StanDep: capturing transcriptomic variability improves context-specific metabolic models

Chintan J. Joshi[1], Song-Min Schinn[1], Anne Richelle[1], Isaac Shamie[2,3], Eyleen J. O'Rourke[5], Nathan E. Lewis[1,2,4*]

[1] Department of Pediatrics, University of California, San Diego, School of Medicine, La Jolla, CA 92093
[2] Novo Nordisk Foundation Center for Biosustainability at the University of California, San Diego, School of Medicine, La Jolla, CA 92093
[3] Bioinformatics and Systems Biology Program, University of California, San Diego, United States
[4] Department of Bioengineering, University of California, San Diego, La Jolla, CA 92093
[5] Department of Biology, University of Virginia, Charlottesville, VA 22903

*Correspondence: nlewisres@ucsd.edu

Keywords: omics data, systems biology, metabolism, data integration

Abstract

Diverse algorithms can integrate transcriptomics with genome-scale metabolic models (GEMs) to build context-specific metabolic models. These algorithms rely on preprocessing - identifying a list of high confidence (core) reactions from transcriptomics. Studies have shown parameters related to preprocessing, such as thresholding of expression profiles, can significantly change model content. Importantly, current thresholding approaches are burdened with setting singular arbitrary thresholds for all the genes; thus, resulting in removal of enzymes needed in small amounts and even many housekeeping genes. Here, we describe StanDep, a novel heuristic method for preprocessing transcriptomics data prior to integration with metabolic models. StanDep clusters enzymes based on their expression pattern across different contexts and determines thresholds for each cluster using data-dependent statistics, specifically standard deviation and mean. Hundreds of models for the NCI-60 cancer cell lines, human tissues, and *C. elegans* cell types were built using StanDep. These models were able to capture higher number of housekeeping genes and improved precision in predicting gene essentiality (CRISPR and RNAi) compared to models built using more established approaches. Our study also provides novel implications for understanding how cells may be dealing with context-specific and ubiquitous functions.

Introduction

Genome-scale metabolic models (GEMs) provide mechanistic insights into an organism's phenotype. Phenotype is strongly influenced by the expression of genes, proteins, and enzymes in an environmental condition. Thus, expression of these have been comprehensively quantified for many tissues and cell types (Uhlen *et al.*, 2015; Cao *et al.*, 2017; Han *et al.*, 2018; Creecy and Conway, 2015), allowing the integration of these omics data with GEMs to answer diverse questions, spanning from the elucidation of molecular mechanisms (Lewis *et al.*, 2010; Mardinoglu *et al.*, 2014) to the identification of drug targets (Lewis and Abdel-Haleem, 2013; Shen *et al.*, 2010; Kim and Lun, 2014); thus, the use of GEMs is common practice for studying context-specific behavior (i.e., tissue, cell type, environmental conditions, or other variations to which cells are exposed to) (Hyduke *et al.*, 2013; Cimini *et al.*, 2009; Machado and Herrgård, 2014).

Manifestation of a phenotype requires not only expression of the genes but expression of proteins, post-translational modifications, assembly of the enzymes, presence of reaction substrates, and reaction occurrence. Further, each of these processes may have component-specific (gene/protein) variability (Li and Biggin, 2015; Winter *et al.*, 2004). Thus, predicting the phenotype from transcriptomic data alone is not trivial. However, recent studies have shown that, despite technological limitations in measuring omics data, transcription can explain larger variance in protein levels than previously thought (Li and Biggin, 2015). Thus, making transcriptomic data a reasonable choice for building cell-type and tissue-specific models (Opdam *et al.*, 2017) and resulting in the development of a wide array of algorithms to facilitate integration (Jerby *et al.*, 2010; Wang *et al.*, 2012; Agren *et al.*, 2012; Becker and Palsson, 2008; Zur *et al.*, 2010; Vlassis *et al.*, 2014). However, many algorithms require that user must come up with a list of active genes or reactions. We refer to this step as preprocessing, and it involves various decisions and; among which thresholding has the largest influence on the content of the context-specific models (Opdam *et al.*, 2017; Richelle, Joshi, *et al.*, 2018).

Different thresholding methods embed different assumptions about transcriptomics data into context-specific models. Global thresholds, for example, assume that each gene is only used if its expression is above a common level (Becker and Palsson, 2008). Similarly, local thresholds (Richelle, Joshi, *et al.*, 2018) wherein a gene/enzyme with moderate expression is considered "on" if it is more highly expressed in one tissue, compared to others. These two paradigms assume genes are either active due to high expression or tissue specificity. However, many enzymes have different substrate binding and catalytic efficiencies, and their activities are needed at different levels (e.g., catabolism of glucose for energy vs. vitamin biosynthesis); there are also enzymes that vary little (e.g., housekeeping genes) (Eisenberg and Levanon, 2013) or highly (e.g. tissue-specific genes) (Winter *et al.*, 2004) between samples. A recent study identified that variance in expression of genes could influence interpretation of transcriptomes across different species and organs (Breschi *et al.*, 2016; Barbosa-Morais *et al.*, 2012). Thus, treating all genes the same and defining a single threshold, in any method (local/global), may not capture the cellular phenotypic space accurately. Furthermore, these threshold values found for one study cannot be extended to other datasets due to lack of biological rationale.

To address the limitations of with global and local thresholding, we propose a novel heuristic approach, StanDep, which calculates thresholds and derives lists of core reactions. StanDep clusters enzymes based on their expression pattern across a variety of conditions. The method then calculates a threshold value for each cluster, followed by derivation of core reaction lists. Using StanDep, we generated lists of core reactions and built models; and compared them with those built using existing methods. We found that core reaction lists and models derived using our method accurately captured a large number of housekeeping genes and reactions that were considered as "off" by existing methods. We found that StanDep-derived models enjoyed high precision and high predictive capacity compared to existing methods without sacrificing accuracy. StanDep implementation, in its current state, worked best with fastCORE (Vlassis *et al.*, 2014), mCADRE (Wang *et al.*, 2012), and MBA (Jerby *et al.*, 2010). Further, using StanDep, we also prepared and validated models of 32 human tissues and 27 *C. elegans* cell types to show that our method preserves cellular properties such as housekeeping genes (Eisenberg and Levanon, 2013) and essential genes (Aguirre *et al.*, 2016; Meyers *et al.*, 2017; Kamath *et al.*, 2003; Doench *et al.*, 2016). Thus, StanDep provides a novel approach to obtain more accurate context-specific models of metabolism from transcriptomics data.

Results

Existing thresholding methods remove many housekeeping reactions

Housekeeping genes are required for cellular maintenance; therefore, these should be expressed across all cells under all conditions (for e.g. cancer cell lines, tissues, etc.). For existing thresholding methods (Richelle, Joshi, *et al.*, 2018) to accurately capture cellular states, they must be able to select housekeeping reactions, i.e. reactions associated with housekeeping genes. To this end, we identified 1000 housekeeping reactions associated with all metabolic housekeeping genes (Eisenberg and Levanon, 2013) in Recon 2.2 (Swainston *et al.*, 2016) (See Methods). We, then, converted expression values of genes in Recon 2.2 and two different datasets, i.e. CellMiner (Reinhold *et al.*, 2012; Shankavaram *et al.*, 2009) and Klijn et al. (hereafter referred to as Klijn data) (Klijn *et al.*, 2015), to enzyme expression values (See Methods). A list of core reactions was then calculated using the following three different thresholding methods: global (top 25th percentile), localT2 (top 25th percentile as upper and top 75th percentile as lower), and localT1 (top 75th percentile) (See Methods). We found that localT2 thresholds resulted in the core reaction lists with the largest fraction of housekeeping reactions retained, compared to localT1 and global thresholds (Table 1, Fig. S1). However, the best dataset-threshold combination (Klijn data with localT2) still eliminated more than 250 housekeeping reactions from the core lists. Thus, housekeeping genes, which produce essential enzymes, are poorly represented by existing thresholding methods.

