos eficiente para metabolizar el metanol. Otra cepa, MC100-3 (*HIS4 ARG4 AOX1 Δ::SCARG4 AOX2 Δ::PPHIS4*) posee deleciones en ambos genes *AOX1* y *AOX2*, por lo cual es totalmente incapaz de crecer en metanol (Mut<sup>r</sup>, fenotipo que no metaboliza el metanol). Estas cepas mutantes Mut<sup>s</sup> y Mut<sup>r</sup>, aunque comprometidas en la capacidad para metabolizar el metanol, mantienen la habilidad para inducir la expresión de altos niveles del promotor *AOX1* (Chiruvolu *et al.*, 1997).

A pesar de que en *P. pastoris* existen dos genes que codifican para la alcohol oxidasa, *AOX1* y *AOX2*, *AOX1* es responsable de la mayoría de la actividad celular de la enzima. La expresión del gen *AOX1* está finamente regulada e inducida fuertemente por metanol. El aislamiento de este último (y su promotor) y el desarrollo de vectores, cepas y métodos para la manipulación genética de *P. pastoris*, ha ayudado al desarrollo del sistema de expresión de proteínas heterólogas de *P. pastoris*. Además, diversos factores han contribuido a su aceptación rápida como sistema de expresión, los más importantes incluyen:

1. Un promotor fuerte derivado del gen que codifica para la alcohol oxidasa I (*AOX1*) específico para *P. pastoris* y que controla la expresión de los genes heterólogos.
2. La similitud de las técnicas necesarias para la manipulación genética de *P. pastoris* con las empleadas en *S. cerevisiae*, uno de las de los sistemas experimentales mejor caracterizados en la biología moderna.
3. Su preferencia por el crecimiento respiratorio, ya que facilita su cultivo a altas densidades celulares relativas.
4. La decisión en 1993 de la compañía Phillips Petroleum Company de lanzar el sistema de expresión de *P. pastoris* para la investigación académica, lo cual amplió los conocimientos del sistema (Gasser *et al.*, 2007).

### **2.1.2.1 *Pichia pastoris*, productora de proteínas recombinantes**

*P. pastoris* es considerada como organismo modelo para determinar las rutas metabólicas implicadas en la asimilación de metanol, involucradas en la biogénesis de peroxisomas (Cereghino & Cregg, 2000; Daly & Hearn, 2005; Macauley & Patrick *et*



*al.*, 2005), además de ser, por mucho, el hospedero más comúnmente utilizado para la producción de proteínas recombinantes, ya que, durante la última década se ha incrementado el uso de *P. pastoris* en el proceso de producción de proteínas heterólogas, y se utiliza en muchos laboratorios alrededor del mundo para investigaciones y aplicaciones médicas (De Schutter *et al.*, 2009).

Por otro lado, *P. pastoris* es conocida como un sistema de expresión altamente eficiente ya que se puede controlar la secreción de la proteína al medio de cultivo con facilidad (De Schutter *et al.*, 2009). En los últimos años se ha demostrado mediante diversos estudios fisiológicos, que muchos procesos incluyendo respuestas de estrés hacia factores ambientales (como pH, temperatura, presencia de oxígeno, fuente de carbono, etc) así como el plegamiento y secreción de proteínas están altamente relacionados (Gasser *et al.*, 2007).

#### **2.1.2.2 Factores que influyen en la producción y secreción de proteínas heterólogas**

Desde el punto de vista industrial, el efecto deseado de un pH bajo es el de reducir la actividad de proteasas nativas que pueden llevar a la degradación de la proteína heteróloga, pero no se ha mostrado una correlación entre el pH y la actividad de las proteasas en el medio de cultivo (Gasser *et al.*, 2008). Los valores de pH sugeridos para la experimentación con *P. pastoris* son entre 3.0 y 7.0, ya que no afectan el crecimiento de la levadura significativamente, lo que confiere una libertad considerable en ajustar el pH a uno que no sea óptimo para una proteasa problema (Macauley & Patrick *et al.*, 2005); sin embargo se ha concluido que el pH óptimo depende significativamente de la proteína y la cepa utilizada (Li *et al.*, 2007).

Las proteínas son susceptibles a plegamientos inadecuados por una variedad de razones como son la formación de enlaces disulfuro intermoleculares e interacciones hidrofóbicas durante el plegamiento del péptido a elevadas temperaturas (Macauley & Patrick *et al.*, 2005).

Cuando *P. pastoris* se cultiva a temperaturas de 15°C, se reducen los niveles de proteasas y aumentan los niveles de producción, presentándose además un mejor plegamiento y reducción de la toxicidad (Li *et al.*, 2007, Sreekrishna, 2010). Lo mismo

se observa a temperaturas de 12°C (Jahic *et al.*, 2003). Sin embargo, aunque a temperaturas bajas se ha mostrado que se mejoran las tasas de secreción, esto depende de la naturaleza de la proteína heteróloga. La temperatura además tiene un efecto importante sobre el metabolismo celular, y en la abundancia y regulación de genes y proteínas relacionados con el plegamiento (Gasser *et al.*, 2008).

La cantidad de oxígeno disuelto es uno de los factores más importantes para el crecimiento de *P. pastoris* y la expresión de las proteínas heterólogas durante los procesos de cultivos en biorreactor, debido a que los procesos de óxido-reducción influyen tanto en la oxidación del producto para formar puentes disulfuro, como en el estrés oxidativo de la célula durante su cultivo. El cultivo con metanol de *P. pastoris* puede provocar un estrés oxidativo significativo, el cual puede ser contrarrestado por la adición de antioxidantes como el ácido ascórbico. Además a una temperatura de crecimiento baja disminuye el estrés oxidativo y se inducen genes codificantes para enzimas claves de la homeostasis de los procesos de óxido-reducción y osmoregulación. Esto puede deberse a que los compartimentos participantes en las vías secretoras mantienen un estado de oxidación elevado en comparación al citosol para permitir una mejor formación de los enlaces disulfuro (Gasser *et al.*, 2008).

El oxígeno disuelto puede controlarse mediante la velocidad de agitación y la velocidad de adición de aire o de O<sub>2</sub>. Para la mayoría de los cultivos en biorreactor de *P. pastoris*, el oxígeno disuelto se debe mantener constante a 30-35%, sin embargo, diferentes proteínas presentan diferentes niveles óptimos de oxigenación los cuales sólo puede determinarse mediante experimentos preliminares (Li *et al.*, 2007).

#### **2.1.2.2.1 Bioproceso optimizado para la sobreproducción de una proteína heteróloga en *P. pastoris***

Recientemente Viader *et al.*, 2013, utilizaron una estrategia de diseño de experimentos (método de optimación Simplex) para optimizar cinco factores ambientales que definieron las condiciones de cultivo en la etapa de inducción y que permitieron a la cepa KM71FTEII, portadora de una secuencia codificante para una fosfohidrolasa (fitasa) y regulada por el promotor AOX1, incrementar la producción a

nivel de biorreactor. Con el proceso de optimización se logro obtener hasta 6.4 g/L de fitasa en el medio de cultivo después de 68h de inducción con metanol.

A partir de éstos resultados se seleccionaron las condiciones de cultivo para bioprocesos que condujeran a altos (CM) y bajos (CP) niveles de producción de la fitasa recombinante y se evaluó, mediante RT-qPCR, la expresión de genes implicados en la secreción, proteólisis y el metabolismo del metanol, además del gen heterólogo (Félix-Castro; 2013).

La expresión de genes fue evaluada para el gen *FTEII*, gen heterólogo de la cepa KM71FTEII de *P. pastoris*, para tres genes de proteasas vacuolares (*PEP4*, *PRB1* y *PRC1*), un gen responsable del metabolismo del metanol (*AOX2*), el gen *KEX2*, proteasa que participa en el procesamiento postraduccional para la secreción de la proteína recombinante, y el gen *KAR2* que codifica para una chaperona del retículo endoplásmico. Además, se emplearon los genes *GAP*, *G6PD*, y *YPT*, como genes normalizadores y reportados como constitutivos en *P. pastoris*

Los resultados de estos estudios indicaron que bajo las condiciones de CM, que condujeron a niveles altos de producción extracelular de fitasa (5.2 veces), se presentó un aumento en la demanda de oxígeno y el consumo de metanol (15 veces), lo que aumentó 4.3 veces la velocidad específica de crecimiento y la densidad celular (1.4 veces), y además ocurrió una disminución de la actividad extracelular de proteasas. Por otro lado, las condiciones de cultivo optimizadas aumentaron la secreción de la proteína recombinante hasta en un 88%, junto con la eficiencia de la producción extracelular de fitasa por célula (rendimiento 2.4 veces).

Los resultados el análisis de la expresión de genes mostraron que las condiciones de cultivo no afectan significativamente la expresión diferencial del gen heterólogo (*FTEII*) bajo las condiciones de los cultivos CM y CP; sin embargo bajo las condiciones de CM se indujo la expresión de los genes *KEX2*, *PEP4*, *PRB1*, *PRC1* y *KAR2* los cuales podrían haber influido a un mejor en el plegamiento y secreción de la fitasa en CM que en CP, aunque con un costo de consumo de energía (Félix-Castro; 2013).

Los autores concluyen que la alta producción extracelular de fitasa en CM fue debida a una mayor utilización del metanol para la generación de energía empleada en la inducción y el crecimiento celular.

### 2.1.3 Vía de secreción de proteínas en *P. pastoris*

Hasta la fecha se conoce muy poco acerca de la fisiología de *P. pastoris* como productora de proteínas heterólogas, en especial, la producción de proteínas complejas ha resultado tener una tasa de éxito bastante baja. Sin embargo, muchos estudios fisiológicos han demostrado que ciertos procesos, como la respuesta a factores ambientales (principalmente pH, osmolaridad, oxígeno disuelto y temperatura) influyen considerablemente en la secreción y plegamiento de proteínas (Gasser *et al.*, 2007).

Con respecto a la secreción, hasta el momento se conocen más de cien genes implicados en ésta vía. (Vanz *et al.*, 2014; De Schutter *et al.*, 2009).

La secreción de proteínas es un proceso esencial para todos los organismos tanto procariotas como eucariotas. *Pichia pastoris* utiliza la vía universal de los organismos eucariotas, es decir, el paso a través del retículo endoplásmico (ER) y Golgi en donde se realizan diversas modificaciones post-traduccionales (Delic *et al.*, 2013; Vanz *et al.*, 2014).

En general *Pichia pastoris* es una herramienta poderosa para la producción individual de numerosas proteínas recombinantes secretadas en cantidades substanciales para su aplicación y su caracterización (Salamin *et al.*, 2010), debido a que una de las ventajas de utilizar esta levadura como hospedero para la producción de proteínas heterólogas que se puedan secretar al medio extracelular, ya que posee un número limitado de proteínas endógenas secretorias y que además, se encuentran disponibles una variedad de cepas deficientes de proteasas. Sin embargo se requiere asegurar la estabilidad de cualquier transformante generado utilizando cassettes de expresión que puedan integrarse dentro del genoma (Cregg *et al.*, 2000).

Además, para que la proteína secretada sea funcional en *P. pastoris* es importante el uso de una secuencia señal que ayude a la proteína naciente a translocarse al RE para que se inicie la vía de secreción universal de los eucariotes; se han reportado un amplio rango de secuencias señal (Tabla 5) incluyendo secuencias nativas de proteínas de levaduras como *Saccharomyces cerevisiae* ( $\alpha$ -MF, prepro secuencia señal del factor  $\alpha$  de apareamiento) y *P. pastoris* (PHO1, secuencia señal de la fosfatasa ácida y SUC2 secuencia señal de la invertasa) (Cregg *et al.*, 2000).

**Tabla 5.** Secuencias señal usadas para secretar la proteína al espacio extracelular (Ahmad *et al.*, 2014).

Secuencia señal	Fuente	Proteína Blanco	Longitud	Referencia
$\alpha$ -MF	<i>S.c.</i> $\alpha$ -factor de apareamiento	Señal de secreción más comúnmente usada en <i>P. pastoris</i>	85 aa, c/s repeticiones EA	Brake <i>et al.</i> , 1984
PHO1	<i>P.p.</i> fosfatasa ácida	Pepsinógeno 5-HT5A de ratón y porcino	15 aa	Payne <i>et al.</i> , 1995; Weiss <i>et al.</i> , 1995; Yoshimasu <i>et al.</i> , 2002
SUC2	<i>S.c.</i> invertasa	Interferón humano, $\alpha$ -amilasa, $\alpha$ -1-antitripsina.	19 aa	Moir & Dumais 1987; Paifer <i>et al.</i> , 1994; Tschopp <i>et al.</i> , 1987b
PHA-E	Fitohemaglutinina	GNA, GFP y proteína nativa	21 aa	Raemaekers <i>et al.</i> , 1999
KILM1	K1 toxina	CM celulosa	44 aa	Skipper <i>et al.</i> , 1985
pGKL	pGKL proteína killer	$\alpha$ -amilasa de ratón	20 aa	Kato <i>et al.</i> , 2001
CLY y CLY-L8	C-lizozima y sin. Péptido rico en leucina	Lizozima humana	18 y 16 aa	Oka <i>et al.</i> , 1999
K28 pre-pro-toxina	K28 toxina viral	Proteína verde fluorescente	36 aa	Eiden-Plach <i>et al.</i> , 2004
Scw, Dse y Exg	<i>P.p.</i> Péptidos señal endógenos	CALB y EGFP	19,20 y 23 aa	Liang <i>et al.</i> , 2013a
Pp Pir1	<i>P.p.</i> Pir1p	EGFP y $\alpha$ 1-antitripsina	61 aa	Khasa <i>et al.</i> , 2011
HBFI y HBFI	Hidrofobinas de <i>Trichoderma reesei</i>	EGFP	16 y 15 aa	Kottmeier <i>et al.</i> , 2011

De manera general, la secreción en levaduras como *P. pastoris* inicia con la translocación de la proteína a través de la membrana del ER con ayuda del complejo SEC61, como ya se había mencionado, se requiere de un péptido señal (Factor  $\alpha$ , secuencia señal utilizada con frecuencia en la producción de proteínas heterólogas) para que dicha translocación ocurra (Vanz *et al.*, 2014; De Schutter *et al.*, 2009).

En *S. cerevisiae* existen dos complejos que funcionan como poros de translocación y sus homólogos se han determinado en *P. pastoris*. Por otro lado, se ha demostrado que ambos complejos interactúan con los ribosomas (mas específico, la subunidad larga de RNAr 28s) a través de los dominios citosólicos de SEC61 y SSH1 respectivamente. Los genes *SEC61* como *SSH1* poseen una secuencia altamente conservada y ambas subunidades son esenciales para la translocación de proteínas y supervivencia de la célula (Delic *et al.*, 2013).

La mayoría de las proteínas secretadas son glicosiladas; como primer paso de la N-glicosilación el oligosacárido precursor  $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$  es ensamblado en el

acarreador de lípidos (dolicil pirofosfato) en los lados lumbinales de la membrana del RE y citosol. Este proceso es mediado por enzimas codificadas por los genes de la familia ARG, altamente revisados por Burda y Aebi (1999).

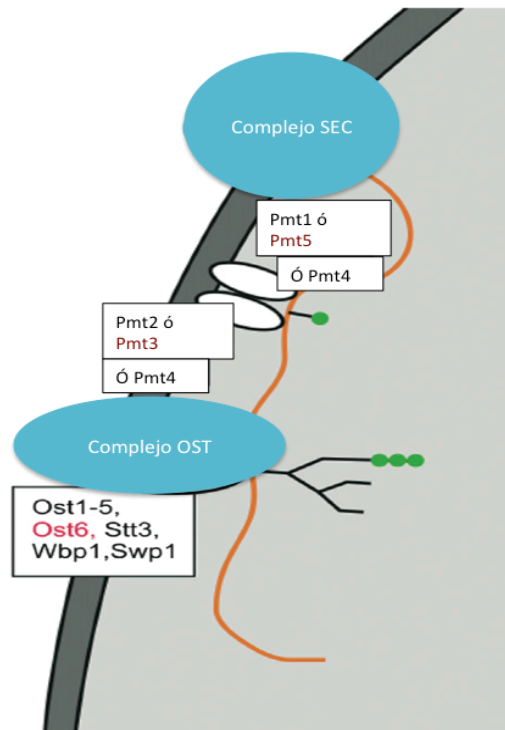
El oligosacárido precursor es transferido por el complejo OST hacia el residuo de asparagina de la secuencia de reconocimiento de la N- glicosilación (Asn-Xxx-Ser/Thr) de la proteína naciente durante la translocación (Vanz *et al.*, 2014); para que después actúe el complejo señal peptidasa para el procesamiento del péptido señal y entonces las proteínas puedan ser plegadas con ayuda de las chaperonas (Vanz *et al.*, 2014; De Schutter *et al.*, 2009).

Los roles de las proteínas codificadas por *SEC63* y *KAR2* en la co- y pos-translocación ha sido un punto de discusión. *SEC63* mantiene tres diferentes roles; el primero, se une a *SEC62* a través de su región ácida C-terminal y establece el complejo SEC heptamérico exclusivo para el transporte post-traducciona; el segundo, se asume que el dominio J localizado en el RE coopera con *KAR2* para regular la apertura del poro de secreción; el tercero, es el rol en el ensamblaje de los complejos SEC heptamérico y SEC' hexamérico, el cuál se atribuye al dominio Brl; además recientemente se observó que está involucrado en el transporte co-traducciona vía el poro con ayuda de la proteína codificada por *SSH1* (Gasser *et al.*, 2008).

El gen *KAR2* codifica para una chaperona del retículo endoplásmico (BiP), perteneciente a la familia de Hsp70 de las proteínas de choque térmico, presente en el lumen del retículo endoplásmico de todos los eucariotes. La unión a BiP previene el desplegamiento de la proteína naciente, ya sea secretoria o transmembranal, hasta que termina la síntesis de la misma. Se ha sugerido que BiP no sólo está involucrada en la translocación de la proteína naciente a través de la membrana del RE al lumen, sino que es un elemento clave por un control de calidad del RE que previene que las proteínas no plegadas salgan del RE. Otras funciones asociadas a BiP son: la solubilización de precursores del plegamiento, estabilización de las subunidades proteicas y la redirección de polipéptidos no plegados al citosol para su marcaje con ubiquitina y posterior degradación (Gasser *et al.*, 2008). *KAR2* se encuentran sobre-expresadas en el proceso de secreción y se requiere igual en ambos procesos de la co y pos-translocación. Además, el mecanismo que empuja el polipéptido naciente al RE involucra no solo a

KAR2 sino también sus co-haperonas LHS1 y SIL1, donde LHS1 es probablemente el factor de intercambio de nucleótidos (NEF) durante el proceso de translocación (Delic *et al.*, 2013).

Después ocurre el procesamiento del N-glicano así como O-glicosilaciones en ER, la proteína se transporta al aparato de Golgi donde tienen un procesamiento mas exhaustivo del N-glicano (Vanz *et al.*, 2014; De Schutter *et al.*, 2009). Las O-glicosilaciones, comienzan por la transferencia de un residuo de manosa del dolícil pirofosfato a un residuo de serina/treonina por la proteína O-manosil transferasa (familia PMT) en el RE. Algunos de los genes involucrados en la glicosilación temprana se muestran en la figura 3.



**Figura 3.** Genes involucrados en los complejos SEC y OST, importantes para la glicosilación temprana (Delic *et al.*, 2013).

La N-glicosilación es muy importante ya que se emplea como control de calidad del plegamiento de proteínas en RE. Aparentemente, las N- y O- glicosilaciones les confiere múltiples funciones a las proteínas secretoras, lo cuál incrementa la solubilidad de la

proteína e impacta en la estabilidad de la pared celular, osmotolerancia y gemación (Delic *et al.*, 2013).

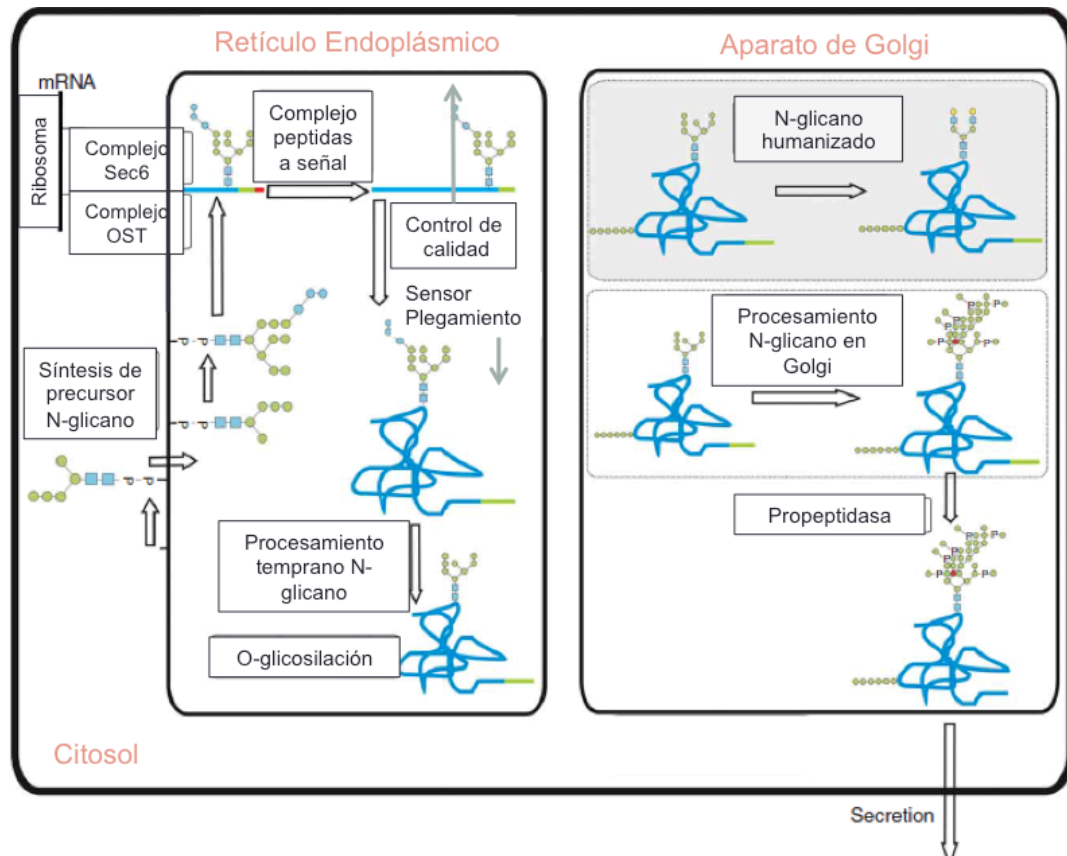
Además existen avances en la glicoingeniería de cepas de *P. pastoris* con una N-glicosilación completamente humanizada (Vervecken *et al.*, 2004; Hamilton *et al.*, 2006; Jacobs *et al.*, 2009), esto se logro al inactivar *OCH1*, un gen que codifica  $\alpha$ -1,6-manosiltransferasa, responsable de la hiper-manosilación en levaduras; segundo, se sobre-expresó una  $\alpha$ -1,2-manosidasa retenida en ER y dos glicosiltransferasas quiméricas, Kre2-GnaTI y Kre2- $\beta$ -1,4-galactosiltransferasa (Hamilton *et al.*, 2006).

Finalmente se inicia el transporte a través de las vesículas hacia el ambiente extracelular, donde pueden ser sustrato de proteasas. Este transporte requiere de un complejo multimérico compuesto de 8 subunidades (Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70 y Exo84) con el cuál atraviesan la membrana plasmática a través de un exoporo necesario para polarizar el crecimiento de la membrana (Delic *et al.*, 2013). En la figura 4 se representa el proceso de secreción.

Por otra parte, se ha detectado que la sobre-expresión de la proteína heteróloga puede implicar un mal plegamiento de la proteína, conformaciones no óptimas y/o una sobre carga en la maquinaria del hospedero. De hecho la acumulación de proteínas mal plegadas en el ER podría explicar los bajos niveles de producción de la proteína heteróloga en *P. pastoris* y además es lo que ocurre con mayor frecuencia en los procesos de producción (Zhang *et al.*, 2006).

Algunos estudios han demostrado que la inducción de la producción de proteínas heterólogas provoca dos principales tipos de respuesta al estrés. La primera es la respuesta al estrés oxidativo desencadenada por las especies reactivas de oxígeno generadas al hacer el intercambio de la fuente de carbono de glicerol a metanol al activar el promotor AOX1; la segunda es la respuesta a las proteínas mal plegadas (UPR), la vía de degradación de proteínas asociada al retículo endoplásmico (ERAD) y la inducción de proteasas vacuolares y autofagia (Hesketh *et al.*, 2013).





**Figura 4.** Representación esquemática de la vía de secreción de proteínas en *P. pastoris* (De Shutter *et al.*, 2009).

### 2.1.3.1 Respuesta a proteínas mal plegadas (UPR)

El ER es un organelo sub-celular donde muchas proteínas se sintetizan y se transfieren hacia diversos destinos sub-celulares así como el espacio extracelular. En éste organelo se encuentra un ambiente óptimo único para el plegamiento de proteínas, ensamblaje y formación de puentes disulfuro. Cuando las células eucariotas recaen en estrés dentro del lumen del ER se activa un proceso llamado Respuesta a las Proteínas mal plegadas (UPR) en el cuál se activan una serie de genes que ayudan a reducir el nivel de estrés (Bernales *et al.* 2006). Esta repuesta tiene 3 objetivos: inicialmente restablecer las funciones normales de la célula al detener la traducción de proteínas, degradar proteínas mal plegadas y activar las vías de señalización que llevan al incremento de la producción de chaperonas involucradas en el plegamiento de proteínas. Si esos objetivos no se logran durante cierto lapso de tiempo o se prolongan, esta

respuesta conduce hacia la disfunción celular y muerte por apoptosis (Osowski & Urano, 2011).

El UPR utiliza tres sensores principales: La proteína kinasa de ER parecida a PKR/Factor 2 de inicio de la traducción de eucariotes (eIF2 $\alpha$ ), es decir, (PERK/eIF2 $\alpha$ ); el Factor 6 de activación de la transcripción (ATF6) y la enzima que requiere inositol 1 (IRE1). Estos tres se localizan en la membrana del ER, y en sus respectivos estados inactivos, estos sensores se asocian con BiP a través de múltiples dominios. Cuando hay un incremento en las proteínas no plegadas o mal plegadas en el lumen del ER, BiP se une a los agregados de esas proteínas y se disocia de los transductores de UPR (Wu & Kaufman, 2006). Esto permite la formación de dímeros (ejemplo: IRE1+IRE1 y PERK+PERK), entonces inicia una serie de cascadas de fosforilación que conllevan a la inducción de genes relacionados con UPR (Harding *et al.*, 2000).

Además BiP también se libera de IRE1, permitiendo su dimerización y autofosforilación. Esto activa un sitio específico de endoribonucleasa (RNasa) localizado en el dominio citosólico. A través de ésta actividad RNasa, se remueve un intron del RNAm de HAC1, el cuál se traduce en un potente factor de transcripción (HAC1p) (Yoshida *et al.*, 2000). Se ha visto que cuando aumenta excesivamente la concentración del RNAm HAC1 significa que hay condiciones extremas de estrés en RE llamado super-UPR (S-UPR), este proceso es necesario para la supervivencia celular y no es necesaria la participación de IRE1. (Bernaes *et al.* 2006).

ATF6 también se activa a través de otra vía de señalización, donde se mueve hacia el compartimiento en cis de Golgi y es escindido por proteasas específicas; este dominio citosólico migra al núcleo y actúa como factor de transcripción junto con HAC1p (Yoshida *et al.*, 2000); los cuáles incrementan la expresión de genes involucrados en UPR como BiP, calreticulina, calnexina, PDI y otras; a su vez, también estimulan la maquinaria del plegamiento de proteínas y aceleran ERAD para remover las proteínas mal plegadas (Wu & Kaufman, 2006).

Por otro lado, debido a la necesidad de obtener nuevas herramientas eficientes para la alta producción de proteínas heterólogas, se han encontrado, mediante diversos estudios, algunos genes que se pueden co-expresar como factores de soporte de la secreción:

1. *PDI*, codifica una enzima del lumen del RE que cataliza el mecanismo e la formación del puente disulfuro en proteínas secretoras y de superficie de la célula que requieren el plegamiento de polipéptidos nacientes.
2. *KAR2* importante en la respuesta a estrés, sobre-expresada de forma única o en conjunto con *PDI*.
3. *SSO1* y *SSO2* genes que codifican homólogos de syntaxina en levaduras necesarios para la fusión de veículas secretoras de la membrana plasmática actuando como t-SNAREs (Factor sensitivo N-etilmaleimida de ataque a receptores de proteína).
4. *ERO1*, esencial para la formación de puente disulfuro, que es oxidado por oxígeno molecular y actúa como oxidante específico de *PDI*, su sobre-expresión con *PDI* incrementan el buen plegamiento de proteínas heterólogas.
5. *IRE1*, *PTC2* y *HAC1*, que codifican proteínas implicadas en la regulación de UPR (Bollock *et al.*, 2009).

#### **2.1.3.2 Vía de degradación de proteínas asociadas al Retículo endoplásmico (ERAD)**

Se cree que la reacción de estrés durante la secreción causada por una proteína no plegada o mal plegada en RE mediada por la activación de UPR, re-establece la homeostasis celular, por ejemplo, al inducir los genes involucrados en el plegamiento de proteínas y en la vía de degradación asociada al ER (ERAD). Durante ERAD las proteínas secretoras mal plegadas son retrotranslocadas al citoplasma del ER, poli-ubiquitinadas, y devueltas al proteosoma para su degradación (Delic *et al.*, 2013). Sin embargo se tiene que atravesar la membrana por lo que es difícil la degradación de proteínas por esta vía (Bernales *et al.* 2006).

Las células eucariotas contienen dos vías principales involucradas con la degradación de proteínas, y organelos; una es a través del proteosoma y la otra es través de la vacuola/lisosoma o autofagia (Yorimitsu & Klionsky, 2005). El proteosoma es un complejo compartimentado de proteasas a donde son dirigidas las proteínas de vida media corta y proteínas mal plegadas del citosol o del núcleo, después de haber sido marcadas con ubiquitina (Huang & Klionsky, 2002). El proteosoma puede ser regulado mediante

distintas vías, según las necesidades celulares, se puede inducir la expresión de todos los genes o solamente la inducción de algunos de los genes que codifican para proteínas del proteosoma; además, para que una proteína sea reconocida y procesada en el proteosoma es necesario que se encuentre marcada con ubiquitina, El proceso de ubiquitinación de una proteína requiere la participación de tres tipos de complejos: la enzima activadora de ubiquitina E1, la enzima conjugadora de ubiquitina E2 y la ubiquitina ligasa E3. Debido a la gran cantidad de procesos en los que participan el proteosoma y el sistema de marcaje por ubiquitinación, la repercusión de sus alteraciones puede tener un gran alcance (Shenoy, 2007).

### **2.1.3.3 Autofagia**

La autofagia es principalmente responsable de la degradación de proteína de vida media larga y de organelos completos (Yorimitsu & Klionsky, 2005). Tres principales pasos operativos caracterizan el proceso de autofagia: 1. El secuestro del material en una vesícula recién formada; 2. Fusión de la vesícula con organelos lisosomales; 3. La degradación del material y el reciclaje de los sustratos. Existen tres formas diferentes descritas, macroautofagia, microautofagia y autofagia mediada por chaperonas (Peracchio *et al.*, 2012). En la autofagia en general, se involucra un re-arreglo dinámico de membranas que termina en una vacuola o lisosoma (Yorimitsu & Klionsky, 2005). En la microautofagia el material citoplasmático y organelos son directamente internalizados por el lisosoma a través de la invaginación de la membrana lisosomal. La macroautofagia (o simplemente autofagia) los agregados macromoleculares, porciones de citoplasma, membranas y organelos completos son secuestrados en nuevas vesículas formadas (autofagosomas) que subsecuentemente se unen a los lisosomas (Peracchio *et al.*, 2012); la formación del autofagosoma ocurre en la estructura pre-autofagosomal (PAS), el autofagosoma terminado se fusiona con la vacuola y el contenido es liberado al lumen, degradado y las macromoléculas generadas son recicladas (Yorimitsu & Klionsky, 2005). En el caso de autofagia mediada por chaperonas, las proteínas citoplasmáticas que llevan la secuencia consenso KFERQ en el C-terminal, son asistidos

para ingresar al lisosoma con ayuda de la chaperona HSC70, que interactúa con la membrana de la proteína lisosomal LAMP2A (Peracchio *et al.*, 2012).

La vía CvT, involucra la formación de vesículas conteniendo el complejo del precursor de aminopeptidas I, las vesículas CvT se fusionan a la vacuola, vertiendo su contenido en lumen (Huang & Klionsky, 2002). Una vía alternativa de proteínas blanco para la degradación es la autofagia, la cuál se describe como una colección de vías en las secciones del citoplasma incluyendo sus organelos, donde las proteínas pueden ser secuestradas en los compartimentos, fusionarse con vacuolas (o lisosomas) y entrar en contacto con hidrolasas ácidas y otras sustancias de degradación (Bernaes *et al.* 2006). Durante éste proceso, los constituyentes son aislados del resto de la célula mediante una vesícula de doble membrana conocida como autofagosoma, la cuál se fusiona con un lisosoma degradando y reciclando el cargo; existen tres formas diferentes descritas, macroautofagia, microautofagia y autofagia mediada por chaperonas.

Además los eventos de fusión de vesículas durante la autofagia es mediado por SNARE's (acrónimo de "SNAP (Soluble NSF Attachment Protein) Receptor"), la fusión dependiente de SNARE es requerida para la biogénesis del autofagosoma y su precursor, el fagóforo (Moreau *et. al.*, 2014). Sin embargo, está bien descrito que SNARE's participa en procesos de fusión de membranas durante el tráfico de proteínas del RE al aparato de Golgi (Delic *et al.*, 2013).

#### **2.1.3.4 Factor sensitivo N-etilmaleimida de fijación a receptores de proteína (SNAREs)**

Las proteínas SNAREs facilitan la fusión vesículas transportadoras con la membrana celular o a vacuolas o lisosomas. Esta familia de pequeñas proteínas son caracterizadas por la presencia de un motif (60-70 aa) organizado en repeticiones heptaméricas. Estas proteínas contienen un solo dominio terminal transmembranal (C-terminal) y múltiples dominios N-terminal que se encuentran plegados independientemente a partir del motif SNARE. Las proteínas SNARE han sido divididas en dos categorías, SNARE de vesícula ó v-SNARE, las cuales están incorporadas en la membrana de la vesícula transportadora durante la gemación y blanco de SNARE ó t-SNARE ("target-SNARE"), incorporadas en las membranas de los compartimentos

celulares blanco, como las vacuolas ó lisosomas. Los complejos SNARE (“trans”) están compuestos de syntaxina 1 y SNAP-25, residentes en las membranas celulares y sinaptobrevina (también referida como proteína asociada a membrana de vesícula o VAMP) anclada a la membrana vesicular (Delic *et al.*, 2013).

#### **2.1.4 Estudio de genes que afectan la producción de proteínas heterólogas en *P. pastoris***

Se han realizado diversos estudios transcripcionales para clarificar el comportamiento de los genes bajo condiciones de estrés específicas. Centrándose en genes sobre-expresados con funciones potenciales en la maquinaria de secreción y/o regulación de estrés, se han detectado 13 genes potenciales, incluyendo *PDII* y *HAC1*, que han mostrado tener efectos beneficiosos en la secreción de anticuerpos utilizando a *P. pastoris* como sistema de expresión (Gasser *et al.*, 2006).

Se ha reportado también que la duplicación de los genes *ERO1* y *PDII* llevan a un incremento significativo en la producción de albúmina sérica humana en *Kluyveromyces lactis* (Lodi *et al.*, 2005).

En otro trabajo, durante la secreción de fragmentos Fab 2F5 usando a *P. pastoris*, se ha observado que la sobre-expresión de genes que codifican a factores que ayudan a la secreción bien caracterizados como *PDII*, *HAC1*, *ERO1* y *KAR2* tienen un efecto positivo en la secreción de proteínas heterólogas. Además la co-sobre-expresión de nuevos genes que codifican a factores que ayudan a la secreción como *SSA4*, *BMH2*, *KIN2*, *SSE1*, y *BFR2* han mostrado causar un incremento elevado altamente significativo en la cantidad de Fab 2F5 secretado; así mismo la co-sobre-expresión de *COG6*, *COY1* Y *SSO2*, han mostrado causar un incremento moderado significativo en la cantidad de anticuerpos heterólogos secretados (Gasser *et al.* 2007).

También se ha visto una sobreexpresión de *KAR2* y *PDI* (genes implicados en el desplegado de proteínas) durante una alta producción de lipasa (Resina *et al.*, 2007).

En otro estudio a diferentes osmolaridades en cultivos con fuente de carbono limitada de *P. pastoris* se identificaron bien 37 proteínas afectadas entre ellas *INO1*, *KAR2*, *RIB3*, *SSA4*, *SSC1*, *SSB1*, etc. (Dragosits *et al.*; 2010).

Además en otro estudio, trabajaron con clonas de *P. pastoris* que co-sobre-expresaran genes de interés y observaron que los mutantes transformantes del gen *WSC4* y los del gen *ERO1* producía una sobre-expresión de la proteína heteróloga (Baumann *et al.*; 2011).

## **2. 2 Estudio de Transcriptomas**

La Transcriptómica es la clave para entender cómo se traduce la información codificada en el genoma y cómo influye en las funciones celulares. Así mismo, ayuda a entender cómo es que dicho proceso de traducción responde a diferentes condiciones dentro de un mismo entorno o el conjunto de éstos (Fang *et al.*, 2012).

Un transcriptoma es una colección de transcritos incluyendo RNAs codificantes y no codificantes en una célula y su cantidad bajo un estado o condición fisiológica específica (Wang *et al.*, 2009; Kogenaru *et al.*, 2012), por lo que es importante cuantificar con precisión la abundancia de cada transcrito dentro de diferentes tejidos y puntos del tiempo, y correlacionar cambios en abundancia de la perturbación genética y ambiental para entender la función del genoma y como éste se adapta al medio ambiente (Fang *et al.*, 2012; Wang *et al.*, 2009).

### **2.2.1 Técnicas utilizadas para obtener perfiles de expresión génica**

Los perfiles de transcriptomas o perfiles de expresión génica permiten determinar la abundancia de estado estacionario de cada transcrito dentro un transcriptoma. El perfil de transcriptoma se realiza tradicionalmente mediante RT-qPCR para analizar unos cuantos genes, o microarreglos (DNAc) para analizar la actividad transcripcional del genoma completo (Fang *et al.*, 2012).

Inicialmente los microarreglos fueron esenciales en el análisis de perfiles transcripcionales o transcriptomas completos (Kogenaru *et al.*, 2012). Recientemente debido a las limitaciones de los microarreglos (Wang *et al.*, 2009; Kogenaru *et al.*, 2012) y al bajo costo de las tecnologías de secuenciación de nueva generación (NGS) o secuenciación profunda debido a su alta cobertura, el perfil de transcriptomas por la técnica de RNA-Seq (secuenciación de todos los transcritos) es el método de elección,

debido a su alta sensibilidad y precisión (Fang *et al.*, 2012).

### **2.2.1.1 Técnica de RNA-Seq**

El estudio a partir de RNA-seq no significa que la secuenciación sea directamente de las moléculas de RNA, el paso de la secuenciación es generalmente la misma que en la secuenciación de DNA, pero el análisis y la preparación de la biblioteca son bastante diferentes, puesto que se involucra generalmente la retro-transcripción (Chu & Corey, 2012).

Aunque los resultados finales son similares a otras tecnologías, los desafíos de generar DNAC y realizar pruebas de expresión diferencial de genes (DGE) mediante análisis bioinformáticos son fundamentalmente distintos a los de medir la intensidad relativa utilizando microarreglos de expresión diferencial (Bullard *et al.*, 2010).

En la obtención de datos mediante RNA-Seq, se mide el número de secuencias de transcrito discretas y no intensidades de fluorescencia continua, disminuyendo el grado de error debido a entrecruzamiento de datos (Bullard *et al.*, 2010). Estos datos se obtienen en un formato FASTQ, basado en formato de texto para almacenar secuencias biológicas (usualmente nucleotídicas) y su puntaje (score) de calidad correspondiente. Este formato es utilizado por las tecnologías NGS, el cuál contiene múltiples secuencias nucleotídicas cortas, obtenidas durante la secuenciación correspondientes a la abundancia de transcritos bajo cierto estado fisiológico y que pueden ser cuantificables mediante múltiples herramientas bioinformáticas.

Además, frecuentemente se requieren diferentes pruebas estadísticas para determinar la expresión diferencial y otros tipos de patrones.

Para los usuarios de RNA-Seq, hay un sinfín de aplicaciones de software bioinformático que simplifican enormemente el proceso de análisis de datos. Debido a que el análisis mediante RNA-Seq no requiere de sondas específicas de especies y transcritos, los datos no están sesgados por anteriores suposiciones sobre la naturaleza del transcriptoma (Bullard *et al.*, 2010).

Hoy en día el análisis de DGE mediante RNA-Seq puede revelar la identidad de la mayoría de los RNA's dentro de una célula, proporcionando decenas o cientos de



millones de lecturas de secuencias e información sobre miles de millones de bases individuales (Chu & Corey, 2012).

Además, a diferencia de las tecnologías anteriores, las NGS permiten ensamblar y analizar transcritos directamente de los datos obtenidos mediante RNA-Seq. Proporciona una oportunidad para hacer diseños experimentales libre de hipótesis, eliminando la necesidad de genomas o transcriptomas de referencia, lo cual es particularmente atractivo para aquellos organismos no-modelo con pobres anotaciones de genes o faltantes de investigar. (Bullard *et al.*, 2010; Fang *et al.*, 2012).

Por otra parte, el método de RNA-Seq tiene otras ventajas claras por encima de las demás tecnologías y se espera que revolucione los métodos utilizados para analizar los perfiles de transcriptoma o transcriptomas completos de eucariotas. Algunas de estas ventajas son que además de analizar la DGE, permiten el análisis de expresión alelo-específica y la identificación de modificaciones post-transcripcionales (como ediciones o variantes de splicing) que conlleva al descubrimiento de nuevos transcritos o fusión de genes (Nookaew *et al.*, 2008; Bullard *et al.*, 2010; Oszolak & Milos, 2011).

Actualmente el análisis transcripcional mediante RNA-Seq se ha utilizado en organismos modelo como *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, *Arabidopsis thaliana*, así como células de *Mus musculus* y *Homo sapiens* (Nookaew *et al.*, 2008); además se ha utilizado en el caso de organismos recombinantes durante la producción de proteínas heterólogas en condiciones de crecimiento específicas y/o diversas fuentes de carbono como *Pichia pastoris* (Liang *et al.*, 2012; Hesketh *et al.*, 2013; Qi *et al.*, 2011), *Saccharomyces cerevisiae* (Feng & Zhao, 2013; Baumann *et al.*, 2011), entre otras especies de levaduras.

#### **2.2.1.1.1 Plataformas utilizadas para RNA-Seq**

Para empezar un experimento de RNA-Seq, primero se debe elegir una plataforma de secuenciación. Los protocolos implementados para la obtención de datos y la comparación entre las distintas plataformas de RNA-Seq varían, y esta variación puede afectar a cómo se interpretan los experimentos, por lo que elegir la plataforma adecuada

para una aplicación específica es un prerrequisito para lograr el éxito experimental (Chu & Corey, 2012).

Existen varias plataformas NGS que están comercialmente disponibles, la mayoría está basada en la tecnología de secuenciación por síntesis, con una DNA polimerasa o ligasa como el componente clave. Las plataformas 454 de Roche, Illumina, Helicos y PacBio (Pacific Biosciences) utilizan una DNA polimerasa para obtener su reacción de secuenciación, mientras SOLID (Life Technologies) y Complete Genomics utilizan una DNA ligasa. Este tipo de secuenciaciones también pueden categorizarse en base a si se secuencia una sola molécula (como Helicos y PacBio) o si se secuencian un conjunto de moléculas (Illumina y SOLID) (Oszolak & Milos, 2011; Chu & Corey, 2012).

La plataforma de Solexa (Illumina) es altamente prometedora debido a que utiliza un método llamado “PCR puente” en la que los fragmentos (generados *de novo*) se amplifican a partir de los cebadores (random primers) unidos a una superficie sólida permitiendo una mayor cobertura, alta fidelidad, sensibilidad y robustez gracias a los protocolos que ofrece en la preparación de bibliotecas de DNAC (Bullard *et al.*, 2010).

La preparación de dichas bibliotecas se caracteriza por requerir una mínima cantidad de RNA mediante un kit (tecnología Tru-seq simple preparation kits); el cuál provee una solución simple y efectiva para generar bibliotecas de RNAm a partir de RNA total (Chu & Corey, 2012).

#### **2.2.1.1.2 Preparación de bibliotecas de DNAC y secuenciación**

Los métodos más actuales de RNA-Seq dependen de la síntesis de DNAC y una serie de pasos posteriores de manipulación, que impone ciertas limitaciones para algunas aplicaciones o tiene cierta susceptibilidad de errores. Para evitar esto, existen protocolos que utilizan la ligación RNA-RNA para construir bibliotecas de DNAC específico de cadena, pero son laboriosos de construir (Oszolak & Milos, 2011).

La preparación de la biblioteca de DNAC para RNA-Seq implica convertir el RNA celular en moléculas que pueden ser secuenciadas. Algunos RNA's abundantes, como el RNA ribosómico (RNAr) que pueden constar de hasta un 80% del RNA celular total, puede desperdiciar recursos y reducir la profundidad de la cobertura de la secuencia,

dando como resultado una menor detección de moléculas de RNA débilmente expresadas (Chu & Corey, 2012). El RNAr se puede eliminar mediante degradación enzimática (como tratamiento duplex específico de Nucleasa) (Yi *et al.*, 2012; Chu & Corey, 2012) o métodos de agotamiento basados en hibridación (Chen & Duan, 2011; Chu & Corey, 2012).

La preparación de la biblioteca variará dependiendo del tamaño que se desea. El RNA puede ser fragmentado, generalmente por hidrólisis química o digestión enzimática, a un tamaño adecuado para la plataforma de secuenciación elegida. En el caso de moléculas de RNA pequeñas como miRNAs (< 200 bases), la fragmentación no es necesaria. Una vez que se obtienen RNA's del tamaño adecuado para la mayoría de las plataformas, se convierten en DNA complementario (DNAc) con una transcriptasa reversa usando cebadores hexámeros al azar (random primers). Posteriormente se utilizan adaptadores de oligonucleótidos que se unen para permitir la amplificación y activar la secuencia del DNAc (Chu & Corey, 2012).

Por ejemplo, el kit de preparación TruSeq de Illumina tiene la capacidad de producir bibliotecas específicas de cadena altamente sensibles. El tamaño del inserto de las bibliotecas pueden ser de tres tipos, 1)  $\pm 130$  pb 2)  $\pm 280$  pb y 3)  $\pm 480$  pb para la generación del cluster de secuenciación.

El kit selecciona específicamente la molécula de RNA que tiene un grupo monofosfato en su extremo 5' y un grupo hidroxilo en el 3' disminuyendo el rango de error y aumentando la robustez. El kit de preparación de muestra Truseq RNA es una solución altamente efectiva en la producción de bibliotecas del RNA total que es compatible con la plataforma de Illumina (2000 HiSeq<sup>TM</sup>, HiSeq 1000, HiScanTMSQ, Genome Analyzer *IIx*) (Figura 5). (Chu & Corey, 2012).



**Figura 5.** Plataformas de Illumina (Illumina, 2011).

Los reactivos de master-mix eliminan la mayoría de pasos de pipeteo y de purificación comparado con otros métodos, minimizando el tiempo de mano de obra. A partir de RNA total, el RNAm se purifica primero usando una selección de la cadena de polyA, se utiliza un método químico para fragmentar las cadenas y se convierte en DNAc de cadena sencilla usando un cebador hexámero al azar. Entonces se sintetiza la cadena complementaria para crear DNAc de doble cadena que está listo para la construcción de bibliotecas TruSeq (Figura 6).

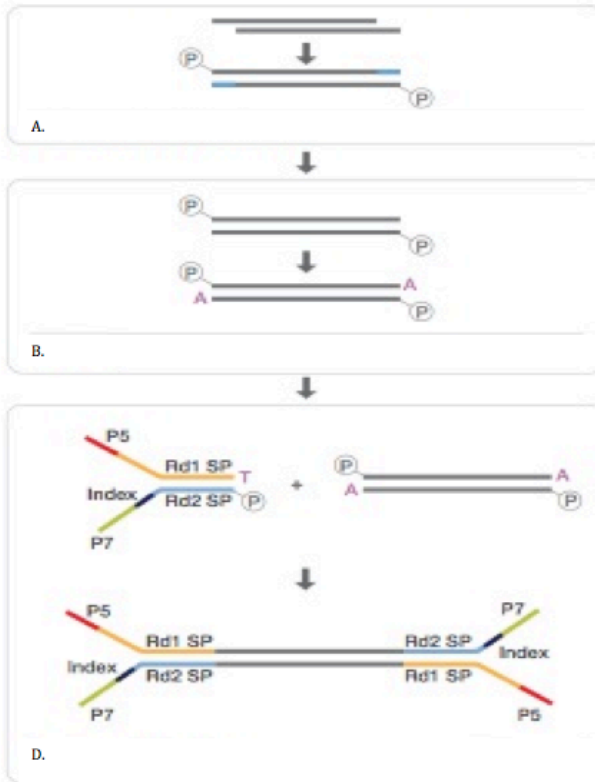


**Figura 6.** Purificación del RNA con el Kit Tru-seq de Illumina.  
A. Selección de Poly-A, fragmentación y Random priming. B. Síntesis de la Primera y segunda cadena

La construcción de la biblioteca comienza con cualquier DNA de doble cadena sintetizado, los fragmentos de DNA con terminación romo se generan utilizando una combinación de reacciones de llenado y actividad de exonucleasa. Una base de adenina se añade a los extremos de cada filamento con terminación romo, preparándolos para la ligadura de los adaptadores de secuenciación. Cada adaptador contiene una base de timina saliente en el extremo 3' proporcionando una proyección complementaria para ligar el adaptador a la cola de poliA del fragmento de DNA (Figura 7).

Estos adaptadores recientemente rediseñados contienen el complemento del sitio de hibridación del cebador para la secuenciación de las terminaciones individuales (“single read”), terminación pareada (“paired-end”) y lecturas múltiples (“multiplexed read”), dependiendo el caso. Esto elimina pasos adicionales de PCR para agregar índices de tag y sitios múltiples de cebadores. Para la generación del cluster se agregan los

primers y las bibliotecas de DNAC en baja densidad para evitar el acumulamiento de varias muestras entre sí.



**Figura 7.** Construcción de la biblioteca de DNAC.

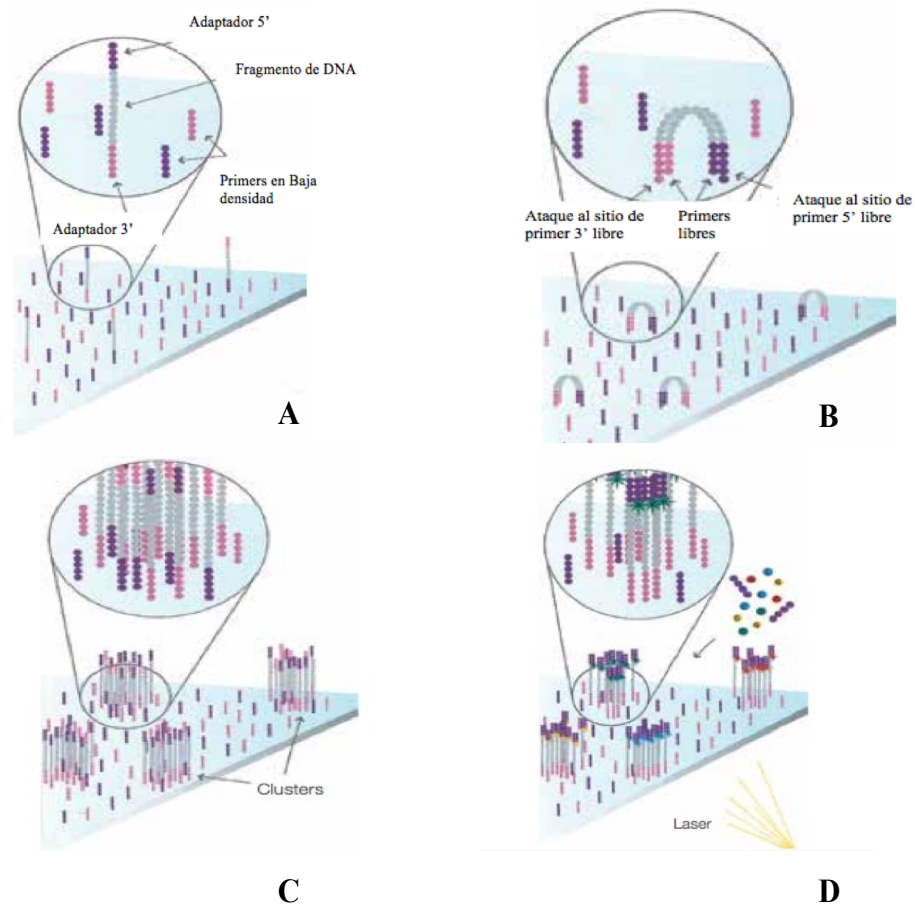
A. Reparación de los extremos y fosforilación. B. Adición de la Adenina (Tailing) en la terminación romo. C. Adición del adaptador indexado que contiene una Timina en el extremo, el cual, se une por complementariedad al extremo que contiene la adenina.

Se recomienda secuenciar con terminación pareada “Paired end” para asegurar una mayor cobertura y precisión de una secuencia a la hora de analizar los datos; esto se logra sintetizando una cadena a partir del extremo 5’ y otra a partir del 3’ de cada fragmento (Figura 8A) de DNA mediante PCR puente (Figura 8B) formando pequeños cluster de DNA’s de ese mismo fragmento (Figura 8C);esto facilita los alineamientos con un genoma de referencia y el ensamblaje, además ayuda a detectar rearrreglos como deleciones, inserciones (ó indels) e inversiones.

Siguiendo los pasos de desnaturalización y amplificación, las bibliotecas pueden combinarse hasta con 12 muestras por línea (96 muestra por flujo celular) para la

generación de clúster o cBot para la secuenciación.

La tecnología de secuenciación por síntesis utiliza nucleótidos marcados fluorescentemente, el equipo es capaz de sintetizar millones de lecturas en paralelo. Durante cada ciclo de secuenciación, un solo deoxinucleótido (dNTP) (Figura 8D) es añadido a la cadena de ácido nucleico y sirve para terminar la polimerización; después de cada incorporación de dNTP, se emite fluorescencia para identificar la base correspondiente en cada cluster (Figura 9) y luego se rompe enzimáticamente para la siguiente incorporación. El equipo quita los adaptadores por default después de la corrida (Illumina, 2011).



**Figura 8.** Secuenciación de cDNA mediante PCR puente (Illumina, 2011).

A. Generación del cluster para la secuenciación. B. PCR tipo puente. C. Cluster's de DNA's de cuatro fragmentos diferentes. D. Incorporación de dNTP al primer nucleótido del fragmento de DNA (Illumina, 2011).



**Figura 9.** Detección de fluorescencia y captación de la primera base correspondiente a cada cluster (Illumina, 2011).

#### **2.2.1.1.4 Análisis de datos obtenidos por RNA-Seq**

Como otras tecnologías de secuenciación de alto rendimiento, el método de análisis transcripcional mediante RNA-Seq enfrenta varios retos, incluyendo el desarrollo de métodos eficientes para almacenar, recuperar y procesar grandes cantidades de datos, que deben ser superados para reducir los errores en el análisis de imagen y eliminar lecturas de baja calidad (Wang *et al.*, 2009).

Por ejemplo, Hiseq2000 (Illumina) puede producir hasta 200 millones de lecturas de 100-nt (aproximadamente 50 GB) de datos en una corrida que proviene de un sólo carril del secuenciador. Estos datos deben ser procesados para identificar no sólo la forma correcta de alinear los transcritos, sino también para el ensamblaje y la cuantificación que pueden dar lugar a resultados con significado biológico. El conjunto de datos experimentales por duplicado o triplicado sirve para corregir la variabilidad de datos y facilitan una interpretación óptima de los datos (Chu & Corey, 2012).

Afortunadamente, la comunidad de Bioinformática ha trabajado intensamente en el desarrollo de las matemáticas, estadística y la informática construyendo así herramientas de software efectivas para el análisis de la expresión diferencial de genes (DEG) mediante RNA-Seq. Las herramientas para el análisis transcripcional mediante RNA-Seq generalmente se dividen en tres categorías: (i) para leer alineaciones; (ii) las de ensamblaje de transcritos o anotación del genoma; y (iii) aquellas para la cuantificación de los transcritos y genes (Trapnell *et al.*, 2012).

#### **2.2.1.1.5 Análisis de calidad**

La calidad se mide en base a la probabilidad que tiene una base de haber sido mal colocada por el secuenciador, como criterio de calidad se emplea un valor de puntaje Phred. La literatura recomienda un puntaje Phred  $\geq 20$ , es decir, que existe un error por cada 100 nucleótidos.

Es importante realizar un reporte de calidad a los archivos FASTQ que contienen las lecturas, esto se logra mediante las herramientas “FastQ Quality Statistics” y “Draw Quality Score Boxplot” para visualizar los datos en forma de gráfico. Este programa otorga un puntaje Phred a cada nucleótido, los ordena y aquellos que estén por debajo del valor recomendado (Phred  $\geq 20$ ) se dice que tiene mala calidad.

Posteriormente se puede realizar un filtrado o un recorte de las secuencias; lo más recomendable es hacer un recorte por calidad (trimming), para esto se requiere de la herramienta “FastQ quality trimmer”, la cuál selecciona todos los nucleótidos con calidad aceptable y aquellos que no estén por arriba del valor deseado los recorta y elimina con el fin de dejar únicamente las lecturas más confiables para los alineamientos (Blankenberg *et al.*, 2010).

#### **2.2.1.1.5 Alineamientos y ensamblaje**

Una vez obtenidas las lecturas de alta calidad, la primera tarea que implica el análisis de datos es alinear las lecturas cortas del RNA-Seq con el genoma de referencia (ensamblaje con base a un genoma de referencia), o bien ensamblarlos en “contigs” (segmentos de DNA superpuestos que en conjunto representan una región consenso de DNA) antes de alinearlos a la secuencia genómica para revelar la estructura de la transcripción (ensamblaje *de novo*), (Chu & Corey, 2012). Por ejemplo, los usuarios que trabajan sin un genoma secuenciado, pero que están interesados en el descubrimiento de genes, deben considerar realizar de ensamblaje de transcriptoma *de novo* (obtención de “contigs”) utilizando una de varias herramientas como Trinity, Trans-abiyss u Oases, entre otras (Trapnell *et al.*, 2012).

Existen varios programas para alinear las lecturas con el genoma, incluyendo ELAND, SOAP31, MAQ32 y RMAP (Chu & Corey, 2012). Por otro lado programas



bioinformáticos como TopHat y Cufflinks proporcionan un flujo de trabajo completo de RNA-Seq (alineamiento y ensamblaje en base a un genoma de referencia), pero existen otros paquetes de análisis de RNA-Seq que pueden utilizarse en lugar de o en combinación con estas herramientas; otra ventaja es la facilidad de manipular estas para analizar DEG gracias a la existencia de la plataforma Galaxy Project (Trapnell *et al.*, 2012).

#### **2.2.1.1.5 Experimentos de RNA-seq con Tophat y Cufflinks**

Tophat y Cufflinks son herramientas bioinformáticas de libre acceso que permiten analizar la DEG y que además ayudan al descubrimiento de nuevos genes y nuevas variantes de splicing de los datos obtenidos mediante la secuenciación de alto rendimiento de RNAm (RNA-seq).

Tophat (<http://tophat.cbcb.umd.edu/>) alinea lecturas al genoma de referencia y descubre sitios de splicing de los transcritos, esos alineamientos (adquiridos en formato BAM) se usan durante el análisis de diversas formas. Cufflinks (<http://cufflinks.cbcb.umd.edu/>) utiliza éste mapeo y lo compara contra el genoma para ensamblar las lecturas en transcritos. Cuffcompare, que es parte del paquete Cufflinks, se puede utilizar para analizar los transcritos previamente ensamblados al compararlos con una anotación de referencia y a través de múltiples experimentos. Cuffdiff, el cuál es también parte del paquete de Cufflinks, toma las lecturas previamente alineadas con Tophat de dos o más condiciones y a su vez los datos de salida obtenidos en Cuffcompare y reporta los genes y transcritos diferencialmente expresados usando un riguroso análisis estadístico (Trapnell *et al.*, 2012). En la tabla 6 se muestra el protocolo que se debe seguir para realizar un análisis DEG completo.

