Supporting Information (SI Appendix)

Framework and resource for more than 11,000 gene-transcript-protein-reaction associations (GeTPRA) in human metabolism

SI Appendix Materials and Methods

Standardization of Metabolite IDs with MNXM IDs Defined in the MNXref Namespace. Information on metabolic contents of the Recon 2Q was standardized using MNXM IDs defined in the MNXref namespace available at MetaNetX (1-3). This standardization was to facilitate the model refinement process described below. Each metabolite ID in the Recon 2Q was converted to MNXM ID accordingly. For metabolite IDs that were not converted to MNXM IDs, they were manually converted to MNXM IDs by comparing their compound structures and synonyms. In the final resulting SBML files, 97 metabolites were assigned with arbitrary IDs (i.e., "MNXMK_" followed by four digits) because they were not covered by the MNXref namespace (i.e., metabolite IDs not converted to MNXM IDs).

Refinement or Removal of Biochemically Inconsistent Reactions. Recon 2 was built upon metabolic genes and reactions collected from EHMN (4, 5), the first genome-scale human liver metabolic model HepatoNet1 (6), an acylcarnitine and fatty-acid oxidation model Ac-FAO (7), and a small intestinal enterocyte model hs_eIEC611 (8). Flux variability analysis (9) of the Recon 2Q identified blocked reactions coming from these four sources of metabolic reaction data. The EHMN caused the greatest number of blocked reactions in the Recon 2Q (1,070 reactions corresponding to 69.3% of all the identified blocked reactions). To refine the EHMN

reactions, following reactions were initially disregarded: 1) reactions having metabolite IDs not convertible to MNXM IDs; and 2) reactions without genes. The remaining reactions were newly assigned with compartments based on experimentally validated information from the HPA version 13 (http://www.proteinatlas.org) (10, 11). Compartment information in the HPA was converted to the compartment IDs available in Recon 2 (Dataset S15). These reactions were proton- and mass-balanced. Finally, the refined reactions that cannot be connected with existing reactions in the Recon 2Q were also disregarded.

For the reactions contributed by HepatoNet1, a total of 54 blocked reactions were identified, and among them, 30 blocked reactions were removed from the Recon 2Q because of the lack of biochemical evidences. The remaining 24 reactions were resolved by adding transporter and/or demand reactions for corresponding dead-end metabolites: e.g., transporter reactions of apolipoprotein B and E, both associated with protein degradation, which connect extracellular space with intracellular lysosome. The hs_eIEC611 model introduced three blocked reactions, NADK, NADtm, and FADtm, in the Recon 2Q, each of which functions as NAD kinase and mitochondrial transporter reactions of NAD and FAD, respectively. These reactions were resolved by adding demand reactions for mitochondrial NAD and FAD. Finally, the Ac-FAO model caused 18 blocked reactions in the Recon 2Q. These 18 blocked reactions constitute a linear pathway of fatty acid oxidation, and a blockage of this pathway was caused by incorrect localization of a reaction FAOXC102C101x, also involved in fatty acid oxidation; correcting compartmentalization of FAOXC102C101x resolved this issue.

After resolution of all the biochemically inconsistent reactions, 157 reactions disqualified to be included in generic human GEM (‘Trivial reaction’ in Dataset S2) and 23 redundant metabolic reactions present in the updated Recon 2Q were further removed. Novel metabolic contents of Recon 2.2 (12), the latest version of a human GEM, were also added to the updated Recon 2Q. All the changes made to the Recon 2Q are listed in Dataset S2.

Validation of Recon 2M.1. Recon 2M.1 was validated by predicting: 1) ATP production rates using varied carbon sources, 2) essential genes, and 3) glucose uptake rate, and lactate and ATP production rates under varied oxygen uptake rates. Recon 2, 2Q and 2.2 were also subjected to the same simulations for a comparison. For the first validation, ATP production rates were calculated under aerobic or anaerobic condition in a defined minimal medium containing one of 35 different carbon sources (Dataset S3 and S4). A purpose of this simulation set is to reproduce the simulation performance of previous Recon models (12) and evaluate consistency of Recon 2M.1. Uptake rate of a carbon source was constrained to 1 mmol/gDCW/h. For the aerobic condition, oxygen uptake rate was set to 1,000 mmol/gDCW/h. For the second validation, essentiality of genes was predicted by constraining the reaction flux value to zero if the reaction has the gene to be knocked out, and implementing flux balance analysis (FBA) with maximization of biomass generation rate as an objective function. This procedure was repeated for all the genes one by one. Information on gene essentiality was obtained from Wang, *et al.* (13) (Fig. 1*C*). A set of essential and non-essential genes were selected if the genes were essential or non-essential in all the four cancer cell lines. In total, 870 essential and 15,425 non-essential genes were obtained. Genes associated with blocked reactions were not considered because inclusion of such genes increases accuracy and sensitivity values. Accuracy, sensitivity and specificity were calculated as follows.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Sensitivity = \frac{TP}{TP + FN}$$

$$Specificity = \frac{TN}{TN + FP}$$

GPR associations of Recon 2M.1 were manually corrected in an iterative manner based on the results of gene essentiality analysis (see Dataset S2 for the correction details). For the last validation simulation set, oxygen uptake rate was varied from 0 to 1 mmol/gCDW/h in order to examine its effects on glucose uptake rate, and lactate and ATP production rates (Fig. 1*D*). Parsimonious FBA (pFBA) (14) was conducted with a pre-calculated maximal biomass generation rate at each oxygen uptake rate as a constraint for each Recon model so as to automatically generate glucose uptake rate. The last two validation simulation sets were conducted under the assumption of RPMI-1640 medium (Fig. 1 *C* and *D* and Dataset S3). RPMI-1640 medium is a frequently used medium for experiments with various human cell lines. Constraints used in this study for the RPMI-1640 medium (Dataset S3) are same as those used in previous *in silico* studies using Recon models (15-18). It should be noted that all these Recon models do not generate biomass *in silico* under the defined minimal medium (Dataset S3) due to requirements of essential nutrients (e.g., essential amino acids).

Conversion of GPR to TPR Associations in Recon 2M.1. As of May 2017, information on 20,244 genes and their 35,310 transcripts with "PRINCIPAL" (greater confidence) or "ALTERNATIVE" (lower confidence) tags was downloaded from the APPRIS database (19). All the 35,310 transcripts tagged with "PRINCIPAL" or "ALTERNATIVE" were considered as PTs, and used in subsequent analyses (Figs. 2-5). Entrez gene IDs described in the GPR associations were converted to their matching transcript IDs of Ensembl (20), RefSeq (21), and UCSC (22) databases, generating TPR associations.

Acquisition of 446 TCGA Personal RNA-Seq Data Across Ten Cancer Types and Statistical Comparative Expression Analyses. All the personal RNA-Seq data were downloaded from TCGA for the ten cancer types (Fig. 1*E*). We updated old UCSC transcript

IDs used in the TCGA RNA-Seq data to the latest version. Metabolic genes defined in the Recon 2M.1 were considered in the RNA-Seq data. Among a total of 1,682 metabolic genes involved in the Recon 2M.1, 85 genes were excluded because they were not available in the collected RNA-Seq data and/or their PTs were not properly defined in the APPRIS database (19). FEs for each metabolic gene were calculated using the remaining 1,597 genes and their corresponding 3,375 PTs defined in the APPRIS. The same sets of genes and their PTs were also subjected to the comparative expression analyses using a R package *edgeR* with recommended default parameter settings (23) (Fig. S4). Changes in (both total and principal) transcript levels between non-tumor and tumor samples with FDR corrected *P*-value < 0.05 were considered significant.

Reconstruction of 1,784 Personal GEMs Across Ten Cancer Types. Personal GEMs were reconstructed using four different types of data available in each TCGA personal RNA-Seq data, including expression levels of total transcripts from non-tumor and tumor samples and expression levels of PTs from non-tumor and tumor samples (Fig. 3*A*). Resulting personal GEMs were called T-GEMs or P-GEMs, depending on the use of data associated with total transcripts or PTs, respectively. Consequently, use of 446 personal RNA-Seq data resulted in the reconstruction of 1,784 personal GEMs. Recon 2M.1 was used as a template model, and integrated with 446 TCGA personal RNA-Seq data through the tINIT method. The tINIT method is an omics integration algorithm that attempts to identify a fully functional metabolic model that is most consistent with expression levels of genes and proteins having biochemical evidences (e.g., expression data) (24, 25). For the implementation of the tINIT method, we only considered metabolic genes in the Recon 2M.1. In this study, the tINIT was conducted with the same objective function as in the original study; however, for a weight score (w_i) for each gene that is used in the objective function, a newly rank-based weight function was

developed in order to minimize the effects of data outliers and sample variations on accuracies of the resulting context-specific GEMs (Fig. 3*A*). In this rank-based weight function, the rank of gene *i* (R_i) is determined based on its relative expression level rather than absolute expression value. Reactions with final negative weight scores from the rank-based weight function are likely to be removed from the context-specific GEM. Final weight score was assigned to each reaction through Boolean calculation of the relevant GPR associations (26). In case of multiple genes with *OR* relationship in the GPR association for a reaction, a final weight score for the reaction was obtained by summing all the weight scores assigned to each gene. For a reaction with genes having *AND* relationship, the minimum weight score among all the weight scores given to each gene was assigned to the reaction. In order to obtain biochemically consistent context-specific GEMs, parameter tests were conducted by changing $\boldsymbol{\varepsilon}$ ranging from 0.15 to 0.30 in order to find a robust threshold (Fig. S6). The parameter $\boldsymbol{\varepsilon}$ was set to be 0.25 in this study ($\boldsymbol{\varepsilon}$ = 0.25 in the equation, Fig. 3*A*; i.e., bottom 25%). Context-specific GEMs generated with the rank-based weight function appeared to have greater accuracy, sensitivity and specificity values than those obtained with an original weight function from Agren, *et al.* (24) (Fig. S7). Further details can be found in Figs. S5-S7.

