

ARTICLE

DOI: 10.1038/s41467-017-00999-2

OPEN

# Efficient protein production by yeast requires global tuning of metabolism

Mingtao Huang<sup>1,2</sup>, Jichen Bao <sup>1,2</sup>, Björn M. Hallström<sup>3</sup>, Dina Petranovic<sup>1,2</sup> & Jens Nielsen <sup>1,2,3,4</sup>

The biotech industry relies on cell factories for production of pharmaceutical proteins, of which several are among the top-selling medicines. There is, therefore, considerable interest in improving the efficiency of protein production by cell factories. Protein secretion involves numerous intracellular processes with many underlying mechanisms still remaining unclear. Here, we use RNA-seq to study the genome-wide transcriptional response to protein secretion in mutant yeast strains. We find that many cellular processes have to be attuned to support efficient protein secretion. In particular, altered energy metabolism resulting in reduced respiration and increased fermentation, as well as balancing of amino-acid biosynthesis and reduced thiamine biosynthesis seem to be particularly important. We confirm our findings by inverse engineering and physiological characterization and show that by tuning metabolism cells are able to efficiently secrete recombinant proteins. Our findings provide increased understanding of which cellular regulations and pathways are associated with efficient protein secretion.

<sup>1</sup>Department of Biology and Biological Engineering, Chalmers University of Technology, SE41296 Gothenburg, Sweden. <sup>2</sup>Novo Nordisk Foundation Center for Biosustainability, Chalmers University of Technology, SE41296 Gothenburg, Sweden. <sup>3</sup>Science for Life Laboratory, KTH Royal Institute of Technology, SE17165 Solna, Sweden. <sup>4</sup>Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, DK2800 Kongens Lyngby, Denmark. Correspondence and requests for materials should be addressed to J.N. (email: [nielsenj@chalmers.se](mailto:nielsenj@chalmers.se))

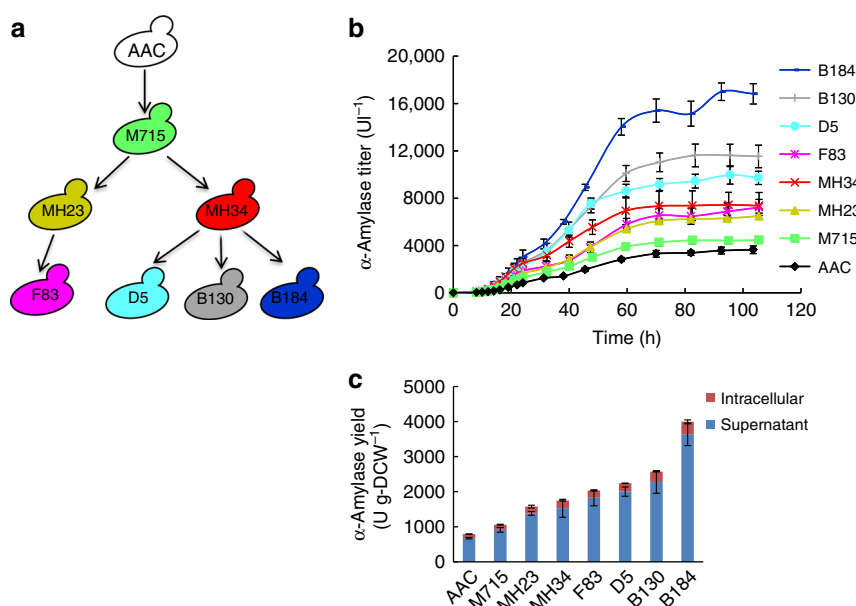
Eukaryal cells have a sophisticated protein secretory system, which ensures proper protein folding, post-translational modification, sorting, trafficking, etc<sup>1–3</sup>. Many other cellular processes interact closely with the protein secretory pathway to ensure supply of building blocks and energy<sup>4</sup>. For this reason, dysfunction of the protein secretory pathway can be lethal to the cell, and indeed many human diseases result from disorders in this pathway<sup>5, 6</sup>. Yeast *Saccharomyces cerevisiae* is a single-cell organism that is widely used as a model to study eukaryal cell biology, including the protein secretory pathway. Indeed, large knowledge about protein secretion in eukarya has been obtained from studies of this yeast<sup>7–9</sup>. Yet, full understanding of the architecture of this pathway and in particular its interaction with other cellular processes is still lacking.

Eukaryal cells are often preferred cell factories for production of many pharmaceutical proteins as they can be engineered to secrete functional proteins with correct fold and modifications into the extracellular medium, which results in reduced costs for downstream purification<sup>10</sup>. Mammalian cells, insect cells, filamentous fungi, and yeasts are, therefore, widely used cell factories for production of recombinant proteins<sup>11</sup>. Many studies have focused on improving protein secretion of these cell factories through metabolic engineering by elimination of bottlenecks at different steps, especially in the secretory processes<sup>12–15</sup>. However, limited understanding of the protein secretory pathway prevents rational engineering of many of these cell factories. There is, therefore, a need for unravelling the underlying mechanisms, and in particular how the secretory pathway and its regulation interact with other cellular processes. We use RNA-seq to perform a transcriptional genome-scale analysis of seven mutant strains of the yeast *S. cerevisiae* having a fivefold varying protein secretion capacity for a recombinant protein. Higher protein secretion may be affected not only by the direct process that a mutant gene is involved in, but also secondary cellular responses to the appearing mutation. The rationale for this study, that even though the mutant strains have many different mutations, mutant strains with higher protein secretion may have similar transcriptional regulatory responses caused by these

different mutations. This hypothesis is confirmed by the present work, which mainly focuses on transcriptional responses to the mutations and, therefore, has less emphasis on actual mutations. From this transcriptional genome-scale analysis we can identify conserved patterns in high-protein secretion mutant strains, and reveal critical factors for efficient protein secretion in yeast.

## Results

**Phenotypic characterization.** Using ultraviolet mutagenesis and microfluidic droplet sorting, we previously isolated several different yeast strains with improved secretion of the heterologous enzyme  $\alpha$ -amylase<sup>16</sup>. Here, we systematically analyzed these strains to reveal the mechanisms of efficient secretion. All the mutant strains, together with the reference strain, were grown in batch cultures in order to obtain quantitative phenotypic information (Fig. 1a). Compared with the reference strain AAC, the mutant strains produced significantly more  $\alpha$ -amylase throughout the culture process resulting in a higher final  $\alpha$ -amylase titer (Fig. 1b, c). An increase in the final  $\alpha$ -amylase titer was associated with an increase in the specific  $\alpha$ -amylase production rate, and this rate was fourfold improved for the best strain B184 compared with AAC (Table 1). Interestingly, increased  $\alpha$ -amylase production was associated with increased specific growth rate, increased glucose uptake rate, increased ethanol production rate (Table 1, Supplementary Fig. 1a, b), and a decreased yield of biomass on glucose and an increased yield of ethanol on glucose (Supplementary Table 1). The final biomass yield of most mutant strains was, therefore, slightly lower (around 10%) (except for the strain F83 that had a 27% lower final biomass yield) compared to AAC (Supplementary Fig. 1e). It was noticed that strain M715 had the lowest specific glycerol production rate and lowest specific acetate production rate, but these rates increased again in descendants of M715 (Table 1, Supplementary Fig. 1c). The higher  $\alpha$ -amylase titer in the medium and lower intracellular  $\alpha$ -amylase percentage of the mutant strains showed that the secretion capacity of the mutant strain was improved (Supplementary Fig. 1d). From this phenotypic characterization of the strains it is clear that increased



**Fig. 1**  $\alpha$ -Amylase secretion yeast strains in batch cultures. **a** Evolutionary relationships among the strains used in this study. Strains were selected for higher  $\alpha$ -amylase production from ultraviolet mutagenesis libraries with microfluidic screening in an earlier study<sup>16</sup>. **b**  $\alpha$ -Amylase titer of strains during batch cultures. Cells were cultured with an initial OD<sub>600</sub> of 0.01 in SD- × SCAA medium in a bioreactor by controlled at 30 °C, 600 rpm agitation, 30 l h<sup>-1</sup> air flow, pH = 6. **c** The final  $\alpha$ -amylase yield. Data shown are mean values  $\pm$  standard deviations of triplicates or quadruplicates

**Table 1 Physiological parameters of the mutant strains**

Strain	$\mu_{\max}$	$r_s$	$r_E$	$r_G$	$r_A$	$r_P$
AAC	0.276 ± 0.010	1.351 ± 0.024	0.347 ± 0.006	0.099 ± 0.002	0.040 ± 0.001	101.37 ± 4.90
M715	0.309 ± 0.017	1.329 ± 0.249	0.360 ± 0.036	0.083 ± 0.011	0.033 ± 0.002	164.01 ± 12.61
MH23	0.304 ± 0.007	1.408 ± 0.147	0.439 ± 0.024	0.091 ± 0.005	0.039 ± 0.003	223.45 ± 10.18
F83	0.296 ± 0.005	1.588 ± 0.158	0.495 ± 0.051	0.117 ± 0.005	0.040 ± 0.005	264.55 ± 69.41
MH34	0.329 ± 0.016	1.649 ± 0.141	0.460 ± 0.023	0.087 ± 0.006	0.034 ± 0.002	262.96 ± 35.56
D5	0.313 ± 0.002	2.023 ± 0.071	0.604 ± 0.024	0.116 ± 0.007	0.040 ± 0.003	283.26 ± 15.40
B130	0.305 ± 0.006	2.045 ± 0.181	0.677 ± 0.058	0.109 ± 0.015	0.040 ± 0.002	314.11 ± 73.15
B184	0.310 ± 0.006	1.969 ± 0.069	0.610 ± 0.049	0.116 ± 0.013	0.047 ± 0.002	386.93 ± 49.01

Data shown are mean values ± standard deviations of triplicates or quadruplicates  
 $\mu_{\max}$  maximum specific growth rate ( $\text{h}^{-1}$ ) on glucose,  $r_s$  specific glucose uptake rate ( $\text{g g-DCW}^{-1} \text{h}^{-1}$ ),  $r_E$  specific ethanol production rate ( $\text{g g-DCW}^{-1} \text{h}^{-1}$ ),  $r_G$  specific glycerol production rate ( $\text{g g-DCW}^{-1} \text{h}^{-1}$ ),  $r_A$  specific acetate production rate ( $\text{g g-DCW}^{-1} \text{h}^{-1}$ ),  $r_P$  specific  $\alpha$ -amylase production rate ( $\text{U g-DCW}^{-1} \text{h}^{-1}$ )

protein secretory capacity does not impose any penalty on growth as the mutant strains were growing faster, but the strains have a higher glucose uptake rate and an increasing fraction of the glucose is directed toward ethanol production. As protein synthesis and secretion are energy consuming processes, increased ratio of ATP production per total cell protein in fermentative growth<sup>17</sup> helps to meet the increased energy demand in strains with higher  $\alpha$ -amylase secretion.