**Tabla 6.** Procesamiento de Archivos FASTQ que se generan después de la secuenciación.

Procesamiento	Archivo de entrada	Herramientas	Archivo(s) de salida
<i>Análisis de calidad</i>	Archivo FASTQ	Conversión de formato ( <b>FastQ Groomer</b> ) Reporte de calidad ( <b>FastQC: Read QC, FastQ Summary statistics</b> ) Recorte de secuencias ( <b>FastQ quality Trimmer</b> )	FastQSanger
<i>Alineamiento usando un genoma de Referencia</i>	FastQSanger Anotaciones del genoma (gtf) Genoma (FASTA)	Alinea las lecturas obtenidas por RNA-seq utilizando un alineador corto utilizando un genoma de referencia y después mapea los resultados para identificar splicing alternativo entre los exones <b>Tophat2 (Bowtie2)</b> <b>IGV (Visualizador integrativo de genomas)</b>	<b>Hits_aceptados.bam</b> Uniones.bed Inserciones.bed Delecciones.bed
<i>Ensamblaje</i>	Hits_aceptados.bam	Ensamblaje de transcritos, Estimación de abundancia relativa de cada transcrito en base a cuantas lecturas soporta cada uno <b>Cufflinks (FPKM)</b>	<b>Transcritos.gtf</b> Isoformas.fpkm_tracking Genes.fpkm_tracking
<i>Expresión diferencial de genes</i>	Transcritos.gtf Anotaciones del genoma (gtf) Genoma (FASTA)	Compara los transcritos ensamblados con un genoma de referencia Y a través de múltiples experimentos <b>Cuffcompare</b>	<b>Transcritos_combinados.gtf</b> Tracking de transcritos Refmap Tmap
	Transcritos_combinados.gtf Hits_aceptados.bam Anotaciones del genoma (gtf) Genoma (FASTA)	Búsqueda de cambios significativos en la expresión de transcritos, splicing y el uso de promotores <b>Cuffdiff</b>	FPKM tracking Conteo tracking <b>Prueba Expresión Diferencial (gtf)</b> Prueba de splicing diferencial CDS Diferencial Uso diferencial de promotores

### 2.2.1.1.6 Análisis de enriquecimiento de genes mediante GO

El análisis de enriquecimiento de genes por GO (Gene Ontology, por sus siglas en inglés) se ha empleado con frecuencia en estudios de análisis masivo de genes (Dragosits et al., 2010; Qi et al., 2011; Liang et al., 2011; Hesketh et al., 2013; Rebnegger et al., 2014), incluyendo los estudios de expresión diferencial de genes mediante la secuenciación masiva por RNA-seq (Qi et al., 2011; Liang et al., 2012; Hesketh et al., 2013) con el fin de identificar las rutas o procesos metabólicos en los que participa cada gen de acuerdo a su actividad molecular, biológica ó celular. El análisis

de enriquecimiento de genes por GO, es una técnica que permite el análisis de un grupo de genes mediante un sistema de clasificación de ontología de genes en el cual se asignan a un conjunto de categorías o “términos” predefinidos dependiendo de sus características funcionales. Los términos se agrupan en tres categorías: la función molecular (que describe la actividad molecular de un gen), proceso biológico (que describe el papel celular o fisiológico más importante llevado a cabo por el gen, coordinado con otros genes) y proceso celular (que describe la ubicación en la célula donde el producto del gen ejecuta su función). Cada gen puede ser descrito (anotado con) en varios términos.

En estos estudios se recupera un perfil funcional del grupo de genes con el fin de comprender mejor los procesos biológicos, celulares o moleculares subyacentes. El análisis se realiza mediante la comparación del grupo de genes de entrada a cada uno de los términos en el GO y con una prueba estadística aplicada para cada término para ver si éste está enriquecido para los genes de entrada. El resultado del análisis es típicamente una lista clasificada de los términos de GO, cada uno asociado con un valor de p, cuando el valor de p es más próximo a cero, es más significativo el término GO particular asociado con el grupo de genes, en otras palabras se considera que un grupo es significativo cuando el número de genes en análisis y asignados a un término resulta elevado al comparar con el total del número de genes asociados con el término GO en cuestión.

### **2.2.1.2 Validación del análisis transcripcional obtenido por la técnica de RNA-seq**

Cuando se realizan éste análisis transcripcionales (RNA-seq) es necesario verificar mediante la validación de algunos genes diferencialmente expresados mediante RT-qPCR; por ejemplo Qi y colaboradores (2011) estudiaron los perfiles de transcriptoma de *P. pastoris* a dos condiciones de cultivo (uno en fase estacionaria y otra en fase exponencial) mediante RNA-seq y observaron la expresión diferencial de 22 genes que codifican chaperonas y 45 genes que codifican proteínas de transporte y secreción; además compararon los niveles de expresión diferencial obtenidos por RNA-seq y RT-qPCR de algunos genes, observaron los mismos niveles de sobre-expresión y sub-expresión de los genes seleccionados con mayor sensibilidad en la técnica de RT-qPCR.

### 2.2.1.2.1 Técnica de RT-qPCR para el análisis de expresión diferencial de genes

La transcripción reversa (RT) seguida de una reacción en cadena de la polimerasa (PCR), representa la tecnología más poderosa para la amplificación y detección de moléculas de RNA específicas. Para cuantificar estos genes expresados en muestras biológicas, el método de RT acoplado al PCR cuantitativo en tiempo real (RT-qPCR) es el método de elección, debido a su alta sensibilidad y especificidad, así como a una buena reproducibilidad y un amplio rango dinámico de cuantificación a diferencia del PCR semicuantitativo (Pfaffl, 2004).

La técnica de qPCR se fundamenta en el uso de un termociclador acoplado a un fluorómetro que permite la determinación del número de copias de DNA que se sintetizan en la reacción para un DNA blanco específico. Mediante una correlación entre el número de copias sintetizadas y el número de ciclos de síntesis, es posible calcular la cantidad o concentración de DNA blanco con el que se inició la reacción. Si el DNA blanco a cuantificar proviene de una reacción de retro-transcripción, esta técnica permite la cuantificación de un transcrito específico y se le denomina RT-qPCR (Bustin, 2002).

Como se mencionó antes, las reacciones en tiempo real se llevan a cabo en un termociclador que permite la medición mediante la detección de una molécula fluorescente, lo cual disminuye los pasos de procesamiento posteriores y minimiza el error experimental. Esto se logra principalmente mediante el uso de tecnologías basadas en fluorescencia como: 1) sondas que fluorescen después de su hidrólisis durante la síntesis del DNA (TaqMan) o hibridación (Tecnología LightCycler); 2) horquillas fluorescentes; o 3) agentes intercalantes (SYBR Green) (Tabla 7). Estos enfoques requieren menos cantidad de RNA y procesan un mayor rango dinámico que la densitometría, y son más resistentes a amplificaciones no específicas (VanGuilder *et al.*, 2008).

En el caso de SYBR Green, muestran un ejemplo de un agente intercalante que fluoresce con la unión a DNA de doble cadena. Después de la síntesis mediada por oligonucleótidos específicos de la secuencia blanco durante el ensayo de PCR, múltiples moléculas de SYBR Green se unen al producto y emiten una fuerte señal de fluorescencia, la cual es fácilmente detectada.

**Tabla 7.** Tecnologías disponibles para RT-qPCR (VanGuilder *et al.*, 2008).

<b>Tecnología</b>	<b>Proveedor</b>	<b>Detección</b>	<b>Ensayos múltiples</b>	<b>Análisis de curva de disociación</b>
<b>TaqMan</b>	Applied Biosystems	Sondas de hidrólisis	Si	No
<b>PrimeTime</b>	IDT Technologies	Sondas de hidrólisis	Si	No
<b>LightCycler</b>	Roche	Sondas de hibridación	Si	No
<b>LUX</b>	Invitrogen	Sonda de horquilla, hibridación	Si	Si
<b>Oligobalizas (Molecular Beacons)</b>	Sigma-Aldrich, ITD	Horquilla bucle, hibridación	Si	Si
<b>SYBR Green</b>	Múltiple	Agente intercalante	No	Si

Debido a que SYBR Green se une indiscriminadamente a DNA de doble cadena, las secuencias de dímeros de oligonucleótidos y errores de amplificación (productos inespecíficos) son detectados en la fluorescencia.

Una consecuencia de la unión de múltiples moléculas, es que la cantidad de señal es dependiente de la masa molecular del producto de DNA de doble cadena creado en la reacción. Si las eficiencias de amplificación son las mismas, la amplificación de un producto más grande, generará una mayor señal que un producto más corto.

Los agentes intercalantes (comúnmente el SYBR Green) se añaden a la mezcla de PCR y se intercalan entre las dos hebras del DNA a lo largo de los ciclos de amplificación, induciendo fluorescencia que posteriormente se mide para cuantificar la cantidad de DNA inicial. Debido a la baja especificidad de estos agentes intercalantes, es recomendable realizar una determinación de la curva de disociación del producto amplificado para comprobar su identidad y la ausencia de productos inespecíficos (VanGuilder *et al.*, 2008).

#### **2.2.1.2.2 Análisis de los datos obtenidos por RT-qPCR**

Antes de realizar los ensayos de qPCR se deben tomar en cuenta una serie de sugerencias propuestas por Bustin y colaboradores (2009) llamada Guía MIQE (Mínima información para la Publicación de Experimentos de PCR Cuantitativo), la cuál sugiere:

- 1) La buena manipulación y obtención de las muestras, donde los pasos más críticos es el tiempo y forma de almacenamiento de las células o tejidos con las cuáles se va a trabajar, de preferencia muestras frescas y que se mantengan a bajas temperaturas en todo momento; así como elegir el método adecuado de extracción de los ácidos nucleicos que permitan la obtención de concentraciones e integridad aceptables para el análisis.
- 2) En el caso de RNA, es importante establecer la misma concentración de RNA que se va a retro-transcribir, utilizar el mismo método de extracción y las mismas condiciones en el termociclador para todas las muestras.
- 3) Se deben diseñar los oligonucleótidos adecuados para cada gen de interés que tengan especificidad y que no generen estructuras secundarias, además es importante utilizar programas que permitan predecir la temperatura de melting ( $T_m$ ) para implementar las condiciones óptimas en el termociclador.
- 4) Es importante utilizar controles negativos de la qPCR para evitar contaminaciones de los reactivos y de la RT para evitar contaminaciones de DNA. Además también es necesario implementar curvas de diluciones seriadas para calibrar los ensayos de qPCR y obtener una eficiencia para corregir la fórmula empleada para el cálculo de expresión relativa.
- 5) También es necesario emplear genes de referencia de expresión constante cuya abundancia muestre una fuerte correlación con las concentraciones de RNAm de las muestras para la normalización de los genes de interés. La elección de éstos genes y el número óptimo se deben determinar experimentalmente y se debe verificar con los métodos reportados.

Un punto clave en los ensayos de qPCR es la estandarización de los componentes para las reacciones. El desarrollo de mezclas maestras (Master Mix) con todos los reactivos para la reacción de qPCR (dNTP's, polimerasa,  $MgCl_2$ ,) y métodos de purificación para mejorar la calidad y la cantidad de las muestras de RNA, han mejorado la eficiencia de los ensayos. Hoebeeck y colaboradores, en el 2007, ensayaron la importancia de estos puntos y vieron cómo influyen en la variación de las reacciones de qPCR, concluyendo que en las reacciones de qPCR deben ser controlados diversos

factores como la homogeneidad del bloque de lectura, homogeneidad de los reactivos en todas las muestras, la calidad de la muestra a evaluar, ausencia de inhibidores e integridad de su secuencia (Hoebeeck *et al.*, 2007).

Una alta reproducibilidad en las reacciones de RT- qPCR y resultados confiables sobre el proceso de cuantificación del RNAm está asociado con pasos de estandarización pre-analíticos (muestreo y conservación, extracción de RNA y conservación, controles de cantidad y calidad del RNA), optimización de ensayos de RT y PCR (en términos de especificidad, sensibilidad, reproducibilidad y robustez) y el procesamiento de datos obtenidos por los ensayos (adquisición de datos, evaluación, cálculos y análisis estadísticos) (Pfaffl *et al.*, 2004).

Para la cuantificación de los valores obtenidos en los ensayos de RT-qPCR existen dos tipos generales de estrategias: medición de genes expresados diferencialmente de forma absoluta o relativa (Tabla 8).

**Tabla 8.** Diferencias entre las dos diferentes estrategias de cuantificación y sus ejemplos.

<b>Cuantificación Absoluta (Método PCR Digital)</b>	<b>Cuantificación Absoluta (Método curva Standard)</b>	<b>Cuantificación Relativa</b>
En éste tipo de cuantificación no se necesitan de estándares conocidos. El blanco de interés puede ser directamente cuantificado con precisión determinado por el número de replicados digitales.	En éste método la muestra no conocida se cuantifica en base a una muestra con cuantificación conocida. Primero se crea una curva estándar; y luego se compara la muestra no conocida con esa curva y se extrapola el valor.	Se analizan cambios en expresión de genes en una muestra relativa con otra muestra de referencia (como una muestra control no tratada).
<b>Ejemplos</b>		
-Cuantificar copias de un alelo presente en mezclas heterogéneas. -Cuantificar el número de equivalentes celulares en una muestra de DNA genómico. -Determinar el número absoluto de copias virales presentes en una muestra sin estándar de referencia.	Correlacionar un número de copias virales con un estado de enfermedad.	La comparación de niveles de expresión génicos de un gen en particular en una muestra químicamente tratada relativa al nivel de expresión de una muestra no tratada.

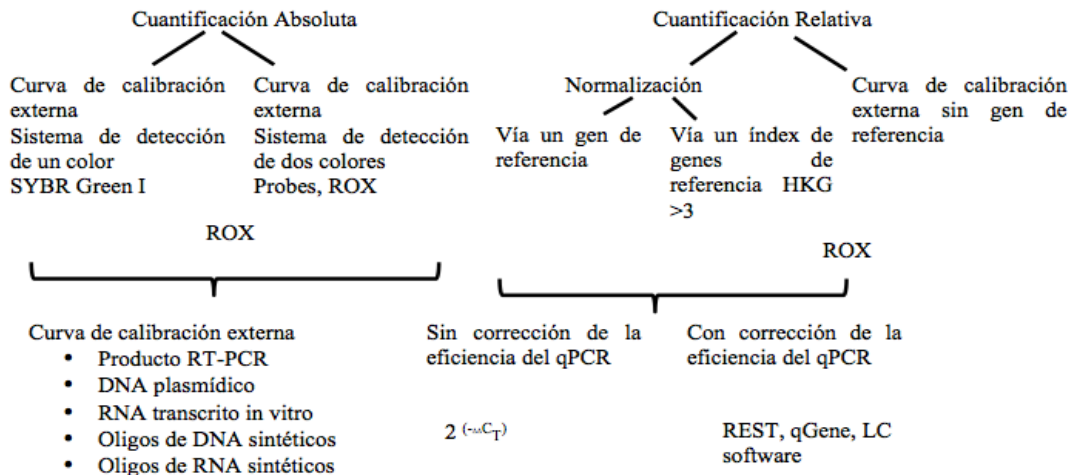
La cuantificación absoluta relaciona la señal de fluorescencia del qPCR con el número de copias empleando una curva de calibración, las cuales pueden derivar de productos de PCR diluidos, DNAc o DNA recombinante, plásmidos linearizados, o muestras de tejido. La confiabilidad de un enfoque absoluto depende tener una eficiencia

de amplificación idéntica para cada ensayo, tanto para el DNAc de la muestra a evaluar como para el blanco empleado en las curvas de calibración. Este enfoque absoluto es erróneo, ya que la cuantificación se muestra relativa al uso de una curva de calibración. El número de copias de RNAm debe ser correlacionado a algunos parámetros biológicos, como la masa de la muestra, la cantidad total de RNA, un número definido de células, o compararlo con el número de copias de un gen de referencia como podría ser RNA ribosomal o genes constitutivos (Pfaffl, 2004).

La cuantificación relativa o comparativa, mide el cambio relativo en los niveles de expresión de RNAm, determina el cambio en los niveles de RNAm de un gen en un estado a través de múltiples muestras y lo expresa relativo a los niveles de otro RNAm. La cuantificación relativa no requiere una curva de calibración o estándares con concentraciones conocidas y la referencia puede ser cualquier transcrito, siempre y cuando se conozca su secuencia para el buen diseño de oligonucleótidos y sondas correspondientes (Bustin, 2002).

Una forma de determinar los niveles relativos de expresión, es midiendo las diferencias ( $\Delta$ ) entre el número de ciclos de amplificación del gen de interés y un gen normalizador en donde la señal de fluorescencia de cada uno cruza un umbral de fluorescencia determinado ( $C_T$ ), y posteriormente se compara el nivel relativo de expresión del gen de interés con una muestra calibradora: control sin tratamiento, el tiempo cero, individuos sanos, etc. Este método se denomina método de  $2^{-\Delta C_T}$  (Pfaffl, 2004), y se basa en dos suposiciones: la primera es que todas las reacciones ocurren con 100% de eficiencia, es decir, que en cada ciclo de qPCR la cantidad de producto se duplica. La segunda suposición es que hay un gen (o genes constitutivos) que se expresa en un nivel constante en todas las muestras. Este control endógeno se emplea para corregir cualquier diferencia en el muestreo. Una vez que se obtienen los datos de  $C_T$  de cada reacción, se pueden generar niveles relativos de expresión (VanGuilder *et al.*, 2008). En la figura 10, donde se muestran las diferencias entre las estrategias de cuantificación.





**Figura 10.** Resumen de las estrategias de cuantificación.

Para obtener resultados con una expresión relativa óptima, se deben generar estrategias de normalización para el control del error experimental y para asegurar un rendimiento idéntico durante el ensayo de RT-qPCR. Estas variaciones pueden ser introducidas por varios procesos, como son la extracción y el procesamiento de RNA, el montaje del ensayo de PCR y por los ciclos del mismo ensayo. Todas las comparaciones relativas deben realizarse con una cantidad constante de RNA o masa de tejido de muestra analizada, o una cantidad idéntica de células seleccionadas. Para asegurar que las condiciones iniciales del ensayo sean idénticas, la expresión relativa de los datos debe ser equilibrada o normalizada de acuerdo al tamaño o masa de las muestras o el volumen de tejido, la cantidad total de RNA obtenida, la cantidad de DNA genómico obtenida, RNAs ribosomales de referencia (18S o 26S RNAr), RNAs mensajeros de referencia, cantidad total de DNA genómico, moléculas de DNA o RNA artificial (Pfaffl, 2004). Para la normalización se emplea el uso de genes de expresión constitutiva (“housekeeping”) o de linaje específico, los cuales también se denominan genes de referencia. La selección de estos genes debe cumplir los siguientes requisitos: expresión similar entre todas las muestras en un estudio dado, expresión resistente a las condiciones experimentales, y expresión que se somete a todos los pasos del ensayo de qPCR con la misma cinética que los genes blanco. Generalmente, la transcripción de estos genes es resistente a la condiciones experimentales, haciéndolos adecuados como controles endógenos para una normalización (Vanguilder *et al.*, 2008).

### 3. JUSTIFICACIÓN

*P. pastoris* es un organismo frecuentemente utilizado para la producción de proteínas heterólogas y el estudio de los procesos metabólicos implicados en la producción de éstas proteínas. Sin embargo, a pesar de los esfuerzos realizados, actualmente se cuenta con un conocimiento limitado sobre la respuesta del hospedero a las alteraciones causadas por los altos niveles de expresión de la proteína recombinante aunado a las condiciones de cultivo que conllevan a estos niveles altos de productividad.

Por esta razón, en este trabajo se busca comprender a nivel transcripcional, la respuesta metabólica causada por factores ambientales como la temperatura, pH y la fuente de carbono, además de la carga que implica la producción de proteínas heterólogas en este hospedero, enfocándose en los mecanismos implicados en los procesos de secreción, proteólisis y metabolismo del metanol.

Para cumplir con este objetivo, en este trabajo se analizaron los transcriptomas de la cepa *P. pastoris* KM71FTEII, que expresa y secreta al medio de cultivo una fosfohidrolasa, cultivada en condiciones que generan una alta y baja producción de la proteína heteróloga modelo, y se validaron los resultados del análisis transcripcional por RT-qPCR mediante la expresión de genes relevantes en la vía de secreción y metabolismo del metanol.

### 4. HIPÓTESIS

Condiciones diferentes de cultivo de la cepa recombinante KM71FTEII (Mut<sup>s</sup>) de *P. pastoris*, influyen en la expresión diferencial de genes que afectan la producción de la proteína heteróloga.

## **5. OBJETIVO GENERAL**

Analizar la expresión diferencial de genes durante la producción de una proteína heteróloga modelo en una cepa Mut<sup>s</sup> de *P. pastoris* bajo dos condiciones de cultivo diferentes.

## **6. OBJETIVOS ESPECÍFICOS**

1. Realizar un análisis transcripcional mediante RNA-seq de la cepa KM71FTEII de *P. pastoris* cultivada a condiciones de cultivo de baja y alta producción de una proteína heteróloga modelo.

1.1. Analizar la expresión diferencial de los genes implicados en el metabolismo de metanol.

1.2. Analizar la expresión diferencial de los genes implicados en la vía de secreción.

2. Validar la expresión de los genes relevantes implicados en la vía de secreción y de la vía del metabolismo del metanol bajo las dos condiciones de cultivo evaluadas mediante RT-qPCR.

## 7. MATERIALES Y MÉTODOS

### 7.1 Área de trabajo

El proyecto se desarrolló en el Laboratorio de Biotecnología Molecular (L5) del Instituto de Biotecnología de la Facultad de Ciencias Biológicas de la Universidad Autónoma de Nuevo León.

### 7.2 Material biológico

Para este estudio se empleó la cepa KM71FTEII de *P. pastoris* (Mut<sup>s</sup>) la cual contiene en su genoma el gen que codifica para la secuencia de una fitasa termoestable, construida previamente en el laboratorio y crecida en un biorreactor a dos condiciones de cultivo diferentes, una con alta producción de fitasa extracelular (cultivo CM o J) y la segunda con baja producción de fitasa extracelular (cultivo CP o C) conservadas en RNA later. Las condiciones de cultivo seleccionadas que condujeron a la alta y baja producción se muestran en la tabla 9.

**Tabla 9.** Condiciones de cultivo y rendimiento en el paso de inducción con MeOH en CM y CP (Viader-Salvadó *et al.*, 2013).

Condiciones de cultivo en la etapa de inducción					
	pH	[MeOH] gL <sup>-1</sup>	Densidad celular inicial (gL <sup>-1</sup> DCW)	Temperatura (°C)	
CM	6	1.5	90	24	
CP	3	0.5	88	30	
68h de inducción con MeOH					
	$\mu$ (h <sup>-1</sup> )	$Y_{p/x}$ (g g <sup>-1</sup> )	Densidad celular final (gL <sup>-1</sup> DCW)	Actividad de fitasa extracelular	
				(U)	(g/L)
CM	0.014	0.114	119	976,102	6.4
CP	-0.007	-0.120	68	174,839	1.9

Los paquetes celulares obtenidos de las diferentes condiciones de cultivo y a diferentes tiempos de inducción con metanol, incluyendo el tiempo cero sin inducción, se procesaron en RNA later y guardaron a -80 °C para su posterior uso.

## **7.3 Materiales y Equipos Empleados**

### **7.3.1 Extracción de RNA**

El RNA se obtuvo a partir de células previamente conservadas con el reactivo RNA Later (*Ambion*) mediante el kit SV Total RNA Isolation System (*Promega*), el cual contiene buffer de lisis de RNA (RLA), buffer de dilución de RNA (RDA),  $\beta$ -mercaptoetanol (48.7%), DNasa I (liofilizada),  $MnCl_2$  0.09 M, buffer “Yellow Core”, solución de paro de DNasa (DSA), solución de lavado de RNA (RWA), agua libre de nucleasas, tubos de elución y columnas de ensamble para centrifugación.

#### **7.3.1.1 Preparación de reactivos extras al kit de extracción**

Para el Buffer de lisis (0.1 M EDTA pH 7.4, 1 M Sorbitol y 0.1%  $\beta$ -mercaptoetanol), los primeros dos reactivos se prepararon por separado con agua tratada con DEPC; para la preparación del agua se añadió 1 mL de DEPC (Dietilpirocarbonato) en 1 L de agua MiliQ se agitó vigorosamente durante 15 min y se dejó incubar en oscuridad a 37°C durante toda la noche, posteriormente se esterilizó para desactivar el DEPC.

#### **7.3.2 Tratamiento DNasa**

Para el tratamiento con DNasa se utilizó el kit de DNasa RQ1 libre de RNasa (*Promega*), el cual contiene buffer de reacción 10 $\times$  (Tris-HCl 400 mM pH 8.0,  $MgSO_4$  100 mM y  $CaCl_2$  10 mM), solución de paro (EGTA 20 mM pH 8.0) y buffer de almacenamiento de enzima (HEPES 10 mM pH 7.5, 50% glicerol (v/v),  $CaCl_2$  10 mM y  $MgCl_2$  10 mM).

#### **7.3.3 Cuantificación e integridad de las muestras de RNA**

Se utilizó el nano-espectrofotómetro P-300 (*IMPLEN*), celda (lid10 *IMPLEN*) para la cuantificación del RNA; así como un bioanalizador de microfluidos 2100 (*Agilent Technologies*), nano chips de RNA (2100 Bioanalyzer, *Agilent Technologies*) para cuantificar y determinar la integridad del RNA.

### 7.3.4 Síntesis de DNAc

Para la síntesis del DNAc se utilizó el kit AffinityScript Multiple Temperature DNAc Synthesis Kit (*Agilent Technologies*) para 50 reacciones, el cual contiene buffer 10× AffinityScript RT, inhibidor de ribonucleasa que bloquea la RNasa (40 U/μL), iniciador oligo (dT) (0.5 μg/μL), iniciadores aleatorios (0.1 μg/μL), dNTPs 100 mM (25 mM cada dNTP), agua libre de RNasa. Además, se utilizó el termociclador (MultiGene Mini, *Labnet International, Inc*) para llevar a cabo la reacción.

### 7.3.5 Ensayos de qPCR

En los ensayos de qPCR se siguieron todas las recomendaciones sugeridas en la guía MIQE (Mínima información para la Publicación de Experimentos de PCR Cuantitativo) descrita por Bustin y colaboradores (2009).

Para la preparación de las muestras se utilizó el Master Mix Brilliant III Ultra-Fast SYBR Green QPCR & QRT-PCR (*Agilent Technologies*), posteriormente el ensayo de qPCR se efectuó en un Termociclador (Stratagene Mx3005P, *Agilent Technologies*).

## 7.4 Diseño experimental

Para el cumplimiento de los objetivos planteados, se diseñó una estrategia general la cual está esquematizada en la figura 11.

El diseño experimental puede resumirse en dos etapas: el análisis transcripcional y la validación de la expresión diferencial de genes involucrados en el proceso de secreción.

Las preparaciones de RNA se realizaron a partir de paquetes celulares previamente cultivados a dos condiciones de cultivo diferentes a las 47 h de inducción con metanol (CM: uno de alta producción de fitasa extracelular y CP: uno de baja producción de fitasa extracelular) y un control (C0: sin inducción con metanol) conservados en RNA later a -80 °C.

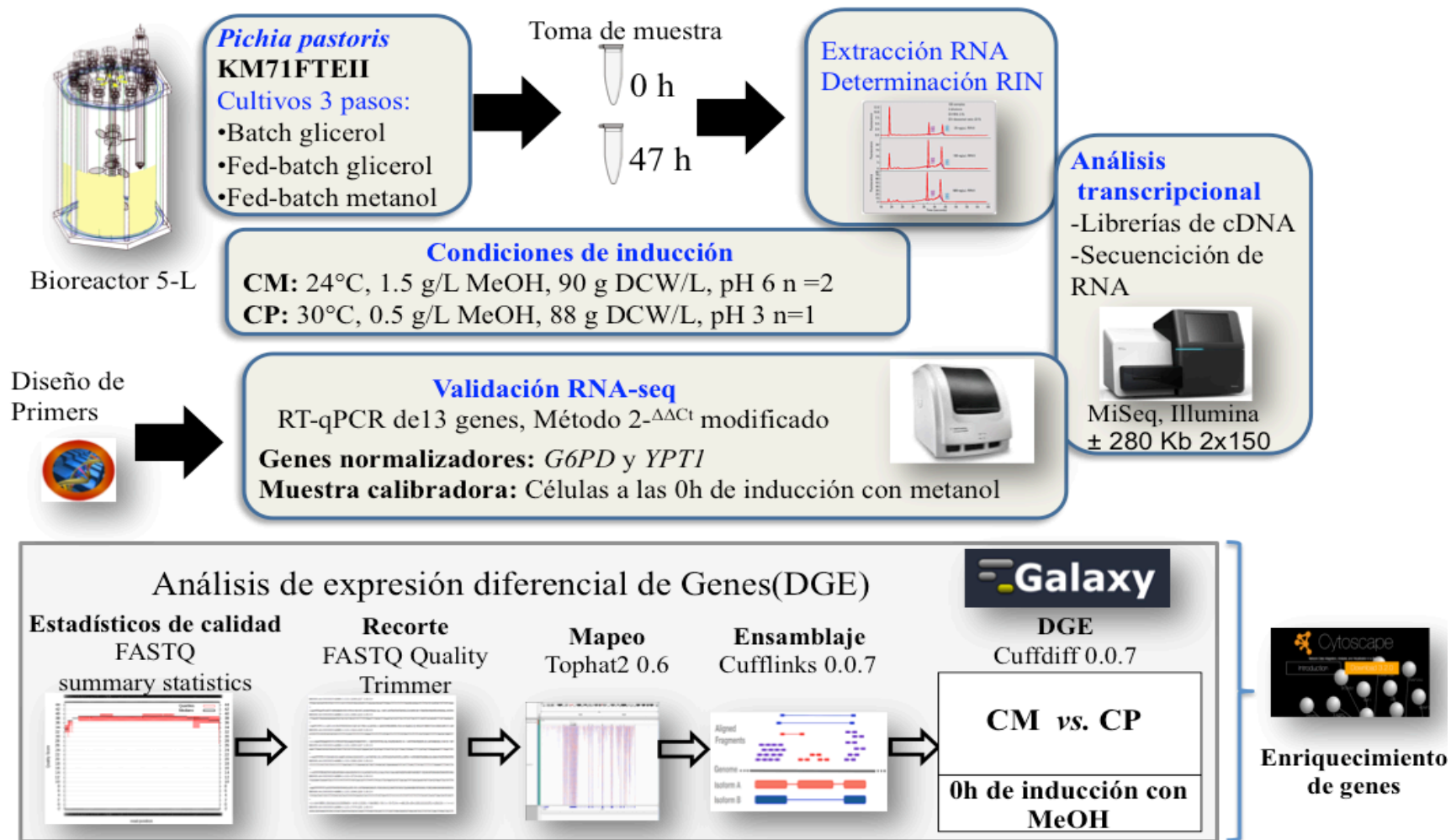


Figura 11. Diseño experimental.

Para el análisis transcripcional y la validación con la técnica RT-qPCR se trabajó únicamente con los paquetes celulares de los tiempos a 0 h (C0: control, única muestra) y 47 h (duplicados biológicos CM47-1 y CM47-2; CP47-1, única muestra).

En la primera etapa, se realizó un análisis de microfluidos con el bioanalizador Agilent 2100 (empleando un chip nano de RNA) para verificar la integridad y calidad del RNA y cumplir con los requisitos ( $RIN > 7$ ) indispensables para la secuenciación. Después de obtener las secuencias se evaluó la calidad de las mismas (punto 7.5.3.1), para posteriormente alinearlas y ensamblarlas en base a un genoma de referencia modificado (descrito en el punto 7.5.3.2) mediante los programas bioinformáticos Tophat2 v0.6 y Cufflinks v0.0.7, respectivamente; se utilizó Cuffcompare v0.0.6 para realizar las comparaciones de las muestras (CM vs C0, CP vs C0 y CM vs CP); finalmente los resultados se cuantificaron y normalizaron mediante el empleo de la prueba de expresión diferencial con el programa Cuffdiff v0.0.7; todos se utilizaron en la plataforma de Galaxy Project (descritos los puntos 7.5.3.2/7.5.3.3). Después de obtener la expresión diferencial de genes, se realizó un análisis de enriquecimiento de genes para identificar los procesos biológicos a los que estaban relacionados los genes sobre- y sub-expresados con el programa ClueGo v2.1.5 de Cytoscape v3.1.0 (descritos en el punto 7.5.4.1).

En la segunda etapa, primero se seleccionaron algunos genes que resultaron diferencialmente expresados en el análisis transcripcional en base a su sobre- y sub-expresión con valores de  $\log_2$  (Fold\_change) más altos en CM vs CP. Para la amplificación de los genes sobre- y sub-expresados de la vía de secreción se diseñaron 5 y 5 pares de oligonucleótidos, respectivamente; además para la amplificación de los genes involucrados en el metabolismo del metanol se diseñaron 3 pares; todos los diseños se realizaron con el programa Primer3 v0.4.0 (descrito en el punto 7.6.3). Posteriormente se realizó un RT a partir del RNA extraído de los paquetes celulares con un previo tratamiento de DNAasa para realizar el ensayo de qPCR para cada uno de los genes seleccionados y obtener los niveles de expresión relativa con el método  $2^{-\Delta\Delta C_T}$  modificado.



## 7.5 Análisis transcripcional

### 7.5.1 Protocolo de Extracción de RNA

Para la obtención de preparaciones de RNA a partir de las muestras conservadas previamente, se utilizó el Kit de Promega SV Total RNA Isolation System®, se siguió el protocolo del manual de Promega modificado, se trabajó en un ambiente libre de polvo u otros contaminantes dentro de una estación de trabajo para PCR de Probiotek® (Productos y equipos biotecnológicos, S.A. de C.V.). Además se utilizaron guantes, puntillas, tubos y agua libres de RNasa.

El protocolo modificado consiste en resuspender el paquete celular (50 mg de biomasa) en 90 µL de un Buffer de lisis (0.1 M EDTA pH 7.4, 1 M Sorbitol y 0.1% β-mercaptoetanol) mediante pipeteo hasta que quede una mezcla homogénea; después de esta re-suspensión de los paquetes celulares, se añadieron 50 U de liticasa, se mezcló con ayuda de un vortex durante 5-10 s y se incubó durante 50 min a 30°C en un agitador termoestático (Thermomixer R, Eppendorf) sin agitación. Es importante mencionar que durante ésta parte las muestras se mantuvieron en hielo el mayor tiempo posible para evitar degradaciones. La enzima de lisis empleada fue obtenida de la compañía Sigma-Aldrich® (Liticasa de *Arthrobacter luteus* L2524).

Después de la incubación se agregaron a la suspensión 75 µL del buffer de lisis, se homogenizó por pipeteo, se agregaron 300 µL del buffer de dilución y se mezcló inmediatamente por inversión.

Las muestras se centrifugaron durante 10 min a 12,000 rpm a 4°C, se formó un paquete celular y una fase acuosa (color azul). La fase acuosa se transfirió a un tubo nuevo de 2 mL, se agregaron 200 µL de etanol al 95%, se mezcló por pipeteo, la mezcla generada se transfirió a un cartucho nuevo de filtración (ensamble de tubo colector y filtro), éste se centrifugó durante 1 min a 12,000 rpm a 4°C y se descartó el líquido del tubo colector. En el siguiente paso se añadieron al cartucho 600 µL de solución de lavado de RNA y éste se centrifugó nuevamente durante 1 min a 12,000 rpm a 4°C.

Posteriormente, se preparó una solución de DNasa, agregando los siguientes reactivos para una preparación: 40 µL de Yellow Core Buffer, 5 µL de MnCl<sub>2</sub> 0.09 M y 5 µL de DNasa I. Esta solución se agregó a los cartuchos de filtración y se incubó

durante 15 min a una temperatura de 20-25°C; es importante preparar la solución de DNasa hasta este paso para evitar la degradación de la enzima. Una vez terminada la incubación, se añaden 250 µL de solución para desactivar la DNasa y el cartucho de filtración se centrifugó durante 1 min a 12,000 rpm a 4°C.

En la siguiente etapa se añadieron 600 µL de la solución de lavado, el cartucho se centrifugó durante 1 min a 12,000 rpm a una temperatura de 4°C y se descartó el líquido del tubo colector. Se añadieron 250 µL de solución de lavado y se centrifugó durante 2 min a 12,000 rpm a una temperatura de 4°C. Por último, se transfirió el cartucho de filtración a un tubo de elución de 1.5 mL, se agregaron 100 µL de agua libre de nucleasas, se centrifugó a 12,000 rpm durante 1 min a 4°C y se descartó el cartucho de filtración.

Se determinó la concentración (ng/µL) de las preparaciones de RNA por espectrofotometría determinando los valores de absorción a 260 y 280 nm y la relación de los mismos (260/ 280) mediante un espectrofotómetro (Nanodrop ND2000). La calidad del RNA se determinó mediante una electroforesis en gel de agarosa al 0.8% en buffer de tris-boratos-EDTA (TBE 1X, 90 mM Tris-HCl, 89 mM ácido bórico, 2 mM EDTA) y además se determinó la integridad mediante el análisis en el Bionalizador Agilent 2100 empleando el estuche “Agilent RNA 6000 nano kit”, el cuál incluye chips de microfluidos, reactivos y consumibles; para obtener los electroferogramas con los picos correspondientes a las bandas 26s y 18s ribosomales y el número de integridad del RNA (RIN).

Finalmente cada tubo de elución se etiquetó y almacenó; una parte a -80°C para los posteriores ensayos y otra parte se almacenó en hielo seco para el envío de las muestras al Laboratorio Nacional de Genómica para de Biodiversidad (LANGEBIO) del Centro de Investigación y de Estudios Avanzados (CINVESTAV) del Instituto Politécnico Nacional.

### **7.5.2 Obtención de las librerías de DNAc y secuenciación de los transcritos**

Las librerías de DNAc de las cinco preparaciones de RNA se obtuvieron empleando el kit “Truseq RNA simple preparation” (Illumina), con el cual se realizó un enriquecimiento de los RNAm con perlas magnéticas unidas a oligos (dT), una posterior

selección con oligonucleótidos al azar (random primers) y una amplificación de los DNAC mediante un PCR puente, además este protocolo incluye una etapa de la normalización de las librerías, la cuál permite un pool con un volumen simple y aleatorio para la secuenciación de los DNAC.

Las librerías de DNAC se analizaron con el Bioanalizador Agilent 2100 empleando el chip de DNA1000 y se secuenciaron empleando una plataforma de secuenciación de nueva generación (NGS) Miseq de Illumina en una modalidad de secuenciación de dos cadenas (paired end) con lecturas de 250 pb (2x250).

### **7.5.3 Análisis Bioinformáticos**

A partir de la secuenciación se obtuvieron archivos FASTQ (2 archivos por muestra correspondientes a las dos cadenas secuenciadas de cada fragmento de DNAC) que fueron analizados mediante la plataforma bioinformática Galaxy, accesible por internet (<http://galaxyproject.org>), siguiendo las recomendaciones descritas en la literatura (Trapnell *et al.*, 2012).

#### **7.5.3.1 Análisis de calidad de los archivos FASTQ**

Se emplearon los programas FASTQ Groomer, FastQC:ReadQC y FASTQ Quality Trimming (Blankenberg *et al.*, 2010) para obtener un formato FASTQ estándar, evaluar la calidad de las lecturas y recortar las secuencias eliminando nucleótidos de baja calidad respectivamente. Como criterio de calidad se empleó un valor de puntaje Phred mayor de 20. Los nuevos archivos FASTQ se evaluaron con el programa FASTQ Summary Statistics (Blankenberg *et al.*, 2010) mediante un gráfico de tipo boxplot de los puntajes de calidad basados en la posición del nucleótido en la secuencia.

#### **7.5.3.2 Mapeo y ensamblaje de los transcritos**

Para comenzar los alineamientos, primero fue necesario realizar una modificación *in silico* del cromosoma 1 del archivo FASTA que contiene el genoma de la cepa GS115

de *Pichia pastoris* (“Ref201010.fa”; BOGAS (Sterck *et al.*, 2012), el cuál se utilizó de referencia para los alineamientos, ensamblaje y cuantificación de los transcritos.

Se añadió la secuencia del plásmido pPIC9FTEII iniciando desde el gen *HIS4* al final del cromosoma 1 y además se modificó el archivo gtf (“pichia\_format.gtf”; BOGAS 01/03/12 (Sterck *et al.*, 2012) que contiene las anotaciones del genoma de la cepa GS115 para añadir las nuevas regiones incorporadas por el plásmido pPIC9FTEII.

Posteriormente con el programa TopHat2 (Trapnell *et al.*, 2012) se realizó un mapeo de los transcritos contra el genoma modificado de la cepa GS115 de *P. pastoris* modificado portador del gen heterólogo *FTEII*. Para esto se tomaron los parámetros estándar sugeridos en la literatura, lo único que se debe seleccionar en la plataforma se muestra en la figura 12.

The image shows a screenshot of the TopHat2 Galaxy interface. The configuration options are as follows:

- Is this library mate-paired?:** Paired-end
- Use a built in reference genome or own from your history:** Use a genome from history
- Select the reference genome:** 2: p\_pastorisFTEII.fa
- TopHat settings to use:** Full parameter list
- Use Own Junctions:** Yes
- Use Gene Annotation Model:** Yes
- Gene Model Annotations:** 1: pichia\_FTEII.gtf

Below the Gene Model Annotations field, there is a note: "TopHat will use the exon records in this file to build a set of known splice junctions for each gene, and will attempt to align reads to these junctions even if they would not normally be covered by the initial mapping."

**Figura 12.** Parámetros que se modificaron en el programa bioinformático TopHat2 localizado en la plataforma Galaxy.

Se utilizó el programa Integrative Genomics Viewer (IGV) (Robinson *et al.*, 2011) para visualizar el mapeo de los transcritos con respecto al genoma modificado. Se

emplearon como archivos de entrada los formatos BAM y BAI generados por el programa TopHat2 de cada archivo FASTQ, así como el genoma de *P. pastoris* junto con el archivo gtf previamente modificados.

Los archivos BAM también sirvieron para hacer el ensamblaje e identificación de los transcritos mediante el programa Cufflinks (Trapnell *et al.*, 2012), también se utilizaron el archivo FASTA del genoma y archivo gtf de *P. pastoris* modificados.

### **7.5.3.3 Analisis diferencial con Cuffdiff**

Las comparaciones entre las condiciones de cultivo se realizaron con el programa Cuffcompare (Trapnell *et al.*, 2012), donde se utilizó como archivos de entrada el archivo gtf generado por Cufflinks de cada condición (CM, CP y C0), así como el genoma y archivo gtf de *P. pastoris* modificados.

Por último se realizó el análisis de expresión diferencial con el programa Cuffdiff (Trapnell *et al.*, 2012) para obtener los genes sobre-expresados y sub-expresados de las muestras CM respecto a C0 (CM/C0), CP respecto a C0 (CP/C0), y CM respecto a CP (CM/CP). Como archivos de entrada se utilizaron el archivo combinado generado por Cuffcompare, el archivo BAM generado por TopHat2, y el archivo FASTA del genoma y gtf de *P. pastoris* modificados, se emplearon los parámetros sugeridos por la literatura (Figura 13, 14 y 15). Se empleó un modelo global único para estimar la dispersión de todos los experimentos (Anders & Huber, 2010).



**Perform quartile normalization:**

Yes ↕

Removes top 25% of genes from FPKM denominator to improve accuracy of differential expression calls for low abundance transcripts.

**Use Reference Annotation:**

Use reference annotation as guide ↕

**Reference Annotation:**  

1: pichia\_FTEII.gtf ↕

Gene annotation dataset in GTF or GFF3 format.

**Perform Bias Correction:**

Yes ↕

Bias detection and correction can significantly improve accuracy of transcript abundance estimates.

**Reference sequence data:**

History ↕

**Using reference file:**

2: p\_pastorisFETII.fa ↕

**Use multi-read correct:**

Yes ↕

Tells Cufflinks to do an initial estimation procedure to more accurately weight reads mapping to multiple locations in the genome.

**Use effective length correction:**



Yes ↕

Cufflinks will not employ its 'effective' length normalization to transcript FPKM.

**Figura 13.** Parámetros que se modificaron en el programa bioinformático Cufflinks localizado en la plataforma Galaxy.

**Use Reference Annotation:**

Yes ↕

**Reference Annotation:**  

1: pichia\_FTEII.gtf ↕

Requires an annotation file in GFF3 or GTF format.

**Ignore reference transcripts that are not overlapped by any transcript in input files:**

**Use Sequence Data:**

Yes ↕

Use sequence data for some optional classification functions, including the addition of the p\_id attribute required by Cuffdiff.

**Choose the source for the reference list:**

History ↕

**Using reference file:**

2: p\_pastorisFETII.fa ↕

**Figura 14.** Parámetros que se modificaron en el programa bioinformático Cuffcompare localizado en la plataforma Galaxy.

**Library normalization method:**

**Dispersion estimation method:**  
  
 If using only one sample per condition, you must use 'blind.'

**Use multi-read correct:**

Tells Cufflinks to do an initial estimation procedure to more accurately weight reads mapping to multiple locations in the genome.

**Perform Bias Correction:**

Bias detection and correction can significantly improve accuracy of transcript abundance estimates.

**Reference sequence data:**

**Using reference file:**

**Figura 15.** Parámetros que se modificaron en el programa bioinformático Cuffdiff localizado en la plataforma Galaxy.

Los niveles de expresión de cada gen mapeado en base al archivo FASTA del genoma modificado se determinó a partir del número de lecturas mapeadas por kilobase de regiones de exones por millón de lecturas mapeadas en el genoma de referencia (FPKM), el cual es un parámetro directamente proporcional al nivel de expresión de un gen e independiente de la cantidad de RNA secuenciado.

Para los resultados más específicos de las vías y procesos metabólicos analizados se consideró que un gen en particular estaba sobre-expresado o sub-expresado si el logaritmo en base 2 de la relación del nivel de expresión de ese gen en una muestra entre el nivel de expresión del mismo gen en la otra muestra que constituye el par de muestras a comparar, es decir, cuando el  $\log_2$  (Fold\_change) fue  $\geq 1.0$  (sobre-expresado más del doble) ó  $\leq -1.0$  (sub-expresado más del doble), respectivamente, y además si la comparación mostró un valor estimado de ( $q \leq 0.05$ ) en base al modelo de dispersión; sin embargo el programa es tolerante e identificó como significativos ( $\log_2$  (Fold\_change)  $\geq 0.7$  y  $\leq -0.7$ ), estos resultados arrojados por el programa se tomaron en cuenta en algunos análisis.

#### 7.5.4 Análisis global de la expresión diferencial de genes

En la expresión diferencial de genes se manejaron tres comparaciones importantes: 1) el cambio en la fuente de carbono de glicerol (C0) a metanol en la condición de cultivo de alta producción (CM), comparación CM/C0; 2) el cambio en la fuente de carbono de glicerol (C0) a metanol en la condición de cultivo de baja producción (CP), comparación CP/C0 3) Las diferencias entre las dos condiciones de cultivo, comparación CM/CP.

Para visualizar el análisis de expresión diferencial de genes global de las tres comparaciones, se emplearon gráficos de tipo volcán graficando en el eje abcisas el valor de  $\log_2(\text{Fold\_change})$  y en el eje de ordenadas el logaritmo del valor P. En el eje del abcisas el punto de corte fue 1.0 y -1.0 (representados con líneas verticales punteadas en los gráficos), mientras que en el de las ordenadas fue 1.3 (línea horizontal punteada), es decir,  $\log_2(\text{Fold\_change}) \geq 1.0$  (sobre-expresado más del doble) ó  $\leq -1.0$  (sub-expresado más del doble) con un valor  $q \leq 0.05$ . Estos puntos de corte delimitan dos cuadrantes que identifican a los genes diferencialmente expresados más significativos (izquierdo = genes sub-expresados y derecho = genes sobre-expresados).

##### 7.5.4.1 Análisis de enriquecimiento de genes

Se utilizó el plugin ClueGo v2.1.5 (Bindea *et al.*, 2009) del programa de cytoscape v3.1.0 (Shannon *et al.*, 2003) para realizar el enriquecimiento de genes y asociarlos a una ruta metabólica específica (términos GO) así como identificar la función o funciones de los genes. Como archivo de entrada se utilizaron todas las claves (ID del gen) de los genes sobre- y sub- expresados considerando un  $\log_2(\text{Fold\_change}) \geq 0.7$  y  $\leq -0.7$  en las tres comparaciones CM/C0, CP/C0 y CM/CP.

Se utilizó la base de datos Go Biological Process (30/01/15) con los parámetros por defecto a excepción de que se activó la opción “Use Go Term Fusion” y se cambió el porcentaje de genes en la opción “Go Term Pathway Selection” a un 10% de genes con la finalidad de mostrar una “network” detallada con términos muy específicos colocados en los niveles de GO 7-15, con pocos genes asociados pero un alto porcentaje



de genes analizados. Estos términos son más específicos y dan una información de los aspectos particulares de los productos de genes estudiados.

Con los resultados del análisis de ClueGo se obtiene un grupo de anotaciones funcionales en una “network” que refleja las relaciones entre los términos GO basados en la similitud de sus genes asociados, el tamaño de los nodos reflejan la significancia estadística de los términos y la conectividad, calculada con estadísticos tipo kappa, sirve para determinar la funcionalidad de los grupos; además un término puede estar en más de un grupo.

Los gráficos de pastel agrupan los diferentes términos GO encontrados basados en la similitud de sus genes asociados y cada sector recibe el nombre del término más significativo (grupo de términos); entre más grande es el tamaño del sector, mayor es el número de términos englobados. La finalidad de los gráficos de pastel es mostrar las categorías de los términos más representativos de procesos biológicos y su relación entre ellos en base a los genes analizados.

Los histogramas muestran los términos desglosados de cada sector del gráfico de pastel. Las barras representan el número de genes encontrados del cluster analizado asociado con el término GO, y la barra representa el porcentaje de los genes encontrados comparados con todos los genes asociados con el término. Los gráficos mostrados en los resultados son parecidos a los histogramas originales obtenidos por el programa. Para ello se graficaron en excel únicamente los términos más representativos (con un valor P corregido más lejano de 1, en la mayoría de los casos  $< 0.05$ )

El análisis estadístico incluye el cálculo automático de los términos y los grupos significantes (corrección de valor P), se muestran con (\*\*) si el término/grupo es muy significativo (valor  $P < 0.001$ ) y (\*) si el término/grupo es significativo (valor  $P < 0.05$ ). Para decir que un grupo es significativo se considera el número de genes únicos encontrados de los genes analizados como parte de una asociación (términos GO) incluidos en el grupo y el total del número único de genes asociados con esos términos, los grupos significativos también se encuentran marcados con asteriscos en los gráficos de pastel.

Debido a que el histograma (resultados desglosados), network (resultado ilustrado en un mapa) y gráfico de pastel (resultados con términos representativos de cada grupo)

muestran lo mismo, en los resultados únicamente se muestran los gráficos pastel y se agregan los histogramas re-estructurados en excel (en ambos casos, con asterisco aparecen los grupos significativos).

#### **7.5.4.1.1 Diagrama de Venn y enriquecimiento de genes**

Además para ver los genes sobre- y sub-expresados en común en las tres comparaciones se empleó el diagrama de Venn. Como archivo de entrada se utilizaron todas las claves (ID del gen) de los genes sobre- y sub- expresados considerando un  $\log_2$  (Fold\_change)  $\geq 0.7$  y  $\leq -0.7$  en las tres comparaciones CM/C0, CP/C0 y CM/CP. A partir de estos resultados, se realizó un análisis de enriquecimiento de genes (ClueGo, con los parámetros estándar), de los genes que se expresaron diferencialmente de manera similar en el cambio de fuente de carbono en CM/C0 y CP/C0, y los genes que se expresaron diferencialmente en CM/CP que no se debieron al cambio de la fuente de carbono.

#### **7.5.4.2 Análisis de expresión diferencial del gen heterólogo FTEII**

Las secuencias aminoacídicas de FTE y FTEII se diseñaron como secuencias de fitasa tipo beta-hélice usando una combinación de dos métodos: 1) La termoestabilidad de ingeniería de proteínas de 15 secuencias aminoacídicas conocidas del género *Bacillus* y 2) un análisis de homología de modelaje de proteínas para determinar la estabilidad de los aminoácidos, usando la estructura de una fitasa de *Bacillus amyloliquefaciens* (Protein Data Bank [PDB] code 2POO) y un programa (Chemical Computing Group Inc., Montreal, Quebec, Canada). La región codificante de FTE se diseño en base a codones preferenciales de *P. pastoris*, posteriormente una mutación N336D dio lugar al gen heterólogo FTEII (Viader-Salvadó *et al.*, 2010).

Se analizó la expresión diferencial del gen heterólogo que se añadió a los archivos gtf y FASTA de las anotaciones y del genoma, respectivamente, descritas en el punto 7.5.3.2 y se cuantificó el gen en las condiciones de cultivo evaluadas de acuerdo a la metodología descrita en el punto 7.5.3.3 en las tres comparaciones evaluadas (CM/C0, CP/C0 y CM/CP).

### **7.5.4.3 Análisis de expresión diferencial de rutas metabólicas específicas**

Se cuantificó la expresión diferencial de genes relacionados al metabolismo del metanol, secreción y genes relacionados con la respuesta al estrés en RE, considerando un  $\log_2(\text{Fold\_change}) \leq -1.0$  ó  $\geq 1.0$ , de acuerdo a la metodología descrita en el punto 7.5.3.3 (CM/C0, CP/C0 y CM/CP).

#### **7.5.4.3.1 Análisis de expresión diferencial de los genes en la vía de asimilación del metanol (vía MUT)**

En la vía del metabolismo del metanol están involucrados 13 genes (De Schutter *et al.*, 2009). *AOX* no se tomo en cuenta debido a que la cepa utilizada fue KM71 Mut<sup>s</sup> (*HIS4 ARG4 AOX1Δ::SCARG4*) haciendolo no funcional e irrelevante para el análisis, el gen *DAK2* no se identificó durante el análisis y además, se agregó el gen *TKL1*, debido a que éste gen parece ser importante en la vía de asimilación del metanol al tener un dominio estructural altamente conservado junto con las enzimas *DAS1* y *DAS2* (Krainer *et al.*, 2012).

De éstos genes evaluados, 7 pertenecen a la vía de asimilación del metanol, 3 a la vía de desasimilación, y 2 genes clave para el inicio del metabolismo del metanol, siendo un total de 12 genes. La cuantificación para la interpretación de la expresión diferencial de estos genes se realizó de acuerdo a la metodología descrita (punto 7.5.3.3) en las tres comparaciones evaluadas (CM/C0, CP/C0 y CM/CP).

Además se calculó la relación *DAS/FLD* para ver si estaba más activa la vía de asimilación o la de desasimilación en ambas condiciones de cultivo. Esto se realizó dividiendo los valores FPKM de CM/C0, CP/C0 y CM/CP de los genes *DAS* y *FLD*, posteriormente se calculó el cociente de *DAS* total (*DAS1* + *DAS2*) entre *FLD*, al resultado se le calculó el  $\log_2$  y se compararon.

#### **7.5.4.3.2 Análisis de expresión diferencial de los genes en la vía de secreción**

En el proceso de secreción están involucrados 98 genes (De Schutter *et al.*, 2009) de los cuáles 10 no se encontraron en ninguna comparación. Además se tomaron en cuenta para éste análisis 22 genes provenientes de la base de datos KEGG y otros

artículos; siendo un total de 110 genes analizados en las tres condiciones de cultivo evaluadas (CM/CP, CP/C0, CM/CP). La cuantificación para la interpretación de la expresión diferencial de estos genes se realizó de acuerdo a la metodología descrita (7.5.3.3) en las tres comparaciones evaluadas (CM/C0, CP/C0 y CM/CP).

#### **7.5.4.3.3 Análisis de expresión diferencial de genes involucrados en el estrés en RE durante la secreción**

Se evaluaron 44 genes correspondientes a SNARE, ERAD, Autofagia, UPR, genes en respuesta a diferentes tipos de estrés ambiental y otros genes que pudieran impactar en la secreción. La cuantificación para la interpretación de la expresión diferencial de estos genes se realizó de acuerdo a la metodología descrita (7.5.3.3) en las tres comparaciones evaluadas (CM/C0, CP/C0 y CM/CP).

### **7.6 Validación del análisis transcripcional**

#### **7.6.1 Tratamiento con DNasa**

Se extrajo RNA, se cuantificó y se verificó la integridad de las muestras por triplicado de la misma forma que se describió en el punto 7.5.1 de cada una de las condiciones de cultivo evaluadas. A partir de estas preparaciones de RNA se realizó un tratamiento con una enzima DNasa, para eliminar cualquier contaminación con DNA; se empleó el Kit de Promega RQ1 RNase-Free DNase, se trabajó en un ambiente limpio y libre de nucleasas, se empleó material nuevo y libre de RNasas.

Primero se mezclaron 24  $\mu$ L de la muestra de RNA a tratar, 2  $\mu$ L de Buffer de reacción y 2  $\mu$ L de enzima RQ1 RNase-Free DNase en tubos de 0.2 mL. La mezcla se incubó durante 1 h a 37°C. Una vez terminada la incubación, se agregaron 2  $\mu$ L de solución de paro de DNasa y se incubó durante 15 min a 65°C.

#### **7.6.2 Protocolo de Síntesis de DNAc**

Para la síntesis del DNAc se utilizó el kit de Agilent AffinityScript Multiple Temperature DNAc Synthesis, se siguió el protocolo para la síntesis de DNAc de una

cadena, se trabajó en un ambiente limpio y libre de nucleasas, se empleó material nuevo y libre de RNAsas. A partir de las muestras de RNA tratadas con DNasa, se realizaron ensayos de retro-transcripción por triplicado para cada una de las muestras.

Primero se mezclaron los siguientes reactivos para una preparación: 150 ng totales de RNA, 1  $\mu\text{L}$  de oligo (dT) (0.5  $\mu\text{g}/\mu\text{L}$ ) y agua libre de RNAsas para obtener un volumen final de 15.7  $\mu\text{L}$ . La mezcla se incubó durante 5 min a 65°C. Se dejó enfriar durante 10 min a temperatura ambiente antes de proseguir con el ensayo. Se realizó una segunda mezcla y se añadió a la primera con los siguientes reactivos para una preparación: 2.0  $\mu\text{L}$  de buffer de reacción, 0.8  $\mu\text{L}$  de mix de dNTP's (25 mM) y 0.5  $\mu\text{L}$  de inhibidor de ribonucleasas (40 U/ $\mu\text{L}$ ). Por último, se agregó 1  $\mu\text{L}$  de enzima transcriptasa reversa AffinityScript Multiple Temperature. Los tubos se incubaron durante 1 h a 42°C y durante 15 min a 70°C. Por cada ensayo de síntesis de DNAc se realizó a la par un ensayo como control negativo de la retro-transcripción siguiendo los pasos descritos previamente, evitando añadir los volúmenes de enzima transcriptasa reversa.

Una vez obtenidas las muestras de DNAc, se realizaron diluciones a 1:25 empleando agua libre de nucleasas en un volumen final de 50  $\mu\text{L}$ . El material se etiquetó y se almacenó a -70°C.

### 7.6.3 Selección y Diseño de oligonucleótidos

Para el diseño de oligonucleótidos específicos, primeramente se seleccionaron 13 genes diferencialmente expresados más representativos de la ruta de secreción, proteólisis de proteínas y metabolismo del metanol (*MPG1*, *PMT2*, *HTMI*, *SSO2*, *SSA2*, *AHP1*, *RSP5*, *E3*, *WSC4*, *NCE103*, *FBA2*, *FGH* y *CAT*) en *P. pastoris* a partir del análisis del transcriptoma. Se utilizaron también genes de expresión constitutiva (*YPT1* y *G6PD*) previamente diseñados en el laboratorio para ser utilizados como genes normalizadores, además del gen *KAR2* que se utilizó como testigo para comparar los resultados con los obtenidos previamente (Félix-Castro; 2013).

Se realizó una búsqueda de las secuencias nucleotídicas de los genes seleccionados en la base de datos de GenBank (<http://www.ncbi.nlm.nih.gov>).

Las secuencias obtenidas fueron utilizadas para el diseño de oligonucleótidos específicos empleando la herramienta primer3 v0.4.0 (<http://bioinfo.ut.ee/primer3-0.4.0/>), el cual permite diseñar oligonucleótidos para ensayos de qPCR. Como información de entrada en el programa se emplearon las secuencias obtenidas previamente en GenBank para obtener las secuencias de los oligonucleótidos específicos 3' y 5' a usar como iniciadores de PCR para cada uno de los genes a analizar, se utilizaron los parámetros sugeridos para el diseño de oligonucleótidos para RT-qPCR basado en SYBR<sup>®</sup> Green (Thornton & Basu, 2011); donde los únicos parámetros que variaron entre cada par de oligonucleótidos diseñados fueron: “Max Self Complementary”, “Max 3' Self-Complimentary” y “Max #”.

