Enzyme levels broadly influence flux throughout the metabolic network

Xuhang Li, L. Safak Yilmaz*, Albertha J.M. Walhout*

Department of Systems Biology, University of Massachusetts Chan Medical School; Worcester, MA, USA.

*Correspondence: marian.walhout@umassmed.edu or lutfu.yilmaz@umassmed.edu

Metabolic reaction flux is regulated in response to nutritional, environmental or pathological conditions by changes in either metabolite or metabolic enzyme levels[1]. Recent studies have proposed that flux is predominately regulated by metabolite, rather than enzyme, levels based on kinetics of selected reactions[2-5]. However, the extent to which enzyme levels affect flux throughout the metabolic network *in vivo* remains unclear[6]. Here, we combine available yeast enzyme level, flux data, and metabolic network modeling to discover three paradigms by which enzyme levels are broadly associated with, and likely influence, flux: cognate reaction, pathway-level coordination, and flux coupling. We find that the architecture of the metabolic network enables the reach of influence for most enzymes. This enzyme reach is implemented as a novel parameter in an enhanced flux potential analysis algorithm to predict flux from enzyme levels in yeast and humans. Our study suggests that metabolic network architecture facilitates a broad physiological impact of changes in enzyme levels and may form a foundation for understanding metabolic states using enzyme expression data for a variety of systems, and eventually, individual cells.

Main

Metabolism is highly dynamic. Metabolic flux, or the rate of metabolic conversion, changes in response to nutritional, environmental, and pathogenic perturbations. Flux can be regulated by multiple mechanisms: metabolite concentrations can directly affect reaction rate; allosteric regulators and covalent modifications can affect the activity of metabolic enzymes; and metabolic enzyme levels can be regulated by changes in gene or protein expression[1]. Previously, it has been proposed that metabolic flux in yeast[2-4] and bacteria[5] is predominantly regulated by metabolite rather than enzyme levels. However, it has also been shown that systems-level enzyme expression can be effectively used to predict cell and tissue metabolism[7-10]. In principle, differential metabolic gene expression can serve two purposes: the first is innate and is to establish enzyme levels in different cell types or during different developmental stages. The second is in response to exogenous conditions, such as changes in diet or exposure to specific compounds. The first is what we refer to as metabolic wiring, while the second can be viewed as rewiring. Overall, the relationships between enzyme levels, wiring and rewiring of metabolic flux at a systems, or network level, remain to be determined[6].

Metabolic enzyme levels influence flux of non-cognate reactions

Ultimately, all reactions in the metabolic network are connected and, therefore, reaction fluxes are likely not independent[11]. In fact, at steady state, the producing and consuming fluxes of the same metabolite need to be equal, which results in flux coupling of associated reactions. Therefore, we hypothesized that metabolic rewiring by gene expression changes may affect not only cognate reactions and that enzyme levels may

have a larger reach through the network. To test this hypothesis, we first asked whether metabolic enzyme levels correlate with flux of cognate reactions using the single available dataset that measured both reaction flux and metabolic enzyme levels for a large number of metabolic reactions in yeast[4] (Supplementary Table 1). In this dataset, flux and enzyme expression measurements are available for 156 metabolic reactions across 25 conditions in which different nutrients (glucose, leucine, uracil, phosphate, nitrogen) were limited and the growth rate was titrated (Supplementary Tables 2-4). We first identified metabolic reactions for which flux correlates with the level of the corresponding enzyme(s) (Fig. 1a). A total of 46 of the 156 reactions (29%) showed a significant positive correlation between flux and enzyme levels (False Discovery Rate (FDR) < 0.05). Hereafter these reactions are referred to as *correlated reactions* (Fig. 1b,c, Extended Data Fig. 1 and Supplementary Table 2). Similar to observations in bacteria[5], most central carbon reactions in yeast are not correlated, while correlated reactions mainly occur in amino acid and nucleic acid metabolism (Fig. 1d, Extended Data Fig. 2a-d). Interestingly, some metabolic pathways have multiple correlated reactions, while others have only one or a few. For instance, all seven reactions in arginine biosynthesis are correlated, while only the last reaction is correlated in threonine, methionine, and cysteine metabolism (Extended Data Fig. 2b,d).

The yeast experiments were performed at steady state, and, therefore, the flux in reactions of the same linear pathway is coordinated[4]. Further, enzymes that catalyze reactions in the same pathway are often coordinately activated or repressed, thereby enabling network rewiring[12-15]. Therefore, we next asked whether pathways with a greater degree of enzyme coexpression exhibit an overall better correlation between enzyme

levels and flux. We manually defined 20 metabolic pathways that consist of at least three connected reactions for which flux measurements were available, and for which the enzyme levels corresponding to at least two reactions were available (Extended Data Fig. 2e and Supplementary Table 2). We found that, except for valine and leucine biosynthesis, there was a significant correlation between the strength of pathway-level enzyme coexpression and the strength of flux-expression correlation ($r=0.61$, $p<0.006$) (Fig. 1e). The exception of leucine biosynthesis pathway is likely due to the use of the leucine auxotrophic strain (MAT**a** *leu2Δ1*)[4]. The correlation indicates that the more different enzymes in a pathway covary in expression under different conditions, the more the flux through the entire pathway correlates with changes in enzyme levels. Based on this observation, we next asked if pathway-level coexpression better predicts flux compared to changes in individual enzymes. We defined pathway-level coexpression as a vector of 25 elements that corresponds to the 25 conditions, with each element containing the median of relative enzyme levels of all enzymes in the pathway in that condition. Indeed, enzyme coexpression correlates better with pathway flux than the enzyme levels for individual reactions, particularly for pathways that are moderately coexpressed (Fig. 1f). Therefore, we conclude that pathway-level regulation of enzyme expression better predicts flux changes compared with changes in enzyme levels of individual reactions. This observation supports our hypothesis that changes in enzyme levels not only affect cognate reactions, but have a reach by which they affect other, non-cognate, reactions in the same pathway as well.

Enzyme reach affects local pathway flux

Which non-cognate reactions are affected by changes in the level of individual enzymes, or, in other words, what is the extent of “enzyme reach”? We hypothesized that enzyme reach would mostly extend to reactions with a linear connection to the cognate reaction and reasoned that such connections could be captured by defining metabolites with a low degree of network connectivity (Fig. 2a). We inspected the “pairwise cross-informing rate” that defines how likely the flux or expression of one reaction is correlated (FDR < 0.05) with the flux or expression of a connected reaction, given the degree of the bridging metabolite that links the two reactions. Indeed, reactions connected by metabolites with a low degree are more likely to cross-inform each other for both flux and expression as well as expression-flux correlation (Fig. 2b).

To capture enzyme reach in the yeast metabolic network, we used metabolic network modeling based on flux balance analysis (FBA). Specifically, we adapted our previously developed Flux Potential Analysis (FPA) algorithm, which calculates the relative flux potential (rFP) of a target reaction based on enzyme mRNA levels[7] (Supplementary Notes 1). FPA is formulated as an FBA problem whose objective function is the maximal flux of the target reaction and whose maximized objective value is named *flux potential*. The key concept of FPA is to convert relative enzyme levels into weight coefficients such that the flux through a reaction is penalized when cognate enzyme levels are low (Extended Data Fig. 3a, Methods)[7]. FPA uses a distance decay function with which enzyme reach decreases with distance to the target reaction. However, the original FPA did not consider the architecture of the network around a target reaction. Therefore, to better capture enzyme reach, we incorporated bridging metabolite degree to utilize flux routes that are both flux-balanced and low in bridging metabolite degree (Supplementary

Notes 1) (Extended Data Fig. 3b). In addition, we formulated a distance decay function with a *distance boundary* parameter to precisely control the maximum length of integrated pathway(s) (Fig. 2c and Extended Data Fig. 3c, Supplementary Notes 1). We refer to this algorithm as enhanced FPA (eFPA).

To study enzyme reach in the yeast metabolic network, we performed eFPA using each one of the 232 reactions for which flux was measured[4] as the target reaction. Interestingly, for most reactions, the correlation between rFP and measured flux varied with distance boundary (Fig. 2d, Extended Data Fig. 4). We refer to the boundary where the correlation is maximal as the *optimal distance boundary* and to the corresponding eFPA as "optimal-boundary eFPA". Overall, the flux of 44% (101/232) reactions could be significantly correlated with rFP (Extended Data Fig. 3d). Importantly, optimal-boundary eFPA improved the correlation for most reactions, further confirming that enzyme levels affect metabolic flux of non-cognate reactions (Fig. 2e). The optimal distance boundary of most reactions was between 2 and 6 (Extended Data Fig. 4), and the total number of significantly correlated reactions peaked at a distance boundary of 4 (Fig. 2f, yellow curve). As expected, when gene expression was solely considered, *i.e.,* without flux balance or metabolite degree, the correlation with flux diminished (Fig. 2f, blue and purple curve). These observations indicate that enzyme reach occurs mostly locally, within a metabolic pathway. We therefore reasoned that we could generate flux predictions that are less sensitive to the choice of distance boundary, which has to be made a priori, by optimizing the decay function in eFPA. To do this, we used an exponential decay function (Extended Data Fig. 3c) and found that this helped to predict relative flux with a broader range of distance boundaries (Fig. 2f, red curve). We refer to the use of exponential decay

function and an a priori single distance boundary of 6 as the “local-pathway eFPA”, which is independent of the fitted optimal distance boundaries. Using local-pathway eFPA, we found that only using levels of enzymes catalyzing non-cognate reactions predicted reaction fluxes equally well as when cognate enzyme levels were included (Fig. 2g). Taken together, our results show that metabolic enzymes have a local reach, i.e., their levels affect not only their cognate reactions but also correlate with reactions to which they are coupled by low-degree metabolites and flux balance.

Remarkably, eFPA predicted the flux of some reactions only with very large distance boundaries (Fig. 2d and Extended Data Fig. 4). For instance, for glutathione synthesis, the levels of enzymes involved in energy generation (glycolysis, oxidative phosphorylation, and the TCA cycle) contributed significantly to accurate flux predictions (Extended Data Fig. 5a). We noticed that high (or low) glutathione synthesis usually coincides with high (or low) level of those energy-generating enzymes, which critically contributed to the accurate prediction of flux (Extended Data Fig. 5b). This observation suggests that glutathione biosynthesis is limited by energy production, which is in agreement with early observations[16]. Thus, eFPA independently revealed the tight coupling between glycolysis and the synthesis of glutathione solely using enzyme levels.

Overall, eFPA correctly predicted the flux for a large proportion of metabolic reactions, including reactions for which enzyme level measurements were not available (*gap-filling predictions*) and reactions for which cognate enzyme levels did not correlate with flux (*overriding predictions*) (Fig. 2h). Finally, a randomization of eFPA by shuffling the reaction labels in the input expression data showed the significance of eFPA

predictions, which suggests that enzyme reach is an intrinsic property of the flux-enzyme relationship and not a result of overfitting (Fig. 2i,j).

Beyond pathway-level coordination

While eFPA predicted flux for about half of the reactions, the other half could not be predicted by enzyme levels (Fig. 2h). We hypothesized that some of the fluxes that were not predicted by eFPA could still be indicated from enzyme levels based on flux coupling, where a set of reactions maintain correlated fluxes because the reactions are directly coupled at steady state. An example of such flux coupling is the control of pathway fluxes by a single rate-limiting enzyme, which by itself is not sufficient to predict the flux by eFPA. To empirically derive coupled reactions, we defined the term *flux collinearity* as the absolute PCC between two fluxes in the yeast dataset, where fully coupled reactions (i.e., reactions in a linear pathway) have perfect collinearity with absolute PCC = 1. The flux of many testable reactions (161/232, 69%) was highly colinear with at least one of the 46 correlated reactions (absolute PCC > 0.8, Fig. 3a). Thus, the flux of these non-correlated reactions can in principle be predicted by the levels of at least one enzyme for which the levels correlate with cognate reaction flux, assuming the empirical flux coupling is generalizable in exponentially growing yeast. We refer to these as “indicator enzymes” (Supplementary Notes 2, Extended Data Fig. 6). As expected, nearly all fluxes predicted by eFPA could also be predicted by an indicator enzyme (96/102, 94%) (Fig. 3a, Venn diagram). Remarkably, 40% of fluxes (65/161) could only be predicted by indicator enzymes, implying that their flux is not controlled by pathway-level coregulation of enzyme expression (Fig. 3a).

Fluxes in pyrimidine biosynthesis reactions were predicted by two indicator enzymes; flux in the upstream branch that synthesizes orotate correlates with levels of aspartate carbamoyltransferase, while flux in the downstream branch that produces UMP was predicted by orotidine-5'-phosphate decarboxylase levels (Fig. 3b, red arrows). The fluxes of the two branches are not highly colinear because orotate is partially secreted[4]. Interestingly, the flux of one of the target reactions in which uracil (ura) is converted to UMP (ump) showed a negative correlation with the levels of the indicator enzyme. This may reflect the alternative use of the salvage and *de novo* pyrimidine biosynthesis pathways (Fig. 3b, reaction 3). Importantly, the reactions catalyzed by both indicator enzymes are key rate-limiting steps in UMP synthesis and are known to be regulated both transcriptionally and allosterically in multiple systems[17,18]. These analyses suggest that yeast pyrimidine biosynthesis is controlled, at least in part, by the levels of two specific rate-limiting enzymes.

