

RESEARCH ARTICLE

Reconstruction of Tissue-Specific Metabolic Networks Using CORDA

André Schultz, Amina A. Qutub*

Department of Bioengineering, Rice University, Houston, Texas, United States of America

* aminaq@rice.edu



CrossMark
click for updates

OPEN ACCESS

Citation: Schultz A, Qutub AA (2016) Reconstruction of Tissue-Specific Metabolic Networks Using CORDA. *PLoS Comput Biol* 12(3): e1004808. doi:10.1371/journal.pcbi.1004808

Editor: Costas D. Maranas, The Pennsylvania State University, UNITED STATES

Received: September 17, 2015

Accepted: February 13, 2016

Published: March 4, 2016

Copyright: © 2016 Schultz, Qutub. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: Data are available in the Human Protein Atlas database (<http://www.proteinatlas.org/about/download>) for both healthy (http://www.proteinatlas.org/download/normal_tissue.csv.zip) and cancerous (<http://www.proteinatlas.org/download/cancer.csv.zip>) tissues. The data used in the mCADRE reconstruction are available in [S1 Table](#).

Funding: This work was funded by: National Science Foundation (http://www.nsf.gov/awardsearch/showAward?AWD_ID=1354390&HistoricalAwards=false) grant number 1354390 (AAQ), and National Science Foundation (<http://www.nsf.gov/awardsearch/showAward?AWD>

Abstract

Human metabolism involves thousands of reactions and metabolites. To interpret this complexity, computational modeling becomes an essential experimental tool. One of the most popular techniques to study human metabolism as a whole is genome scale modeling. A key challenge to applying genome scale modeling is identifying critical metabolic reactions across diverse human tissues. Here we introduce a novel algorithm called Cost Optimization Reaction Dependency Assessment (CORDA) to build genome scale models in a tissue-specific manner. CORDA performs more efficiently computationally, shows better agreement to experimental data, and displays better model functionality and capacity when compared to previous algorithms. CORDA also returns reaction associations that can greatly assist in any manual curation to be performed following the automated reconstruction process. Using CORDA, we developed a library of 76 healthy and 20 cancer tissue-specific reconstructions. These reconstructions identified which metabolic pathways are shared across diverse human tissues. Moreover, we identified changes in reactions and pathways that are differentially included and present different capacity profiles in cancer compared to healthy tissues, including up-regulation of folate metabolism, the down-regulation of thiamine metabolism, and tight regulation of oxidative phosphorylation.

Author Summary

Cellular metabolism is defined by a large, intricate network of thousands of components, and plays a fundamental role in many diseases. To study this network in its entirety, metabolic models have been built which encompass all known biochemical reactions in the human metabolism. However, since not all metabolic reactions take place in any given tissue, these generalized models need to be tailored to study specific cell types. Algorithms developed to date to perform this tailoring process have focused on keeping tissue-specific models as concise as possible. This approach, however, can remove essential reactions from the model and hamper subsequent analysis. Here we present CORDA, a tissue-specific building algorithm that yields concise, but not minimalistic, tissue-specific models. CORDA has many advantages over previous methods, including better agreement with experimental data and better model functionality. Using CORDA, we developed a library of 76 healthy and 20 cancer-specific models of metabolism, which we used to identify

([ID=1150645&HistoricalAwards=false](#)) grant number 1150645 (AAQ). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

similarities between healthy and cancerous tissues, as well as metabolic pathways that are unique to cancer. Results of this work provide a broadly applicable tool to model cell- and tissue-specific metabolism, while highlighting potential new pathway targets for cancer therapies.

Introduction

Genome-wide Metabolic Reconstructions (GEMs) computationally model the molecules and reactions responsible for metabolism in any given organism, and have been applied across a variety of fields including metabolic engineering and evolutionary analysis [1]. Computational methods developed to study GEMs [2] have generated novel hypotheses about the structure of metabolic networks in microorganisms, and helped elucidate gaps in our knowledge of metabolism [3, 4]. Since the publication of the comprehensive human metabolic reconstruction Recon1 [5], human GEMs have enabled the study of human metabolism at a genome level [6]. These studies include the prediction of novel metabolic functions [7], prediction of metabolic biomarkers for congenital genetic disorders [8, 9], context analysis of omics data [10–12], comparison between humans and other mammals through gene homolog mapping [13, 14], and prediction of suitable cancer drugs [15, 16] and drug targets [17–19].

A particularly prolific subfield of human GEMs is the development of tissue-specific reconstructions. Different groups of metabolic reactions occur in different cell types. Hence, numerous studies have been dedicated to generating tissue specific or cell specific models of metabolism [20, 21]. These tissue-specific reconstructions can be built by piecing together the model based on previously established biological evidence obtained by reviewing the literature [22–26], through the integration of omics data and computational methods in order to tailor generic, published human reconstructions [5, 9, 27–29] to the desired cell type [15, 16, 30–33], or through a combination of computational algorithms and manual curation [27, 28, 34–36].

Automated tissue-specific reconstruction algorithms developed to date can be broadly categorized into two groups [20]: “flux-dependent” and “pruning” methods. Flux dependent methods find an optimal flux distribution through the general reconstruction which contains the maximum number of high confidence reactions (i.e. reactions whose presence is supported by significant experimental data) [15, 31, 32, 37–39]. These algorithms have been successfully used to predict gene essentiality in cancer tissues [19, 33], cancer specific metabolic pathways [31], metabolic biomarkers for congenital genetic disorders [8, 9], and cancer specific anti-growth factors [15, 16]. One of the main advantages of flux-dependent methods is the fact that they predict a flux distribution along with the tissue-specific model [20]. While this characteristic can be desirable, it also renders flux-dependent reconstructions “snapshots” of the metabolic state defined by the data, as opposed to comprehensive, functional metabolic models [15, 20].

The second category of tissue-specific reconstruction methods are pruning algorithms, which include MBA [34], mCADRE [30] and fastCORE [40]. Models generated using these algorithms have been used to calculate metabolic flux values in hepatocytes [34], identify pathways specific to cancer [30], and predict cancer drug targets [17, 18]. These algorithms start with a core set of reactions, obtained through literature review or experimental data, and proceed by removing the remaining reactions in the generalized human reconstruction while maintaining functionality in the core set. In these algorithms, a tradeoff can be defined between maintaining the model as concise as possible and including all core reactions. That is, if a core reaction requires too many undesirable reactions to carry flux, the algorithm may remove this core reaction from the tissue model, a tradeoff referred to as *flexible core*.

There are two main advantages to defining a core set of reactions before performing the tissue-specific algorithm. The first advantage is the possible inclusion of multiple sources of data and biochemical information [20, 34]. The definition of the reactions core is left to the user's discretion, allowing for both the combination of data sources and the manual inclusion of reactions. Secondly, reactions with overwhelming evidence are always included in the final tissue model, since a non-flexible set of high confidence reactions can be defined [20]. This pruning approach then allows for the construction of comprehensive tissue models, containing all reactions that may be in a tissue's metabolism, as opposed to a snapshot of the metabolic state returned by the flux-dependent methods [15, 20].

Current pruning methods are also accompanied by two major limitations, however. First, the order in which reactions are removed from the model plays a major role in the final reconstruction. Second, similar to flux-dependent methods, current algorithms aim to keep the final tissue-reconstruction as concise as possible, an approach referred to as *parsimonious*. These algorithms aim to remove from the tissue-specific model all reactions for which experimental data is unresponsive or unavailable, such as reactions with low levels of gene expression or non-gene associated reactions. While a concise tissue-specific reconstruction is desirable, keeping the reconstruction as parsimonious as possible may lead to the removal of fundamental reactions and physiologically unlikely flux distributions. In Recon 1, for instance, oxygen and H₂O exchange reactions can be removed from the reconstruction with no effect on model functionality (Fig 1A). During simulations, however, these would be replaced by the uptake of the toxic metabolites superoxide anion and hydrogen peroxide respectively, leading to the prediction of physiologically inaccurate flux distributions (Fig 1A). The oxygen exchange reaction is in fact not present in the MBA and mCADRE liver reconstructions, and the water exchange reaction is not present in the mCADRE liver reconstruction.

Hence, in order to ensure our algorithm did not rely on alternative, physiologically unlikely pathways, and that it was independent of any ordering assignments, we chose to take an approach which was not parsimonious. Here we introduce a novel tissue-specific reconstruction algorithm based on Cost Optimization Reaction Dependency Assessment (CORDA). CORDA returns a concise, functional tissue-specific reconstruction, and features a flexible reactions core. CORDA does not depend on Flux Variability Analysis [41] or Mixed Integer Linear Programming (MILP) problems, but only on Flux Balance Analysis [42] (FBA), which is dependent on Linear Programming (LP). This characteristic renders CORDA considerably faster than previous, similar methods. Finally, the CORDA algorithm returns reaction associations that assist in any manual curation to be performed following the automated reconstruction process.

In line with previous studies [43], we apply CORDA to generate a library of 76 healthy and 20 cancer-specific metabolic reconstructions. These reconstructions enabled us to identify metabolic similarities amongst healthy tissues as well as key differences between healthy and cancerous tissues. Furthermore, by sampling the feasible solution space in cancer and healthy models, this library can be used to predict the up- and down-regulation of cancer-specific pathways in cancer metabolism.

Results

The CORDA algorithm

The CORDA algorithm is based on a novel approach to identify the dependency of desirable reactions (i.e. reactions with high experimental evidence) on undesirable reactions (i.e. reactions with no experimental evidence), a method referred to here as *dependency assessment*. In the *dependency assessment* approach, the metabolic network is modified in four ways (Fig 1B).

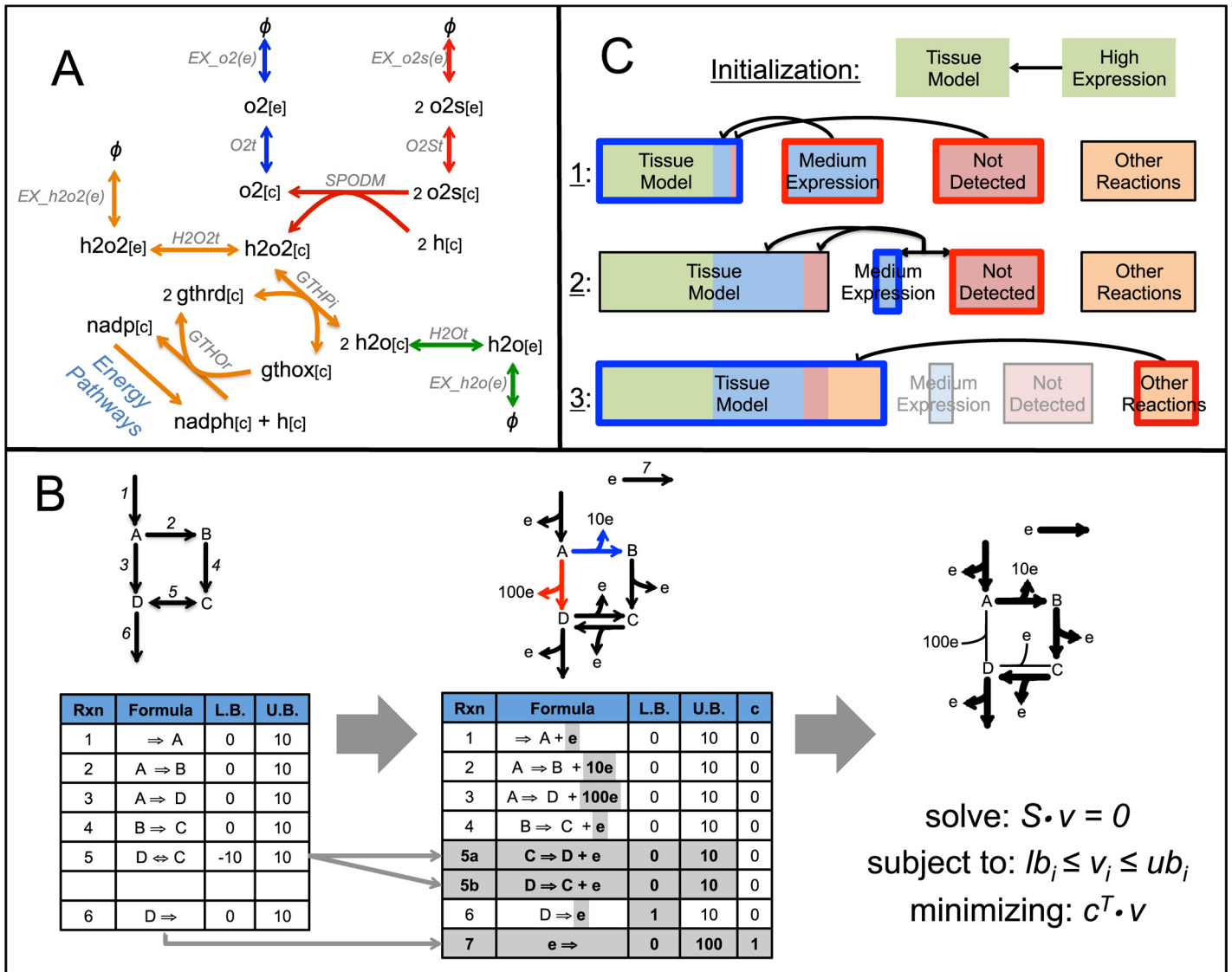


Fig 1. Overview of the CORDA algorithm. (A) Recon1 subnetwork involving water (h2o), oxygen (o2), hydrogen peroxide (h2o2) and superoxide anion (o2s) illustrating how standard oxygen (blue) and water (green) import pathways can be substituted by alternative, physiologically unlikely pathways (red and orange respectively). All metabolites and reactions are labeled as in Recon1. (B) Overview of the *dependency assessment* method. Each reaction in the reconstruction is associated with a specific cost through the addition of a pseudo-metabolite to the model. FBA is then performed while minimizing the cost production in order to identify high cost reactions which are favorable to the reaction being tested. (C) The CORDA tissue-specific algorithm. During each step, reaction groups being tested are outlined in blue, while reaction groups associated with a high cost are outlined in red.

doi:10.1371/journal.pcbi.1004808.g001

First, reversible reactions are split into forward and backward components. Second, a pseudo-metabolite is added as a product for every reaction in the model. At this point, undesirable reactions will carry a higher stoichiometric coefficient for this added metabolite, assigning these reactions a higher “cost”. Third, a reaction consuming this pseudo-metabolite is added to the model. Finally, a positive lower bound is set for the reaction being tested in order to force that reaction to carry flux. After modifying the network, FBA ([Materials and Methods](#)) is performed while minimizing the flux through the cost-consuming reaction ([Fig 1B](#)). The flux distribution returned will then use high cost, undesirable reactions only as necessary for the reaction being tested to carry flux. Throughout the manuscript, we will refer to high cost

reactions predicted to carry flux as *associated* with the reaction being tested. In order to identify pathways with the same cost (i.e. same number of undesirable reactions), multiple dependency assessment can be performed while adding a small amount of noise to the cost of each reaction.

Using this *dependency assessment*, we have developed the CORDA algorithm for the reconstruction of tissue-specific models (Fig 1C). CORDA takes as input the reactions in the generalized human reconstruction separated into high (HC), medium (MC), and negative (NC) confidence groups (see [Materials and Methods](#) section for a detailed description). All remaining reactions in the reconstruction (i.e. non gene associated reactions or reactions for which no data is available) are designated as others (OT). All HC reactions are included in the model, and the maximum number of MC reactions is included while minimizing the inclusion of NC reactions. While the definition of these four reaction groups are left to the user's discretion, here we categorize them according to proteomics data from the Human Protein Atlas (HPA) [44, 45] and a methodology used in previous studies [30, 32, 37] ([Materials and Methods](#)). To begin the algorithm, all HC reactions are moved into the tissue reconstruction (RE). In a first step, MC and NC reactions associated with each RE reaction (which are the same as the HC group at this point) are identified using the dependency assessment and moved into the RE group. In a second step, NC reactions associated with a high number of MC reactions are identified and moved into the tissue model, and all remaining NC reactions are blocked (upper and lower bounds set to zero). Next, all MC reactions still able to carry flux are also moved to the RE group. Finally, in the final step of the algorithm, all OT reactions associated with any RE reaction are moved to the RE group for the final tissue-specific model. A detailed description of the CORDA method, including detailed steps, algorithm parameters, and categorization of model reactions is available in the [Materials and Methods](#) section.

Validation of the CORDA algorithm

Parameter sensitivity analysis. As a first step in the validation of the CORDA algorithm, we generated 108 hepatocyte specific models using a wide range of algorithm parameters ([Materials and Methods](#)). The data used in this step, as well as the generalized human reconstruction, were the same used during the mCADRE liver reconstruction to allow for a fair and direct comparison between models. The 108 calculated models have an average of 1,857.3 (± 21.0) reactions, 1,760 (94.8%) of which are present in all models. Also, 98.3% of all MC reactions are present in all models, and 96.2% of the flexible MC and NC reactions core is unambiguously determined as either present or not present in all 108 models. A small number of NC reactions (20.25%) was also present in all reconstructions. Interestingly, the protein or expression evidence for half of those NC reactions has changed in the HPA since the publication of the mCADRE model, and they are no longer considered not detected. This demonstrates the ability of CORDA to include essential, significant NC reactions in the tissue-specific model, as well as the importance of a flexible core.

The main difference between the 108 calculated models stems from the number of OT reactions included. Reconstructions calculated using multiple dependency assessments, in order to identify pathways with the same cost, led to the inclusion of more OT reactions during the final step, defining a tradeoff between model size and robustness. No other parameter generated significantly different reconstructions, and all reconstructions demonstrated significant robustness to parameter values. More information on this analysis is available in the supplemental information ([S1 Text](#)).