Posteriormente, las secuencias de los oligonucleótidos diseñados fueron evaluadas para determinar su especificidad sobre el DNA blanco y la formación de estructuras secundarias.

Para evaluar la especificidad, se realizaron alineamientos con la herramienta Blastn, accesible por internet (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>), empleando como información de entrada las secuencias 3' y 5' de cada uno de los oligonucleótidos diseñados, contra el genoma de *P. pastoris*. Para determinar la especificidad de los oligonucleótidos se tomaron en cuenta dos criterios: hibridación de ambos oligonucleótidos en la misma secuencia, es decir, uno en la secuencia directa y otro en la secuencia complementaria, que los oligonucleótidos no hibridaran en otra parte del genoma y que tuvieran un valor ( $e \leq 0.099$ ).

Para evaluar la formación de estructuras secundarias, se realizaron análisis empleando la herramienta OligoAnalyzer 3.1 (<http://www.idtdna.com/analyzer/applications/oligoanalyzer/>). Este programa permite analizar la formación de estructuras secundarias de tipo horquilla, homodímeros y heterodímeros, entre otras propiedades. Como información de entrada se emplearon las secuencias 3' y 5' de cada uno de los oligonucleótidos diseñados y se evaluó el valor de  $\Delta G$  (kcal/mol). Para determinar que no hubiera formación de estructuras secundarias se evaluó que el valor de  $\Delta G$  no fuera menor a -10 kcal/mol, tomando en cuenta las siguientes concentraciones: oligonucleótido 3.3  $\mu\text{M}$ , Na 100 mM, Mg 5.5 mM y dNTPs

1 mM. Una vez que se evaluaron las propiedades de los oligonucleótidos diseñados, estos se sintetizaron en la compañía Integrated DNA Technologies, Inc (Coralville, Iowa).

#### 7.6.4 Descripción de las mezclas de trabajo utilizadas para la estandarización de los ensayos de qPCR mediante SYBR Green

Se emplearon dos muestras, el DNAc en dilución 1:25 (la dilución se realizó con agua MilliQ libre de nucleasas) y el control negativo de la síntesis de DNAc (sin enzima retro-transcriptasa). También se realizaron mezclas de qPCR sin DNAc, como control negativo del ensayo de qPCR.

Los diferentes DNAc empleados se etiquetaron para su identificación: símbolo (+) para las muestras con DNAc (reacción positiva de retro-transcripción) y símbolo RT (-) para las muestras del DNAc sin retrotranscriptasa (reacción negativa de la retrotranscripción) y PCR (-) para el control negativo del ensayo de qPCR.

Primero se realizó una mezcla de reacción con los reactivos mostrados en la tabla 10 para una preparación; esta mezcla se utilizó para cada reacción correspondiente a cada uno de los genes evaluados.

**Tabla 10.** Mezcla de trabajo utilizada en el ensayo de qPCR con SYBR Green de todos los DNAc (dilución 1:25).

Reactivos	Volúmen 1X (µL)
Brilliant III Ultra-Fast SYBR Green	10
DNAc dilución 1:25 (+)	2
Oligonucleótido 5'	2
Oligonucleótido 3'	2
Agua libre de nucleasas	4

Posteriormente se realizó una segunda mezcla de reacción para analizar las muestras de control negativo de la retro-transcripción, para una preparación se utilizaron los mismos reactivos con la mitad del volumen (Tabla 11).

**Tabla 11.** Mezcla de trabajo utilizada en el ensayo de qPCR con SYBR Green de los controles negativos de la qPCR.

<b>Reactivos</b>	<b>Volúmen 1X (µL)</b>
Brilliant III Ultra-Fast SYBR Green	5
DNAc RT (-)	1
Oligonucleótido 5'	1
Oligonucleótido 3'	1
Agua libre de nucleasas	2

Además se realizó una tercera mezcla de reacción para analizar las muestras de control negativo de la qPCR, para una preparación se utilizaron los mismos reactivos con la mitad del volumen sin DNAc (Tabla 12).

**Tabla 12.** Mezcla de trabajo para el ensayo de qPCR con SYBR Green de los controles negativos de la qPCR.

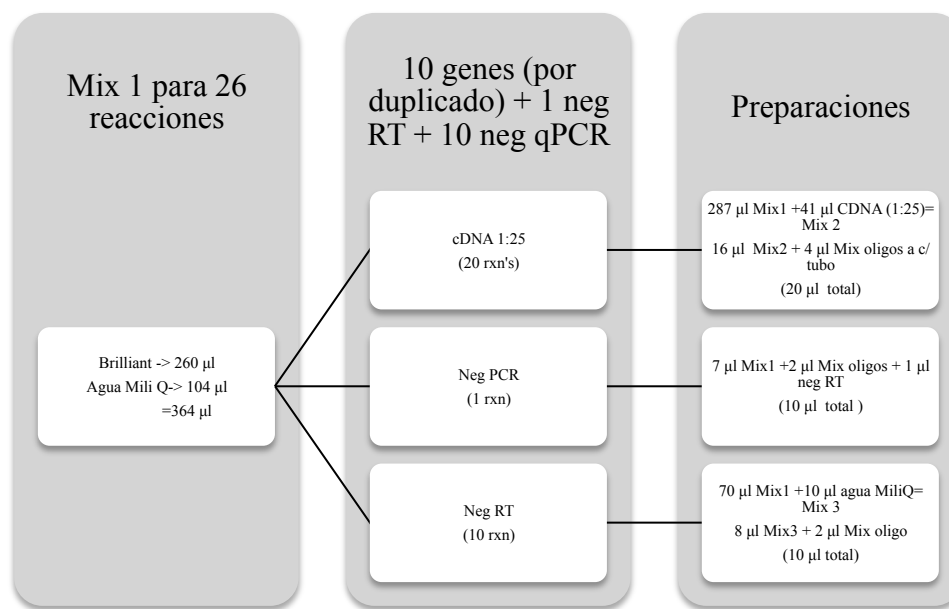
<b>Reactivos</b>	<b>Volúmen 1X (µL)</b>
Brilliant III Ultra-Fast SYBR Green	5
Oligonucleótido 5'	1
Oligonucleótido 3'	1
Agua libre de nucleasas	2

En la figura 16 se muestra un ejemplo de la preparación de mix para 10 genes por duplicado de una sola muestra.

Todas las reacciones se colocaron en tubos ópticos especiales de 8X cerrados con tiras de tapas ópticas especiales de 8X, además todos los tubos con las mezclas realizadas se colocaron en una minicentrífuga durante 10 segundos y se pasaron a la placa del equipo Mx3005P QPCR System de Agilent Technologies.

Las condiciones para cada ciclo de qPCR se ajustaron en base a los oligonucleótidos que se diseñaron, se seleccionaron las condiciones del termociclador mostradas en la tabla 13 para todos los genes. Además se agregó un tercer segmento que corresponde a las curvas de disociación, las cuáles sirven para diferenciar amplicones específicos de no-específicos basándose en la Temperatura de “melting” ó curva de disociación.





**Figura 16.** Estrategia para la preparación del mix para 10 genes por duplicado de una sola muestra utilizado en los ensayos de qPCR.

Cuando se trabajan con mas de 6 genes es preferible añadir media reacción más para los qPCR 1:25 para evitar que algun tubo no se compete con la reacción (20 µl totales). Se agrega un solo Mix de primer (mezcla del 5' y 3') correspondiente a cada gen por cada tubo de reacción por duplicado, y en los neg de qPCR (reacción a la mitad del volumen) no es necesario que se agreguen duplicados.

El gen *SSA2*, se ajustó una temperatura de 65° C en el último paso del segundo bloque de la etapa de qPCR correspondiente a la temperatura de alineamiento y extensión, debido a que con la temperatura de 60° C mostró amplicones no específicos.

**Tabla 13.** Condiciones empleadas en el termociclador para los ensayos de qPCR de los genes seleccionados.

Etapas	Segmento	Ciclo	Duración del ciclo	Temperatura	Pasos de la qPCR
qPCR	1	1	3 min	95°C	Incubación
	2	40	20 s	95°C	Desnaturalización
			20 s	60°C*	Alineamiento, hibridación y Extensión
Curva de disociación	3	1	1 min	95°C	Desnaturalización de productos amplificados
			30 s	55°C	Alineamiento
			30 s	95°C	Incremento gradual hasta llegar a 95°C

\* El gen *SSA2* utilizó una temperatura de 65°C en alineamiento, hibridación y extensión

#### **7.6.4.1 Curvas estándar para la corrección de la eficiencia en los ensayos de qPCR para cada uno de los genes seleccionados**

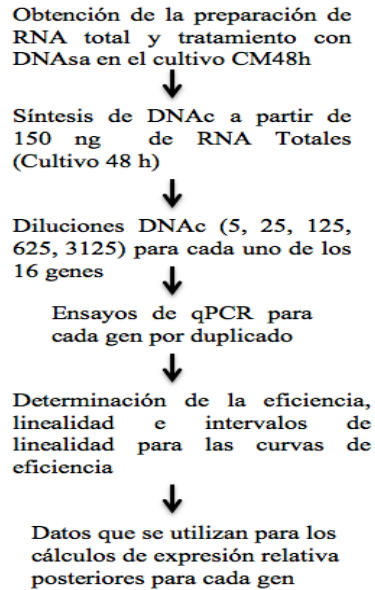
Se validó empíricamente la especificidad de los oligonucleótidos mediante el análisis de las curvas de disociación generadas por el equipo (segmento 3 mencionado en la tabla 13), para comprobar la amplificación de un sólo producto generado por la qPCR observable como un único pico en el gráfico.

Para determinar la eficiencia de amplificación (E) de cada DNAc se realizaron ensayos de qPCR de cinco diluciones seriadas del DNAc (1:5, 1:25, 1:125, 1:625 y 1:3125) para cada uno de los genes a excepción de los genes *SSA2*, *RSP5* y *E3*, cuyas diluciones fueron (DNAc directo, 1:5, 1:25, 1:125 y 1:625). En todos los casos las diluciones se prepararon con agua MilliQ libre de nucleasas por duplicado partiendo del DNAc directo, el cuál se sintetizó a partir de 150 ng de RNA total tratado con DNAsas de las células del cultivo CM a las 48 h de inducción con metanol (Figura 17).

Los ensayos de qPCR se realizaron siguiendo los lineamientos establecidos en la previa estandarización (7.6.4) para cada uno de los genes en estudio empleando los oligonucleótidos diseñados.

Para cada dilución se estimó el valor de  $C_T$  y se ajustó el valor del umbral de fluorescencia. El valor  $C_T$  se graficó en el eje de las ordenadas y en el eje de las abscisas la cantidad equivalente de RNA. A partir de la curva obtenida se determinaron la eficiencia, linealidad e intervalos de linealidad para los genes en estudio en los ensayos de qPCR. Se consideró un valor de  $r^2$  más cercano a uno y los intervalos limitantes de la curva de dilución fueron considerados como los valores de  $C_T$  aceptables para calcular los niveles de expresión de cada gen.

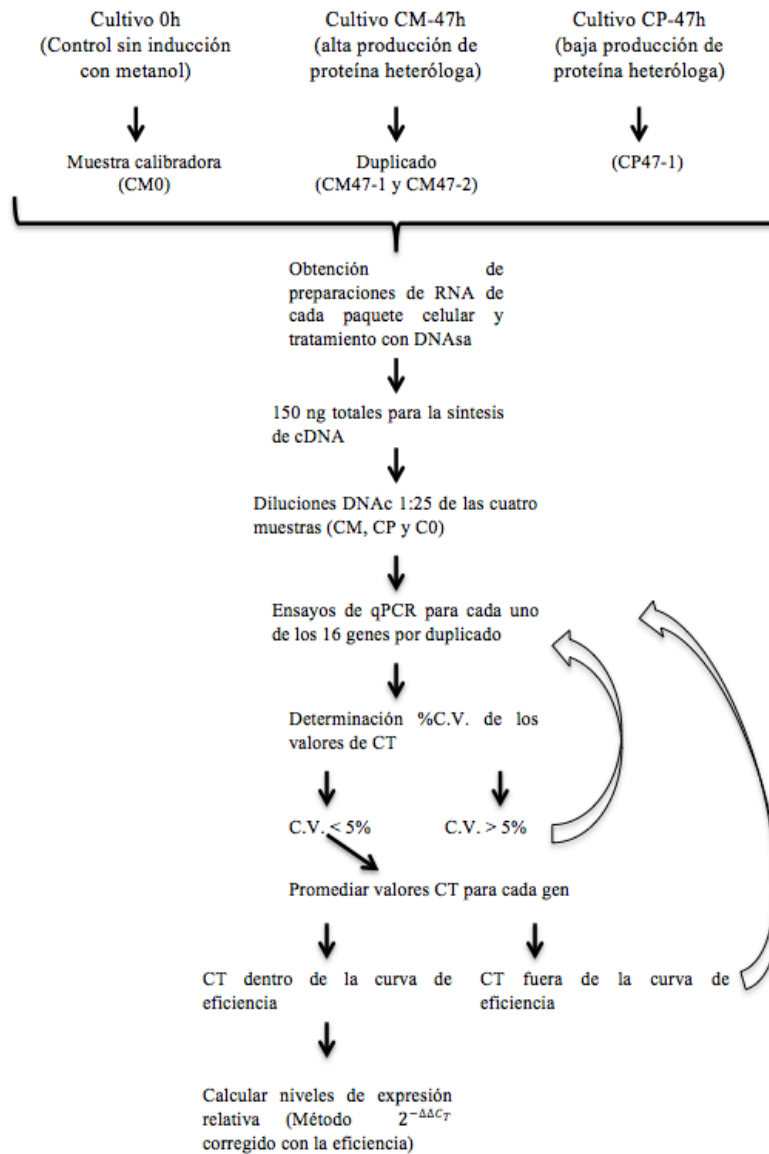
La eficiencia de amplificación (E) se determinó a partir de la pendiente de la curva en el intervalo lineal, considerando como criterio de aceptación un valor superior al 90% y menor del 110%. Los valores de todos los parámetros ( $r^2$ , E, y valores de  $C_T$  que delimitan la curva de dilución) fueron estimados de forma automática por el programa del equipo. Por último estos valores se registraron para los cálculos posteriores de expresión relativa para cada gen.



**Figura 17.** Estrategia para la generación de la curva estándar para la corrección de la eficiencia en el análisis de expresión relativa de cada gen en el cultivo CM48h.

#### 7.6.4.2 Ensayos de RT-qPCR para la determinación de los niveles de expresión de cada gen

Una vez generadas las curvas de eficiencia, el análisis de los niveles de expresión de cada gen se realizó a partir del RNA de las mismas muestras enviadas a secuenciar; es decir, por duplicado biológico para la condición CM (CM47-1, CM47-2): de alta producción de proteína heteróloga; una única muestra CP (CP47-1): de baja producción de proteína heteróloga; y C0 (C0-1): muestra calibradora s/inducción con metanol; para cada uno de los genes evaluados. El DNAc correspondiente a cada condición de cultivo evaluada se amplificó por duplicado siguiendo las mezclas previamente descritas (7.6.4) y se determinó el valor de CT, la media, la desviación estándar y el coeficiente de variación. Se tomó en cuenta un valor de  $(C.V \leq 5\%)$ , se promediaron los valores de CT obtenidos y se verificó que el valor promedio de CT de cada gen se encontrara dentro de su curva de eficiencia correspondiente. Este proceso se realizó para cada uno de los genes evaluados (*MPG1*, *PMT2*, *HTM1*, *SSO2*, *SSA2*, *AHP1*, *RSP5*, *E3*, *WSC4*, *NCE103*, *FBA2*, *FGH* y *CAT*) y para los genes de referencia empleados como normalizadores (*G6PD* y *YPT1*), los cuáles ya se tenían sintetizados, para determinar la variación global en los niveles de expresión (Figura 18).



**Figura 18.** Estrategia para el cálculo de expresión relativa de cada gen.

#### 7.6.4.3 Cálculo de expresión relativa para cada uno de los genes

Para realizar el cálculo de la expresión relativa de los genes en estudio, se tomó el método de la  $2^{-\Delta\Delta C_T}$  (Livak & Schmittgen, 2001), el cual calcula la cantidad de blanco, normalizado a un control endógeno y relativo a un calibrador. Se empleó la fórmula corregida de la  $2^{-\Delta\Delta C_T} 2^{-\Delta\Delta C_T}$  para incluir en el cálculo los valores de las eficiencias de amplificación obtenidas para cada gen y los factores de dilución de los DNAc correspondientes a las curvas de eficiencia en los ensayos de qPCR (ecuación 1).

$$\frac{X_{N,q}}{X_{N,cb}} = \frac{FD_{X,q} (1 + E_X)^{-(C_{T,X,q})}}{FD_{X,cb} (1 + E_X)^{-(C_{T,X,cb})}} \div \frac{FD_{R,q} (1 + E_R)^{-(C_{T,R,q})}}{FD_{R,cb} (1 + E_R)^{-(C_{T,R,cb})}}$$

**Ecuación 1.** Cálculo de la expresión relativa, incluye eficiencias de amplificación y factores de dilución.

De esta manera se calcularon los niveles de expresión relativa de los 14 genes en estudio empleando 2 genes normalizadores y la muestra de 0 h de inducción con metanol como calibradora para la validación del análisis transcripcional. Los cálculos se realizaron para las dos condiciones de cultivo (CM y CP de alta y baja producción de la proteína heteróloga) a evaluar calculando la media de los niveles de expresión de los replicados únicamente en la condición CM. Se compararon los niveles de expresión en las dos condiciones de cultivo cada una relativa a los niveles de expresión a las 0 h de inducción con metanol para ambas y se determinaron los niveles de expresión relativa en la condición CM y CP.

#### **7.6.5 Comparación de las técnicas de análisis de expresión diferencial de genes (RNA-seq vs RT-qPCR)**

Se promediaron los valores correspondientes a los niveles de expresión relativa en la comparación CM, se calculó el cociente de los valores de expresión relativa en las condiciones CM con respecto a CP y se calculo el  $\log_2$  del cociente.

Se graficaron los valores  $\log_2$  (Fold\_change) correspondientes a los niveles de expresión del análisis transcripcional en el eje de las ordenadas y en el eje de las abscisas los valores  $\log_2$  correspondientes a los niveles de expresión relativa obtenidos por el método de la  $2^{-\Delta\Delta C_T}$  modificado, posteriormente se analizó el valor de la pendiente y  $r^2$ .

## 8. RESULTADOS

### 8.1 Obtención de preparaciones de RNA a partir de células de *P. pastoris* KM71FTEII cultivadas en biorreactor para la secuenciación de transcritos

Los resultados de la extracción de RNA de los paquetes celulares de *P. pastoris* KM71FTEII cultivadas en biorreactor a dos condiciones de cultivo (0 y 47 h de inducción con metanol) conservados en RNAlater y obtenidos en estudios previos en el laboratorio, tuvieron una concentración comprendida entre 65.1 y 185.8 ng/ $\mu$ L, con valores de absorbancias comprendidos entre 1.83 a 2.10 de la relación 280/280 nm (Tabla 14).

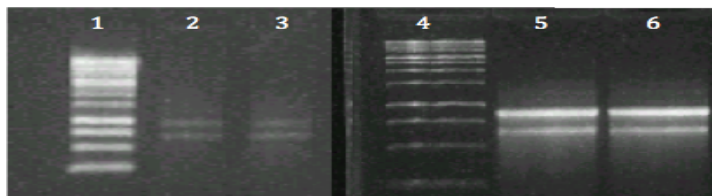
**Tabla 14.** Concentración de RNA y relación A260/A280 en las preparaciones de RNA que provienen de células de *P. pastoris* KM71FTEII cultivadas en biorreactor.

Muestra	Concentración (ng/ $\mu$ L)	A280/A260
C0	69.7	1.83
CM47-1	159.4	2.10
CM47-2	185.8	2.01
CP47-1	65.1	1.99

C0: células de tiempo cero de inducción con metanol; CM47-1 y CM47-2: replicados biológicos de células de 47 h de inducción con metanol a las condiciones de alta producción; CP47-1: células de 47 h de inducción con metanol a las condiciones de baja producción.

Los geles de agarosa de todas las preparaciones de RNA (Figura 19) revelaron dos bandas predominantes correspondientes a las fracciones 26S (banda superior) y 18S de RNA ribosomal (banda inferior), lo que indica una buena integridad en todas las preparaciones realizadas, también se puede visualizar un gel de electroforesis digital obtenido con el bionalizador Agilent 2100 (Figura 20). Los electroferogramas de todas las preparaciones de RNA realizado con el Bioanalizador Agilent 2100 fueron muy similares mostrando dos picos mayoritarios correspondientes a fracciones 18S (retención a 40 s) y 26S (retención a 45 s) de RNA ribosomal, además de un pico al inicio del electroferograma correspondiente a un buffer que contiene dos marcadores moleculares (upper y lower marker) para la identificación del tamaño de los picos (Figura 21). Los

valores del número de integridad de RNA (RIN) obtenidos a partir de los electroferogramas comprendieron de 7.8 a 8.5 (Tabla 15).



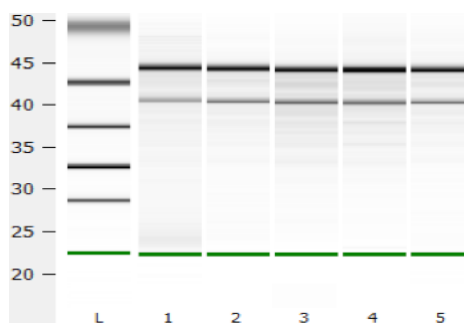
**Figura 19.** Gel de agarosa al 0.8% de las preparaciones de RNA.

Carril 1: marcador de tamaño molecular; carril 2: C0 células sin inducción con metanol; carril 3: CP47-1 células a las 47 h de inducción con metanol a las condiciones de baja producción; carril 4: marcador de tamaño molecular; carril 5: CM47-1 y carril 6: CM47-2 replicados biológicos de las células a las 47 h de inducción con metanol a las condiciones de alta producción.

**Tabla 15.** Número de integridad de RNA (RIN) de la preparaciones de RNA que provienen de las células de *P. pastoris* KM71FTEII cultivadas en biorreactor.

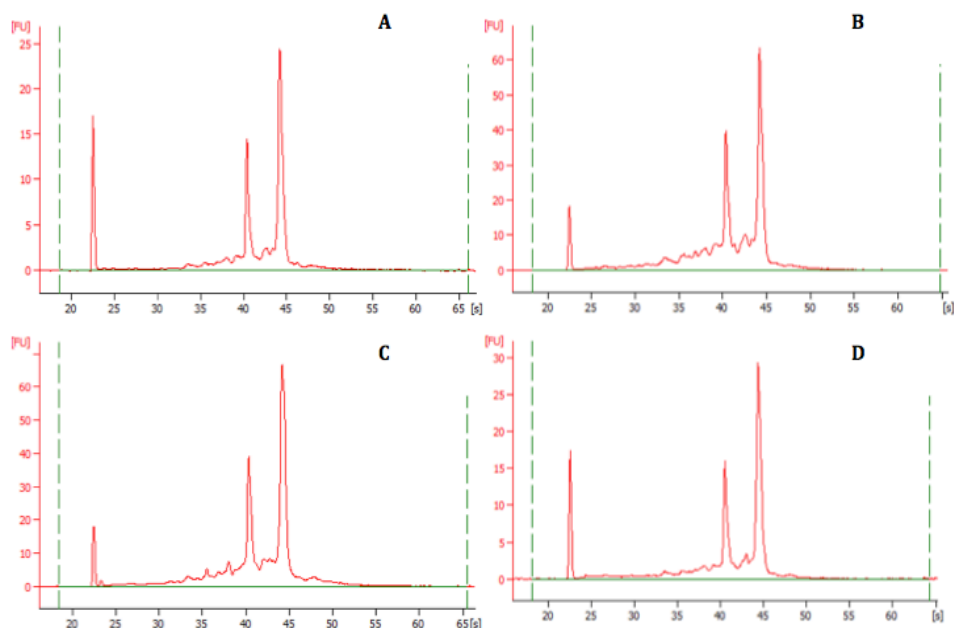
Muestra	Número de integridad de RNA (RIN)
C0	8.5
CM47-1	7.8
CM47-2	8.4
CP47-1	8.3

C0: células de tiempo cero de inducción con metanol; CM47-1 y CM47-2: replicados biológicos de células de 47 h de inducción con metanol a las condiciones de alta producción; CP47-1: células de 47 h de inducción con metanol a las condiciones de baja producción.



**Figura 20.** Gel de electroforesis digital de las preparaciones de RNA de las muestras obtenido por el Bioanalizador Agilent 2100.

Carril 1: Marcador molecular; carril 2: C0 células de tiempo cero de inducción con metanol; carril 3: CP47-1 células a las 47 h de inducción con metanol a las condiciones de baja producción; carril 4: CM47-1 y carril 5: CM47-2 replicados biológicos de las células a las 47 h de inducción con metanol a las condiciones de alta producción.



**Figura 21.** Electroferogramas de las preparaciones de RNA.

A: C0 células sin inducción con metanol; B-C: CM47-1 y CM47-2, replicados biológicos de las células a las 47 h de inducción con metanol a las condiciones de alta producción; D: CP47-1 células a las 47 h de inducción con metanol a las condiciones de baja producción.

### 8.1.1 Obtención de las librerías de DNAc y secuenciación de transcritos

Como resultado de la secuenciación de los cDNA's sintetizados a partir de los transcritos (RNA-seq), se obtuvieron un total de ocho archivos en formato FASTQ, dos por cada librería correspondiente a la secuenciación a partir del extremo 5' y 3' de la cadena de DNAc. En la tabla 16 se muestra un resumen de las lecturas y la calidad realizadas por el secuenciador de las cuatro librerías de DNAc obtenidas.

En todos los casos, las lecturas de las dos cadenas fueron iguales, sin embargo la calidad de las lecturas de la cadena 5' (R1) fue superior a la calidad de la cadena 3' (R2). El total de lecturas entre las dos cadenas de cada muestra comprendió de 2,806,730 a 3,141,558. En total se obtuvieron 11,943,162 de lecturas.

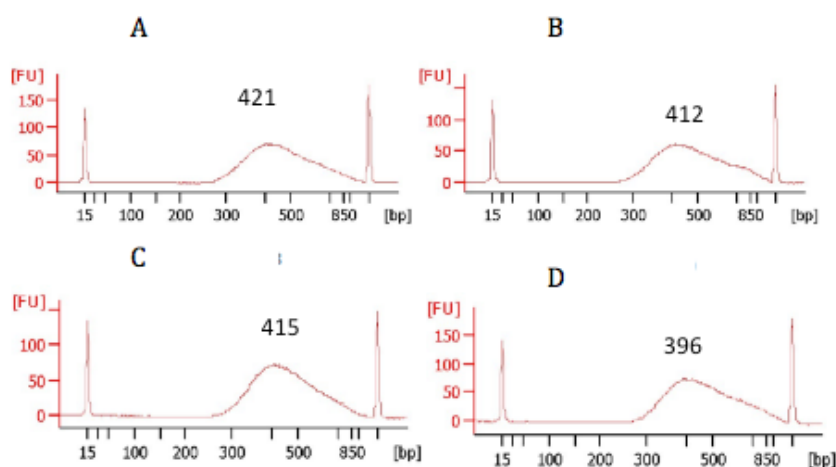
Todos los electroferogramas de las librerías de DNAc, obtenidas a partir de las cinco preparaciones de RNA, tuvieron una distribución similar de tamaños moleculares con un máximo próximo a los 400 pb (Figura 22).



**Tabla 16.** Resumen de las lecturas obtenidas por el secuenciador y la calidad de las cinco librerías de DNAc obtenidas.

Muestra	Clave LANGEBIO	R1	Calidad $\geq$ 30 (%)	R2	Calidad $\geq$ 30 (%)	Total
C0	MG1AT1SS01.10	1,403,250	96.2	1,403,250	90.9	2,806,500
CM47-1	MG1AT1SS02.08	1,553,156	96.2	1,553,156	90.9	3,106,312
CM47-2	MG1AT1SS03.03	1,443,677	96.2	1,443,677	90.9	2,887,354
CP47-1	MG1AT1SS05.20	1,567,171	96.2	1,567,171	90.9	3,134,342
<b>Total</b>						11,934,508

C0: células de tiempo cero de inducción con metanol, CM47-1 y CM47-2: replicados biológicos de células de 47 h de inducción con metanol a las condiciones de alta producción, CP47-1 células de 47 h de inducción con metanol a las condiciones de baja producción.



**Figura 22.** Electroferogramas de las librerías de DNAc obtenidas a partir de las cinco preparaciones de RNA.

A: C0 células de tiempo cero de inducción con metanol; B y C: CM47-1 y CM47-2 replicados biológicos de células de 47 h de inducción con metanol a las condiciones de alta producción extracelular; D: CP47-1 células de 47 h de inducción con metanol a las condiciones de baja producción.

## 8.2 Análisis Bioinformático de los transcritos secuenciados

### 8.2.1 Análisis de calidad de las secuencias

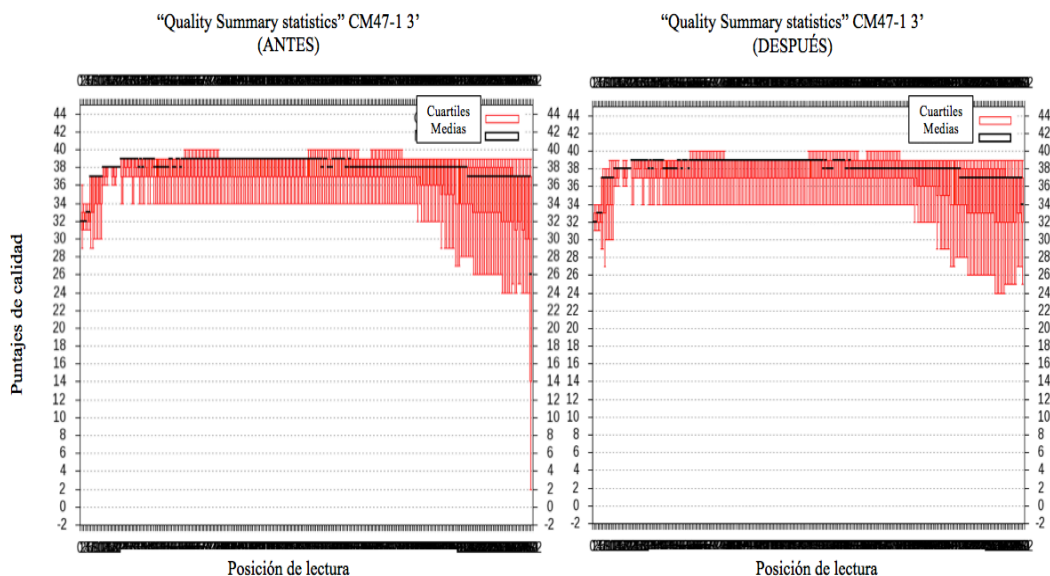
Después del recorte de secuencias, mediante el programa “FASTQ Quality Trimming” (considerando un valor Phred  $\geq$  20) se descartaron algunas lecturas de mala calidad sumando un total de 0.007 a 0.032 % de lecturas removidas (Tabla 17). A continuación se muestra un ejemplo de los gráficos obtenidos con la herramienta

“FASTQ Summary Statistics” (Figura 23) antes y después del recorte de las secuencias, con ayuda de éste grafico se verificaron las lecturas descartadas y se aseguró la ausencia de lecturas de mala calidad en todos los archivos FASTQ que contenían las lecturas de los transcritos secuenciados.

**Tabla 17.** Número de lecturas finales y removidas después del recorte de secuencias con el criterio de calidad de un valor de puntaje Phred  $\geq 20$ .

Muestra	Cadena	Lecturas finales	Lecturas removidas	Lecturas removidas (%)
C0	R1	1,402,806	444	0.032
	R2	1,402,814	436	0.031
CM47-1	R1	1,552,564	592	0.038
	R2	1,552,591	565	0.036
CM47-2	R1	1,443,418	259	0.018
	R2	1,443,431	246	0.017
CP47-1	R1	1,567,054	117	0.007
	R2	1,567,059	112	0.007

C0: lecturas correspondientes a las células de tiempo cero de inducción con metanol; CM47-1 y CM47-2: lecturas correspondientes a los replicados biológicos de células de 47 h de inducción con metanol a las condiciones de alta producción; CP47-1: lecturas correspondientes a las células de 47 h de inducción con metanol a las condiciones de baja producción.



**Figura 23.** Ejemplo de los gráficos obtenidos con la herramienta “FASTQ Summary Statistics” de las lecturas correspondientes a la muestra CM47-1 R2 (3’), antes y después del recorte de las secuencias.

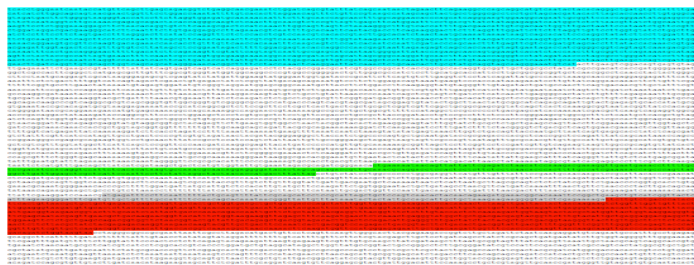
Lecturas correspondientes a las células CM47-1 3’, después de 47 h de inducción con metanol, antes (izquierda) y después (derecha) del recorte de secuencias con puntaje Phred  $\geq 20$  realizado con la herramienta FASTQ Summary Statistics. Los gráficos que corresponden al resto de las lecturas son similares.

### 8.2.2 Modificación del cromosoma antes del mapeo con el genoma de referencia

En los archivos FASTA y GTF originales del genoma y anotaciones, respectivamente, de *P. pastoris* GS115 (archivos descritos en el punto 7.5.3.2), se agregaron 9,064 nucleótidos *in silico* correspondientes a la secuencia del plásmido pPIC9fteII al final del cromosoma 1 desde el inicio del gen *HIS4*, el cual se muestra en la figura 24. Además se agregaron las anotaciones correspondientes al gen *FTEII* con su secuencia señal y una segunda copia del gen *HIS4* al archivo gtf que contiene las anotaciones del genoma de la cepa GS115 (tabla 18).

**Tabla 18.** Anotaciones que se añadieron al archivo gtf que contiene las anotaciones del genoma de la cepa GS115 para agregar las regiones incorporadas del plásmido pPIC9fteII.

Gen/segmento	Tamaño (pb)	Anotación	Región cromosómica
<i>HIS4</i>	2,031	Rec_ <i>His4</i> _pPIC9	2,878,649-2,880,679
Región plásmido	4,457	----	2,880,680-2,885,136
Secuencia pre-pro factor alpha	253	Rec_ <i>FTEII</i> _pPIC9	2,885,137-2,886,457
<i>FTEII</i>	1,068		
Región plásmido	1,255	----	2,886,458-2,887,712



**Figura 24.** Secuencia del plásmido que contiene el gen heterólogo de interés insertado *in silico* al final del cromosoma 1.

La primera parte de color azul se muestra la región correspondiente al gen *His4*, la segunda parte verde la región promotora 5' y la tercera parte roja el gen heterólogo así como su secuencia señal de gris.

### 8.2.3 Mapeo de los transcritos con un genoma de referencia

El mapeo realizado con los archivos FASTA y GTF modificados del genoma y anotaciones, respectivamente, de la cepa de *P. pastoris* GS115 (7.5.3.2) se presenta en la tabla 19, se puede visualizar el número de lecturas mapeadas con el programa TopHat2 y el porcentaje de lecturas mapeadas respecto a las lecturas iniciales. En todos los casos, el

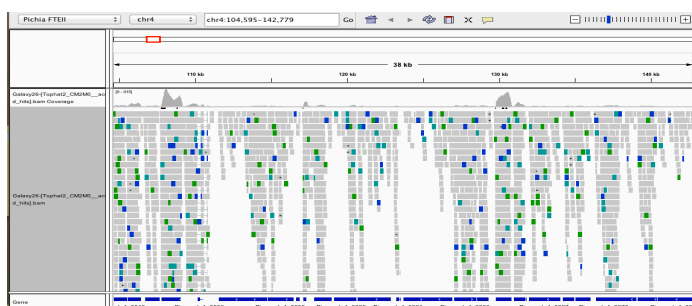
porcentaje de lecturas mapeadas fue mayor en las cadenas 5' (88-90%) que en las cadenas 3' (84-88%).

**Tabla 19.** Número y porcentaje de lecturas mapeadas en base al genoma modificado de referencia de la cepa GS115 de *P. pastoris*.

Muestra	R1		R2	
	Lecturas mapeadas	Porcentaje de lecturas mapeadas (%)	Lecturas mapeadas	Porcentaje de lecturas mapeadas (%)
C0	1266340	90	1233494	88
CM47-1	1376692	89	1300131	84
CM47-2	1283055	89	1243880	86
CP47-1	1378246	88	1341892	86

C0: Lecturas mapeadas correspondientes a las células de tiempo cero de inducción con metanol; CM47-1 y CM47-2: lecturas mapeadas correspondientes a los replicados biológicos de células de 47 h de inducción con metanol a las condiciones de alta producción ; CP47-1: lecturas mapeadas correspondientes a las células de 47 h de inducción con metanol a las condiciones de baja producción.

Los transcritos contenidos en los archivos FASTQ provenientes de la muestra C0 del cromosoma 4 mapeado en base al genoma de referencia GS115 modificado de *P. pastoris*, se visualizaron con el programa IGV (Figura 25). En todos los casos se observó una diferencia en el número de transcritos mapeados en las diferentes regiones del genoma. Todos los transcritos mapeados correspondientes a cada muestra secuenciada se observan de manera similar a la indicada en la figura.



**Figura 25.** Visualización con el programa IGV de los transcritos mapeados correspondientes a la muestra C0 del cromosoma 4 respecto al genoma de *P. pastoris*.

Los rectángulos azules de la parte inferior corresponden a los genes de referencia, mientras que los rectángulos grises son los alineamientos de transcritos en cada gen.

### 8.2.4 Análisis diferencial con cufflinks

Después del ensamblaje (Cufflinks), comparación a través de múltiples experimentos (Cuffcompare) y cuantificación de los transcritos (Cuffdiff) se obtuvo una tabla de datos fácilmente visualizables en Excel, mostrada en la figura 26. Esta tabla representa de manera independiente cada una de las comparaciones evaluadas, con los siguientes datos descritos en orden: 1) La identificación y clave de los genes mapeados. 2) Su posición el genoma de referencia. 3) La clave de la muestra 1 y 2. 5) El nivel de expresión en FPKM de cada una de las dos muestras. 6) El valor de  $\log_2$  (Fold\_change) de los valores FPKM de la muestra 2 entre los de la muestra 1. 7) El valor de P de la comparación entre las dos muestras. 8) El valor Q (valor P corregido), indicador real de la significancia de la comparación.

Todos los genes que resultaron sobre-expresados o sub-expresados según el valor de  $\log_2$  (Fold\_change) mostraron un valor de  $q < 0.005$ .

The table below is a screenshot of an Excel spreadsheet showing data from Cuffdiff. It has 13 columns labeled B through N. Annotations with arrows point to specific columns: 'Fpkm CP' points to column H, 'Fpkm CM' points to column I, 'Valor P' points to column L, and 'Log<sub>2</sub> (Fold\_change)' points to column J.

B	C	D	E	F	G	H	I	J	K	L	M	N
gene_id	gene	locus	sample_1	sample_2	status	value_1	value_2	log2(fold_ch.test_stat	p_value	q_value	significant	
XLOC_00008	PAS_chr1-3	chr1:306287	cp	cm	OK	94.5434	19.8335	-2.25304	-5.27393	5.00E-05	0.0012554	yes
XLOC_00011	PAS_chr1-3	chr1:394736	cp	cm	OK	452.504	70.4734	-7.68278	-7.20693	5.00E-05	0.0012554	yes
XLOC_00017	PAS_chr1-1	chr1:606302	cp	cm	OK	666.472	196.257	1.7638	-5.17077	5.00E-05	0.0012554	yes
XLOC_00030	PAS_chr1-1	chr1:111652	cp	cm	OK	34.809	93.2246	1.42125	3.47926	5.00E-05	0.0012554	yes
XLOC_00031	PAS_chr1-1	chr1:117810	cp	cm	OK	12.372	78.3603	2.66305	5.35593	5.00E-05	0.0012554	yes
XLOC_00032	PAS_chr1-1	chr1:119670	cp	cm	OK	19.7278	89.8131	2.18669	5.38968	5.00E-05	0.0012554	yes

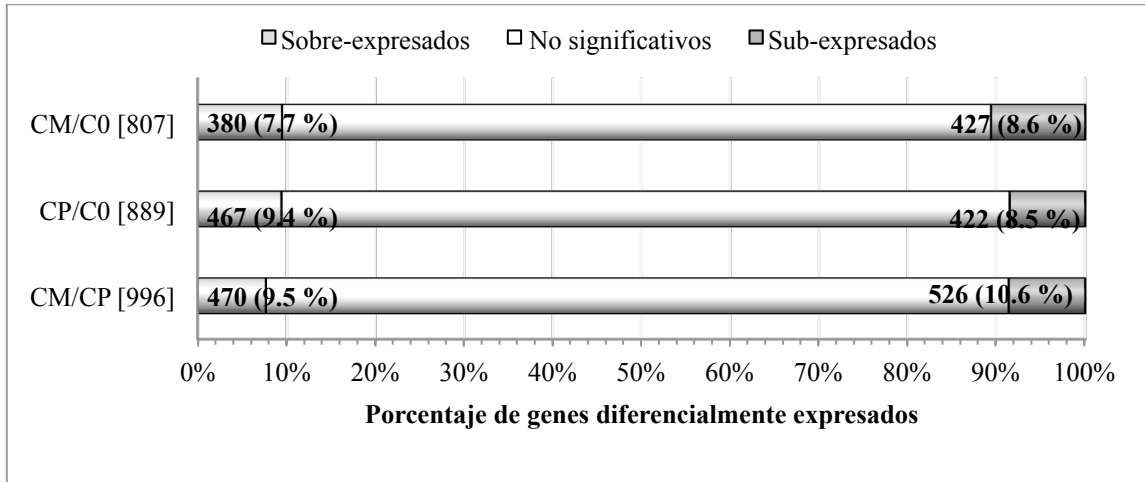
**Figura 26.** Datos generados con la herramienta Cuffdiff que corresponden a la cuantificación y normalización de los transcritos durante la prueba de expresión diferencial de CM respecto a CP visualizada en Excel.

### 8.2.5 Análisis global de expresión diferencial de genes

Los datos obtenidos por Cuffdiff (mencionado en 8.2.4) mostraron la identificación de 4,950 genes respecto a los 5,315 genes totales anotados del archivo GTF modificado (mencionado en 7.5.3.2) en las tres comparaciones CM/C0, CP/C0 y CM/CP, correspondiente al 93.1% del total.

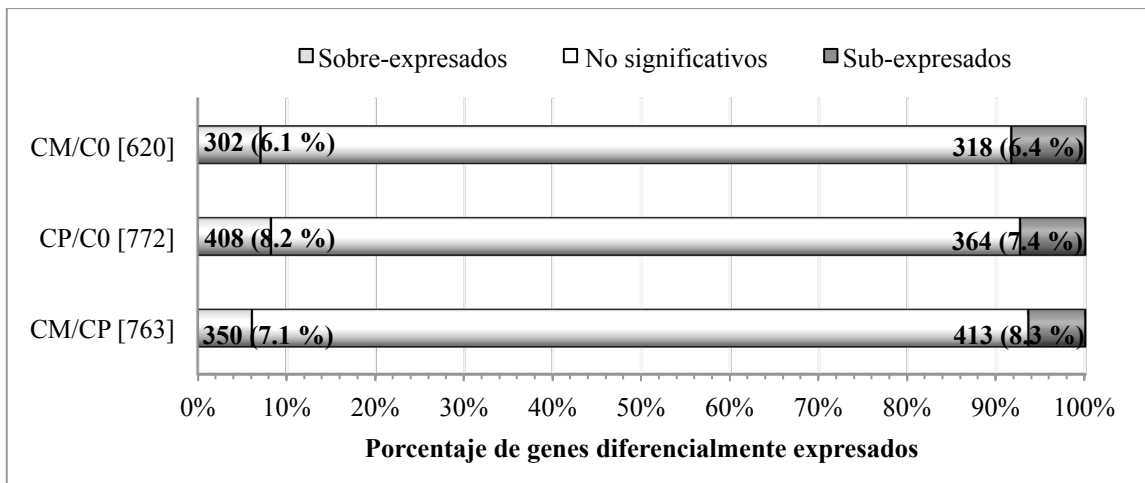
El número de genes sobre-expresados y sub-expresados encontrados en la prueba de expresión diferencial (Cuffdiff)  $q \leq 0.05$  y  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$  se

muestra en la figura 27; y considerando la condición estricta con valor  $q \leq 0.05$  y  $\log_2(\text{Fold\_change}) \leq -1.0$  y  $\geq 1.0$  se muestra en la figura 29.



**Figura 27.** Expresión global diferencial de genes encontrados con Cuffdiff, considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

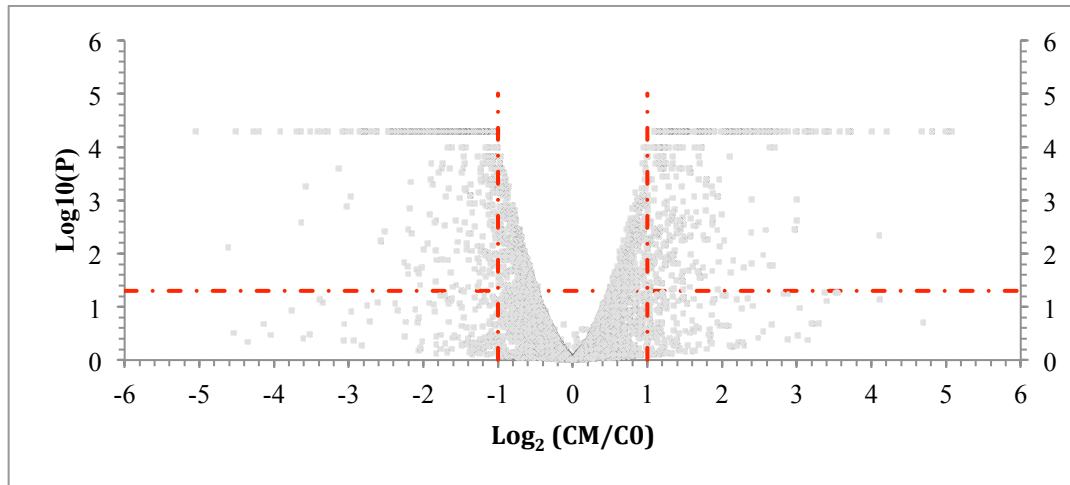
En el eje de las ordenadas se muestra el total de genes diferencialmente expresados correspondientes a cada comparación, además se muestra el número de genes sobre- (gris claro) y sub- (gris oscuro) expresados, seguido del porcentaje de genes sobre- y sub- expresados respecto al total de genes identificados.



**Figura 28.** Expresión global diferencial de genes considerando un  $\log_2(\text{Fold\_change}) \leq -1.0$  y  $\geq 1.0$ .

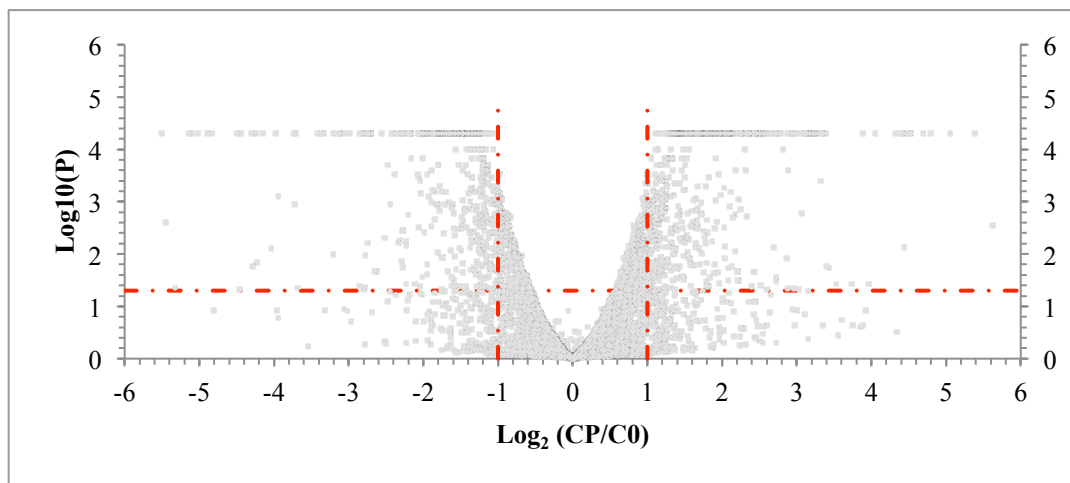
En el eje de las ordenadas se muestra el total de genes diferencialmente expresados correspondientes a cada comparación, además se muestra el número de genes sobre- (gris claro) y sub- (gris oscuro) expresados, seguido del porcentaje de genes sobre- y sub- expresados respecto al total de genes identificados.

Los gráficos tipo volcán que muestran el análisis global de la expresión diferencial de genes de las tres comparaciones (CM/C0, CP/C0 y CM/CP) totales considerando un valor  $q \leq 0.05$  y  $\log_2(\text{Fold\_change}) \leq -1.0$  y  $\geq 1.0$ , éstos gráficos se muestran en las figuras 29, 30 y 31, respectivamente.



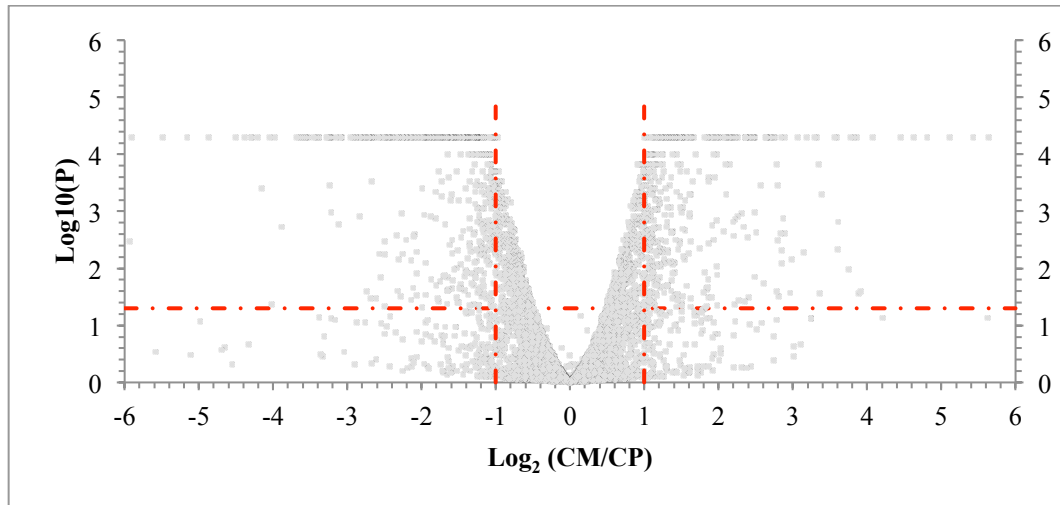
**Figura 29.** Análisis global de la expresión diferencial de genes en la comparación CM/C0 considerando un  $\log_2(\text{Fold\_change}) \geq 1.0$  y  $\leq -1.0$ .

Los puntos ubicados en los cuadrantes superior izquierda y derecha delimitados por las líneas (punto de corte) representan a genes sub-expresados (302, 6.1%) y sobre-expresados (318, 6.4%), respectivamente.



**Figura 30.** Análisis global de la expresión diferencial de genes en la comparación CP/C0 considerando un  $\log_2(\text{Fold\_change}) \geq 1.0$  y  $\leq -1.0$ .

Los puntos ubicados en los cuadrantes superior izquierda y derecha delimitados por las líneas (punto de corte) representan a genes sub-expresados (408, 8.2%) y sobre-expresados (364, 7.4%), respectivamente.



**Figura 31.** Análisis Global de la expresión diferencial de genes en la comparación CM/CP considerando un  $\log_2(\text{Fold\_change}) \geq 1.0$  y  $\leq -1.0$ .

Los puntos ubicados en los cuadrantes superior izquierda y derecha delimitados por las líneas (punto de corte) representan a genes sub-expresados (350, 7.1%) y sobre-expresados (413, 8.3%), respectivamente.

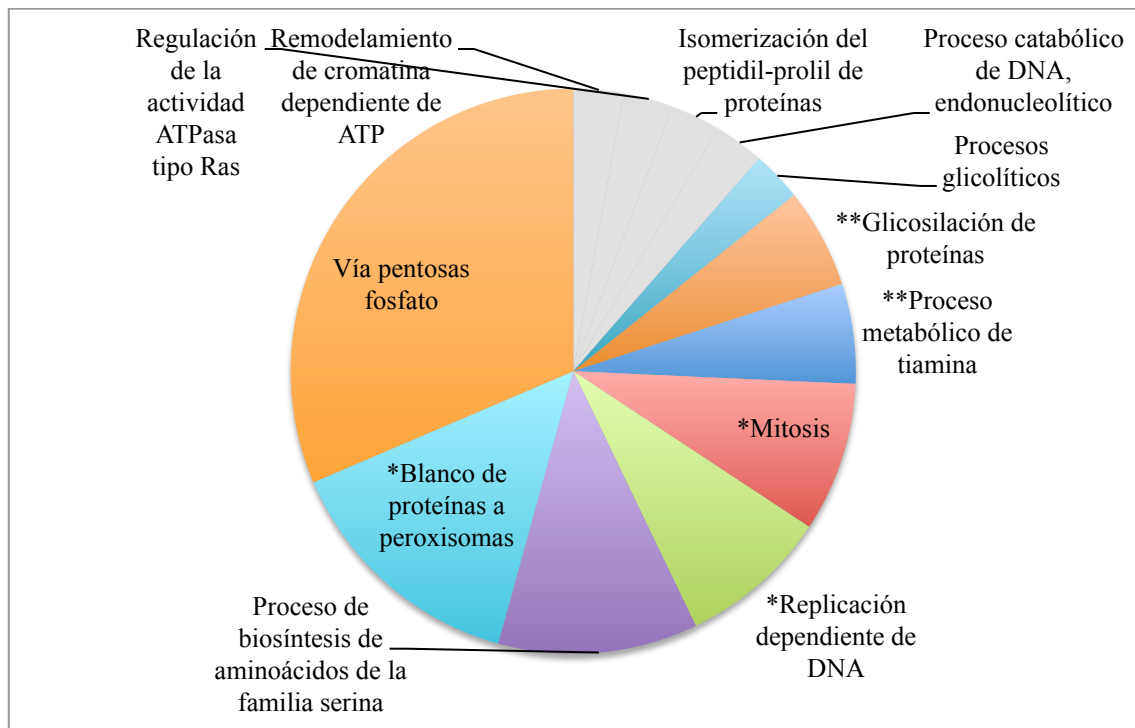
### 8.2.5.1 Análisis global de enriquecimiento de genes

En los gráficos de pastel comprendidos en ésta sección, cada sector engloba las categorías con al menos 2 genes y 10% de los genes de la categoría encontrada. El tamaño de cada sector es proporcional al número de términos agrupados, el nombre de cada sector corresponde al término más representativo de cada grupo. En el gráfico se indica el grado de significancia de la totalidad de cada agrupación (Valor P del grupo:  $p \leq 0.05$  (significativos) \* y  $p \leq 0.01$  (altamente significativos) \*\* corregidos con “Bonferroni step down”). De color gris se representan los términos no agrupados.

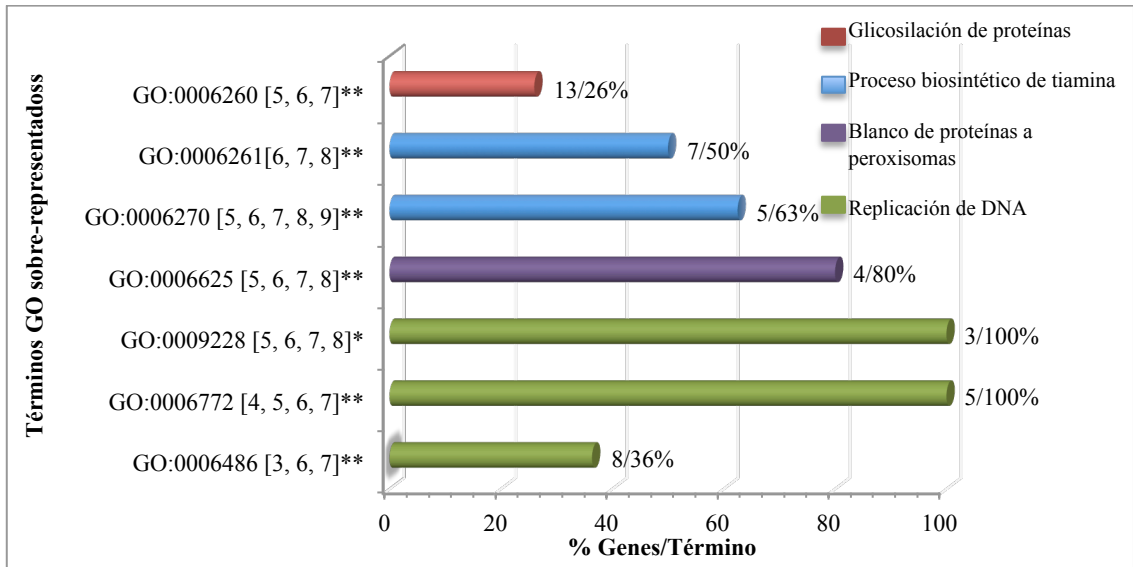
En cuanto a los gráficos de barras, en el eje de las ordenadas se encuentran los Términos GO seguido por los niveles GO entre corchetes y el grado de significancia ( $p \leq 0.05$  \* y  $p \leq 0.01$  \*\*, corregidos de igual manera con “Bonferroni step down”), éstos gráficos fueron creados en base a los términos con valores P más lejanos de 1, es decir, más significativos. Además a un lado de las barras se encuentra el número de genes encontrados y el porcentaje con respecto al total de genes en dicho término. El nombre de cada agrupación (color de las barras) está encabezado por el término más significativo.



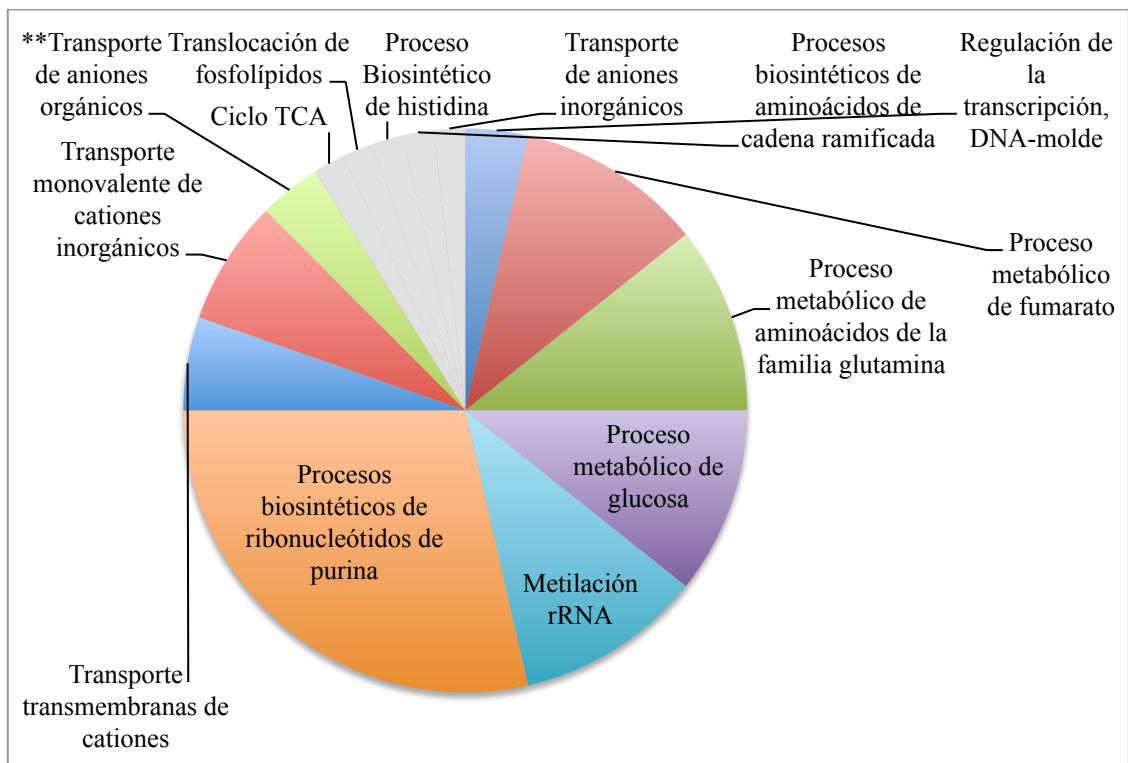
Del primer análisis de enriquecimiento de genes utilizando los genes sobre- y sub-expresados totales de cada comparación, considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$  y utilizando parámetros estrictos mencionados en el punto 7.5.4.1 se describen a continuación. En la comparación CM/C0 se formaron 8 grupos además de 4 términos no agrupados; en total se identificaron 36 términos y 25 conexiones entre algunos de ellos. La distribución de los grupos de términos sobre-representados de los genes sobre-expresados se muestra en la figura 32. Además se muestran los términos GO más importantes con el porcentaje de genes encontrados asociados a cada término en la comparación CM/C0 en la figura 33. También se formaron 9 grupos y 5 términos no agrupados, con un total de 56 términos identificados sobre-representados de los genes sub-expresados y 71 conexiones entre algunos de ellos, cuya distribución se muestra en la figura 34, mientras que los términos GO más importantes con el porcentaje de genes encontrados asociados a cada término en CM/C0 se muestran en la figura 35.



