IDENTIFYING CANCER SPECIFIC METABOLIC SIGNATURES USING CONSTRAINT-BASED MODELS

A. SCHULTZ[1], S. MEHTA[1], C.W. HU[1], F.W. HOFF[2], T.M. HORTON[3], S.M. KORNBLAU[2] and A.A. QUTUB[1*]

[1]*Department of Bioengineering, Rice University, Houston, Texas 77005, U.S.A*
[*]*E-mail: aminaq@rice.edu*

[2]*Department of Leukemia, The University of Texas M.D. Anderson Cancer Center, Houston, Texas 77030, U.S.A*

[3]*Department of Pediatrics, Baylor College of Medicine and Texas Children's Hospital, Houston, Texas 77030, U.S.A*

Cancer metabolism differs remarkably from the metabolism of healthy surrounding tissues, and it is extremely heterogeneous across cancer types. While these metabolic differences provide promising avenues for cancer treatments, much work remains to be done in understanding how metabolism is rewired in malignant tissues. To that end, constraint-based models provide a powerful computational tool for the study of metabolism at the genome scale. To generate meaningful predictions, however, these generalized human models must first be tailored for specific cell or tissue sub-types. Here we first present two improved algorithms for (1) the generation of these context-specific metabolic models based on omics data, and (2) Monte-Carlo sampling of the metabolic model flux space. By applying these methods to generate and analyze context-specific metabolic models of diverse solid cancer cell line data, and primary leukemia pediatric patient biopsies, we demonstrate how the methodology presented in this study can generate insights into the rewiring differences across solid tumors and blood cancers.

Keywords: Genome-scale metabolic reconstructions, constraint-based models, tissue-specific models, Flux Balance Analysis, cancer metabolism.

Introduction

Cancer tissues exhibits significant metabolic differences when compared to their healthy counterparts, such as the *Warburg effect*[1] and *glutamine addiction*.[2] In recent years it has been revealed that these metabolic transformations are largely driven by oncogenes and subdued by tumor suppressor genes.[3,4] This regulation suggests that cancer metabolism plays an important role in tumor progression, as opposed to being a consequence of the tumor microenvironment.[5] These findings have led to a renewed interest in the field of cancer metabolism,[6] with particular interest in exploiting metabolic differences as therapeutic targets.[7] Cancer metabolism, however, is also extremely heterogeneous across cancer types,[8] and treatments targeting metabolic pathways need to be carefully tailored to specific cancer phenotypes. Consequently, a better understanding of the metabolic differences across cancer sub-types, and between healthy and cancerous tissues will greatly assist the development of novel therapeutic strategies.[7,8]

Genome-Scale Models: To help elucidate the metabolic differences between cancer and healthy tissues, computational approaches can be extremely helpful. In particular, genome-scale models (GEMs) have proven extremely useful in studying human metabolism at the genome level,[9,10] with many studies dedicated specifically to cancer metabolism.[11–13] These

studies have, for example, identified glycosaminoglycans as a marker for clear cell renal cell carcinoma,[14] identified carnitine palmitoyltransferase 1 as a potential target for hepatocellular carcinoma,[15] and identified MLYCD as a potential target for leukemia and kidney cancer.[16]

GEMs are defined at the core by a *stoichiometric matrix* S, where each row corresponds to a metabolite, each column to a metabolic reaction, and each entry to the stoichiometric coefficient of that particular metabolite in that particular reaction.[17] For any given *stoichiometric matrix*, flux distribution column vectors (v) can be defined where each element v_i gives the metabolic flux (e.g. rate of metabolite conversion) through each reaction i. The matrix multiplication $S \cdot v = m$ then yields a vector m where each element m_j gives the rate of change of concentration of metabolite j given the reaction fluxes defined by v. A steady-state flux distribution is one where $S \cdot v = 0$. A more detailed description of the constraint-based model formulation is available in the *supplemental information*.

Metabolic Model Analysis: Although a wide array of methods have been developed to study GEMs,[18] many of them are dependent on an *objective function*, which is most often assumed to be cellular growth.[19] Mammalian cells, however, do not have a well established objective, and do not seek to optimize biomass production. One prominent unbiased and objective-independent method for GEM analysis, suited for the study of mammalian cells, is Monte-Carlo sampling (MCS). This method finds normally distributed steady-state flux distributions inside the solution space of $S \cdot v = 0$ defined by lower (lb) and upper (ub) reaction bounds, such that $lb_i \leq v_i \leq ub_i$. Valuable insight into the metabolic capabilities of the model in question can be obtained by analyzing how different MCS conditions (e.g. different lower and upper bounds) affect the sampled reaction flux values. This approach has been used, for example, to model the metabolic exchange between *M. tuberculosis* and human macrophages,[20] and between different cell types in the human brain;[21] to study aspirin resistance in platelet cells;[22] and to characterize metabolic differences between healthy and cancerous tissues.[23]

Mammals also have a complex and compartmentalized metabolism, where not every metabolic reaction takes place in all cells of the body. In order to generate predictions specific to different cell types, cancer categories or patients, generalized human GEMs then need to be tailored to specific contexts.[24] We recently introduced the Cost Optimization Reaction Dependency Assessment (*CORDA*) tissue-specific algorithm,[23] which builds tissue-specific metabolic models based on omics data and a generalized human metabolic reconstruction. The algorithm is based on a *dependency assessment* (DA), where reactions associated with little experimental evidence, called negative confidence reactions (NC), are assigned an arbitrarily high cost. This cost is then minimized while enforcing a small flux through medium (MC) or high (HC) confidence reactions (i.e. reactions with medium or considerable experimental evidence) in order to identify which NC reactions are beneficial for MC or HC reactions to carry flux. This DA is then used to build a tissue-specific model including all HC reactions and as many MC reactions as possible, while minimizing the inclusion of NC reactions. For additional details on the original algorithm we refer readers to the original *CORDA* publication.[23]