Cross-validation. In order to assess CORDA's ability to include relevant reactions in the tissue-specific model, we performed an additional 100 cross validation reconstructions using randomly sampled subsets of each reaction group. For each reconstruction performed in this

analysis, a subset of 80% of each reaction group used to calculate the previous 108 reconstructions was sampled and used in the same reconstruction process. For each model generated, a hypergeometric p-value was calculated for each reaction group based on how many of the reactions randomly left out during the reconstruction process were ultimately included in the tissue model. The 100 p-values obtained were then combined using Fisher’s method, showing that the tissue-specific models generated here were enriched in a statistically significant manner with the HC and MC reactions left out of the reconstruction process, but not with the NC reactions. This analysis demonstrates the ability of CORDA to selectively include reactions with supportive experimental data. Further information on the cross-validation analysis is available in [S1 Text](#).

Comparison to previous models. As further validation, the CORDA algorithm was compared to two previously published methods: MBA [34] and mCADRE [30]. MBA and mCADRE were selected for comparison because they both contain a flexible core feature, and are both pruning algorithms, returning a comprehensive tissue-specific reconstruction like CORDA. Since both of these models were generated using the generalized human reconstruction Recon1 [5], here we use one of the 108 reconstructions generated during the parameter sensitivity analysis to allow for a direct comparison.

As a first step, the size and composition of the different hepatocyte specific reconstructions were compared ([Table 1](#)). We find that all three reconstructions have similar size and composition when considering reactions, metabolites, and genes. The mCADRE reconstruction has considerably fewer reactions, the difference stemming mostly from exchange reactions (i.e. the mCADRE reconstruction has 63 fewer reactions than the MBA reconstruction, but 50 fewer exchange reactions). The CORDA reconstruction contains only 6% more reactions than the mCADRE reconstruction and 2.4% more than MBA, which is surprising considering this algorithm does not take a parsimonious approach. This fact is even more significant considering the CORDA reconstructions performed using a single dependency assessment have an average size of 1,828.7 reactions ([S1 Text](#)), 3.7% larger than mCADRE and 0.15% larger than MBA, demonstrating the ability of CORDA to perform nearly as concise reconstructions despite not being parsimonious.

The difference in number of reactions between CORDA and other reconstructions stems mostly from a larger number of transport reactions. CORDA contains only 43 more reactions than the MBA reconstruction, but 79 more transport reactions. Similarly, CORDA contains 106 more reactions than the mCADRE reconstruction, but 140 more transport and exchange reactions. This discrepancy indicates that the parsimonious pruning methods are more likely than CORDA to exclude exchange and transport reactions.

When considering the similarity between models, there are 1,231 reactions present in all three reconstructions, accounting for 67.4% of MBA, 69.8% of mCADRE, and 65.86% of

Table 1. Comparison between CORDA, MBA and mCADRE hepatocyte specific reconstructions.

	MBA	mCADRE	CORDA
Reactions	1,826	1,763	1,869
Exchange reactions	247	197	249
Transport reactions	539	530	618
Metabolites	1,360	1,402	1,334
Unique metabolites	729	762	722
Genes	1,333	1,267	1,242
Unique genes	1,051	994	972

doi:10.1371/journal.pcbi.1004808.t001

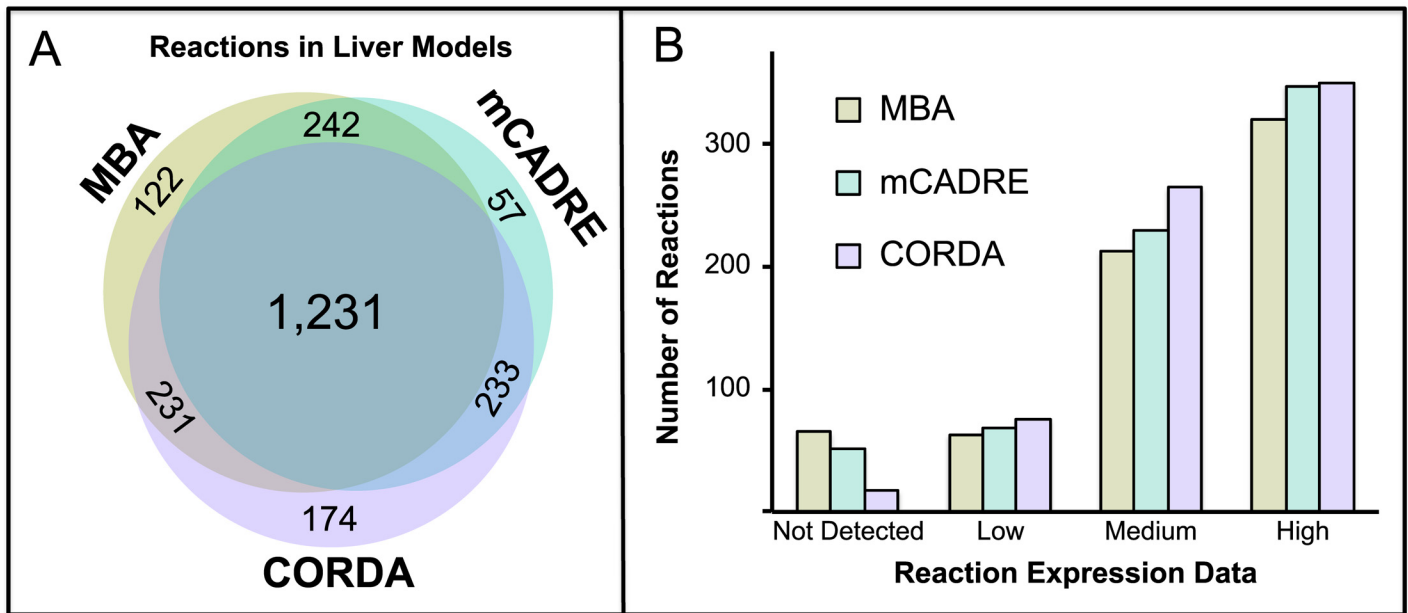


Fig 2. Comparison between MBA, mCADRE and CORDA hepatocyte-specific reconstructions. (A) Venn diagram of reactions included in each model. A total of 1,231 reactions are present in all models. (B) Number of reactions included in each model according to the protein expression data associated with each reaction. The CORDA algorithm includes more low and medium confidence reactions, while including considerably fewer reactions with no protein evidence.

doi:10.1371/journal.pcbi.1004808.g002

CORDA reactions (Fig 2A). Despite this relatively low overlap, no two models seem to be more similar than the other possible pairs. Furthermore, a higher level of similarity between models is observed when considering unique genes and metabolites, with at least 77% of genes and at least 84% of metabolites in each model shared across all models (S1 Text). Further comparison between the models is available in S1 Text.

Next, the number of HC, MC and NC reactions included in each of the models was analyzed. Here, Not Detected, Low, Medium and High corresponds to the experimental evidence associated with each reaction (Materials and Methods). The CORDA reconstruction showed better agreement with experimental data in all reaction categories (Fig 2B). Particularly, CORDA contained a significantly higher number of medium confidence reactions, 264 as opposed to 229 in mCADRE and 212 in MBA, while including significantly fewer negative confidence reactions, 17 as opposed to 51 in mCADRE and 65 in MBA. It is worth noting that the MBA reactions core was chosen manually and based on data sources different than the one used for the mCADRE and CORDA reconstructions. The difference in reactions core used during the reconstruction process could explain the much lower agreement of MBA with this particular data set.

As a final validation of the CORDA algorithm, the ability of each of the models to perform a series of metabolic tasks was analyzed. These metabolic functions were divided into three categories: (1) amino-acid and ammonium recycling, (2) glucogenic production, and (3) nucleotide production. Briefly, during each test the model was allowed to freely exchange basic metabolites (i.e. water and oxygen), while the remaining exchange reactions were set to mirror the particular test (i.e. uptake of ammonium and release of urea during ammonium recycling test). The model was then forced to produce (1) urea, (2) glucose, or (3) specific nucleotides. If the model was able to do so, the test was considered passed, otherwise, the test was considered failed. If the appropriate exchange reactions were not present to perform the test, the result was

Table 2. Metabolic test results.

	MBA	mCADRE	CORDA
Amino-acid recycling	21/0/0	8/5/8	19/1/1
Glucogenic	19/0/0	5/5/9	18/1/0
Nucleotide production	4/4/-	8/0/-	6/2/-

Test results are reported as number of tests passed/ tests failed/ and inconsistent test results, meaning the appropriate exchange reactions were not included in the model. Test-specific exchange reactions are not required for nucleotide production tests.

doi:10.1371/journal.pcbi.1004808.t002

considered inconsistent. Further details on how these metabolic tasks were calculated are available in the Materials and Methods section. Test results are summarized in [Table 2](#), and additional information is provided in [S1 Text](#).

The MBA reconstruction passed all of the amino-acid recycling and glucogenic tests, which is not surprising given the core set of MBA reactions, and therefore the necessary pathways for these tasks, were manually included in the reconstruction. The mCADRE reconstruction passed all eight nucleotide production tests, since these were included in the model building process. Where the tests were not manually included, however, MBA passed only 50% of the nucleotide production tests, and mCADRE passed only 13 of the amino-acid recycling and glucogenic tests, with an additional 17 tests being inconclusive. On the other hand, CORDA passed a total of 43 of the combined 48 metabolic tests (89.6%). This result is significant considering none of these tests were included in the reconstruction process.

Metabolic tests were also performed for all 108 hepatocyte specific reconstructions generated during the parameter sensitivity analysis section. All reconstructions had the same results for all amino-acid recycling, glucogenic, and five of the nucleotide production tests, demonstrating that task results are not heavily dependent on the CORDA parameters or noise ([S1 Text](#)). The inability of 39 of the 108 models (36.1%) to produce three of the nucleotides was traced back to the reaction *TRDR*, an MC reaction dependent on the NC reaction *RNDR1*, which was included in some but not all models. This reaction dependency, however, was returned by the CORDA algorithm, and hence the reactions needed to produce all the nucleotides can be easily included upon manual curation.

These analyses demonstrate that the CORDA algorithm provides a reconstruction with better agreement to experimental data and better metabolic functionality when compared to previous, similar methods. Furthermore, the analysis of the 108 models to perform metabolic tasks highlights the importance of the reaction dependencies returned by CORDA in subsequent manual curation, indicating which NC reaction needs to be added back into the model for the desired MC reaction to carry flux.

Monte-Carlo sampling. Monte-Carlo Sampling can be used to find a uniform distribution of steady-state flux vectors throughout the metabolic model, providing insight into the shape and size of the model's solution space [46, 47]. This uniform random sampling technique allows for the unbiased estimation of probability distributions of flux values for each reaction in the model. While the sampled flux values do not necessarily correlate with physiological metabolic fluxes, the sampled distribution can estimate the capability and flexibility of each reaction in the model given the network constraints [46]. This technique has been used to study pathological states in the human red blood cell [46, 47] and mitochondria [48], as well as the interaction between cell types in the human brain [49] and between *M. tuberculosis* and macrophages [25].

Here we performed Monte-Carlo sampling on the CORDA, MBA and mCADRE hepatocyte models, as well as the generalized human reconstruction Recon1. We also introduce a second CORDA reconstruction (CORDA2), calculated using the latest, most up-to-date data from the HPA. To allow for a direct comparison between CORDA and previous algorithms, the CORDA reconstruction used in the previous sections (referred in this section as CORDA1) was calculated using the same data used in the mCADRE reconstruction. This was done to ensure that differences in model functionality, capacity, and differential inclusion of high confidence reactions stemmed from the difference in algorithms used in the reconstruction process, and not from different datasets used to calculate each model. The CORDA2 model, on the other hand, has been calculated to exemplify how the most recent HPA data leads to better model capacity predictions. The protein expression data and reaction groups used in the reconstruction of both CORDA models is available in [S1 Table](#), and both models are available in [S1 File](#). Details of how the flux values were sampled are outlined in the Materials and Methods section.

The distribution of sampled flux values for reactions representing several hepatocyte specific functions, including production of urea from arginine, production of bilirubin, production of pyruvate from lactate (Cahill cycle), gluconeogenesis, and cholesterol production are plotted in [Fig 3](#). CORDA1 showed a higher capacity than all other reconstructions for bilirubin production and lactate recycling, and a similar capacity for gluconeogenesis and cholesterol efflux. This model only shows a lesser capacity than MBA and Recon1 in the production of urea. Overall, however, these results show that the CORDA algorithm better captures model capability for hepatocyte specific functions when compared to previous models and the generalized human reconstruction. In other words, the subset of Recon1 defined by the CORDA model better represents hepatocyte functions than the subset defined by the MBA and mCADRE reconstructions.

In addition, the sampled fluxes for all reactions considered here showed a significant shift towards higher values in the CORDA2 model when compared to MBA, mCADRE, and the generalized human reconstruction Recon1 ($p < 10^{-20}$), including in the production of urea. While CORDA1 showed better functionality in bilirubin production and lactate recycling than CORDA2, CORDA2 outperformed CORDA1 in gluconeogenesis and cholesterol efflux capability. These results suggest that the most recent data from the HPA captures a wider range of tissue-specific functionalities. It is also worth noting that the CORDA models were the only models where the flux through lactate dehydrogenase was highly biased towards positive values, converting lactate to pyruvate (Cahill cycle). All other models considered here showed either an even distribution between positive and negative values, or mostly negative values leading to the production of lactate.

Another interesting result of this analysis is the flux values sampled for HMG-CoA reductase in the MBA reconstruction. This enzyme represents the rate-limiting step in the *de novo* synthesis of cholesterol and other isoprenoids. Flux values sampled for this reaction using the MBA reconstruction are extremely high when compared to other models, and they were never close to or equal to zero. This distribution suggests that this reaction might also be used in other cellular processes, and thus carries a higher flux more frequently during sampling. To investigate this possibility, we analyzed which reactions are dependent on the HMG-CoA reductase reaction (*HMGCOARx*) in each of the hepatocyte models. This was done by evaluating which reactions lose their ability to carry flux upon setting the upper and lower bounds of *HMGCOARx* to zero. While in the CORDA models only reactions in the cholesterol metabolism, endoplasmic reticulum transport, and peroxisomal transport pathways are blocked, a much larger number of reactions are blocked in the MBA and mCADRE reconstructions. In both of these models, a high number of bile acid biosynthesis reactions lose functionality upon

Monte-Carlo Sampling of Hepatocyte Models

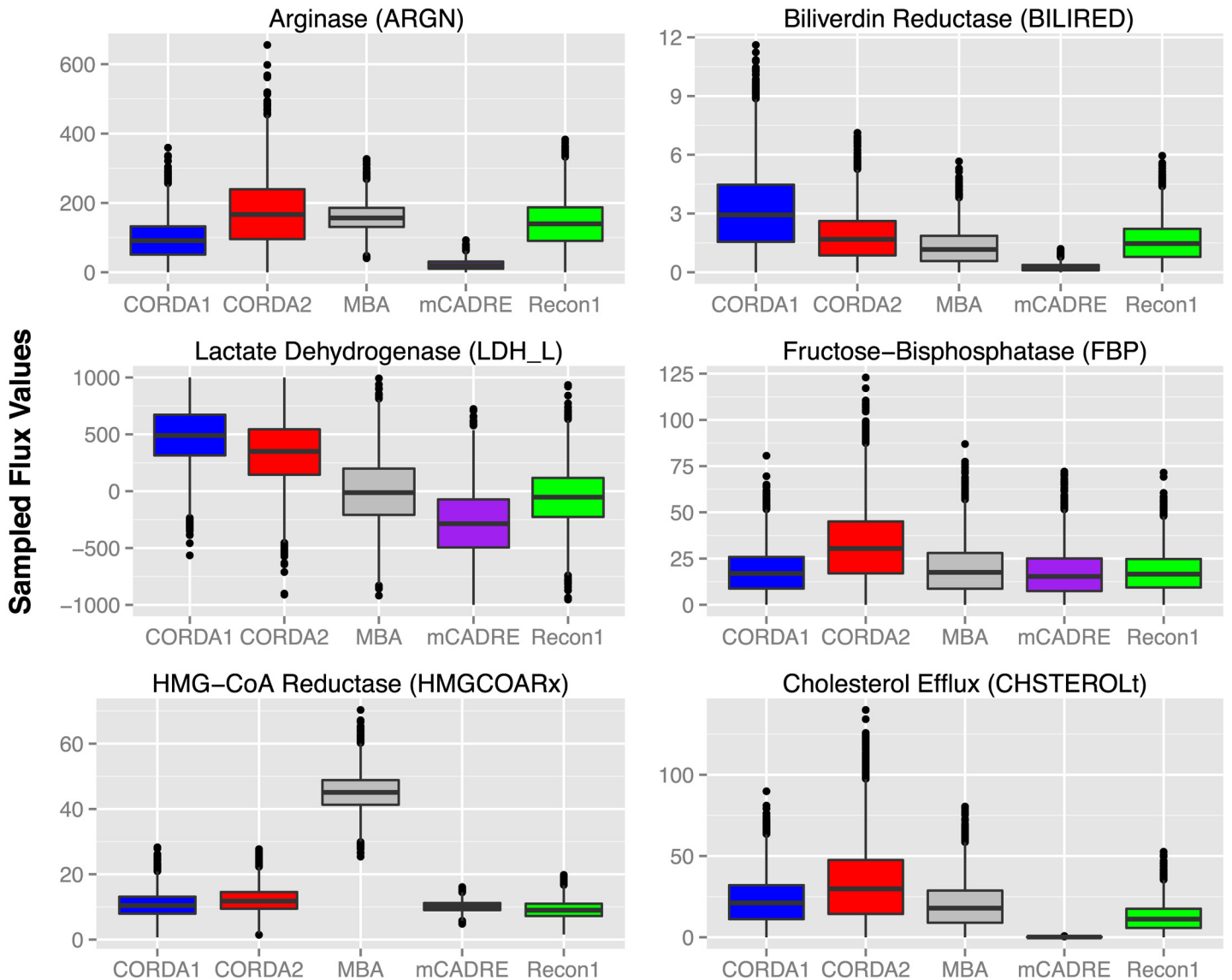


Fig 3. Monte-Carlo sampling of Recon1 and hepatocyte specific models. Distribution of flux values sampled for selected reactions representing hepatocyte specific functions in each of the hepatocyte models and the generalized human reconstruction Recon1. The name of each reaction plotted, as defined in the Recon1 reconstruction, is presented in parenthesis.

doi:10.1371/journal.pcbi.1004808.g003

blockage of *HMGCOARx*. Reactions in other sub-systems, such as Lysine metabolism and purine catabolism in MBA, and taurine and hypotaurine metabolism, vitamin D metabolism, and CoA biosynthesis in mCADRE, also have a loss of function (S1 Text). Overall, upon blockage of *HMGCOARx*, 188 additional reactions in the MBA model, and 156 in the mCADRE model lose function, compared to 30 in CORDA1 and 33 in the CORDA2 model (S1 Text). This analysis demonstrates how parsimonious algorithms are likely to remove alternative pathways from the model, conferring very high levels of influence over the network to particular reactions.

