

Article

Genome-Scale Metabolic Modeling with Protein Expressions of Normal and Cancerous Colorectal Tissues for Oncogene Inference

Feng-Sheng Wang *, Wu-Hsiung Wu, Wei-Shiang Hsiu, Yan-Jun Liu and Kuan-Wei Chuang

Department of Chemical Engineering, National Chung Cheng University, Chiayi 62102, Taiwan; ww@cs.ccu.edu.tw (W.-H.W.); math30512@gmail.com (W.-S.H.); kid5911027@gmail.com (Y.-J.L.); tomjuicf@gmail.com (K.-W.C.)

* Correspondence: chmfs@ccu.edu.tw; Tel.: +886-5-2720411 (ext. 33404)

Received: 11 November 2019; Accepted: 21 December 2019; Published: 25 December 2019



Abstract: Although cancer has historically been regarded as a cell proliferation disorder, it has recently been considered a metabolic disease. The first discovery of metabolic alterations in cancer cells refers to Otto Warburg's observations. Cancer metabolism results in alterations in metabolic fluxes that are evident in cancer cells compared with most normal tissue cells. This study applied protein expressions of normal and cancer cells to reconstruct two tissue-specific genome-scale metabolic models. Both models were employed in a tri-level optimization framework to infer oncogenes. Moreover, this study also introduced enzyme pseudo-coding numbers in the gene association expression to avoid performing posterior decision-making that is necessary for the reaction-based method. Colorectal cancer (CRC) was the topic of this case study, and 20 top-ranked oncogenes were determined. Notably, these dysregulated genes were involved in various metabolic subsystems and compartments. We found that the average similarity ratio for each dysregulation is higher than 98%, and the extent of similarity for flux changes is higher than 93%. On the basis of surveys of PubMed and GeneCards, these oncogenes were also investigated in various carcinomas and diseases. Most dysregulated genes connect to *catalase* that acts as a hub and connects protein signaling pathways, such as those involving *TP53*, *mTOR*, *AKT1*, *MAPK1*, *EGFR*, *MYC*, *CDK8*, and *RAS* family.

Keywords: oncogene; flux balance analysis; cancer cell metabolism; tissue-specific metabolic models; constraint-based modeling; multi-level optimization

1. Introduction

Recent developments in omics data, such as genomics [1], transcriptomics [2], proteomics [3], metabolomics [4], and fluxomics [5], can be observed by consulting numerous biological databases. Such advancement in high-throughput data acquisition has shifted the focus from data generation to processing and understanding how to integrate the collected information. Because the amount of available omics data is increasing rapidly, advanced computational methods are required to mine the relevant knowledge from these data sources to boost the development of genome-scale metabolic models (GSMMs) of several organisms, including humans. Metabolism is the primary biological mechanism that is linked from genotype to phenotype; it can help us understand cell physiology and certain disease phenotypes caused by metabolic dysregulation [6,7]. GSMMs composed of metabolites and reactions allow the representation of the full set of metabolic processes within a cell to be curated using knowledge of cellular functions from the literature. GSMMs combined with constraint-based approaches lead to the development of methods to simulate cell behavior, such as flux balance analysis (FBA) [8].

A better understanding of the genome-scale human metabolic network may enable the identification of disease genes and related pathways, thereby offering better targets for drug development. The release of genome-scale human metabolic networks, such as Recon 1 and 2 [9,10], Edinburgh human metabolic network (EHMN) [11], and human metabolic reactions [12], has led to the emergence of network medicine. Recon 2.2 and Recon 3D are the most comprehensive human genome-scale network reconstructions [13,14]. Recon 3D includes three-dimensional metabolite and protein structure data and enables an integrated analysis of metabolic functions in humans [13]. Recon 2.2 ensures the charge balance in all reactions and improves the simulation of energy metabolism [14].