Table 1. Fraction of housekeeping reactions retained following application of different thresholds

	Klijn, 2014	**CellMiner**
Global 75th (global)	0.59	0.47

Local T2 (25th, 75th) (localT1)	0.74	0.67
Local T1 25th (localT2)	0.53	0.50

There are various issues with existing thresholding methods. First, all genes are assumed to have the same expression patterns, regulation, and efficiencies. This, however, is not true; gene expression is governed by transcriptional efficiency, transcription factors, organization of regulation machinery, etc. Second, existing methods require a user-defined threshold above which genes/reactions are considered on; however, without direct experimental evidence (which would anyway be context dependent) there is no way to determine an exact threshold above which a gene can be classified as active. Third, some genes are expressed at lower levels because the mRNAs are more stable, the downstream cellular machinery involved in processing the gene products is more efficient, or their metabolic products are needed in lower quantities. A thresholding method dependent on a single arbitrary expression value may therefore exclude such genes. Fourth, genes such as housekeeping genes may be more consistently expressed across all cellular contexts as their gene products are required for cellular maintenance. However, the magnitude of expression of such genes may be high or low depending on their function or stability. Thus, there is a need for thresholding methods that consider not only the magnitude of the gene expression but also the patterns of variability in gene expression for every gene, or at least class of genes, across different cellular contexts.

Magnitude alone is insufficient to capture housekeeping genes

In the analysis above, we hypothesized that not only magnitude, but also patterns of variability in gene expression, is an important criterion for capturing genes required for cellular maintenance. To test this hypothesis, we clustered the patterns of expression variability for 1666 metabolic genes in Recon 2.2 across NCI-60 cell lines from the Klijn dataset, using hierarchical clustering into 25 clusters (Fig. S2A; please see Methods section). We found that seven clusters were enriched in housekeeping genes. Cluster 3 contained 53% of all housekeeping genes which represented >50% of the genes in the cluster (Fig. 1A). A similar analysis of the CellMiner data (Fig. S2B; see *Methods* and *Supplementary text*) revealed that some of the housekeeping genes were in a low-expression cluster (Cluster 14, Fig. S2B); a phenomenon that was not observed in the Klijn dataset. This may explain why CellMiner consistently yielded fewer housekeeping reactions for all thresholding methods (Table 1).

Though the housekeeping genes were expressed in all cancer cell lines, their absolute expression varied across the Klijn and CellMiner datasets. Thus, when calculating core reaction lists, magnitude alone led to exclusion of housekeeping gene functions. All housekeeping genes, in either of the datasets, showed enrichment in low variation clusters across cell lines. Therefore, we hypothesized that accounting for standard deviation as a measure for gene-specific variability may more accurately select active gene or reaction sets including housekeeping genes, which would result in more accurate tissue or cell type-specific metabolic models.

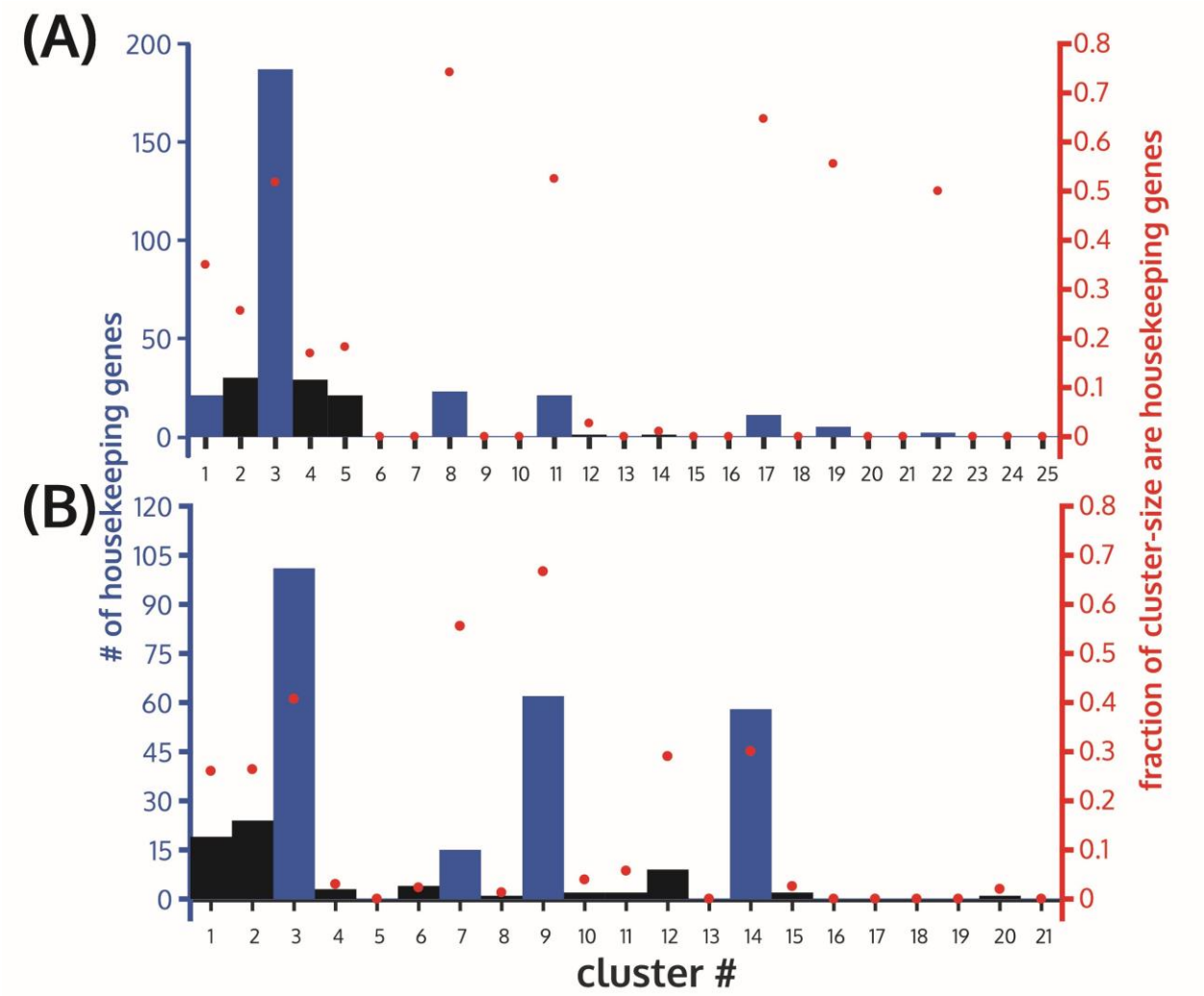

Figure 1: Enrichment analysis for presence of housekeeping genes in clusters of (A) Klijn et al. NCI-60 cancer cell line data and (B) CellMiner NCI-60 cancer cell line data. (A) Clusters 1, 3, 8, 11, 17, 19, and 22 are enriched for housekeeping genes. (B) Clusters 3, 7, 9, and 14 are enriched for housekeeping genes. The bars show number of housekeeping genes (left y-axis). Clusters not enriched for housekeeping genes are shown by black bars, while clusters enriched in housekeeping genes are shown in blue bars. The red dots show the fraction of number of genes in each cluster are housekeeping genes (right y-axis). Significant enrichment set as a hypergeometric p-value ≤ 0.05.

