nature microbiology

ARTICLES

PUBLISHED: 1 FEBRUARY 2016 | ARTICLE NUMBER: 15030 | DOI: 10.1038/NMICROBIOL.2015.30

The metabolic background is a global player in *Saccharomyces* gene expression epistasis

Mohammad Tauqeer Alam[1†], Aleksej Zelezniak[1,2†], Michael Mülleder[1†], Pavel Shliaha[1,3], Roland Schwarz[4], Floriana Capuano[1], Jakob Vowinckel[1], Elahe Radmaneshfar[2], Antje Krüger[5], Enrica Calvani[1], Steve Michel[5], Stefan Börno[5], Stefan Christen[6], Kiran Raosaheb Patil[7], Bernd Timmermann[5], Kathryn S. Lilley[1,3] and Markus Ralser[1,2*]

The regulation of gene expression in response to nutrient availability is fundamental to the genotype–phenotype relationship. The metabolic-genetic make-up of the cell, as reflected in auxotrophy, is hence likely to be a determinant of gene expression. Here, we address the importance of the metabolic-genetic background by monitoring transcriptome, proteome and metabolome in a repertoire of 16 *Saccharomyces cerevisiae* laboratory backgrounds, combinatorially perturbed in histidine, leucine, methionine and uracil biosynthesis. The metabolic background affected up to 85% of the coding genome. Suggesting widespread confounding, these transcriptional changes show, on average, 83% overlap between unrelated auxotrophs and 35% with previously published transcriptomes generated for non-metabolic gene knockouts. Background-dependent gene expression correlated with metabolic flux and acted, predominantly through masking or suppression, on 88% of transcriptional interactions epistatically. As a consequence, the deletion of the same metabolic gene in a different background could provoke an entirely different transcriptional response. Propagating to the proteome and scaling up at the metabolome, metabolic background dependencies reveal the prevalence of metabolism-dependent epistasis at all regulatory levels. Urging a fundamental change of the prevailing laboratory practice of using auxotrophs and nutrient supplemented media, these results reveal epistatic intertwining of metabolism with gene expression on the genomic scale.

Metabolism is the largest functional system within a cell, and as metabolic reactions are connected over a flux of metabolites, it assembles in a highly connected network[1–7]. Metabolic activity needs to be adapted constantly to match cellular physiology, nutrition, growth rate and stress situations. This dual dependency on both cell-extrinsic and -intrinsic properties renders metabolism a key mediator of gene–environment interactions, while its size represents a quantitative factor in physiology and gene expression[8–10]. Enumerating the total compendium of metabolism-responsive genes is a difficult task, but transcriptional changes that follow the metabolic oscillations of *Saccharomyces cerevisiae* suggest that it could comprise more than 50% of the genome[11].

One difficulty in studying genetic–metabolic interactions is caused by the minimal redundancy within the metabolic network. Other than secondary metabolic pathways, most metabolic systems cannot be perturbed without system-wide consequences. Exceptions to this are some metabolic pathways of amino-acid and nucleobase biosynthesis, for which cells possess a preference for uptake over self-synthesis for the product metabolites. These biosynthetic pathways can be perturbed as long as the product is provided extracellularly[12]. Single-gene auxotrophies in such pathways have been established as effective selection markers for genetic experiments, so they have been crossed into a large number of laboratory strains. In the present work we exploit four such markers to study the importance of the metabolic background on gene expression in *S. cerevisiae* and study the combinatorial effects resulting from *HIS3* (ref. 13), *LEU2* (ref. 14), *URA3* (ref. 15) or *MET15* (ref. 16) deletions on the transcriptome, proteome and metabolome.

We report that metabolic background differences induce strong but adaptive molecular signatures, even when growth is restored by external nutrient supplementation. Gene expression is affected in a metabolism-dependent manner, and on 88% of transcriptional events involving 77% of the differentially expressed transcripts we detect evidence for epistatic interactions occurring between metabolic genes. These interactions are found to have a fundamental impact on the gene expression response that follows gene deletions, and reveal the metabolic genotype (or metabotype) to be, on a global scale, responsible for context-dependent biological responses on the transcriptome, proteome and metabolome. The global dependency on metabotype substantiates an upstream, dynamic and key role for cellular metabolic make-up in gene expression regulation. Metabolic–genetic background differences, dismissively considered to be harmless in pre-genomic times, could therefore have affected the outcome of a large number of experiments.

Results

Molecular signatures of the metabolic–genetic background. Previously, we have reported that even when *S. cerevisiae* histidine, leucine, uracil and methionine auxotrophies are complemented, as it is done in a typical laboratory experiment, some physiological

[1]Department of Biochemistry and Cambridge Systems Biology Centre, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK. [2]The Francis Crick Institute, Mill Hill Laboratory, London NW7 1AA, UK. [3]Cambridge Centre for Proteomics, Department of Biochemistry, University of Cambridge, 80 Tennis Court Road, Cambridge CB2 1GA, UK. [4]European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, UK. [5]Max Planck Institute for Molecular Genetics, Ihnestrasse 73, Berlin, Germany. [6]Department of Molecular Systems Biology, Eidgenoessische Technische Hochschule, 8093 Zürich, Switzerland. [7]European Molecular Biology Laboratory, EMBL, 69117 Heidelberg, Germany. [†]These authors contributed equally to this work. *e-mail: markus.ralser@crick.ac.uk

