Perspective

Pharmaceutical BIOPROCESSING

Insights on biomarkers from Chinese hamster ovary 'omics' studies

As efforts for selecting high-producing cells during cell line development are moving toward the use of high-throughput, automated methods, there is considerable promise for identifying suitable markers arising from transcriptomic, proteomic and metabolomic studies. Herein we discuss the opportunities presented from recent characterization studies of high producers as well as comparisons with non- and lower-producing clones. We further put forward unique characteristics of cell lines with high productivity, which can be used for cell line screening and genetic engineering.

Sarantos Kyriakopoulos[1] & Cleo Kontoravdi[*,1]

[1]Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, UK

[*]Author for correspondence:
Tel.: +44 (0)20 7594 6655
cleo.kontoravdi98@imperial.ac.uk

The production of glycoprotein-based therapeutics involves a complicated, expensive and lengthy development process, a critical step of which is arguably generating the producer cell line [1–3]. Two obvious properties are required for this cell line: good growth characteristics, in other words, maintaining a high integral of viable cell concentration in culture; and sustained high specific protein productivity (q_p). Most cell line development platforms select for cell lines with good growth characteristics at the early stages of the process, usually picking the fastest growers. As previous studies have shown [4], this approach does not necessarily identify the best cell lines in terms of overall characteristics and often misses the top producers. It is therefore important to screen cell lines for a range of properties including growth, metabolism and specific protein productivity. Current research is focusing on developing assays for Chinese hamster ovary (CHO) cell line selection based on known characteristics of high producers from the wealth of published studies, in which producer cell lines from the same parental cell line exhibiting different characteristics in terms of growth rate and recombinant protein productivity are compared using –omics technologies.

In this review, we summarize recent comparative studies on CHO cells using **transcriptome**, **proteome** and metabolome profiling and highlight prominent targets that have been found to be correlated with cell growth and recombinant protein productivity. We further compare these targets with current strategies for genetically engineering high producing cell lines. We expect that these characterization efforts will shed light on appropriate selection criteria during cell line development, especially given the industry's investment in automated high-throughput devices, or provide targets for genetic modification of host cell lines.

Overview of most interesting genes & proteins

Table 1 summarizes recent studies involving the comparison of different producer cell lines from the same parent cell line. It focuses on genes that have been found to be positively correlated with recombinant protein productivity at the mRNA or protein level in two or more studies as classified by their biological role. The complete list of studies can be found in Supplementary File 1 (see online at www.futuremedicine.com/doi/suppl/10.2217/pbp.14.45). As expected, the majority of genes identified in these studies are involved in aspects of protein synthesis and processing. These range from genes involved in transcription regulation (*E2f6, Eif3*) [5,6] and mRNA processing (*Hspa8*) [7], to protein synthesis (*Ccdc122*, *Ccdc72*, *Rps20*,

FUTURE SCIENCE part of fsg

Table 1. Genes positively correlated with recombinant protein productivity as identified in more than one study based on their biological role.

Gene ID	Protein name	Biological role†	-omics technique; sampling phase; process mode	Cell line	Product	Ref.
Acaa2	Acetyl–Coenzyme A Acyltransferase 2	Fatty acid oxidation in mitochondria	Transcriptomics (Microarray), Proteomics (LC–MS/MS); Exponential and Stationary; Fed-batch	17 CHO cell lines including parental	mAb	[9]
Acs14	Acyl CoA synthetase	Lipid metabolism	Transcriptomics (Microarray); Exponential and Stationary; Fed-batch	CHO-DUXB-11 (nine different subclones)	mAb	[5]
Ccdc122	Coiled-coil domain-containing protein 122	Protein synthesis	Transcriptomics (Microarray); Exponential and Stationary; Fed-batch with temperature shift	CHO-K1 (two cell lines with sustained q_p and two with varying q_p)	mAb	[8]
Ccdc72	Coiled–coil domain-containing protein 72	Protein synthesis	Proteomics (LC–MS/MS); Exponential and Stationary; Fed-batch	17 CHO cell lines including parental	mAb	[9]
E2f6	Transcription factor E2F6	Regulation of transcription	Transcriptomics (Microarray); Exponential and Stationary; Fed-batch	CHO-DUXB-11 (nine different subclones)	mAb	[5]
Eif3	40S ribosome, eukaryotic translation initiation factor 3 (EIF3)	Regulation of transcription	Proteomics (LC–MS); Exponential, Stationary, Decline; Fed-Batch	CHO-DG44 (high producer transfected with antiapoptotic Bcl-X_L)	Fusion protein	[6]
Hspa5	Glucose-regulated protein 78 (BiP)	Unfolded protein response	Proteomics (LC–MS); Exponential, Stationary, Decline; Fed-Batch	CHO-DG44 (high producer transfected with antiapoptotic Bcl-X_L)	Fusion protein	[6]
Hspa8	Heat shock cognate 71 kDa protein	mRNA processing; protein folding	Transcriptomics (Microarray); Stationary; Fed-Batch with temperature shift	CHO-DUX and CHO-K1 (ten cell lines varying in q_p between 0.81 and 50.4 pg protein/cell/day)	mAb and Fc fusion protein	[7]
Rps20	40S ribosomal protein S20	Protein synthesis	Transcriptomics (Microarray); Exponential and Stationary; Fed-batch	CHO-DUXB-11 (nine different subclones)	mAb	[5]
Rps6	Ribosomal protein S6	Protein synthesis; mTOR pathway	Transcriptomics (Pathway-focused PCR array for mouse genes) and Proteomics (LC–MS); Exponential and Stationary and Decline; Batch and Fed-batch	CHO-K1SV (two cell lines, 17.4-fold change in q_p, initially, then six varying from 2 to 50 pg/cell/day) and CHO-DG44 (high producer transfected with antiapoptotic Bcl-X_L)	mAb	[6,10]

Note that cell lines CHO-DUXB-11, CHO-DUX and CHO-DUKX are the same and originate by the mutagenesis of CHO-K1 to remove DHFR activity.
CHO: Chinese hamster ovary; q_p: Specific recombinant protein productivity; mAb: Monoclonal antibodies.
†Biological role was cross-validated using UniprotKB database [13].

Rps6) [5,8–10]. A gene related to the **unfolded protein response** (*Hspa5*, encoding for BiP) also features in the list. This is unsurprising since the encoded endoplasmic reticulum (ER) chaperone (BiP) and foldases (for example PDI), associated with protein folding in the ER, have long been targets for genetic modification to improve productivity in CHO [11,12] but also yeast cells albeit with varying results.

Interestingly, only one gene was identified in more than one study: *Rps6* encoding for the 40S ribosomal protein S6 identified in [10] and [6]. The majority of recombinant protein products evaluated are monoclonal antibodies (mAbs), although there are also comparative studies between cell lines expressing mAbs and fusion proteins with similar findings [7]. In terms of metabolic markers, genes related to fatty acid oxidation (*Acaa2*) and lipid (*Acs14*) metabolism have been found to be positively correlated to recombinant protein productivity [5,9], as identified with analysis of mAb-producing CHO cell lines at both the transcriptomic and proteomic level.

