FEBS openbio

FEBS PRESS
science publishing by scientists

Optimization of a modeling platform to predict oncogenes from genome-scale metabolic networks of non-small-cell lung cancers

You-Tyun Wang, Min-Ru Lin, Wei-Chen Chen, Wu-Hsiung Wu and Feng-Sheng Wang

Department of Chemical Engineering, National Chung Cheng University, Chiayi, Taiwan

Keywords
cancer cell metabolism; constraint-based modeling; flux balance analysis; tissue-specific metabolic models; trilevel optimization

Correspondence
F.-S. Wang, Department of Chemical Engineering, National Chung Cheng University, Chiayi, Taiwan
E-mail: chmfsw@ccu.edu.tw

(Received 16 March 2021, revised 19 May 2021, accepted 16 June 2021)

doi:10.1002/2211-5463.13231

Cancer cell dysregulations result in the abnormal regulation of cellular metabolic pathways. By simulating this metabolic reprogramming using constraint-based modeling approaches, oncogenes can be predicted, and this knowledge can be used in prognosis and treatment. We introduced a trilevel optimization problem describing metabolic reprogramming for inferring oncogenes. First, this study used RNA-Seq expression data of lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC) samples and their healthy counterparts to reconstruct tissue-specific genome-scale metabolic models and subsequently build the flux distribution pattern that provided a measure for the oncogene inference optimization problem for determining tumorigenesis. The platform detected 45 genes for LUAD and 84 genes for LUSC that lead to tumorigenesis. A high level of differentially expressed genes was not an essential factor for determining tumorigenesis. The platform indicated that pyruvate kinase (PKM), a well-known oncogene with a low level of differential gene expression in LUAD and LUSC, had the highest fitness among the predicted oncogenes based on computation. By contrast, pyruvate kinase L/R (PKLR), an isozyme of PKM, had a high level of differential gene expression in both cancers. Phosphatidylserine synthase 1 (*PTDSS1*), an oncogene in LUAD, was inferred to have a low level of differential gene expression, and overexpression could significantly reduce survival probability. According to the factor analysis, *PTDSS1* characteristics were close to those of the template, but they were unobvious in LUSC. Angiotensin-converting enzyme 2 (*ACE2*) has recently garnered widespread interest as the SARS-CoV-2 virus receptor. Moreover, we determined that *ACE2* is an oncogene of LUSC but not of LUAD. The platform developed in this study can identify oncogenes with low levels of differential expression and be used to identify potential therapeutic targets for cancer treatment.

Abbreviations
ACAA2, acetyl-CoA acyltransferase 2; ACE2, angiotensin-converting enzyme 2; BCAT1, branched chain amino acid transaminase 1; COBRA, constraint-based reconstruction and analysis toolbox; CORDA, cost optimization reaction dependency assessment; COSMIC, Catalogue Of Somatic Mutations In Cancer; DEG, differentially expressed gene; ENO1, enolase 1; FVA, flux variability analysis; GLO1, glyoxalase I; GPR, gene–protein–reaction; GSMN, genome-scale metabolic network; HPA, human protein atlas; iMAT, integrative metabolic analysis tool; LUAD, lung adenocarcinoma; LUSC, lung squamous cell carcinoma; MFVA, metabolite-metabolic networflow variability analysis; NCI-60, 60 human tumor cell line anticancer drug screen from the US National Cancer Institute; NF-κB, nuclear factor-κB; NHDE, nested hybrid differential evolution; NSCLC, non-small-cell lung carcinoma; PKLR, pyruvate kinase L/R; PKM, pyruvate kinase; PSPH, phosphoserine phosphatase; PTDSS1, phosphatidylserine synthase 1; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; SHMT1, serine hydroxymethyltransferase 1; SLC, solute carrier; TCGA, The Cancer Genome Atlas; TLOP, trilevel optimization problem; VMH, Virtual Metabolic Human.

FEBS Open Bio **11** (2021) 2078–2094 © 2021 The Authors. *FEBS Open Bio* published by John Wiley & Sons Ltd on behalf of Federation of European Biochemical Societies
This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

Lung carcinoma is one of the most common malignancies, resulting in the largest number of cancer-related deaths worldwide [1]. Two main subtypes of lung cancer exist, namely small-cell lung carcinoma and non-small-cell lung carcinoma (NSCLC), accounting for 15% and 85% of all lung cancers, respectively [2]. Lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), two main subtypes of NSCLC, are predominant lung cancers, accounting for 40% and 33%, respectively, of cancer deaths worldwide [3]. Lung carcinoma is related to genetic and epigenetic dysregulation, and understanding its biological mechanism is crucial for developing effective treatment.

Systems biology approaches and big database mining have been applied to construct genetic and epigenetic networks with next-generation sequencing data of LUAD and LUSC for comparing the differential genetic and epigenetic progression mechanisms [4]. Such approaches might enable the deciphering of genotype discrepancy for both tissues. However, genes and their expression alone do not always constitute a reliable indicator of cellular phenotype. The recent availability of omics datasets allows the analysis of cellular characteristics at the levels of genes, mRNAs, proteins, and metabolites. A genome-scale metabolic network (GSMN) can offer a biological mechanism that links genotype to phenotype; it can help us understand cell physiology and certain disease phenotypes caused by metabolic dysregulation [5,6]. Human metabolism is complex and specialized in different tissue and cell types. The reprogramming of tissue-specific metabolism in GSMNs will provide deeper insights into the metabolic basis of various physiological and pathological processes. Such metabolic reprogramming approaches have been applied to predict oncogenes, essential enzymes, and drug targets for developing novel medical treatments [7–11].

Cancer metabolism is an emerging hallmark of cancer [12]. Genetic alterations and epigenetic modifications of cancer cells result in the abnormal regulation of cellular metabolic pathways that differ from normal cells. GSMNs combined with constraint-based modeling approaches can predict the metabolic reprogramming of cancer cells to reveal oncogenes, essential enzymes, and drug targets for developing novel medical treatments [13–22]. The first large-scale reconstructed metabolic model for cancer was built based on the gene expression data of all cancer cell lines in the NCI-60 collection and the human general metabolic network (Recon 1) [14,23]. This model was applied on the identification of essential genes and cytostatic drug targets of cancer cell lines [14] and prediction of metabolic targets for inhibiting cancer migration [23]. The GSMN (*iHepatocytes2322*) for hepatocytes was reconstructed by extending Recon 1 using data from Human Metabolic Reaction 2.0 database and proteomics data in Human Protein Atlas (https://metabolicatlas.org/). This GSMN was used to identify PSPH, SHMT1, and BCAT1 as potential therapeutic targets for the treatment of nonalcoholic steatohepatitis using the transcriptomics data obtained from patients with nonalcoholic fatty liver disease [16]. Another small-scale constraint-based model was created and combined with machine-learning techniques to investigate the mechanism of pyruvate dehydrogenase under hypoxia [13].

Due to the complexity and specialization of human metabolism in different tissue and cancer cells, mapping tissue-specific metabolisms in GSMNs can advance the understanding of cancer metabolism [24]. Recon 2.2 and Recon 3D are the most comprehensive human genome-scale network reconstructions [25,26]. Recon 2.2 was incorporated with the Human Protein Atlas (HPA) [27] to reconstruct GSMNs of the colorectal tissue [18] and head-and-neck tissue [21] at normal and cancerous states. Recon 2M.2 [28,29] integrated with RNA-Seq data from NCI-60 cell lines presented a systematic framework for the generation of gene–transcript–protein–reaction that enables the accurate prediction of metabolic behaviors. Recon 3D is a human general genome-scale network reconstruction that includes three-dimensional metabolite and protein structure data and enables an integrated analysis of metabolic functions in humans. In this study, we first applied the CORDA method [30] integrated with The Cancer Genome Atlas (TCGA) database [31] and Recon 3D to reconstruct GSMNs for LUAD, LUSC, and their healthy cells. Multivariate analysis was used to analyze the reactions, metabolites, and enzyme-encoding genes of these GSMNs to discriminate differential expressions between normal and cancer cells. The oncogene inference optimization formulation [18,21] was used to mimic gene screening procedures in a wet laboratory to evaluate the mechanism by which gene dysregulations induce tumorigenesis.

Materials and methods

Reconstruction of tissue-specific metabolic models

This study applied RNA-Seq data from TCGA database to reconstruct genome-scale metabolic models (Fig. 1) for LUAD and LUSC and their corresponding healthy tissues.

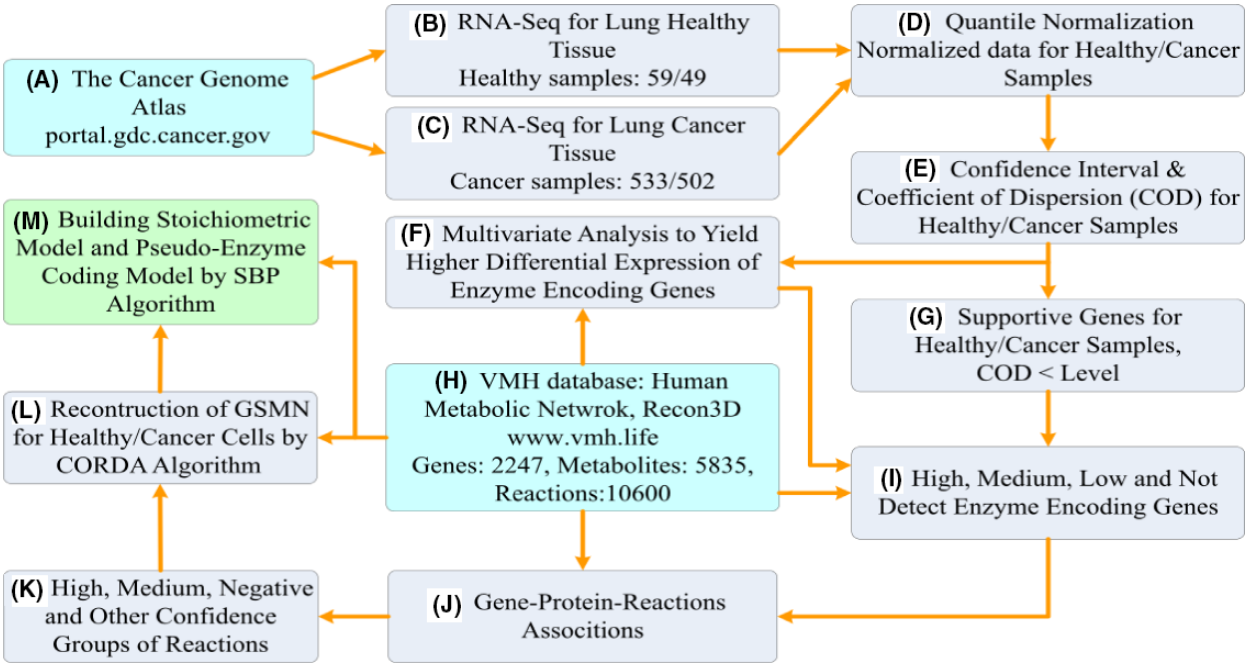

Fig. 1. Roadmap of the reconstruction of genome-scale metabolic models. Roadmap of the reconstruction of genome-scale metabolic models for LUAD and LUSC and their corresponding healthy tissues. (A) Download RNA-Seq data of LUAD and LUSC from the TCGA database. (B–G) Statistical analysis of download RNA-Seq data to generate input information for the CORDA algorithm. (H) The general human GSMN (Recon 3D) was downloaded from VHM database (https://www.vmh.life) and used as a base model. (I) Classify enzyme-encoding genes into four classes. (J) Compute gene–protein–reaction associations using the enzyme-encoding genes and Recon 3D general model. (K) Identify reactions having various confidence indices. (L) Reconstruct a tissue-specific metabolic model using the CORDA algorithm and Recon 3D general model. (M) Build the stoichiometric and GPR models in GAMS format for simulation.