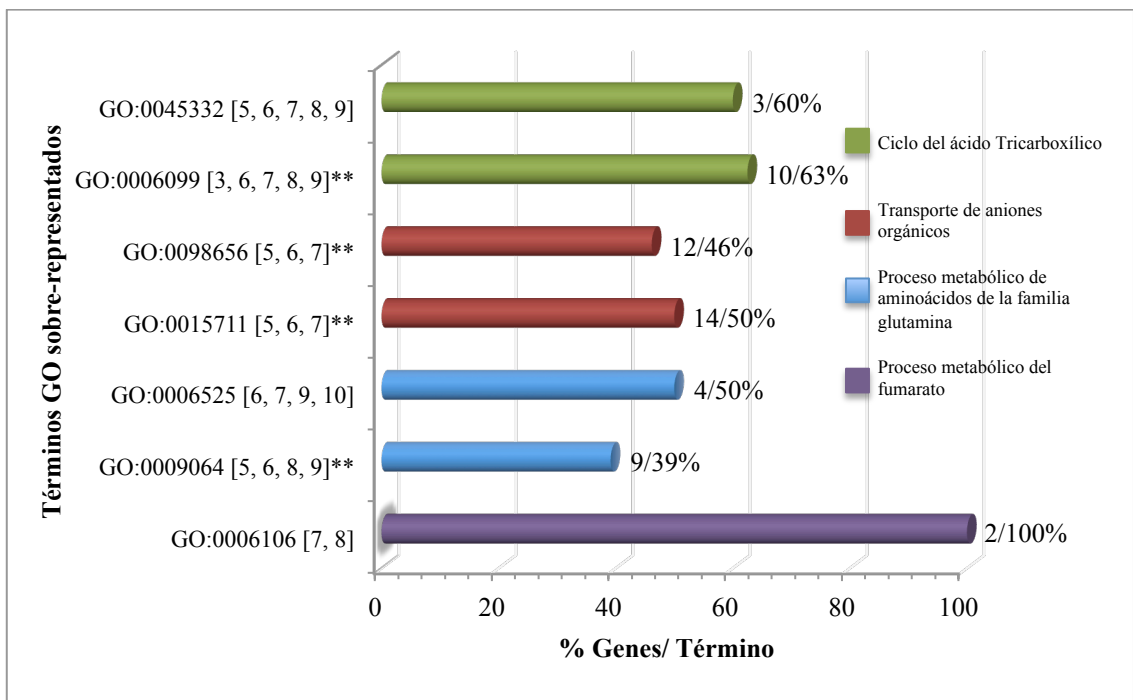
**Figura 32.** Términos GO sobre-representados de los genes sobre-expresados en la comparación CM/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .



**Figura 33.** Porcentaje de genes asociados a los términos GO sobre-representados de los genes sobre-expresados más importantes en la comparación CM/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

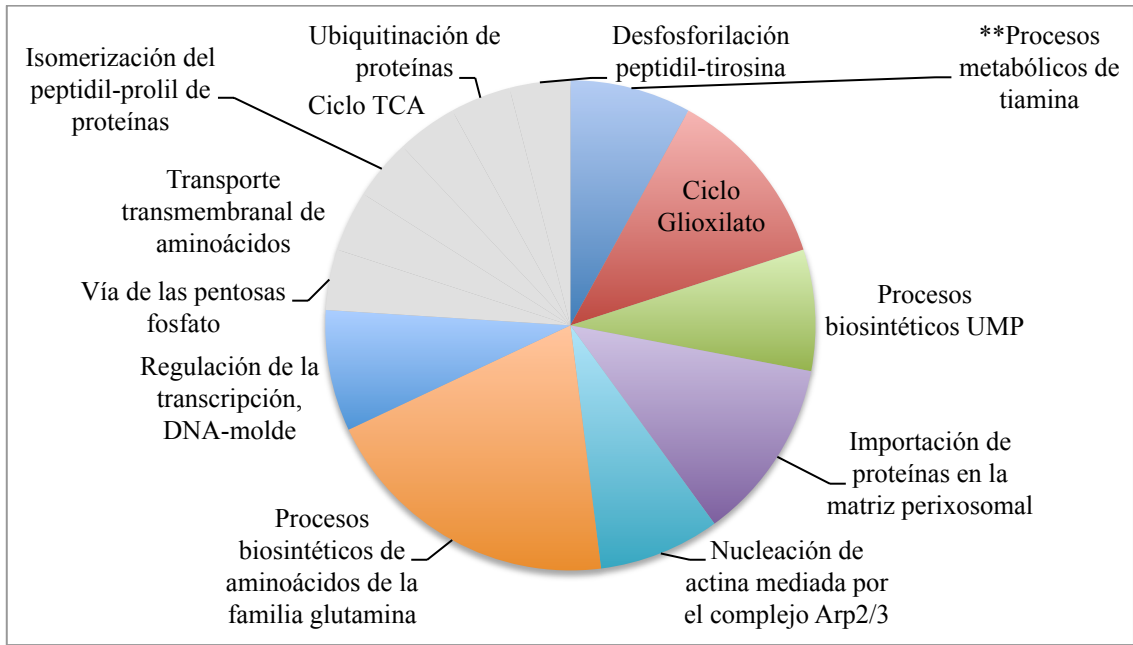


**Figura 34.** Términos GO sobre-representados de los genes sub-expresados en la comparación CM/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

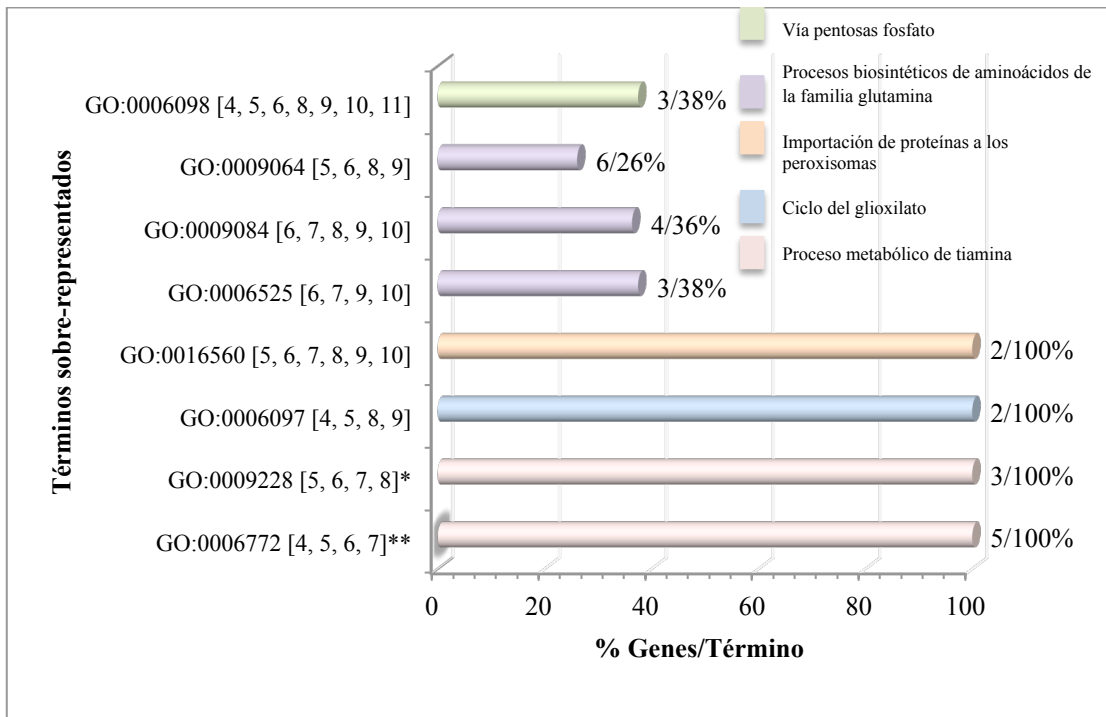


**Figura 35.** Porcentaje de genes asociados a los términos GO sobre-representados de los genes sub-expresados más importantes en la comparación CM/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .

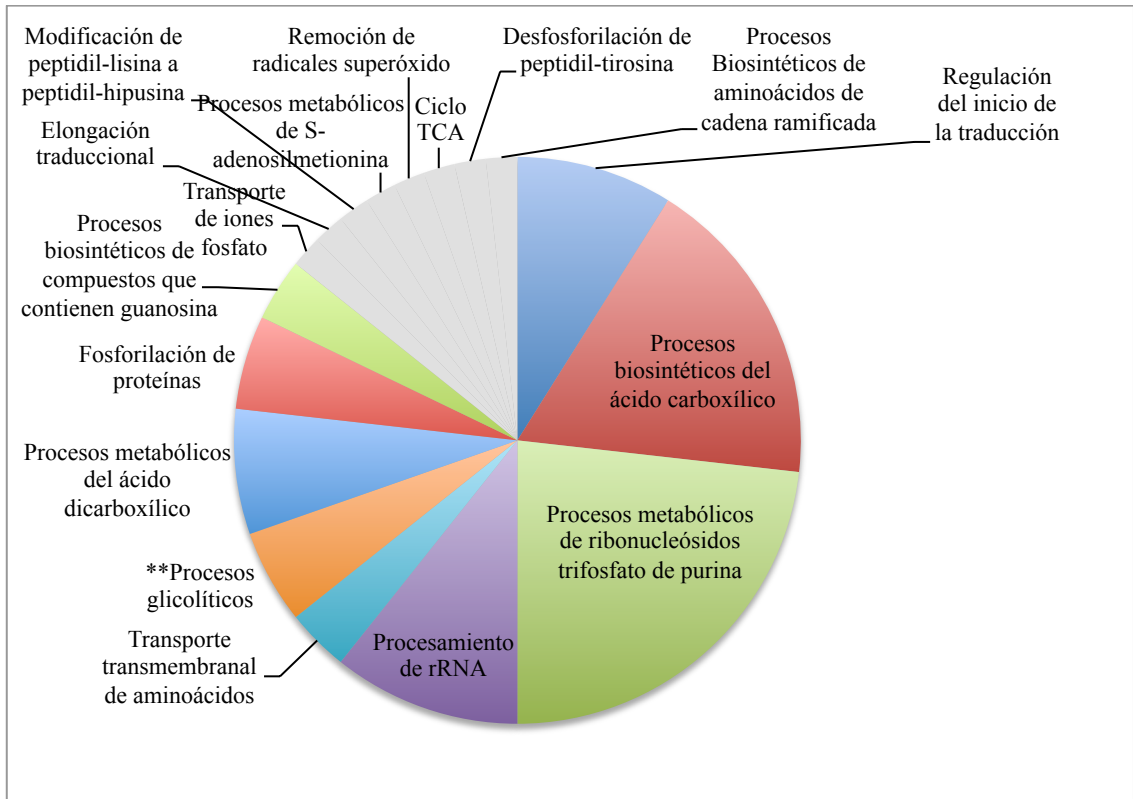
En la comparación CP/C0, se formaron 7 grupos y 6 términos no agrupados sobre-representados, en total se encontraron 25 términos con 8 conexiones entre algunos de ellos. Esta distribución de las agrupaciones de términos se muestra en la figura 36 y los términos más importantes en la figura 37. Además, se formaron 9 grupos y 8 términos no agrupados sobre-representados en base a los genes sub-expresados, con un total de 56 términos con 52 conexiones entre algunos de ellos; la distribución se muestra en la figura 38 y los términos más importantes en la figura 39.



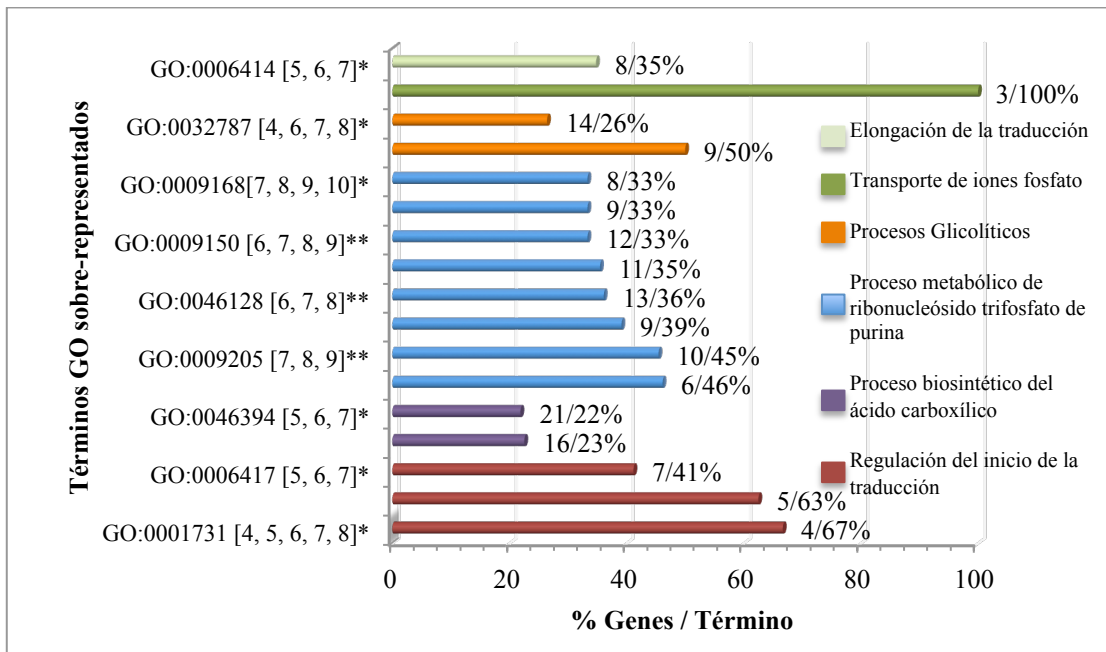
**Figura 36.** Términos GO sobre-representados de los genes sobre-expresados en la comparación CP/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .



**Figura 37.** Porcentaje de genes asociados a los términos GO sobre-representados de los genes sobre-expresados más importantes en la comparación CP/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

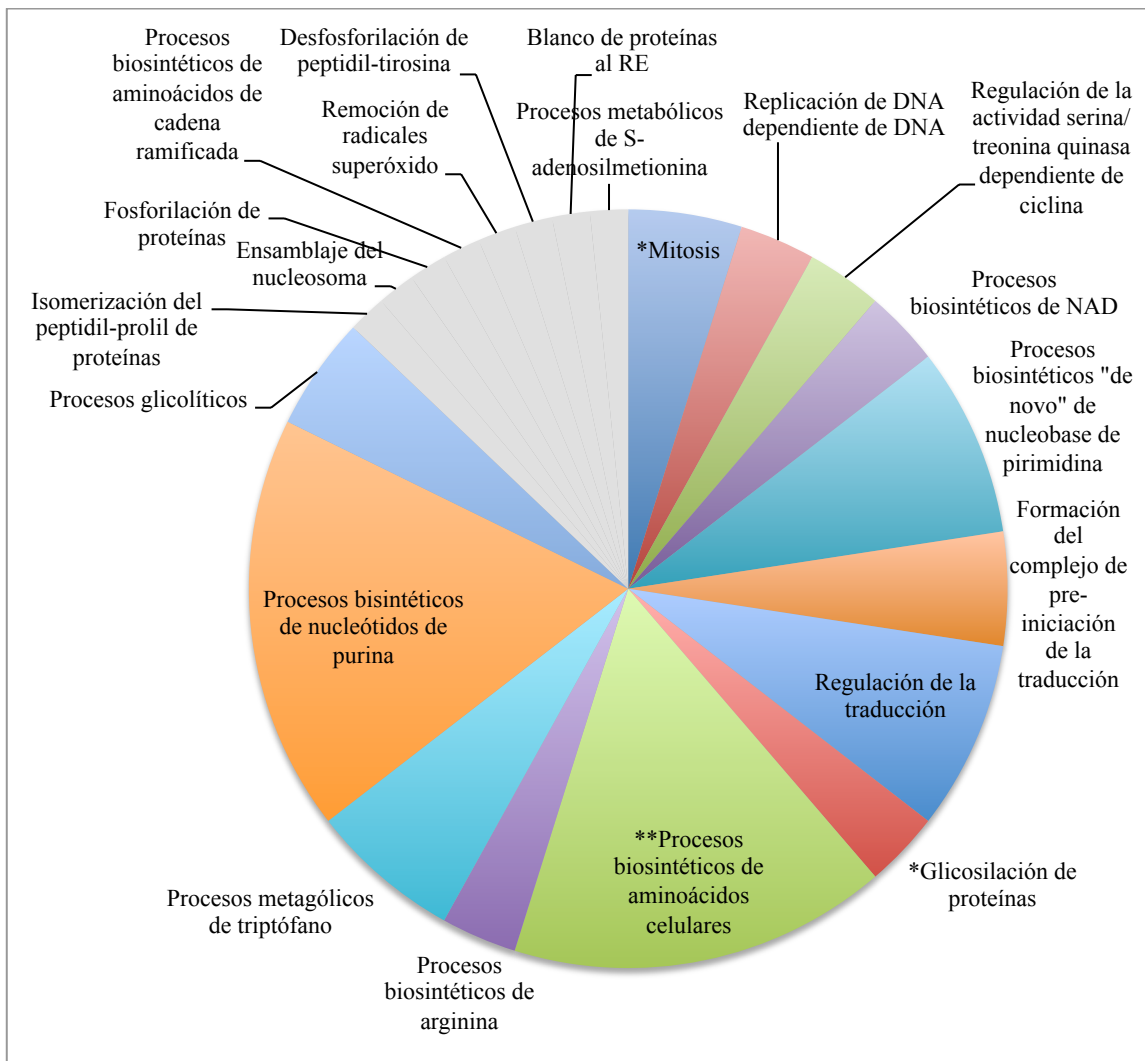


**Figura 38.** Términos GO sobre-representados de los genes sub-expresados en la comparación CP/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

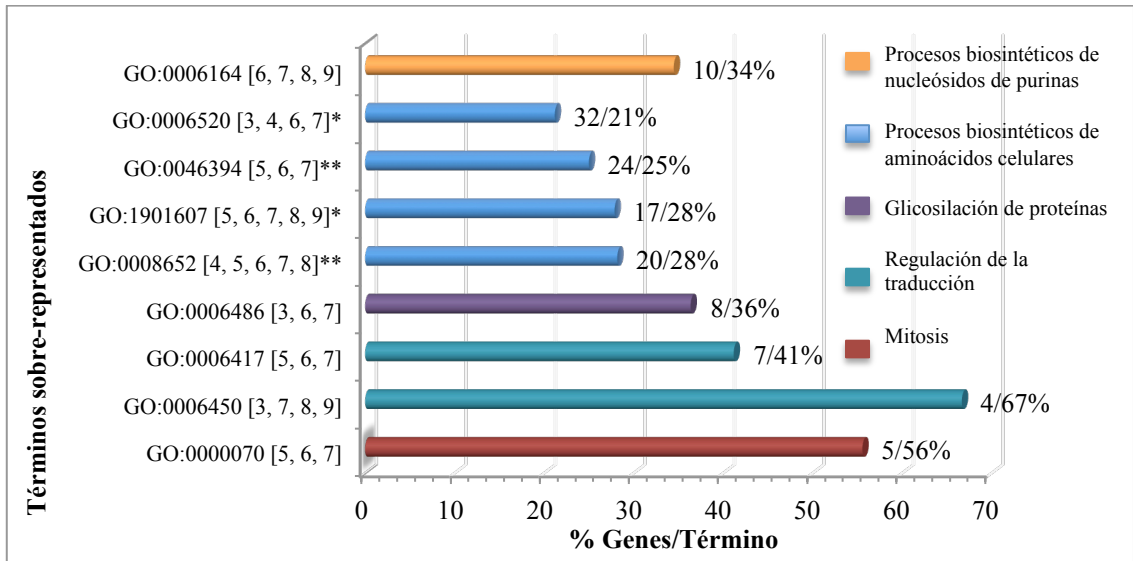


**Figura 39.** Porcentaje de genes asociados a los términos GO sobre-representados de los genes sub-expresados en la comparación CP/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

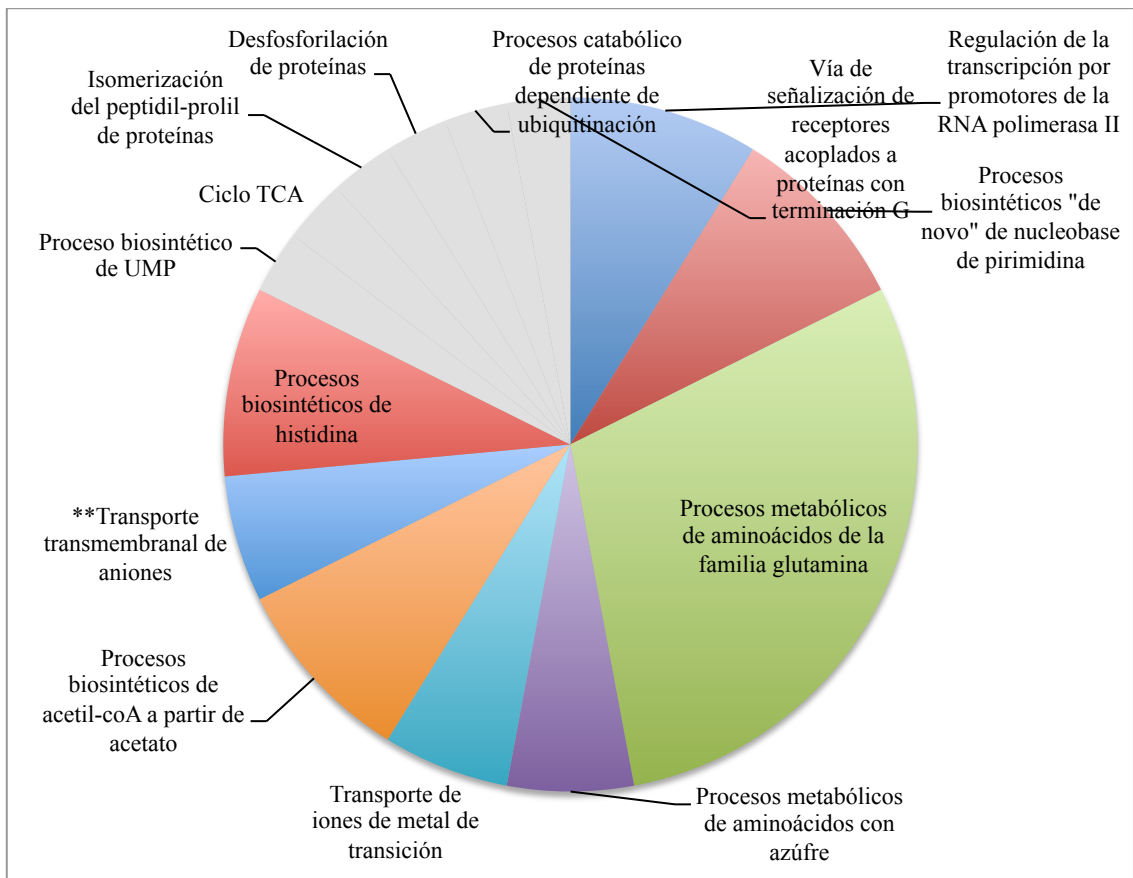
En la comparación CM/CP, se formaron 13 grupos y 8 términos no agrupados sobre-representados, en total se encontraron 62 términos con 46 conexiones entre algunos de ellos, la distribución de la agrupación de términos se muestra en la figura 40, así como los términos más importantes en la figura 41. Además, se formaron 8 grupos y 5 términos no agrupados sub-representados, en total se encontraron 34 términos con 17 conexiones entre algunos de ellos, la distribución de dichas agrupaciones se muestra en la figura 42, mientras que los términos más importantes se encuentran en la figura 43.



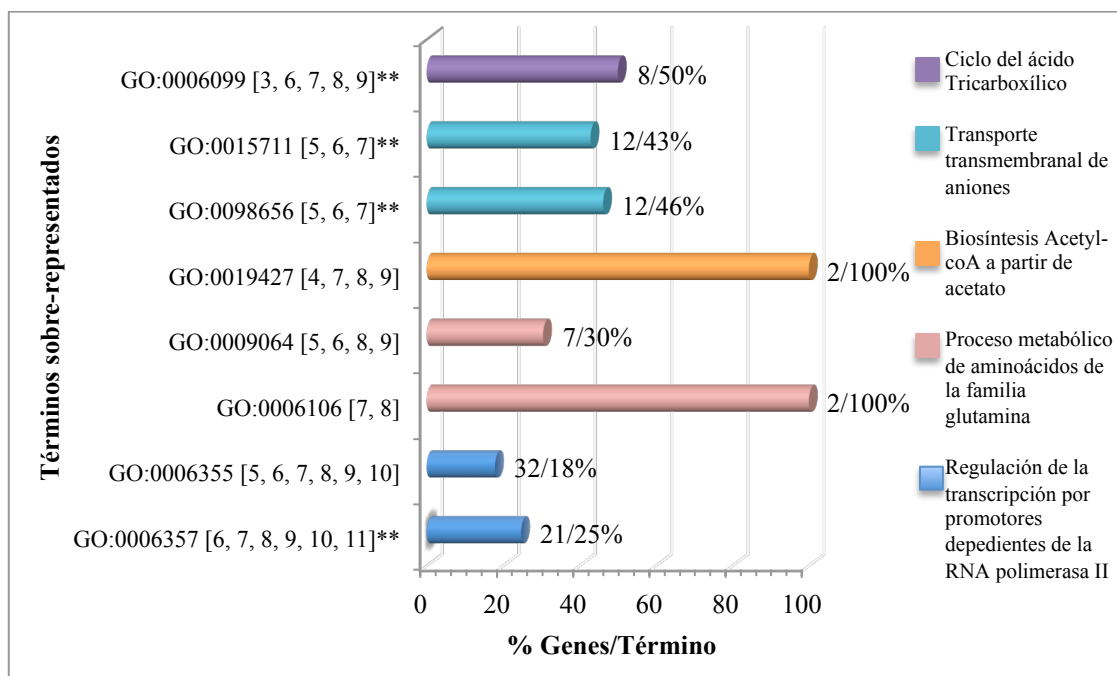
**Figura 40.** Términos GO sobre-representados de los genes sobre-expresados en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .



**Figura 41.** Porcentaje de genes asociados a los términos GO sobre-representados de los genes sobre-expresados en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .



**Figura 42.** Términos GO sobre-representados de los genes sub-expresados en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .



**Figura 43.** Porcentaje de genes asociados a los términos GO sobre-representados de los genes sub-expresados en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .

En la comparación CM/CP los términos más significativos con mayor cantidad de genes involucrados a los término respecto a los genes que resultaron sobre-expresados, correspondieron a la biosíntesis de aminoácidos, regulación de la traducción, mitosis, glicosilación de proteínas y biosíntesis de purinas. Esto indica que en la condición de alta producción se conduce a una mayor producción de aminoácidos, mayor actividad en la traducción y en la glicosilación de proteínas lo que representa procesos anabólicos activos para satisfacer la demanda de la producción y secreción de la proteína heteróloga y la producción de biomasa, éste último evento se puede también fundamentar en la sobre-expresión de los genes relacionados con la síntesis de purinas y una mitosis más activa. Estos resultados correlacionan con los resultados obtenidos por Viader-Salvadó y colaboradores (2013), donde indicaron que al final del cultivo crecido bajo las condiciones de alta producción (CM) se obtuvieron valores de  $\mu$  ( $0.014$  vs  $-0.007$   $h^{-1}$ ) y densidad celular ( $119$  vs  $68$   $g L^{-1}DCW$ ) más altos que el de la condición de baja producción (CP).

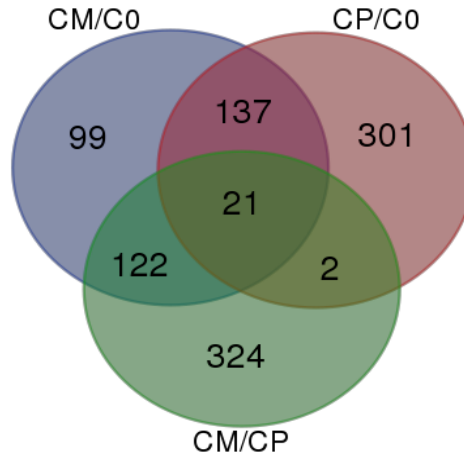


### **8.2.5.2 Genes sobre- y sub-expresados comunes en las tres comparaciones determinados mediante el diagrama de Venn**

El análisis a través del diagrama de Venn permitió identificar los genes que se encontraron sobre-expresados (Figura 44) y sub-expresados (Figura 45) en más de una comparación y en las distintas comparaciones; se utilizaron las claves de genes totales encontrados por el programa con un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ . Este diagrama permitió discriminar la expresión diferencial de los genes comunes en ambas condiciones de cultivo debida al cambio de la fuente de carbono de la expresión de aquellos genes que solo se da por una comparación específica.

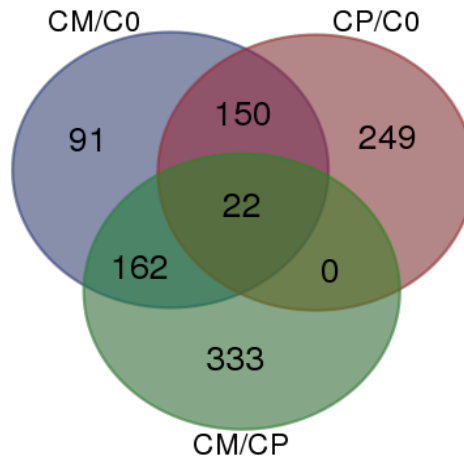
Este gráfico identificó elementos comunes (genes sobre- y sub-expresados) entre las distintas comparaciones. La expresión de los genes comunes tanto CM/C0 y CP/C0 al cambiar la fuente de carbono , brindan información importante de las rutas metabólicas que se activan en respuesta al metanol sin importar las condiciones de cultivo empleadas. Sin embargo, en éste trabajo también se enfocó en determinar, aquellos genes que se expresan diferencialmente bajo las condiciones de alta y baja producción, excluyendo a los genes que fueron influenciados por el cambio de la fuente de carbono.

Los resultados reflejan una expresión diferencial entre las dos condiciones de cultivo evaluadas (324 genes sobre-expresados y 333 genes sub-expresados en CM) excluyendo aquellos expresados en ambas condiciones debido del cambio de la fuente de carbono. Además el número de genes que se activan o se reprimen de manera diferencial en ambas condiciones de cultivo al cambiar la fuente de carbono tiende a ser mayor en las condiciones de baja producción.



**Figura 44.** Diagrama de Venn que muestra los genes sobre-expresados en las tres comparaciones considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

En el gráfico se representa la sobre-expresión de genes debido el cambio de la fuente de carbono y excluyentes en la comparación CM/C0 y CP/C0 (99 genes vs 301 genes, respectivamente), así como los genes comunes que sobre-expresaron al cambio de la fuente de carbono en las tres comparaciones CM/C0, CP/C0 y CM/CP (21 genes), genes comunes en comparaciones CM/C0 y CP/C0 (137 genes), genes comunes en las comparaciones CM/C0 y CM/CP (122 genes) y en las comparaciones CP/C0 y CM/CP (2 genes). Además, los genes cuya sobre-expresión no se debió al cambio de la fuente de carbono en CM/CP (324 genes).



**Figura 45.** Diagrama de Venn que muestra los genes sub-expresados en las tres comparaciones considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

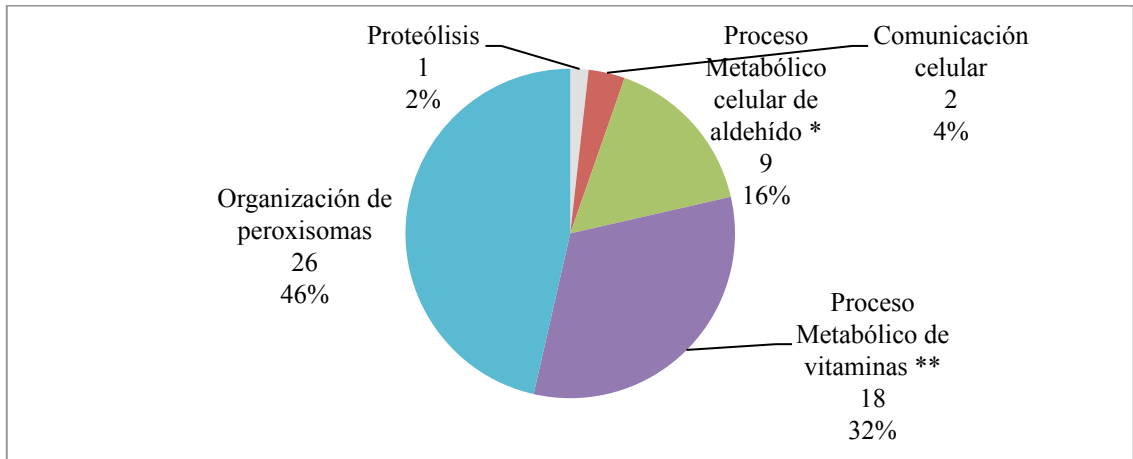
En el gráfico se representa la sub-expresión de genes debido el cambio de la fuente de carbono en CM/C0 y CP/C0 (91 genes vs 249 genes, respectivamente), así como los genes comunes que responden de manera similar al cambio de la fuente de carbono en las tres comparaciones CM/C0, CP/C0 y CM/CP (22 genes), genes comunes en comparaciones CM/C0 y CP/C0 (150 genes), en las comparaciones CM/C0 y CM/CP (162 genes) y en las comparaciones CP/C0 y CM/CP (0 genes). Además, los genes la comparación CM/CP que excluyen aquellos genes comunes expresados debido de la fuente de carbono (333 genes).

Debido a la importancia de conocer las rutas metabólicas que se encuentran asociadas a los genes sobre- y sub-expresados, se realizaron análisis de enriquecimiento de genes del grupo de genes comunes que se expresaron diferencialmente frente al cambio de la fuente de carbono en ambos cultivos CM/C0 y CP/C0, así como los genes que presentaron una expresión diferencial en CM respecto a CP excluyendo aquellos regulados por cambio en la fuente de carbono. En ambos casos se utilizó ClueGO con los parámetros estándar.

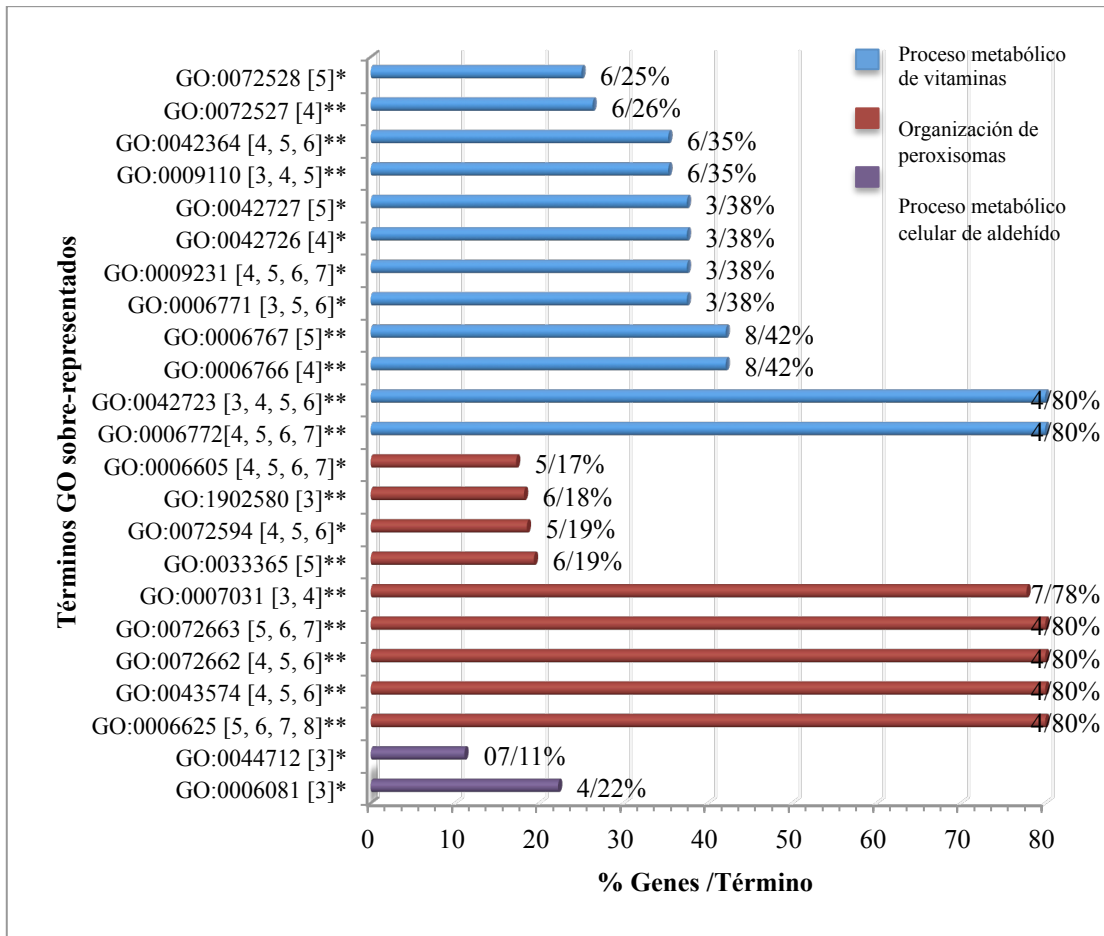
En los gráficos de pastel cada sector es proporcional al número de términos agrupados, el nombre de cada sector corresponde al término más representativo de cada grupo (resultados obtenidos con parámetros estándar). En el gráfico se indica el grado de significancia de la totalidad de cada agrupación (Valor P del grupo:  $p \leq 0.05$  (significativos) \* y  $p \leq 0.01$  (altamente significativos) \*\* corregidos con “Bonferroni step down”). De color gris se representan los términos no agrupados.

En cuanto a los gráficos de barras, en el eje de las ordenadas se encuentran los Términos GO seguido por los niveles GO entre corchetes y el grado de significancia ( $p \leq 0.05$  \* y  $p \leq 0.01$  \*\*, corregidos de igual manera con “Bonferroni step down”), éstos gráficos fueron creados en base a los términos con valores P más lejanos de 1, es decir, más significativos. Además a un lado de las barras se encuentra el número de genes encontrados y el porcentaje con respecto al total de genes en dicho término. El nombre de cada agrupación (color de las barras) está encabezado por el término más significativo.

En relación a la distribución del número y porcentaje de términos sobre-representados correspondientes a los genes comunes sobre-expresados al cambiar la fuente de carbono de glicerol a metanol en ambas condiciones de cultivo (CM/C0 y CP/C0), se formaron 4 grupos de términos relacionados (0-3) y 1 término no agrupado (Figura 46). Además en la figura 47 se representa el porcentaje de genes encontrados asociados a los Términos GO, con valor P modificado más lejano a 1, sobre-representados de los genes comunes sobre-expresados en CM/C0 y CP/C0 .



**Figura 46.** Porcentaje de términos agrupados sobre-representados de los genes sobre-expresados debido al cambio en la fuente de carbono en CM/C0 y CP/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .



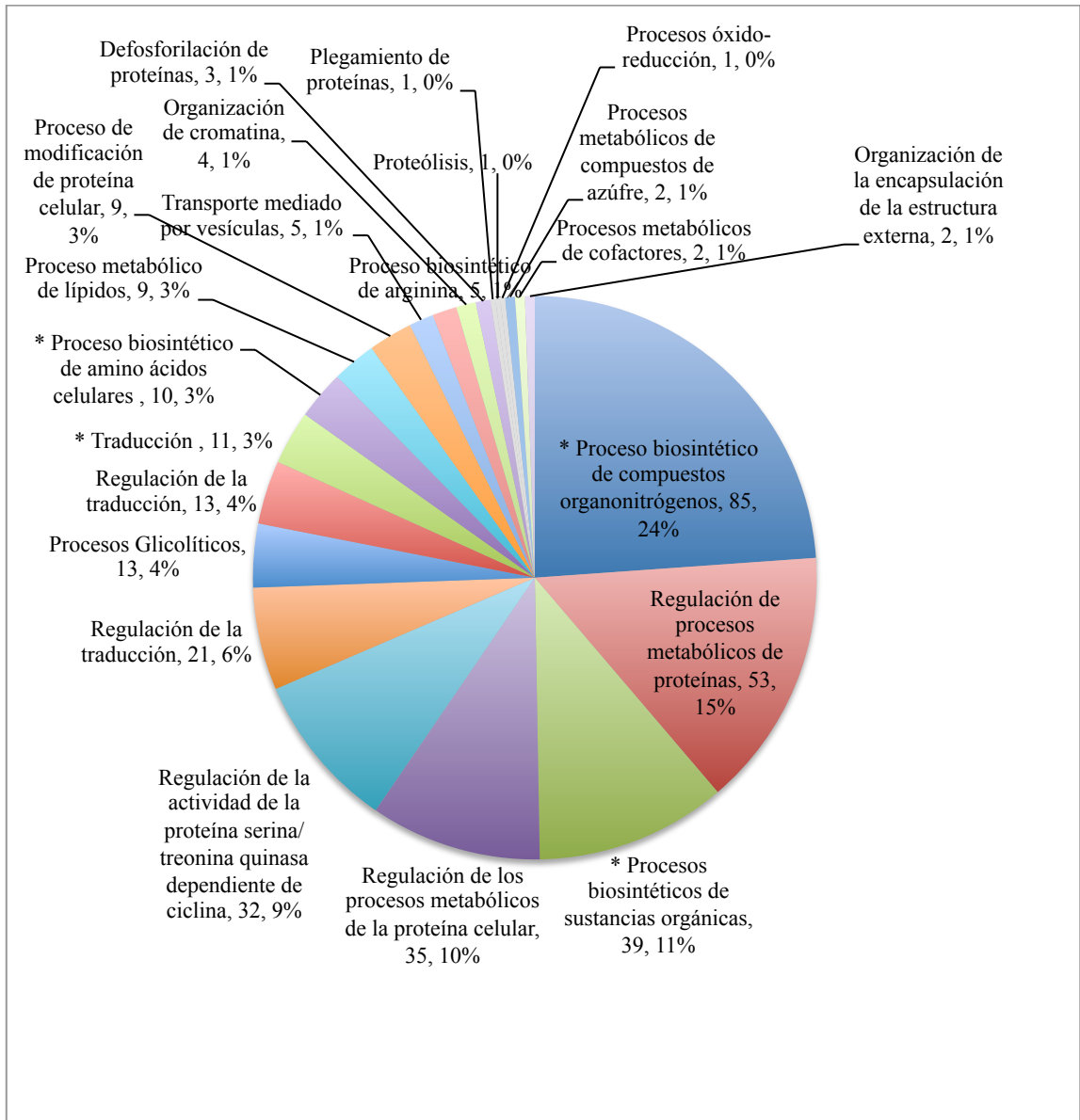
**Figura 47.** Porcentaje de genes asociados a los Términos GO más importantes sobre-representados de los genes sobre-expresados debido al cambio en la fuente de carbono en CM/C0 y CP/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

Respecto la distribución del número y porcentaje de términos sobre-representados correspondientes a los genes sobre-expresados excluyendo a los regulados por el cambio en la fuente de carbono en la comparación CM/CP, se encontraron 19 grupos de términos (0-18) y 2 términos no agrupados (Figura 48). Además, en la figura 49 se representa el porcentaje de genes encontrados asociados a los Términos GO, con valor P modificado más lejano a 1, sobre-representados de los genes sobre-expresados en CM/CP.

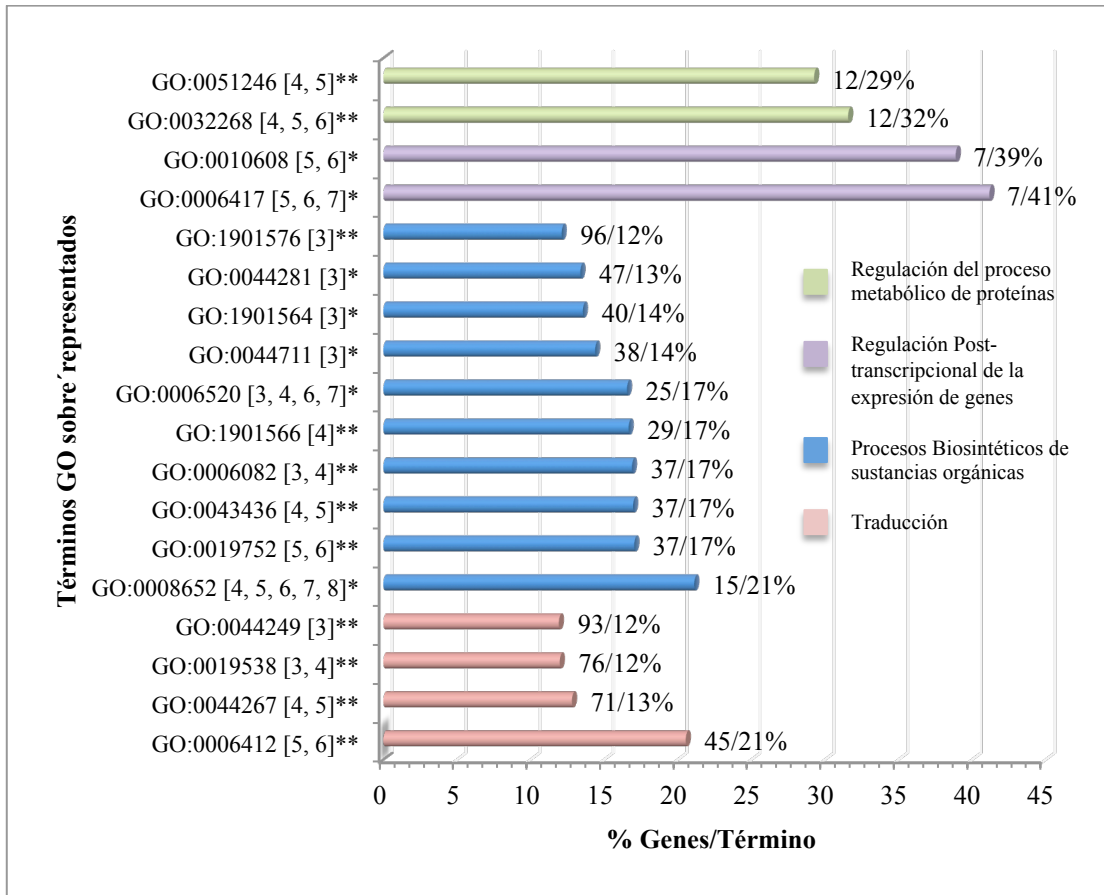
Los resultados muestran que hay una respuesta evidente como resultado del cambio de la fuente de carbono de glicerol a metanol en ambas condiciones de cultivo (CM/C0 y CP/C0) debido a que los genes sobre-expresados mostraron una sobre-representación significativa de términos relacionados a la organización de peroxisomas; ésta respuesta correlaciona con la adaptación a metabolizar el metanol por la cepa empleada (KM71 pPIC9FTEII), comportamiento esperado para una cepa metilotrófica . En la comparación CM/CP, excluyendo los genes sobreexpresados en ambos cultivos como respuesta al metanol, los genes sobre-expresados mostraron la sobre-representación significativa de términos relacionados con la traducción, procesos biosintéticos de sustancias orgánicas, regulación post-transcripcional de la expresión de genes y regulación del proceso metabólico de proteínas.

Los términos significativos sobre-expresados relacionados con la traducción, y los procesos biosintéticos de sustancias orgánicas son indicadores de un proceso anabólico activo, además los términos significativos sobre-expresados involucrados en procesos de regulación de la expresión de genes y regulación del metabolismo de proteínas indican un proceso de regulación activo para responder a la demanda metabólica generada por los altos niveles de producción de la proteína heteróloga generados en el cultivo CM.

Esto complementa la información obtenida en el primer análisis global de enriquecimiento de genes (8.2.5.1) y la expresión diferencial de genes que favorecen al crecimiento celular observado en CM durante las fermentaciones realizadas por Viader y colaboradores (2013).

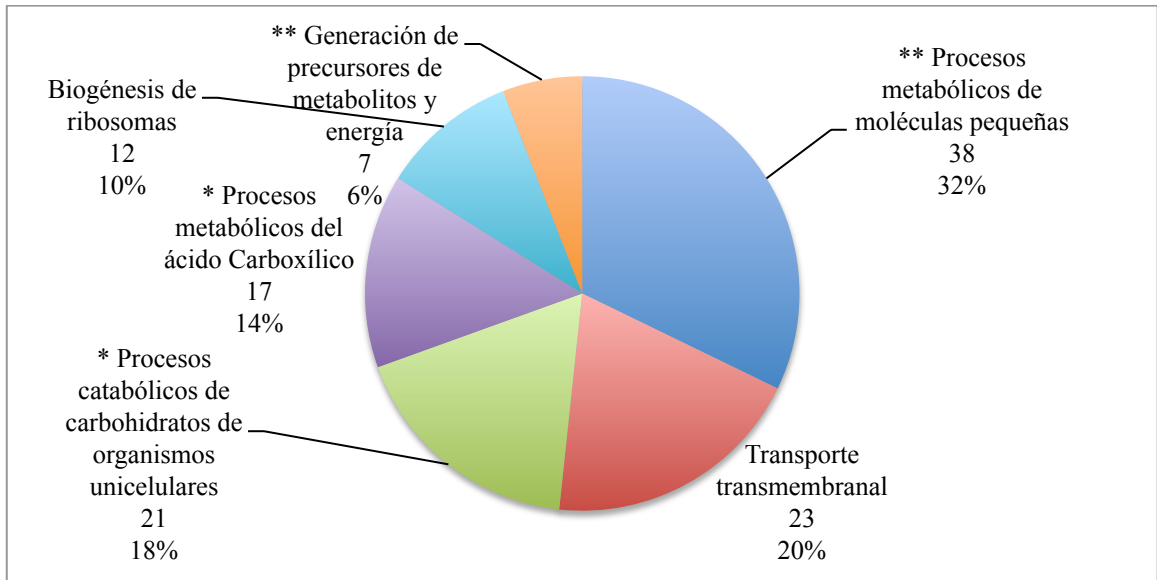


**Figura 48.** Porcentaje de términos agrupados sobre-representados de los genes sobre-expresados en la comparación CM/CP que excluye a los genes sobre-expresados tanto en CM como en CP debido al cambio en la fuente de carbono utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .

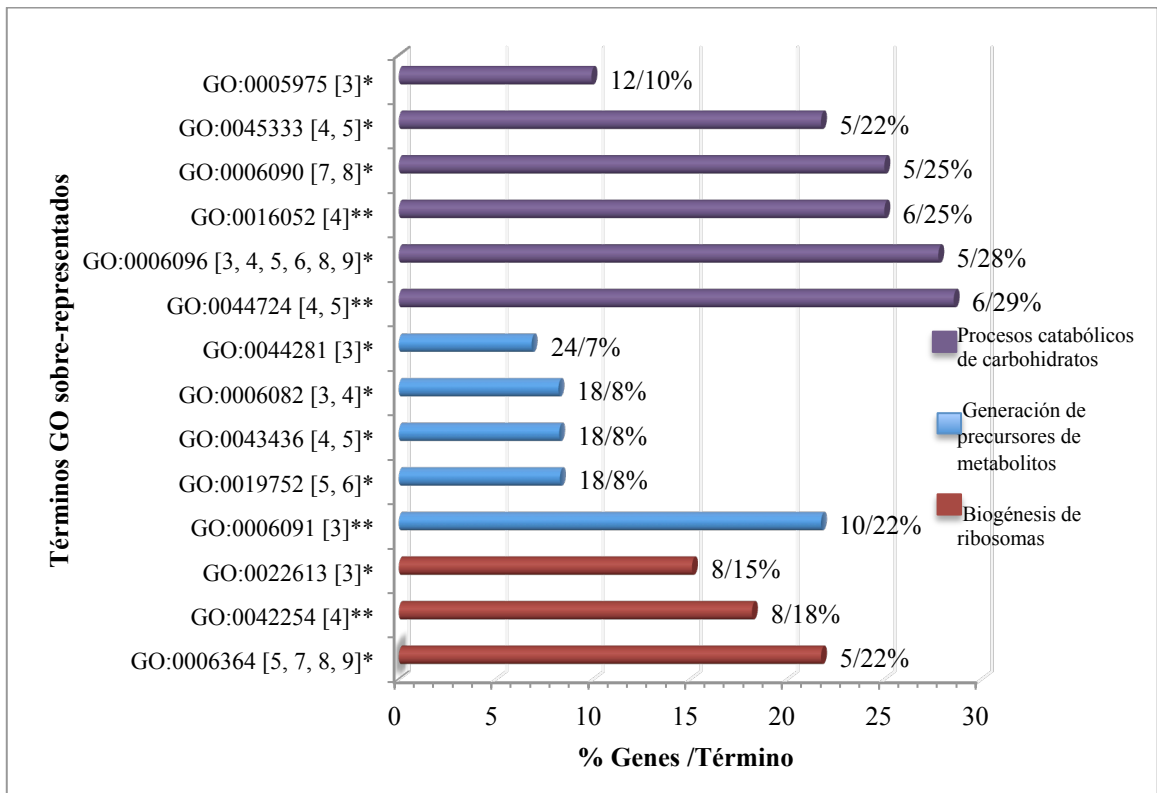


**Figura 49.** Porcentaje de genes asociados a los términos GO más importantes sobre-representados de los genes sobre-expresados en la comparación CM/CP que excluye a los genes sobre-expresados tanto en CM como en CP debido al cambio en la fuente de carbono utilizando la ontología de procesos biológicos considerando un  $\log_2(\text{Fold\_change}) \leq -0.7$  y  $\geq 0.7$ .

Por otro lado, en relación a la distribución de términos sobre-representados de los genes sub-expresados al cambiar la fuente de carbono de glicerol a metanol en ambas condiciones de cultivo, se encontraron 6 grupos (0-5) (Figura 50). En la figura 51 se representa el porcentaje de genes encontrados asociados a los Términos GO, con valor P modificado más lejano a 1, sobre-representados de los genes sub-expresados en CM/C0 y CP/C0.



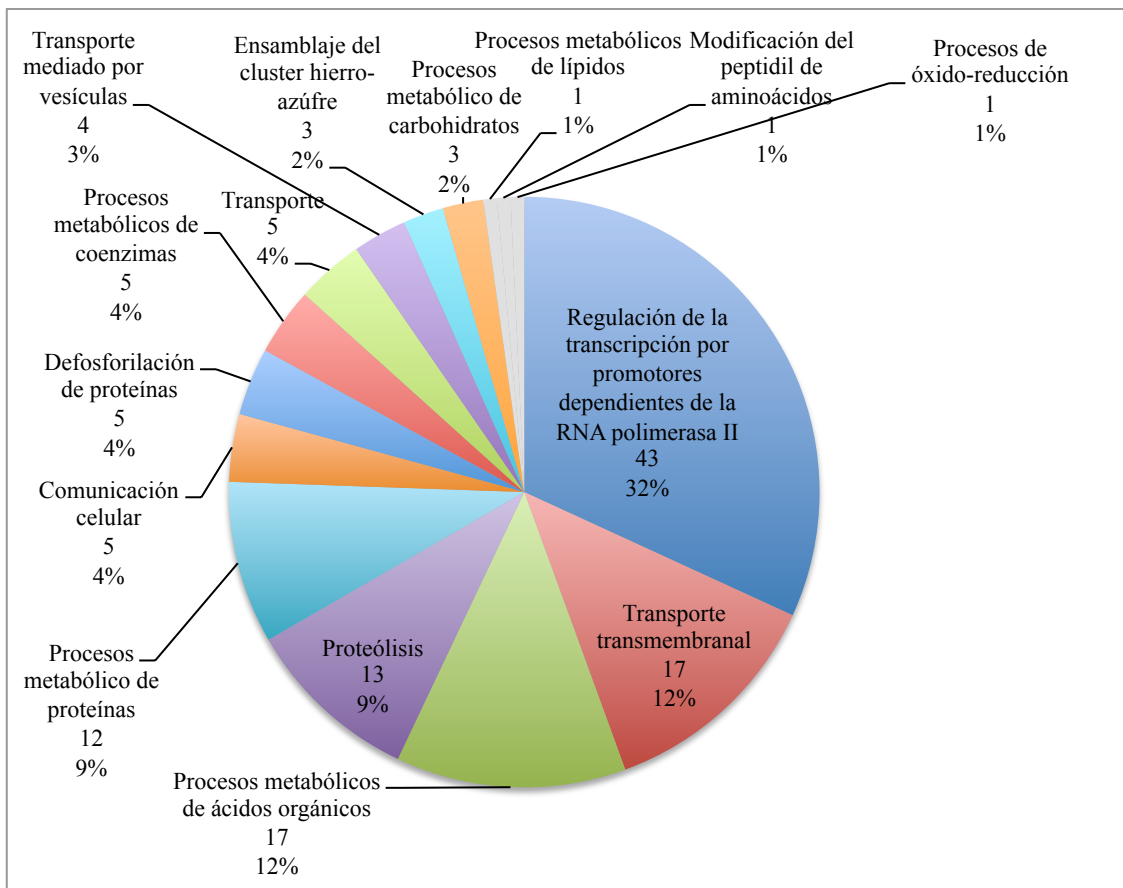
**Figura 50.** Porcentaje de términos agrupados sobre-representados de los genes sub-expresados debido al cambio en la fuente de carbono en las comparaciones CM/C0 y CP/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .



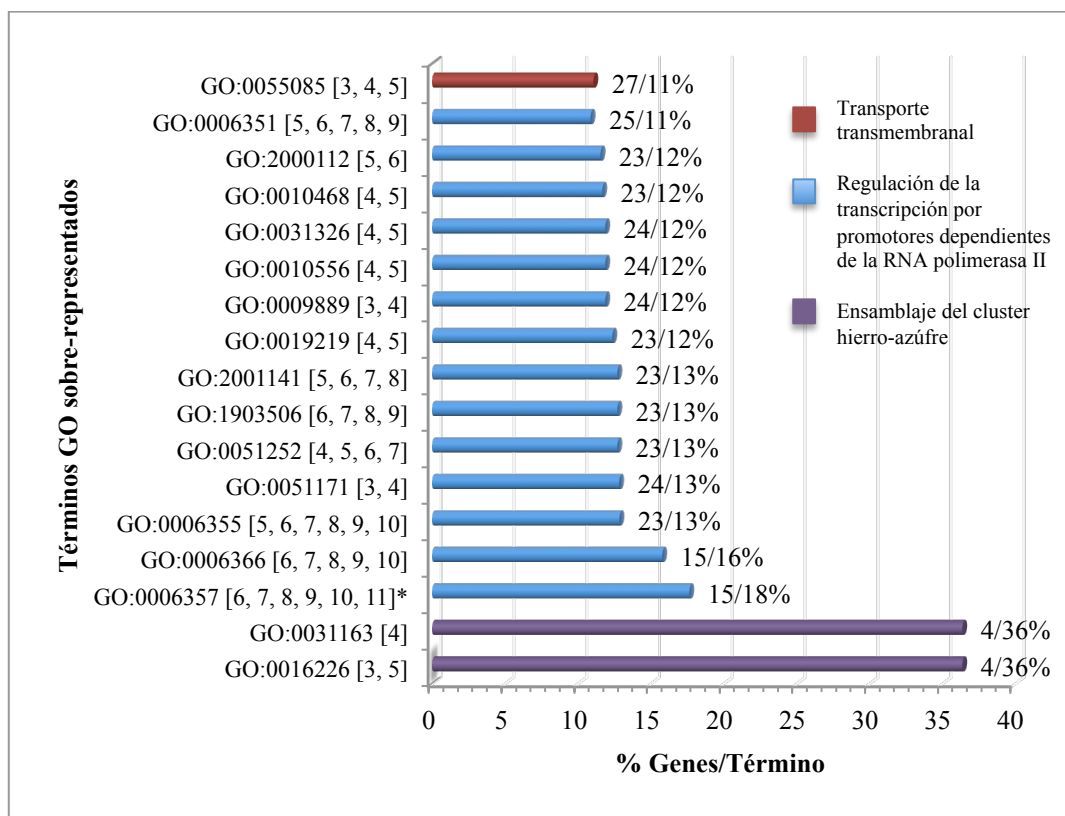
**Figura 51.** Porcentaje de genes asociados a los términos GO mas importantes sobre-representados de los genes sub-expresados debido al cambio en la fuente de carbono en las comparaciones CM/C0 y CP/C0 utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .



Los resultados obtenidos por la sub-expresión de genes debido al cambio de la fuente de carbono mostraron una sobre-representación significativa de términos relacionados con una disminución de la actividad catabólica reflejado con la sub-expresión de genes relacionados con la generación de precursores de metabolitos y energía y los procesos catabólicos de carbohidratos de organismos unicelulares. También se ven afectados los procesos relacionados el metabolismo de moléculas pequeñas y de ácidos carboxílicos. Estos resultados conducen a plantear que la disminución de la actividad de los procesos catabólicos en procesos del metabolismo intermediario debido al cambio de la fuente de carbono son compensados con el metabolismo del metanol en la generación de energía y fuente de carbono necesaria para la generación de biomasa.



**Figura 52.** Porcentaje de términos agrupados sobre-representados de los genes sub-expresados que excluye a los genes sub-expresados tanto en CM como en CP debido al cambio en la fuente de carbono en las comparaciones CM/CP utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .



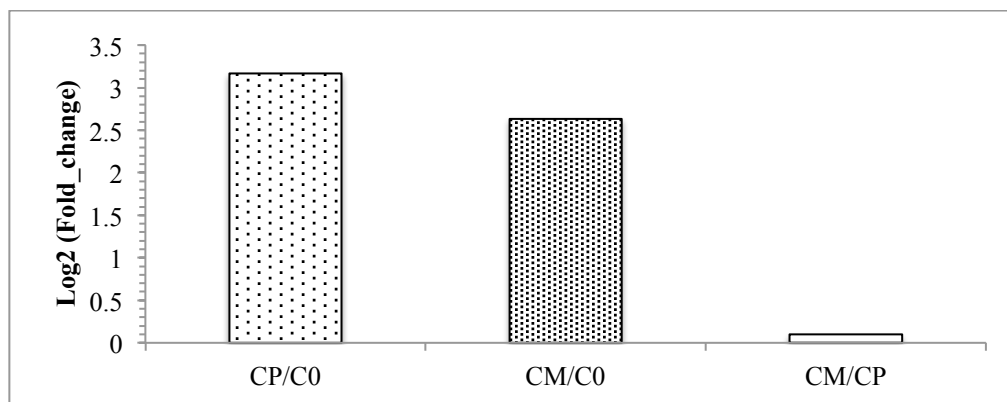
**Figura 53.** Porcentaje de genes asociados a los términos GO sobre-representados de los genes sub-expresados que excluye a los genes sub-expresados tanto en CM como en CP debido al cambio en la fuente de carbono en la comparación CM/CP utilizando la ontología de procesos biológicos considerando un  $\log_2$  (Fold\_change)  $\leq -0.7$  y  $\geq 0.7$ .

La sub-expresión de genes en CM debido solo al cambio de las condiciones de cultivo, sin considerar los genes sub-expresados tanto en CM como en CP debido el cambio de la fuente de carbono, se sobre-representó un solo termino significativo, éste fue relacionado con la regulación de la transcripción por promotores dependientes de la RNA polimerasa II, la sub-expresión de estos genes bajo las condiciones de CM pueden estar relacionados con una disminución en la carga metabólica debido a funciones no implicadas con la síntesis de la proteína recombinante.

### 8.2.5.2 Análisis de expresión diferencial del gen Heterólogo

El gen heterólogo *FTEII* se sobre-expresó debido al cambio de la fuente de carbono con valores de  $\log_2$  (Fold\_change) de 2.6 y 3.2 para las comparaciones CM vs C0 y CP vs C0 respectivamente y no se presentó una expresión diferencial de los

transcritos de FTEII por el cambio de las condiciones de cultivo, obteniendo un valor de  $\log_2$  (Fold\_change de -0.53 para la comparación CM vs CP (Figura 54).



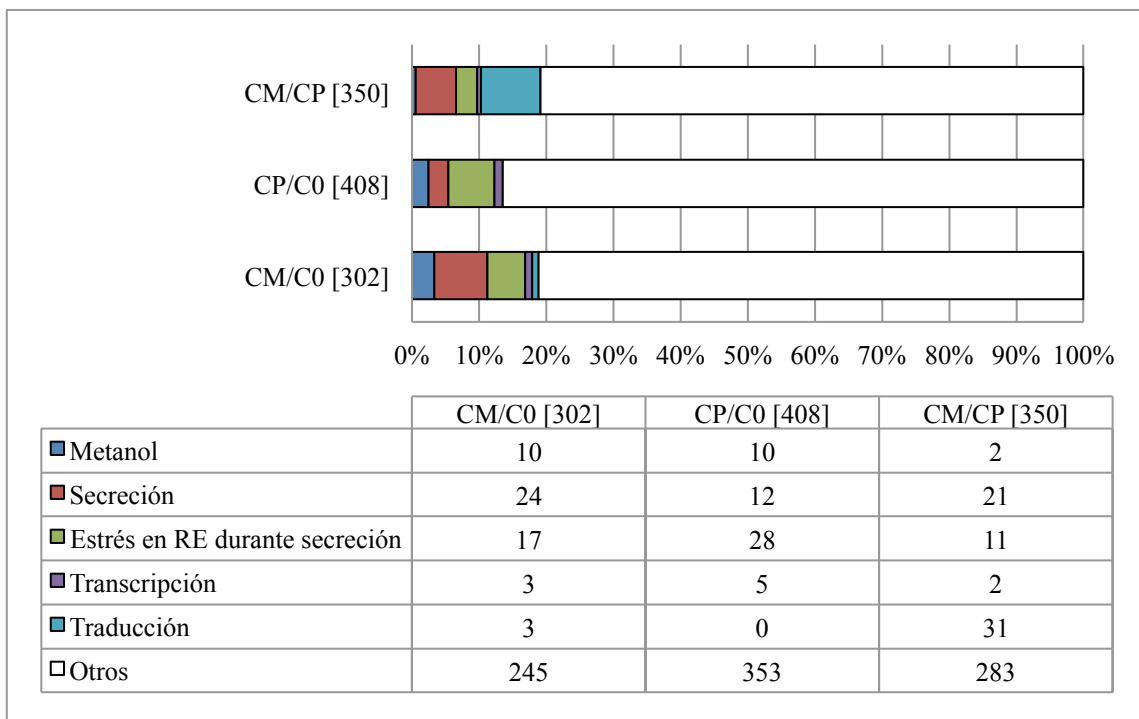
**Figura 54.** Expresión diferencial del gen heterólogo *FTEII* en las tres comparaciones (CM/C0, CP/C0 y CM/CP).

Los resultados indican que el gen heterólogo se expresó sin diferencias en ambas condiciones de cultivo después de la inducción con metanol. Debido a que los niveles de producción condujeron a valores  $Y_{p/x}$  más altos en CM que CP (Viader-Salvadó *et al.*, 2013), y a que el nivel de transcritos fue igual en las dos condiciones de cultivo se infiere que los altos niveles de producción de la proteína heteróloga no fueron el resultado de una regulación diferencial del promotor AOX1 y que las diferencias condiciones de cultivo en CM y CP no fueron determinantes en la regulación de este promotor, así que el cambio de la fuente de carbono impactó en el aumento de transcritos de FTEII, tal como se esperaba, sin embargo la diferencia en los niveles de producción fueron el resultado de la regulación de genes que afectaron el comportamiento metabólico diferente entre ambas condiciones de cultivo.

### 8.2.5.3 Análisis de expresión diferencial de las Rutas metabólicas específicas

La distribución de los genes sobre-expresados (Figura 55) y sub-expresados (Figura 56) del análisis de genes involucrados en la vía del metanol, secreción y de genes que responden al estrés en el retículo endoplásmico durante la secreción mencionados por De Schutter y colaboradores (2009) y otros autores, se representan a

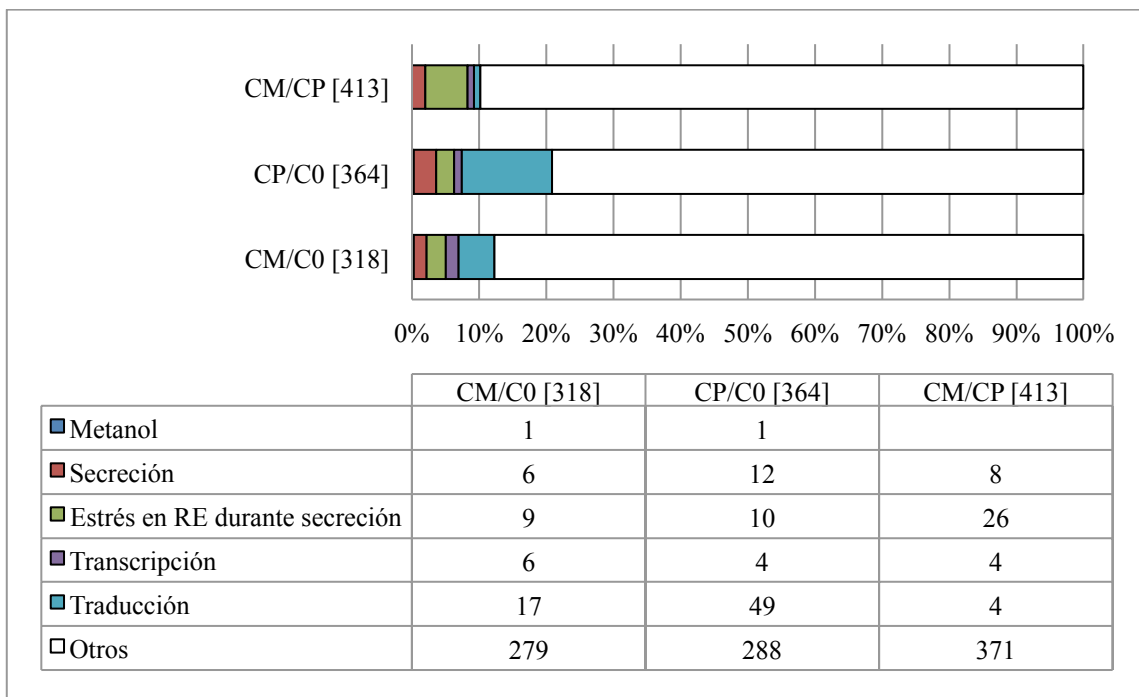
continuación. En el eje de las ordenadas se encuentran entre corchetes el número de genes totales sobre- ó sub-expresados correspondientes a cada comparación. En el eje de las abscisas se representan en las barras los porcentajes de genes sobre-expresados de las rutas metabólicas analizadas. Además, la tabla indica el nombre de cada ruta y el número de genes sobre-expresados correspondiente a cada comparación. Estos resultados nos ubican en la pequeña proporción de genes que se analizan en éste trabajo.



**Figura 55.** Distribución del número de genes sobre-expresados considerando un  $\log_2$  (Fold\_change)  $\leq 1.0$  ó  $\geq 1.0$  en las rutas metabólicas analizadas.

En general, de las rutas metabólicas analizadas, los genes relacionados directa o indirectamente con la secreción presentan una mayor participación en la expresión diferencial resultado de las dos condiciones de cultivo.

Con los resultados obtenidos, se puede inferir que la regulación de los genes del proceso de secreción impactaron en los niveles altos de producción bajo las condiciones CM.

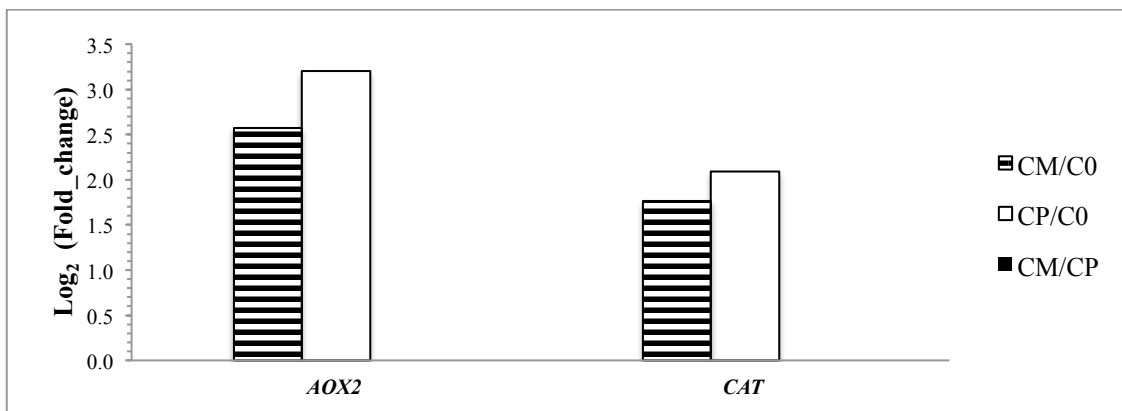


**Figura 56.** Distribución del número de genes sub-expresados considerando un  $\log_2(\text{Fold\_change}) \leq 1.0$  ó  $\geq 1.0$  en las rutas metabólicas analizadas.

Los genes relacionados con la secreción, además de expresarse diferencialmente entre CM y CP estos se sobre-expresaron en CM en presencia del metanol. También es importante enfatizar la participación de los genes relacionados con el estrés en RE durante la secreción los cuales se sub-expresaron en CM vs CP.

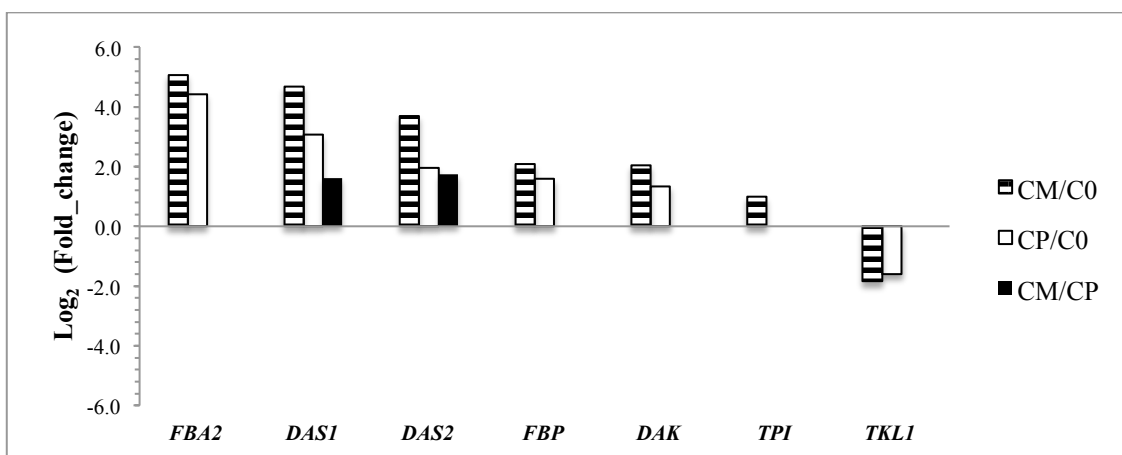
### 8.2.5.3.1 Análisis de expresión diferencial de genes de la vía del metabolismo del metanol (vía MUT)

En las comparaciones CM vs C0 y CP vs C0 se sobre-expresaron los genes *AOX2* y *CAT*, dos genes clave para el inicio del metabolismo del metanol (Figura 57). A pesar de que presentan valores de expresión diferencial ligeramente diferentes (Tabla 20), estos no fueron significativos; lo que indica que la regulación de estos genes se comporta igual frente al cambio de la fuente de carbono y no es afectada por las condiciones de cultivo.



**Figura 57.** Expresión diferencial considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$  de los genes importantes en el inicio del metabolismo del metanol (*AOX2* y *CAT*) en las tres comparaciones.

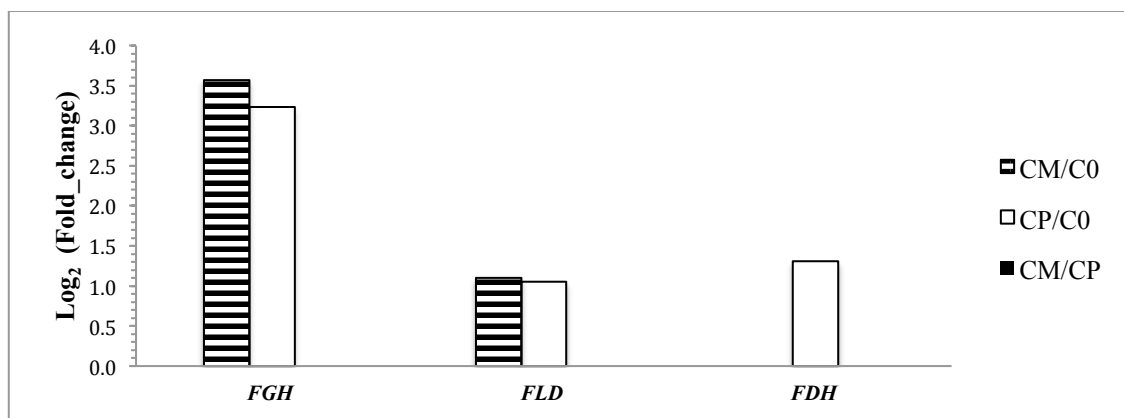
Respecto a la vía de asimilación del metanol (Tabla 21), se sobre-expresaron todos los genes de ésta vía en la comparación CM vs C0 y CP vs C0 (Figura 58), a excepción del gen *TKL1*, el cuál se sub-expresó para ambas comparaciones, y *TPI* que no se sobre-expresó en CP vs C0. A pesar de todo los genes de la vía de asimilación del metanol presentaron niveles más altos de expresión en CM que en CP por el cambio de la fuente de carbono, solo los genes *DAS1* y *DAS2*, se expresaron diferencialmente de forma significativa en CM vs CP (Figura 58).



**Figura 58.** Genes diferencialmente expresados considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$  involucrados en la vía de asimilación del metanol en las tres comparaciones.

En la vía de desasimilación del metanol (Tabla 22), los genes *FGH* y *FLD* se sobre-expresaron por el cambio de la fuente de carbono, mientras que el gen *FDH* solo

se sobre-expresó en la comparación CP vs C0 (Figura 59). Sin embargo no hubo diferencias significativas en la expresión de estos tres genes entre ambas condiciones.



**Figura 59.** Expresión diferencial de genes considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$  involucrados en la vía de desasimilación del metanol en las tres comparaciones.

Los resultados muestran que de los 7 genes que participan en vía de asimilación del metanol, 6 genes se sobre-expresaron al cambiar la fuente de carbono de glicerol a metanol en el cultivo CM y 5 genes en CP. Un hecho importante a resaltar, es que el gen *TKL1*, a pesar de codificar para una enzima importante para la asimilación del metanol en levaduras, se sub-expresó en ambas condiciones, estos resultados reflejan que este gen no está implicado en el metabolismo del metanol en *P. pastoris*, bajo las condiciones estudiadas. El gen *TPI*, sólo se sobre-expresó en la comparación CM vs C0 y no así en CP vs C0; estudios realizados por Solà y colaboradores (2007), a través de experimentos con marcaje isotópico de <sup>13</sup>C en células de *P. pastoris*, revelaron que la triosa fosfato formada a partir de metanol puede ser utilizada tanto anabólicamente (síntesis de componentes celulares) como catabólicamente (obtención de energía vía ciclo de los TCA). La sobre-expresión de *TPI*, bajo las condiciones del cultivo CM podría justificar la mayor generación de biomasa y energía obtenida en el cultivo CM que en el CP. Por otro lado, la sobre-expresión de *DAS1* y *DAS2* en la comparación CM vs CP, podría justificar una mayor actividad metabólica, la cuál impactó en una alta demanda de metanol del cultivo CM y condujo a valores de velocidad específica (0.014 vs -0.007 h<sup>-1</sup>) y densidad celular (119 vs 68 g L<sup>-1</sup>DCW) más altos en el cultivo CM que en el cultivo CP (Viader-Salvadó *et al.*, 2013).

**Tabla 20.** Nombre y función de las proteínas codificadas por los genes involucrados en el inicio del metabolismo del metanol.

Gen	ID del Gen	log <sub>2</sub> (Fold_change)			Enzima	Función
		CM/C0	CP/C0	CM/CP		
<i>AOX2</i>	PAS_chr4_0152	4.7	3.2	-	Alcohol oxidasa	Conversión de metanol a aldehído y peróxido de hidrógeno
<i>CAT</i>	PAS_chr2-2_0131	1.8	2.1	-	Catalasa peroxisomal	Hidrólisis el peróxido de hidrógeno en la matriz peroxisomal

**Tabla 21.** Nombre y función de las proteínas codificadas por los genes involucrados en la vía de asimilación del metanol.

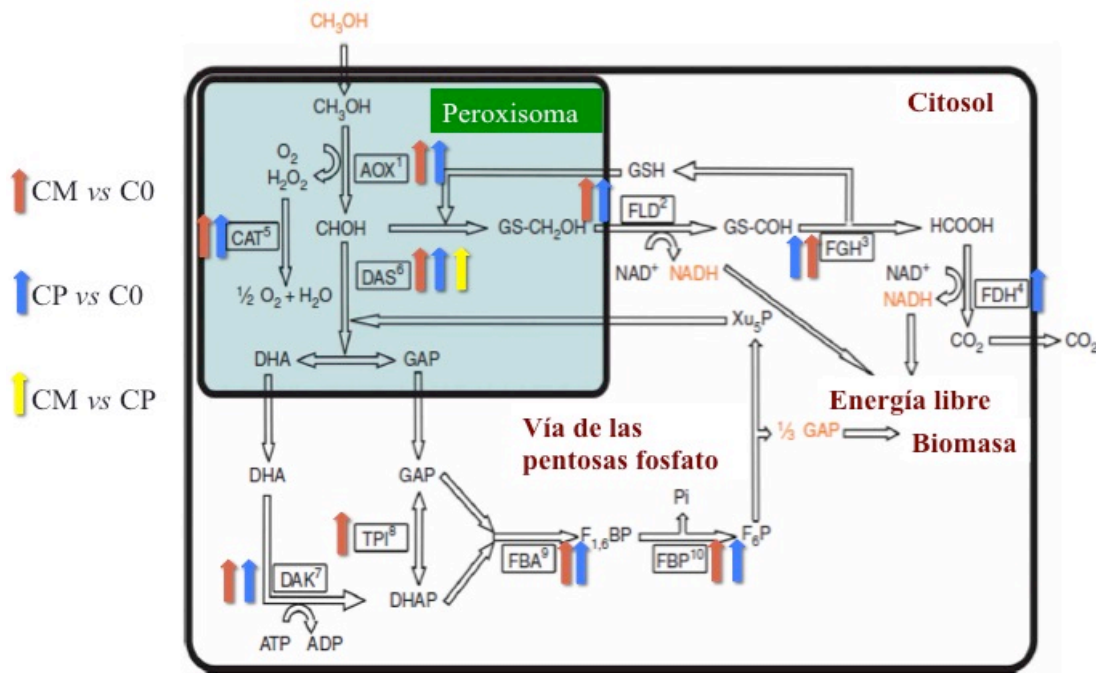
Gen	ID del Gen	log <sub>2</sub> (Fold_change)			Enzima	Función
		CM/C0	CP/C0	CM/CP		
<i>DAS1</i>	PAS_chr3_0832	4.7	3.1	1.6	Dihidroxiacetona sintasa 1	Transcetolasa, transfiere glicolaldehído de la xilulosa 5-fosfato al formaldehído para formar gliceraldehído fosfato (GAP) y dihidroxiacetona (DHA). Participa en la vía de pentosas fosfato
<i>DAS2</i>	PAS_chr3_0834	3.7	2	1.7	Dihidroxiacetona sintasa 1	Transcetolasa, transfiere glicolaldehído de la xilulosa 5-fosfato al formaldehído para formar gliceraldehído fosfato (GAP) y dihidroxiacetona (DHA). Participa en la vía de pentosas fosfato
<i>DAK</i>	PAS_chr3_0841	2	1.3	-	Dihidroxiacetona quinasa	Fosforila DHA para formar dihidroxiacetona fosfato. Requerida para la detoxificación de DHA.
<i>TPI</i>	PAS_chr3_0951	1	-	-	Tirosefosfato isomerasa	Enzima glicolítica regulada por disponibilidad de hierro
<i>FBP</i>	PAS_chr3_0868	2.1	1.6	-	Fructosa 1,6 bifosfatasa	Cataliza la desfosforilación de la frutosa 1,6-fosfato para formar frutosa 6-fosfato. Enzima reguladora en la vía de gluconeogénesis
<i>FBA2</i>	PAS_chr1-1_0319	5.1	4.4	-	Fructosa 1,6-bisfosfato aldolasa	Catálisis de la formación de fructosa 1,6-bisfosfato por la condensación de DHAP y GAP, Requerida para la gluconogénesis y glicólisis
<i>TKL1</i>		-1.8	-1.6	-	Transcetolasa	Encargada de la conversión de xilulosa 5-fosfato y ribosa 5-fosfato en sedoheptulosa 5-fosfato y gliceraldehído 3-fosfato en la vía de las pentosas fosfato.



**Tabla 22.** Nombre y función de las proteínas codificadas por los genes involucrados en la vía de desasimilación del metanol.

Gen	ID del Gen	log <sub>2</sub> (Fold_change)			Enzima	Función
		CM/C0	CP/C0	CM/CP		
<i>FLD</i>	PAS_chr3_1028	1.1	1.1	-	S-(hidroximetil) glutatión deshidrogenasa	Formaldehido deshidrogenasa citosólica dependiente de NAD y glutatión, transformación del formaldehido a CO <sub>2</sub> y NADH, producción de energía requerida para el crecimiento en metanol. Detoxificación del formaldehido.
<i>FGH</i>	PAS_chr3_0867	3.6	3.2	-	S-formilglutatión hidrolasa	Hidroliza el S-formilglutatión en formato y glutatión. Detoxificación del formaldehido y regeneración del glutatión.
<i>FDH</i>	PAS_chr3_0932	-	1.3	-	Formato deshidrogenasa dependiente de NAD(+)	Formaldehido deshidrogenasa citosólica dependiente de NAD, transformación del formaldehido a CO <sub>2</sub> y NADH, producción de energía requerida para el crecimiento en metanol. Detoxificación del formato.

Con el cálculo la actividad real del gen clave en la vía de asimilación *DAS* (suma de los niveles de expresión *DAS1* y *DAS2*) y su relación con los niveles de expresión con el gen clave en la vía de desasimilación *FLD* se determinó que vía fue más activa en términos de expresión. Los cocientes *DAS/FLD* en CM vs C0, CP vs C0 y CM vs CP fueron de 5.3, 2.5 y 2.6 respectivamente, lo cuál refleja que la vía de asimilación fue más activa durante el metabolismo del metanol en ambas condiciones de cultivo, y fue aun mayor bajo las condiciones de CM. Una representación gráfica de los diferentes patrones de expresión de todos los genes de la vía del metabolismo del metanol bajo las tres comparaciones estudiadas se muestra en la figura 60.



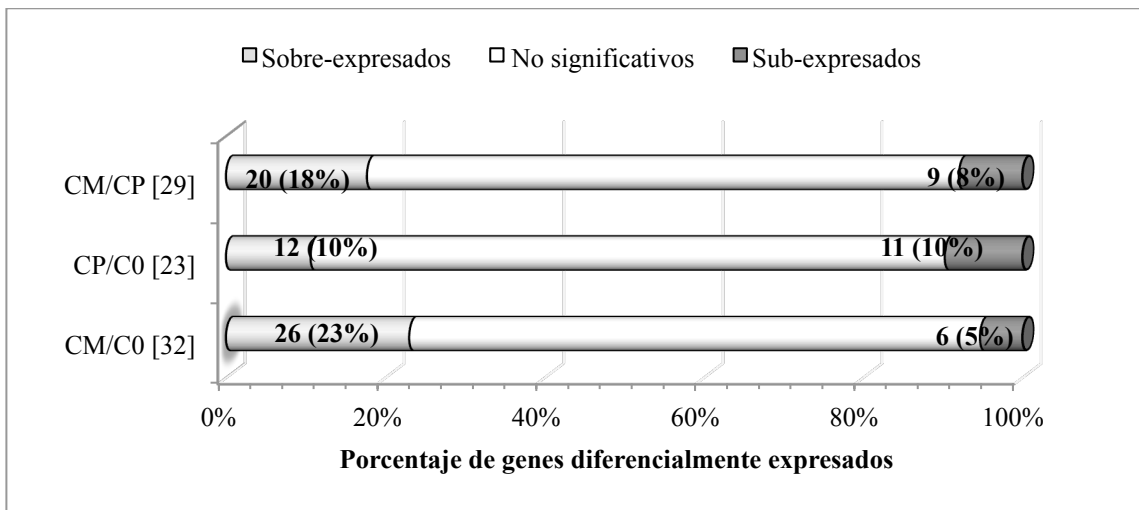
**Figura 60.** Vía del metabolismo del metanol para la generación de biomasa y energía.

Se indican en flechas rojas y azules los genes que se encontraron sobre-expresados en las comparaciones CM/C0 y CP/C0, respectivamente; mientras de color amarillo se muestran los genes sobre-expresados en la comparación CM/CP.

### 8.2.5.3.2 Análisis de expresión diferencial de genes involucrados en la vía de secreción

Los genes de la vía de secreción, resultaron diferencialmente expresados en las tres comparaciones realizadas (CM vs C0, CP vs C0 y CM vs CP), encontrándose

diferentes patrones de expresión de los 110 genes evaluados, los cuáles están representados en la figura 61.

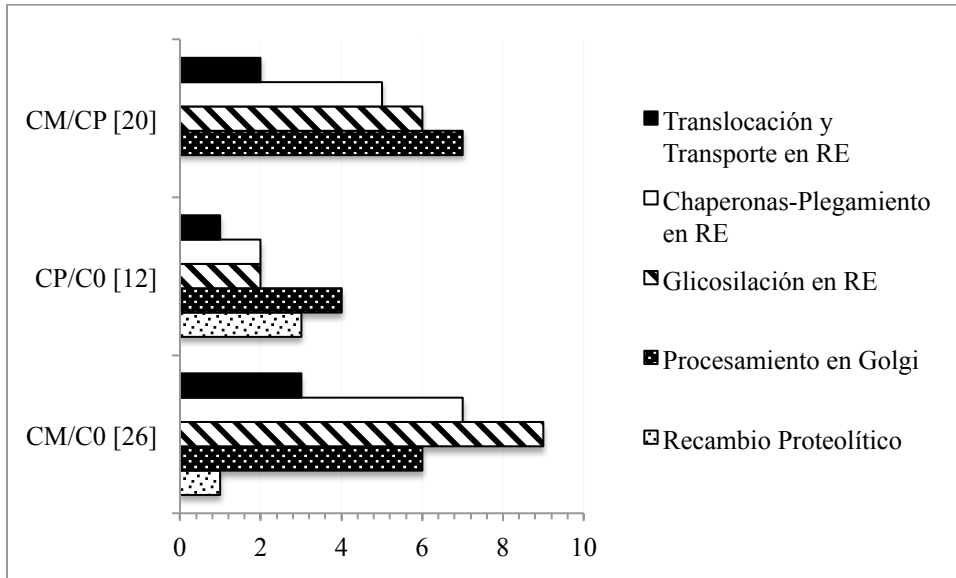


**Figura 61.** Expresión Global Diferencial de genes de la vía de secreción considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$ .

En el eje de las ordenadas se encuentra entre corchetes el número de genes totales diferencialmente expresados en cada comparación, además también se encuentra el número de genes sobre- y sub-expresados con su respectivo porcentaje entre paréntesis dentro de cada barra. En el eje de las abscisas se encuentra el porcentaje de genes sobre- ó sub-expresados.

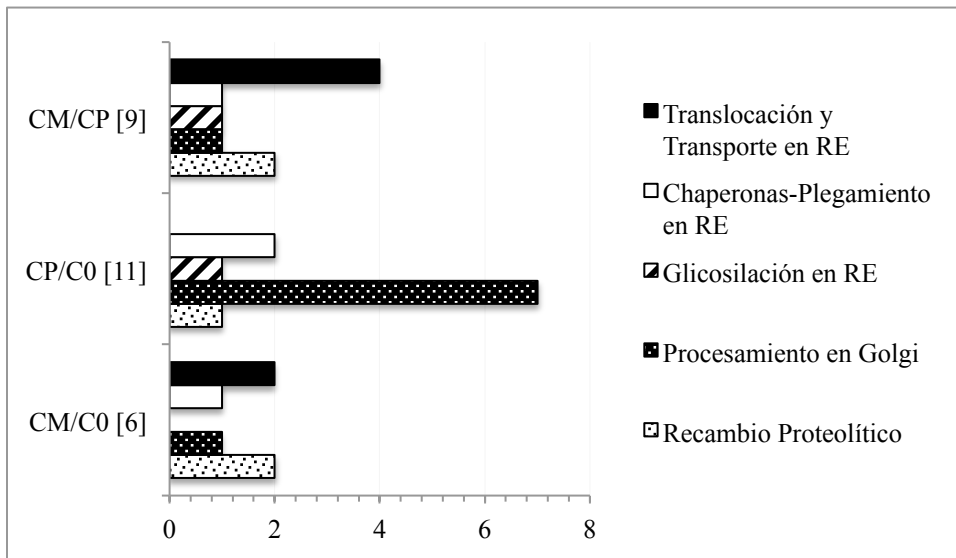
La comparación CM vs CP se caracteriza por tener un mayor número de genes sobre-expresados que en las demás comparaciones, indicando que la condición de alta producción presenta un mayor número de genes involucrados en secreción y activos durante la transcripción.

Los genes diferencialmente expresados se organizaron en 5 grupos diferentes con respecto a la función que realizan, en la figura 62 se muestra la distribución de los genes sobre-expresados en las tres comparaciones, mientras que en la figura 63 se muestra la distribución de los genes subexpresados.



**Figura 62.** Distribución del número de genes sobre-expresados correspondientes a la vía de secreción considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$  en las tres comparaciones.

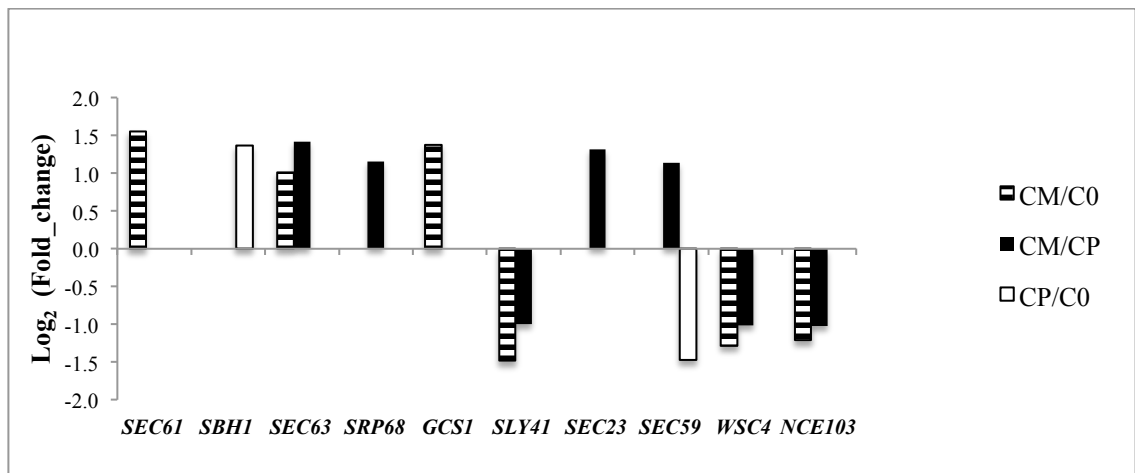
En el eje de las ordenadas se encuentra entre corchetes el número de genes totales sobre-expresados de la vía de secreción en cada comparación. En el eje de las abscisas se encuentra el número de genes de cada grupo.



**Figura 63.** Distribución del número de genes sub-expresados correspondientes a la vía de secreción considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$  en las tres comparaciones.

En el eje de las ordenadas se encuentra entre corchetes el número de genes totales sub-expresados de la vía de secreción en cada comparación. En el eje de las abscisas se encuentra el número de genes de cada grupo.

El grupo conformado por los genes que codifican para proteínas involucradas en la translocación y transporte de proteínas en el RE se muestra en la tabla 23. Los resultados muestran que mientras que *SEC61* se sobre-expresó en CM y *SBH1* se sobre-expresó en CP, sin diferencias entre las dos condiciones, *SEC63* se sobre-expresó en metanol y con diferencias entre las condiciones de cultivo (Figura 64). *SEC61* y *SBH1* forman parte del complejo *SEC61* que conforma el poro en el proceso de translocación en el RE, mientras *SEC63* junto con *SEC61*, forman parte del complejo SEC requerido en la translocación post-traduccional en lumen del RE. Por otro lado, la sobre-expresión de *SRP68* en CM, cuya proteína forma parte del complejo SRP, indica actividad presente en la translocación co-traduccional en lumen del RE. Estos resultados indican que tanto los procesos de translocación co-traduccional y post-traduccional presentan mayor actividad en las condiciones de CM.



**Figura 64.** Genes diferencialmente expresados considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$  involucrados en la translocación y transporte de proteínas.

*SLY41* se sub-expresó bajo las condiciones de CM por el cambio de la fuente de carbono y en CM vs CP, mientras que *GCSI* se sobre-expresó en CM en metanol, sin diferencia con CP, las proteínas codificadas por ambos genes participan en el transporte de RE a Golgi. Los genes SEC, implicados en el complejo translocación, *SEC23* y *SEC59* se sobre-expresaron en CM vs CP y además *SEC59* se sub-expresó en CP. *SEC23*, codifica una proteína activadora de GTPasa que estimula al actividad GTPasa de

SAR1p, componente del heterodímero SEC23p-SEC24p de la capa vesicular COPII involucrado en el transporte de RE a Golgi y SEC59 codifica una dolicol quinasa que cataliza el paso terminal en la biosíntesis de dolicol monofosfato (Dol-P), requerido para la viabilidad celular y para obtención de tasas normales de síntesis de intermediarios de lípidos, además de ser importante en la N-glicosilación de proteínas. Estos resultados sugieren que la sobre-expresión de *SEC59*, podría estar regulando de manera importante la N-glicosilación de proteínas impactando en la secreción en CM a diferencia de CP.

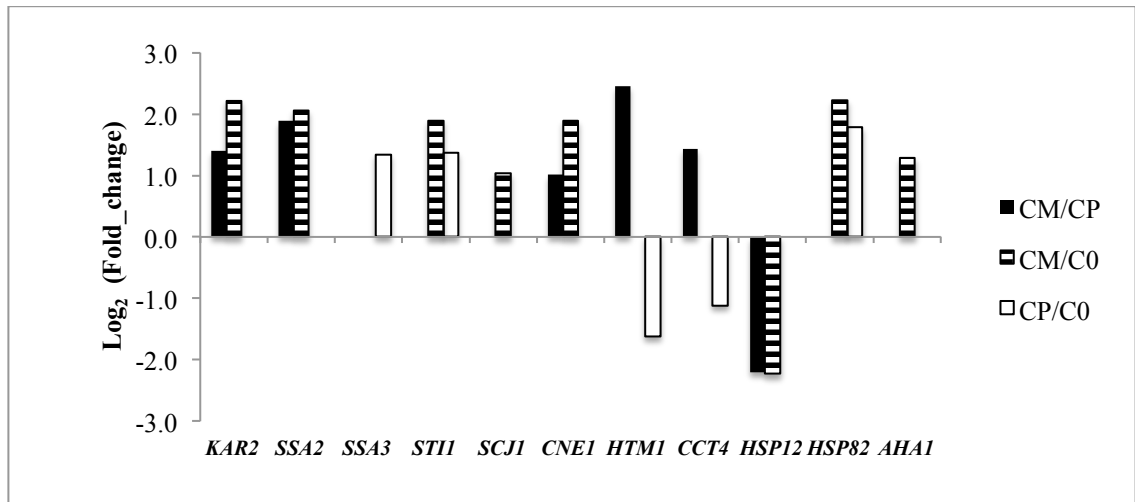
Tanto *WSC4* como *NCE103* se sub-expresaron en el cambio de la fuente de carbono en CM y bajo las condiciones de CM vs CP. *WSC4* además de participar en la translocación de las proteínas en RE, se piensa que es un componente de respuesta al estrés, por lo cual se puede asumir que la sub-expresión de este gen puede indicar un estrés no significativo bajo las condiciones de CM. *NCE103* está involucrado en la protección contra el daño oxidativo y el transporte no clásico de proteínas, se induce en la respuesta a bajas concentraciones de CO<sub>2</sub>.

En general se puede concluir que el proceso de translocación en RE esta más activo en CM justificando la sobre-expresión de muchos genes involucrados en la secreción de proteínas en esta condición y que la respuesta al estrés está menos activa en CM.

**Tabla 23.** Nombre y función de proteínas codificadas por los genes involucrados en la translocación y transporte de proteínas.

Gen	ID del gen	Log <sub>2</sub> (Fold_change)			Proteína	Función
		CM/C0	CP/C0	CM/CP		
<i>SEC61</i>	PAS_chr1-3_0202	1.6			Subunidad $\alpha$ del complejo de de translocación SEC61.	Forma el complejo SEC61 del poro de translocación de proteínas en el RE y retro-translocación de proteínas mal plegadas.
<i>SBH1</i>	PAS_chr2-2_0210		1.4		Subunidad $\beta$ del complejo del poro de translocación SEC61.	Forma parte del complejo SEC61 del poro de translocación de proteínas en el RE
<i>SEC63</i>	PAS_chr4_0395	1.0		1.4	Subunidad del complejo SEC63.	Forma parte del complejo SEC en la translocación post-traduccional junto con SEC61 en el RE.
<i>SRP68</i>	PAS_chr1-4_0134			1.2	Subunidad del complejo SRP (Partícula de reconocimiento de señal).	Forma parte del complejo SRP y participa en la translocación co-traduccional en lumen del RE.
<i>GCSI</i>	PAS_chr3_0974	1.4			GTPasa activadora del factor de ADP- ribosilación (ARF GAP).	Involucrada en el transporte RE-Golgi, requerida para la formación del poro de membrana.
<i>SLY41</i>	PAS_chr2-1_0457	-1.5		-1.0	Proteína transportadora no caracterizada.	Involucrada en el transporte RE a Golgi.
<i>SEC23</i>	PAS_chr4_0600			1.3	Proteína activadora de GTPasa	Involucrada en transporte de RE a Golgi y formación de vesículas COPII. Proceso de translocación en RE
<i>SEC59</i>	PAS_chr2-1_0498		-1.5	1.1	Dolicol kinasa	Biosíntesis de dolicol monofosfato (Dol-P), para la síntesis de oligosacárido para la N-glicosilación de proteínas. Proceso de translocación en RE
<i>WSC4</i>	PAS_chr3_0626	-1.3		-1.0	Proteína residente de la membrana del RE.	Involucrada en la translocación de proteínas solubles a secretarse y la inserción de proteínas en la membrana del RE, participa en la respuesta al estrés.
<i>NCE103</i>	PAS_chr4_0578	-1.2		-1.0	Anhidrasa carbónica.	Involucrada en la protección contra el daño oxidativo, transporte por la vía no convencional de proteínas carentes de secuencia señal.

Para el grupo de las proteínas chaperonas involucradas en el control de calidad del plegamiento de proteínas, se incluyeron genes que codifican chaperonas y chaperoninas residentes de la membrana del RE (Tabla 24) y algunas presentes en el citosol. La figura 65 muestra una representación gráfica de los genes sobre- ó sub- expresados estadísticamente significativos en las tres comparaciones.



**Figura 65.** Genes diferencialmente expresados que codifican chaperonas involucradas en el plegamiento de proteínas, con un  $\log_2(\text{Fold\_change}) \leq -1.0$  ó  $\geq 1.0$ .

El gen *KAR2* y *SSA2*, se sobre-expresaron en CM vs CP, ambos genes involucrados en el plegamiento, control de calidad y translocación en el retículo endoplásmico. *KAR2* codifica para una chaperona residente en RE que participa en la translocación a través del RE, en el plegamiento de proteínas en el lumen del RE, así como también regula la respuesta UPR mediante la interacción de IRE1p. Mientras que *SCJ1*, chaperona también residente en RE, coopera con *KAR2* en el plegamiento y ensamblaje de proteínas en el lumen del RE, no mostró diferencias entre CM y CP, pero si se sobre-expresó por el cambio de la fuente de carbono en CM. Los genes *STI1* y *SSA3* que codifican para proteínas de choque térmico importantes durante el plegamiento de proteínas en RE y que se activan en respuesta al estrés, se sobre-expresaron en CP, *SSA3* además juega un papel importante en la translocación co-traduccional de proteínas; *STI1* se sobre-expresó también en CM, éste gen activa la actividad ATPasa del grupo SSA que forman parte de las chaperonas HSP70 e inactiva



la actividad ATPasa de chaperonas HSP90. Además, no se mostro expresión diferencial de éstos genes entre las condiciones de cultivo.

*CNEI* y *HTMI(MNLI)* se sobre-expresaron en la comparación CM vs CP; la chaperona codificada por *CNEI*, participa en el ciclo de la calnexina, ciclo involucrado en el control de calidad de las proteínas N-glicosiladas plegadas en su conformación nativa, y la chaperona codificada por *HTMI(MNLI)*, es una exomanosidasa que extrae glicoproteínas mal plegadas, pero no glicoproteínas sometidas a plegamiento productivo del ciclo calnexina, está además directamente implicada en la degradación asociada al retículo endoplásmico (ERAD) mediante la unión a glicoproteínas mal plegadas para la degradación de una manera N-glicano dependiente; ambos genes podrían estar contibuyendo en los altos niveles de proteínas secretadas en CM. Es importante enfatizar que *HTMI (MNLI)*, sub-expresado en CP, podría indicar que el control de calidad en el ciclo de la calnexina puede estar afectado negativamente y con esto el proceso de secreción afectando negativamente en CP.

El gen *CCT4* que codifica para una chaperonina que ayuda al plegamiento de polipéptidos recién traducidos se sobre-expresó en la comparación CM vs CP y se sub-expresó en CP debido al cambio de la fuente de carbono, estos resultados indican que la expresión de *CCT4* juega un papel importante en los resultados obtenidos en CM.

*HSP12*, proteína de la membrana plasmática importante en el mantenimiento de la organización de la membrana durante condiciones de estrés, se sub-expresó en CM y en CM vs CP; y podría significar una mejor respuesta durante la adaptación a las condiciones de cultivo en CM a diferencia que CP.

El gen *HSP82* que codifica para una chaperona requerida durante el plegamiento de proteínas desnaturalizadas y encargada de regresarlas a su conformación nativa, constituye del 1-2% de las proteínas en el citosol y es inducida por calor y estrés; y la co-chaperona AHA1 se une a HSP82 para estimular su actividad ATPasa, aparentemente no contribuyeron al comportamiento diferencial de los cultivos CM y CP, sin embargo, parece haber mayor actividad de ambas en presencia del metanol en CM.

La sobre-expresión de los genes *SSA2*, *CNEI*, *KAR2*, *CCT4* y *HTMI* bajo las condiciones de CM vs CP, indican que en CM las condiciones de cultivo y la fuente de carbono actúan de forma conjunta para desencadenar una regulación diferencial hacia la

inducción de la expresión de genes que codifican para proteínas involucradas en el proceso de plegamiento y que este factor pudo haber sido determinante en una regulación post-traducciona que favoreciera a un mejor plegamiento, secreción y por lo tanto impacten en los niveles de producción bajo las condiciones de CM.

Es importante enfatizar que *CCT4* y *HTM1* podrían estar jugando un papel preponderante, ya que estos genes se sub-expresan bajo las condiciones de CP en presencia de metanol. Con estos resultados se podría asumir que la regulación de estos dos genes en CP podrían estar relacionados con los bajos niveles de producción bajo ésta condición.

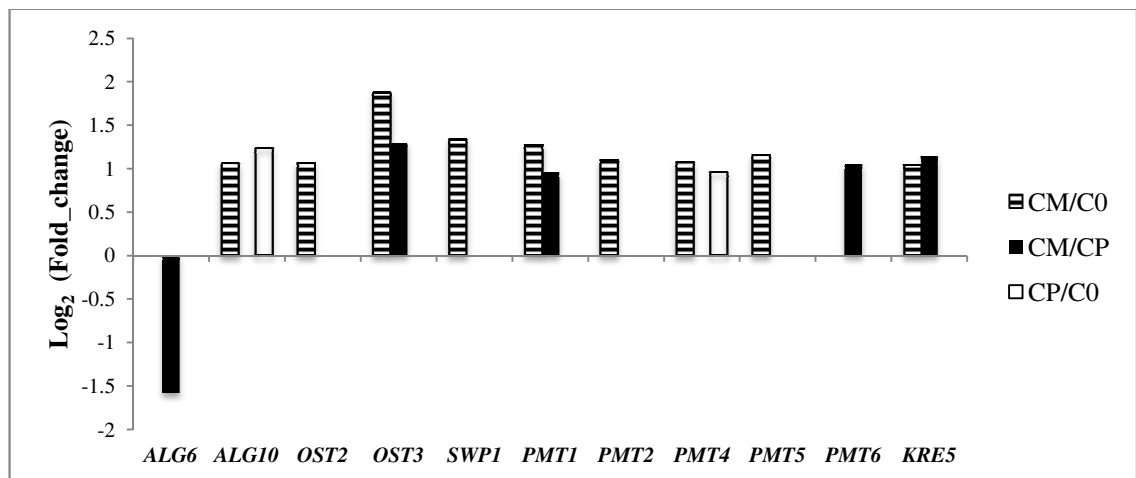
Otra observación que refuerza el concepto antes propuesto, es que con excepción de *SSA3*, *HSP12*, *CCT4* y *HTM1*, todos los demás genes analizados (*KAR2*, *SSA2*, *STII*, *SCJ1*, *CNE1* y *HSP82*), se sobre-expresan bajo las condiciones de CM, en el que el metanol está presente.

En conclusión la sobre-expresión de los genes *KAR2*, *SSA2*, *SCJ1*, *STII*, *CNE1* y *HSP82* en CM favorecen los niveles altos de producción de FTEII y su regulación es determinada por las condiciones de cultivo en CM, mientras que *HSP82* se sobre-expresa solo debido al cambio de la fuente de carbono, además *CCT4* y *HTM1* no afectaron a CM por el cambio de la fuente de carbono, sin embargo la sub-expresión de ambos en CP podrían estar influyendo en los niveles bajos de producción en estas condiciones.

**Tabla 24.** Nombre y función de genes que codifican chaperonas involucradas en el plegamiento de proteínas en el RE

Gen	ID del gen	log <sub>2</sub> (Fold_change)			Proteína	Función
		CM/C0	CP/C0	CM/CP		
<i>KAR2</i>	PAS_chr2-1_0140	2.2		1.4	ATPasa de la familia HSP70	Chaperona de membrana del RE, sobre-expresada en respuesta de proteínas mal plegadas, se requiere para la translocación de proteínas en lumen del RE para facilitar su plegamiento, es inducida por acumulación de proteínas en el RE y en respuesta al choque térmico; además regula UPR vía IRE1p
<i>SSA2</i>	PAS_chr4_0552	2.1		1.9	ATPasa de la familia HSP70	Chaperona citosólica involucrada en el plegamiento de proteínas y la señal de localización nuclear (NSL) - transporte nuclear. Además
<i>SSA3</i>	PAS_chr3_0230		1.3		ATPasa de la familia HSP70	Involucrada en el plegamiento de proteínas y la respuesta al estrés. Además juega un papel importante en la translocación co-traduccional de proteínas dependiente de SRP.
<i>STI1</i>	PAS_chr2-1_0518	1.9	1.4		Proteína del choque térmico familia HSP90	Co-chaperona que interactúa con las chaperonas de la familia HSP90 inhibiendo su actividad ATPasa e interactúa con la familia SSA (HSP70) induciendo su actividad ATPasa. Inducible con el estrés
<i>SCJ1</i>	PAS_chr1-3_0174	1.0			Chaperona residente en el RE, relacionada a Dna-J	Posee un dominio J soluble en el lumen del RE, coopera con KAR2 en el plegamiento y ensamblaje de proteínas en el lumen del RE.
<i>CNE1</i>	PAS_chr2-1_0322	1.9		1.0	Calnexina	Chaperona de membrana del RE involucrada en el plegamiento y control de calidad de glicoproteínas
<i>HTM1/ MNL1</i>	PAS_chr3_0891	-	-1.6	2.5	Exomanosidasa $\alpha$	Chaperona sensora de plegamiento que actúa sobre proteínas glicosiladas en el RE marcando a las proteínas destinadas a la degradación a los proteosomas por la vía de ubiquitina (ERAD)
<i>CCT4</i>	PAS_chr3_0571	-	-1.1	1.4	Subunidad 1 $\Delta$ del complejo de anillo de la chaperonina Cct	Chaperonina que ayuda al plegamiento de polipéptidos recién traducidos.
<i>HSP12</i>	PAS_chr4_0627	-2.2		-2.2	Proteína de choque térmico 12 kDa	Localizada en la membrana plasmática importante en el mantenimiento de la organización de la membrana durante condiciones de estrés; protege de la desecación y es inducida por choque térmico, estrés oxidativo y otros factores.
<i>HSP82</i>	PAS_chr1-4_0130	2.2	1.8		Proteína del choque térmico de la familia HSP90	Chaperona requerida para el plegamiento de proteínas desnaturalizadas y volverlas a su conformación nativa. Constituye del 1-2% de las proteínas en el citosol y es inducida por calor y stress. Interviene en la activación de proteínas reguladoras
<i>AHA1</i>	PAS_chr1-4_0072	1.3			co-chaperona de la familia HSP90	Se une a HSP82p e inicia la actividad ATPasa

El grupo de genes involucrados en la glicosilación de las proteínas en el Retículo Endoplásmico está formado principalmente por el complejo OST, que glicosila los sitios de N-glicosilación, los genes involucrados en el procesamiento temprano del N-glicano (glicanos tipo  $\text{Man}_8\text{GlcNAc}_2$ ) y los genes que codifican las O-manosiltransferasas involucradas en la O-glicosilación. En la figura 66 se representan los genes sobre- y sub-expresados estadísticamente significativos en las tres comparaciones, además se encuentran descritos en la tabla 25.



**Figura 66.** Genes diferencialmente expresados involucrados en la síntesis de glicoproteínas en RE, considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$ .

*ALG10* se sobre-expresó en ambas condiciones de cultivo en el cambio de la fuente de carbo, sin diferencias significativas en CM vs CP, por otro lado *ALG6* se sub-expresó en CM vs CP. *ALG10* codifica una glucosiltransferasa involucrada en el paso terminal de la glicosilación de los oligosacáridos  $\text{Glc}(2)\text{Man}(9)\text{GlcNAc}(2)$  unidos a lípidos y *ALG6*, codifica una  $\alpha$ -1,3 glucosiltransferasa, involucrada en la transferencia de oligosacáridos del dolicil-pirofosfato a los residuos de asparagina de las proteínas durante la N-glicosilación.

Los tres genes relacionados con la N-glicosilación analizados, *OST2*, *OST3* y *SWP1*, se sobre-expresaron en CM debido al cambio de la fuente de carbono, pero solo *OST3* mostró expresión diferencial significativa en CM vs CP. *OST3*, codifica la subunidad  $\gamma$  del complejo oligosacariltransferasa del lumen de RE, cataliza la glicosilación ligada a

asparagina de proteínas recién sintetizadas, además OST3p es importante en la N-glicosilación.

Los resultados indican que la actividad de N-Glicosilación fue más activa en CM que en CP y que *OST3* puede ser determinante en la regulación diferencial de esta vía que contribuyó a la diferencia de los niveles de secreción entre CM y CP.

Cuatro de los genes (*PMT1*, *PMT2*, *PMT4* y *PMT5*) implicados en la O-glicosilación evaluados, se sobre-expresaron en CM, *PMT1* y *PMT6* se sobre-expresaron en CM vs CP, y *PMT4* se sobre-expresó en CP. *PMT1*, *PMT4* y *PMT6*, codifican una O-manosiltransferasa que transfiere residuos de manosa del dolicil fosfato-D-manosa a residuos serina/treonina de proteínas; además, *PMT1* está involucrada en el control de calidad del RE. Los resultados indican una mayor actividad de O-glicosilación en CM vs CP, además *PMT1* podría estar jugando un papel muy importante en el control de calidad de ésta vía contribuyendo en los niveles de secreción alcanzados en CM.

Estudios (Ecker *et al.*, 2003) sobre la manosilación de residuos serina/treonina (O-glicosilación) llevada a cabo por las proteínas PMT, específicamente *PMT1*, *PMT2* y *PMT4*, indican que la o-glicosilación, precede y potencialmente controla la N-glicosilación de las glicoproteínas de la pared celular de las levaduras. Por otro lado Goder y Melero (2011) sugieren que el complejo *PMT1/PMT2* tienen además función de chaperona involucradas en dirigir proteínas mal plegadas a ERAD.

Éstos resultados sugieren que las proteínas PMT podrían estar jugando un papel importante en el plegamiento de las proteínas y en condiciones fisiológicas más favorables en CM que en CP .

*KRE5* codifica una proteína requerida para la biosíntesis de  $\beta$ -1,6-glucanos en *Saccharomyces cerevisiae* y es un homólogo putativo de la UDP-glucosa: glucoproteína glucosiltransferasa presente en *Pichia pastoris* y que participa en el ciclo de calnexina, ciclo relacionado con el control de calidad del plegado de las proteínas. *KRE5* fue sobre-expresado bajo las condiciones de CM, indicando que el proceso de control de calidad de plegamiento de proteínas a través del ciclo de calnexina está más activo en CM que en CP.

**Tabla 25.** Nombre y función de las proteínas codificadas por los genes involucrados en la síntesis de glicoproteínas en RE.

Gen	ID del gen	log <sub>2</sub> (Fold_change)			Proteína	Función
		CM/C0	CP/C0	CM/CP		
<i>ALG6</i>	PAS_chr2-1_0549			-1.6	Alfa-1,6 glucosiltransferasa	Biosíntesis del precursor N-glicanos en el RE (lípidos-dolicol oligosacáridos, LLO) para la N-glicosilación de proteínas nacientes.
<i>ALG10</i>	PAS_chr1-4_0475	1.1	1.2		Alfa-1,2 glucosiltransferasa	Biosíntesis del precursor N-glicanos en el RE (lípidos-dolicol oligosacáridos, LLO) para la N-glicosilación de proteínas nacientes.
<i>OST2</i>	PAS_chr2-2_0346	1.1			Subunidad del complejo oligosacariltransferasa en ER	Transferencia de oligosacárido de manosa a partir de un lípido-dolicol oligosacárido a residuos de asparagina de proteínas sintetizadas en el RE.
<i>OST3</i>	PAS_chr4_0610	1.9		1.3	Subunidad del complejo oligosacariltransferasa en ER	Transferencia de oligosacárido de manosa a partir de un lípido-dolicol oligosacárido a residuos de asparagina de proteínas sintetizadas en el RE.
<i>SWP1</i>	PAS_chr1-3_0248	1.3			Subunidad del complejo oligosacariltransferasa en ER	Transferencia de oligosacárido de manosa a partir de un lípido-dolicol oligosacárido a residuos de asparagina de proteínas sintetizadas en el RE.
<i>PMT1</i>	PAS_chr2-1_0212	1.3		1.0	Manosiltransferasa- Dolicil--fosfato-manosa 1	Transfiere residuos de manosa del dolicil fosfato D-manosa a residuos de serina/treonina (O-glicosilación).
<i>PMT2</i>	PAS_chr2-1_0256	1.1			O-manosiltransferasa	O-glicosilación
<i>PMT4</i>	PAS_chr1-4_0033	1.1	1.0		O-manosiltransferasa	O-glicosilación, además actúa en complejo con PMT1p
<i>PMT5</i>	PAS_chr1-1_0286	1.2			O-manosiltransferasa 1	O-glicosilación
<i>PMT6</i>	PAS_chr4_0777			1.0	Proteína O-manosiltransferasa	O-glicosilación
<i>KRE5</i>	PAS_chr3_0929	1.0		1.1	UDP-glucosa, glicoproteína glucosiltransferasa	Biosíntesis de beta-1,6 glucano

En conclusión, los resultados muestran que bajo las condiciones de CM se presenta una mayor actividad transcripcional de los genes implicados en el procesamiento de las proteínas a través del RE durante la N- y O-glicosilación que permiten un estado fisiológico favorable para el plegamiento y secreción de proteína heteróloga.