It is also interesting to identify genes that are negatively correlated with recombinant productivity, which can be used in cell line selection as well. The complete list is again presented in Supplementary File 1 of the supplementary material. A summary of genes with the same or similar biological role reported in two or more studies is shown in Table 2, and comprises only two ribosomal proteins (S12 and SA) and two ribosomal protein kinases (ribosomal protein S6 kinase α-1 and α-2) identified at the transcriptomic level [10,14].

Since recombinant protein titer is a function of both specific productivity and cell growth, identifying genes and functions that are positively correlated with cell growth is also of interest. Differentially expressed genes found to be positively correlated with growth in two or more studies are presented in Table 3 (the complete list can be found in Supplementary File 2). Genes involved in mRNA processing (*Hnrnpa2b1*, *Hnrnpc*) [15,16], the unfolded protein response (*Hspd1*) [16,17], intracellular transport of proteins (*Pcna*, *Vcp*) [7,15] and cell cycle (*Mcm5*) [7,15] were identified as positively correlated with cell growth. A similar analysis has also been performed in [15].

Table 4 (the complete list can be found in Supplementary File 2) presents genes reported to be negatively correlated with cell growth. None of these genes were identified in more than one study, so instead we report genes with similar biological role identified in two or more studies. Genes associated with various types of motility (*Actb*, *Actg1*) [16,17], with pathways involved in glycosylation (*Man2a1*, *Uap1*) [16,17], with cell redox homeostasis (*Txndc4*, *Txndc5*) [7,16] and with mitochondrial electron transport (*Uqcrb*) [16] and the tricarboxylic acid (TCA) cycle (*Idh3a*) [17] were all identified as negatively correlated with cell growth.

Overall, a closer look at the global findings of transcriptomic and proteomic studies reveals that relatively few events are correlated to productivity compared with those linked to cell size and growth, as also concluded in [9]. It is also interesting that although some trends emerge, certain classes of genes, for example, those encoding for ribosomal proteins, can be either positively (*Rps20* in [5]) or negatively (*Rps12* in [14]) correlated with productivity.

Key Term

Transcriptome: Collection of all RNA molecules, the monitoring of which allows characterization of transcriptional activities.

Proteome: Set of all proteins expressed, in this case, by a single cell or a population of cultured cells.

Unfolded protein response: Multibranched pathway that is caused by the accumulation of un/misfolded proteins in the endoplasmic reticulum and acts via three mechanisms to restore homeostasis.

Transcriptomic & proteomic traits of high producers

Several groups have conducted comparative transcriptomic and proteomic studies between high and low producers derived from the same parental cell line. Vishwanathan *et al.* [5] performed transcriptomic analysis of nine mAb-producing DHFR-CHO (CHO-DUXB-11 derived) cell lines. Their analysis revealed that the characteristics of the parent cell line are different than that of producers and that the characteristics of certain clones are carried forward through selection and methotrexate amplification to their subclones regardless of the productivity levels of the latter. A set of genes were found to be differentially expressed between parent and transfected clones: genes associated with EDG-1, PDGF, mTOR, Toll-like receptor and cytokine signaling pathways were upregulated. Enriched EDG-1, PDGF, Ras and extracellular matrix receptor interaction was also found to be differentially expressed between clones along the amplification process. Finally, a subset of 15 genes was found to be differentially expressed between low and high producers. These include cell cycle (*Ncapd2*, *Tfdp1*), transcription regulation (*E2f6*), chromatin organization (*H2afy*), lipid metabolism (*Hmgcr*, *Acs14*, *Hsd17b12*), protein processing (*Dnpep*, *Fn1*, *Timp2*), protein secretion (*Arfrp1*, *Ckap4*), protein synthesis (*Rps20*) and signaling pathways (*Sar1b*, *Pask*).

Carlage *et al.* [6] performed proteomics (LC–MS) analysis of low and a high producing CHO (CHO-DG44 derived) cell lines expressing a recombi-

Table 2. Genes with very close biological role that were found to be negatively correlated with recombinant protein productivity in more than one study.

Gene ID	Protein name	Biological role†	-omics technique; sampling phase; process mode	Cell line	Product	Ref.
Rps12	Ribosomal protein S12	Protein synthesis	Transcriptomics (Microarray and Coexpression analysis); Multiple phases; multiple process modes	CHO-DUX and CHO-K1; 167 cell lines (ranging from -0.0058 to 0.0467 h^{-1})	mAb, fusion protein, growth factors, coagulation factors, nonproducing	[14]
Rpsa	Ribosomal protein SA					
Rps6ka1	Ribosomal protein S6 kinase α-1	mTOR pathway related; Intracellular signal transduction	Transcriptomics (Pathway-focused PCR array for mouse genes); Exponential; Batch	CHO-K1SV (two cell lines, 17.4-fold change in q_p, Initially, then six varying from 2 to 50 pg/cell/day)	mAb	[10]
Rps6ka2	Ribosomal protein S6 kinase α-2					

Note that cell lines CHO-DUXB-11, CHO-DUX and CHO-DUKX are the same and originate by the mutagenesis of CHO-K1 to remove DHFR activity.
CHO: Chinese hamster ovary; mAb: Monoclonal antibodies.
†Biological role was cross-validated using UniprotKB database [13].

nant fusion protein. The high producing cell line was also engineered to overexpress the *Bcl-x_L* gene and samples for the analysis were taken at exponential, stationary phase and decline phases of cell culture. Thirty-two differentially expressed proteins were identified from the total of 392 resolved.

Clarke *et al.* [7] developed a partial least squares model able to predict cell-specific productivity within 4.44 pg/cell/day based on microarray data including 287 genes (212 of which were annotated). Data were collected at stationary phase of fed-batch cell culture with temperature shift of ten different CHO (CHO-K1 and CHO-DXB11 derived) cell lines producing an mAb or Fc fusion protein. Prioritized genes involved protein folding (*Canx*, *Hspa8*, *Pdia4*, *Ppid*), protein processing (*Anxa2*, *Ctsl*, *Psmd4*, *Nedd4*), intracellular protein transport (*Lman2*, *Nsf*, *Rab6a*, *Rtn3*), lipid metabolism (*Npc1*), transcription regulation (*App*, *Pparbp*) and regulation of mitosis (*Cdc20*) biological processes, which were all found to be positively correlated with productivity.

Doolan *et al.* [8] performed microarray analysis of mAb-producing CHO (CHO-K1 derived) cell lines varying in their ability to sustain q_p during the course of fed-bath culture. In total, 22 gene transcripts were found to be differentially expressed between cell lines that sustained high productivity versus cell lines that did not sustain this phenotype. Two of those genes were also verified by quantitative reverse transcription-PCR: *Cryab*, associated with protein folding and *Mgst1,* associated with antioxidative activity.