Human tissues have diverse metabolic functions, which are complex and specialized in different tissues and cell types. Mapping out these tissue-specific metabolisms in GSMs will advance our understanding of the metabolic basis of various physiological and pathological processes. Nevertheless, for the in-depth study of specific cell phenotypes, the development of tissue- or cell-type-specific metabolic models is necessary. Automated reconstruction algorithms developed to date can be broadly categorized into flux-dependent and pruning methods [15–22]. Flux-dependent methods identify an optimal genome-scale metabolic network (GSMN) and include the maximum number of high confidence reactions supported by substantial experimental data. In contrast, pruning methods start with a core set of reactions obtained through literature reviews or experimental data and proceed by removing the remaining reactions in the general reconstruction while maintaining functionality in the core set. Nevertheless, both algorithms aim to keep the final tissue-specific reconstruction as concise as possible. Schultz and Qutub [23] recently introduced a cost optimization reaction dependency assessment (CORDA) algorithm to build concise, but not minimalistic, tissue-specific metabolic models based on omics data and a general human metabolic reconstruction. Many biological databases, such as HPA [3], HMDB [4], and TCGA [24], are available and enable the reconstruction of tissue-specific metabolic models. Yizhak et al. have reviewed the challenges regarding to genome-scale modeling of cancer metabolism, and provided valuable insights for clinically relevant applications [25]. In this study, we first established an algorithm by integrating the CORDA method [23] with the HPA database [3] and the human metabolic network Recon 2.2 [14] to reconstruct GSMs both healthy and cancerous colorectal tissues. Colorectal cancer (CRC), a malignant tumor of the large intestine with an incidence of approximately 10%, is the second and third most commonly occurring cancer in females and males, respectively. An estimated 50,630 deaths from CRC were reported in 2018 in the United States [26]. The Taiwan Cancer Registry report revealed that CRC was the most frequently diagnosed cancer in men and the second most frequently diagnosed cancer in women in 2017 [27]. Molecular biologists have discovered several genes that contribute to the susceptibility of the two types of colon cancer, namely familial adenomatous polyposis (FAP) and hereditary nonpolyposis colon cancer (HNPCC). Notably, genes causing HNPCC and FAP are relatively easy to discover, unlike those that cause susceptibility to CRC. In this study, we introduced a tri-level optimization framework to infer details on dysregulated genes that induce CRC through metabolic reprogramming in cancer cells. The GSMs for the cancerous and healthy cells of the colorectal tissue were used to build templates of flux alterations between cells. The tri-level optimization problem is difficult to solve; therefore, the NHDE algorithm was applied to solve this problem.

2. Materials and Methods

2.1. Reconstruction of GSMs

The metabolic network reconstruction processes are described in Figure 1. The process has seven steps and involves the integration of four databases to establish constraint-based models of cancerous and healthy tissue cells. Protein expression data for all genes that appeared in the gene association of Recon 2.2 were retrieved from the HPA database (Figure 1A). All reactions were classified into high, medium, and negative confidence groups based on the dependency assessment (Figure 1B). Exchange reactions for required nutrition uptake and reactions of well-known cancer pathways were

involved in the high confidence group (Figure 1B). Recon 2.2, a human general metabolic network, and data on the dependence of tissue-specific reactions were provided as the input information for the CORDA algorithm to construct a tissue-specific metabolic network (Figure 1C), which was saved in Systems Biology Markup Language (SBML) format (Figure 1D) for model exchange. Both cancer (CA) and healthy (HT) models were built by using the CORDA algorithm and merged into a basal (BL) model. Each tissue-specific metabolic network comprised thousands reactions and species, rendering it difficult to modify and analyze manually. We developed a systems biology program (SBP) platform that supported the SBML file to perform the simulation and analysis of metabolic networks in the general algebraic modeling system (GAMS) environment. The SBP platform can transfer a metabolic network in SBML format to its GAMS model automatically (Figure 1E). The BL model in Figure 1F was used to investigate how healthy cells undergo metabolic reprogramming to become cancer cells. Finally, we can employ the BL model to compute network topology and conduct a flux variability analysis (Figure 1G).