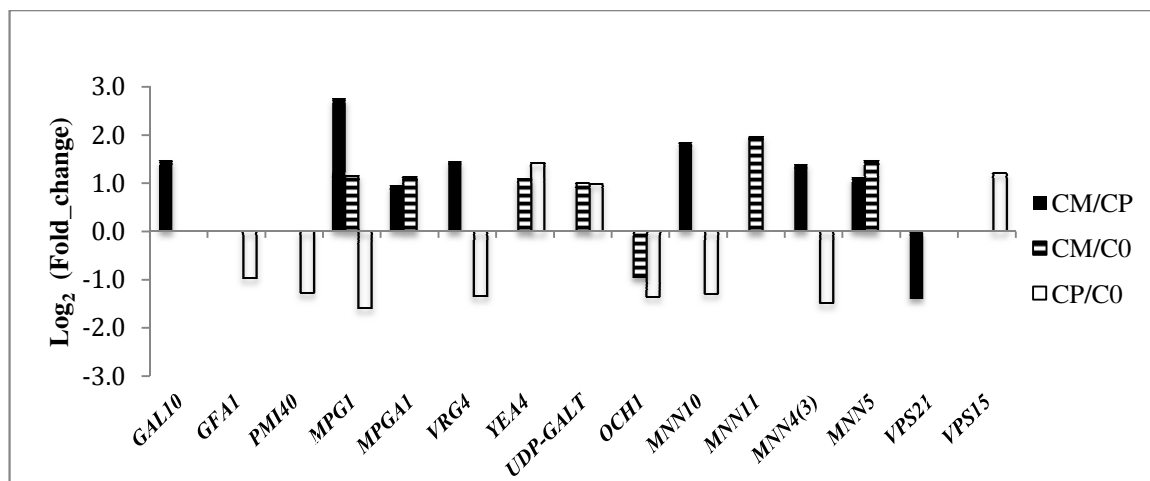
El análisis del grupo conformado por los genes involucrados en el procesamiento de glicoproteínas y algunos genes involucrados en el tráfico de proteínas en Golgi (Tabla 26), demuestra una alta actividad transcripcional de genes relacionados con el procesamiento de glicoproteínas en Golgi (Figura 67).

La transcripción de los genes analizados involucrados con la síntesis de nucleótidos glicosilados fue más activa en CM que en CP, *GAL10*, *MPG1* *MPGAI* se sobre-expresaron en CM vs CP, mientras que *MPG1* (síntesis de GDP-manosa.), *GAF1* (síntesis de UDP- GlcNAc), *PMI40* (síntesis de GDP-manosa) se sub-expresaron en CP en presencia de metanol, *MPG1*, codifica una GDP manosa pirofosforilasa importante durante la biosíntesis de la pared celular, que parece de gran importancia en CM.

Mientras que los genes *YEA4* y *UDP-GalT*, genes involucrados en el transporte de N-acetilglucosamina y galactosa, respectivamente, se sobre-expresaron en ambos casos, *VRG4*, transportador de GDP-manosa que regula la función de Golgi y la glicosilación se sobre-expresó diferencialmente en CM vs CP y se sub-expresó bajo las condiciones de CP en metanol. Estos resultados indican que el proceso de manosilación es afectado, probablemente con menor actividad, en las condiciones de CP; además *VRG4* podría jugar un papel importante en la regulación del aparato de Golgi, influyendo en la glicosilación en CM.

Mientras que *OCH1* (inicio de la ramificación  $\alpha$ -1,6 en la hipermanosilación) se sub-expresó y *MNN11* (complejo de la manan polimerasa II) se sobre-expresó en metanol en ambas condiciones sin diferencias. *MNN10*, codifica una subunidad del complejo manosil transferasa que media la elongación de los esqueletos de manano de los oligosacáridos, *MNN4*, que es un regulador positivo de MNN6p durante la correcta fosforilación de oligosacáridos ligados a N y *MNN5*, importante en el inicio de la ramificación  $\alpha$ -1,2 en la hipermanosilación, se sobre-expresaron en CM vs CP. Estos resultados indican que la manosilación de glicoproteínas con ramificaciones  $\alpha$ -1,2

podrían estar presentes con mayor frecuencia en CM que en CP. En particular, la sobre-expresión de *MNN4* en CM, podría indicar una manosilación de O-glicoproteínas más activa bajo esta condición.



**Figura 67.** Genes diferencialmente expresados involucrados en el procesamiento y tráfico de glicoproteínas en aparato de Golgi, considerando un  $\log_2(\text{Fold\_change}) \leq -1.0$  ó  $\geq 1.0$ .

Los genes *VPS21* y *VPS15* codifican para proteínas involucradas en la clasificación de proteínas vacuolares, definen el tráfico de hidrolasas solubles vacuolares en la ruta del Golgi tardío a la vacuola. El gen *VSP21* se sub-expresó en CM vs CP, *VPS15* se sobre-expresó en CP en respuesta al metanol, lo cual sugiere que el tráfico y distribución de proteínas de Golgi a vacuolas es más activo en CP. La menor actividad de estos genes en CM, podrían ser algunos de los factores que expliquen la menor actividad proteolítica que se presentó en ésta condición influyendo la mayor producción de la proteína.

En conclusión, la regulación del procesamiento en el aparato de Golgi en CM, influyó en la sobre-producción de la proteína heteróloga en ésta condición; además un factor que contribuyó en explicar la menor actividad proteolítica en CM, fue el comportamiento de los genes que codifican proteínas que detienen el tráfico de hidrolasas solubles vacuolares en Golgi tardío.

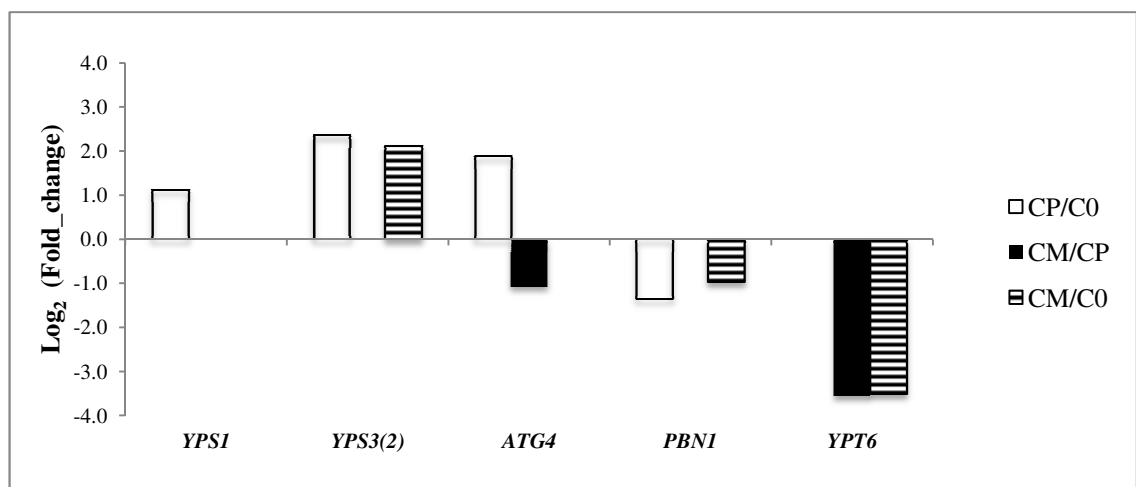


**Tabla 26.** Nombre y función de proteínas codificadas por genes involucrados en el procesamiento y tráfico de glicoproteínas en aparato de Golgi.

Gen	ID del gen	log <sub>2</sub> (Fold_change)			Proteína	Función
		CM/C0	CP/C0	CM/CP		
<i>GAL10</i>	PAS_chr4_0839			1.5	UDP-glucosa-4-epimerasa	Síntesis de nucleótidos glicosilados, síntesis de UDP-Gal.
<i>GFA1</i>	PAS_chr2-1_0626		-1.0		Glutamina-fructosa-6-fosfato amidotransferasa	Síntesis de nucleótidos glicosilados, síntesis de UDP- GlcNAc.
<i>PMI40</i>	PAS_chr3_1115		-1.3		Manosa-6-fosfato isomerasa	Síntesis de nucleótidos glicosilados, síntesis de GDP-manosa.
<i>MPG1</i>	PAS_chr3_0870	1.2	-1.6	2.8	GDP- manosa pirofosforilasa	Síntesis de nucleótidos glicosilados, síntesis de GDP-manosa.
<i>MPGA1</i>	PAS_chr2-1_0093	1.1		1.0	Manosa-1-fosfato guaniltransferasa alfa	Síntesis de nucleótidos glicosilados, síntesis de GDP-manosa.
<i>VRG4</i>	PAS_chr3_0916		-1.3	1.5	Transportador de GDP-manosa	Transporte de GDP-manosa en la glicosilación en Golgi.
<i>YEA4</i>	PAS_chr1-3_0163	1.1	1.4		Transportador de UDP-GlcNAc	Transporte de UDP-GlcNAc en la glicosilación en Golgi.
<i>UDP-GALT</i>	PAS_chr4_0810	1.0	1.0		Transportador de UDP-Gal	Transporte de UDP-Gal en la glicosilación en Golgi.
<i>OCH1</i>	PAS_chr1-3_0251	-1.0	-1.4		$\alpha$ -1,6 Manosiltransferasa	Manosilación de glicoproteínas, incorpora la manosa inicial de la ramificación $\alpha$ -1,6 en la hipermanosilación.
<i>MNN10</i>	PAS_chr2-2_0185		-1.3	1.9	Subunidad del complejo de la manan polimerasa II	La manan polimerasa II cataliza la elongación del oligosacárido de manano (hipermanosilación) en glicoproteínas (maduración de glicanos).
<i>MNN11</i>	PAS_chr2-2_0125	2.0	2.0		Subunidad del complejo de la manan polimerasa II	Elongación del oligosacárido de manano (hipermanosilación) en glicoproteínas .
<i>MNN4(3)</i>	PAS_chr2-1_0706		-1.5	1.4	Subunidad del complejo manosiltransferasa en Golgi	Media la elongación del oligosacárido de manano en O-glicanos través de la regulación de KTR6.
<i>MNN5</i>	PAS_chr3_0370	1.5		1.1	$\alpha$ -1,2 manosiltransferasa	Manosilación de glicoproteínas, incorpora la manosa inicial de la ramificación $\alpha$ -1,2 en la hipermanosilación.
<i>VPS21</i>	PAS_chr4_0219			-1.4	GTPasa homóloga a Rab5 de mamíferos	Transporte y distribución de proteínas de Golgi tardío a la vacuola, está implicada en el tráfico prevacuolar al compartimento endosomal tardío.
<i>VPS15</i>	PAS_chr3_0042		1.2		Serina/treonina proteína cinasa	Distribución de proteínas vacuolares en Golgi tardío a la vacuola, está implicada en el tráfico prevacuolar y el compartimento endosomal tardío.

El grupo de genes analizados involucrados en el mantenimiento de la homeostasis intracelular y en la degradación o recambio proteico en sistema vacuolares son descritos en la tabla 27 y su expresión se muestra en la figura 68. *YPS1* se sobre-expresó solo en CP por el cambio de la fuente de carbono, mientras que *YPS3(2)* se sobre-expresó en ambas condiciones sin diferencias, indicando un comportamiento similar de estos genes de la familia de proteasas yapsina involucradas en el mantenimiento de la homeostásis e integridad de la pared celular. *YPS1* está relacionado con la participación de *KEX2* en el procesamiento del factor alfa y algunos autores (Warten & De Wolf, 2005) indican que la disfunción de *YPS1* disminuye la proteólisis celular y favorece al incremento de la producción de proteínas heterólogas, en esta caso la sobre-expresión de este gen en CP podría estar afectando los niveles de producción de la proteína heteróloga.

Por otro lado *ATG4*, gen involucrado en el proceso autofagia por la vía CVT (citoplasma-vacuola), presenta una expresión diferencial CM vs CP sub-expresado en CM (Figura 68), asumiendo con este resultado una mayor actividad autofágica en CP. *PBNI* se sub-expresó tanto en CM como en CP sin diferencias, con los cual se puede inferir que la actividad de la proteasa B (PrB) podría estar afectada en ambas condiciones.



**Figura 68.** Genes diferencialmente expresados que codifican para proteínas implicadas mantenimiento de la homeostasis intracelular y en la degradación o recambio proteolítico en sistemas vacuolares, considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$ .

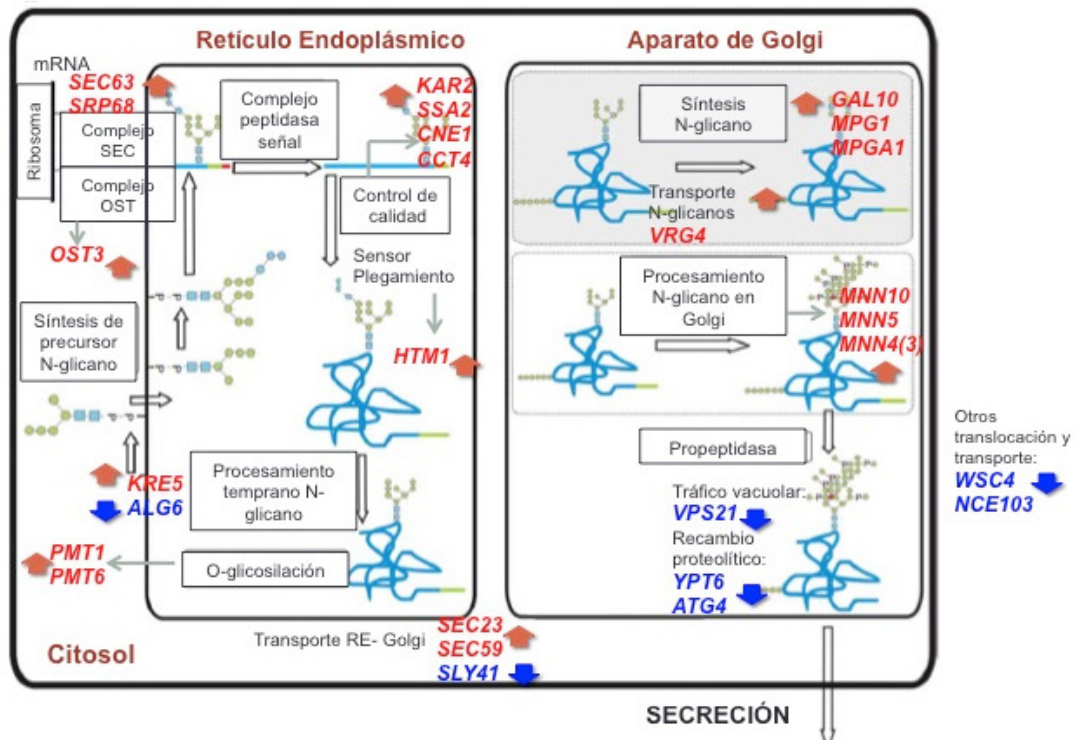
*YPT6* sub-expresado en CM en el cambio de la fuente de carbono y en CM vs CP, está involucrado en el transporte vesicular entre el compartimento endosomal y aparato de Golgi tardío, también está involucrado con *KEX2* y el procesamiento del factor alfa. De acuerdo a los resultados, la sub-expresión de *YPT6* podría indicar que la actividad vacuolar esta disminuida en CM, influyendo en el proceso de secreción a través de la misma.

En general se puede concluir que los genes relacionados con la actividad vacuolar en el proceso de recambio o degradación de proteínas pueden estar disminuidos en CM.

**Tabla 27.** Nombre y función de proteínas codificadas por genes involucrados en la homeostásis intracelular y degradación o recambio proteolítico en sistemas vacuolares.

Gen	ID del gen	log <sub>2</sub> (Fold_change)			Proteína	Función
		CM/C0	CP/C0	CM/CP		
<i>YPS1</i>	PAS_chr4_0584		1.1		Aspartil proteasa de la vía de secreción (ligada a glicosilfosfatidil inositol) de la familia de las yapsinas.	Mantenimiento de la integridad de la pared celular homeostasis de glucanos de la pared celular, implicada junto con KEX2 en el procesamiento del precursor del factor alfa.
<i>YPS3(2)</i>	PAS_chr3_0866	2.1	2.4		Aspartil proteasa de la vía de secreción (ligada a glicosilfosfatidil inositol) de la familia de las yapsinas	Mantenimiento de la integridad de la pared celular homeostasis de los glucanos de la pared celular.
<i>ATG4</i>	PAS_chr1-4_0522		1.9	-1.1	Cisteín proteasa	Involucrada en el procesamiento de proteínas y en la autofagia por la vía CVT (citoplasma-vacuola targeting).
<i>PBN1</i>	PAS_chr4_0463	-1.0	-1.4		Componente esencial de la glicosilfosfatidilinositol-manosiltransferasa I	Interviene en el proceso autocatalítico post-traducciona de PrB1p, precursor de la proteasa B (PrB), una serin endoproteinasa vacuolar análoga a KEX2. PrB interviene en la degradación de proteínas y péptidos durante la inanición por fuente de nitrógeno o carbono, autofagia, meiosis y esporulación y en la degradación de proteínas de membrana estimulada de forma constitutiva o por condiciones ambientales.
<i>YPT6</i>	PAS_chr4_0164	-3.5		-3.6	Homologo en levaduras a la familia de las GTPaseas Rab de mamíferos	Participa en procesos implicados en el tráfico de membrana, como son la maduración de vacuolas y el reciclaje de proteínas de membrana, asociado al aparato de Golgi interviene en la fusión de vesículas de los endosomas al aparato de Goli.

Una representación gráfica de los diferentes patrones de expresión de todos los genes de la vía de secreción bajo las tres comparaciones evaluadas se muestra en la figura 69.



**Figura 69.** Genes sobre-expresados con un log en la vía de secreción de proteínas.

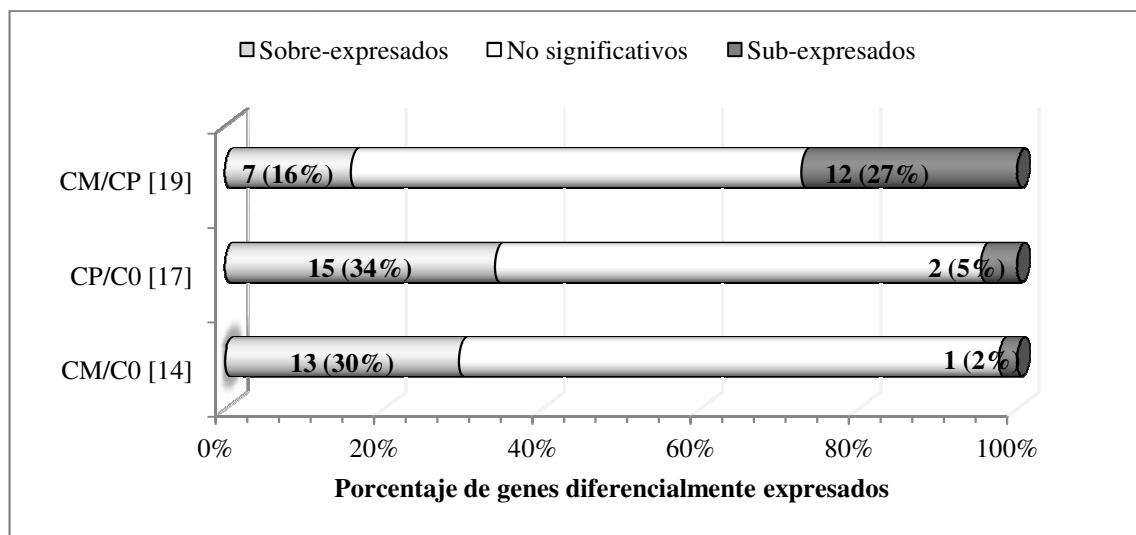
Se indican en flechas rojas aquellos genes sobre-expresados en la comparación CM/CP; las flechas azules indican aquellos genes sub-expresados en la comparación CM/CP.

### 8.2.5.3.3 Análisis de expresión diferencial de genes involucrados en el estrés en RE durante la secreción

Hasta el momento es bien establecido que la sobre-expresión de proteínas heterólogas está conectada con diferentes procesos en respuesta al estrés en RE durante la secreción, una de estas reacciones es la respuesta a las proteínas mal plegadas o UPR (Mattanovich et al., 2004).

La respuesta al estrés celular debido a las proteínas mal plegadas o UPR involucra la participación de chaperonas relacionadas con el control de calidad y el soporte al proceso de plegamiento y está conectada con la degradación de proteínas vía ERAD.

ERAD es un proceso por el cual, las proteínas mal plegadas en el RE, se redirigen hacia el traslocón para su translocación hacia el citosol, en donde son sujetas a la degradación proteosomal. El proceso de autofagia y los eventos de fusión de vesículas durante la autofagia mediado por SNARE's son también eventos relacionados con el proceso de secreción. Los genes relacionados con UPR, en el complejo proteico del proteosoma mediado por ERAD, SNARE y autofagia fueron considerados en este grupo. En la figura 70 se muestran los patrones de expresión de los 44 genes evaluados en las tres comparaciones realizadas.

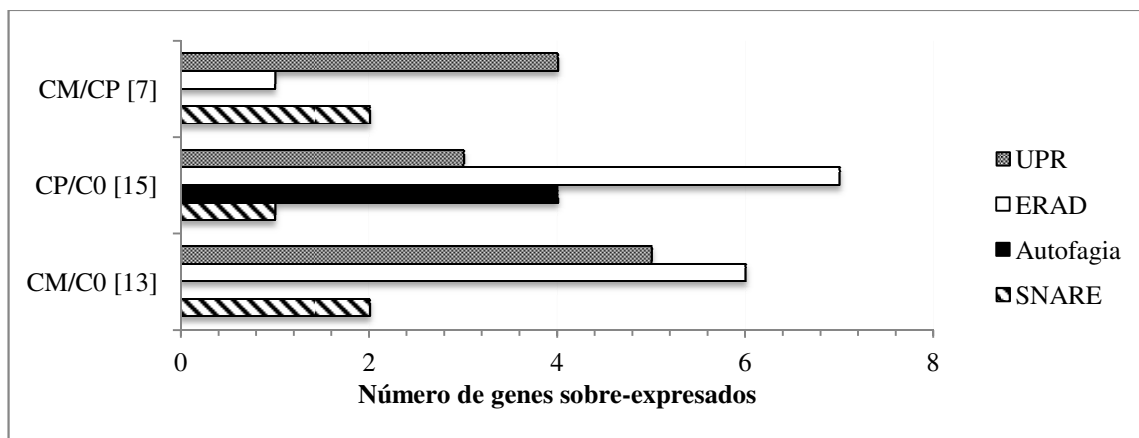


**Figura 70.** Expresión Global diferencial de otros genes indirectamente relacionados con la secreción considerando un  $\log_2(\text{Fold\_change}) \leq -1.0$  ó  $\geq 1.0$ .

En el eje de las ordenadas se encuentra entre corchetes el número de genes totales diferencialmente expresados en cada comparación, además también se encuentra el número de genes sobre- y sub-expresados con su respectivo porcentaje entre paréntesis dentro de cada barra. En el eje de las abscisas se encuentra el porcentaje de genes sobre- ó sub-expresados.

En el gráfico se muestra que hubo mayor número de genes sobre-expresados en la condición CP por efecto del cambio a metanol y genes sub-expresados en CM respecto a CP, estos resultados podrían indicar que el estrés en RE es menor en CM que en CP.

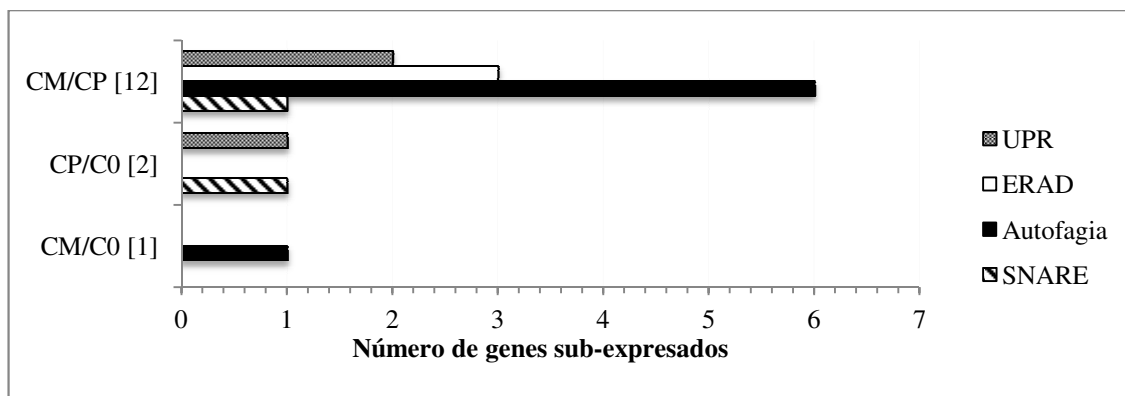
Estos 44 genes se organizaron en 6 grupos diferentes con respecto los procesos o vías involucrados que realizan, los genes sobre- y sub-expresados (Figura 71 y 72) se representan a continuación.



**Figura 71.** Distribución del número de genes sobre-expresados correspondientes a otros genes indirectamente relacionados con la secreción considerando un  $\log_2 ) \leq -1.0$  ó  $\geq 1.0$  en las tres comparaciones.

En el eje de las ordenadas se encuentra entre corchetes el número de genes totales sobre-expresados de otros genes indirectamente relacionados a la secreción en cada comparación. En el eje de las abscisas se encuentra el número de genes de cada grupo.

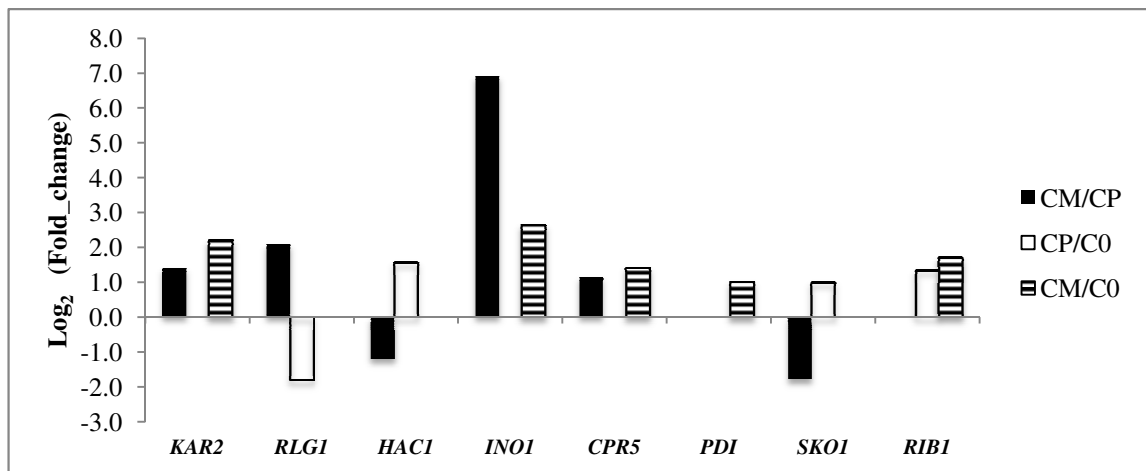
Los resultados muestran que los genes involucrado en los procesos de UPR y ERAD se sobre-expresan en CM por el cambio a metanol y a su vez una expresión diferencial de estos genes al comparar CM vs CP. Los procesos de ERAD, autofagia y respuesta al estrés ambiental fueron sobresalientes en CP por el cambio de la fuente de carbono, los resultados confirman un mayor estado de estrés en CP que en CM.



**Figura 72.** Distribución del número de genes sub-expresados correspondientes a otros genes indirectamente relacionados con la secreción considerando un  $\log_2 (\text{Fold\_change}) \leq -1.0$  ó  $\geq 1.0$  en las tres comparaciones.

En el eje de las ordenadas se encuentra entre corchetes el número de genes totales sub-expresados de otros genes indirectamente relacionados a la secreción en cada comparación. En el eje de las abscisas se encuentra el número de genes de cada grupo.

En la célula, la sobrecarga de la ruta de secreción es sensada en el RE por la unión de BiP, producto de *KAR2*, a proteínas en proceso de plegamiento o proteínas mal plegadas, activándose la vía de la respuesta a proteínas mal plegadas (UPR). La UPR puede ser activada por diversas condiciones que causan estrés en el RE, unos de los principales componentes de UPR son *KAR2* (a través de Ire1p), *RLG1* y *HAC1* (Tabla 28). De estos genes, *KAR2* y *RLG1* se sobre-expresaron diferencialmente en CM vs CP, mientras que *KAR2* se sobre-expresó en CM, *RLG1* se sobre-expresó en CP al cambiar al metanol (Figura 73). Para completar el proceso se requiere de *HAC1*, que codifica para *HAC1p* siendo ésta la proteína que finalmente regula la UPR, la cual se sub-expresó en CM vs CP, efecto obtenido por la sobre-expresión en CP por el cambio de la fuente de carbono. Por otro lado, *CPR5*, se sobre-expresó en CM vs CP, expresión diferencial causada al cambio de la fuente de carbono en CM (Figura 73). Estos resultados indican que UPR está más activa en CM que en CP, promoviendo el plegamiento de proteínas y probablemente evitando el acumulamiento de proteínas mal plegadas en CM.



**Figura 73.** Genes diferencialmente expresados involucrados en la respuesta a proteínas mal plegadas (UPR), considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$ .

*HAC1* no se induce en CM por el cambio de la fuente de carbono y por lo tanto durante la expresión de la proteína heteróloga, al menos en el tiempo evaluado, este comportamiento indica una regulación diferente entre los tres principales genes que participan en el proceso de la UPR. Algunos autores (Yoshida *et al.*, 2000) afirman que el factor transcripcional codificado por el gen *HAC1* es importante para activar la



respuesta UPR y otros autores (Gasser *et al.*, 2007) afirman que su co-expresión tiene un efecto positivo en la secreción de proteínas heterólogas. La sobre-expresión en CP podría causar una respuesta conocida como respuesta super-UPR mencionada por algunos autores (Bernaes *et al.*, 2006).

El gen *INO1* se sobre-expresó en CM en respuesta al cambio de la fuente de carbono y a las condiciones de cultivo. Éste gen está involucrado en la síntesis de inositol fosfato, además, se ha visto que *HAC1* e *IRE1* regulan la inducción de *INO1* y el metabolismo de fosfolípidos en respuesta al estrés en ER de una manera compleja en levaduras. Este gen podría ser un indicador de la regulación de UPR por parte de los genes *HAC1* e *IRE1* en CM vs CP a pesar de que ambos genes no se sobre-expresan en ésta condición.

*PDII* presenta mayor actividad transcripcional en CM, sin ser significativa al comparar con CP. Estudios previos (Kauffman *et al.*, 2001; Mattanovich *et. al.*, 2004) han reportado que HAC1p actúa como factor transcripcional para la inducción de PDI y al menos al tiempo evaluado no concuerda esta relación tanto en CM como en CP. Por otro lado, el plegamiento que compromete la formación de enlaces disulfuro, no ha sido determinante bajo las condiciones evaluadas en términos de expresión diferencial significativa.

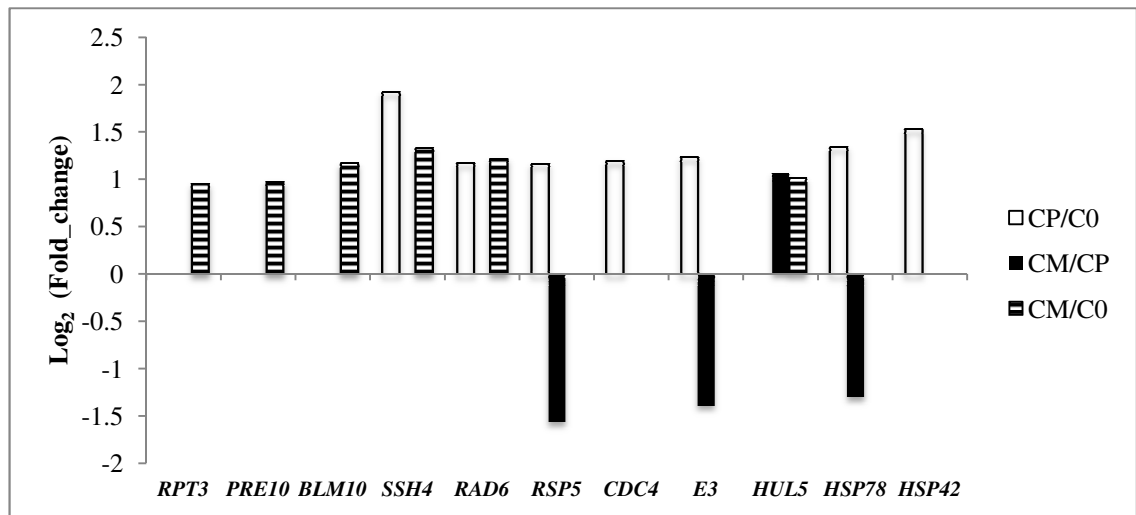
*SKO1* se sobre-expresó en CP por al cambio de la fuente de carbono, y se sub-expresó diferencialmente en CM vs CP, este gen codifica para un factor transcripcional de la familia ATF1 y está involucrado en respuestas de estrés oxidativo y osmótico a través de la represión por Hog1, esta proteína participa en la vía alta osmolaridad por glicerol. Además pertenece a la familia ATF/CREB, donde también se encuentra el factor ATF6, importante en UPR. Estos resultados reflejan la presencia de estrés oxidativo y osmótico en CP.

*RIB1* se sobre-expresó en ambas condiciones sin diferencias significativa, indicando la demanda de la síntesis de riboflavina por el cambio de la fuente de carbono.

**Tabla 28.** Nombre y función de las proteínas codificadas por los genes involucrados en UPR.

Gen	ID del gen	log <sub>2</sub> (Fold_change)			Proteína	Función
		CM/C0	CP/C0	CM/CP		
<i>KAR2</i>	PAS_chr2-1_0140	2.2		1.4	Proteína BIP, ATPasa de la familia HSP70	Principal proteína residente del RE P, actúa como chaperona en la translocación de proteína en lumen del RE, regula el plegamiento de proteínas; además regula UPR vía IRE1p
<i>RLG1</i>	PAS_chr3_0622		-1.8	2.1	RNA ligasa codificada por RLG1	Requerida para splicing del RNA de HAC1, involucrado con la regulación de UPR
<i>HAC1</i>	PAS_chr1-1_0381		1.6	-1.2	Factor transcripcional Hac1p	Factor de transcripción de la familia ATF/CREB, regula la respuesta de proteínas mal plegadas (UPR). El estrés en ER induce la vía de splicing de HAC1 por IRE1p que facilita la síntesis eficiente de Hac1p
<i>INO1</i>	PAS_chr2-2_0113	2.6		6.9	Inositol 3-fosfato sintasa	Síntesis de inositol fosfato e inositol que contiene fosfolípidos
<i>CPR5</i>	PAS_chr1-1_0267	1.4		1.1	Peptidil-prolil cis-trans isomerasa B (ciclofilina)	Cataliza la isomerización cis-trans de los enlaces peptídicos de los residuos de prolina N-terminal, es inducido transcripcionalmente durante UPR en el RE.
<i>PDI</i>	PAS_chr4_0844	1.0			Proteína disulfuro isomerasa	Proteína multifuncional residente del lumen del RE, responsable para la formación, rearrreglo e inducción de enlaces de disulfuros de las proteínas secretadas sintetizadas de novo.
<i>SKO1</i>	PAS_chr1-1_0165		1.0	-1.8	Represor transcripcional de la familia de genes ATF/CREB	Factor de transcripción de la familia ATF/CREB, está involucrado en respuestas de estrés oxidativo y osmótico a través de la represión por Hog1, éste participa en la vía alta osmolaridad por glicerol.
<i>RIB1</i>	PAS_chr2-1_0736	1.7	1.3		GTP ciclohidrolasa II	Importante en el primer paso de la biosíntesis de riboflavina

El grupo de genes diferencialmente expresados involucrados en el complejo proteico del proteosoma mediado por la degradación de proteínas mediada por el retículo endoplásmico (ERAD) se muestran en la figura 74 y se representan en la tabla 29.



**Figura 74.** Genes diferencialmente expresados involucrados en el complejo proteico del proteosoma mediado por ERAD, considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$ .

En las condiciones CM y CP se sobre-expresaron 6 y 7 genes, respectivamente, frente al cambio de la fuente de carbono. Los genes *RPT3*, *PRE10* y *BLM10* involucrados en la conformación estructural y funcional del proteosoma 26S se sobre-expresaron en CM, por el cambio de la fuente de carbono, sin embargo no se mostró expresión diferencial significativa al comparar con CP (Figura 74). Los genes *SSH4* y *RAD6*, involucrados en el proceso de ubiquitinación en el proteosoma se sobre-expresaron en ambas condiciones (Figura 74). *RSP5*, *E3* y *CDC4* se sobre-expresaron en CP por el cambio de la fuente de carbono, mientras que *RSP5* y *E3* se sub-expresaron en CM vs CP, genes involucrados en la ubiquitinación de la proteína que será procesada en el proteosoma.

**Tabla 29.** Nombre y función de las proteínas codificadas por los genes involucrados en el complejo proteico del proteosoma mediado por ERAD.

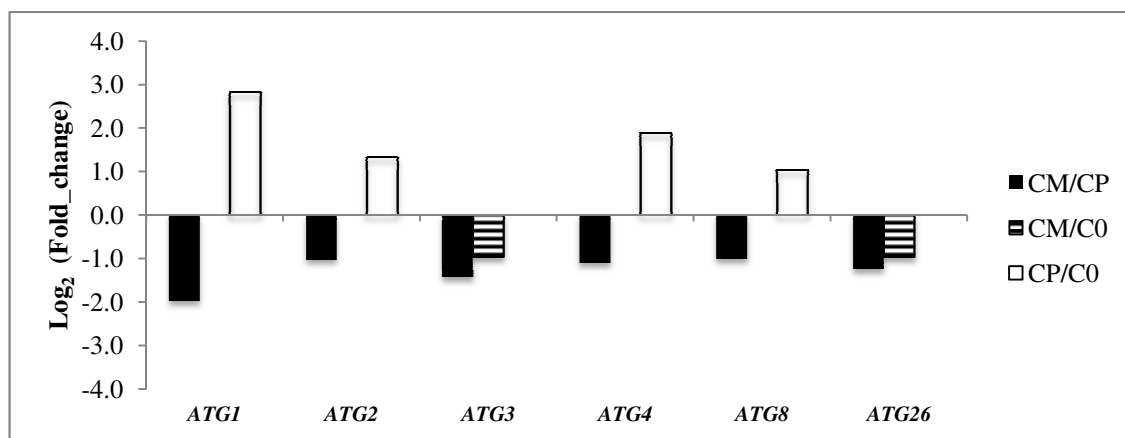
Gen	ID del gen	Log <sub>2</sub> (Fold_change)			Proteína	Función
		CM/C0	CP/C0	CM/CP		
<i>RPT3</i>	PAS_chr3_0881	1.0			ATPasa de la partícula reguladora 19S del proteosoma 26S	Una de las 6 ATPasas de la partícula reguladora 19S del proteosoma 26S involucrado en la degradación de sustratos ubiquitinados
<i>PRE10</i>	PAS_chr3_0699	1.0			Subunidad 7 alfa de la partícula catalítica 20 S del proteosoma 26S	Forma parte del núcleo de la partícula 20S del proteosoma 26S , involucrado en la degradación se sustratos ubiquitinados
<i>BLM10</i>	PAS_chr3_0776	1.2			Subunidad activadora de la partícula catalítica 20S del proteosoma 26S	Activador de la formación del núcleo del la partícula catalítica 20S del proteosoma 26S, involucrado en la degradación se sustratos ubiquitinados.
<i>SSH4</i>	PAS_chr3_0508	1.3	1.9		Factor de ubiquitinación dependiente de Rsp5p	Requerida para la ubiquitinación dependiente de RSP5p y clasifica las proteínas a descargar hacia los cuerpos multivesiculares.
<i>RAD6</i>	PAS_chr3_0359	1.2	1.2		Complejo E2 conjugada a ubiquitina	Transportadora de ubiquitina, involucrada en la degradación de proteínas mediada por ubiquitina por la vía del proteosoma
<i>RSP5</i>	PAS_chr2-1_0354		1.2	-1.6	Ubiquitina ligasa E3	Involucrada en la degradación de proteínas mediada por ubiquitinación por la vía del proteosoma.
<i>CDC4</i>	PAS_chr2-1_0666		1.2		Complejo de ubiquitin ligasa	Forma parte de un complejo ubiquitina ligasa dependiente del proteosoma, implicado en el catabolismo proteico.
<i>E3</i>	PAS_chr4_0444		1.2	-1.4	Ubiquitina ligasas E3 sinoviolina	Involucrada en la degradación de proteínas mediada por ubiquitinación por la vía del proteosoma.
<i>HUL5</i>	PAS_chr2-1_0164	1.0		1.1	Ubiquitina ligasa E4	Elonga cadenas poliubiquitinadas de los sustratos proteicos, requerida para el transporte retrógrada hacia el citoplasma durante ERAD.
<i>HSP78</i>	PAS_chr2-1_0324		1.3	-1.3	Chaperona oligomérica de la matriz mitocondrial	Componente de la maquinaria de del control de calidad de proteínas remueve proteínas dañadas y mal plegadas. Bajo condiciones de estrés térmico y resolubiliza agregados de proteínas recuperando propiedades. Termo tolerancia durante un choque térmico.
<i>HSP42</i>	PAS_c034_0039		1.5		Proteína de choque térmico 42	Forma agregados con las proteínas mal plegadas para su clasificación durante un estrés fisiológico.

Interesantemente *HUL5* se sobre-expresó en CM vs CP, gen implicado en la elongación de las cadenas poliubiquitinadas de los sustratos proteicos y requerida para el transporte retrógrada hacia el citoplasma durante ERAD.

*HSP78* y *HSP42* codifican para proteínas de choque térmico, *HSP78* se sub-expresó en CM vs CP, componente de la maquinaria del control de calidad de proteínas, importante en impedir los agregados de las proteínas mal plegadas; y *HSP42* se sobre-expresó en CP en respuesta al metanol. Ambos genes están involucrados durante procesos de estrés fisiológico, indicando con esto un mayor estado de estrés en CP que en CM.

Los resultados indican que el recambio proteico mediado por el proteosoma se encuentra activo en ambas condiciones, sin embargo, se presentó expresión diferencial entre los genes involucrados, en dónde la actividad de la vía de ubiquitinación (*RSP5* y *E3*) y la de control de calidad y estrés (*HSP78*) se sub-expresaron en CM vs CP, indicando una menor actividad de degradación proteosomal asociada a ERAD en CM a diferencia de CP.

El grupo de genes diferencialmente expresados involucrados en autofagia se muestran en la figura 75 y se representan en la tabla 30.



**Figura 75.** Genes diferencialmente expresados involucrados en autofagia, considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$ .

En las condiciones de CP se sobre-expresaron los genes *ATG1*, *ATG2*, *ATG4* y *ATG8* debido al cambio de la fuente de carbono (Figura 75), todos ellos implicados en la

biogénesis de vesículas en la autofagia y vía del transporte direccional de citosol a la vacuola (vía Cvt) relacionada con el sitio de ensamblaje del fagóforo (PAS). Los genes *ATG3* y *ATG26* se sub-expresaron en CM debido al cambio de la fuente de carbono, *ATG3* incorpora a *ATG8* a la fosfatidiletanolamina (PE) en la biogénesis de vesículas, mientras que *ATG26* participa en la peroxifagia.

Algunos autores (Asakura *et al.* 2009) han identificado al gen *ATG26* como un factor que activa pexofagia (degradación de peroxisomas) en la levadura *P. pastoris*, *ATG26* es reclutada a un complejo de nucleación lípido-proteína citosólica a través de su PBD (dominio de unión a fosfoinosítido) y activa el alargamiento y la maduración de una estructura tipo autofagosoma. De acuerdo a estos resultados, este gen se encuentra significativamente sub-expresado en CM, demostrando una estabilidad en los peroxisomas y supresión en la pexofagia que beneficia la condición de cultivo empleada en CM, a diferencia de CP.

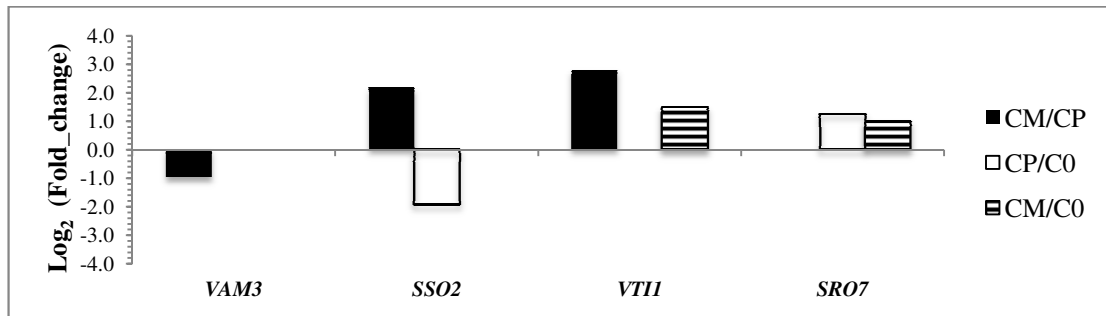
Todos los genes relacionados con la atofagia se sub-expresaron diferencialmente en CM *vs* CP indicando con esto, una menor actividad autofágica en CM respecto a CP. En levaduras, la autofagia es inducida bajo condiciones de estrés celular, como la falta de nutrientes, aumento de temperatura o estrés oxidativo; y está implicada en fenómenos de muerte celular no apoptóticos.

Por ésta razón los resultados indican que, en especial el factor temperatura, es determinante durante la respuesta de adaptación al estrés ambiental, así entonces la sub-expresión de los genes involucrados en autofagia en CM, indicadores de estrés y proteólisis, fue un factor que contribuyó a la sobre-producción de la proteína heteróloga en CM *vs* CP.

**Tabla 30.** Nombre y función de las proteínas codificadas por los genes involucrados en autofagia.

Gen	ID del gen	log <sub>2</sub> (Fold_change)			Proteína	Función
		CM/C0	CP/C0	CM/CP		
<i>ATG1</i>	PAS_chr2-1_0641		2.8	-2.0	Serina/treonina proteína kinasa	Requerida para la formación vesicular durante la autofagia y la vía del transporte direccional de citosol a la vacuola (Cvt pathway) relacionada con el sitio de ensamblaje del fagóforo (PAS). Reticulofagia por estrés en RE.
<i>ATG2</i>	PAS_chr1-3_0168		1.3	-1.0	Proteína relacionada con autofagia 2	Requerida durante la formación vesicular durante la autofagia, la vía del transporte direccional de citosol a la vacuola (Cvt pathway) y peroxifagia.
<i>ATG3</i>	PAS_chr3_0688			-1.4	Un análogo de una enzima de conjugación de ubiquitina	Fijación de la fosfatidiletanolamina (PE) a Atg8 que conduce a la interacción con PAS, formación de Cvt vesículas y autofagosoma
<i>ATG4</i>	PAS_chr1-4_0522		1.9	-1.1	Cisteína proteasa	Requerida en autofagia, hidrólisis de ATG8p para formar autofagosoma y generación de vesículas Cvt
<i>ATG8</i>	PAS_chr4_0704		1.0	-1.0		Interacción con PAS, formación de Cvt vesículas y autofagosoma, participa en autofagia no específica. Reticulofagia por estrés en RE.
<i>ATG26</i>	PAS_chr4_0167	-1.0		-1.2	UDP-glucosa esteroil glucosiltransferasa	Involucrada en formación de membranas de lípidos, relacionada con autofagia específica de peroxisomas.

En el análisis de los genes asociados a SNARE, se encontró que *VAM3* se sub-expresó en CM vs CP, esta proteína de tipo t-SNARE está asociada con la fusión de membrana de vesículas durante la autofagia (Tabla 31), corroborando que este proceso se encuentra reprimido en CM (Figura 76).



**Figura 76.** Genes diferencialmente expresados involucrados en los receptores de ataque a proteínas solubles NSF (SNARE), considerando un  $\log_2$  (Fold\_change)  $\leq -1.0$  ó  $\geq 1.0$ .

Los resultados muestran la sobre-expresión de *SSO2* y *VTII* en CM vs CP; además *SSO2* se sub-expresó en CP en respuesta al cambio en la fuente de carbono. El gen *SSO2* está involucrado en la formación de vesículas secretoras y transporte de proteínas de forma intracelular, mientras que *VTII* está involucrado en el tráfico de membrana cis-Golgi hacia la membrana plasmática. Reportes previos (Ruohonen *et al.*, 1997) han demostrado que la sobre-expresión de *SSO2*, en *Saccharomyces cerevisiae*, ha incrementado de 4 a 6 veces la secreción de dos proteínas heterólogas (alfa-amilasa e invertasa) en este hospedero. La sobre-expresión de *SSO2* y *VTII* parecen estar favoreciendo el transporte intracelular y la secreción de las proteínas y la proteína heteróloga en CM, además *SRO7* podría estar contribuyendo con este proceso, así la sobre-expresión de estos genes en CM podría ser otro factor que explique una mayor producción bajo estas condiciones.



**Tabla 31.** Nombre y función de las proteínas codificadas por los genes involucrados en SNARE.

<b>Gen</b>	<b>ID del gen</b>	<b>Log<sub>2</sub> (Fold_change)</b>			<b>Proteína</b>	<b>Función</b>
		<b>CM/C0</b>	<b>CP/C0</b>	<b>CM/CP</b>		
<i>VAM3</i>	PAS_chr2-2_0350			-1.0	Homólogo a syntaxina (t-SNARE)	Proteína t-SNARE requerida en la fusión de membrana de vacuolas homotípicas, en la autofagia y en la fusión de las vesículas transportadores tardíos a la vacuola.
<i>SSO2</i>	PAS_chr1-4_0294		-1.9	2.2	Homólogo a syntaxina (t-SNARE)	Proteína t-SNARE involucrada en la fusión de vesículas secretoras en la membrana plasmática
<i>VTII</i>	PAS_chr3_1085	1.5		2.8	Proteína v-SNARE pro	Involucrada en el tráfico de membrana cis-Golgi, v-SNARE que interactúa con t-SNARE`s involucrados múltiples vías de clasificación vacuolar. Transporte vesicular entre aparato de Golgi y membrana plasmática
<i>SRO7</i>	PAS_chr1-4_0066	1.0	1.2		Proteína SNI2	Involucrada en exocitosis, acoplamiento y fusión de vesículas post-Golgi con la membrana plasmática

### 8.3 Validación del análisis transcripcional mediante RT-qPCR de los genes seleccionados

#### 8.3.1 Selección y Diseño de oligonucleótidos específicos para los genes involucrados en la vía de secreción y metabolismo del metanol.

Se seleccionaron 10 genes relacionados con la vía de secreción, los cuáles mostraron una sobre-expresión (*SSA2*, *HTM1*, *MPG1*, *PMT2* y *SSO2*) y sub-expresión (*WSC4*, *NCE103*, *WSC4* y *NCE103*, *RSP5*, *E3* y *AHP1*) relevante en el análisis transcripcional en la comparación CM vs CP, estos genes se indican en la tabla 32.

**Tabla 32.** Genes utilizados en la validación transcripcional involucrados en la vía de secreción seleccionados a partir de los resultados obtenidos en la comparación CM/CP.

Genes	Nombre	Proceso involucrado
<b>Sobre-expresados</b>	<i>MPG1</i>	Síntesis de GTP Manosa
	<i>PMT2</i>	O-Glicosilación
	<i>HTM1</i>	Sensor del Plegamiento
	<i>SSO2</i>	tSNARE
	<i>SSA2</i>	Chaperona implicada en el plegamiento de proteínas
<b>Sub-expresados</b>	<i>AHP1</i>	Enlace Sulfuro
	<i>RSP5</i>	Degradación de proteínas mediada por ubiquitinación
	<i>E3</i>	Degradación de proteínas mediada por ubiquitinación
	<i>WSC4</i>	Translocación proteínas secretorias solubles
	<i>NCE103</i>	Exportación de proteínas no clásica

Se diseñaron un total de 10 pares de oligonucleótidos específicos en base a las secuencias de transcritos reportadas en GenBank. Todos los oligonucleótidos diseñados hibridaron en la región CDS de la secuencia de cada gen y mostraron productos amplificados comprendido entre 103 a 150 pb (Tabla 41, Anexo 1). Los resultados muestran de manera detallada los oligonucleótidos diseñados que amplifican los genes involucrados en la vía de secreción sobre- y sub-expresados en el análisis transcripcional en las figuras 79 y 80 (Anexo1); de manera general, los oligonucleótidos 5' y 3' diseñados presentaron un tamaño comprendido entre 20 y 26 pb y un rango entre 40 y 57 %GC. En la evaluación de la especificidad, todos los alineamientos de los oligonucleótidos diseñados, mostraron una especificidad única del 100% con cada secuencia blanco codificante de los genes de *P. pastoris* sobre-expresados y sub-expresados (Figuras 82 y 83, respectivamente, Anexo 3) involucrados en la vía de secreción; además, en la evaluación de la formación de estructuras secundarias de tipo

horquilla, homodímeros y heterodímeros, no se encontraron valores de  $\Delta G$  menores a -10 kcal/mol en ninguno de los oligonucleótidos (Tabla 43, Anexo 3).

En el caso de la vía del metabolismo del metanol, se seleccionaron 3 genes en base a su sobre-expresión: *FBA2*, *FGH* y *CAT* (Tabla 33).

**Tabla 33.** Genes utilizados para la validación transcripcional involucrados en la vía del metabolismo del metanol seleccionados a partir de los resultados obtenidos en la comparación CM/CP.

Gen	Nombre	Parte de la Ruta de metanol
<i>FBA2</i>	Fructosa bisfosfato aldolasa	Vía pentosa fosfato--- Biomasa
<i>FGH</i>	S-Formilglutación Hidrolasa	Energía Libre
<i>CAT</i>	Catalasa	Hidrólisis del peróxido de hidrógeno

Se diseñaron un total de 3 pares de oligonucleótidos específicos en base a las secuencias de transcritos reportadas en GenBank. Todos los oligonucleótidos diseñados hibridaron en la región CDS de la secuencia de cada gen y mostraron productos amplificados de 134, 139 y 144 pb correspondientes a cada gen (Tabla 42, Anexo1). Los resultados muestran de manera detallada los oligonucleótidos diseñados que amplifican los genes involucrados en la vía del metabolismo del metanol sobre-expresados en el análisis transcripcional (Figura 81, Anexo 2); de manera general, los oligonucleótidos 5' y 3' diseñados presentaron un tamaño comprendido entre 21 y 23 pb y un rango entre 47 y 55 %GC. En la evaluación de la especificidad, todos los alineamientos de los oligonucleótidos diseñados mostraron una especificidad única del 100% con cada secuencia blanco codificante de los genes de *P. pastoris* sobre-expresados (Figura 84, Anexo 3) involucrados en la vía del metabolismo del metanol. En la evaluación de la formación de estructuras secundarias de tipo horquilla, homodímeros y heterodímeros, no se encontraron valores de  $\Delta G$  menores a -10 kcal/mol en ninguno de los oligonucleótidos (Tabla 44, Anexo 3).

En conclusión todos los genes seleccionados para la validación del análisis transcripcional mostraron especificidad hacia su secuencia blanco y se predijo una estabilidad aceptable al presentar valores de  $\Delta G$  menores a -10 kcal/mol.

### 8.3.3 Curva Estándar para la obtención de los intervalos de linealidad y eficiencia de cada uno de los genes seleccionados para los ensayos de qPCR.

De acuerdo a las curvas de disociación de cada uno de los genes evaluados, se obtuvo un único pico de emisión de fluorescencia a una sola temperatura de desnaturalización en todos los casos (Figura 85, Anexo 4), garantizando la ausencia de la amplificación de productos inespecíficos y de dímeros de oligonucleótidos, y por lo tanto la especificidad de los mismos. La temperatura de hibridación utilizada para la amplificación de todos los genes fue de 60°C, a excepción del gen *SSA2*, el cuál utilizó una temperatura de 65°C

Las curvas de amplificación correspondientes a cada gen evaluado mostraron un comportamiento similar (Figura 86, Anexo 4), lo que generó valores de  $C_T$  que delimitan el intervalo de linealidad, además de una correlación de  $C_T$  respecto al logaritmo de la cantidad equivalente de RNA en ng, con  $R^2$  mayores de 0.986 (Figura 87, Anexo 4), y eficiencias de amplificación comprendidas entre 90% a 107.9% (Tabla 34).

**Tabla 34.** Eficiencia, linealidad, intervalos de linealidad y umbral de fluorescencia de los amplicones evaluados.

Gen	Eficiencia (%)	Linealidad ( $r^2$ )	Intervalo de linealidad ( $c_t$ )		Umbral (dr)
<i>MPG1</i>	96.40%	0.996	22.47	32.17	2000
<i>PMT2</i>	90.50%	0.999	25.72	35.8	2500
<i>HTM1</i>	92.30%	0.986	27.74	37.33	1500
<i>SSO2</i>	91.90%	0.996	23.11	33.26	2000
<i>SSA2**</i>	92.90%	0.994	20.35	29.87	2500
<i>AHP1</i>	90.00%	0.998	24.74	34.63	3500
<i>RSP5*</i>	107.90%	0.997	24.77	33.41	2000
<i>E3*</i>	90.80%	0.995	23.77	33.67	2000
<i>WSC4</i>	90.50%	0.993	26.33	36.01	2500
<i>NCE103*</i>	91.40%	0.993	22.06	31.98	2500
<i>FBA2</i>	100.90%	0.997	19.99	29.14	2000
<i>FGH</i>	95.40%	0.999	20.82	30.36	2000
<i>CAT</i>	99.30%	1	20.06	29.33	2500

\*Genes con con DNAc directo antes de las diluciones (1:5 a 1:625)

\*\*Gen con temperatura 65°C en el paso 2 del segmento 2 y con DNAc directo antes de las diluciones (1:5 a 1:625)

### 8.3.4 Ensayos de qPCR mediante SYBR Green para cada uno de los genes evaluados

Los valores de  $C_T$  obtenidos para cada gen se encontraron dentro del intervalo de linealidad y mostraron valores de  $CV \leq m\%$ , indicando una buena reproducibilidad de los duplicados analíticos en las tres condiciones de cultivo evaluadas C0 (rango comprendido entre 0.04 a 0.74), CM47-1 (rango comprendido entre 0.02-0.85), CM47-2 (rango comprendido entre 0.02-1.56), CP47-1 (rango comprendido entre 0.02-1.32); en las tablas 35-38 se muestran la media de los valores  $C_T$ , desviación estándar y coeficientes de variación de cada uno de los genes evaluados en los cultivos C0, CM47-1, CM47-2 y CP47-1. La curva de amplificación de cada uno de los genes evaluados se muestra en las figuras 88-91 (Anexo 4).

**Tabla 35.** Valores de ciclo umbral ( $C_T$ ), media y CV (%) de cada uno de los genes evaluados en el cultivo C0 antes de la inducción con metanol.

Gen	Cultivo C0			
	CT1	CT2	Media	CV(%)
<i>YPT1</i>	28.3	28.5	28.4	0.42
<i>G6PD</i>	26.7	26.6	26.6	0.06
<i>KAR2</i>	27.9	28.3	28.1	0.69
<i>MPG1</i>	26.5	26.3	26.4	0.36
<i>PMT2</i>	30.2	30.5	30.3	0.59
<i>HTM1</i>	32.0	32.5	32.3	0.74
<i>SSO2</i>	26.3	26.4	26.3	0.15
<i>SSA2</i>	29.7	29.5	29.6	0.24
<i>AHP1</i>	27.8	27.6	27.7	0.45
<i>RSP5</i>	30.5	30.1	30.3	0.59
<i>E3</i>	29.2	29.4	29.3	0.34
<i>WSC4</i>	29.2	29.0	29.1	0.29
<i>NCE103</i>	26.8	26.9	26.9	0.15
<i>FBA2</i>	27.8	28.1	27.9	0.50
<i>FGH</i>	27.6	27.6	27.6	0.04
<i>CAT</i>	25.7	25.4	25.5	0.63

**Tabla 36.** Valores de ciclo umbral ( $C_T$ ), media y CV(%) de cada uno de los genes evaluados en el cultivo CM47-1 a las 47 horas de inducción con metanol.

Gen	Cultivo CM47-1			
	CT1	CT2	Media	CV(%)
<i>YPT1</i>	28.5	28.3	28.4	0.44
<i>G6PD</i>	26.4	26.8	26.6	0.75
<i>KAR2</i>	24.4	24.7	24.6	0.61
<i>MPG1</i>	24.9	24.8	24.8	0.38
<i>PMT2</i>	28.9	28.9	28.9	0.07
<i>HTM1</i>	31.2	31.4	31.3	0.32
<i>SSO2</i>	25.8	25.6	25.7	0.23
<i>SSA2</i>	27.2	27.1	27.1	0.22
<i>AHP1</i>	28.2	28.5	28.3	0.49
<i>RSP5</i>	29.3	29.0	29.1	0.53
<i>E3</i>	28.5	29.0	28.8	0.85
<i>WSC4</i>	30.1	29.8	30.0	0.45
<i>NCE103</i>	27.8	28.0	27.9	0.47
<i>FBA2</i>	23.4	23.4	23.4	0.02
<i>FGH</i>	24.3	24.0	24.1	0.66
<i>CAT</i>	23.3	23.4	23.4	0.09

**Tabla 37.** Valores de ciclo umbral ( $C_T$ ), media y CV(%) de cada uno de los genes evaluados en el cultivo CM47-2 a las 47 horas de inducción con metanol.