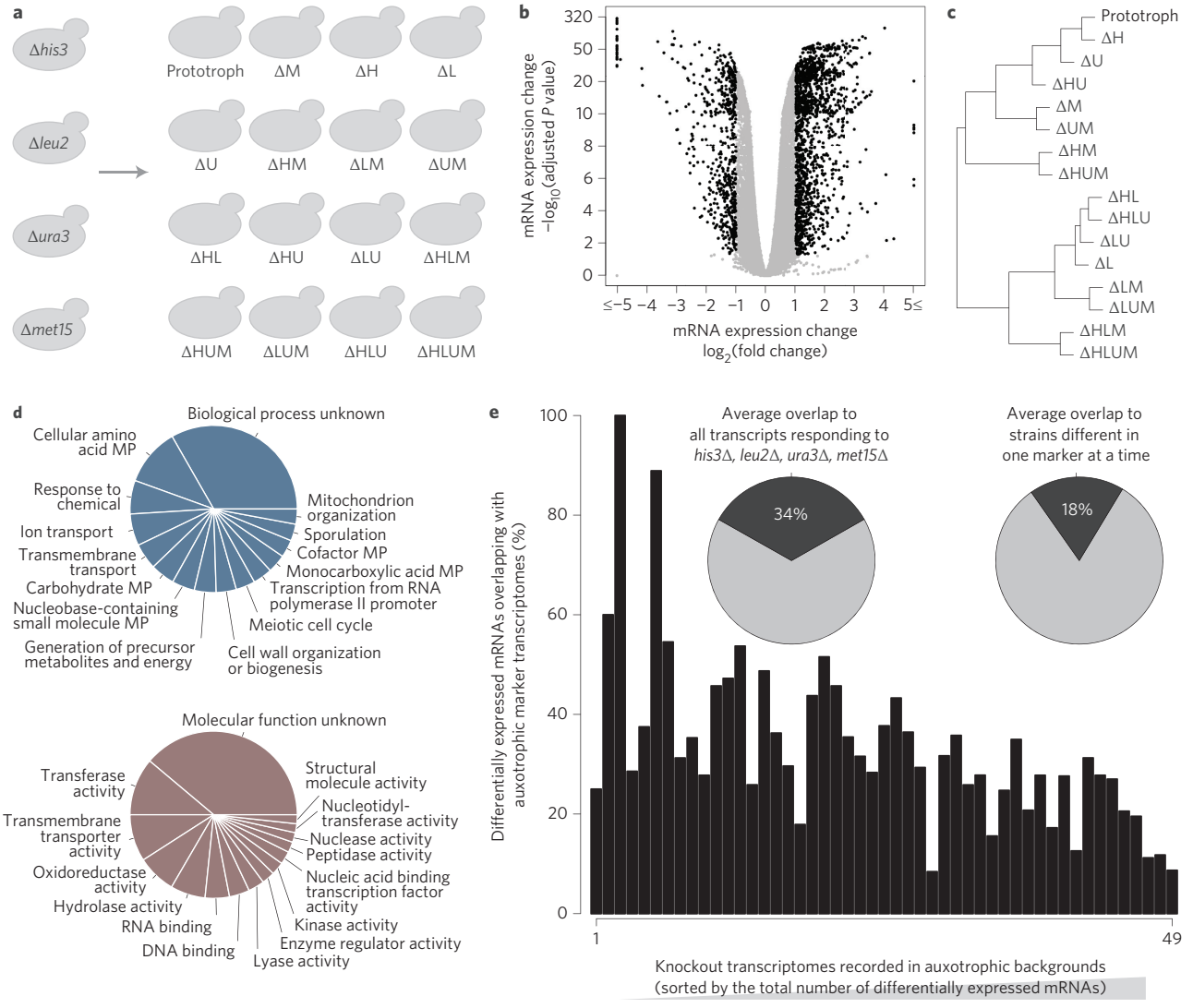

Figure 1 | Gene expression response to 16 combinatorial differences in the metabolic–genetic background. a, Schematic overview of otherwise isogenic 16 *S. cerevisiae* strains possessing nutrient-complemented auxotrophies in histidine (*his3*Δ), leucine (*leu2*Δ), uracil (*ura3*Δ) and methionine (*met15*Δ) biosynthesis. **b**, Differential gene expression with reference to the prototrophic yeast (volcano plot) illustrating gene expression ($\log_2$ scaling) and significance ($-\log_{10}$ scaling, BH adj. *P* value) for the 16 metabotypes grown in synthetic complete medium, as determined by RNA sequencing ($n = 3$/strain). **c**, Hierarchical clustering of mRNA expression profiles by means of Euclidean distance and complete linkage agglomeration, dividing the 16 strains first by leucine, followed by methionine auxotrophy. **d**, Top 15 enriched Gene Ontology (GO):Process slim mapper (top) and GO:Function slim terms (bottom), across 573 differentially expressed transcripts with greater than twofold change, adj. $P < 0.05$ with reference to the prototrophic strain. MP, metabolic process. **e**, Confounding effects of metabotype on transcriptome. Overlap between genes highly and significantly (fold change >2, adj. $P < 0.05$) differentially expressed in the auxotrophic strains and those detected to be differentially expressed across previously published microarray experiments conducted on single gene deletions in the BY4741 auxotrophic background. Expression profiles are sorted according to number of differentially expressed transcripts (ascending). Insets: Average overlaps between the ArrayExpress data and the transcriptomes obtained for all four auxotrophic markers present in BY4741, or those differing in just one marker at a time.

differences are retained. These manifest as small to moderate differences in growth rate, some of which show evidence for epistatic interactions between the metabolic markers[17]. Starting with an in-depth analysis of this growth data, we found that in supplemented medium, *LEU2* has a consistent effect on growth rate, and the other markers exert minor effects that reveal themselves only in a context- (or background-) dependent manner. The overall growth rates are therefore well explained by the leucine effect using either an additive model (Supplementary Fig. 1a) or a multiple linear regression model that uses *HIS3*, *LEU2*, *URA3* and *MET15* as categorical predictors (adj. $R^2 = 0.86$, $P = 2.18 \times 10^{-5}$; Supplementary Fig. 1a,b).

The molecular levels revealed a much more differentiated picture. mRNA expression profiles were obtained from 16 strains in triplicate exponentially grown cultures, each with an identical cell density (optical density at 600 nm (OD_{600}) of 0.8), by mRNA sequencing, resulting in highly reproducible expression profiles (Supplementary Fig. 2). A total of 5,011 transcripts from 5,923 expressed mRNAs (85% of the transcriptome) were significantly differentially expressed (adj. *P* values (Benjamini and Hochberg, BH, method) of <0.05) in auxotrophic strains compared with the prototroph (Fig. 1b). A global transcriptional signature was corroborated by robust median normalization (Supplementary Fig. 3a). Hierarchical clustering revealed the strongest separation by the *LEU2* gene, followed by the *MET15* gene, indicating that these two perturbations leave the most consistent signature in the transcriptome (Fig. 1c). In total, 573 transcripts (9.7% of the transcriptome) were differentially expressed, not only significantly,

but more than twofold (Fig. 1b). These were enriched for metabolic activity (Gene Ontology (GO) Process terms) and enzymatic function (GO Function terms) (Fig. 1d) and, according to hypergeometric testing, for amino acid and carbohydrate metabolic pathways (Supplementary Fig. 3b). Thus, auxotrophic background differences are reflected by gene expression differences detected on three-quarters of the coding genome, with ~600 mostly metabolism-associated genes being strongly differentially expressed.

As *HIS3*, *LEU2*, *MET15* and *URA3* have frequently been exploited as genetic selection markers, these results suggest that metabolic transcriptional signatures could have confounded previous gene expression experiments. mRNA expression data collected for a variety of single gene knockouts (the vast majority being non-metabolic genes) in different contexts and laboratories, but all generated in auxotrophic BY4741 backgrounds (listed in Supplementary File 1), were obtained from ArrayExpress[18] and reprocessed to achieve identical cutoff criteria (fold change > 2, BH adj. $P < 0.05$). On average, 34% of differentially expressed mRNAs overlapped with those induced by *HIS3*, *LEU2*, *MET15* or *URA3* deletions greater than twofold (Fig. 1e). Even if strains that differ only by a single marker are compared, 18% of transcriptional changes, on average, overlap (Fig. 1e). A similar picture was obtained for a further set of ~70 microarray experiments that studied different conditions (Supplementary Fig. 4a and Supplementary File 1). We notice a significant correlation ($r = -0.52$, $P = 0.0001186$), such that transcription profiles with a larger number of differentially expressed genes are better distinguished from the metabolic background signature (Fig. 1e).

Next, we questioned whether this overlap would be the same, or larger, for metabolic genes. We compared each combination of *HIS3*, *LEU2*, *URA3* or *MET15* deletions with differential gene expression induced by the other markers. A substantially larger overlap of 83% was detected (Supplementary Fig. 4b). Hence, a notable proportion of differential gene expression overlaps with the transcriptional signatures of the metabolic markers. On comparing more than a hundred transcriptomes, we detected, on average, an overlap of approximately one-third for cases where a non-metabolic gene was deleted and, on the basis of our data set, approximately three-quarters in cases where one of the auxotrophy-causing metabolic genes was deleted.

Metabolism-induced gene expression signatures are context-dependent and correlate with flux. We noted that expression signatures were qualitatively and quantitatively dependent not only on the metabolic deficiencies, but also on their combination. First, strains possessing three or four auxotrophies did not have more transcripts induced than strains possessing one or two deficiencies (Fig. 2a and Supplementary Fig. 3c). This result was confirmed by normalization strategies referred to the median, ruling out this result being a consequence of wild-type bias (Supplementary Fig. 3e,f). In addition, different transcripts responded to individual or to combinatorial perturbation. For instance, 112 transcripts were differentially expressed exclusively in the combinatorial knockouts, but not in the corresponding single knockouts (adj. $P < 0.05$, fold change of 2, Fig. 2c), while 128 transcripts were differentially expressed solely in a unique strain (Fig. 2b). Qualitatively similar results were obtained without fold-change cutoffs (Supplementary Fig. 3g,h).

These results raised the question of whether specific transcriptionally responsive genes could be assigned to *HIS3*, *URA3*, *LEU2* and *MET15* deletions, or whether the transcriptome responds differently depending on the metabolic background. We compared, four times, eight complementary strain pairs that differed from each other only in one auxotrophy (graphically exemplified for *HIS3* in Fig. 2d). Each gene deletion induced a strong transcriptional response (Supplementary Fig. 6a,b), but in each background, different gene sets responded to the same gene deletion (Fig. 2d). Indeed, universal targets were the exception. Virtually the only consistent hits were the deleted genes themselves (Fig. 2d). Consistently, transcriptional changes induced by deletion of the same gene in different backgrounds did not correlate with each other, except for a small subset of *LEU2* responsive genes (Supplementary Fig. 6d). An analogous picture was obtained when considering all significantly differentially expressed transcripts, ruling out a thresholding bias (Supplementary Fig. 6b,c).

We speculated that strain-specific transcriptional profiles might originate from different metabolic flux. Context-dependent gene expression changes did correlate strongly ($r = 0.78$, $P = 5.77 \times 10^{-4}$) with flux, as determined by flux variability analysis[19,20] on constraining the model with experimentally measured amino-acid uptake and growth rates for the 16 strains (Fig. 2e, Supplementary Note 2 and Supplementary File 2). Metabolic reactions with the highest correlation between gene expression and calculated fluxes were enriched for intermediate metabolic pathways (Supplementary Fig. 7).

Metabolic perturbations interact epistatically. In principle, a target transcript could respond irrespective of whether a metabolic pathway is perturbed alone or in combination, or the response could be sensitive to epistatic interactions between the pathways. Of the differentially expressed transcripts, 77% responded in more than one auxotroph and thus may fall into one of these categories (Fig. 2c). To identify the epistatic interactions, we applied both additive models, as introduced by Fisher[21], and multiplicative models. On our large and systematic data set, both strategies yielded to a large extent (97%) the same transcripts when applied on gene expression data scaled to fold changes and considering only genes that were both significantly (adj. $P < 0.05$) and highly (greater than twofold) differentially expressed (Fig. 3a). In accordance with the growth rates, we therefore subdivided the epistatic interactions using the additive model (Fig. 3b). Epistatic interactions were detected on 88% of transcripts differentially expressed in two or more auxotrophs. Most frequent were suppressive or masking (dominant) interactions (Fig. 3d). Suppressive interactions were distributed among all markers, while masking interactions were dominated by *LEU2* (Fig. 3d, right panel). The second most frequent were positive and negative interactions, in which the effect of two alleles acting in concert was stronger or weaker than the value expected from the individual alleles (Fig. 3d). Their case was complemented by a special case of suppression: we applied the term 'pseudo-masking' when a second mutation weakened the effect of a masking allele. Illustrating the dynamic and context-dependent nature of epistasis, 37% of transcripts falling into a particular epistatic category in one case could fall into any category in another, depending on the metabolic background (Fig. 3e).