As an additional input for the tINIT, a total of 162 essential metabolic reactions (Dataset S6) were obtained from the Recon 2M.1 through reaction knockout simulation. These essential metabolic reactions were selected if following 20 metabolic tasks were not satisfied: 1) cell growth rate, and production rates of 2) ten key intermediates (i.e., 2-oxoglutarate, 3-phospho-D-glycerate, D-erythrose 4-phosphate, D-fructose 6-phosphate, D-glucopyranose 6-phosphate, glyceraldehyde 3-phosphate, oxaloacetate, phosphoenolpyruvate, pyruvate and α-D-ribose 5-phosphate), 3) eight nucleotides (i.e., ATP, CTP, GTP, UTP, dATP, dCTP, dGTP and dTTP), and 4) ATP production rate (more than 10% decrease upon removal of a reaction). The first three simulation sets were conducted under the assumption RPMI-1640 medium,

while the last simulation set was conducted under aerobic or anaerobic growth in defined minimal media containing 21 different carbon sources including glucose and 20 amino acids (Dataset S3). These 162 metabolic reactions were set to be active in all the resulting personal GEMs because of their importance in the model functionality.

Simulation of Cancer Metabolism Using T-GEMs Built with Recon 2.2, 2M.1 and 2M.2. Non-tumor and tumor T-GEMs were first built with Recon 2.2, 2M.1 (a subset of the 1,784 personal GEMs; Fig. 3*A*) and 2M.2 as template models. These generic Recon models were integrated with non-tumor and tumor samples of 446 TCGA personal RNA-Seq data and by using tINIT method. Next, fluxes of non-tumor and tumor T-GEMs built with Recon 2.2, 2M.1 and 2M.2 were predicted by setting the expression data from non-tumor and tumor samples of the 446 TCGA personal RNA-Seq data as constraints and running the least absolute deviation method (27, 28). Here, expression values of genes or transcripts were mapped to reactions through GPR (for Recon 2.2) or TPR (for Recon 2M.1 and 2M.2) associations, respectively. For T-GEMs built with Recon 2M.1 and 2M.2, the GeTPRA dataset serves to specifically map transcripts to their corresponding reactions with correct compartments (Fig. 5). In case of T-GEMs built with Recon 2.2, gene information was mapped to all the relevant reactions as previously (24). Finally, the least absolute deviation minimizes the Euclidean distance between the mapped expression values and the reaction flux value, thereby generating intracellular flux distributions (27, 28).

Prediction of Anticancer Targets Using Tumor T-GEMs Built with Recon 2.2 and 2M.2. Metabolic fluxes of tumor T-GEMs derived from Recon 2.2 and 2M.2 obtained above (Fig. 5) were first compared with fluxes of the counterpart non-tumor T-GEMs. The metabolic reactions with fluxes predicted to be significantly increased in tumor T-GEMs in comparison

with non-tumor T-GEMs across the ten cancer types were considered as potential anticancer targets (Student's *t*-test, FDR corrected *P*-value < 0.01). These reactions were subjected to single knockouts to calculate relative growth rates of T-GEMs. The relative growth rate was calculated by dividing perturbed growth rate under each reaction knockout condition by normal growth rate without the knockout. The perturbed growth rate was calculated using 'minimization of metabolic adjustment' (MOMA) (29). Reactions were considered as anticancer targets if they generated growth rates less than 5% of the normal growth rates in more than 10% of T-GEMs for each cancer type. Reactions were also considered as anticancer targets that reduce the ratio of glycolytic to oxidative ATP flux (AFR) (30) if their single knockouts led to AFR values less than 50% of the normal AFR value without the knockout. Information on the approved drugs and their targets was obtained from DrugBank 5.0 (31). Cytoscape was used to generate networks that show relationships among reactions predicted as anticancer targets, their corresponding pathways, and approved drugs inhibiting the corresponding reactions (32).

Metabolic Simulations in General. All the metabolic simulations were conducted in Python environment with Gurobi Optimizer 6.0 and *GurobiPy* package (Gurobi Optimization, Inc., Houston, TX). Reading, writing, and manipulation of the COBRA-compliant SBML files were implemented using *COBRApy* (33). Constraints describing defined minimal medium and RPMI-1640 medium (34) are available in Dataset S3.

SI Appendix Figures

Fig. S1. Examples of standardized reaction and metabolite information in the resulting COBRA-compliant SBML file format. (*A*) Examples of a standardized reaction having gene-protein-reaction (GPR) associations described with gene IDs (Entrez) and transcript-protein-reaction (TPR) associations with transcript IDs (Ensembl, RefSeq and UCSC) in the SBML file. (*B*) An example of standardized metabolite information with MNXM IDs in the SBML file.

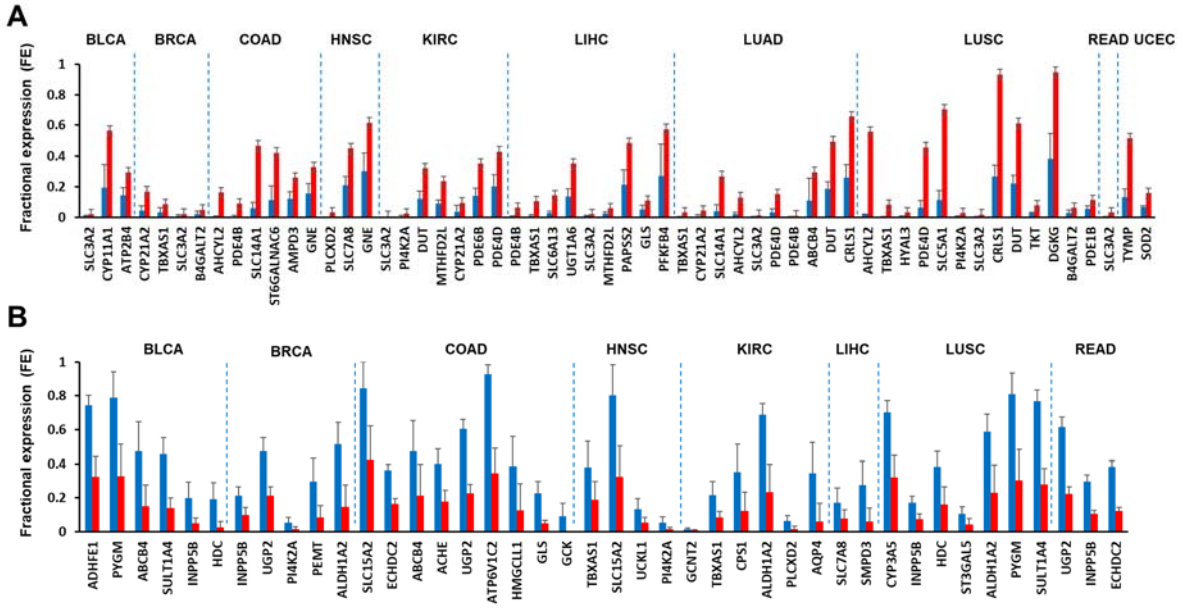


Fig. S2. Genes having differential fractional expressions (FEs) between non-tumor (blue bars) and tumor samples (red bars) of the 446 TCGA personal RNA-Seq data across ten cancer types. (*A*) Genes are shown that have FEs significantly increased in tumor samples compared with non-tumor samples across ten cancer types (FDR corrected *P*-value < 0.05 from Student's *t*-test; absolute changes > 2.0-fold). (*B*) Genes are shown that have FEs significantly decreased in tumor samples compared with non-tumor samples across ten cancer types (FDR corrected *P*-value < 0.05 from Student's *t*-test; absolute changes > 2.0-fold). Error bars mean ± s.d. Ten cancer type names are: bladder urothelial carcinoma (BLCA), breast invasive carcinoma (BRCA), colon adenocarcinoma (COAD), head and neck squamous cell carcinoma (HNSC), kidney renal clear cell carcinoma (KIRC), liver hepatocellular carcinoma (LIHC), lung adenocarcinoma (LUAD), lung squamous cell carcinoma (LUSC), rectum adenocarcinoma (READ) and uterine corpus endometrial carcinoma (UCEC). BLCA (*n*=19), BRCA (*n*=114), COAD (*n*=26), HNSC (*n*=43), KIRC (*n*=72), LIHC (*n*=50), LUAD (*n*=58), LUSC (*n*=51), READ (*n*=6), and UCEC (*n*=7).

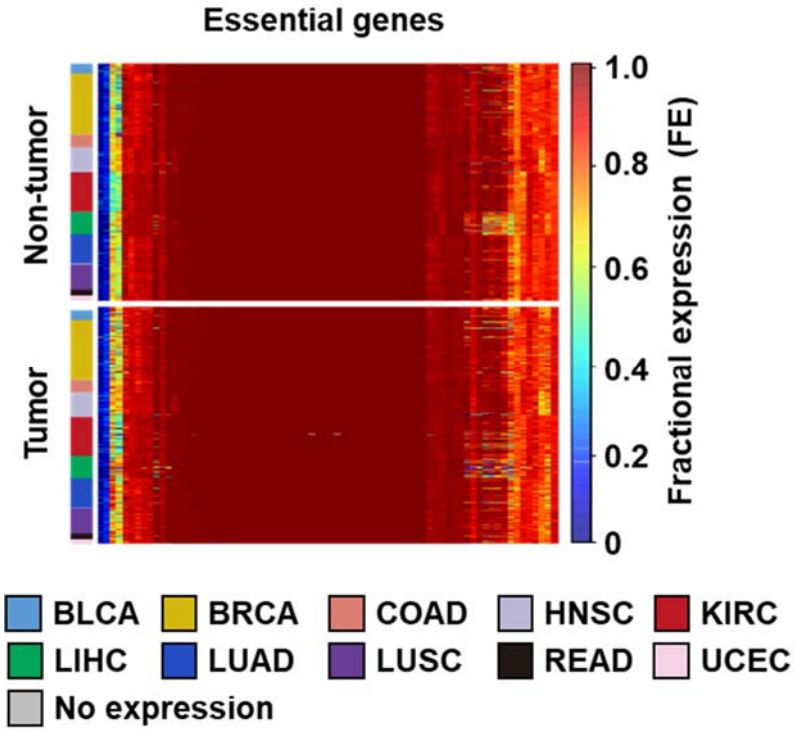


Fig. S3. A heat map showing the distribution of fractional expressions (FEs) for essential metabolic genes from non-tumor and tumor samples of the 446 TCGA personal RNA-Seq data across ten cancer types. Information on essential metabolic genes was obtained from Wang, *et al.* (13). FEs were calculated by dividing principal transcript levels (PTLs) by total transcript levels (TTLs) for each metabolic gene (Fig. 2*A*). Unexpressed genes are shown in grey.