Kang *et al.* [9] used proprietary microarray chip-based transcriptomics and LC–MS/MS proteomics analysis to study 17 different DHFR-CHO cell lines varying in productivity during a 13-day fed-batch culture. Samples for transcriptomic and proteomics analyses were collected at exponential phase and correlated with the titer on days 10 or 13. The researchers identified several proteins positively correlated with q_p, including proteins associated with DNA repair (DDB1), intracellular trafficking (AP3D1, AP2B2) and ER translocation efficiency (SRPR). DHFR protein levels were also well correlated with productivity. Similarly with an aforementioned study [5], no correlation was observed with heavy or light chain of the recombinant mAb and productivity. Interestingly, the only top ranked protein positively correlated with q_p at the proteomic level that was also top ranked at the transcriptomic level was ACAA2. This is an enzyme involved in fatty acid oxidation in the mitochondria, which is associated with TCA cycle activity. Proteins negatively correlated with q_p include folding chaperones (CCT3, TCP1, CCT4, CCT7 and CCT8), suggesting poorer folding efficiency in low producers. Another negatively correlated protein is MTHFD2, which is involved in folate metabolism and may be associated with efforts of cells with low levels of DHFR to survive. This is because MTHFD2 supplies precursors for metabolites involved with the DHFR enzyme. On the transcriptomics side, along with the aforementioned *Acaa2* transcript, other genes positively correlated with q_p include calcium sensing and signal activation related genes (*Tmem20*, anticorrelated with *Rcan1*). Finally, the transcriptomic and proteomic data were analyzed with principal component analysis and were found to correlate much better with cell size and doubling time than with productivity.

Table 3. Genes positively correlated with cell growth as identified in more than one study based on their biological role[†].

Gene ID	Protein name	Biological role[‡]	-omics technique; sampling phase; process mode	Cell line	Product	Ref.
Hnrnpa2b[§]	Heterogeneous nuclear ribonucleoprotein A2/B1	RNA processing; RNA localization	Translatome (Microarray of different RNA pools), Transcriptomics (Microarray), miRNA (Microarray), Proteomics (LC–MS/MS); exponential; batch	CHO-DG44 and CHO-K1	mAb	[15,16]
Hnrnpc[¶]	Heterogeneous nuclear ribonucleoprotein C	RNA processing	Translatome (Microarray of different RNA pools), Transcriptomics (Microarray), miRNA (Microarray), Proteomics (LC–MS/MS); exponential; batch	CHO-DG44 and CHO-K1	mAb	[15,16]
Hspd1[#]	60 kDa heat shock protein, mitochondrial	Response to unfolded protein; Protein refolding; Mitochondrial protein import and macromolecular assembly	Transcriptomics (Microarray), miRNA (Microarray), Proteomics (LC–MS/MS and 2D DIGE) exponential; batch	CHO-K1 and CHO-DUKX	mAb and nonproducing (CHO-DUKX)	[16,17]
Mcm5	Minichromosome maintenance deficient 5, cell division cycle	Cell cycle; DNA metabolic processes/DNA repair	Translatome (Microarray of different RNA pools), Transcriptomics (Microarray and Coexpression analysis); multiple phases; multiple modes	CHO-DG44 and CHO-DUX and CHO-K1	mAb, fusion protein, growth factors, coagulation factors, nonproducing	[14,15]
Pcna	Proliferating cell nuclear antigen	Nucleic transport of proteins; DNA metabolic processes/DNA repair; Cell cycle	Translatome (Microarray of different RNA pools) and Transcriptomics (Microarray and Coexpression analysis) [multiple phases; multiple modes]	CHO-DG44 and CHO-DUX and CHO-K1	mAb, fusion protein, growth factors, coagulation factors, nonproducing	[14,15]
Vcp	Valosin containing protein	Nucleic transport of proteins; ER to Golgi vesicle-mediated transport; Cellular response to DNA damage stimulus	Translatome (Microarray of different RNA pools) and Transcriptomics (Microarray) and Proteomics (LC–MS/MS and 2D DIGE) [exponential, batch]	CHO-DG44 and CHO-DUX and CHO-K1	mAb and nonproducing (CHO-DUKX)	[15,17]

Note that cell lines CHO-DUXB-11, CHO-DUX and CHO-DUKX are the same and originate by the mutagenesis of CHO-K1 to remove DHFR activity.
DIGE: Differential in-gel electrophoresis; CHO: Chinese hamster ovary; ER: Endoplasmic reticulum; mAb: Monoclonal antibodies.
[†]Part of this review is also presented in Courtes ***et al.*** [15].
[‡]Biological role was cross-validated using UniprotKB database [13].
[§]Associated miRNAs from Clarke ***et al.*** [16]: miR-190b, miR-204, miR-497.
[¶]Associated miRNAs from Clarke ***et al.*** [16]: miR-30e, miR-455–5p.
[#]Associated miRNAs from Clarke ***et al.*** [16]: miR-206.

Driven by the importance of the mTOR pathway, Edros *et al.* [10] used a pathway-specific PCR array to quantify the levels of associated genes in two mAb-producing CHO cell lines with a 17.4-fold difference in q_p. Eight genes (*Pik3cd*, *Pik3cg*, *Pld1*, *Rragc*, *Ins2*, *Telo2*, *Rps6*, *Prkab1*) were found to be differentially expressed out of the total of 84 genes associated with the mTOR pathway that were tested. The transcript (*Pik3cd*) that presented the greatest upregulation (more than 70-fold), which encodes for the ρ110δ protein, was also confirmed at the protein level using western blot. Tests on additional cell lines revealed that the ρ110δ protein presented a Pearson correlation coefficient of 0.835 with the q_p of six different mAb-producing CHO cell lines with q_p ranging from 3 to 51 pg/cell/day.

Clarke *et al.* [14] performed coexpression analysis of microarray data resulting from a wide range of processes (fed-batch with different media, temperature and feeds), products (mAb and fc fusion protein) and 167 cell lines CHO (CHO-DXB11 and CHO-K1) cell lines with q_p ranging from 0.52 to 55.39 pg/cell/day. Such a study adds one more level of information to the typical microarray data analysis, as correlation between the different transcripts was also employed independently of phenotype (in this case q_p). Coexpression analysis identified 70 genes to be positively correlated with productivity, whereas about 110 were negatively correlated. Positively correlated genes included lipid metabolism (*Lpl*) and intracellular protein transport (*Sec22b*, *Vamp3*) related processes. Genes with negative correlation with productivity included cholesterol biosynthesis (*Fdps*, *Mvd*, *Cyp51a1*, *Dhcr7*, *Hmgcs1*) and fatty acid metabolism (*Acat2*) related processes.

Kyriakopoulos *et al.* [18] used qRTPCR to analyze the expression of 40 amino acid transporters in the parental and two mAb-producing GS-CHO (CHO-K1 derived) cell lines with a twofold difference in q_p at exponential and stationary phase of batch culture. They observed that genes *slc1a2* (acidic amino acid transport) and *slc43a2* (branched chain amino acid transport) were positively associated with productivity as were found to be constantly higher in producer cell lines than in parental. Gene *slc6a6*, which encodes for a taurine transporter, was found to be differentially expressed and positively correlated with productivity during exponential phase of batch culture. The overexpression of the latter gene has been found to enhance productivity in a previous study [19].

Edros *et al.* [20] used qRTPCR and flow cytometry to analyze a panel of six mAb-producing CHO cell lines with q_p ranging from 3 to 51 pg/cell/day. They identified that heavy and light mAb chain mRNA, heavy chain polypetides and intracellular mAb presented the greatest correlation with q_p. Growth rate, biomass, ER content and light chain mAb polypeptide were also correlated with q_p, but not for all time points tested. Also molecular markers, such as total protein content, cell size and Golgi apparatus content were not found to be correlated with productivity.

