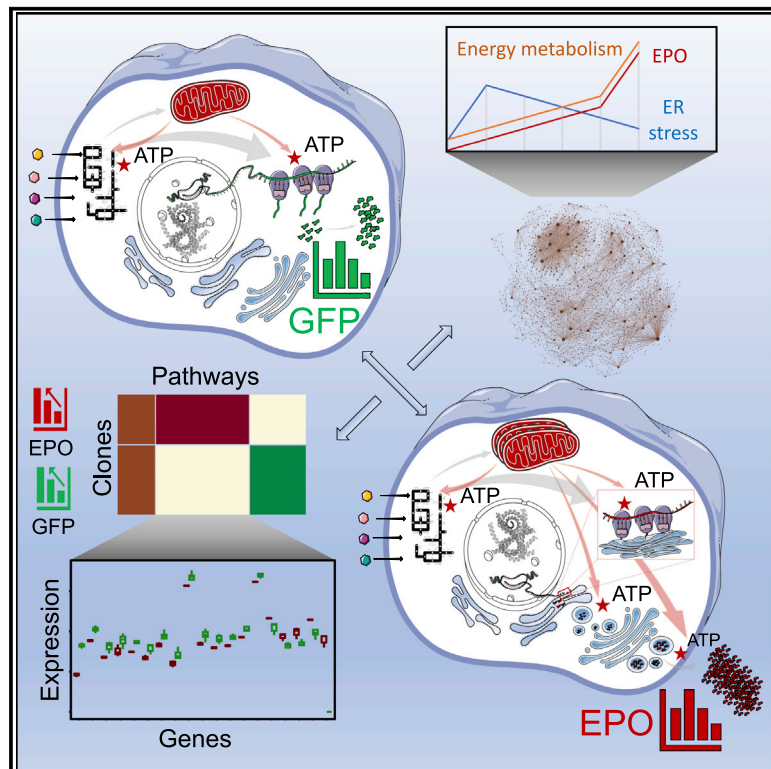# Enhanced metabolism and negative regulation of ER stress support higher erythropoietin production in HEK293 cells

## Graphical abstract



## Authors

Rasool Saghaleyni, Magdalena Malm, Noah Moruzzi, ..., Jens Nielsen, Jonathan L. Robinson, Johan Rockberg

## Correspondence

jonrob@chalmers.se (J.L.R.),
johan.rockberg@biotech.kth.se (J.R.)

## In brief

Finding bottlenecks in protein secretion is necessary for improvements in recombinant protein production. By investigating the transcriptomic profiles of different HEK293 clones, expressing either erythropoietin (EPO) or GFP, Saghaleyni et al. show that higher metabolic rates and negative regulation of ER stress are associated with increased EPO production.

## Highlights

- Transcriptomics analysis of HEK293 cells expressing GFP or EPO at varying rates

- Ribosomal genes show recombinant protein-specific patterns of expression

- Higher metabolic rates are observed in EPO-expressing cells compared with controls

- ATF6B facilitates higher EPO production through a moderated ER stress response

# Cell Reports

## Report

# Enhanced metabolism and negative regulation of ER stress support higher erythropoietin production in HEK293 cells

Rasool Saghaleyni,[1,9] Magdalena Malm,[2,9] Noah Moruzzi,[3] Jan Zrimec,[1] Ronia Razavi,[2] Num Wistbacka,[2] Hannes Thorell,[2] Anton Pintar,[2] Andreas Hober,[4] Fredrik Edfors,[4] Veronique Chotteau,[5] Per-Olof Berggren,[3] Luigi Grassi,[6] Aleksej Zelezniak,[1] Thomas Svensson,[1,7] Diane Hatton,[6] Jens Nielsen,[1,8] Jonathan L. Robinson,[1,7,*] and Johan Rockberg[2,10,*]

[1]Department of Biology and Biological Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden
[2]KTH - Royal Institute of Technology, School of Engineering Sciences in Chemistry, Biotechnology, and Health, Department of Protein Science, 106 91 Stockholm, Sweden
[3]The Rolf Luft Research Center for Diabetes and Endocrinology, Department of Molecular Medicine and Surgery, Karolinska Institute, 17176 Stockholm, Sweden
[4]Science for Life Laboratory, KTH - Royal Institute of Technology, 171 65 Solna, Sweden
[5]KTH - Royal Institute of Technology, School of Engineering Sciences in Chemistry, Biotechnology, and Health, Department of Industrial Biotechnology, 106 91 Stockholm, Sweden
[6]Cell Culture & Fermentation Sciences, BioPharmaceutical Development, BioPharmaceuticals R&D, AstraZeneca, Cambridge, UK
[7]Department of Biology and Biological Engineering, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Chalmers University of Technology, Kemivägen 10, 41258 Gothenburg, Sweden
[8]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
[9]These authors contributed equally
[10]Lead contact
*Correspondence: jonrob@chalmers.se (J.L.R.), johan.rockberg@biotech.kth.se (J.R.)
https://doi.org/10.1016/j.celrep.2022.110936

## SUMMARY

Recombinant protein production can cause severe stress on cellular metabolism, resulting in limited titer and product quality. To investigate cellular and metabolic characteristics associated with these limitations, we compare HEK293 clones producing either erythropoietin (EPO) (secretory) or GFP (non-secretory) protein at different rates. Transcriptomic and functional analyses indicate significantly higher metabolism and oxidative phosphorylation in EPO producers compared with parental and GFP cells. In addition, ribosomal genes exhibit specific expression patterns depending on the recombinant protein and the production rate. In a clone displaying a dramatically increased EPO secretion, we detect higher gene expression related to negative regulation of endoplasmic reticulum (ER) stress, including upregulation of ATF6B, which aids EPO production in a subset of clones by overexpression or small interfering RNA (siRNA) knockdown. Our results offer potential target pathways and genes for further development of the secretory power in mammalian cell factories.

## INTRODUCTION

The demand for greater efficiency and quality of protein production in biotechnology is rapidly increasing due to substantial advances in drug discovery (Tambuyzer et al., 2020) and the need for highly effective pharmaceutical proteins for the treatment of severe diseases, such as cancer (Kintzing et al., 2016). Chinese hamster ovary (CHO) cells are the current standard host for the production of a wide range of recombinant proteins partly due to the ability to generate similar post-translational modifications (PTMs) to those in humans, which is often a requirement for complex therapeutic proteins (Orellana et al., 2015; Meleady 2017; Davy et al. 2017). However, the PTM pattern from CHO cells is not identical to human PTMs (Dumont et al., 2016) and the

incompatibility with some types of proteins negatively affects drug efficacy, potency, or stability (Kuriakose et al. 2016; Goh and Ng 2017). Therefore, besides further development of the CHO cell line to meet the high-quality PTM requirements and increased yields (Datta et al. 2013; Koffas et al., 2018; Liang et al., 2020; Tejwani et al., 2018; Fouladiha et al., 2020; Wang et al., 2020), a lot of focus on improving hosts for biopharmaceutical production is on cell factories derived from human cells with the natural ability of generating human PTMs, such as the human embryonic kidney 293 (HEK293) cells (Almo and Love 2014; Malm et al., 2020; Tegel et al., 2020).

Although human-derived cell lines benefit from the ability to generate human PTMs, challenges still remain, such as increasing the protein production titer (Chin et al., 2019; Dietmair

et al., 2012; Mori et al., 2020) and creating a genetic engineering toolbox with specialized tools for human cells (Xu and Qi 2019). Recent publications have pursued some of these challenges, including the aim to increase the protein production and secretion power either by cell-line development approaches (Chin et al., 2019; Rahimpour et al., 2013) or cell culture process optimization (Schwarz et al., 2019), as well as to increase the quality of the secreted proteins by engineering folding and PTM pathways (Meuris et al., 2014; Del Val et al., 2016; Liang et al., 2020; Behrouz et al., 2020). However, despite the current knowledge of protein production and secretion in eukaryotic cells, there is still notable ambiguity in understanding and predicting the production and secretion rates as well as product quality under different conditions (Kafri et al., 2016; Liu et al. 2016). This is due to the complexity and presence of many specific biochemical steps across multiple cell organelles that orchestrate, as well as define, the rates of production and secretion of each protein (Kaufman and Popolo 2018; Kafri et al., 2016). Accordingly, the limited understanding of the biology behind the protein production process, combined with the continuously increasing demands on production quantity from industry and product quality from regulatory bodies, result in a high risk of production failure for many therapeutic proteins.

In the present study, we conducted a transcriptomic comparative analysis to capture physiological differences caused by protein production and secretion in HEK293F cells. In order to understand which differences are caused by protein production and which arise from the secretion-related processes, we generated two groups of cells producing either the erythropoietin (EPO) or the non-secretory protein GFP and compared each of these groups with each other and with their parental cell lines. EPO is naturally produced via the conventional secretory pathway mainly in kidney cells in response to hypoxia and stimulates red blood cell production in the bone marrow (Scholz et al., 1990). Recombinant human EPO (rhEPO) is an important treatment for chronic kidney disease and anemia (Santoro and Canova 2005). Moreover, EPO serves as a model recombinant protein with complex glycostructures and PTMs (Salgado et al., 2015). In our study, we identified genes whose expression correlated with recombinant EPO or GFP production and then explored the biological functions associated with these genes. Furthermore, we detected ribosomal genes with specific patterns of expression correlating with EPO and GFP production. Since the generation of single clones resulted in one clone with greatly increased EPO production titer—a 3-fold increase compared with the other clones—we set out to identify the reasons behind these improved protein titers, highlighting genes that can potentially facilitate increased protein production and secretion in future studies.
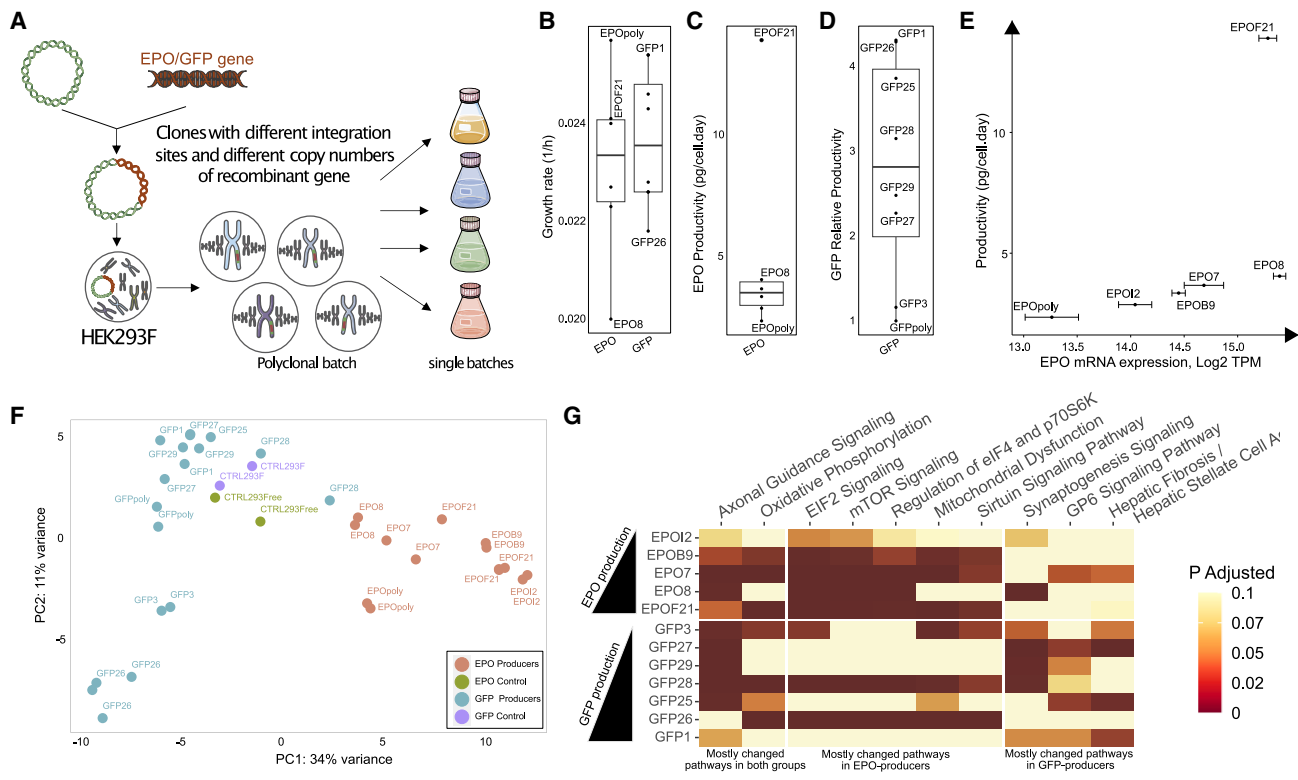
## RESULTS