Generation of multiple healthy and cancer tissue-specific models

Following the algorithm validation, we generated a library of 76 healthy and 20 cancer tissue-specific models using CORDA. In order to generate the most comprehensive models possible, we used the generalized human reconstruction Recon2 [9] in the calculation of this library. Recon2 is one of the most comprehensive human reconstructions performed to date, containing approximately twice the amount of reactions than Recon1, 1.7 times more unique metabolites, and 1.2 times more unique genes. Details of how the reconstructions were calculated can be found in the Materials and Methods section.

Identification of essential metabolites. As an initial validation of this library, we analyzed the reconstructions for the identification of essential metabolites specific to cancer. Essential metabolites are necessary to carry out specific cellular functions, and can be used for the identification of possible antimetabolites. Antimetabolites, in turn, are structurally similar to essential metabolites but cannot be used by the cell, thus stalling enzymes consuming the essential metabolite through competitive inhibition [50]. By targeting cellular functions specific to cancer, antimetabolites have been widely used in the treatment of multiple types of cancer [51–53].

The identification of essential metabolites can be predicted computationally using GEMs [15, 16, 54]. Given a metabolite to be tested, all reactions consuming that metabolite are blocked in the forward direction, and reactions producing the metabolite are blocked in the backwards direction. An array of essential metabolic functions is then tested, and failure to complete any of those functions renders the metabolite essential for that GEM. All 76 healthy and 20 cancer specific reconstructions calculated here were tested for all 271 unique metabolites present in all models, using the 32 metabolic functions included in the reconstruction process (S1 Table). Two metabolites selectively targeted cancer over healthy reconstructions: phosphatidylethanolamine (*pe_hs*), essential in 1.3% of healthy and 10% of cancer specific reconstructions, and triglyceride (*tag_hs*), essential in 31.6% of healthy and 70% of cancer specific reconstructions. Both of these metabolites are involved in fatty-acid and glycerophospholipid pathways, which have been previously identified as specifically essential to cancer tissues [15, 16].

Composition analysis of cancer and healthy tissue models. Next, the 96 tissue-specific models were clustered according to the reactions present in each of them (Materials and Methods). The clustering results are summarized in Fig 4. We find that the tissue-specific models largely cluster according to tissue type, with most of the cancer models clustering together. The only exception stems from the liver and prostate cancer models, which cluster with their healthy counterparts (Fig 4).

Following the clustering results described above, we divided the models into seven categories according to the results presented in Fig 4. These categories are glandular (green), epithelial (orange), lymphoid (cyan), cancer (blue), muscle (yellow), brain (magenta) and miscellaneous (black), which includes all of the remaining models. The presence of reactions in 89 different subsystems was then calculated and clustered for each of the model categories (Materials and Methods). Results are summarized in Fig 5. Evidence for the up- or down-regulation of many metabolic pathways differentially included or excluded from cancer models as opposed to healthy tissues is available in the literature [55–82] and highlighted in the discussion section. Overall, this clustering analysis further confirms the ability of CORDA to generate reconstructions in agreement with experimental data, and validates the library of healthy and cancer tissue reconstructions generated here.

Single reactions that are present most often in cancer but not in healthy, and in healthy but not in cancer tissue-specific reconstructions, were also analyzed (Table 3). All reactions that are present most often in cancer tissues have been shown to be up-regulated in at least one type of cancer. Reactions that are not gene-associated are part of sarcosine or folate metabolism

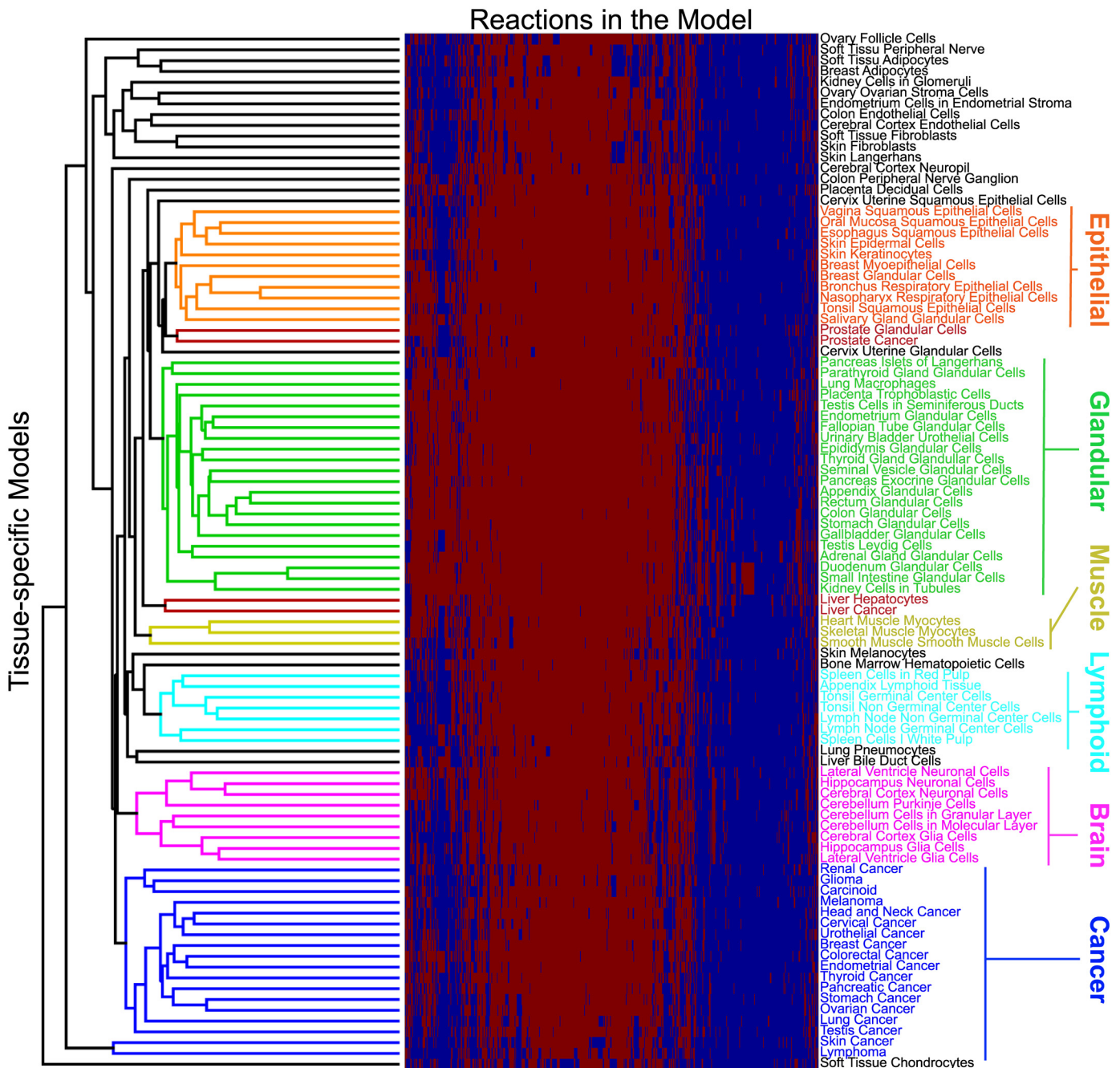


Fig 4. Clustering of tissue specific models. Healthy- and cancer-specific models clustered according to reactions present in each model. Highlighted clusters indicate epithelial and myoepithelial like tissues (orange), glandular like tissues (green), muscle tissues (yellow), lymphoid like tissues (cyan), brain cells (magenta), cancer tissues (blue), and cancers clustered with their healthy counterpart (red). Red squares indicate reactions present in the model and blue squares indicate reactions that are absent.

doi:10.1371/journal.pcbi.1004808.g004

pathways, both of which have been shown to be up-regulated in cancer [70–72, 83, 84]. Similarly, reactions that are present most often in healthy tissue reconstructions, but not cancer, have largely been shown to be down-regulated in cancer (Table 3). Many of these reactions are part of D-glucosamine and n-3 polyunsaturated fatty acids metabolisms, both of which have been shown to selectively target and kill cancer cells [58–61, 85–87]. An additional three

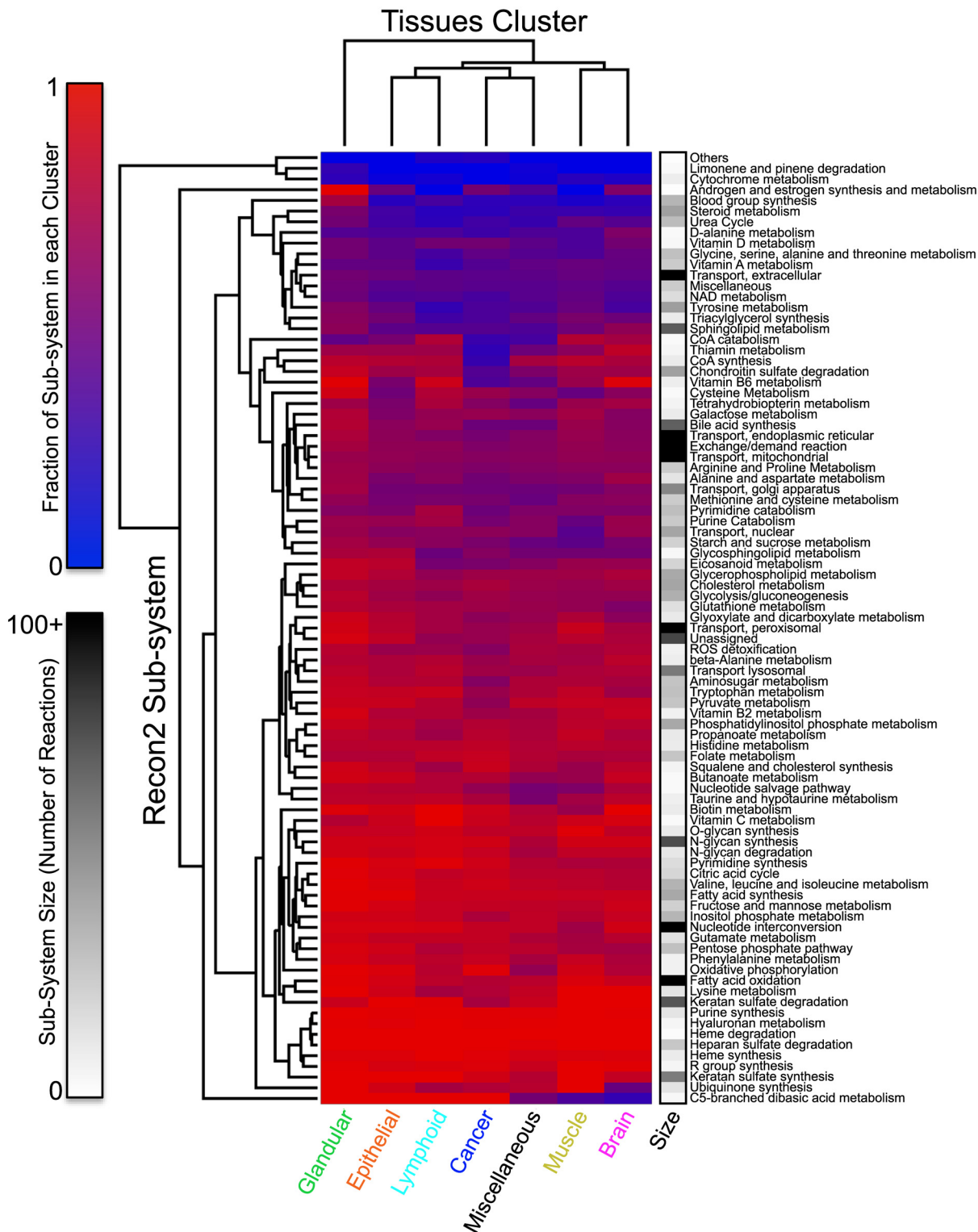


Fig 5. Clustering of subsystems for each model category. The presence of reactions in each subsystem, for each model category, was calculated and clustered. Heat map represents the fraction of reactions in each subsystem included on average in models of the specific category. Number of reactions in each subsystem considered is also included.

doi:10.1371/journal.pcbi.1004808.g005

Table 3. Reactions differentially included in healthy versus cancer tissue models.

Reaction	Subsystem	Associated Genes	Cancer Models	Healthy Models	Literature Evidence
Reactions present in cancer tissue models most often:					
<i>DNDPt15m</i>	Transport, mitochondrial	60386	95%	32.9%	Breast Cancer [88]
<i>EX_sarcs(e)</i>	Exchange/demand reaction	-	85%	22.4%	Prostate Cancer [83, 84]
<i>H2O2syn</i>	Tyrosine metabolism	53905, 50506	90%	39.5%	Prostate Cancer [89]
<i>MTHFD2m</i>	Folate metabolism	10797	65%	7.9%	Multiple [71, 73], Breast [70, 72]
<i>SARCStex</i>	Transport, extracellular	-	85%	22.4%	Prostate Cancer [83, 84]
<i>r0514</i>	Folate metabolism	1719	85%	10.5%	Multiple [71, 90]
<i>r0961</i>	Transport, extracellular	1356	70%	9.2%	Multiple [91, 92]
<i>r0962</i>	Transport, mitochondrial	-	85%	10.5%	Multiple [71], Breast [70, 72]
<i>r2085</i>	Transport, extracellular	6566	70%	9.2%	Neuroblastoma [93]
<i>r2086</i>	Transport, extracellular	6566	90%	35.5%	Neuroblastoma [93]
Reactions present in healthy tissue models most often:					
<i>ACOA07p</i>	Fatty acid oxidation	51	15%	97.4%	-
<i>ALCD22_D</i>	Pyruvate metabolism	124–128, 130, 131, 137872, 284273	25%	97.4%	Colorectal Cancer [94, 95]
<i>DADA</i>	Nucleotide interconversion	100	5%	86.8%	Multiple [96], Renal adenocarcinoma [97], Gastric [98]
<i>DADAE</i>	Nucleotide interconversion	100	5%	86.8%	Multiple [96], Renal adenocarcinoma [97], Gastric [98]
<i>DESAT24_1</i>	Fatty acid synthesis	9415	5%	80.3%	Multiple [99]
<i>EX_gam(e)</i>	Exchange/demand reaction	-	20%	94.7%	Multiple [58–61]
<i>EX_tethex3(e)</i>	Exchange/demand reaction	-	5%	76.3%	Multiple [85–87]
<i>EX_tetpent3(e)</i>	Exchange/demand reaction	-	5%	80.3%	Multiple [85–87]
<i>FACOAL245_2</i>	Fatty acid oxidation	2180	5%	80.3%	Hepatoma [100], Lung [101], Colon [102]
<i>FACOAL246_1</i>	Fatty acid oxidation	2180	5%	76.3%	Hepatoma [100], Lung [101], Colon [102]
<i>GAMt1r</i>	Transport, extracellular	6513, 6514, 6517	20%	94.7%	Multiple [58–61]
<i>HEX10</i>	Aminosugar metabolism	2645, 3098, 3099, 3101, 80201	20%	94.7%	Multiple [58–61]
<i>O2Stx</i>	Transport, peroxisomal	-	25%	98.7%	-
<i>PNTK</i>	CoA synthesis	53354, 55229, 79646, 80025	20%	98.7%	-
<i>PPCDC</i>	CoA synthesis	60490	20%	98.7%	-
<i>PPDOx</i>	Pyruvate metabolism	8574	20%	97.4%	Hepatocellular Carcinoma [103]
<i>PPNCL3</i>	CoA synthesis	79717	20%	98.7%	-
<i>SPODMx</i>	ROS detoxification	6647	25%	98.7%	Multiple [104]
<i>TETHEX3t</i>	Transport, extracellular	-	5%	76.4%	Multiple [85–87]
<i>TETPENT3t</i>	Transport, extracellular	-	5%	80.3%	Multiple [85–87]

References for reactions largely present in cancer models refer to up-regulation in cancer metabolism, while references for reactions largely present in healthy tissue models refer to their down regulation in cancer cells.

doi:10.1371/journal.pcbi.1004808.t003

reactions, *PNTK*, *PPCDC* and *PPNCL3* are part of coenzyme-A synthesis pathways, a pathway also implicated as significantly excluded from cancer models during the clustering analysis (Fig 5). These reactions are largely excluded from cancer reconstructions due to the gene *PPCS*, the only gene from these three reactions with supportive information in the HPA, and thus included in the reconstruction process. *PPCS* is expressed in all healthy tissues, with medium or high confidence in 33 of them, but is largely low or negatively expressed in most cancer tissues. Similarly, the reaction *ACOA07p*, dependent on the gene *ACOX1*, was largely included in healthy tissues and not cancer tissues according to experimental data.

