Supplementary information

Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data

Johan Gustafsson[1,2], Mihail Anton[3], Fariba Roshanzamir[1], Rebecka Jörnsten[4], Eduard J. Kerkhoven[1], Jonathan L. Robinson[1,5], Jens Nielsen[1,2,5,*]

[1] Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg, Sweden.

[2] Wallenberg Center for Protein Research, Chalmers University of Technology, Gothenburg, Sweden.

[3] Department of Biology and Biological Engineering, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Chalmers University of Technology, Sweden.

[4] Mathematical Sciences, University of Gothenburg and Chalmers University of Technology, Gothenburg, Sweden

[5] BioInnovation Institute, Copenhagen, Denmark

* Corresponding author

E-mail: nielsenj@chalmers.se

Supplementary Figures

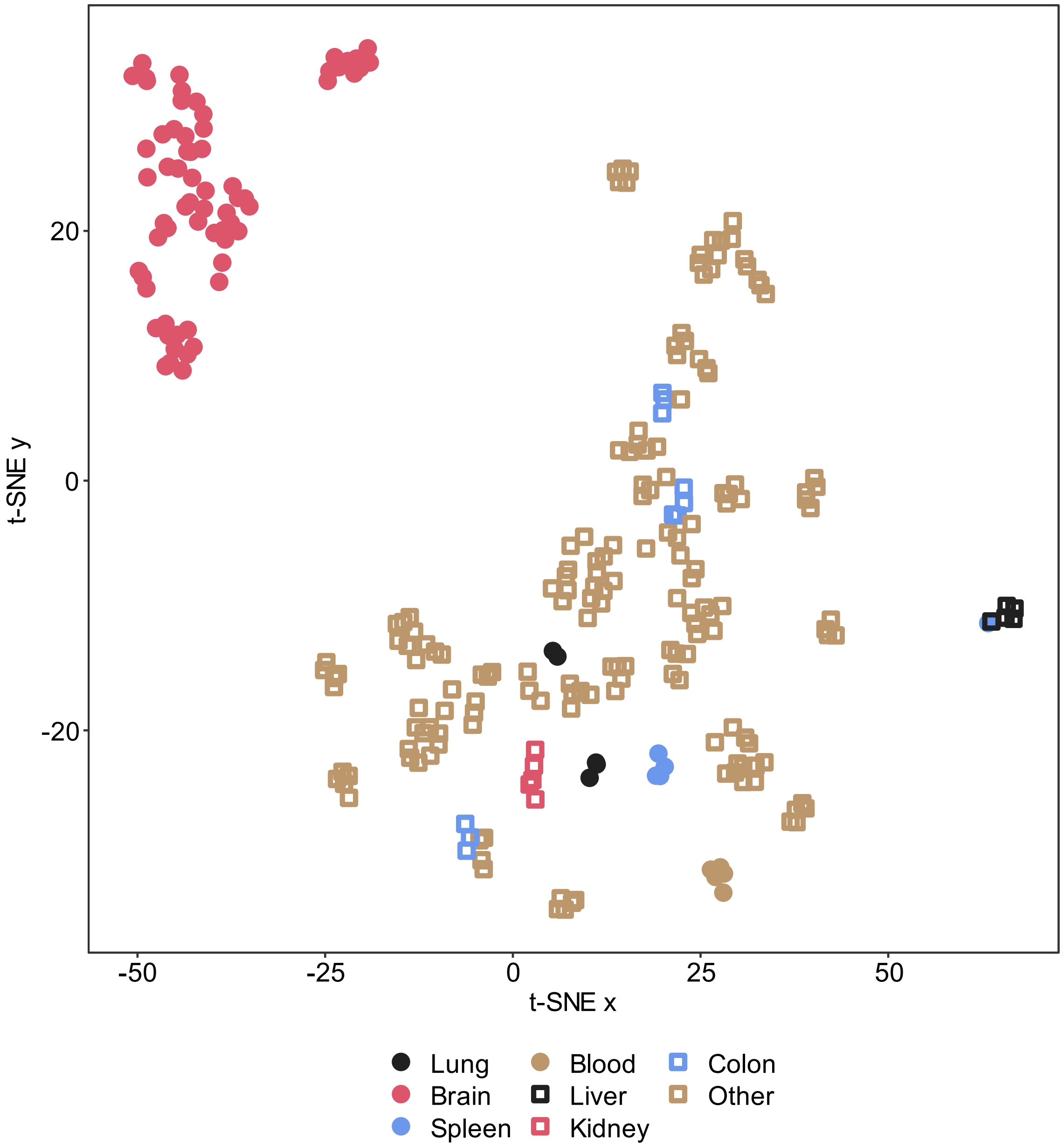


Fig S1: Grouping per tissue for the previous version of tINIT. *Structural comparison of genome-scale metabolic models generated by tINIT from RNA-Seq profiles from GTEx (bulk data), 5 samples per tissue, displayed as a t-SNE projection.*

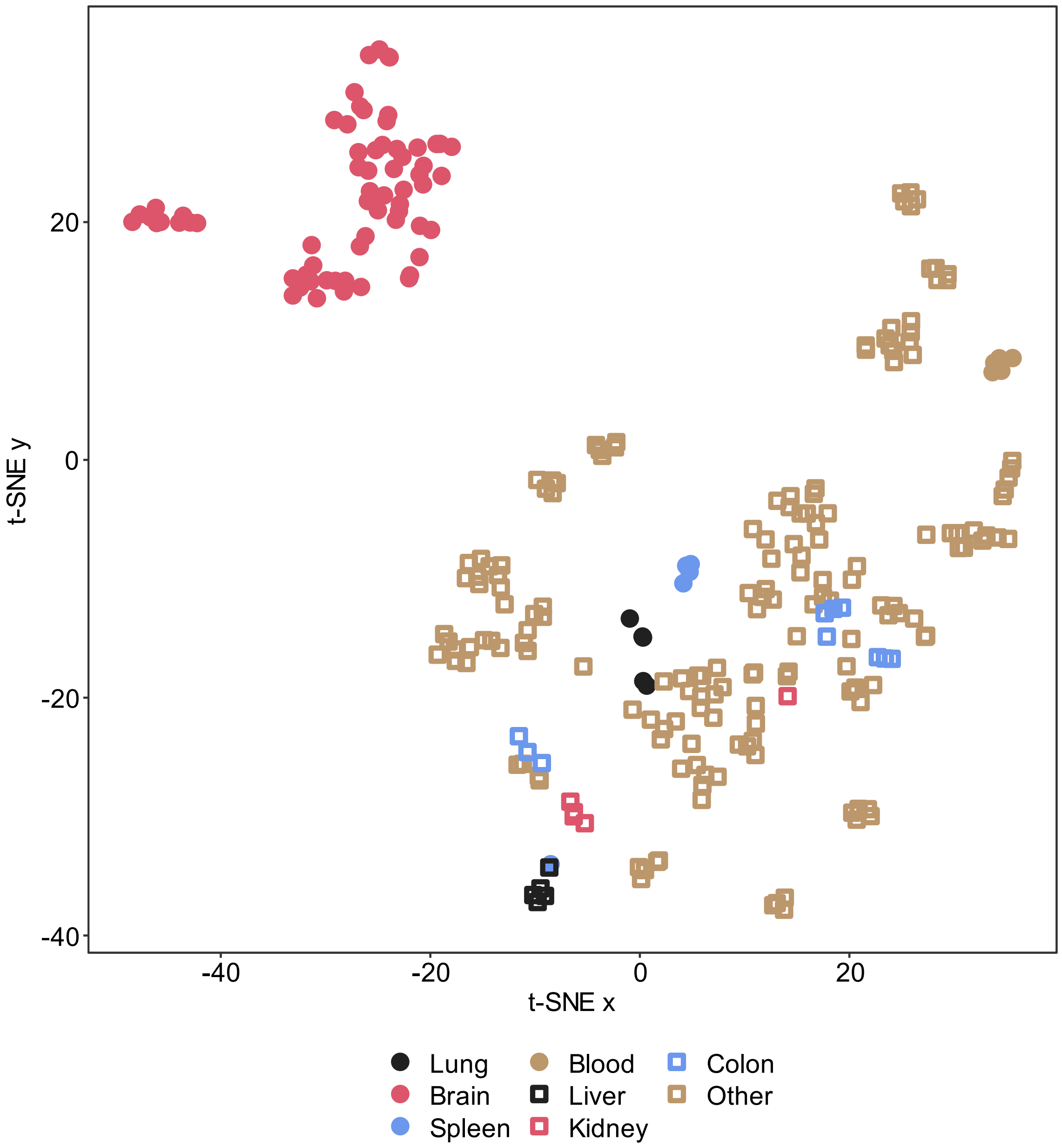


Fig S2: Grouping per tissue for ftINIT, the new faster version of tINIT. *Structural comparison of genome-scale metabolic models generated by ftINIT from RNA-Seq profiles from GTEx (bulk data), 5 samples per tissue, displayed as a t-SNE projection. The grouping is very similar to that of the old tINIT, but with slightly more spread for kidney and lung samples.*