Need for New Analyses: MCS of large metabolic networks is computationally expensive, and static approaches are only feasible for extremely small networks.[25] For MCS of higher dimensional networks, the Artificially Centered Hit and Run (ACHR) algorithm[26] is most

frequently used. Given a set of points, or steady-state flux distributions, inside the solution space, ACHR calculates a center point as the average of all points, then moves each point i randomly along the directional vector defined by the trajectory between the center and another random point j. ACHR sampling of large networks can be extremely time consuming, however, and even small relative increments in computational efficiency can lead to fewer hours of computational time. Although alternatives to ACHR have been proposed, many of these methods are limited by sample distributions that are significantly different than ACHR outputs,[27–29] by their dependence on objective functions,[27] by long computational times,[30] or by lack of validation and parametrization in larger metabolic networks.[31]

Introduction of CORDA2 and mfACHR: Here we present two improved algorithms for the study of human GEMs. We first introduce an improved version of the *CORDA* algorithm to build tissue-specific metabolic models,[23] referred to here as *CORDA2*. *CORDA2* yields tissue models very similar to the ones given by the previous algorithm, but it is considerably faster than *CORDA* computationally. *CORDA2* is also noise-independent, thus providing unique model outputs for any given set of parameters, which facilitates the comparison of metabolic models across different modeling conditions (i.e. different cancer categories). We next introduce a new formulation of the ACHR algorithm,[26] referred to here as the matrix-form ACHR (mfACHR), which performs significantly faster than previous formulations.

Integrating the two new methods, we generate a panel of cell-line specific metabolic models using *CORDA2* and experimental data from the Human Protein Atlas[32] (HPA), and illustrate how flux samples generated using *mfACHR* can provide valuable insights into the metabolic profile of different cancer types, including pediatric leukemia. While we had previously shown that MCS of CORDA models can identify metabolic differences between healthy and cancerous tissues, here we show that this framework can also pinpoint metabolic differences between different cancer categories. The methods presented in this study provide significant advances in the generation and analysis of context-specific metabolic models.

Methods

Cost Optimization Reaction Dependency Assessment 2

In this work we present two modifications to *CORDA*, defining a new version of the algorithm referred to here as *CORDA2*. First, in the original algorithm, reversible reactions were split into forward and backward rates during every DA to ensure cost production regardless of directionality. That is, a reaction '$A \Leftrightarrow B$' was split into '$A \Rightarrow B + cost$' and '$B \Rightarrow A + cost$'. Since thousands of DAs are performed throughout the model building process, this modification was then repeated thousands of times during the algorithm. In *CORDA2*, this modification is performed at the beginning of the algorithm, and forward and backward rates are treated separately throughout the model building process, speeding the computational time. Furthermore, while in *CORDA* the reaction directionality in the tissue-model was imported from the generalized human reconstruction, *CORDA2* assigns directionality based on whether the forward, backward, or both reaction parts are included in the final tissue model.

Second, pathways with similar costs are captured in *CORDA* by adding a small amount of noise to reaction costs during every DA. This noise-driven approach leads to different