These metabolic models were entirely flux-consistent inspected by the 'findFluxConsistentSubset' function of COBRA toolbox. In total, 533 and 502 samples for LUAD and LUSC and 59 and 49 corresponding healthy samples were collected, respectively. Quantile normalization was applied to normalize the raw data for healthy and cancerous samples to compute the mean, confidence interval, and coefficient of dispersion for each gene. Such data were then used to evaluate supportive genes and obtain a high differential expression of enzyme-encoding genes between the cancer and healthy cells. Recon 3D consisted of 2247 enzyme-encoding genes, which were classified into four levels based on their participation, namely high, medium, low, and not detected. Four groups of confidence reactions, namely high, medium, negative, and others, were obtained through gene–protein–reaction association in Recon 3D. The tissue-specific GSMNs of healthy and cancer cells were reconstructed using the CORDA algorithm and saved in SBML format. We developed a systems biology program (SBP) platform to automatically develop the stoichiometric models and GPR model in GAMS format to perform the simulation.

Oncogene inference optimization

The oncogene inference optimization framework was modified from a trilevel optimization problem (TLOP) that has been applied to analyze colorectal cancer [18]. In this study, the objectives in the TLOP considered the consistency between the change trends of mutant fluxes/metabolite flows compared with the template and that of cancer cells, and applied fuzzy equal constraints to quantitatively limit the change ratios of the mutant as close as possible to the template. These objectives in the modified approach can qualitatively and quantitatively optimize the flux pattern of a mutant as close as possible to the template. This oncogene inference optimization framework has been used to detect tumor suppressor genes in head-and-neck squamous cell carcinoma [21] and can be used as a mimic experiment for gene screening to predict oncogenes, similar to metabolic transformation algorithm for identification of drug targets [10,32]. The outer optimization aims to infer the dysregulation of enzyme-encoding genes that alter the metabolism of normal cells leading to cancer, and the inner optimization problems present perturbed behaviors of

mutant cells. The mathematical formulation is expressed as follows:

$$\begin{cases} \text{Outer optimization problem:} \\ \text{Similarity ratio of metabolite}-\text{flow rates and fluxes to the template:} \\ \max\limits_{\delta,\, z_i} SR_M,\ \max\limits_{\delta,\, z_i} SR_F \\ \text{Fuzzy equal grade of metabolite}-\text{flow rates and fluxes compared to the template:} \\ \widetilde{Equal}\limits_{\delta,\, z_i} LFC_M^{MUBL} \approx LFC_M^{CABL},\ \widetilde{Equal}\limits_{\delta,\, z_i} LFC_F^{MUBL} \approx LFC_F^{CABL} \\ \text{subject to the inner optimization problems:} \\ \begin{cases} \text{Flux balance analysis (FBA) problem} \\ \max\limits_{\mathbf{v}_{f/b}} obj \equiv (w_{ATP} v_{ATP} + w_{biomass} v_{biomass}) \\ \text{subject to} \\ \mathbf{N}(\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\ v_{f/b,i}^{LB} \le v_{f/b,i} \le v_{f/b,i}^{UB},\ z_i \notin \Omega^{MU} \\ v_{f/b,j}^{LB,MU} \le v_{f/b,j} \le v_{f/b,j}^{UB,MU},\ z_j \in \Omega^{MU} \end{cases} \\ \begin{cases} \text{Uniform flux distribution (UFD) problem} \\ \min\limits_{\mathbf{v}_{f/b}} \sum\limits_{i \in \Omega^{Int}} (v_{f,k})^2 + (v_{b,k})^2 \\ \text{subject to} \\ \mathbf{N}(\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\ v_{f/b,i}^{LB} \le v_{f/b,i} \le v_{f/b,i}^{UB},\ z_i \notin \Omega^{MU} \\ v_{f/b,j}^{LB,MU} \le v_{f/b,j} \le v_{f/b,j}^{UB,MU},\ z_j \in \Omega^{MU} \\ obj \ge obj^* \end{cases} \end{cases} \tag{1}$$

where $v_{f/b}$ is the forward/backward flux vector of reversible reactions; $\mathbf{N}$ is an $m \times n$ stoichiometric matrix where m is the number of metabolites, and n is the number of reactions; $v_{f/b,i}^{LB}$ and $v_{f/b,i}^{UB}$ are the positive lower and upper bounds of the ith forward/backward flux, respectively; $v_{f/b,i}^{LB,MU}$ and $v_{f/b,i}^{UB,MU}$ are the positive lower and upper bounds of the ith upregulation, downregulation, or knockout flux in the set of mutated reactions $\mathbf{\Omega}^{MU}$ due to the ith enzyme dysregulation, which is determined using the GPR model; obj^* is the maximum cellular objective obtained from the flux balance analysis (FBA) problem; $\widetilde{Equal}$ is the fuzzy equal objective function that represent the fuzzy goals. For example, the $LFC_m{}^{MUBL}$ and $LFC_m{}^{CABL}$ should be restored to a state that is as close as possible; the integer vector z is used to determine mutated enzymes; and δ is the regulated strength parameter for the mutants with a value within (0, 1].

The GPR model used a pseudo-enzyme coding number strategy to represent GPR associations in Recon 3D [18]. It identified redundant pseudoenzymes and isozymes in the model such that the reactions were catalyzed through reduced association. Therefore, pseudoenzymes were applied to determine modulated genes, and the level of the mutated bounds was computed using the following equations:

$$\begin{cases} \text{Upregulation:} \\ \begin{cases} (1-\delta) v_{f,i}^{basal} + \delta v_{f,i}^{UB} \le v_{f,i} \le v_{f,i}^{UB} \\ v_{b,i}^{LB} \le v_{b,i} \le (1-\delta) v_{b,i}^{basal} + \delta v_{b,i}^{LB};\ i \in \Omega^{MU} \end{cases} \\ \text{Downregulation:} \\ \begin{cases} v_{f,i}^{LB} \le v_{f,i} \le (1-\delta) v_{f,i}^{basal} + \delta v_{f,i}^{LB} \\ (1-\delta) v_{b,i}^{basal} + \delta v_{b,i}^{UB} \le v_{b,i} \le v_{b,i}^{UB};\ i \in \Omega^{MU} \setminus \Omega^{IZ} \end{cases} \\ \begin{cases} v_{f,i}^{LB} \le v_{f,i} \le v_{f,i}^{UB} \\ v_{b,i}^{LB} \le v_{b,i} \le v_{b,i}^{UB};\ i \in \Omega^{MU} \cap \Omega^{IZ} \end{cases} \\ \text{Knockout:} \\ \begin{cases} v_{f,i} = 0 \\ v_{b,i} = 0;\ i \in \Omega^{MU} \setminus \Omega^{IZ} \end{cases} \\ \begin{cases} v_{f,i}^{LB} \le v_{f,i} \le v_{f,i}^{UB} \\ v_{b,i}^{LB} \le v_{b,i} \le v_{b,i}^{UB};\ i \in \Omega^{MU} \cap \Omega^{IZ} \end{cases} \end{cases} \tag{2}$$

where $v_{f/b,i}^{basal}$ is the basal flux in the normal state, and $\mathbf{\Omega}^{IZ}$ is the set of reactions regulated by isozymes represented in the GPR model.

Multiple objectives are considered in the outer optimization problem in Eqn (1). In the first and second objectives, the similarity ratios of metabolite-flow rates and fluxes (SR_M and SR_F) are maximized for determining a dysregulated metabolite-flow/flux pattern that is as similar as possible to the template. The third and fourth objectives are used to obtain a mutant log2 fold change, $LFC_{M/F}^{MUBL}$, of metabolite-flow rates/fluxes as close as possible to that of the template, $LFC_{M/F}^{CABL}$. The similarity ratios of the metabolite-flow rates/fluxes (SR_M and SR_F) for a mutant are evaluated as follows:

$$SR_{M/F} = \frac{\sum_{m=1}^{N_{M/F}} \left|\mu_m^{M/F}\right|}{N_{M/F}} \quad (3)$$

where the similarity indicator ($\mu_m^{M/F}$) for each metabolite-flow rate or flux in the metabolic network is defined as follows:

$$\mu_m^{M/F} = \begin{cases} 1, & \text{if } LFC_{M/F,m}^{MUBL} > tol_+ \text{ and } LFC_{M/F,m}^{CABL} > tol_+ \\ -1, & \text{if } LFC_{M/F,m}^{MUBL} < tol_- \text{ and } LFC_{M/F,m}^{CABL} < tol_- \\ 0, & \text{otherwise} \end{cases} \quad (4)$$

where log2 fold changes, $LFC_{M/F}^{CABL}$, of the metabolite-flow rates/fluxes of templates are computed from the reconstructed GSMNs for cancer and normal cells in advance.

The tolerances for increase or decrease are defined as $tol_+ = \log_2(1+\varepsilon)$ and $tol_- = \log_2(1-\varepsilon)$, respectively, and ϵ is the percentage of flux alteration. A numerical example is provided (Doc. S1) to illustrate the computation of flux template, similarity ratio, and logarithmic fold change. The log2 fold change between the metabolite-flow rate of the mth metabolite in cancer or dysregulated (denoted as MU) and normal states (denoted as BL) is computed as follows:

$$LFC_{M,m}^{MUBL} = \log_2\left(\frac{r_{m,MU}}{r_{m,BL}}\right) \quad (5)$$

where the metabolite-flow rate is a pool of flux-sum synthesis rates of the mth metabolite in the dysregulated or normal cells, and expressed as follows:

$$r_m = \sum_{s\in\Omega^c}\left(\sum_{N_{ij}>0,\,j} N_{ij}v_{f,j} - \sum_{N_{ij}<0,\,j} N_{ij}v_{b,j}\right),\ m\in\Omega^m \quad (6)$$

where N_{ij} is the stoichiometric coefficient of the ith metabolite participating in the jth reaction; $\mathbf{\Omega}^c$ is the set of metabolites located in different compartments; and $\mathbf{\Omega}^m$ is the set of metabolites in the GSMN. The bracket in Eqn (6) indicates the synthesis rates of the mth metabolite at its located compartment (i.e., sum up the forward fluxes, $v_{f,j}$, and backward fluxes, $v_{b,j}$, of the metabolite). The log2 fold change of the forward/backward flux in dysregulated and normal states is defined as follows:

$$LFC_{F,f/b}^{MUBL} = \log_2\left(\frac{v_{f/b,MU}}{v_{f/b,BL}}\right) \quad (7)$$

The fold changes of the template, $LFC_{M/F,m}^{CABL}$ can be obtained by applying to above definition of $LFC_{M/F,m}^{MUBL}$ on a reconstructed cancer model instead of dysregulated models. Note that the templates are the flux distribution patterns for cancer and normal tissue. The templates can obtain from clinical data if they are available; otherwise, they were computed from the FBA and UFD problems without the dysregulated restrictions.