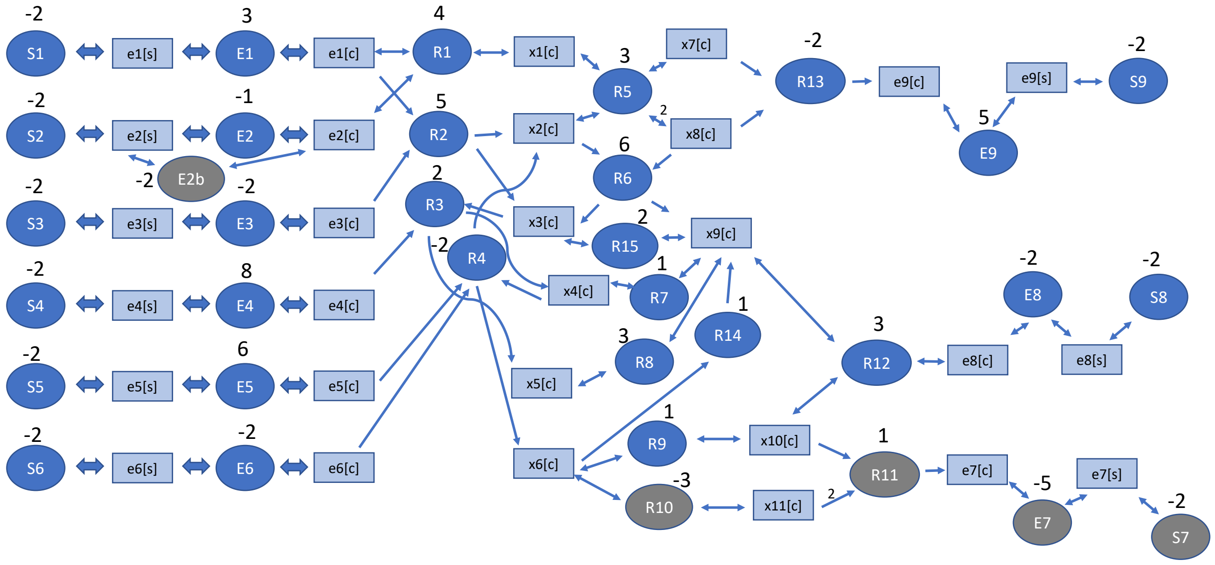


Fig S3: Model quality evaluation – ground truth. *tINIT (previous version) run on a small test model without any simplification flags (see Note S2 for details). This way of running tINIT is not practically useful for large models such as Human1 since the execution time is very long. We use this resulting model as ground truth to evaluate the performance of ftINIT and tINIT with simplification flags. Rectangles represent metabolites, ellipses represent reactions. Blue reactions were included and gray reactions were excluded in the final model. The numbers indicate the reaction scores used by the optimization, where the sum of the scores of the included reactions are maximized in the algorithm.*

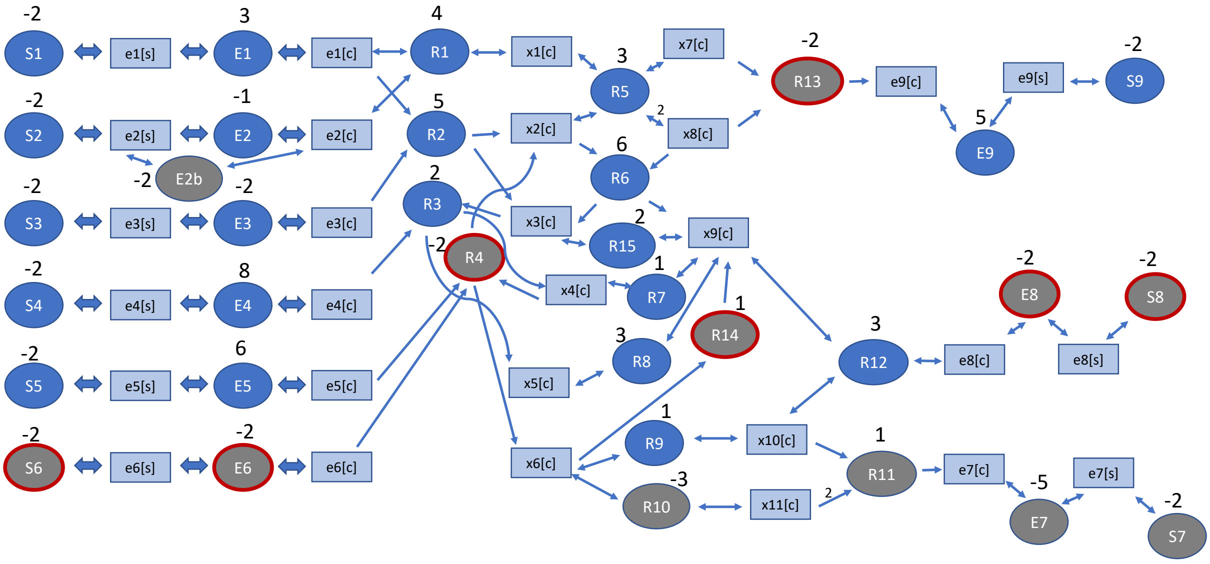


Fig S4: Model quality evaluation – tINIT with simplifications. *tINIT (previous version) run on a small test model with standard simplification flags, which allows for secretion of all metabolites and allows for fluxes in both directions through reversible reactions (see Note S2 for details). Rectangles represent metabolites, ellipses represent reactions. Blue reactions were included and gray reactions were excluded in the final model. Red ellipses around reactions indicate discrepancies between the resulting model and ground truth, emphasizing that the simplification algorithm creates unwanted gaps in the model. The numbers indicate the reaction scores used by the optimization, where the sum of the scores of the included reactions are maximized in the algorithm.*

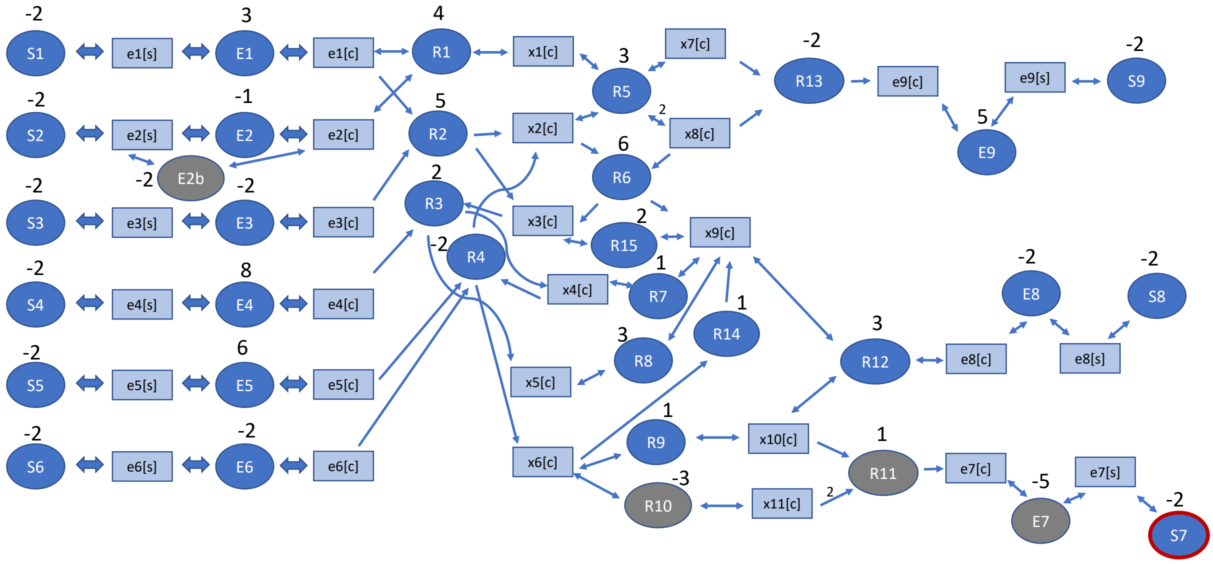


Fig S5: Model quality evaluation – ftINIT. *ftINIT (here including step 2) run on a small test model. Rectangles represent metabolites, ellipses represent reactions. Blue reactions were included and gray reactions were excluded in the final model. Red ellipses around reactions indicate discrepancies between the resulting model and ground truth. The resulting model shows good agreement with ground truth, differing only in an exchange reaction, which is no concern for any modeling results because it is disconnected from the rest of the network – to keep the exchange reactions is a choice to simplify constraining of the metabolite uptake rates the same way for all models. The numbers indicate the reaction scores used by the optimization, where the sum of the scores of the included reactions are maximized in the algorithm.*

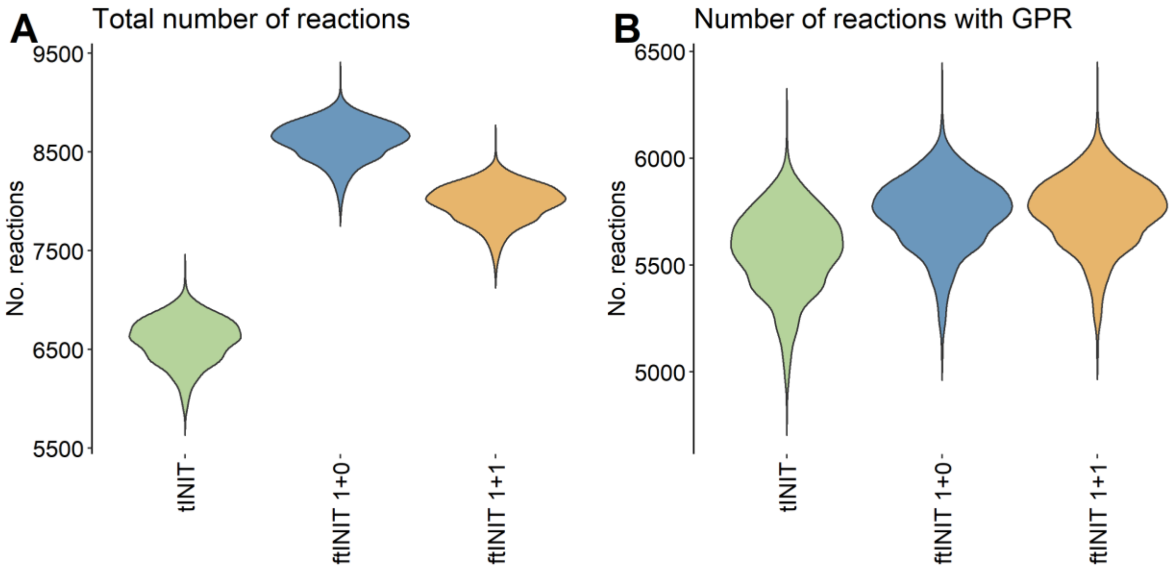


Fig. S6: Model statistics for tINIT and ftINIT. *The figure shows data for 891 models generated using tINIT and ftINIT in "1+0" and "1+1" mode. A. Total number of reactions for the models. B. Total number of reactions that have a valid gene association (GPR). ftINIT gives less penalty to including reactions without GPRs, which yields a higher inclusion of reactions with GPRs that are dependent on reactions without GPRs, yielding a slightly higher total number of reactions with GPRs. The large difference is for reactions without GPRs though, where two factors are important; 1) tINIT leaves gaps in the model, which ftINIT does not, and 2) ftINIT includes more reactions with GPRs that are dependent on reactions without GPRs, and these reactions without GPRs are also included in the model.*