Some other indicator enzymes are less likely to catalyze a rate-limiting step. For instance, fluxes in the main energy generating pathways, the TCA cycle and oxidative phosphorylation, were best predicted by levels of succinyl-CoA synthetase, which converts succinyl-CoA to succinate in the TCA cycle, rather than by dehydrogenases such as isocitrate dehydrogenase that catalyze irreversible and rate-limiting reactions in the cycle[19] (Fig. 3c). The strong collinearity between TCA cycle flux and other mitochondrial energy production reactions in yeast has been observed previously[20]. The levels of succinyl-CoA synthetase may serve as a general indicator of the fluxes related to mitochondrial ATP production in yeast.

The flux in reactions of a few pathways, such as the pentose phosphate pathway and fatty acid biosynthesis, could be predicted neither by eFPA nor by indicator enzymes (Fig. 3a). This observation agrees with previous work where the pentose phosphate pathway was found to be regulated allosterically and by substrate concentration[5,21,22]. Fluxes in other pathways, such as aspartate and glutamate metabolism and the TCA cycle, could only be predicted by indictor enzymes, which may be due to either the prevalence of hub metabolites, the lack of enzyme coregulation in these pathways, or both.

Enzyme reach in human metabolism

Next, we asked if local-pathway enzyme reach also applies to human metabolism. We performed local-pathway eFPA on a human dataset containing protein and mRNA levels for 32 tissues[23] and the human metabolic network model, Human 1[24] (Fig. 4a). Since systems-level flux measurements are not available for human tissues, we could not systematically validate flux predictions like we did with the yeast dataset. However, in the yeast dataset, eFPA captures boundary fluxes, i.e., metabolite exchange with the environment and drainage into biomass, equally well as fluxes within the metabolic network (Fig. 4b, Extended Data Fig. 4 and Extended Data Fig. 7a). In human tissues, such boundary fluxes represent metabolite secretion or uptake and may provide a proxy for the enrichment of metabolites in different tissues. Therefore, we hypothesized that we could validate boundary flux predictions with available tissue-enriched metabolite annotations[25].

To interpret rFPs for tissue enrichment, we defined the tissue enrichment score as the difference of an rFP value from the median rFP of 32 tissues (ΔrFP). For a metabolite, we used the maximum ΔrFP of all relevant transporter reactions that take up or secrete this metabolite. These scores were compared to a collection of 827 tissue-enriched metabolite annotations across 17 of the 32 human tissues for which metabolite enrichment annotations are available [25]. We found that local-pathway eFPA better recalled the reference set than using only the levels of relevant transporter(s) (Fig. 4c, Extended Data Fig. 7b). This observation suggests that local-pathway enzyme reach also occurs in human metabolism.

So far, our analyses used protein levels as input for eFPA. However, mRNA levels obtained from RNA-seq are much more readily available and, therefore, we repeated the eFPA of human metabolites using mRNA levels[23]. Protein and mRNA levels performed similarly (Extended Data Fig. 7c,d), and some metabolites were only predicted by mRNA levels, which could be because of the higher coverage of RNA-seq data compared to proteomic data (Extended Data Fig. 7c,e). Therefore, mRNA levels can effectively be used in eFPA.

eFPA generated a comprehensive landscape of human tissue-enriched metabolic fluxes using either protein or mRNA levels as input (Fig. 4d and Extended Data Fig. 8, Supplementary Table 6). Tissues were frequently predicted to have high flux in specific reactions and therefore have specialized metabolic functions (Fig. 4d and Extended Data Fig. 8). Many of these are consistent with existing knowledge, such as bile acid biosynthesis in liver[26], active metabolism of membrane lipids and signaling molecules in brain[27,28], and energy metabolism in muscle[29]. Our predictions further yielded many

hypotheses, for instance, a high metabolic similarity between lung and spleen that was driven by reactions involved in immunometabolism (e.g., synthesis of leukotrienes) (Fig. 4d). Finally, some predictions were strictly derived by local-pathway eFPA. For example, peroxisomal oxidation of palmitoyl-CoA was predicted to be enriched in heart, although this organ was not enriched for the expression of associated enzymes (Extended Data Fig. 9). However, this prediction is likely valid because the heart generates energy by burning fatty acids[30]. Altogether, these results indicate that eFPA can derive meaningful predictions that are enabled by the inclusion of enzyme reach.

Discussion

We discovered that metabolic enzyme levels can influence reaction flux of both cognate and non-cognate reactions. The reach of enzyme levels is usually local, at the level of individual pathways. However, longer distance enzyme reach can also play a role in predicting metabolism. Our data suggest that network architecture plays a key role that can be unveiled by carefully considering the degree of metabolites in the context of flux balance: reactions that form a flux-balanced route via metabolites with a low degree, i.e., that are produced and consumed by few reactions, provide high predictive power for which enzymes can have reach to which reactions. Overall, we uncover three principles of enzyme reach: a simple correlation where the expression of an enzyme (or transporter) affects flux through its cognate reaction, a distance effect where the coordinated expression of multiple enzymes in a pathway overall affects pathway flux, and a distance effect where the expression of a single indicator enzyme associates with multiple reactions. Our findings highlight that, in addition to flux control by metabolite levels,

enzyme expression also plays a key role. Although it is possible that enzyme levels change to accommodate instead of controlling the change of flux, our observation is in agreement with theoretical models of Metabolic Control Analysis (MCA)[11] that propose that the physiological control of flux is more likely achieved by coordinated changes in multiple enzyme levels and that the controlling effect of each individual enzyme is relatively small[11,31].

Importantly, eFPA is based on a reasonably designed mechanistic modeling framework with only one free parameter (distance boundary). This is in contrast to multi-parameter statistical modeling that is prone to overfitting. The unique advantage of eFPA is that it exploits coordinated enzyme expression in the most relevant pathway(s) of a target reaction without prior pathway definitions. eFPA will provide a powerful tool to predict metabolic flux for organisms for which both expression profiles and a metabolic network model are available. The gap-filling predictions (Fig. 2h) indicate the method is robust even with low-coverage data and therefore may provide a useful tool for the understanding of the metabolism of individual cells, for instance by using single-cell RNA-seq and proteomics data.

References

1 Fell, D. A. *Understanding the control of metabolism*. (Portland Press Ltd, 1997).

2 Daran-Lapujade, P. *et al.* Role of transcriptional regulation in controlling fluxes in central carbon metabolism of Saccharomyces cerevisiae. A chemostat culture study. *J Biol Chem* **279**, 9125-9138, doi:10.1074/jbc.M309578200 (2004).

3 Daran-Lapujade, P. *et al.* The fluxes through glycolytic enzymes in Saccharomyces cerevisiae are predominantly regulated at posttranscriptional levels. *Proc Natl Acad Sci U S A* **104**, 15753-15758, doi:10.1073/pnas.0707476104 (2007).

4 Hackett, S. R. *et al.* Systems-level analysis of mechanisms regulating yeast metabolic flux. *Science* **354**, doi:10.1126/science.aaf2786 (2016).

5 Chubukov, V. *et al.* Transcriptional regulation is insufficient to explain substrate-induced flux changes in Bacillus subtilis. *Mol Syst Biol* **9**, 709, doi:10.1038/msb.2013.66 (2013).
6 Kochanowski, K., Sauer, U. & Chubukov, V. Somewhat in control--the role of transcription in regulating microbial metabolic fluxes. *Curr Opin Biotechnol* **24**, 987-993, doi:10.1016/j.copbio.2013.03.014 (2013).
7 Yilmaz, L. S. *et al.* Modeling tissue-relevant Caenorhabditis elegans metabolism at network, pathway, reaction, and metabolite levels. *Mol Syst Biol* **16**, e9649, doi:10.15252/msb.20209649 (2020).
8 Yilmaz, L. S. & Walhout, A. J. A *Caenorhabditis elegans* Genome-Scale Metabolic Network Model. *Cell Syst* **2**, 297-311, doi:10.1016/j.cels.2016.04.012 (2016).
9 Shlomi, T., Cabili, M. N., Herrgard, M. J., Palsson, B. O. & Ruppin, E. Network-based prediction of human tissue-specific metabolism. *Nature biotechnology* **26**, 1003-1010, doi:10.1038/nbt.1487 (2008).
10 Wagner, A. *et al.* Metabolic modeling of single Th17 cells reveals regulators of autoimmunity. *Cell*, doi:10.1016/j.cell.2021.05.045 (2021).
11 Kacser, H. & Burns, J. A. The control of flux. *Symp Soc Exp Biol* **27**, 65-104 (1973).
12 Watson, E. *et al.* Metabolic network rewiring of propionate flux compensates vitamin B12 deficiency in *C. elegans. Elife* **5**, pii: e17670 (2016).
13 Bulcha, J. T. *et al.* A persistence detector for metabolic network rewiring in an animal. *Cell Rep* **26**, 460-468 (2019).
14 Giese, G. E. *et al.* Caenorhabditis elegans methionine/S-adenosylmethionine cycle activity is sensed and adjusted by a nuclear hormone receptor. *Elife* **9**, doi:10.7554/eLife.60259 (2020).
15 Kochanowski, K. *et al.* Global coordination of metabolic pathways in Escherichia coli by active and passive regulation. *Mol Syst Biol* **17**, e10064, doi:10.15252/msb.202010064 (2021).
16 Yanari, S., Snoke, J. E. & Bloch, K. Energy sources in glutathione synthesis. *J Biol Chem* **201**, 561-571 (1953).
17 Traut, T. W. & Jones, M. E. Inhibitors of orotate phosphoribosyl-transferase and orotidine-5'-phosphate decarboxylase from mouse Ehrlich ascites cells: a procedure for analyzing the inhibition of a multi-enzyme complex. *Biochem Pharmacol* **26**, 2291-2296, doi:10.1016/0006-2952(77)90293-3 (1977).
18 Santoso, D. & Thornburg, R. Uridine 5'-Monophosphate Synthase Is Transcriptionally Regulated by Pyrimidine Levels in Nicotiana plumbaginifolia. *Plant Physiol* **116**, 815-821, doi:10.1104/pp.116.2.815 (1998).
19 Stueland, C. S., Gorden, K. & LaPorte, D. C. The isocitrate dehydrogenase phosphorylation cycle. Identification of the primary rate-limiting step. *J Biol Chem* **263**, 19475-19479 (1988).
20 Blank, L. M. & Sauer, U. TCA cycle activity in Saccharomyces cerevisiae is a function of the environmentally determined specific growth and glucose uptake rates. *Microbiology (Reading)* **150**, 1085-1093, doi:10.1099/mic.0.26845-0 (2004).
21 Stanton, R. C. Glucose-6-phosphate dehydrogenase, NADPH, and cell survival. *IUBMB Life* **64**, 362-369, doi:10.1002/iub.1017 (2012).
22 Ralser, M. *et al.* Metabolic reconfiguration precedes transcriptional regulation in the antioxidant response. *Nat Biotechnol* **27**, 604-605, doi:10.1038/nbt0709-604 (2009).

23 Jiang, L. *et al.* A Quantitative Proteome Map of the Human Body. *Cell* **183**, 269-283 e219, doi:10.1016/j.cell.2020.08.036 (2020).
24 Robinson, J. L. *et al.* An atlas of human metabolism. *Sci Signal* **13**, doi:10.1126/scisignal.aaz1482 (2020).
25 Pang, Z. *et al.* MetaboAnalyst 5.0: narrowing the gap between raw spectra and functional insights. *Nucleic Acids Res* **49**, W388-W396, doi:10.1093/nar/gkab382 (2021).
26 Russell, D. W. The enzymes, regulation, and genetics of bile acid synthesis. *Annu Rev Biochem* **72**, 137-174, doi:10.1146/annurev.biochem.72.121801.161712 (2003).
27 Barber, C. N. & Raben, D. M. Lipid Metabolism Crosstalk in the Brain: Glia and Neurons. *Front Cell Neurosci* **13**, 212, doi:10.3389/fncel.2019.00212 (2019).
28 Hanley, M. R. *et al.* Neural function: metabolism and actions of inositol metabolites in mammalian brain. *Philos Trans R Soc Lond B Biol Sci* **320**, 381-398, doi:10.1098/rstb.1988.0083 (1988).
29 Westerblad, H., Bruton, J. D. & Katz, A. Skeletal muscle: energy metabolism, fiber types, fatigue and adaptability. *Exp Cell Res* **316**, 3093-3099, doi:10.1016/j.yexcr.2010.05.019 (2010).
30 Lopaschuk, G. D. & Saddik, M. The relative contribution of glucose and fatty acids to ATP production in hearts reperfused following ischemia. *Mol Cell Biochem* **116**, 111-116, doi:10.1007/BF01270577 (1992).
31 Fell, D. A. Enzymes, metabolites and fluxes. *J Exp Bot* **56**, 267-272, doi:10.1093/jxb/eri011 (2005).
32 Lu, H. *et al.* A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun* **10**, 3586, doi:10.1038/s41467-019-11581-3 (2019).
33 King, Z. A. *et al.* BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* **44**, D515-522, doi:10.1093/nar/gkv1049 (2016).