Fitness evaluation

The TLOP in Eqn (1) is a mixed-integer optimization problem that is NP-hard [33]. Classical algorithms for solving bilevel optimization problems use duality theory to convert the inner-level optimization problem into constraints in the outer-level problem. However, duality transformation is difficult for multilevel optimization problems, such as the TLOP in this study. We applied the NHDE algorithm (Doc. S2), which has been used to solve oncogene inference problems [18,21], to infer the oncogenes of LUAD and LUSC. The problem [Eqn (1)] consisted of the crisp objectives and fuzzy equal objective to introduce a combination of weighted-sum and minimum decisions for evaluating the fitness, η_D, which was used in the NHDE algorithm as follows:

$$\eta_D = [(\eta_S + \eta_E)/2 + \min\{\eta_S, \eta_E\}]/2 \quad (8)$$

where η_S is the average similarity ratios of SR_M and SR_F, and the membership grade, η_E, is used to measure how close the fuzzy equal objective (logarithmic fold change of the metabolite-flow rates/fluxes) of the mutant is to the template.

The fuzzy equal objective for each metabolite is quantified by eliciting a membership function. In this study, the membership function is a combination of the left-hand (η_m^L) and right-hand (η_m^R) side linear membership functions, as shown in Fig. 2. The mathematical expressions are respectively formulated as follows:

$$\eta_m^L\left(LFC_m^{MUBL}\right) = \frac{LFC_m^{MUBL} - LFC_m^{CABL,LB}}{LFC_m^{CABL} - LFC_m^{CABL,LB}} \quad (9)$$

$$\eta_m^R\left(LFC_m^{MUBL}\right) = \frac{LFC_m^{CABL,UB} - LFC_m^{MUBL}}{LFC_m^{CABL,UB} - LFC_m^{CABL}} \quad (10)$$

where $LFC_m^{CABL,LB}$ and $LFC_m^{CABL,UB}$ are the lower and upper bounds of the log2 fold change of metabolite-flow rate/flux in the cancer and basal states for the mth metabolite, and their levels can be provided by the user in advance as follows:

$$LFC_m^{CABL,LB} = \begin{cases} LFC_m^{CABL}/4, & \text{if } LFC_m^{CABL} > 0 \\ 4LFC_m^{CABL}, & \text{if } LFC_m^{CABL} < 0 \end{cases} \quad (11)$$

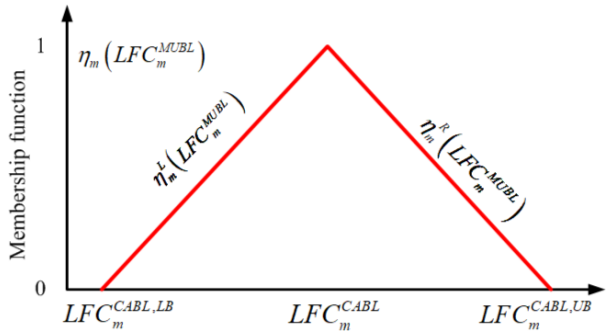

Fig. 2. Fuzzy equal membership function. Fuzzy equal membership grade, $\eta_m\left(LFC_m^{MUBL}\right)$, of a mutant consisting of a left-hand side membership function, η_m^L, and right-hand side membership function, η_m^R, applied to evaluate the closeness for LFC_m^{MUBL} of the mutant to the template. The membership grade is zero if LFC_m^{MUBL} exceeds the lower bound ($LFC_m^{CABL,LB}$) or upper bound ($LFC_m^{CABL,UB}$). Conversely, the membership grade is between 0 and 1 if the membership function is within its bounds.

$$LFC_m^{CABL,UB} = \begin{cases} 4LFC_m^{CABL}, & \text{if } LFC_m^{CABL} \geq 0 \\ LFC_m^{CABL}/4, & \text{if } LFC_m^{CABL} \leq 0 \end{cases} \quad (12)$$

The fuzzy equal membership grade for each metabolite/flux is elicited as follows:

$$\eta_m\left(LFC_m^{MUBL}\right) = \max\left\{\min\left[\eta_m^L\left(LFC_m^{MUBL}\right), \eta_m^R\left(LFC_m^{MUBL}\right), 1\right], 0\right\} \quad (13)$$

The decision grade of the network sums the membership grades for all metabolites/fluxes as $\eta_E = \frac{1}{M}\sum_{m=1}^{M}\eta_m\left(LFC_m^{MUBL}\right)$, which is between 0 and 1. The decision grade differs from a least-square error criterion in regression methods. According to the definition of the fuzzy equal membership function in Eqn (13), the grade is zero if LFC_m^{MUBL} exceeds the lower bound ($LFC_m^{CABL,LB}$) or upper bound ($LFC_m^{CABL,UB}$). Conversely, the membership grade is between 0 and 1 if the membership function is within its bounds.

Metabolite-flow variability analysis

Generally, the optimal fluxes of FBA in the TLOP problem could have many distributions with an identical objective value. Bias in similarity ratios may be yielded through such an evaluation. To overcome such a drawback, flux variability analysis (FVA) can be applied in a posterior inspection to determine the maximum and minimum values of all fluxes that satisfy the constraints and allow for the same optimal objective value. FVA can be applied to compute minimum and maximum fluxes to yield a flux space of a metabolic network [34]. Moreover, it must cover all objective values of the cell growth because cancer cell growth may not proliferate sustainably at its maximum rate. In this study, we introduced metabolite-flow variability analysis (MFVA) to compute the minimum and maximum quantities of each metabolite for the normal model and the mutants, respectively. The MFVA formulation was expressed as follows:

$$\begin{cases} \text{MFVA Problem for cancer, normal, and mutant cases} \\ \max_{\zeta\in(0,\,1]} / \min r_m \\ \text{subject to the inner optimization problems:} \\ \begin{cases} \text{FBA Problem:} \\ \max_{\mathbf{v}_{f/b}} obj \equiv \left(w_{ATP}v_{ATP} + w_{biomass}v_{biomass}\right) \\ \mathbf{N}^{CA/BL}\left(\mathbf{v}_f - \mathbf{v}_b\right) = \mathbf{0} \\ v_{f/b,j}^{LB} \leq v_{f/b,j} \leq v_{f/b,j}^{UB} \end{cases} \\ \begin{cases} \text{UFD problem:} \\ \min_{\mathbf{v}_{f/b}} \left(\sum_f v_f^2 + \sum_b v_b^2\right) \\ \mathbf{N}^{CA/BL}\left(\mathbf{v}_f - \mathbf{v}_b\right) = \mathbf{0} \\ v_{f/b,i}^{LB} \leq v_{f/b,i} \leq v_{f/b,i}^{UB} \\ obj \geq \zeta obj^* \end{cases} \end{cases} \quad (14)$$

The metabolite-flow intervals, $\left[r_m^{\min}, r_m^{\max}\right]$, for cancer and normal cell and each mutant can be obtained through MFVA, and were used to determine the trend of flux change between dysregulated case and normal situation in terms of seven categories of classification [21]. However, the categories were a qualitative measure to determine the trend of flux change between mutant and normal case. In this study, we introduced the interval arithmetic [35,36] for the fuzzy equal membership grades in Eqn (13) to yield the interval membership grade as a quantitative measure to determine how much close to the template for each mutant. The interval decision grade, $[\eta_E]_i$, for GSMN of the ith mutant was calculated as follows:

$$[\boldsymbol{\eta}_E]_i = \left[\eta_{E,\min}, \eta_{E,\max}\right]_i = \frac{1}{M}\left[\sum_{m=1}^{M}\eta_{m,\min}, \sum_{m=1}^{M}\eta_{m,\max}\right]_i \quad (15)$$

The computational procedures of the minimum and maximum log2 fold changes ($\eta_{E,\min}$ and $\eta_{E,\max}$) are explained in detail in Doc. S3.

Results and discussions

Analysis of tissue-specific metabolic models

The GSMN of Recon 3D were downloaded from VHM database (https://www.vmh.life) and consisted of 5835 metabolites, 10600 reactions, and 2247 associated genes. For LUAD and LUSC, 533 and 502 cancer samples and 59 and 49 corresponding healthy samples, respectively, were obtained from TCGA database. The GSMNs for healthy and cancerous lung tissues were reconstructed using the CORDA algorithm, and statistics of the metabolites and reactions for LUAD and LUSC are presented in Fig. 3. The four models had 3360 metabolites, 5125 reactions, and 1747 genes in common, as shown in the overlapping region in Fig. 3. The cancer models comprised 3773 metabolites, 6158 reactions, and 1901 genes for LUAD and 3836 metabolites, 6290 reactions, and 1962 genes for LUSC (Fig. 3). The corresponding heathy models consisted of 4227 metabolites, 6803 reactions, and 1907 genes for LUAD and 4254 metabolites, 6761 reactions, and 1916 genes for LUSC. Twelve metabolic pathways with top-ranked number of metabolites and reactions for both tissues are shown in Fig. 4. From the classification, we observed more than 900 and 300 metabolites for fatty acyls and carboxylic acids, respectively, for both GSMNs; more than 1500 and 900 reactions for extracellular transports and fatty acid oxidation, respectively, were observed.

Identifying differentially expressed genes (DEGs) is critical in exploring molecular mechanisms of biological conditions [37]. We assessed P values and fold changes ($\log_2(CA/HT)$) using ANOVA in the SAS® software (https://www.sas.com/) to determine the differential expressions of enzyme-encoding genes for LUAD and LUSC between the normal and tumor samples from TCGA database (Doc. S4). In total, 159 and 241 enzyme-encoding genes for LUAD and LUSC, respectively, were within the absolute values of log2 fold change > 2 and $P < 0.05$, as shown in the volcano plots (Doc. S4). Such DEGs were used as a set of candidate genes to solve the oncogene inference optimization problem.