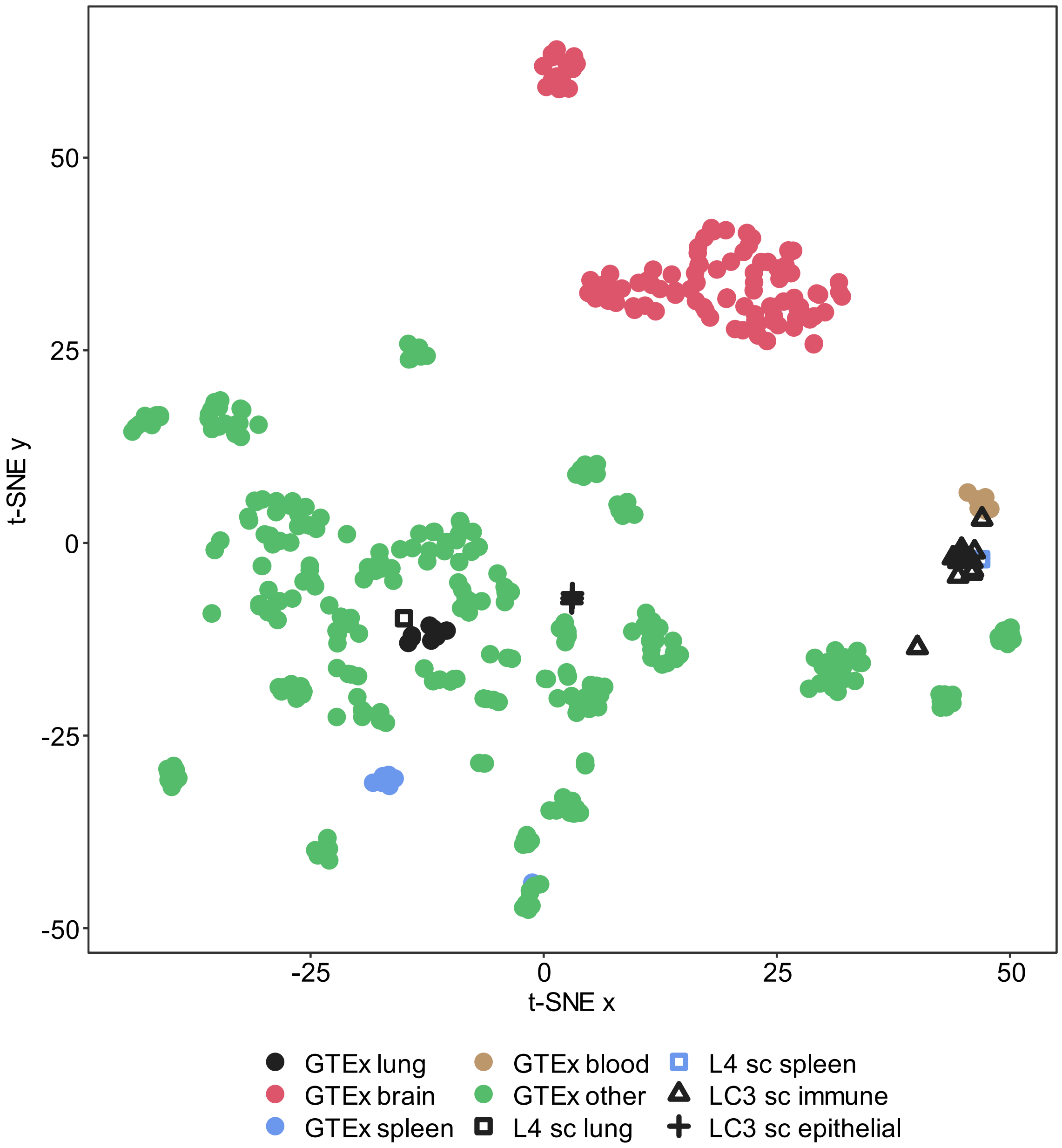


Fig. S7: Effect of TMM normalization on structural comparison. *Structural comparison of genome-scale metabolic models generated by ftINIT from RNA-Seq profiles from GTEx (bulk data) and several single-cell RNA-Seq datasets, displayed as a t-SNE projection. The results are similar to those based on TPM/CPM-normalized data. See Fig. 2D in the main text for details about the datasets used.*

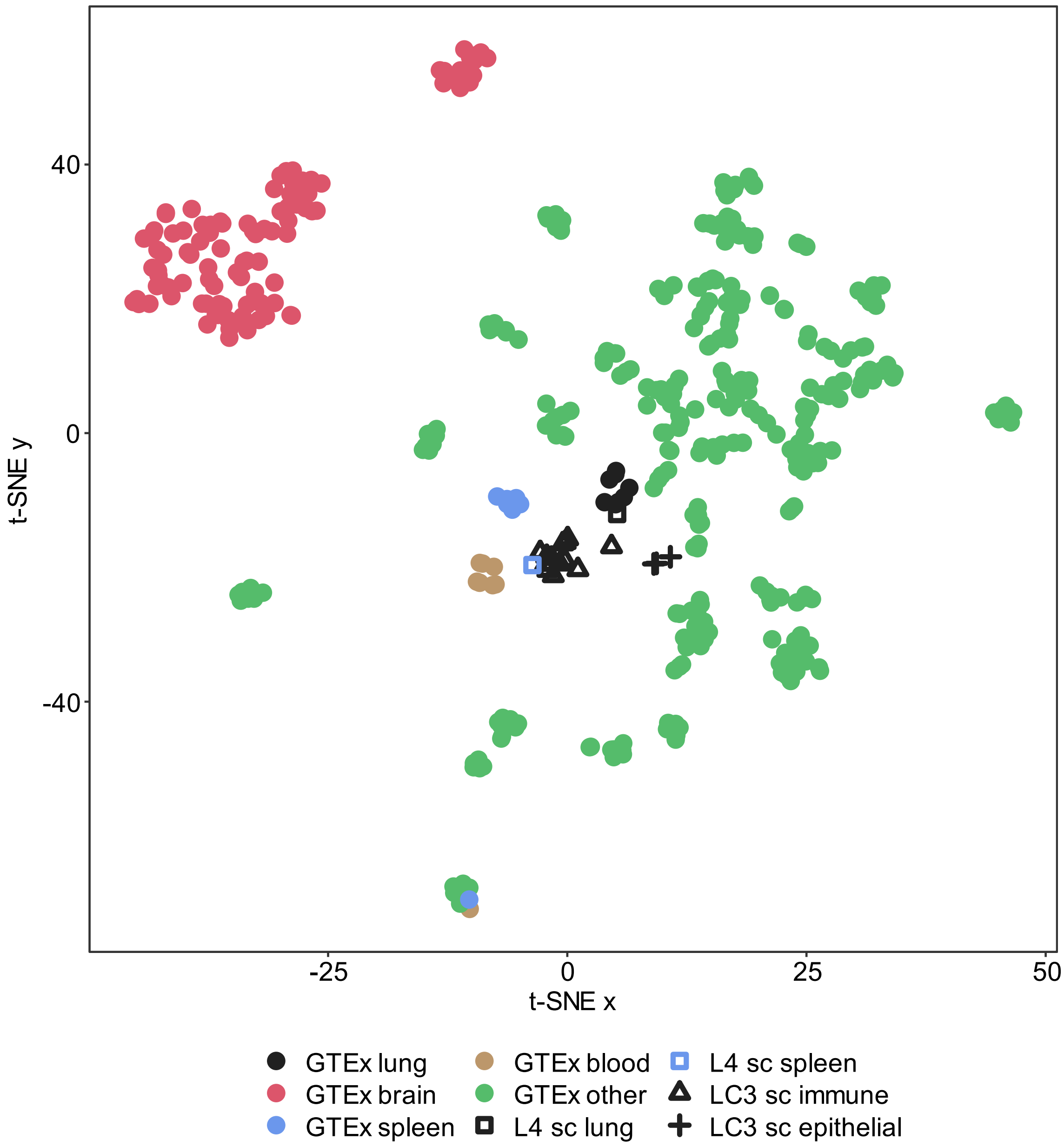


Fig. S8: Effect of quantile normalization on structural comparison. *Structural comparison of genome-scale metabolic models generated by ftINIT from RNA-Seq profiles from GTEx (bulk data) and several single-cell RNA-Seq datasets, displayed as a t-SNE projection. The GTEx blood and spleen samples are no longer grouped (outliers at the bottom of the figure). See Fig. 2D in the main text for details about the datasets used.*

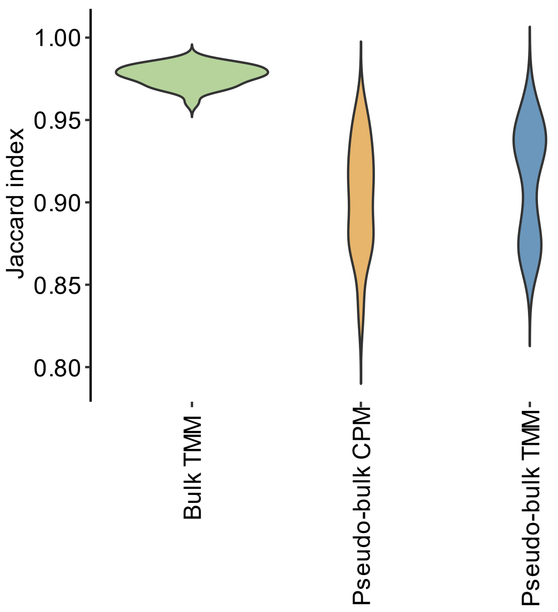


Fig. S9: Variation across samples in single-cell data. *The Jaccard index is calculated between all pairs of 8 samples of each category (bulk T cells, TMM normalized, the Bulk T cell dataset; single-cells pooled per sample, T cells from the HCA CB dataset, CPM normalized; single-cells pooled per sample, T cells from the HCA CB dataset, TMM normalized).*