METHODS

Brief descriptions of the main methods used in this study are included here. Details for each of the following sections are provided in Supplementary Methods, following the same order of related sections listed here as well as in the main text.

Processing of SIMMER dataset

Fluxomic and proteomic data from Hackett *et al*.[4] were transformed to facilitate correlation analysis and metabolic network modeling. To obtain the relative flux level (referred to as *relative flux*), we normalized absolute flux levels by the growth rate. To be consistent with the Flux Potential Analysis (FPA)[7], where unscaled expression levels are used, we exponentially transformed the log2-scale proteomics data to obtain the unscaled relative abundances (referred to as *relative protein level*). Finally, to correlate the flux of a reaction with expression level of enzymes associated with it, the *relative protein levels* need to be converted to a single value that represents the expression level of the reaction according to the Gene-Protein-Reaction (GPR) associations[7]. We followed a method described previously[7] that produces a normalized, reaction-level, relative expression across conditions that varies from 0 to 1. We refer to this value as *relative expression of reaction*.

The flux and protein measurements were mapped to the most recent consensus metabolic model of yeast (yeastGEM_v8.3.5[32]). Out of the 233 flux measurements, 232 were mapped to the corresponding reactions, with one reaction, r_1099, discarded because of changes in the reaction formula in the new yeast model. Proteins encoded by 486 genes in the model were found to be quantified by proteomics, accounting for 42% of all model genes and associated with 809 reactions (20%). Out of these 809 reactions, 658 had complete expression measurement (i.e., all associated proteins were quantified).

A total of 156 reactions had both the flux and enzyme expression levels determined. These reactions were used in the correlation analysis. Such dual-omics dataset with fluxes and enzyme expression levels was generated for each one of the 25 conditions where media composition and dilution rate were varied.

Correlation analysis of flux and enzyme expression level

For each of the 156 reactions whose flux and expression data are both available, we calculated the Pearson correlation coefficient (*PCC*) between *relative expression of reaction* and *relative flux* using the 25 conditions as 25 datapoints. The *p-value* of each correlation was calculated based on a two-tailed hypothesis test using the *corr* function in MATLAB 2019a and adjusted for multiple testing by *mafdr* function of MATLAB with the '*BHFDR'* method (Benjamini and Hochberg (BH) FDR correction). The resulting FDR values were used to evaluate the correlation. A correlation was considered significant if the FDR is less than or equal to 0.05.

Analysis of pathway-level coexpression

We derived the *PCCs* of *relative expression of reaction* (over the 25 conditions) for every pairwise combination of reactions in a defined pathway (Extended Data Fig. 2e and Supplementary Table 2) and took the median value of these *PCCs* to define the strength of pathway coexpression. Similarly, we defined the strength of pathway-level flux-expression correlation as the median of the flux-expression *PCCs* for all reactions in a pathway. Pathway-level coexpression patterns were defined as the median of *relative expression of reaction* over pathway reactions which formed a 25-element vector for each studied pathway.

Analysis of pairwise cross-informing rate

We considered all connected reaction pairs with measured flux and/or protein levels for the cross-informing rate analysis. For instance, to calculate the flux-flux cross-informing rate, we collected every connected reaction pair with fluxes determined for both reactions. We calculated *PCC* and *p-values* for each collected pair and considered reaction pairs with FDR less than 0.05 and *PCC* greater than 0 as *cross-informed*. The cross-informing rate was defined as the proportion of cross-informed pairs in a given set of pairs. The sets of pairs were determined based on the network degrees of bridging metabolites, i.e., metabolites that connect the paired reactions. Importantly, due to the limited number of data points (i.e., only 232 reactions with measured fluxes), we grouped the pairs with approximate bridging metabolite degrees in the calculation of cross-informing rate. Each group produced a single data point in Fig. 2b in which the x-axis refers to the average bridging metabolite degree of pairs in the group and y-axis refers to the proportion of *cross-informed* reaction pairs, i.e., the cross-informing rate for the group.

Enhanced Flux Potential Analysis (eFPA)

FPA (and eFPA) is a specific flux balance analysis (FBA) problem that calculates the maximum flux potential (FP) of a target reaction under certain constraints that address relative expression of reactions and their distance from the target[7]. The mathematical details of FPA can be found in Supplementary Methods and in our previous publication[7]. In brief, as in a regular FBA problem, FPA is done under the steady-state assumption with reaction fluxes constrained between prescribed upper and lower boundaries. Specifically, a weighted sum of flux in the network is further constrained to be less than or equal to a constant that is referred to as *flux allowance* while the weight of each flux is

inversely proportional (but not linearly, see the next section) to the *relative expression of (the corresponding) reaction.* The flux of the target reaction is selected as the objective function, which is maximized to find FP as the objective value. Since the flux allowance is a constant, reactions with smaller weight coefficients (i.e., higher relative expression) in the weighted sum are more likely to carry larger flux to maximize the flux left for the target reaction, thereby conveying the influence of their expression changes to the flux potential of the target. The enhanced flux potential analysis (eFPA) algorithm stays the same as original FPA except for the new distance decay functions and the use of weighted metabolic distance instead of naïve metabolic distance (see below).

Weight coefficients and distance decay function

The key component of FPA is the calculation of the *weight coefficients*. Weight coefficient is proportional to the reciprocal of the *relative expression of reaction* and is scaled by a distance decay function. The distance decay function represents the decrease of the influence of network reactions (enzyme reach) on FP as their distance from the target reaction increases. Thus, the distance function downscales the weight coefficients as the distance to the target increases, which would allow a reaction to take large flux values without significantly affecting target FP calculations even if its expression coefficient is high (i.e., if it is poorly expressed). The eFPA employed two redesigned distance decay functions compared with the original FPA. Please refer to the Supplementary Methods for mathematical details.

To integrate the bridging metabolite degree (Fig. 2a) with eFPA, we weighted the metabolic distance between two adjacent reactions based on the number of connections to the metabolites that connect them (Extended Data Fig. 3b), such that the distance

between reactions connected by hub metabolites is upscaled. Thus, the distance between a pair of reactions can be greater than the original distance that is the number of reactions between them plus one. We refer to the new distance measure as *weighted metabolic distance*. Weighted metabolic distance effectively encourages the reach of influence from enzymes in linear pathways to the target reaction (i.e., connected by metabolites of lower network degrees), thus achieving the integration of enzyme changes of interest. The distance boundary parameter in eFPA is measured in the scale of weighted metabolic distance. To relate this parameter to the actual length of the integrated pathway for interpretation, we further converted it to an interpretable metabolic distance from the target (i.e., the maximum distance of integrated reactions to the target) that was used for data visualization in Fig. 2d and Extended Data Fig. 4. Please refer to Supplementary Methods (sections *Mathematical formulation of Flux Potential Analysis (FPA)*, *Weight coefficients and distance decay function*, *Weighted metabolic distance* and *Calculation of effective distance boundary*) for the mathematical details.

eFPA of SIMMER dataset

As in our previous FPA analysis[7], eFPA does not rely on any quantitative boundary constraints as we are focused on the effect of enzyme expression changes. All available nutrients based on the definition of culturing media were made freely exchangeable by setting the lower boundary of pertaining exchange reactions to -1000. These nutrients include phosphate, glucose, ammonium, uracil and leucine (Supplementary Table 5). No non-growth-associated maintenance (NGAM) was imposed (lower boundary to 0) during eFPA. This constrained model was used in all eFPA analyses.

To qualitatively account for media differences across conditions, we blocked the uptake of unavailable nutrients in each condition. For example, uracil and leucine uptakes were blocked in the eFPA of phosphate-limiting, carbon-limiting and nitrogen-limiting conditions. In addition, to address the extreme low abundance of nutrients (e.g., glucose concentration is over 20-fold lower in carbon-limiting conditions), we set an arbitrarily large gene expression coefficient on the pertaining exchange reactions when applicable. Please refer to Supplementary Methods for details.

eFPA was performed by a modified version of the generic eFPA function (https://github.com/WalhoutLab/eFPA) to enable highly parallel computation in a computer cluster. The distance boundary parameter was changed according to the question of interest as indicated in the text. All 232 reactions (Supplementary Table 2) with determined fluxes in SIMMER dataset were analyzed with eFPA.

Other details regarding parameterization, analysis and interpretation of eFPA in yeast can be found in Supplementary Methods (sections *Correlation analysis of eFPA results and measured fluxes*, *Titration of the distance boundary*, *Calculation of effective distance boundary*, *Deciphering prediction mechanisms of eFPA* and *eFPA on the drainage flux of biomass precursors*).

Randomization test of eFPA

To assess the statistical significance of eFPA modeling, we shuffled the rows (reaction labels) of the *relative expression of reaction* matrix (658 reactions by 25 conditions), i.e., randomized the association between the expression levels and reaction labels. After shuffling, eFPA was performed with randomized data following the same procedure as described above. This randomization was performed for 1000 times. It is noteworthy that

we were unable to perform a greater scale of randomization due to the overwhelming computational demand.

Processing of human tissue dataset

The quantitative proteomics and transcriptomics of human tissues contain RNA and protein levels of more than 12,000 genes across 32 normal human tissues quantified based on 201 individual primary samples[23]. We analyzed all 32 tissues using the *tissue median* provided in the referred study. We rescaled both RNA and protein tissue medians to make them suitable for system-level modeling, including the conversion of log2-scale data to unscaled values and a variance-stabilizing transformation based on Tissue Specificity score reported in the referred study[23]. Please refer to Supplementary Methods, section *Processing of tissue expression data*, for details about the data transformation.

Processing of tissue-enriched metabolite set

The tissue enriched metabolite sets were obtained from MetaboAnalystR software (version 4.93, www.metaboanalyst.ca)[25]. There were 73 tissue-enriched or subcellular-enriched metabolite sets in the original data table. Tissue labels in MetaboAnalystR metabolite set were manually matched with the 32 tissues measured in Jiang *et al.* [23] (Supplementary Table 7). In the end, 331 metabolites from the human metabolic model were defined with an enrichment in at least one of 17 matched tissue types, resulting in 827 tissue-metabolite pairs in total (Supplementary Table 8).

Human metabolic network model

The newest consensus human metabolic model, Human1 (version 1.5.0), was downloaded from metabolicatlas.org[24]. To increase the numerical stability, the stoichiometry matrix of the model was adjusted to avoid reactions with overly large coefficients or small flux capacities. To model the differential nutrient availability in blood stream, a set of specialized uptake reactions were added to control the mass influx of each type of imported nutrient using a flux balance method that we developed previously[7]. Detailed descriptions on model modifications can be found at Supplementary Methods, section *Human metabolic network model*.

eFPA analysis for human tissues

To comprehensively model the tissue metabolism in humans we followed the modeling pipeline MERGE, which we previously developed for modeling *C. elegans* tissue metabolism[7]. In brief, two stages of modeling were performed in a sequential manner. In the first step, a semi-quantitative modeling of on/off status and direction of reaction fluxes was performed using iMAT++ algorithm[7]. This step is to globally fit the distinct expression levels of enzymes in tissues to the metabolic network, to derive (1) a flux distribution that we named as Optimized Flux Distribution (OFD) and that can be used to assign the flux directions for reversible reactions; and (2) a tissue-specific metabolic network in which inactive reactions that cannot carry flux according to flux variability analysis[7] were removed from the network. After iMAT++ modeling, the eFPA analysis was performed on the derived tissue networks. In addition, the rFP predictions generated from eFPA was overlayed with OFD to derive high-confidence predictions with reaction directionality. This step is important for the interpretation of eFPA predictions of reversible reactions that show high flux potential in both directions, such that the direction predicted by iMAT++ is

chosen as the most likely metabolic function with high potential. Together, this tissue modeling pipeline provides a comprehensive collection of high-quality predictions about tissue-enriched metabolic fluxes.

Two sets of tissue networks were built for the 32 tissues based on either proteomics data (protein-based tissue network; PTN) or RNA-seq data (RNA-based tissue network; RTN). PTN and RTN were used in eFPA based on proteomics and RNA-seq data, respectively. The decay function of eFPA algorithm used in human tissue modeling was the exponential decay with a distance boundary of 6 (i.e., the local-pathway eFPA was used), and the distance measurement is the weighted metabolic distance. When not specified (i.e., referred to as *"eFPA"* or *"Protein eFPA"*), the input data for eFPA were protein levels of genes that are commonly detected by both proteomics and RNA-seq (12,121 genes, covering 2869/3627, 79%, genes in the Human 1 model). For eFPA using RNA expression as input (referred to as *"RNA eFPA"* in figures and texts), we performed eFPA using either commonly detected genes (12,121, labeled as *"common genes"* or not specified) or all detected genes by RNA-seq (19,273, labeled as *"all genes"*).

To derive the tissue-metabolic landscape, we analyzed all regular reactions in human 1 model except for transport, exchange, and custom uptake reactions (see above). To model the tissue-enriched metabolites, we performed eFPA on transporter reactions that take up or secrete metabolites from or into extracellular space, respectively. The computation was performed in Massachusetts Green High Performance Computing Center (GHPCC) with 512 cores. To exemplify the computational performance, eFPA of all internal regular reactions (7,248) and 32 tissues took less than 30 minutes.