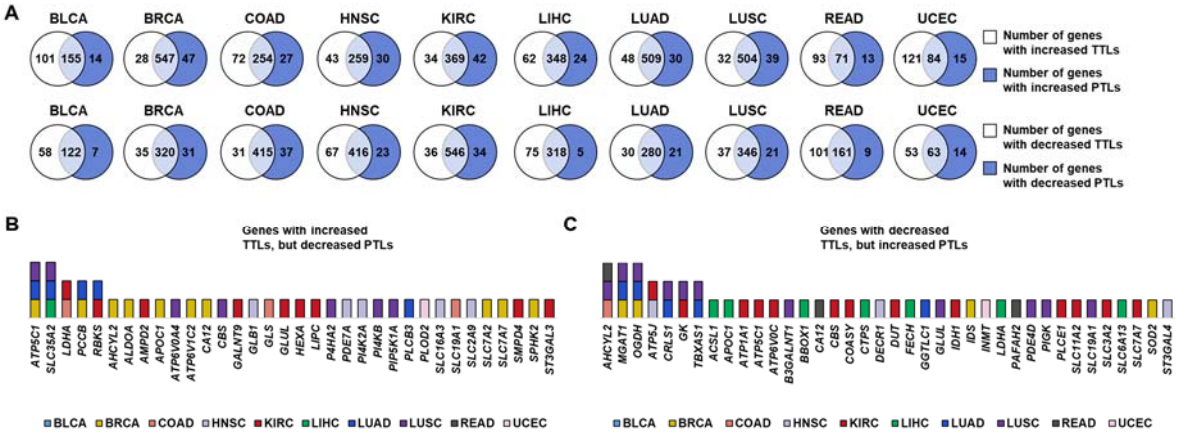


Fig. S4. Evidences of changes in some principal transcript levels (PTLs) decoupled from changes in total transcript levels (TTLs) in non-tumor and tumor samples of the 446 TCGA personal RNA-Seq data across ten cancer types. (*A*) Number of genes with significant changes in their TTLs and PTLs in non-tumor samples versus their matched tumor samples (FDR corrected *P*-value < 0.05). Number of genes with significantly up-regulated TTLs and PTLs are indicated in the white and blue circles, respectively in the first row. Numbers of genes with significantly down-regulated TTLs and PTLs are shown in the white and blue circles, respectively in the second row. Numbers in the overlapping regions of the white and blue circles indicate genes showing consistent changes in their TTLs and PTLs. (*B*) List of genes with increased TTLs, but decreased PTLs (in at least one of their PTs) in tumor samples, compared with non-tumor samples (FDR corrected *P*-value < 0.05). In total, 34 genes had significantly increased TTLs, but showed at least one of PTs decreased in tumor samples, compared to non-tumor samples, across the ten cancer types. (*C*) List of genes with decreased TTLs, but increased PTLs (in at least one of their PTs) in tumor samples, compared with non-tumor samples across the ten cancer types (FDR corrected *P*-value < 0.05). In total, 38 genes had significantly decreased TTLs, but having at least one of PTs increased in tumor samples, compared to non-tumor samples, across the ten cancer types. Ten cancer type names and number of samples for each cancer type are available in Fig. S2.

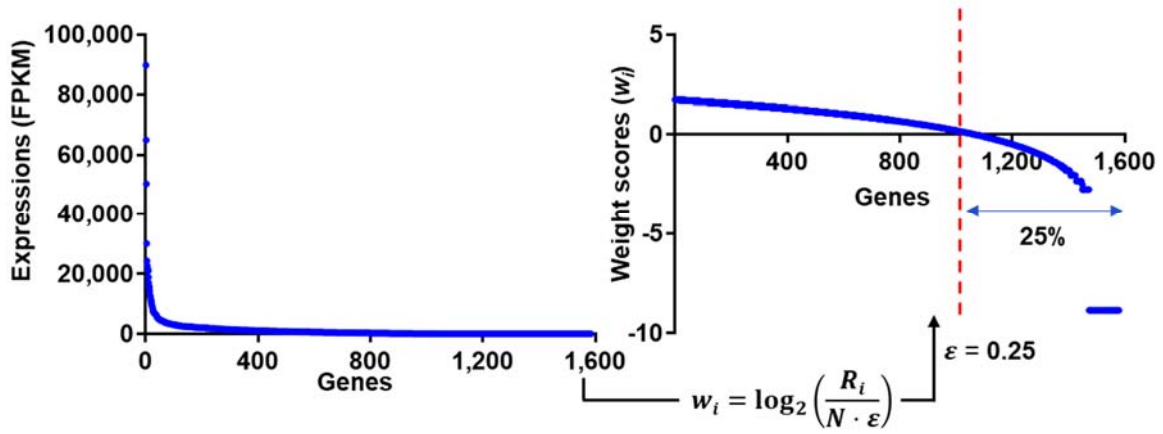


Fig. S5. Definition and characterization of the rank-based weight function. Rank-based weight function was coined to make sure that accuracy of a context-specific genome-scale metabolic model (GEM) built through integration with omics data is robust to the presence of data outliers and sample variations. All the genes are first sorted in descending order based on their FPKM values where FPKM stands for "fragments per kilobase of exon per million fragments mapped" (left scatterplot). Genes with "FPKM >= 1" and "FPKM < 1" are considered to be "expressed" and "not expressed", respectively. A gene *i* with the greatest FPKM value (highest rank) receives the value *N* (the total number of genes) for its R_i, and following genes with lower ranks receive reduced *N* values accordingly for their R_i values: e.g., $N - 1$ for R_i of the second ranked gene, $N - 2$ for R_i of the third ranked gene, etc. Distribution of weight scores (w_i) from the rank-based weight function is shown on the right scatterplot. Reactions with final negative weight scores are likely to be removed from the context-specific GEMs (e.g., tissue/cell-specific GEMs) based on Boolean calculations for their gene-protein-reaction (GPR) associations (Materials and Methods).

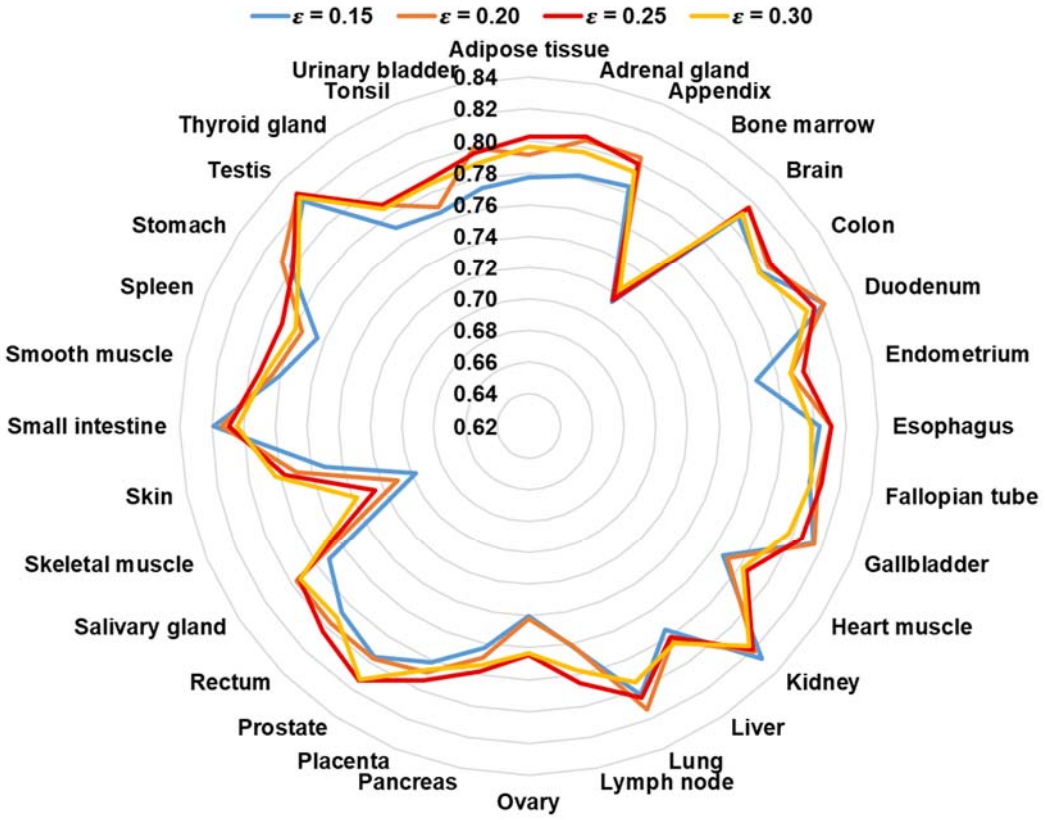


Fig. S6. Accuracy comparison of 32 tissue-specific GEMs built using the rank-based weight function with varied $\boldsymbol{\varepsilon}$ values. Model accuracies obtained at $\boldsymbol{\varepsilon}$ = 0.15, 0.20, 0.25 and 0.30 are shown herein. To obtain an optimal parameter $\boldsymbol{\varepsilon}$ that gives context-specific GEMs most consistent with the expression data, following steps took place. First, RNA-Seq data of 32 normal tissues from Uhlen, *et al.* (10) were used to reconstruct 32 tissue-specific GEMs using the rank-based weight function with different $\boldsymbol{\varepsilon}$ values ranging from 0.15 to 0.30; 32 tissue-specific GEMs were reconstructed for each different $\boldsymbol{\varepsilon}$ value. Recon 2M.1 having GPR associations with Entrez gene IDs was used as a template. Next, all the metabolic genes in the RNA-Seq data of 32 tissues were split into two groups based on their expression levels: FPKM >=1 or FPKM < 1. Finally, accuracies of the 32 tissue-specific GEMs were calculated based on the following equation: accuracy = $(TP + TN) / (TP + TN + FP + FN)$ where *TP* indicates the case of genes with FPKM >=1 and present in the GEMs, *FN* for genes with FPKM >=1,

but absent in the GEMs, *FP* for genes with FPKM < 1, but present in the GEMs, and *TN* for genes with FPKM < 1, and absent in the GEMs. As a result, $\boldsymbol{\varepsilon} = 0.25$ (i.e., 25%) gave the tissue-specific GEMs with the highest accuracies for 17 out of 32 tissues.