### EPO and GFP producer clones have an altered metabolism compared with the host cell line

To investigate recombinant protein production in HEK293 cells, we transfected 293-F cell lines to generate stable clones producing either a secretory protein (EPO) or a non-secretory protein (GFP) (Figure 1A). Initially, polyclonal pools of cells pro-

ducing either EPO or GFP (EPOpoly and GFPpoly, respectively) were generated by random integration of plasmid DNA into the host genome resulting in collections of clones with various transgene integration sites and copy numbers. From these pools, five EPO-producing and seven GFP-producing clones were isolated. We observed that the growth rates of the EPO and GFP producers were lower than those of their respective host cell lines, but the decrease was no more than 22% and 14% compared with the hosts (Figure 1B). Recombinant protein productivity was, however, markedly different among clones and varied by almost 6-fold for the EPO producers and up to 4-fold for the GFP producers (Figures 1C and 1D). The most productive EPO clone (EPOF21) had a cell-specific productivity of 13.9 pg/cell/day, which was over 3-fold higher than the second highest producing clone EPO8 (4.05 pg/cell/day), and, interestingly, its EPO mRNA abundance was up to 20% lower than for EPO8 (Figures 1E and S1A). Except for this extraordinary production clone, we could measure significant correlation (Pearson's $r = 0.99$, $p = 0.001$) between EPO mRNA amount and the secreted EPO productivity for all clones (Figures 1E and S1B). There was a similarly high correlation observed between mRNA copy number and GFP productivity (Pearson's $r = 0.88$, $p = 0.004$; Figure S1C) for the GFP producer clones. For both EPO and GFP clones, no significant correlation was found between protein productivity and gene copy number or clone growth rate (Figures S1A–S1C).

We analyzed and compared transcriptomic data (Illumina HiSeq) to find how the protein producer cell lines differ from their respective parental host cell lines (Table S1). Principal component analysis (PCA) clustered clones in the first component based on their respective recombinant protein (EPO or GFP Figures 1F and S1D). We performed pairwise differential expression analysis between each recombinant protein producer clone and its respective host (Figures S1E–S1G; Table S1). Results of this analysis (Figures S1E and S1F) showed, while in the EPO producers EPOI2 had the highest number of differentially expressed (B.H. adj. $p < 0.05$, $|L2FC| > 1$) genes (1,137 genes upregulated and 82 genes downregulated), in the GFP producers the GFP26 clone was the most different compared with control cell line (487 genes downregulated, 466 genes upregulated). We also found 45 and 10 common differentially expressed genes in EPO and GFP producers with their control cell line, respectively (Figure S1G).

Pathway enrichment analysis using the Ingenuity Pathway Analysis (IPA) (Krämer et al., 2014) database (Figure 1G) showed that genes associated with axonal guidance signaling and oxidative phosphorylation exhibited consistent upregulation by most of the recombinant protein producer clones. Apart from these two pathways, eIF2 signaling and mTOR signaling pathways were significantly (B.H. adj. $p < 0.05$) altered across all EPO producers (Figure 1G). We observed different functional enrichment patterns between EPO and GFP clones and also clones in each group of EPO and GFP versus their parental cell line. However, axonal guidance signaling and oxidative phosphorylation exhibited a more similar pattern of change in comparison of producer clones with their corresponding parental cell (Figure 1G). Mitochondrial dysfunction, regulation of eIF4, p7056K signaling, and sirtuin signaling pathways were significantly (B.H. adj.

**Figure 1. Isolated EPO and GFP single clones showed differences in productivity and altered gene expression related to oxidative phosphorylation and axonal guidance signaling**
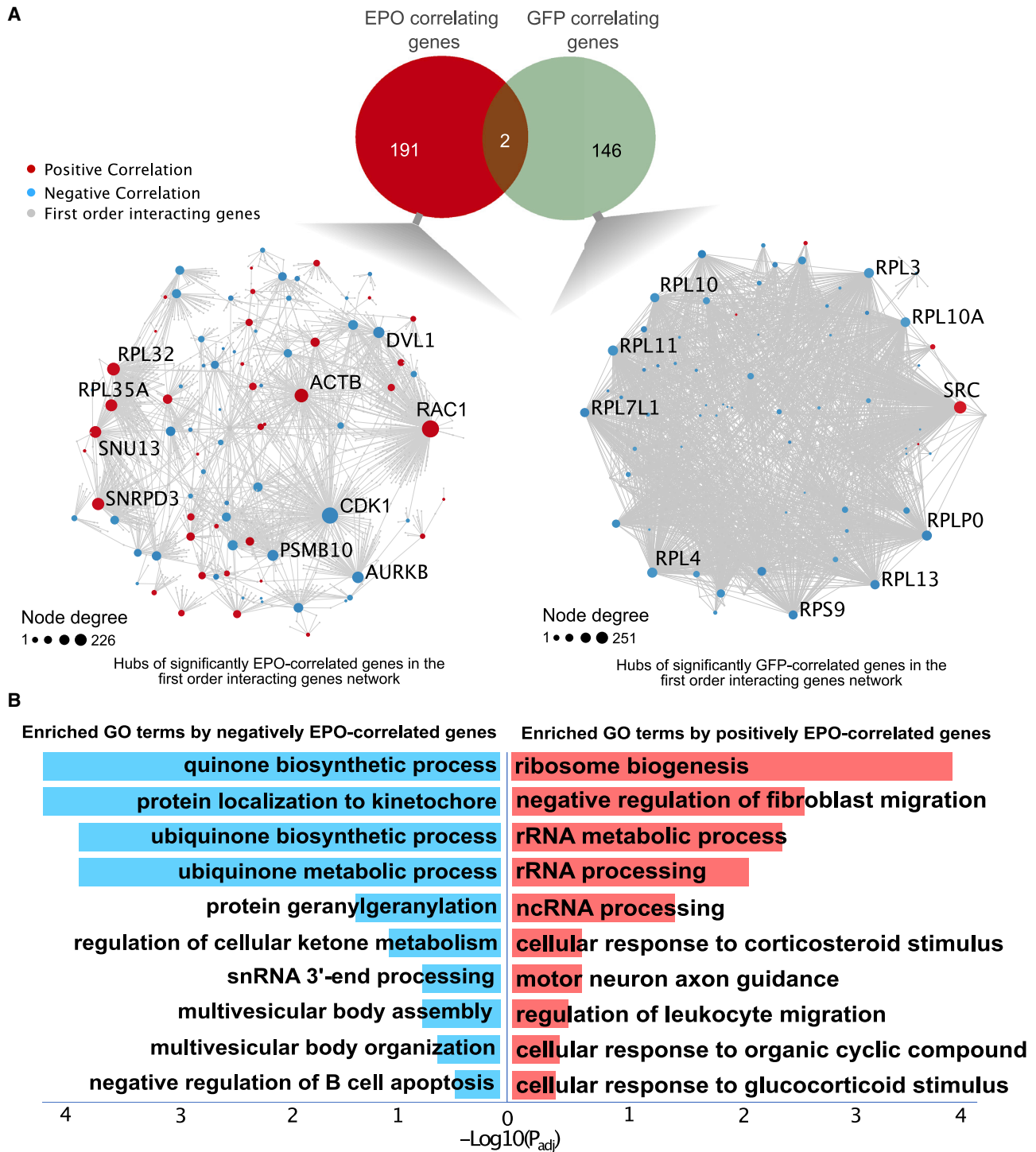
(A) Schematic diagram of the cell line development resulting in random integration and copy number of the EPO and GFP genes across the cell genomes.

(B) Boxplot of growth rates of EPO and GFP producer clones.

(C) Boxplot of specific productivity of EPO in different HEK293 clones.

(D) Scatter plot of relative productivity of GFP in the different clones in comparison to the polyclonal batch of the GFP clones (GFPpoly).

(E) Specific EPO productivity versus its mRNA expression in duplicate or triplicate in different clones.

(F) Principal component 1 in PCA analysis RNA sequencing (RNA-seq) data generated in duplicate and triplicate separates producer clones based on their recombinant EPO or GFP.

(G) Most significantly (B.H. adj. p < 0.05) enriched pathways in pairwise comparison of EPO or GFP producers against the control hosts.

p < 0.05) altered in more than three out of five EPO producers. On the other hand, gene enrichment analysis of GFP producers against their parental 293-F cell line did not reveal any common pathways enriched across all the cell lines. However, synaptogenesis signaling and GP6 signaling pathways indicated a significant change (B.H. adj. p < 0.05) in at least four out of seven of the GFP producers (Figure 1G). IPA indicated some of the significantly upregulated (B.H. adj. p < 0.05) genes related to translational and post-translational pathways were enriched in eIF2 and mTOR signaling pathways in EPO producers compared with the control cell line (Figures 1G and S2). An increase in the expression of genes associated with translation was predicted to activate downstream processes, including protein folding, ER stress, and apoptosis, as well as upstream processes, such as amino acid biosynthesis (Figure S2).

## Expression of ribosomal genes differs between EPO and GFP producers

We sought to find genes that significantly correlated with EPO and GFP production and investigate their roles in the process of recombinant protein production. For this purpose, we first extracted genes with an average transcript per million (TPM) above 10 across all cell lines in each group of EPO and GFP clones and then considered positively and negatively correlated (|Spearman r| > 0.5, p < 0.05) genes. Altogether, we found 67 and 126 genes with positive and negative correlation with EPO production and 18 and 130 genes revealed positive and negative correlation with GFP production (Figure 2A; Table S2). To find the major regulators among the genes highly correlated (mean TPM > 10, | Spearman r| > 0.5, p < 0.05) with either EPO or GFP productivity, we generated interaction networks between the genes and their first-order interacting partners based on experimental evidence (confidence score >900) extracted from the STRING database (Szklarczyk et al., 2019) (Figures 2A and Table S2). We excluded those interacting genes that are not expressed in our dataset or have very low expression (mean TPM < 10) and then ranked the genes based on their node degree (k), which measures the interactivity of each gene based on the number of observed interacting gene partners. The k of the top 10 most interactive genes (Figure 2A: network hubs) ranged from 61 to 226 and 94 to 251

**A**



EPO correlating genes

GFP correlating genes

● Positive Correlation
● Negative Correlation
● First order interacting genes

191 | 2 | 146

Node degree
1 ● ● ● 226

Hubs of significantly EPO-correlated genes in the
first order interacting genes network

Node degree
1 ● ● ● 251

Hubs of significantly GFP-correlated genes in the
first order interacting genes network

**B**

Enriched GO terms by negatively EPO-correlated genes | Enriched GO terms by positively EPO-correlated genes



quinone biosynthetic process
protein localization to kinetochore
ubiquinone biosynthetic process
ubiquinone metabolic process
protein geranylgeranylation
regulation of cellular ketone metabolism
snRNA 3'-end processing
multivesicular body assembly
multivesicular body organization
negative regulation of B cell apoptosis

ribosome biogenesis
negative regulation of fibroblast migration
rRNA metabolic process
rRNA processing
ncRNA processing
cellular response to corticosteroid stimulus
motor neuron axon guidance
regulation of leukocyte migration
cellular response to organic cyclic compound
cellular response to glucocorticoid stimulus

4  3  2  1  0  1  2  3  4

$-\mathrm{Log}10(P_{adj})$

**Figure 2. Translational and post-translational genes are strongly correlated with EPO production**

(A) Most significantly (mean TPM >10, |Spearman's r| > 0.5, p < 0.05) correlated genes with EPO or GFP production. Red and blue circles show positive and negative correlated genes, respectively.

(B) Gene set enrichment analysis of positively correlated genes with EPO production highlights biological GO terms that mostly associated with ribosome biogenesis, rRNA processing, and cytoskeleton reorganization, while negatively correlated genes with EPO production are mostly associated with ubiquitin biosynthesis.

in the networks of EPO- and GFP-correlating genes, respectively (Table S2).