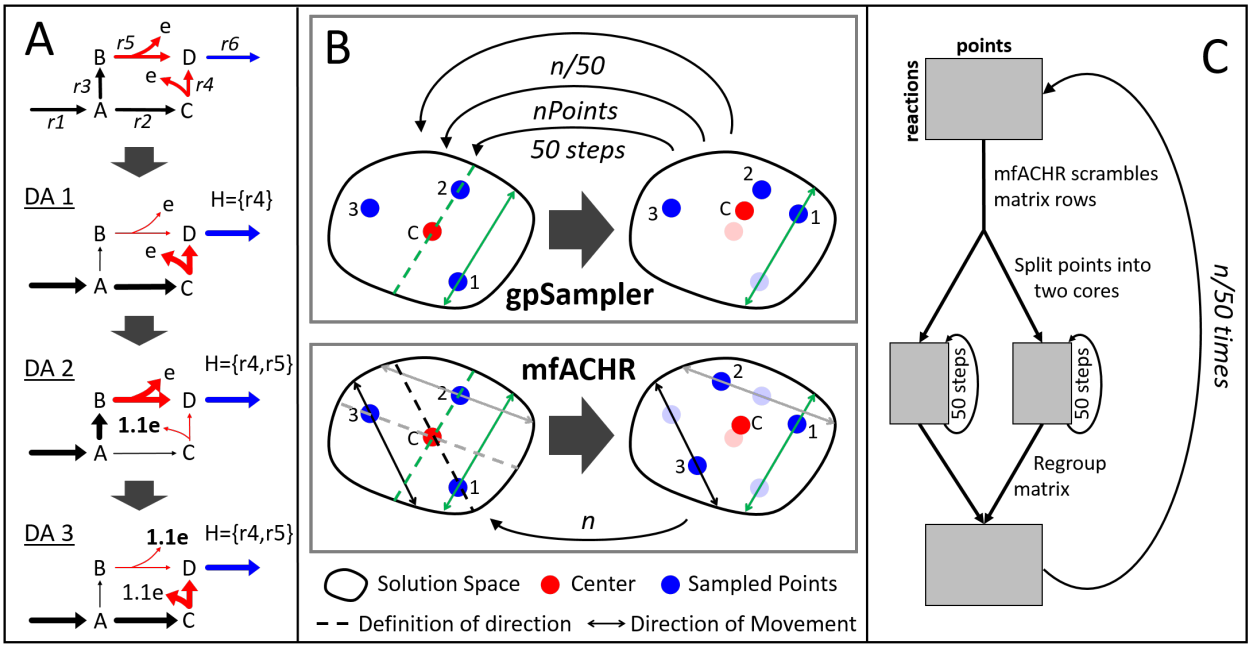

Fig. 1. **Representation of the CORDA2 and mfACHR algorithms**. (A) Identification of undesirable reactions (red) beneficial for the desirable reaction (blue) to carry flux through three DAs. Pathways taken during each DA are highlighted, and H represents the set of undesirable reactions taken up to that point. After an undesirable reaction is used, its cost (e) is increased. The process is repeated until H is unchanged. (B) gpSampler moves one point at a time, 50 steps at a time. The mfACHR algorithm identifies all possible directions of movement at once and moves all points simultaneously. Vectors defining the trajectory of movement, taken as the difference between j and the center point, and the corresponding path of movement of i are color-coded. (C) During parallelization of the MCS process, the matrix of sampled points is divided into 2 cores, which are sampled for 50 steps, then re-combined.

reconstructions after every run of the algorithm, and it is not guaranteed to include every alternative pathway. This approach is also inefficient since the same pathway can be sampled multiple times. In *CORDA2*, only undesirable reactions are assigned an arbitrarily high cost (while in *CORDA* all reactions received a basal cost value). This cost is then minimized during the DA, and the high cost reactions used are saved in a set H. The cost associated with the reactions in H is then increased, and the DA is performed again (**Fig. 1A**). This process is repeated iteratively until H is unchanged. This way, once a pathway is used, its cost is increased and another pathway with similar but now slightly lower cost is identified in the next DA. Additional details of the *CORDA2* formulation, as well as the MATLAB code for its implementation, can be found in the *supplemental information*.

Matrix-Form Artificially Centered Hit and Run

One of the most widely implemented ACHR formalisms is *gpSampler*.[33] *GpSampler* starts by moving a given point 50 steps as described by the ACHR algorithm, then repeats the process for each point being sampled. This whole process is then repeated $\frac{n}{50}$ times for a total of n ACHR steps (**Fig. 1B**). Here we propose a slightly different ACHR formulation, termed matrix-form ACHR (mfACHR). In *mfACHR*, all possible directions of movement are first

calculated as the directional vectors defined by each sampled point and the center (dashed lines in **Fig. 1B**). These trajectories are then randomly assigned to each point, and each point is moved randomly along its assigned direction of movement (solid lines in **Fig. 1B**) within the bounds of the solution space. This whole process is repeated a total of n times for a desired number of steps. Both *gpSampler* and *mfACHR* can also be implemented in multiple cores. For that, the points being sampled are first divided into i groups, i being the number of cores used. Each group is then assigned to a core and mixed for 50 steps. All points are then re-combined and the process is repeated $\frac{n}{50}$ times for a total of n steps (**Fig. 1C**).

Cancer Cell Proteomics and Model Generation

Cell line gene and protein expression data were obtained from the HPA[32] in order to build the cell-line specific models. Gene expression data was measured using RNA-seq and protein expression was measured by immunohistochemistry using an extensive library of well validated antibodies. Forty-four models were generated using gene expression data and fifty-two models were generated using the proteomics data. Protein expression was available for 523 (35.0%) gene products, and gene expression data was available for 1,474 (98.7%) of the 1,494 unique genes in the generalized human reconstruction Recon1.[34] All gene and protein expression values were categorized into not detected, low/medium, and high expression in line with threshold values from the HPA, then used to categorize reaction confidence values used in the *CORDA2* algorithm. Following the reconstruction all models were sampled using *mfACHR*. Details of how these models were generated and sampled can be found in the *supplemental information*. For additional details on how the dataset was collected we refer readers to the HPA.[32]

Leukemia Patient Samples: Pediatric leukemia data was obtained from bone marrow biopsies of 95 acute myeloid leukemia (AML), 57 B-cell acute lymphoblastic leukemia (B-ALL), and 16 T-cell acute lymphoblastic leukemia (T-ALL) pediatric patients, and were collected at the Texas Children's Hospital. Protein expression level was measured using reverse phase protein array (RPPA) using 194 strictly validated antibodies.[35] Additional information on the pediatric leukemia data is available in the *supplemental information*.

Results and Discussion

Results of our study demonstrate the robustness of the *CORDA2* and *mfACHR* methods, and their utility in analyzing diverse cell line and primary leukemia cancer metabolism. A summary of the *CORDA2* and *mfACHR* validation is provided below, while a complete description of the algorithm validation and analysis is provided in the *supplemental information*.

CORDA2 Validation

In order to validate the *CORDA2* algorithm, outputs of this formulation were compared to 108 tissue-specific metabolic models generated using *CORDA* and similar model parameters (e.g. same dataset and overlapping algorithm parameters). Overall, at least 99.7% of MC reactions, 88.9% of NC reactions, and 93% of unclassified reactions included in each of the previous 108 models are also included in the *CORDA2* model, showing significant overlap between the

output of both algorithms. Furthermore, *CORDA2* was approximately 2.5 times faster than *CORDA* when the later was performed with five DAs for every reaction tested. Although performing fewer DAs in *CORDA* led to computational times comparable to *CORDA2*, the reconstructions returned in that case are not as comprehensive. In the original *CORDA* publication, models reconstructed using one DA were on average 2.3% smaller than models built using multiple DAs. The *CORDA2* algorithm also showed very similar results across multiple metabolic tests when compared to the previous formulation. This analysis shows that *CORDA2* yields models similar to *CORDA* in composition and behavior, while being faster and noise independent.