Gen	Cultivo CM47-2			
	CT1	CT2	Media	CV(%)
<i>YPT1</i>	28.8	29.3	29.1	0.83
<i>G6PD</i>	27.2	27.3	27.2	0.06
<i>KAR2</i>	25.3	25.7	25.5	0.81
<i>MPG1</i>	26.0	25.5	25.7	1.01
<i>PMT2</i>	29.9	29.2	29.6	1.18
<i>HTM1</i>	32.7	31.8	32.2	1.43
<i>SSO2</i>	26.3	26.3	26.3	0.04
<i>SSA2</i>	28.2	28.1	28.2	0.28
<i>AHP1</i>	28.2	27.9	28.1	0.64
<i>RSP5</i>	29.7	29.8	29.7	0.02
<i>E3</i>	29.0	29.0	29.0	0.05
<i>WSC4</i>	31.5	31.5	31.5	0.03
<i>NCE103</i>	29.0	28.8	28.9	0.38
<i>FBA2</i>	23.9	23.6	23.7	0.67
<i>FGH</i>	24.4	24.6	24.5	0.39
<i>CAT</i>	24.5	23.7	24.1	1.56

**Tabla 38.** Valores de ciclo umbral ( $C_T$ ), media y CV(%) de cada uno de los genes evaluados en el cultivo CP47-1 a las 47 horas de inducción con metanol.

Gen	Cultivo CP47-1		Media	CV(%)
	CT1	CT2		
<i>YPT1</i>	28.29	28.08	28.19	0.37
<i>G6PD</i>	27.54	27.48	27.51	0.11
<i>KAR2</i>	27.06	27.47	27.27	0.75
<i>MPG1</i>	28.68	28.56	28.62	0.21
<i>PMT2</i>	29.25	29.55	29.40	0.51
<i>HTM1</i>	28.68	28.48	28.58	0.35
<i>SSO2</i>	28.43	28.42	28.43	0.02
<i>SSA2</i>	27.67	27.85	27.76	0.32
<i>AHP1</i>	28.04	28.79	28.42	1.32
<i>RSP5</i>	28.43	28.17	28.30	0.46
<i>E3</i>	28.02	28.27	28.15	0.44
<i>WSC4</i>	28.35	28.56	28.46	0.37
<i>NCE103</i>	28.09	28.46	28.28	0.65
<i>FBA2</i>	24.14	24.02	24.08	0.25
<i>FGH</i>	25.39	25.38	25.39	0.02
<i>CAT</i>	23.05	23.09	23.07	0.09

#### 8.3.4.1 Calculo de los niveles de expresión relativa

Los resultados de los niveles de expresión relativa de cada uno de los genes en las condiciones evaluadas C0, CM47-1, CM47-2 y CP47-1 se muestran en la tabla 39, Los genes que presentaron niveles de expresión relativa u 10 en el cultivo CM47h fueron FBA2, FGH y KAR2, mientras que en el cultivo CP47h fueron *FBA2* y *HTM1*. En las muestras correspondientes a CM47 se muestra la media y su desviación estándar comprendidos entre 0.01 y 3.89.

**Tabla 39.** Valores de Expresión relativa de cada uno de los genes en las condiciones de cultivo evaluadas.

<i>Gen</i>	CM47-1	CM47-2	MEDIA	CP47-1
<i>KAR2</i>	10.84	9.18	10.01 ± 1.17	2.20
<i>MPG1</i>	2.84	2.37	2.61 ± 0.33	0.28
<i>PMT2</i>	2.55	2.44	2.50 ± 0.08	2.25
<i>HTMI</i>	1.90	1.56	1.73 ± 0.25	13.61
<i>SSO2</i>	1.48	1.50	1.49 ± 0.01	0.31
<i>SSA2</i>	5.05	3.90	4.48 ± 0.81	4.11
<i>AHP1</i>	0.65	1.20	0.93 ± 0.39	0.77
<i>RSP5</i>	2.30	2.26	2.28 ± 0.03	5.30
<i>E3</i>	1.38	1.81	1.60 ± 0.30	2.59
<i>WSC4</i>	0.56	0.33	0.44 ± 0.16	1.84
<i>NCE103</i>	0.51	0.40	0.45 ± 0.08	0.49
<i>FBA2</i>	23.09	28.59	25.84 ± 3.89	18.12
<i>FGH</i>	10.46	12.38	11.42 ± 1.36	5.56
<i>CAT</i>	4.48	4.11	4.29 ± 0.27	6.74

#### 8.3.4.2 Comparación de las técnicas de RT-qPCR y RNA-seq

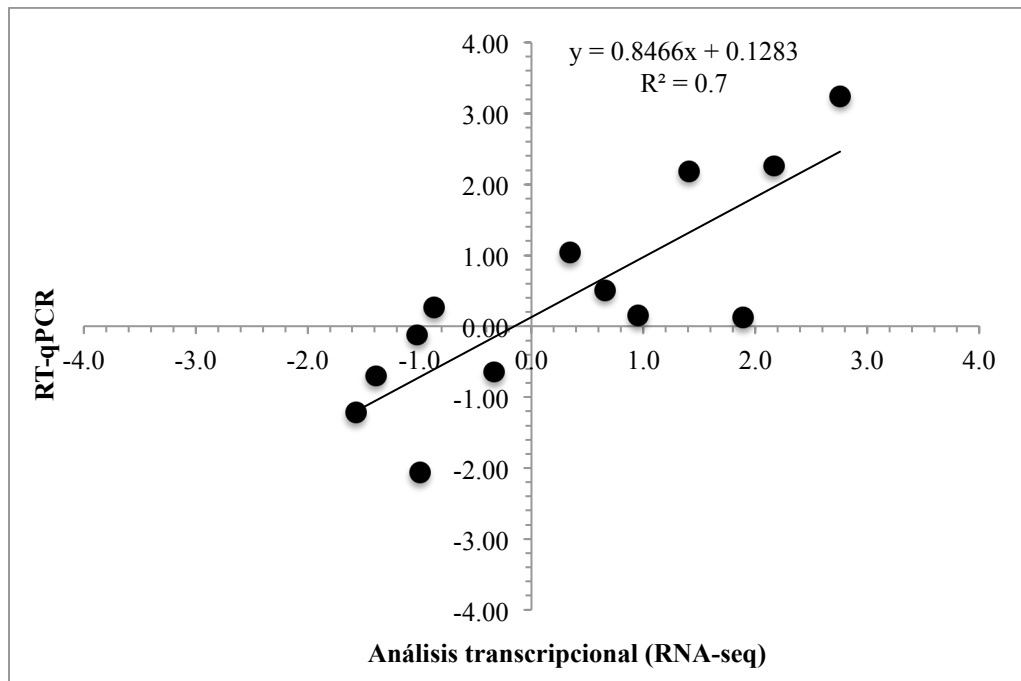
Los resultados del  $\log_2$  calculado a partir de los niveles de expresión relativa se muestra en la tabla 40, además también se incluyen los resultados  $\log_2$  (Fold\_change) obtenidos en el análisis transcripcional.

En general, todos los genes mostraron una correlación cuando se compararon los resultados obtenidos con la técnica de RT-qPCR y RNAseq, a excepción de los genes *HTMI* y *AHP1*, los cuáles no mostraron el mismo comportamiento. El gen *HTMI* presentó mayor variación, posiblemente debido a un evento de splicing alternativo (evidencia mostrada en las anotaciones y en uno de los archivos de salida del análisis transcripcional), por lo que se descartó para el cálculo del  $R^2$ . En la figura 77 se muestra la correlación entre ambas técnicas, su  $R^2$  y la pendiente; sin embargo, para poder visualizar estos resultados se muestran en el gráfico de barras (Figura 78).



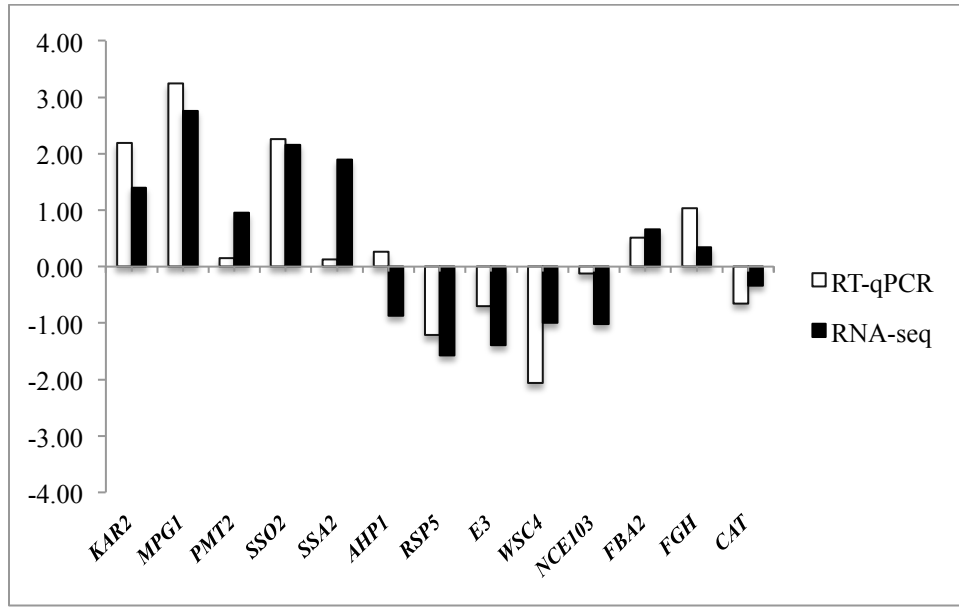
**Tabla 40.** Comparación del  $\log_2$  (Fold\_change) obtenido en el análisis transcripcional y el  $\log_2$  calculado a partir de los niveles de expresión relativa correspondiente a cada uno de los genes.

<i>Gen</i>	Media CM47	CP47-1	Cociente CM/CP	$\log_2$ calculado	$\log_2$ (Fold_change) Análisis Transcripcional
<i>KAR2</i>	10.01	2.20	4.56	2.19	1.4
<i>MPG1</i>	2.61	0.28	9.47	3.24	2.8
<i>PMT2</i>	2.50	2.25	1.11	0.15	1.0
<i>HTM1</i>	1.73	13.61	0.13	-2.98	2.5
<i>SSO2</i>	1.49	0.31	4.79	2.26	2.2
<i>SSA2</i>	4.48	4.11	1.09	0.12	1.9
<i>AHP1</i>	0.93	0.77	1.20	0.26	-0.9
<i>RSP5</i>	2.28	5.30	0.43	-1.22	-1.6
<i>E3</i>	1.60	2.59	0.62	-0.70	-1.4
<i>WSC4</i>	0.44	1.84	0.24	-2.06	-1.0
<i>NCE103</i>	0.45	0.49	0.92	-0.12	-1.0
<i>FBA2</i>	25.84	18.12	1.43	0.51	0.7
<i>FGH</i>	11.42	5.56	2.06	1.04	0.3
<i>CAT</i>	4.29	6.74	0.64	-0.65	-0.3



**Figura 77.** Correlación entre las dos técnicas utilizadas en el análisis transcripcional.

En la parte de arriba se muestra el  $R^2$  y la pendiente. Además se muestra que solo un gen no tiene el mismo comportamiento de expresión en ambas técnicas (Gen AHP1), pero la variación es muy poca.



**Figura 78.** Comportamiento de los genes evaluados en las dos técnicas utilizadas RNA-seq vs RT-qPCR.

## 9. DISCUSIÓN

Los avances en la nueva era de la genómica, transcriptómica y proteómica han sido determinantes en el entendimiento de procesos celulares, y no es excepción en el caso de la levadura metilotrófica *P. pastoris*. Esta levadura es de gran importancia industrial sobre todo para la producción de proteínas heterólogas.

A pesar de que se conocen más de cien genes involucrados en el plegamiento, proteólisis y secreción en esta levadura y que la expresión de estos genes están directamente influenciados por factores ambientales como son temperatura, pH, oxígeno disuelto, etc. (Vanz *et al.*, 2014; Delic *et al.*, 2013; Baumann *et al.*, 2010; De Shutter *et al.*, 2009), existen escasos estudios sobre la regulación de estos genes en *P. pastoris* cultivada bajo condiciones de producción a nivel de bioreactor y considerando parámetros ambientales que permitan obtener altos y bajos niveles de producción de proteínas heterólogas.

Recientemente Viader y colaboradores (2013) optimizaron cinco factores ambientales que definieron condiciones de cultivo en la etapa de inducción con metanol y permitieron incrementar el nivel de producción de una fitasa en *P. pastoris* (cepa KM71FTEII) a nivel de biorreactor. A partir de estos resultados se definieron condiciones de cultivo del bioproceso que condujo a altos niveles de producción (CM) y que produjeron 6.4 g/L de fitasa y condiciones de cultivo que condujeron a bajos niveles de producción (CP) y que produjeron 1.9 g/L de fitasa, ambos a las 68 h de inducción. Las condiciones CM mostraron, mayor demanda de oxígeno y consumo de metanol (14.7 veces más), valores mas altos en la velocidad específica de crecimiento (0.013 vs 0.003), biomasa (1.4 veces más) y rendimiento ( $Y_{p/x}$ ) que en CP, hasta 87% vs 53% de la proteína heteróloga secretada al medio de cultivo, y una disminución en la actividad de proteasas extracelulares. Además, se estudió la expresión de genes relacionados con la secreción, proteólisis, metabolismo del metanol y la expresión del gen heterólogo mediante RT-qPCR, con el fin de evaluar su comportamiento bajo las dos condiciones de cultivo. Los resultados mostraron que ambas condiciones de cultivo no afectaron significativamente la expresión diferencial del gen *AOX2*, responsable del primer paso en el metabolismo del metanol en las cepas Mut<sup>s</sup>, y tampoco afectaron la expresión del

gen heterólogo (*FTEII*); sin embargo, fueron determinantes en CM para la inducción de la expresión de genes de la ruta de secreción y proteólisis, lo cual podría haber influido en los niveles de producción extracelular de la proteína heteróloga en el cultivo CM (Félix-Castro, 2013).

Con el fin de comprender los mecanismos de regulación transcripcional que afectan a la fisiología de *P. pastoris* por la sobre-producción de la proteína recombinante y las condiciones de cultivo, en particular sobre la vía del metabolismo del metanol y secreción, en el presente trabajo se realizó un análisis de expresión diferencial de genes empleando la tecnología de secuenciación masiva de mRNA (RNA-seq) de la cepa KM71FTEII Mut<sup>s</sup> de *P. pastoris* crecida bajo condiciones que conducen a altos (CM) y bajos (CP) niveles de producción extracelular de la proteína recombinante. Adicionalmente, se analizó la expresión diferencial de los genes *MPG1*, *PMT2*, *SSO2*, *SSA2*, *AHP1*, *RSP5*, *E3*, *WSC4*, *NCE103*, *FBA2*, *FGH* y *CAT* mediante RT-qPCR con el fin de validar los resultados obtenidos mediante el análisis transcripcional.

### **Metodología experimental**

este trabajo se emplearon células de cultivos en biorreactor que estaban preservadas en RNA later a -70°C por aproximadamente un año, dificultando obtener un RNA íntegro y con las concentraciones adecuadas en algunas muestras. Además, el pH bajo de las células del cultivo CP afectó las condiciones estándar del protocolo empleado para la obtención del RNA, por lo cual se realizó manteniendo las muestras en hielo durante todo el proceso para evitar degradaciones y con esto aumentar el rendimiento de la preparación. Otro factor importante a considerar durante el proceso de preparación de las muestras de RNA, fue la calidad de la enzima utilizada para romper la pared celular de la levadura (lítica), la cual debe de ser evaluada en su actividad, antes de utilizarla en la preparación de muestras valiosas como fue el caso de este trabajo. En el análisis de expresión de genes, su interpretación y validez es absolutamente dependiente de la calidad de la muestra, por tal razón el criterio de la selección de las muestras sometidas a secuenciación fue la calidad del RNA, en este trabajo se empleó el criterio del valor de RIN, el más aceptado actualmente en el ámbito de secuenciación de RNA. Los valores

de RIN obtenidos fueron mayores a 7.5, asegurando de esta manera la integridad del RNA que se sometió a secuenciación.

El uso de RNA-seq para el estudio de la expresión diferencial de genes se ha empleado para el estudio de la fisiología de *P. pastoris* (Hesketh *et al.*, 2013; Liang *et al.*, 2012; Qi *et al.*, 2011) empleando diferentes herramientas bioinformáticas para su análisis e interpretación.

La existencia de tres genomas secuenciados de *P. pastoris*, permitió la selección del genoma de la cepa GS115 de *P. pastoris* para realizar un ensamblaje en base al genoma de referencia evitando un ensamblaje *de novo*. Otra ventaja en este trabajo fue la disponibilidad de herramientas bioinformáticas empleadas en el análisis de la expresión diferencial de genes (DEG) con la técnica de RNA-seq. Muchas de estas herramientas bioinformáticas son de libre acceso, con abundante literatura reciente al respecto, además de la posibilidad de evitar el uso de computadoras de alto rendimiento, tal como lo ofrece el empleo de la plataforma Galaxy, que permite enviar los archivos de trabajo e instrucciones para que los análisis se realicen en computadoras de alto rendimiento localizadas en un centro de genómica comparativa y bioinformática de la universidad “Penn State” en Pensilvania, EU .

Para la técnica de RT-qPCR se empleó el protocolo previamente establecido por Félix-Castro en el 2013, adecuando condiciones para el caso de algunos de los genes evaluados. Tal fue el caso de las condiciones de amplificación del gen *SSA2* y las diluciones empleadas para la determinación de las curvas estándar requeridas en el cálculo de la eficiencia de la amplificación de los genes *SSA2*, *RSP5*, *E3* y *NCE103*, debido a la concentración del DNAC en la muestra utilizada.

En el presente trabajo se utilizaron los paquetes bioinformáticos Tophat y Cufflinks (Trapnell *et al.*, 2012) para los alineamientos, el ensamblaje, la cuantificación de transcritos y la expresión diferencial de genes. Adicionalmente se utilizó el plug-in “ClueGo” de Cytoscape para el enriquecimiento de genes. Estos programas han sido citados en numerosos estudios transcripcionales de levaduras y otros organismos eucariotas (Feldmesser *et al.*, 2014; Hesketh *et al.*, 2013b; Merrick *et al.*, 2013; Nookaew *et al.*, 2012; Papini *et al.*, 2012).

El mapeo de las lecturas obtenidas mediante RNA-seq varía dependiendo de la herramienta utilizada. Por ejemplo, Liang y colaboradores (2012) obtuvieron una cobertura del 94.21% del genoma de *P. pastoris* utilizando SOAP2, además el 93.5% de los 5,315 genes anotados fueron identificados con el programa estadístico DESeq. Merrick y colaboradores (2013), obtuvieron una cobertura del 65 y 64.3% del genoma de *Rattus norvegicus* utilizando Tophat2; además se identificaron el 85% de exones anotados y el 10% de exones desconocidos con los programas estadísticos Cuffdiff y Deseq. Nookaew y colaboradores (2012), obtuvieron una cobertura en promedio del 96% del genoma de *S. cerevisiae* utilizando tres programas para realizar alineamientos Gsnap, TopHat y Stampy; además de los programas Cuffdiff, baySeq, DESeq y NOISeq para la cuantificación de los transcritos y se observó una correlación del >70% de los genes diferencialmente expresados entre los cuatro programas utilizados, lo cual indica que al menos estos programas se pueden utilizar para obtener resultados confiables.

La cobertura obtenida en el presente trabajo fue de 89-84% y 89-86% para CM47-1 y CM47-2, respectivamente, 88-86% para CP, 90-88% para C0, y 93.1% de los genes identificados en las comparaciones CM/C0, CP/C0 y CM/CP, fue similar a los trabajos previamente reportados (Qi *et al.*, 2011, Liang *et al.*, 2012); sin embargo, no se obtuvo el mapeo y ensamblaje del 100% de los genes, una explicación a esto podría ser que los transcritos se mapearon en base a un genoma de referencia, el cual puede tener errores en sus anotaciones o bien por la presencia de splicings alternativos; además el genoma empleado de referencia para el mapeo de los transcritos corresponde a la cepa GS115 e incluye algunas modificaciones bioinformáticas (adición de las secuencias del gen *FTEII* y el gen *HIS4* funcional) el cuál discrepa parcialmente con las bibliotecas de transcritos secuenciadas de la cepa KM71FTEII (Mut<sup>s</sup>) que tiene eliminado parte de la secuencia del gen *AOX1*. Otra explicación podría ser que algunos genes podrían estar expresándose en niveles muy bajos haciendo difícil su detección.

En el análisis de la DEG se encontraron 620, 772 y 763 genes expresados diferencialmente en las comparaciones CM vs C0, CP vs C0 y CM vs CP respectivamente, resultados que indican la respuesta fisiológica al cambio de la fuente de carbono y a la producción de la proteína heteróloga aunada a las diferencias en la condiciones de cultivo CM y CP. Liang y colaboradores (2012), empleando el programa

DESeq, encontraron hasta 1885 genes diferencialmente expresados en una cepa GS115 de *P. pastoris* (GS115/KNS-Rml ) por el cambio de la fuente de carbono de glicerol a metanol. Las diferencias encontradas en el número de genes se puede deber a múltiples factores, tal como son la cepa y la proteína heteróloga en estudio, las condiciones de cultivo, así como las herramientas bioinformáticas empleadas para realizar el análisis.

Otro factor que influye en el número de genes diferencialmente expresados es el punto de corte que determina la DEG, en el presente trabajo se utilizó el programa estadístico Cuffdiff para la cuantificación de transcritos y se utilizó el criterio de  $\text{Log}_2 \leq -1.0$  ó  $\geq 1.0$ , es decir sub- ó sobre-expresados más del doble, siendo este criterio el más utilizado en éste tipo de análisis.

A pesar de que existen diversos programas para este tipo de análisis, Cuffdiff se ha utilizado en diversos trabajos publicados (Merrick *et al.*, 2013 ; Nookaew *et al.*, 2012; Feldmesser *et al.*, 2014; Feng & Zhao, 2013), dando certidumbre a los análisis realizados con esta herramienta en el presente trabajo.

### **Expresión del gen heterólogo**

La regulación de la expresión del gen heterólogo *FTEII* a nivel de transcripción, no mostró diferencias significativas entre CM vs CP, sin embargo la inducción con metanol (cambio de la fuente glicerol a metanol) en CP condujo a un incremento en la abundancia de los transcritos de *FTEII* mayor a 3.2 veces ( $\log_2$  change), mientras que en CM fue menor (2.6). Estos resultados concuerdan con los obtenidos por Félix-Castro (2013), donde la expresión de *FTEII* determinada mediante RT-qPCR no presentó diferencias significativas entre ambas condiciones y además demuestran que las condiciones de cultivo en CM y CP no afectan la regulación del promotor *AOX1* inducible con metanol. Sin embargo, los niveles de producción de *FTEII* condujeron a valores  $Y_{p/x}$  más altos en CM que CP (Viader-Salvadó *et al.*, 2013), lo cual indica que los altos niveles de producción de la proteína heteróloga fueron el resultado de procesos post-transcripcionales.

## Expresión diferencial global

El análisis de enriquecimiento de genes por GO se ha empleado con frecuencia en estudios de análisis masivo de genes (Dragosits *et al.*, 2010; Qi *et al.*, 2011; Liang *et al.*, 2011; Hesketh *et al.*, 2013; Rebnegger *et al.*, 2014), incluyendo los estudios de expresión diferencial de genes mediante la secuenciación masiva por RNA-seq (Qi *et al.*, 2011; Liang *et al.*, 2012; Hesketh *et al.*, 2013). En estos estudios se recupera un perfil funcional del grupo de genes con el fin de comprender mejor los procesos biológicos subyacentes.

En el análisis de enriquecimiento de genes en la categoría de proceso biológico y con la prueba estadística aplicada (“Bonferroni step down”) se encontró que los genes sobre-expresados en CM y CP por el efecto del cambio de la fuente de carbono, mostraron en comun una sobre-representación de términos relacionados la actividad peroxisomal (importación de proteínas a peroxisomas) y procesos metabólicos de la tiamina. Estos resultados se confirma con los resultados obtenidos por el análisis del diagama de Venn, en donde se identifica que en ambas condiciones, los genes sobre-expresados están sobre-representados genes relacionados con la actividad peroxisomal, procesos metabólicos relacionados con aldehidos, como es el caso del ciclo del glioxilato (variación del ciclo de los ácidos tricarbóxicos (TCA) que emplea compuestos monocarbonados como fuente de carbono y sintetiza carbohidratos) y metabolismo de vitaminas, principalmente la tiamina.

La actividad peroxisomal es una respuesta que correlaciona con la adaptación de la cepa metilotrófica empleada (KM71 pPIC9FTEII) para metabolizar el metanol. Liang y colaboradores (2012) realizaron un análisis transcripcional bajo condiciones de producción en *P. pastoris* crecida en quimiostato y señalan que el cambio de glicerol a metanol da a lugar a la sobre-expresión de casi todos los genes que controlan el desarrollo y la función de los peroxisomas.

El grupo de genes relacionado con el metabolismo de vitaminas, ha sido ya reportado como sobre-representado en genes sobre-expresados por el cambio de la fuente de carbono a metanol (Liang *et al.*, 2012). En este grupo, los genes relacionados con el metabolismo de la tiamina toma un papel central, ya que ésta participa en mecanismos de regulación de genes a través de la activación de TPP-riboswitches



(Liang *et al.*, 2012). A pesar de que Liang y colaboradores, trabajaron con cepas de GS115, otra proteína y bajo condiciones de quimiostato, al parecer la expresión de este grupo de genes está estrechamente relacionado con el empleo de metanol como fuente de carbono.

Los procesos metabólicos relacionados con aldehidos, como es el caso del ciclo del glioxilato o el metabolismo del gliceraldehido 3-fosfato, es probable que esten relacionados con el empleo de compuestos monocarbonados para sintetizar carbohidratos, como es el caso de la síntesis de 1, 5 ribulosa a partir de 3-fosfo glicerato y que por lo tanto estarían relacionados con el metabolismo de metanol, un substrato monocarbonado.

A diferencia de CP, en CM al cambiar de glicerol a metanol se sobre-expresaron genes, que fueron sobre-representados con términos relacionados con la replicación del DNA y glicosilación de proteínas

Hesketh y colaboradores (2013) analizaron la respuesta fisiológica asociada con la producción inducida con metanol de dos variantes de lizosima (una producida intra y otra extra-celular y diferente grado de plegamiento) en *P. pastoris* crecida en quimiostato. Comparaciones a nivel transcripcional y la ontología de genes mostraron que bajo las condiciones de inducción y comparando contra una cepa control para el caso de la lizosima secretada, los genes sobre-expresados, fueron asociados a la respiración, a la replicación de DNA, septin-citoesqueleto, a la o-glicosilación, metabolismo de glioxilato y TCA. Por otro lado los genes sub-expresados fueron asociados a la biogénesis de ribosomas y procesamiento de RNAr

Al comparar las condiciones CM y CP, se sobre-expresaron genes que fueron sobre-representados en los términos de procesos biosintéticos de aminoácidos y purinas regulación de la traducción, mitosis y glicosilación de proteínas. Al excluir los genes sobre-expresados en ambos cultivos como respuesta al metanol, mediante el diagrama de Venn, los genes sobre-expresados mostraron la sobre-representación significativa de términos relacionados con la traducción, procesos biosintéticos de sustancias orgánicas, regulación post-transcripcional de la expresión de genes y regulación del proceso metabólico de proteínas.

Estos resultados indican que en la condición de CM, con niveles más altos de producción, están activos procesos anabólicos que satisfacen la demanda metabólica requerida para la producción y secreción de la proteína heteróloga y la producción de biomasa, este último evento se puede también fundamentar con la sobre-expresión de los genes relacionados con la síntesis de purinas y una mitosis más activa. Los resultados obtenidos, correlacionan con los resultados obtenidos por Viader-Salvadó y colaboradores (2013), donde indicaron que al final del cultivo (68 h) crecido bajo las condiciones de alta producción (CM) se obtuvieron valores de  $\mu$  (0.014 vs -0.007 h<sup>-1</sup>) y densidad celular (119 vs 68 g L<sup>-1</sup>DCW) más altos que el de la condición de baja producción.

Los genes sub-expresados debido al cambio a la fuente de carbono fue más acentuada para CP que en CM, en CM estos fueron asociados a la vía del TCA (metabolismo del fumarato y glutamina) y transporte de aniones orgánicos, mientras que para el caso de CP fueron asociados a la regulación del inicio de la traducción, proceso biosintético del ácido carboxílico, proceso metabólico de ribonucleosidos de trifosfato de purina, procesos glicolíticos, transporte de iones de fosfato y elongación de la traducción. Considerando solo el análisis de Venn se obtuvo una sobre-representación significativa de términos relacionados con la generación de precursores de metabolitos de energía, los procesos catabólicos de carbohidratos, el metabolismo de moléculas pequeñas y de ácidos carboxílicos y biogénesis de ribosomas.

Estos resultados conducen a considerar una disminución de la actividad los procesos catabólicos del metabolismo central debido al cambio de la fuente de carbono, los cuales son compensados con el metabolismo del metanol en la generación de energía y fuente de carbono necesaria para la generación de biomasa. En un estudio comparativo de flujos metabólicos de cepas productoras de dos proteínas recombinantes crecidas en glicerol y mezclas de glicerol y metanol (Unrean, 2013), se reportó la inactivación o reducción del flujo glucolítico, algunas reacciones del TCA y el metabolismo de glicerol (glicerol 3-fosfato deshidrogenasa) debido al cambio de glicerol a metanol o por el incremento de la proporción de metanol en el cultivo, en este mismo trabajo se reportan activas la vía del las pentosas fosfato, piruvato, acetato que impactan en la generación de

biomasa que compensaría las rutas catabólicas inactivas del metabolismo central en presencia de metanol.

Por otro lado Liang y colaboradores (2012) reportan la sub-expresión de genes del metabolismo del glicerol y genes relacionados con la regulación transcripcional, sugieren que *Pichia pastoris* puede ser propensa a reducir la transcripción de genes menos importantes para disponer de energía y de sustratos para la expresión del gen heterólogo.

Respecto a la sub-expresión de genes relacionado con la biogénesis de ribosomas, ha sido reportado que la disminución del crecimiento conduce a la sub-expresión de genes con la síntesis de ribosomas cuando se cambia de glicerol a metanol (Sauer et al., 2004). Estos resultados están acorde con los estudios realizados por Hesketh y colaboradores (2013) y antes mencionados, en donde reportan que los genes sub-expresados por el cambio de glicerol a metanol fueron asociados a la biogénesis de ribosomas y procesamiento de RNAr, lo cual impacta al proceso traduccional.

En la comparación de CM vs CP los genes sub-expresados sobre-representaron términos relacionados con el TCA, incluyendo la biosíntesis de acetil CoA a partir de acetato, proceso metabólico de aminoácidos de la familia de glutamina, transporte de aniones y regulación transcripcional por promotores dependientes de la RNA polimerasa II, sin embargo mediante el análisis de Venn, al excluir los genes que se afectan por el cambio a metanol, los genes sub-expresados se asociaron a procesos implicados con la regulación de la transcripción por promotores dependientes de la RNA polimerasa II, esto podría estar relacionados con una disminución en la carga metabólica reduciendo la transcripción de genes con funciones no implicadas con la síntesis de la proteína recombinante.

Los resultados confirman que el cambio de glicerol a metanol causa cambios metabólicos importantes afectando la expresión diferencial de 620 genes en CM y 772 genes en CP, que en general los genes sobre-expresados están relacionados con la actividad peroxisomal y procesos anabólicos relacionados con el metabolismos central que conducen a la generación de biomasa. Sin embargo las diferencias de las condiciones de cultivo en CM y CP (pH, temperatura, concentración de metanol) condujeron en CM a estimular, la replicación del DNA, procesos biosintéticos

relacionados con proteínas, regulación de la traducción y la glicosilación de proteínas, manifestando una actividad anabólica más alta en CM que en CP.

Por otro lado el cambio de la fuente de carbono condujo a la sub-expresión de genes relacionados con vías del metabolismo central, glucólisis y TCA y con la regulación de la transcripción, lo cual se puede relacionar con la velocidad de crecimiento menor en metanol que en glicerol.

### **Expresión diferencial de genes del metabolismo del metanol**

La especies de levaduras metilotróficas, en general, utilizan una vía de asimilación y desasimilación de metanol (Yurimoto *et al.*, 2011) para la generación de biomasa y energía, respectivamente.

Respecto al metabolismo del metanol, los transcritos de *AOX2* y *CAT* incrementaron en valores mayores a 2.5 y 1.5 veces ( $\log_2$  Fold-change) respectivamente, por el cambio de la fuente de carbono y sin diferencias en CM vs CP. Liang & colaboradores (2012), en la comparación del cambio de carbono de glicerol a metanol en cultivos en quimiostato de la cepa GS115/KNS-Rml, reportaron la sobre-expresión de *AOX2* y *CAT* y de todos los genes involucrados en el metabolismo de metanol.

A pesar que bajo las condiciones de CP la demanda (0.57 vs 0.03 mL/min) y el consumo (1.5 vs 0.5 L) de metanol fue menor que en CM, la regulación de la expresión de ambos genes a las dos condiciones de cultivo no fue diferente y por lo tanto independiente de la concentración de metanol. Estos resultados indican que las condiciones de cultivo no afectaron la regulación de estos genes de forma diferenciada.

Los genes la vía de desasimilación del metanol (*FLD*, *FGH*, y *FDH*) se sobre-expresaron en respuesta al cambio de la fuente de carbono en CP y sin diferencias en la comparación de CM vs CP. *FGH* (S-formilglutathion hidrolasa) fue el gen con el mayor incremento en la abundancia de transcritos en esta vía, con valores de hasta 3.5 veces  $\log_2$  Fold-change, indicando la relevancia de la regeneración del glutathion para mantener un equilibrio en el ambiente oxidativo. Por otro lado *FDH* (formato deshidrogenasa, dependiente de NAD, el NADH recuperado en esta reacción se considera que se emplea para la generación de energía y producción de biomasa con metanol como fuente de carbono) no se sobre-expresa en CM por el cambio de metanol (bajo el requerimiento de

$\log_2(\text{Fold\_change}) \leq -1.0$  ó  $\geq 1.0$ ), indicando que la conversión del ácido fórmico a  $\text{CO}_2$  podría estar más limitada y como consecuencia emplear una ruta alternativa para la asimilación de este substrato monocarbonado.

Todos los genes de la vía de asimilación del metanol se sobre-expresaron por el cambio de la fuente de carbono en CM y sin diferencia en ambas condiciones de cultivo, con excepción de los genes DAS. Los genes , *DAS1* y *DAS2* se sobre-expresaron en CM, DAS convierte  $\text{Xu}_5\text{P}$  y formaldehído a compuestos de 3 carbonos, DHA y GAP, para la generación de biomasa en esta vía (Krainer *et al.*, 2012). La sobre-expresión de estos genes, está relacionados con la generación de energía a través del metabolismo central requerida para una mayor actividad anabólica en CM

El gen *TPI* sobre-expresado en CM, participa en la síntesis de triosas fosfato a partir de metanol y éstas pueden ser utilizada tanto anabólicamente (síntesis de componentes celulares) como catabólicamente (obtención de energía vía ciclo de los TCA) implicando la generación de energía a través del TCA (Solà *et al.*, 2007).

El gen *TKL1*, el cuál fue sub-expresado en ambas condiciones sin diferencias, se relaciona con actividad transcetolasa (*TKL1*), en *S. Cerevisiae* ha demostrado ser una reacción limitante en la parte no oxidativa de la vía de las pentosas fosfato (Krainer *et al.*, 2012), este mismo autor señala que *TKL1* podría estar interviniendo en funciones equivalentes a DAS durante el metabolismo del metanol, sin embargo bajo las condiciones estudiadas este gen fue reprimido , y por lo tanto sin participar en ésta vía.

La relación de los niveles de expresión de los genes *DAS* (suma de los niveles de expresión *DAS1* y *DAS2*) de la vía de asimilación con los niveles de expresión de *FLD*, gen clave en la vía de desasimilación, determinó que la vía de asimilación fue más activa en términos de expresión. Los cocientes *DAS/FLD* en CM/C0, CP/C0 y CM/CP fueron de 5.3, 2.5 y 2.6 respectivamente, demostrando una vía de asimilación más activa durante el metabolismo del metanol en ambas condiciones de cultivo y obteniendo el valor más alto bajo las condiciones de CM. Una vía de asimilación más activa en CM, permite suponer, que las condiciones de cultivo favorecieron a que el substrato monocarbonado, el metanol, se convirtiera en triosas fosfato preferentemente (en lugar de  $\text{CO}_2$  y NADH por la vía de desasimilación), y proseguir por la vía de las pentosas y el TCA y así permitir la síntesis de intermediarios metabólicos, generación de energía y

por lo tanto biomasa. Estos planteamientos explicarían los valores más altos en la velocidad específica de crecimiento en CM, que condujeron a una mayor demanda de metanol y generación nuevamente de energía empleada para cumplir con la carga metabólica requerida por la célula cuando está bajo las condiciones de producción y secreción de la proteína recombinante.

### **Expresión deferencial de genes involucrados en la secreción**

Uno de los cuellos de botella en la producción de proteínas recombinantes es el plegamiento de la forma nativa de la proteína heteróloga. Las altas tasas de producción y la agregación de proteínas recombinantes mal plegadas conducen a condiciones de estrés a las cuales la célula responde para disminuir la toxicidad provocadas por estas proteínas y restaurar la homeostasis celular relacionada con el proceso de plegamiento (Gasser *et al.*, 2008). El plegamiento tanto de proteínas endógenas como recombinantes ocurre en el RE y es mediado por proteínas residentes en las que se incluyen calnexina, chaperonas de la familia hsp70 y hsp90 (como es el caso de BiP codificada por *KAR2*) y PDI (codificada por *PDII*). Tras las N-glicosilaciones, que se inicia en el lumen de RE, las proteínas que están bien plegadas son exportadas a Golgi mediante vesículas de transporte y de Golgi serán transportadas para su secreción. La acumulación de proteínas (agregación de proteínas mal plegadas) en el RE conduce a la síntesis de chaperonas residentes en el RE, como BiP y PDI y se induce el proceso denominado UPR, que a su vez induce al mecanismo mediado por ERAD, para recuperar la conformación nativa de las proteínas a través del transporte retrógrada de las proteínas mal plegadas o bien para provocar su degradación, en este proceso se involucra el transporte a través de vesículas (Gasser *et al.*, 2008; Damasceno *et al.*, 2007), cuya fusión es mediada por SNARE (Gasser *et al.*, 2007).

Se analizaron 110 genes involucrados con el proceso de secreción, agrupados en 5 procesos: procesos de translocación y transporte de proteínas en el RE, chaperonas involucradas en el control de calidad del plegamiento de proteínas en el RE, glicosilación de proteínas en RE, procesamiento y tráfico de proteínas en Golgi y mantenimiento de la homeostasis intracelular y recambio proteico en sistema vacuolares.

La mayor sobre-expresión de estos genes se presentó en CM (26) al cambiar la fuente de carbono y en la comparación de CM vs CP (20). De los genes sobre-expresados en CM predominaron los relacionados con la glicosilación en RE, seguido de las proteínas involucradas en el plegamiento y control de calidad de proteínas y el procesamiento en Golgi. En el caso de la comparación de CM vs CP, predominó procesamiento en Golgi, glicosilación en RE y el control de calidad y plegamiento. En el caso de CP, los genes relacionados con el recambio proteolítico fue sobresaliente.

En el caso de los genes sub-expresados, por el cambio la fuente de carbono en CM y respecto a CP se destaca los genes relacionados con la translocación y el recambio proteolítico y en CP se destaca los genes relacionados con el procesamiento en Golgi por el cambio de la fuente de carbono.

Estos resultados indican claramente que las condiciones de CM condujeron a la regulación de genes que permitieron un mejor plegamiento y procesamiento de las proteínas en el RE y Golgi que en CP. Además, la sobre-expresión de genes relacionados con el recambio proteolítico y la reducción de la actividad transcripcional del procesamiento en Golgi podrían haber participado en las causas de bajo rendimiento provocado en CP.

Analizando procesos de translocación y transporte de proteínas en el RE, los resultados indican que los procesos de translocación co-traduccional (sobre-expresión de *SRP68*) y post-traduccional (sobre-expresión de *SEC63* y *SEC61*) presentan mayor actividad en las condiciones de CM, de la misma manera que la translocación y N-glicosilación en RE (sobre-expresión de *SEC23* y *SEC59*). El transporte de RE a Golgi podría estar afectado en CM por la sub-expresión de *SLY41* y la sub-expresión de *WSC4* y *NCE103* en CM son indicativos de niveles bajos de estrés, ya que el estrés ocasiona la activación de estos genes. Por otro lado, bajo las condiciones del análisis, en CP se sobre-expresó solo *SBH1* que participa en complejo *SEC61* que conforma el poro en el proceso de translocación en el RE y se sub-expresó *SEC59* requerido para la viabilidad celular y para obtención de tasas normales de síntesis de intermediarios de lípidos, además de ser importante en la N-glicosilación de proteínas, indicando que las condiciones de CP reprime la expresión de genes involucrados en mantener condiciones metabólicas que favorezcan la viabilidad celular.

Las proteínas chaperonas involucradas en el control de calidad y plegamiento de proteínas en RE codificadas por *KAR2*, *SSA2*, *SCJ1*, *STH1*, *CNE1* y *HSP82* fueron sobre-expresadas en CM y de éstas, *KAR2*, *SSA2*, *CNE1* fueron sobre-expresadas diferencialmente en CM vs CP, indicando que estos genes estuvieron implicados en los niveles altos de producción de FTEII y su regulación es determinada por las condiciones de cultivo en CM. Por otro lado, *CCT4* y *HTM1*, expresados diferencialmente en CM y CP, y sin cambio significativo en CM causado por el cambio de la fuente de carbono, sin embargo sub-expresados en CP podrían estar influyendo en los niveles bajos de producción en CP, *HTM1* participa en replegamiento de proteínas y está directamente implicado con ERAD. Por último, *HSP82*, sobre-expresado sin diferencias en ambos genes, codifica para una chaperona que constituye del 1-2% de las proteínas en el citosol y es inducida por calor y estrés y es requerida durante el plegamiento de proteínas desnaturalizadas y encargada de regresarlas a su conformación nativa, puede ser una proteína determinante resultado de la respuesta al estrés provocado por las condiciones de cultivo y la expresión del gen heterólogo inducido por metanol.

De los genes involucrados en los procesos de glicosilación en RE, *ALG10* sobre-expresado en CM y *ALG6* sub-expresado diferencialmente en CM vs CP, indican actividad en el procesamiento temprano del N-glicano. Los tres genes relacionados con la N-glicosilación analizados, *OST2*, *OST3* y *SWP1*, se sobre-expresaron solo en CM debido al cambio de la fuente de carbono, pero solo *OST3* mostró sobre-expresión diferencial significativa en CM vs CP, indicando mayor actividad transcripcional de estos genes en CM. Con excepción de *PMT6*, todos los genes evaluados (*PMT1*, *PMT2*, *PMT4* y *PMT5*) implicados en la O-glicosilación en RE, se sobre-expresaron en CM, *PMT1* y *PMT6* se sobre-expresaron en CM vs CP, y solo *PMT4* se sobre-expresó en CP, indicando con esto, que la O-glicosilación presentó mayor actividad en CM que en CP. Cabe señalar que *PMT1*, *PMT2* se relaciona con el control de la N-glicosilación de las glicoproteínas de la pared celular de las levaduras (Ecker *et al.*, 2003) y con funciones de chaperona involucradas en dirigir proteínas mal plegadas a ERAD (Goder & Melero, 2011). *KRE5* sobre-expresado bajo las condiciones de CM, participa en el ciclo de calnexina relacionado con el control de calidad del plegado de las proteínas, indicando



bajo este contexto que el proceso de control de calidad de plegamiento de proteínas a través del ciclo de calnexina está más activo en CM que en CP.

En conclusión, los resultados muestran que bajo las condiciones de CM se presenta una mayor actividad transcripcional de los genes implicados en el procesamiento de las proteínas a través del RE durante la N- y O-glicosilación que permiten un estado fisiológico favorable para el plegamiento y secreción de proteína heteróloga.

De los genes involucrados en el procesamiento de glicoproteínas y en el tráfico de proteínas en Golgi, la transcripción de los genes involucrados con la síntesis de nucleótidos glicosilados fue más activa en CM que en CP, evento constatado por la sobre-expresión de *GAL10*, *MPG1* *MPGAI* en CM y la sub-expresión de *MPG1*, *GAF1*, *PMI40* en CP. Es importante señalar que *MPG1*, codifica una GDP manosa pirofosforilasa importante durante la biosíntesis de la pared celular, que parece de gran importancia en CM y que pudo haber afectado en CP favoreciendo a la lisis celular.

Los genes involucrados en el transporte de azúcares requeridos en la glicosilación en Golgi, *YEA4*, *UDP-GalT* y *VRG4* mantuvieron una transcripción más activa en CM. Por otro lado, la sobre-expresión en CM de *MNN11*, *MNN10*, *MNN4*, *MNN5*, es indicativo de que los procesos de la manosilación de glicoproteínas se mantuvieron con mayor actividad en CM que en CP. *OCHI* fue sub-expresado en ambas condiciones lo cual puede ser indicativo de que la ramificación  $\alpha$ -1,6 en la hipermanosilación, puede estar reprimida. Los genes *VPS21* y *VPS15*, las cuales codifican para proteínas que definen el tráfico de hidrolasas solubles vacuolares en la ruta del Golgi tardío a la vacuola, presentaron una mayor actividad transcripcional en CP, sugiriendo que el tráfico y distribución de proteínas de Golgi a vacuolas es más activo en CP que en CM, lo cual podría explicar la menor actividad proteolítica que se presentó en CM influyendo en la mayor producción de la proteína heteróloga.

El análisis de los genes involucrados (*YPS1*, *ATG4*, *PBN1* y *YPT6*) en la degradación o recambio proteico en sistemas vacuolares, indica que en general su actividad transcripcional está disminuida en CM. Habría que enfatizar que *PBN*, sub-expresado en CM, se relaciona con la activación de *PRB1*, proteasa vacuolar que fue activamente transcrita a lo largo del cultivo CM (Félix-Castro, 2013), esto sugiere que la

activación de PRB1 podría estar disminuida (como consecuencia de niveles bajos de PBN) y por lo tanto como consecuencia actividad proteolítica disminuida en CM.

En conclusión, la regulación de los genes involucrados en el procesamiento de las proteínas en RE y Golgi en CM, indican en general una mayor actividad transcripcional como respuesta a la adaptación requerida para la producción de la proteína recombinante, la cual es a su vez es inducida por el metanol pudieron haber influido en la sobre-producción de la proteína heteróloga en ésta condición.

Otros autores con análisis transcripcionales globales realizados en *P. pastoris* han reportado la sobre-expresión de genes involucrados en la translocación, procesamiento en RE y Golgi y la respuesta a UPR y ERAD debido a la secreción de proteínas inducidas con metanol. Así es el caso reportado por Hesketh y colaboradores (2013), en donde la inducción de la síntesis y secreción de la lisozima T70N mostró la sobre-expresión de estos genes y estos autores sugieren que los cultivos en lote favorecen a un mayor estado de estrés que conducen a la respuesta UPR y ERAD. Por otro lado, estudios realizados por Rebnegger y colaboradores (2014), en un cepa que produce y secreta albumina humana, cuya expresión fue regulada por el promotor GAP y por lo tanto bajo una expresión constitutiva, encontraron que la sobre-expresión de genes involucrados en la translocación de proteínas en el RE, plegamiento y secreción de proteínas fue dependiente de la velocidad de crecimiento, lo cual es congruente con una mayor actividad anabólica, demanda por el crecimiento.

### **Expresión diferencial de genes involucrados en la respuesta UPR, ERAD y autofagia.**

La respuesta de las levaduras al estrés ambiental está altamente relacionada y conectada con reacciones de estrés metabólico, y por tanto, cuando la célula entra en un estado fisiológico adaptativo en respuesta a algún estrés ambiental se genera resistencia a otros tipos de estrés, esto puede llevar a la señalización UPR que limita el plegamiento de proteínas en ER y perjudicar la producción de proteínas heterólogas (Mattanovich *et al.*, 2004).

Para el análisis de los genes relacionados a la respuesta al estrés por la aglomeración de proteínas mal plegadas inducida por las condiciones ambientales y una

tasa alta de producción de proteínas recombinante, se analizaron 44 genes. Los resultados mostraron un mayor número de genes sobre-expresados en la condición CP por efecto del cambio a metanol y un mayor número de genes sub-expresados en CM respecto a CP en el análisis de este grupo de genes.

En CM, por el cambio a metanol y al comparar CM vs CP, en los genes sobre-expresados predominaron los genes asociados a UPR, ERAD y SNARE. Por otro lado, en los genes sobre-expresados en CP predominaron los genes asociados a los procesos de UPR, ERAD y autofagia. Un resultado sobresaliente es la sub-expresión de genes relacionados con la autofagia en CM, sobre-expresados en CP y por lo tanto predominante en la sub-expresión diferencial en CM respecto a CP, estos los resultados reflejan un mayor estado de estrés en CP que en CM.

La sobre-expresión diferencial en CM vs CP de dos de los principales componentes de UPR, *KAR2* y *RLG1* y la sub-expresión de *HAC1* en CM por el cambio de la fuente de carbono, son resultados controvertidos en relación a la activación de la respuesta UPR. Algunos autores (Yoshida *et al.*, 2000; Graf *et al.*, 2008) afirman que el gen *HAC1* regula la activación de la respuesta UPR y otros (Hesketh *et al.*, 2013), indican que la sobre-expresión de genes relacionados a UPR, ERAD y tráfico de proteínas fueron independientes de la expresión y regulación de *HAC1*, mientras que Rebnegger y colaboradores (2014), indican que la activación de la respuesta UPR estuvo asociada a la sobre-expresión de *HAC1* y a su vez a la velocidad específica de crecimiento, aunque en este caso la expresión del gen heterólogo, dependiente de GAP, fue constitutiva y en un sistema de quimiostato.

En este trabajo, al menos en el tiempo evaluado, *HAC1* fue sub-expresado en CM y por lo tanto su influencia en la regulación de la expresión de genes relacionados con UPR, es cuestionable. Sin embargo es claro que *KAR2*, *RLG1*, *INO1*, *CPR5* y *PDII* fueron sobre-expresados en CM y diferencialmente en CM vs CP, con excepción de *PDII*, todos ellos relacionados con el control de calidad y la respuesta al estrés por el plegamiento de proteínas (Gasser *et al.*, 2008); y por lo tanto se puede inferir que este proceso fue más activo en CP que en CM. *SKO1* relacionado con estrés oxidativo y osmótico se sobre-expresó en CP pero no en CM reflejando la presencia de estrés oxidativo y osmótico en CP.

Diversas investigaciones han reportado la inducción de la expresión de BiP producto de *KAR2*, durante la expresión de proteínas recombinantes en diferentes especies de levaduras (Kauffman *et al.*, 2002; Hohenblum *et al.*, 2004) y que BiP y PDI que incrementan la secreción de éstas en *P. pastoris*, además la transcripción de *KAR2* y *PDII* activan el factor de transcripción HAC1p (Hesketh *et al.*, 2013).

Con respecto a los genes relacionados con la respuesta a ERAD, los resultados indicaron que el recambio proteico mediado por el proteosoma se encuentra activo en ambas condiciones, sin embargo, se presentó expresión diferencial entre los genes involucrados en este proceso, en donde la actividad de la vía de ubiquitinación (*RSP5* y *E3* genes involucrados en la ubiquitinación de la proteína que será procesada en el proteosoma) y la de control de calidad y estrés (*HSP78*, involucrada en el control de calidad del plegamiento de proteínas y procesos de estrés fisiológico) se sub-expresaron en CM vs CP, indicando una actividad transcripcional asociada a ERAD en ambas condiciones pero diferenciada entre ellas.

Todos los genes relacionados con la autofagia se sub-expresaron diferencialmente en CM vs CP indicando una menor actividad autofágica en CM respecto a CP. En levaduras, la autofagia es inducida bajo condiciones de estrés celular, como la falta de nutrientes, aumento de temperatura o estrés oxidativo; y está implicada en fenómenos de muerte celular no apoptóticos.

Por ésta razón, los resultados indican que, tanto la temperatura, el pH y la concentración de metanol, podrían haber jugado un papel determinante en la respuesta de adaptación al estrés ambiental, así entonces la sobre-expresión en CP de los genes involucrados en autofagia son una consecuencia de la respuesta al estrés que pudieron haber contribuido a una mayor actividad en el recambio proteico y celular, conduciendo a una lisis celular activa en CP y resultando en el pobre desempeño fisiológico en estas condiciones. Por otro lado la sub-expresión en CM de los genes involucrados en autofagia podría ser el resultado de una mejor adaptación en CM como consecuencia de un estrés poco inducido o mejor controlado bajo estas condiciones.

## **Validación de resultados obtenidos del análisis transcripcional mediante RT-qPCR**

La estandarización de la técnica de RT-qPCR es clave para el análisis de los niveles de expresión ya que permite obtener resultados confiables. La estandarización incluye, el muestreo y manejo de muestras; la preparación de RNA que incluye calidad, concentración y su conservación; la optimización de los ensayos de síntesis de DNAc y qPCR; y finalmente la selección de al menos dos genes normalizadores y el empleo de una muestra calibradora (Bustin *et al.*, 2010). El método de  $2^{-\Delta\Delta CT}$  representa el modelo matemático mayormente utilizado para la determinación de los niveles de expresión relativos (Pfaffl, 2006). Sin embargo, debido a que se obtuvieron diferentes eficiencias de amplificación cercanos al 100% para cada gen de estudio, se realizó el cálculo de los niveles de expresión relativa empleando el método de la  $2^{-\Delta\Delta CT}$  corregida (Félix-Castro, 2013).

La validación del análisis de RNA-seq mediante RT-qPCR, no siempre implica una correlación del 100 % entre los resultados obtenidos por las dos técnicas, ya que factores como errores en la secuenciación, el manejo de los programas informáticos e incluso por la presencia de splicing diferencial pueden intervenir en un resultado diferente. Por ejemplo, Qi y colaboradores (2011), analizaron veinte genes para la validar los resultados de un análisis transcripcional, mediante RT-qPCR, de un estudio de *P. pastoris* cultivada bajo condiciones de microgravedad comparada con el cultivo a la gravedad normal, encontraron que dos de los genes analizados no correlacionaron en los resultados obtenidos por las dos técnicas. Merrick y colaboradores (2013) validaron un análisis transcripcional realizado en ratas tratadas con AFB1 comparadas con las no tratadas y de once genes analizados, dos no correlacionaron entre las dos técnicas. En el presente trabajo, de los trece genes analizados por RT-qPCR, dos no correlacionaron con los resultados del análisis transcripcional. Uno de los genes fue *HTMI*, el cual se descartó debido a que presentó splicing alternativo en el análisis de DEG; además como testigo se introdujeron los resultados obtenidos con *KAR2* para verificar que los resultados fueran confiables, ya que se comparó la abundancia de transcritos de *KAR2* en ambas condiciones de cultivo con la obtenida por Félix-Castro (2013). Los resultados obtenidos en la correlación de los doce genes evaluados mediante la técnica RT-qPCR y el análisis transcripcional mostró un  $R^2 = 0.7$ , con solo

divergencia en la abundancia de transcritos de un solo gen, *AHP1*, al comparar los resultados obtenidos entre las dos técnicas. Por lo cual, la metodología empleada para el análisis transcripcional se puede asumir como confiable.

Los resultados del análisis transcripcional realizado en el presente trabajo, refleja que las condiciones empleadas en CM, (temperatura, pH, concentración de metanol y concentración inicial) promovieron la expresión de genes relacionados con la secreción, el plegamiento y la traducción que conducen a una alta producción extracelular de la proteína recombinante no atribuible a la regulación diferenciada a nivel transcripcional del gen heterólogo. Las condiciones de cultivo empleadas en CM permitieron la regulación de la ruta del metanol hacia la vía de asimilación y una mejor respuesta de adaptación al estrés que condujeron a un mayor crecimiento y producción de biomasa en CM que en CP.

Las condiciones empleadas en CP promovieron un estado de estrés celular, al cual la célula no fue capaz de adaptarse, conduciendo a la célula a una autólisis, lo cual se reflejó en el bajo consumo de metanol, menor actividad metabólica y por lo tanto menor crecimiento que en CM que impactaron en los bajos niveles de producción y rendimientos obtenidos.

## 10. CONCLUSIONES

- a) Este trabajo explora por primera vez los procesos metabólicos, a nivel transcripcional, que generan una alta producción de una proteína heteróloga en cultivos de *P. pastoris* a nivel de biorreactor.
- b) Las condiciones de cultivo en CM produjeron una actividad anabólica más alta que en CP reflejada por la sobre-expresión diferencial de genes relacionados con la replicación del DNA, procesos biosintéticos relacionados con proteínas, regulación de la traducción y glicosilación de proteínas.
- c) Todos los genes del metabolismo del metanol se sobre-expresaron en ambas condiciones sin diferencia, con excepción los genes DAS que se sobre-expresaron bajo las condiciones de CM.
- d) La sobre-expresión de los genes DAS de la vía de asimilación del metanol puede explicar la mayor actividad metabólica, velocidad específica de crecimiento y biomasa generada en CM respecto a las obtenidas en CP.
- e) Los niveles de transcritos del gen heterólogo fueron iguales en CM y CP y por lo tanto no influyeron en la diferencia en los niveles de producción de la proteína recombinante y rendimientos obtenidos en CM vs CP.
- f) Las condiciones en CM condujeron una mayor actividad transcripcional de los genes implicados en el procesamiento de las proteínas a través del RE, durante la N- y O-glicosilación y el procesamiento en Golgi que en las condiciones de CP, permitiendo un estado fisiológico favorable para el plegamiento y secreción de proteína la heteróloga
- g) La sobre-expresión diferencial de genes asociados a UPR, ERAD y SNARE en CM respecto a CP están relacionados con el estrés presente en ambas condiciones, ocasionado por las condiciones ambientales y la alta tasa de expresión del gen heterólogo.
- h) Las condiciones de CP condujeron a una pobre adaptación reflejado en la sobre-expresión de genes relacionados con el recambio proteolítico, la reducción de la actividad transcripcional del procesamiento en Golgi y la actividad transcripcional

de genes relacionados con la autofagia, la cual fue reprimida o no inducida en CM, reflejando un mayor estado de estrés en CP que en CM.

- i) Las condiciones empleadas en CP promovieron un estado de estrés celular, al cual la célula no fue capaz de adaptarse, conduciendo a la célula a una autólisis, lo cual se reflejó en el bajo consumo de metanol, menor actividad metabólica y por lo tanto menor crecimiento que en CM que impactaron en los bajos niveles de producción y rendimientos obtenidos.

### **CONCLUSIÓN FINAL:**

A pesar que la expresión del gen heterólogo no fue diferente en ambas condiciones de cultivo, las condiciones de CM favorecieron a la regulación de expresión de genes que condujeron a una mejor adaptación fisiológica y a la actividad transcripcional de genes que favorecieron el plegamiento y secreción de la proteína impactando en los rendimientos obtenidos bajo estas condiciones

## **11. PERSPECTIVAS**

Con la contribución de este trabajo se podrán realizar estudios más detallados de mecanismos poco estudiados en *Pichia pastoris* que puedan ser útiles en el desarrollo de nuevos diseños de ingeniería metabólica y la optimización de procesos para la producción de proteínas heterólogas u otros metabolitos.