Among the top 10 interacting genes in the network of EPO-correlating genes, three of them were directly involved in translation. RPL32 (r = 0.88, k = 94) and RPL35A (r = 0.88, k = 77), both components of cytosolic ribosomal subunits, were positively correlated with EPO production (Colombo et al. 1996; Anger et al., 2013) (r = 0.88, k = 73). Also, SNU13, a highly conserved nuclear protein involved in pre-mRNA splicing, showed a positive correlation with EPO production (Bertram et al., 2017). To investigate enriched pathways by significantly correlated genes, we used biological Gene Ontology (GO) terms and performed gene enrichment analysis using detected genes with a significant correlation with EPO production (Figure 2B, HyperGSA, p < 0.05). Pathways associated with ribosome biogenesis (p = 0.00004076) and RNA processing (0.0001966 < p < 0.0005302) exhibited significant enrichment by positively EPO-correlated genes. Likewise in GFP-correlated genes, nine out of the top 10 hub genes were associated with the ribosomal compartment, but all indicated negative trends of expression with increasing GFP production (−0.88 < r < −0.78, 93 < k < 134; Figure 2A; Table S2).

We also performed Pearson correlation analysis to find overlaps and differences compared with the Spearman rank-based correlation approach (Figures S3A and S3B). Results of the Pearson correlation analysis revealed, altogether, 223 and 93 genes were positively and negatively correlated with EPO production, respectively, and 99 and 19 genes were positively and negatively correlated with GFP production, respectively (Figure S3C, mean TPM >10, |Pearson's r| > 0.5, p < 0.05). Similar to the Spearman analysis, we found major regulators among the genes highly correlated (mean TPM >10, |Pearson's r| > 0.5, p < 0.05) with either EPO or GFP productivity by generating interaction networks (Figures S3D and S3E; Table S2). Four out of the top 10 interacting genes in the network of EPO-correlating genes were involved in translation. RPL38 (r = −0.89, p = 0.01, k = 73) and RPS9 (r = −0.91, p = 0.01, k = 88), both components of cytosolic ribosomal subunits, were negatively correlated with EPO production (Kondrashov et al., 2011), while MRPS11 (r = 0.9, p = 0.01, k = 71), a mitochondrial ribosomal gene, and EFL1 (p = 0.82, r = 0.04, k = 48), involved in 60S ribosomal subunit biogenesis (Thomson et al., 2013), were positively correlated. Similarly, with GFP producers, we observed a negative trend of expression in genes involved in translation with increasing GFP production. Likewise, four out of the top 10 genes (EEF2, EEF1G, RPL3, and RPL4) with the highest number of interactions in the network of GFP-correlating genes, serving as translation factors or components of the ribosomal large subunit, were negatively correlated (Pearson's r < −0.7, p ≤ 0.05, 94 < k < 202) with GFP production. Analysis of the corresponding biological GO terms (HyperGSA, p < 0.05; Figures S3F and S3G) for the EPO-correlating genes identified protein N-linked glycosylation as the top enriched GO term with positively EPO-correlated genes (HyperGSA, p = 0.01366; Figure S3F). Investigation of associated genes with this pathway highlighted seven genes that are positively correlated with EPO production and are involved in N-linked glycosylation (Figure S3F). KRTCAP2 (p = 0.02, Pearson's r = 0.85) and OST4 (p = 0.003, Pearson's

r = 0.95) both are subunits of oligosaccharyltransferase (OST) complex that catalyzes the initial transfer of a defined glycan from dolichol-pyrophosphate to the nascent polypeptide chains (Roboti and High 2012; Dumax-Vorzet et al. 2013). MGAT4A (p = 0.03, Pearson's r = 0.84) also regulates the formation of multi antennary branching structures in the Golgi apparatus (López-Orduña et al. 2007). MPDU1 (p = 0.01, Pearson's r = 0.89), ALG12 (p = 0.007, Pearson's r = 0.92), and ALG3 (p = 0.009, Pearson's r = 0.91) are involved in mannose transfer in the process of protein glycosylation (Kranz et al., 2001). However, UBE2J1 (p = 0.01, Pearson's r = 0.88) is involved in modification of proteins with ubiquitin and targeting abnormal proteins for degradation (Elangovan et al., 2017). Positive correlation of all these genes with EPO production may highlight some changes that happen in PTM processes coordinated with higher EPO production.

Although the individual correlated genes detected by the Pearson and Spearman approaches exhibited differences (Figure S3A), pathways related to translation and ribosome biogenesis were among the most significantly enriched (HyperGSA, p < 0.05) by both approaches (Figures 2B, S3F, and S3G). Furthermore, nine out of 10 hubs of the Spearman GFP-correlation network were ribosomal proteins (Figure 2A). This suggested that, in general, gene expression associated with translation processes is adopted in each group of EPO and GFP producers to support their specific needs imparted by their particular recombinant protein. To follow this observation, we investigated the expression of ribosomal genes in pairwise comparison of each clone with its respective control (Figures S4A and S4B). GFP producers did not share common differentially expressed ribosomal genes (B.H. adj. p < 0.05; Figure S4C). However, EPO producers showed 22 common differentially expressed (p < 0.05) ribosomal genes with at least 50% increase in their expression in comparison with control (Figure S4D). The functionality of these genes is mostly related to signal recognition particle (SRP)-dependent co-translational protein targeting the membrane. This suggests EPO producers have increased the share of genes related to co-translational protein targeting to membrane in their ribosomes to facilitate EPO production.

## EPO production results in a restructured cellular metabolism

We sought to understand which changes in the protein producers were due to the production of secretory EPO and which were common to all producers. PCA of the EPO and GFP transcriptomics data identified marked differences between EPO and GFP producers (Figure 1F), where a complete separation of EPO and GFP clones with the first principal component (34% of variance explained) showed that the transcriptomics data could capture differences between the secretory EPO and non-secretory GFP producers. We found 986 (922 up- and 64 downregulated) differentially expressed genes (adj p < 0.05, | Log2 fold change [L2FC]| > 1) between EPO and GFP producers. Gene set enrichment analysis (Figure S5A) showed that, for the EPO producers, beside gene sets specific for secretory protein production such as proteins targeting the ER or the cell membrane, genes associated with the oxidative phosphorylation

(legend on next page)

pathway were significantly (B.H. adj. p < 0.05) upregulated (Figure 3A).

To further explore the increased oxidative phosphorylation in gene expression among EPO producers, we investigated expression changes in genes associated with mitochondria (Figure 3A). All differentially expressed genes (B.H. adj. p < 0.05) involved in oxidative phosphorylation exhibited increased expression in EPO producers. The upregulated genes are associated with all complexes in the electron transport chain except for complex II, which facilitates the donation of electrons from FADH2. Higher gene expression in NADH dehydrogenase (complex I) followed by an increase in genes associated with complex III and complex V, could lead to higher ATP production (Figure 3A). All genes of mitochondrial origin across the electron transport chain (ETC) complexes exhibited at least a 2-fold increase in their expression (Figure S3A: genes with names starting with MT-). Apart from upregulation of mitochondrial genes, other genes that are expressed from the nuclear genome and are associated with oxidative phosphorylation also had a significant (B.H. adj. p < 0.05) expression increase in EPO producers (Figure 3A). To translate the findings to biological function and to investigate other possible differences in metabolism between EPO-producing cells versus GFP producers, we measured cellular metabolism using extracellular flux analysis (Figures 3B, 3C, S5B, and S5C). Using this technique, we could simultaneously monitor both metabolic branches of cellular ATP production, namely mitochondria and aerobic glycolysis. Here, we found that EPO-producing cells indeed have higher basal and maximal oxygen consumption rate (OCR) compared with GFP-producing cells (Figures 3B and S5B). Moreover, we found that aerobic glycolysis and pyruvate conversion to lactate (extracellular acidification rate [ECAR]) was higher in EPO-producing clones compared with the GFP clones investigated as well as the parental cells (Figures 3C and S5C). This suggests that the findings at the gene expression level regarding oxidative phosphorylation are translated into an increase in metabolic function in EPO-producing cells and, together with an increase in aerobic glycolysis, an increase in energy production.

### Genes negatively regulating ER stress are upregulated in EPO high-producer clone

Besides the observations regarding differences between mRNA copy number and protein productivity in EPOF21 in comparison with other EPO producers (Figure 1E), differences in transcription and translation as well as post-translational pathways might have affected protein productivity. In order to find genes and

pathways with altered patterns of expression, we first conducted pairwise differential expression analysis to find which genes are differentially expressed between EPOF21 and other EPO producers (Figure S6A). All clones revealed a statistically significant change (B.H. adj. p < 0.05) in EPO gene expression compared with EPOF21. However, logarithmic fold changes of EPO were different in each pairwise comparison of EPO clones with EPOF21, and only in EPOI2 did the EPO gene expression pass our arbitrary cutoff (L2FC > 1). The number of downregulated differentially expressed genes (B.H. adj. p < 0.05, L2FC < −1) varied between 186 (EPOF21 versus EPOI2) and 491 (EPOF21 versus EPO8), and the number of upregulated genes (B.H. adj. p < 0.05, L2FC > 1) showed a variation between 113 (EPOF21 versus EPOI2) and 792 (EPOF21 versus EPO8) genes (Figure S6A). We also found 44 common differentially expressed genes in comparisons of each of the EPO producer clones against EPOF21. Gene set enrichment analysis between EPOF21 and the other EPO producer cell lines (Figures S6B and S6C) indicated that many common changes between EPOF21 and the other EPO producers were related to post-translational pathways. To obtain extended gene sets spanning pathways related to protein secretion, we used a set of secretory protein machinery genes that are defined as core genes involved in protein secretion in human cells (Feizi et al., 2017; Gutierrez et al., 2020), and then searched for all GO pathways (Liberzon et al., 2011) that contained a significant (B.H. adj. p < 0.05) number of these genes (HyperGSA). This collection of gene sets was then used to compare the different EPO producer cell lines (Figure 4A: top altered pathways in at least one of the comparisons). Interestingly, when comparing EPOF21 with all other cell lines, we observed marked differences between the EPO8 cell line and the other EPO producers, which could have resulted from differences in EPO mRNA expression levels between EPO8 and the other cell lines (Figure 1E). Furthermore, the upregulated pathways in EPOF21 compared with all other EPO producers except EPO8 indicated a marked increase in expression of gene sets related to the ER and handling of misfolded proteins (Figure 4A).

To further investigate specific differences between EPOF21 and the other EPO producers, we analyzed which genes exhibited a significant pattern of differential expression (B.H adj. p < 0.05, L2FC > 1, mean TPM > 10) between EPOF21 and other EPO clones. Of the 19 genes that displayed a consistent pattern of change, five were upregulated and 14 downregulated (Figure 4B). Among the upregulated genes, ATF6B is a transcription factor active under ER stress conditions due to accumulation of