mfACHR Validation

To assess the performance of *mfACHR* when compared to *gpSampler*, flux distributions and convergence speed of both formulations were compared for three different metabolic models: a red blood cell (RBC) model,[36] a platelet model,[22] and the generalized human reconstruction Recon1.[34] These models have 453, 1,008, and 2,473 active reactions respectively, and were sampled for $3{\cdot}10^4$, $7{\cdot}10^4$, $3{\cdot}10^5$ steps respectively. As an initial step in this validation, MCS outputs of four algorithm formulations (*mfACHR*, *mfACHR* parallel, *gpSampler*, and *gpSampler* parallel) were compared, and all four formulations were shown to converge to similar steady states (*supplemental information*).

Next, convergence speed was assessed by computational time and number of algorithm steps. Convergence based on number of steps was measured as the percentage of reactions at any given point with a Kullback-Leibler divergence (KLD) of sampled flux values below 0.05 of the final distribution. KLD represents the expected logarithmic difference between two probability distributions, and it has been previously used with a similarity threshold of 0.05 to compare sets of sampled flux distributions in metabolic models.[31] The four tested formula-

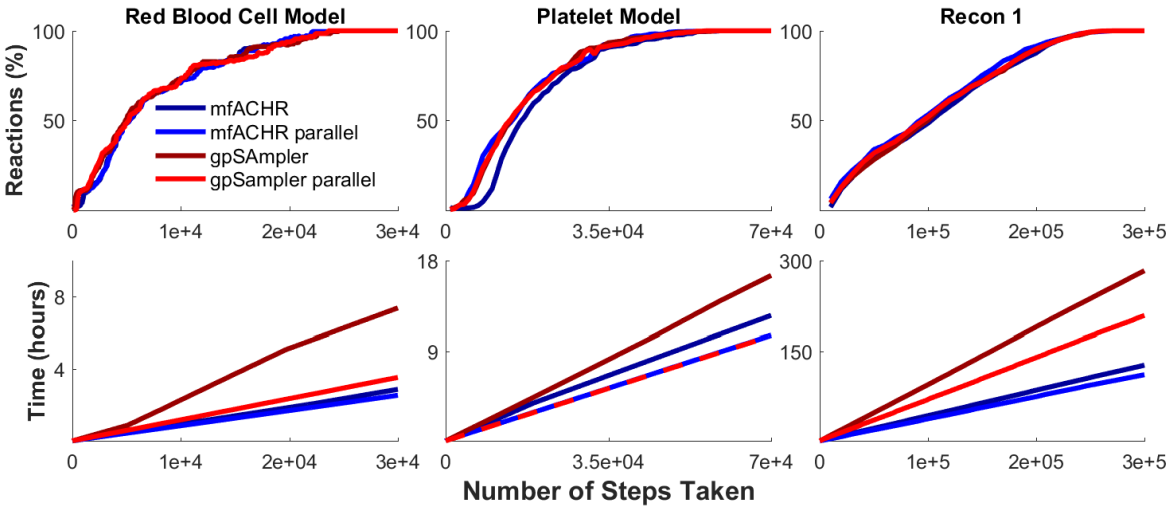

Fig. 2. **Conversion speed of *mfACHR* and *gpSampler*.** (Top) Percentage of reactions in the model with a KLD below 0.05 when compared to the final set of sampled points. (Bottom) Computational running time per number of algorithm steps.

tions showed nearly identical conversion curves when considering the number of steps taken (**Fig. 2**). When considering computational times, *mfACHR* performed significantly better than *gpSampler* when both methods were performed without parallelization. When considering parallelization, *mfACHR* showed very similar computational times in the platelet model, slightly better times in the RBC model, and significantly better times in Recon1. Differences in computational time can be partially attributed to the fact that matrix operations performed by *mfACHR* are automatically parallelized in MATLAB, while the *for* loops performed by *gpSampler* are not. This allows for *mfACHR* to perform significantly faster than *gpSampler* even when the latter is performed with parallelization, and explains the low relative increase in efficiency when explicit parallelization is implemented in *mfACHR*. Overall, *mfACHR* showed consistently faster computational times when compared to *gpSampler*, often in the order of hours, while converging at the same speed in terms of number of algorithm steps.

Cancer Cell Models

Following the validation of both algorithms, a series of cell-line specific models were generated using *CORDA2* and sampled using *mfACHR*, as described in the methods section. Twenty-six of the cancer metabolic models were combined into four tissue categories as presented in the HPA: myeloid, lymphoid, brain, and female reproductive system (FRS) cancer cell lines. These cancer types were chosen since they had the most number of cell lines. We then identified metabolic reactions that have significantly different sampled flux distributions between the four cancer categories (**Fig. 3**). MCS of *CORDA* models previously highlighted metabolic differences between healthy and cancerous tissues.[23] That is, using *CORDA* we correlated high sampled flux values with metabolic pathways known to take place in healthy or cancerous phenotypes. Analogously, in this study we demonstrate that *mfACHR* sampling of *CORDA2* models generated using HPA expression data can also highlight metabolic characteristics between different cancer categories. These characteristics include:

Brain tumors produce high levels of triglyceride: Lipid synthesis is an important factor for cancer survival and progression, and it has been previously suggested as a therapeutic target.[37–40] However, while most cancer types divert fatty-acids predominantly towards the production of phospholipids, not triglycerides,[39,41] glioma cells have been shown to synthesize triacylglycerol at high rates for membrane complex lipids.[42,43] Glioma cells, as well as healthy astrocytes and neurons, can also produce fatty acids from ketone-bodies,[44,45] a metabolic characteristic of brain cells which can further explain the high rate of fatty acid production in glioma cells. In the MCS results presented here, brain tumors present a significantly higher flux through glycerol-3-phosphate acyltransferase (**Fig. 3**) and 1-acylglycerol-3-phosphate O-acyltransferase, enzymes responsible for triacylglycerol synthesis.