Inferred oncogenes

The Catalogue Of Somatic Mutations In Cancer (COSMIC) database (https://cancer.sanger.ac.uk/cosmic) have collected 723 cancer genes that are somatically mutated and causally implicated in human cancer, including 45 enzyme-encoding genes involved in Recon 3D. Total 25 out of the 45 enzyme-encoding genes regulate reactions in Recon 3D according to the GPR association. To demonstrate the effectiveness of the oncogene inference optimization algorithm, the similarity ratio (η_S) and membership grade (η_E) of each dysregulation of the 25 enzyme-encoding genes for LUAD and LUSC tissue-specific GSMNs reconstructed based on data from different databases (TCGA and HPA) were computed (Fig. 5). The results show most of the similarity ratios for LUAD are greater than 0.76, except *CANT1* and *SLC34A2*, and greater than 0.8 for LUSC. However, the membership grade varies from case to case (with value ranging from 0.16 to 0.76).

The NHDE algorithm in [18,21] was applied to evaluate the fitness in inferring carcinogenesis for all candidate enzyme-encoding genes. High DEGs (159 genes for LUAD and 241 genes for LUSC) for both tissues were first applied individually to enable the TLOP to determine the carcinogenicity of each gene. We obtained 9 of 159 genes for LUAD and 21 of 241 genes for LUSC that have high levels of differential

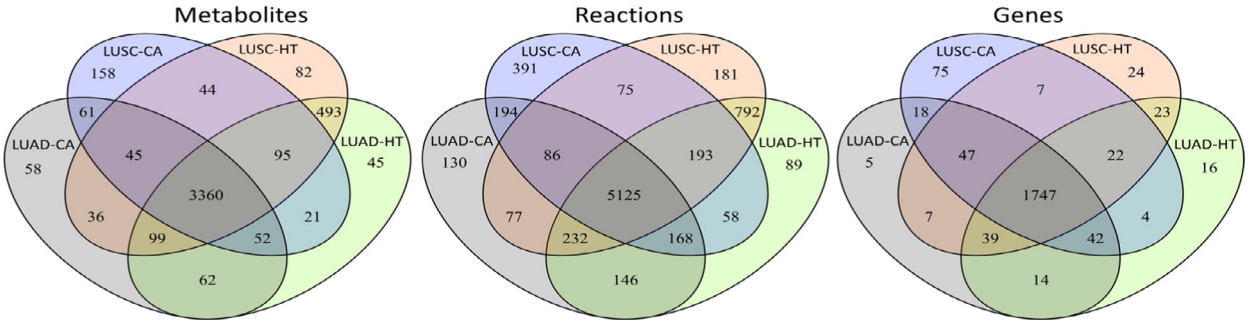

Fig. 3. Statistics of reconstructed metabolic models. Statistics of reconstructed metabolic models for LUAD and LUSC and their corresponding healthy models. LUAD-CA and LUSC-CA indicate the cancer models for LUAD and LUSC, respectively, and LUAD-HT and LUSC-HT denote their corresponding healthy models, respectively. The number in the overlapping regions of two, three, and four models indicates their common elements.

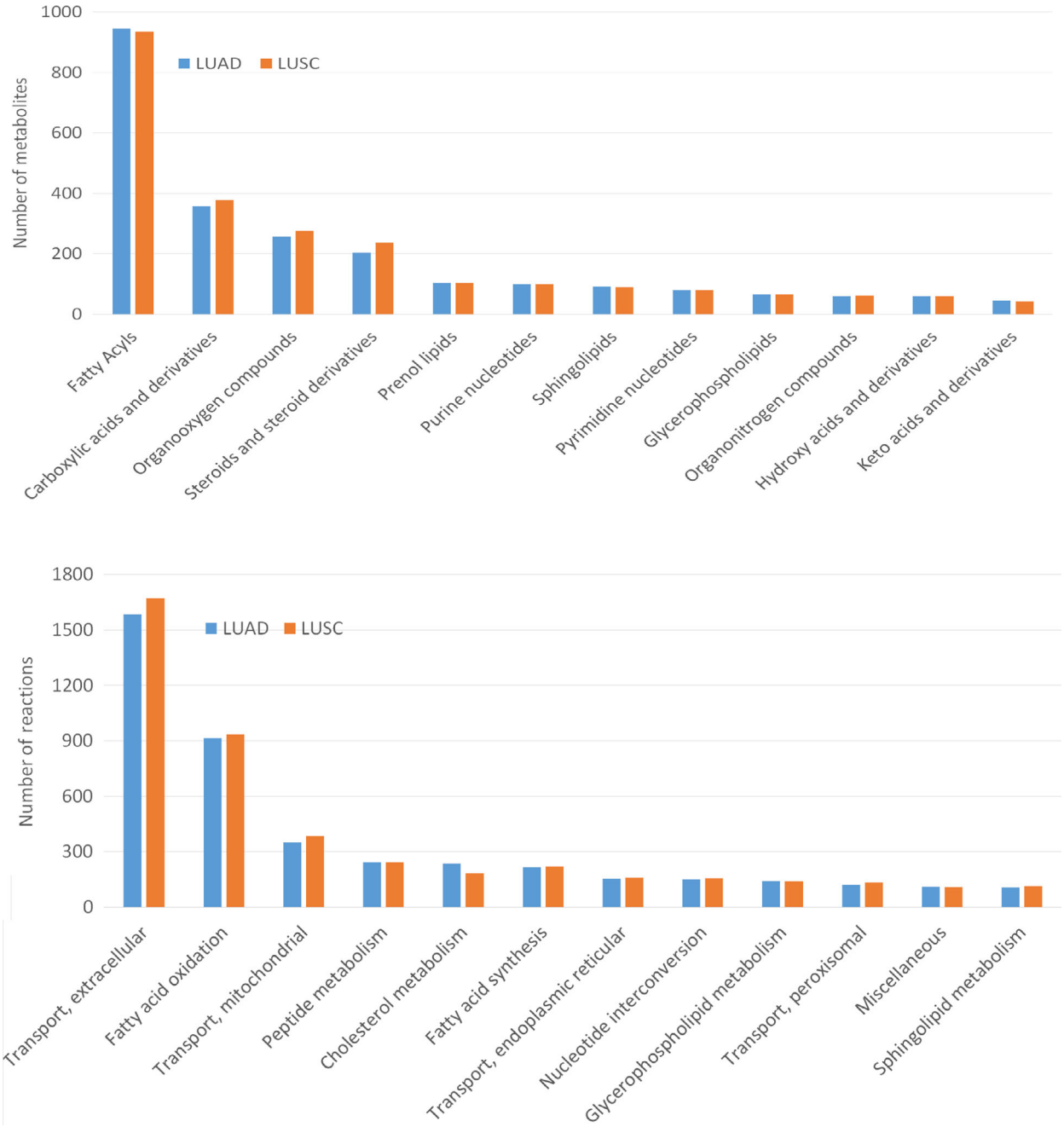

Fig. 4. Statistics of major metabolic pathways. Classification of metabolites and reactions for the reconstructed GSMNs for LUAD and LUSC. The classifications for metabolites and reactions were sought through the definitions in the HMDB database (https://hmdb.ca/) and the VMH database (https://www.vmh.life/), respectively.

expression in the computation. The optimal fitness, η_D, for these dysregulations were within [0.69, 0.76] for LUAD and [0.76, 0.81] for LUSC (Doc. S5). Additionally, the NHDE algorithm was also applied to identify oncogenes that have a low level of differential expression. We determined that 36 and 63 oncogenes for LUAD and LUSC had > 0.72 and > 0.76 optimal fitness, respectively (Doc. S5), in the computation.

Tables 1 and 2 list the 15 top-ranked oncogenes for LUAD ($\eta_D > 0.75$) and LUSC ($\eta_D > 0.85$), respectively. Pyruvate kinase (PKM) had the highest fitness ($\eta_D = 0.781$), and it plays a pivotal role in regulating

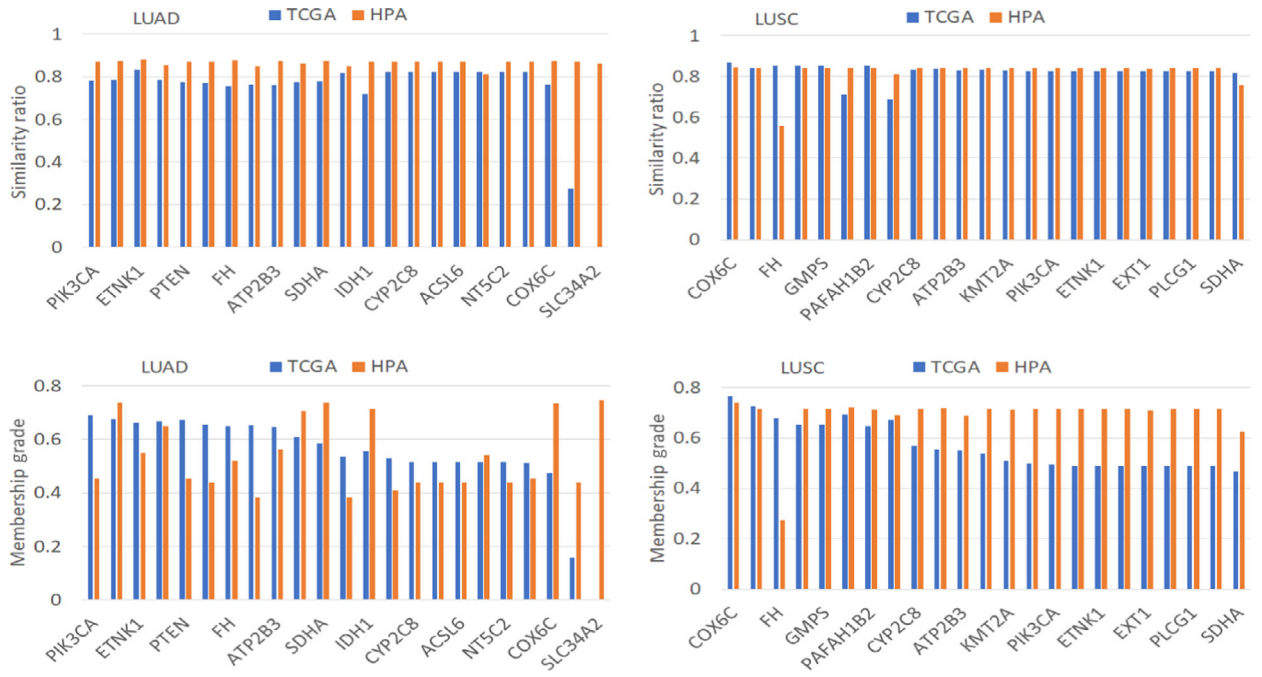

Fig. 5. Similarity ratios and membership grades for dysregulated genes. The dysregulated genes were retrieved from the COSMIC database (https://cancer.sanger.ac.uk/cosmic). TCGA and HPA indicate the similarity ratios and membership grades computed using the GSMNs reconstructed based on data from the TCGA and HPA databases, respectively. *ATP1A1* and *CHST11* genes are not available in the GPR association of the reconstructed GSMNs.

glucose-derived carbon from catabolic to biosynthetic pathways. Its dysregulation is a hallmark of tumorigenesis [38–40] and leads to several cancers, such as breast cancer, renal cell carcinoma, hepatocellular carcinoma, and colorectal cancer. There are four isozymes of pyruvate kinase in mammals (L, R, M1, and M2) encoded by two different genes: *PKLR* and *PKM*. The L and R isozymes are generated from the *PKLR* by differential splicing of RNA; the M1 and M2 forms are produced from the *PKM* gene by differential splicing. The *PKLR* gene achieved the ninth highest inferred oncogene, $\eta_D = 0.744$, in Table 1. The optimal fitness for the dysregulated enzymes was > 0.744, which indicated that metabolic alterations for both dysregulations were $> 74\%$ consistent with the template. We also reconstructed the tissue-specific GSMNs of LUAD and LUSC using gene expression data from the HPA database. Such models were applied to inspect the fitness of the inferred oncogenes from TCGA database, as shown in Tables 1 and 2. In comparison with data obtained from both approaches, we could decipher that higher fitness value indicates higher possibility of tumorigenesis.