Finally, reactions that are present in three or fewer healthy models, and reactions that are present in a single cancer model, were analyzed in order to assess CORDA's ability to include tissue specific reactions (S1 Table). We found that the inclusion of tissue specific reactions are largely aligned with our knowledge of metabolism. Bile acid and urea cycle reactions are concentrated on the hepatocyte and intestinal mucosa reconstructions [105]; tyrosine metabolism, a precursor to melanin, is highly active in skin melanocytes and epidermal models; the soft tissue adipocyte model contains a specific vitamin D metabolic reaction, a pathway highly associated with this tissue [106, 107]; steroid metabolism reactions are concentrated in adrenal, duodenum and prostate glandular, as well as testis Leydig cell reconstructions; and the iodine exchange reaction is unique to the thyroid gland model. Similarly, bile acid synthesis and cholesterol metabolic reactions are highly concentrated in the liver cancer reconstruction; chondroitin sulfate degradation reactions are highly concentrated in the prostate cancer model [108–111]; and tyrosine metabolism reactions are largely present in the thyroid [112, 113] and melanoma [114] cancer reconstructions (S1 Table).

Monte-Carlo sampling of cancer and healthy tissue models. Monte-Carlo Sampling was also performed on each of the 96 cancer and healthy tissue models. Due to the heterogeneity between cell types, flux values were analyzed between all cancer and healthy tissue models together (Materials and Methods). The distribution of sampled flux values for selected reactions or group of reactions are plotted in Fig 6. Plots for individual healthy and cancer models are available in S1 Text. Similar to the hepatocyte specific analysis, the sampled values showed good agreement with experimental data and our understanding of cancer metabolism. Cancer models showed an overall greater capacity for lactate secretion (in accordance with the Warburg effect [115, 116]), glycolysis [71], the pentose phosphate pathway [116] (particularly through the up-regulation of *TKT1* [71]), and through methylene-THF dehydrogenase (*MTHFD2*) in the conversion of 5,10-methylene-THF to 5,10-methenyl-THF (positive direction)[71]. On the other hand, pathways that have been shown to be down-regulated in cancer demonstrate a decreased capacity in cancer models when compared to healthy tissue models, including mitochondrial respiration (Complex IV)[115] and superoxide dismutase [71, 104].

These results demonstrate the ability of Monte Carlo sampling to predict the capacity of cancer versus healthy tissue models beyond simple topological analysis (i.e., simply looking at the presence or absence of certain reactions or pathways in the model). For instance, this analysis is able to capture a higher capability of cancer models for lactate secretion, and a lower capacity through superoxide dismutase, based on model topology alone, even though these reactions are present in all the models in the library. This analysis is made possible by the faster, high throughput capability of the CORDA algorithm, allowing for the generation of a library of tissue-specific models based solely on experimental data. While mCADRE allows for a relatively fast and high throughput generation of tissue-specific models, results shown in Fig 3 suggest that these reconstructions show poor functionality predictions. MBA, on the other hand, yields much better predictions of model functionality than mCADRE, but still falls short of the CORDA reconstructions according to our hepatocyte specific analysis (Fig 3). The computational cost of the MBA algorithm also renders the generation of a library of models extremely

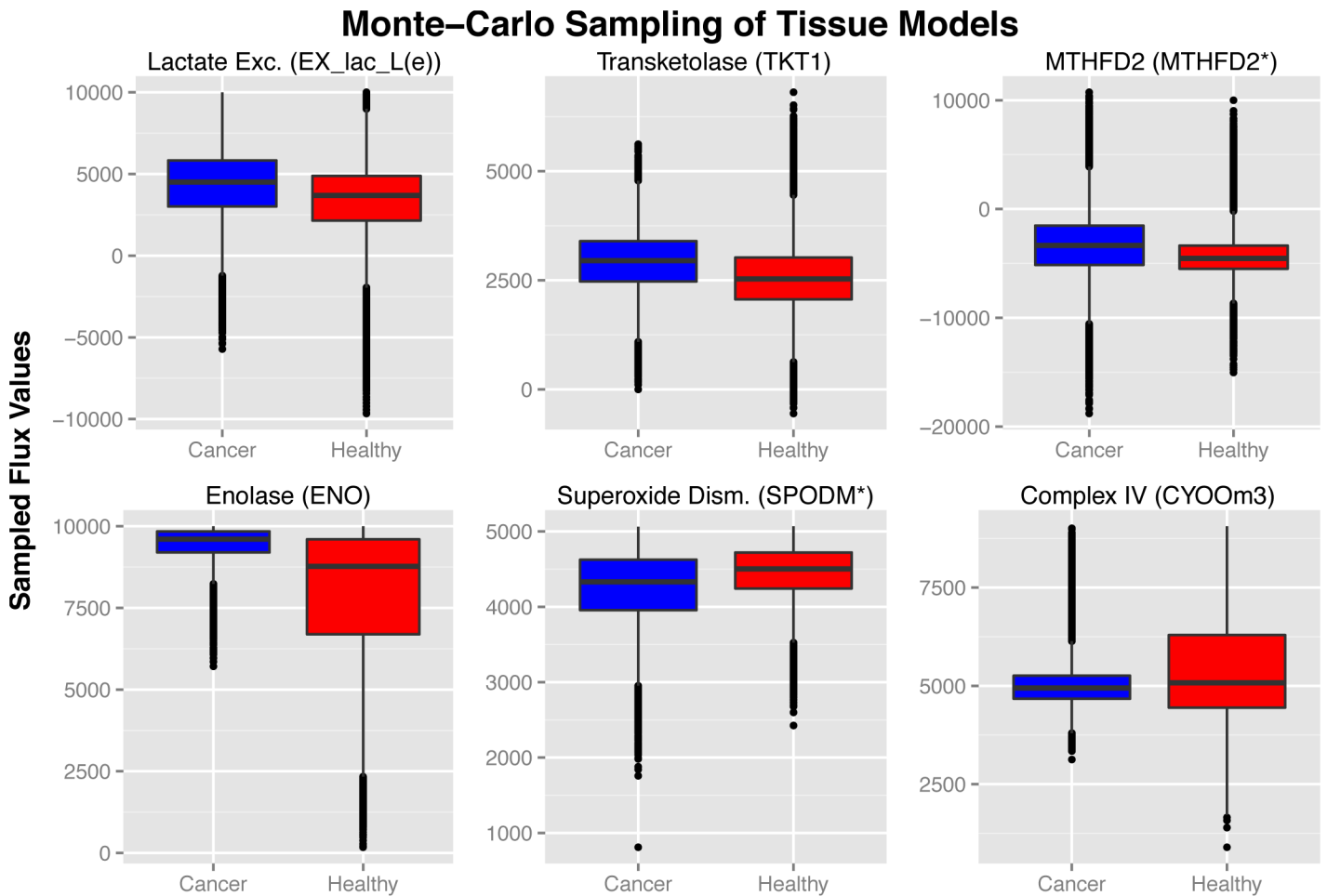


Fig 6. Monte-Carlo sampling of cancer and healthy tissue models. Flux values sampled from all cancer and healthy tissue models. Recon2 reactions plotted in each boxplot are indicated in parenthesis. Asterisk indicates groups of reactions taking place in different cellular compartments. According to general cancer metabolism, cancer models showed a higher capacity through lactate production, pentose phosphate pathway, MTHFD2, and glycine hydroxymethyltransferase, while showing a lower capacity through oxidative phosphorylation and superoxide dismutase.

doi:10.1371/journal.pcbi.1004808.g006

expensive and time consuming, especially if the core set of reactions were to be selected manually. In sum, CORDA resulted in direct comparisons of cancer versus healthy tissue metabolism, as well as more accurate reconstructions of hepatocyte function compared to prior tissue-specific metabolic models, in a computationally efficient manner.

Discussion

Here we introduced a novel tissue-specific algorithm based on Cost Optimization Reaction Dependency Assessment (CORDA). CORDA relies solely on FBA, rendering it more computationally efficient than previous methods. CORDA takes a non-parsimonious approach to the reconstruction process, based on the addition of valuable reactions to the reconstruction as opposed to the removal of non-essential reactions. We showed that the CORDA algorithm provides reconstructions that agree better with experimental data, and that demonstrate better metabolic functionality than prior methods like MBA and mCADRE. Furthermore, CORDA provides reaction associations that can greatly assist subsequent manual curation, while maintaining the reconstructions only slightly larger than previous parsimonious approaches.

Monte-Carlo sampling analysis also demonstrates that the CORDA generated models provide better predictions of tissue-specific functionality.

In addition to the algorithm validation, we generated a library of 76 healthy and 20 cancer tissue-specific reconstructions, which show considerable agreement with our current knowledge of healthy tissue and cancer metabolism. First, as an initial validation of our cancer and healthy tissue models, we computationally predicted metabolites that are more frequently essential in cancer models than healthy tissues [15, 16, 54]. Two metabolites were implicated in this analysis: phosphatidylethanolamine (*pe_hs*) and triglyceride (*tag_hs*), both of which are part of metabolic pathways previously implicated as cancer specific [15, 16]. While future work is merited to identify more specific essential metabolites (e.g. through the inclusion of more comprehensive metabolic tasks in the tissue reconstruction process, and more metabolites in the essential metabolite identification algorithm), these results help validate the cancer and healthy tissue reconstructions presented here.

Following this analysis, we demonstrated that the tissue models calculated by CORDA cluster largely according to tissue type. Similar clustering patterns, based on gene expression and proteomics data, have been observed experimentally. In particular, based on the expression of over 30,000 genes across multiple individuals and tissues, one study found that brain, muscle, and liver tissues, as well as Epstein-Barr virus-transformed lymphocytes, form well defined groups, while skin, adipocytes, and nerve tissues cluster closely together [117]. A separate study used in the generation of the HPA, based on protein evidence from almost 17,000 protein-coding genes in 44 major tissues and organs, also showed that tonsils, spleen, appendix, and lymph node tissues cluster closely together, and that bone marrow clusters separately, but close to these lymphoid tissues [45].

Evidence supporting many of the apparent exceptions identified by our clustering analysis is also available. For instance, Uhlén et. al. found that brain and liver tissues, along with testis, cluster considerably separate from other tissues and closer to each other, which is what we observed by clustering the CORDA models. The same study found that prostate tissue clusters closely with salivary glands [45]. It is worth noting that good agreement with the data by Uhlén et. al. is expected, given that a subset of this data was used to generate the tissue-specific models. This agreement, however, suggests that the similarities between tissues shown by Uhlén et. al. [45] and Melé et. al. [117] at the gene expression and protein level are also present in the metabolic enzymes level.

Additionally, breast and salivary glands are known to share many morphological features, and studies have shown that both can give rise to tumors with similar morphology [118, 119] and myoepithelial differentiation [120]. These finding can explain why breast and salivary glands clustered with epithelial and myoepithelial cells, as opposed to glandular cells. Finally, skin cancer and non-Hodgkin's lymphoma appear frequently as secondary cancers in immunosuppressed individuals [121, 122]. This could lead to cancers with significantly different metabolic profiles, supporting their separation from the remaining cancer models.

Clustering of tissue-specific models according to subsystems has also highlighted many differences between healthy and cancerous tissues at the pathway level (Fig 5). Evidence for many of these differences are also available in the literature, including:

- **Thiamine metabolism:** Thiamine deficiency is common in advanced cancer patients [55, 56]. Thiamine supplementation has been shown to increase tumor proliferation *in vitro* through transketolase activation [55], and multiple studies have linked thiamine metabolism to cancer through numerous mechanisms [57].
- **Aminosugar metabolism:** D-glucosamine, the most abundant aminosugar and an important precursor in many biological pathways, has been shown to reduce tumor proliferation [58].

Although the mechanism of this inhibition is unknown, studies have suggested it may be through the targeting of cellular membranes [59], protein synthesis through p70S6K regulation [60], or protein N-glycosylation [61].

- **Pyruvate metabolism:** Pyruvate connects many metabolic pathways, and alterations in these pathways play an important role in cancer metabolism [62, 63]. Recent studies have shown that mitochondrial pyruvate carrier function is lost in cancer, and low expression of this protein is associated with poor survival [64]. This loss of function was also shown to induce the Warburg effect, a hallmark of cancer [64]. Additionally, the pyruvate kinase isoenzyme PKM2 has been shown to play a key role in cancer metabolism, diverting glucose to anaerobic pathways [65].
- **Glycosaminoglycans degradation:** The accumulation of glycosaminoglycans in cancer cells has been widely established, and therapies targeting these polysaccharides have been proposed [66, 67]. In particular, chondroitin sulfate proteoglycans have been shown to accumulate in cancer cells and promote tumorigenesis by interacting with surface receptors [68]. Furthermore, keratan sulfate has been shown to accumulate in the serum of patients with cartilage tumors, and serum levels were shown to decrease upon tumor removal [69]. Our results suggest that this accumulation is coupled with a down-regulation of glycosaminoglycans degradation.
- **Folate Metabolism:** Folate metabolism, in particular the enzyme methylenetetrahydrofolate dehydrogenase (MTHFD2), has been shown to be up-regulated in cancer cells [70–73], and it has been shown to contribute to energy and purine requirements in cancer [73]. Folate pathways have also been associated with poor prognosis [71, 72] and increased cell proliferation *in vitro* [73]. Also, anti-folate agents have been shown to reduce cancer proliferation, and have been proposed as anti-cancer agents [73, 74]. Monte-Carlo sampling analysis has also shown a more positive flux distribution through the reactions *MTHFD2* and *MTHFD2r* in cancer over healthy tissue models (Fig 6), in accordance with our knowledge of cancer metabolism [71].
- **Squalene and cholesterol synthesis:** Squalene is an important precursor of cholesterol. Squalene synthase expression has been shown to be increased in prostate cancer [76], and inhibition of this enzyme has been demonstrated to cause cell death in prostate cancer [75, 76]. Furthermore, squalene oxidase has been indicated as an oncogene in both pancreatic [77] and breast [78] cancer.
- **Oxidative phosphorylation:** Oxidative phosphorylation pathways have long been considered down-regulated in cancer tissues. Recent studies have found, however, that these pathways remain intact [79] or even increase during metastasis [80–82]. The activity of this pathway in cancer versus healthy tissue models is further discussed during the Monte-Carlo sampling analysis discussion.

Single reactions included most often in cancer or healthy tissue models were also analyzed, and again literature evidence has been found to support many of them (Table 3). Two surprising findings stemmed from this analysis. First is the predicted down-regulation of CoA synthesis reactions, implicated in both the subsystem and single reaction analyses. Upon further inspection, we traced this differential inclusion to the gene *PPCS*, the only gene related to this pathway included in the reconstruction process, which is significantly down-regulated in cancer cells [44, 45]. Second, the exclusion of *ACOAO7p* from most cancer models is also unexpected, since this reaction is part of the fatty-acid oxidation pathway, which has been shown to be up-regulated in cancer tissues [123, 124]. Protein evidence of this reaction's associated gene,

ACOX1, supports this exclusion from cancer models [44, 45], suggesting an alternate pathway for palmitoyl-CoA oxidation in cancer tissues.

Finally, Monte-Carlo sampling was also performed in all healthy and cancer tissue models. Sampling results demonstrate that cancer models show an increased capacity through pathways that are largely up-regulated in cancer metabolism, and a reduced capacity through pathways previously shown to be down-regulated. Interestingly, mitochondrial respiration showed a slightly reduced and tightly constrained capacity in cancer over healthy tissue models, despite the presence of a larger number of oxidative phosphorylation reactions in cancer models (Fig 5). For decades, the role of mitochondrial respiration was thought to be decreased in cancer tissues due to their high glycolytic capacity. In recent years, however, researchers have shown that this pathway actually plays an important role in cancer metabolism [125, 126]. Our results suggest that although a larger number of oxidative phosphorylation reactions are present in cancer models, the activity of this pathway is tightly regulated by cancer metabolism topology (Fig 6). On one hand, the low probability of cancer models reaching high cytochrome c oxidase flux values compared to healthy tissues is in line with cancer's high glycolytic potential. At the other extreme, the low probability of cancer models reaching relatively low cytochrome c oxidase sampled fluxes is in line with the key role played by mitochondrial respiration in cancer metabolism uncovered in recent years.

We have also investigated the differences in glycine hydroxymethyltransferase capacity in cancer versus healthy tissue models (S1 Text). This reaction is dependent on two proteins, *SHMT1* and *SHMT2*, which correspond the cytosolic and mitochondrial isozymes respectively. Both these proteins have been shown to be up-regulated in cancer over healthy tissue models [127], although *SHMT2* has been so to a greater extent [71, 127]. The over expression of these proteins, however, has been shown to be heavily dependent on cancer type [127]. This claim is supported by the protein expression of *SHMT2* in the HPA, where half the cancer types considered have samples with both high and not detected *SHMT2* expression. This variability could explain why the distribution of reactions associated with these genes is similar between cancer and healthy tissue models (S1 Text). Some cancer types, however, show a considerable increase in *SHMT2* expression when compared to their healthy counterparts, including breast, glioma, head and neck, lung, stomach, testicular, and thyroid cancer. In all but one of these models (glioma), the flux distribution of glycine hydroxymethyltransferase was shown to be considerably shifted towards higher values when compared to their healthy counterparts (S1 Text). These results demonstrate CORDA's ability to predict cancer type specific functionality, and not only differences between all cancer and healthy tissues taken together.

The CORDA tissue-specific reconstruction algorithm, as well as the healthy and cancer tissue-specific reconstructions presented here, introduce a new approach for the development of comprehensive tissue-specific metabolic reconstructions. These reconstructions can generate novel insights into both healthy and diseased human metabolic behavior. Furthermore, the ability of CORDA to generate models based solely on experimental data, along with the computational efficiency of this algorithm, allows for continuous updates of this library of tissue-specific models, both as more experimental data is updated and made available, and as more comprehensive human metabolic reconstructions are developed.