Brain and lymphoid tumors have highly active glutamine metabolism: Glutamine plays an essential role in cancer metabolism,[46,47] and different tumors have been shown to utilize glutamine differently.[47] Brain tumors, in particular, have been shown to accumulate glutamine both *in vitro* and *in vivo*.[48,49] Glutamine metabolism has also been shown to play an important role in lymphoid tissues.[50] The role of this pathway in breast cancer, on the other hand, is not well defined, since basal but not luminal breast cancer cells show glutamine-

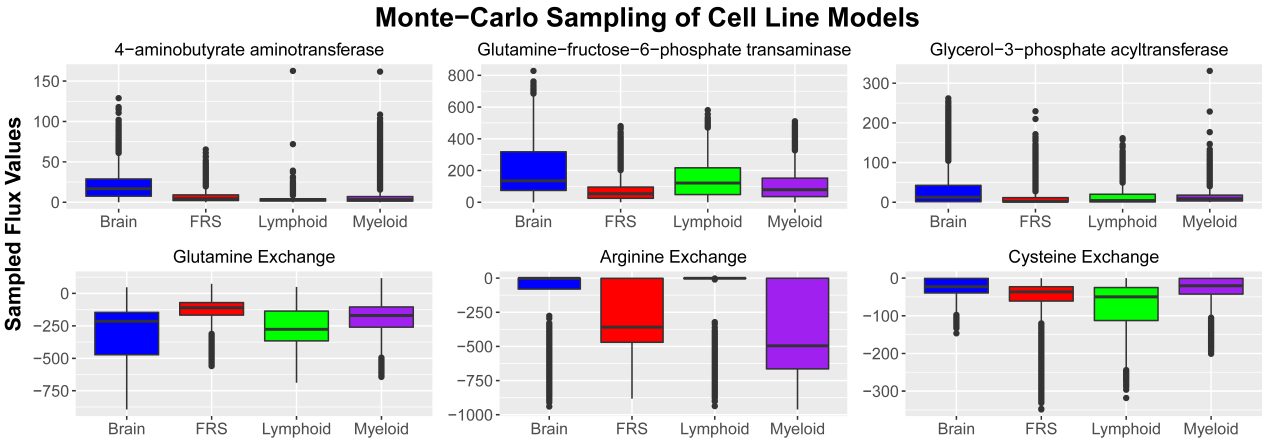

Fig. 3. **MCS results**. Sampled flux values for six different reactions across four model categories. Boxplots represent combined flux values for a particular reaction in all models in that cancer category. For exchange reactions, negative values represent uptake of the particular metabolite, while positive values represent secretion. Colored boxes represent values within the interquartile range (IQR), ranging from the 25th to the 75th percentile. Horizontal line represented the median value (50th percentile), and vertical lines indicate values within 1.5 IQR of the 25th and 75th percentiles. Outliers are represented by dots.

dependence.[51] In the results presented here, brain and lymphoid cell lines show high levels of glutamine uptake, while cell lines of the FRS show relatively low levels (**Fig. 3**).

Lymphoid tissues are cysteine dependent: While cysteine is not considered an essential amino-acid, lymphoid tumors have been shown to contain much lower levels of cystathionase, the last enzyme in the cysteine production pathway, when compared to healthy lymphoid tissues, and are dependent on cysteine for growth.[52] Targeting cysteine transporters has also been shown to selectively target lymphoma cells,[53] and cysteine uptake has been associate with malignant progression in lymphoma cells.[54] In this study, lymphoid models presented much higher levels of cysteine uptake (**Fig. 3**).

Tumors show different levels of arginine dependence: Different types of cancer respond differently to arginine deprivation.[55] A study performed on 26 healthy and cancerous cell lines found that tumor cells are much more sensitive to arginine deprivation than healthy cells.[56] Furthermore, while premyelocytic and lymphoblastic leukaemia cell lines die in about two days of arginine deprivation, cell lines of the FRS died largely in three to four days, and glioma cell lines died in four to five days.[56] Interestingly, levels of arginine dependence presented in the study by Scott et. al.[56] correspond to sampled flux values of arginine uptake in the present study. Myeloid cancers, the most arginine dependent, were predicted to uptake the largest amounts of arginine, followed by models of the FRS, then brain tumors, the least arginine dependent. Acute myeloid leukemia tumors have also been shown to be dependent on arginine for proliferation.[57]

Brain tumors were also predicted to have higher fluxes through the enzyme glutamine-fructose-6-phosphate transaminase (GF6PTA) (**Fig 3**), the rate limiting step in the hexosamines synthesis pathway (HSP), a nutrient sensor pathway.[58,59] When excess nutrients such as glucose and free fatty-acids are available, the HSP prevents cells from uptaking excess amounts from the bloodstream.[60] Furthermore, overweight and obese patients, which have

excess amounts of nutrients in the bloodstream, are at an overall increased risk of mortality due to cancer.[61] Interestingly, sampled flux values through the HSP presented here are anti-correlated with the increase in risk of mortality in cancer patients. According to a study of over 57,000 cancer patients, obese patients with brain tumors have a modest increase in mortality compared to non-obese glioma patients, while patients with cancer of the FRS have a high increase in risk, and patients with Non-Hodgkins lymphoma, multiple myeloma, and leukemia have a medium increase.[62] Accordingly, brain tumor models in this study present high GF6PTA flux values, while tumors of the FRS present low fluxes, and lymphoid and myeloid tumors present intermediate values (**Fig. 3**). One possible explanation for this correlation is that higher fluxes through the HSP can prevent cells from uptaking excess amounts of nutrients, which in turn leads to a lower relative increase in malignancy. Further work should help elucidate these observations in context.[63,64]