StanDep seeds core reaction lists with housekeeping reactions

The limitations of current methods exemplified above show the way to more accurate methods, that is the inclusion of expression magnitude and variability across different contexts. Thus, we present StanDep, a novel thresholding method that considers both aspects of the datasets. To implement StanDep, we first converted gene expression data of the NCI60 cell lines (from Klijn et al.) to enzyme expression values (see *Methods*). We then clustered enzyme expression across all the cell lines (Fig. S3), and calculated cluster-specific thresholds to identify core reactions (Fig S3). Our formulation contained two terms: (a) the mean term, and (b) the standard deviation term (see *Methods*). Of the 4738 reactions with enzyme expression data, 187 reactions were deemed inactive in all cell lines. Meanwhile, on average, more than

80% of housekeeping reactions were deemed active in each cell line (Fig. S4). Furthermore, many of the cell lines from similar cancer types cluster together based on enzyme selection (Fig. 2A) and core reaction lists (Fig. 2B).

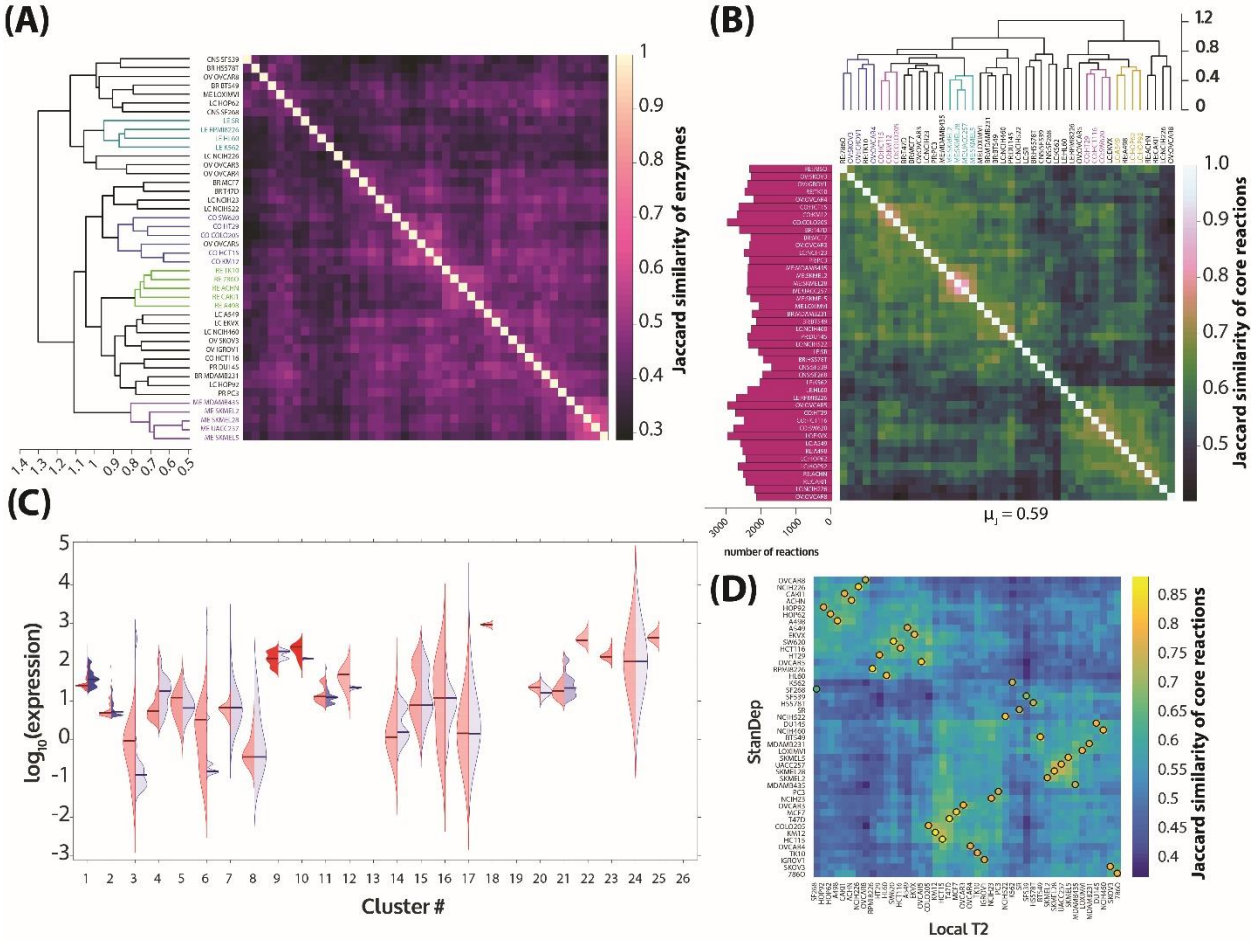

Figure 2: Properties StanDep-derived core reaction lists of Klijn et al. dataset and comparison with localT2-derived core reaction lists. (A) Jaccard similarity of *enzymes* selected using StanDep between 44 different NCI-60 cancer cell lines. The dendrogram on the left panel shows clustering of the cell lines based their Jaccard similarity ratios across all the other cell lines. The color of dendrogram lines and texts represent cell lines of same subtype clustered together (e.g., the purple cluster belongs to models of cancer cell lines of melanoma). (B) Jaccard similarity of *reactions* selected using StanDep between 44 different NCI-60 cancer cell lines. The dendrogram on the top panel shows clustering of the cell lines based on their Jaccard similarity ratios across all the other cell lines. The color of dendrogram lines and texts represent cell lines of the same subtype clustered together (e.g., the cyan cluster belongs to cancer cell lines of melanoma). The panel on the left shows the number of reactions selected in each of the cancer cell line models. (C) Violin map of comparison of reaction lists when compared to localT2 thresholding method. Red distributions show the distribution of reaction expression values in that cluster that were not present in StanDep-derived reaction lists but were present in localT2, while blue distributions show expression values for reactions in that cluster that were not present in localT2 but were present in StanDep. The opacity of the distribution represents the number of reaction-cell line pairs used to make the distributions. StanDep increases coverage for reactions that are expressed at low to medium expression levels as shown by larger coverage and opacity of the blue distribution for clusters 1 & 2. (D) StanDep-derived core reaction lists are compared with localT2-derived core reaction lists using the Jaccard similarity ratio. Almost all StanDep-derived core reaction lists are most similar to their localT2 derived counterparts with over 80% similarity.

We then compared StanDep to other methods that use thresholds to identify core reactions *(see Methods)*. The number of reactions in the StanDep-based core reaction lists was comparable to that of the localT2

method (Fig. S5). Though the core reaction lists for global thresholds are smaller, the similarity among cell lines is higher than StanDep and other methods (global, 0.64; localT2, 0.54, localT1, 0.34; and StanDep, 0.59). We then analyzed the reactions that differed between StanDep and the other methods. We found that StanDep favors reactions in clusters with more moderately expressed (11-100 FPKM) housekeeping reactions (Clusters 1 and 2; Fig. 2C, Fig. S6). By contrast, localT2 and global approaches favored reactions in clusters with fewer housekeeping reactions and reactions with higher expression (>100 FPKM) (Clusters 9 and 10; Fig. 2C, Fig. S6B). Further, we compared Jaccard similarity (J) between StanDep and existing methods across all cell lines and found that StanDep agreed more with localT2 ($\mu_J = 0.80$; Fig. 2D) compared to localT1 ($\mu_J = 0.61$; Fig. S7A) and global ($\mu_J = 0.57$; Fig. S7B). In addition, across methods, we found the core reaction lists for the same cell line to be more similar than that from a different cell line (black circles; Fig. 2D, Fig. S7). Altogether, the results suggest that StanDep seeds core reaction lists with housekeeping reactions that are not captured by existing methods, while showing largest similarity to the localT2 approach.