Detailed parameters, settings and modifications can be found in Supplementary Methods, sections from *Building human tissue networks* to *eFPA analysis: boundary reactions*.

Linking boundary flux predictions to tissue-enriched metabolites

eFPA produced an rFP matrix with columns representing the 32 tissues and rows the reactions (i.e., internal reactions and transporters). We next row-wise centered this matrix by subtracting the row median, so that we obtained the delta rFP from the median (ΔrFP). The maximum ΔrFP among the transporter reactions that import or export a target metabolite was defined as the tissue enrichment score of the metabolite. This yielded predictions for 289/331 metabolites that we had reference data for. These 289 metabolites accounted for 740/827 metabolite-tissue pairs that were annotated as tissue-enriched, and were used in further evaluations. To convert the ΔrFPs into binary calls of tissue enrichment (*enriched* or *not enriched*), we set a threshold on the ΔrFPs, where metabolite-tissue pairs with ΔrFPs higher than this cutoff were labeled as *tissue enriched*. Hypergeometric test was used to check if the predicted *tissue enriched* metabolite-tissue pairs significantly overlap with the reference *tissue enriched* metabolite-tissue pairs under various ΔrFP threshold. To compare the ΔrFP values for known tissue-enriched metabolites with that for all metabolites (approximately null distribution for non-tissue-enriched metabolites), the ΔrFP values for all transportable metabolites (i.e., the maximum ΔrFP of all pertaining transporter reactions given a metabolite of interest) in the network (n=1,379) were computed. We compared this overall distribution with the ΔrFP distribution of the 740 reference tissue-metabolite pairs using boxplots. We also tested

whether the median ΔrFP of tissue-enriched metabolites is greater than the median of the background via rank-sum test.

Clustering rFPs of internal reactions

To generate Fig. 4d, reactions were filtered unless their ΔrFP were greater than 0.2 (tissue-enriched) or lower than -0.2 (tissue-depleted) in at least one of the 32 tissues. To address the cases when eFPA predicts tissue enrichment for both directions of a reversible reaction, we selected the eFPA predictions based on the relevant flux distribution (OFD) predicted by iMAT++ (see above). The reactions that passed these criteria were merged at the end. We termed these final filtered predictions *high-confidence predictions of human tissue metabolism* and generated the clustergram with them using the *clustergram* function in MATLAB with 'cosine' distance for both row and column clustering. These high-confident predictions were starred (i.e., labeled with "*") in their reaction ID in Supplementary Table 6.

Clustering tissue-enrichment of subsystems

To visualize the subsystem-level metabolic specialization in tissues, we generated a matrix where rows are subsystems and columns are the 32 tissues. The values in the matrix are number of tissue-enriched reactions (ΔrFP greater than 0.2) assigned to each subsystem and tissue. The matrix was row-wise normalized by dividing all values with the maximum, which yielded the relative tissue-enrichment for subsystems shown in the heatmap (Extended Data Fig. 8).

Acknowledgments We thank members of the Walhout lab for supports and discussions on this project. We thank Caryn Navarro, Hefei Zhang, Job Dekker, Olga Ponomarova, Robert Brewster and Shivani Nanda for discussions on the manuscript. This work was supported by a grant from the National Institutes of Health GM122502 to A.J.M.W.

Author contributions X.L., L.S.Y. and A.J.M.W. conceived the project and wrote the manuscript. X.L. and L.S.Y. developed the eFPA algorithm. X.L. performed all investigations under supervision of L.S.Y. and A.J.M.W.

Competing interests Authors declare that they have no competing interests.

Data and materials availability Core data are available in the main text or the supplementary materials. The complete datasets and codes to reproduce the presented work can be found at https://github.com/WalhoutLab/eFPA. A tutorial for applying eFPA is also included in the Github repository.

Correspondence Please address correspondence to Marian Walhout, marian.walhout@umassmed.edu or Safak Yilmaz, lutfu.yilmaz@umassmed.edu.

Supplementary information

Supplementary Methods

Supplementary Notes

Supplementary Tables 1-9

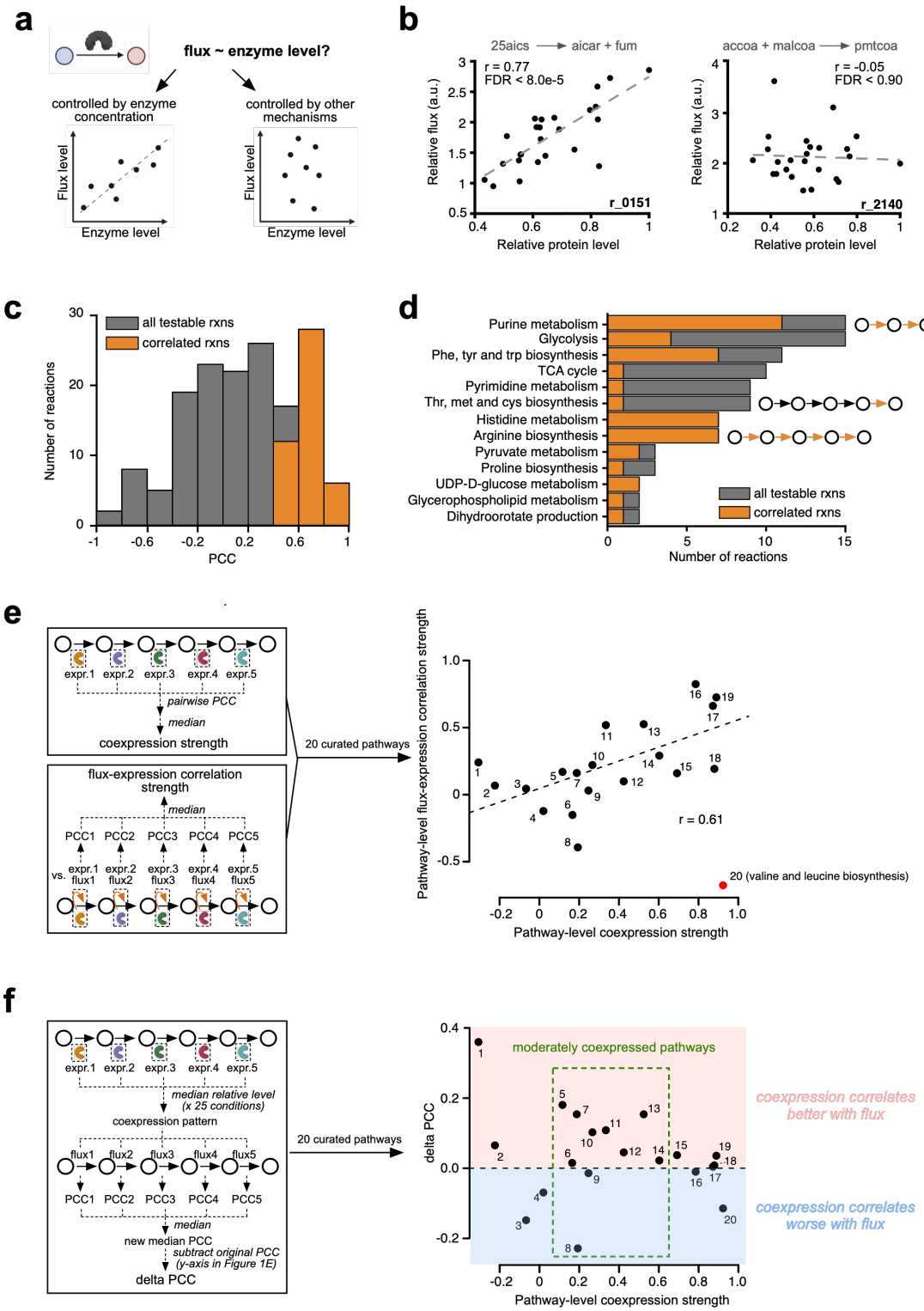


Fig. 1: Pathway-level coordinated regulation of enzyme levels dictates metabolic flux in yeast. a, Models for quantitative relations between enzyme levels and flux. **b**, Representative examples for flux-enzyme level correlation. Each datapoint represents a

measurement (25 total) of relative flux and relative enzyme level for the indicated reaction. The metabolite abbreviations are defined by BiGG model nomenclature[33]. a.u., arbitrary units. **c**, Pearson Correlation Coefficient (PCC) distribution for the 156 reactions for which flux and enzyme levels were both measured. Reactions with significant positive correlation (FDR < 0.05, PCC > 0) are indicated in orange. **d**, Metabolic pathway annotations of the 156 reactions. **e**, Correlation between pathway-level enzyme coexpression and flux-expression concordance. Pathway information corresponding to the number indices is provided in Extended Data Fig. 2e. The trend line was fitted without the outlier (red dot). The correlation coefficient (r) of the trend line is shown in the plot. **f**, Comparing flux concordance with individual and pathway-level enzyme expression.

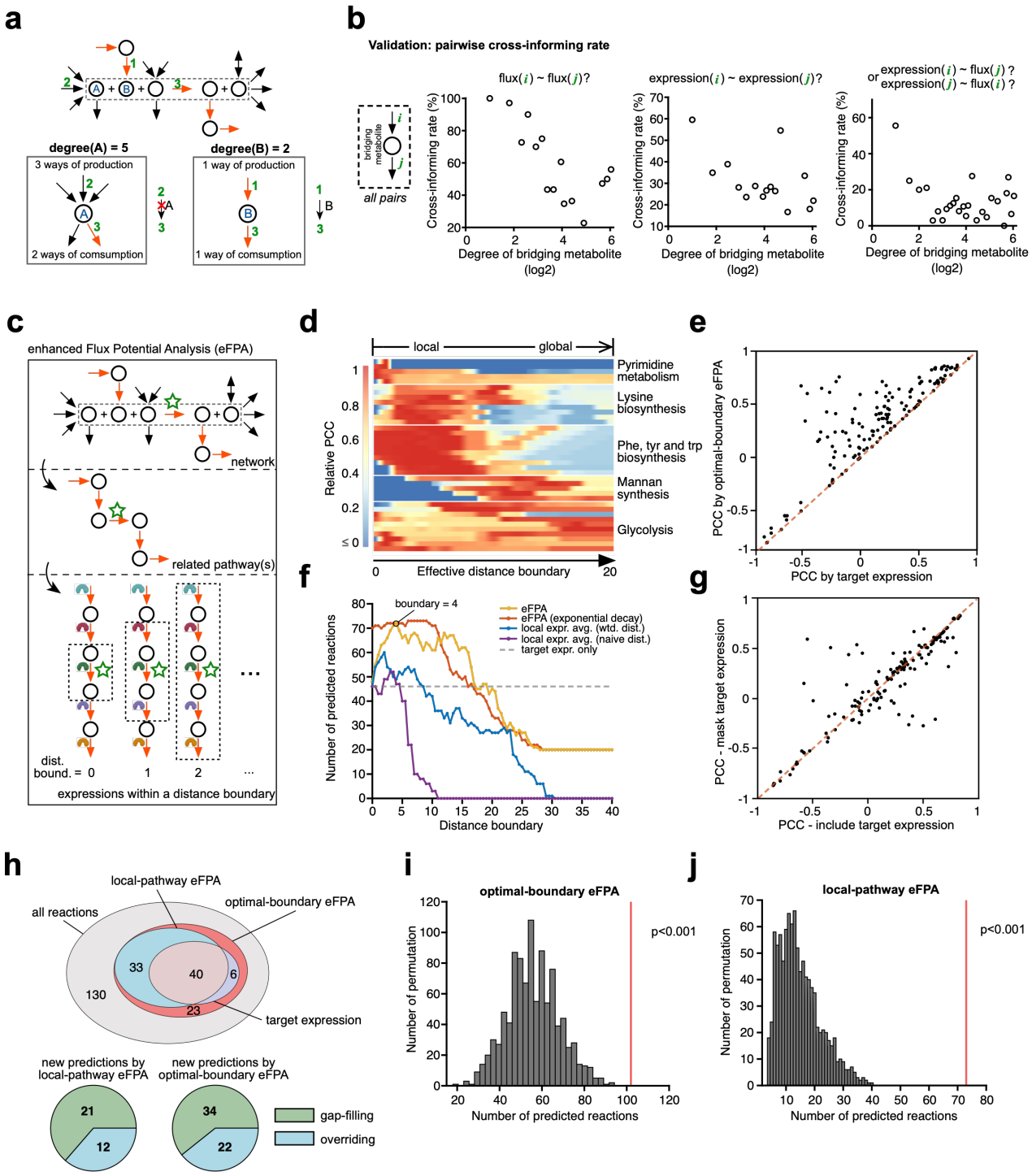


Fig. 2: Enhanced flux potential analysis (eFPA) predicts flux by leveraging pathway-level enzyme expression. a, Toy network illustrating the degree of bridging metabolites. Circles are metabolites and arrows are reactions. The red arrows indicate a

hypothetical pathway for the target reaction inside the dotted box. **b**, Analysis of pairwise cross-informing rate. **c**, Schematic for eFPA. The target reaction is labeled by a green star. **d**, Representative results of eFPA for different reactions. Values are row-normalized flux-rFP PCCs. Rows correspond to different target reactions and columns to the effective distance boundary that indicates the maximum length of pathway(s) integrated (see Methods). **e**, Comparison of the correlation of flux with target rFP and target expression. **f**, Total number of predicted reactions with different approaches. Local expression average was defined by the average expression of reactions that are connected to target reaction within a distance boundary measured by naïve or weighted distance. **g**, Comparison of local-pathway eFPA predictions with and without inclusion of enzyme levels corresponding to the target reaction. Values are flux-rFP PCCs. **h**, Classification of fluxes predicted by different types of eFPA. **i**, **j**, Randomization test of optimal-boundary eFPA (**i**) and local-pathway eFPA (**j**). The histograms show the distribution of the number of reactions predicted in 1000 permutations and the red line indicates the real observation.