Davies *et al.* [21] performed a large-scale study assessing functional heterogeneity and heritability of 100 cell line clones. The selection process involved selecting for cell lines that presented the least variation in growth rates during extensive subculture. These cell lines were then transiently transfected with a vector encoding for both mAb and green fluorescent protein (GFP), utilizing two different transfection methods: chemical treatment with lipofectamine and electroporation. Differences were found when comparing titers of the different clones and of the two products. MAb titer variation between the clones, for example, was much less than GFP titer variation. This was attributed to the fact that GFP is a nonmammalian protein. MAb and GFP expression was found to be better correlated for the cells transfected by lipofection suggesting that the transfection method also plays an important role in observed product titer variations. Transient mAb expression of individual clones was not found to be heritable, however clones were identified that presented differences in growth rate, endocytotic transfectability and N-glycan processing after multiple subcultures. In conclusion, the results in [21] show the broad diversity of subclones during selection and extensive cultivation attributed to the dynamic nature of their functional genome.

Finally, O'Callaghan *et al.* [22] performed a transcriptomic and proteomic analysis of heavy and light chains in seven mAb-producing CHO cells varying over 350-fold in q_p. Their findings suggest that heavy chain controlled mAb production, as light chain mRNA levels were in excess in all cell lines tested. There was also a positive correlation between q_p and heavy chain mRNA transcription and translation rates. Overall, a variable expression of heavy and light chain between different cell lines was observed, leading to the conclusion that cell line engineering should be cell line specific.

Metabolic traits of high producers

Comparative studies of cell metabolism have been carried out, yielding a higher degree of converging findings in terms of characteristics of high producers compared with transcriptomic and proteomic studies. Specifically, Ghorbaniaghdam *et al.* [23] used a cumate-inducible vector for expressing an mAb product in CHO cells in batch culture and performed metabolic analysis of resulting q_p phenotypes. A kinetic metabolic model was also employed to assess metabolic regulation. Their findings suggest that the high producer cell line presented a higher global flux in TCA cycle, result-

Table 4. Genes negatively correlated with cell growth classified according to their biological role.

Gene ID	Protein name	Biological role†	-omics technique; sampling phase; process mode	Cell line	Product	Ref.
Actb	α-actin	Various types of cell motility	Transcriptomics (Microarray), Proteomics (LC–MS/MS, 2D DIGE); exponential; batch	CHO-K1 and CHO-DUKX	mAb (CHO-K1); nonproducing (CHO-DUKX)	[17]
Actg1‡	Actin, cytoplasmic 2	Various types of cell motility	Transcriptomics (Microarray), miRNA (Microarray), Proteomics (LC–MS/MS); exponential; batch	CHO-K1 (ten clones varying in growth rate from 0.01 to 0.044 h^{-1})	mAb	[16]
Man2a1§	α-mannosidase 2	N-glycan processing; Mannose metabolic process	Transcriptomics (Microarray), miRNA (Microarray), Proteomics (LC–MS/MS); exponential; batch	CHO-K1 (ten clones varying in growth rate from 0.01 to 0.044 h^{-1})	mAb	[16]
Uap1	UDP-N-acetyl-glucosamine pyro-phosphorylase 1	UDP-N-acetylglucosamine biosynthetic process	Transcriptomics (Microarray), Proteomics (LC–MS/MS, 2D DIGE); exponential; batch	CHO-K1 and CHO-DUKX	mAb (CHO-K1); nonproducing (CHO-DUKX)	[17]
Txndc4	Thioredoxin domain containing 4 (endoplasmic reticulum)	Cell redox homeostasis	Transcriptomics (Microarray, Coexpression analysis), multiple phases; multiple modes	CHO-DUX and CHO-K1; 167 cell lines (growth rate ranging from -0.0058 to 0.0467 h^{-1})	mAb, fusion protein, growth factors, coagulation factors, nonproducing	[14]
Txndc5¶	Thioredoxin domain-containing protein 5	Cell redox homeostasis	Transcriptomics (Microarray), miRNA (Microarray), Proteomics (LC–MS/MS); exponential; batch	CHO-K1 (ten clones varying in growth rate from 0.01 to 0.044 h^{-1})	mAb	[16]
Uqcrb#	Cytochrome b-c1 complex subunit 7	Mitochondrial electron transport, ubiquinol to cytochrome c	Transcriptomics (Microarray), miRNA (Microarray), Proteomics (LC–MS/MS); exponential; batch	CHO-K1 (ten clones varying in growth rate from 0.01 to 0.044 h^{-1})	mAb	[16]
Idh3a	Isocitrate dehydrogenase 3 (NADþ)-α	Tricarboxylic acid cycle	Transcriptomics (Microarray) and Proteomics (LC–MS/MS, 2D DIGE); exponential; batch	CHO-K1 and CHO-DUKX	mAb (CHO-K1); nonproducing (CHO-DUKX)	[17]

Note that cell lines CHO-DUXB-11, CHO-DUX and CHO-DUKX are the same and originate by the mutagenesis of CHO-K1 to remove DHFR activity.
CHO: Chinese hamster ovary; mAb: Monoclonal antibodies.
†Biological role was cross-validated using UniprotKB database [13].
‡Associated miRNAs from Clarke ***et al.*** [16]: miR-10a.
§Associated miRNAs from Clarke ***et al.*** [16]: miR-92a.
¶Associated miRNAs from Clarke ***et al.*** [16]: miR-636.
#Associated miRNAs from Clarke ***et al.*** [16]: miR-18a, miR-18b, miR-92a.

Key Term

Principal component analysis: Statistical procedure used to find patterns in data of high dimension.

ing in better redox and increased metabolic efficiency both before and during induction.

Kyriakopoulos *et al.* [18] performed a time-course analysis of extracellular and intracellular amino acid concentrations during the batch culture of a parental and two mAb-producing GS-CHO cell lines (varying twofold in q_p). Specific extracellular rates of amino acid, lactate and ammonia were not found to present any specific trends in the producer cell lines. Along with the transcriptomic data for amino acid transporters presented in the same study which showed that none of the proteinogenic amino acid transporters were differentially expressed between the producer cell lines, the researchers suggested that transport of amino acids is not the limiting step for recombinant protein formation.

Chong *et al.* [24] performed metabolomic analysis (LC–MS) of different mAb-producing CHO cell lines (CHO-DG44 derived). The researchers identified that oxidized (GSSG) and reduced (GSH) glutathione, FAD, NADH, UDP-Glc or UDP-Gal (could not be distinguished with their method), UDP-GlcNAc or UDP-GalNAc (could not be distinguished with their method) pools were increased for high producers. The results suggest that high producing cell lines regulate their redox status (glutathione, FAD, NADH) better than lower producers. In general, it appears that high producing cell lines present an increased capacity for oxidative metabolism and, consequently, more efficient energy utilization mechanisms. This was also recently observed in a study where maximum mAb production along the course of a typical fed-batch culture was found to be correlated with increased oxidative metabolism [25].