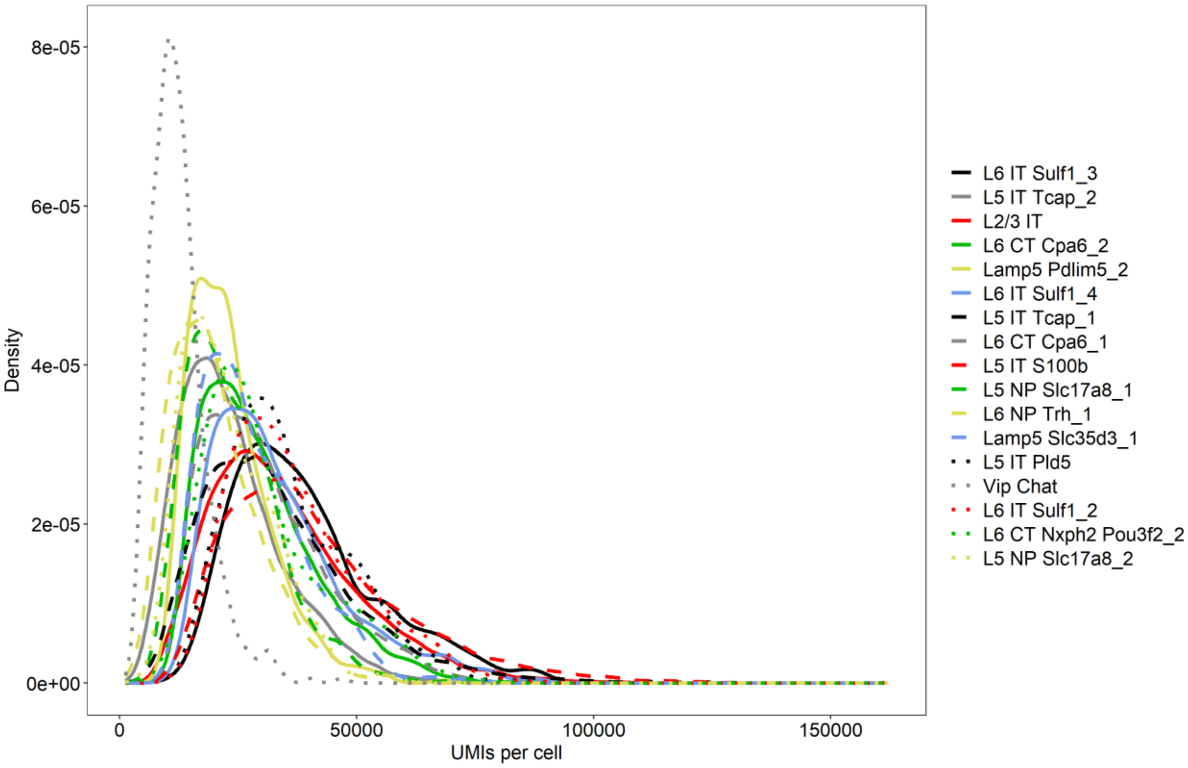


Fig. S10: Distribution of UMIs per cell per cluster in the MCOR3 dataset.

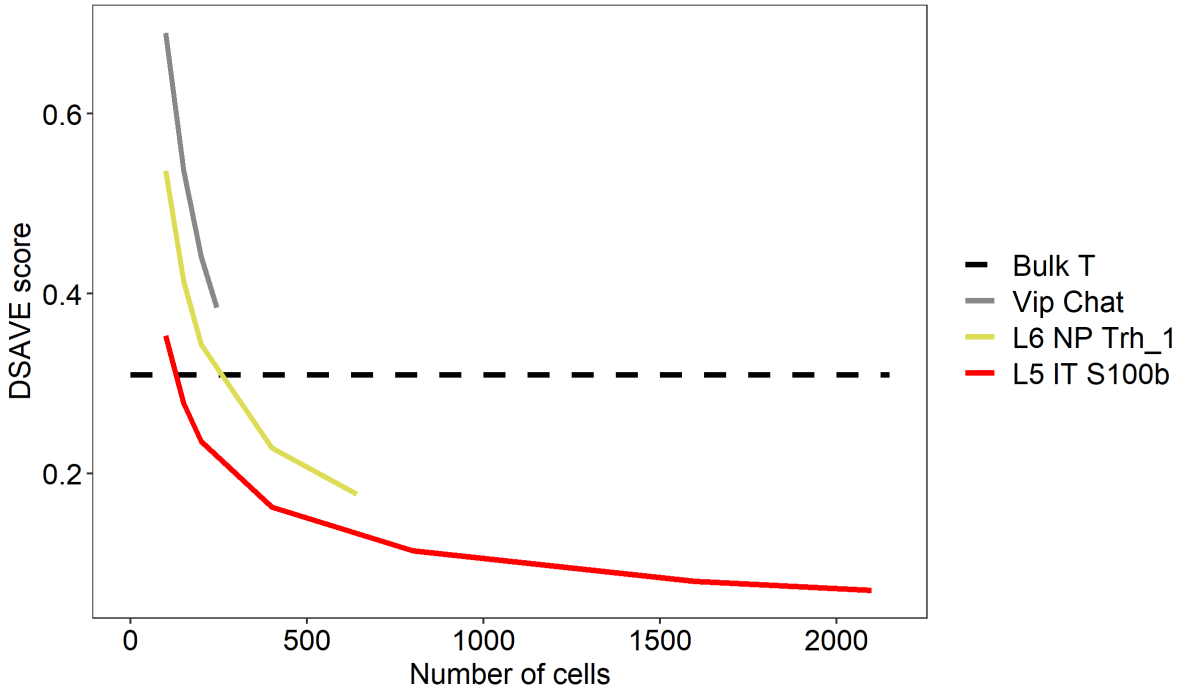


Fig. S11: DSAVE analysis of the MCOR3 dataset. *Only a subset of the clusters is shown for clarity; the Vip Chat is the cluster with the least UMIs per cell (see Fig. S6). Although we do not have enough cells for prediction when this cluster reaches the bulk variation, since DSAVE requires 2 times the number of cells to be able to measure the variation, we estimate it to be less than 450 cells.*

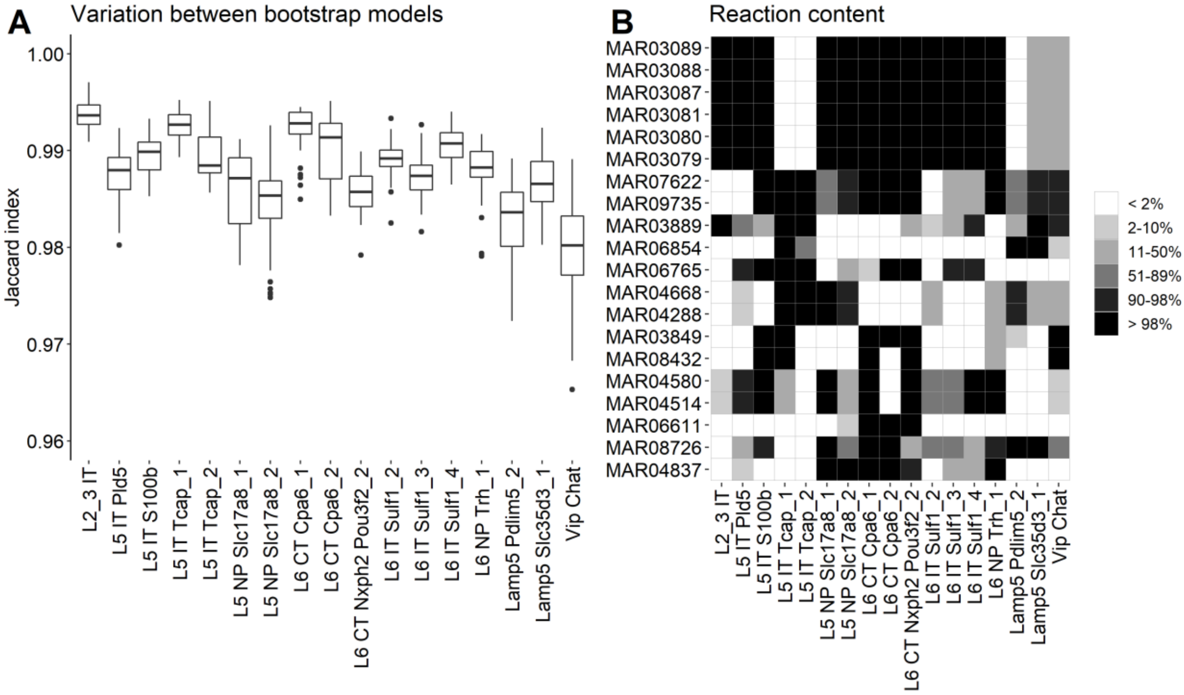


Fig. S12 Variation between bootstrap models. *A. Pairwise comparison of reaction content between bootstrap models from each cell type. B. Reaction content for 100 bootstraps of each cell. The 20 first reactions (out of 240) for which at least two cell types had more than 98% success rate and at least two had less than 2% such rate are shown. Notably, the Vip Chat cell type, for which DSAVE registered the largest variation (see Fig. S11), also shows most uncertainty here.*