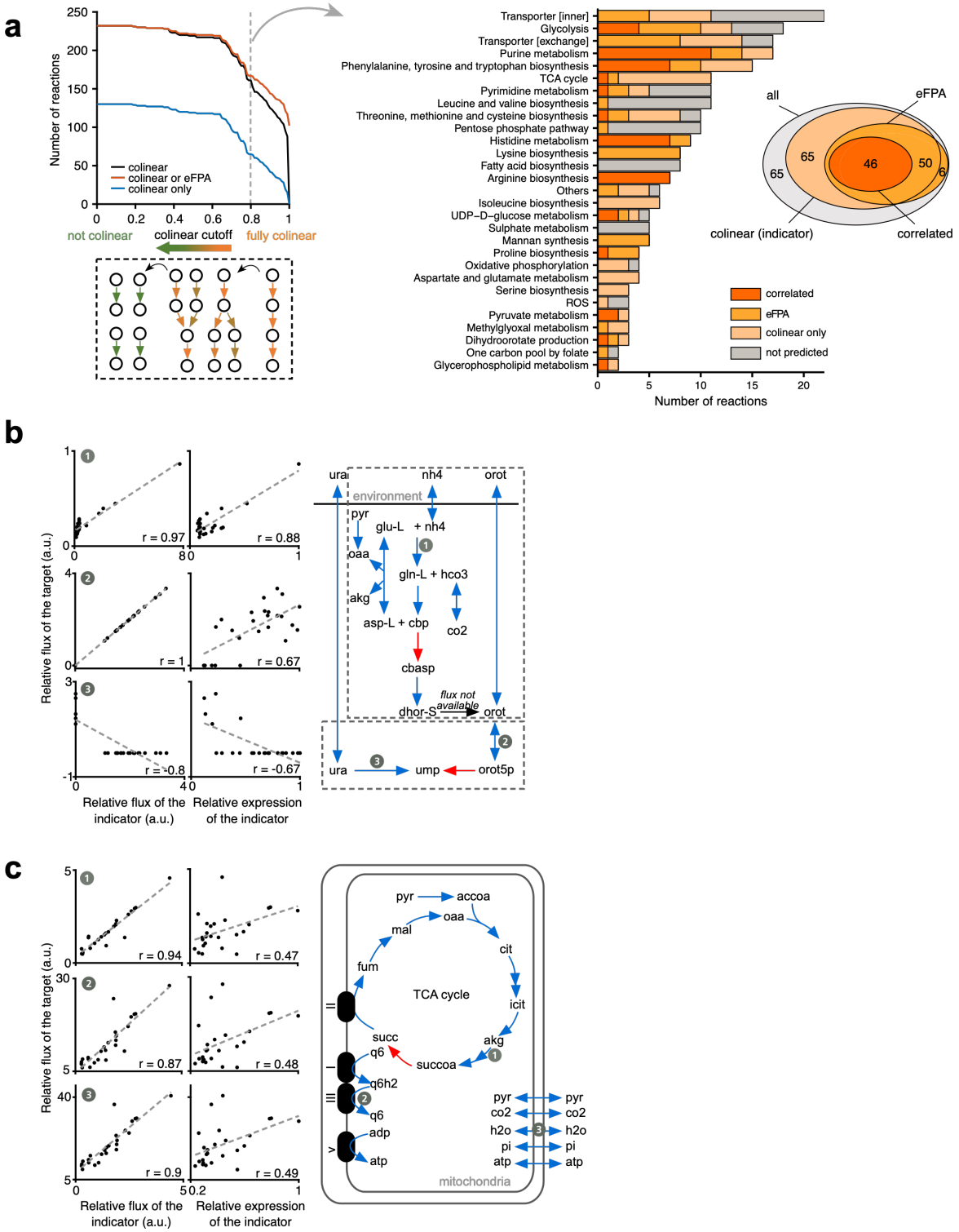


Fig. 3: Indicator enzymes complement eFPA and may identify rate-limiting enzymes. **a**, Flux collinearity analysis. Collinearity is defined as the absolute PCC between two reaction fluxes. Given a collinearity cutoff (x-axis), the black line shows the number of

reactions that were colinear with any of the 46 correlated reactions. The number of reaction fluxes predicted by either collinearity or eFPA are shown in red, while those only predicted by collinearity are shown in blue. The cartoon illustrates flux coupling between reactions indicated by different levels of collinearity. Pathway annotations of reactions predicted by either correlation, eFPA, or collinearity are shown on the right. **b**, The indicator of pyrimidine metabolism fluxes. In the network cartoon, red arrows refer to the indicator and blue arrows to the target reactions. Two distinct sets of colinear fluxes were identified and are labeled by the two dashed boxes. The correlation between the flux of the target reactions and the flux or expression of the indicator is plotted for representative reactions. **c**, As in (**b**), but for mitochondrial energy metabolism fluxes.

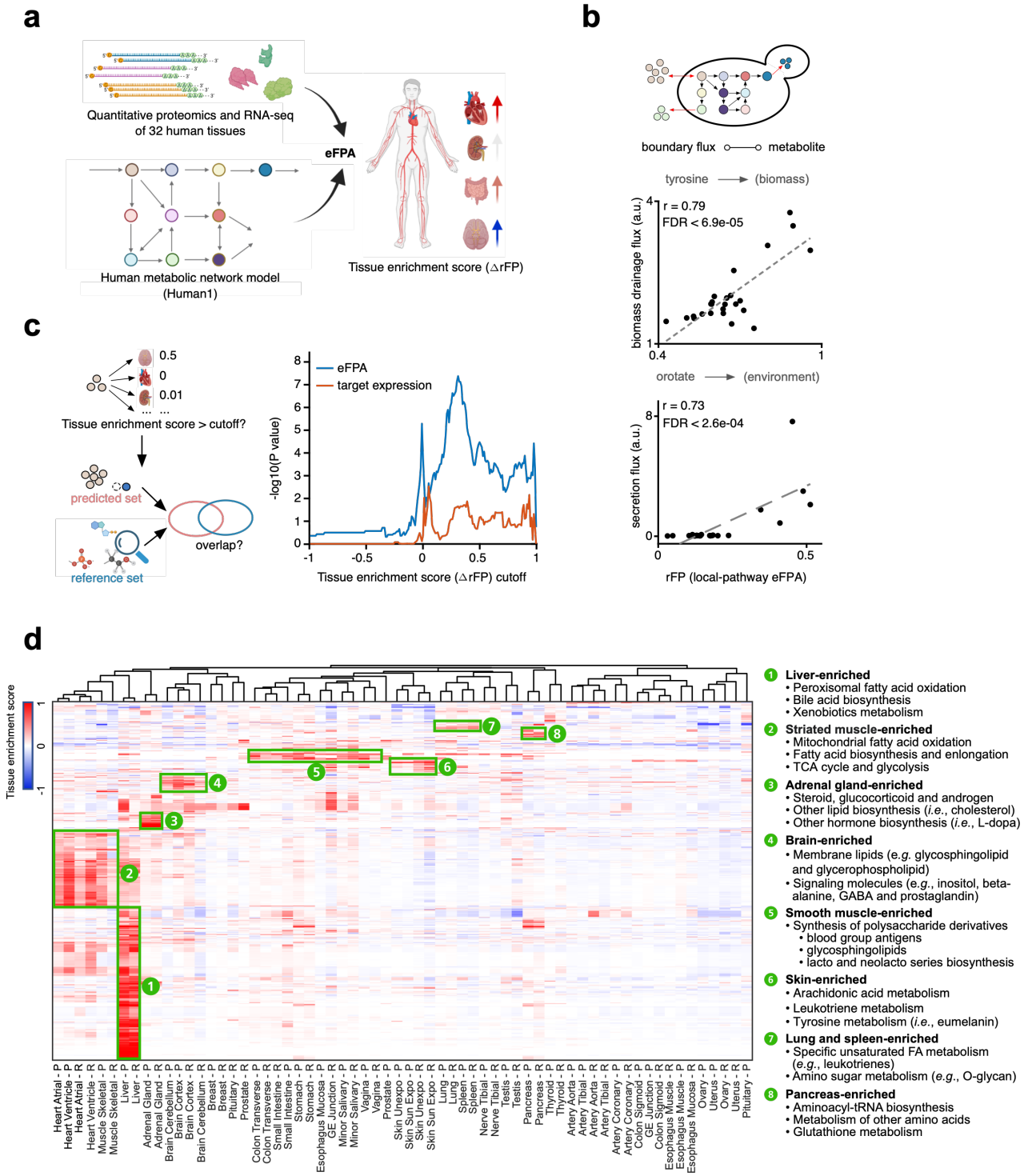
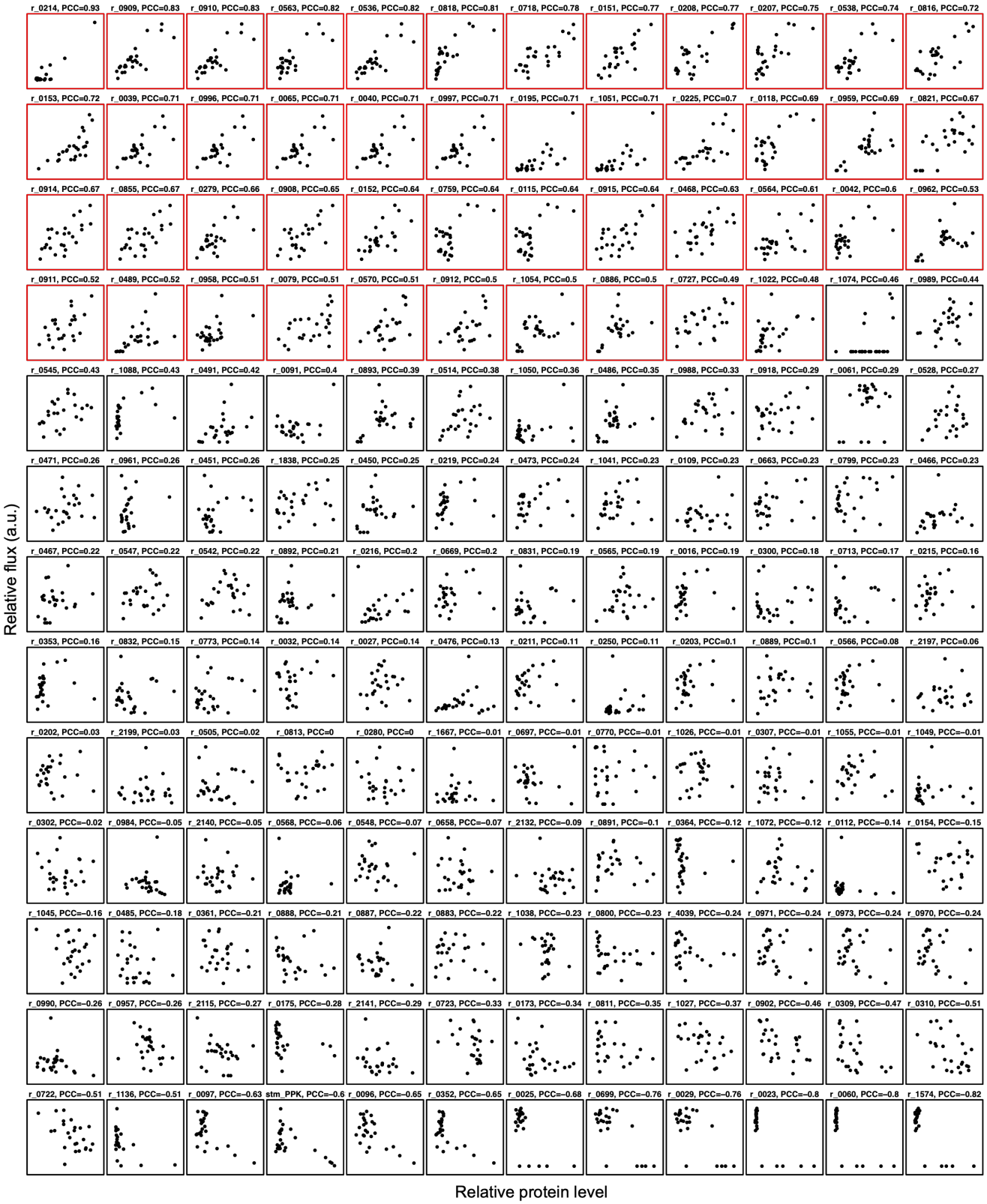
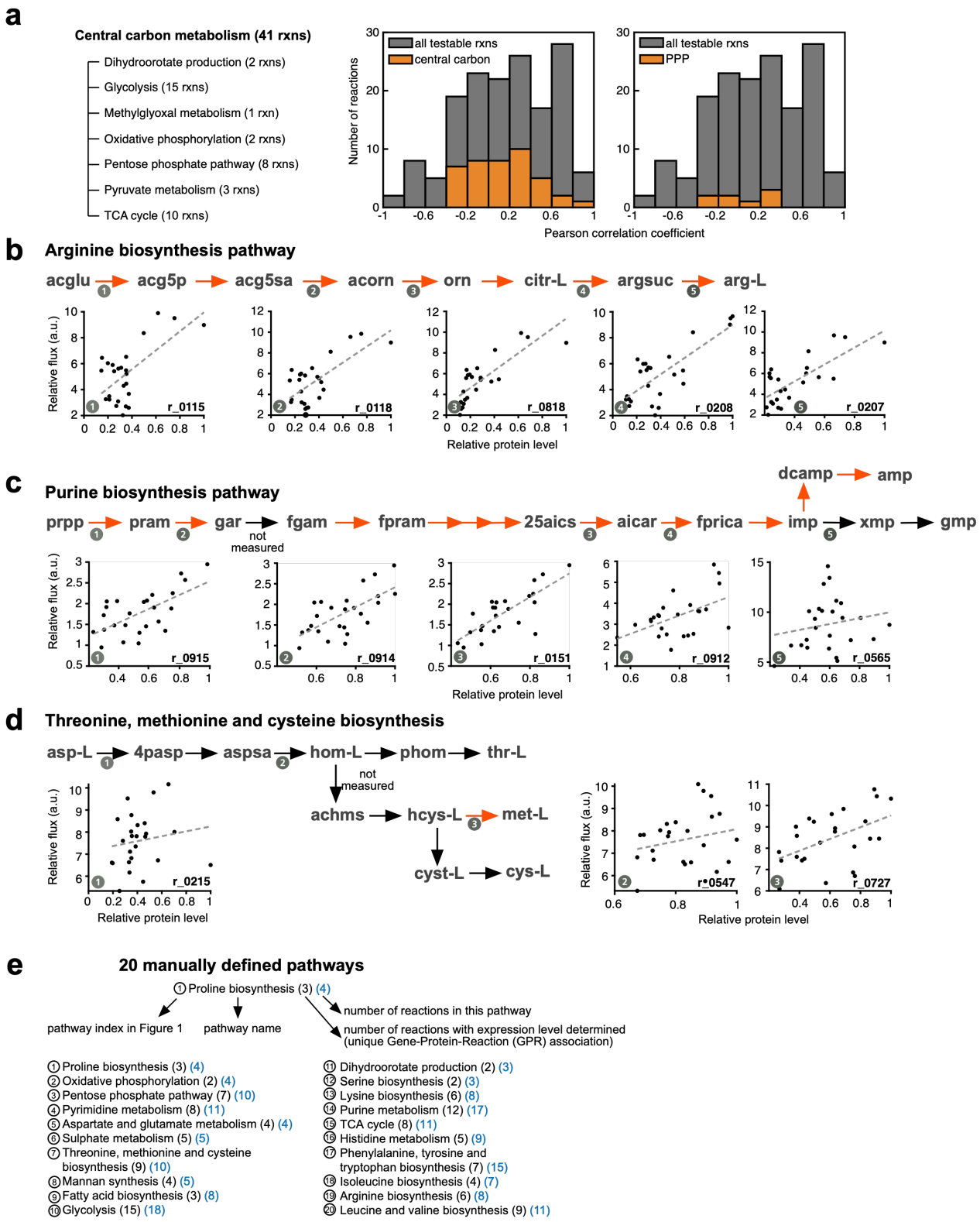