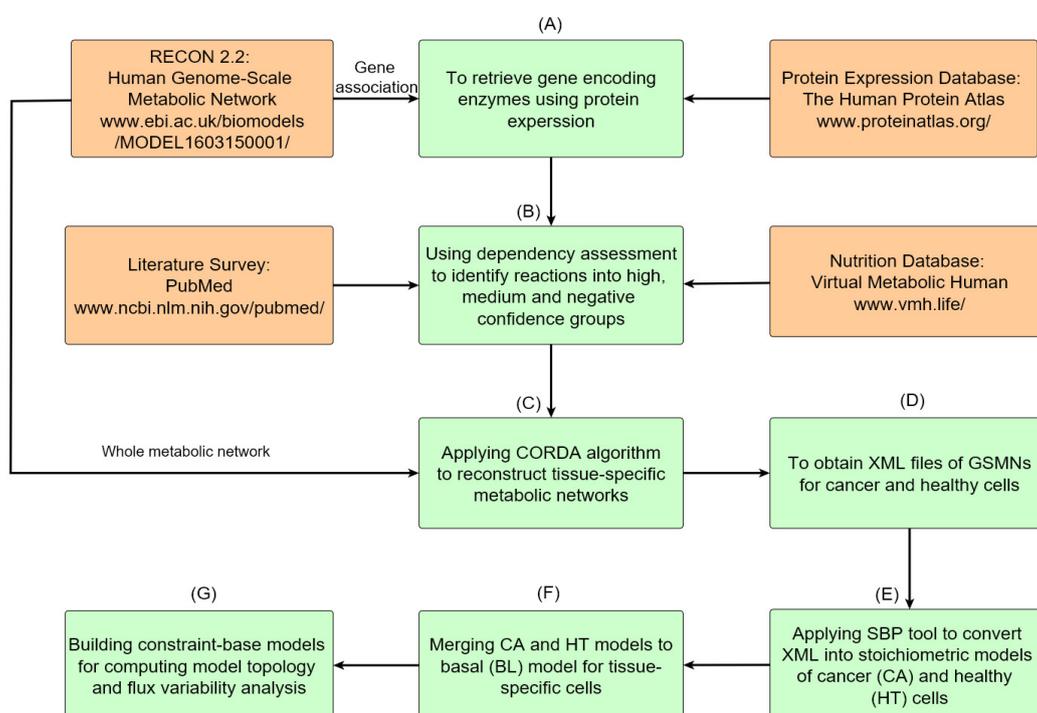


Figure 1. Roadmap of reconstruction of genome-scale metabolic networks for normal and cancer tissues. (A) Protein expressions of normal and cancerous colorectal tissues are accessed from HPA, and gene encoding enzymes are obtained through the gene association of Recon 2.2. (B) The acquired data are used to determine high, medium, and negative confidence sets of reactions. (C) The metabolic networks of cancer and healthy cells for the colorectal tissue are reconstructed using the CORDA algorithm. (D) Metabolic networks are stored in XML format. (E) The SBP tool is used to transfer metabolic networks to their stoichiometric models and gene-protein-reaction models. (F) Both cancer and healthy models are merged into a basal model. (G) The basal model can be used for further analysis and simulation.

2.2. Flux Balance Analysis

FBA is a constraint-based modeling approach where the stoichiometry of the underlying biochemical network constrains the solution. The stoichiometric matrix of a typical metabolic system is underdetermined, and infinite solutions are possible. The FBA formulates a metabolic network as a linear programming problem, as shown in Equation (1), wherein the solution of the underdetermined

system is a member of the solution space and optimizes an objective function of choice, such as maximal growth and maximal ATP synthesis rate.

$$\left\{ \begin{array}{l} \max_{\mathbf{v}_{f/b}} (w_{ATP}v_{ATP} + w_{biomass}v_{biomass}) \\ \text{subject to} \\ \mathbf{N}(\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\ v_{f/b,i}^{LB} \leq v_i \leq v_{f/b,i}^{UB}, i \in \Omega \end{array} \right. \quad (1)$$

where $\mathbf{v}_{f/b}$ 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 i^{th} backward/forward flux, respectively; Ω is the set of forward and backward fluxes. The coefficients (w_{ATP} and $w_{biomass}$) are the weighting factors for computing the linear objective function. The objective of FBA is to maximize the biomass formation rate ($v_{biomass}$) for the cancer model. However, normal cells may have different objectives depending on growth signals [28]; thus, the maximum ATP synthesis rate (v_{ATP}) is applied in this situation.

FBA assumes that metabolic networks will reach a steady state constrained by the stoichiometry. Even though the stoichiometric constraints lead to an underdetermined system, the bounded solution space of all feasible fluxes can be identified, that is, a large set of alternative flux distributions with an identical objective value. We minimized the squared sum of all internal fluxes for FBA to ensure efficient channeling of all fluxes through all pathways to eliminate numerous flux distributions in Equation (1). The Euclidean norm problem for minimization is expressed as follows:

$$\left\{ \begin{array}{l} \min_{\mathbf{v}_{f/b}} \sum_{i \in \Omega^{Int}} (v_{f,i})^2 + (v_{b,i})^2 \\ \text{subject to} \\ \mathbf{N}(\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\ v_{f/b,j}^{LB} \leq v_{f/b,j} \leq v_{f/b,j}^{UB}, j \in \Omega \\ w_{ATP}v_{ATP} + w_{biomass}v_{biomass} \geq obj^* \end{array} \right. \quad (2)$$

where obj^* is the optimal objective value obtained from Equation (1), and Ω^{Int} is the set of intracellular reactions. The problem expressed in Equation (2) is a quadratic programming problem that has a unique solution.