Transcriptomic & proteomic traits of fast growers

Given the synergistic effect of specific productivity and cell growth toward increasing recombinant protein titers, it is also interesting to review recent studies that identify genes associated with high specific growth rates. Translatome analysis was performed for the first time by Courtes *et al.* [15] to identify the traits of a mAb-producing CHO (CHO-DG44 derived) cell line during exponential phase of batch culture. The technique involves microarray analysis of different pools of mRNA that were separated on sucrose gradient. Results were reported by taking into account the resulting pools that correspond to ribosome loading, in other words, RNA bound to polysomes and to monosomes. Overall, 43 genes were found to be well expressed at exponential phase. Highly expressed genes were associated with RNA processing and RNA localization (*Hnrnpc* and *Hnrnpa2b1* encoding for heterogeneous nuclear ribonucleoproteins), cell-cycle regulation (*Prc1* encoding for the protein regulator of cytokinesis 1), carbohydrate metabolic processes and cofactor metabolism (*G6pdh* encoding for glucose-6-phosphate dehydrogenase), RNA processing and ribosome biogenesis (*Utp6* encoding for UTP6 small subunit processome) and chromosomes organization and modification, DNA metabolic processes and DNA repair (*Ruvbl1* encoding for RuvB-like protein1). The researchers concluded that transcript level and transcriptional efficiency are not correlated for 95% of the genes, a finding that highlights the importance of transcriptional control mechanisms in CHO cells.

In the aforementioned study by Carlage *et al.* [6], in which proteomic analysis of CHO (CHO-DG44 derived) cells expressing a recombinant fusion protein was performed, certain classes of proteins were also found to be correlated with cell growth. Specifically, eukaryotic translation initiation factor 3 (*Eif3*) and the 40S ribosomal protein S6 (*Hspa6*), a molecular chaperone associated with ER stress and unfolded protein response (BiP) and cell-cycle regulation proteins (RACK1, GTP-binding nuclear protein Ran, calcyclin) were all found to be positively correlated with growth. Proteins including ribosomal intermediate filament proteins (vimentin, annexin, histone H1.2, histone H2A) and a growth inhibitor (galectin-1) were negatively correlated with growth.

In Clarke *et al.* [16] three -omics levels, transcriptomics including miRNA analysis and proteomics, were used to study ten mAb-producing CHO (CHO-K1 derived) cell lines varying in growth rate from 0.01 to 0.044 h^{-1}. Such an analysis has the advantage of allowing for cross-validation of the target genes and proteins identified as differentially expressed with the regulation of the miRNA levels. The analysis focused on the exponential phase of batch culture. Forty-one genes and proteins (29 positively correlated and 12 negatively correlated with growth) were found to be differentially expressed and were also verified with the miRNA profiling data (22 miRNA differentially upregulated and 14 differentially downregulated with growth). The genes and proteins that best correlated with high specific growth rate were involved in mRNA processing and protein synthesis. The same transcriptomic data (not including the miRNA and proteomic analyses) were also analyzed in another study of the same group [26]. Therein, cell cycle and translation genes were found to be upregulated for cell lines presenting high growth rate, while genes associ-

ated with cellular homeostasis were downregulated. Overall, 416 transcripts (190 upregulated; 226 downregulated) were found to be differentially expressed across the cell lines tested.

The same group have also performed a large-scale microarray and coexpression analysis of cell lines with different growth rates [14]. This study was also mentioned in the *transcriptomic and proteomic traits of high producers* part as the researchers performed correlation analysis for both growth and productivity. In total, 167 cell lines with specific growth rates ranging from -0.0058 to 0.0467 h^{-1} were studied which allowed them to report results for several desired traits of production cell lines. Genes that were positively correlated with growth were involved in cell cycle and DNA replication (*Mcm7*, *Mcm5*, *Mcm4*, *Cdc20*, *Mad2l1*, *Nolc1*, *Pcna*, *Npm1*, *Ran*) and folate metabolism (*Atic*, *Shmt2*, *Gart*, *Mthfd2*). Genes that were negatively correlated with growth were involved in lipid metabolism (*Pgs1*, *Gpaa1*, *Pigt*) and vesicle-mediated transport (*Txndc4*, *Cope*, *Copb2*, *Tmed10*).

Finally, one more study by the same group [17] involved the use of transcriptomics and proteomics to profile fast versus slow growing mAb-producing CHO (CHO-K1 derived) cell lines as well as fast versus slow growing CHO-DXB11 cells not expressing a recombinant product. Overall 118 gene transcripts and 58 proteins were found to be differentially expressed when comparing slow and fast growers. Genes found to be positively correlated with growth were associated with the unfolded protein response (*Hspb1*, *Hspd1*) and ER to Golgi vesicle mediated transport (*Vcp*). In contrast, genes involved in cell motility (*Actb*, *Lasp1*), glycolysis (*Eno1*), nucleotide sugar biosynthesis (*Uap1*), TCA cycle (*Idh3a*) and glutathione biosynthesis (*Gss*) were found to be negatively correlated with growth.

Metabolic traits of fast growers

Only one study has focused on performing metabolomic analysis of CHO cells with different growth characteristics so far. Dietmair *et al.* [27] analyzed the performance of a CHO cell line expressing a recombinant growth hormone growing in three different media and hence presenting three different growth rates. Intracellular samples were analyzed using HPLC and GC-MS in exponential and stationary phase of batch culture. Metabolites such as dCTP, CTP, ATP, GTP, NAD, UDP-GlcA, glutamine, glutamate and thymine were found to be in higher levels intracellularly for the faster growing cells, while UDP-Glc was lower. During stationary phase, the researchers identified that CTP was the main metabolite limiting cell proliferation rate and hypothesized that its depletion was the reason for cells entering that phase during batch culture.

Key Term

Translatome: Polysome-associated mRNAs, which serve as a measure of translational efficiency and form the background of the proteome.

Genetic engineering strategies for CHO cells

Current genetic engineering strategies for CHO cells include mainly gene overexpression, with unfolded protein response (UPR) markers, for example, spliced form of X-box-binding protein 1 (*Xbp1s*), having been widely targeted [11,28–32]. Other UPR markers include *Ppib*, which encodes for cyclophilin b, an ER residence protein, in [11], co-overexpression of *Hspa5*, which encodes for BiP, *Atf6c*, which encodes for cleaved activating transcription factor 6, a UPR transactivator and *Xbp1s* in the same study [11] and co-overexpression of ERO1-Lα and *Xbp1s* in [32]. Additionally, the gene encoding for growth arrest and DNA damage inducible protein GADD34 was targeted in [33], while Ohya *et al.* [34] overexpressed the gene encoding for ATF4 and Mohan and Lee [35] overexpressed calnexin.