Fig. 4: eFPA predictions for human tissue metabolism. a, Cartoon of eFPA analysis. rFP values were centered by subtracting the median rFP of a reaction across all 32 tissues, and the difference from median (ΔrFP) was referred to as the tissue enrichment

score. **b**, Representative predictions of boundary flux in yeast. The biomass drainage flux of tyrosine and the secretion flux of orotate are shown. **c**, Enrichment p-values (hypergeometric test) between predicted tissue-enriched transportable metabolite set and the reference set under different tissue enrichment score (ΔrFP) cutoffs. **d**, Heatmap of the ΔrFP of 3441 tissue-enriched reactions in 32 human tissues. Tissue-enriched reactions were defined with ΔrFP greater or equal to 0.2 in at least one tissue based on either protein or mRNA expression data as the input. Only high-confidence predictions are shown (see Methods). Column label shows tissue names with a single letter suffix indicating the prediction was made from protein (-P) or mRNA (-R) data.

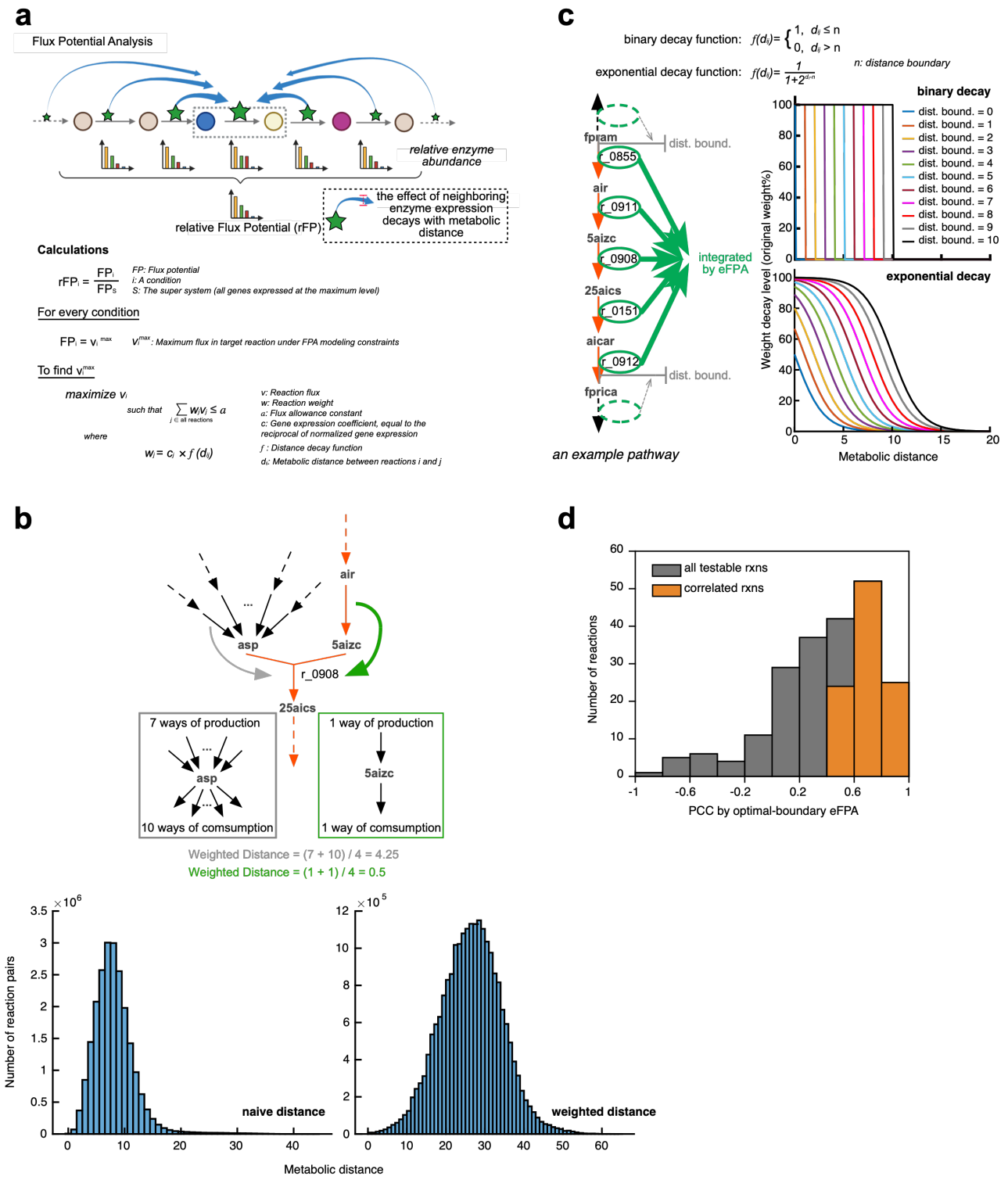


Extended Data Figure 1: Scatter plots showing the flux-enzyme level correlation for all 156 testable reactions in yeast.



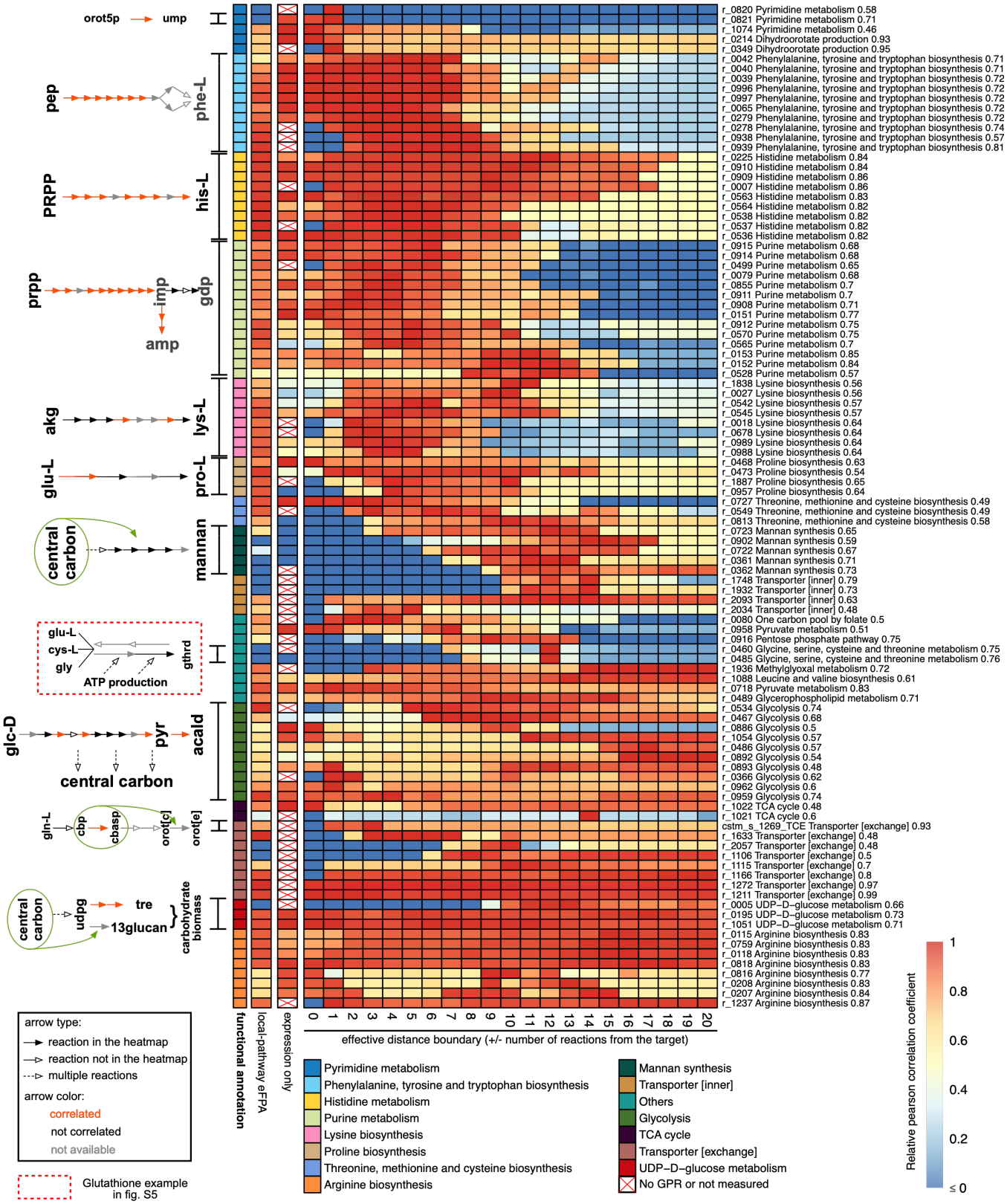
Extended Data Figure 2: Analysis of flux-enzyme level correlation in metabolic pathways. a, The *PCC* distribution for reactions of central carbon metabolism. The

distribution of reactions of the pentose phosphate pathway (PPP) is shown on the right. **b**,**c**,**d,** overlay of the flux-enzyme level correlations on selected metabolic pathways. The three selected pathways include ones that are fully (**b**) or partially (**c**) composed of correlated reactions, and one that has only one correlated reaction (**d**). Orange arrows indicate reactions that show significant correlation (FDR < 0.05, *PCC* > 0) and black arrows indicate uncorrelated reactions. **e**, Twenty manually-defined pathways that are labeled by their indices in Fig. 1e,f.



