Transcriptome analysis of EPO and GFP HEK293 Cell-lines Reveal Shifts in Energy and ER Capacity Support Improved Erythropoietin Production in HEK293F Cells

Rasool Saghaleyni[1], Magdalena Malm[2], Jan Zrimec[1], Ronia Razavi[2], Num Wistbacka[2], Veronique Chotteau[2], Diane Hatton[3], Luigi Grassi[3], Aleksej Zelezniak[1], Thomas Svensson[1,4], Jens Nielsen[1,5], Jonathan L. Robinson[1,4*] and Johan Rockberg[2*]

1 Department of Biology and Biological Engineering, Chalmers University of Technology, SE-412 96 Gothenburg, Sweden
2 KTH - Royal Institute of Technology, School of Engineering Sciences in Chemistry, Biotechnology, and Health, Dept. of Protein Science, SE-106 91 Stockholm, Sweden
3 Cell Culture & Fermentation Sciences, BioPharmaceuticals Development R&D, AstraZeneca, Milstein Building, Granta Park, Cambridge CB21 6GH UK
4 Department of Biology and Biological Engineering, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Chalmers University of Technology, Kemivägen 10, SE-41258 Gothenburg, Sweden.
5 Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

Corresponding authors:

Johan Rockberg, KTH - Royal Institute of Technology, School of Engineering Sciences in Chemistry, Biotechnology and Health, S- 106 91 Stockholm, Sweden
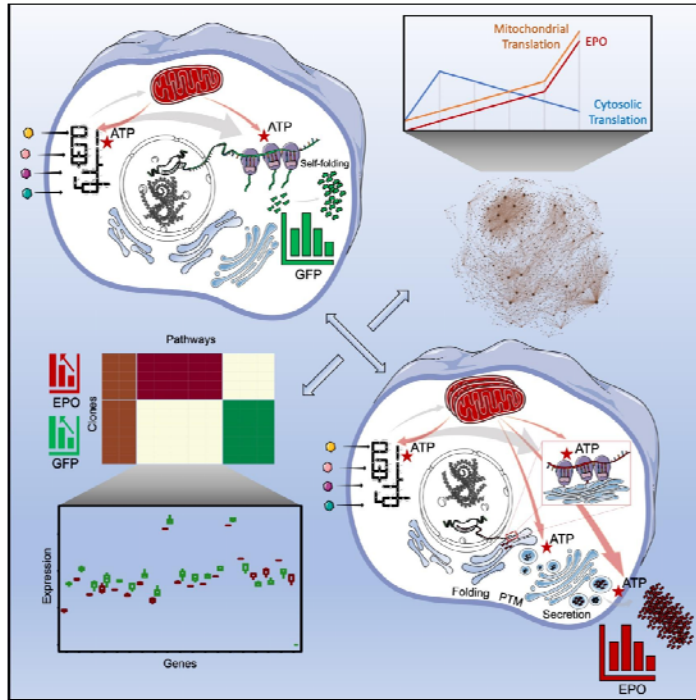E-mail: johan.rockberg@biotech.kth.se

Office: +46 8 790 99 88
Fax: +46 8 5537 8481

Jonathan L. Robinson, Department of Biology and Biological Engineering, Chalmers University of Technology, 41296 Gothenburg, Sweden
E-mail: jonrob@chalmers.se

Graphical Abstract

In Brief

Although the protein secretion process has been widely studied, the complexity of it leaves many questions with regards to defining bottlenecks for successful protein secretion to be answered. By investigating the transcriptomic profiles of different HEK293 clones with varying translational rates producing either the secreted protein erythropoietin or the intracellular GFP, we reveal that high ATP production and improved capacity of specific post-translational pathways are key factors associated with boosting erythropoietin production.

Highlights

- Transcriptomics analysis of a panel of HEK293 stable cell lines expressing GFP or erythropoietin (EPO) at varying translational rates

- Expression of mitochondrial ribosomal genes is positively correlated with EPO secretion

- Expression of different cytosolic ribosomal genes are correlated with productivity in a recombinant-protein specific manner

- High EPO producing clones have significant upregulation of ATF6B, potentially enabling a beneficial ER stress response to cope with high protein secretion

Summary

Higher eukaryotic cell lines like HEK293 are the preferred hosts for production of therapeutic proteins requiring human post translational processing. However, recombinant protein production can result in severe stress on the cellular machinery, resulting in limited titre and product quality. To investigate the cellular and metabolic characteristics associated with these limitations, we compared erythropoietin (secretory) and GFP (non-secretory) protein producer HEK293 cell-lines using transcriptomics analysis. Despite the high demand for ATP in all protein producer clones, a significantly higher capacity for ATP production was observed with erythropoietin producers as evidenced by the enrichment of upregulated genes in the oxidative phosphorylation pathway. In addition, ribosomal genes exhibited specific patterns of expression depending on the recombinant protein and the production rate. In a clone displaying a dramatically increased erythropoietin secretion, we detected higher ER stress, including upregulation of the ATF6B gene. Our results are significant in recognizing key pathways for recombinant protein production and identifying potential target genes for further development of secretory power in mammalian cell factories.

Keywords

1. 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) as those in humans, which is often a requirement for complex therapeutic proteins (Davy et al., 2017; Orellana et al., 2015; 2017). However, the PTM pattern of 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 (Goh and Ng, 2017; Kuriakose et al., 2016). Therefore, besides further development of the CHO cell-line to meet the high-quality PTM requirements and increased yields (Datta et al., 2013; Fouladiha et al., 2020; Koffas et al., 2018; Liang et al., 2020; Tejwani et al., 2018; 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 (Behrouz et al., 2020; Del Val et al., 2016; Liang et al., 2020; Meuris et al., 2014). However, despite the current knowledge of protein production and secretion in mammalian cells, there is still notable ambiguity in understanding and predicting the production and secretion rates and product quality under different conditions. 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). 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 secretory protein erythropoietin (EPO) or the non-secretory protein GFP and compared each of these groups with each other and with their parental cell-lines. We found genes with a high correlation with EPO or GFP production and investigated biological functions associated with them. Furthermore, we investigated ribosome heterogeneity between EPO and GFP producers and detected ribosomal genes with specific patterns of expression correlating with EPO and GFP production. Since the generation of single clones by random integration enabled us to isolate a 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.

2. Results

2.1 Single cell cloning generates EPO and GFP producer clones with an altered metabolism from the host cell-line
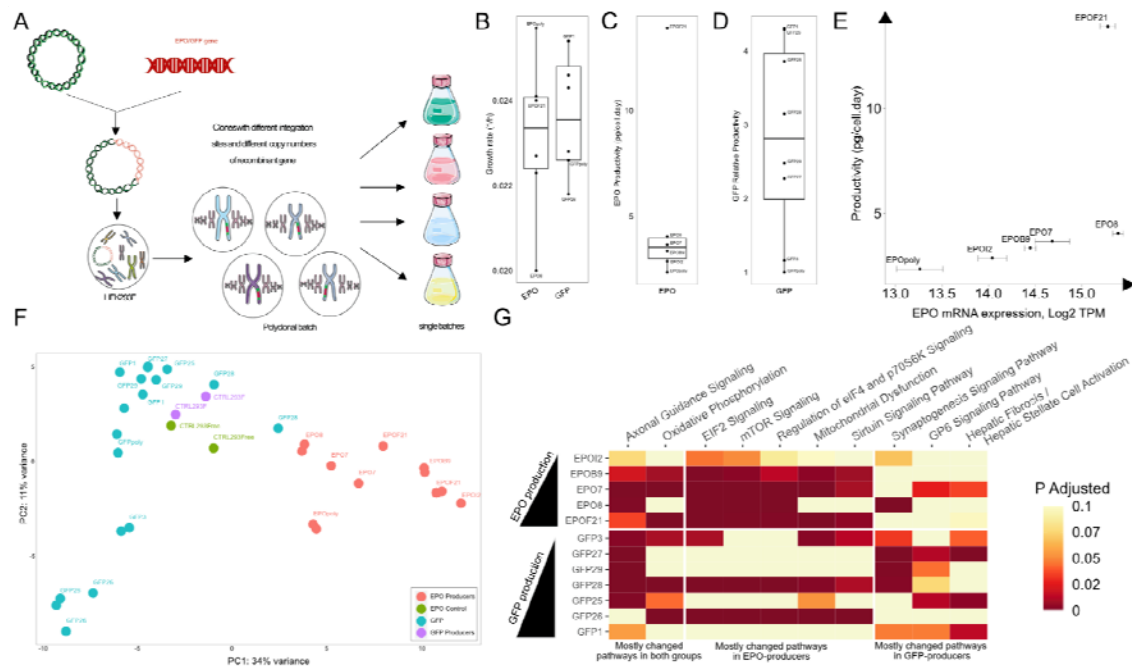
To investigate recombinant protein production in HEK293 cells, we transformed 293-F cell-lines to generate secretory protein (EPO) producer clones and non-secretory protein (GFP) producer clones, respectively (**Figure 1**A, **M1**-**3**). Initially, polyclonal batches of cells producing either EPO or GFP (EPOpoly and GFPpoly, respectively) were generated through random integration of plasmid DNA into the host genome resulting in collections of clones with various transgene integration sites and copy numbers. From these batches 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 to the hosts, respectively (**Figure 1**B). Protein productivity was however markedly different among clones and varied by almost 6-fold with EPO producers and up to 4-fold with GFP producers (**Figure 1**C-D, **M2**). The most productive EPO clone (EPOF21) had a titer of 13.9 pg/cell.day, over 3-fold higher than the second highest producing clone EPO8 (productivity = 4.05 pg/cell.day) and, interestingly its mRNA copy number was up to 20% lower (**Figure 1**E, **Figure S1**A, **M4**). Except for this extraordinary production clone, we could measure significant correlation (Pearson's $r = 0.99$, $p = 0.001$) between EPO mRNA amount and secreted EPO productivity for all clones (**Figure 1**E, **Figure S1**B), which was similarly high as the correlation observed between mRNA copy number and GFP productivity (Pearson's $r = 0.88$, $p = 0.004$, **Figure S1**C). For both EPO and GFP clones, no significant correlation was found between protein productivity and gene copy number or clone growth rate (**Figure S1**A-C).

We analysed and compared transcriptomic data (Illumina HiSeq, **M1**) 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, **Figure 1**F, **Figure S1**D). We performed pairwise differential expression analysis between each recombinant protein producer clone and their respective host (**Figure S1**E-G, **M5-6**, **Table S2**). Results of this analysis (**Figure 1**E-F) showed, while in EPO producers EPOI2 had the highest number of differentially expressed (B.H. adj. p-value < 0.05, |L2FC| > 1) genes (1137 genes upregulated and 82 genes down regulated), in GFP producers GFP26 clone was the most different clone compared to control cell-line (487 genes down regulated, 466 genes upregulated) based on the number of differentially expressed genes. We also found 45 and 10 common differentially expressed

genes in pairwise comparison of EPO producers and GFP producers with their respective control cell line, respectively (**Figure S1G**).