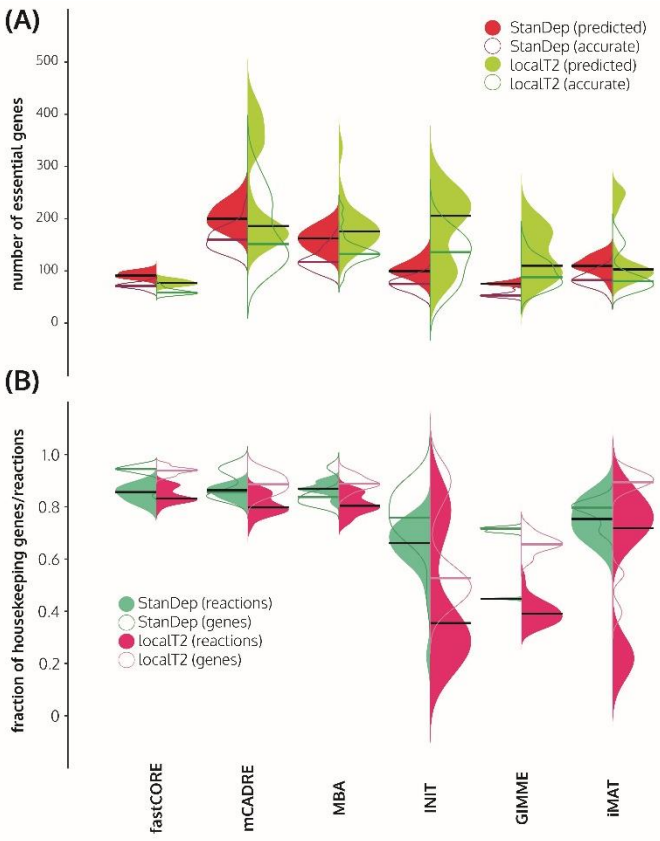

Figure 3: Validation of (A) predictions of gene essentiality with CRISPR gene essentiality data from DepMap, and (B) model content with existing list of housekeeping genes and reactions. (A) Distribution of number of essential genes predicted (filled) and accurately predicted (unfilled) in models built using StanDep (red) and localT2 (green). (B) Distribution of fraction of housekeeping genes (unfilled) and reactions (filled) present in models built using StanDep (cyan) and localT2 (pink). Constrained models and unconstrained models predicted high and similar coverage for housekeeping reactions, but slightly lower and bimodal coverage for housekeeping genes, while predictive power for MBA-like methods was higher than the other three methods. Constrained models can predict more essential genes than unconstrained models. mCADRE accurately predicted largest number of essential genes.

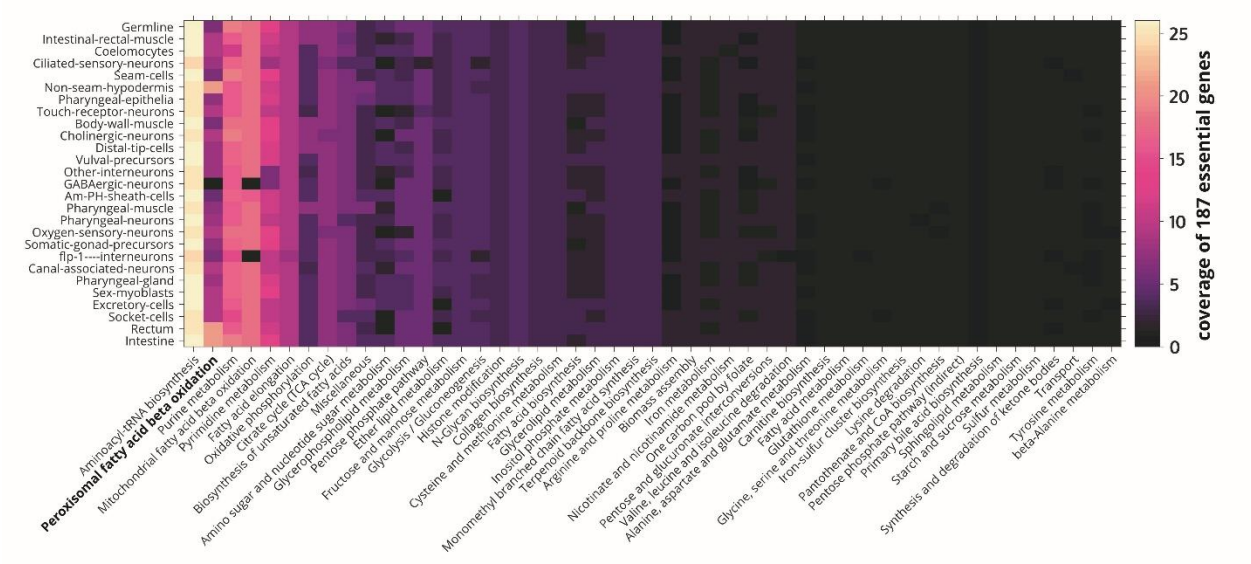

Figure 4: Cell types-specific distribution of 187 essential genes across pathways in StanDep-derived mCADRE models. Comparison with randomly permuted gene labels for presence of essential genes in the cell type models is presented in Fig S22. Large fraction of essential genes predicted are present in all cell types.

Models extracted using StanDep accurately predict essential genes

Using StanDep, we built hundreds of models by varying 4 model uptake/secretion constraint types (Jain *et al.*, 2012) and 6 model extraction methods (MEMs) (Agren *et al.*, 2012; Becker and Palsson, 2008; Jerby *et al.*, 2010; Vlassis *et al.*, 2014; Wang *et al.*, 2012; Zur *et al.*, 2010). We found that models built using similar MEMs clustered together (Fig. S8A). However, the presence/absence of constraints made little difference to the overall content of the models (Fig. S9A).

One characteristic of an effective thresholding method would be to generate a list of core reactions that highly overlaps and contains the same number of reactions than the model, suggesting independence from the extraction method; a quality we here describe as self-consistency. We then use a Jaccard dissimilarity test to compare the core reaction lists for different cell types generated by StanDep and localT2 to the reactions present in models generated using six extraction methods . For models extracted using fastCORE and mCADRE, localT2 and StanDep led to models of similar size (Fig. S10). However, core reactions identified using StanDep were more self-consistent than those identified using localT2 (Fig. S11B). Thus, models built using StanDep-derived core reactions had fewer unsupported reactions compared to localT2. Please see supplementary text for more information on model comparisons.

A second attractive capability of a thresholding method would be to extract models that can accurately predict context-specific essential genes t (i.e. different cancer cell lines). To evaluate our StanDep-based NCI-60 models, we looked at the capacity of StanDep-derived models to predict previously reported essential genes (Aguirre *et al.*, 2016; Doench *et al.*, 2016; Meyers *et al.*, 2017). StanDep-mCADRE models predicted essential genes more accurately (Wilcoxon, p-value = 7.57×10^{-4}, Fig. 3A); while StanDep-fastCORE models predicted a higher number of essential genes (p-value = 5.36×10^{-6}) than their

localT2 counterparts. Further, we also found that the prediction capacity (i.e. total number of predictions generated; left-tailed F-test, maximum p-value = 3.66×10^{-4} for MBA) and accuracy (maximum p-value = 0.0396 for MBA) of localT2-derived models had higher variations than StanDep-derived models. Thus, both metrics (prediction capacity and accuracy) suggest that StanDep achieves better precision than localT2.