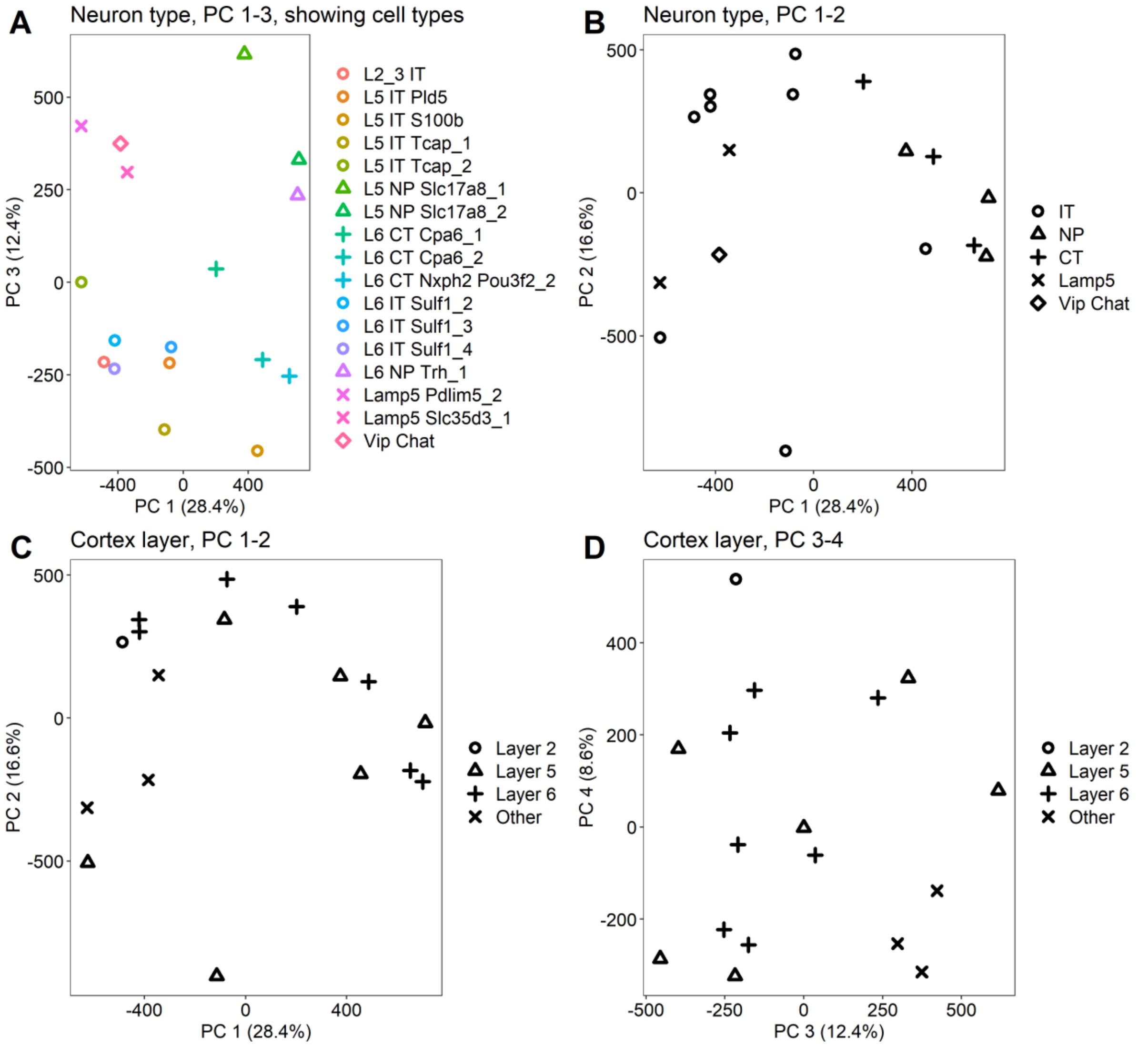


Fig. S13: Exploration of factors affecting the top PCs from the neuron metabolic networks. *Structural comparison of context-specific models generated by ftINIT from pooled single-cell clusters from the MCOR3 dataset, visualized as a PCA projection. A. PC1 and PC3, where the symbols reflect neuron type. Same as Fig. 3C in the main text but showing the individual cell types. B. The two first principal components, where the symbols reflect neuron type. PC2 does not contribute well to separating the cell populations by neuron type (IT, NP, CT, Lamp5, Vip Chat). C. The two first principal components, where the symbols reflect cortex layer. No clear dependence on cortex layer. D. PC 3 and 4, where the symbols reflect cortex layer. No clear dependence on cortex layer.*

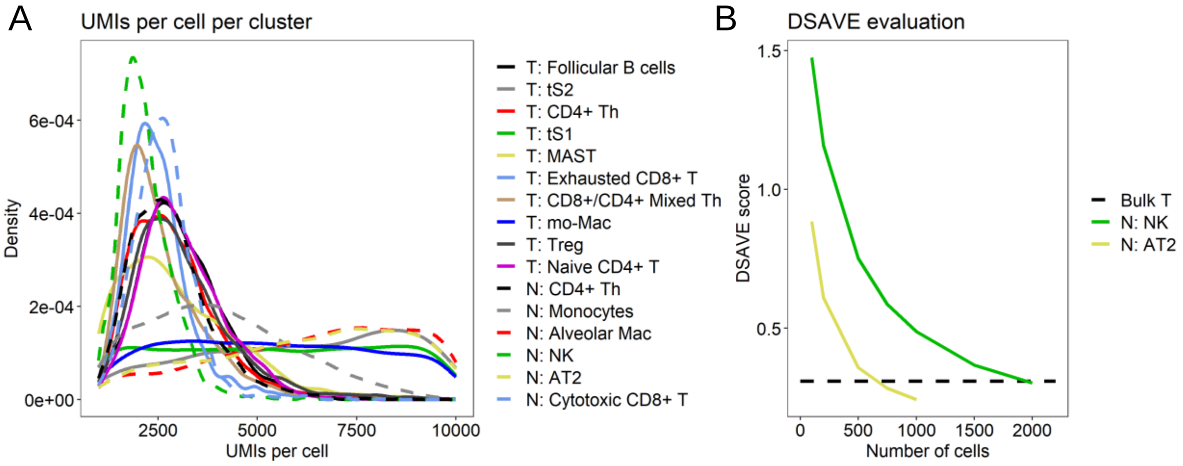


Fig S14: Evaluation of cluster size for the LC3 dataset. *A. The distribution of UMIs per cell across clusters. B. DSAVE total variation score for two clusters with high (N: AT2) and low (N: NK) average number of UMIs per cell.*
]

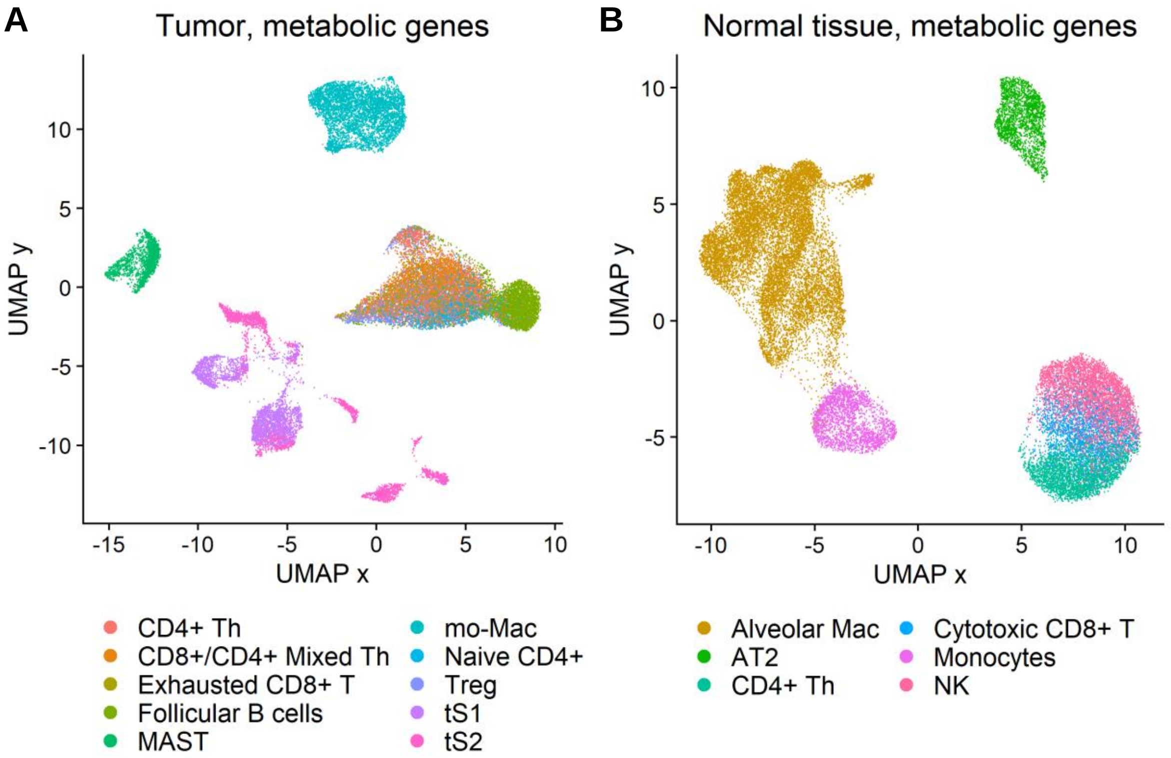


Fig S15: UMAP projection using only metabolic genes for the LC3 dataset. *These figures are similar to Fig. 4 A-B in the main text, with the difference that only metabolic genes were included in the data processing. A. Tumor samples. B. Samples from normal tissue.*