Both PKM and PKLR isozymes are involved in the catalysis reaction (R_PYK) of phosphoenolpyruvate-to-pyruvate conversion. Furthermore, PKM not only catalyzes R_PYK but also the reaction of R_RE2954C, that is, the conversion of phosphoenolpyruvate and deoxythymidine-5'-diphosphate to pyruvate and deoxythymidine-5'-triphosphate. Moreover, PKLR catalyzes the R_r0280 reaction to convert phosphoenolpyruvate and deoxyadenosine diphosphate to pyruvate and deoxyadenosine triphosphate. The three aforementioned reactions used different metabolites to transform phosphoenolpyruvate to pyruvate. In addition, the products, namely deoxythymidine-5'-triphosphate, adenosine triphosphate, and deoxyadenosine triphosphate, of the three reactions acted as precursors of biomass reactions. We applied a reaction-based approach discussed in the oncogene inference optimization problem [18] to determine which reaction was a dominant malfunction in LUAD and LUSC. The optimal similarity ratios (η_S) and membership grades (η_E) for these dysregulated reactions were similar (Table 3). *PKM* dysregulated R_PKY and R_RE2954C, which yielded a slightly higher η_S and η_E than did *PKLR*. Both *PKM* and *PKLR* dysregulated three reactions to obtain nearly the same characteristic. The flux fold change of each reaction was greater than twofold in pyruvate synthesis, which was consistent with the Warburg effect of enhanced pyruvate formation to increase lactate production (Table 3).

Table 1. Top 15 inferred oncogenes of LUAD. η_S^{TCGA} and η_S^{HPA} are the average similarity ratios for reconstructed GSMNs based on the data from the TCGA and HPA databases, respectively. η_E^{TCGA} and η_E^{HPA} are the membership grades of the fuzzy equal function for reconstructed GSMNs based on the data from the TCGA and HPA databases, respectively. Higher η_E values indicate higher consistency of the flux alterations with the template. DEG and *P* value were calculated in SAS® software. The pathway for each gene was found from the GeneCards (https://www.genecards.org/) and VMH (https://www.vmh.life/) databases. A gene is biological significant if IDEGI > 2 and *P* value < 0.05.

Gene	DEG[a]	*P* value[b]	(η_S^{TCGA}, η_E^{TCGA})	(η_S^{HPA}, η_E^{HPA})	Pathway	Disease (score)[c]
PKM	0.98	1.01E−62	(0.842, 0.761)	(0.874, 0.734)	Abacavir pathway	Breast adenocarcinoma (0.90)
ENO1	1.36	4.78E−78	(0.839, 0.745)	(0.856, 0.747)	HIF-1-α transcription factor network	Lung cancer susceptibility 3 (0.72)
PTDSS1	0.17	1.29E−06	(0.835, 0.741)	(0.851, 0.725)	Glycerophospholipid biosynthetic pathway	Polyneuropathy (1.34)
GLTP	−0.17	6.10E−06	(0.834, 0.736)	(0.872, 0.737)	Sphingolipid metabolism	Cervical squamous cell carcinoma (0.44)
OCRL	0.52	3.91E−49	(0.831, 0.738)	NA	3-phosphoinositide degradation	Lowe oculocerebrorenal syndrome (2.11)
CA12	1.60	5.26E−12	(0.831, 0.736)	(0.870, 0.439)	Nitrogen metabolism	Hemangioma of subcutaneous tissue (1.42), lung cancer (0.37)
SLC22A7	0.79	3.10E−03	(0.837, 0.728)	(0.873, 0.737)	Zidovudine pathway	renal cell carcinoma (0.64)
PLPP1	−0.17	3.30E−02	(0.825, 0.737)	(0.545, 0.165)	Triacylglycerol biosynthesis	Myxosarcoma (1.33)
PKLR	2.17	4.34E−04	(0.830, 0.715)	(0.875, 0.740)	Abacavir pathway	Intracortical osteogenic sarcoma (1.50)
MTHFD2	1.53	1.15E−47	(0.832, 0.727)	(0.878, 0.757)	Nucleotide metabolism	Mitochondrial complex I deficiency (0.87)
SLC25A11	−0.13	6.69E−03	(0.842, 0.716)	(0.868, 0.734)	Glucose metabolism	Paragangliomas 6 (2.83)
ALDH4A1	0.23	8.64E−04	(0.835, 0.723)	(0.858, 0.731)	Alanine, aspartate and glutamate metabolism	Hyperprolinemia, type Ii (2.48)
GAPDH	1.93	4.52E−61	(0.826, 0.731)	(0.855, 0.733)	Cori cycle	Angioimmunoblastic T-cell lymphoma (1.14)
SLC13A5	3.24	2.64E−07	(0.835, 0.721)	(0.860, 0.731)	Transport of glucose and other sugars	Nasal cavity benign neoplasm (1.50)
SLC20A1	0.92	2.65E-22	(0.834, 0.721)	(0.867, 0.736)	Glucose/energy metabolism	Leukemia (0.86)

[a]DEG = log2(CA/HT) denotes a differential expression gene and is computed from cancer and healthy samples of TCGA datasets.; [b]*P* value is computed from cancerous and healthy samples of TCGA datasets.; [c]Diseases and scores are obtained from the GeneCards database.

Enolase 1 (ENO1), along with *PKM* and *PKLR*, is a glycolytic enzyme. Glycolysis is an ATP-generating step that is pivotal in cancer cell proliferation and metastasis. *ENO1* is overexpressed in several tumor types, including NSCLC [41–43]. ENO1 is an upstream enzyme of PKM and PKLR that catalyzes 2-phosphoglycerate to form phosphoenolpyruvate. We observed that its DEG for LUAD obtained from RNA-Seq datasets in TCGA was not statistically significant (DEG = 1.36). Furthermore, the computation results revealed that the flux fold change increased by 1.33 times from normal to cancer states. The average similarity ratio ($\eta_S = 0.839$) and membership grade ($\eta_E = 0.745$) of ENO1 were slightly smaller than those of PKM.

Phosphatidylserine synthase, encoded by *PTDSS1*, is involved in phosphatidylserine biosynthetic pathway, which is a part of phospholipid metabolism. Phosphatidylserine is a precursor in the biomass reaction of the reconstructed metabolic model. Its dysregulation causes Lenz–Majewski syndrome, which is a rare disease characterized by complex craniofacial, dental, cutaneous, and limb abnormalities combined with intellectual disability [44]. Phosphatidylserine signaling is highly dysregulated in the tumor microenvironment and autoimmune diseases [45]. According to the computation results, 1.1% upregulation of *PTDSS1* in LUAD could increase the production of phosphatidylserine that exhibited carcinogenic effects that yielded the average similarity ratio ($\eta_S = 0.835$) and membership grade ($\eta_E = 0.741$). By contrast, *PTDSS1* was one of the 34 top-ranked oncogenes for LUSC that yielded η_S and η_E values of 0.899 and 0.789 (Doc. S5), respectively. A survival analysis for oncogenes can be applied to investigate the clinical significance of metabolic alterations. In this study, we surveyed a survival analysis from the HPA database to explain the survival significance of the inferred oncogenes. The high expression of *PTDSS1* in LUAD could significantly reduce the survival probability compared with that of PKM, PKLR, and ENO1 (Doc. S6).

Moreover, we also reconstructed LUAD and LUSC GSMNs for comparison using the iMAT algorithm [46] based on data from different databases (TCGA

Table 2. Top 15 inferred oncogenes of LUSC. η_S^{TCGA} and η_S^{HPA} are the average similarity ratios for reconstructed GSMNs based on the data from the TCGA and HPA databases, respectively. η_E^{TCGA} and η_E^{HPA} are the membership grades of the fuzzy equal function for reconstructed GSMNs based on the data from the TCGA and HPA databases, respectively. Higher η_E values indicate higher consistency of the flux alterations with the template. DEG and *P* value were calculated in SAS® software. The pathway for each gene was found from the GeneCards (https://www.genecards.org/) and VMH (https://www.vmh.life/) databases. A gene is biological significant if IDEGI > 2 and *P* value < 0.05.

Gene	DEG[a]	*P* value[b]	(η_S^{TCGA}, η_E^{TCGA})	(η_S^{HPA}, η_E^{HPA})	Pathway	Disease (score)[c]
SLCO2B1	−2.22	8.80E−20	(0.917, 0.829)	(0.871, 0.842)	Atenolol pathway	Ileum cancer (1.31)
SLC9A1	−0.22	3.08E−03	(0.911, 0.822)	(0.847, 0.826)	Osteoclast signaling	Gastroesophageal reflux (1.00)
SLC7A10	2.68	1.14E−02	(0.911, 0.819)	(0.862, 0.835)	Differentiation of white and brown adipocyte	Follicular lymphoma (0.91)
SLC20A1	0.42	3.45E−04	(0.904, 0.822)	(0.870, 0.843)	Glucose/energy metabolism	Leukemia (0.86)
AQP8	−1.27	1.85E−05	(0.906, 0.817)	(0.878, 0.854)	Detoxification of reactive oxygen species	Colorectal adenoma (0.89)
AGL	0.42	1.49E−12	(0.908, 0.814)	NA	Glycogen metabolism	Bladder lateral wall cancer (1.26)
SLC12A4	−1.11	8.74E−40	(0.906, 0.815)	(0.871, 0.841)	Transport of glucose and other sugars	Fish-eye disease (1.50)
KYAT1	0.74	2.25E−36	(0.909, 0.812)	(0.865, 0.838)	Selenocompound metabolism	Schizophrenia (0.74)
SLC6A2	4.35	1.46E−16	(0.897, 0.818)	NA	Methylphenidate pathway	Adrenal medulla cancer (1.21)
SLC5A12	4.93	1.70E−63	(0.898, 0.816)	(0.867, 0.631)	NRF2 pathway	Follicular lymphoma (0.91)
SLC4A2	−0.32	3.49E−07	(0.903, 0.810)	(0.867, 0.631)	Bile secretion	Hepatocellular carcinoma (0.68)
SLCO1C1	−0.91	1.11E−04	(0.898, 0.813)	(0.867, 0.631)	Transport of vitamins and nucleosides	Allan–Herndon–Dudley syndrome (1.20)
SLC23A2	−0.26	1.75E−04	(0.897, 0.814)	(0.866, 0.839)	Metabolism of water-soluble vitamins and cofactors	Hepatitis C virus (0.87)
ACE2	0.77	2.46E−06	(0.896, 0.815)	(0.867, 0.631)	A-beta plaque formation and APP metabolism	Renal oncocytoma (0.72)
SLC43A1	−0.97	9.28E−14	(0.897, 0.810)	(0.866, 0.840)	Amino acid transport across the plasma membrane	Seminoma (1.07)

[a]DEG = log2(CA/HT) denotes a differential expression gene and is computed from cancer and healthy samples of TCGA datasets.; [b]*P* value is computed from cancerous and healthy samples of TCGA datasets.; [c]Diseases and scores are obtained from the GeneCards database.