Highly connected and conserved genes are buffered against metabolism-induced epistasis. It has been suggested that most connected nodes ('hubs') in a genetic interaction network are more stable in terms of expression change compared with less connected genes[22] and are thus less likely to be affected by epistatic interactions. Indeed, epistatically responding transcripts were significantly less connected in genetic interaction[23] and protein–protein interaction networks (Fig. 3f), but were more likely to be co-expressed[24] (Supplementary Fig. 8g). In contrast, GO terms and functional classification (Supplementary Fig. 8a), co-citation frequency, co-occurrence of protein domains and three-dimensional protein structure networks of interacting orthologous proteins as obtained from YeastNet.v3[24] (Supplementary Fig. 8b–f) did not differ between epistatically responding and non-responding transcripts.

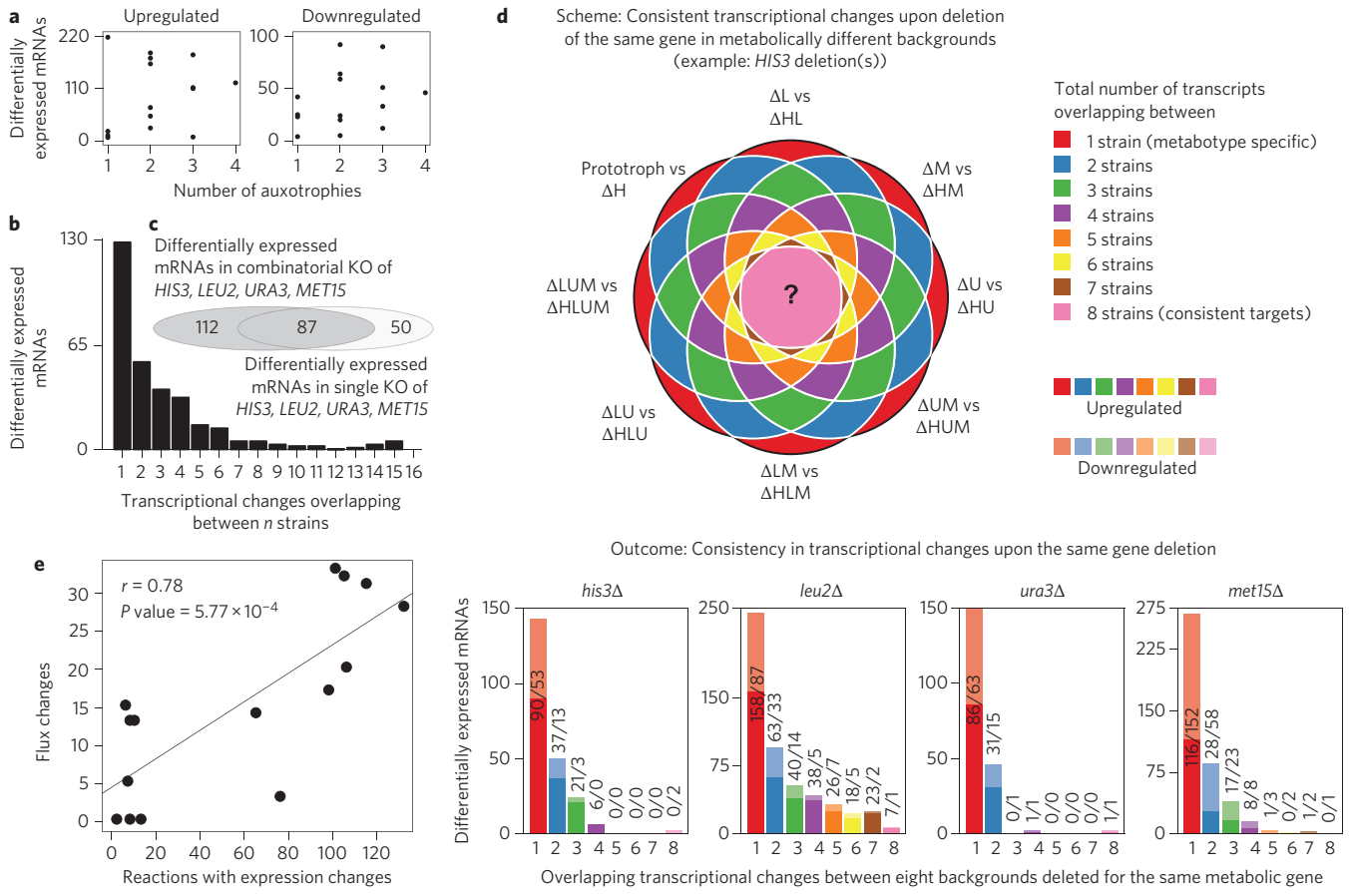

Figure 2 | Transcriptional response to a metabolic gene deletion is sensitive to the metabolic-genetic background. a, The number of auxotrophies is no predictor for total differential gene expression quantities. Numbers are given in relation to the prototrophic strain; the analysis relative to the expression median is given in Supplementary Fig. 3c. **b**, Specific transcriptional profiles dominate over commonly regulated genes; the majority of transcripts respond specifically in one or two of the 16 strains, despite differing in just four metabolic genes. **c**, Overlap of gene expression changes (adj. $P < 0.05$, fold change 2) induced when *HIS3*, *LEU2*, *URA3* or *MET15* were deleted separately or in combination (adj. $P < 0.05$, fold change >2; see Supplementary Fig. 3g, without fold-change cutoff). **d**, Transcriptional response on deleting the same gene (illustrated for *HIS3*) in a different metabotype reveals the dominance of context-specific transcriptional responses. Genes that consistently respond to the same deletion (in up to eight cases) are the exception. Lighter colours indicate downregulated genes, and opaque colours represent upregulated mRNAs (adj. $P < 0.05$, fold change >2. Supplementary Fig. 6c shows the same analysis without fold-change cutoff). **e**, Flux variability analysis predicts context-dependent flux changes that correlate with strain-specific gene expression. There is a significant correlation ($r = 0.78$, $P = 5.77 \times 10^{-4}$) between the number of reactions in which at least one gene is differentially expressed and the number of reactions for which the flux range is significantly different.

It has also been observed that more important and conserved genes are more stable to expression changes and that epistasis determines sequence conservation[25,26]. We compared the evolutionary conservation and essentiality[27] of affected and unaffected transcripts. Transcripts sensitive to metabolic epistasis had significantly fewer orthologues as identified by PSI-BLAST ($P = 1.6 \times 10^{-7}$; Fig. 3g) and were significantly less likely to be essential ($P = 1.04 \times 10^{-17}$; Fig. 3h). Epistatic metabolic interactions thus manifest prevalently on the least connected and conserved genes, while highly conserved genes and most connected genes appear to be buffered.

Metabolic epistasis translates into the proteome and amplifies at the metabolome. Next, we generated protein expression profiles via liquid chromatography tandem mass spectrometry (LC-MS/MS). We chose a data-independent acquisition (DIA) strategy ($HDMS^E$)[28,29] that identifies fewer proteins than shotgun methods, but is advantageous in large-sample series as the same peptides are consistently quantified in each injection replicate. Furthermore, facilitated by the large systematic data set, we improved the precision compared with conventional strategies by using covariance statistics to choose peptides best suited for protein quantification. Across the 48 proteomes, this strategy yielded precise and reproducible quantities for 442 proteins associated with 446 genes (Supplementary Fig. 9). Applying the same criteria as used for transcriptomics, 11% of proteins were found to be differentially expressed (Fig. 4a). Hierarchical clusters of transcriptome and proteome were fully correlated (co-phylogenetic correlation = 0.92) and divided the 16 strains in a similar fashion (Fig. 4b). The correlation of principal components revealed coherent regulation at the transcriptome and proteome, and the first principal components, explaining 26.36 and 46.52% of the overall variation, were fully correlated ($r = 0.96$, $P = 6.3 \times 10^{-9}$; Fig. 4c). Both regulatory layers were correlated with growth rates (Supplementary Fig. 10), suggesting a common regulatory response for growth and metabolism.