Pathway enrichment analysis (**Figure 1G**) showed that axonal guidance signaling and oxidative phosphorylation were the two pathways in the IPA pathways database exhibiting constant change 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*-value < 0.05) altered across all EPO producers (**Figure 1G**). Moreover, mitochondrial dysfunction, regulation of eIF4, p7056K signaling and Sirtuin signaling pathways were significantly (B.H. adj. *p*-value < 0.05) altered in more than half of the EPO producers. On the other hand, gene enrichment analysis of GFP producers against their parental HEK293 cell-line did not reveal any common pathways enriched across all the cell lines (**M6**). Instead, we observed generally different phenotypes between clones, with the exception of axonal guidance signaling, oxidative phosphorylation, synaptogenesis signaling and GP6 signaling pathways with a significant change (B.H. adj. *p*-value < 0.05) in more than half of the GFP producers (**Figure 1G**).

To further investigate the direction of change and the downstream effects of gene expression changes in altered signaling pathways, we performed Ingenuity Pathway Analysis (IPA, QIAGEN Inc.) (Krämer et al., 2014), to infer downstream links of gene expression effects. IPA indicated that significantly upregulated (B.H. adj. *p*-value < 0.05) genes related to translational and post-translational pathways were enriched in eIF2 and mTOR signalling pathways in EPO producers compared to the control cell line (**Figure 1G**, **Figure S2**, **M6**). 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**).

Figure 1. Single cell cloning procedure generates EPO and GFP producer clones with an altered metabolism from the host cell-line. (A) Schematic diagram of the cell line development resulting in random integration sites and copy number of the EPO and GFP genes across the genomes. (B) Growth rates of EPO and GFP producer clones. (C) Specific productivity of EPO in different HEK293 clones. (D) 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 different clones. (F) Principal component one in PCA analysis separates producer clones based on their recombinant EPO or GFP protein. (G) Most significantly (B.H. adj. *p*-value < 0.05) enriched pathways in pairwise comparison of EPO or GFP producers against the control hosts.

2.2 Expression of associated genes with mitochondrial ribosomal proteins is positively correlated with EPO production

We sought to find genes that significantly correlate with EPO and GFP production and investigate their roles in the process of protein production (**M4**). For this purpose, we first extracted genes with an average TPM above 10 across all cell lines for both EPO and GFP and then considered positively and negatively correlated ($|$Pearson's $r| > 0.5$, $p < 0.05$) genes. 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 ($Mean_{TPM} > 10$, $|r| > 0.5$, $p < 0.05$) (**Figure 2A, Table S3**). Amongst these genes, 6 genes were found to correlate ($|r| > 0.5$, $p < 0.05$) with both EPO and GFP (**Figure 2B-C**). Their known functions suggested that they might generally be involved in regulation of protein production; the function of MVP is related to generating

ribonucleoproteins (Zheng et al., 2005), B3GNT5 is involved in posttranslational modifications (Togayachi et al., 2001), and for the less characterized gene products of WDR53 and NPAS1, evidence suggests that they act as regulatory elements in cell metabolism (Wu et al., 2016). Analysis of the corresponding biological GO terms (HyperGSA, $p < 0.05$, **M4**) for the EPO-correlating genes showed that many of the genes with a positive correlation ($r > 0.5$) are involved in secretory pathway components such as protein folding, post-translational protein modification and post-Golgi mediated transport (**Figure 2D, M4**). Moreover, a major node enriched with both negative and positive EPO-correlating genes was associated with mitochondrion organization. Conversely, the GO terms (HyperGSA, $p < 0.05$, **M4**) associated with GFP-correlating genes spanned genes that were mostly negatively correlated ($r < -0.5$) with GFP production and covered a wide range of GO terms including RNA catabolic processes, RNA export from the nucleus and translation (**Figure 2E)**. One of the detected enriched pathways by both positive and negative GFP-correlating genes was translation (HyperGSA p-value = 0.029) and interestingly, almost half of the genes in this pathway (4 out of 9) were components of the large subunit of cytosolic ribosomes (RPL3, RPL4, RPL29, RPL36A).

To find the major regulators among the genes highly correlated ($Mean_{TPM} > 10$, $|r| > 0.5$, $p < 0.05$) 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) (**Figure 2F-G**, **Table S3**, **M4**). 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 node degree (k) of the top ten most interactive genes (**Figure 2F-G**: network hubs) ranged from 53 to 144 and 79 to 202 in the networks of EPO- and GFP-correlating genes, respectively (**Table S3)**. Among the top ten interacting genes in the network of EPO-correlating genes, four of them were directly 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 ten 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 ($r < -0.7$, $p \leq 0.05$, $94 < k < 202$) with GFP production. When looking into all the ribosomal genes correlating ($|r| > 0.5$, $p < 0.05$) with EPO production **(Figure S3**, **M4**), besides RPL38, RPS9 and MRPS11 mentioned above, there were additionally three ribosomal genes (MRPS18A, MRPL40 and MRPL49) belonging to mitochondrial ribosomal genes positively correlating ($r > 0.8$, $p < 0.05$) with EPO production **(Figure S3A)**. Similarly in GFP producers, beside RPL3 and RPL4, three ribosomal genes (EEF2, RPL29 and RPL36A, all components of cytosolic ribosome), were significantly negatively correlated ($r < -0.7$, $p \leq 0.04$) with GFP production **(Figure S3B)**.

Among the other (non-ribosomal) top ten interacting genes in the network of EPO-correlating ($|r| > 0.5$, $p < 0.05$) genes, RAC1 ($r = 0.84$, $p = 0.03$, $k = 144$) is a member of the rho family of GTPases and function as molecular switches that regulate a variety of different processes within the cell, including extracellular organization and cell division (Lin et al., 2005; Thomas et al., 2019). PAK2 ($r = 0.84$, $p = 0.03$, $k = 53$), one of the other genes with a high and positive correlation with EPO production, is a direct target of RAC1 and links Rho GTPases to cytoskeleton reorganization and nuclear signaling (May et al., 2014). Furthermore, we observed positive correlation in EPO production and expression of AURKB ($r = 0.84$, $p = 0.03$, $k = 62$), a serine/threonine kinase, which with RAC1 and PAK2 is active in the regulation of chromosomes segregation during mitosis and meiosis through association with microtubules (May et al., 2014). The wide range of targets for these three genes (RAC1, PAK2 and AURKB) and positive correlation of their expression with EPO production (with exactly the same correlation coefficients for all of them) suggests coordinated expression of these genes could be related to EPO production process. We also detected positive correlation of HSPA8 expression with EPO production ($r = 0.84$, $p = 0.03$, $k = 84$). HSPA8 is a well-known chaperone acting in the protein folding process and previously shown to increase protein production by improving the folding processes in CHO cells (Bonam et al., 2019; Lee et al., 2009). In the network of GFP correlated genes we SRSF5 ($r = -0.71$, $p = 0.04$, $k = 87$) and HNRNPA1 ($r = -0.7$, $p = 0.04$, $k = 83$), which are involved in the metabolism of pre-mRNA both shown negative correlation with EPO production (Chen et al., 2018; Yang et al., 2018). This suggests that a reverse trend in expression of translation-associated genes with GFP production could be coupled with upstream pathways involved in mRNA maturation.
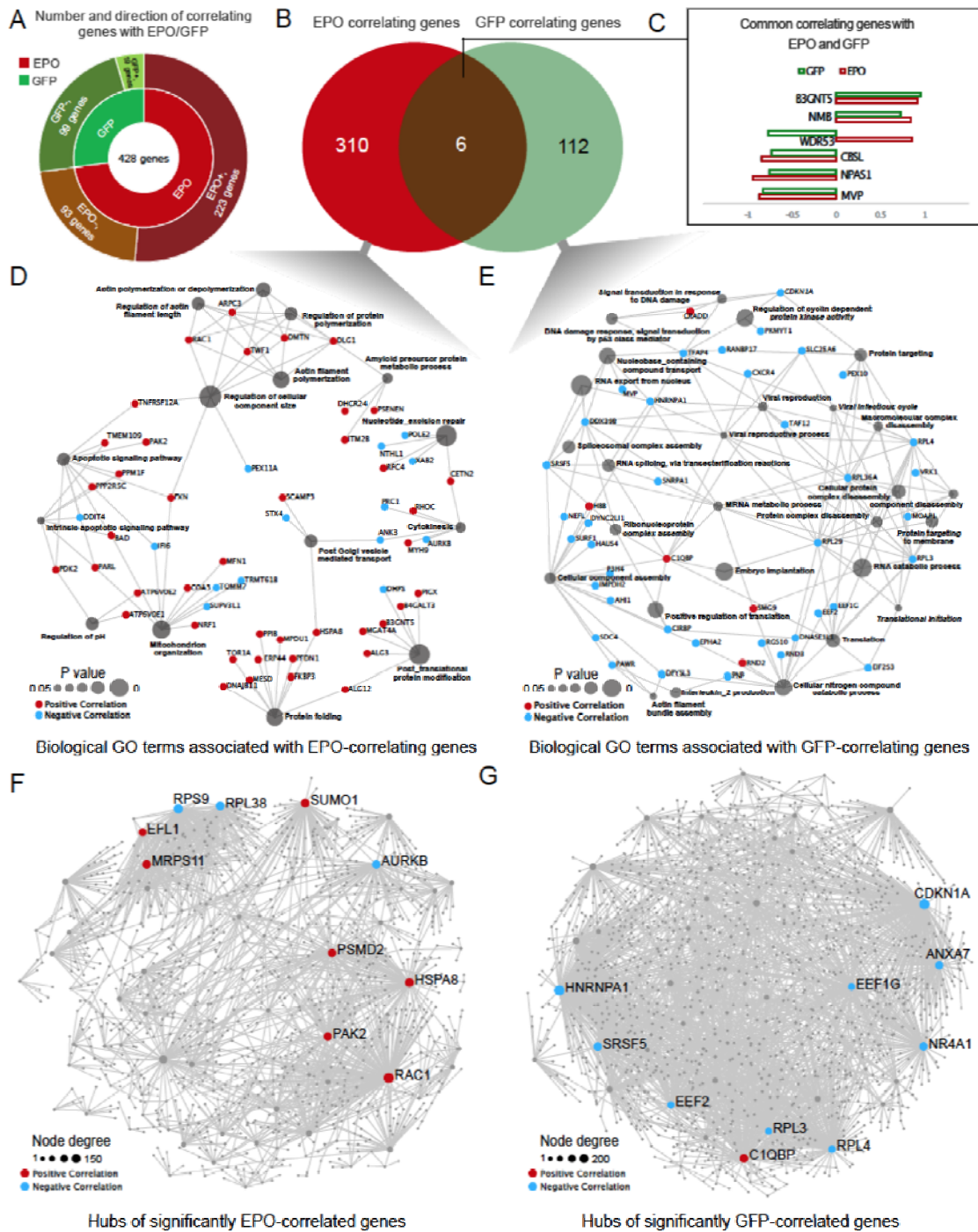
Figure 2. Translational and post-translational genes are strongly correlated with EPO and GFP production. (A) Most significantly ($Mean_{TPM} > 10$, |Pearson's r| > 0.5, p-value < 0.05) correlated genes with EPO or GFP production. Positive and negative signs show positive and negative correlation, respectively (B) Common and specific correlated genes with EPO and GFP production (C) Correlation coefficient of common correlated genes with both EPO and GFP production. (D-E) Enrichment analysis of significantly correlated ($Mean_{TPM} > 10$, |r| > 0.5, p < 0.05) genes with EPO (D) and GFP production (E). (F-G) Top ten hubs with the highest number of interactions (supported by experimental

evidence outlined in STRING database) among significantly correlated ($Mean_{TPM} > 10$, $|r| > 0.5$, $p < 0.05$) genes with EPO (F) and GFP (G) productivity.