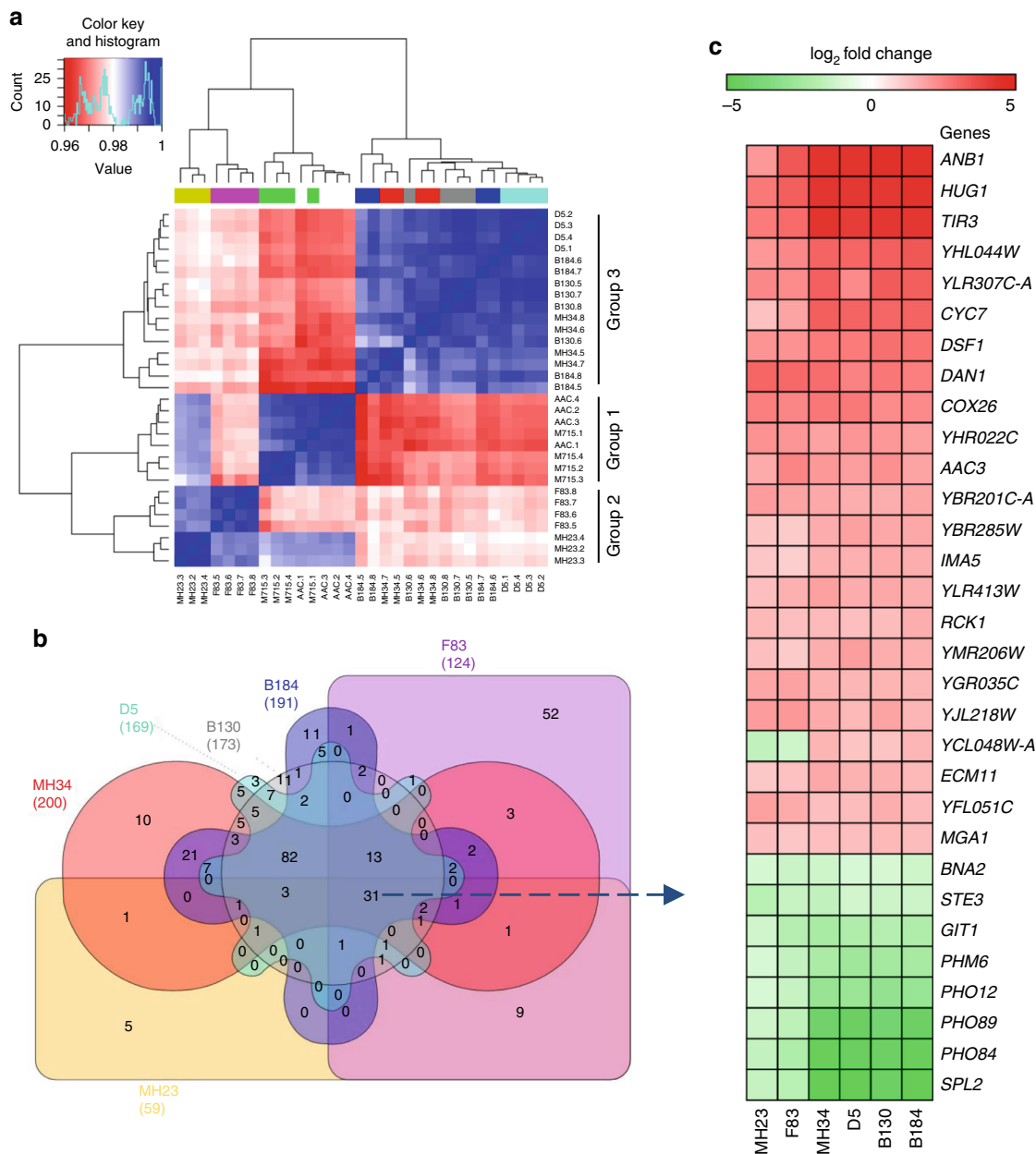
**Transcriptional profiling.** The relative amylase yield was calculated for the exponential phase and the end of cultivation, respectively. Mutant strains showed higher amylase yield in exponential phase compared with the reference strain AAC (Supplementary Fig. 2a, b). A transcriptional steady state was reported in exponentially growing yeast cells, and analysis of the transcriptome in the exponential phase is, therefore, reliable<sup>18</sup>. Therefore, cells were sampled at early exponential phase ( $\text{OD}_{600} \approx 1$ ) for RNA sequencing to reveal important factors influencing protein secretion. The Spearman's correlation coefficient among samples by pairwise comparison showed strong reproducibility of biological replicates (Fig. 2a). Furthermore, this analysis showed that the strains can be classified into three groups: group 1 contained AAC and M715; group 2 contained MH23 and F83; group 3 contained MH34, D5, B130, and B184. This grouping was consistent with the evolutionary pedigree of the strains (Fig. 1a) and revealed by principal component analysis (Supplementary Fig. 3). These findings indicated that group 2 and group 3 had different evolutionary paths toward increased protein secretion. In contrast to strain M715, more genes were significantly upregulated or downregulated in higher  $\alpha$ -amylase producing strains from group 2 and group 3 (Supplementary Fig. 4). This indicated that the significant improvement in  $\alpha$ -amylase production obtained in strains from these two groups was associated with a global modulation of gene expression, which may be due to a requirement for adjustment of many different cellular processes to support the increased  $\alpha$ -amylase production. Chromosome III was duplicated in MH34 and its descendants D5, B130, and B184, and in these strains, many genes located on this chromosome showed an about twofold increase at the transcription level.

To identify common expression changes in the strains in group 2 and group 3, significantly differentially expressed genes were plotted in a Venn diagram (Fig. 2b). Hereby a total of 31 commonly differentially expressed genes within all mutant strains in these two groups were identified, and the differential expression levels of these 31 genes compared with the reference strain are summarized in Fig. 2c. Several genes, including *ANB1*, *TIR3*, *CYC7*, *DANI*, and *AAC3*, which are expressed under anaerobic/hypoxic conditions and/or required for anaerobic growth<sup>19, 20</sup>, are significantly upregulated in the mutant strains.

As the dissolved oxygen levels in the medium at the time point of cell sampling for RNA extraction were around 90% for all strains (Supplementary Fig. 2c), it implied that the mutant strains exhibited anaerobic characteristics despite the aerobic environment. Most of the significantly downregulated genes were phosphate responsive genes. *PHO12*, *PHO84*, *PHO89*, and *SPL2*, and these are all related to phosphate utilization and regulation. *GIT1* is involved in phospholipid metabolism and *PHM6* is regulated by phosphate levels<sup>21</sup>.

**Reporter TFs analysis and reporter gene ontology (GO) terms analysis.** To identify underlying transcriptional regulatory responses in the mutant strains, the transcriptome data were integrated with a network of transcription factors (TFs) and associated genes to identify so-called reporter TFs for mutant strains from group 2 and group 3<sup>22, 23</sup>. From this analysis, 308 TFs were scored. The top five reporter TFs of each strain supposedly represented an important regulatory network associated with increased protein secretion and were presented with directional significances of their target genes that were found to be upregulated or downregulated (Fig. 3a). The expression level of the reporter TFs themselves were also evaluated (Fig. 3b). In order to study whether expression level of genes on the duplicated chromosome III influenced the reporter analysis, we did a reporter TFs analysis where we removed all genes on the chr III from the analysis, and then redid the reporter TF analysis for the strains MH34, D5, B130, and B184 (Supplementary Fig. 5). Identified TFs by the reporter analysis without including genes on chr III were almost the same as the identified TFs by the reporter analysis using data for all genes (Fig. 3a), although the significances were different. Only *SNF2* is not identified in the reporter analysis not including genes on chr III. This suggests that key information in global transcriptional regulatory responses, caused by mutations and chr III duplication, was not determined by expression level of genes on chr III.