Overexpression strategies have also targeted anti-apoptotic proteins, for example, the caspase-inhibitor XIAP in [29] and co-overexpression of E1B-19K (BCL2/adenovirus E1B 19 kDa protein-interacting protein 3-like) and *Aven*, a cell death regulator, in [36]. Genes facilitating intracellular protein transport have also been used as genetic engineering targets, for example, the gene encoding for ceramide transfer protein (CERT) in [37] and [38], or the genes encoding for Sec1 family domain-containing protein 1 (SLY1) and Munc18c (mammalian uncoordinated) that were successfully co-overexpressed in Peng and Fussenegger [30]. In a previous study, Fussenegger *et al.* attempted to arrest cell proliferation in high cell density cultures by coexpression of p21 and the differentiation factor CCAAT/enhancer-binding protein α using an inducible vector [39]. The resulting production of recombinant secreted alkaline phosphatase was increased 10–15-times on a per cell basis, which was further increased by coexpressing a third gene, bcl-xL, to prevent the onset of apoptosis. A few studies have also successfully overexpressed cell membrane transporters, for example, the *slc6a6* taurine transporter in [19] and the *Glut5* gene encoding for a fructose transporter in [40]. Overexpression of miRNA targets arises as a more recent trend and has been shown to result in high q_p phenotypes, for example, miR-17 overexpression in [41], and miR-557 co-overexpression with miR-1287 in [42].

Gene silencing or knock-out strategies have also been successful. As such, miRNA targets have been employed, for example, miR-7 in [43] and mmu-miR-466h-5p in [44]. Other targets include the glutamine

synthetase (GS) gene knockouts in [45] and genes *Ldha*, encoding for lactate dehydrogenase A catalyzing the pyruvate to lactate reaction, and *Pdhk*s, encoding for pyruvate dehydrogenase kinases 1, 2 and 3 that catalyze the suppression of the enzyme catalyzing the reaction from pyruvate to acetyl-CoA, associated with pyruvate metabolism and lactate production. Apoptotic genes have also been targeted successfully in silencing studies, for example, genes encoding for apoptosis regulator BAX and Bak (Bcl-2 homologous antagonist/killer) in [46] and [47]. A list of recent genetic engineering strategies can be found in Supplementary File 3.

Conclusions & future perspective

A review of recent –omics studies comparing CHO cell lines with different characteristics has identified a small subset of genes that are correlated either positively or negatively with recombinant protein productivity and cell growth. These are mainly related to oxidative metabolism, the cell cycle, transcription and translation regulation, protein processing and the unfolded protein response with respect to recombinant protein productivity. This points to the need of engineering greater protein synthesis and processing capacity in professional secretors. Intervention could involve a variety of targets, which have, however, yielded variable results when evaluated with different protein products [11] as also reviewed in [48]. In parallel, genes related to the cell cycle, DNA repair, RNA processing and mitochondrial function have been found to predominantly correlate with cell growth. However, there are no genes consistently being identified as positive or negative indicators of growth or productivity across cell lines and recombinant products. In fact, the clusters of genes identified by two or more studies of different CHO cell lines producing different mAb and non-mAb products is substantially smaller than the list of genes that have been identified as being differentially expressed at the mRNA, or protein level in individual studies. Interestingly, some studies demonstrate contradicting results, for example, overexpression of *Xbp1s* does not always result in increased q_p phenotypes [34,37]. Similarly, ribosomal proteins have been found to be both positively and negatively correlated with productivity in different studies. Overall, the UPR, apoptosis, cell cycle and lactate metabolism appear to be the pathways that are most frequently targeted in genetic engineering attempts, usually with positive outcomes. Although there are studies targeting more than one gene [11,39], these usually target a single pathway. Additional targets identified in –omics studies involve the overexpression of the *Rps6* gene (involved in protein synthesis and mTor pathway, see Table 1) and the overexpression of the six genes (*Hnrnpa2b1*, *Hnrnpc*, *Hspd1*, *Mcm5*, *Pcna*, *Vcp*) found to be associated with good growth characteristics (Table 3). Our analysis does not identify any obvious silencing targets that have been reported in more than one study (Tables 2 & 4).

These observations point to the fact that genes correlated with cell growth and recombinant protein productivity are largely cell line dependent and that findings relating to cellular events other than metabolism could be the result of clonal variation rather than a global indicator of performance. This is exemplified in [21], which showed that the breadth of clonal variation in terms of recombinant protein titer varies depending on the product itself. The argument is reinforced by a recent study on difficult-to-express proteins, in which the same host was used to transiently express a panel of eight mAbs that reached low process titers [11]. The researchers examined the effect of coexpressing a variety of molecular chaperones involved in the unfolded protein response and their results showed that the effectiveness of each chaperone coexpression varied depending on product. They therefore concluded that the optimal strategy is different for each product, despite using the same host and working with the same class of difficult-to-express proteins. The same group has further suggested that predicting the expression levels of different proteins requires knowledge of the product sequence and could be aided by computational tools [11,13].

Additional complications arise from the inherent genetic variability of different CHO cell hosts [49]. CHO cells have been the workhorse of the biopharmaceutical industry thanks to their amenability to genetic modification. Over the decades they have been adapted to different processing environments (suspension, serum-free cultivation, proprietary media and process conditions) and have undoubtedly gone through an evolution process that results in the documented genetic diversity of hosts existing in academic and industrial laboratories [50].

An additional challenge is the lack of commercially available standardized tools for –omics analyses specific to CHO cells. The bioprocessing community would benefit from a universal CHO cell DNA microarray or sequencing method that would standardize analyses performed in different laboratories and ensure transferability of research methodologies. To establish such tools, it is necessary to have a well-defined reference genome that is relevant to all cell lines. This would also help identify any mutations with adverse effects, which would point to the need of re-engineering these functions into the specific cell lines. Standardization of data analysis methods

would also help ensure comparability of studies and results across different hosts and studies, even though it would be difficult to carry out microarray analysis for each cell line development campaign due to the large number of cell lines involved. Finally, transcriptomic results arising from such analyses require validation at the proteomic level on an individual cell line basis. This prompts the use of multiomics approaches for obtaining a global view of strategies for attaining higher levels of recombinant protein production, as also advocated by [15,16], or aiding the production of difficult-to-express proteins.

Despite the disparity of current '-omics' characterization efforts and genetic engineering strategies, knowledge of differentially expressed genes has already been demonstrated to benefit genetic engineering approaches. A prominent example is that of the *slc6a6* taurine transporter, which was shown to be highly expressed at the transcriptomic level during late stage culture, and was subsequently overexpressed resulting in prolonged viability and a 47% increase in mAb titer [19]. The question then becomes how best to assess the effect of different potential strategies if the outcome is cell line- and product-dependent. Once targets for overexpression have been identified through –omics studies, it is possible to use the transient expression set-up put forward in [11] to identify the most suitable methodology for the particular experimental system. The advantage of transient gene expression is that it is fast and amenable to high-throughput studies. The final outcome can be a tailored cell line co-overexpressing the most promising target stably or even transiently to reduce the metabolic burden.

Contrary to the variable results of transcriptomic and proteomic studies, there appears to be wider consensus in the literature with regards to the metabolic traits of high producers compared with their transcriptomic and proteomic characteristics. Given the established methodologies for high-throughput automated assays, for example, FACS, ClonePix™ (Molecular Devices, CA, USA), Cell Xpress™ (Cyntellect Inc., CA, USA) reviewed in [51], metabolic profiling offers a promising and robust strategy both for selection and cell engineering. This could involve, for example, a process selecting for increased glutathione pools or other redox-related metabolites (FAD, NADH), nucleotide sugars (UDP-GlcNAc or UDP-GalNAc) or nucleotides as described in [24] and [27]. Due to the complex nature of the system, it is expected that more than one metabolite would need to be screened. In terms of process engineering, time-course –omics data would need to be available for refining process conditions. In addition, these data would need to arise from protocols and materials that closely mimic the production process, after a suitable cell line has been identified. This point is also emphasized in a recent review by Dickson [52] who stresses the importance of aligning media and feed compositions to specific cell lines and recombinant products (particularly non-mAbs) by using metabolomics to achieve a molecular-level understanding of individual metabolic requirements.