2.3 Erythropoietin production requires restructuring of cellular metabolism to meet its energy demands

We sought to understand which changes in protein producers were due to the production of secretory EPO and which were common to all hosts. Principal component analysis of the EPO and GFP transcriptomics data (**M5**) marked differences between EPO and GFP producers (**Figure 1**F), where a complete separation of EPO and GFP clones with the 1st 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 down regulated) differentially expressed genes (adj p-value < 0.05, |Log2 Fold Change| > 1) between EPO and GFP producers. Gene set enrichment analysis (**Figure S4**A, **M7**) showed that with EPO producers, beside gene sets specific for secretory protein production, such as proteins targeting the endoplasmic reticulum or the cell membrane, the oxidative phosphorylation pathway was significantly (B.H. adj. p-value < 0.05) enriched by up regulated genes (**Figure 3**A).

To further investigate the increased oxidative phosphorylation activity among EPO producers, we used IPA to elucidate how expression changes in genes associated with this pathway impart their effect and which mitochondrial membrane complexes are more active in increasing ATP production (**Figure 3**B). Indeed, all differentially expressed genes (B.H. adj. p-value < 0.05) involved in oxidative phosphorylation exhibited increased expression in EPO producers. The up regulated genes are associated with all complexes in the electron transport chain except for complex II, which facilitates the donation of electrons from FADH2. IPA clearly indicated higher activity in NADH dehydrogenase (complex I) to create an electrochemical gradient across the inner mitochondrial membrane followed by an increased activity of complex III to complex V and finally higher ATP production. All genes of mitochondrial origin across the electron transport chain (ETC) complexes exhibited at least a 2-fold increase in their expression (**Figure S4**B: genes with names starting with 'MT-'). Apart from up regulation 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-value < 0.05) expression increase in EPO producers.

Moreover, pathways related to translation and ribosome biogenesis were among the most significantly enriched (B.H. adj. p-value < 0.05) by upregulated genes in both groups of EPO and GFP producers (**Figure 3**A). 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 (**Figure S5**A-B). Although GFP producers did not share common differentially expressed ribosomal genes (B.H. adj. *p*-value < 0.05, **Figure S5C**), EPO producers showed 22 common differentially expressed (*p*-value < 0.05) ribosomal genes with at least 50% increase in their expression in comparison to control, **Figure S5D**). Functionality of these genes is mostly related to SRP-dependent co-translational protein targeting to 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.
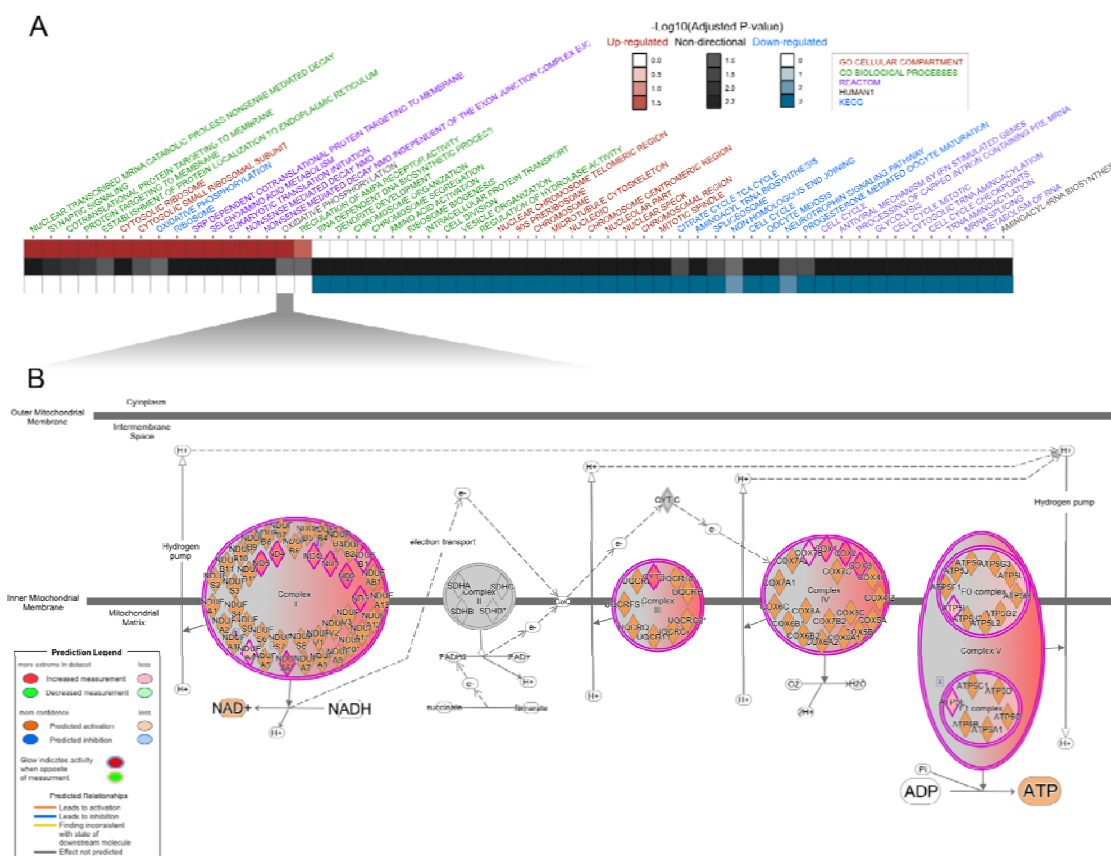

Figure 3. Cellular metabolism is restructured to meet the energy demands of EPO production. (A) Heatmap of enriched (B.H. adj. *p*-value < 0.05) pathways in comparison of EPO producers with GFP producers. Red and blue color ranges show up- and down regulation in EPO producers, respectively. Also, column color shows corresponding gene set collection for enriched pathway (B) The oxidative phosphorylation pathway was significantly (B.H. adj. *p*-value = 0.03) up regulated in EPO producers compared to GFP producers. Detailed visualization of IPA prediction indicates up regulated genes and their role in increasing ATP production.

2.4 Post-translational pathways can limit the EPO protein production rate

Besides the observations regarding differences between mRNA copy number and protein productivity in EPOF21 in comparison to 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**, **M5**). Results of differential expression analysis indicated EPO expression is not significantly changing between clones. Number of down regulated differentially expressed genes (B.H. adj. *p*-value < 0.05, L2FC < -1) varied between 186 (EPOF21 vs. EPOI2) and 491 (EPOF21 vs. EPO8) and number of up regulated genes (B.H. adj. *p*-value < 0.05, L2FC > 1) showed a variation between 113 (EPOF21 vs. EPOI2) and 792 (EPOF21 vs. EPO8) genes (**Figure S6A)**. We also found 44 commonly differentially expressed genes in comparisons of each of EPO producer clones against EPOF21. Gene set enrichment analysis (**M7**) between EPOF21 and other EPO producer cell-lines (**Figure S6B-C**), indicated many common changes between EPOF21 and the other EPO producers were related to post-translational pathways. To obtain an extended gene set 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*-value < 0.05) number of these genes (HyperGSA, **M8**). 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 to 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, comparison of the up regulated pathways in EPOF21 against other EPO producers except EPO8 indicated a marked expression increase in gene sets related to the endoplasmic reticulum and handling of misfolded proteins in EPOF21 (**Figure 4A**). Although we did not observe any significantly down regulated pathways in EPOF21 in comparison to EPOB9 and EPOI2, some pathways related to protein localization to organelles and also cellular macromolecular catabolic processes were down regulated in EPOF21 in comparison to EPO8 and EPO7. This may indicate the potential problematic steps in the process of folding and targeting of secretory proteins in EPO8 and EPO7 that result in the lower EPO production rates in these cell-lines relative to EPOF21.

To further investigate specific differences of EPOF21 with other EPO producers, we analysed which differentially expressed genes exhibited a significant and solid pattern of change (B.H adj. p-value < 0.05, L2FC > 1, $Mean_{TPM}$ > 10). Of the 19 genes that displayed a consistent pattern of change, 5 were up regulated and 14 down regulated (**Figure 4**B). Among the upregulated genes, ATF6B is a transcription factor active under ER stress conditions due to accumulation of unfolded proteins (Thuerauf et al., 2004), whereas FN1 and INA have 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). Down regulated 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-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).
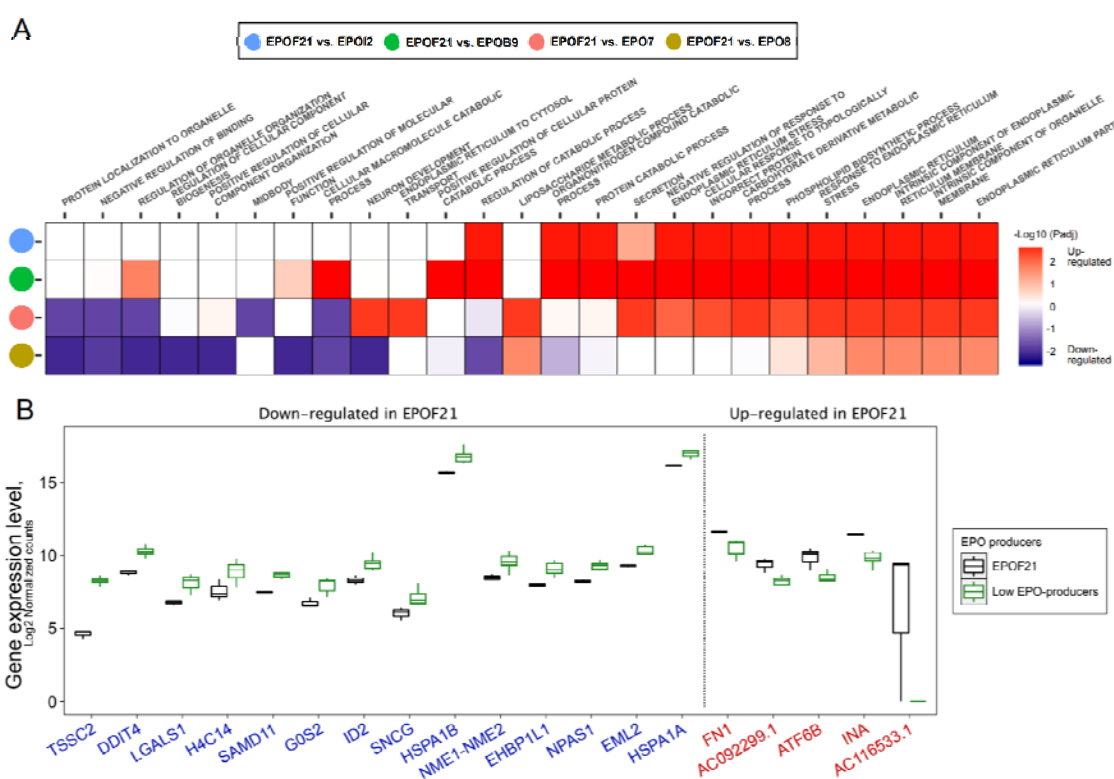

Figure 4. Post-translational pathways are different between EPO producer clones. (A) Significantly enriched GO terms in pairwise comparisons of EPOF21 and other EPO producers using

GO slim secretion. (B) Gene expression levels of the most significant (B.H. adj. p-value < 0.05) consistently differentially expressed genes between EPOF21 and other EPO producers.