2.3. Oncogene Inference Problem

We introduced a tri-level optimization problem (TLOP) to simulate gene screening procedures in a wet lab to infer oncogenes. The flowchart of the in silico experiment is presented in Figure 2. GSMMS of cancerous and normal cells were reconstructed, as shown in Figure 2A, the procedures of which are discussed in Figure 1. Both models were then applied to compute the flux distribution patterns at the cancer and normal situations (Figure 2B). The flux template, which acted as a control, was built according to the flux distributions of cancer and normal models (Figure 2C). Flux alterations were computed from each mutant (Figure 2D) and then compared with the control (Figure 2E). The template was incorporated with the TLOP to infer oncogenes, as shown in Figure 2E–I. The outer optimization problem was employed to decide which genes were modulated (Figure 2F). The mutated genes were provided for the inner optimization problem to compute the flux distribution pattern to evaluate the flux alteration (Figure 2D–G). The TLOP framework can be formulated by using the procedures from Figure 2C–I, and is expressed as follows:

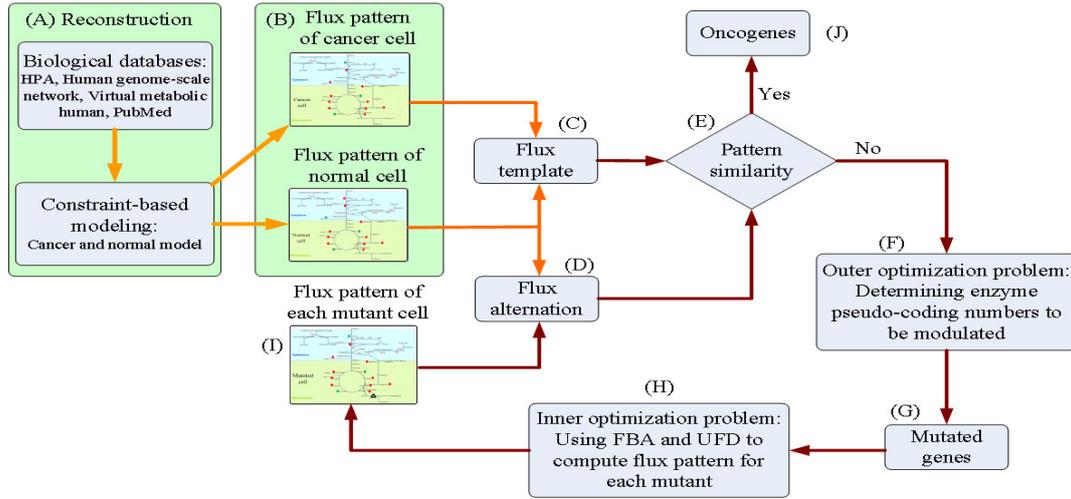


Figure 2. Flowchart of the in silico experiment for inferring oncogenes. (A) Reconstruct the cancer and normal models. (B) Compute the flux distributions of cancer and normal models. (C) Build the flux template according to the flux distributions of cancer and normal models. (D)–(I) Simulation of a wet lab experiment for determining oncogenes. Orange arrows indicate the building processes of the flux template that acted as the control in the oncogene inference problem. Red arrows present the mutant schemes for formulating the tri-level oncogene inference problem.