Materials and Methods

The Cost Optimization Reaction Dependency Assessment (CORDA) algorithm

While previous methods determined reaction dependencies using Flux Variability Analysis (FVA), the CORDA algorithm takes a different approach, referred here as *dependency*

assessment. The novelty of this method lies not in the LP formulation itself, which is the same as the widely established Flux Balance Analysis (FBA), but in the model modifications performed prior to the application of FBA, as well as the interpretation of the flux distribution returned. Assuming we want to test whether a given reaction, x , is dependent on the presence of a group of reactions, Y , to carry flux, CORDA proceeds in five steps. The parameters required for the CORDA algorithm are summarized in [Table 4](#):

1. As a first step, x is constrained to carry a strictly positive or strictly negative flux $\pm\epsilon$, depending on whether the reaction is reversible or not, and which direction we wish to test. The magnitude of ϵ can be arbitrarily small.
2. Each reaction in Y is associated with a high “cost” γ . This cost is added to each reaction as a pseudo metabolite, while splitting reversible reactions into forward and backwards reactions. That is, a reaction “ $A \rightleftharpoons B$ ” is split into “ $A \Rightarrow B+cost$ ” and “ $B \Rightarrow A+cost$ ”. This way, the cost is positively produced whether the reaction is taking place in the forward or backwards direction. At this step, reactions not in Y are assigned a cost of zero.
3. Each reaction cost is increased by a small random value sampled uniformly between zero and κ . The parameter κ is several orders of magnitude lower than γ . This noise is added to account for multiple pathways which allow x to carry a flux ϵ , and are associated with the same cost. With the added noise, these pathways will have slightly different costs.
4. A reaction consuming this pseudo metabolite is added to the reconstruction and set as the model objective. This added reaction is then the only mean of cost consumption, while all other model reactions produce the pseudo metabolite regardless of directionality.
5. FBA is performed while minimizing the flux through the cost consuming reaction. The flux distribution obtained is then the flux distribution with minimal cost needed to maintain the strictly positive or strictly negative flux $\pm\epsilon$ of x . Any reaction in Y predicted to carry a flux in this flux distribution will be referred to as *associated* with x .

It is worth noting that the high cost reactions implicated in step five are not necessarily essential for x to carry a flux $\pm\epsilon$, but are the set of reactions in Y that combined carry the minimal amount of flux. That is, no flux distribution through the metabolic network allows for the predefined flux through x with a lower combined flux through the reactions of Y . For instance, if one of the reactions in Y deemed associated with x were to be removed from the reconstruction, x could still be able to carry a flux $\pm\epsilon$, but the combined flux through the reactions in Y would be larger than before. This way, this dependency assessment does not minimize the number of undesirable reactions to allow x to carry flux, but instead the combined flux through

Table 4. Parameters required for the dependency assessment and CORDA tissue-building algorithm.

Parameter	Parameter Use
γ	Cost associated with undesirable reactions (Y).
κ	Order of magnitude of noise added to reaction costs. Each reaction cost is increased by a value sampled uniformly between zero and κ .
ϵ	Magnitude of flux value by which the reaction being tested (x) is constrained.
n	Number of times each reaction dependency is assessed in order to sample from pathways with equivalent costs.
ρ	Threshold for the inclusion of Negative Confidence reactions in step 2 of the tissue-specific algorithm.

doi:10.1371/journal.pcbi.1004808.t004

them. Naturally, however, a lower number of reactions would more easily allow for a lower combined flux. It is also for this reason that throughout the manuscript we use the term *associate* instead of *dependent*. Throughout the literature, referring to one reaction as *dependent* on another means the removal of the later from the model negates the former's ability to carry flux, which is not necessarily the case for the reaction *associations* defined here.

Another significant advantage of this dependency assessment over previous pruning algorithms is that it requires only the LP problem solved during FBA, rendering it much faster than previous methods. While MBA and mCADRE used a much faster variation of FVA, it is still considerably more computationally expensive than LP. Although mCADRE is up to three orders of magnitude faster than MBA [30], the mCADRE model used in this study took about 4 hours to be calculated in a 2.34 GHz CPU with 4G RAM using the IBM CPLEX solver [30]. The CORDA reconstruction, on the other hand, using the same data and general human reconstruction, took under 30 minutes in a 2.66 GHz CPU with 4G RAM using the Gurobi solver [128].

In order to obtain a tissue-specific metabolic reconstruction using this dependency assessment, we define the Cost Optimization Reaction Dependency Assessment (CORDA) algorithm. This algorithm takes as input the reactions in the generalized human reconstruction divided into four categories:

- **High Confidence (HC) reactions:** Reactions that are sure to be included in the tissue-specific reconstruction.
- **Medium Confidence (MC) reactions:** These reactions will be included in the final reconstruction if they are not dependent on negative confidence reactions associated with few MC reactions.
- **Negative Confidence (NC) reactions:** Reactions not to be included in the final tissue-specific reconstruction. These will only be included if they are associated with any HC or a high number of MC reactions. The NC and MC reactions core is then flexible, while the HC core is not.
- **Other (OT) reactions:** All remaining reactions in the generalized human reconstruction not included in the HC, MC or NC groups.

Here, we also allow for the inclusion of metabolic tasks in the HC group. That is, during the CORDA algorithm, sinks can be specified for given metabolites, and added to the model when tested to ensure the final tissue model can produce these metabolites. These reactions are added when being tested then immediately removed from the model, so that none of these metabolic task reactions are present when other reactions are being tested, and no two test reactions are present in the model at the same time. The 32 metabolic tasks included in all CORDA reconstructions in this manuscript are available in [S1 Table](#).

While the definition of these reaction groups can be left to the user's discretion, here we defined the four groups according to proteomics data from the HPA [44, 45], and boolean gene-reaction rules included in the generalized reconstructions Recon1 and Recon2. In the HPA, each protein is classified as being Not Detected, or present at Low, Medium or High levels in each tissue. The gene-reaction association rules are composed of gene names and "AND" and "OR" boolean associations. For instance, the reaction *r0634* in Recon2 has the boolean rule "*HADHB AND (ACAA2 OR ACAA1)*", and can therefore be considered active if the gene HADHB, as well as ACAA2 or ACAA1, are active.

Using this boolean mapping, gene IDs were first replaced by the numerical values -1, 1, 2, and 3, corresponding to Not detected, Low, Medium and High protein expression levels respectively. Genes not included in the dataset were assigned a numerical value of zero. Next, AND

boolean associations were replaced by the function MIN; OR boolean associations were replaced by the function MAX; and the expression was evaluated. Reactions with a final score of 3 were assigned to the HC group; reactions with scores of 1 or 2 were assigned to the MC group; and reactions with a score of -1 were assigned to the NC group. Reaction scores of -1, 1, 2, and 3 also correspond to Not Detected, Low, Medium, and High expression levels expressed in Fig 2.

As an example, HADHB is expressed at low levels in cerebellum Purkinje cells; ACAA2 is not detected; and ACAA1 is expressed at high levels. The *r0634* gene-reaction rule mentioned above was then replaced by “ $MIN(1,MAX(-1,3))$ ”, which evaluates to 1. During the Purkinje cells reconstruction, this reaction was then placed in the MC group. Similar approaches have been used by previous studies to assign reaction confidence scores [30, 32, 37].

Aside from the four reaction groups, the CORDA algorithm also requires 5 parameters to operate, which are summarized in Table 4. To begin the algorithm, all HC reactions are moved into the tissue-specific reconstruction (RE), since these are sure to be included in the final model. Given the remaining three reaction groups, the CORDA algorithm proceeds in three steps:

1. MC and NC reactions associated with each RE reaction (which are the same as HC reactions at this point) are obtained using the dependency assessment n times. Here, each NC reaction is given a cost of γ and each MC reaction a cost of $\sqrt{\gamma}$, in order to favor the inclusion of MC over NC reactions. Any MC or NC reaction associated with any RE reaction, during any of the n dependency assessments, is then moved from the MC and NC groups to the RE group. These reaction associations are returned by the algorithm in order to assist in any subsequent manual curation.
2. NC reactions associated with each MC reaction are obtained using the dependency assessment n times. At this point, in order to maximize the inclusion of MC reactions and minimize the inclusion of NC reactions, we take a different approach than simply moving MC reactions and their associated NC reactions to RE. Instead, any NC reaction associated with p or more MC reactions is first moved from the NC group to the RE group. Subsequently, all remaining NC reactions are blocked (upper and lower bounds set to zero) and any MC reaction still able to carry flux in any direction, thus not depending on the blocked NC reactions, is moved from the MC to the RE group. This is done since, as described above, the blockage of NC reactions associated with a particular MC reaction does not necessarily remove that reaction's ability to carry flux. All MC reactions not included in the reconstruction, as well as their associated NC reactions, are then returned by the algorithm to assist in any subsequent manual curation.
3. Lastly, all MC and NC reactions not yet added to the RE group are blocked. OT reactions associated with each RE reaction are then obtained, again using the dependency assessment n times. Any OT reaction associated with any RE reaction during any of the n dependency assessments is then moved from the OT to the RE group. This final RE group then defines the tissue-specific reconstruction.

It is worth noting that one of the main advantages of CORDA over pruning algorithms is the fact that it is independent of how reactions are ordered. This is due to the fact that reaction associations are calculated for each step, and at the end of each step a decision is made as to which reactions are added to the tissue reconstruction. This way, the order in which reaction dependencies are calculated does not affect the final tissue reconstruction.

The CORDA reconstructions used for comparison to previous methods were generated using $\gamma = 10^5$, the highest cost value tested, $\kappa = 10^{-2}$, the lowest noise value tested, $\epsilon = 1$, a

threshold similar to a previous study [32], $n = 5$, to allow for the inclusion of a larger number of OT reactions, and $p = 2$.

For a direct comparison to previous methods, the CORDA reconstructions used during the parameter sensitivity analysis, cross-validation, and comparison to previous methods were performed using the same data used for the mCADRE hepatocyte reconstruction. For the Monte-Carlo sampling analysis, a new reconstruction was generated using the most up-to-date data from the HPA. Both of these reconstructions are available in the supplemental material (S1 File). All calculations in this study were performed using the COBRA toolbox [129] and the Gurobi optimizer [128]. The MATLAB function file used for CORDA reconstructions is also available in the supplemental material (S2 File). Finally, an example of the CORDA algorithm, applied to small sample networks, is available in S2 Text.

Parameter sensitivity analysis

While CORDA requires a number of different parameters, many of these values can be arbitrarily assigned. For instance, γ can be arbitrarily large, while ϵ and κ can be arbitrarily small. In order to demonstrate that the CORDA algorithm is robust to a wide range of parameters, we performed 108 hepatocyte specific reconstructions varying all parameters but p (which was set to be equal to two) to a wide range of values. A separate sensitivity analysis of p was performed and is included in S1 Text. The parameter p can be set in order to define a more or less flexible MC and NC core, and can be set to the user's discretion.

These 108 reconstructions were based on the generalized human reconstruction Recon1 [5], using the same set of protein expression data (total of 560) and 32 of the metabolic tests used in the mCADRE hepatocyte specific reconstruction [30]. The data used in this step, as well as the metabolic tests and calculated reaction groups, are available in the supplemental information (S1 Table). Metabolic tests were included as single reactions in the reconstruction in order to assure the model was able to produce certain metabolites. Each metabolic test was added to the model when being tested then immediately removed, so that no two tests were present in the model at the same time, and no metabolic test reaction was included when other reactions were being assessed. Details of this analysis are available in S1 Text.

Metabolic tests analysis

During the metabolic tasks validation analysis, the exchange rate of the *basal inputs* carbon dioxide (co2[e]), water (h2o[e]), protons (h[e]), oxygen (o2[e]), phosphate (pi[e]), hydrogen peroxide (h2o2[e]), superoxide anion (o2s[e]), bicarbonate (hco3[e]) and carbon monoxide (co[e]) were unconstrained. All other uptake reactions were blocked unless otherwise specified.

For each of the 20 amino-acid recycling tests, the uptake rate of the given amino acid and glucose were set to an arbitrary value, so that the amino-acid being tested was the only source or nitrogen. Next, the production of urea was set to a strictly positive value, and FBA was performed while optimizing the production of urea. The same test was also performed for ammonium. For each of the 21 glucogenic tests, the uptake rate of the given metabolite was set to an arbitrary value, and the production of glucose was optimized. For both the amino-acid and glucogenic tests, if the model returned a feasible flux distribution the test was considered passed, otherwise it was considered failed. If the exchange reaction of the given metabolite was not present in the model, the result was considered inconsistent. The generalized Recon1 reconstruction failed two of the glucogenic tests, so the results of the remaining 19 tests are reported in the main text.

For the eight nucleotide production tests, a sink consuming the given nucleotide was added to the cytosolic compartment. The model was allowed to uptake glucose and ammonium (as a

source of nitrogen), and the flux through the sink was optimized. If the model was able to produce the given nucleotide, the test was considered passed.

Generation of healthy and cancer tissue models

Following the validation of the CORDA algorithm, we generated a library of 76 healthy and 20 cancer tissue-specific reconstructions using the generalized human reconstruction Recon2 [9] and the most recent proteomics data from the HPA [44, 45]. All reactions used to generate the tissue-specific models are available in [S1 Table](#), and tissue-specific models are available in SBML and MATLAB format at [130]. The healthy tissue models were calculated using the same classification as described in the algorithm description section, since data for each protein was categorized as not detected, low, medium or highly expressed in each cell type. For cancer models, the same classification was available for any number of samples for each protein in each cancer type. In this case, values of -1, 1, 2 and 3 were assigned to each sample according to not detected, low, medium or high expression levels respectively, and these values were averaged for a final protein score in that particular cancer type. These protein values were then used in the gene-reaction boolean association as described in the algorithm description for a final reaction score. Reactions with a score equal to or greater than 2.5 were assigned to the HC group, less than 2.5 but greater than 1 to the MC group, and less than or equal to -0.5 to the NC group.

For instance, in renal cancer samples, protein *HADHB* has been analyzed in 12 different samples in the HPA, and was found to be expressed in high levels in 2 of them, medium levels in 8, and in low levels in 2. The protein score associated with *HADHB* in renal cancer is then calculated as $\frac{(2 \cdot 3) + (8 \cdot 2) + (2 \cdot 1)}{12} = 2$. Similarly, *ACAA1* expression was calculated as medium in 5 samples, low in two samples, and not detected in four samples of renal cancer, yielding a score of $\frac{(5 \cdot 2) + (2 \cdot 1) + (4 \cdot (-1))}{11} = 0.73$. Finally, *ACAA2* is present in high levels in one sample, medium level in 5 samples, low levels in one sample and not detected in 3 samples of renal cancer, giving this protein a score of $\frac{(1 \cdot 3) + (5 \cdot 2) + (1 \cdot 1) + (3 \cdot (-1))}{10} = 1.1$. With that, the score for *r0634* is calculated as “*MIN* (2, *MAX*(0.73, 1.1))”, which is 1.1, putting this reaction in the MC group during the renal cancer reconstruction. Data and reaction distributions used during these calculations can be found in [S1 Text](#).

Clustering of healthy and cancer tissue models

Healthy and cancer specific models were clustered according to reactions present in each model. For that, 4,205 reactions present in at least one, but not all models were obtained. A binary vector was then calculated for each model indicating whether reactions were present (1) or not present (-1). These vectors were then clustered using hierarchical clustering with Hamming distance as the similarity metric, and average linkage. Leaf orders were also calculated in order to maximize the similarity between neighbors in the hierarchical binary cluster tree dendrogram. These results are summarized in [Fig 4](#).

Next, in order to divide the clusters according to subsystem expression, a total of 4,751 reactions present in any of the models was obtained. These reactions were then divided by subsystem according to their classification in the Recon2 reconstruction. For each of the clusters of models calculated in the previous step, the average number of reactions from each subsystem included in the cluster’s models was then calculated. Finally, this number was divided by the total number of reactions in that subsystem which were included in any of the models for a final score between zero and one. These values were then clustered using hierarchical clustering with Euclidean distance as the similarity metric, and average linkage. Leaf orders were again organized to maximize similarity between neighbors to yield [Fig 5](#).

Flux Balance Analysis

Perhaps the most widely used method to analyze GEMs is Flux Balance Analysis (FBA) [42]. FBA predicts a flux distribution through the metabolic network which optimizes (maximizes or minimizes) a given *objective function*, defined as a single reaction or group of reactions in the network. This flux distribution is subject to upper- and lower-bound constraints, which include exchange reactions, and a steady state assumption for all model metabolites, so that no metabolite has a net production or consumption rate.

The mathematical formulation of GEMs are defined at the core by a stoichiometric matrix S , where each row defines a metabolite, each column defines a reaction, and each entry the stoichiometric coefficient of that metabolite in that particular reaction. Vectors defining lower (lb) and upper (ub) bounds for each reaction, as well as an objective vector (c) of the same length, are also defined. Given this model, FBA finds a flux vector v through all reactions in the GEM such that:

$$\begin{aligned} S \cdot v &= 0 \\ lb_i &\leq v_i \leq ub_i \\ \text{optimize} &: v \cdot c^T \end{aligned}$$

During the dependency assessment described here, the stoichiometric matrix S is altered to reflect the changes described above. Given a reaction j being tested, a group of undesirable reactions Y , and a matrix S of size m by n , let $\tilde{\kappa}$ denote a random number drawn uniformly between 0 and κ . The GEM is modified in the following ways:

1. Split all reversible reactions: *if* ($lb_i \leq 0$ & $i \neq j$) $\rightarrow S(:, n+1) = -S(:, i)$, $lb_{n+1} = 0$, $ub_{n+1} = -lb_i$ and $lb_i = 0$, where $S(:, n+1)$ denotes the addition of a column (reaction) to S .
2. Add cost to reactions: *if* ($i \in Y$) $\rightarrow S(m+1, i) = \gamma + \tilde{\kappa}$; *else* $\rightarrow S(m+1, i) = \tilde{\kappa}$, where $S(m+1, :)$ is the row denoting the cost pseudo-metabolite.
3. Add cost consuming reaction: $S(:, n+1) = [0 \ 0 \ 0 \ \dots \ 0 \ -1]$
4. Set bound of reaction j to desired value: *if* $\epsilon > 0 \rightarrow lb_j = \epsilon$; *else* $\rightarrow ub_j = \epsilon$
5. Set objective to cost consumption: $c = [0 \ 0 \ 0 \ \dots \ 0 \ 1]$

With these constraints in place, FBA is performed as described above while minimizing the objective function. For each reaction in the reconstruction, if $i \in Y$ and $v_i \neq 0$, the reaction i is deemed associated with j .