3. Discussion

In the present study, we generated two groups of clones producing either EPO or GFP protein at different levels (**Figure 1**A). EPO and GFP were chosen as well studied model proteins with different characteristics such as 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 1**F), 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 production titers as well as different gene expression profiles (**Figure 1**C-E, **Figure S1**A), both transcription and translation were very well synced (**Figure S1**B). Thus, all of the clones except EPOF21, a clone with over 3-fold higher production levels compared to the next highest EPO producer (**Figure 1**E), had an almost equal ratio of secretion of the recombinant protein versus transcription of the recombinant gene transcripts (**Figure 1**E). The high increase of this ratio in EPOF21 (in relation to the other clones) indicated that major translational and post-translational processes were affected. This suggests that, whereas 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 is a highly useful candidate to further pinpoint major limiting parameters within secretory protein production, which can in general help to improve the overall productivity of cell factories.

Analysis of transcriptomic data between protein producer hosts and control parental cells showed up regulation of oxidative phosphorylation in all clones regardless of transgene or level of recombinant protein production (**Figure 1**G), which indicates a high energy demand is required to support transgene expression. Moreover, significant upregulation of genes linked to oxidative phosphorylation was observed when comparing EPO producers with GFP producers (**Figure 3**B), suggesting that EPO clones have an even higher energy demand imposed by post-translatory secretory pathways (Gutierrez et al., 2020). Simultaneous up regulation of genes from both the mitochondria and nucleus highlight that the increase in energy production is not governed merely by the change in the mitochondrial genome, but also by the upstream regulatory pathways of the cell (**Figure S4**B). The demand for more energy in EPO producers compared to GFP producers, could be related to the intrinsic molecular differences between EPO and GFP or could be 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).

We note that the EPO protein itself acts as a signaling molecule and could therefore 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 erythropoietin receptor (EpoR) which is generally restricted to erythroid progenitor cells. Evidence supporting protein expression of EpoR and/or EPO activity via EpoR interaction in other (non-hematopoietic) cell types is lacking (Elliott and Sinclair, 2012), with the exception of skeletal muscle (Rundqvist et al., 2009). Furthermore, a study of EpoR expression and activity by Ott et al. (Ott et al., 2015) detected no EpoR protein in HEK293 cells unless they were transfected with an EpoR overexpression vector. In line with this the EPOR expression in our dataset is not correlated with EPO production (**Figure S7**A) and furthermore there is no significant difference in the expression of this gene between EPO producers and GFP producers and also between EPO producers and their control cell-line (**Figure S7**A-D). Moreover, we didn't detect a solid and significant difference (B.H adj. *p*-value < 0.05, L2FC > 1) in expression of downstream genes in the EPOR signaling pathway (**Figure S7**E) in comparing the expression of these genes between EPOF21 and other EPO producers (**Figure S7**F). 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, though some confounding effect from signaling cannot be ruled out entirely.

In EPO-clones, significant up regulation of genes belonging to the eIF2 and mTOR signaling pathways could lead to activation of translation (**Figure S2**A). 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 pre-initiation complex (Hinnebusch, 2011; Schmitt et al., 2010; Stolboushkina and Garber, 2011; Wek et al., 2006). Likewise, activation of mTORC1 module in mTOR signaling pathway (**Figure S2**B) could promote protein synthesis and profoundly increase cellular ATP level by controlling mitochondrial biogenesis (Laplante and Sabatini, 2009). However, when looking into associated GO terms with all positively and negatively correlated genes with recombinant protein production, only GFP production showed a positive correlation with translation-associated genes, whereas EPO production was positively correlated with members of protein folding and post-translational modification pathways (**Figure 2**D). We also observed a positive correlation between expression of genes associated with apoptosis signaling and EPO productivity, which emphasize how protein production pressure could stimulate pathological ER stress and activate apoptotic related pathways (Iurlaro and Muñoz-Pinedo, 2016; Sano and Reed, 2013). Three genes RAC1 and

PAK2 detected between the top 10 genes with the highest number of interactions (based on experimental evidence) with other correlating genes (**Figure 2**F, **M4**), has frequently been reported in previous studies as a module that greatly regulates cytoskeleton reorganization, intracellular trafficking and apoptosis (Chi et al., 2013; Coleman and Olson, 2002; Croisé et al., 2014; Embade et al., 2000). Considering their upstream regulatory roles in controlling cell behaviour and the fact that all these three genes have a significant positive trend of expression with increase in EPO production, it is very likely that upregulation of these genes in high producer cells is associated with their potential role in boosting the level of EPO production.

Among other top ten interacting genes in the network of EPO correlating genes (**Figure 2**F), HSPA8, a member of the heat shock proteins family A, has also been previously reported as a target for improving recombinant protein production in CHO cells (Lee et al., 2009). Likewise, RPL38 and RPS9 (cytosolic ribosomal) and MRPS4 (mitochondrial ribosomal) genes with negative and positive trends of expression with EPO production, respectively, were found among the top ten EPO correlating genes with highest interactions with other genes. Due to similar observations for cytosolic ribosomal genes in GFP producers (**Figure 2G**) and since a negative trend of expression for ribosomal genes with increase in protein production was unexpected, we further investigated this observation by analysing the expression trends for each ribosomal gene that significantly correlated with protein production (**Figure S3**). Remarkably, we detected similar negative and positive trends of expression across all cytosolic and mitochondrial ribosomal genes, respectively. This consistent pattern of change between ribosomal genes belonging to two different compartments indicates the presence of an upstream regulation in favor of a specific need by the cell. The observations that EPO production imposes a high energy burden on the cells (**Figure 1**G, **Figure 3**A-B and **Figure S5**) and the elevated level of expression of mitochondrial ribosomal genes with higher rates of EPO production and also lack of such observation in GFP producers, suggests that secretory EPO-producing cells allocate more protein resources to produce mitochondrial ribosomes and increase the production of mitochondrial proteins involved in the energy production process through oxidative phosphorylation. These results further highlight the importance of energy metabolism for secreting higher levels of EPO and could explain a general strategy employed by cells to support higher levels of secretory protein production by increasing energy production and compensating for this increase by downregulating the production of cytosolic ribosomes.

We also detected a negative trend in the expression of some cytosolic ribosomal genes (**Figure 2**E, **Figure 2**G and **Figure S3**B) 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 ribosomal components. The latter is known as ribosome heterogeneity (Genuth and Barna, 2018a) and 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, 2018b). It has previously been shown that specialized ribosomes can preferentially translate different subsets of mRNAs (Shi et al., 2017). Our observations highlight the presence of both up regulated and down regulated ribosome-related pathways in EPO and GFP producers (**Figure 3**A) and pinpoint transgene-specific correlations between expression levels of different ribosomal genes and the respective recombinant protein titers (**Figure 2**D-E, **Figure S3**). 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 of high producer EPOF21 clone with other EPO producers, we observed that the majority of differences among the EPO producer clones are in the post-translational steps, including protein folding, post-translational modifications and handling of misfolded proteins (**Figure 4**, **Figure S6**B-C). Indeed, in accordance with the increase of EPO transcripts, post-translational and ER-related pathways in EPOF21 were upregulated (**Figure 4**A). However, the EPO8 clone, with almost the same level of EPO transcripts as EPOF21 (**Figure 1**E), did not display an increased activity in some of ER-associated pathways. 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.

To find regulatory elements behind 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 up regulated genes in EPOF21 was ATF6B (**Figure 4**B). 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 (Iurlaro and Muñoz-Pinedo, 2016; Thuerauf et al., 2004). 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 to 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). So, downregulation of the stress-inducible chaperones HSPA1A and HSPA1B in EPOF21 (**Figure 4**B) could suggest moderate level of ER stress in the this clone, capable of supporting a high level of protein secretion, potentially governed by the higher expression of ATF6B. So, the difference in the expression levels of ATF6B transcription factor between EPOF21 and the low EPO-producers suggests that tuning the expression of this gene could potentially be a useful strategy for controlling EPO production in HEK293F cells.

In conclusion, the present study offers important insights into transcriptomic changes during secretory EPO and non-secretory GFP protein production as well as key parameters influencing the different rates of protein production across a variety of protein producers. The results are thus valuable for improved understanding of the biology behind protein secretion in mammalian cells and also the behaviour 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.

Acknowledgements

Author Contributions

Conceptualization, JN, JLR, and JR; Methodology, RS, JZ, MM, JLR and JR; Formal analysis, RS, JZ and MM; Investigation, RS, MM, JZ, VC, RR, NW, JLR and JR; Writing – Original Draft, RS, JZ, MM, JLR and JR; Writing – Review & Editing, RS, JZ, MM, DH, JN, LG, VC, JLR and JR; Visualization, RS, JZ, LG and JLR; Supervision, DH, TS, JN, JLR and JR, Funding Acquisition, JN and JR.

Declaration of Interests

The authors declare no competing interests.

Figure legends

Figure 1: Single cell cloning procedure generates EPO and GFP producer clones with an altered metabolism from the host cell-line.

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

Figure 3: Cellular metabolism is restructured to meet the energy demands of erythropoietin production.

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

Tables with legends

STAR Methods

Key Resources Table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cell lines		
293-F		https://www.thermofisher.com
Freestyle 293-F		https://www.thermofisher.com
Experimental Reagents and Resources		
Illumina HiSeq		www.illumina.com
Targeted Locus Amplification	((de Vree et al., 2014))	www.cergentis.com
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
Trinity	(Grabherr et al., 2011)	github.com/trinityrnaseq
IPA®	QIAGEN Inc.	digitalinsights.qiagen.com/
MSigDB database	(Liberzon et al., 2011)	www.gsea-msigdb.org/gsea/msigdb

Contact for Reagent and Resource Sharing

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).

Method Details

M1. Cell cultivation and isolation of single clones

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 limited 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 of cells 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™).

M2. Productivity measurements

Specific 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_HH-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).

M3. Genome copy number estimation of GFP and EPO using TLA technology

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 number of plasmid integrations into the genome, number of fusion reads and ratio between coverage of transgene and surrounding genomic region.

M4. 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 correlation coefficient, 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 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.

M5. RNA preparation and 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 x 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*-values 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.

M6. 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).

M7. 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 (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*-values 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).

M8. 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*-values were adjusted to control for the false discovery rate (*FDR*) with the Benjamini-
Hochberg procedure.

M9. Software

All computational analyses were performed using R v3.6 (www.r-project.org) and Matlab
R2017b (www.mathworks.com). Code and datasets to reproduce the figures presented here
as well as all analysis outputs, are available on GitHub repository:
https://github.com/SysBioChalmers/EPO_GFP. Data files too large to host on GitHub were
deposited on Zenodo: https://zenodo.org/record/4004264#.X0iqstMzbuM

Supplemental Information titles and legends

Figure S1: Correlation analysis between growth rate, protein productivity, gene and mRNA copy number.

Figure S2: eIF2 signaling and mTOR signaling pathways show activation of translation in EPO producers in comparison to host parental cell-line.

Figure S3: Cytosolic and mitochondrial ribosomal genes have a reverse trend of expression in EPO producers.

Figure S4: Increase in energy production in EPO clones is governed by the upstream regulatory pathways of the cell.

Figure S5: Ribosomal genes show recombinant protein specific patterns of expression.

Figure S6: Gene enrichment analysis between EPOF21 and other low EPO producers highlights differences in post-translational pathways.

Figure S7: The difference in Expression of erythropoietin receptor (EPOR) gene is not significant between EPO producers vs. control cell-line and also EPO producers vs. GFP producers.