**Figure 3. Cellular metabolism is restructured to meet the energy demands of EPO production**
(A) The oxidative phosphorylation pathway was significantly (B.H. adj. p = 0.03) upregulated in EPO producers compared with GFP producers. Detailed visualization of upregulated genes and their role in increasing ATP production shows, apart from upregulation of mitochondrial encoded genes, other genes like NDUFA6, NDUFB1, and NDUFA4 in complex I; COX7B and COX7A2 in complex III; and ATP5MD and ATP5F1E in complex V have a significant (B.H. adj. p < 0.05) expression increase in EPO producers compared with GFP producers.
(B) Comparison of oxygen consumption rate (OCR) of clones shows an increase in basal, oligomycin, and FCCP (carbonyl cyanide-p-trifluoromethoxyphenylhydrazone) sensitive respiration in EPO clones versus GFP clones.
(C) Comparison of extracellular acidification rate (ECAR) highlights an increased basal and oligomycin induced aerobic metabolism of EPO versus GFP clones. Data are represented by mean + confidence interval [CI], and statistical significance was calculated between each EPO and GFP clones and represented by using a (significance versus GFP polyclone), b (GFP1), and c (GFP27). Each single dot represents a single technical replicate and different dot colors (white, black, and gray) represent single experiments (n = 3).
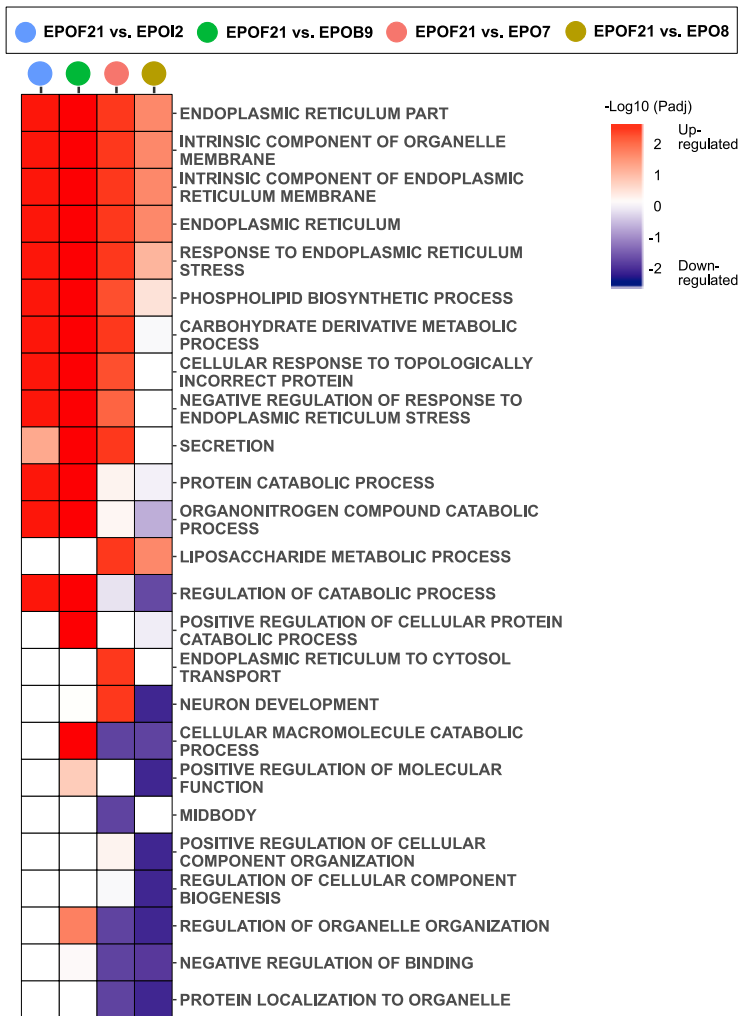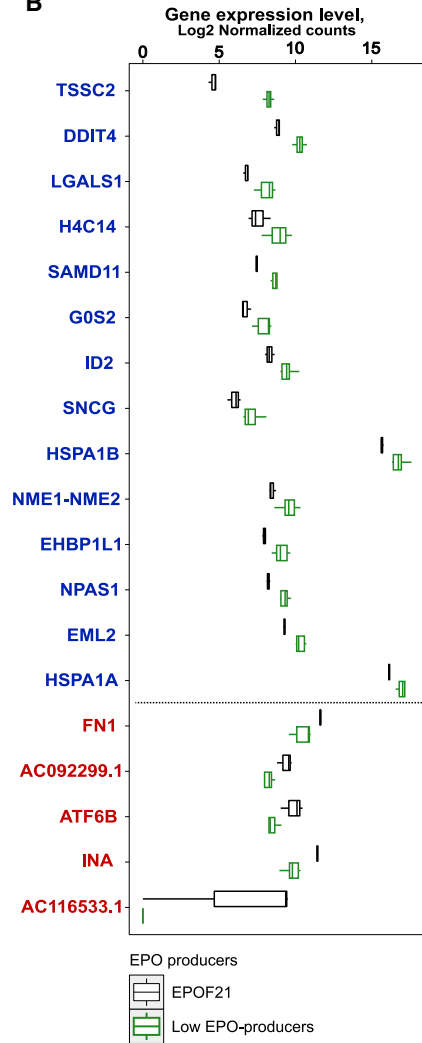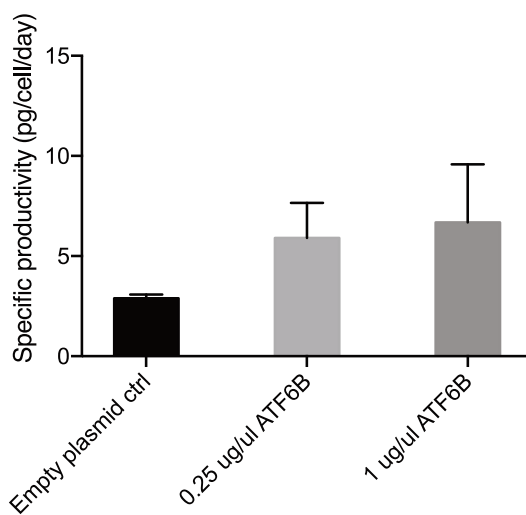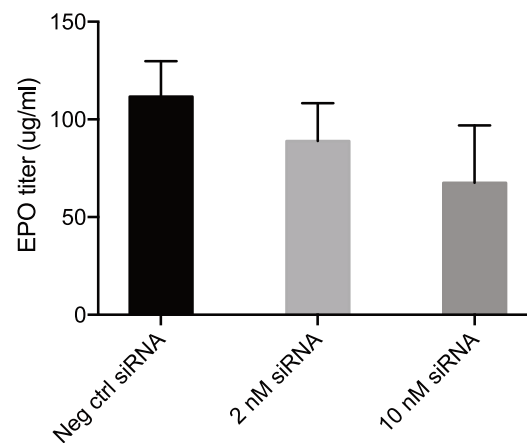
unfolded proteins (Thuerauf et al. 2004), whereas FN1 and INA share a role in extracellular matrix assembly (Singh et al. 2010). The gene with the highest positive fold change between EPOF21 and other EPO producers, AC116533.1, is a pseudogene of the ribosomal gene RPL36A, a ribosomal protein shown to play a role in ribosome biogenesis in yeast (Wan et al. 2015). Downregulated genes covered a wider spectrum of pathways, including apoptosis and growth regulation (LGALS1, G0S2, and EML2), nucleosome organization (HIST2H4A and SAMD1), metabolism of nucleotides (NME1 and NME2), protein folding (HSPA1A and HSPA1B), and vesicle trafficking (EHBP1L1). Also, ID1 and NPAS1 play roles in transcription regulatory pathways (Erbel-Sieler et al., 2004; Sikder et al., 2003), and TSSC2 is a pseudogene and a homologue to the Asparagine-Linked Glycosylation 1 (ALG1) gene in yeast (Jaeken et al. 2015).

As gene sets related to ER stress and/or unfolded proteins were identified as significantly enriched among secretion GO slims, we investigated the common significantly up- or down-regulated (B.H adj. $p < 0.05$) genes within these categories (response to endoplasmic reticulum [ER] stress, cellular response to topologically incorrect protein. and negative regulation of response to ER stress) between EPOF21 and the other EPO clones (Table S3). The most upregulated gene in EPOF21 compared with the other clones within these gene sets was ATF6B, also identified as one of the top most upregulated genes among all genes (Figure 4B). Interestingly, nine out of 10 genes belonging to the gene set of negative regulation of response to ER stress were found to be upregulated in EPOF21 compared with the other EPO producers. Besides ATF6B, other significantly upregulated genes in EPOF21 of this gene set included CLU, SYVN1, WFS1, HYOU1, PPP1R15A, UBE2J1, DERL2, and TMBIM6. As ATF6B stands out as one of the most differentially upregulated genes in EPOF21, we evaluated the effect of ATF6B on EPO production in high and low EPO producers. Overexpression of ATF6B in the low-producing clone EPOI2 resulted in an increased trend of specific EPO productivity (Figures 4C and S7A), whereas small interfering RNA (siRNA) silencing of ATF6B expression in the high producer EPOF21 showed a negative trend in volumetric EPO titers with lower ATF6B expression levels (Figures 4E and S7C). In the case of siRNA experiments, a lot of cell cluster formation, regardless of siRNA, prevented reliable cell counting and hence the specific productivity was not possible to determine. Even though changes in EPO expression were not significant (ANOVA one-way followed by Dunnett's test) in any of the experiments, collectively the trends in EPO production indicates that ATF6B may play a role in regulating the ER stress as a result of a high level of recombinant EPO entering the ER, aiding increased EPO secretion in the EPOF21 clone. Hence, ATF6B and other genes negatively regulating ER stress serve as potential cell line

engineering targets for improved recombinant productivity in HEK293 cells.

## DISCUSSION

In the present study, we generated two groups of clones producing either EPO or GFP at different levels (Figure 1A). EPO and GFP were chosen due to their different characteristics (i.e., final cellular location) in order to study the pathways behind protein production and differences caused by protein secretion. Indeed, analysis of the clonal transcriptomics data, where both clonal variation and the recombinant protein type were captured by the most informative first principal component (Figure 1F), suggests that the design of the experiment was appropriate for exploring both the protein production and secretion stages. However, despite individual HEK293 clones showing different productivities as well as different gene expression profiles (Figures 1C–1E and S1A), both transcription and translation were very well synced (Figure S1B). Thus, all of the clones except EPOF21, a clone with over 3-fold higher production levels compared with the next highest EPO producer (Figure 1E), had an almost equal ratio of secretion of the recombinant protein versus transcription of the recombinant gene (Figure 1E). The higher ratio in EPOF21 (in relation to the other clones) indicated that major translational and post-translational processes were affected. This suggests that, although all clones were useful to study the effects related to non-secretory and secretory protein production as well as their differences, the extraordinary EPOF21 clone was particularly valuable in further pinpointing major limiting parameters within secretory protein production, which can in turn help to improve the general overall productivity of cell factories.

Analysis of transcriptomic data between protein producer hosts and control parental cells showed upregulation of oxidative phosphorylation in the majority of the clones regardless of the transgene or level of recombinant protein production (Figure 1G), which suggests a high energy metabolism requirement to support transgene expression. Moreover, significant upregulation of genes linked to oxidative phosphorylation was observed when comparing EPO producers with GFP producers (Figure S5A), suggesting that the post-translatory secretory pathways in EPO clones may impose an even higher energy demand (Gutierrez et al., 2020). Simultaneous upregulation of genes encoded in both the mitochondria and nucleus highlight that the increase in transcription of genes involved in energy production is governed by some cellular metabolic master regulator (Figure 3A). The potentially increased demand for energy in EPO producers compared with GFP producers is reflected by an increase in respiration and aerobic glycolytic pathway, which can also increase energy production. Alternatively, the increased energy production in EPO clones could be also related to the

---

**Figure 4. Post-translational pathways are different between EPO producer clones**

(A) Significantly enriched GO terms in pairwise comparisons of EPOF21 and other EPO producers using GO slim secretion.

(B) Gene expression levels of the most significant (B.H. adj. $p < 0.05$) consistently differentially expressed genes between EPOF21 and other EPO producers. Red shows genes with upregulation in EPOF21 and blue genes are downregulated in EPOF21.

(C and D) EPO protein volumetric titers in supernatants (mean ± SD, n = 3) at day 3 in cultivation after transfection of the low-producer clones EPOI2 (C) and EPO7 (D) with an expression vector encoding rATF6B or an empty vector control, or (D) the high-producer EPOF21 clone with ATF6B siRNA or a negative control siRNA.

intrinsic molecular differences between EPO and GFP or the result of the general absence of secretory energy requirements in GFP producers. For instance, there are 114 different N-linked and O-linked reported structures for EPO (Alocci et al., 2019), but the complexity of post translational processing for GFP is simpler in general (Barondeau et al., 2003). Alternatively, as EPO itself acts as a signaling molecule, it could potentially impart regulatory effects on EPO-producing clones, such as increased oxidative phosphorylation activity (Plenge et al., 2012). However, EPO activity relies on the presence of the EPO receptor (EPOR), which is generally restricted to erythroid progenitor cells. Furthermore, a study of EPOR expression and activity by Ott et al. (2015) detected no EPOR protein in HEK293 cells unless they were transfected with an EPOR overexpression vector. Accordingly, the EPOR expression in our dataset did not correlate with EPO production (Figure S8A), and, furthermore, there was no significant difference in the expression of this gene between EPO producers and GFP producers or between EPO producers and their control cell line (Figures S8A–S8D). Moreover, we did not detect any significant difference (B.H adj. $p < 0.05$, L2FC $> 1$) in expression of downstream genes in the EPOR signaling pathway (Figure S8E) in comparing the transcriptomes of EPOF21 and other EPO producers (Figure S8F). We therefore reason that the transcriptional changes observed in EPO-producing clones are likely to be associated primarily with differences in secreted protein production rather than EPO-EPOR signaling, although an effect from signaling cannot be entirely ruled out.

In EPO-clones, significant upregulation of genes belonging to the eIF2 and mTOR signaling pathways could lead to activation of translation (Figure S2A). The eIF2 initiation complex is active in translation initiation in eukaryotic cells and plays a role in stabilizing the preinitiation complexes through binding to mRNA, GTP, methionine tRNA, and finally the 40S ribosomal subunit, to generate the 43S preinitiation complex (Hinnebusch 2011; Wek et al. 2006; Stolboushkina and Garber 2011; Schmitt et al. 2010). Likewise, activation of the mTORC1 module in the mTOR signaling pathway (Figure S2B) could promote protein synthesis and profoundly increase cellular ATP level by controlling mitochondrial biogenesis (Laplante and Sabatini 2009).