Sampled flux values also predict a high flux through the enzyme 4-aminobutyrate aminotransferase in brain cancer cells. This result is expected since this enzyme is responsible for GABA production, a pathway highly active in brain tissues. In brain cancer cells, however, this enzyme can help produce acetyl-CoA for energy production, since larger amounts of nutrients are diverted away from glycolysis and into the HSP. A diagram of this proposed mechanism is presented in **Fig 4A**.

Primary pediatric leukemia models: We next analyzed sampled flux values in three different types of leukemia blood sample models (AML, T-ALL, and B-ALL) to clinical pro-

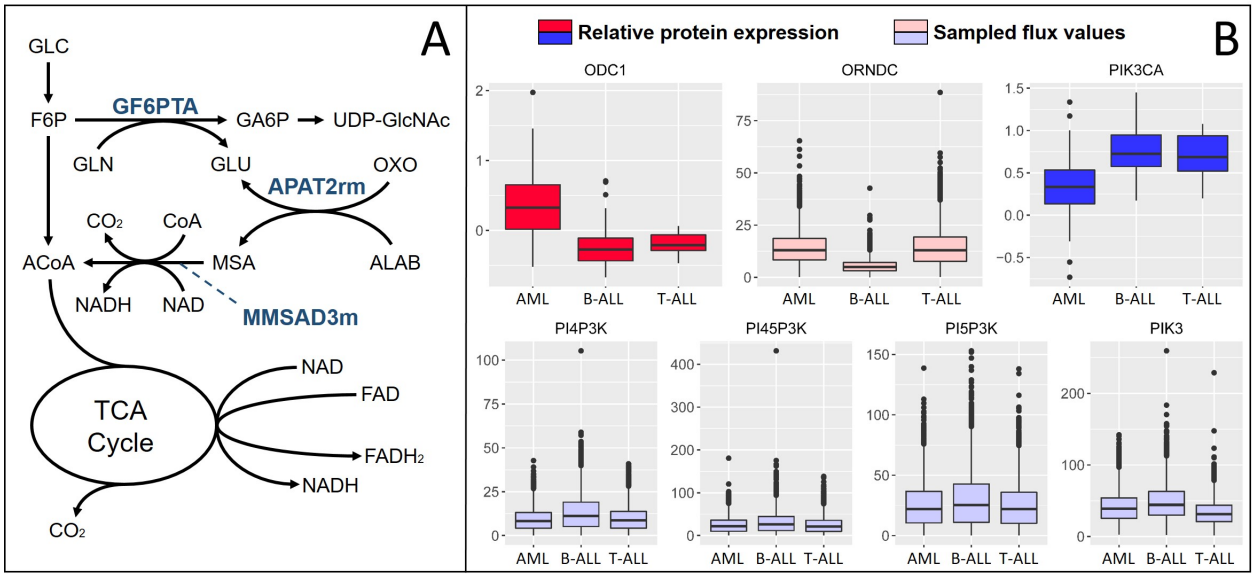

Fig. 4. **Model Predictions**. (A) Pathways with increased activity in brain tumors. Metabolites are glucose (GLC), fructose-6-phosphate (F6P), acetyl-CoA (ACoA), glutamine (GLN), glutamate (GLU), glucosamine-6-phosphate (GA6P), Uridine diphosphate N-acetylglucosamine (UDP-GlcNAc), oxoglutarate (OXO), beta-alanine (ALAB), and malonate semialdehyde (MSA). (B) Relative protein expression and sampled flux values for proteins differentially expressed between AML and ALL pediatric patients. ODC1 participates in the reaction Ornithine Decarboxylase (ORNDC), and PIK3CA participates in reactions PI4P3K, PI45P3K, PI5P3K, and PIK3. All reactions are labeled as in the BiGG database.[65]

teomics data collected from 168 pediatric leukemia patients as described in the methods section. Seven proteins were present both in the leukemia blood sample models and the clinical dataset, of which two were significantly differentially expressed between AML and ALL patients. The relative protein expression of these two proteins, along with the sampled flux values of reactions associated with these proteins, are presented in **Fig 4B**.

Sampled flux values follow trends that correlate with protein expression in both the B-ALL and AML models. That is, while AML patients show significantly higher expression levels of ODC1, the AML model showed significantly higher fluxes through Ornithine Decarboxylase (ORNDC), an ODC1 participating reaction, when compared to the B-ALL model. Likewise, while AML patients showed significantly lower expression of PIK3CA, the AML model also showed significantly lower sampled flux values through the PIK3CA reactions (**Fig 4B**). Sampled flux values between the AML and T-ALL model did not seem to match the differential protein expression, however. One possible explanation for this is the fact that there were considerably fewer T-ALL patients in the clinical dataset, and fewer T-ALL samples were used to generate the proteomics data used in the models building process (2 compared to 3 B-ALL and 4 AML). For instance, in the HPA, T-ALL ODC1 and PIK3CA protein scores are in between B-ALL and AML values, as opposed to much closer to B-ALL values like we see in the pediatric clinical data. This first example application to integrating RPPA leukemia data with metabolic pathway analysis demonstrates how *CORDA2* and *mfACHR* can also be used to analyze clinical data and provide insight into patient-specific metabolic behaviors.