Table S1: Table of raw counts for transcriptomics analysis

Table S2: List of enriched pathways in pairwise comparison of EPO producers against EPO control host.

Table S3: List of significantly correlated genes with EPO and GFP production and statistics for EPO- and GFP-correlating genes networks

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.

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.

Barondeau, D.P., Putnam, C.D., Kassmann, C.J., 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.

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. Pharm. Biotechnol. *21*, 226–235.

Bonam, S.R., Ruff, M., and Muller, S. (2019). HSPA8/HSC70 in Immune Disorders: A Molecular Rheostat that Adjusts Chaperone-Mediated Autophagy Substrates. Cells *8*.

Bray, N.L., Pimentel, H., Melsted, P., and Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. *34*, 525–527.

Chen, Y., Huang, Q., Liu, W., Zhu, Q., Cui, C.-P., Xu, L., Guo, X., Wang, P., Liu, J., Dong, G., et al. (2018). Mutually exclusive acetylation and ubiquitylation of the splicing factor SRSF5 control tumor growth. Nat. Commun. *9*, 2464.

Chi, X., Wang, S., Huang, Y., Stamnes, M., and Chen, J.-L. (2013). Roles of rho GTPases in intracellular transport and cellular transformation. Int. J. Mol. Sci. *14*, 7089–7108.

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.

Coleman, M.L., and Olson, M.F. (2002). Rho GTPase signalling pathways in the morphological changes associated with apoptosis. Cell Death Differ. *9*, 493–504.

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.

Croisé, P., Estay-Ahumada, C., Gasman, S., and Ory, S. (2014). Rho GTPases, phosphoinositides, and actin: a tripartite framework for efficient vesicular trafficking. Small GTPases *5*, e29469.

Datta, P., Linhardt, R.J., and Sharfstein, S.T. (2013). An 'omics approach towards CHO cell engineering. Biotechnology and Bioengineering *110*, 1255–1271.

Davy, A.M., Kildegaard, H.F., and Andersen, M.R. (2017). Cell Factory Engineering. Cell Syst *4*, 262–275.

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.

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.

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.

Elliott, S., and Sinclair, A.M. (2012). The effect of erythropoietin on normal and neoplastic cells. Biologics *6*, 163–189.

Embade, N., Valerón, P.F., Aznar, S., López-Collazo, E., and Lacal, J.C. (2000). Apoptosis induced by Rac GTPase correlates with induction of FasL and ceramides production. Mol. Biol. Cell *11*, 4347–4358.

Erbel-Sieler, C., Dudley, C., Zhou, Y., Wu, X., Estill, S.J., Han, T., Diaz-Arrastia, R., Brunskill, E.W., Potter, S.S., 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.

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 Systems Biology and Applications *3*, 22.

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. Bioprocess Biosyst. Eng.

Gargalovic, P.S., Imura, M., Zhang, B., Gharavi, N.M., Clark, M.J., Pagnon, J., 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.

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.

Genuth, N.R., and Barna, M. (2018a). The Discovery of Ribosome Heterogeneity and Its Implications for Gene Regulation and Organismal Life. Mol. Cell *71*, 364–374.

Genuth, N.R., and Barna, M. (2018b). Heterogeneity and specialized functions of translation machinery: from genes to organisms. Nat. Rev. Genet. *19*, 431–452.

Goh, J.B., and Ng, S.K. (2017). Critical Reviews in Biotechnology Impact of host cell line choice on glycan profile Impact of host cell line choice on glycan profile.

Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. *29*, 644–652.

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.

Hinnebusch, A.G. (2011). Molecular mechanism of scanning and start codon selection in eukaryotes. Microbiol. Mol. Biol. Rev. *75*, 434–467, first page of table of contents.

Iurlaro, R., and Muñoz-Pinedo, C. (2016). Cell death induced by endoplasmic reticulum

stress. FEBS J. *283*, 2640–2652.

Jaeken, J., Lefeber, D., and Matthijs, G. (2015). Clinical utility gene card for: ALG1 defective congenital disorder of glycosylation. Eur. J. Hum. Genet. *23*.

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.

Kaufman, R.J., and Popolo, L. (2018). Chapter 5 - Protein Synthesis, Processing, and Trafficking. In Hematology (Seventh Edition), R. Hoffman, E.J. Benz, L.E. Silberstein, H.E. Heslop, J.I. Weitz, J. Anastasi, M.E. Salama, and S.A. Abutalib, eds. (Elsevier), pp. 45–58.e1.

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.

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.

Kol, S., Kallehauge, T.B., Adema, S., and Hermans, P. (2015). Development of a VHH-Based Erythropoietin Quantification Assay. Mol. Biotechnol. *57*, 692–700.

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.

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–A467.

Krämer, A., Green, J., Pollard, J., Jr, and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics *30*, 523–530.

Kuriakose, A., Chirmule, N., and Nair, P. (2016). Immunogenicity of Biotherapeutics: Causes and Association with Posttranslational Modifications. J Immunol Res *2016*, 1298473.

Laplante, M., and Sabatini, D.M. (2009). mTOR signaling at a glance. J. Cell Sci. *122*, 3589–3594.

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.

Lee, Y.Y., Wong, K.T.K., Tan, J., Toh, P.C., Mao, Y., Brusic, V., and Yap, M.G.S. (2009). Overexpression of heat shock proteins (HSPs) in CHO cells for extended culture viability and improved recombinant protein production. J. Biotechnol. *143*, 34–43.

Liang, C., Chiang, A.W.T., Hansen, A.H., Arnsdorf, J., 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. Curr Res Biotechnol *2*, 22–36.

Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., and Mesirov, J.P. (2011). Molecular signatures database (MSigDB) 3.0. Bioinformatics *27*, 1739–1740.

Lin, M., Van Golen, C., and Van Golen, K.L. (2005). RhoC and Rac1 GTPases in prostate cancer cell motility, invasion, and metastasis. Cancer Res. *65*, 239–240.

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.

Malm, M., Saghaleyni, R., Lundqvist, M., and Giudici, M. (2020). Evolution from adherent to suspension–systems biology of HEK293 cell line development. bioRxiv.

May, M., Schelle, I., Brakebusch, C., Rottner, K., and Genth, H. (2014). Rac1-dependent recruitment of PAK2 to G2 phase centrosomes and their roles in the regulation of mitotic entry. Cell Cycle *13*, 2211–2221.

Meuris, L., Santens, F., Elson, G., Festjens, N., Boone, M., Dos Santos, A., 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.

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.

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.

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.

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.

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

Rahimpour, A., Vaziri, B., Moazzami, R., Nematollahi, L., Barkhordari, F., Kokabee, L., Adeli, 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.

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*.

Rundqvist, H., Rullman, E., Sundberg, C.J., Fischer, H., Eisleitner, K., Ståhlberg, M., Sundblad, P., Jansson, E., and Gustafsson, T. (2009). Activation of the erythropoietin receptor in human skeletal muscle. Eur. J. Endocrinol. *161*, 427–434.

Sano, R., and Reed, J.C. (2013). ER stress-induced cell death mechanisms. Biochim. Biophys. Acta *1833*, 3460–3470.

Schmitt, E., Naveau, M., and Mechulam, Y. (2010). Eukaryotic and archaeal translation initiation factor 2: a heterotrimeric tRNA carrier. FEBS Lett. *584*, 405–412.

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.

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.

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.

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.

Singh, P., Carraher, C., and Schwarzbauer, J.E. (2010). Assembly of fibronectin extracellular matrix. Annu. Rev. Cell Dev. Biol. *26*, 397–419.

Soneson, C., Love, M.I., and Robinson, M.D. (2015). Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res. *4*, 1521.

Stolboushkina, E.A., and Garber, M.B. (2011). Eukaryotic type translation initiation factor 2: structure–functional aspects. Biochemistry .

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.

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.

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. Biotechnol. *58*, 45–54.

Tejwani, V., Andersen, M.R., Nam, J.H., and Sharfstein, S.T. (2018). Glycoengineering in CHO Cells: Advances in Systems Biology. Biotechnology Journal *13*, 1700234.

Thomas, P., Pranatharthi, A., Ross, C., and Srivastava, S. (2019). RhoC: a fascinating journey from a cytoskeletal organizer to a Cancer stem cell therapeutic target. J. Exp. Clin. Cancer Res. *38*, 328.

Thomson, E., Ferreira-Cerca, S., and Hurt, E. (2013). Eukaryotic ribosome biogenesis at a glance. J. Cell Sci. *126*, 4815–4821.

Thuerauf, D.J., Morrison, L., and Glembotski, C.C. (2004). Opposing Roles for ATF6α and ATF6β in Endoplasmic Reticulum Stress Response Gene Induction. J. Biol. Chem. *279*, 21078–21084.

Togayachi, A., Akashima, T., Ookubo, R., Kudo, T., Nishihara, S., Iwasaki, H., Natsume, A., Mio, H., Inokuchi, J., Irimura, T., et al. (2001). Molecular cloning and characterization of UDP-GlcNAc:lactosylceramide beta 1,3-N-acetylglucosaminyltransferase (beta 3Gn-T5), an essential enzyme for the expression of HNK-1 and Lewis X epitopes on glycolipids. J. Biol. Chem. *276*, 22032–22040.

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.

de Vree, P.J.P., de Wit, E., Yilmaz, M., van de Heijning, M., Klous, P., Verstegen, M.J.A.M., Wan, Y., Teunissen, H., Krijger, P.H.L., Geeven, G., et al. (2014). Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. Nat. Biotechnol. *32*, 1019–1025.

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.

Wang, H., Marcišauskas, S., Sánchez, B.J., Domenzain, I., Hermansson, D., Agren, R., Nielsen, J., and Kerkhoven, E.J. (2018). RAVEN 2.0: A versatile toolbox for metabolic network reconstruction and a case study on Streptomyces coelicolor. PLoS Comput. Biol. *14*, e1006541.

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.

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.

Wu, D., Su, X., Potluri, N., Kim, Y., and Rastinejad, F. (2016). NPAS1-ARNT and NPAS3-ARNT crystal structures implicate the bHLH-PAS family as multi-ligand binding transcription factors. Elife *5*.

Xu, X., and Qi, L.S. (2019). A CRISPR–dCas Toolbox for Genetic Engineering and Synthetic Biology. J. Mol. Biol. *431*, 34–47.

Yang, S., Jia, R., and Bian, Z. (2018). SRSF5 functions as a novel oncogenic splicing factor and is upregulated by oncogene SRSF3 in oral squamous cell carcinoma. Biochim. Biophys. Acta Mol. Cell Res. *1865*, 1161–1172.

Zheng, C.-L., Sumizawa, T., Che, X.-F., Tsuyama, S., Furukawa, T., Haraguchi, M., Gao, H., Gotanda, T., Jueng, H.-C., Murata, F., et al. (2005). Characterization of MVP and VPARP assembly into vault ribonucleoprotein complexes. Biochem. Biophys. Res. Commun. *326*, 100–107.

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.