Monte-Carlo sampling

Monte-Carlo sampling was performed in a manner similarly to Bordbar et. al. [25] and Lewis et. al. [49]. This sampling method is a slight variation of the Artificially Centered Hit and Run (ACHR) algorithm developed by Kaufman and Smith [131]. In this algorithm, warmup points are initially generated at random corners of the solution space by solving an LP problem with objective vectors containing randomly generated ones and negative ones. The center point between all points is then computed. Next, for each point sampled, a random direction is selected as the difference between a randomly selected point and the center point. By selecting the direction this way, the direction is biased in the longer direction of the solution space, speeding up the rate of mixing while maintaining uniformity. After a direction is chosen, the limit of how far the current point can travel in that direction is calculated, and a new point is randomly chosen along that line. After several iterations, the

set of generated points will be well mixed and approach a uniform sampling of the solution space.

The termination condition imposed on the ACHR algorithm here is the same imposed by Bordbar et. al. [25] and Lewis et. al. [49], introducing the concept of mixed fraction. For that, a partition is created over the set of points by drawing a line at the median value, with half the points on either side of the partition. The mixed fraction is the number of points that cross this line during mixing. Initially, the mixed fraction is one as all the points are on their original side of the line. As the sample solutions are mixed, the probability of each point crossing the median line approaches 0.5 asymptotically. The sampled points were initially mixed using the warmup points created as described above until the mixed fraction reached a particular threshold. Following that, the samples were mixed two more times, using the previous iteration's final points as warmup points, until the same mixed fraction was reached. For the comparison between CORDA and other tissue-specific algorithms, a mixed fraction threshold of 0.52 was chosen as the termination condition. For the cancer and healthy tissue-specific models, a mixed fraction threshold of 0.6 was chosen to make the 96 sampling experiments computationally feasible.

Due to the heterogeneity between tissue-specific models, sampled flux values were evaluated between all cancer and healthy tissue models separately. That is, all sampled flux values for the given reaction were obtained from all cancer models that contain that reaction, and compared to all sampled values from healthy tissue models that contain the reaction. Results of this analysis are presented in Fig 6. In some cases, two or more reactions were combined: MTHFD2* combines reactions MTHFD2 and MTHFD2m, GHMT2r* combines reactions GHMT2r and GHMT2rm, and SPODM* combines reactions SPODM, SPODMe, SPODMm, SPODMn and SPODMx. These are the same reactions taking place in different cellular compartments. For these, flux values from each of these groups of reactions were added within each sampled flux distribution when plotting Fig 6.

Supporting Information

S1 Text. Supplementary Text 1. Supplementary text includes details of parameter sensitivity analysis, hypergeometric p-value calculations, parameter p sensitivity analysis, comparison between CORDA, MBA and mCADRE, healthy and cancer tissue reconstructions process, and Monte-Carlo sampling analysis.

(PDF)

S2 Text. Supplementary Text 2. Example of application of CORDA to small sample networks.

(PDF)

S1 Table. Supplementary Tables. Table contains the proteomics data and reaction groups used to calculate the Hepatocyte reconstructions, 32 metabolic tests used throughout the manuscript, reactions used to calculate tissue-specific models, reactions uniquely included in healthy and cancer reconstructions, and essential metabolite results.

(XLSX)

S1 File. Liver CORDA models. Reference hepatocyte specific reconstructions calculated using CORDA and Recon1. Two reconstructions are available, one using the same protein dataset used in the mCADRE reconstruction, and one using the most recent protein data in the HPA (S1 Table). The parameters used in this reconstruction are $\gamma = 10^5$, $\kappa = 10^{-2}$, $\epsilon = 1$, $p = 2$, and

$n = 5$.

(ZIP)

S2 File. CORDA file. MATLAB function used for CORDA reconstructions. (ZIP)

Acknowledgments

We would like to thank Dr. K.W. Lin, Dr. B. Long, Dr. D. Noren, A. Mahadevan, C.W. Hu and G. Britton for helpful discussions on the manuscript. We would also like to thank Dr. Jay Storz, from University of Nebraska, Dr. Zac Cheviron, from University of Montana, and Dr. Grant McClelland and Dr. Graham Scott, from McMaster University, for helpful discussions.

Author Contributions

Conceived and designed the experiments: AS AAQ. Performed the experiments: AS. Analyzed the data: AS. Contributed reagents/materials/analysis tools: AAQ. Wrote the paper: AS AAQ.

References

1. Oberhardt MA, Palsson BØ, Papin JA. Applications of genome-scale metabolic reconstructions. *Mol Syst Biol.* 2009; 5:320. doi: [10.1038/msb.2009.77](https://doi.org/10.1038/msb.2009.77) PMID: [19888215](https://pubmed.ncbi.nlm.nih.gov/19888215/)
2. Lewis NE, Nagarajan H, Palsson BO. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol.* 2012 Apr; 10(4):291–305. doi: [10.1038/nrmicro2737](https://doi.org/10.1038/nrmicro2737) PMID: [22367118](https://pubmed.ncbi.nlm.nih.gov/22367118/)
3. Bordbar A, Monk JM, King ZA, Palsson BO. Constraint-based models predict metabolic and associated cellular functions. *Nat Rev Genet.* 2014 Feb; 15(2):107–20. doi: [10.1038/nrg3643](https://doi.org/10.1038/nrg3643) PMID: [24430943](https://pubmed.ncbi.nlm.nih.gov/24430943/)
4. Hyduke DR, Lewis NE, Palsson BØ. Analysis of omics data with genome-scale models of metabolism. *Mol Biosyst.* 2013 Feb; 9(2):167–74. doi: [10.1039/c2mb25453k](https://doi.org/10.1039/c2mb25453k) PMID: [23247105](https://pubmed.ncbi.nlm.nih.gov/23247105/)
5. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, et al. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc Natl Acad Sci U S A.* 2007 Feb; 104(6):1777–82. doi: [10.1073/pnas.0610772104](https://doi.org/10.1073/pnas.0610772104) PMID: [17267599](https://pubmed.ncbi.nlm.nih.gov/17267599/)
6. Bordbar A, Palsson BO. Using the reconstructed genome-scale human metabolic network to study physiology and pathology. *J Intern Med.* 2012 Feb; 271(2):131–41. doi: [10.1111/j.1365-2796.2011.02494.x](https://doi.org/10.1111/j.1365-2796.2011.02494.x) PMID: [22142339](https://pubmed.ncbi.nlm.nih.gov/22142339/)
7. Rolfsson O, Palsson BØ, Thiele I. The human metabolic reconstruction Recon 1 directs hypotheses of novel human metabolic functions. *BMC Syst Biol.* 2011; 5:155. doi: [10.1186/1752-0509-5-155](https://doi.org/10.1186/1752-0509-5-155) PMID: [21962087](https://pubmed.ncbi.nlm.nih.gov/21962087/)
8. Shlomi T, Cabili MN, Ruppin E. Predicting metabolic biomarkers of human inborn errors of metabolism. *Mol Syst Biol.* 2009; 5:263. doi: [10.1038/msb.2009.22](https://doi.org/10.1038/msb.2009.22) PMID: [19401675](https://pubmed.ncbi.nlm.nih.gov/19401675/)
9. Thiele I, Swainston N, Fleming RMT, Hoppe A, Sahoo S, Aurich MK, et al. A community-driven global reconstruction of human metabolism. *Nat Biotechnol.* 2013 May; 31(5):419–25. doi: [10.1038/nbt.2488](https://doi.org/10.1038/nbt.2488) PMID: [23455439](https://pubmed.ncbi.nlm.nih.gov/23455439/)
10. Blazier AS, Papin JA. Integration of expression data in genome-scale metabolic network reconstructions. *Front Physiol.* 2012; 3:299. doi: [10.3389/fphys.2012.00299](https://doi.org/10.3389/fphys.2012.00299) PMID: [22934050](https://pubmed.ncbi.nlm.nih.gov/22934050/)
11. Kim MK, Lun DS. Methods for integration of transcriptomic data in genome-scale metabolic models. *Comput Struct Biotechnol J.* 2014 Aug; 11(18):59–65. doi: [10.1016/j.csbj.2014.08.009](https://doi.org/10.1016/j.csbj.2014.08.009) PMID: [25379144](https://pubmed.ncbi.nlm.nih.gov/25379144/)
12. Saha R, Chowdhury A, Maranas CD. Recent advances in the reconstruction of metabolic models and integration of omics data. *Curr Opin Biotechnol.* 2014 Oct; 29:39–45. doi: [10.1016/j.copbio.2014.02.011](https://doi.org/10.1016/j.copbio.2014.02.011) PMID: [24632194](https://pubmed.ncbi.nlm.nih.gov/24632194/)
13. Seo S, Lewin HA. Reconstruction of metabolic pathways for the cattle genome. *BMC Syst Biol.* 2009; 3:33. doi: [10.1186/1752-0509-3-33](https://doi.org/10.1186/1752-0509-3-33) PMID: [19284618](https://pubmed.ncbi.nlm.nih.gov/19284618/)
14. Sigurdsson MI, Jamshidi N, Steingrimsson E, Thiele I, Palsson BØ. A detailed genome-wide reconstruction of mouse metabolism based on human Recon 1. *BMC Syst Biol.* 2010; 4:140. doi: [10.1186/1752-0509-4-140](https://doi.org/10.1186/1752-0509-4-140) PMID: [20959003](https://pubmed.ncbi.nlm.nih.gov/20959003/)
15. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J. Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling. *Mol Syst Biol.* 2014; 10:721. doi: [10.1002/msb.145122](https://doi.org/10.1002/msb.145122) PMID: [24646661](https://pubmed.ncbi.nlm.nih.gov/24646661/)

16. Ghaffari P, Mardinoglu A, Asplund A, Shoaie S, Kampf C, Uhlen M, et al. Identifying anti-growth factors for human cancer cell lines through genome-scale metabolic modeling. *Sci Rep*. 2015; 5:8183. doi: [10.1038/srep08183](https://doi.org/10.1038/srep08183) PMID: [25640694](https://pubmed.ncbi.nlm.nih.gov/25640694/)
17. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol*. 2011; 7:501. doi: [10.1038/msb.2011.35](https://doi.org/10.1038/msb.2011.35) PMID: [21694718](https://pubmed.ncbi.nlm.nih.gov/21694718/)
18. Frezza C, Zheng L, Folger O, Rajagopalan KN, MacKenzie ED, Jerby L, et al. Haem oxygenase is synthetically lethal with the tumour suppressor fumarate hydratase. *Nature*. 2011 Sep; 477(7363):225–8. doi: [10.1038/nature10363](https://doi.org/10.1038/nature10363) PMID: [21849978](https://pubmed.ncbi.nlm.nih.gov/21849978/)
19. Gatto F, Miess H, Schulze A, Nielsen J. Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism. *Sci Rep*. 2015; 5:10738. doi: [10.1038/srep10738](https://doi.org/10.1038/srep10738) PMID: [26040780](https://pubmed.ncbi.nlm.nih.gov/26040780/)
20. Robaina Estévez S, Nikoloski Z. Generalized framework for context-specific metabolic model extraction methods. *Front Plant Sci*. 2014; 5:491. doi: [10.3389/fpls.2014.00491](https://doi.org/10.3389/fpls.2014.00491) PMID: [25285097](https://pubmed.ncbi.nlm.nih.gov/25285097/)
21. Ryu JY, Kim HU, Lee SY. Reconstruction of genome-scale human metabolic models using omics data. *Integr Biol (Camb)*. 2015 Mar; doi: [10.1039/C5IB00002E](https://doi.org/10.1039/C5IB00002E)
22. Gille C, Bölling C, Hoppe A, Bulik S, Hoffmann S, Hübner K, et al. HepatoNet1: a comprehensive metabolic reconstruction of the human hepatocyte for the analysis of liver physiology. *Mol Syst Biol*. 2010 Sep; 6:411. doi: [10.1038/msb.2010.62](https://doi.org/10.1038/msb.2010.62) PMID: [20823849](https://pubmed.ncbi.nlm.nih.gov/20823849/)
23. Sahoo S, Thiele I. Predicting the impact of diet and enzymopathies on human small intestinal epithelial cells. *Hum Mol Genet*. 2013 Jul; 22(13):2705–22. doi: [10.1093/hmg/ddt119](https://doi.org/10.1093/hmg/ddt119) PMID: [23492669](https://pubmed.ncbi.nlm.nih.gov/23492669/)
24. Chang RL, Xie L, Xie L, Bourne PE, Palsson BØ. Drug off-target effects predicted using structural analysis in the context of a metabolic network model. *PLoS Comput Biol*. 2010; 6(9):e1000938. doi: [10.1371/journal.pcbi.1000938](https://doi.org/10.1371/journal.pcbi.1000938) PMID: [20957118](https://pubmed.ncbi.nlm.nih.gov/20957118/)
25. Bordbar A, Lewis NE, Schellenberger J, Palsson BØ, Jamshidi N. Insight into human alveolar macrophage and M. tuberculosis interactions via metabolic reconstructions. *Mol Syst Biol*. 2010 Oct; 6:422. doi: [10.1038/msb.2010.68](https://doi.org/10.1038/msb.2010.68) PMID: [20959820](https://pubmed.ncbi.nlm.nih.gov/20959820/)
26. Bordbar A, Feist AM, Usaite-Black R, Woodcock J, Palsson BO, Famili I. A multi-tissue type genome-scale metabolic network for analysis of whole-body systems physiology. *BMC Syst Biol*. 2011; 5:180. doi: [10.1186/1752-0509-5-180](https://doi.org/10.1186/1752-0509-5-180) PMID: [22041191](https://pubmed.ncbi.nlm.nih.gov/22041191/)
27. Mardinoglu A, Agren R, Kampf C, Asplund A, Nookaew I, Jacobson P, et al. Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol Syst Biol*. 2013; 9:649. doi: [10.1038/msb.2013.5](https://doi.org/10.1038/msb.2013.5) PMID: [23511207](https://pubmed.ncbi.nlm.nih.gov/23511207/)
28. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun*. 2014; 5:3083. doi: [10.1038/ncomms4083](https://doi.org/10.1038/ncomms4083) PMID: [24419221](https://pubmed.ncbi.nlm.nih.gov/24419221/)
29. Pornputtpong N, Nookaew I, Nielsen J. Human metabolic atlas: an online resource for human metabolism. *Database (Oxford)*. 2015; 2015:bav068.
30. Wang Y, Eddy JA, Price ND. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol*. 2012; 6:153. doi: [10.1186/1752-0509-6-153](https://doi.org/10.1186/1752-0509-6-153) PMID: [23234303](https://pubmed.ncbi.nlm.nih.gov/23234303/)
31. Agren R, Bordel S, Mardinoglu A, Pornputtpong N, Nookaew I, Nielsen J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol*. 2012; 8(5):e1002518. doi: [10.1371/journal.pcbi.1002518](https://doi.org/10.1371/journal.pcbi.1002518) PMID: [22615553](https://pubmed.ncbi.nlm.nih.gov/22615553/)
32. Shlomi T, Cabili MN, Herrgård MJ, Palsson BØ, Ruppin E. Network-based prediction of human tissue-specific metabolism. *Nat Biotechnol*. 2008 Sep; 26(9):1003–10. doi: [10.1038/nbt.1487](https://doi.org/10.1038/nbt.1487) PMID: [18711341](https://pubmed.ncbi.nlm.nih.gov/18711341/)
33. Gatto F, Nookaew I, Nielsen J. Chromosome 3p loss of heterozygosity is associated with a unique metabolic network in clear cell renal carcinoma. *Proc Natl Acad Sci U S A*. 2014 Mar; 111(9):E866–75. doi: [10.1073/pnas.1319196111](https://doi.org/10.1073/pnas.1319196111) PMID: [24550497](https://pubmed.ncbi.nlm.nih.gov/24550497/)
34. Jerby L, Shlomi T, Ruppin E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol Syst Biol*. 2010 Sep; 6:401. doi: [10.1038/msb.2010.56](https://doi.org/10.1038/msb.2010.56) PMID: [20823844](https://pubmed.ncbi.nlm.nih.gov/20823844/)
35. Kumar A, Harrelson T, Lewis NE, Gallagher EJ, LeRoith D, Shiloach J, et al. Multi-tissue computational modeling analyzes pathophysiology of type 2 diabetes in MKR mice. *PLoS One*. 2014; 9(7):e102319. doi: [10.1371/journal.pone.0102319](https://doi.org/10.1371/journal.pone.0102319) PMID: [25029527](https://pubmed.ncbi.nlm.nih.gov/25029527/)
36. Våremo L, Scheele C, Broholm C, Mardinoglu A, Kampf C, Asplund A, et al. Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes. *Cell Rep*. 2015 May; 11(6):921–33. doi: [10.1016/j.celrep.2015.04.010](https://doi.org/10.1016/j.celrep.2015.04.010) PMID: [25937284](https://pubmed.ncbi.nlm.nih.gov/25937284/)