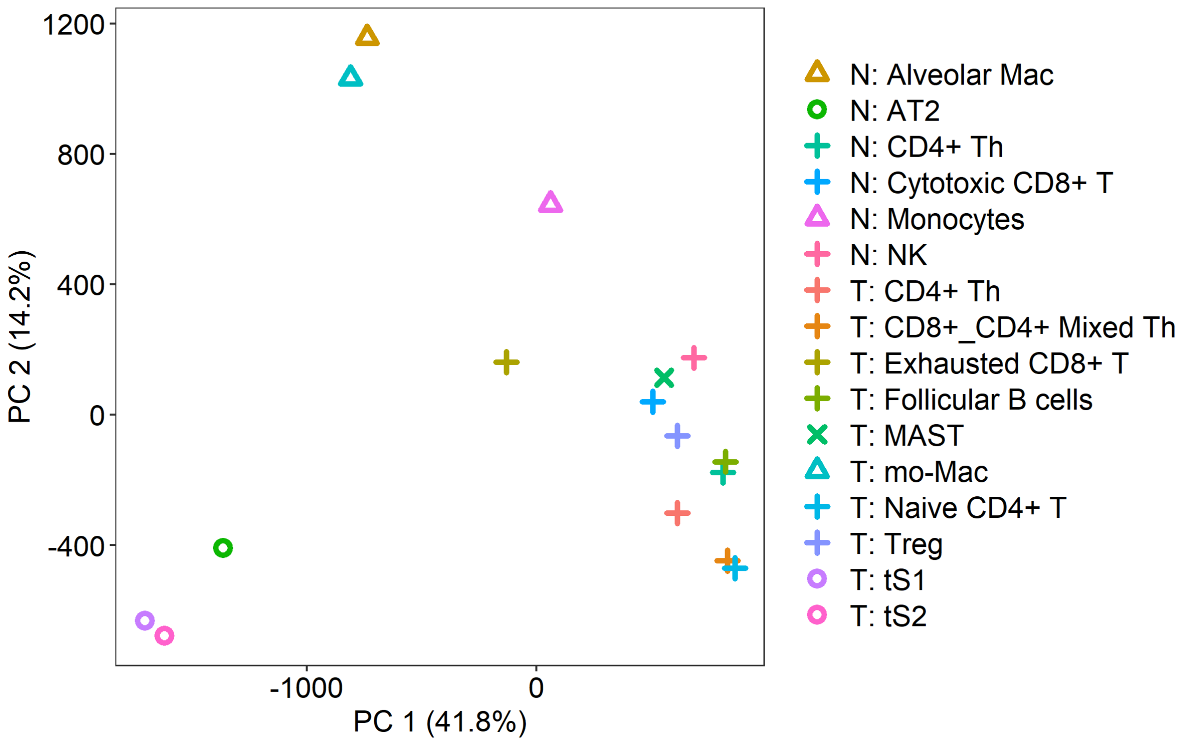


Fig S16: Structural comparison of cell types in the tumor microenvironment. *This figure is similar to Fig. 4C in the main text, with the difference that the individual cell type names are displayed here.*

Supplementary Tables

Table S1. Datasets

ID	Information
HCA CB	Cord blood data from Census of Immune cells, Human Cell Atlas (1,2), in total ~254,000 cells from 8 patients, 10X Chromium v2. The data can be downloaded from Census of immune cells. https://data.humancellatlas.org/ The T and B cells of this dataset, identified using the cell type classification of the authors, are referred to as HCA CB T and HCA CB B, respectively. For practical reasons, these populations have been reduced to 25,000 cells each.
PBMC68k	~68,000 blood (PBMC) cells from a single patient, 10X Chromium v1, denoted as Fresh 68k PBMC (3). The T cells of this dataset, as identified by the authors, are referred to as PBMC68k T. The data is available at 10X Genomics home page (https://www.10xgenomics.com/resources/datasets/fresh-68-k-pbm-cs-donor-a-1-standard-1-1-0).
Mel	4,600 cells from the tumors of 19 melanoma patients, Smart-Seq2 (4). The T cells of this dataset, as identified by the authors, are referred to as Mel T. The data is available for download at the GEO data repository, accession number 72056.
LC	~39,000 cells from the tumor microenvironment of lung cancers and ~13,000 cells from adjacent healthy lung tissue, 10X Chromium (mix of v1 and v2) (5). The T cells and macrophages, as identified by the authors, are referred to as LC T and LC Mac, respectively. The data is available for download in ArrayExpress under accessions E-MTAB-6149 and E-MTAB-6653.
TCD8	~10,000 CD8 positive FACS-sorted T cells from the blood (PBMC) of a single patient, 10X Chromium v2 (6). For symmetry, the cells in this dataset are referred to as TCD8 T. The data is available for download at the GEO data repository, accession number 112845.
LC3	~100,000 cells from lung tumors and ~100,000 cells from adjacent healthy lung tissue, in total 44 patients, 10X Chromium v2 (7). The data is available for download at the GEO data repository, accession code GSE131907.
L4	~57,000 cells from lung tissue and ~94,000 cells from spleen tissue, 10X Chromium v2 (8). All cells from the lung and spleen sample, respectively, were pooled to form two pseudo-bulk samples. The data is available through the Human Cell Atlas Data Coordination Platform and NCBI BIOPROJECT accession code PRJEB31843.
MCOR3	~94,000 deeply sequenced cells from the mouse primary motor cortex, primarily neurons, 10X Chromium v3 (9). The data is available at http://data.nemoarchive.org/biccn/lab/zeng/transcriptome/scell/
Bulk T	8 T cell bulk RNA-Seq samples from human PBMC, available from the BLUEPRINT Epigenome Project(10). The samples were taken from the project EGAD00001001173 and have the following sample IDs: S002EV11, S004M711, S007DD11, S007G711, S008H111, S009W411, S001FRB1, and S0041C11. The FASTQ files

	were processed as previously described (11): The FASTQ files were first processed using kallisto (12) to produce gene counts (estimated counts produced by kallisto). The data was then normalized using TMM (13), and scaled to an average count of 10^6 per sample.
DepMap	Bulk RNA-Seq TPM-normalized data and CRISPR screening data for gene essentiality from DepMap (14,15) (version 21Q3, specifically “CCLE_expression_full.csv” and “Achilles_gene_effect.csv”). Only 15 samples in the RNA-Seq data with matching gene essentiality data were used. The data is available from https://depmap.org/portal/.
GTEx	Bulk RNA-Seq data from in total 53 tissues from GTEx (16) (version 8, RNA-Seq v. 1.1.9, both TPM and raw counts files). Only 5 samples from each tissue were used. For the execution time evaluation, the median expression within each tissue was used. The data is available from the GTEx portal, https://gtexportal.org/home/.
HPA	Single-cell data from multiple human organs and cell types gathered from multiple datasets (17). The data is made available through the Human Protein Atlas (https://www.proteinatlas.org/about/download).

Supplementary notes

Note S1 – Statistical considerations for generation of context-specific models from single-cell RNA-Seq data

A central question when generating context-specific models from single-cell RNA-Seq data is how to assess the uncertainty both in the data and in the algorithm. What we primarily seek to estimate when generating context-specific models from single-cell data is the uncertainty of the presence of reactions (or other binary aspects, such as the ability to perform metabolic tasks). Given several cell populations per cell type, it is possible to do a pairwise comparison between two cell types, where a statistical test could explain if the tendency to include a reaction is higher in models generated from one cell type than the other. However, for cases where we seek to compare multiple samples, it would also be beneficial to estimate the uncertainty in reaction presence in a sample without comparing it to other samples. In this note, we explain how we address these two cases statistically, especially the latter.

In the ideal case, an adequate number of biological samples with enough cells per cell type is available in the dataset. Differences in reaction presence between two samples can then be tested on pooled pseudo-bulk samples using a two-sided Fisher's exact test, followed by correction for multiple testing by for example the Benjamini-Hochberg method. However, most datasets do not contain enough cells and samples to obtain stable models and significance using such a method, which poses a challenge for the statistical analysis.

The variation between two pools of cells originating from different samples can be divided into two components: 1) systematic variation (technical and biological) between the biological samples and 2) cell-to-cell variation (mainly sampling effects but also effects such as transcriptional bursting), which we assume is mostly independent of sample. The first component is independent of pool size, while pool size heavily influences the second. At an infinite pool size, the second component is zero, whereas it dominates the variation for small pool sizes. The importance of the two components also varies with gene expression; for highly expressed genes, the second component is less important than for lowly expressed genes. As an example of the importance of these components, Squair et al found that differential expression analysis methods that operate on pseudo-bulk pools of single cells per sample are more accurate compared to methods operating directly on single cells, ignoring sample origin (18). They show the main source of the error to be the inability to account for systematic variation across samples in the cell-wise methods, leading to an underestimation of the total variation, giving rise to false positives. The effect was only detected for highly expressed genes, which is likely due to the greater contribution of sampling effects (component 2) to the total variation compared to systematic across-sample variation (component 1) for lowly expressed genes, and therefore the total variation is only slightly underestimated by mixing samples for such genes.

For datasets not containing enough cells and samples to enable statistical tests across pseudo-bulk samples, we propose to treat all cells belonging to a certain cell type as a single population, regardless of sample origin. As an approximation, we assume that the available cells in the population are a perfect representation of the cell type, containing all possible cell variants in proportions representative of the cell type. We can then create bootstraps from the cell population to estimate the variation in the cell population and pool each bootstrap to generate a pseudo-bulk sample. This approach makes it possible to statistically assess the presence of reactions in smaller datasets and is a good option in cases where insufficient data is available for tests across samples. However, this approach comes with two drawbacks: 1) the variation is underestimated since component 1 of the variation is not included, which could lead to false positives; 2) the variation is underestimated since the available cells in the population is not a perfect representation of the cell type, which also could lead to false positives. The latter effect worsens with small pool sizes. As an example, a pool of 10 cells will have many genes that are falsely not expressed due to sampling effects. Bootstrapping will not yield any variation in gene expression for such genes, and associated reactions will confidently, but falsely, be considered non-present. So, while the bootstrapping method is useful, it is important to understand that it will produce more false positives than a test across samples and that the pool size has an effect. If the pool size is small, many genes will falsely be zero (or very high), while if it is large, the variation considered to originate from component 2 may be considerably smaller than that of component 1, which is largely ignored. It is also worth noting that in the case where uncertainty is estimated across samples, differences in the number of cells (and counts) between cell types could lead to biases in a similar way to drawback 2 explained above. An alternative explanation is that the uncertainty in the estimated variables (on/off) cannot be estimated directly from the data, in contrast to the case of bulk RNA-Seq differential expression analysis, where the uncertainty can be directly estimated from the counts assuming they follow a negative binomial distribution for each gene.