Rox1p is a transcriptional regulator that represses hypoxia induced genes<sup>20</sup>. The reporter TFs analysis showed upregulation of genes repressed by Rox1p, which was consistent with the identification of commonly significantly upregulated genes in the mutant strains. Furthermore, expression of *ROX1* gene was found to be decreased in most of the mutant strains. Gene regulated by TUP1p, an element of Tup1p–Cyc8p complex mediating repression of anaerobic genes with Rox1p<sup>24</sup>, were also found to be upregulated in the mutant strains. Hap1p is a heme-responsive transcriptional activator and transcriptional repressor under aerobic and anaerobic conditions, respectively<sup>25</sup>. *HAP2*, *HAP3*, *HAP4*, and *HAP5* encode proteins to form the Hap2p/3p/4p/5p CCAAT-binding complex, which is a heme-activated and glucose-repressed transcriptional activator and global regulator of respiratory gene expression<sup>26</sup>. Genes regulated by Hap1p and



**Fig. 2** Overview of transcriptome data. **a** Heatmap of Spearman's correlation coefficient for the similarities in the expression profiles of exponential phase between the different samples by pairwise comparison. **b** Common very significantly differentially expressed genes ( $P$ -adj < 0.05 (Benjamini-Hochberg method) and abs (log<sub>2</sub> Fold change) > 1) in the mutant strains compared with the reference strain AAC. The number of very significantly differentially expressed genes in each strain is specified under the strain name. **c** Expression levels of common significantly differentially expressed genes in mutant strains

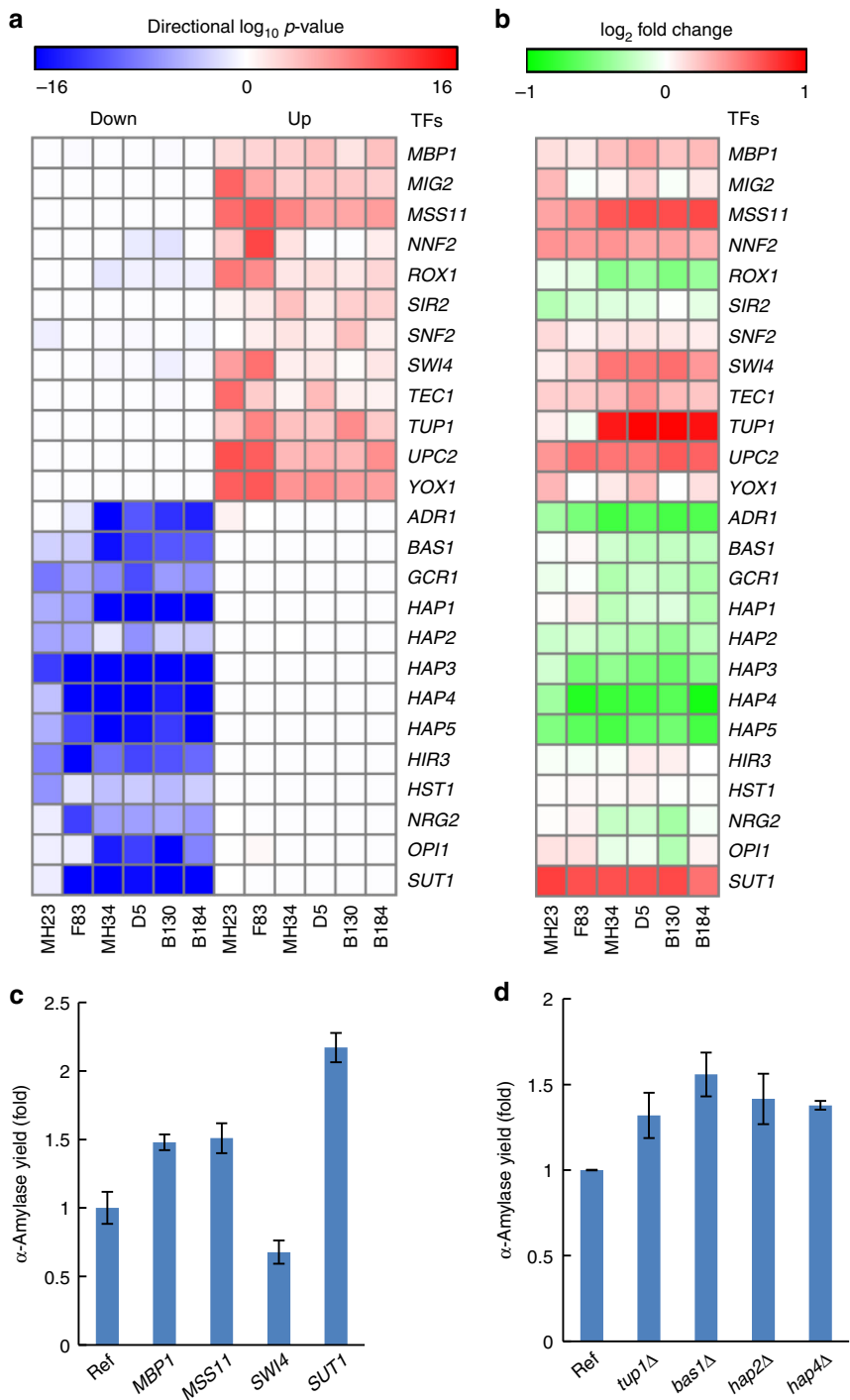
the Hap2p/3p/4p/5p CCAAT-binding complex were found to be downregulated in the mutant strains. All these results indicated that the mutant strains were expressing genes as if they were in hypoxia, which may facilitate efficient protein secretion.

Besides anaerobic metabolism, other types of regulation were identified, including nutrient signaling, nucleotide synthesis, and phosphate metabolism, cell cycle, etc. by the reporter TFs analysis (Fig. 3a). *MSS11* and *TEC1* are involved in response to starvation and responsible for nutrient regulation<sup>27</sup>, and both represent reporter TFs for upregulated genes. As production of heterologous protein competes with intracellular resources that could be used for cell growth, slightly lower final biomass yields were found for the efficient protein secretion strains (Supplementary Fig. 1e). Actually, the  $\alpha$ -amylase produced by the best production

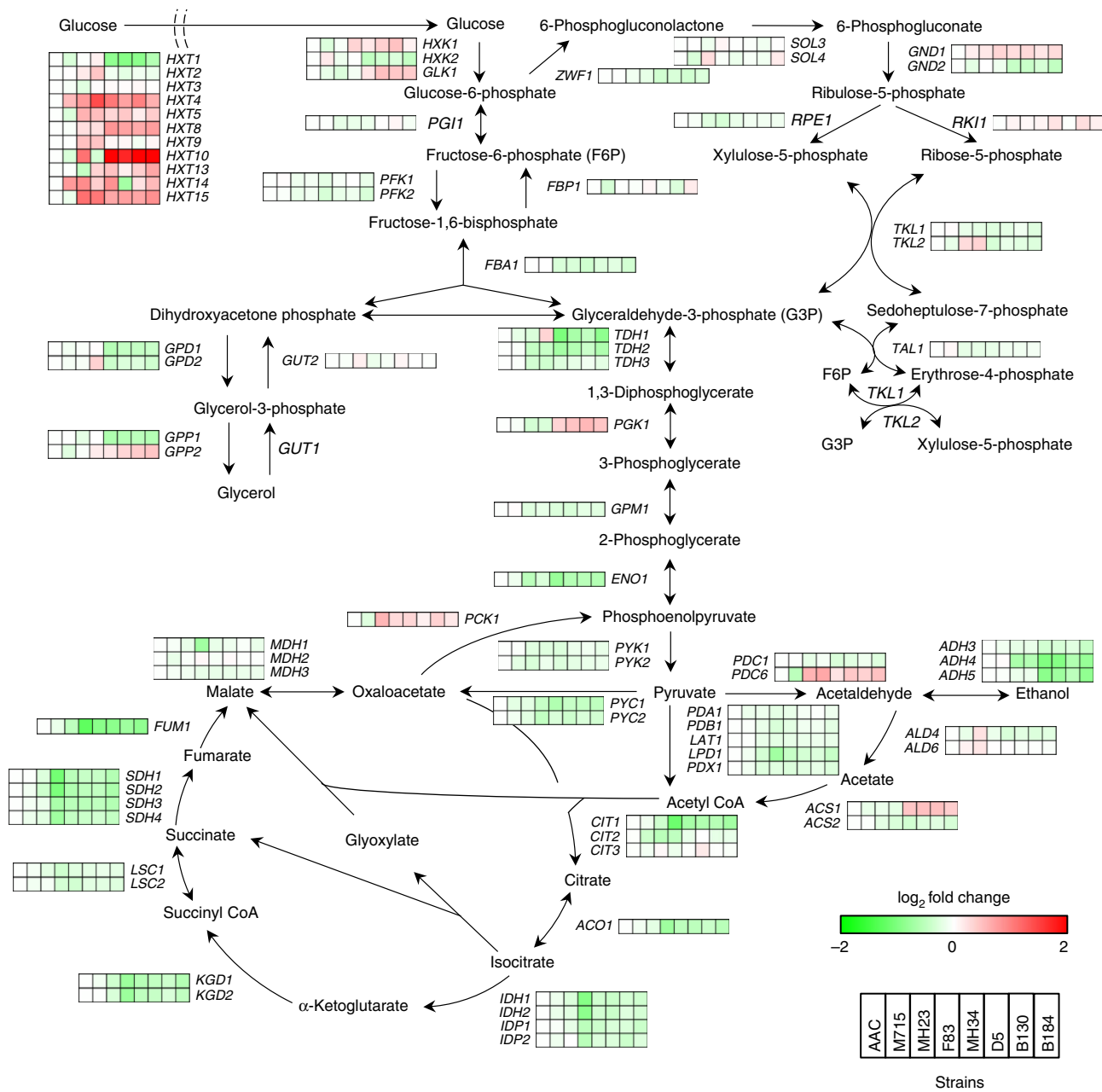
strain B184 accounted for about 13% of total cellular protein produced (Supplementary Fig. 2d). Strong competition for resources may result in induction of a partial nutrient starvation response, hence activated nutrient related regulators like Mss11p and Tec1p. Bas1p is responsible for inducing expression of genes involved in nucleotide synthesis and phosphate consumption by interaction with Pho2p<sup>28</sup>. Downregulation of genes regulated by Bas1p in the mutant strains hinted that protein secretion may benefit from this kind of response. The first two steps of de novo synthesis of pyrimidine are catalyzed by Ura2p, which is feedback inhibited by the level of UTP. The *URA2* gene was found to be downregulation ( $P$ -adj < 0.05, Benjamini-Hochberg method) in our mutant strains. *FUR1*, which encodes enzyme for conversion of uracil into UMP in the salvage pathway, was found to be

upregulation in the mutant strains, which indicates that nucleotide biosynthesis may shift to use of the salvage pathway. The reporter TFs analysis also revealed that genes regulated by Mbp1p and Swi4p were upregulated. Both Mbp1p and Swi4p can interact with Swi6p to form a MBF (Mbp1p/Swi6p-dependent cell cycle box Binding Factor) complex and a SBF (Swi4p/Swi6p-dependent cell cycle box Binding Factor) complex, respectively,

which regulate gene expression during the G1/S transition of the cell cycle<sup>29</sup>. To validate if the identified reporter TFs represent key points in regulation of protein secretion, some of the TFs were selected for evaluation. Previously engineering of TFs through gene deletion or over-expression has shown to impact protein secretion, and our TF reporter analysis of the mutant strains is consistent with