We detected a negative trend in the expression of some cytosolic ribosomal genes (Figures 2A and S3E) with increasing GFP production levels. The observed decrease in the expression of some ribosome-associated genes could be a result of an induced stress due to higher levels of translation, or it could indicate a rearrangement in the profile of ribosomal components. The latter is known as ribosome heterogeneity (Genuth and Barna 2018b), which considers ribosomes as dynamic macromolecular complexes that use a variation of different components in their structure to fit with desired specialized functions (Genuth and Barna 2018a). It has previously been shown that specialized ribosomes can preferentially translate different subsets of mRNAs (Shi et al., 2017). Moreover, looking into associated biological GO terms of positively correlated genes with EPO production showed an enrichment of ribosome biogenesis and rRNA metabolic processes (Figure 2B). Our observations highlight the presence of both upregulated and downregulated ribosome-related pathways in EPO and GFP producers

(Figures 2 and S3) and pinpoint transgene-specific correlations between expression levels of different ribosomal genes and the respective recombinant protein titers (Figure S4). These results could suggest that cells decrease the expression of some ribosomal components in a transgene-specific manner, in favor of more convenient production of the specific recombinant protein.

Focusing on the differences between high producer EPOF21 clone and other EPO producers, we observed that the majority of differences are in the post-translational steps, including protein folding, PTMs, and handling of misfolded proteins (Figures 4, S6B, and S6C). Indeed, in accordance with the increase of EPO transcripts in the EPOF21 clone, post-translational and ER-related pathways, such as negative regulation of ER stress, were upregulated in EPOF21, compared with other EPO clones with lower EPO transcript levels (Figure 4A). However, the EPO8 clone, with almost 20% higher EPO transcript than EPOF21 (Figures 1E and S1A), did not display an increased activity in some of the ER-associated pathways compared with EPOF21. This indicates that a higher expression of the recombinant gene without the support of post-translational steps is not only inefficient but can even cause problems with protein expression and lead to lower protein productivity. Furthermore, protein N-linked glycosylation was the highest enriched pathway with positively EPO-correlating genes (Figure S3D). Since EPO is a highly glycosylated protein (Falck et al., 2017) and these genes have a role in N-linked glocalization and also indicate a significant positive correlation with EPO expression, their higher expression could support an increase in the EPO production level.

To find the regulatory elements behind the higher activity of post-translational pathways in EPOF21, we investigated genes with a consistent pattern of change between EPOF21 and other EPO producers. Among the top upregulated genes in EPOF21 was ATF6B (Figure 4B). ATF6B is integrated within the ER membrane under normal conditions, but, during ER stress conditions, the cytoplasmic N-terminal domain is cleaved and the protein enters the nucleus to activate ER stress response genes (Thuerauf et al. 2004; Iurlaro and Muñoz-Pinedo 2016). Although ATF6B and its isomer ATF6A both activate the ER stress response genes (ERSRG), ATF6B represses the strong effect of ATF6A, and through this regulation causes a moderated activation of the ER stress response in comparison with ATF6A (Correll et al., 2019; Koul et al., 2017). Moreover, a previous study indicated that targeting ATF6A using a microRNA (miR-1287) enhances productivity in CHO cells producing a therapeutic antibody (Pieper et al., 2017), where miR-1287 had a very similar role to ATF6B in competing with and suppressing ATF6A to decrease the ER stress response. Additionally, it has previously been reported that continued ER stress causes higher expression of genes involved in the folding process (Jäger et al., 2012). Accordingly, activation of ATF6A during unfolded protein response (UPR) condition induces upregulation of heat shock proteins such as HSPA1A and HSPA1B (Gargalovic et al., 2006; Lee et al., 2003). Thus, downregulation of the stress-inducible chaperones HSPA1A and HSPA1B in EPOF21 (Figure 4B) could suggest a moderate level of ER stress in this clone, capable of supporting a high level of protein secretion, potentially governed by the higher expression of ATF6B. Indeed,

ATF6B expression knockdown in the high-producer clone, and rATF6B overexpression in low producers, suggested a role for the ATF6B transcription factor in enhancing EPO production in 293-F cells (Figures 4C and 4D). Other genes with negative regulatory roles during ER stress, which were upregulated in the high-producer clone compared with low-producers, were, for instance, CLU, encoding clusterin, and SYVN1, encoding E3 ubiquitin-protein ligase synoviolin. Clusterin is suggested to protect cells from apoptosis during ER stress, for instance by inducing autophagy (Lee et al., 2019), whereas SYVN1 also protects cells from apoptosis and is involved in degradation of misfolded proteins (Yamasaki et al., 2007; Kaneko et al., 2002). Another interesting example of an upregulated gene with negative regulatory effect on ER stress was PPP1R15A or GADD34, shown to reverse eIF-2$\alpha$-induced inhibition of protein translation upon ER stress (Brush et al. 2003). The over-representation of genes involved in negative regulation of ER stress that are upregulated compared with downregulated in EPOF21 compared with low producer clones, combined with the functional validation of the effect of ATF6B on EPO productivity, suggest that EPOF21 may enable higher EPO production through a more balanced ER stress response.

In conclusion, the present study offers important insights into transcriptomic changes during secretory EPO and non-secretory GFP production as well as key parameters influencing the different rates of protein production across a variety of producer clones. The results are thus valuable for improving our understanding of the biology behind protein secretion in mammalian cells and the behavior of cells under ER stress conditions. Moreover, the key differences uncovered between high- and low-producing EPO clones are potentially useful for future targeted cell-line engineering to improve therapeutic protein production.

### Limitations of the study

We report enhanced metabolism and negative regulation of ER stress to support higher EPO production through studies of stable HEK293 clones and pools of varying production levels of EPO and GFP. While experimentally confirming transcriptomics findings by both upregulation and silencing of genes and functional measurements, we believe further fine-tuning of key genes could strengthen our claims further. As alluded to in the discussion, part of the observed differences between clones could reflect the intrinsic functionality of recombinant proteins in cell metabolism and might not be a direct result of the secretion properties of the recombinant protein. Despite our reported attempts to pinpoint such effects for EPO, a more complete analysis could allow for such elucidation, which would make our conclusions more widely applicable for other proteins.

### STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHOD DETAILS
  - Cell line generation and cultivation
  - Protein expression analysis
  - Genome copy number estimation
  - Correlation analysis
  - RNA sequencing and data analysis
  - Ingenuity pathway analysis
  - Gene set analysis
  - Extracellular flux analysis (SeaHorse)
  - Generating GO slim secretion
- QUANTIFICATION AND STATISTICAL ANALYSIS

#### REFERENCES

Almo, S.C., and Love, J.D. (2014). Better and faster: improvements and optimization for mammalian recombinant protein production. Curr. Opin. Struct. Biol. 26, 39–43. https://doi.org/10.1016/j.sbi.2014.03.006.

Alocci, D., Mariethoz, J., Gastaldello, A., Gasteiger, E., Karlsson, N.G., Kolarich, D., Packer, N.H., and Lisacek, F. (2019). GlyConnect: glycoproteomics goes visual, interactive, and analytical. J. Proteome Res. 18, 664–677. https://doi.org/10.1021/acs.jproteome.8b00766.

Anger, A.M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and Drosophila 80S ribosome. Nature 497, 80–85. https://doi.org/10.1038/nature12104.

Barondeau, D.P., Putnam, C.D., Kassmann, C.J., Tainer, J.A., Getzoff, E.D., Tainer, J.A., and Getzoff, E.D. (2003). Mechanism and energetics of green

fluorescent protein chromophore synthesis revealed by trapped intermediate structures. Proc. Natl. Acad. Sci. U S A. *100*, 12111–12116. https://doi.org/10.1073/pnas.2133463100.

Behrouz, H., Molavi, B., Tavakoli, A., Askari, M., Maleknia, S., Mahboudi, F., and Khodadadian, M. (2020). Multivariate optimization of the refolding process of an incorrectly folded fc-fusion protein in a cell culture broth. Curr. Pharmaceut. Biotechnol. *21*, 226–235. https://doi.org/10.2174/1389201020666191002144424.

Bertram, K., Agafonov, D.E., Dybkov, O., Haselbach, D., Leelaram, M.N., Will, C.L., Urlaub, H., Kastner, B., Lührmann, R., and Stark, H. (2017). Cryo-EM structure of a pre-catalytic human spliceosome primed for activation. Cell *170*, 701–713.e11. https://doi.org/10.1016/j.cell.2017.07.011.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527. https://doi.org/10.1038/nbt.3519.

Brush, M.H., Weiser, D.C., and Shenolikar, S. (2003). Growth arrest and DNA damage-inducible protein GADD34 targets protein phosphatase 1 alpha to the endoplasmic reticulum and promotes dephosphorylation of the alpha subunit of eukaryotic translation initiation factor 2. Mol. Cell Biol. *23*, 1292–1303. https://doi.org/10.1128/mcb.23.4.1292-1303.2003.

Chin, C.L., Goh, J.B., Srinivasan, H., Liu, K.I., Gowher, A., Shanmugam, R., Lim, H.L., Choo, M., Tang, W.Q., Tan, A.H.M., et al. (2019). A human expression system based on HEK293 for the stable production of recombinant erythropoietin. Sci. Rep. *9*, 16768. https://doi.org/10.1038/s41598-019-53391-z.

Colombo, P., Read, M., and Fried, M. (1996). The human L35a ribosomal protein (RPL35A) gene is located at chromosome band 3q29-Qter. Genomics *32*, 148–150. https://doi.org/10.1006/geno.1996.0093.

Correll, R.N., Grimes, K.M., Prasad, V., Lynch, J.M., Khalil, H., and Molkentin, J.D. (2019). Overlapping and differential functions of ATF6α versus ATF6β in the mouse heart. Sci. Rep. *9*, 2059. https://doi.org/10.1038/s41598-019-39515-5.

Datta, P., Linhardt, R.J., and Sharfstein, S.T. (2013). An 'omics approach towards CHO cell engineering. Biotechnol. Bioeng. *110*, 1255–1271. https://doi.org/10.1002/bit.24841.

Davy, A.M., Kildegaard, H.F., and Andersen, M.R. (2017). Cell factory engineering. Cell Syst. *4*, 262–275. https://doi.org/10.1016/j.cels.2017.02.010.

Del Val, I.J., Polizzi, K.M., and Kontoravdi, C. (2016). A theoretical estimate for nucleotide sugar demand towards Chinese hamster ovary cellular glycosylation. Sci. Rep. *6*, 28547. https://doi.org/10.1038/srep28547.

Dietmair, S., Hodson, M.P., Quek, L.-E., Timmins, N.E., Gray, P., and Nielsen, L.K. (2012). A multi-omics analysis of recombinant protein production in Hek293 cells. PLoS One *7*, e43394. https://doi.org/10.1371/journal.pone.0043394.

Dumax-Vorzet, A., Roboti, P., and High, S. (2013). OST4 is a subunit of the mammalian oligosaccharyltransferase required for efficient N-glycosylation. J. Cell Sci. *126*, 2595–2606. https://doi.org/10.1242/jcs.115410.

Dumont, J., Euwart, D., Mei, B., Estes, S., and Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. Crit. Rev. Biotechnol. *36*, 1110–1122. https://doi.org/10.3109/07388551.2015.1084266.

Elangovan, M., Chong, H.K., Park, J.H., Yeo, E.J., and Yoo, Y.J. (2017). The role of ubiquitin-conjugating enzyme Ube2j1 phosphorylation and its degradation by proteasome during endoplasmic stress recovery. J. Cell Commun. Signal. *11*, 265–273. https://doi.org/10.1007/s12079-017-0386-6.

Erbel-Sieler, C., Dudley, C., Zhou, Y., Wu, X., Estill, S.J., Han, T., Ramon Diaz-Arrastia, E.W.B., Steven Potter, S., Steven, L., and McKnight, S.L. (2004). Behavioral and regulatory abnormalities in mice deficient in the NPAS1 and NPAS3 transcription factors. Proc. Natl. Acad. Sci. U S A *101*, 13648–13653. https://doi.org/10.1073/pnas.0405310101.

Falck, D., Haberger, M., Plomp, R., Hook, M., Bulau, P., Wuhrer, M., and Reusch, D. (2017). Affinity purification of erythropoietin from cell culture supernatant combined with MALDI-TOF-MS analysis of erythropoietin N-glycosylation. Sci. Rep. *7*, 5324. https://doi.org/10.1038/s41598-017-05641-1.