We seek to be confident in that a reaction considered present in one cell type statistically shows a larger tendency to be present in that cell type compared to cell types where it is considered non-present. Since we test all reactions, the result needs to be corrected for multiple testing. In this work, we used the bootstrapping strategy explained above with 100 bootstraps, and therefore seek to statistically compare the number of bootstraps for which the reaction is on between two cell types. Ideally, we would like to find a number *x* of bootstraps such that if a reaction is present in x bootstraps in one cell type, the reaction should with statistical significance be more available in that cell type compared to another cell type where the reaction is present in 100-*x* bootstraps. Finding such a number would enable us to define reactions as “on” and “off” in cell types and move away from relative pairwise comparison of cell types, where we are just be able to determine if a reaction is more present in one cell type than another. However, when using the Benjamini-Hochberg procedure to adjust p values for multiple testing, the results of such an adjustment will depend on the p values from other reactions for the pair of cell types compared, and the value of *x* would be different for different pairs of cell types. It is thus difficult with this definition to determine if a reaction is present with a significant p value without relating it to another cell type, which would be desired. We have therefore adopted the strategy to define fixed thresholds for when a reaction is considered present, and similarly non-present, where the thresholds are chosen at values that guarantee significance also after correction for multiple testing, regardless of the p values for other reactions. For a reaction to be considered confidently present, we have chosen that > 98% of bootstrap results should include the reaction, and < 2 % for non-present. The thresholds are purposely selected conservatively to keep the false positives at a low level at the expense of more false negatives. To estimate the statistical significance at these levels, we consider a pair of cell types and apply a Fisher’s exact test with the reaction present in 99 bootstraps for one cell type present in 1 bootstrap for the other, yielding $p < 2.2*10^{-16}$. We can then estimate the minimal number of reactions n_{min} required in the model to render such a p value non-significant after correction for multiple testing using the Benjamini-Hochberg method:

$$n_{min} = \frac{i}{p}Q$$

where *i* is the p-value rank, *p* is the p value, and *Q* is the false positive rate, which we set to 0.05. In the worst-case scenario, where the adjusted p-value will reach its maximum value, the reaction we are investigating will have the lowest p value of all reactions, giving the rank *i*=1. Furthermore, we assume that $p= 2.2*10^{-16}$ (and not smaller). Using these values yields $n_{min} = 2.3*10^{14}$ reactions. With fewer reactions, which in the metabolic model is on the order of 10,000, it is not possible that a correction for multiple testing will yield an adjusted non-significant p value, and hence, there is no need to perform a pairwise comparison of the p values of cell types when using the thresholds.

Note S2 – ftINIT

The purpose of the ftINIT (fast task-driven Integrative Network Inference for Tissues) algorithm is to determine the available metabolic subnetwork in a sample based on a template model and omics data. The ftINIT algorithm is as previous versions based on assigning scores to reactions, in this case from RNA-Seq data and gene rules (GPRs) as described previously, where a TPM threshold value per gene determines if the score should be positive or negative (19,20). The previous version of the algorithm was based on two steps: The network minimization step and gap filling based on essential metabolic tasks.

The network minimization step strives to determine which reactions to include in the model by maximizing the sum of the reaction scores of the included reactions, while ensuring that all included reactions can carry flux. The optimization problem in this step is based on mixed integer linear programming (MILP) and is a computationally expensive problem for a large model such as Human-GEM. To speed up the process, the optimization can be run with two flags that allow for secretion of all metabolites and for reversible reactions to carry flux in both directions. In practice, the latter means that all reversible reactions can form loops within themselves and can thereby always carry flux, and any such reactions with positive scores will automatically be included in the resulting model. The major downside of using these flags is that unwanted gaps can be left in the model, but for use with Human1, they are required to reach an acceptable execution time.

In the gap-filling step, the algorithm traverses a list of metabolic tasks, where a certain product(s) should be possible to produce from a given substrate(s). For each task the algorithm ensures that the task can be performed by adding reactions at a minimal cost to the total reaction score of the included reactions.

The purpose of ftINIT is to provide a substantially faster software for generating context-specific models from RNA-Seq data with comparable model output quality. To accomplish this goal, we have applied 3 strategies to the network minimization step: 1) Optimization of the existing code (not further described), 2) Reformulation of the MILP, and 3) A split of the network minimization step into two sub-steps. In addition, the gap-filling code was optimized, reducing the execution time of that step by an order of magnitude (not further described). ftINIT is primarily optimized for use with the Human1 model.

Reformulation of the MILP

A key factor for reducing the MILP solve time is to reduce the number of variables in the problem, especially the number of integer (Boolean) variables. In the previous versions of tINIT, the model was converted to an irreversible model, in which each reaction was given a boolean variable describing if the reaction should be included or not in the final model. In ftINIT, we have reduced the number of integer variables using approaches such as: 1) We do not convert the model to an irreversible model, which leads to fewer reactions. It also solves the problem with loops caused by reversible reactions. 2) Irreversible reactions with a positive score do not need an integer variable; the problem can be defined using only continuous variables for these reactions. A Boolean variable is usually used to determine if the reaction is included in the model or not; it is important for the optimization that the value of this variable is either 0 (or almost 0) or 1, and does not take on any value in between these values. For reactions with positive score, the value can however be allowed to instead be a continuous value from 0 to 1, since the optimization leads to that in-between values are unfavorable and will thus not be chosen. 3) All linearly dependent reactions are merged (only when running the MILP). 4) Some reversible reactions are made irreversible (such as reactions identified as essential, or reactions that can only carry flux in one direction for topological reasons).

5) When the MILP is calculated, the stoichiometry of the model is modified to reduce the differences in magnitude between fluxes.

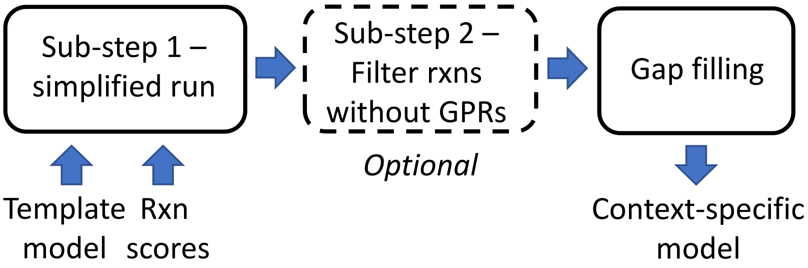
Division of the network minimization step into sub-steps

An obvious method to shorten the execution time of the MILP in the network minimization step of the problem is to reduce the size of the problem. The Human-GEM model contains many reactions without GPRs, which means that RNA-Seq data cannot be used to score these reactions. Such reactions are by default assigned a score of -2, and will in tINIT/ftINIT not be added unless they are needed to provide flux for other reactions with positive scores. In total, there are several thousand reactions without GPRs, and an alternative method to scoring them with a -2 is to not score some of them at all, and thereby not include them in the optimization, while still allowing them to carry flux. In ftINIT, we have implemented this possibility. To enable the user to select which reactions to omit from the optimization problem, they are divided into 8 partly overlapping categories, and selected individually with a vector of 8 Booleans, for example [1 1 1 1 1 1 1 0], where the reactions selected is the union of the specified categories (Table A). The vector of Booleans is referred to as the 'rxns to ignore mask'.

Table A. The 8 different categories of reactions without GPRs that can be selected for exclusion from the optimization problem. *The number of custom reactions represent the number of such reactions used in this study.*

Id	Name	Description	No. rxns
1	Exchange	All exchange reactions.	1,110
2	Import	Reactions importing metabolites into the cell.	707
3	Simple transport	Reactions that transport one metabolite from one compartment to another.	826
4	Advanced transport	Transport reactions that involve multiple metabolites, e.g., antiporters.	129
5	Spontaneous	Reactions marked as spontaneous in the model.	14
6	EC	Extracellular reactions, i.e., reactions in the e (or s depending on model) compartment.	32
7	Custom	Reactions defined by the user.	52
8	All	All reactions without GPRs.	3,475

We divided the network minimization step of ftINIT into 2 sub-steps, of which the second is optional (Fig. A). In the first sub-step, the reactions without GPRs as selected by the user are excluded from the problem. In the second step, which is optional, we flag all previously included reactions as essential, and allow for a different selection of reactions without GPRs that should be excluded from the problem. If for example only the exchange reactions are omitted from the problem in this step, the algorithm will for the rest of the reactions without GPRs determine if they should be included or not.


Sub-step 1 – simplified run

Sub-step 2 – Filter rxns without GPRs

Optional

Gap filling

Template model

Rxn scores

Context-specific model

Fig. A: Sub-steps in the ftINIT algorithm.

At the end of the network minimization step, all reactions without GPRs that were not part of the problem in the calculations will simply be included in the model.