Furthermore, in strains with a high level of differential gene expression, proteome and transcriptome abundance values corresponded with one another (Pearson correlation coefficient (PCC) > 0.7; Fig. 4d). A weak correlation could be observed in the strains in which few differentially expressed proteins were captured (Fig. 4d). Intriguingly, for differentially expressed genes also, the fold changes significantly correlated. This quantitative correlation

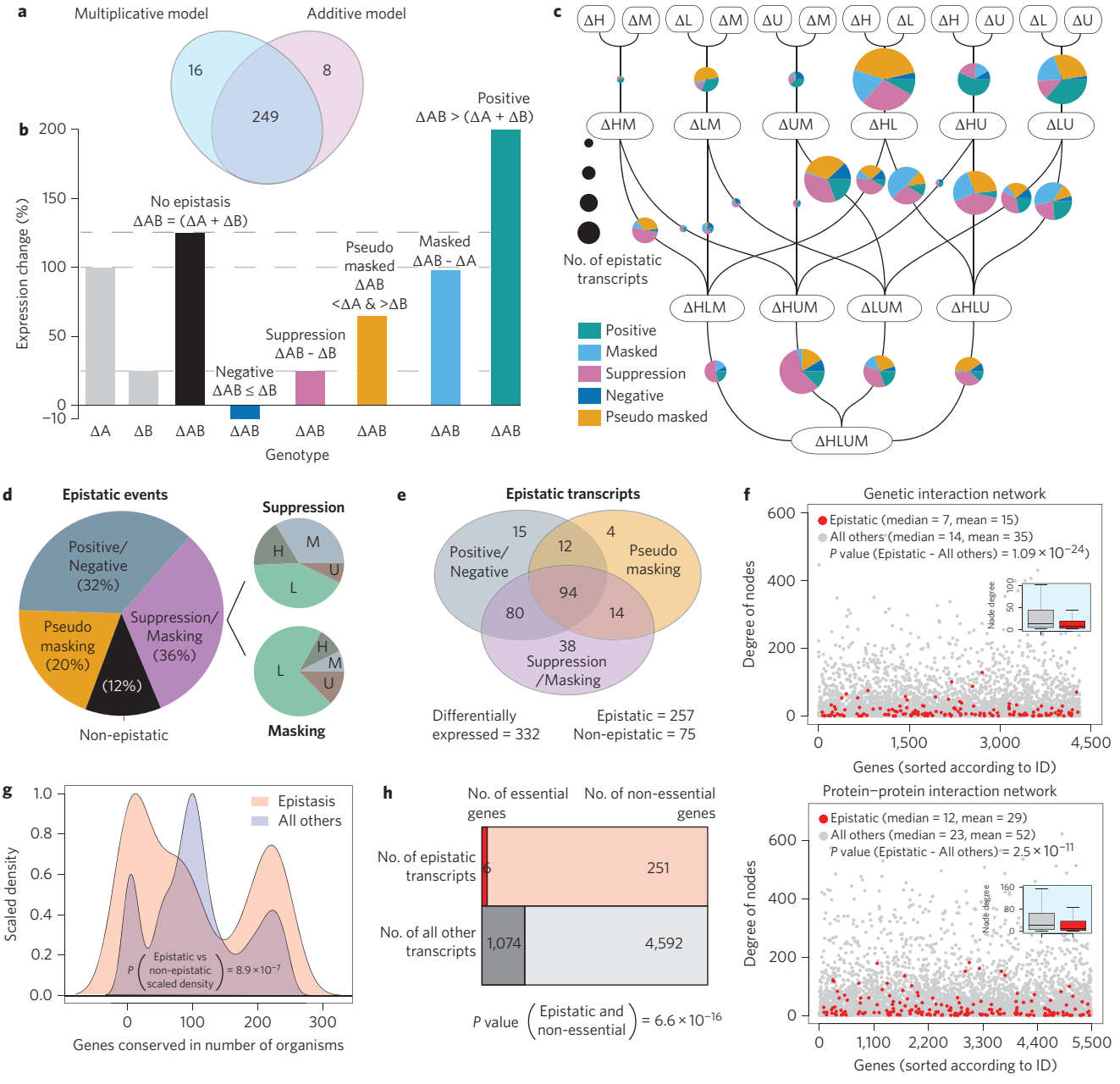

Figure 3 | Metabolic perturbations interact epistatically and dominate quantitative expression profiles. a, Multiplicative and additive models identify 97% of the same transcripts being affected by metabolism-induced epistasis, when comparing differentially expressed transcripts (greater than twofold change; adj. $P < 0.05$) with the median expression value of each gene. **b**, Schematic illustration of the classification of epistatic phenotypes on the basis of a linear model. **c**, The 16 strains aligned in a neighbourhood network to illustrate the 38 neighbourhood pairs that differ from each other in just one metabolic gene at a time. mRNA expression profiles were compared pairwise to identify and classify the epistatic gene expression changes in **b**. Pie charts, where size represents the number of epistatically responding transcripts identified in each neighbourhood and their distribution between the categories as defined in **b**. **d**, Classification of epistatic events. Additive events (12%), considered non-epistatic according to ref. 21, contrast with epistatic interactions classified as masking and suppression, positive and negative epistasis and pseudo-masking. Values are given as the sum of all events across the 38 neighbourhood pairs illustrated in Fig. 3c. Suppressive interactions were found to be frequent for all four alleles, and masking interactions are dominated by *LEU2*. **e**, Transcripts affected by epistatic interactions switch between subcategories depending on the metabotype. Compared are transcripts differentially expressed according to the median (greater than twofold change; adj. $P < 0.05$). **f**, Degree of distribution of transcripts affected by metabolism-induced epistasis and all other transcripts in a genetic interaction network[23] and a protein–protein interaction network[24]. Genes encoding for epistatic transcripts are significantly less connected compared with the average gene. **g**, Density distribution of PSI-BLAST identified orthologues of epistatically responding and all other genes. Those affected by metabolism-induced epistasis are significantly less conserved. **h**, Distribution of essential genes across epistatically responding and all other transcripts, according to viability in *S. cerevisiae*. Genes affected by metabolism-induced epistasis are less often essential.

was better for metabolic enzymes than it was for other genes (Fig. 4e). On all assessed levels, the proteome was found to be similarly dependent on the metabolic background as the transcriptome. Also, metabolic epistasis was observed at the proteome level to a similar extent as found for the transcriptome. Of 257 epistatic transcripts differentially expressed more than twofold, we were able to