An important metric that has not been widely considered in comparative studies carried out to date is that of protein product quality. Strategies that increase specific protein productivity or alter cell metabolism have the potential of changing the extent of post-translational modification, with the most obvious example being that of protein glycosylation (as reviewed in [53,54]). It is therefore imperative that cell line selection or engineering approaches are compared in terms of their effect on cell growth, specific productivity and product quality to arrive at the optimal strategy for each cell line and recombinant protein product.

Financial & competing interests disclosure

The authors have no relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript. This includes employment, consultancies, honoraria, stock ownership or options, expert testimony, grants or patents received or pending or royalties.

No writing assistance was utilized in the production of this manuscript.

Executive summary

- Correlation of gene expression with cell growth and recombinant protein productivity appears to be cell line dependent.
- Biological clusters of genes positively correlated with productivity include oxidative metabolism, the unfolded protein response, transcription and translation regulation.
- Biological clusters of genes positively correlated with cell growth include cell cycle, DNA repair, RNA processing and mitochondrial function.
- Multiomics approaches are required since transcript levels do not necessarily correlate with protein levels.
- Metabolic profiling has been more successful at identifying characteristics of high producers.

References

Papers of special note have been highlighted as:
• of interest; •• of considerable interest

1 Li F, Vijayasankaran N, Shen AY, Kiss R, Amanullah A. Cell culture processes for monoclonal antibody production. *MAbs* 2(5), 466–479 (2010).

2 Hacker DL, De Jesus M, Wurm FM. 25 years of recombinant proteins from reactor-grown cells - where do we go from here? *Biotechnol. Adv.* 27(6), 1023–1027 (2009).

3 Bandaranayake AD, Almo SC. Recent advances in mammalian protein production. *FEBS Lett.* 588(2), 253–260 (2014).

4 Porter AJ, Racher AJ, Preziosi R, Dickson AJ. Strategies for selecting recombinant CHO cell lines for cGMP manufacturing: improving the efficiency of cell line generation. *Biotechnol. Prog.* 26(5), 1455–1464 (2010).

5 Vishwanathan N, Le H, Jacob NM *et al.* Transcriptome dynamics of transgene amplification in Chinese hamster ovary cells. *Biotechnol. Bioeng.* 111(3), 518–528 (2014).

6 Carlage T, Hincapie M, Zang L *et al.* Proteomic profiling of a high-producing Chinese hamster ovary cell culture. *Anal. Chem.* 81(17), 7357–7362 (2009).

7 Clarke C, Doolan P, Barron N *et al.* Predicting cell-specific productivity from CHO gene expression. *J. Biotechnol.* 151(2), 159–165 (2011).

8 Doolan P, Barron N, Kinsella P *et al.* Microarray expression profiling identifies genes regulating sustained cell specific productivity (S-Qp) in CHO K1 production cell lines. *Biotechnol. J.* 7(4), 516–526 (2012).

9 Kang S, Ren D, Xiao G *et al.* Cell line profiling to improve monoclonal antibody production. *Biotechnol. Bioeng.* 111(4), 748–760 (2014).

10 Edros R, McDonnell S, Al-Rubeai M. The relationship between mTOR signalling pathway and recombinant antibody productivity in CHO cell lines. *BMC Biotechnol.* 14(1), 15 (2014).

11 Pybus LP, Dean G, West NR *et al.* Model-directed engineering of "difficult-to-express" monoclonal antibody production by Chinese hamster ovary cells. *Biotechnol. Bioeng.* 111(2), 372–385 (2013).

•• **The authors test a set of tools known to enhance productivity on difficult to express proteins and find that a tailored strategy is necessary for each product.**

12 Borth N, Mattanovich D, Kunert R, Katinger H. Effect of increased expression of protein disulfide isomerase and heavy chain binding protein on antibody secretion in a recombinant CHO cell line. *Biotechnol. Prog.* 21(1), 106–111 (2005).

13 Pybus LP, James DC, Dean G *et al.* Predicting the expression of recombinant monoclonal antibodies in Chinese hamster ovary cells based on sequence features of the CDR3 domain. *Biotechnol. Prog.* 30(1), 188–197 (2014).

14 Clarke C, Doolan P, Barron N *et al.* Large scale microarray profiling and coexpression network analysis of CHO cells identifies transcriptional modules associated with growth and productivity. *J. Biotechnol.* 155(3), 350–359 (2011).

15 Courtes FC, Lin J, Lim HL *et al.* Translatome analysis of CHO cells to identify key growth genes. *J. Biotechnol.* 167(3), 215–224 (2013).

• **First report of translatome analysis of CHO cells demonstrates a significant discrepancy between transcript level and transcriptional efficiency, prompting the need for multiomics approaches.**

16 Clarke C, Henry M, Doolan P *et al.* Integrated miRNA, mRNA and protein expression analysis reveals the role of post-transcriptional regulation in controlling CHO cell growth rate. *BMC Genomics* 13, 656 (2012).

•• **The authors present a comprehensive multiomics study that has identified markers for genetic engineering and selection strategies.**

17 Doolan P, Meleady P, Barron N *et al.* Microarray and proteomics expression profiling identifies several candidates, including the valosin-containing protein (VCP), involved in regulating high cellular growth rate in production CHO cell lines. *Biotechnol. Bioeng.* 106(1), 42–56 (2010).

18 Kyriakopoulos S, Polizzi KM, Kontoravdi C. Comparative analysis of amino acid metabolism and transport in CHO variants with different levels of productivity. *J. Biotechnol.* 168(4), 543–551 (2013).

19 Tabuchi H, Sugiyama T, Tanaka S, Tainaka S. Overexpression of taurine transporter in Chinese hamster ovary cells can enhance cell viability and product yield, while promoting glutamine consumption. *Biotechnol. Bioeng.* 107(6), 998–1003 (2010).

20 Edros RZ, McDonnell S, Al-Rubeai M. Using molecular markers to characterize productivity in Chinese hamster ovary cell lines. *PloS ONE* 8(10), e75935 (2013).

21 Davies SL, Lovelady CS, Grainger RK, Racher AJ, Young RJ, James DC. Functional heterogeneity and heritability in CHO cell populations. *Biotechnol. Bioeng.* 110(1), 260–274 (2013).

22 O'Callaghan PM, McLeod J, Pybus LP *et al.* Cell line-specific control of recombinant monoclonal antibody production by CHO cells. *Biotechnol. Bioeng.* 106(6), 938–951 (2010).

23 Ghorbaniaghdam A, Chen J, Henry O, Jolicoeur M. Analyzing clonal variation of monoclonal antibody-producing CHO cell lines using an in silico metabolomic platform. *PloS ONE* 9(3), e90832 (2014).

24 Chong WP, Thng SH, Hiu AP, Lee DY, Chan EC, Ho YS. LC-MS-based metabolic characterization of high monoclonal antibody-producing Chinese hamster ovary cells. *Biotechnol. Bioeng.* 109(12), 3103–3111 (2012).