37. Becker SA, Palsson BO. Context-specific metabolic networks are consistent with experiments. *PLoS Comput Biol*. 2008 May; 4(5):e1000082. doi: [10.1371/journal.pcbi.1000082](https://doi.org/10.1371/journal.pcbi.1000082) PMID: [18483554](https://pubmed.ncbi.nlm.nih.gov/18483554/)
38. Schmidt BJ, Ebrahim A, Metz TO, Adkins JN, Palsson BØ, Hyduke DR. GIM3E: condition-specific models of cellular metabolism developed from metabolomics and expression data. *Bioinformatics*. 2013 Nov; 29(22):2900–8. doi: [10.1093/bioinformatics/btt493](https://doi.org/10.1093/bioinformatics/btt493) PMID: [23975765](https://pubmed.ncbi.nlm.nih.gov/23975765/)
39. Robaina Estévez S, Nikoloski Z. Context-Specific Metabolic Model Extraction Based on Regularized Least Squares Optimization. *PLoS One*. 2015; 10(7):e0131875. doi: [10.1371/journal.pone.0131875](https://doi.org/10.1371/journal.pone.0131875) PMID: [26158726](https://pubmed.ncbi.nlm.nih.gov/26158726/)
40. Vlassis N, Pacheco MP, Sauter T. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput Biol*. 2014 Jan; 10(1):e1003424. doi: [10.1371/journal.pcbi.1003424](https://doi.org/10.1371/journal.pcbi.1003424) PMID: [24453953](https://pubmed.ncbi.nlm.nih.gov/24453953/)
41. Mahadevan R, Schilling CH. The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab Eng*. 2003 Oct; 5(4):264–76. doi: [10.1016/j.ymben.2003.09.002](https://doi.org/10.1016/j.ymben.2003.09.002) PMID: [14642354](https://pubmed.ncbi.nlm.nih.gov/14642354/)
42. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010 Mar; 28(3):245–8. doi: [10.1038/nbt.1614](https://doi.org/10.1038/nbt.1614) PMID: [20212490](https://pubmed.ncbi.nlm.nih.gov/20212490/)
43. Lewis NE, Abdel-Haleem AM. The evolution of genome-scale models of cancer metabolism. *Front Physiol*. 2013; 4:237. doi: [10.3389/fphys.2013.00237](https://doi.org/10.3389/fphys.2013.00237) PMID: [24027532](https://pubmed.ncbi.nlm.nih.gov/24027532/)
44. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, et al. Towards a knowledge-based Human Protein Atlas. *Nat Biotechnol*. 2010 Dec; 28(12):1248–50. doi: [10.1038/nbt1210-1248](https://doi.org/10.1038/nbt1210-1248) PMID: [21139605](https://pubmed.ncbi.nlm.nih.gov/21139605/)
45. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Proteomics. Tissue-based map of the human proteome. *Science*. 2015 Jan; 347(6220):1260419. doi: [10.1126/science.1260419](https://doi.org/10.1126/science.1260419) PMID: [25613900](https://pubmed.ncbi.nlm.nih.gov/25613900/)
46. Price ND, Schellenberger J, Palsson BO. Uniform sampling of steady-state flux spaces: means to design experiments and to interpret enzymopathies. *Biophys J*. 2004 Oct; 87(4):2172–86. doi: [10.1529/biophysj.104.043000](https://doi.org/10.1529/biophysj.104.043000) PMID: [15454420](https://pubmed.ncbi.nlm.nih.gov/15454420/)
47. Wiback SJ, Famili I, Greenberg HJ, Palsson BØ. Monte Carlo sampling can be used to determine the size and shape of the steady-state flux space. *J Theor Biol*. 2004 Jun; 228(4):437–47. doi: [10.1016/j.jtbi.2004.02.006](https://doi.org/10.1016/j.jtbi.2004.02.006) PMID: [15178193](https://pubmed.ncbi.nlm.nih.gov/15178193/)
48. Thiele I, Price ND, Vo TD, Palsson BØ. Candidate metabolic network states in human mitochondria. Impact of diabetes, ischemia, and diet. *J Biol Chem*. 2005 Mar; 280(12):11683–95. doi: [10.1074/jbc.M409072200](https://doi.org/10.1074/jbc.M409072200) PMID: [15572364](https://pubmed.ncbi.nlm.nih.gov/15572364/)
49. Lewis NE, Schramm G, Bordbar A, Schellenberger J, Andersen MP, Cheng JK, et al. Large-scale in silico modeling of metabolic interactions between cell types in the human brain. *Nat Biotechnol*. 2010 Dec; 28(12):1279–85. doi: [10.1038/nbt.1711](https://doi.org/10.1038/nbt.1711) PMID: [21102456](https://pubmed.ncbi.nlm.nih.gov/21102456/)
50. Garg D, Henrich S, Salo-Ahen OMH, Myllykallio H, Costi MP, Wade RC. Novel approaches for targeting thymidylate synthase to overcome the resistance and toxicity of anticancer drugs. *J Med Chem*. 2010 Sep; 53(18):6539–49. doi: [10.1021/jm901869w](https://doi.org/10.1021/jm901869w) PMID: [20527892](https://pubmed.ncbi.nlm.nih.gov/20527892/)
51. Fridley BL, Batzler A, Li L, Li F, Matimba A, Jenkins GD, et al. Gene set analysis of purine and pyrimidine antimetabolites cancer therapies. *Pharmacogenet Genomics*. 2011 Nov; 21(11):701–12. doi: [10.1097/FPC.0b013e32834a48a9](https://doi.org/10.1097/FPC.0b013e32834a48a9) PMID: [21869733](https://pubmed.ncbi.nlm.nih.gov/21869733/)
52. Valenzuela MMA, Neidigh JW, Wall NR. Antimetabolite Treatment for Pancreatic Cancer. *Chemotherapy (Los Angel)*. 2014 Dec; 3(3).
53. Kaye SB. New antimetabolites in cancer chemotherapy and their clinical impact. *Br J Cancer*. 1998; 78 Suppl 3:1–7. doi: [10.1038/bjc.1998.747](https://doi.org/10.1038/bjc.1998.747) PMID: [9717984](https://pubmed.ncbi.nlm.nih.gov/9717984/)
54. Kim HU, Kim SY, Jeong H, Kim TY, Kim JJ, Choy HE, et al. Integrative genome-scale metabolic analysis of *Vibrio vulnificus* for drug targeting and discovery. *Mol Syst Biol*. 2011 Jan; 7:460. doi: [10.1038/msb.2010.115](https://doi.org/10.1038/msb.2010.115) PMID: [21245845](https://pubmed.ncbi.nlm.nih.gov/21245845/)
55. Comín-Anduix B, Boren J, Martinez S, Moro C, Centelles JJ, Trebukhina R, et al. The effect of thiamine supplementation on tumour proliferation. A metabolic control analysis study. *Eur J Biochem*. 2001 Aug; 268(15):4177–82. PMID: [11488910](https://pubmed.ncbi.nlm.nih.gov/11488910/)
56. Basu TK, Dickerson JW. The thiamin status of early cancer patients with particular reference to those with breast and bronchial carcinomas. *Oncology*. 1976; 33(5–6):250–2. PMID: [1026857](https://pubmed.ncbi.nlm.nih.gov/1026857/)
57. Lu'o'ng KVQ, Nguyen LTH. The role of thiamine in cancer: possible genetic and cellular signaling mechanisms. *Cancer Genomics Proteomics*. 2013; 10(4):169–85. PMID: [23893925](https://pubmed.ncbi.nlm.nih.gov/23893925/)
58. QUASTEL JH, CANTERO A. Inhibition of tumour growth by D-glucosamine. *Nature*. 1953 Feb; 171(4345):252–4. doi: [10.1038/171252a0](https://doi.org/10.1038/171252a0) PMID: [13036842](https://pubmed.ncbi.nlm.nih.gov/13036842/)

59. Friedman SJ, Skehan P. Membrane-active drugs potentiate the killing of tumor cells by D-glucosamine. *Proc Natl Acad Sci U S A*. 1980 Feb; 77(2):1172–6. doi: [10.1073/pnas.77.2.1172](https://doi.org/10.1073/pnas.77.2.1172) PMID: [6928667](https://pubmed.ncbi.nlm.nih.gov/6928667/)
60. Oh HJ, Lee JS, Song DK, Shin DH, Jang BC, Suh SI, et al. D-glucosamine inhibits proliferation of human cancer cells through inhibition of p70S6K. *Biochem Biophys Res Commun*. 2007 Sep; 360(4):840–5. doi: [10.1016/j.bbrc.2007.06.137](https://doi.org/10.1016/j.bbrc.2007.06.137) PMID: [17624310](https://pubmed.ncbi.nlm.nih.gov/17624310/)
61. Chesnokov V, Gong B, Sun C, Itakura K. Anti-cancer activity of glucosamine through inhibition of N-linked glycosylation. *Cancer Cell Int*. 2014; 14:45. doi: [10.1186/1475-2867-14-45](https://doi.org/10.1186/1475-2867-14-45) PMID: [24932134](https://pubmed.ncbi.nlm.nih.gov/24932134/)
62. Currie E, Schulze A, Zechner R, Walther TC, Farese RV Jr. Cellular fatty acid metabolism and cancer. *Cell Metab*. 2013 Aug; 18(2):153–61. doi: [10.1016/j.cmet.2013.05.017](https://doi.org/10.1016/j.cmet.2013.05.017) PMID: [23791484](https://pubmed.ncbi.nlm.nih.gov/23791484/)
63. Tennant DA, Durán RV, Gottlieb E. Targeting metabolic transformation for cancer therapy. *Nat Rev Cancer*. 2010 Apr; 10(4):267–77. doi: [10.1038/nrc2817](https://doi.org/10.1038/nrc2817) PMID: [20300106](https://pubmed.ncbi.nlm.nih.gov/20300106/)
64. Schell JC, Olson KA, Jiang L, Hawkins AJ, Van Vranken JG, Xie J, et al. A role for the mitochondrial pyruvate carrier as a repressor of the Warburg effect and colon cancer cell growth. *Mol Cell*. 2014 Nov; 56(3):400–13. doi: [10.1016/j.molcel.2014.09.026](https://doi.org/10.1016/j.molcel.2014.09.026) PMID: [25458841](https://pubmed.ncbi.nlm.nih.gov/25458841/)
65. Wong N, Ojo D, Yan J, Tang D. PKM2 contributes to cancer metabolism. *Cancer Lett*. 2015 Jan; 356(2 Pt A):184–91. doi: [10.1016/j.canlet.2014.01.031](https://doi.org/10.1016/j.canlet.2014.01.031) PMID: [24508027](https://pubmed.ncbi.nlm.nih.gov/24508027/)
66. Afratis N, Gialeli C, Nikitovic D, Tseggenidis T, Karousou E, Theocharis AD, et al. Glycosaminoglycans: key players in cancer cell biology and treatment. *FEBS J*. 2012 Apr; 279(7):1177–97. doi: [10.1111/j.1742-4658.2012.08529.x](https://doi.org/10.1111/j.1742-4658.2012.08529.x) PMID: [22333131](https://pubmed.ncbi.nlm.nih.gov/22333131/)
67. Yip GW, Smollich M, Götte M. Therapeutic value of glycosaminoglycans in cancer. *Mol Cancer Ther*. 2006 Sep; 5(9):2139–48. doi: [10.1158/1535-7163.MCT-06-0082](https://doi.org/10.1158/1535-7163.MCT-06-0082) PMID: [16985046](https://pubmed.ncbi.nlm.nih.gov/16985046/)
68. Asimakopoulou AP, Theocharis AD, Tzanakakis GN, Karamanos NK. The biological role of chondroitin sulfate in cancer and chondroitin-based anticancer agents. *In Vivo*. 2008; 22(3):385–9. PMID: [18610752](https://pubmed.ncbi.nlm.nih.gov/18610752/)
69. Kliner DJ, Gorski JP, Thonar EJ. Keratan sulfate levels in sera of patients bearing cartilage tumors. *Cancer*. 1987 Jun; 59(11):1931–5. doi: [10.1002/1097-0142\(19870601\)59:11%3C1931::AID-CNCR2820591116%3E3.0.CO;2-7](https://doi.org/10.1002/1097-0142(19870601)59:11%3C1931::AID-CNCR2820591116%3E3.0.CO;2-7) PMID: [2952259](https://pubmed.ncbi.nlm.nih.gov/2952259/)
70. Liu F, Liu Y, He C, Tao L, He X, Song H, et al. Increased MTHFD2 expression is associated with poor prognosis in breast cancer. *Tumour Biol*. 2014 Sep; 35(9):8685–90. doi: [10.1007/s13277-014-2111-x](https://doi.org/10.1007/s13277-014-2111-x) PMID: [24870594](https://pubmed.ncbi.nlm.nih.gov/24870594/)
71. Nilsson R, Jain M, Madhusudhan N, Sheppard NG, Strittmatter L, Kampf C, et al. Metabolic enzyme expression highlights a key role for MTHFD2 and the mitochondrial folate pathway in cancer. *Nat Commun*. 2014; 5:3128. doi: [10.1038/ncomms4128](https://doi.org/10.1038/ncomms4128) PMID: [24451681](https://pubmed.ncbi.nlm.nih.gov/24451681/)
72. Lehtinen L, Ketola K, Mäkelä R, Mpindi JP, Viitala M, Kallioniemi O, et al. High-throughput RNAi screening for novel modulators of vimentin expression identifies MTHFD2 as a regulator of breast cancer cell migration and invasion. *Oncotarget*. 2013 Jan; 4(1):48–63. doi: [10.18632/oncotarget.756](https://doi.org/10.18632/oncotarget.756) PMID: [23295955](https://pubmed.ncbi.nlm.nih.gov/23295955/)
73. Tedeschi PM, Markert EK, Gounder M, Lin H, Dvorzhinski D, Dolfi SC, et al. Contribution of serine, folate and glycine metabolism to the ATP, NADPH and purine requirements of cancer cells. *Cell Death Dis*. 2013; 4:e877. doi: [10.1038/cddis.2013.393](https://doi.org/10.1038/cddis.2013.393) PMID: [24157871](https://pubmed.ncbi.nlm.nih.gov/24157871/)
74. Bertino JR. Cancer research: from folate antagonism to molecular targets. *Best Pract Res Clin Haematol*. 2009 Dec; 22(4):577–82. doi: [10.1016/j.beha.2009.09.004](https://doi.org/10.1016/j.beha.2009.09.004) PMID: [19959110](https://pubmed.ncbi.nlm.nih.gov/19959110/)
75. Brusselmans K, Timmermans L, Van de Sande T, Van Veldhoven PP, Guan G, Shechter I, et al. Squalene synthase, a determinant of Raft-associated cholesterol and modulator of cancer cell proliferation. *J Biol Chem*. 2007 Jun; 282(26):18777–85. doi: [10.1074/jbc.M611763200](https://doi.org/10.1074/jbc.M611763200) PMID: [17483544](https://pubmed.ncbi.nlm.nih.gov/17483544/)
76. Fukuma Y, Matsui H, Koike H, Sekine Y, Shechter I, Ohtake N, et al. Role of squalene synthase in prostate cancer risk and the biological aggressiveness of human prostate cancer. *Prostate Cancer Prostatic Dis*. 2012 Dec; 15(4):339–45. doi: [10.1038/pcan.2012.14](https://doi.org/10.1038/pcan.2012.14) PMID: [22546838](https://pubmed.ncbi.nlm.nih.gov/22546838/)
77. Harada T, Chelala C, Crnogorac-Jurcevic T, Lemoine NR. Genome-wide analysis of pancreatic cancer using microarray-based techniques. *Pancreatol*. 2009; 9(1–2):13–24. doi: [10.1159/000178871](https://doi.org/10.1159/000178871) PMID: [19077451](https://pubmed.ncbi.nlm.nih.gov/19077451/)
78. Kim S, Kon M, DeLisi C. Pathway-based classification of cancer subtypes. *Biol Direct*. 2012; 7:21. doi: [10.1186/1745-6150-7-21](https://doi.org/10.1186/1745-6150-7-21) PMID: [22759382](https://pubmed.ncbi.nlm.nih.gov/22759382/)
79. Zheng J. Energy metabolism of cancer: Glycolysis versus oxidative phosphorylation (Review). *Oncol Lett*. 2012 Dec; 4(6):1151–1157. doi: [10.3892/ol.2012.928](https://doi.org/10.3892/ol.2012.928) PMID: [23226794](https://pubmed.ncbi.nlm.nih.gov/23226794/)
80. Moreno-Sánchez R, Rodríguez-Enríquez S, Marín-Hernández A, Saavedra E. Energy metabolism in tumor cells. *FEBS J*. 2007 Mar; 274(6):1393–418. doi: [10.1111/j.1742-4658.2007.05686.x](https://doi.org/10.1111/j.1742-4658.2007.05686.x) PMID: [17302740](https://pubmed.ncbi.nlm.nih.gov/17302740/)