A third sought-after aspect of a thresholding method would be to be able to extract models that effectively capture housekeeping genes and housekeeping reactions as defined by Eisenberg and Levanon (Eisenberg and Levanon, 2013). Though the presence of housekeeping genes in localT2 and StanDep-derived models was similar, StanDep-derived models included more housekeeping reactions than localT2-derived models (Fig. 3B) for mCADRE (Wilcoxon, p-value = 8.25×10^{-5}), MBA (p-value = 1.49×10^{-5}), and GIMME (p-value = 1.56×10^{-5}). Further, we found the ratio of housekeeping genes to housekeeping reactions captured in StanDep models was closer to 1 than localT2 models for all MEMs except fastCORE (Wilcoxon, maximum p-value = 0.0114 for INIT). A difference in the number of housekeeping genes and housekeeping reactions is not surprising because each gene can contribute several reactions. Nevertheless, the better match between genes and reactions suggest that StanDep facilitates better enzyme usage across housekeeping functions. Altogether, the analyses suggest that StanDep-derived models extracted using MBA-like MEMs (Robaina Estevez and Nikoloski, 2014) achieve better precision without sacrificing accuracy.

Reconstruction of worm cell types reveal distinct cell-type biology

Existing thresholding methods rely on a single cutoff for all genes; however, such thresholds are often chosen somewhat arbitrarily. StanDep avoids this by calculating thresholds from within the data. To demonstrate the general applicability of StanDep, we built cell-type specific models of a wild-type animal using a published GEM (Yilmaz and Walhout, 2016). In absence of exometabolomics data, we chose to build models using fastCORE and mCADRE as they performed better than other MEMs (Fig. S13). As our dataset, we used published whole-body single cell RNA-Seq data for *C. elegans* (Cao *et al.*, 2017). As with the NCI-60 cell lines, we found that the StanDep-derived models displayed higher reaction similarity within the same cell type ($\mu_J = 0.82$) across extraction methods (Fig. S14) than cell type models ($\mu_J = 0.70$), and the reaction content differed more than gene content ($\mu_J = 0.91$) among same cell type models. However, higher difference in reaction content than gene content is not surprising as our analysis of ratio of number of housekeeping genes to housekeeping reactions suggested different enzyme usage among fastCORE and mCADRE (Fig. 3). Nevertheless, given the high similarity in gene and reaction content across the two MEMs, our results suggest that when using StanDep-derived models have lower influence from extraction methods than the cell types.

We then tested the ability of the StandDep-derived cell type-specific models of *C. elegans* to predict 187 essential genes identified in a whole-body RNAi screen (Kamath *et al.*, 2003) which were also present in the *C. elegans* GEM (Yilmaz and Walhout, 2016). For multicellular organisms, there are two levels of essentiality, organismal and cellular; thus, a gene might be essential for a cell type but not for the whole organism and vice versa. Therefore, due to the lack of availability of cell type-specific essential genes, we

tested the presence of these genes in our models. Interestingly, for both mCADRE and fastCORE cell type-specific models, we found that over 80% of animal-level essential genes were present in all models and had over 99% coverage across models of each MEMs (Fig. 4, Fig. S15). However, this is not surprising given that these are metabolic essential genes belonging to core cellular processes such as glycolysis, aminoacyl-tRNA biosynthesis, oxidative phosphorylation, citrate cycle, pentose phosphate pathway, amino acid metabolism etc. Thus, the majority of these genes should be expected to be present across all cell types.

Interestingly, we found that reactions for β-oxidation, the second most represented pathway in animal-level essential genes, are differentially localized across neuron models; for e.g. peroxisomal oxidase and acyltransferase activities are present in ciliated-sensory, touch-receptor, cholinergic, and canal-associated neurons but not in other-interneurons, oxygen-sensory and pharyngeal neurons. Recently, a study showed that peroxisomal β-oxidation is required in ciliated-sensory neurons for pheromone-induced dauer development (Park and Paik, 2017). In addition to enrichment of peroxisomal β-oxidation in neuronal cell types, we also found it was enriched in intestine, hypodermis, and rectum. Fatty acid β-oxidation is known to be important in the intestine and hypodermis (Park and Paik, 2017). This consistency with the cell-type specific models suggests that the models contain the distinct biology of different cell types despite their high inclusion rate amongst the animal-level essential genes. Supporting this notion, the presence of the 900 nonessential genes from Kamath et al. (Kamath *et al.*, 2003) varied more widely among different models (Fig. S15B). Further supporting this notion, the reaction content similarity among models clustered similar cell types together (Fig. S14A). Thus, StanDep combined with either mCADRE or fastCORE yields models that capture phenotypically relevant aspects of cell type-specific metabolism.

Discussion

Several methods exist for extracting context-specific models by integrating transcriptomic data into the genome scale models (Wang *et al.*, 2012; Agren *et al.*, 2012; Vlassis *et al.*, 2014; Zur *et al.*, 2010; Jerby *et al.*, 2010; Becker and Palsson, 2008). Identification of core reaction lists from omics data (which we refer here to as preprocessing), commonly precedes the application of these methods. For preprocessing one applies a threshold to the transcriptomics dataset at a gene- or enzyme-level to decide if the gene or enzyme is expressed to sufficient levels to be considered active. Previous work identified thresholding as the most influential parameter impacting model content in context-specific models (Opdam *et al.*, 2017; Richelle, Joshi, *et al.*, 2018). However, it remains unclear how such thresholds be decided. Here, we present StanDep, a novel heuristic approach for determining thresholds. We made hundreds of models using StanDep and evaluated them against models constructed using an existing thresholding method (Richelle, Joshi, *et al.*, 2018).

We established that existing thresholding methods poorly capture housekeeping genes (Fig. S1). In part, because by only considering the absolute magnitude of expression, they disregard that housekeeping genes are also characterized by low variability (Eisenberg and Levanon, 2013, 2003). Further, fewer housekeeping genes are expressed at high levels (Fig. S2). Therefore, capturing housekeeping genes in all

models, built using existing thresholding methods, requires setting low threshold values. However, setting low thresholds leads to larger core reaction lists (Richelle, Joshi, *et al.*, 2018), which may result in the ultimately higher false positives. StanDep overcomes this problem by making standard deviation, as a measure of variability, part of the equation to calculate thresholds.

Recently, a multi-organism study identified that different subsets of genes have different patterns of transcriptional variation across organs and species (Breschi *et al.*, 2016). Breschi et al. also showed that variability in expression of genes may have had significant impact on interpreting transcriptomes by treating all genes uniformly (Barbosa-Morais *et al.*, 2012; Breschi *et al.*, 2016). Model extraction methods typically build context-specific models by placing the connectivity delineated by the genome scale metabolic network on top of the core reaction lists. However, genome scale metabolic networks are known to have alternate pathways (Zanghellini *et al.*, 2013), and an extraction method may favor selection of an incorrect one when lacking sufficient evidence within the core reaction list. Thus, treating all genes the same to derive core reaction lists could significantly impact the final model. StanDep avoids treating all enzymes the same by clustering genes based on the pattern of their own expression across multiple contexts and calculating a threshold for each cluster that is more likely to reflect the real ON/OFF status of any reaction across contexts. Importantly, StanDep-derived core reaction lists increased coverage of housekeeping reactions (Fig. S4) and added fewer unsupported reactions to the models when used with MEMs (Fig. S11). Further, StanDep, compared to localT2, showed increased precision without compromising accuracy across all extraction methods (Fig. 3A). Consistent with localT2-derived models, StanDep-derived mCADRE models performed well in predicting essential and housekeeping genes (Fig. 3). This is consistent with our previously published findings across four cell lines (Opdam *et al.*, 2017). Thus, StanDep can be adapted to work with existing context-specific extraction methods and used with diverse datasets.