Table 3. Comparison of carcinogenicity caused by reaction-based and enzyme-based dysregulations in LUAD and LUSC, respectively. Flux fold change (LFC) is denoted as log2 fold change between cancer and healthy states; LFC was computed as $\log2(r_{cancer}/r_{normal})$. η_S and η_E indicate the average similarity ratio and fuzzy equal membership grade, respectively. H, proton; PEP, phosphoenolpyruvate; DTDP, deoxythymidine-5′-diphosphate; ADP, adenosine diphosphate; DADP, deoxyadenosine diphosphate; PYR, pyruvate; ATP, adenosine triphosphate; DTTP, deoxythymidine-5′-triphosphate; DATP, deoxyadenosine triphosphate.

		LUAD		LUSC	
Dysregulated reactions/genes	Reactions	LFC	(η_S, η_E)	LFC	(η_S, η_E)
R_PYK	H + ADP + PEP → ATP + PYR	3.753	(0.829, 0.723)	3.128	(0.876, 0.754)
R_RE2954C	H + PEP + DTDP ↔ PYR + DTTP	3.671	(0.843, 0.745)	3.687	(0.867, 0.738)
R_r0280	H + PEP + DADP → PYR + DATP	3.901	(0.830, 0.714)	3.361	(0.882, 0.760)
R_RE2954C + R_r0280	H + PEP + D PYR + DTTP	2.620	(0.841, 0.743)	2.60	(0.868, 0.740)
	H + PEP + DADP → PYR + DATP	2.324		2.239	
PKM	H + ADP + PEP → ATP + PYR	2.260	(0.842, 0.761)	1.857	(0.860, 0.728)
	H + PEP + DTDP ↔ PYR + DTTP	2.868		2.424	
PKLR	H + ADP + PEP → ATP + PYR	2.563	(0.830, 0.715)	2.395	(0.869, 0.752)
	H + PEP + DADP → PYR + DATP	2.878		2.680	
PKM + *PKLR*	H + ADP + PEP → ATP + PYR	1.749	(0.848, 0.754)	1.735	(0.863, 0.717)
	H + PEP + DTDP ↔ PYR + DTTP	2.300		2.286	
	H + PEP + DADP → PYR + DATP	2.019		1.984	

and HPA). The similarity ratios and membership grades for each dysregulated gene are shown in Doc. S7. The results show that the similarity ratios of LUAD and LUSC could fulfill the prediction. Because the GSMNs reconstructed by the iMAT algorithm are more parsimonious than by the CORDA algorithm, some genes are not available in the GPR association of the GSMNs.

Dysregulation of membrane transporters

The computation results (Doc. S5) reveal that 12 out of 45 genes for LUAD and 45 out of 84 genes for LUSC served as solute carriers (SLCs). SLC genes could easily cause carcinogenesis, and we found that 11 SLC genes for LUAD (except *SLC6A4*) are common to those for LUSC. These SLC gene-encoding proteins are categorized into SLC families and SLC anion transporter families. Such transmembrane transporters could mediate the influx and efflux of substances such as ions, nucleotides, and sugars across biological membranes. The dysregulation of these genes can drive metabolic diseases, such as type II diabetes. Reports have indicated that the mediation of *SLC5A2* and *SLC13A5* genes may be therapeutic targets for treating type II diabetes and nonalcoholic fatty liver diseases [47,48]. In the computation, *SLC13A5* was determined to be an oncogene in LUAD and LUSC, and it had high gene differential expression (DEG = 3.239 for LUAD and 5.126 for LUSC) between cancer tissues and healthy tissues. However, *SLC5A2* was identified in LUSC only, but it possessed low gene differential expression (DEG = −1.1228). *SLC5A2* encodes a member of the sodium glucose cotransporter family, which is a sodium-dependent glucose transport protein. The dysregulation of *SLC5A2* could lead to non-small-cell lung cancer and pancreatic [49] and prostate adenocarcinomas [50].

SLC13A5 is a sodium-coupled citrate transporter that plays a key role in importing citrate from the bloodstream into human cells. Surveys conducted using PubMed and GeneCards indicated that SLC13A5 is related to nasal cavity neoplasm and hepatocellular carcinoma [51]. The computation revealed that the fold change of its mediated flux increased more than sixfold. Other common genes, namely *SLC25A11*, *SLC20A1*, and *SLC22A7*, achieved high fitness (Table 1). The oxoglutarate carrier SLC25A11 is important for ATP production in cancer during NADH transportation from the cytosol to mitochondria as a malate. The dysregulation of SLC25A11 could lead to non-small-cell lung cancer [52] and liver

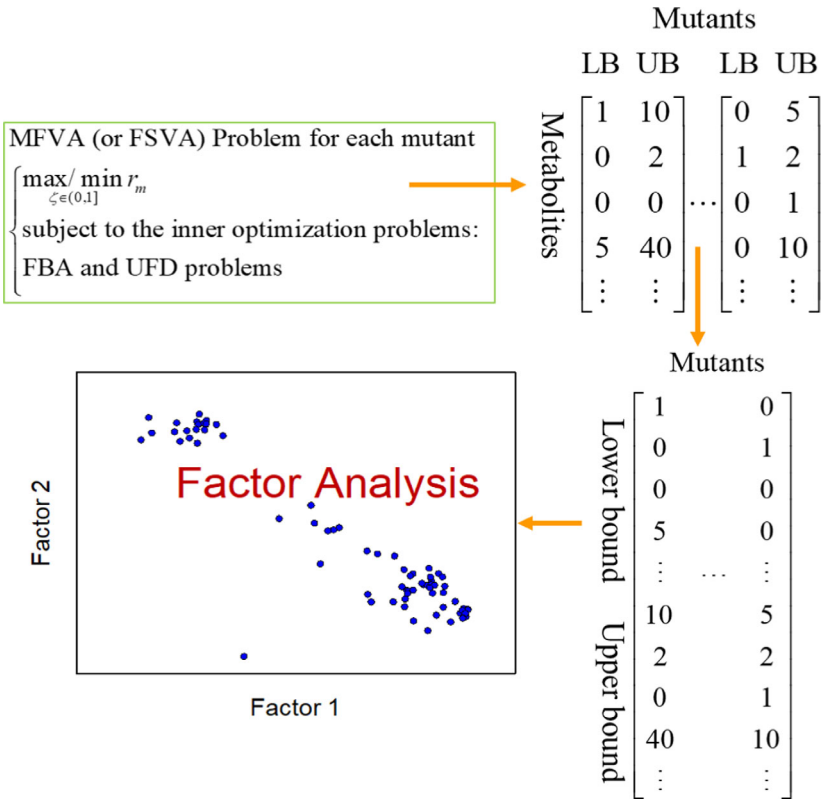

Fig. 6. Concept of MFVA. Concept for establishing flux-sum bounds of each mutant through MFVA for factor analysis.

cancer [53]. Sodium-dependent phosphate transporter 1 (SLC20A1) plays a fundamental housekeeping role in phosphate transport, such as absorbing phosphate from interstitial fluid for normal cellular functions such as cellular metabolism, signal transduction, and nucleic acid and lipid synthesis. Some review articles [54] have indicated that the phosphate transporter is overexpressed in tumor cells, which was consistent with the fold change of DEG (0.92 for LUAD and 0.42 for LUSC) evaluated from TCGA database; therefore, it has been considered a key promoter of tumorigenesis.

Angiotensin-converting enzyme 2, encoded by *ACE2*, has recently garnered widespread interest as the SARS-CoV-2 receptor. It appears to be an infective agent responsible for coronavirus disease 2019 and associated cardiovascular diseases [33,55]. We determined that *ACE2* was an oncogene of LUSC and achieved an average similarity ratio of 0.896 and membership grade of 0.815, making it one of the 15 top-ranked oncogenes. However, *ACE2* was not implicated to cause tumorigenesis in LUAD because although its average similarity ratio could reach 0.82, its membership grade was 0.385, which was smaller than those of the 15 top-ranked oncogenes. Furthermore, we inspected RNA-Seq data from TCGA and found that fold changes in DEG (1.67 for LUAD and 0.77 for LUSC) for both cancers increased nonsignificantly. According to the HPA database, the high expression of *ACE2* in LUSC indicated high survival probability during the initial stages, but the results after 11 years were identical (Doc. S6). By contrast, high or low *ACE2* expression in LUAD was not differentiated.

Results of MFVA

In this study, we applied the TLOP to infer oncogenes in GSMNs of lung adenocarcinoma and lung squamous cell carcinoma. FBA was involved in the inner optimization problem of TLOP. FBA can calculate steady-state metabolic fluxes for GSMNs in a reasonable computational time with modern personal computers, but it is a biased method in constraint-based modeling approaches for yielding optimal flux distributions. Monte Carlo sampling methods for GSMNs can cope with such a biased prediction, but still spend a lot of computer time [56]. In this study, we introduced an interval arithmetic [35,36] for MFVA, an extension of FBA, to determine the robustness of metabolic models in various simulation conditions. However, its use has been somewhat limited by the long computation time compared with FBA. MFVA is generally incapable of embedding in the oncogene inference problem [Eqn (1)] due to the computational burden, but it could be applied to investigate whether the optimal results were achieved. MFVA was applied to calculate the lower and upper flux-sum bounds (i.e., 4366 observed components) for each dysregulated gene. The metabolic flux-sum bound matrix was formed by representing each dysregulated gene as a column of lower and upper flux-sum bounds (Fig. 6) as calculated through MFVA. Moreover, the flux-sum bound column of the template was added to the matrix. Such metabolic flux-sum bound matrices for all mutants of LUAD and LUSC were then used to perform a factor analysis to yield two factors (Fig. 7A and B).