Outer optimization problem:

$$\left\{ \begin{array}{l}
 \text{Fuzzy equal of logarithmic fold change to the template:} \\
 \text{equal } LFC_m^{MUBL} \approx LFC_m^{CABL} \\
 \text{equal } LFC_m^{MUHT} \approx LFC_m^{CAHT} \\
 \text{Similarity ratio of flux-sum synthesis for metabolites:} \\
 \max_{\delta, z} SR^{T,HT}, \max_{\delta, z} SR^{T,BL} \\
 \text{Similarity ratio of flux for reactions:} \\
 \max_{\delta, z} SR^{T_{rxn},HT}, \max_{\delta, z} SR^{T_{rxn},BL} \\
 \text{subject to the inner optimization problems:} \\
 \text{FBA problem} \\
 \left\{ \begin{array}{l}
 \max_{\mathbf{v}_{f/b}} (w_{ATP}v_{ATP} + w_{biomass}v_{biomass}) \\
 \text{subject to} \\
 \mathbf{N}^{BL} (\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\
 v_{f/b,i}^{LB,MU} \leq v_{f/b,i} \leq v_{f/b,i}^{UB,MU}, z_i \in \Omega^{MU} \\
 v_{f/b,j}^{LB} \leq v_{f/b,j} \leq v_{f/b,j}^{UB}, z_j \notin \Omega^{MU}
 \end{array} \right. \\
 \text{UFD problem} \\
 \left\{ \begin{array}{l}
 \min_{\mathbf{v}_{f/b}} \sum_{i \in \Omega^{Int}} (v_{f,i})^2 + (v_{b,i})^2 \\
 \text{subject to} \\
 \mathbf{N}^{BL} (\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\
 v_{f/b,i}^{LB,MU} \leq v_{f/b,i} \leq v_{f/b,i}^{UB,MU}, z_i \in \Omega^{MU} \\
 v_{f/b,j}^{LB} \leq v_{f/b,j} \leq v_{f/b,j}^{UB}, z_j \notin \Omega^{MU} \\
 (w_{ATP}v_{ATP} + w_{biomass}v_{biomass}) \geq (w_{ATP}v_{ATP}^* + w_{biomass}v_{biomass}^*)
 \end{array} \right.
 \end{array} \right. \quad (3)$$

where $\widetilde{\text{equal}}$ 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; \mathbf{N}^{BL} is the stoichiometric matrix of basal models; the integer vector \mathbf{z} is used to determine mutated enzymes; δ is the regulated strength parameter for the mutants; and Ω^{MU} is the set of mutated reactions.

The hierarchical multiple objectives are considered in the outer optimization problem in Equation (3). The first priority is to employ the fuzzy equal measure to determine that the mutant logarithmic flux changes (LFC_m^{MUBL} and LFC_m^{MUHT}) are as close to the templates (LFC_m^{CABL} and LFC_m^{CAHT}) as possible. The logarithmic fold change (LFC_m) between the synthesis rates or fluxes of the m^{th} metabolite in cancerous or dysregulated (denoted as deficient) and basal or healthy (denoted as normal) states is computed as follows:

$$LFC_m = \log_2 \left(\frac{r_{m,deficient}}{r_{m,normal}} \right) \quad (4)$$

where the overall synthesis rate (r_m) of the m^{th} intracellular compound in deficient and normal states is evaluated as follows:

$$r_m = \sum_{i \in \Omega^c} \left(\sum_{N_{ij} > 0, j} N_{ij} v_{f,j} - \sum_{N_{ij} < 0, j} N_{ij} v_{b,j} \right), m \in \Omega^m \quad (5)$$

Here Ω^c is the set of metabolites located in different compartments, and Ω^m is the set of metabolites. The expression enclosed by brackets in Equation (5) indicates the synthesis rate of the i^{th} metabolite that summed the influxes of the forward reactions and backward reactions. Each intracellular metabolite may exist in different compartments of the metabolic network, e.g., nine compartments in Recon 2.2, so that the overall synthesis rate of the m^{th} compound is computed by Equation (5).

The second and third objectives are determining that the similarity ratios of the flux reprogramming are maximized. The similarity ratios ($SR^{T,BL/HT}$ and $SR^{T_{rxn},BL/HT}$) of the flux-sum synthesis for metabolites and ratios of fluxes for all reactions are evaluated as follows:

$$SR^{T/T_{rxn}} = \frac{\sum_{m=1} |\mu_m^{T/T_{rxn}}|}{N_{T/T_{rxn}}} \quad (6)$$

where the similarity indicator ($\mu_m^{T/T_{rxn}}$) for the m^{th} metabolite is defined as follows:

$$\mu_m^{T/T_{rxn}} = \begin{cases} 1, & \text{if } LFC_m > tol_+ \text{ and } LFC_m^{T/T_{rxn}} > tol_+ \\ -1, & \text{if } LFC_m < tol_- \text{ and } LFC_m^{T/T_{rxn}} < tol_- \\ 0, & \text{otherwise.} \end{cases} \quad (7)$$

where the logarithmic fold change ($LFC_m^{T/T_{rxn}}$) of the m^{th} metabolite for the templates are provided 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 (Supplementary Materials, Figure S1) to illustrate the computation of flux template, similarity ratio, and logarithmic fold change.