We benchmarked StanDep using NCI-60 cancer cell lines. However, the metabolism of cancer cells may be seen as examples of outliers within the spectrum of normal physiology, and hence easier to capture. Therefore, in this study, we also created models of human tissues (Fig. S17) and *C. elegans* cell types (Fig. 4) and show that they have large coverage of housekeeping and essential genes. Furthermore, we showed that StanDep can capture relevant aspects of cell type-specific metabolism, such as the presence of peroxisomal fatty acid β-oxidation in neurons, intestine, and hypodermis. However, when building animal-level models, these cell type-specific models will need to be stitched together in the context of the whole animal to understand presence of pathways in non-canonical cell types or tissues (for e.g. peroxisomal β-oxidation in rectum). Thus, novel algorithms are needed for stitching together such cell type-specific models.

The comparisons made in this study between the model predictions and the experimentally-defined essential genes have broader implications on how different cellular processes may be correlated across different layers of regulation. Our results suggest that variability is globally determined and maintained across levels of regulation to define context-specific cellular function; otherwise, these transcriptomics-based models could not accurately predict phenotypes. In other words, as cells move through different processes such as transcription, translation, post-translational modification, and so on, the biology of the cell would preserve the minimally necessary variability to maintain cell-specific functionalities

(phenotypes) and homeostasis, and that variability converges on the same players being regulated in similar way (i.e. in general, the most transcribed genes would not correspond to the fastest degraded proteins). Thus, maintenance of a specific pattern of variability across levels of regulation would be a strategy by which the genome reaches context-specific function and homeostasis. In addition to perturbation through genetic, chemical, physical or other means, this notion of restricted variability across levels of regulation, may be of consequence when cells in a multicellular organism differentiate and new specialized variabilities must be dynamically coordinated across levels of regulation.

Acknowledgements

This work was supported by the NIGMS (grant no. R35 GM119850), a Lilly Innovation Fellows Award to CJ, and funding from the Keck Foundation.

METHODS

Datasets used

For this study, we used 2 sets of gene expression data for NCI60 cancer cell lines (Reinhold *et al.*, 2012; Klijn *et al.*, 2015), HPA gene expression data for tissues (Uhlen *et al.*, 2015), and gene expression data for *C. elegans* cell types (Cao *et al.*, 2017). For validating our models, we used a list of housekeeping genes (Eisenberg and Levanon, 2013), CRISPR data for 20 NCI60 cancer cell lines (Aguirre *et al.*, 2016; Doench *et al.*, 2016; Meyers *et al.*, 2017), and RNAi phenotypic data for *C. elegans* (Kamath *et al.*, 2003). For further details on data extraction, please see supplementary methods.

Data processing

We selected genes that are part of the human metabolic reconstruction, Recon 2.2 (Swainston *et al.*, 2016). This included for NCI60 data, 1416 genes; for HPA data, 1661 genes out of 1673 genes in Recon 2.2; and for *C. elegans* cell type data, 1248 genes were part of the global expression datasets and *C. elegans* GEM (Yilmaz and Walhout, 2016). We then converted gene expression values into enzyme expression values using gene mapping. Gene mapping involved extracting gene-protein-reaction (GPR) relationships from the model and calculating enzyme expression. The extraction of GPR was done using the COBRA function, *GPRparser.m*. For enzymes that have only one subunit, the value of enzyme expression is same as the value of gene expression. For multimeric enzymes, these relationships share an “AND” relationship; thus, the minimum value amongst genes part of the enzymes were set as enzyme expression value. The assumption for multimeric enzymes was that gene with lowest expression will govern the amount of functional enzyme expressed. It should be noted that we did not resolve OR relationships representing isoenzymes and allowed all functional enzymes to be represented in the

enzyme expression dataset. The enzyme expression data spanned 1325 enzymes (4133 reactions) for NCI60 data, 1792 enzymes for HPA data, and 2533 enzymes for *C. elegans* data.

Hierarchical Clustering

Clustering distribution patterns of gene expression

We $\log_{10}$-transformed the calculated enzyme expression dataset and counted the number of samples expressed with each bin width. Bin width were set based on the $\log_{10}$-transformed minimum and maximum enzyme expression values. This resulted in a matrix with rows representing each enzyme, columns representing bins, the value within the matrix representing number of samples from the dataset which were expressed within each bin range. We then performed hierarchical clustering with *Euclidean* distance metric and *complete* linkage metric to cluster genes based on distribution pattern of gene expression. We also show the comparison between using other distance and linkage methods (Fig. S18, S20, Supplementary Results).

Deciding number of clusters

Clustering in our work is used as a tool to divide genes into categories based on distribution patterns of their expression across different conditions. These clusters are then responsible for generating their own threshold. Therefore, number of clusters were determined such that all pathway is enriched in at least one cluster. The pathways were extracted from the GEMs, Recon 2.2 (for NCI60 and human tissues) and iCEL1273 (for *C. elegans* cell types). Only pathways which contained at least one gene-associated reaction were considered. For the NCI60 Klijn et al. dataset, we used 26 and 25 clusters for enzyme expression and gene expression respectively; for NCI60 CellMiner dataset, we used 21 clusters; for HPA dataset, we used 19 clusters; and for *C. elegans*, we used 14 and 18 clusters for enzyme expression and gene expression data respectively. We also show the comparison of choosing different number of clusters (Fig. S19; Supplementary Results).

Clustering core reaction sets or models

For analysis of models, we calculated Jaccard similarity of reaction content across different models which were part of any given analysis. We then performed hierarchical clustering to see how tissues are grouped. Hierarchical clustering was performed with the *Euclidean* distance metric and *complete* linkage metric. The interpretation of clustering Jaccard similarity is that models that are most similar to each other are likely to be equally far from other models.

Pathway enrichment

Pathway enrichment was performed by calculating hypergeometric p-value (p-value < 0.05) for the number of enzymes belonging to a given pathway present within a given cluster. Pathway association of an enzyme was calculated based on pathway association of the reactions being catalyzed by an enzyme.

Identification of Core Reactions

StanDep

StanDep applies thresholds specific to each cluster of genes. In the StanDep threshold formulation, we included two terms: (i) standard deviation, and (ii) mean term. Fine-tuned expression level of genes is represented as the Standard deviation term; and is dependent on the difference between standard deviation of the cluster and the dataset. Lower standard deviation favors the selection of enzymes in all contexts while higher standard deviation term reflects context-specificity of the enzymes. The mean term, interpretation of second assumption, is dependent on the magnitude of the expression of enzymes in that cluster. In both cases, we used the difference between cluster and overall data to address inconsequential variations that maybe occurring in expression. The standard deviation is always positive but logarithmic mean may be negative and sometime be even quite large. Therefore, we introduced normalization to make the standard deviation term and mean term at par. The threshold for each cluster is given by the following equations:

$$\Theta_c = (\theta_c - min(\theta_c)) * 100/max(\theta_c - min(\theta_c));\ \Theta_c \in [0,100] \quad \textbf{(1)}$$

$$\theta_c = f(\sigma_c) + g(\mu_c); \quad \textbf{(2)}$$

$$f(\sigma_c) = (\sigma_c - \Delta)/max(\sigma_c - \Delta); \quad \textbf{(3)}$$

$$g(\mu_c) = -(\mu_c - \boldsymbol{M}); \quad \textbf{(4)}$$

In the above set of equations, $\boldsymbol{\Theta_c}$ is the processed threshold value for a given cluster *c*; $\boldsymbol{\theta_c}$ is the raw value of threshold for cluster *c*; $\boldsymbol{\sigma_c}$ is the standard deviation of the cluster *c*; $\boldsymbol{\Delta}$ is the standard deviation of the dataset; $\boldsymbol{\mu_c}$ is the mean of the cluster *c*; and $\boldsymbol{M}$ is the mean of the dataset. The equation is derived by penalizing cluster-specific thresholds based on: (i) how low the cluster mean is compared to the mean of the dataset; (ii) how far the standard deviation of the cluster is from the standard deviation of the dataset. The final normalization was done to ensure that the clusters-specific thresholds are between 0 and 100. The $\boldsymbol{\Theta}$ is the top percentile value of the cluster-specific data above which an enzyme in that cluster in a given context is qualified active. If the value of $\boldsymbol{\Theta_c}$ is 100, we set the threshold value of the cluster as the mean of the data.