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

Supplementary Information
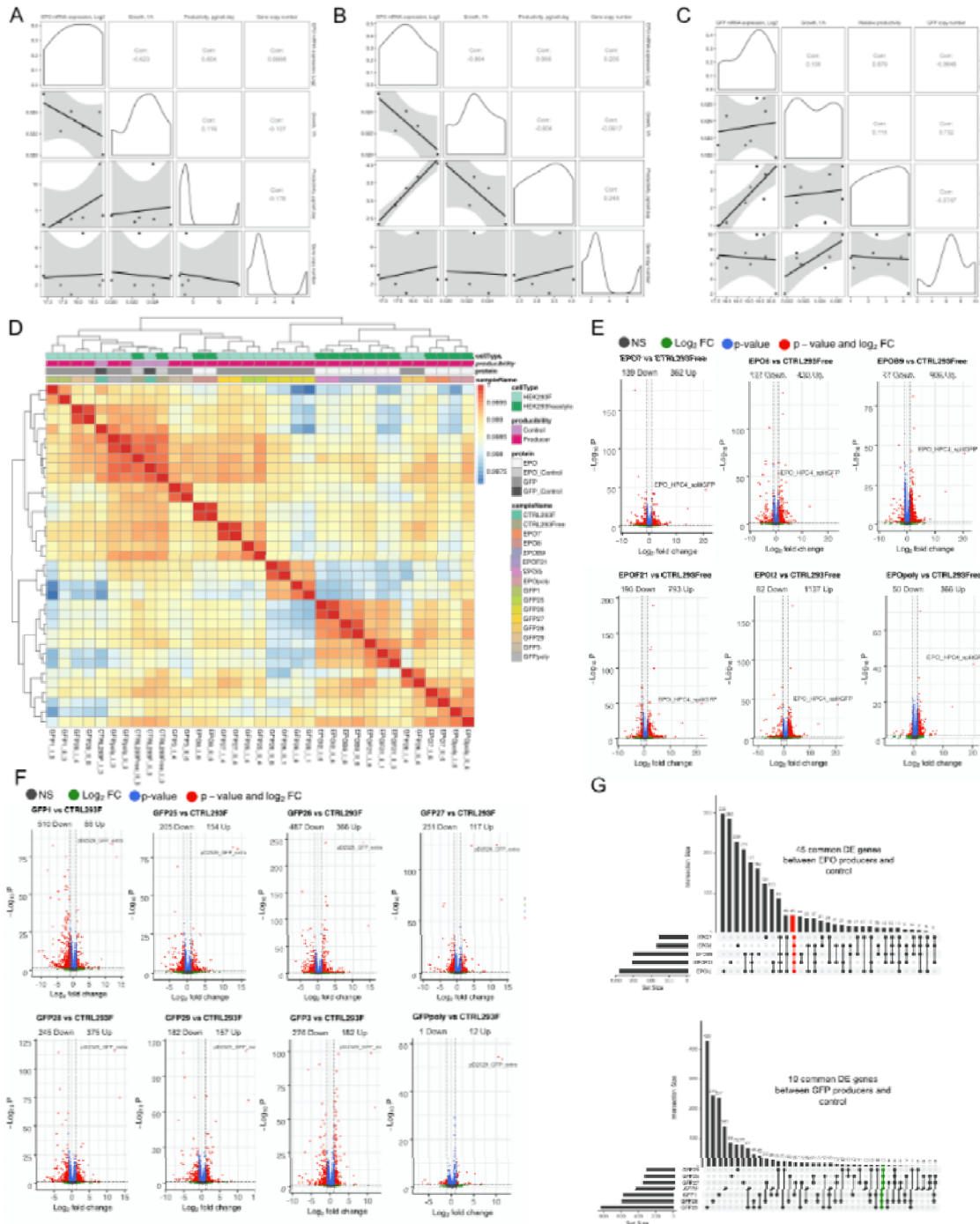
Figure S1. Correlation analysis between growth rate, protein productivity, gene and mRNA copy number. (A) EPO producer clones and (B) EPO producers without the highest producing EPOF21 clone. (C) Correlation analysis between growth rate, protein productivity, gene and mRNA copy number for GFP producer clones. (D) Samples correlation plot for all replicates. (E) Differentially expressed genes (B.H. adj. *p*-value < 0.05, | L2FC | > 1) between EPO producers vd. control (E) and GFP producers vs control (F). (G) Common differentially expressed genes (*p*-value < 0.05, | L2FC | >

1) in pairwise comparison of EPO producer clones wirth control (45 genes) and pairwise comparison of GFP producer clones wirth control (10 genes).
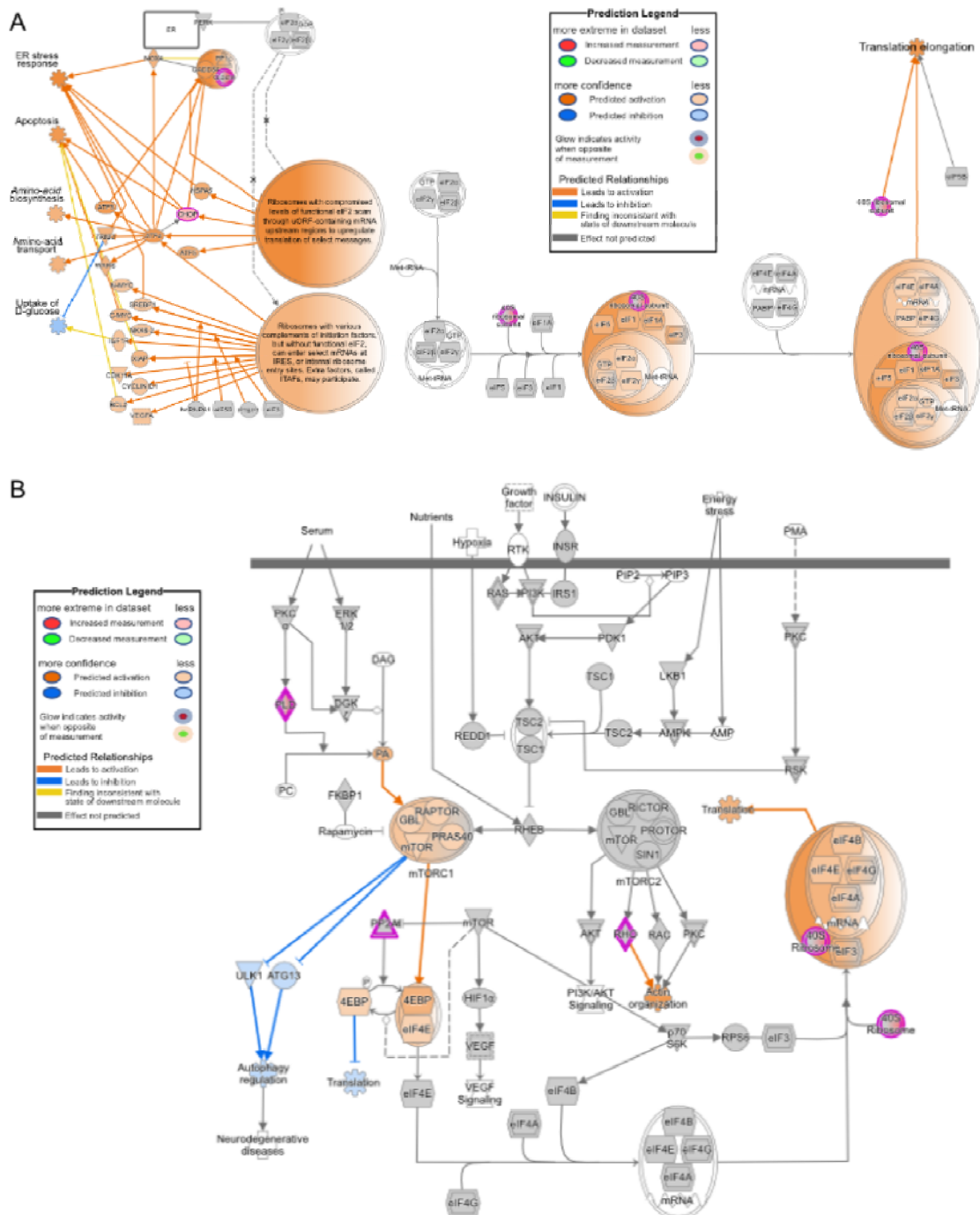
Figure S2. **eIF2 signaling and mTOR signaling pathways show activation of translation in EPO producers in comparison to host parental cell-line.** Detailed visualization of mTOR signaling and