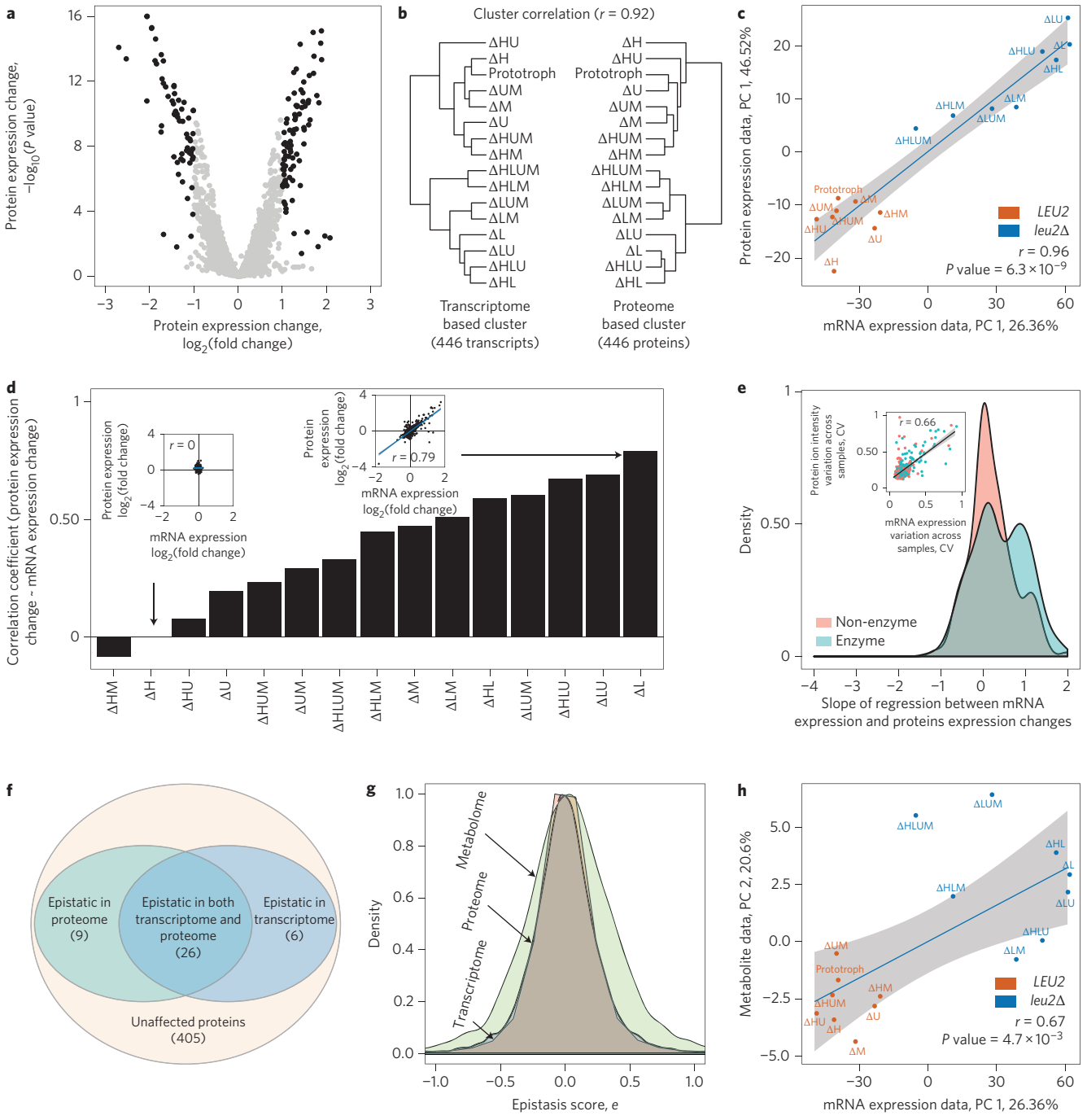

Figure 4 | Metabolism-induced epistasis propagates to the proteome and increases at the metabolome. a, Differential protein expression in 16 metabotypes grown in supplemented SC medium determined by HDMS^E, illustrated in a volcano plot, with $\log_2$ expression values on the *x* axis and $-\log_{10}$ adj. *P* values on the *y* axis. $n = 3$/strain. **b**, Hierarchical clusters based on Euclidean distance and complete linkage agglomeration of gene and protein expression profiles correspond to each other ($r = 0.92$). The transcriptome and proteome cluster the strains in a similar fashion. **c**, Correlation of the major principal components of mRNA and protein expression. Strains are colour-coded according to the presence or absence of the *LEU2* gene, which has the most consistent (least epistatic) profile. **d**, Pearson's correlation coefficient of mRNA and protein expression values for 446 proteins and transcripts as identified in **a**. The transcriptome and proteome correlate quantitatively in strains with differential protein expression (insets). **e**, The transcript and protein dynamic range correspond for metabolic enzymes. Regression slope of gene and protein expression dynamic range distinguishes enzymatic and non-enzymatic genes. Inset: Fold-change across all differentially expressed genes, comparing mRNA and protein level change. **f**, A majority of genes affected by epistasis in the transcriptome are also epistatic in the proteome; $P < 0.05$, fold change > 2, *Z* score $>2\sigma$ or $<-2\sigma$. **g**, Metabolic epistasis affects the transcriptome and proteome to a similar extent, but has a broader impact on metabolite concentrations (represented by the absolute quantities in 50 essential metabolites; Supplementary Fig. 11); the distribution is expressed as 'epistasis score' *e*, describing the relative deviation of expression quantity from a linear relationship in pairwise interactions, and shown as a density plot of all quantified transcripts, proteins and metabolites. The transcriptome and proteome show the same distribution of *e*, but it is increased in the metabolome. **h**, Correlation between the first principal component of transcriptome data (26.36%) and the second principal component of metabolome data (20.6%). Strains are coloured according to the presence and absence of the *LEU2* gene, which had the strongest loading.

capture quantities for 32 proteins, of which 26 were significantly epistatic ($P < 0.05$; Fig. 4f).

To relate these results to the metabolome, we used liquid chromatography selected reaction monitoring (LC-SRM) to absolutely quantify 50 important metabolites (Supplementary Fig. 11a and Supplementary Note 1). Of all metabolites differentially concentrated more than twofold, 75% were found to be affected by epistatic interactions between the metabolic markers (Supplementary Fig. 11b,c). Because of the different biologies of the metabolite concentrations, a direct comparison with transcriptome/proteome data is limited by nature. However, when calculating an epistasis score for all quantified transcripts, proteins or metabolites, metabolite concentrations revealed the broadest distribution (Fig. 4g). Metabolism-induced epistasis therefore appears to manifest more strongly on metabolite concentration than on transcriptome and proteome. Indeed, metabolite concentrations have, with allostery and post-translational regulation, additional layers where epistasis can manifest. Consistently, transcriptome and proteome were only the secondary determinant of metabolite concentration changes. The first principal components (PCs) of neither the proteome nor transcriptome correlated with the first PC of the metabolome (Supplementary Fig. 11d). Instead, the second principal component, explaining 20.6% of the variation in metabolite level, did show a significant agreement with the transcriptome ($r = 0.67$, $P = 4.7 \times 10^{-3}$) and proteome ($r = 0.72$, $P = 1.6 \times 10^{-3}$; Fig. 4h).

Discussion

Metabolic genes are often nicknamed 'housekeeping' genes, which incorrectly implies that the metabolism is static in nature. Indeed, it is increasingly recognized that the chemical–physical environment of a cell is dynamic in terms of metabolite load, redox potential or pH and that metabolic networks are flexible and dynamically regulated. The size and physiological importance of the metabolic network exclude the possibility that gene expression is inert to these adaptations[30–32]. Here, we attempt a system-wide analysis of the impact of the metabolic–genetic background on transcriptome, proteome and metabolome. We exploit 16 combinatorial auxotrophs that would be tolerated as the background in a typical *S. cerevisiae* genetic and genomic experiment. Several studies have found evidence of the physiological importance of metabolic background deficiencies[17,33–36], but in the absence of a comparative, systems-scale analysis, the magnitude and nature of these effects has remained unclear. One objective of the present work was thus to provide systematic high-quality data to allow the impact of metabolic–genetic backgrounds to be accessed by the scientific community. The obtained data sets are valuable for elaborating the relationship between transcriptome, proteome and metabolome in response to metabolic perturbation and reveal a surprisingly high agreement between transcriptional and proteomics results. Transcriptome and proteome did agree in terms of the level of mRNA and protein expression, in the dynamic range, in the correlation of the first principal component, as well as in a concordant level of epistasis (Fig. 4). At least, on inducing metabolic perturbations in exponentially growing cells and when using high precision technology, the correlation of transcriptome and proteome can thus achieve higher values as frequently assumed[37,38].

Despite moderate growth rate differences, which were mostly driven by the *LEU2* gene, all 16 metabotypes and all four auxotrophic markers had a strong molecular impact. In fact, when studying just these four out of many possible metabolic perturbations, we find three-quarters of the coding genome to be affected. Moreover, metabolism-induced gene expression was characterized by a high degree of epistasis. Despite the deletion of *HIS3*, *LEU2*, *MET15* or *URA3* causing strong signatures with hundreds of differentially expressed genes, there were virtually no transcripts that would always respond to their deletion; transcriptional profiles were hence critically sensitive to the metabolic–genetic background. The analysis of 32 strain pairs that differed in one marker at a time allowed the nature of these transcriptional interactions to be categorized. A total of 77% induced transcripts and 88% transcriptional events reflected epistatic interactions between the metabolic genes. The 'epistatic transcriptome' was dominated by the less conserved and less connected genes, which were most commonly affected by masking/suppressive interactions followed by positive or negative quantitative interactions. We note, however, that the nature of epistasis is also dynamic; a transcript falling into one category in one strain pair could have another type of epistasis in another situation. These context-dependent gene expression interactions correlated with flux, implying that the molecular signatures are concordant with metabolic activity (Supplementary Note 2). The metabolic make-up of the cell is hence a systemic determinant of the outcome of a gene expression response.

The association with flux implied that metabolic background effects are strongest for metabolic genes. Indeed, the overlap of transcriptional changes between metabolic gene deletion and the background (83%) was substantially larger than for a random (non-metabolic) gene and the background (34%; Fig. 1e and Supplementary Fig. 4). The metabolic background hence confounds gene expression profiles both for non-metabolic and metabolic genes, but for the latter category, the effects are stronger. Potential consequences of this observation are illustrated by the following gedankenexperiment. If a subset of our omics data sets were analysed in isolation (that is, one analysis would just use the data recorded in the *his3*Δ backgrounds, and the other the *HIS3* data) one would identify different gene expression changes following deletion of *URA3*, *MET15* or *LEU2*. These differences could propagate; that is, two studies could report different Gene Ontology, signalling or gene regulatory networks and eventually claim different functions for *URA3*, *MET15* and *LEU2*. Differences in the metabolic–genetic background could thus negatively impact cross-laboratory reproducibility. This finding is consistent with the importance of the genetic background in defining gene essentiality and phenotype penetrance. A total of 10% of gene deletions have a different phenotype in the closely related *S. cerevisiae* strains S288c and Sigma 1278b[39], with 57 being essential in just one of the two strains[40]. We have shown previously that metabolism is implicated in such effects, with 13 synthetic lethal phenotypes in the S288c background rescued on repairing just three auxotrophies[17].

In summary, we have addressed the importance of the metabolic–genetic background for gene expression interactions found on the transcriptome, proteome and metabolome. Sixteen combinatorial deficiencies in histidine, leucine, uracil and methionine biosynthesis—four, in principle, unrelated model metabolic pathways—retained a strong molecular signature upon complementation and caused a system-wide response involving up to 85% of the coding genome. The metabolic genes interacted epistatically on 77% of differentially expressed genes, and these interactions determined to a large extent the transcriptional outcome of a (metabolic and non-metabolic) gene deletion. Metabolism-induced epistasis manifested across all molecular levels, and transcriptome and proteome were highly correlated and did in part influence an increasingly variable metabolome. It is thus essential to examine the role of background and metabolism-induced epistasis in genetic and genomic experiments and to elucidate its role in gene regulatory networks. Overall, the metabolic make-up of a cell is a key molecular factor in defining the consequence of gene loss across the transcriptome, proteome and metabolome.

Methods

Sampling for transcriptomics and proteomics. Main cultures were inoculated into 50 ml synthetic complete medium as described in ref. 17 at a starting OD_{600} of 0.15 (30 °C, 180 r.p.m.). Strains were then grown for 6–19 h until reaching an OD_{600} of 0.8, at which point the cells were collected by centrifugation (2 min, 3,000*g*)

and the supernatant discarded. The pellet was resuspended in 1 ml H_2O and aliquoted. The cells were then collected (0.5 min, 5,000*g*), the supernatant removed with a pipette, and the sample snap-frozen and stored at −80 °C until further processing.