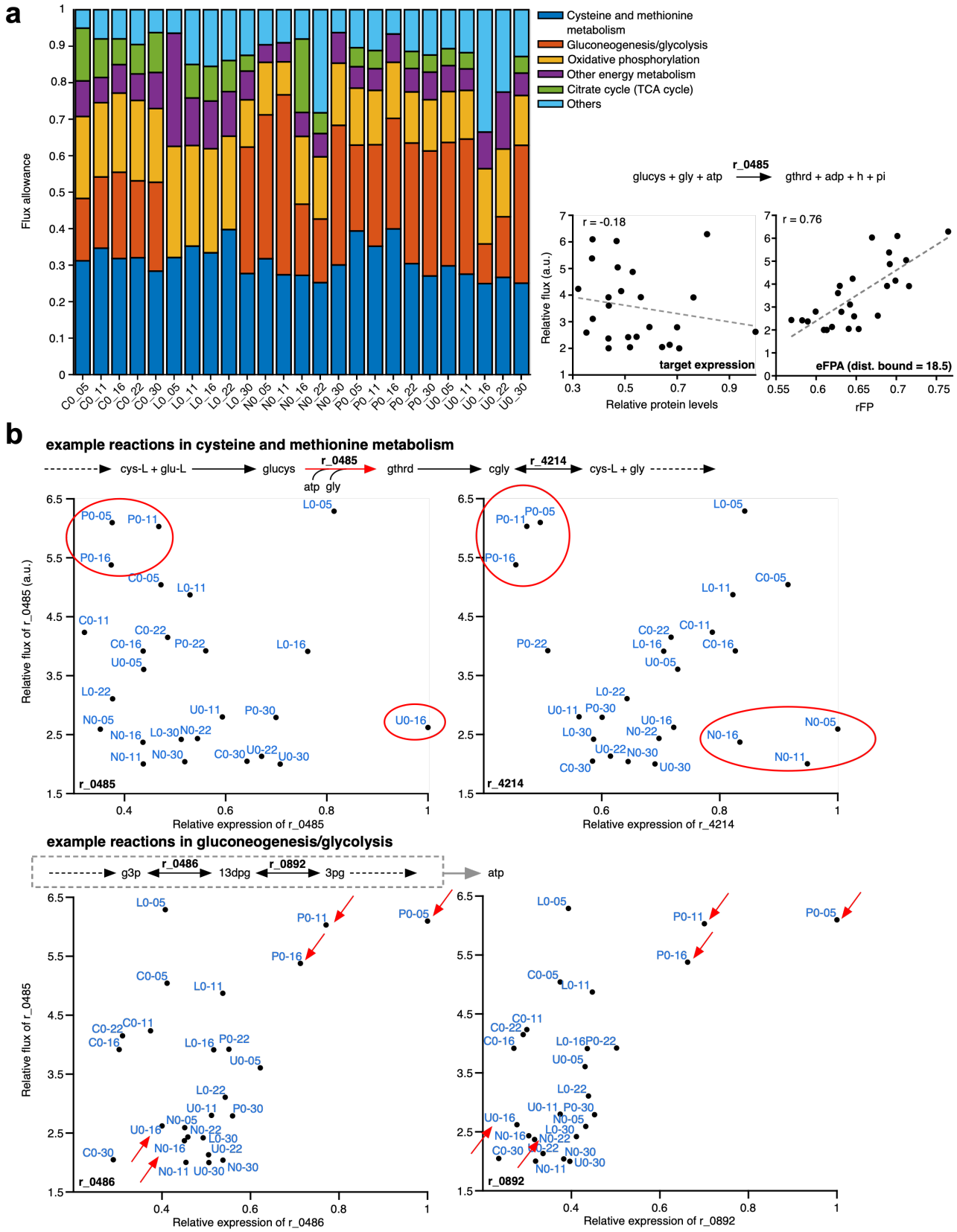
Extended Data Figure 3: Principles of FPA and eFPA. a, Schematic of Flux Potential Analysis (FPA). FPA and eFPA integrate the levels of enzymes that catalyze reactions

surrounding a target reaction to predict relative flux potential (rFP) of the target reaction. The contribution of surrounding enzymes can be tuned by a distance decay function such that FPA is versatile to integrate expression information from a local subnetwork to the entire network. The mathematical formulation of FPA is briefly summarized (also see Supplementary Notes 1). **b**, The weighted metabolic distance. A cartoon illustrating the calculation of weighted metabolic distance is shown on the top. On the bottom, the distributions of naïve (unweighted) and weighted metabolic distance are shown for all reaction pairs in the yeast metabolic model. **c**, The distance decay functions of eFPA. The formulas of binary (used in optimal-boundary eFPA) and exponential decay functions (used in local-pathway eFPA) are shown in the figure. The left panel shows a cartoon to illustrate the concept of distance boundary. The right panel shows the decay curve of the two functions. **d**, The distribution of *PCC* between experimental flux and rFP predicted by optimal-boundary eFPA for the 232 flux-measured reactions. Significantly correlated reactions (FDR < 0.05, *PCC* > 0, min-max rFP range > 0.2) are colored in orange.



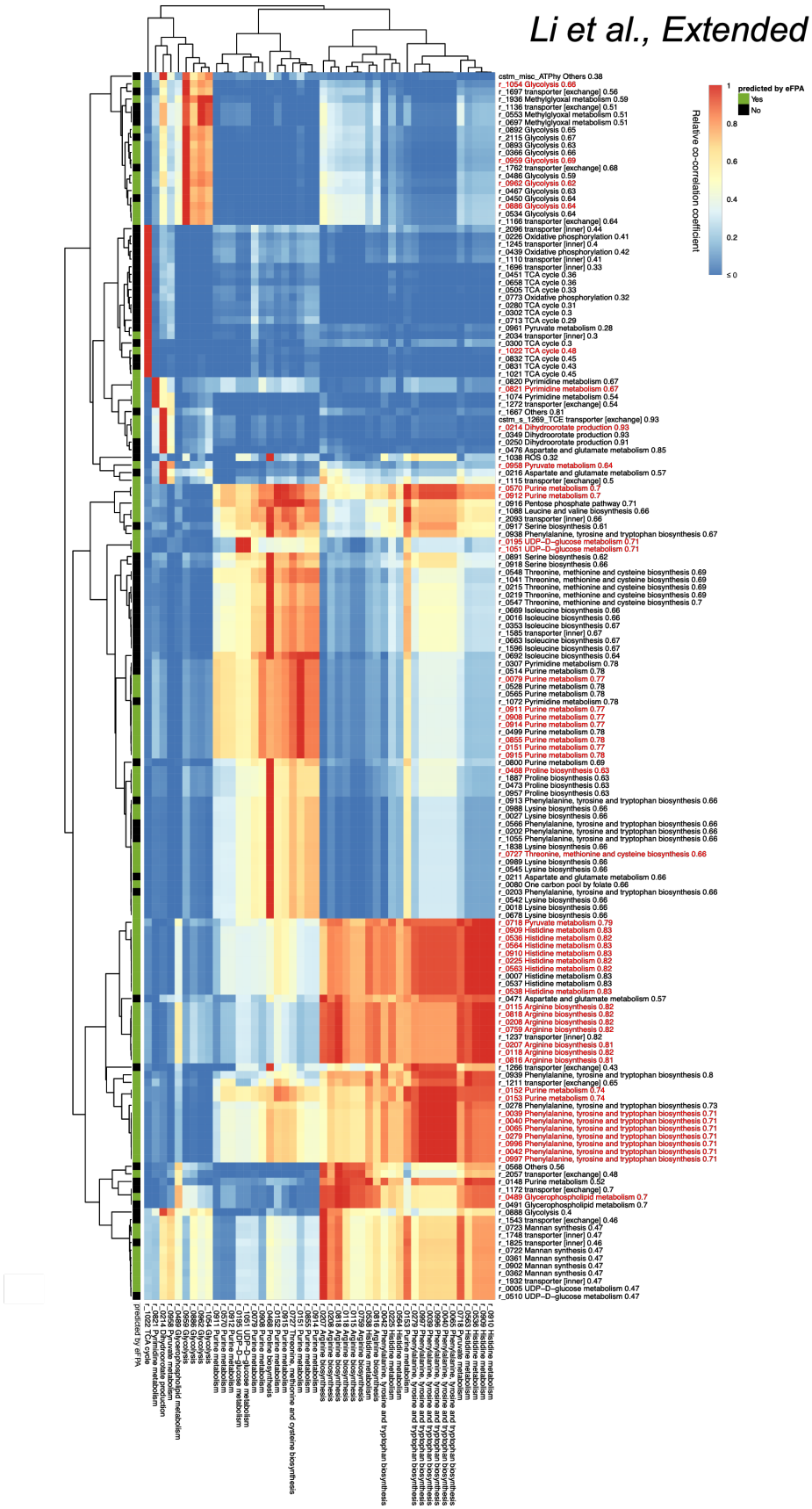
Extended Data Figure 4: Heatmap of PCC between flux and rFP with a titration of distance boundary. The relative PCC (row-wise normalized by dividing each value by

the row maximum) for 102 predicted reactions. The numbers on x-axis indicate the effective distance boundary, which represents the calculated actual length of the integrated pathway based on the distance boundary parameter in the scale of weighted metabolic distance. Target reactions (rows) are arranged based on their position in the pathway they are associated with, and reaction IDs are indicated on y-axis together with the associated pathway and the optimal-boundary PCC. Significant contributors (e.g., relevant local pathways) to eFPA analysis of some targets are depicted on the left.

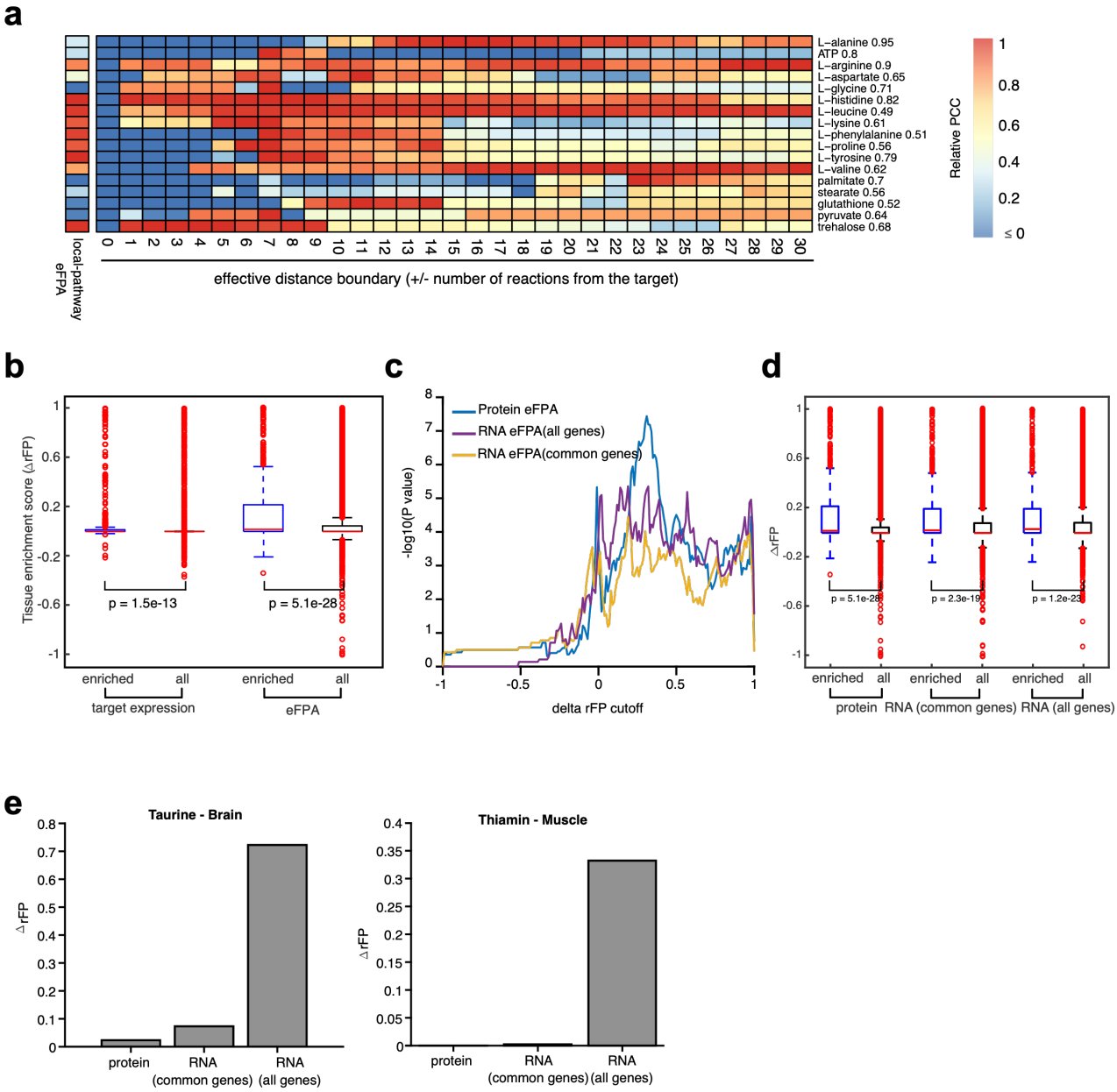


Extended Data Figure 5: Mechanisms for the prediction of glutathione synthesis flux a, Left panel: flux allowance contributions of different metabolic pathways (colors) in

different conditions (x-axis) during the optimal-boundary eFPA of the glutathione synthesis reaction r_0485, which belongs to the *cysteine and methionine metabolism* in the pathway annotations. Flux allowance contribution is defined by the product of reaction weight coefficient and flux (see Supplementary Methods for details). Since total allowance is held constant, major contributors to this value have the biggest influence on the prediction of the flux of the target. Right panel: scatter plot showing the correlation between rFP and relative flux of glutathione synthesis. **b**, Scatter plot showing the correlation between the relative protein levels of key contributing reactions and the flux of glutathione synthesis. The red circles or arrows are used to highlight in all graphs the conditions in which the enzyme levels of glycolysis reactions are concordant with the flux of glutathione synthesis while that of reactions in close proximity to the target are not.