In total, 23 genes of LUAD in Factor 1 had factor loadings > 0.73 (Fig. 7A; numerical data are listed in Doc. S8). Acetyl-CoA acyltransferase 2, encoded by *ACAA2*, and lactoylglutathione lyase, encoded by

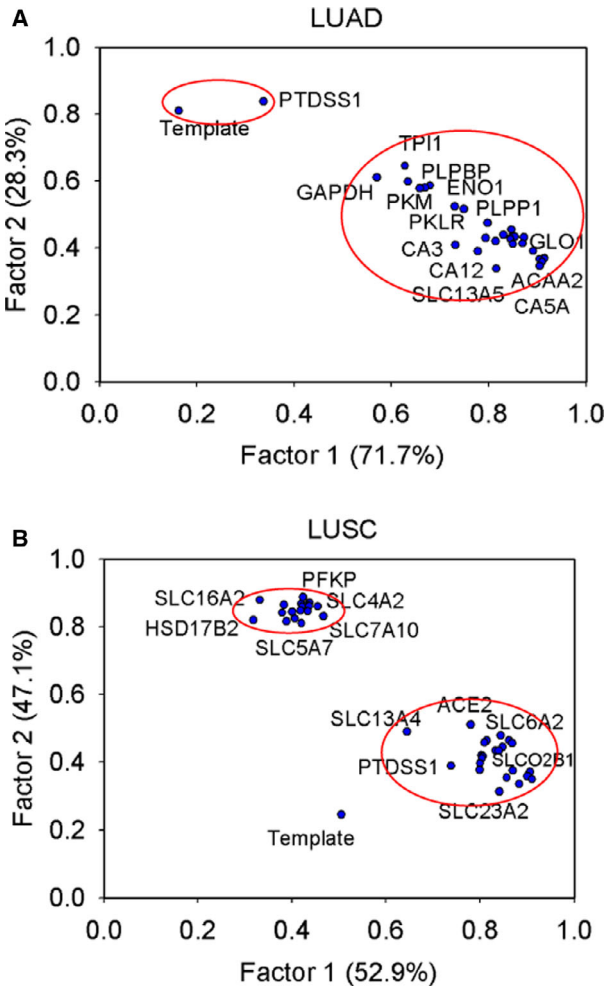

Fig. 7. Factor analysis. Factor analysis for analyzing the flux-;bound matrices of (A) LUAD and (B) LUSC to yield two factors.

GLO1, had the two highest factor loadings (0.914) in Factor 1. ACAA2 catalyzes the final step of the mitochondrial fatty acid beta-oxidation pathway. To date, clinical diseases caused by mutations or variants of ACAA2 have not been identified. However, the ACAA2 locus has been associated with abnormal blood lipid levels, particularly HDL and LDL cholesterol levels [55]. GLO1 participates in pyruvate metabolism to convert S-lactoylglutathione into methylglyoxal and glutathione, and regulates TNF-induced transcriptional activity of NF-κB. For Factor 2, phosphatidylserine synthase 1, encoded by *PTDSS1*, had a factor loading of 0.84, which was close to the template (0.81). PTDSS1 mainly catalyzes the conversion of phosphatidylcholine in the phosphatidylserine biosynthesis pathway, which is part of phospholipid metabolism.

The factor analysis of 42 separate dysregulated genes for LUSC in two groups is shown in Fig 7B (Doc. S8). The 23 genes of Factor 1 in the first group had factor loadings > 0.64 (Fig. 7B). The template was still close to PTDSS1 (0.74) and yielded a factor loading of 0.51. The 18 genes of Factor 2 in the second group had factor loadings > 0.81 (Fig. 7B).

The lower and upper flux-sum bounds obtained through MFVA were applied to compute the minimum and maximum membership grades (Fig. 8). In other words, the interval range of each mutant was compared with the lower and upper bounds of the template through interval computation (Doc. S9). A numerical example in Doc. S3 describes the computation of interval numbers to yield the minimum and maximum membership grades (Fig. 8). The optimal

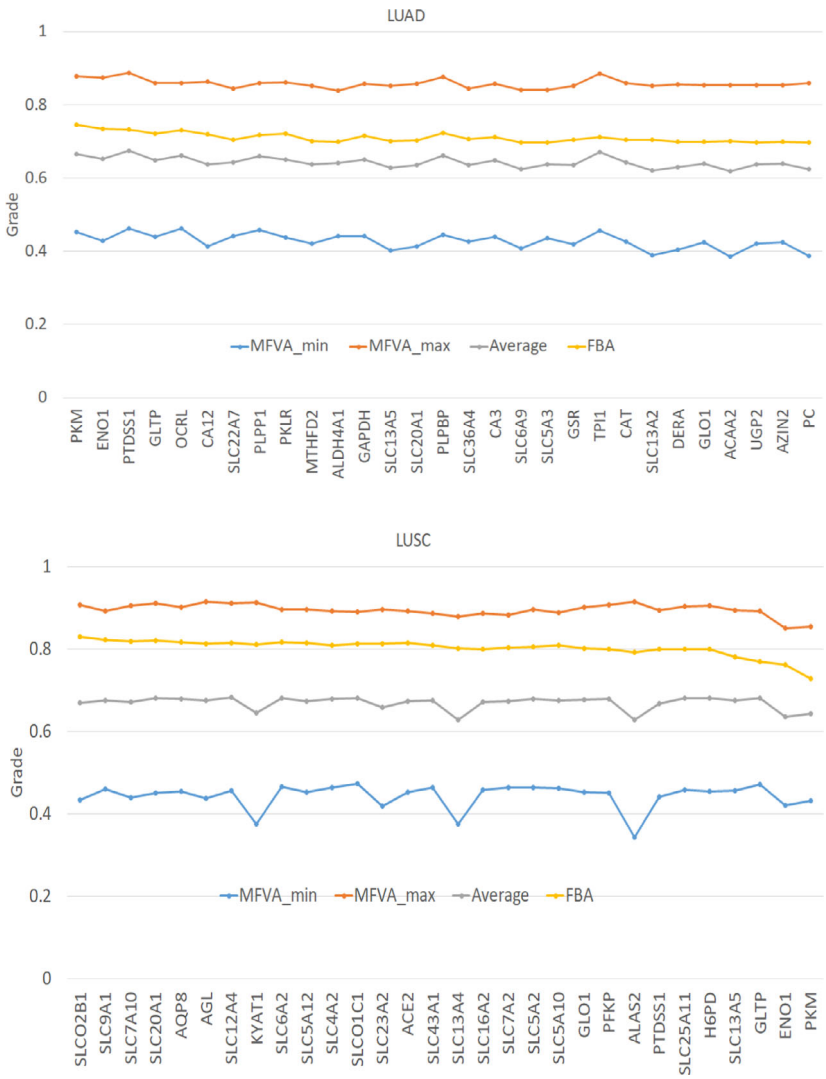

Fig. 8. Minimum and maximum membership grades obtained through MFVA. MFVA_min and MFVA_max are the minimum and maximum fuzzy equal membership grades for each gene computed through MFVA, respectively. FBA is the optimal membership grade obtained by solving the inner optimization problem. Average is the average of MFVA_min and MFVA_max.

membership grade obtained through FBA in the oncogene inference problem was within the range of each mutant. PTDSS1 achieved the highest range among the mutants in LUAD, and the maximum membership grade of the mutants for LUSC was smoothly variant. According to the factor analysis, LUAD characteristics were similar to those of PTDSS1 but unobvious to those of LUSC.

Conclusion

This study integrated the RNA-Seq data of healthy and cancerous lung tissues downloaded from TCGA database with the genome-scale metabolite and protein structure data of Recon 3D to reconstruct tissue-specific GSMNs. The models were first used to generate flux patterns as a template/control in the trilevel oncogene inference optimization framework for inferring tumorigenic genes. The similarity ratios and fuzzy equal membership grades are the objectives in the oncogene inference optimization platform. The similarity ratio was used as a quality criterion to evaluate the similarity between the dysregulated flux pattern and that of the template. The fuzzy equal membership grade was used as a quantity metric to measure how close to the template the fold change of a dysregulated flux pattern is. The platform involved with the template could detect 45 and 84 tumorigenic genes for LUAD and LUSC, respectively. We observed that a high level of DEGs was not an essential factor for determining tumorigenesis. Nine out of 45 and 21 out of 84 genes for LUAD and LUSC, respectively, with high levels of differential expression based on cancer and healthy samples in TCGA database; the other genes, such as *PKM* and *PTDSS1*, have low levels of differential expression. The MFVA used the interval arithmetic to yield the interval membership grade as a quantitative measure of biased predictions from FBA.

Acknowledgements

This study was supported by the Ministry of Science and Technology of Taiwan (Grant No. MOST106-2221-E-194-049-MY3 and MOST109-2320-B-194-003).

Conflict of interest

The authors declare no conflict of interest.

Data accessibility

The data that support the findings of this study are available in the Supporting information of this article.

Author contributions

FSW was responsible for study conception and design and drafted the original manuscript. WHW wrote the program and revised the manuscript. YTW, MRL, and WCC performed the data analysis and database survey. All authors have read and approved the final manuscript.

References

1 Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA and Jemal A (2018) Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* **68**, 394–424.

2 Zappa C and Mousa SA (2016) Non-small cell lung cancer: current treatment and future advances. *Trans Lung Cancer Res* **5**, 288–300.

3 Herbst RS, Morgensztern D and Boshoff C (2018) The biology and management of non-small cell lung cancer. *Nature* **553**, 446–454.

4 Yeh S-J, Chang C-A, Li C-W, Wang LH-C and Chen B-S (2019) Comparing progression molecular mechanisms between lung adenocarcinoma and lung squamous cell carcinoma based on genetic and epigenetic networks: big data mining and genome-wide systems identification. *Oncotarget* **10**, 3760–3806.

5 Asgari Y, Zabihinpour Z, Salehzadeh-Yazdi A, Schreiber F and Masoudi-Nejad A (2015) Alterations in cancer cell metabolism: The Warburg effect and metabolic adaptation. *Genomics* **105**, 275–281.

6 Jalili M, Gebhardt T, Wolkenhauer O and Salehzadeh-Yazdi A (2018) Unveiling network-based functional features through integration of gene expression into protein networks. *Biochim Biophys Acta Mol Basis Dis* **1864**, 2349–2359.

7 Auslander N, Cunningham CE, Toosi BM, McEwen EJ, Yizhak K, Vizeacoumar FS, Parameswaran S, Gonen N, Freywald T, Bhanumathy KK *et al.* (2017) An integrated computational and experimental study uncovers FUT9 as a metabolic driver of colorectal cancer. *Mol Syst Biol* **13**, 956.

8 Gatto F, Ferreira R and Nielsen J (2020) Pan-cancer analysis of the metabolic reaction network. *Metab Eng* **57**, 51–62.

9 Hu J, Locasale JW, Bielas JH, O'Sullivan J, Sheahan K, Cantley LC, Heiden MGV and Vitkup D (2013) Heterogeneity of tumor-induced gene expression changes in the human metabolic network. *Nat Biotechnol* **31**, 522–529.

10 Yizhak K, Gabay O, Cohen H and Ruppin E (2013) Model-based identification of drug targets that revert disrupted metabolism and its application to ageing. *Nat Commun* **4**, 2632.