• **Metabolomics study of CHO clones with varying levels of specific mAb productivity identified a small number metabolites involved in key cellular pathways that are highly correlated with specific mAb productivity.**

25 Templeton N, Dean J, Reddy P, Young JD. Peak antibody production is associated with increased oxidative metabolism in an industrially relevant fed-batch CHO cell culture. *Biotechnol. Bioeng.* 110(7), 2013–2024 (2013).

26 Doolan P, Clarke C, Kinsella P *et al.* Transcriptomic analysis of clonal growth rate variation during CHO cell line development. *J Biotechnol.* 166(3), 105–113 (2013).

27 Dietmair S, Hodson MP, Quek LE *et al.* Metabolite profiling of CHO cells with different growth characteristics. *Biotechnol. Bioeng.* 109(6), 1404–1414 (2012).

28 Gulis G, Simi KC, de Toledo RR, Maranhao AQ, Brigido MM. Optimization of heterologous protein production in Chinese hamster ovary cells under overexpression of spliced form of human X-box binding protein. *BMC Biotechnol.* 14(1), 26 (2014).

29 Becker E, Florin L, Pfizenmaier K, Kaufmann H. Evaluation of a combinatorial cell engineering approach to overcome apoptotic effects in XBP-1(s) expressing cells. *J. Biotechnol.* 146(4), 198–206 (2010).

30 Peng RW, Fussenegger M. Molecular engineering of exocytic vesicle traffic enhances the productivity of Chinese hamster ovary cells. *Biotechnol. Bioeng.* 102(4), 1170–1181 (2009).

31 Tigges M, Fussenegger M. Xbp1-based engineering of secretory capacity enhances the productivity of Chinese hamster ovary cells. *Metab. Eng.* 8(3), 264–272 (2006).

32 Cain K, Peters S, Hailu H *et al.* A CHO cell line engineered to express XBP1 and ERO1-Lalpha has increased levels of transient protein expression. *Biotechnol. Prog.* 29(3), 697–706 (2013).

33 Omasa T, Takami T, Ohya T *et al.* Overexpression of GADD34 enhances production of recombinant human antithrombin III in Chinese hamster ovary cells. *J. Biosci. Bioeng.* 106(6), 568–573 (2008).

34 Ohya T, Hayashi T, Kiyama E *et al.* Improved production of recombinant human antithrombin III in Chinese hamster ovary cells by ATF4 overexpression. *Biotechnol. Bioeng.* 100(2), 317–324 (2008).

35 Mohan C, Lee GM. Calnexin overexpression sensitizes recombinant CHO cells to apoptosis induced by sodium butyrate treatment. *Cell Stress Chaperon.* 14(1), 49–60 (2009).

36 Figueroa B Jr, Ailor E, Osborne D, Hardwick JM, Reff M, Betenbaugh MJ. Enhanced cell culture performance using inducible anti-apoptotic genes E1B-19K and Aven in the production of a monoclonal antibody with Chinese hamster ovary cells. *Biotechnol. Bioeng.* 97(4), 877–892 (2007).

37 Rahimpour A, Vaziri B, Moazzami R *et al.* Engineering the cellular protein secretory pathway for enhancement of recombinant tissue plasminogen activator expression in Chinese hamster ovary cells: effects of CERT and XBP1s genes. *J. Microbiol. Biotechnol.* 23(8), 1116–1122 (2013).

38 Florin L, Pegel A, Becker E, Hausser A, Olayioye MA, Kaufmann H. Heterologous expression of the lipid transfer protein CERT increases therapeutic protein productivity of mammalian cells. *J. Biotechnol.* 141(1–2), 84–90 (2009).

39 Fussenegger M, Schlatter S, Datwyler D, Mazur X, Bailey JE. Controlled proliferation by multigene metabolic engineering enhances the productivity of Chinese hamster ovary cells. *Nat. Biotechnol.* 16(5), 468–472 (1998).

•• **Multigene cell engineering approach increases recombinant protein titer 30-times.**

40 Le H, Vishwanathan N, Kantardjieff A *et al.* Dynamic gene expression for metabolic engineering of mammalian cells in culture. *Metab. Eng.* 20, 212–220 (2013).

41 Jadhav V, Hackl M, Klanert G *et al.* Stable overexpression of miR-17 enhances recombinant protein production of CHO cells. *J. Biotechnol.* 175(0), 38–44 (2014).

42 Strotbek M, Florin L, Koenitzer J *et al.* Stable microRNA expression enhances therapeutic antibody productivity of Chinese hamster ovary cells. *Metab. Eng.* 20(0), 157–166 (2013).

43 Sanchez N, Kelly P, Gallagher C *et al.* CHO cell culture longevity and recombinant protein yield are enhanced by depletion of miR-7 activity via sponge decoy vectors. *Biotechnol. J.* 9(3), 396–404 (2014).

44 Druz A, Son YJ, Betenbaugh M, Shiloach J. Stable inhibition of mmu-miR-466h-5p improves apoptosis resistance and protein production in CHO cells. *Metab. Eng.* 16, 87–94 (2013).

45 Fan L, Kadura I, Krebs LE, Hatfield CC, Shaw MM, Frye CC. Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnol. Bioeng.* 109(4), 1007–1015 (2012).

46 Macaraeg NF, Reilly DE, Wong AW. Use of an anti-apoptotic CHO cell line for transient gene expression. *Biotechnol. Prog.* 29(4), 1050–1058 (2013).

47 Lim SF, Chuan KH, Liu S *et al.* RNAi suppression of Bax and Bak enhances viability in fed-batch cultures of CHO cells. *Metab. Eng.* 8(6), 509–522 (2006).

48 Dietmair S, Nielsen LK, Timmins NE. Engineering a mammalian super producer. *J. Chem. Technol. Biotechnol.* 86(7), 905–914 (2011).

49 Wurm FM, Hacker D. First CHO genome. *Nat. Biotechnol.* 29(8), 718–720 (2011).

50 Lewis NE, Liu X, Li Y *et al.* Genomic landscapes of Chinese hamster ovary cell lines as revealed by the Cricetulus griseus draft genome. *Nat. Biotechnol.* 31(8), 759–765 (2013).

• **Genome sequencing of six Chinese hamster ovary (CHO) cell lines revealed a large number of polymorphisms and variations, including mutations in key pathways such as apoptosis.**

51 Lai T, Yang Y, Ng SK. Advances in mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals* 6(5), 579–603 (2013).

52 Dickson AJ. Enhancement of production of protein biopharmaceuticals by mammalian cell cultures: the metabolomics perspective. *Curr. Opin. Biotechnol.* 30(0), 73–79 (2014).

53 Jedrzejewski PM, Jimenez del Val I, Polizzi KM, Kontoravdi C. Applying quality by design to glycoprotein therapeutics: experimental and computational efforts of process control. *Pharm. Bioproc.* 1(1), 51–69 (2013).

54 del Val IJ, Kontoravdi C, Nagy JM. Towards the implementation of quality by design to the production of therapeutic monoclonal antibodies with desired glycosylation patterns. *Biotechnol. Prog.* 26(6), 1505–1527 (2010).