Feizi, A., Gatto, F., Uhlen, M., and Nielsen, J. (2017). Human protein secretory pathway genes are expressed in a tissue-specific pattern to match processing demands of the secretome. Npj Syst. Biol. Appl. *3*, 22. https://doi.org/10.1038/s41540-017-0021-4.

Fouladiha, H., Marashi, S.-A., Torkashvand, F., Mahboudi, F., Lewis, N.E., and Vaziri, B. (2020). A metabolic network-based approach for developing feeding strategies for CHO cells to increase monoclonal antibody production. Bioproc. Biosyst. Eng. *43*, 1381–1389. https://doi.org/10.1007/s00449-020-02332-6.

Gargalovic, P.S., Imura, M., Zhang, B., Nima, M., Gharavi, M.J., Clark, J.P., Yang, W.-P., Yang, W.P., He, A., Truong, A., Patel, S., et al. (2006). Identification of inflammatory gene modules based on variations of human endothelial cell responses to oxidized lipids. Proc. Natl. Acad. Sci. U S A *103*, 12741–12746. https://doi.org/10.1073/pnas.0605457103.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80. https://doi.org/10.1186/gb-2004-5-10-r80.

Genuth, N.R., and Barna, M. (2018a). Heterogeneity and specialized functions of translation machinery: from genes to organisms. Nat. Rev. Genet. *19*, 431–452. https://doi.org/10.1038/s41576-018-0008-z.

Genuth, N.R., and Barna, M. (2018b). The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. Mol. Cell *71*, 364–374. https://doi.org/10.1016/j.molcel.2018.07.018.

Goh, J.B., and Ng, S.K. (2017). Impact of host cell line choice on glycan profile. Crit. Rev. Biotechnol. *38*, 851–867. https://doi.org/10.1080/07388551.2017.1416577.

Gutierrez, J.M., Feizi, A., Li, S., Kallehauge, T.B., Hefzi, H., Grav, L.M., Ley, D., Baycin Hizal, D., Betenbaugh, M.J., Voldborg, B., et al. (2020). Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. Nat. Commun. *11*, 68. https://doi.org/10.1038/s41467-019-13867-y.

Hinnebusch, A.G. (2011). Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiol. Mol. Biol. Rev. *75*, 434–467. https://doi.org/10.1128/mmbr.00008-11.

Iurlaro, R., and Muñoz-Pinedo, C. (2016). Cell death induced by endoplasmic reticulum stress. FEBS J. *283*, 2640–2652. https://doi.org/10.1111/febs.13598.

Jaeken, J., Lefeber, D., and Matthijs, G. (2015). Clinical utility gene card for: ALG1 defective congenital disorder of glycosylation. Eur. J. Hum. Genet. *23*, 1431. https://doi.org/10.1038/ejhg.2015.9.

Jäger, R., Bertrand, M.J.M., Gorman, A.M., Vandenabeele, P., and Samali, A. (2012). The unfolded protein response at the crossroads of cellular life and death during endoplasmic reticulum stress. Biol. Cell *104*, 259–270. https://doi.org/10.1111/boc.201100055.

Kafri, M., Metzl-Raz, E., Jona, G., and Barkai, N. (2016). The cost of protein production. Cell Rep. *14*, 22–31. https://doi.org/10.1016/j.celrep.2015.12.015.

Kaneko, M., Ishiguro, M., Niinuma, Y., Uesugi, M., and Nomura, Y. (2002). Human HRD1 protects against ER stress-induced apoptosis through ER-associated degradation. FEBS (Fed. Eur. Biochem. Soc.) Lett. *532*, 147–152. https://doi.org/10.1016/s0014-5793(02)03660-8.

Kaufman, R.J., and Popolo, L. (2018). Protein synthesis, processing, and trafficking. In Hematology, Seventh Edition, R. Hoffman, E.J. Benz, L.E. Silberstein, H.E. Heslop, I. Jeffrey, J.A. Weitz, M.E. Salama, and S.A. Abutalib, eds. (Elsevier), pp. 45–58.e1. https://doi.org/10.1016/b978-0-323-35762-3.00005-6.

Kintzing, J.R., Filsinger Interrante, M.V., and Cochran, J.R. (2016). Emerging strategies for developing next-generation protein therapeutics for cancer treatment. Trends Pharmacol. Sci. *37*, 993–1008. https://doi.org/10.1016/j.tips.2016.10.005.

Koffas, M.A.G., Linhardt, R.J., Wang, Q., and Betenbaugh, M.J. (2018). Metabolic engineering of CHO cells to prepare glycoproteins. Emerg. Top. Life Sci. *2*, 433–442. https://doi.org/10.1042/etls20180056.

Kol, S., Kallehauge, T.B., Adema, S., and Hermans, P. (2015). Development of a VHH-based erythropoietin quantification assay. Mol. Biotechnol. *57*, 692–700. https://doi.org/10.1007/s12033-015-9860-7.

Kondrashov, N., Pusic, A., Stumpf, C.R., Shimizu, K., Hsieh, A.C., Ishijima, J., Shiroishi, T., and Barna, M. (2011). Ribosome-mediated specificity in Hox mRNA translation and vertebrate tissue patterning. Cell *145*, 383–397. https://doi.org/10.1016/j.cell.2011.03.028.

Koul, S., Jin, J.-K., Hogan, C.M., and Glembotski, C.C. (2017). ATF6B and ATF6A play complimentary roles in mediating adaptive and maladaptive signaling in cardiac myocytes. Circ. Res. *121*, A467.

Krämer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in ingenuity pathway analysis. Bioinformatics *30*, 523–530. https://doi.org/10.1093/bioinformatics/btt703.

Kranz, C., Denecke, J., Lehrman, M.A., Ray, S., Kienz, P., Kreissel, G., Sagi, D., Peter-Katalinic, J., Freeze, H.H., Schmid, T., et al. (2001). A mutation in the human MPDU1 gene causes congenital disorder of glycosylation type if (CDG-If). J. Clin. Investig. *108*, 1613–1619. https://doi.org/10.1172/jci13635.

Kuriakose, A., Chirmule, N., and Nair, P. (2016). Immunogenicity of biotherapeutics: causes and association with posttranslational modifications. J. Immunol. Res. *2016*, 1–18. https://doi.org/10.1155/2016/1298473.

Laplante, M., and Sabatini, D.M. (2009). mTOR signaling at a glance. J. Cell Sci. *122*, 3589–3594. https://doi.org/10.1242/jcs.051011.

Lee, A.-H., Iwakoshi, N.N., and Glimcher, L.H. (2003). XBP-1 regulates a subset of endoplasmic reticulum resident chaperone genes in the unfolded protein response. Mol. Cell Biol. *23*, 7448–7459. https://doi.org/10.1128/mcb.23.21.7448-7459.2003.

Lee, J., Hong, S.-W., Kwon, H., Park, S.E., Eun-Jung, Rhee, Park, C.-Y., Oh, K.-W., Park, S.-W., Lee, W.-Y., and Lee, W.Y. (2019). Resveratrol, an activator of SIRT1, improves ER stress by increasing clusterin expression in HepG2 cells. Cell Stress Chaperones *24*, 825–833. https://doi.org/10.1007/s12192-019-01012-z.

Liang, C., Austin, W.T., Chiang, A.H.H., Johnny, A., Schoffelen, S., Sorrentino, J.T., Kellman, B.P., Bao, B., Voldborg, B.G., and Lewis, N.E. (2020). A markov model of glycosylation elucidates isozyme specificity and glycosyltransferase interactions for glycoengineering. Current Res. Biotechnol. *2*, 22–36. https://doi.org/10.1016/j.crbiot.2020.01.001.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdottir, H., Tamayo, P., Mesirov, J.P., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740. https://doi.org/10.1093/bioinformatics/btr260.

Liu, Y., Beyer, A., and Aebersold, R. (2016). On the dependency of cellular protein levels on mRNA abundance. Cell *165*, 535–550. https://doi.org/10.1016/j.cell.2016.03.014.

López-Orduña, E., Cruz, M., and García-Mena, J. (2007). The transcription of MGAT4A glycosyl transferase is increased in white cells of peripheral blood of type 2 diabetes patients. BMC Genet. *8*, 73. https://doi.org/10.1186/1471-2156-8-73.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. *15*, 550. https://doi.org/10.1186/s13059-014-0550-8.

MacLean, B., Tomazela, D.M., Shulman, N., Chambers, M., Finney, G.L., Frewen, B., Kern, R., Tabb, D.L., Liebler, D.C., and MacCoss, M.J. (2010). Skyline: an open source document editor for creating and analyzing targeted proteomics experiments. Bioinformatics *26*, 966–968. https://doi.org/10.1093/bioinformatics/btq054.

Malm, M., Saghaleyni, R., Lundqvist, M., Giudici, M., Chotteau, V., Field, R., Varley, P.G., Hatton, D., Grassi, L., Svensson, T., et al. (2020). Evolution from adherent to suspension–systems biology of HEK293 cell line development. Preprint at bioRxiv. 18996. https://doi.org/10.1038/s41598-020-76137-8. https://www.biorxiv.org/content/10.1101/2020.01.29.924894v1.abstract.

P. Meleady, ed. (2017). Heterologous Protein Production in CHO Cells: Methods and Protocols (New York, NY: Humana Press).

Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Anaëlle Dos, S., Devos, S., Rousseau, F., Plets, E., Houthuys, E., et al. (2014). GlycoDelete engineering of mammalian cells simplifies N-glycosylation of recombinant proteins. Nat. Biotechnol. *32*, 485–489. https://doi.org/10.1038/nbt.2885.

Mori, Y., Yoshida, Y., Satoh, A., and Moriya, H. (2020). Development of an experimental method of systematically estimating protein expression limits in HEK293 cells. Sci. Rep. *10*, 4798. https://doi.org/10.1038/s41598-020-61646-3.

Orellana, C.A., Marcellin, E., Munro, T., Gary, P.P., and Nielsen, L.K. (2015). Multi-omics approach for comparative studies of monoclonal antibody producing CHO cells. BMC Proc. *9*, O8. https://doi.org/10.1186/1753-6561-9-s9-o8.

Ott, C., Martens, H., Hassouna, I., Oliveira, B., Erck, C., Zafeiriou, M.-P., Peteri, U.-K., Hesse, D., Gerhart, S., Altas, B., et al. (2015). Widespread expression of erythropoietin receptor in brain and its induction by injury. Mol. Med. *21*, 803–815. https://doi.org/10.2119/molmed.2015.00192.

Pieper, L.A., Strotbek, M., Wenger, T., Olayioye, M.A., and Hausser, A. (2017). ATF6β-Based fine-tuning of the unfolded protein response enhances therapeutic antibody productivity of Chinese hamster ovary cells. Biotechnol. Bioeng. *114*, 1310–1318. https://doi.org/10.1002/bit.26263.

Plenge, U., Belhage, B., Guadalupe-Grau, A., Andersen, P.R., Lundby, C., Flemming Dela, Stride, N., Pott, F.C., Christian Pott, F., Helge, J.W., and Boushel, R. (2012). Erythropoietin treatment enhances muscle mitochondrial capacity in humans. Front. Physiol. *3*, 50. https://doi.org/10.3389/fphys.2012.00050.

Rahimpour, A., Vaziri, B., Moazzami, R., Nematollahi, L., Barkhordari, F., Kokabee, L., Ahmad, A., and Mahboudi, F. (2013). 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*, 1116–1122. https://doi.org/10.4014/jmb.1302.02035.

Robinson, J.L., Kocabaş, P., Wang, H., Cholley, P.-E., Cook, D., Nilsson, A., Anton, M., Ferreira, R., Domenzain, I., Billa, V., et al. (2020). An Atlas of human metabolism. Sci. Signal. *13*. https://doi.org/10.1126/scisignal.aaz1482.

Roboti, P., and High, S. (2012). Keratinocyte-associated protein 2 is a bona fide subunit of the mammalian oligosaccharyltransferase. J. Cell Sci. *125*, 220–232. https://doi.org/10.1242/jcs.094599.

Salgado, E.R., Montesino, R., Jiménez, S.P., González, M., Hugues, F., Cabezas, O.I., Maura-Perez, R., Saavedra, P., Lamazares, E., Salas-Burgos, A., et al. (2015). Post-translational modification of a chimeric EPO-fc hormone is more important than its molecular size in defining its in vivo hematopoietic activity. Biochim. Biophys. Acta *1850*, 1685–1693. https://doi.org/10.1016/j.bbagen.2015.04.012.

Santoro, A., and Canova, C. (2005). Anemia and erythropoietin treatment in chronic kidney diseases. Minerva Urol. Nefrol. *57*, 23–31.