11 Zielinski DC, Jamshidi N, Corbett AJ, Bordbar A, Thomas A and Palsson BO (2017) Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism. *Sci Rep* **7**, 41241.
12 Pavlova NN and Thompson CB (2016) The emerging hallmarks of cancer metabolism. *Cell Metab* **23**, 27–47.
13 Eyassu F and Angione C (2017) Modelling pyruvate dehydrogenase under hypoxia and its role in cancer metabolism. *R Soc Open Sci* **4**, 170360.
14 Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E and Shlomi T (2011) Predicting selective drug targets in cancer through metabolic networks. *Mol Syst Biol* **7**, 501.
15 Gottstein W, Olivier BG, Bruggeman FJ and Teusink B (2016) Constraint-based stoichiometric modelling from single organisms to microbial communities. *J R Soc Interface* **13**, 20160627.
16 Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M and Nielsen J (2014) Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat Commun* **5**, 3083.
17 Wang Y, Eddy JA and Price ND (2012) Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst Biol* **6**, 153.
18 Wang F-S, Wu W-H, Hsiu W-S, Liu Y-J and Chuang K-W (2019) Genome-scale metabolic modeling with protein expressions of normal and cancerous colorectal tissues for oncogene inference. *Metabolites* **10**, 16.
19 Wu M and Chan C (2012) Human metabolic network: reconstruction, simulation, and applications in systems biology. *Metabolites* **2**, 242–253.
20 Wu H-Q, Cheng M-L, Lai J-M, Wu H-H, Chen M-C, Liu W-H, Wu W-H, Chang PM-H, Huang C-YF, Tsou A-P *et al.* (2017) Flux balance analysis predicts Warburg-like effects of mouse hepatocyte deficient in miR-122a. *PLoS Comput Biol* **13**, e1005618.
21 Wu W-H, Li F-Y, Shu Y-C, Lai J-M, Chang PM-H, Huang C-YF and Wang F-S (2020) Oncogene inference optimization using constraint-based modelling incorporated with protein expression in normal and tumour tissues. *R Soc Open Sci* **7**, 191241.
22 Yizhak K, Chaneton B, Gottlieb E and Ruppin E (2015) Modeling cancer metabolism on a genome scale. *Mol Syst Biol* **11**, 817.
23 Yizhak K, Le Dévédec SE, Rogkoti VM, Baenke F, de Boer VC, Frezza C, Schulze A, van de Water B and Ruppin E (2014) A computational study of the Warburg effect identifies metabolic targets inhibiting cancer migration. *Mol Syst Biol* **10**, 744.
24 Aurich MK, Paglia G, Rolfsson Ó, Hrafnsdóttir S, Magnúsdóttir M, Stefaniak MM, Palsson BØ, Fleming RMT and Thiele I (2015) Prediction of intracellular metabolic states from extracellular metabolomic data. *Metabolomics* **11**, 603–619.
25 Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, Mih N, Gatto F, Nilsson A, Preciat Gonzalez GA, Aurich MK *et al.* (2018) Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat Biotechnol* **36**, 272–281.
26 Swainston N, Smallbone K, Hefzi H, Dobson P, Brewer J, Hanscho M, Zielinski D, Ang KS, Gardiner N, Gutierrez J *et al.* (2016) Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* **12**, 109.
27 Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, Forsberg M, Zwahlen M, Kampf C, Wester K, Hober S *et al.* (2010) Towards a knowledge-based human protein atlas. *Nat Biotechnol* **28**, 1248–1250.
28 Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, Wallqvist A and Papin JA (2017) Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat Commun* **8**, 14250.
29 Ryu JY, Kim HU and Lee SY (2017) Framework and resource for more than 11,000 gene-transcript-protein-reaction associations in human metabolism. *Proc Natl Acad Sci USA* **114**, E9740–E9749.
30 Schultz A and Qutub AA (2016) Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Comput Biol* **12**, e1004808.
31 National Cancer Institute of U.S. (2019) Department of Health and Human Services. The Cancer Genome Atlas Program, Washington, D.C.
32 Valcárcel LV, Torrano V, Tobalina L, Carracedo A and Planes FJ (2019) rMTA: robust metabolic transformation analysis. *Bioinformatics* **35**, 4350–4355.
33 Pozo C, Miró A, Guillén-Gosálbez G, Sorribas A, Alves R and Jiménez L (2015) Gobal optimization of hybrid kinetic/FBA models via outer-approximation. *Comput Chem Eng* **72**, 325–333.
34 Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, Haraldsdóttir HS, Wachowiak J, Keating SM, Vlasov V *et al.* (2019) Creation and analysis of biochemical constraint-based models using the COBRA Toolbox vol 3.0. *Nat Protoc* **14**, 639–702.
35 Kaufmann A and Gupta MM (1991) *Introduction to Fuzzy Arithmetic: Theory and Applications*. International Thomson Computer Press, MI.
36 Zettervall H (2014) *Fuzzy Set Theory Applied to Make Medical Prognoses for Cancer Patients*. Blekinge Institute of Technology, Sweden.
37 Crow M, Lim N, Ballouz S, Pavlidis P and Gillis J (2019) Predictability of human differential gene expression. *Proc Natl Acad Sci* **116**, 6491–6500.
38 Luo W and Semenza GL (2012) Emerging roles of PKM2 in cell metabolism and cancer progression. *Trends Endocrinol Metab* **23**, 560–566.

39 Morita M, Sato T, Nomura M, Sakamoto Y, Inoue Y, Tanaka R, Ito S, Kurosawa K, Yamaguchi K, Sugiura Y *et al.* (2018) PKM1 confers metabolic advantages and promotes cell-autonomous tumor cell growth. *Cancer Cell* **33**, 355–367.
40 Zahra K, Dey T, Ashish M, Mishra SP and Pandey U (2020) Pyruvate kinase M2 and cancer: the role of PKM2 in promoting tumorigenesis. *Front Oncol* **10**, 159.
41 Altenberg B and Greulich KO (2004) Genes of glycolysis are ubiquitously overexpressed in 24 cancer classes. *Genomics* **84**, 1014–1020.
42 Fu Q-F, Liu Y, Fan Y, Hua S-N, Qu H-Y, Dong S-W, Li R-L, Zhao M-Y, Zhen Y, Yu X-L *et al.* (2015) Alpha-enolase promotes cell glycolysis, growth, migration, and invasion in non-small cell lung cancer through FAK-mediated PI3K/AKT pathway. *J Hematol Oncol* **8**, 22.
43 Ji M, Wang Z, Chen J, Gu L, Chen M, Ding Y and Liu T (2019) Up-regulated ENO1 promotes the bladder cancer cell growth and proliferation via regulating β-catenin. *Biosci Rep* **39**.
44 Sohn M, Ivanova P, Brown HA, Toth DJ, Varnai P, Kim YJ and Balla T (2016) Lenz-Majewski mutations in PTDSS1 affect phosphatidylinositol 4-phosphate metabolism at ER-PM and ER-Golgi junctions. *Proc Natl Acad Sci USA* **113**, 4314–4319.
45 Birge RB, Boeltz S, Kumar S, Carlson J, Wanderley J, Calianese D, Barcinski M, Brekken RA, Huang X, Hutchins JT *et al.* (2016) Phosphatidylserine is a global immunosuppressive signal in efferocytosis, infectious disease, and cancer. *Cell Death Differ* **23**, 962–978.
46 Zur H, Ruppin E and Shlomi T (2010) iMAT: an integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–3142.
47 Schumann T, König J, Henke C, Willmes DM, Bornstein SR, Jordan J, Fromm MF and Birkenfeld AL (2020) Solute carrier transporters as potential targets for the treatment of metabolic disease. *Pharmacol Rev* **72**, 343–379.
48 Superti-Furga G, Lackner D, Wiedmer T, Ingles-Prieto A, Barbosa B, Girardi E, Goldmann U, Gürtl B, Klavins K, Klimek C *et al.* (2020) The RESOLUTE consortium: unlocking SLC transporters for drug discovery. *Nat Rev Drug Discov* **19**, 429–430.
49 Taira N, Atsumi E, Nakachi S, Takamatsu R, Yohena T, Kawasaki H, Kawabata T and Yoshimi N (2018) Comparison of GLUT-1, SGLT-1, and SGLT-2 expression in false-negative and true-positive lymph nodes during the ^{1}F-FDG PET/CT mediastinal nodal staging of non-small cell lung cancer. *Lung Cancer* **123**, 30–35.
50 Scafoglio C, Hirayama BA, Kepe V, Liu J, Ghezzi C, Satyamurthy N, Moatamed NA, Huang J, Koepsell H, Barrio JR *et al.* (2015) Functional expression of sodium-glucose transporters in cancer. *Proc Natl Acad Sci USA* **112**, E4111–E4119.
51 Li Z, Li D, Choi EY, Lapidus R, Zhang L, Huang S-M, Shapiro P and Wang H (2017) Silencing of solute carrier family 13 member 5 disrupts energy homeostasis and inhibits proliferation of human hepatocarcinoma cells. *J Biol Chem* **292**, 13890–13901.
52 Lee J-S, Lee H, Lee S, Kang JH, Lee S-H, Kim S-G, Cho ES, Kim NH, Yook JI and Kim S-Y (2019) Loss of SLC25A11 causes suppression of NSCLC and melanoma tumor formation. *EBioMedicine* **40**, 184–197.
53 Pan G, Wang R, Jia S, Li Y, Jiao Y and Liu N (2020) SLC25A11 serves as a novel prognostic biomarker in liver cancer. *Sci Rep* **10**, 9871.
54 Lacerda-Abreu MA, Russo-Abrahão T, de Queiroz Monteiro R, Rumjanek FD and Meyer-Fernandes JR (2018) Inorganic phosphate transporters in cancer: Functions, molecular mechanisms and possible clinical applications. *Biochim Biophys Acta, Rev Cancer* **1870**, 291–298.
55 Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, Cooper GM, Roos C, Voight BF, Havulinna AS *et al.* (2008) Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat Genet* **40**, 189–197.
56 Schellenberger J, Que R, Fleming RMT, Thiele I, Orth JD, Feist AM, Zielinski DC, Bordbar A, Lewis NE, Rahmanian S *et al.* (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* **6**, 1290–1307.

Supporting information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Doc. S1. A numerical example to illustrate the computation of the template, similarity ratio, and LFC.

Doc. S2. Detail description of the NHDE algorithm.

Doc. S3. A numerical example to illustrate the interval computation of membership grade of fuzzy equal function.

Doc. S4. Differential expressions of enzyme-encoding genes and volcano plots for LUAD and LUSC.

Doc. S5. Inferred oncogenes for LUAD and LUSC.

Doc. S6. Survival analysis obtained from the HPA database to explain survival significance of the inferred oncogenes.

Doc. S7. Similarity ratios and membership grades of the dysregulated genes from Tables 1 and 2 for the LUAD and LUSC GSMNs reconstructed by integrating the iMAT algorithm with data from different databases (TCGA or HPA).

Doc. S8. Results of factor analysis.

Doc. S9. The lower and upper flux-sum bounds obtained through MFVA.