**Fig. 3** Reporter transcription factors (TFs) analysis revealed important transcriptional network responses in mutant strains. TFs were scored by the modulation in expression level of genes that controlled by TFs. **a** The top five scored reporter TFs for each strain in distinct-directional up class (red) and distinct-directional down class (blue) are chosen and presented by their significance. **b** Expression levels of reported TFs in the mutant strains compared with the reference strain AAC. Enhanced production of  $\alpha$ -amylase by overexpression of TFs **c** or deletion of TFs **d**, data shown are mean values  $\pm$  standard deviations of duplicates



**Fig. 4** Transcriptional changes of genes involved in carbohydrate metabolism. Most genes of the central carbon metabolism were downregulated and genes encoding high-affinity hexose transporters were upregulated in the mutant strains

these studies. Thus, the TF reporter analysis indicates that deletion of *ROX1* or upregulation of *UPC2* should improve protein production, which is indeed the case<sup>30</sup>, and that deletion of *HAP1* should improve protein production<sup>31</sup>. To evaluate if some of the additional reporter TFs identified in this study have an impact on protein secretion, we deleted or over-expressed some of the respective genes according to their altered expression level in the mutant strains. Hence, *MBP1*, *MSS11*, *SWI4*, and *SUT1* were over-expressed and *BAS1*, *HAP2*, and *HAP4* were deleted in strain AAC (Fig. 3c, d). *TUP1* is located on chromosome III and had high expression level in the mutant strains, due to duplication of the entire chromosome III. As mentioned above *Tup1p* interacts with *Rox1p*, and as deletion of *ROX1* has been found to have a positive impact on protein secretion we also deleted *TUP1* (Fig. 3d). Except for over-expression of *SWI4*, all the engineering strategies resulted in

increased protein secretion, and this shows that efficient protein secretion can be achieved by global tuning of gene expression by engineering of TF expression. Furthermore, these findings provide a validation that computationally identified reporter TFs for mutant strains can be used as targets for improving protein secretion.

We also calculated reporter GO terms for the mutant strains from group 2 and group 3. Hereby we found that genes associated with GO terms related to mitochondrial function, generation of precursor metabolites and energy, cellular respiration, cellular amino-acid metabolic process, etc. were downregulated. In contrast, genes associated with GO terms related to ribosome function, cytoplasmic translation, regulation of organelle organization, Golgi vesicle transport, protein lipidation, lipid metabolic process etc. were upregulated (Supplementary Fig. 6). These results were consistent with the phenotypic analysis that showed



that the mutant strains exhibited reduced respiration and increased protein synthesis and protein trafficking. Furthermore, changes of lipid metabolism in the mutant strains may contribute to increased protein secretion<sup>32, 33</sup>.

We also experimentally verified some of the nonsense mutations in genes associated with identified GO terms. Thus, *PGM2*, encoding phosphoglucomutase involved in glycogen biosynthesis, and *PXA1*, encoding one of the subunits for a fatty-acid transporter to the peroxisomes, carry nonsense mutations and are hence non-functional. We, therefore, evaluated the impact of deleting these genes on protein secretion in the reference strain AAC and found that these deletions results in 30–45% improvement in protein production (Supplementary Fig. 7), which hereby supported our GO term analysis.

### Transcription analysis of genes in carbohydrate metabolism.

As macroscopic fluxes such as glucose uptake rate and ethanol production rate were changed in the mutant strains, and at the same time GO terms associated with cellular respiration and generation of precursor metabolites and energy metabolism were found to be downregulated, we investigated the transcriptional levels of genes related to carbohydrate metabolism in the mutant strains. The expression levels of genes were presented on a map of the central carbon metabolism (Fig. 4). *HXT1*, encoding a low-affinity glucose transporter<sup>34</sup>, was found to be downregulated in mutant strains from group 3. In contrast, genes encoding high-affinity hexose transporters were upregulated in almost all mutant strains. This altered transcriptional level of hexose transporter genes was inconsistent with the high glucose concentration in the medium, and, therefore, indicated abnormal glucose sensing in the mutant strains, but still resulted in an increased glucose uptake rate (Table 1). By GO slim mapper analysis of mutations and differential expressed genes in the mutant strains<sup>16</sup> (Supplementary Data 1), a mutant gene *SNF3* was found to be associated with the GO term carbohydrate transport. As *SNF3p* is a plasma membrane glucose sensor involved in regulation of expression of hexose transporters<sup>35</sup>, abnormal glucose sensing and altered transcriptional level of hexose transporter genes were likely affected by the mutation in *SNF3*. Most genes of the central carbon metabolism were downregulated in the mutant strains. For the TCA cycle, this was consistent with the reduced respiration and increased ethanol production in these strains, whereas decreased expression of glycolytic genes seemed inconsistent with the increased glucose uptake rate. However, glycolytic flux is well known to not primarily be controlled at the transcriptional level, contrary to the TCA cycle and respiration<sup>35</sup>.

Amino acids are building blocks for protein synthesis. Many genes involved in amino acid biosynthesis pathway were downregulated in the mutant strains (Supplementary Fig. 8). Genes *CHA1*, *HIS4*, *ILV6*, and *THR4* were found to be upregulated in strains MH34, D5, B130, and B184, but this can be ascribed to the location of these genes on the duplicated chromosome III. *GDH3*, *GLN1*, and *GLT1* involved in glutamate and glutamine biosynthesis were also upregulated<sup>36</sup>. Both glutamate and glutamine serve as amino donors. Furthermore, genes *AAT1*, *ALT1*, and *BAT1*, which encode transaminases for reversible conversion between glutamate and other amino acids<sup>37–39</sup>, were upregulated. These results indicated that glutamate and glutamine play important roles in protein synthesis. Conversion of amino acids through transaminases to balance the intracellular amino acids pool is critical for efficient protein production, in particular when amino acids are supplied via the medium as in our case. When comparing the amino acid composition of  $\alpha$ -amylase with that of yeast cell proteins, it was found that there is a 9.3-fold higher requirement for cysteine for production of the same amount of  $\alpha$ -

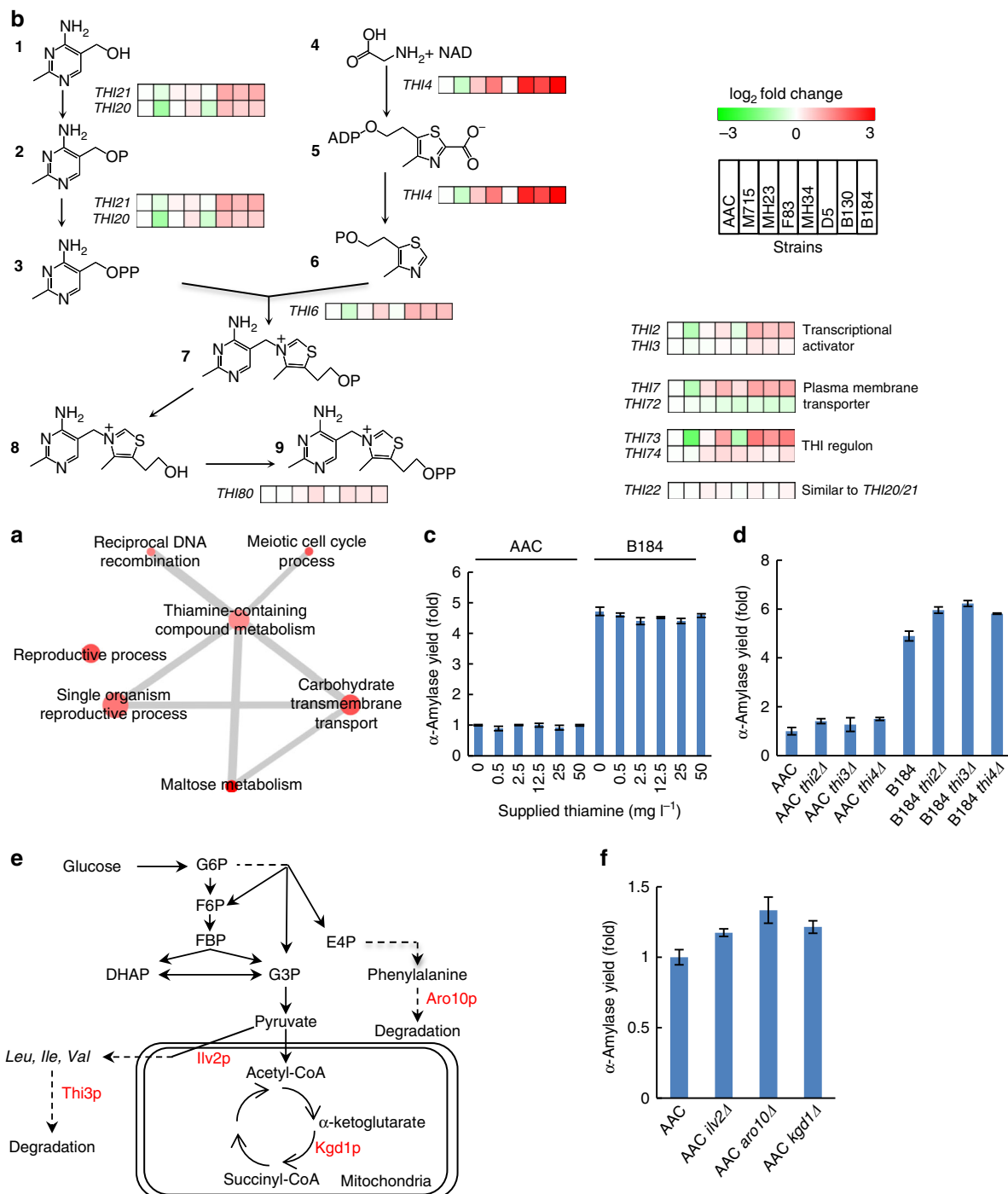
amylase compared with the average biosynthesis of the same amount of yeast cell proteins (Supplementary Table 2). In fact, the requirement of cysteine in the mutant strains increased by 25–85% compared with the reference strain AAC (Supplementary Table 3). However, genes of the cysteine biosynthetic pathway still did not show increased expression (Supplementary Fig. 8). Yet, from analysis of genes involved in amino acid transport, we found that *YCT1* and *ERS1* responsible for cysteine transport were significantly upregulated<sup>40, 41</sup> (Supplementary Fig. 9), and this can help the mutant strains to meet the increased demand for cysteine when  $\alpha$ -amylase production is increased (Supplementary Table 3).