## 12. BIBLIOGRAFÍA

1. Ahmad, M., Hirz, M., Pichler, H., and Schwab, H. (2014). Protein expression in *Pichia pastoris*: recent achievements and perspectives for heterologous protein production. *Appl. Microbiol. Biotechnol.* *98*, 5301–5317.
2. Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. *Genome Biol.* *11*, R106.
3. Asakura, M., Ninomiya, S., Sugimoto, M., Oku, M., Yamashita, S., Okuno, T., Sakai, Y., and Takano, Y. (2009). Atg26-Mediated Pexophagy Is Required for Host Invasion by the Plant Pathogenic Fungus *Colletotrichum orbiculare*. *Plant Cell* *21*, 1291–1304.
4. Baumann, K., Carnicer, M., Dragosits, M., Graf, A., Stadlmann, J., Jouhten, P., Maaheimo, H., Gasser, B., Albiol, J., and Mattanovich, D. (2010). A multi-level study of recombinant *Pichia pastoris* in different oxygen conditions. *BMC Syst. Biol.* *4*, 141.
5. Baumann, K., Dato, L., Graf, A.B., Frascotti, G., Dragosits, M., Porro, D., Mattanovich, D., Ferrer, P., and Branduardi, P. (2011). The impact of oxygen on the transcriptome of recombinant *S. cerevisiae* and *P. pastoris*-a comparative analysis. *BMC Genomics* *12*, 218.
6. Bernales, S., McDonald, K.L., and Walter, P. (2006). Autophagy Counterbalances Endoplasmic Reticulum Expansion during the Unfolded Protein Response. *PLoS Biol.* *4*.
7. Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.-H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinforma. Oxf. Engl.* *25*, 1091–1093.
8. Blankenberg D., Gordon A., Von Kuster G., Coraor N., Taylor J., Nekrutenko A. (2010) Galaxy Team. Manipulation of FASTQ data with Galaxy. *Bioinformatics.* *26*, 1783-1785.
9. Bollok, M., Resina, D., Valero, F., and Ferrer, P. (2009). Recent patents on the *Pichia pastoris* expression system: expanding the toolbox for recombinant protein production. *Recent Patents Biotechnol.* *3*, 192–201.
10. Bullard, J.H., Purdom, E., Hansen, K.D. and Dudoit, S. (2010) Evaluation of statistical methods for normalization and differential expression in RNAm-Seq experiments. *BMC Bioinf.* *11*, 94.

11. Bustin, S. A. (2002) Quantification of RNAm using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol.* 29, 23-29.
12. Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al. (2009). The MIQE Guidelines: Minimum Information for Publication of Quantitative Real-Time PCR Experiments. *Clin. Chem.* 55, 611–622.
13. Cereghino, J.L., Cregg, J.M. (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev.* 24, 45-66.
14. Chen, Z., and Duan, X. (2011). Ribosomal RNA depletion for massively parallel bacterial RNA-sequencing applications. *Methods Mol. Biol.* 733, 93–103.
15. Chiruvolu, V., Cregg, J. M., and Meagher, M. M. (1997) Recombinant protein expression in an alcohol oxidase-defective strain of *Pichia pastoris* in fedbatch fermentations. *Enzyme Microbiol Technol. Enzyme Microb. Technol.* 21, 277–283.
16. Chu, Y., and Corey, D.R. (2012). RNA Sequencing: Platform Selection, Experimental Design, and Data Interpretation. *Nucleic Acid Ther.* 22, 271–274.
17. Cregg, J.M. and Madden, K.R. (1987) Development of yeast transformation systems and construction of methanol utilization defective mutants of *Pichia pastoris* by gene disruption, in. *Biological Research on Industrial Yeasts* (Stewart, G. G., Russell, I., Klein, R. D. and Hiebsch, R. R., eds.), CRC Press, Boca Raton, FL 1–18.
18. Daly, R., Hearn, M.T.W. (2005). Expression of heterologous proteins in *Pichia pastoris*: a useful experimental tool in protein engineering and production. *J. Mol. Recognit.* 18, 119–138.
19. Damasceno, L.M., Anderson, K.A., Ritter, G., Cregg, J.M., Old, L.J., and Batt, C.A. (2007). Coexpression of chaperones for enhanced secretion of a single-chain antibody fragment in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* 74, 381–389.
20. Damasceno, L.M., Huang, C.-J., and Batt, C.A. (2012). Protein secretion in *Pichia pastoris* and advances in protein production. *Appl. Microbiol. Biotechnol.* 93, 31–39.
21. Delic, M., Valli, M., Graf, A.B., Pfeffer, M., Mattanovich, D., and Gasser, B. (2013). The secretory pathway: exploring yeast diversity. *FEMS Microbiol. Rev.* 37, 872–914.

22. De Schutter, K., Lin, Y.C., Tiels, P., Van Hecke, A., Glinka, S., Weber-Lehmann, J., Rouzé, P., Van de Peer, Y., and Callewaert, N. (2009). Genome sequence of the recombinant protein production host *Pichia pastoris*. *Nature Biotechnology* 27, 561–566.
23. Ecker, M., Mrsa, V., Hagen, I., Deutzmann, R., Strahl, S., and Tanner, W. (2003). *O*-Mannosylation precedes and potentially controls the *N*-glycosylation of a yeast cell wall glycoprotein. *EMBO Rep.* 4, 628–632.
24. Fang, Z., Martin, J.A., and Wang, Z. (2012). Statistical methods for identifying differentially expressed genes in RNA-Seq experiments. *Cell Biosci.* 2, 26.
25. Feldmesser, E., Rosenwasser, S., Vardi, A., and Ben-Dor, S. (2014). Improving transcriptome construction in non-model organisms: integrating manual and automated gene definition in *Emiliania huxleyi*. *BMC Genomics* 15, 148
26. Félix-Castro M. B. (2013) Análisis de la expresión de genes involucrados en los procesos de secreción, proteólisis y metabolismo del metanol de la cepa KM71FTEII de *P. pastoris* cultivada a dos condiciones de cultivo (tesis de pregrado). Universidad Autónoma de Nuevo León, México.
27. Feng, X., and Zhao, H. (2013). Investigating host dependence of xylose utilization in recombinant *Saccharomyces cerevisiae* strains using RNA-seq analysis. *Biotechnol. Biofuels* 6, 96.
28. Gasser, B., Maurer, M., Gach, J., Kunert, R., and Mattanovich, D. (2006). Engineering of *Pichia pastoris* for improved production of antibody fragments. *Biotechnology and Bioengineering* 94, 353–361.
29. Gasser, B., Sauer, M., Maurer, M., Stadlmayr, G., and Mattanovich, D. (2007a). Transcriptomics-Based Identification of Novel Factors Enhancing Heterologous Protein Secretion in Yeasts. *Applied and Environmental Microbiology* 73, 6499–6507.
30. Gasser, B., Maurer, M., Rautio, J., Sauer, M., Bhattacharyya, A., Saloheimo, M., Penttilä, M., and Mattanovich, D. (2007b). Monitoring of transcriptional regulation in *Pichia pastoris* under protein production conditions. *BMC Genomics* 8, 179.
31. Gasser B., Saloheimo M., Rinas U, Dragosits M., Rodríguez-Carmona E., Baumann K., Giuliani M., Parrilli E., Branduardi P., Lang C., Porro D., Ferrer P., Tutino M.L., Mattanovich D. and Villaverde A. (2008). Protein folding and conformational stress in microbial cells producing recombinant proteins: a host comparative overview. *Microb Cell Fact.* 7, 11.

32. Goder, V., and Melero, A. (2011). Protein O-mannosyltransferases participate in ER protein quality control. *J. Cell Sci.* *124*, 144–153.
33. Graf, A., Gasser, B., Dragosits, M., Sauer, M. et al., Novel insights into the unfolded protein response using *Pichia pastoris* specific DNA microarrays. *BMC Genomics* 2008, *9*, 39
34. Hamilton, S.R., Davidson, R.C., Sethuraman, N., Nett, J.H., Jiang, Y.W., Rios, S., Bobrowicz, P., Stadheim, T.A., Li, H.J., Choi, B.K., Hopkins, D., Wischnewski, H., Roser, J., Mitchell, T., Strawbridge, R.R., Hoopes, J., Wildt, S., Gerngross, T.U. (2006) Humanization of yeast to produce complex terminally sialylated glycoproteins. *Scien.* *313*,1441–1443
35. Harding, H.P., Novoa, I., Zhang, Y., Zeng, H., Wek, R., Schapira, M., and Ron, D. (2000). Regulated translation initiation controls stress-induced gene expression in mammalian cells. *Mol. Cell* *6*, 1099–1108.
36. Hartner, F., and Glieder, A. (2006). Regulation of methanol utilisation pathway genes in yeasts. *Microb Cell Fact.* *5*, 39.
37. Hartner, F.S., Ruth, C., Langenegger, D., Johnson, S.N., Hyka, P., Lin-Cereghino, G.P., Lin-Cereghino, J., Kovar, K., Cregg, J.M., and Glieder, A. (2008). Promoter library designed for fine-tuned gene expression in *Pichia pastoris*. *Nucleic Acids Res.* *36*, e76–e76
38. Hesketh, A.R., Castrillo, J.I., Sawyer, T., Archer, D.B., and Oliver, S.G. (2013). Investigating the physiological response of *Pichia* (*Komagataella*) *pastoris* GS115 to the heterologous expression of misfolded proteins using chemostat cultures. *Appl. Microbiol. Biotechnol.* *97*, 9747–9762.
39. Higgins, D.R., and Cregg, J.M. (1998). Introduction to *Pichia pastoris*. In *Pichia Protocols*, Spring. 1–15.
40. Hohenblum, H., Gasser, B., Maurer, M., Borth, N., Mattanovich, D. (2004). Effects of gene dosage, promoters, and substrates on unfolded protein stress of recombinant *Pichia pastoris*. *Biotechnol Bioeng.* *85*(4), 367-375.
41. Huang, W.P., and Klionsky, D.J. (2002). Autophagy in yeast: a review of the molecular machinery. *Cell Struct. Funct.* *27*, 409–420.
42. Idiris, A., Tohda, H., Kumagai, H., and Takegawa, K. (2010). Engineering of protein secretion in yeast: strategies and impact on protein production. *Appl. Microbiol. Biotechnol.* *86*, 403–417.
43. Illumina® Sequencing. TruSeq™ RNA and DNA Sample Preparation Kits v2. 20011.pdf (<http://www.illumina.com/applications/sequencing/rna.ilmn>).

44. Jacobs, P.P., Geysens, S., Vervecken, W., Contreras, R., Calleweert, N. (2009) Engineering complex-type N-glycosylation in *Pichia pastoris* using GlycoSwitch technology. *Nat Protocols* 4, 58-70.
45. Kamei, Y., Tamura, T., Yoshida, R., Ohta, S., Fukusaki, E., and Mukai, Y. (2011). GABA metabolism pathway genes, UGA1 and GAD1, regulate replicative lifespan in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* 407, 185–190.
46. Kauffman, K., Pridgen, E., Doyle, F., Dhurjati, P., Robinson, A. (2002) Decreased protein expression and intermittent recoveries in BiP levels result from cellular stress during heterologous protein expression in *Saccharomyces cerevisiae*. *Biotechnol Prog* 18(5), 942-950.
47. Kogenaru, S., Yan, Q., Guo, Y., and Wang, N. (2012). RNA-seq and microarray complement each other in transcriptome profiling. *BMC Genom.* 13, 629.
48. Krainer, F.W., Dietzsch, C., Hajek, T., Herwig, C., Spadiut, O., and Glieder, A. (2012). Recombinant protein expression in *Pichia pastoris* strains with an engineered methanol utilization pathway. *Microb. Cell Fact.* 11, 22.
49. Li, P., Anumanthan, A., Gao, X., Ilangovan, K., Suzara, V.V., Düzgünes, N., Renogopalakrishnan, V. (2007). Expression of recombinant proteins in *Pichia pastoris*. *Appl Biochem Biotechnol.* 142, 105-124.
50. Liang, S., Wang, B., Pan, L., Ye, Y., He, M., Han, S., Zheng, S., Wang, X., and Lin, Y. (2012). Comprehensive structural annotation of *Pichia pastoris* transcriptome and the response to various carbon sources using deep paired-end RNA sequencing. *BMC Genom.* 13, 738.
51. Livak, K.J., and Schmittgen, T.D. (2001). Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the  $2^{-\Delta\Delta CT}$  Method. *Meth.* 25, 402–408.
52. Lodi, T., Neglia, B., and Donnini, C. (2005) Secretion of human serum albumin by *Kluyveromyces lactis* overexpressing *KIPDII* and *KIERO1*. *Appl. Environ. Microbiol.* 71, 4359–4363.
53. Lussier M., Gentsch M., Sdicu A.M., Bussey H., and Tanner W. (1995). Protein O-Glycosylation in Yeast: the PMT2 gene specifies a second protein o-mannosyltransferase that functions in addition to the PMT1-encoded activity. *J. Biol. Chem.* 270, 2770–2775.
54. Macauley-Patrick, S., Fazenda, M.L., McNeil, B. and Harvey, L.M. 2005. Heterologous protein production using the *Pichia pastoris* expression system. *Yeast.* 22, 249– 270.

55. Mattanovich, D., Gasser, B., Hohenblum, H., and Sauer, M. (2004). Stress in recombinant protein producing yeasts. *J. Biotechnol.* *113*, 121–135.
56. Merrick, B.A., Phadke, D.P., Auerbach, S.S., Mav, D., Stiegelmeier, S.M., Shah, R.R. and Tice, R.R. (2013). RNA-Seq Profiling Reveals Novel Hepatic Gene Expression Pattern in Aflatoxin B1 Treated Rats. *PLoS ONE* *8*, e61768.
57. Moreau, K., Puri, C. and Rubinsztein, D.C. (2015). Methods to analyze SNARE-dependent vesicular fusion events that regulate autophagosome biogenesis. *Methods* *75*, 19–24.
58. Nett, J.H. and Gerngross, T.U. (2003) Cloning and disruption of the *Pp URA5* gene and construction of a set of integration vectors for the stable genetic modification of *Pichia pastoris*. *Yeast* *20*, 1279–1290.
59. Nookaew, I., Papini, M., Pornputtpong, N., Scalcinati, G., Fagerberg, L., Uhlen, M., and Nielsen, J. (2012). A comprehensive comparison of RNA-Seq-based transcriptome analysis from reads to differential gene expression and cross-comparison with microarrays: a case study in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* *40*, 10084–10097.
60. Osowski, C.M., and Urano, F. (2011). Measuring ER stress and the unfolded protein response using mammalian tissue culture system. *Meth. Enzymol.* *490*, 71–92.
61. Ozsolak, F., and Milos, P.M. (2010). RNA sequencing: advances, challenges and opportunities. *Nat. Rev. Genet.* *12*, 87–98.
62. Papini, M., Nookaew, I., Uhlen, M., and Nielsen, J. (2012). *Scheffersomyces stipitis*: a comparative systems biology study with the Crabtree positive yeast *Saccharomyces cerevisiae*. *Microb. Cell Factories* *11*, 136.
63. Pathway Analysis of *Pichia pastoris* to Elucidate Methanol Metabolism and its Regulation for Production of Recombinant Proteins. *Pornkamol Unrean Biotechnol. Prog.*, 2014, Vol. 30, No. 1
64. Peracchio, C., Alabiso, O., Valente, G., and Isidoro, C. (2012). Involvement of autophagy in ovarian cancer: a working hypothesis. *J. Ovarian Res.* *5*, 22.
65. Persengiev, S.P., and Green, M.R. (2003). The role of ATF/CREB family members in cell growth, survival and apoptosis. *Apoptosis Int. J. Program. Cell Death* *8*, 225–228.

66. Pfaffl, M.W. Relative quantification. Real-time PCR. Internat. Univer. Line. 63-82.
67. Rebnegger, C., Graf, A.B., Valli, M., Steiger, M.G., Gasser, B., Maurer, M., and Mattanovich, D. (2014). In *Pichia pastoris*, growth rate regulates protein synthesis and secretion, mating and stress response. *Biotechnol.* 9, 511–525.
68. Resina, D., Bollók, M., Khatri, N.K., Valero, F., Neubauer, P., and Ferrer, P. (2007). Transcriptional response of *P. pastoris* in fed-batch cultivations to *Rhizopus oryzae* lipase production reveals UPR induction. *Microb. Cell Fact.* 6, 21.
69. Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. *Nat. Biotechnol.* 29, 24–26.
70. Ruohonen, L., Toikkanen, J., Tieaho, V., Outola, M., Soderlund, H., and Keranen, S. (1997). Enhancement of protein secretion in *Saccharomyces cerevisiae* by overproduction of Sso protein, a late-acting component of the secretory machinery. *Yeast Chichester Engl.* 13, 337–351.
71. Salamin, K., Sriranganadane, D., Lechenne, B., Jousson, O., and Monod, M. (2010). Secretion of an Endogenous Subtilisin by *Pichia pastoris* Strains GS115 and KM71. *Appl. Environ. Microbiol.* 76, 4269–4276.
72. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13, 2498–2504.
73. Shenoy, S.K. (2007). Seven-Transmembrane Receptors and Ubiquitination. *Circ. Res.* 100, 1142–1154.
74. Solà A., Jouhten P., Maaheimo H., Sánchez-Ferrando, F., Szyperski T., Ferrer P. (2007). Metabolic flux profiling of *Pichia pastoris* grown on glycerol/methanol mixtures in chemostat cultures at low and high dilution rates. *Microb.* 153, 281-290.
75. Sreekrishna, K. (2010). *Pichia*, optimization of protein expression. *Encyclopedia of Industrial Biotechnology: Bioprocess, Bioseparation, and Cell Technology*, edited by Michael C. Flockinger.
76. Sterck, L., Billiau, K., Abeel, T., Rouzé, P., Van de Peer, Y. (2012). ORCAE: online resource for community annotation of eukaryotes. *Nat Methods* 9, 1041.  
 Online Resource: *Pichia pastoris*:

<http://bioinformatics.psb.ugent.be/orcae/overview/Picpa> and <https://bioinformatics.psb.ugent.be/gdb/pichia/>

77. Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D.R., Pimentel, H., Salzberg, S.L., Rinn, J.L., and Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat. Protoc.* 7, 562–578.
78. Thornton, B., and Basu, C. (2011). Real-time PCR (qPCR) primer design using free online software. *Biochem. Mol. Biol. Educ.* 39, 145–154.
79. Van der Klei, I.J., Yurimoto, H., Sakai, Y., and Veenhuis, M. (2006). The significance of peroxisomes in methanol metabolism in methylotrophic yeast. *Biochimica et Biophysica Acta (BBA). Mol. Cell Res.* 1763, 1453–1462.
80. VanGuilder H.D., Vrana K.E., Freeman W.M. (2008). Twenty-five years of quantitative PCR for gene expression analysis. *BioTechniques.* 44, 619-626.
81. Vanz, A.L., Lunsdorf, H., Adnan, A., Nimtz, M., Gurramkonda, C., Khanna, N., and Rinas, U. (2012). Physiological response of *Pichia pastoris* GS115 to methanol-induced high level production of the Hepatitis B surface antigen: catabolic adaptation, stress responses, and autophagic processes. *Microb. Cell Factories* 11, 103.
82. Vanz, A.L., Nimtz, M., and Rinas, U. (2014). Decrease of UPR-and ERAD-related proteins in *Pichia pastoris* during methanol-induced secretory insulin precursor production in controlled fed-batch cultures. *Microb. Cell Factories* 13, 23.
83. Vervecken W, Kaigorodov V, Callewaert N, Geysens S, De Vusser K, Contreras R. (2004) *In vivo* synthesis of mammalian-like, hybrid-type N-glycans in *Pichia pastoris*. *Appl Environ Microbiol* 70, 2639-2646.
84. Viader-Salvadó, J.M., Castillo-Galván, M., Fuentes-Garibay, J.A., Iracheta-Cárdenas, M.M., and Guerrero-Olazarán, M. (2013). Optimization of five environmental factors to increase beta-propeller phytase production in *Pichia pastoris* and impact on the physiological response of the host. *Biotechnol. Prog.* 29, 1377–1385.
85. Wang, Z., Gerstein, M., Snyder, M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nat Rev Genet.* 10,57–63.
86. Werten, M.W.T., and de Wolf, F.A. (2005). Reduced proteolysis of secreted gelatin and Yps1-mediated alpha-factor leader processing in a *Pichia pastoris* *kex2* disruptant. *Appl. Environ. Microbiol.* 71, 2310–2317.



87. Wu, J., Kaufman, R.J. (2006). From acute ER stress to physiological roles of the Unfolded Protein Response. *Cell Death Differ.* *13*, 374–384.
88. Wu, M., Shen, Q., Yang, Y., Zhang, S., Qu, W., Chen, J., Sun, H., and Chen, S. (2013). Disruption of YPS1 and PEP4 genes reduces proteolytic degradation of secreted HSA/PTH in *Pichia pastoris* GS115. *J. Ind. Microbiol. Biotechnol.* *40*, 589–599.
89. Yorimitsu, T., and Klionsky, D.J. (2005). Autophagy: molecular machinery for self-eating. *Cell Death Differ.* *12 Suppl 2*, 1542–1552.
90. Yoshida, H., Okada, T., Haze, K., Yanagi, H., Yura, T., Negishi, M., and Mori, K. (2000). ATF6 Activated by Proteolysis Binds in the Presence of NF-Y (CBF) Directly to the cis-Acting Element Responsible for the Mammalian Unfolded Protein Response. *Mol. Cell. Biol.* *20*, 6755–6767.
91. Yurimoto, H., Oku, M., and Sakai, Y. (2011). Yeast Methylophony: Metabolism, Gene Regulation and Peroxisome Homeostasis. *Internat. Jour. of Microb.* 1–8.
92. Zhang, W., Zhao, H.L., Xue, C., Xiong, X. H., Yao, X.Q., Li, X.Y., Chen, H.P., and Liu, Z.M. (2006). Enhanced Secretion of Heterologous Proteins in *Pichia pastoris* Following Overexpression of *Saccharomyces cerevisiae* Chaperone Proteins. *Biotechnol. Prog.* *22*, 1090–1095.
93. Zhang, A.L., Luo, J.X., Zhang, T.Y., Pan, Y.W., Tan, Y.H., Fu, C.Y., and Tu, F. (2009). Recent advances on the GAP promoter derived expression system of *Pichia pastoris*. *Mol. Biol. Rep.* *36*, 1611–1619.

## ANEXO 1. OLIGONUCLEÓTIDOS DISEÑADOS

**Tabla 41.** Oligonucleótidos específicos utilizados para la amplificación de los genes seleccionados de la vía de secreción.

Clave Gen	Gen	Proteína	Tamaño Secuencia	Oligo 5'	Oligo 3'	Amplicón (pb)
PAS_chr3_0870	<i>MPG1</i>	GDP-manosil pirofosforilasa	1095	GAACCTGTCTTTATCTCTCATC CCT	ACCTATCTTACAGTCCTTACCA ACC	134
PAS_chr2-1_0256	<i>PMT2</i>	O- manosiltransferasa	2253	TCCAGCCAACAACAGATTACA ACC	TGAACGAGACGAACAGACATA CCA	140
PAS_chr3_0891	<i>HTM1</i>	Alfa-manosidasa	2130	CACATCAAGCCAGCAGTCAA A	TCCAAAGCCATAGAACCCTCG	122
PAS_chr1-4_0294	<i>SSO2</i>	t-SNARE	861	TCAACACGTTCAAGAGACCGA GA	TGATAAGGACACAGATCCAGAA GCA	146
PAS_chr4_0552	<i>SSA2</i>	ATPasa	1938	AGGTCAAGGCTACTGCTGGT	CACAGGCGGTTCTCAATCTTCT C	150
PAS_chr1-4_0659	<i>AHP1</i>	Alquil hidroperoxido reductasa	513	TCGTTATTACCACTGACTC CTTT	CTTTAGATAGCCCAACTCTTTCG AC	125
PAS_chr2-1_0354	<i>RSP5</i>	Ubiquitina proteína ligasa	2304	GATGGGAAATGAGGCTGACT AACAC	TGGTCTAAGGACGAAGGCAATC T	103
PAS_chr4_0444	<i>E3</i>	Ubiquitina proteína ligasa	1095	CATCGTGAACATAATACGGCT TGGA	AGGCAAATGTAAAGAGAGGGT AGCA	108
PAS_chr3_0626	<i>WSC4</i>	Proteína de membrana RE	1200	TGTCAGGTAGTGAATGTTTCT GCTC	GAATGTATCCGTAATAGCCGTT TCCA	129
PAS_chr4_0578	<i>NCE103</i>	Anhidrasa carbónica	906	TCAACAAGCAGTCGGTCCTTT C	GAGTCAGAGCAGCCAATCCAC	142

**Tabla 42.** Oligonucleótidos específicos utilizados para la amplificación de los genes seleccionados de la vía del metabolismo del metanol.

Clave Gen	Gen	Proteína	Tamaño	Oligo 5'	Oligo 3'	Amplicón (pb)
PAS_chr1-1_0319	<i>FBA2</i>	Fructosa 1,6- bisfosfato aldolasa	1062	AGAGCCTTTGTTCTCCTCCCA	CCCACCAGTTATCCCAATCTC C	139
PAS_chr3_0867	<i>FGH</i>	S-formilglutation hidrolasa	900	GATGAGACGAAGACTGACAT GGA	TCTGAGGCATTGTTGGGAGTG	134
PAS_chr2-2_0131	<i>CAT</i>	Catalasa	1533	CCCACATTTCCACACCCA	AGAGAAGCAGGAGTTCCACG A	144

## ANEXO 2. CARACTERÍSTICAS DE OLIGONUCLEÓTIDOS DISEÑADOS

• <i>MPG1</i>							
Clave	PAS_chr3_0870						
Tamaño del producto	134						
Tamaño de la secuencia	1095						
Pair any comp	4.00						
3' comp	0.00						
Primer	inicio	long	TM	%GC	any	3'	secuencia
5'	686	25	62.20	44	2.00	0.00	GAACCTGTC TTTATCTCTCATCCCT
3'	819	25	62.14	44	3.00	0.00	ACCTATCTTACAGTCCTTACCAACC
• <i>PMT2</i>							
Clave	PAS_chr2-1_0256						
Tamaño del producto	140						
Tamaño de la secuencia	2253						
Pair any comp	2						
3' comp	0						
Primer	Inicio	long	TM	%GC	Any	3'	secuencia
5'	1060	24	64.06	45.83	2.00	2.00	TCCAGCCAACAACAGATTACAACC
3'	1199	24	64.32	45.83	2.00	0.00	TGAACGAGACGAACAGACATACCA
• <i>HTM1</i>							
Clave	PAS_chr3_0891						
Tamaño del producto	122						
Tamaño de la secuencia	2130						
Pair any comp	2						
3' comp	0						
Primer	inicio	long	TM	%GC	any	3'	secuencia
5'	895	21	61.85	47.62	2.00	0.00	CACATCAAGCCAGCAGTCAAA
3'	1016	21	62.29	52.38	2.00	2.00	TCCAAAGCCATAGAACCCTCG
• <i>SSO2</i>							
Clave	PAS_chr1-4_0294						
Tamaño del producto	146						
Tamaño de la secuencia	861						
Pair any comp	3						
3' comp	0						
Primer	inicio	long	TM	%GC	any	3'	secuencia
5'	657	23	64.30	47.83	4.00	0.00	TCAACACGTTCAAGAGACCCGAGA
3'	802	25	64.15	44.00	4.00	0.00	TGATAAGGACACAGATCCAGAAGCA
• <i>SSA2</i>							
Clave	PAS_chr4_0552						
Tamaño del producto	150						
Tamaño de la secuencia	1938						
Pair any comp	3						
3' comp	0						
Primer	inicio	long	TM	%GC	any	3'	secuencia
5'	644	20	63.35	55.00	3.00	0.00	AGGTCAAGGCTACTGCTGGT
3'	793	23	63.91	52.17	3.00	0.00	CACAGCGGTTCTCAATCTTCTC

**Figura 79.** Características de los oligonucleótidos que amplifican los genes seleccionados por su sobre-expresión de la vía de secreción diseñados con Primer3.

• <i>AHP1</i>							
Clave	PAS_chr1-4_0659						
Tamaño del producto	125						
Tamaño de la secuencia	513						
Pair any comp	3						
3' comp	0						
Primer	inicio	long	TM	%GC	any	3'	secuencia
5'	248	25	62.54	40	4.00	0.00	TCGTTATTACCACTGACACTCCTTT
3'	372	26	63.16	42.31	4.00	2.00	CTTAGATAGCCCAACTCTTTCGAC
• <i>RSP5</i>							
Clave	PAS_chr2-1_0354						
Tamaño del producto	103						
Tamaño de la secuencia	2304						
Pair any comp	3						
3' comp	0						
Primer	inicio	long	TM	%GC	any	3'	secuencia
5'	1049	25	64.15	48.00	3.00	0.00	GATGGGAAATGAGGCTGACTAACAC
3'	1151	23	63.91	47.83	3.00	0.00	TGGTCTAAGGACGAAGCAATCT
• <i>E3</i>							
Clave	PAS_chr4_0444						
Tamaño del producto	108						
Tamaño de la secuencia	1095						
Pair any comp	4.00						
3' comp	0						
Primer	inicio	long	TM	%GC	any	3'	secuencia
5'	705	25	63.98	44.00	3.00	0.00	CATCGTGAACATAATACGGCTTGA
3'	812	25	64.64	44.00	2.00	0.00	AGGCAAATGTAAAGAGAGGGTAGCA
• <i>WSC4</i>							
Clave	PAS_chr3_0626						
Tamaño del producto	129						
Tamaño de la secuencia	1200						
Pair any comp	3						
3' comp	0						
Primer	inicio	long	TM	%GC	any	3'	secuencia
5'	182	25	63.55	44.00	3.00	0.00	TGTCAGGTAGTGAATGTTTCTGCTC
3'	310	26	63.76	42.31	3.00	0.00	GAATGTATCCGTAATAGCCGTTTCCA
• <i>NCE103</i>							
Clave	PAS_chr4_0578						
Tamaño del producto	142						
Tamaño de la secuencia	906						
Pair any comp	2						
3' comp	0						
Primer	inicio	long	TM	%GC	any	3'	Secuencia
5'	248	22	63.54	50.00	3.00	0.00	TCAACAAGCAGTCGGTCCTTTC
3'	389	21	63.48	57.14	3.00	0.00	GAGTCAGAGCAGCCAATCCAC

**Figura 80.** Características de los oligonucleótidos que amplifican los genes seleccionados por su sub-expresión de la vía de secreción diseñados con Primer3.

- |                        |                 |
|------------------------|-----------------|
| <i>FBA2</i>            |                 |
| Clave                  | PAS_chr1-1_0319 |
| Tamaño del producto    | 139             |
| Tamaño de la secuencia | 1062            |
| Pair any comp          | 3.00            |
| 3' comp                | 0.00            |

Primer	Inicio	long	TM	%GC	Any	3'	Secuencia
5'	363	21	63.3	52.38	3.00	0.00	AGAGCCTTTGTTCTCCTCCA
3'	501	22	62.7	54.55	2.00	0.00	CCCACCAGTTATCCCAATCTCC

- |                        |               |
|------------------------|---------------|
| <i>FGH</i>             |               |
| Clave                  | PAS_chr3_0867 |
| Tamaño del producto    | 134           |
| Tamaño de la secuencia | 900           |
| Pair any comp          | 3.00          |
| 3' comp                | 0.00          |

Primer	Inicio	long	TM	%GC	Any	3'	Secuencia
5'	88	23	62.36	47.83	4.00	0.00	GATGAGACGAAGACTGACATGGA
3'	221	21	62.78	52.38	3.00	0.00	TCTGAGGCATTGTTGGGAGTG

- |                        |                 |
|------------------------|-----------------|
| <i>CAT</i>             |                 |
| Clave                  | PAS_chr2-2_0131 |
| Tamaño del producto    | 144             |
| Tamaño de la secuencia | 1533            |
| Pair any comp          | 3.00            |
| 3' comp                | 0.00            |

Primer	Inicio	long	TM	%GC	Any	3'	Secuencia
5'	456	21	63.08	52.38	2.00	0.00	CCCACATTTTCATCCACACCCA
3'	599	21	63.35	52.38	4.00	1.00	AGAGAAGCAGGAGTCCACGA

**Figura 81.** Características de los oligonucleótidos que amplifican los genes seleccionados por su sobre-expresión de la vía del metabolismo del metanol con Primer3.

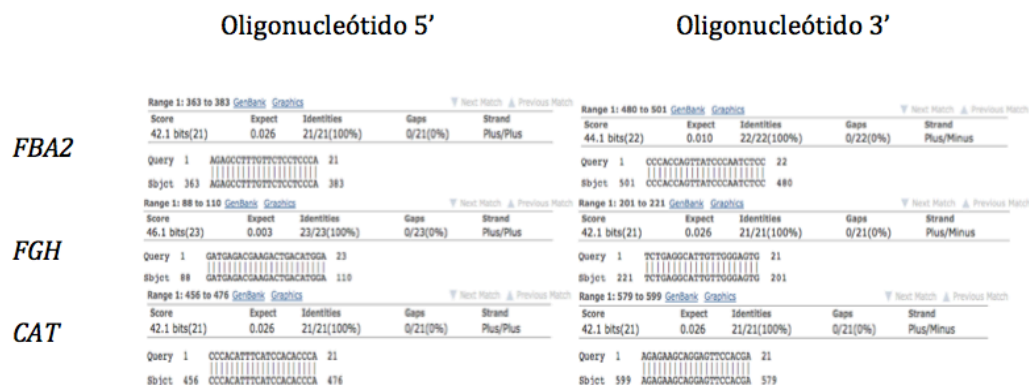
## ANEXO 3. ALINEAMIENTOS Y FORMACIÓN DE ESTRUCTURAS SECUNDARIAS



**Figura 82.** Alineamientos de los oligonucleótidos 5' y 3' que amplifican los genes seleccionados por su sobre-expresión de la vía de secreción.



**Figura 83.** Alineamientos de los oligonucleótidos 5' y 3' que amplifican los genes seleccionados por su sub-expresión de la vía de secreción.



**Figura 84.** Alineamientos de los oligonucleótidos 5' y 3' que amplifican los genes seleccionados por su sobre-expresión de la vía del metabolismo del metanol.

**Tabla 43.** Formación de estructuras secundarias de tipo horquilla, homodímero y heterodímero de las secuencias oligonucleotídicas que amplifican los genes involucrados en la vía de secreción.

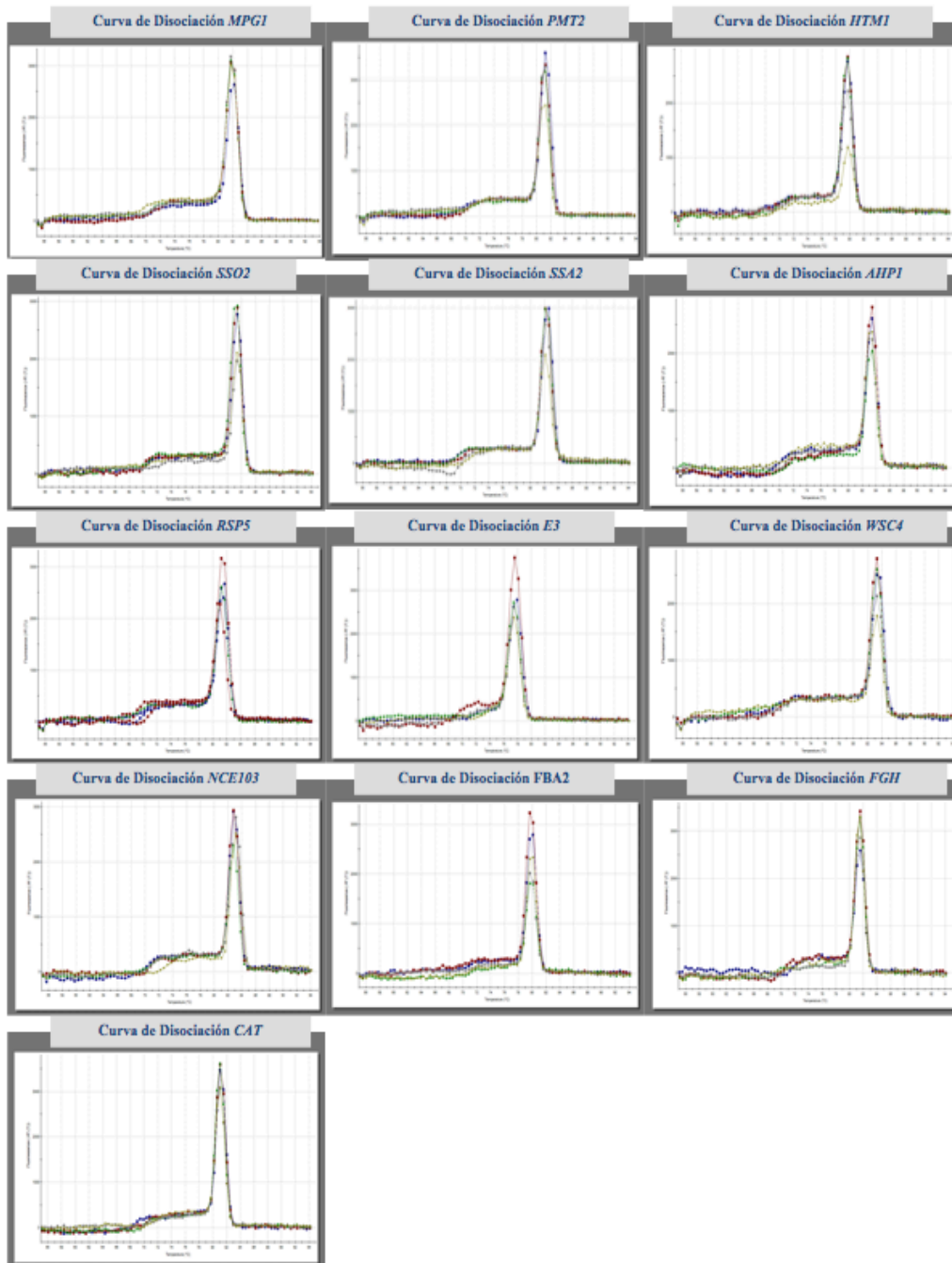
Gen	Oligonucleótido	$\Delta G$ (Kcal/mol)		
		Horquilla	Homodímero	Heterodímero
<i>MPG1</i>	5'	0.51	-1.95	-4.89
	3'	0.2	-1.94	
<i>PMT2</i>	5'	0.93	-3.14	-1.95
	3'	-1.03	-3.61	
<i>HTM1</i>	5'	1.03	-3.14	-3.14
	3'	0.36	-3.61	
<i>SSO2</i>	5'	-1.53	-6.3	-3.53
	3'	-1.83	-4.64	
<i>SSA2</i>	5'	-1.4	-3.14	-3.55
	3'	0.3	-3.61	
<i>AHP1</i>	5'	0.54	-3.61	-5.19
	3'	-0.07	-6.76	
<i>RSP5</i>	5'	-0.24	-3.14	-3.14
	3'	-1.48	-3.61	
<i>E3</i>	5'	-1.58	-4.95	-4.77
	3'	-0.27	-3.14	
<i>WSC4</i>	5'	-0.7	-3.55	-4.74
	3'	0.02	-3.61	
<i>NCE103</i>	5'	0.22	-3.61	-3.14
	3'	0.95	-3.14	

**Tabla 44.** Formación de estructuras secundarias de tipo horquilla , homodímero y heterodímero de las secuencias oligonucleotídicas que amplifican los genes involucrados en la vía del metabolismo del metanol.

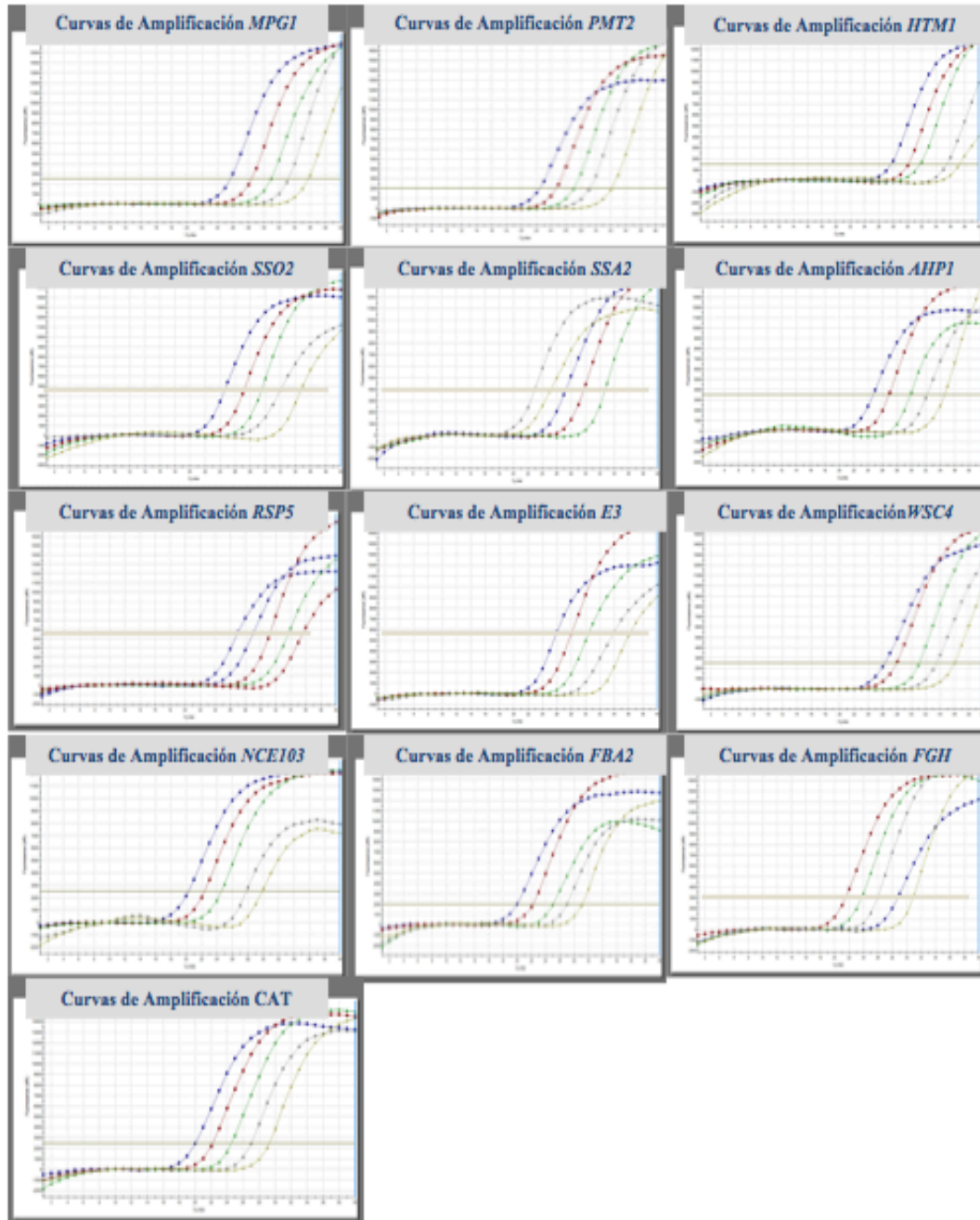
Gen	Oligonucleótido	$\Delta G$ (Kcal/mol)		
		Horquilla	Homodímero	Heterodímero
<i>FBA2</i>	5'	-0.52	-3.17	-3.9
	3'	1.02	-1.94	
<i>FGH</i>	5'	-0.32	-5.38	-3.43
	3'	-0.07	-3.14	
<i>CAT</i>	5'	2.04	-1.47	-4.64
	3'	-1.39	-4.64	



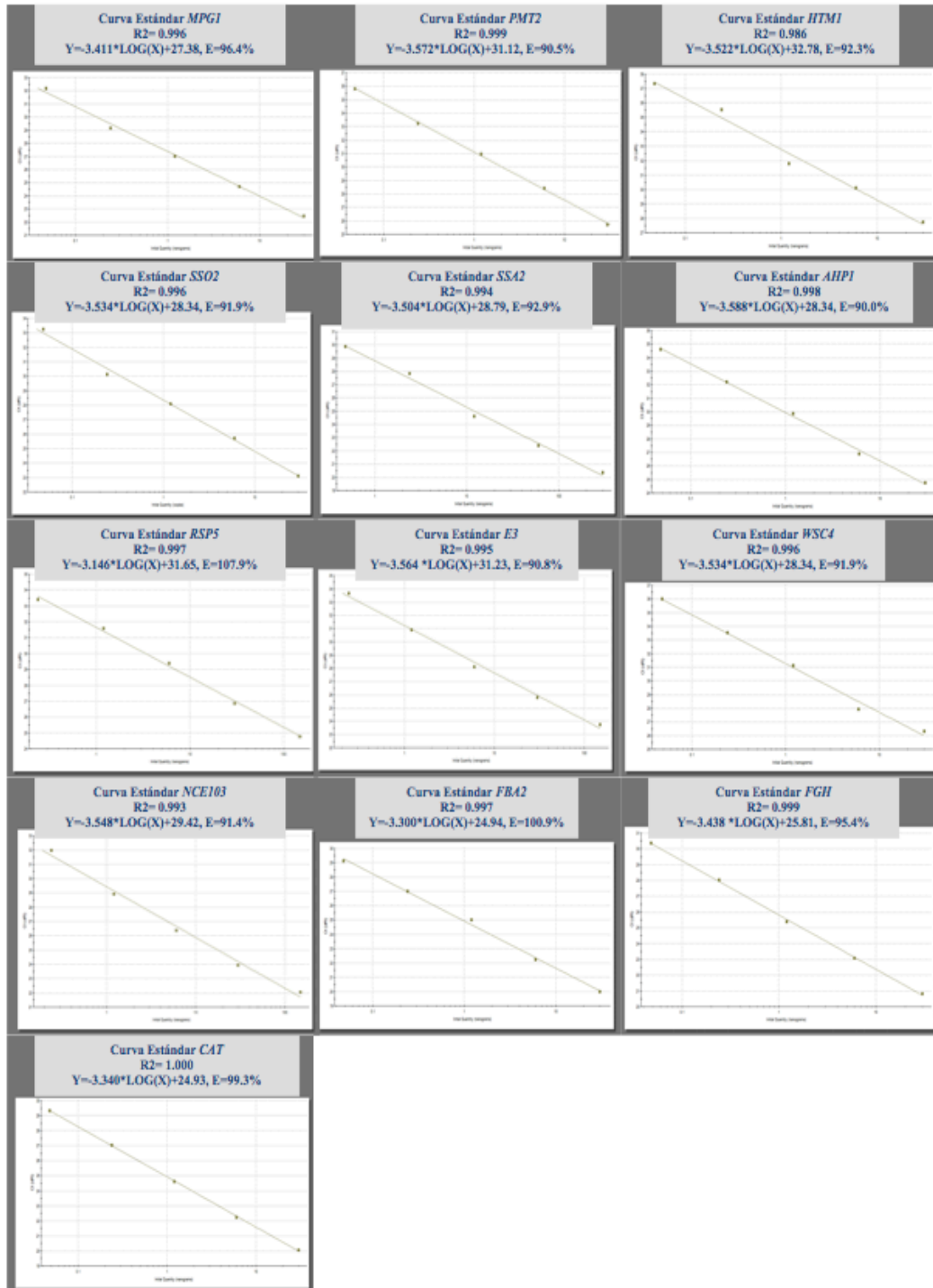
## ANEXO 4. CURVAS DE DISOCIACIÓN, AMPLIFICACIÓN Y ESTÁNDAR



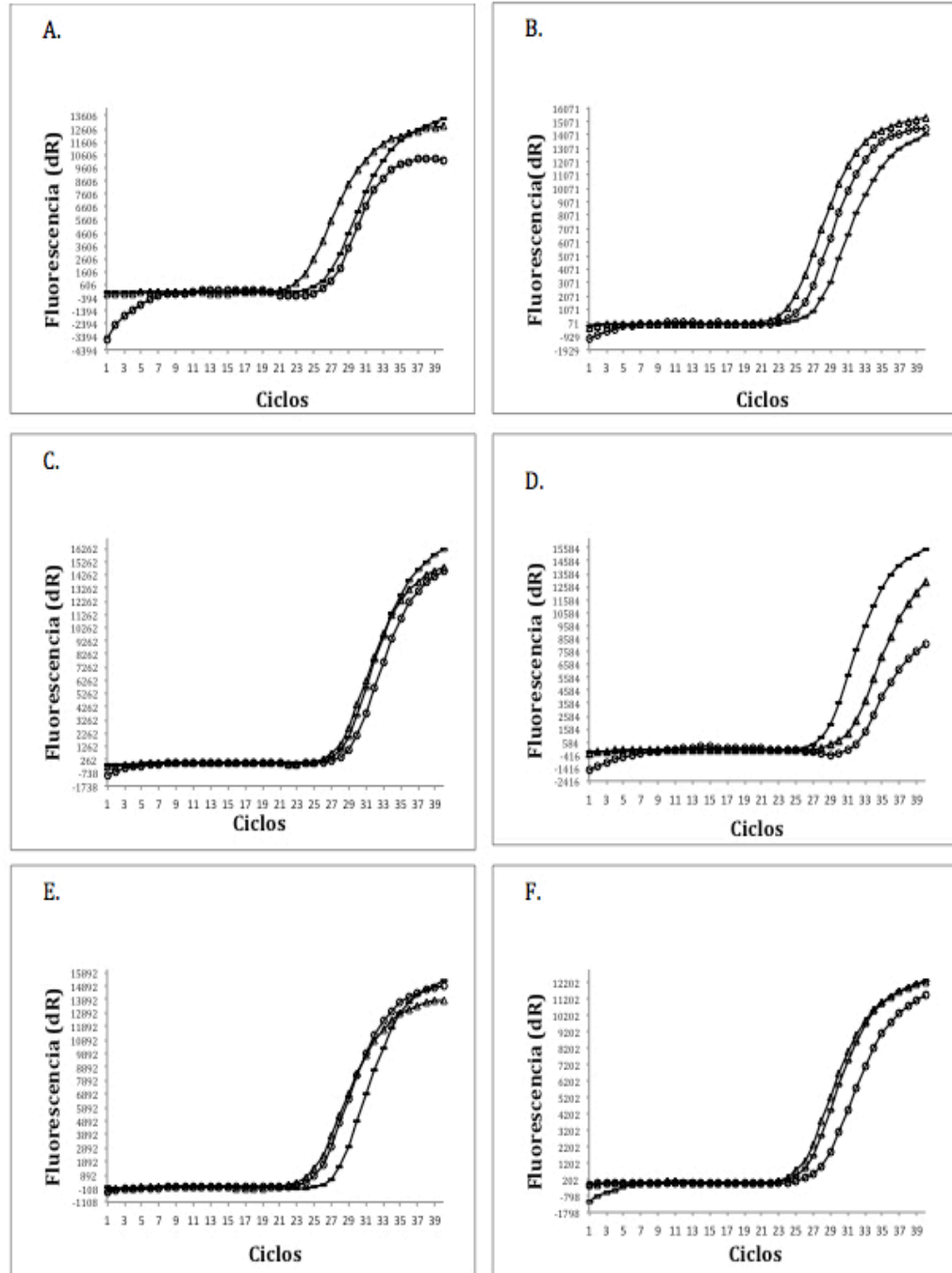
**Figura 85.** Curvas de disociación de todos los productos amplificados que corresponden a los genes seleccionados.



**Figura 86.** Curvas de amplificación correspondientes a los factores de dilución empleados para el cálculo de la eficiencia, linealidad e intervalos de linealidad correspondientes a los genes seleccionados.

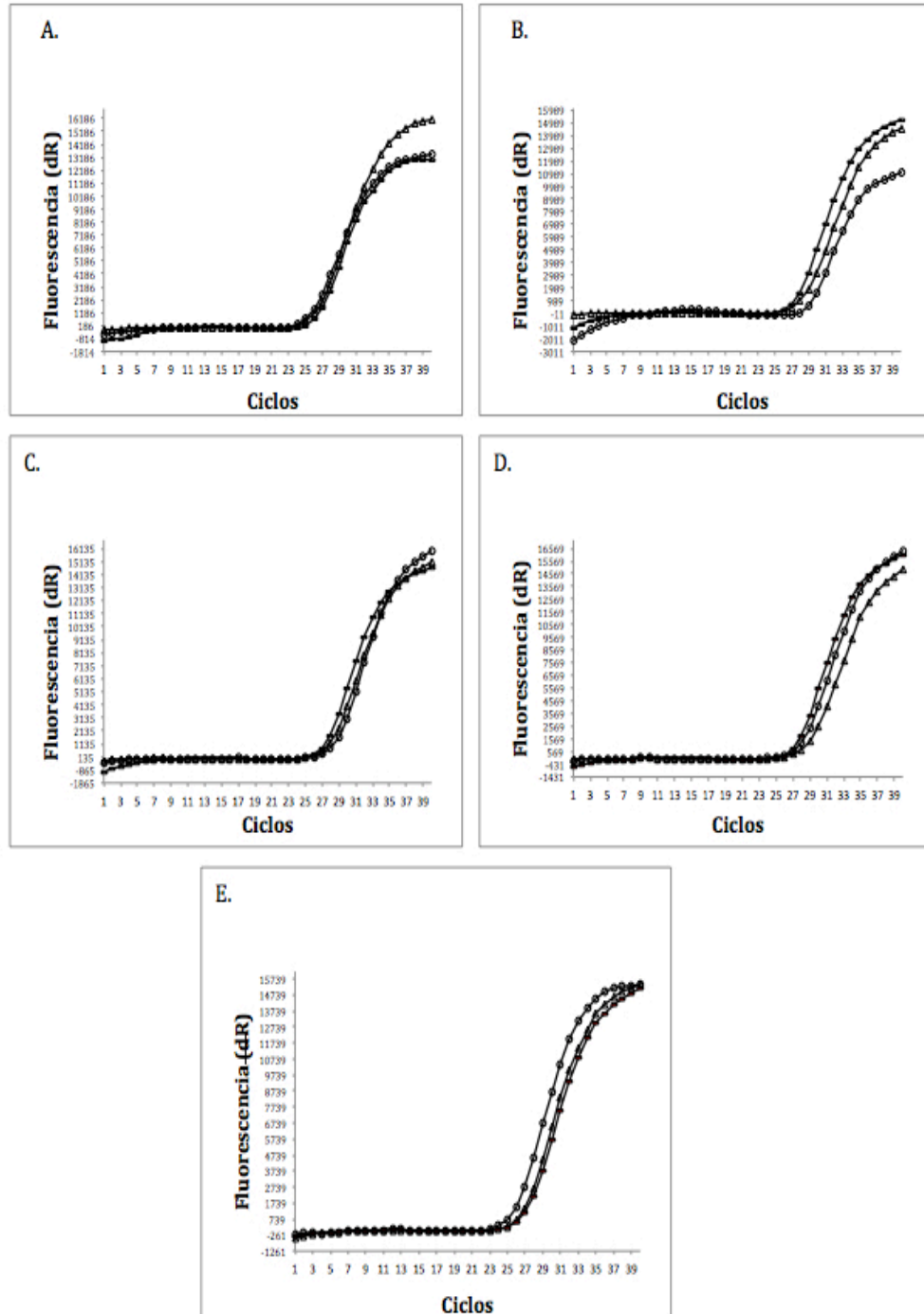


**Figura 87.** Curva estándar de cada uno de los productos de amplificación correspondientes a cada uno de los genes evaluados.



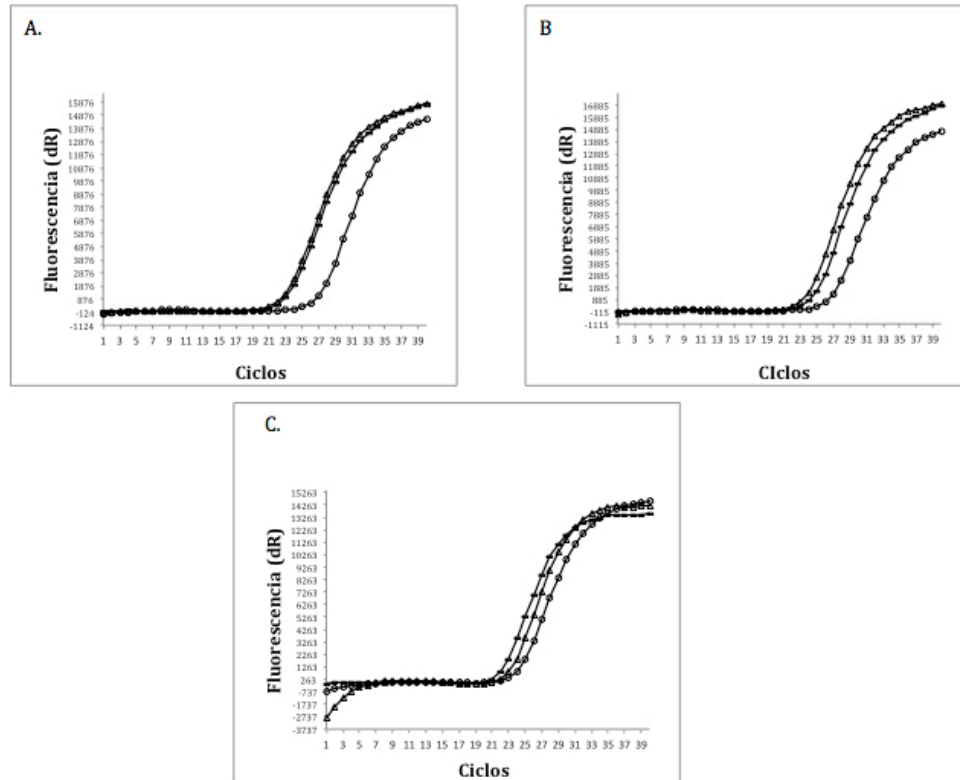
**Figura 88.** Curvas de amplificación de los genes sobre-expresados de la vía de secreción en las condiciones CM, CP y C0.

La figura o representa la condición C0 (control sin inducción con metanol),  $\Delta$  representa la condición CM de alta producción y  $\square$  representa la condición CP de baja producción. A. Gen *KAR2*, B. Gen *MPG1*, C. Gen *PTM2*, D. Gen *HTM1*, E. Gen *SSO2* y F. Gen *SSA2*.



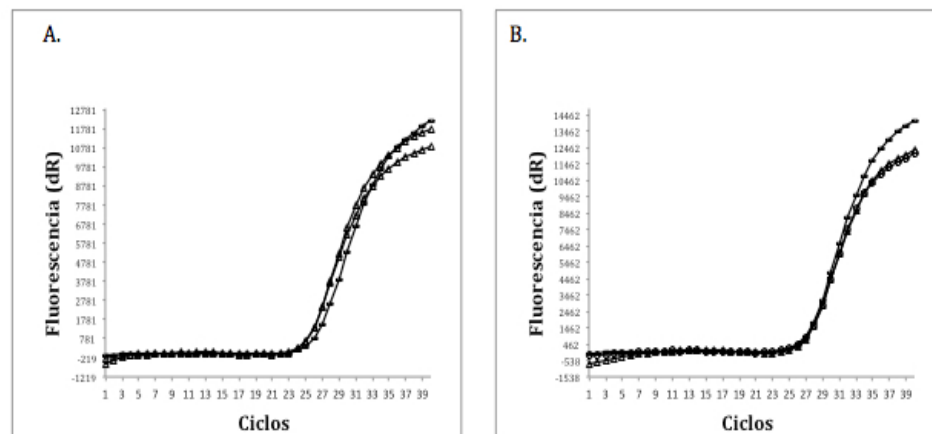
**Figura 89.** Curvas de amplificación de los genes seleccionados por su sub-expresión en la vía de secreción en las condiciones CM,CP y C0.

La figura  $\circ$  representa la condición C0 (control sin inducción con metanol),  $\Delta$  representa la condición CM de alta producción y  $\square$  representa la condición CP de baja producción. A. Gen *AHP1*, B. Gen *RSP5*, C. Gen *E3*, D. Gen *WSC4* y E. Gen *NCE103*.



**Figura 90.** Curvas de amplificación de los genes seleccionados por su sobre-expresión en la vía del metanol en las condiciones CM, CP y C0.

La figura  $\circ$  representa la condición C0 (control sin inducción con metanol),  $\Delta$  representa la condición CM de alta producción y  $\square$  representa la condición CP de baja producción. A. Gen *FBA2*, B. Gen *FGH* y C. Gen *CAT*.



**Figura 91.** Curvas de amplificación de los normalizadores empleados para el calculo de la expresión relativa en las condiciones CM, CP y C0.

La figura  $\circ$  representa la condición C0 (control sin inducción con metanol),  $\Delta$  representa la condición CM de alta producción y  $\square$  representa la condición CP de baja producción. A. Gen *G6PD* y B. Gen *YPT1*.

# RESUMEN AUTOBIOGRÁFICO

Karla Beatríz Fernández Cano

Candidata para el Grado de

Maestro en Ciencias con acentuación en Microbiología

**Tesis: ANÁLISIS TRANSCRIPCIONAL DE LA VÍA DE SECRECIÓN Y METABOLISMO DEL METANOL DURANTE LA PRODUCCIÓN DE UNA PROTEÍNA HETERÓLOGA EN *Pichia pastoris***

**Campo de estudio:** Ciencias Naturales y Exactas

**Datos Personales:** Nacida en Orizaba, Veracruz el 15 de septiembre de 1988, hija de Carlos Fernández Armida y Sonia Cano García.

**Educación:** Egresada de la Universidad Autónoma de Nuevo León, grado obtenido Químico Bacteriólogo Parasitólogo en el año 2012.

**Experiencia Profesional:**

Elaboración de Tesis de Licenciatura en el Laboratorio de virología (L7) del departamento de Inmunología y Virología de la Facultad de Ciencias Biológicas (UANL) 2011-2012.

Elaboración de tesis de Maestría en el laboratorio de Biotecnología Molecular (L5) del Instituto de Biotecnología de la Facultad de Ciencias Biológicas (UANL) 2013-2015