Eukaryotic Initiation Factor 2 (eIF2) signaling pathways that were the most significantly differentially up regulated pathways across all EPO producer clones compared to control cell-line.
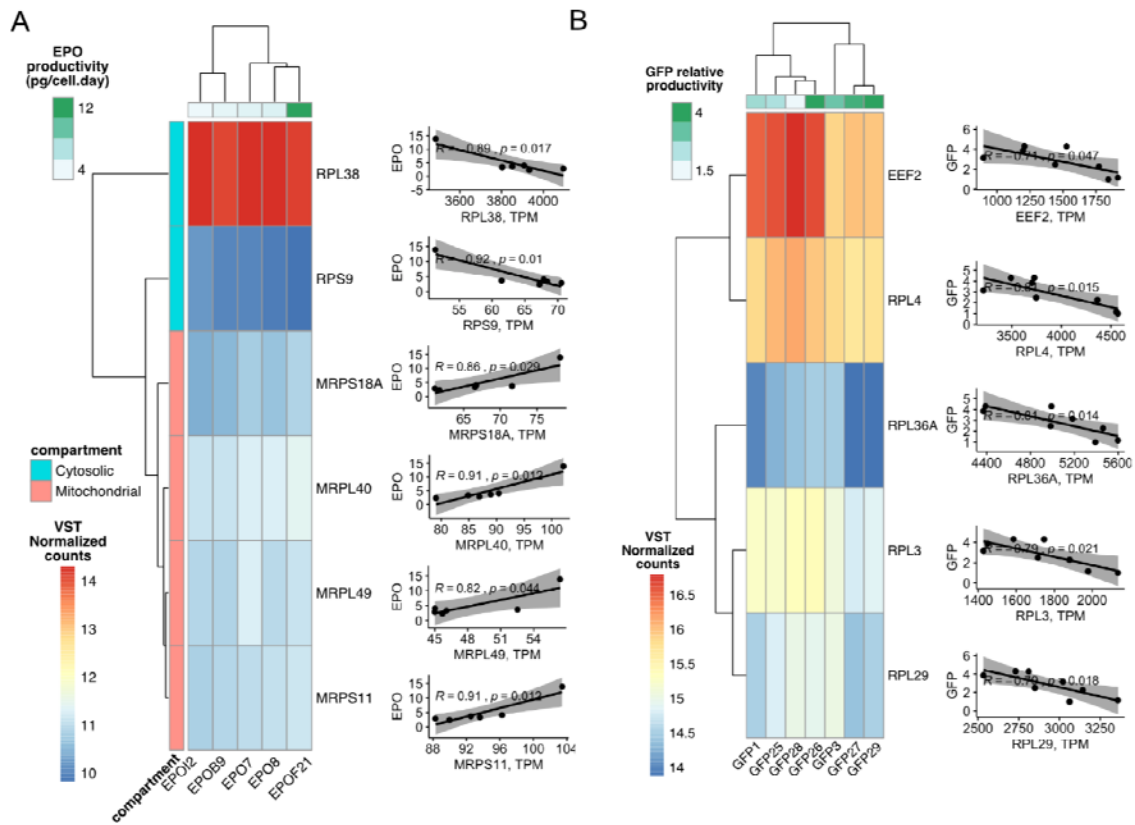
Figure S3. Cytosolic and mitochondrial ribosomal genes have a reverse trend of expression in EPO producers. (A) RPL38 and RPS9 ribosomal genes have significant negative correlation with EPO production and both are components of cytosolic ribosomes while MRPS18A, MRPL40, MRPL49 and MRPS11 are components of mitochondrial ribosomes and all have significant positive correlation with EPO production. (B) All ribosomal genes with significant correlation with GFP are components of cytosolic ribosome and all have negative correlation with GFP production.
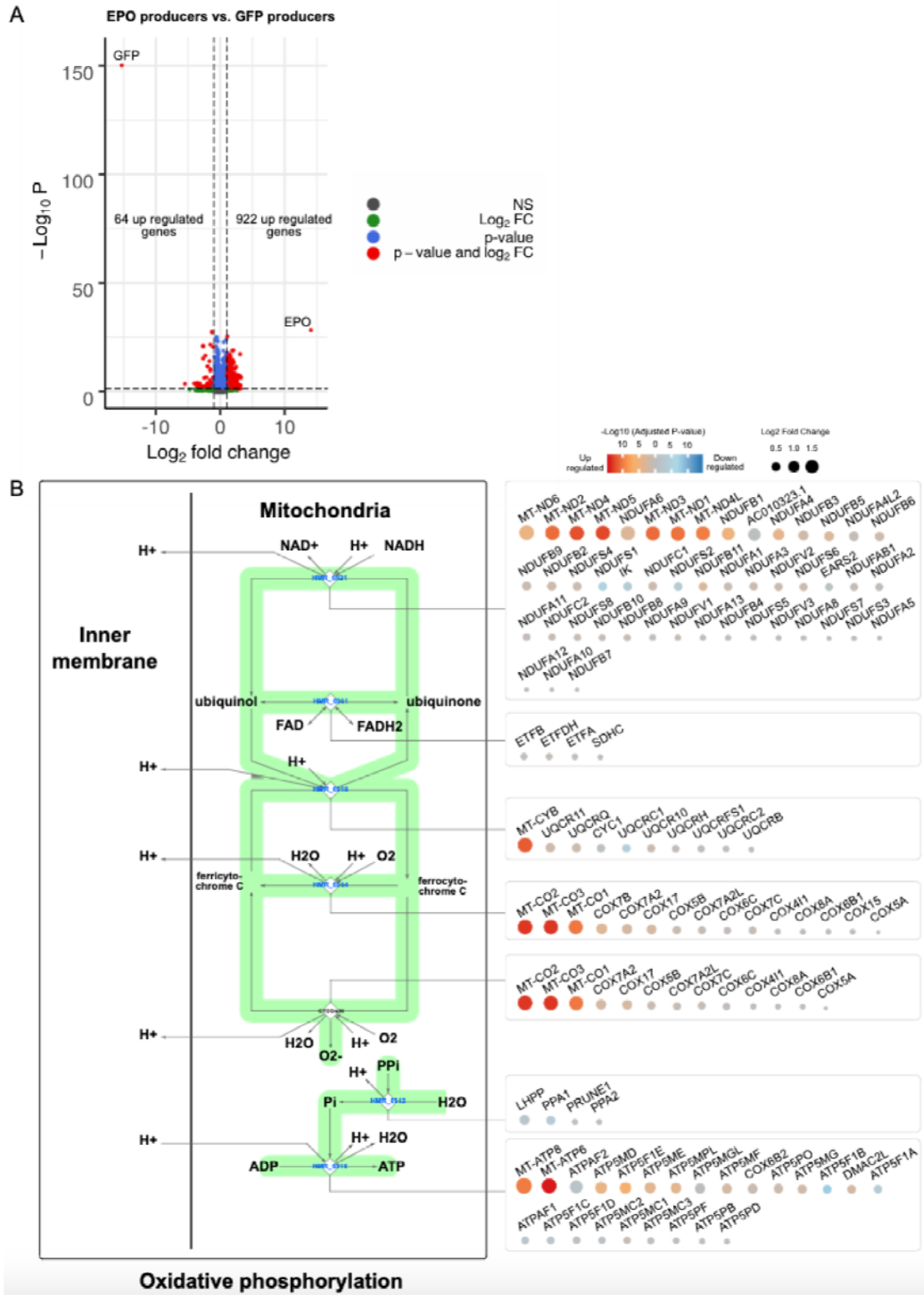
Figure S4. Increase in energy production in EPO clones is governed by the upstream regulatory pathways of the cell. Apart from up regulation of mitochondrial encoded genes, other genes like NDUFA6, NDUFB1 and NDUFA4 in complex I, COX7B and COX7A2 in complex III and

ATP5MD and ATP5F1E in complex V have a significant (B.H. adj. *p*-value < 0.05) expression increase in EPO producers compared to GFP producers.
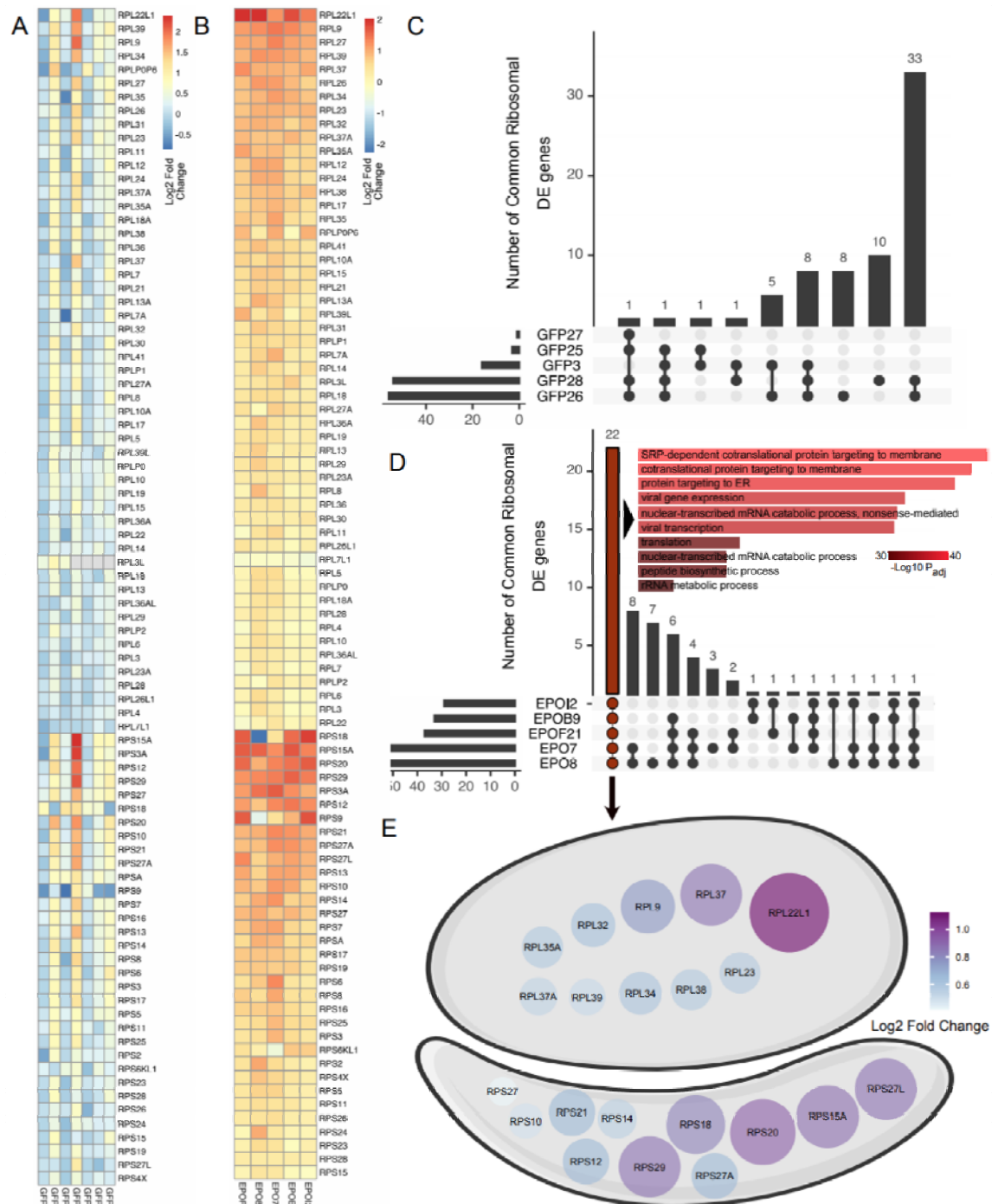
Figure S5. Ribosomal genes show recombinant protein specific patterns of expression. (A) Pairwise comparison of ribosomal genes between GFP producers versus control. (B) Pairwise comparison of ribosomal genes between EPO producers versus control. (C) Common significantly differentially expressed (B.H. adj. *p*-value < 0.05, | L2FC | > 0.58) ribosomal genes across all GFP