**Analysis and engineering of thiamine biosynthesis.** In our previous study<sup>16</sup>, the *Rhizopus oryzae* glucan-1,4- $\alpha$ -glucosidase was shown also to have increased secretion by the mutant strains, and as for  $\alpha$ -amylase strain B184 was by far the best in terms of protein secretion. Here, we, therefore, tested this strain for its ability to produce two other proteins, human serum albumin, and *Trichoderma reesei* endo-1,4-beta-xylanase II. Compared with the reference strain CEN.PK 530.1C, both proteins showed higher protein yield in B184 (Supplementary Fig. 10). These results suggested our findings were not limited to a specific protein and may be generally applicable. We, therefore, performed a more detailed analysis of the transcriptome of the best strain B184 in terms of protein secretion.

Significantly upregulated and downregulated genes in this strain were analyzed by GO bioprocess enrichment (Fig. 5a and Supplementary Fig. 11), and this analysis pointed to genes involved in thiamine-biosynthesis having significantly increased expression (Fig. 5a). The GO term associated with thiamine-containing compound metabolism includes thiamine biosynthetic process, thiamine-containing compound biosynthetic process and thiamine metabolic process. By overlaying gene expression for the strains on a pathway map for thiamine biosynthesis it was further clear that the expression levels of genes related to this pathway were upregulated in strain B184 (Fig. 5b). It was noticed that these genes were also upregulated in strains F83, D5, and B130. All these four strains were derived from the second round of screening and showed higher protein secretion level compared with their ancestral strains. Based on this information, we speculated that thiamine may be required for efficient protein secretion. We, therefore, performed experiments with additional thiamine supplemented to the medium, but from these experiments it was found that the reference strain AAC showed no difference in  $\alpha$ -amylase secretion and the best mutant strain B184 even exhibited a small decrease in  $\alpha$ -amylase secretion upon additional thiamine supplementation (Fig. 5c). We then further investigated the transcriptional control of thiamine biosynthetic genes. It has been reported that expression of thiamine biosynthetic genes is controlled by Thi2p together with Thi3p and Pdc2p, and transcription of *THI2* is regulated negatively by the intracellular thiamine level<sup>42</sup>. As the expression levels of *THI2* and *THI3* in the mutant strains were upregulated, this implied that the cells secreting efficiently  $\alpha$ -amylase may be in a low cellular thiamine status. Yet the low thiamine status triggered thiamine response mechanism, which attempted to counter-balance thiamine level by elevated the transcription of thiamine biosynthetic genes. Hence, *THI2*, *THI3*, and *THI4* were deleted in strains AAC and B184 to reduce thiamine biosynthesis, and the single gene deletion resulted in increased protein secretion in both strain backgrounds (Fig. 5d). Several thiamine containing proteins (thiamin diphosphate-dependent enzymes), encoded by *ARO10*, *ILV2*, *KGD1*, and *THI3*, are involved in amino-acid metabolism (Fig. 5e)<sup>42, 43</sup>, and their expression levels were lower

in the mutant strains. Hence, a decrease thiamine concentration may additionally reduce the activity of these enzymes, suggesting that attenuating activity of these pathways may benefit protein secretion. We, therefore, deleted these genes in AAC, and deletion of each of these genes resulted in increased protein secretion (Fig. 5f).

**Reduced ER stress analysis.** Heterologous protein production often brings a protein folding burden to the cell and, therefore, causes oxidative stress in the endoplasmic reticulum (ER)<sup>44</sup>. The unfolded protein response (UPR) pathway is activated by arising oxidative stress in the ER to assist in reducing cellular stress. Hac1p is a key UPR-induced TF for transcriptional activation of



**Fig. 5** Low thiamine status is favorable for  $\alpha$ -amylase production. **a** GO bioprocess enrichment of most upregulation genes ( $P$ -adj < 0.05 (Benjamini-Hochberg method) and  $\log_2$  fold change > 0.5) in strain B184. **b** Transcriptional levels of genes related to thiamine biosynthesis. 1: hydroxymethylpyrimidine; 2: hydroxymethylpyrimidine phosphate; 3: 2-methyl-4-amino-5-hydroxymethylpyrimidine diphosphate; 4: L-glycine; 5: adenylated thiazole; 6: 4-methyl-5-( $\beta$ -hydroxyethyl)thiazole phosphate; 7: thiamine phosphate; 8: thiamine; 9: thiamine diphosphate. **c**  $\alpha$ -Amylase production in the SD-2  $\times$  SCAA medium supplied with different amount of thiamine. **d** Deletion of *THI2*, *THI3*, and *THI4* enhances  $\alpha$ -amylase production. **e** Thiamine diphosphate-dependent enzymes involved in amino-acid metabolism. DHAP dihydroxyacetone phosphate; E4P erythrose-4-phosphate; F6P, fructose-6-phosphate; FBP, fructose-1,6-bisphosphate; G6P, glucose-6-phosphate; G3P, glyceraldehyde-3-phosphate. **f** Deletion of *ILV2*, *ARO10*, and *KGD1* enhances  $\alpha$ -amylase production. Data shown in **c**, **d**, **f** are mean values  $\pm$  standard deviations of duplicates



ER chaperone encoding genes, including *KAR2*, *ERO1*, etc.<sup>45</sup>. Overexpression of *HAC1* has been reported to enhance protein secretion<sup>46</sup>. Mutant strains MH23 and F83 from group 2 were found to have increased expression levels of *HAC1* and *ERO1*. In contrast to group 2, mutant strains MH34, D5, B130, and B184 from group 3 were found to have *HAC1*, *ERO1*, and *KAR2* downregulated (Supplementary Fig. 12a). However, it was noticed that both *PDI1* and *EMC1* responsible for efficient folding of proteins in the ER were upregulated in strains MH34, D5, B130, and B184. The upregulation of *PDI1* and *EMC1* was due to their location on the duplicated chromosome III. Several other genes involved in protein folding, *EMC2-6*, were found to be upregulated, and the expression extent was more significant in strains from group 3. These results suggested that strains from group 2 and group 3 used different mechanisms to release oxidative stress from heterologous protein production. Group 2 mainly relied on the UPR pathway, whereas group 3 relied on direct up-regulation of key chaperones in the ER enabling more efficient protein folding.

In order to evaluate whether the altered expression of chaperones in the ER affected the oxidative stress, we measured the reactive oxygen species (ROS) level in the strains. Compared with a non-producing control strain, all  $\alpha$ -amylase production strains had higher ROS level, indicating increased ER stress due to heterologous protein production (Supplementary Fig. 12b). Within these  $\alpha$ -amylase production strains, the wild-type strain AAC showed the highest ROS level. This indicated that a reduction of oxidative stress can be achieved by both activation of UPR and enhancement of the protein folding capacity. Accumulated ROS was also observed visually using fluorescence microscopy (Supplementary Fig. 12d). Considering that the mutant strains produce more  $\alpha$ -amylase, the data on ROS accumulation point to that the ROS generation associated with heterologous protein production, i.e., ROS per  $\alpha$ -amylase yield unit, was reduced in the mutant strains (Supplementary Fig. 12c). A reduced oxidative stress in mutant strains was also supported by the decreased expression levels of oxidative stress response genes (Supplementary Fig. 12a).

The mutant strains carry mutations in several genes related to ER stress or organelle transport, such as *EMC1*, *ERV29*, *USO1*, *VPS10*, *SNC2*, which is consistent with our findings from the transcriptome analysis. In our previous study<sup>16</sup> we confirmed protein secretion was associated with *ERV29* and *SNC2*; and deletion of *SNC2* resulted in improved protein production. To further evaluate whether there is a general impact of modulating ER stress or organelle transport we further evaluated *EMC1*, *USO1*, and *VPS10*. These genes were deleted in AACK and deletion of *EMC1* and *VPS10* was found to improve protein production by 10–40%, whereas deletion of *USO1* did not impact protein production (Supplementary Fig. 7). The mutant strains also have gene duplications in genes involved in ER stress, e.g., *PDI1*, and we, therefore, over-expressed this in AACK and found that this results in improvement of protein production.

The reduced oxidative stress could also be related to an improved supply of NADPH. Increased supply of NADPH via the pentose phosphate pathway (PPP) was shown beneficial for higher protein production in Chinese Hamster Ovary (CHO) cells and *Pichia pastoris*<sup>47, 48</sup>. Although no increased expression was seen for PPP associated genes, we found reduced biomass yields of the mutant strains. As many biosynthesis pathways require NADPH as reducing equivalents, a decreased biomass yield will allow the cells to save more NADPH for maintaining redox balance (protection against oxidative stress). The mutant strains, therefore, seemed to meet the demands for NADPH associated with increased amylase production by redistribution of resources. This is supported by the total protein content in the

mutant strains. As amylase content increased, the total protein (yeast cell protein and amylase) did not increase in the mutant strains compared with the reference strain AAC (Supplementary Fig. 13).