81. Ralph SJ, Rodríguez-Enríquez S, Neuzil J, Moreno-Sánchez R. Bioenergetic pathways in tumor mitochondria as targets for cancer therapy and the importance of the ROS-induced apoptotic trigger. *Mol Aspects Med.* 2010 Feb; 31(1):29–59. doi: [10.1016/j.mam.2009.12.006](https://doi.org/10.1016/j.mam.2009.12.006) PMID: [20026172](https://pubmed.ncbi.nlm.nih.gov/20026172/)
82. Zu XL, Guppy M. Cancer metabolism: facts, fantasy, and fiction. *Biochem Biophys Res Commun.* 2004 Jan; 313(3):459–65. doi: [10.1016/j.bbrc.2003.11.136](https://doi.org/10.1016/j.bbrc.2003.11.136) PMID: [14697210](https://pubmed.ncbi.nlm.nih.gov/14697210/)
83. Khan AP, Rajendiran TM, Ateeq B, Asangani IA, Athanikar JN, Yocum AK, et al. The role of sarcosine metabolism in prostate cancer progression. *Neoplasia.* 2013 May; 15(5):491–501. PMID: [23633921](https://pubmed.ncbi.nlm.nih.gov/23633921/)
84. Sreekumar A, Poisson LM, Rajendiran TM, Khan AP, Cao Q, Yu J, et al. Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression. *Nature.* 2009 Feb; 457(7231):910–4. doi: [10.1038/nature07762](https://doi.org/10.1038/nature07762) PMID: [19212411](https://pubmed.ncbi.nlm.nih.gov/19212411/)
85. Bégin ME, Das UN, Ells G, Horrobin DF. Selective killing of human cancer cells by polyunsaturated fatty acids. *Prostaglandins Leukot Med.* 1985 Aug; 19(2):177–86. doi: [10.1016/0262-1746\(85\)90084-8](https://doi.org/10.1016/0262-1746(85)90084-8) PMID: [2864701](https://pubmed.ncbi.nlm.nih.gov/2864701/)
86. Bégin ME, Ells G, Das UN, Horrobin DF. Differential killing of human carcinoma cells supplemented with n-3 and n-6 polyunsaturated fatty acids. *J Natl Cancer Inst.* 1986 Nov; 77(5):1053–62. PMID: [3464797](https://pubmed.ncbi.nlm.nih.gov/3464797/)
87. Bougnoux P. n-3 polyunsaturated fatty acids and cancer. *Curr Opin Clin Nutr Metab Care.* 1999 Mar; 2(2):121–6. PMID: [10453342](https://pubmed.ncbi.nlm.nih.gov/10453342/)
88. Zastre JA, Hanberry BS, Sweet RL, McGinnis AC, Venuti KR, Bartlett MG, et al. Up-regulation of vitamin B1 homeostasis genes in breast cancer. *J Nutr Biochem.* 2013 Sep; 24(9):1616–24. doi: [10.1016/j.jnutbio.2013.02.002](https://doi.org/10.1016/j.jnutbio.2013.02.002) PMID: [23642734](https://pubmed.ncbi.nlm.nih.gov/23642734/)
89. Pettigrew CA, Clerkin JS, Cotter TG. DUOX enzyme activity promotes AKT signalling in prostate cancer cells. *Anticancer Res.* 2012 Dec; 32(12):5175–81. PMID: [23225414](https://pubmed.ncbi.nlm.nih.gov/23225414/)
90. Askari BS, Krajcinovic M. Dihydrofolate reductase gene variations in susceptibility to disease and treatment outcomes. *Curr Genomics.* 2010 Dec; 11(8):578–83. doi: [10.2174/138920210793360925](https://doi.org/10.2174/138920210793360925) PMID: [21629435](https://pubmed.ncbi.nlm.nih.gov/21629435/)
91. Zowczak M, Iskra M, Paszkowski J, Mańczak M, Torliński L, Wysocka E. Oxidase activity of ceruloplasmin and concentrations of copper and zinc in serum of cancer patients. *J Trace Elem Med Biol.* 2001; 15(2–3):193–6. doi: [10.1016/S0946-672X\(01\)80066-3](https://doi.org/10.1016/S0946-672X(01)80066-3) PMID: [11787988](https://pubmed.ncbi.nlm.nih.gov/11787988/)
92. Senra Varela A, Lopez Saez JJ, Quintela Senra D. Serum ceruloplasmin as a diagnostic marker of cancer. *Cancer Lett.* 1997 Dec; 121(2):139–45. PMID: [9570351](https://pubmed.ncbi.nlm.nih.gov/9570351/)
93. Fang J, Quinones QJ, Holman TL, Morowitz MJ, Wang Q, Zhao H, et al. The H⁺-linked monocarboxylate transporter (MCT1/SLC16A1): a potential therapeutic target for high-risk neuroblastoma. *Mol Pharmacol.* 2006 Dec; 70(6):2108–15. doi: [10.1124/mol.106.026245](https://doi.org/10.1124/mol.106.026245) PMID: [17000864](https://pubmed.ncbi.nlm.nih.gov/17000864/)
94. Moon JW, Lee SK, Lee YW, Lee JO, Kim N, Lee HJ, et al. Alcohol induces cell proliferation via hypermethylation of ADHFE1 in colorectal cancer cells. *BMC Cancer.* 2014; 14:377. doi: [10.1186/1471-2407-14-377](https://doi.org/10.1186/1471-2407-14-377) PMID: [24886599](https://pubmed.ncbi.nlm.nih.gov/24886599/)
95. Tae CH, Ryu KJ, Kim SH, Kim HC, Chun HK, Min BH, et al. Alcohol dehydrogenase, iron containing, 1 promoter hypermethylation associated with colorectal cancer differentiation. *BMC Cancer.* 2013; 13:142. doi: [10.1186/1471-2407-13-142](https://doi.org/10.1186/1471-2407-13-142) PMID: [23517143](https://pubmed.ncbi.nlm.nih.gov/23517143/)
96. Uberti J, Johnson RM, Talley R, Lightbody JJ. Decreased lymphocyte adenosine deaminase activity in tumor patients. *Cancer Res.* 1976 Jun; 36(6):2046–7. PMID: [1268856](https://pubmed.ncbi.nlm.nih.gov/1268856/)
97. Sufrin G, Tritsch GL, Mittelman A, Moore RH, Murphy GP. Adenosine deaminase activity in patients with renal adenocarcinoma. *Cancer.* 1977 Aug; 40(2):796–802. doi: [10.1002/1097-0142\(197708\)40:2%3C796::AID-CNCR2820400230%3E3.0.CO;2-O](https://doi.org/10.1002/1097-0142(197708)40:2%3C796::AID-CNCR2820400230%3E3.0.CO;2-O) PMID: [890659](https://pubmed.ncbi.nlm.nih.gov/890659/)
98. Kojima O, Majima T, Uehara Y, Yamane T, Fujita Y, Takahashi T, et al. Alteration of adenosine deaminase levels in peripheral blood lymphocytes of patients with gastric cancer. *Jpn J Surg.* 1985 Mar; 15(2):130–3. doi: [10.1007/BF02469742](https://doi.org/10.1007/BF02469742) PMID: [4010093](https://pubmed.ncbi.nlm.nih.gov/4010093/)
99. Park WJ, Kothapalli KSD, Lawrence P, Brenna JT. FADS2 function loss at the cancer hotspot 11q13 locus diverts lipid signaling precursor synthesis to unusual eicosanoid fatty acids. *PLoS One.* 2011; 6(11):e28186. doi: [10.1371/journal.pone.0028186](https://doi.org/10.1371/journal.pone.0028186) PMID: [22140540](https://pubmed.ncbi.nlm.nih.gov/22140540/)
100. Cui M, Wang Y, Sun B, Xiao Z, Ye L, Zhang X. MiR-205 modulates abnormal lipid metabolism of hepatoma cells via targeting acyl-CoA synthetase long-chain family member 1 (ACSL1) mRNA. *Biochem Biophys Res Commun.* 2014 Feb; 444(2):270–5. doi: [10.1016/j.bbrc.2014.01.051](https://doi.org/10.1016/j.bbrc.2014.01.051) PMID: [24462768](https://pubmed.ncbi.nlm.nih.gov/24462768/)
101. Huang W, Jin Y, Yuan Y, Bai C, Wu Y, Zhu H, et al. Validation and target gene screening of hsa-miR-205 in lung squamous cell carcinoma. *Chin Med J (Engl).* 2014; 127(2):272–8.
102. Cao Y, Dave KB, Doan TP, Prescott SM. Fatty acid CoA ligase 4 is up-regulated in colon adenocarcinoma. *Cancer Res.* 2001 Dec; 61(23):8429–34. PMID: [11731423](https://pubmed.ncbi.nlm.nih.gov/11731423/)

103. Chang Q, Zhang Y, Beezhold KJ, Bhatia D, Zhao H, Chen J, et al. Sustained JNK1 activation is associated with altered histone H3 methylations in human liver cancer. *J Hepatol*. 2009 Feb; 50(2):323–33. doi: [10.1016/j.jhep.2008.07.037](https://doi.org/10.1016/j.jhep.2008.07.037) PMID: [19041150](https://pubmed.ncbi.nlm.nih.gov/19041150/)
104. Huang P, Feng L, Oldham EA, Keating MJ, Plunkett W. Superoxide dismutase as a target for the selective killing of cancer cells. *Nature*. 2000 Sep; 407(6802):390–5. doi: [10.1038/35030140](https://doi.org/10.1038/35030140) PMID: [11014196](https://pubmed.ncbi.nlm.nih.gov/11014196/)
105. Morris SM Jr. Regulation of enzymes of the urea cycle and arginine metabolism. *Annu Rev Nutr*. 2002; 22:87–105. doi: [10.1146/annurev.nutr.22.110801.140547](https://doi.org/10.1146/annurev.nutr.22.110801.140547) PMID: [12055339](https://pubmed.ncbi.nlm.nih.gov/12055339/)
106. Ding C, Gao D, Wilding J, Trayhurn P, Bing C. Vitamin D signalling in adipose tissue. *Br J Nutr*. 2012 Dec; 108(11):1915–23. doi: [10.1017/S0007114512003285](https://doi.org/10.1017/S0007114512003285) PMID: [23046765](https://pubmed.ncbi.nlm.nih.gov/23046765/)
107. Mutt SJ, Hyppönen E, Saarnio J, Järvelin MR, Herzig KH. Vitamin D and adipose tissue—more than storage. *Front Physiol*. 2014; 5:228. PMID: [25009502](https://pubmed.ncbi.nlm.nih.gov/25009502/)
108. Ricciardelli C, Mayne K, Sykes PJ, Raymond WA, McCaul K, Marshall VR, et al. Elevated stromal chondroitin sulfate glycosaminoglycan predicts progression in early-stage prostate cancer. *Clin Cancer Res*. 1997 Jun; 3(6):983–92. PMID: [9815775](https://pubmed.ncbi.nlm.nih.gov/9815775/)
109. Ricciardelli C, Mayne K, Sykes PJ, Raymond WA, McCaul K, Marshall VR, et al. Elevated levels of versican but not decorin predict disease progression in early-stage prostate cancer. *Clin Cancer Res*. 1998 Apr; 4(4):963–71. PMID: [9563891](https://pubmed.ncbi.nlm.nih.gov/9563891/)
110. Ricciardelli C, Quinn DI, Raymond WA, McCaul K, Sutherland PD, Stricker PD, et al. Elevated levels of peritumoral chondroitin sulfate are predictive of poor prognosis in patients treated by radical prostatectomy for early-stage prostate cancer. *Cancer Res*. 1999 May; 59(10):2324–8. PMID: [10344737](https://pubmed.ncbi.nlm.nih.gov/10344737/)
111. Quinn DI, Henshall SM, Sutherland RL. Molecular markers of prostate cancer outcome. *Eur J Cancer*. 2005 Apr; 41(6):858–87. doi: [10.1016/j.ejca.2004.12.035](https://doi.org/10.1016/j.ejca.2004.12.035) PMID: [15808955](https://pubmed.ncbi.nlm.nih.gov/15808955/)
112. Kim DW, Jo YS, Jung HS, Chung HK, Song JH, Park KC, et al. An orally administered multitarget tyrosine kinase inhibitor, SU11248, is a novel potent inhibitor of thyroid oncogenic RET/papillary thyroid cancer kinases. *J Clin Endocrinol Metab*. 2006 Oct; 91(10):4070–6. doi: [10.1210/jc.2005-2845](https://doi.org/10.1210/jc.2005-2845) PMID: [16849418](https://pubmed.ncbi.nlm.nih.gov/16849418/)
113. Massicotte MH, Brassard M, Claude-Desroches M, Borget I, Bonichon F, Giraudet AL, et al. Tyrosine kinase inhibitor treatments in patients with metastatic thyroid carcinomas: a retrospective study of the TUTHYREF network. *Eur J Endocrinol*. 2014 Apr; 170(4):575–82. doi: [10.1530/EJE-13-0825](https://doi.org/10.1530/EJE-13-0825) PMID: [24424318](https://pubmed.ncbi.nlm.nih.gov/24424318/)
114. Kloss G, Leven M. Accumulation of radioiodinated tyrosine derivatives in the adrenal medulla and in melanomas. *Eur J Nucl Med*. 1979 Jun; 4(3):179–86. doi: [10.1007/BF00620482](https://doi.org/10.1007/BF00620482) PMID: [499239](https://pubmed.ncbi.nlm.nih.gov/499239/)
115. Vander Heiden MG, Cantley LC, Thompson CB. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science*. 2009 May; 324(5930):1029–33. doi: [10.1126/science.1160809](https://doi.org/10.1126/science.1160809) PMID: [19460998](https://pubmed.ncbi.nlm.nih.gov/19460998/)
116. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. *Nat Rev Cancer*. 2011 Feb; 11(2):85–95. doi: [10.1038/nrc2981](https://doi.org/10.1038/nrc2981) PMID: [21258394](https://pubmed.ncbi.nlm.nih.gov/21258394/)
117. Melé M, Ferreira PG, Reverter F, DeLuca DS, Monlong J, Sammeth M, et al. Human genomics. The human transcriptome across tissues and individuals. *Science*. 2015 May; 348(6235):660–5. doi: [10.1126/science.aaa0355](https://doi.org/10.1126/science.aaa0355) PMID: [25954002](https://pubmed.ncbi.nlm.nih.gov/25954002/)
118. Piscuoglio S, Hodi Z, Katabi N, Guerini-Rocco E, Macedo GS, Ng CKY, et al. Are acinic cell carcinomas of the breast and salivary glands distinct diseases? *Histopathology*. 2015 Feb;. doi: [10.1111/his.12673](https://doi.org/10.1111/his.12673)
119. Shen TK, Teknos TN, Toland AE, Senter L, Nagy R. Salivary gland cancer in BRCA-positive families: a retrospective review. *JAMA Otolaryngol Head Neck Surg*. 2014 Dec; 140(12):1213–7. doi: [10.1001/jamaoto.2014.1998](https://doi.org/10.1001/jamaoto.2014.1998) PMID: [25257187](https://pubmed.ncbi.nlm.nih.gov/25257187/)
120. Pia-Foschini M, Reis-Filho JS, Eusebi V, Lakhani SR. Salivary gland-like tumours of the breast: surgical and molecular pathology. *J Clin Pathol*. 2003 Jul; 56(7):497–506. doi: [10.1136/jcp.56.7.497](https://doi.org/10.1136/jcp.56.7.497) PMID: [12835294](https://pubmed.ncbi.nlm.nih.gov/12835294/)
121. Hemminki K, Jiang Y, Steineck G. Skin cancer and non-Hodgkin's lymphoma as second malignancies. markers of impaired immune function? *Eur J Cancer*. 2003 Jan; 39(2):223–9. PMID: [12509955](https://pubmed.ncbi.nlm.nih.gov/12509955/)
122. Birkeland SA, Storm HH, Lamm LU, Barlow L, Blohmé I, Forsberg B, et al. Cancer risk after renal transplantation in the Nordic countries, 1964–1986. *Int J Cancer*. 1995 Jan; 60(2):183–9. PMID: [7829213](https://pubmed.ncbi.nlm.nih.gov/7829213/)
123. Baenke F, Peck B, Miess H, Schulze A. Hooked on fat: the role of lipid synthesis in cancer metabolism and tumour development. *Dis Model Mech*. 2013 Nov; 6(6):1353–63. doi: [10.1242/dmm.011338](https://doi.org/10.1242/dmm.011338) PMID: [24203995](https://pubmed.ncbi.nlm.nih.gov/24203995/)

124. Mashima T, Seimiya H, Tsuruo T. De novo fatty-acid synthesis and related pathways as molecular targets for cancer therapy. *Br J Cancer*. 2009 May; 100(9):1369–72. doi: [10.1038/sj.bjc.6605007](https://doi.org/10.1038/sj.bjc.6605007) PMID: [19352381](https://pubmed.ncbi.nlm.nih.gov/19352381/)
125. Gaude E, Frezza C. Defects in mitochondrial metabolism and cancer. *Cancer Metab*. 2014; 2:10. doi: [10.1186/2049-3002-2-10](https://doi.org/10.1186/2049-3002-2-10) PMID: [25057353](https://pubmed.ncbi.nlm.nih.gov/25057353/)
126. Viale A, Corti D, Draetta GF. Tumors and Mitochondrial Respiration: A Neglected Connection. *Cancer Res*. 2015 Sep;. doi: [10.1158/0008-5472.CAN-15-0491](https://doi.org/10.1158/0008-5472.CAN-15-0491)
127. Amelio I, Cutruzzolá F, Antonov A, Agostini M, Melino G. Serine and glycine metabolism in cancer. *Trends Biochem Sci*. 2014 Apr; 39(4):191–8. doi: [10.1016/j.tibs.2014.02.004](https://doi.org/10.1016/j.tibs.2014.02.004) PMID: [24657017](https://pubmed.ncbi.nlm.nih.gov/24657017/)
128. Gurobi Optimizer Reference Manual [Website]; 2014. Available from: <http://www.gurobi.com/index>.
129. Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc*. 2011 Sep; 6(9):1290–307. doi: [10.1038/nprot.2011.308](https://doi.org/10.1038/nprot.2011.308) PMID: [21886097](https://pubmed.ncbi.nlm.nih.gov/21886097/)
130. Qutub AA. The Qutub Lab [Website]; 2015. Available from: <http://qutublab.org/apps-code-tools.html#Metabolic>.
131. Kaufman DE, Smith RL. Direction choice for accelerated convergence in hit-and-run sampling. *Operations Research*. 1998; 46(1):84–95. doi: [10.1287/opre.46.1.84](https://doi.org/10.1287/opre.46.1.84)