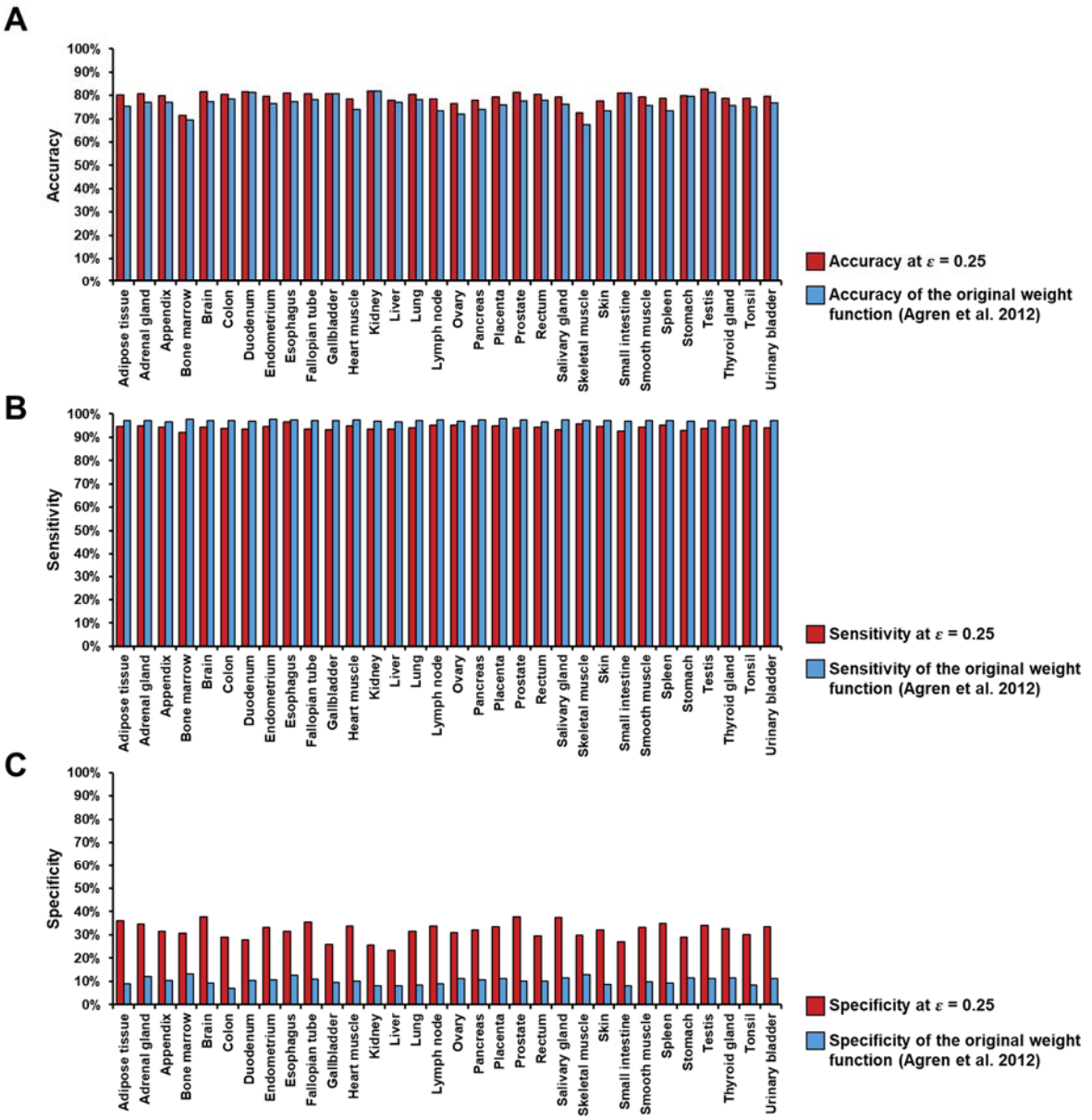


Fig. S7. Comparison of accuracy, sensitivity, and specificity values of the rank-based weight function with $\boldsymbol{\varepsilon}$ = 0.25 and a weight function reported by Agren, *et al.* (24). (*A*) The 32 tissue-specific GEMs reconstructed using the rank-based weight function with $\boldsymbol{\varepsilon}$ = 0.25 appeared to be slightly more accurate for 29 out of 32 tissues (i.e., greater value of the accuracy defined in Fig. S6) than those built with the original weight function. (*B*) The 32 tissue-specific GEMs built with the original weight function showed slightly greater sensitivity values compared with

those built with the rank-based weight function. (*C*) The 32 tissue-specific GEMs built with the rank-based weight function showed greater specificity values compared with those built with the original weight function.

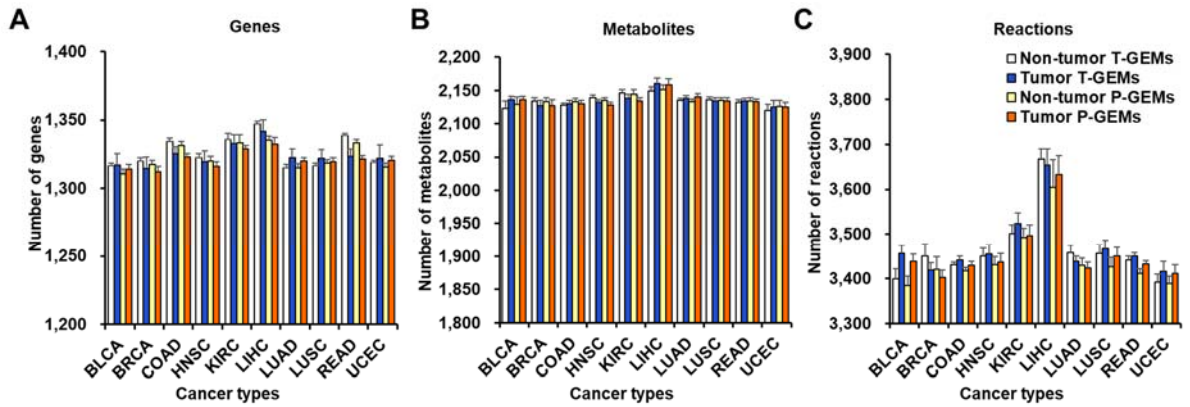
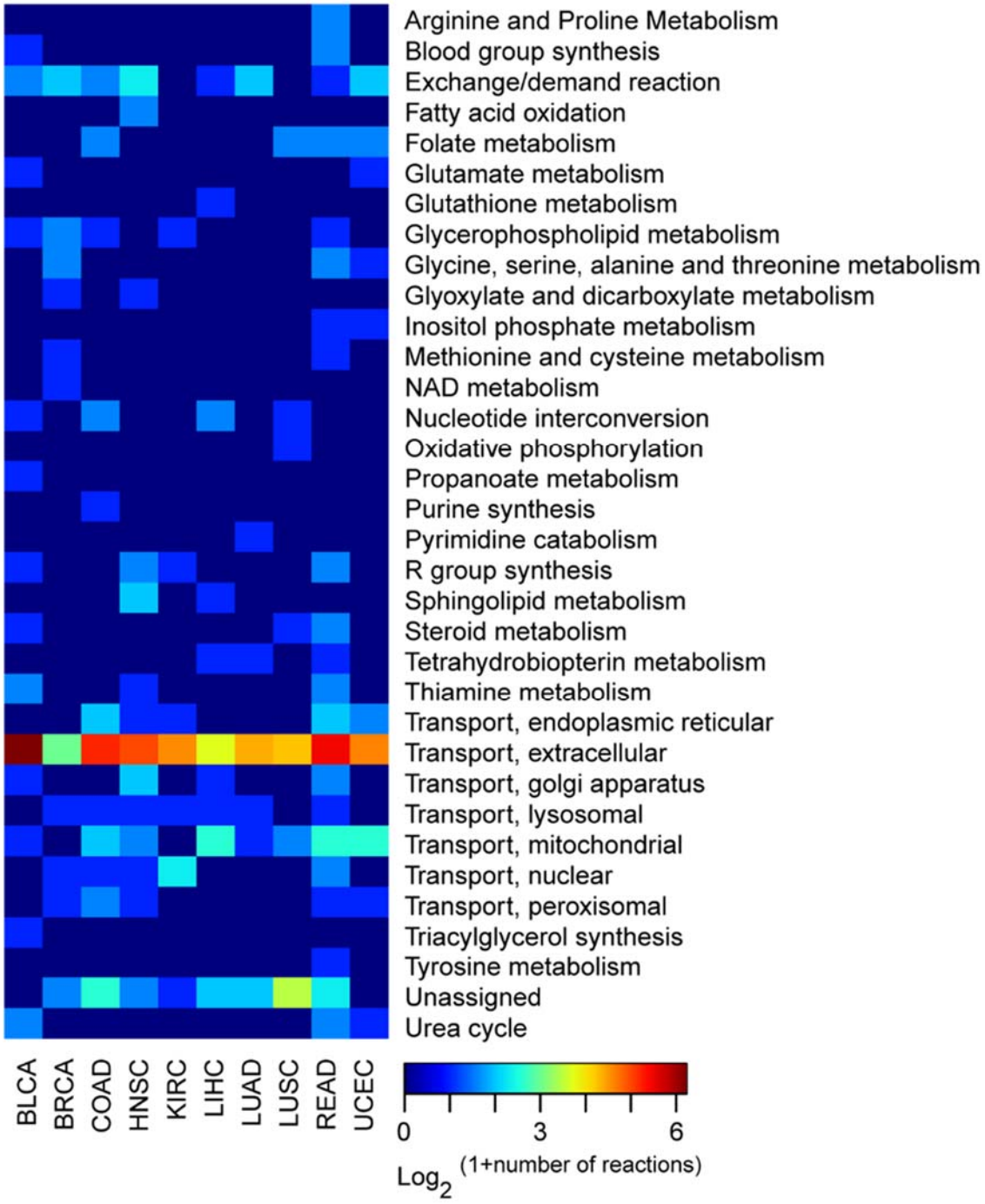
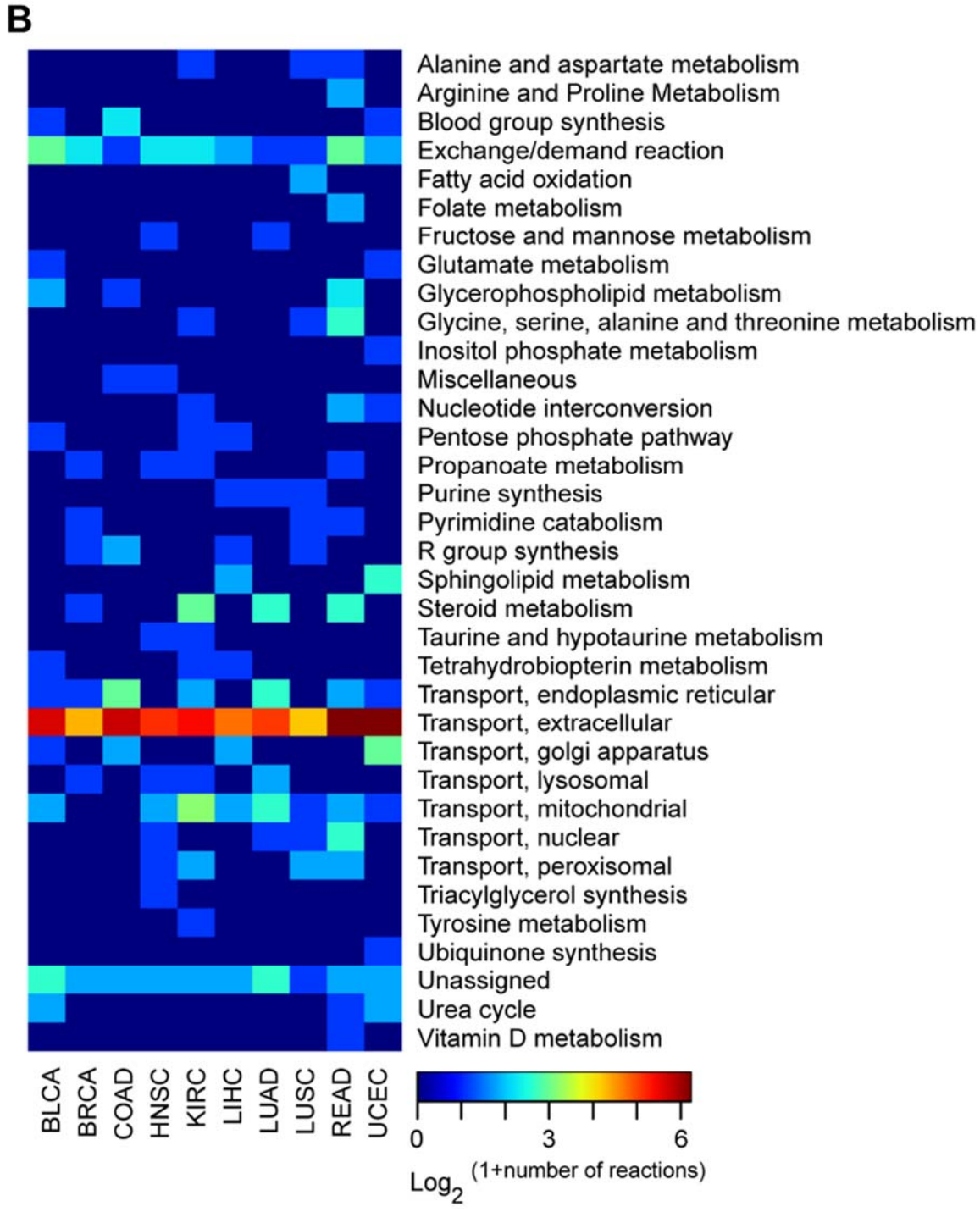