Schmitt, E., Naveau, M., and Mechulam, Y. (2010). Eukaryotic and archaeal translation initiation factor 2: a heterotrimeric tRNA carrier. FEBS (Fed. Eur. Biochem. Soc.) Lett. *584*, 405–412. https://doi.org/10.1016/j.febslet.2009.11.002.

Schneider, V.A., Graves-Lindsay, T., Howe, K., Bouk, N., Chen, H.-C., Kitts, P.A., Murphy, T.D., Pruitt, K.D., Thibaud-Nissen, F., Albracht, D., et al. (2017). Evaluation of GRCh38 and de Novo haploid genome assemblies demonstrates the enduring quality of the reference assembly. Genome Res. *27*, 849–864. https://doi.org/10.1101/gr.213611.116.

Scholz, H., Schurek, H.J., Eckardt, K.U., and Bauer, C. (1990). Role of erythropoietin in adaptation to hypoxia. Experientia *46*, 1197–1201. https://doi.org/10.1007/bf01936936.

Schwarz, H., Zhang, L., Andersson, N., Nilsson, B., and Chotteau, V. (2019). Small-scale end-to-end mAb platform with a continuous and integrated design. Integrated Continuous Biomanuf. IV. https://dc.engconfintl.org/biomanufact_iv/4/.

Shi, Z., Fujii, K., Kovary, K.M., Genuth, N.R., Röst, H.L., Teruel, M.N., and Barna, M. (2017). Heterogeneous ribosomes preferentially translate distinct

subpools of mRNAs genome-wide. Mol. Cell 67, 71–83.e7. https://doi.org/10.1016/j.molcel.2017.05.021.

Sikder, H.A., Devlin, M.K., Dunlap, S., Ryu, B., and Alani, R.M. (2003). Id proteins in cell growth and tumorigenesis. Cancer Cell 3, 525–530. https://doi.org/10.1016/s1535-6108(03)00141-7.

Singh, P., Carraher, C., and Schwarzbauer, J.E. (2010). Assembly of fibronectin extracellular matrix. Annu. Rev. Cell Dev. Biol. 26, 397–419. https://doi.org/10.1146/annurev-cellbio-100109-104020.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research 4, 1521. https://doi.org/10.12688/f1000research.7563.1.

Stolboushkina, E.A., and Garber, M.B. (2011). Eukaryotic type translation initiation factor 2: structure–functional aspects. Biochemistry 76, 283–294. https://doi.org/10.1134/s0006297911030011. https://idp.springer.com/authorize/casa?redirect_uri=https://link.springer.com/article/10.1134/S0006297911030011&casa_token=wkSdgPayzwUAAAAA:n85484QBZ5ilkxBRBaFfb3yNwT6Ra4XlLcWaXGg4D9wsoRpxGu_UvacITITXFMfsu1sVk2euXT_YaGVbwg.

Szklarczyk, D., Gable, A.L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N.T., Morris, J.H., Bork, P., et al. (2019). STRING v11: protein-Protein Association Networks with Increased Coverage, Supporting Functional Discovery in Genome-Wide Experimental Datasets. Nucleic Acids Res. 47, D607–D613. https://doi.org/10.1093/nar/gky1131.

Tambuyzer, E., Vandendriessche, B., Austin, C.P., Brooks, P.J., Larsson, K., Miller Needleman, K.I., Valentine, J., Davies, K., Groft, S.C., Preti, R., et al. (2020). Therapies for rare diseases: therapeutic modalities, progress and challenges ahead. Nat. Rev. Drug Discov. 19, 93–111. https://doi.org/10.1038/s41573-019-0049-9.

Tegel, H., Dannemeyer, M., Kanje, S., Sivertsson, Å., Berling, A., Svensson, A.-S., Hober, A., Enstedt, H., Volk, A.L., Lundqvist, M., et al. (2020). High throughput generation of a resource of the human secretome in mammalian cells. N. Biotech. 58, 45–54. https://doi.org/10.1016/j.nbt.2020.05.002.

Tejwani, V., Andersen, M.R., Nam, J.H., and Sharfstein, S.T. (2018). Glycoengineering in CHO cells: advances in systems biology. Biotechnol. J. 13, e1700234. https://doi.org/10.1002/biot.201700234.

Thomson, E., Ferreira-Cerca, S., and Hurt, E.J. (2013). Eukaryotic ribosome biogenesis at a glance. Cell Sci. 126, 4815–4821. https://doi.org/10.1242/jcs.111948.

Thuerauf, D.J., Morrison, L., and Glembotski, C.C. (2004). Opposing roles for ATF6$\alpha$ and ATF6$\beta$ in endoplasmic reticulum stress response gene induction. J. Biol. Chem. 279, 21078–21084. https://doi.org/10.1074/jbc.m400713200.

Väremo, L., Nielsen, J., and Nookaew, I. (2013). Enriching the gene set analysis of genome-wide data by incorporating directionality of gene expression and combining statistical hypotheses and methods. Nucleic Acids Res. 41, 4378–4391. https://doi.org/10.1093/nar/gkt111.

de Vree, P.J.P., Teunissen, H., Krijger, P.H.L., Geeven, G., Eijk, P.P., Sie, D., Ylstra, B., Hulsman, L.O.M., van Dooren, M.F., van Zutven, L.J.C.M., et al. (2014). Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. Nat. Biotechnol. 32, 1019–1025. https://doi.org/10.1038/nbt.2959.

Wang, H., Marcišauskas, S., Sánchez, B.J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., Kerkhoven, E.J., Agren, R., and Nielsen, J. (2018). RAVEN 2.0: a versatile toolbox for metabolic network reconstruction and a case study on streptomyces coelicolor. PLoS Comput. Biol. 14, e1006541. https://doi.org/10.1371/journal.pcbi.1006541.

Wang, Q., Wang, T., Yang, S., Sha, S., Wu, W.W., Chen, Y., Paul, J.T., Shen, R.-F., Cipollo, J.F., and Betenbaugh, M.J. (2020). Metabolic engineering challenges of extending N-glycan pathways in Chinese hamster ovary cells. Metab. Eng. 61, 301–314. https://doi.org/10.1016/j.ymben.2020.06.007.

Wan, K., Yabuki, Y., and Mizuta, K. (2015). Roles of Ebp2 and ribosomal protein L36 in ribosome biogenesis in Saccharomyces cerevisiae. Curr. Genet. 61, 31–41. https://doi.org/10.1007/s00294-014-0442-1.

Wek, R.C., Jiang, H.-Y., and Anthony, T.G. (2006). Coping with stress: eIF2 kinases and translational control. Biochem. Soc. Trans. 34, 7–11. https://doi.org/10.1042/bst0340007.

Xu, X., and Qi, L.S. (2019). A CRISPR–dCas toolbox for genetic engineering and synthetic biology. J. Mol. Biol. 431, 34–47. https://doi.org/10.1016/j.jmb.2018.06.037.

Yamasaki, S., Yagishita, N., Sasaki, T., Nakazawa, M., Kato, Y., Yamadera, T., Bae, E., Toriyama, S., Ikeda, R., Zhang, L., et al. (2007). Cytoplasmic destruction of p53 by the endoplasmic reticulum-resident ubiquitin ligase 'synoviolin. EMBO J. 26, 113–122. https://doi.org/10.1038/sj.emboj.7601490.

Zhou, G., Soufan, O., Ewald, J., Hancock, R.E.W., Basu, N., and Xia, J. (2019). NetworkAnalyst 3.0: a visual analytics platform for comprehensive gene expression profiling and meta-analysis. Nucleic Acids Res. 47, W234–W241. https://doi.org/10.1093/nar/gkz240.

# STAR★METHODS

## KEY RESOURCES TABLE

| REAGENT or RESOURCE | SOURCE | IDENTIFIER |
|---|---|---|
| **Antibodies** | | |
| Rabbit anti-ATF6B, HPA046871 | Kindly provided by the Human Protein Atlas | https://www.proteinatlas.org |
| Monoclonal anti-FLAG M2 antibody produced in mouse | Sigma Aldrich | Cat# F3165, clone M2 |
| Anti-GAPDH (5H11) antibody | AbClon | Cat# AbC-1001 |
| Swine anti-rabbit immunoglobulins/HRP | Agilent | Cat# p039901-2 |
| Goat anti-mouse immunoglobulins/HRP | Agilent | Cat# P044701-2 |
| **Chemicals, peptides, and recombinant proteins** | | |
| Immobilon Western Chemiluminescent HRP Substrate | Merck Millipore | Cat# WBKLS0100 |
| **Critical commercial assays** | | |
| Illumina HiSeq | | www.illumina.com |
| Targeted Locus Amplification | (de Vree et al. 2014) | www.cergentis.com |
| RNAlater(TM) stabilization solution, Invitrogen | Thermo Fisher Scientific | Cat# AM7024 |
| 293Fectin(TM) Transfection reagent | Thermo Fisher Scientific | Cat# 12347019 |
| DsiRNA hs.Ri.ATF6B.13.1 | Integrated DNA Technologies (IDT) | Cat# hs.Ri.ATF6B.13.1 |
| Negative control DsiRNA | Integrated DNA Technologies (IDT) | Cat# 51-01-14-04 |
| Mem-PER(TM) plus membrane protein extraction kit | Thermo Fisher Scientific | Cat# 89842 |
| **Deposited data** | | |
| Transcriptome data | This paper | Sequence Read Archive (SRA), BioProject: PRJNA834597 |
| Codes for analysis | This paper | Zenodo: https://zenodo.org/record/6519745#.YnN_TPNBwUE https://doi.org/10.5281/zenodo.6519745 |
| **Experimental models: Cell lines** | | |
| 293-F | Thermo Fisher Scientific | Cat# 11625019 |
| Freestyle 293-F | Thermo Fisher Scientific | Cat# R79007 |
| **Softwares and algorithms** | | |
| The R Project | | www.r-project.org |
| MATLAB R2017b | The MathWorks, Inc. | mathworks.com/products/matlab.html |
| Bioconductor | (Gentleman et al., 2004) | www.bioconductor.org/ |
| DESeq2 | (Love et al. 2014) | R Bioconductor |
| PIANO | (Väremo et al. 2013) | R Bioconductor |
| IPA® | QIAGEN Inc. | digitalinsights.qiagen.com/ |
| MSigDB database | (Liberzon et al., 2011) | www.gsea-msigdb.org/gsea/msigdb |

## RESOURCE AVAILABILITY

### Lead contact
Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Johan Rockberg (johan.rockberg@biotech.kth.se).

### Materials availability
Generated cell lines can be shared upon request but may be limited due to our need to maintain the stocks.

**Data and code availability**

- Code and datasets to reproduce the figures presented here as well as all analysis outputs, are available on GitHub repository: https://zenodo.org/record/6519745#.YnN_TPNBwUE. DOI is listed in the key resources table (https://doi.org/10.5281/zenodo.6519745).
- Data files too large to host on GitHub were deposited on Sequence Read Archive (SRA), BioProject: PRJNA834597.
- Any additional information required to reanalyze the data reported in this work paper is available from the Lead contact upon request.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

The HEK293 derived suspension cell lines 293-F and Freestyle 293-F (Thermo Fisher Scientific) have been used in this study to generate stable cell clones expressing GFP or EPO.

## METHOD DETAILS

### Cell line generation and cultivation

HEK293 cell lines 293-F and Freestyle 293-F were cultivated in Freestyle 293 expression medium (Gibco, Thermo Fisher Scientific) at 37°C, 125 rpm and 8% $CO_2$. Stable cell clones expressing EPO or GFP were generated by transfection of linearized pD2529 plasmids (Atum), expressing either recombinant human EPO fused to a C-terminal HPC4-tag or recombinant GFP, into Freestyle 293-F (for EPO clones) or 293-F (for GFP clones) cell lines. Transfections were carried out using PEI at a DNA:PEI ratio of 1:3 and 1 μg plasmid per 1 million cells. Polyclonal batches of cells expressing GFP and EPO, respectively, were generated by puromycin selection. Single clones of HEK293 cells expressing EPO or GFP were generated from the polyclonal batches by seeding single cells per well of 384-well plates by either limiting dilution or FACS (Astrios, Beckman Coulter). In case of GFP-expressing cells, sorting by FACS was performed based on the GFP-signal. Verification of cell monoclonality was performed by microscopy (Leica DMI6000B). Single cells were expanded in 1.5% HEPES and growth media at 37°C and 8% $CO_2$. Cells of single clones were seeded at 0.3 million cells/mL in duplicates per clone and cultivated for 72 hours in 125 mL Erlenmeyer shake flasks with vented caps at 125 rpm, 37°C and 5% $CO_2$. Every 24 hours cell growth and viability were determined using a TC20 cell counter (Bio-Rad Laboratories). At 72 hours post inoculation, cell and supernatant samples were collected for downstream analysis. Cell samples for downstream RNA isolation were stored in RNAlater™ stabilization solution (Invitrogen™). Over-expression of ATF6B was performed by transfection of the pKTH16 plasmid expressing ATF6B fused to a FLAG-tag, alone or combined with an empty pKTH16 plasmid to get a final concentration of 1 μg DNA per ml cell cultivation at 1 million cells/mL in a 2.5 mL total volume in 24 deep-well plates. Transfection was performed using PEImax at a DNA to PEI ratio of 1:3. Plates were incubated at 225 rpm, 37°C and 5% $CO_2$. ATF6B silencing was performed by transfecting a dsiRNA targeting ATF6B (hs.Ri.ATF6B.13.1, Integrated DNA Technologies) or a negative control dsiRNA (Integrated DNA Technologies) at a final concentration of 2 or 10 nM using 293Fectin transfection reagent at 2 μg/mL (Thermo Fisher Scientific) in 125 mL Ehrlenmeyer shake flasks as described previously. Cell samples were collected after 48 h and both cell and supernatant samples were harvested after 72 h.