Transcriptomics

Sample preparation. RNA for transcriptomics was prepared using the yeast RNA mini kit (Zymo Research) followed by DNase treatment, and processed using the TruSeq RNA library preparation kit (Illumina) following the manufacturer's instructions. Twelve samples each were pooled, loaded on a full lane, and sequenced with 2 × 50 bp paired end on a HiSeq 2000 (Illumina), yielding cluster densities of between 730,000 and 806,000 (that is, ∼10,000,000 fragments/1 GB of data per sample). The paired-end reads were aligned using tophat with default parameters against the yeast genome (ENSEMBL Version EF4). To generate gene-wise read counts for gene expression estimation, the htseq tool was applied with the following parameters (htseq-count–mode=intersection-nonempty,–stranded=no), on average achieving a median coverage of between 294 and 942 per covered gene of iGenomes EF4 gene annotation (Supplementary Table 1).

RNAseq data processing. RNAs with very low read count values (<50 counts across all replicates, both raw and normalized read counts) were removed, so that, in total, 5,923 genes were considered expressed in at least one strain. '*DEseq*'[41] in R was used for normalization (leading to a median coverage of 757 to 788 per RNA) and calculation of the *P* values for differential expression of RNAseq data. The gene expression fold change between strains was calculated by dividing the average normalized read count values from all replicates. Hierarchical clustering was based on Euclidean distance and complete linkage agglomeration.

Differential mRNA expression. Differential mRNA expression was calculated with reference (1) to the prototrophic strain, (2) to median values of each transcript across all 16 strains or (3) the fully auxotrophic strain BY4741, as indicated. *P* values were adjusted for multiple test correction using the BH method[42], using the p.adjust function in R. The threshold values for differentially expressed mRNAs were adj. $P < 0.05$, fold change of >2 for upregulated genes and fold change of <0.5 for downregulated genes, and a read count of >50 in both strains. A list of differentially expressed genes from any strain were annotated with GO terms (GO:Process and GO:Functions) by the GO slim mapper tool (http://www.yeastgenome.org/cgi-bin/GO/goSlimMapper.pl; ref. 43).

Proteomics

Sample preparation. For data-independent acquisition proteomics, the 16 strains were grown and analysed in three biological replicates, adding to 48 samples. Proteins were extracted using SDS-containing extraction buffer as described previously[44], combined with protein precipitation with 10% trichloroacetic acid (TCA)[45]. Protein pellets were resuspended in 0.1% Rapigest and 50 mM triethylammonium bicarbonate buffer. Protein (100 µg) was reduced by 5 mM dithiothreitol (DTT) for 30 min at 37 °C, alkylated with 15 mM iodoacetamide at room temperature for 1 h and digested with trypsin at a 1:20 ratio of protein to trypsin overnight, as described previously[46].

LC-MS/MS proteomics. Approximately 500 ng of peptides were separated chromatographically (NanoACQUITY, Waters). The LC aqueous mobile phase (buffer A) contained 0.1% formic acid in water and the organic mobile phase (buffer B) contained 0.1% formic acid in 100% acetonitrile. The samples were trapped on column (Symmetry C18 5 µm, 180 µm × 20 mm, Waters) and desalted for 5 min at a 5 µl min^{-1} flow rate of aqueous mobile phase. The separation was performed on a T3 1.8 µM, 75 µm × 250 mm column (Waters) at 300 nl min^{-1} flow rates using a 90 min linear gradient elution (from 3% to 35% organic mobile phase). The column was then washed with 80% organic mobile phase for 5 min and re-equilibrated with 3% organic mobile phase for 15 min. The analytical column temperature was maintained at 40 °C. Eluting peptides were analysed on SYNAPT G2 hybrid IMS-MS system (Waters). Data were acquired in IMS-MS^E mode with low and high energy scans of 900 ms. The collision energy was linearly ramped from 21 to 44 V in the Transfer region of TriWave during high energy scans. The emitters employed were manufactured by etching a fused-silica line with hydrofluoric acid as described in ref. 47. [Glu-1]-fibrinopeptide B (500 fmol $µl^{-1}$) was infused with a lockspray ion probe at a flow rate of 500 nl min^{-1} using an auxiliary pump and was acquired once every 30 s for a 1 s period.

Proteome data processing. The raw data were initially processed with the ProteinLynx Global Server (PLGS) 2.5.2 to generate a list of precursor and fragment exact mass retention times (EMRTs) and associations between them based on a similarity of retention time and drift time. The thresholds for low-energy ions, high-energy ions and low-energy EMRT pairs were set to 100, 15 and 750 counts, respectively. The data were lock-mass corrected post-acquisition using [Glu-1]-fibrinopeptide as a lock mass compound with 785.84 *m/z* for $z = 2$ and 0.25 Da tolerance window. The EMRTs were then searched against the UniProt *S. cerevisiae* database using an ion accounting algorithm as described previously[48]. A peptide required at least one fragment and a protein required at least three fragments and one peptide for identification. The database search was performed at 100% false discovery rate (FDR) to identify real and decoy peptides for subsequent filtering by *q*-value. Peptide identifications from strain 1 replicates 1 and 3 and strain 16 replicate 1 were combined to produce a master *in silico* run, which was used to transfer identifications to all other acquisitions using the synapter algorithm as described previously[49]. To account for potential sources of technical variation in proteomic experiments, we removed peptides, which were not detected in all 48 samples. For each protein a Spearman's rank correlation coefficient was then calculated between each pair of peptides across all samples. Peptides displaying overall low co-correlation (with a shorth[50,51] correlation of <0.3) were removed from subsequent analysis. Such a procedure assumes that the signals of peptides coming from the same protein have to be correlated. This selection thus identifies non-specific peptides, or peptides that are not linear for other reasons (that is, post-translational modifications present to a varying extend). Finally, for each protein, the signals of all detected peptides were geometrically averaged and subsequently accounted for differential expression using modified eBayes *t*-test using the limma package[52] implemented in R.

Metabolomics. Pre-cultures of the 16 strains were prepared in synthetic complete (SC) medium and cultured for 12–20 h in a 96-well plate (30 °C, 250 r.p.m.). Main cultures were inoculated into fresh SC medium, and cultivation, sample collection, quenching and metabolite extraction were performed according to ref. 53. Specifically, the yeast strains were cultured in 96-well fritted plates (1 strain per two wells; Nunc) with a 4 mm glass bead in each well for mixing. For quenching, a 48-well plate containing 3.6 ml of −40 °C quenching solution (60% methanol, 10 mM NH_4-acetate) per well was prepared and put into a vacuum manifold. Mid-exponential cultures were put on top of the 48-well plate and sucked into the quenching solution. The quenched cultures were centrifuged for 5 min at 4,000 r.p.m. and −9 °C and the supernatant was discarded. Subsequently, the plate with the cell pellets was transferred into a −50 °C ethanol bath with dry ice. For extraction, the cell pellets were resuspended in 1 ml precooled extraction solution (75% ethanol, 10 mM NH_4-acetate) and 50 µl ^{13}C-yeast internal standard was added. To complete extraction, the plate was transferred into an 80 °C water bath for 3 min with vortexing steps every 30–45 s. The extracts were stored at −80 °C until they were dried overnight with a vacuum centrifuge (Christ-RVC 2–33 CD plus, Kuehner AG). LC-MS measurements were performed as described previously[54]. Specifically, the dried extracts were dissolved in 50–100 µl ddH_2O and 15 µl were injected for LC-MS analysis. The metabolites were separated with a Waters Acquity T3 end-capped reversed-phase column (150 mm × 2.1 mm × 1.8 µm) on an Acquity UPLC (Waters) system. For mass spectrometric detection of the metabolites we used a TSQ Quantum Ultra triple quadrupole mass spectrometer (Thermo Fisher Scientific) with a heated electrospray ionization source operating in negative mode with selective reaction monitoring. Peak areas were determined by summing raw intensities from peak start to peak end using MatLab. The metabolite peak areas were further normalized to the ^{13}C internal standard and to biomass as determined by optical density (OD_{600}).

Metabolic pathway enrichment. Metabolic pathway enrichment was determined by hypergeometric testing using the *phyper* function in R. The FDR was calculated according to ref. 55. Thresholds for significant enrichment were an FDR of <0.05, $P < 0.05$ and a number of enriched genes of one-third of the pathway size.

Epistasis. To avoid an error propagation problem in the epistasis analysis, we normalized the gene expression values given in Fig. 3 and in several of the Supplementary figures as indicated to the median expression value of all strains. This strategy yielded largely similar results to normalization to the prototrophic strain, as its expression profile is close to that of the median (Supplementary Fig. 3i). Normalization to the median is, however, better suited for the epistasis analysis in our experimental setup as it (1) avoids unequal error propagation of the wild-type strain measurement error and (2) allows a comparison of all 16 strains with one another.