## Discussion

Here, we studied some of the underlying mechanisms of efficient protein secretion through comparative systems biology analysis of efficient  $\alpha$ -amylase secretion mutant strains and a reference strain. From genome-wide transcription analysis we found that the majority of genes related to glycolysis and TCA cycle were downregulated in the mutant strains, but still the final biomass yield was only slightly decreased and the maximum specific growth rate was even increased. Previously, it was found that glycolytic enzymes represent a large proportion (30–60%) of the soluble proteins in the cell<sup>49</sup>, and a recent study reported that there are strong redundancies in glycolytic enzymes so that reducing the level of several glycolytic proteins only has a minor impact on yeast growth<sup>50</sup>. Downregulation of glycolytic genes may, therefore, have occurred in order to allocate more proteome mass and protein synthesis capacity for heterologous protein production. Attenuation of phosphoglucosyltransferase activity (encoded by *PGM2*) through inactivation of *PGM2* may have further contributed to this effect, as confirmed by a 40% increase in protein production by deletion of this single gene (Supplementary Fig. 7). Furthermore, allocation of more resources to protein secretion resulted in a reduction of cellular stress, allowing the cells to grow slightly faster. Faster growth is associated with increased ethanol production, and, therefore, a slightly lower biomass yield. It is also interesting that even though the cells have an increased amino acid demand for production of  $\alpha$ -amylase, genes associated with amino acid biosynthesis were downregulated in the mutant strains, but this was compensated by increased expression of amino acid transporters. Thus, the mutant strains seem to shift from amino acid biosynthesis to increased amino acid uptake. Yet genes involved in glutamate/glutamine biosynthesis and amino-group transfer (transaminases) were upregulated, which is necessary to balance the requirement for amino acids for cell growth and  $\alpha$ -amylase production to the provision of amino acids from the extracellular medium. In addition to serving as building blocks of proteins, glutamate and glutamine play important roles in many other processes (e.g., oxidative stress responses, nucleotide metabolism), and increased expression of genes involved in glutamate/glutamine biosynthesis may, therefore, also be due to demands from other processes. The gene *ARO10* encodes a protein (decarboxylase) involved in amino acid degradation. Improvement of amylase production by deletion of *ARO10* also highlights the importance of maintaining amino acid pool by reduced amino acid degradation for efficient protein production.

It is noteworthy that many identified reporter TFs were related to anaerobic conditions. Hypoxic genes controlled by reporter TFs were upregulated in the mutant strains, whereas respiratory genes were down-regulated. This was consistent with reporter GO terms analysis, which showed that cellular respiration and mitochondrial function were downregulated. A previous study showed that deletion of *ROX1*, which results in de-repressed expression of anaerobic genes at aerobic conditions, caused increased heterologous protein secretion<sup>30</sup>. Genes controlled by *Upc2p* and *UPC2* itself were found upregulated in the *rox1 $\Delta$*  strain, which was also found in our mutant strains. Increased protein secretion was also shown by overexpression of the *UPC2-1* allele, which constitutively activates ergosterol biosynthesis genes. These results indicated that anaerobic characteristics of the mutant strains were favorable for protein

secretion. This was supported by improved amylase production in a *KGD1* deletion strain, which has a decreased respiratory growth. Elevated oxidative stress due to an increased burden from protein folding hinders protein production, and our analysis shows that this can be overcome either by increased expression of protein folding chaperones or activation of the UPR<sup>51</sup>. Our results emphasize the importance of reducing oxidative stress associated with protein production regardless of the pathway used for this. Again, herein reducing oxidative stress in the mutant strains with evidences at transcriptional levels and by experimental validation (ROS level) supported the importance of UPR and HSR found in our previous study<sup>16</sup>. Hence, mutant genes and duplicated genes related to ER or organelle transport (such as *EMC1*, *ERV29*, *PDI1*, *USO1*, *VPS10*, *SNC2*) in the mutant strains may contribute to reduced oxidative stress and facilitate protein secretion. This was confirmed by deletion or over-expression of these genes in the reference strain (Supplementary Fig. 7). It has also been reported that a mutant *EMC1* causes induction of the UPR<sup>8</sup>, and this can partly explain our finding of a UPR in strain MH23 and F83, both of which had a mutant *EMC1*<sup>16</sup>. Our results also supported the findings from another previous study<sup>52</sup>, in which Wentz et al.<sup>52</sup> used a rapid flow cytometric sorting method to screen yeast complementary DNA overexpression libraries for increasing protein surface display. *ERO1* was one of hits in the study by Wentz et al. and overexpression of *ERO1* increased protein secretion through assisting protein folding. As improved supply of NADPH was beneficial for protein production by protection against oxidative stress<sup>47, 48</sup>, although no increased expression changes were seen in PPP in our mutant strains, it would be interesting to metabolic engineer the PPP in our mutant strains for improved protein production in future studies. Besides alteration of intracellular processes by changes in gene expression, cells can also alter processes by regulating the activity of enzymes<sup>53, 54</sup> and our results showed that several pathways catalyzed by thiamine-dependent enzymes were affected by a low thiamine status in the mutant strains<sup>55</sup>, which may provide another solution to tune the activity of metabolic pathways. Although valuable findings were found based on transcriptional data from the exponential phase, it may be worthwhile to investigate the transcriptome from other growth phases as follow-up studies.

We believe our study is the first systems level analysis of strains with various levels of protein secretion. As the strains used in our analysis possess a large number of different genomic mutations causing changes in many different intracellular processes it would be difficult to identify the effects of these mutations. However, from our systems level analysis we could identify common regulation patterns and hereby we could specify some general rules for efficient protein secretion. We confirmed several of these findings through inverse metabolic engineering where we altered the expression of identified reporter TFs. As *S. cerevisiae* is both an industrially important cell factory for recombinant protein production and a key eukaryal model organism, our findings can most likely be used as guidelines for design of other cell factories for efficient protein secretion, e.g., filamentous fungi used for production of industrial enzymes and CHO cells used for production of pharmaceutical proteins. Furthermore, our findings may also contribute gaining improved insight into the mechanism of the human protein secretory pathway. Dysfunction of this pathway is associated with many different diseases<sup>56</sup>, and it may even be possible to use our strains as a platform for screening for drugs that can cure diseases related to protein secretion<sup>57, 58</sup>.

## Methods

**Strains and plasmids.** All strains and plasmids used in this study are listed in Supplementary Table 4. All primers used in this study are listed in Supplementary

Table 5. Plasmids for gene overexpression were constructed by insertion of the gene fragment, which was amplified from the yeast genome by corresponding primer pairs and digested with restriction enzymes, to the expression vector pSPGM1 (Supplementary Fig. 14). Synthesized human serum albumin (HSA) gene with alpha factor leader and *Trichoderma reesei* endo-1,4-beta-xylanase II gene with alpha factor leader were cloned in plasmid CPOTud, resulting in plasmids pCP-AHSA and pCP-AXYN2, respectively (Supplementary Fig. 10a, b). Single gene deletion and promoter replacement were performed using amdS as a selection marker, and transformants were selected on SM-Ac plate<sup>59</sup>. The deletion cassette was constructed by amplification of the amdS marker from the plasmid pUG-amdSYM with primer pairs containing regions homologous to the target gene. The standard LiAc/SS DNA/PEG method was used for yeast transformation<sup>60</sup>.

**Media and culture conditions.** Yeast strains were grown in YPD medium, YPE medium or SD-URA medium at 30 °C according to the auxotrophy of the cells, phenotypes of which were described in Supplementary Table 4<sup>61</sup>. Single gene deletion strains were selected on SM-Ac plate<sup>59</sup>. For  $\alpha$ -amylase production in tubes or shake flasks, yeast strains were cultured for 96 h at 200 rpm, 30 °C in the SD-2  $\times$  SCAA medium<sup>61</sup> containing 20 g l<sup>-1</sup> glucose, 6.9 g l<sup>-1</sup> yeast nitrogen base without amino acids, 190 mg l<sup>-1</sup> Arg, 400 mg l<sup>-1</sup> Asp, 1260 mg l<sup>-1</sup> Glu, 130 mg l<sup>-1</sup> Gly, 140 mg l<sup>-1</sup> His, 290 mg l<sup>-1</sup> Ile, 400 mg l<sup>-1</sup> Leu, 440 mg l<sup>-1</sup> Lys, 108 mg l<sup>-1</sup> Met, 200 mg l<sup>-1</sup> Phe, 220 mg l<sup>-1</sup> Thr, 40 mg l<sup>-1</sup> Trp, 52 mg l<sup>-1</sup> Tyr, 380 mg l<sup>-1</sup> Val, 1 g l<sup>-1</sup> BSA, 5.4 g l<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub>, and 8.56 g l<sup>-1</sup> NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O (pH = 6.0 by NaOH). For bioreactor batch cultures, 5.4 g l<sup>-1</sup> Na<sub>2</sub>HPO<sub>4</sub> and 8.56 g l<sup>-1</sup> NaH<sub>2</sub>PO<sub>4</sub>·H<sub>2</sub>O in the SD-2  $\times$  SCAA medium were replaced by 2 g l<sup>-1</sup> KH<sub>2</sub>PO<sub>4</sub> (pH = 6.0 by NaOH). Seed cultures were used to inoculated 500 ml SD-2  $\times$  SCAA medium in 1 l bioreactor vessels (DasGip, Germany) with an initial OD<sub>600</sub> of 0.01. The bioreactor system was run at 30 °C, 600 rpm agitation, 30 l h<sup>-1</sup> air flow, pH = 6 (controlled by NaOH). Biological quadruplicate (in some cases triplicate) experiments were conducted for each strain.

**Analytical procedures.** For dry cell weight (DCW) determination, yeast cells in 5 ml culture were harvested by a 0.45  $\mu$ m nitrocellulose filter and washed with distilled water. The DCW was measured until the cells were dried to a constant by 15 min microwave heating and then 3 days in a silica gel drier. The concentration of metabolites (glucose, ethanol, glycerol, etc.) in the culture was measured by loading the supernatant to a HPLC-87H column (Bio-Rad, USA) on a Dionex Ultimate 3000 HPLC system (Dionex Softron GmbH, Germany). The HPLC system was run at 45 °C with 5 mM H<sub>2</sub>SO<sub>4</sub> as mobile phase at a flow rate of 0.6 ml min<sup>-1</sup>.