Fig. S8. Model statistics of the 446 reconstructed personal GEMs for each type: non-tumor T-GEMs, tumor T-GEMs, non-tumor P-GEMs, and tumor P-GEMs. (*A*) Number of genes, (*B*) metabolites, and (*C*) reactions in the four different types of 446 personal GEMs for each cancer type. Ten cancer type names and number of samples for each cancer type are available in Fig. S2. Error bars mean ± s.d.

A


Arginine and Proline Metabolism
Blood group synthesis
Exchange/demand reaction
Fatty acid oxidation
Folate metabolism
Glutamate metabolism
Glutathione metabolism
Glycerophospholipid metabolism
Glycine, serine, alanine and threonine metabolism
Glyoxylate and dicarboxylate metabolism
Inositol phosphate metabolism
Methionine and cysteine metabolism
NAD metabolism
Nucleotide interconversion
Oxidative phosphorylation
Propanoate metabolism
Purine synthesis
Pyrimidine catabolism
R group synthesis
Sphingolipid metabolism
Steroid metabolism
Tetrahydrobiopterin metabolism
Thiamine metabolism
Transport, endoplasmic reticular
Transport, extracellular
Transport, golgi apparatus
Transport, lysosomal
Transport, mitochondrial
Transport, nuclear
Transport, peroxisomal
Triacylglycerol synthesis
Tyrosine metabolism
Unassigned
Urea cycle
BLCA
BRCA
COAD
HNSC
KIRC
LIHC
LUAD
LUSC
READ
UCEC
0
3
6
Log2 (1+number of reactions)

(Continued)


B
Alanine and aspartate metabolism
Arginine and Proline Metabolism
Blood group synthesis
Exchange/demand reaction
Fatty acid oxidation
Folate metabolism
Fructose and mannose metabolism
Glutamate metabolism
Glycerophospholipid metabolism
Glycine, serine, alanine and threonine metabolism
Inositol phosphate metabolism
Miscellaneous
Nucleotide interconversion
Pentose phosphate pathway
Propanoate metabolism
Purine synthesis
Pyrimidine catabolism
R group synthesis
Sphingolipid metabolism
Steroid metabolism
Taurine and hypotaurine metabolism
Tetrahydrobiopterin metabolism
Transport, endoplasmic reticular
Transport, extracellular
Transport, golgi apparatus
Transport, lysosomal
Transport, mitochondrial
Transport, nuclear
Transport, peroxisomal
Triacylglycerol synthesis
Tyrosine metabolism
Ubiquinone synthesis
Unassigned
Urea cycle
Vitamin D metabolism
BLCA
BRCA
COAD
HNSC
KIRC
LIHC
LUAD
LUSC
READ
UCEC
0
3
6
Log2 (1+number of reactions)

(Continued)

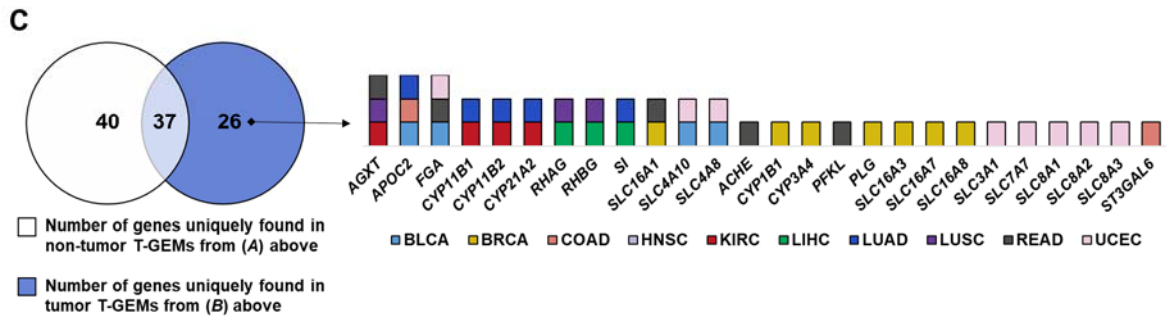


Fig. S9. Distribution of metabolic reactions (represented by the color of each cell in the heat map) uniquely present in GEMs built with (*A*) non-tumor and (*B*) tumor samples of total transcript-level data ('T-GEMs'). These metabolic reactions are absent in GEMs built with principal transcript levels ('P-GEMs'). *x*- and *y*-axes of each heat map represent the cancer type and metabolic pathway name, respectively. (*C*) Number of genes uniquely found in non-tumor and/or tumor T-GEMs from (*A*) and (*B*), respectively. Twenty six metabolic genes were exclusively present in tumor T-GEMs. Ten cancer type names and number of samples for each cancer type are available in Fig. S2.