To calculate epistasis from the median-normalized data, Fisher's additive model was used as a basis to obtain an epistasis score (*e*) for mRNA expression, protein expression, metabolite concentration and growth data[21]: epistasis score (*e*) = observed (W_{AB}) − expected ($W_A + W_B$), where W_{AB}, W_A and W_B are per cent changes in either gene expression, protein expression, metabolite concentration or growth rate of knockout strain expressed in per cent[56–58]. The calculation was performed by comparing strains representing closest neighbours in a genetic dependence network (Fig. 3c), comparing all 38 possible strain pairs in total. *P* values (BH adjusted) for the differential expression of genes in all 38 strains pairs were calculated with reference to the respective product strains. Expression fold change was calculated by dividing the average expression values in each strain with the respective wild-type, median value or any other reference strain as indicated. The epistasis score was transformed to a standard epistasis score (*Z* score). mRNAs, proteins or metabolites with significant *P* values (<0.05) in at least one strain, plus a *Z* score of $>2\sigma$ or $<-2\sigma$ were considered epistatic. As illustrated in Fig. 3b, the standard epistasis score and *P* value (adjusted) for differential expression were used to subgroup epistatic genes into masking, suppression, pseudo-masking or positive and negative epistasis. Categorization was conducted for differential expression in all

38 strain pairs representing closest neighbourhoods (Fig. 3c), and fold changes with respect to the median value were used for normalization as follows: (a) *masking*, where the expression value is explained by the allele with quantitative stronger effect; (b) *suppression*, where the expression value is explained by the allele with quantitative lower effect; (c) *pseudo-masking*, where the expression value depends on both alleles, but the effect is lower than expected from the strong allele and larger than expected from the weak allele; (d) *positive epistasis*, where the expression value is larger than expected from additivity of the single alleles; and (e) *negative epistasis*, where the expression value is lower than expected from additivity of the two single alleles and lower than each individual value.

Gene essentiality and conservation. A list of essential *S. cerevisiae* genes was taken from the Database of Essential Genes (DEG) (http://www.essentialgene.org)[59]. The result of PSI-BLAST[60] analysis of *S. cerevisiae* from the SGD database[43] was used to count the number of species in which a gene is conserved.

Confounding effects of metabolism-induced transcriptional changes in transcriptomics. The gene expression data for a total of 118 yeast knockout microarray experiments, conducted in 14 studies in different context and laboratories[61–72] (see 'Accession numbers' section) was downloaded from ArrayExpress. All of these studies are based on the BY4741 background, auxotroph in histidine, leucine, uracil and methionine, as used in our study. We categorize these data sets into single gene knockout experiments (49 arrays, as used in Fig. 1e) or all other conditions (69 arrays from four studies, Supplementary Fig. 4a). The raw data from Affymetrix GeneChip Yeast Genome 2.0 and S98 Arrays were re-analysed by applying the same stringent quality filtering criteria as used on our RNASeq data (FDR-adjusted $P < 0.05$ and fold change >2). The lists of differentially expressed genes in 49 microarray experiments were compared with differentially expressed genes affected by the auxotrophic markers in the respective yeast strain background.

Flux variability analysis. The NAD-corrected yeast genome-scale metabolic model iMM904[73,74] was used in a flux variability analysis (FVA)[19,20]. The model was constrained by setting the biomass flux according to the experimentally determined growth rates[17] and metabolite uptake rates for amino acids and glucose as measured in the present study (Supplementary File 2). On including both growth rate and nutrient uptake rates, the FVA predicts a feasible flux range for every reaction of the model and for every strain. Differences in metabolic flux (maximum flux – minimum flux) are expressed as fold change and are calculated with reference to the prototrophic wild-type strain. Flux range is considered to be significantly different when it deviates by two standard deviations from the mean.

Accession numbers. Transcriptome data has been deposited at ArrayExpress[18] (https://www.ebi.ac.uk/arrayexpress/) under accession no. E-MTAB-3991. Proteomics data has been deposited at ProteomeXchange (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository[75] with the data set identifier PXD001491. Metabolome data has been deposited at Metabolights[76] (http://www.ebi.ac.uk/metabolights/) with the accession no. MTBLS168. Processed data are provided in Supplementary File 2. Previously reported gene expression data that was reanalysed in this study can be found in ArrayExpress under accession numbers E-GEOD-31774 (ref. 61), E-GEOD-31326 (ref. 62), E-TABM-638 (ref. 63), E-GEOD-18994 (ref. 64), E-GEOD-29530 (ref. 65), E-MTAB-1059 (ref. 66), E-GEOD-21571 (ref. 67), E-GEOD-56702 (ref. 68), E-GEOD-31176 (ref. 69), E-MTAB-2539 (ref. 70), E-MEXP-3150 (ref. 71), E-GEOD-18644 (ref. 72), E-GEOD-25582, E-GEOD-28794.

Received 12 September 2015; accepted 17 December 2015; published 1 February 2016

References

1. Albert, R. Scale-free networks in cell biology. *J. Cell Sci.* **118,** 4947–4957 (2005).
2. Barabási, A.-L. & Oltvai, Z. N. Network biology: understanding the cell's functional organization. *Nature Rev. Genet.* **5,** 101–113 (2004).
3. Herrgård, M. J. *et al.* A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nature Biotechnol.* **26,** 1155–1160 (2008).
4. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabási, A.-L. The large-scale organization of metabolic networks. *Nature* **407,** 651–654 (2000).
5. Newman, M. E. J. Modularity and community structure in networks. *Proc. Natl Acad. Sci.* **103,** 8577–8582 (2006).
6. Romero, P. *et al.* Computational prediction of human metabolic pathways from the complete human genome. *Genome Biol.* **6,** R2 (2005).
7. Thiele, I. *et al.* A community-driven global reconstruction of human metabolism. *Nature Biotechnol.* **31,** 419–425 (2013).
8. Clark, A. G. & Fucito, C. D. Stress tolerance and metabolic response to stress in *Drosophila melanogaster*. *Heredity* **81,** 514–527 (1998).
9. Ihmels, J., Levy, R. & Barkai, N. Principles of transcriptional control in the metabolic network of *Saccharomyces cerevisiae*. *Nature Biotechnol.* **22,** 86–92 (2004).
10. Liu, L., Li, Y. & Tollefsbol, T. O. Gene–environment interactions and epigenetic basis of human diseases. *Curr. Issues Mol. Biol.* **10,** 25–36 (2008).
11. Tu, B. P., Kudlicki, A., Rowicka, M. & McKnight, S. L. Logic of the yeast metabolic cycle: temporal compartmentalization of cellular processes. *Science* **310,** 1152–1158 (2005).
12. Campbell, K. *et al.* Self-establishing communities enable cooperative metabolite exchange in a eukaryote. *eLife* http://dx.doi.org/10.7554/eLife.09943 (2015).
13. Fink, G. R. Gene–enzyme relations in *Histidine biosynthesis* in yeast. *Science* **146,** 525–527 (1964).
14. Satyanarayana, T., Umbarger, H. E. & Lindegren, G. Biosynthesis of branched-chain amino acids in yeast: regulation of leucine biosynthesis in prototrophic and leucine auxotrophic strains. *J. Bacteriol.* **96,** 2018–2024 (1968).
15. Lacroute, F. Regulation of pyrimidine biosynthesis in *Saccharomyces cerevisiae*. *J. Bacteriol.* **95,** 824–832 (1968).
16. Masselot, M. & De Robichon-Szulmajster, H. Methionine biosynthesis in *Saccharomyces cerevisiae*. I. Genetical analysis of auxotrophic mutants. *Mol. Gen. Genet.* **139,** 121–132 (1975).
17. Mülleder, M. *et al.* A prototrophic deletion mutant collection for yeast metabolomics and systems biology. *Nature Biotechnol.* **30,** 1176–1178 (2012).
18. Brazma, A. *et al.* ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Res.* **31,** 68–71 (2003).
19. Mahadevan, R. & Schilling, C. H. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* **5,** 264–276 (2003).
20. Schellenberger, J. *et al.* Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nature Protoc.* **6,** 1290–1307 (2011).
21. Fisher, R. A. The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edin.* **52,** 399–433 (1918).
22. Park, S. & Lehner, B. Epigenetic epistatic interactions constrain the evolution of gene expression. *Mol. Syst. Biol.* **9,** 645 (2013).
23. Costanzo, M. *et al.* The genetic landscape of a cell. *Science* **327,** 425–431 (2010).
24. Kim, H. *et al.* YeastNet v3: a public database of data-specific and integrated functional gene networks for *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **42,** D731–D736 (2014).
25. Breen, M. S., Kemena, C., Vlasov, P. K., Notredame, C. & Kondrashov, F. A. Epistasis as the primary factor in molecular evolution. *Nature* **490,** 535–538 (2012).
26. Kemmeren, P. *et al.* Large-scale genetic perturbations reveal regulatory networks and an abundance of gene-specific repressors. *Cell* **157,** 740–752 (2014).
27. Alam, M. T., Medema, M. H., Takano, E. & Breitling, R. Comparative genome-scale metabolic modeling of actinomycetes: the topology of essential core metabolism. *FEBS Lett.* **585,** 2389–2394 (2011).
28. Shliaha, P. V., Bond, N. J., Gatto, L. & Lilley, K. S. Effects of traveling wave ion mobility separation on data independent acquisition in proteomics studies. *J. Proteome Res.* **12,** 2323–2339 (2013).
29. Silva, J. C. *et al.* Quantitative proteomic analysis by accurate mass retention time pairs. *Anal. Chem.* **77,** 2187–2200 (2005).
30. Grüning, N.-M., Lehrach, H. & Ralser, M. Regulatory crosstalk of the metabolic network. *Trends Biochem. Sci.* **35,** 220–227 (2010).
31. Patel, A. *et al.* A liquid-to-solid phase transition of the ALS protein FUS accelerated by disease mutation. *Cell* **162,** 1066–1077 (2015).
32. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33,** 245–254 (2003).
33. Hashimoto, S. *et al.* Isolation of auxotrophic mutants of diploid industrial yeast strains after UV mutagenesis. *Appl. Environ. Microbiol.* **71,** 312–319 (2005).
34. Kokina, A., Kibilds, J. & Liepins, J. Adenine auxotrophy—be aware: some effects of adenine auxotrophy in *Saccharomyces cerevisiae* strain W303-1A. *FEMS Yeast Res.* **14,** 697–707 (2014).
35. Low, B. Rapid mapping of conditional and auxotrophic mutations in *Escherichia coli* K-12. *J. Bacteriol.* **113,** 798–812 (1973).
36. Pronk, J. T. Auxotrophic yeast strains in fundamental and applied research. *Appl. Environ. Microbiol.* **68,** 2095–2100 (2002).
37. Hack, C. J. Integrated transcriptome and proteome data: the challenges ahead. *Brief. Funct. Genom. Proteom.* **3,** 212–219 (2004).
38. Payne, S. H. The utility of protein and mRNA correlation. *Trends Biochem. Sci.* **40,** 1–3 (2015).
39. Ryan, O. *et al.* Global gene deletion analysis exploring yeast filamentous growth. *Science* **337,** 1353–1356 (2012).
40. Dowell, R. D. *et al.* Genotype to phenotype: a complex problem. *Science* **328,** 469 (2010).
41. Anders, S. & Huber, W. Differential expression analysis for sequence count data. *Genome Biol.* **11,** R106 (2010).
42. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B Methodol.* **57,** 289–300 (1995).
43. Cherry, J. M. *et al.* Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* **40,** D700–D705 (2012).