The current published literature on the below thresholding methods does not address how the threshold values should be derived. Therefore, we used some of the most commonly used percentile values in previously published studies (Richelle, Joshi, *et al.*, 2018; Opdam *et al.*, 2017).

Other thresholding methods

We used three of the existing thresholding methods: (i) global, (ii) localT1, and (iii) localT2. The implementation for each of them was same as in a previous study(Richelle, Joshi, *et al.*, 2018; Richelle, Chiang, *et al.*, 2018). However, they have also been described in detail in supplementary methods.

Constraining Pre-extraction Models and Model reduction

Exometabolomic constraints

Exometabolomic data of the NCI60 cell line were obtained from previous work (Jain *et al.*, 2012) and further processed as previously described (Opdam *et al.*, 2017). After processing, we added 23 new demand reactions, wherein each reaction is secreting a different metabolite. These were added to reflect the experimental observations by Jain et al. The biomass reaction was changed to one that contains precursor molecules from the one that contains macromolecules like DNA, RNA, protein, lipids, carbohydrate, and others. The replacement of the biomass reaction was done to all the models. The global lower and upper bounds for all reactions except biomass and ATP demand were set to -1000 and 1000 respectively. The lower bounds of the biomass reaction and ATP demand were constrained to relatively small values of the order of 1e-2 and 1.833 mmol gDW^{-1} h^{-1} (Kilburn *et al.*, 1969) respectively. The cell line specific constraints on 78 demand and exchange reactions were applied on the modified Recon 2.2, followed by making flux consistent constrained genome-scale models for each of the cell lines. This was done by identifying and removing flux-inconsistent reactions using *fastcc.m* in COBRA Toolbox. The flux tolerance was always set to 1e-8.

No constraints

To make unconstrained models, we did not apply exometabolomic constraints but only applied constraints on lower bounds of biomass and ATP demand reaction as described above. The global lower and upper bounds were set to -1000 and 1000 respectively. This was followed by identifying and removing flux inconsistent reactions. The flux tolerance was always set to 1e-8.

Semi constrained

To make semi-constrained models, we applied directional constraints on demand and exchange reactions of each cell line, applied constraints on lower bounds of biomass and ATP demand as described above. The global lower and upper bounds were set to -1000 and 1000 respectively. This was followed by identifying and removing flux-inconsistent reactions. The flux tolerance was always set to 1e-8.

Relaxed constraints

To make relaxed models, we constrained the direction of flow to 10 mmol gDW^{-1} h^{-1} on demand and exchange reactions as suggested by exometabolomic data. The order of magnitude of original constraints on these reactions was 1e-3 to 1e-6. The global lower and upper bounds were set to -1000 and 1000 respectively. This was followed by identifying and removing flux-inconsistent reactions. The flux tolerance was always set to 1e-8.

Implementation with model extraction methods (MEMs)

In this study, we compared the models derived using localT2 and StanDep. This section describes the extraction of StanDep-derived models by tailoring each of the MEMs. Models derived using localT2 were not constructed in this study, rather we extracted those models from a previous study (Richelle, Chiang, *et al.*, 2018). Therefore, for implementation of each of the extraction methods with these thresholding methods, please see the methods for that study.

To construct models using 6 of the extraction methods these inputs were common to all: (i) a flux-consistent Recon 2.2 genome-scale model was used, and (ii) epsilon, a.k.a. flux tolerance, was set to 1e-8. Inputs specific to a given MEM are described below.

FASTCORE

To construct models using FASTCORE (Vlassis *et al.*, 2014), we used *fastcore.m* in the COBRA toolbox. Other inputs needed for the algorithm are requires core reaction lists. Please see above on how we identified them. The biomass reaction was manually added to the core reaction list.

iMAT

To construct models using iMAT (Zur *et al.*, 2010), we used *iMAT.m* in the COBRA toolbox. Other inputs needed for the algorithm are: (i) core reactions (i.e., list of reactions identified to be active, including the biomass reaction) and (ii) non-core reactions, which are not part of core reactions (reactions not associated to a gene were not included in non-core reactions).

MBA

To construct models using MBA (Jerby *et al.*, 2010), we used *MBA.m* in the COBRA toolbox. Other inputs needed for the algorithm are: (i) high expression set, list of reactions which are highly expressed and (ii) medium expression set, list of reactions which are moderately expressed. We generated 10% interval around threshold for each cluster. We defined high expression set as the list of reactions catalyzed by enzymes which are above 110% of the threshold value, and medium expression set as the list of reactions catalyzed by enzymes which are between 90% and 110% of the threshold value. For instances where a reaction was present in both high and medium expression set, we interpreted it as at least enzyme

associated to the reaction being able to express at high levels. Thus, we put these reactions in high expression sets. The biomass reaction was given the highest value.

mCADRE

To construct models using mCADRE (Wang *et al.*, 2012), we used *mCADRE.m* in the COBRA toolbox. Other inputs needed for the algorithm are: (i) ubiquity score (i.e., how often a reaction is expressed across samples of the same context); (ii) confidence scores quantifying level of evidence for a reaction to be present in the model; (iii) protected reactions; and (iv) since we did not protect any reactions, we set the functionality check to 0. To calculate ubiquity score ($\boldsymbol{U_{c,i}}$), we calculated threshold distances ($\boldsymbol{D_{c,i}}$), here defined as distance of a given enzyme expression ($\boldsymbol{x_{i,c}}$) in the context *i* from the threshold ($\boldsymbol{\Theta_c}$) of the cluster *c* where the enzyme belongs. The threshold distances and ubiquity scores were calculated using the eqs (5-7).

$$D_{c,i} = x_{i,c} - \Theta_c \quad \textbf{(5)}$$

$$if\ D_{c,i} > 0;\ U_{c,i} = 1 \quad \textbf{(6)}$$

$$if\ D_{c,i} < 0;\ U_{c,i} = 1 - (D_{c,i}/min(D_{c,i})) \quad \textbf{(7)}$$

We used the ubiquity score to quantify how often an enzyme is expressed in samples of the same context. For isoenzymatic reactions, the reaction ubiquity score was set to the enzyme with maximum ubiquity score. For reactions which do not have an associated gene, the ubiquity score was set to -1. Since, we did not have confidence scores, we assigned a confidence of 0 to all reactions, as suggested in COBRA toolbox tutorial for mCADRE. However, we also tried using our list of core reactions as a binary vector specifying whether a reaction is in the core set and if it did not have any effect of the final model. The biomass reaction was manually assigned a ubiquity score of 1. The confidence score of 1 is associated with transcriptomics evidence and our metric ubiquity score already has this information.

INIT

To construct models using INIT (Agren *et al.*, 2012), we used *INIT.m* in the COBRA toolbox. Other inputs needed for the algorithm are reaction weights, varying between -1 and 1. To calculate enzyme weights, we calculated the threshold distance for each enzyme as described previously, without

normalizing. Weights for all reactions catalyzed by an enzyme were same as the enzyme weight. Here, we used a different normalizing scheme. We scaled our threshold distances to a maximum threshold distance for any of the enzymes within the data. For isoenzymatic reactions, the weights of each enzyme were added. We set the weights for non-gene associated reactions to 0. The biomass reaction was manually assigned a weight of 1.