**Protein quantification.** The  $\alpha$ -amylase activity was measured using the  $\alpha$ -amylase assay kit (Cat No. K-CERA; Megazyme, Ireland) with a commercial  $\alpha$ -amylase from *Aspergillus oryzae* (Cat No. 86250; Sigma, USA) as a standard. The weight of  $\alpha$ -amylase can be calculated with 69.6 U mg<sup>-1</sup> as protein conversion coefficient<sup>62</sup>. Five-hundred microliters cell cultures were centrifuged at 16,000 $\times$ g for separation of supernatant and cell pellet. Then the culture supernatant was used for determination of  $\alpha$ -amylase activity. For intracellular  $\alpha$ -amylase measurements, the cell pellet was washed with distilled water and resuspended in 500  $\mu$ l phosphate-buffered saline (PBS) with 5  $\mu$ l halt protease inhibitor cocktail (Cat No. 87786; Thermo Fisher, USA). The cell suspension was added to lysing matrix tube and cell lysis was performed using a FastPrep-24 tissue and cell homogenizer (MP Bio-medicals, USA) by two 60 s cycles at a speed of 6.5 m s<sup>-1</sup> (samples were put on ice for 5 min between the two cycles). Cell debris was removed by centrifugation and the supernatant fraction was used for  $\alpha$ -amylase quantification. Cells expressing HSA or XYN were cultured in SD-2  $\times$  SCAA medium without BSA. After cultivation, the supernatant was collected by centrifugation at 16,000 $\times$ g for sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS/PAGE) analysis<sup>16</sup>. SDS/PAGE gel was analyzed by the software ImageJ and concentration of HSA and XYN was estimated<sup>63</sup>.

**Transcriptome profiling.** Cell samples for RNA-seq were taken at the early exponential phase (OD<sub>600</sub>  $\approx$  1) and stored at -80 °C until processing<sup>64</sup>. RNA was extracted using the RNeasy Mini kit (Cat No. 74104; Qiagen, Germany) and prepared for sequencing using the Illumina TruSeq samples preparation kit v2, with poly-A enrichment. The fragments were clustered on cBot and sequenced on a single lane on a HiSeq 2500 with paired ends (2  $\times$  125 bp), according to the manufacturer's instructions. The number of read pairs obtained for each sample ranged from 5.9–10.1 million. The raw data can be downloaded from the European Nucleotide Archive with access number ERP019558. The raw reads from each samples were mapped to the CEN.PK 113-7D reference genome (<http://cenpk.tudelft.nl>) using TopHat (v 2.0.12)<sup>65</sup>, with 84.0–87.6% of the reads successfully mapped. Cufflinks version 2.2.1<sup>66</sup> was used to calculate FPKM values, and the feature Counts module of the Subread package (v 1.4.6)<sup>67</sup> was used to determine raw read counts. Analysis of differential expression was performed using DESeq<sup>68</sup>. Reporter GO terms and reporter TFs analysis were performed using the Platform for Integrative Analysis of Omics (PIANO) R package<sup>23</sup> with GO terms information from the Saccharomyces Genome Database (<http://www.yeastgenome.org>) and information on regulatory interactions between TFs and genes from Yeasttract (<http://www.yeasttract.com>). Differential expression level of genes

(log<sub>2</sub>FoldChange) and corresponding significant levels (adjusted *p*-values, calculated by the Benjamini–Hochberg method) were used as input. The reporter TFs analysis was used to calculate the significance for expression change of gene sets controlled by TFs, which were scored by the modulation in expression level of genes that are controlled by a given TFs<sup>22</sup>. The top five TFs, in distinct-directional up class and in distinct-directional down class, were chosen and presented in significances. The reporter GO terms analysis was similar to reporter TFs analysis, but gene sets were classified by GO terms. All GO terms were scored by modulation in expression level of genes within the same GO term. The Venn diagram was generated by using InteractiVenn<sup>69</sup>. GO slim mapper analysis was performed by using the SGD online tool according to the instruction (<http://www.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl>).

**ROS measurements.** The yeast cells were grown in SD-2 × SCAA medium and harvested at an OD<sub>600</sub> of 1–2 by centrifugation for ROS measurements. The cell pellets were washed with PBS twice, 50 mM sodium citrate buffer (SCB) once and resuspended in SCB to an OD<sub>600</sub> of 1. One milliliter cell suspension was combined with 1 µl 50 mM dihydrorhodamine 123 (DHR123) and incubated in the dark at room temperature for 30 min. Then, the cell pellet was washed twice with SCB and resuspended in 1 ml SCB for fluorescence intensity measurement and imaging. The fluorescence intensity was measured in a 96-well plate using a FLUOstar Omega microplate reader (BMG Labtech, Germany) at excitation wavelength 485 nm and emission wavelength 520 nm. Cells were imaged using a fluorescence microscope (Leica DMI4000B, Germany) with DIC and YFP filters.

**Data availability.** The RNA-seq raw data of the mutant strains and the reference strain can be downloaded from the European Nucleotide Archive with the access number ERP019558. The data that support the findings of this study are available from the corresponding author upon reasonable request.

Received: 23 January 2017 Accepted: 9 August 2017

Published online: 25 October 2017

## References

- Ellgaard, L., Molinari, M. & Helenius, A. Setting the standards: Quality control in the secretory pathway. *Science* **286**, 1882–1888 (1999).
- Hou, J., Tyo, K. E. J., Liu, Z. H., Petranovic, D. & Nielsen, J. Metabolic engineering of recombinant protein secretion by *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **12**, 491–510 (2012).
- Feizi, A., Osterlund, T., Petranovic, D., Bordel, S. & Nielsen, J. Genome-scale modeling of the protein secretory machinery in yeast. *PLoS ONE* **8**, e63284 (2013).
- Klein, T., Niklas, J. & Heinzel, E. Engineering the supply chain for protein production/secretion in yeasts and mammalian cells. *J. Ind. Microbiol. Biotechnol.* **42**, 453–464 (2015).
- Gitler, A. D. et al. The Parkinson's disease protein alpha-synuclein disrupts cellular Rab homeostasis. *Proc. Natl Acad. Sci. USA* **105**, 145–150 (2008).
- Gregersen, N., Bross, P., Vang, S. & Christensen, J. H. Protein misfolding and human disease. *Annu. Rev. Genomics Hum. Genet.* **7**, 103–124 (2006).
- Schekman, R. & Novick, P. 23 genes, 23 years later. *Cell* **116**, S13–S15 (2004).
- Jonikas, M. C. et al. Comprehensive characterization of genes required for protein folding in the endoplasmic reticulum. *Science* **323**, 1693–1697 (2009).
- Stalder, D., Mizuno-Yamasaki, E., Ghasseman, M. & Novick, P. J. Phosphorylation of the Rab exchange factor Sec2p directs a switch in regulatory binding partners. *Proc. Natl Acad. Sci. USA* **110**, 19995–20002 (2013).
- Huang, M., Bao, J. & Nielsen, J. Biopharmaceutical protein production by *Saccharomyces cerevisiae*: current state and future prospects. *Pharm. Bioprocess.* **2**, 167–182 (2014).
- Demain, A. L. & Vaishnav, P. Production of recombinant proteins by microbes and higher organisms. *Biotechnol. Adv.* **27**, 297–306 (2009).
- Grav, L. M. et al. One-step generation of triple knockout CHO cell lines using CRISPR/Cas9 and fluorescent enrichment. *Biotechnol. J.* **10**, 1446–1456 (2015).
- de Ruijter, J. C., Koskela, E. V. & Frey, A. D. Enhancing antibody folding and secretion by tailoring the *Saccharomyces cerevisiae* endoplasmic reticulum. *Microb. Cell Fact.* **15**, 87 (2016).
- Tang, H. et al. Engineering protein folding and translocation improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **112**, 1872–1882 (2015).
- Bao, J., Huang, M., Petranovic, D. & Nielsen, J. Moderate expression of SEC16 increases protein secretion by *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **83**, e03400–e03416 (2017).
- Huang, M. et al. Microfluidic screening and whole-genome sequencing identifies mutations associated with improved protein secretion by yeast. *Proc. Natl Acad. Sci. USA* **112**, E4689–E4696 (2015).
- Nilsson, A. & Nielsen, J. Metabolic trade-offs in yeast are caused by F1F0-ATP synthase. *Sci. Rep.* **6**, 22264 (2016).
- Pelechano, V. & Perez-Ortin, J. E. There is a steady-state transcriptome in exponentially growing yeast cells. *Yeast* **27**, 413–422 (2010).
- Abramova, N. E. et al. Regulatory mechanisms controlling expression of the DAN/TIR mannoprotein genes during anaerobic remodeling of the cell wall in *Saccharomyces cerevisiae*. *Genetics* **157**, 1169–1177 (2001).
- Kwast, K. E., Burke, P. V., Brown, K. & Poyton, R. O. *REO1* and *ROX1* are alleles of the same gene which encodes a transcriptional repressor of hypoxic genes in *Saccharomyces cerevisiae*. *Curr. Genet.* **32**, 377–383 (1997).
- Ljungdahl, P. O. & Daignan-Fornier, B. Regulation of amino acid, nucleotide, and phosphate metabolism in *Saccharomyces cerevisiae*. *Genetics* **190**, 885–929 (2012).
- Oliveira, A. P., Patil, K. R. & Nielsen, J. Architecture of transcriptional regulatory circuits is knitted over the topology of bio-molecular interaction networks. *BMC Syst. Biol.* **2**, 17 (2008).
- Varemo, L., Nielsen, J. & Nookaew, I. 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 (2013).
- Sertil, O., Kapoor, R., Cohen, B. D., Abramova, N. & Lowry, C. V. Synergistic repression of anaerobic genes by Mot3 and Rox1 in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **31**, 5831–5837 (2003).
- Hickman, M. J. & Winston, F. Heme levels switch the function of Hap1 of *Saccharomyces cerevisiae* between transcriptional activator and transcriptional repressor. *Mol. Cell Biol.* **27**, 7414–7424 (2007).
- Zhang, L. & Hach, A. Molecular mechanism of heme signalling in yeast: the transcriptional activator Hap1 serves as the key mediator. *Cell. Mol. Life. Sci.* **56**, 415–426 (1999).
- Zaman, S., Lippman, S. I., Zhao, X. & Broach, J. R. How *Saccharomyces* responds to nutrients. *Annu. Rev. Genet.* **42**, 27–81 (2008).
- Pinson, B. et al. Metabolic intermediates selectively stimulate transcription factor interaction and modulate phosphate and purine pathways. *Genes Dev.* **23**, 1399–1407 (2009).
- Iyer, V. R. et al. Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* **409**, 533–538 (2001).
- Liu, L., Zhang, Y., Liu, Z., Petranovic, D. & Nielsen, J. Improving heterologous protein secretion in aerobic conditions by activating hypoxia induced genes in *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **15**, fov070 (2015).
- Yu, J. et al. Enhanced expression of heterologous inulinase in *Kluyveromyces fragilis* by disruption of hap1 gene. *Biotechnol. Lett.* **32**, 507–512 (2010).
- Klein, T. et al. Overcoming the metabolic burden of protein secretion in *Schizosaccharomyces pombe*—A quantitative approach using <sup>13</sup>C-based metabolic flux analysis. *Metab. Eng.* **21**, 34–45 (2014).
- Tate, E. W., Kalesh, K. A., Lanyon-Hogg, T., Storck, E. M. & Thinon, E. Global profiling of protein lipidation using chemical proteomic technologies. *Curr. Opin. Chem. Biol.* **24**, 48–57 (2015).
- Lewis, D. A. & Bisson, L. F. The *HXT1* gene product of *Saccharomyces cerevisiae* is a new member of the family of hexose transporters. *Mol. Cell Biol.* **11**, 3804–3813 (1991).
- Fendt, S. M. et al. Tradeoff between enzyme and metabolite efficiency maintains metabolic homeostasis upon perturbations in enzyme capacity. *Mol. Syst. Biol.* **6**, 356 (2010).
- Guillamon, J. M., van Riel, N. A. W., Giuseppe, M. L. F. & Verrips, C. T. The glutamate synthase (GOGAT) of *Saccharomyces cerevisiae* plays an important role in central nitrogen metabolism. *FEMS Yeast Res.* **1**, 169–175 (2001).
- Morin, P. J., Subramanian, G. S. & Gilmore, T. D. *AAT1*, a gene encoding a mitochondrial aspartate aminotransferase in *Saccharomyces cerevisiae*. *Biochim. Biophys. Acta* **1171**, 211–214 (1992).
- Penalosa-Ruiz, G. et al. Paralogous *ALT1* and *ALT2* retention and diversification have generated catalytically active and inactive aminotransferases in *Saccharomyces cerevisiae*. *PLoS ONE* **7**, e45702 (2012).
- Kispal, G., Steiner, H., Court, D. A., Rolinski, B. & Lill, R. Mitochondrial and cytosolic branched-chain amino acid transaminases from yeast, homologs of the *myc* oncogene-regulated Eca39 protein. *J. Biol. Chem.* **271**, 24458–24464 (1996).
- Kaur, J. & Bachhawat, A. K. Yct1p, a novel, high-affinity, cysteine-specific transporter from the yeast *Saccharomyces cerevisiae*. *Genetics* **176**, 877–890 (2007).
- Gao, X. D., Wang, J., Keppler-Ross, S. & Dean, N. *ERS1* encodes a functional homologue of the human lysosomal cystine transporter. *FEBS J.* **272**, 2497–2511 (2005).
- Nosaka, K. et al. Genetic regulation mediated by thiamin pyrophosphate-binding motif in *Saccharomyces cerevisiae*. *Mol. Microbiol.* **58**, 467–479 (2005).
- Vuralhan, Z., Morais, M. A., Tai, S. L., Piper, M. D. W. & Pronk, J. T. Identification and characterization of phenylpyruvate decarboxylase genes in *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **69**, 4534–4541 (2003).