44. von der Haar, T. Optimized protein extraction for quantitative proteomics of yeasts. *PLoS ONE* **2,** e1078 (2007).
45. Fic, E., Kedracka-Krok, S., Jankowska, U., Pirog, A. & Dziedzicka-Wasylewska, M. Comparison of protein precipitation methods for various rat brain structures prior to proteomic analysis. *Electrophoresis* **31,** 3573–3579 (2010).
46. Vowinckel, J. *et al.* The beauty of being (label)-free: sample preparation methods for SWATH-MS and next-generation targeted proteomics. *F1000Research* **2, 272** (2014).
47. Kelly, R. T. *et al.* Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Anal. Chem.* **78,** 7796–7801 (2006).
48. Li, G.-Z. *et al.* Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* **9,** 1696–1719 (2009).
49. Bond, N. J., Shliaha, P. V., Lilley, K. S. & Gatto, L. Improving qualitative and quantitative performance for MSE-based label-free proteomics. *J. Proteome Res.* **12,** 2340–2353 (2013).
50. Gentleman, R., Carey, V. J., Huber, W., Irizarry, R. A. & Dudoit, S. (eds) *Bioinformatics and Computational Biology Solutions Using R and Bioconductor* (Springer, 2005).
51. Andrews, D. *Robust Estimates of Location* (Princeton Univ. Press, 1972).
52. Ritchie, M. E. *et al.* Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43,** e47 (2015).
53. Ewald, J. C., Heux, S. & Zamboni, N. High-throughput quantitative metabolomics: workflow for cultivation, quenching, and analysis of yeast in a multiwell format. *Anal. Chem.* **81,** 3623–3629 (2009).
54. Buescher, J. M. *et al.* Global network reorganization during dynamic adaptations of *Bacillus subtilis* metabolism. *Science* **335,** 1099–1103 (2012).
55. Boyle, E. I. *et al.* GO::TermFinder—open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinformatics* **20,** 3710–3715 (2004).
56. Dixon, S. J., Costanzo, M., Baryshnikova, A., Andrews, B. & Boone, C. Systematic mapping of genetic interaction networks. *Annu. Rev. Genet.* **43,** 601–625 (2009).
57. Mani, R., St. Onge, R. P., Hartman, J. L., Giaever, G. & Roth, F. P. Defining genetic interaction. *Proc. Natl Acad. Sci. USA* **105,** 3461–3466 (2008).
58. Segrè, D., Deluna, A., Church, G. M. & Kishony, R. Modular epistasis in yeast metabolism. *Nature Genet.* **37,** 77–83 (2005).
59. Giaever, G. *et al.* Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* **418,** 387–391 (2002).
60. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25,** 3389–3402 (1997).
61. Ansari, S. A. *et al.* Distinct role of Mediator tail module in regulation of SAGAdependent, TATA-containing genes in yeast. *EMBO J.* **31,** 44–57 (2012).
62. Dymond, J. S. *et al.* Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature* **477,** 471–476 (2011).
63. Fournier, M. L. *et al.* Delayed correlation of mRNA and protein expression in rapamycin-treated cells and a role for Ggc1 in cellular sensitivity to rapamycin. *Mol. Cell. Proteom.* **9,** 271–284 (2010).
64. Jimeno, S. *et al.* New suppressors of THO mutations identify Thp3 (Ypr045c)-Csn12 as a protein complex involved in transcription elongation. *Mol. Cell. Biol.* **31,** 674–685 (2011).
65. Lu, L., Roberts, G. G., Oszust, C. & Hudson, A. P. The YJR127C/ZMS1 gene product is involved in glycerol-based respiratory growth of the yeast *Saccharomyces cerevisiae*. *Curr. Genet.* **48,** 235–246 (2005).
66. Miller, C. *et al.* Mediator phosphorylation prevents stress response transcription during non-stress conditions. *J. Biol. Chem.* **287,** 44017–44026 (2012).
67. Morillo-Huesca, M., Clemente-Ruiz, M., Andújar, E. & Prado, F. The SWR1 histone replacement complex causes genetic instability and genome-wide transcription misregulation in the absence of H2A.Z. *PloS ONE* **5,** e12143 (2010).
68. Santos-Pereira, J. M., García-Rubio, M. L., González-Aguilera, C., Luna, R. & Aguilera, A. A genome-wide function of THSC/TREX-2 at active genes prevents transcription–replication collisions. *Nucleic Acids Res.* **42,** 12000–12014 (2014).
69. Sanz, A. B. *et al.* Chromatin remodeling by the SWI/SNF complex is essential for transcription mediated by the yeast cell wall integrity MAPK pathway. *Mol. Biol. Cell* **23,** 2805–2817 (2012).
70. Schulz, D., Pirkl, N., Lehmann, E. & Cramer, P. Rpb4 functions mainly in mRNA synthesis by RNA polymerase II. *J. Biol. Chem.* **289,** 17446–17752 (2014).
71. Seizl, M., Larivière, L., Pfaffeneder, T., Wenzeck, L. & Cramer, P. Mediator head subcomplex Med11/22 contains a common helix bundle building block with a specific function in transcription initiation complex stabilization. *Nucleic Acids Res.* **39,** 6291–6304 (2011).
72. Tauber, E. *et al.* Functional gene expression profiling in yeast implicates translational dysfunction in mutant huntingtin toxicity. *J. Biol. Chem.* **286,** 410–419 (2011).
73. Mo, M. L., Palsson, B. O. & Herrgård, M. J. Connecting extracellular metabolomics measurements to intracellular flux states in yeast. *BMC Syst. Biol.* **3,** 37 (2009).
74. Szappanos, B. *et al.* An integrated approach to characterize genetic interaction networks in yeast metabolism. *Nature Genet.* **43,** 656–662 (2011).
75. Vizcaíno, J. A. *et al.* The PRoteomics IDEntifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* **41,** D1063–D1069 (2013).
76. Haug, K. *et al.* MetaboLights—an open-access general-purpose repository for metabolomics studies and associated meta-data. *Nucleic Acids Res.* **41,** D781–D786 (2013).

Acknowledgements

The authors thank U. Sauer (ETH Zurich) for support with metabolite measurements and scientific discussions and M. Werber and S. Klages (Max Planck Institute for Molecular Genetics) for support with RNA sequencing analysis. The authors acknowledge the Wellcome Trust (RG 093735/Z/10/Z), the ERC (starting grant 260809), the Isaac Newton Trust (RG 68998) and the Darwin Trust of Edinburgh for a studentship for P.V.S. A.Z. is an EMBO fellow. M.R. is a Wellcome Trust Research Career Development and Wellcome-Beit Prize fellow.

Author contributions

M.T.A., A.Z., R.S., E.R. and S.B. performed data analysis. M.M., P.S. and S.C. carried out raw data processing. M.M., P.S., F.C., J.V., A.K., E.C., S.M. and S.C. conducted the experiments. K.R.P., B.T., K.S.L. and M.R. conceived the study. M.R. wrote the first draft. M.T.A., A.Z. and M.R. wrote the paper. All authors contributed to preparing the final version.

Additional information

Supplementary information is available online. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to M.R.

Competing interests

The authors declare no competing financial interests.