GIMME

To construct models using GIMME (Becker and Palsson, 2008), used *GIMME.m* in the COBRA toolbox. Other inputs needed for the algorithm are: (i) a reaction expression vector representing gene expression values associated with the reactions; and (ii) a threshold determining whether reaction expression is considered active. We calculated the reaction expression vector in the same way as we calculated enzyme weights for INIT. However, unlike INIT weights, we set the expression level of reactions that do not have a gene association to 0. The thresholds were set to 1. The biomass reaction was given a value of 1.

Gene essentiality in NCI60

NCI60 data

To test our essentiality predictions of NCI60 models with CRISPR screen data, we downloaded pooled CRISPR knockout screen data from DepMap.org (Doench *et al.*, 2016; Aguirre *et al.*, 2016; Meyers *et al.*, 2017) for 20 NCI-60 cell lines. Essential genes were identified based on the CRISPR score. The CRISPR score was calculated as the ratio of abundance of single guide RNA (sgRNA) of a knock out after and before growth selection. A negative CRISPR score suggests a higher probability that the gene is essential. The accuracy was estimated using the percentage of predicted essential genes that have a negative score (Tobalina *et al.*, 2016). We then used 1-tailed Wilcoxan rank sum test to identify if the CRISPR scores for genes predicted to be essential in the metabolic model and CRISPR scores of genes predicted to be non-essential are coming from the same populations.

RNAi phenotypic data

To get the list of essential genes in *C. elegans*, we extracted genes that presented a *Nonv* or *Gro* RNAi phenotype. As described by the authors (Kamath *et al.*, 2003), *Nonv* phenotype refers to all phenotypic classes that result in lethality or sterility (1170 essential genes); and *Gro* refers to phenotypic classes that result in growth defects, slow post-embryonic growth or larval arrest (276 essential genes). Out of these, the iCEL1273 (Yilmaz and Walhout, 2016) model contained 187 genes. Similarly, we found 900 non-essential genes in iCEL1273.

REFERENCES

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**, e1002518.
Aguirre,A.J. *et al.* (2016) Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting. *Cancer Discov.*, **6**, 914–29.
Barbosa-Morais,N.L. *et al.* (2012) The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science (80-.).*, **338**, 1587–1593.
Becker,S.A. and Palsson,B.O. (2008) Context-specific metabolic networks are consistent with experiments. *PLoS Comput. Biol.*, **4**, e1000082.
Breschi,A. *et al.* (2016) Gene-specific patterns of expression variation across organs and species. *Genome Biol.*, **17**, 151.
Cao,J. *et al.* (2017) Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science*, **357**, 661–667.
Cimini,D. *et al.* (2009) Global transcriptional response of Saccharomyces cerevisiae to the deletion of SDH3. *BMC Syst. Biol.*, **3**, 17.
Creecy,J.P. and Conway,T. (2015) Quantitative bacterial transcriptomics with RNA-seq. *Curr. Opin. Microbiol.*, **23**, 133–40.
Doench,J.G. *et al.* (2016) Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.*, **34**, 184–191.
Eisenberg,E. and Levanon,E.Y. (2013) Human housekeeping genes, revisited. *Trends Genet.*, **29**, 569–574.
Eisenberg,E. and Levanon,E.Y. (2003) Human housekeeping genes are compact. *Trends Genet.*, **19**, 362–365.
Han,X. *et al.* (2018) Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell*, **172**, 1091–1107.e17.
Hyduke,D.R. *et al.* (2013) Analysis of omics data with genome-scale models of metabolism. *Mol. Biosyst.*, **9**, 167–74.
Jain,M. *et al.* (2012) Metabolite Profiling Identifies a Key Role for Glycine in Rapid Cancer Cell Proliferation. *Science (80-.).*, **336**, 1040–1044.
Jerby,L. *et al.* (2010) Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.*, **6**, 401.
Kamath,R.S. *et al.* (2003) Systematic functional analysis of the Caenorhabditis elegans genome using RNAi. *Nature*, **421**, 231–237.
Kilburn,D.G. *et al.* (1969) The Energetics of Mammalian Cell Growth. *J. Cell Sci.*, **4**.
Kim,M.K. and Lun,D.S. (2014) Methods for integration of transcriptomic data in genome-scale metabolic models. *Comput. Struct. Biotechnol. J.*, **11**, 59–65.
Klijn,C. *et al.* (2015) A comprehensive transcriptional portrait of human cancer cell lines. *Nat. Biotechnol.*, **33**, 306–312.
Lewis,N.E. *et al.* (2010) Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat. Biotechnol.*, **28**, 1279–1285.
Lewis,N.E. and Abdel-Haleem,A.M. (2013) The evolution of genome-scale models of cancer metabolism. *Front. Physiol.*, **4**, 237.
Li,J.J. and Biggin,M.D. (2015) Statistics requantitates the central dogma. *Science (80-.).*
Machado,D. and Herrgård,M. (2014) Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism. *PLoS Comput. Biol.*, **10**, e1003580.
Mardinoglu,A. *et al.* (2014) Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.*, **5**, 3083.
Meyers,R.M. *et al.* (2017) Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. *Nat. Genet.*, **49**, 1779–1784.
Opdam,S. *et al.* (2017) A Systematic Evaluation of Methods for Tailoring Genome-Scale Metabolic Models. *Cell Syst.*, **4**, 318–329.e6.

Park,S. and Paik,Y.-K. (2017) Genetic deficiency in neuronal peroxisomal fatty acid β-oxidation causes the interruption of dauer development in Caenorhabditis elegans. *Sci. Rep.*, **7**, 9358.
Reinhold,W.C. *et al.* (2012) CellMiner: A web-based suite of genomic and pharmacologic tools to explore transcript and drug patterns in the NCI-60 cell line set. *Cancer Res.*, **72**, 3499–3511.
Richelle,A., Joshi,C., *et al.* (2018) Assessing key decisions for transcriptomic data integration in biochemical networks. *bioRxiv*.
Richelle,A., Chiang,A.W.T., *et al.* (2018) Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *bioRxiv*, 384099.
Robaina Estevez,S. and Nikoloski,Z. (2014) Generalized framework for context-specific metabolic model extraction methods. *Front. Plant Sci.*, **5**, 491.
Shankavaram,U.T. *et al.* (2009) CellMiner: A relational database and query tool for the NCI-60 cancer cell lines. *BMC Genomics*.
Shen,Y. *et al.* (2010) Blueprint for antimicrobial hit discovery targeting metabolic networks. *Proc. Natl. Acad. Sci. U. S. A.*, **107**, 1082–7.
Swainston,N. *et al.* (2016) Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics*, **12**, 109.
Tobalina,L. *et al.* (2016) Assessment of FBA Based Gene Essentiality Analysis in Cancer with a Fast Context-Specific Network Reconstruction Method. *PLoS One*, **11**, e0154583.
Uhlen,M. *et al.* (2015) Tissue-based map of the human proteome. *Science (80-.).*, **347**, 1260419–1260419.
Vlassis,N. *et al.* (2014) Fast Reconstruction of Compact Context-Specific Metabolic Network Models. *PLoS Comput. Biol.*, **10**, e1003424.
Wang,Y. *et al.* (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.*, **6**, 153.
Winter,E.E. *et al.* (2004) Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. *Genome Res.*, **14**, 54–61.
Yilmaz,L.S. and Walhout,A.J.M. (2016) A Caenorhabditis elegans Genome-Scale Metabolic Network Model. *Cell Syst.*, **2**, 297–311.
Zanghellini,J. *et al.* (2013) Elementary flux modes in a nutshell: Properties, calculation and applications. *Biotechnol. J.*, **8**, 1009–1016.
Zur,H. *et al.* (2010) iMAT: an integrative metabolic analysis tool. *Bioinformatics*, **26**, 3140–3142.