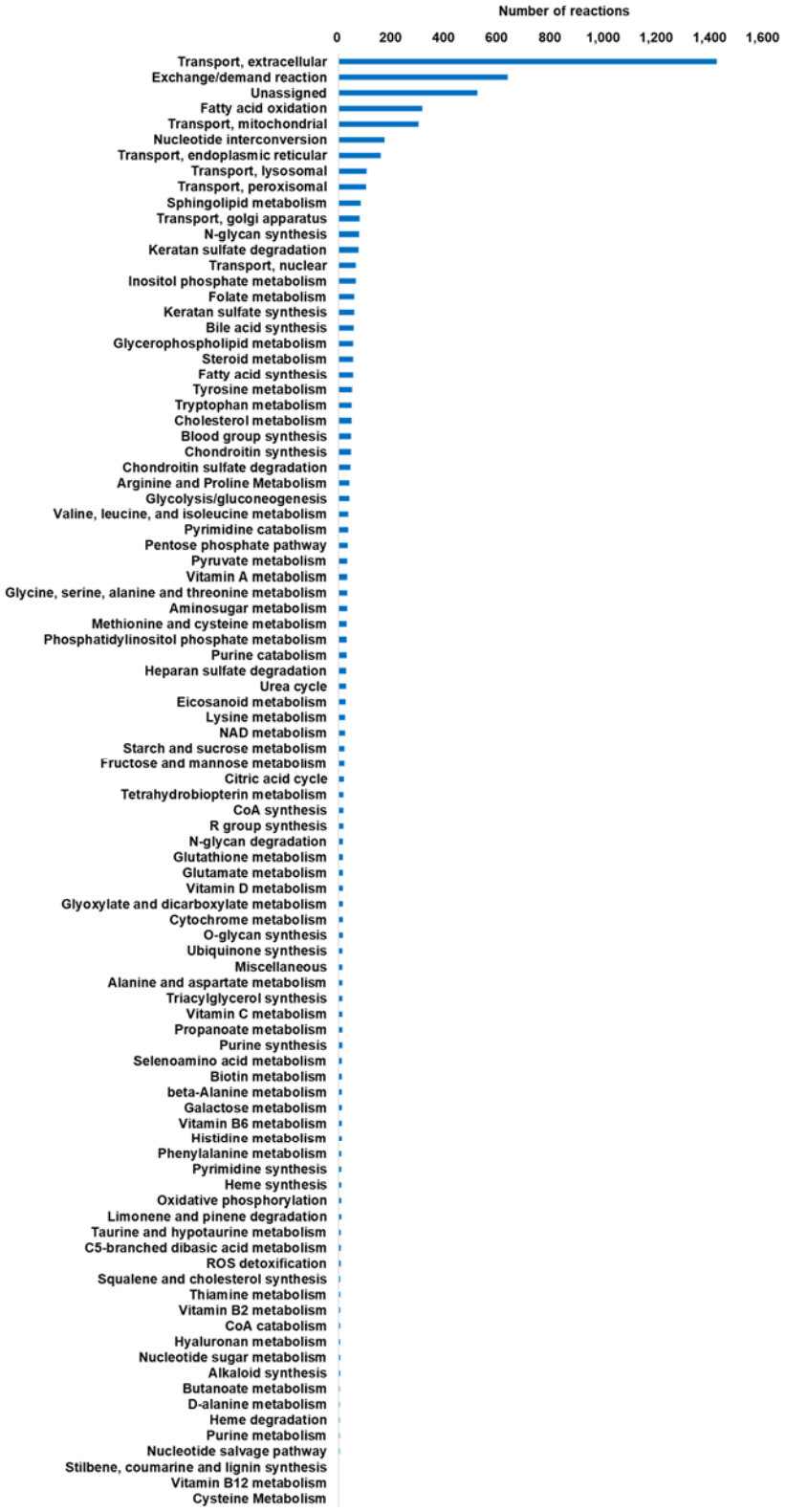
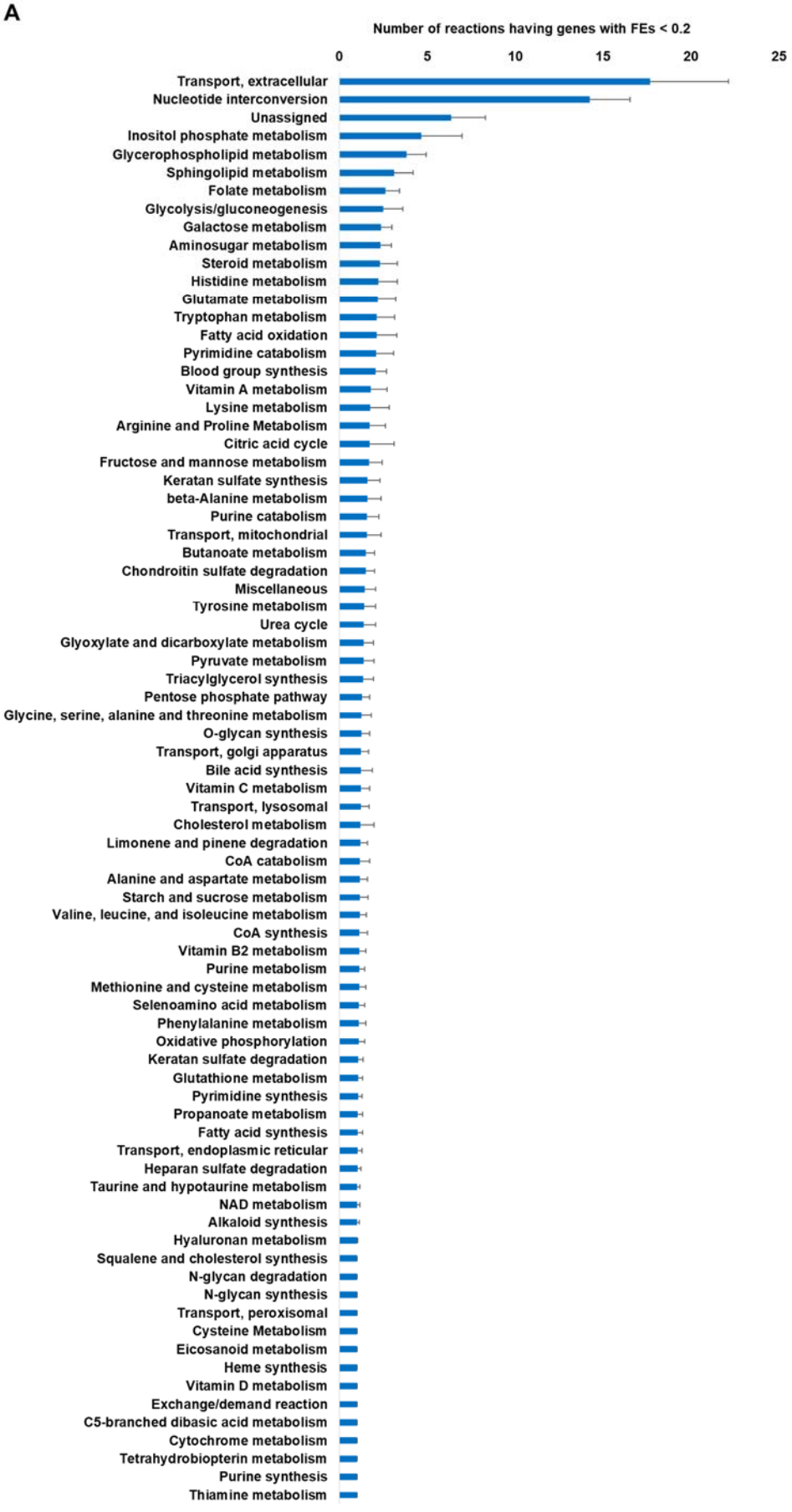


Fig. S10. Number of metabolic reactions in each pathway in Recon 2M.1.

A


Number of reactions having genes with FEs < 0.2
0
5
10
15
20
25
Transport, extracellular
Nucleotide interconversion
Unassigned
Inositol phosphate metabolism
Glycerophospholipid metabolism
Sphingolipid metabolism
Folate metabolism
Glycolysis/gluconeogenesis
Galactose metabolism
Aminosugar metabolism
Steroid metabolism
Histidine metabolism
Glutamate metabolism
Tryptophan metabolism
Fatty acid oxidation
Pyrimidine catabolism
Blood group synthesis
Vitamin A metabolism
Lysine metabolism
Arginine and Proline Metabolism
Citric acid cycle
Fructose and mannose metabolism
Keratan sulfate synthesis
beta-Alanine metabolism
Purine catabolism
Transport, mitochondrial
Butanoate metabolism
Chondroitin sulfate degradation
Miscellaneous
Tyrosine metabolism
Urea cycle
Glyoxylate and dicarboxylate metabolism
Pyruvate metabolism
Triacylglycerol synthesis
Pentose phosphate pathway
Glycine, serine, alanine and threonine metabolism
O-glycan synthesis
Transport, golgi apparatus
Bile acid synthesis
Vitamin C metabolism
Transport, lysosomal
Cholesterol metabolism
Limonene and pinene degradation
CoA catabolism
Alanine and aspartate metabolism
Starch and sucrose metabolism
Valine, leucine, and isoleucine metabolism
CoA synthesis
Vitamin B2 metabolism
Purine metabolism
Methionine and cysteine metabolism
Selenoamino acid metabolism
Phenylalanine metabolism
Oxidative phosphorylation
Keratan sulfate degradation
Glutathione metabolism
Pyrimidine synthesis
Propanoate metabolism
Fatty acid synthesis
Transport, endoplasmic reticular
Heparan sulfate degradation
Taurine and hypotaurine metabolism
NAD metabolism
Alkaloid synthesis
Hyaluronan metabolism
Squalene and cholesterol synthesis
N-glycan degradation
N-glycan synthesis
Transport, peroxisomal
Cysteine Metabolism
Eicosanoid metabolism
Heme synthesis
Vitamin D metabolism
Exchange/demand reaction
C5-branched dibasic acid metabolism
Cytochrome metabolism
Tetrahydrobiopterin metabolism
Purine synthesis
Thiamine metabolism

(Continued)