Conclusion

This work illustrates how Monte-Carlo sampling of metabolic models generated using *CORDA2* can generate valuable predictions about context specific cancer metabolism. In applying these new optimized methods to different cancer systems, we show how this work goes beyond the identification of metabolic differences between healthy and cancerous tissues. It identifies differences in metabolism between different cancer types, paving the way to patient-specific metabolic models of cancer. In sum, the *CORDA2* platform elucidates metabolic differences across cancers and provides valuable knowledge of context-specific metabolic behavior that can help guide future directed cancer therapies.

Acknowledgments

This work was funded by NSF grant numbers 1354390 and 1150645, and NIH grant numbers GM106027 and CA164024.

Author Contributions

AS and AAQ developed the computational methods. AS, SM, and AAQ applied and validated the methods. TMH and SMK collected the AML and ALL clinical data. FWH and CWH processed the clinical data.

Supplemental Information

Supplemental files are available at www.qutublab.org/psb

References

1. O. Warburg *et al.*, *Science* **123**, 309 (1956).
2. H. Eagle, *Science* **122**, 501 (1955).
3. R. J. DeBerardinis, N. Sayed, D. Ditsworth and C. B. Thompson, *Current opinion in genetics & development* **18**, 54 (2008).
4. R. D. Michalek and J. C. Rathmell, *Immunological reviews* **236**, 190 (2010).
5. C. Munoz-Pinedo, N. El Mjiyad and J. Ricci, *Cell death & disease* **3**, p. e248 (2012).
6. R. A. Cairns, I. S. Harris and T. W. Mak, *Nature Reviews Cancer* **11**, 85 (2011).
7. M. G. Vander Heiden, *Nature reviews Drug discovery* **10**, 671 (2011).
8. J. R. Cantor and D. M. Sabatini, *Cancer discovery* **2**, 881 (2012).
9. A. Bordbar and B. O. Palsson, *Journal of internal medicine* **271**, 131 (2012).
10. A. Mardinoglu and J. Nielsen, *Journal of internal medicine* **271**, 142 (2012).
11. K. Yizhak, B. Chaneton, E. Gottlieb and E. Ruppin, *Molecular systems biology* **11**, p. 817 (2015).
12. I. Goldstein, K. Yizhak, S. Madar, N. Goldfinger, E. Ruppin and V. Rotter, *Cancer Metab* **1**, 10 (2013).
13. C. Frezza, L. Zheng, O. Folger, K. N. Rajagopalan, E. D. MacKenzie, L. Jerby, M. Micaroni, B. Chaneton, J. Adam, A. Hedley *et al.*, *Nature* **477**, 225 (2011).
14. F. Gatto, N. Volpi, H. Nilsson, I. Nookaew, M. Maruzzo, A. Roma, M. E. Johansson, U. Stierner, S. Lundstam, U. Basso *et al.*, *Cell reports* **15**, 1822 (2016).
15. R. Agren, A. Mardinoglu, A. Asplund, C. Kampf, M. Uhlen and J. Nielsen, *Molecular systems biology* **10**, p. 721 (2014).
16. K. Yizhak, E. Gaude, S. Le Dévédec, Y. Y. Waldman, G. Y. Stein, B. van de Water, C. Frezza and E. Ruppin, *Elife* **3**, p. e03641 (2014).
17. J. D. Orth, I. Thiele and B. Ø. Palsson, *Nature biotechnology* **28**, 245 (2010).
18. N. E. Lewis, H. Nagarajan and B. O. Palsson, *Nature Reviews Microbiology* **10**, 291 (2012).
19. A. M. Feist and B. O. Palsson, *Current opinion in microbiology* **13**, 344 (2010).
20. A. Bordbar, N. E. Lewis, J. Schellenberger, B. Ø. Palsson and N. Jamshidi, *Molecular systems biology* **6**, p. 422 (2010).
21. N. E. Lewis, G. Schramm, A. Bordbar, J. Schellenberger, M. P. Andersen, J. K. Cheng, N. Patel, A. Yee, R. A. Lewis, R. Eils *et al.*, *Nature biotechnology* **28**, 1279 (2010).
22. A. Thomas, S. Rahmanian, A. Bordbar, B. Ø. Palsson and N. Jamshidi, *Scientific reports* **4** (2014).
23. A. Schultz and A. A. Qutub, *PLoS Comput Biol* **12**, p. e1004808 (2016).
24. S. R. Estévez and Z. Nikoloski, *Front. Plant Sci* **5**, 10 (2014).
25. N. D. Price, J. Schellenberger and B. O. Palsson, *Biophysical journal* **87**, 2172 (2004).
26. D. E. Kaufman and R. L. Smith, *Operations Research* **46**, 84 (1998).
27. N. Chaudhary, K. Tøndel, J. Puchałka, V. A. M. dos Santos and R. Bhatnagar, *Molecular BioSystems* (2016).
28. W. Megchelenbrink, M. Huynen and E. Marchiori, *PloS one* **9**, p. e86587 (2014).
29. S. Bordel, R. Agren and J. Nielsen, *PLoS Comput Biol* **6**, p. e1000859 (2010).
30. P. A. Saa and L. K. Nielsen, *Bioinformatics* , p. btw132 (2016).
31. D. De Martino, M. Mori and V. Parisi, *PloS one* **10**, p. e0122670 (2015).
32. M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund *et al.*, *Science* **347**, p. 1260419 (2015).
33. J. Schellenberger, R. Que, R. M. Fleming, I. Thiele, J. D. Orth, A. M. Feist, D. C. Zielinski, A. Bordbar, N. E. Lewis, S. Rahmanian *et al.*, *Nature protocols* **6**, 1290 (2011).
34. N. C. Duarte, S. A. Becker, N. Jamshidi, I. Thiele, M. L. Mo, T. D. Vo, R. Srivas and B. Ø. Palsson, *Proceedings of the National Academy of Sciences* **104**, 1777 (2007).