44. Malhotra, J. D. et al. Antioxidants reduce endoplasmic reticulum stress and improve protein secretion. *Proc. Natl Acad. Sci. USA* **105**, 18525–18530 (2008).
45. Travers, K. J. et al. Functional and genomic analyses reveal an essential coordination between the unfolded protein response and ER-associated degradation. *Cell* **101**, 249–258 (2000).
46. Valkonen, M., Penttilä, M. & Saloheimo, M. Effects of inactivation and constitutive expression of the unfolded-protein response pathway on protein production in the yeast *Saccharomyces cerevisiae*. *Appl. Environ. Microbiol.* **69**, 2065–2072 (2003).
47. Sengupta, N., Rose, S. T. & Morgan, J. A. Metabolic flux analysis of CHO cell metabolism in the late non-growth phase. *Biotechnol. Bioeng.* **108**, 82–92 (2011).
48. Nocon, J. et al. Increasing pentose phosphate pathway flux enhances recombinant protein production in *Pichia pastoris*. *Appl. Microbiol. Biotechnol.* **100**, 5955–5963 (2016).
49. Herrero, P., Galindez, J., Ruiz, N., Martinezcampa, C. & Moreno, F. Transcriptional regulation of the *Saccharomyces cerevisiae* *HXK1*, *HXK2* and *GLK1* genes. *Yeast* **11**, 137–144 (1995).
50. Solis-Escalante, D. et al. A minimal set of glycolytic genes reveals strong redundancies in *Saccharomyces cerevisiae* central metabolism. *Eukaryot. Cell* **14**, 804–816 (2015).
51. Kimata, Y., Ishiwata-Kimata, Y., Yamada, S. & Kohno, K. Yeast unfolded protein response pathway regulates expression of genes for anti-oxidative stress and for cell surface proteins. *Genes Cells* **11**, 59–69 (2006).
52. Wentz, A. E. & Shusta, E. V. A novel high-throughput screen reveals yeast genes that increase secretion of heterologous proteins. *Appl. Environ. Microbiol.* **73**, 1189–1198 (2007).
53. Harbison, C. T. et al. Transcriptional regulatory code of a eukaryotic genome. *Nature* **431**, 99–104 (2004).
54. Hochstrasser, M. Origin and function of ubiquitin-like proteins. *Nature* **458**, 422–429 (2009).
55. Kern, D. et al. How thiamine diphosphate is activated in enzymes. *Science* **275**, 67–70 (1997).
56. Treusch, S. et al. Functional links between A $\beta$  toxicity, endocytic trafficking, and alzheimer's disease risk factors in yeast. *Science* **334**, 1241–1245 (2011).
57. Kritzer, J. A. et al. Rapid selection of cyclic peptides that reduce alpha-synuclein toxicity in yeast and animal models. *Nat. Chem. Biol.* **5**, 655–663 (2009).
58. Chung, C. Y. et al. Identification and rescue of  $\alpha$ -synuclein toxicity in parkinson patient-derived neurons. *Science* **342**, 983–987 (2013).
59. Solis-Escalante, D. et al. amdSYM, a new dominant recyclable marker cassette for *Saccharomyces cerevisiae*. *FEMS Yeast Res.* **13**, 126–139 (2013).
60. Gietz, R. D. & Woods, R. A. Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method. *Methods Enzymol.* **350**, 87–96 (2002).
61. Hou, J., Tyo, K., Liu, Z., Petranovic, D. & Nielsen, J. Engineering of vesicle trafficking improves heterologous protein secretion in *Saccharomyces cerevisiae*. *Metab. Eng.* **14**, 120–127 (2012).
62. Liu, Z., Tyo, K. E. J., Martinez, J. L., Petranovic, D. & Nielsen, J. Different expression systems for production of recombinant proteins in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **109**, 1259–1268 (2012).
63. Abramoff, M. D., Magalhães, P. J. & Ram, S. J. Image processing with ImageJ. *Biophoton. Int.* **11**, 36–42 (2004).
64. Usaite, R., Patil, K. R., Grotkjær, T., Nielsen, J. & Regenberg, B. Global transcriptional and physiological responses of *Saccharomyces cerevisiae* to ammonium, L-alanine, or L-glutamine limitation. *Appl. Environ. Microbiol.* **72**, 6194–6203 (2006).
65. Kim, D. et al. TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* **14**, R36 (2013).
66. Trapnell, C. et al. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat. Biotechnol.* **28**, 511–515 (2010).
67. Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).
68. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11**, R106 (2010).
69. Heberle, H., Meirelles, G. V., da Silva, F. R., Telles, G. P. & Minghim, R. InteractVenn: a web-based tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* **16**, 169 (2015).

## Acknowledgements

This work was funded by the Novo Nordisk Foundation, Vetenskapsrådet, FORMAS, and Knut and Alice Wallenberg Foundation. We would like to acknowledge the Science for Life Laboratory, the National Genomics Infrastructure (NGI), and Uppmax for providing assistance in massive parallel sequencing and computational infrastructure. We also thank Verena Siewers, Yongjin Zhou, Guokun Wang, Jiufu Qin, Mark Bishops, Yun Chen, and José L. Martínez for useful discussions and comments.

## Author contributions

M.H. and J.N. conceived and designed the study. M.H. and J.B. performed experiments. M.H. and B.M.H. performed bioinformatics analysis. M.H., D.P., and J.N. analyzed the data and wrote the paper.

## Additional information

**Supplementary Information** accompanies this paper at doi:10.1038/s41467-017-00999-2.

**Competing interests:** The authors declare no competing financial interests.

**Reprints and permission** information is available online at <http://npg.nature.com/reprintsandpermissions/>

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017