producers compared to control. (D) 22 common significantly differentially expressed (B.H. adj. *p*-value < 0.05, | L2FC | > 0.58) ribosomal genes across all EPO producers in comparison to control. Most significantly associated pathways with these genes belong to SRP (signal recognition particle) dependent cotranslational protein targeting to the membrane. (E) Fold change of common differentially expressed ribosomal genes between EPO producers versus control.
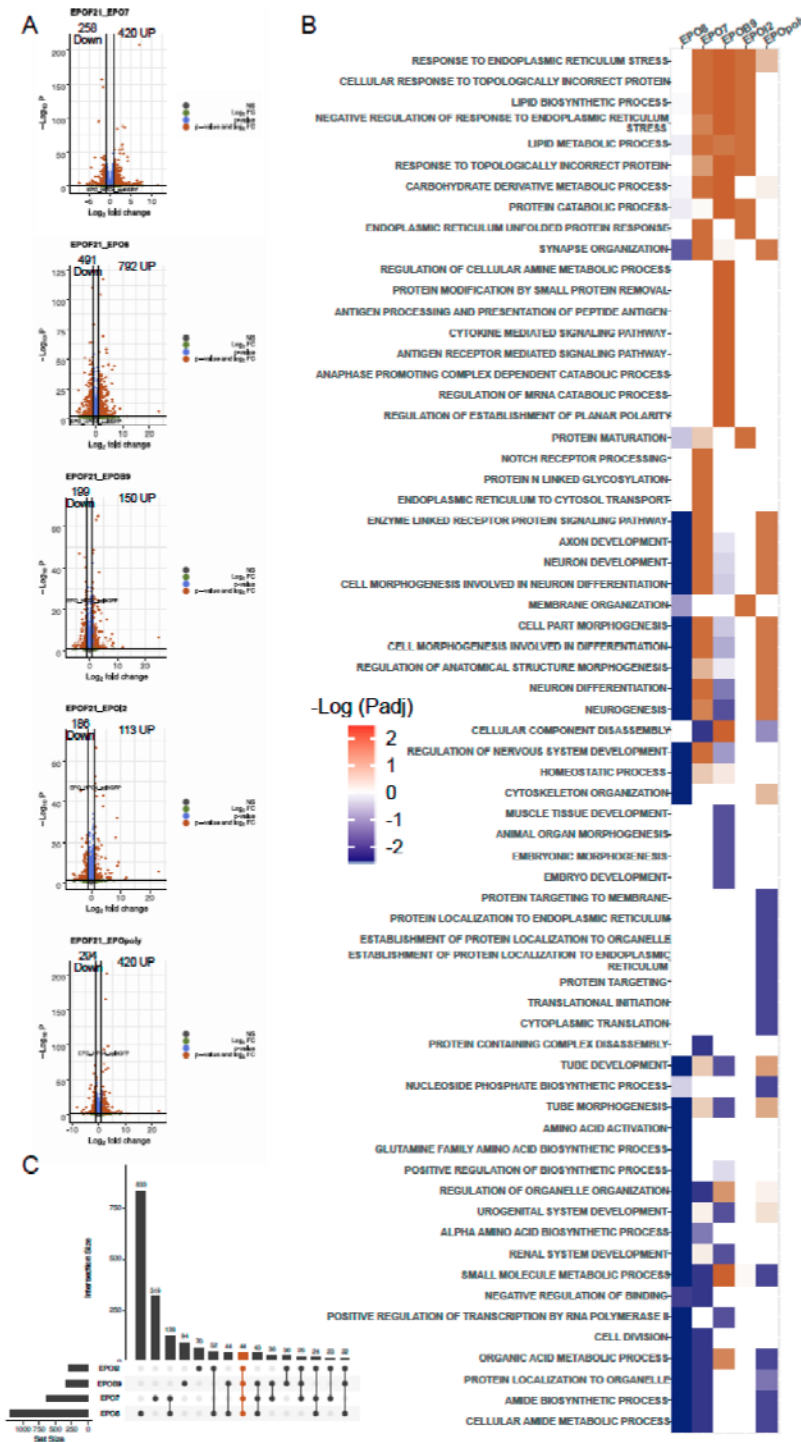
Figure S6. Gene enrichment analysis between EPOF21 and other low EPO producers highlights differences in post-translational pathways. (A) Differentially expressed genes between EPO producers vs. EPOF21 high producer clone. EPO expression is not highly changed between EPO producers (B) Enriched pathways for differentially expressed genes in pairwise comparison of EPOF21 with each of low EPO producer clones. (C) There are 44 common differentially expressed genes between EPOF21 and other EPO producers.
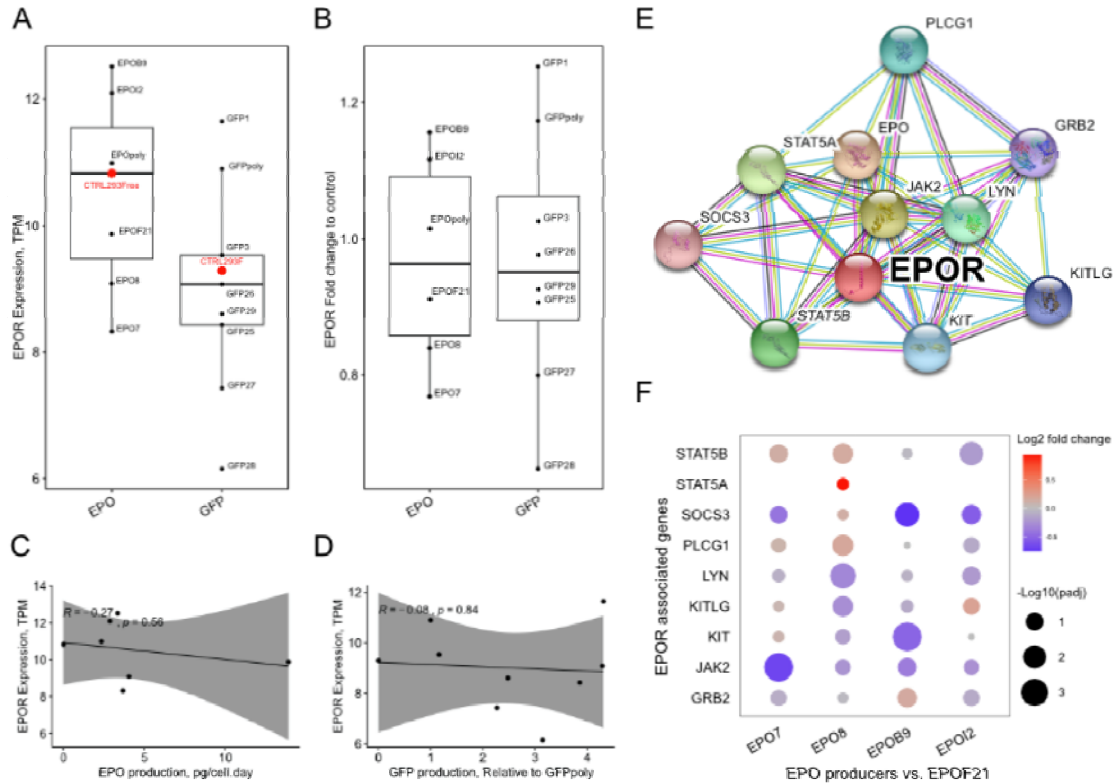
Figure S7. The difference in expression of Erythropoietin receptor (EPOR) gene is not significant between EPO producers vs. control cell-line and also EPO producers vs. GFP producers. (A) Distribution of EPOR expression between EPO producers and GFP producers. (B) Normalized expression of EPOR by considering control cell-lines background effect shows no difference in expression of EPOR between EPO producers and GFP producers. (C) EPOR expression is not significantly correlated with EPO production across EPO producer clones including control cell-line (D) There is no significant correlation in EPOR expression and GFP production in GFP producer clones. (E) Network of interacting genes with EPOR. (F) Comparison of expression of EPOR interacting genes indicates expression of these genes is not significantly different (B.H adj. p-value < 0.05, L2FC > 1) between EPOF21 and other EPO producer clones.

A

Legend: EPOF21 vs. EPOI2 (blue), EPOF21 vs. EPOB9 (green), EPOF21 vs. EPO7 (red), EPOF21 vs. EPO8 (olive)

Column categories: PROTEIN LOCALIZATION TO ORGANELLE, NEGATIVE REGULATION OF BINDING, REGULATION OF ORGANELLE ORGANIZATION, REGULATION OF CELLULAR COMPONENT BIOGENESIS, POSITIVE REGULATION OF CELLULAR COMPONENT ORGANIZATION, MIDBODY, POSITIVE REGULATION OF MOLECULAR FUNCTION, CELLULAR MACROMOLECULE CATABOLIC PROCESS, NEURON DEVELOPMENT, ENDOPLASMIC RETICULUM TO CYTOSOL TRANSPORT, POSITIVE REGULATION OF CELLULAR PROTEIN CATABOLIC PROCESS, REGULATION OF CATABOLIC PROCESS, LIPOSACCHARIDE METABOLIC PROCESS, ORGANONITROGEN COMPOUND CATABOLIC PROCESS, PROTEIN CATABOLIC PROCESS, SECRETION, NEGATIVE REGULATION OF RESPONSE TO ENDOPLASMIC RETICULUM STRESS, CELLULAR RESPONSE TO TOPOLOGICALLY INCORRECT PROTEIN, CARBOHYDRATE DERIVATIVE METABOLIC PROCESS, PHOSPHOLIPID BIOSYNTHETIC PROCESS, RESPONSE TO ENDOPLASMIC RETICULUM STRESS, ENDOPLASMIC RETICULUM, INTRINSIC COMPONENT OF ENDOPLASMIC RETICULUM MEMBRANE, INTRINSIC COMPONENT OF ORGANELLE MEMBRANE, ENDOPLASMIC RETICULUM PART

Scale: -Log10 (Padj), Up-regulated (red, 2) to Down-regulated (blue, -2)

B

Down-regulated in EPOF21 | Up-regulated in EPOF21

Gene expression level, Log2 Normalized counts

Genes (down-regulated, blue): TSSC2, DDIT4, LGALS1, H4C14, SAMD11, G0S2, ID2, SNCG, HSPA1B, NME1-NME2, EHBP1L1, NPAS1, EML2, HSPA1A

Genes (up-regulated, red): FN1, AC092299.1, ATF6B, INA, AC116533.1

EPO producers: EPOF21 (black), Low EPO-producers (green)

A Number and direction of correlating genes with EPO/GFP

EPO
GFP

428 genes

GFP+, 10 genes
GFP-, 99 genes
GFP
EPO
EPO+, 223 genes
EPO-, 93 genes

B EPO correlating genes / GFP correlating genes

310   6   112

C Common correlating genes with EPO and GFP

GFP   EPO

B3GNT5
NMB
WDR53
CBSL
NPAS1
MVP

-1   -0.5   0   0.5   1

D

Actin polymerization or depolymerization
Regulation of actin filament length
ARPC3
RAC1
TWF1
DMTN
DLG1
Regulation of protein polymerization
Regulation of cellular component size
Actin filament polymerization
Amyloid precursor protein metabolic process
TNFRSF12A
TMEM109
PAK2
DHCR24
PSENEN
Nucleotide_excision repair
POLE2
TM2B
NTHL1
XAB2
RFC4
CETN2
PPM1F
PPP2R5C
Apoptotic signaling pathway
PEX11A
SCAMP3
PRC1
RHOC
Cytokinesis
FXN
DDIT4
STX4
ANK3
AURKB
MYH9
Intrinsic apoptotic signaling pathway
IFI6
BAD
Post Golgi vesicle mediated transport
PDK2
PARL
MFN1
TRMT61B
DHPS
PIGX
COA3
TOMM7
SUPV3L1
B4GALT3
MGAT4A
B3GNT5
ATP6V0E2
NRF1
PPIB
MPDU1
HSPA8
ALG3
ATP6V0E1
TOR1A
ERP44
PFDN1
FKBP3
ALG12
Regulation of pH
Mitochondrion organization
MESD
DNAJB11
Post_translational protein modification
Protein folding

P value
0.05   0

Positive Correlation
Negative Correlation

Biological GO terms associated with EPO-correlating genes

E

Signal transduction in response to DNA damage
CDKN1A
CRADD
Regulation of cyclin dependent protein kinase activity
DNA damage response, signal transduction by p53 class mediator
TFAP4
PKMYT1
RANBP17
SLC25A6
Protein targeting
CXCR4
PEX10
Nucleobase_containing compound transport
MVP
Viral reproduction
HNRNPA1
TAF12
Viral infectious cycle
DDX39B
Macromolecular complex disassembly
RNA export from nucleus
RPL4
VRK1
Spliceosomal complex assembly
Viral reproductive process
SRSF5
RNA splicing, via transesterification reactions
SNRPA1
RPL36A
Cellular protein complex disassembly
HBB
Protein complex disassembly
component disassembly
NEFL
MRNA metabolic process
MOAP1
DYNC2LI1
Ribonucleoprotein complex assembly
SURF1
RPL29
Protein targeting to membrane
HAUS4
C1QBP
RPL3
P3H4
Embryo implantation
IMPDH2
EEF1G
RNA catabolic process
AHI1
Positive regulation of translation
SMG9
EEF2
Translational initiation
CIRBP
DNASE1L1
SDC4
EPHA2
RGS10
Translation
PAWR
RND3
RND2
EIF2S3
DPYSL3
PNP
Interleukin_2 production
Actin filament bundle assembly
Cellular nitrogen compound catabolic process

P value
0.05   0

Positive Correlation
Negative Correlation

Biological GO terms associated with GFP-correlating genes

F

RPS9   RPL38   SUMO1
EFL1
MRPS11
AURKB
PSMD2
HSPA8
PAK2
RAC1

Node degree
1   150

Positive Correlation
Negative Correlation

Hubs of significantly EPO-correlated genes

G

CDKN1A
ANXA7
HNRNPA1
EEF1G
SRSF5
NR4A1
EEF2
RPL3
C1QBP   RPL4

Node degree
1   200

Positive Correlation
Negative Correlation

Hubs of significantly GFP-correlated genes