35. T. M. Horton, Y. Qiu, G. Jenkins and S. M. Kornblau, *Blood* **124**, 3784 (2014).
36. A. Bordbar, N. Jamshidi and B. O. Palsson, *BMC systems biology* **5**, p. 110 (2011).
37. T. Mashima, H. Seimiya and T. Tsuruo, *British journal of cancer* **100**, 1369 (2009).
38. J. A. Menendez and R. Lupu, *Nature Reviews Cancer* **7**, 763 (2007).
39. F. P. Kuhajda, *Cancer research* **66**, 5977 (2006).
40. T. Migita, S. Okabe, K. Ikeda, S. Igarashi, S. Sugawara, A. Tomida, R. Taguchi, T. Soga and H. Seimiya, *The American journal of pathology* **182**, 1800 (2013).
41. F. P. Kuhajda, K. Jenner, F. D. Wood, R. A. Hennigar, L. B. Jacobs, J. D. Dick and G. R. Pasternack, *Proceedings of the National Academy of Sciences* **91**, 6379 (1994).
42. H. Cook and M. Spence, *Canadian Journal of Biochemistry and Cell Biology* **63**, 919 (1985).
43. H. W. Cook and M. W. Spence, *Biochimica et Biophysica Acta (BBA)-Lipids and Lipid Metabolism* **918**, 217 (1987).
44. M. S. Patel, J. J. Russell and H. Gershman, *Proceedings of the National Academy of Sciences* **78**, 7214 (1981).
45. L. M. Roeder, S. E. Poduslo and J. T. Tildon, *Journal of neuroscience research* **8**, 671 (1982).
46. R. J. DeBerardinis and T. Cheng, *Oncogene* **29**, 313 (2010).
47. C. T. Hensley, A. T. Wasti and R. J. DeBerardinis, *The Journal of clinical investigation* **123**, 3678 (2013).
48. E. A. Maher, I. Marin-Valencia, R. M. Bachoo, T. Mashimo, J. Raisanen, K. J. Hatanpaa, A. Jindal, F. M. Jeffrey, C. Choi, C. Madden *et al.*, *NMR in biomedicine* **25**, 1234 (2012).
49. I. Marin-Valencia, C. Yang, T. Mashimo, S. Cho, H. Baek, X.-L. Yang, K. N. Rajagopalan, M. Maddie, V. Vemireddy, Z. Zhao *et al.*, *Cell metabolism* **15**, 827 (2012).
50. M. Ardawi and E. Newsholme, Glutamine metabolism in lymphoid tissues, in *Glutamine metabolism in mammalian tissues*, (Springer, 1984) pp. 235–246.
51. H.-N. Kung, J. R. Marks and J.-T. Chi, *PLoS Genet* **7**, p. e1002229 (2011).
52. J. Iglehart, R. M. York, A. P. Modest, H. Lazarus and D. Livingston, *Journal of Biological Chemistry* **252**, 7184 (1977).
53. P. Gout, A. Buckley, C. Simms and N. Bruchovsky, *Leukemia (08876924)* **15** (2001).
54. P. Gout, Y. Kang, D. Buckley, N. Bruchovsky and A. Buckley, *Leukemia* **11**, 1329 (1997).
55. D. N. Wheatley, *Seminars in Cancer Biology* **15**, 247 (2005).
56. L. Scott, J. Lamb, S. Smith and D. Wheatley, *British journal of cancer* **83**, p. 800 (2000).
57. F. Mussai, S. Egan, J. Higginbotham-Jones, T. Perry, A. Beggs, E. Odintsova, J. Loke, G. Pratt, A. Lo, M. Ng *et al.*, *Blood* **125**, 2386 (2015).
58. M. G. Buse, *American Journal of Physiology-Endocrinology And Metabolism* **290**, E1 (2006).
59. M.-J. J. Pouwels, C. J. Tack, P. N. Span, A. J. Olthaar, C. Sweep, F. C. Huvers, J. A. Lutterman and A. R. Hermus, *The Journal of Clinical Endocrinology & Metabolism* **89**, 5132 (2004).
60. L. Wells, K. Vosseller and G. Hart, *Cellular and Molecular Life Sciences CMLS* **60**, 222 (2003).
61. E. E. Calle and R. Kaaks, *Nature Reviews Cancer* **4**, 579 (2004).
62. E. E. Calle, C. Rodriguez, K. Walker-Thurmond and M. J. Thun, *New England Journal of Medicine* **348**, 1625 (2003).
63. T. N. Sergentanis, G. Tsivgoulis, C. Perlepe, I. Ntanasis-Stathopoulos, I.-G. Tzanninis, I. N. Sergentanis and T. Psaltopoulou, *PloS one* **10**, p. e0136974 (2015).
64. T. Niedermaier, G. Behrens, D. Schmid, I. Schlecht, B. Fischer and M. F. Leitzmann, *Neurology* **85**, 1342 (2015).
65. Z. A. King, J. Lu, A. Dräger, P. Miller, S. Federowicz, J. A. Lerman, A. Ebrahim, B. O. Palsson and N. E. Lewis, *Nucleic acids research* **44**, D515 (2016).