While the sub-step setup is highly configurable, allowing for any number of steps, we recommend using one of two predefined setups, which we call modes: 1) "1+0" - runs the algorithm without sub-step 2 with the rxns to ignore mask set to [1 1 1 1 1 1 1 0]. This setup will include most reactions without GPRs, and is suitable for running for example structural comparisons. 2) "1+1" – identical to setup 1, with the difference that sub-step 2 is included, with the rxns to ignore mask for sub-step 2 set to [1 0 0 0 1 0 0 0]. This will remove many reactions without GPRs that are not needed for the other included reactions to be able to carry flux, but still include most exchange reactions and all spontaneous reactions. We recommend this sub-step for cases when it is desired to remove many of the reactions without GPRs. This setup takes a bit longer to run, typically 1-2 minutes per run, and may add some randomness since there may be many possible ways to remove reactions while retaining an optimal MILP score.

ftINIT will with the used parameterization (the rxns to ignore mask described in the standard modes above together with the commonly used RNA-Seq threshold value of 1 TPM) produce larger models than its predecessor (tINIT). If smaller models are desired, these parameters can be modified, although it may affect the execution time of the algorithm.

ftINIT/tINIT parameterization for this project

ftINIT was run with with a MipGap of 0.04% and a time limit of 2 minutes per step, and has not been tested with other solvers. The MipGap parameter describes a limit for the maximal possible error of the objective function as calculated by the solver – if the maximal possible error becomes smaller than this value the solver stops, reporting that a solution is found. The time limit parameter makes the solver stop after a certain time, regardless of the status of the solution. Should a solution not be found within the time limit of 2 minutes, the allowed MipGap is increased to 0.30%. If the MipGap after two minutes is higher than the allowed value, the optimization will be rerun with a time limit of 5,000 seconds – though based on our observations this only happens on rare occasions. The parameters for the previous tINIT could be freely specified but was often used with only a time limit parameter of either 1,000 s, 5,000 s, or 10,000 s depending on the complexity of the samples.

Note S3 – Detailed methods description

Model preparation

To reduce the model size, we removed in total 2,680 reactions from Human1 (Human-GEM v. 1.12) that were deemed unnecessary, leaving a model with 10,390 reactions. The reactions removed included reactions related to drug metabolism, amino acid triplets, a list of reactions that were identified as duplicates, and dead-end reactions. A similar strategy was used for preparing Mouse1 (Mouse-GEM v. 1.3).

Pooling single-cell samples to simulate pseudo-bulk

The single-cell data was pooled into pseudo-bulk RNA-Seq profiles for use with ftINIT/tINIT. We chose a strategy that assigned each mRNA molecule equal importance, and therefore summed all UMIs from all cells in a population into a single pseudo-bulk sample. The sample counts were then scaled to a total sum of 10^6 (counts per million, CPM, normalization), except in cases where other normalization methods such as TMM or quantile normalization were applied.

Evaluation of tINIT execution time

tINIT and ftINIT were run for the 10 first tissues in the GTEx dataset, where each tissue was represented by the median expression values of all RNA-Seq samples from that tissue. The time was measured per run, performed on a standard laptop computer (Intel Core i7-6600U, 2.60 GHz, 2+2 cores). ftINIT uses a MipGap of 0.04% for the first two minutes of each step, which is increased to 0.30% after two minutes if the solver has not found a solution within that time. Thus, since the worst-case scenario of ftINIT can lead to an acceptance of a MipGap of 0.30%, we set the optimization for tINIT to end at a MipGap of 0.30%. This is not the set of parameters by which tINIT is normally run (where only a time parameter is set), but this set of parameters makes the execution times comparable between the methods, although in favor of tINIT.

Evaluation of required pool size and cell type contamination

For each pool size and dataset, we generated 20 pairs of random cell populations of the examined pool size, where the two populations in each pair did not share any cells. The cells in the populations were pooled to generate RNA-Seq profiles and processed using ftINIT, generating 400 models per dataset. Each pair was then compared for reaction content using Jaccard similarity coefficient. The mean value of the Jaccard similarity coefficients was then calculated for all pairs within the same pool size and dataset.

For the cell type contamination we used single cell data from patient 4 of the LC dataset. 8 different pool sizes were determined and in addition 5 different fractions of contamination. 60 pairs of random cell populations were then generated for each pool size and level of contamination. The first of the cell populations in each pair was uncontaminated, drawn completely from the LC T cell population, while the other was drawn from both LC T and LC M, such that

$$n_m = f_c p$$

where n_m is the number of cells drawn from the LC M population, f_c is the fraction of contaminated cells to include, and p is the pool size. The remaining cells up to p cells were drawn from the LC T population. The cells in the populations were pooled to generate RNA-Seq profiles and processed using ftINIT, generating in total 5,000 models. The mean value of the Jaccard similarity coefficients (comparing reaction presence) was then calculated for all model pairs within the same pool size and level of contamination.

t-SNE and PCA

The “tsne” function in MATLAB was used to generate t-distributed stochastic neighbor embedding (t-SNE) (21) coordinates from the model structure. PCA was performed using the function “prcomp” in R (except for the Seurat workflow). The t-SNE figures in Fig. 1 and 2 were based on the presence of each reaction in a single model, and the Hamming distance was used as distance between samples. For the PCA plots and the t-SNE plots in Fig. 5, each reaction was scored between 0 and 100 depending on how many bootstrap models included the reaction. The PCA was then performed with these scores per reaction as variables across the cell subtypes.

Model group comparisons

The model groups used in the comparisons in Fig. 2 E-F can be described as follows: “Across bulk tissues” – comparison of all GTEx sample pairs that origin from different tissues; “Within bulk tissues“ – comparison of all GTEx sample pairs that origin from the same tissue; “Within LC3 immune” - comparison of all pairs of immune cell types from the LC3 dataset (both healthy tissue and tumor); “Immune across tech” – comparison of all immune cell types from the LC3 dataset to all whole blood samples in GTEx; “Across tissues and tech” – comparison of all immune cell types from the LC3 dataset to all GTEx samples.

Single-cell analysis using Seurat

Seurat v. 4.04 (22) was used to analyze the single-cell datasets MCOR3 and LC3, using a standard Seurat pipeline, including log normalization (scale.factor=10,000), finding of variable genes(vst, 2,000 features), data scaling, PCA, finding of neighbors (using 15 PCs), clustering (resolution = 0.5, otherwise standard parameters), followed by UMAP (using 15 PCs) and DimPlot for visualization. The cell type classifications provided as part of the publications were used, and all cells that had a classification matching one of the cell types/subtypes selected for study were included in the processing. No other cell filtering was performed. For the LC3 dataset the healthy tissue data and tumor data were processed separately. The metabolic genes for human and mouse were extracted from the respective models, where all genes included in a gene rule was considered metabolic. Genes from 5 reactions were excluded ('MAR09577', 'MAR09578', 'MAR09579', 'MAR07617', and 'MAR07618') since these reactions describe phosphorylation of proteins, which we consider to be signaling reactions. In total, we classified 2,784 genes in MCOR3 dataset and 2,912 in the LC3 dataset as metabolic.

Statistical information

The statistical tests used for the bootstrap models are described in Note S1. For Fig 2C, a one-sided Wilcoxon rank sum test was performed to compare the results from the 60 contaminated samples to the results from the 60 pure samples, using the R function wilcox.test with the parameters alternative = "greater" and paired = FALSE.

The boxplots were generated using the R function geom_boxplot in ggplot2 with standard parameters. This means that the center line represents the median, the hinges correspond to the 25th and 75th percentiles, and the whiskers stretch to the furthest point within the distance of 1.5 times the size of the box from the hinge. Outliers are presented in Fig. 1C but not in Fig. 2F to avoid cluttering of the figure.

References

1. Li B, Kowalczyk MS, Dionne D, Ashenberg O, Tabaka M, Tickle T, et al. Census of Immune Cells [Internet]. Human Cell Atlas Data Portal. 2018 [cited 2019 Feb 19]. Available from: https://data.humancellatlas.org/explore/projects/cc95ff89-2e68-4a08-a234-480eca21ce79

2. Rozenblatt-Rosen O, Stubbington MJT, Regev A, Teichmann SA. The Human Cell Atlas: from vision to reality. Nat News. 2017 Oct 26;550(7677):451.

3. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Jan 16;8:14049.

4. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016 Apr 8;352(6282):189–96.

5. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. Nat Med. 2018 Aug;24(8):1277–89.

6. Chen J, Cheung F, Shi R, Zhou H, Lu W, Candia J, et al. PBMC fixation and processing for Chromium single-cell RNA sequencing. J Transl Med. 2018 Jul 17;16(1):198.

7. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. Nat Commun. 2020 May 8;11(1):2285.

8. Madissoon E, Wilbrey-Clark A, Miragaia RJ, Saeb-Parsy K, Mahbubani KT, Georgakopoulos N, et al. scRNA-seq assessment of the human lung, spleen, and esophagus tissue stability after cold preservation. Genome Biol. 2019 Dec 31;21(1):1.

9. Booeshaghi AS, Yao Z, Velthoven C van, Smith K, Tasic B, Zeng H, et al. Isoform cell type specificity in the mouse primary motor cortex [Internet]. 2020 Mar [cited 2021 Sep 15] p. 2020.03.05.977991. Available from: https://www.biorxiv.org/content/10.1101/2020.03.05.977991v3

10. Blueprint Epigenome Project, 2016. [Internet]. [cited 2019 Mar 4]. Available from: http://dcc.blueprint-epigenome.eu/#/home

11. Gustafsson J, Robinson J, Inda-Díaz JS, Björnson E, Jörnsten R, Nielsen J. DSAVE: Detection of misclassified cells in single-cell RNA-Seq data. Sengupta D, editor. PLOS ONE. 2020 Dec 3;15(12):e0243360.

12. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. Nat Biotechnol. 2016 May;34(5):525–7.

13. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol. 2010 Mar 2;11:R25.

14. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019 May;569(7757):503–8.

15. Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nat Genet. 2017 Dec;49(12):1779–84.

16. Carithers LJ, Ardlie K, Barcus M, Branton PA, Britton A, Buia SA, et al. A Novel Approach to High-Quality Postmortem Tissue Procurement: The GTEx Project. Biopreservation Biobanking. 2015 Oct;13(5):311–9.

17. Karlsson M, Zhang C, Méar L, Zhong W, Digre A, Katona B, et al. A single–cell type transcriptomics map of human tissues. Sci Adv. 7(31):eabh2169.

18. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nat Commun. 2021 Sep 28;12(1):5692.

19. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. Mol Syst Biol. 2014 Mar 1;10(3):721.

20. Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT. PLOS Comput Biol. 2012 maj;8(5):e1002518.

21. van der Maaten L, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;9:2579–605.

22. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun 24;184(13):3573-3587.e29.