### Protein expression analysis

Productivity of EPO in cell supernatants was determined by Octet RED96 biolayer interferometry (ForteBio, Fremont, CA, USA) as described by (Kol et al., 2015). Briefly, biotinylated $V_H$H-anti EPO (Capture Select™, Thermo Scientific) was immobilized on streptavidin sensors and used to measure EPO binding directly in cell supernatants in citric acid (20 mM), 0.1% BSA, 0.1% tween-20, 0.5 M NaCl. Signals were compared to an EPO standard curve of known concentration. Regeneration of sensors was performed using 10 mM $NaH_2PO_4$ (pH 12). GFP productivity was determined by measuring the GFP signal (FL-1 channel) of cells by flow cytometry (Gallios, Beckman Coulter). EPO productivity following siRNA ATF6B silencing or ATF6B overexpression was determined by LC-MS/MS. Fifty microliter supernatant from each cultivation was spiked with 2.1 pmol stable isotope labeled protein fragment mapping to the EPO sequence. The samples were precipitated by addition of 200 μL cold acetone and incubated at −20°C overnight. The samples were centrifuged (20,000 x $g$, 4°C) for 30 min and the supernatant was discarded. The pellet was washed twice with cold acetone with subsequent centrifugations (20,000 x $g$, 4°C, 10 min). The pellets were resuspended in 20 μL 7 M urea, 2 M thiourea and subsequently diluted by addition of 140 μL 100 mM triethylammonium bicarbonate. Five microliters of each sample was diluted in 1xPBS (10 mM phosphate, 150 mM NaCl) for tryptic digestion. The samples were reduced by addition of dithiothreitol added to the samples to a final concentration of 10 mM. The samples were incubated for 1 hour at 30°C. This was followed by alkylation by addition of 2-chloroacetamide to a final concentration of 50 mM and incubated for 30 min, in darkness at room temperature. Trypsin (P/N 90058, Thermo Fisher Scientific, Waltham, MA, USA) was added to the samples in a 1:50 ratio (enzyme:substrate) and the samples were incubated at 37°C overnight. Digestion was quenched by addition of formic acid (FA) to a final concentration of 0.5%. The samples were analyzed using a TSQ Altis (Thermo Fisher Scientific) coupled to an Ultimate 3000 LC-system (Thermo Fisher), equipped with a 15 cm EASY-Spray analytical column (P/N ES906A, particle size: 2 μm, pore size: 100Å, 150 μm × 15 cm, Thermo Fischer Scientific) and an Acclaim PepMap 100 trap cartridge column (PN 160454, particle size: 5 μm, pore size: 100 Å, 0.3 mm × 5 mm,

Thermo Fischer Scientific). Ten microliters of each sample was loaded onto the column and the peptides were separated over a 10 min long gradient (1-30% solvent B; solvent A [3% acetonitrile, 0.1% FA], solvent B [95% acetonitrile, 0.1% FA], 3 μL/min). The mass spectrometer was operated in a scheduled SRM mode with 3 min retention time windows and a cycle time of 0.5 s, targeting 51 transitions in total (Table S3).

All data were processed using Skyline (v. 21.1.0.146) (MacLean et al., 2010), from which the ratio to standard was extracted for all peptides (Table S3). Quantification for all samples was based on the peptide VNFYAWK++.

Intracellular protein extraction prior to western blotting was performed by Mem-PER plus membrane protein extraction kit (Thermo Fisher Scientific). ATF6B expression after over-expression or siRNA silencing was evaluated by SDS-PAGE and western blotting using a mouse anti-FLAG antibody (Sigma Aldrich) at a 1:2000 dilution, or rabbit anti-ATF6B antibody (HPA046871, Atlas Antibodies) at a 1:320 dilution, respectively for detection. For loading control, membranes were stripped and stained with a mouse anti-GAPDH antibody (5H11, AbClon) at a 1:10000 dilution. Primary rabbit antibodies were detected with a HRP-conjugated swine anti-rabbit antibody (p039901-2, Dako) at a 1:4000 dilution, and mouse primary antibodies were detected with an HRP-conjugated polyclonal goat anti-mouse antibody (P0447, Dako) at a 1:10000 dilution. For chemiluminescent detection, Immobilon Western Chemiluminescent HRP Substrate (Millipore) was added to membranes and images were acquired using a ChemiDoc Imaging system (Bio-rad).

## Genome copy number estimation

Cryopreserved cell stocks, in cell growth medium with 10% DMSO, of each cell clone were sent to Cergentis B.V. (Utrecht, The Netherlands) for Targeted Locus Amplification (TLA) (de Vree et al., 2014) and next-generation sequencing by Illumina MiniSeq. EPO and GFP transgene sequences were mapped to the pD2529 plasmid sequences used for generating stable clones. Target-specific sequences were mapped to the hg19 genome (Genome Reference Consortium Human Build 37 (GRCh37). Estimations of copy numbers were based on the number of plasmid integrations into the genome, number of fusion reads and ratio between coverage of the transgene and the surrounding genomic region.

## Correlation analysis

For correlation analyses between growth rates, protein productivity, gene and transcript copy number and final visualizations, the R package GGally v1.5.0 (https://github.com/ggobi/ggally) was used with default settings. Correlation analysis between gene expression and EPO or GFP production was performed using the Pearson and Spearman methods, where results were filtered based on the significance threshold of 0.05 and absolute correlation coefficients surpassing 0.5. An additional filter was applied to exclude genes with a mean of TPM lower than 10 across all samples of the same producing group. Interaction between correlating genes was based on the STRING database (Szklarczyk et al., 2019) and filtered according to experimental evidence and a confidence interval higher than 900. First-order interacting partners of correlating genes with mean of TPM higher than 10, were also included in the networks. For network visualization and gene set analysis of highly correlated genes, NetworkAnalyst (Zhou et al., 2019) was used with default settings and GO biological process version v7.1 was used as the gene set collection.

## RNA sequencing and data analysis

Total RNA was extracted from cells using RNeasy plus Mini Kit (Qiagen) according to the protocol provided by the manufacturer. RNA integrity was verified by RNA 6000 Nano chips on a 2100 Bioanalyzer instrument (Agilent Technologies). Extracted RNA samples were shipped to GATC (Konstanz, Germany) for mRNA sequencing by Illumina HiSeq instrument using the Inview Transcriptome Discover service (paired end, 2 × 150 bp read length, >30 million read pairs).

Transcript quantification was performed using the standalone package Kallisto v0.43.1_1 (Bray et al., 2016) with default settings and version GrCh38 of the human genome (Schneider et al., 2017) was used for transcript mapping. To import raw counts data into R, the tximport package v1.14.2 (Soneson et al. 2015). For PCA analysis, the R package DESeq2 v1.26.0 (Love et al. 2014) was used with default settings and log transformed normalized counts. Differential expression analysis was also performed using DESeq2 v1.26.0, where the Wald test was used for calculating logarithmic fold changes and *p-value*s were adjusted by the Benjamini-Hochberg method. To find genes with a robust expression change between EPOF21 and other EPO producer clones, we filtered genes with an absolute log2 fold change higher than 1, adjusted *p-value* smaller than 0.05 and average TPM higher than 10 across all samples.

## Ingenuity pathway analysis

Enrichment analysis for finding enriched pathways was performed using Ingenuity Pathway Analysis (IPA®) software (Ingenuity Systems, http://ingenuity.com) with default settings. For all comparisons, gene names were first mapped to the Ingenuity database and then statistically significant differentially expressed genes (B.H. adj. *p-value* < 0.05) with at least 50% change in their expression were selected for finding significantly enriched pathways (Benjamini-Hochberg corrected *p-value* < 0.05).

## Gene set analysis

To measure the enrichment of different gene groups, gene set analysis (GSA) was performed. The procedure for running GSA was the same across the study. First, gene set collections were retrieved from the Molecular Signature Database (MSigDB) (Liberzon et al., 2011). For gene sets related to Human1 (Robinson et al., 2020), the MATLAB toolbox RAVEN v2.3.1 was used (H. Wang et al., 2018)

with default settings to group genes based on their associations with reactions in different subsystems of the Human1 v1.3.1 model. To calculate test statistics for each given gene set, the Wilcoxon rank-sum from the R package PIANO v2.2.0 (Väremo et al. 2013) was used, where results for each gene set based on *p-value*s and log2 fold-change of genes from the DE analysis were compared with 100,000 randomly shuffled gene sets of the same size (random permutations = 100,000).

### Extracellular flux analysis (SeaHorse)

Metabolic flux analysis was performed on different clones of EPO and GFP producing cells as well as their parental controls 293F and 293Freestyle using Seahorse XF96 Extracellular Flux Analyzer (Agilent Technologies, Santa Clara, CA, U.S.). The same day of the experiment a XF96 cell plate was coated with Cell-Tak (Corning, NY, U.S.) at 22.4 µg/mL and 25 µL per well for 30 m. After washing each well twice with 200 µL $H_2O$, the plate was allowed to dry in the cell culture hood for 30 m before cell seeding. HEK cells were washed once with assay media (XF DMEM containing 25 mM glucose, 4 mM glutamine, 1 mM pyruvate, pH 7.4), resuspended and counted. 30000 cells/well were seeded in 50 µL and the plate was spun at 200xg for 1 m without brake. The plate was then incubated at 37°C without $CO_2$ for 30 m and a further 130 µL of assay media was added to each well.

   The oxygen consumption rate (OCR) and acidification rate (ECAR) were recorded at a basal level and following injection of oligomycin (3 µM final), carbonyl-cyanide 4-trifluoromethoxy-phenylhydrazone FCCP (0.2-0.6 µM final), and a mixture of rotenone and antimycin A (5 µM). The running template was 2 m mix, 1 m wait and 3 m measure. All chemicals were purchased from Sigma-Aldrich. Data were normalized on the number of cells per well and against basal OCR and ECAR. Normalization for cell number was carried out by staining the nuclei with Hoescht 33342 (Molecular Probes) injected together rotenone and antimycin A and then imaging each well using a BD pathway 855 instrument(BD Biosciences, Franklin Lakes, U.S.) with 10x objective and montage 5×4. Cell number was counted with Cell profiler software.

   Data were analyzed after outlier removal (GraphPad software San Diego, U.S.) by using ANOVA (non parametric) considering all the technical replicates of 3 different experiments.

### Generating GO slim secretion

To estimate the extent to which the protein secretory subsystems differed between EPOF21 and other EPO producers, a GO slim for protein secretion-related gene subsets was generated for use in gene set analysis. The GO slim consisted of a list of 590 previously reported genes involved in human protein secretion and their association to core components of the secretory pathway (Feizi et al., 2017). In addition, gene sets were retrieved from the GO biological processes, GO cellular components and GO molecular functions MSigDB collections (Liberzon et al., 2011) if they were significantly enriched in secretion-associated genes (B.H. adj. *p-value* < 0.05). A one-tailed Fisher's exact from the R package PIANO v2.2.0 was used as the statistical test for calculating the enrichment significance of gene sets and *p-value*s were adjusted to control for the false discovery rate (*FDR*) with the Benjamini-Hochberg procedure.

### QUANTIFICATION AND STATISTICAL ANALYSIS

All computational analyses were performed using R v3.6 (www.r-project.org) and Matlab R2017b (www.mathworks.com). More details about each of the analyses are provided under the method details section.