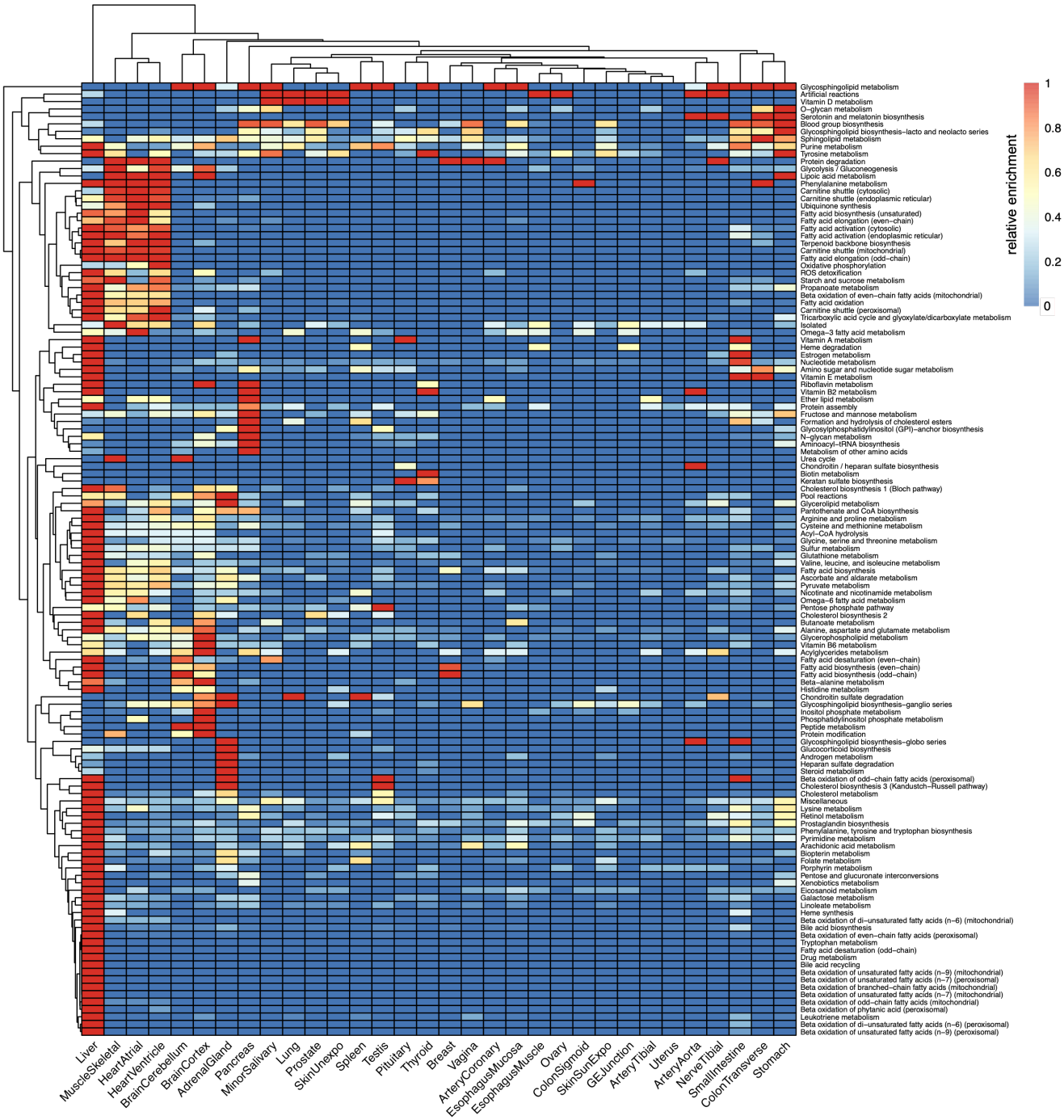

predicted by eFPA
Yes
No
Relative co-correlation coefficient
1
0.8
0.6
0.4
0.2
≤ 0
r_1022 TCA cycle
r_0821 Pyrimidine metabolism
r_0489 Glycerophospholipid metabolism
r_0488 Glycolysis
r_0886 Glycolysis
r_0959 Glycolysis
r_0962 Glycolysis
r_0964 Glycolysis
r_0886 Glycolysis
r_0911 Purine metabolism
r_0912 Purine metabolism
r_0570 Purine metabolism
r_0195 UDP–D–glucose metabolism
r_0079 Purine metabolism
r_0468 Proline biosynthesis
r_0152 Purine metabolism
r_0727 Threonine, methionine and cysteine biosynthesis
r_0151 Purine metabolism
r_0914 Purine metabolism
r_0908 Purine metabolism
r_0855 Purine metabolism
r_0536 Histidine metabolism
r_0554 Histidine metabolism
r_0910 Histidine metabolism
r_0225 Histidine metabolism
r_0563 Histidine metabolism
r_0007 Histidine metabolism
r_0537 Histidine metabolism
r_0538 Histidine metabolism
r_0115 Arginine biosynthesis
r_0818 Arginine biosynthesis
r_0208 Arginine biosynthesis
r_0759 Arginine biosynthesis
r_0207 Arginine biosynthesis
r_0118 Arginine biosynthesis
r_0816 Arginine biosynthesis
r_0758 Histidine metabolism
r_0816 Histidine metabolism
r_0225 Histidine metabolism
r_0564 Histidine metabolism
r_0564 Histidine metabolism
r_0079 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0086 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0039 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0040 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0065 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0279 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0042 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0997 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0005 Phenylalanine, tyrosine and tryptophan biosynthesis
r_0718 Pyruvate metabolism
r_0536 Histidine metabolism
r_0538 Histidine metabolism
r_0010 Histidine metabolism
r_0010 Histidine metabolism
r_0010

with any of the 46 correlated reactions (indicators, columns) is greater than 0.8 are shown. The co-correlation coefficient is defined as the product of the PCC between indicator and target fluxes and the PCC between target flux and indicator enzyme expression (see Supplementary Notes 2). This metric indicates how well the enzyme levels of an indicator reaction can be used to predict the target flux. The coefficients were row-wise normalized (i.e., by dividing the values with the row maximum) for visualization. The maximum co-correlation coefficient of each row is indicated next to the reaction name. Target reactions that belong to the set of 46 correlated reactions are highlighted in red on the y-axis.



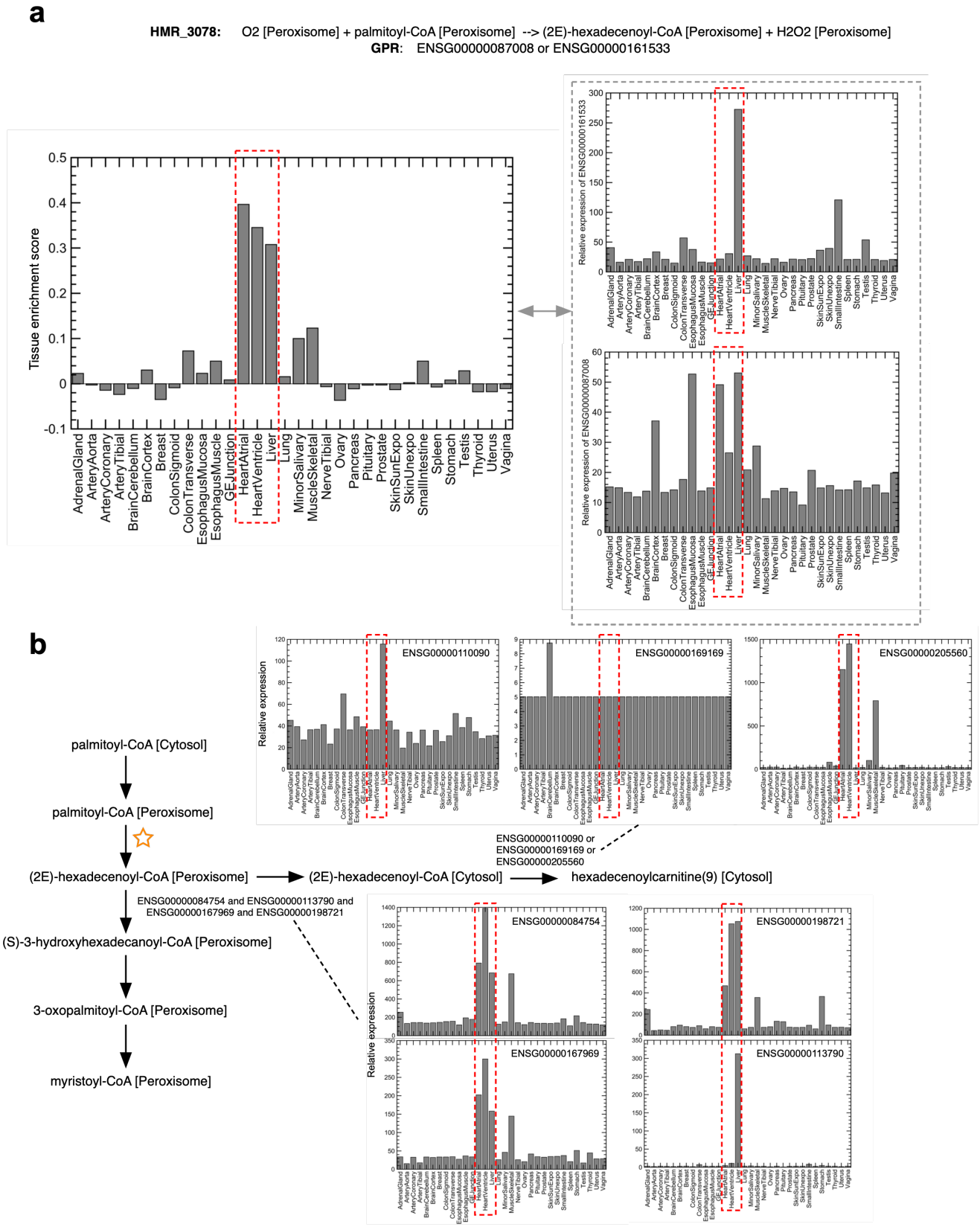
Extended Data Figure 7: Investigating local-pathway enzyme reach by predicting boundary fluxes. a, Heatmap of flux-rFP PCC with a titration of distance boundary for the drainage fluxes of biomass precursors in the yeast model, which serve as boundary

fluxes for these metabolites[4]. The relative (row-wise normalized) PCCs of 17 predicted (FDR < 0.05, PCC > 0, min-max rFP range > 0.2) reactions are shown in the heatmap. The layout of the heatmap is identical to Extended Data Fig. 4. **b**, Boxplots of metabolite tissue enrichment scores (ΔrFP) based on transporter flux potentials calculated by transporter expression only or by eFPA. The scores for the reference tissue-enriched metabolites (*enriched*) were compared with that of all transportable metabolites in Human 1 (*all*). P-values indicate the significance of difference between the median score of enriched metabolites and that of all metabolites (rank-sum test). Both the box statistics and p-values show a larger increase of *enriched* with respect to *all* in the case of eFPA. **c**,**d**, Comparison of eFPA results obtained using protein and mRNA levels. Common genes and all genes refer to the gene set that is shared in protein and RNA-seq data and the set of all genes covered in the RNA-seq data alone, respectively. Results were evaluated both by the overlap with the reference set (**c**) and by the levels of tissue-enrichment scores (**d**). **e**, Two examples showing that the higher coverage of RNA-seq data enabled useful predictions that were missed in the protein data modeling.



Extended Data Figure 8: Heatmap showing the relative tissue-enrichment of different subsystems in human 1 model. The levels of relative enrichment for indicated subsystems (rows) in each human tissue analyzed (columns) are shown. A reaction is

considered enriched in a tissue if its tissue enrichment score (ΔrFP) in that tissue is greater than 0.2. The number of tissue-enriched reactions for a subsystem and tissue is row-wise (across tissues) normalized by dividing with the row maximum to derive the relative enrichment. This heatmap is based on the eFPA predictions with the protein data.



Extended Data Figure 9: An example of eFPA-driven prediction of human tissue metabolism. a, ΔrFP prediction for palmitoyl-CoA peroxisomal oxidation. The relative protein levels of two associated enzymes are plotted on the right. Although liver

expression was consistently high in both enzymes, the heart expression was not. **b**, Enzyme levels of surrounding reactions for palmitoyl-CoA peroxisomal oxidation. The product of the target reaction is either further oxidized in peroxisome or exported into cytosol. The gene expressions in both routes show significant enrichment in heart.