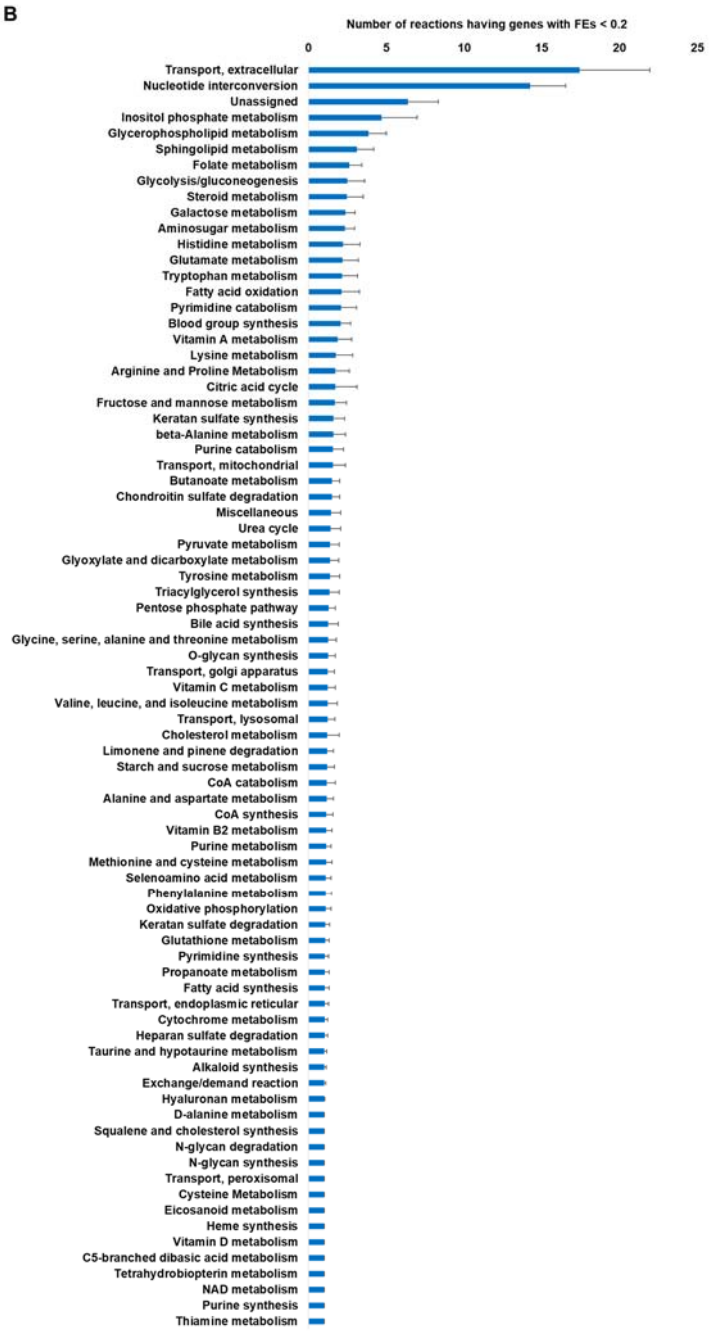


Fig. S11. Number of reactions having genes with FEs < 0.2 in each pathway of non-tumor (*A*) and tumor (*B*) samples from 446 TCGA personal RNA-Seq data. Error bars mean ± s.d.

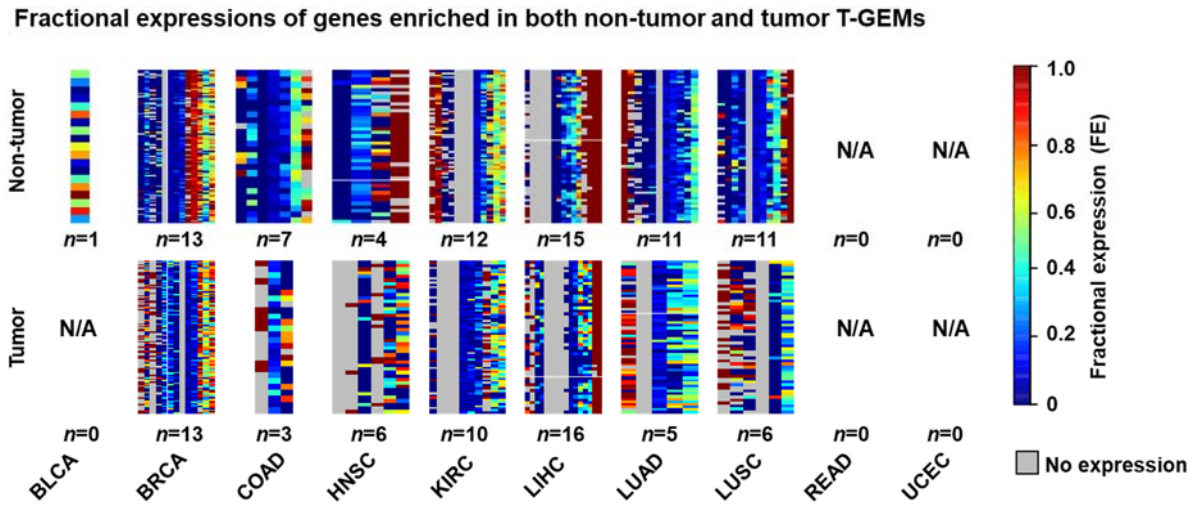
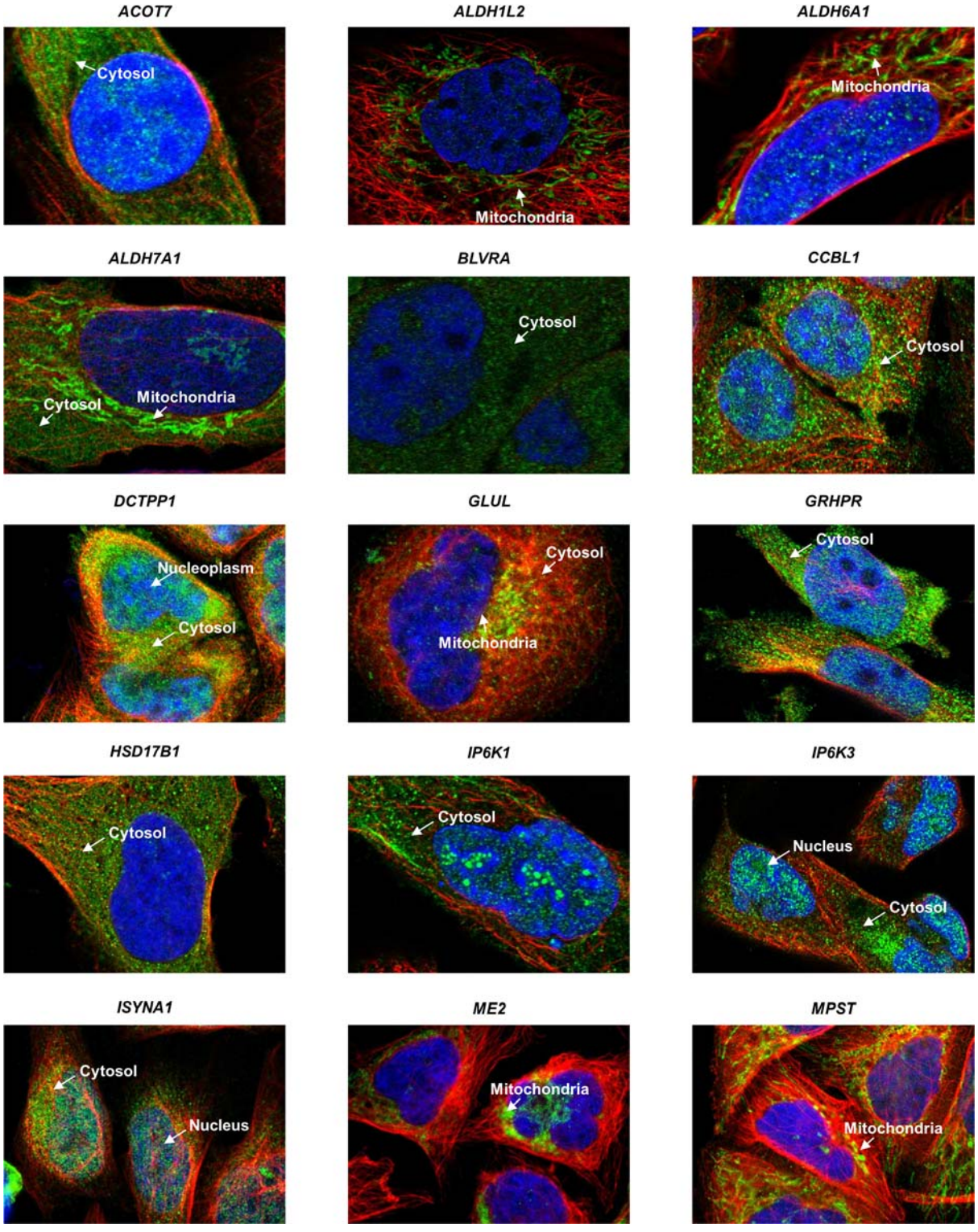


Fig. S12. Heat maps showing the distribution of fractional expressions (FEs) for metabolic genes enriched in non-tumor and/or tumor T-GEMs in comparison with P-GEMs for the ten cancer types. x- and y-axes of each heat map represent the number of enriched genes (indicated with 'n=' followed by number in the figure) and patients (n = 446) considered, respectively. Comparative gene enrichment analysis was conducted using Fisher's exact test (FDR corrected P-value < 0.05). Ten cancer type names are available in Fig. S2.


ACOT7
Cytosol
ALDH1L2
Mitochondria
ALDH6A1
Mitochondria
ALDH7A1
Cytosol
Mitochondria
BLVRA
Cytosol
CCBL1
Cytosol
DCTPP1
Nucleoplasm
Cytosol
GLUL
Cytosol
Mitochondria
GRHPR
Cytosol
HSD17B1
Cytosol
IP6K1
Cytosol
IP6K3
Nucleus
Cytosol
ISYNA1
Cytosol
Nucleus
ME2
Mitochondria
MPST
Mitochondria

(Continued)