2.4. Association of Gene-Protein-Reaction

The TLOP framework must decide which genes are selected for mutation. Most optimal strain design methods [29–33] involve directly applying the reactions of the stoichiometric matrix to determine the optimal modulations. Posterior decision-making must be conducted to determine the corresponding gene that each optimal reaction integrates with the gene association. Such decision-making is tedious and incapable of identifying the reactions that are regulated by isozymes

or redundant genes. A regulatory FBA framework [34] has been used to apply gene association incorporated with the stoichiometric matrix to formulate a connective matrix, wherein logical operations “AND” and “OR” were applied to build a gene-protein-reaction (GPR) expression, as shown in Figure 3A, for conceptual description. Nonetheless, such a GPR model is still unable to discriminate redundant genes, and the regulatory FBA becomes a discrete nonlinear optimization problem. This study introduced a strategy for enzyme pseudo-coding numbers to separate the original gene association into two parts, as shown in Figure 3B. The first part can identify the redundant pseudo-enzymes and isozymes, such that the reduced association catalyzes the reactions. The second part represents single gene or complex gene corresponding to an enzyme pseudo-coding number.

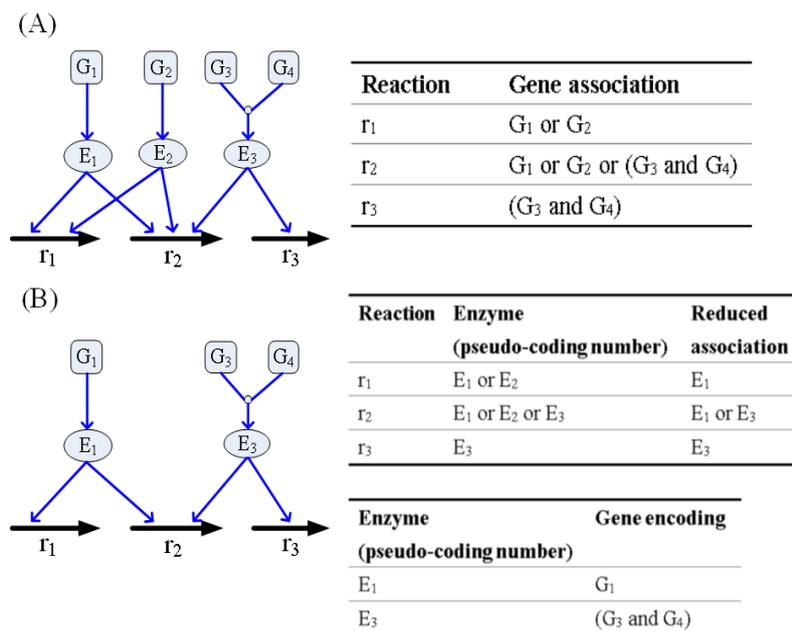


Figure 3. Example for building a gene-protein-reaction model using the enzyme pseudo-coding numbers. (A) Three reactions and their gene associations. (B) Reduced gene associations and encoded genes of the enzyme. E_2 is a redundant enzyme that is identical to E_1 and catalyzed the same reactions. r_2 is catalyzed by the isozymes (E_1 and E_3). E_1 is encoded by G_1 , and E_3 is encoded by a complex of G_3 and G_4 .

The pseudo-enzymes are then used as upper-level variables in Equation (3) to select modulated genes and the value for the regulated bounds is computed using the following equations:

Upregulation:

$$\begin{cases} (1 - \delta)v_{f,i}^{basal} + \delta v_{f,i}^{UB} \leq v_{f,i} \leq v_{f,i}^{UB} \\ v_{b,i}^{LB} \leq v_{b,i} \leq (1 - \delta)v_{b,i}^{basal} + \delta v_{b,i}^{LB}, i \in \Omega^{MU} \end{cases}$$