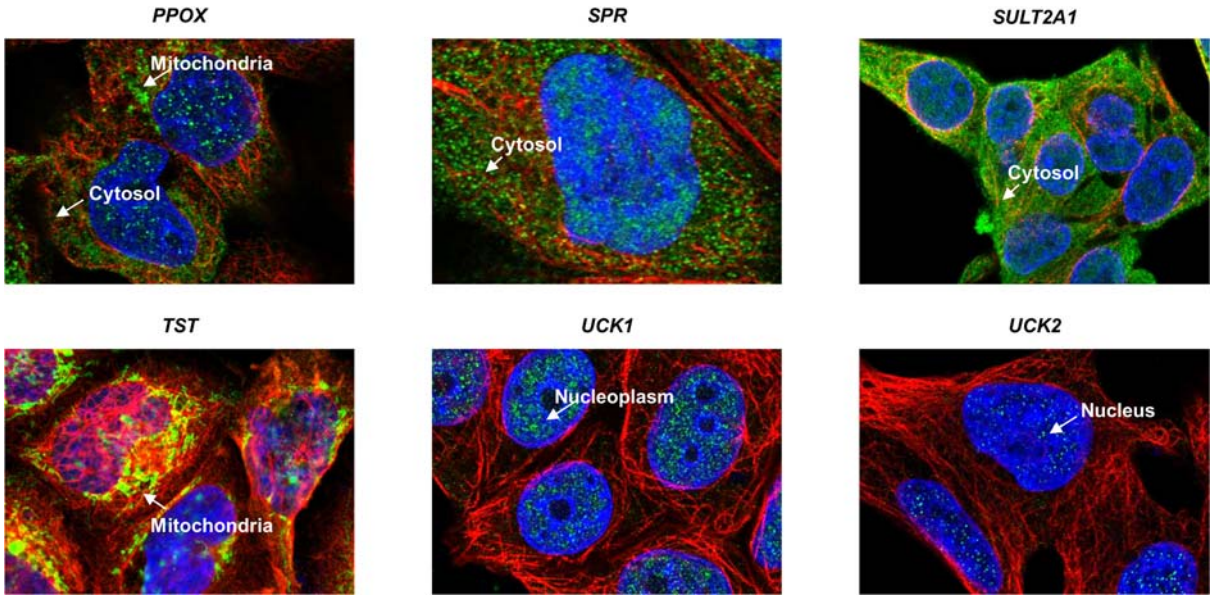


Fig. S13. Immunohistochemistry images for subcellular localizations of protein isoforms encoded by *ACOT7, ALDH1L2, ALDH6A1, ALDH7A1, BLVRA, CCBL1, DCTPP1, GLUL, GRHPR, HSD17B1, IP6K1, IP6K3, ISYNA1, ME2, MPST, PPOX, SPR, SULT2A1, TST, UCK1* and *UCK2*. All the shown immunohistochemistry images were obtained from the Human Protein Atlas (HPA; http://v16.proteinatlas.org) (10, 11). Green stains within each image indicate proteins of interest for the 21 genes. Antibody IDs of the shown immunohistochemistry images are: HPA025735 for *ACOT7*; HPA039481 for *ALDH1L2*; HPA029073 for *ALDH6A1*; HPA023296 for *ALDH7A1*; HPA019709 for *BLVRA*; HPA021176 for *CCBL1*; HPA002832 for *DCTPP1*; HPA007571 for *GLUL*; HPA022971 for *GRHPR*; HPA021032 for *HSD17B1*; HPA040825 for *IP6K1*; HPA053644 for *IP6K3*; HPA007931 for *ISYNA1*; HPA008247 for *ME2*; HPA001240 for *MPST*; HPA030123 for *PPOX*; HPA039505 for *SPR*; HPA041487 for *SULT2A1*; HPA003044 for *TST*; HPA050969 for *UCK1*; and HPA057128 for *UCK2*.

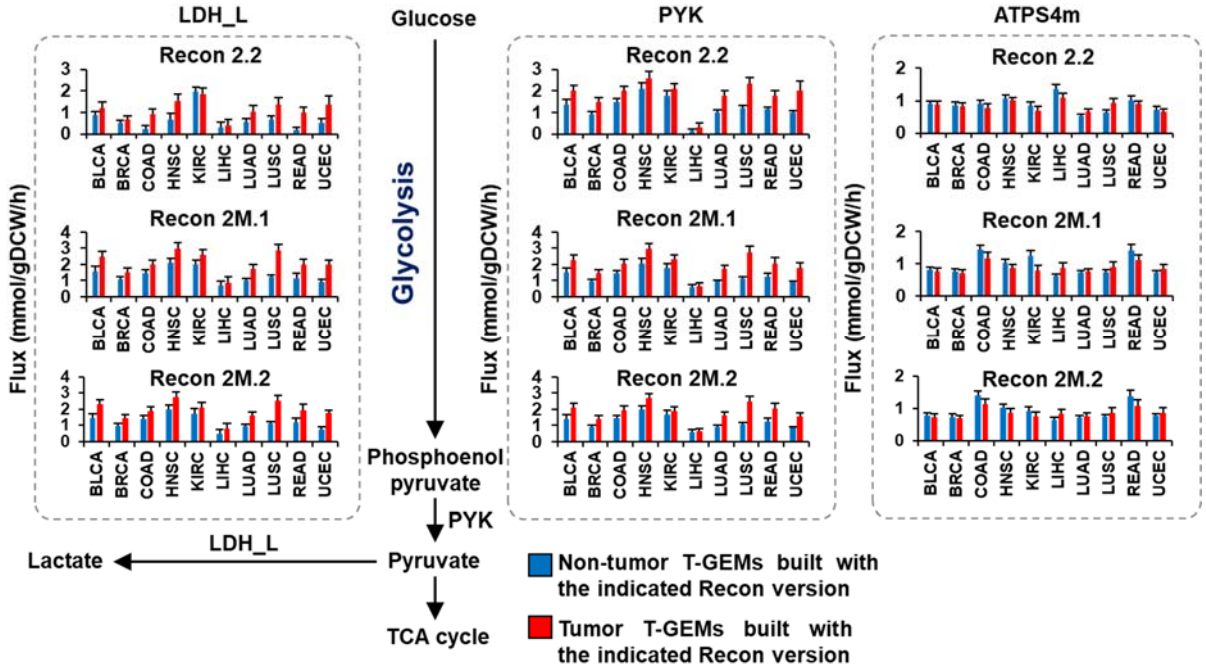


Fig. S14. Metabolic fluxes in central carbon metabolism predicted using non-tumor and tumor T-GEMs built with the three Recon models, Recon 2.2, Recon 2M.1 and Recon 2M.2. Error bars mean ± s.d.

References

1. Ganter M, Bernard T, Moretti S, Stelling J, & Pagni M (2013) MetaNetX.org: a website and repository for accessing, analysing and manipulating metabolic networks. *Bioinformatics* 29(6):815-816.
2. Bernard T, *et al.* (2014) Reconciliation of metabolites and biochemical reactions for metabolic networks. *Brief. Bioinform.* 15(1):123-135.
3. Moretti S, *et al.* (2016) MetaNetX/MNXref - reconciliation of metabolites and biochemical reactions to bring together genome-scale metabolic networks. *Nucleic Acids Res.* 44(D1):D523-526.
4. Hao T, Ma HW, Zhao XM, & Goryanin I (2010) Compartmentalization of the Edinburgh Human Metabolic Network. *BMC Bioinformatics* 11:393.
5. Ma H, *et al.* (2007) The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol. Syst. Biol.* 3:135.
6. Gille C, *et al.* (2010) HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol. Syst. Biol.* 6:411.
7. Sahoo S, Franzson L, Jonsson JJ, & Thiele I (2012) A compendium of inborn errors of metabolism mapped onto the human metabolic network. *Mol. Biosyst.* 8(10):2545-2558.
8. Sahoo S & Thiele I (2013) Predicting the impact of diet and enzymopathies on human small intestinal epithelial cells. *Hum. Mol. Genet.* 22(13):2705-2722.
9. Mahadevan R & Schilling CH (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng.* 5(4):264-276.
10. Uhlen M, *et al.* (2015) Proteomics. Tissue-based map of the human proteome. *Science* 347(6220):1260419.
11. Thul PJ, *et al.* (2017) A subcellular map of the human proteome. *Science* 356(6340).
12. Swainston N, *et al.* (2016) Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* 12:109.
13. Wang T, *et al.* (2015) Identification and characterization of essential genes in the human genome. *Science* 350(6264):1096-1101.
14. Lewis NE, *et al.* (2010) Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol. Syst. Biol.* 6:390.
15. Wang Y, Eddy JA, & Price ND (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* 6:153.
16. Nam H, *et al.* (2014) A systems approach to predict oncometabolites via context-specific genome-scale metabolic networks. *PLoS Comput. Biol.* 10(9):e1003837.
17. Shlomi T, Benyamini T, Gottlieb E, Sharan R, & Ruppin E (2011) Genome-scale metabolic modeling elucidates the role of proliferative adaptation in causing the Warburg effect. *PLoS Comput. Biol.* 7(3):e1002018.
18. Folger O, *et al.* (2011) Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* 7:501.
19. Rodriguez JM, *et al.* (2013) APPRIS: annotation of principal and alternative splice isoforms. *Nucleic Acids Res.* 41(Database issue):D110-117.
20. Cunningham F, *et al.* (2015) Ensembl 2015. *Nucleic Acids Res.* 43(Database issue):D662-669.
21. Pruitt KD, *et al.* (2014) RefSeq: an update on mammalian reference sequences. *Nucleic Acids Res.* 42(Database issue):D756-763.
22. Karolchik D, *et al.* (2003) The UCSC Genome Browser Database. *Nucleic Acids Res.* 31(1):51-54.

23. Robinson MD, McCarthy DJ, & Smyth GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26(1):139-140.
24. Agren R, *et al.* (2012) Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* 8(5):e1002518.
25. Agren R, *et al.* (2014) Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol. Syst. Biol.* 10(3):721.
26. Colijn C, *et al.* (2009) Interpreting expression data with metabolic flux models: predicting Mycobacterium tuberculosis mycolic acid production. *PLoS Comput. Biol.* 5(8):e1000489.
27. Lee D, *et al.* (2012) Improving metabolic flux predictions using absolute gene expression data. *BMC Syst. Biol.* 6:73.
28. Kim HU, Kim TY, & Lee SY (2011) Framework for network modularization and Bayesian network analysis to investigate the perturbed metabolic network. *BMC Syst. Biol.* 5 Suppl 2:S14.
29. Segre D, Vitkup D, & Church GM (2002) Analysis of optimality in natural and perturbed metabolic networks. *Proc. Natl. Acad. Sci. USA* 99(23):15112-15117.
30. Yizhak K, *et al.* (2014) A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration. *Mol. Syst. Biol.* 10:744.
31. Law V, *et al.* (2014) DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res.* 42(Database issue):D1091-1097.
32. Shannon P, *et al.* (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13(11):2498-2504.
33. Ebrahim A, Lerman JA, Palsson BO, & Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst. Biol.* 7:74.
34. Wang YL, Eddy JA, & Price ND (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* 6.