Downregulation:

$$\begin{cases} v_{f,i}^{LB} \leq v_{f,i} \leq (1 - \delta)v_{f,i}^{basal} + \delta v_{f,i}^{LB} \\ (1 - \delta)v_{b,i}^{basal} + \delta v_{b,i}^{UB} \leq v_{b,i} \leq v_{b,i}^{UB}, i \in \Omega^{MU} \setminus \Omega^{IZ} \\ v_{f,i}^{LB} \leq v_{f,i} \leq v_{f,i}^{UB} \\ v_{b,i}^{LB} \leq v_{b,i} \leq v_{b,i}^{UB}, i \in \Omega^{MU} \cap \Omega^{IZ} \end{cases} \quad (8)$$

Knockout:

$$\begin{cases} v_{f,i} = 0 \\ v_{b,i} = 0, i \in \Omega^{MU} \setminus \Omega^{IZ} \\ v_{f,i}^{LB} \leq v_{f,i} \leq v_{f,i}^{UB} \\ v_{b,i}^{LB} \leq v_{b,i} \leq v_{b,i}^{UB}, i \in \Omega^{MU} \cap \Omega^{IZ} \end{cases}$$

where Ω^{IZ} is the set of reactions regulated by isozymes. A bound for down-regulation or knockout is assigned to its original region if the reaction is catalyzed by isozymes.

2.5. Nested Hybrid Differential Evolution Algorithm

The candidate genes in Equation (3) are represented by integer variables such that it is a mixed-integer optimization problem that is NP-hard [35]. Classical algorithms for solving bi-level optimization problems have applied duality theory to convert the inner level optimization problem into constraints in the outer-level problem. However, the duality transformation is difficult for multilevel optimization problems, such as TLOP in this study. In this study, we extended the NHDE algorithm to solve the TLOP [36]. The computational concept of NHDE (Supplementary Materials, Figure S2) is based on a hybrid differential evolution (HDE), which was extended from the original differential evolution (DE) algorithm. The basic operations of the NHDE algorithm are similar to those of the DE and HDE algorithms, except for coding representation, selection, and evaluation operations. The NHDE algorithm has been applied to solve metabolic engineering [33] and biomedical problems [36,37]. In the outer optimization problem, the NHDE algorithm is used to determine which pseudo-enzymes are selected to be modulated, and the inner optimization problems (FBA and UFD) are then solved using a linear and quadratic optimization solver. An optimal solution for each candidate individual is achieved when the UFD problem is convergent, and the set of these individual solutions comprises a feasible solution to TLOP.

The TLOP framework in Equation (3) considers the hierarchical multiple objectives. We introduced the equal sum and minimum decision method to evaluate the fitness of individuals in NHDE. The evaluation procedures regarding the fitness η for individuals are expressed as follows:

$$\eta = \min \{ \eta_{CABL}, \eta_{CAHT} \} \quad (9)$$

where η_{CABL} and η_{CAHT} denote the fitness of CA model to BL and HT models, respectively, and they are evaluated using the hierarchical objectives as follows:

$$\eta_{CABL} = \eta_{MUBL} + \min \left\{ \eta_{MUBL}, \left(SR^{T,BL} + SR^{Trxn,BL} \right) / 2 \right\} \quad (10)$$

The η_{CAHT} is calculated using a similar procedure. The first priority of the objectives is to employ the fuzzy equal operation to determine that the mutant flux change ratios are as close to the templates as possible. The membership grades, η_{MUBL} and η_{MUHT} , for each fuzzy equal of logarithmic fold change are described in Supplementary Materials (Figure S3).

3. Results and Discussion

3.1. Templates of Flux Patterns for Cancer and Normal Cells

The metabolic model of Recon 2.2 consists of 5324 species, 7785 reactions, and 1675 associated genes (Figure 1A). Until date, Recon 2.2 has not been employed to reconstruct a tissue-specific GSMN model. Overall, 12865 gene expression profiles for normal and tumorous human tissues were acquired from HPA in the first step of the proposed procedures (Figure 1A). Based on the protein expressions and literature survey (Figure 1B), 1365 (including 74 reactions from literature) and 721 high confidence reactions for normal and cancerous colorectal tissues, respectively, were retrieved. The CORDA algorithm was applied to reconstruct the HT and CA colorectal models. The HT model comprised 2591 species, 4034 reactions, and 1483 genes, and the CA model comprised 2404 species, 3725 reactions, and 1454 genes. Figure 4 presents the intersections of species, reactions, and genes for both models. Both models were merged into the BL model to investigate how the healthy cell could be metabolically reprogrammed to become cancerous. Thus, the BL model is the union set of HT and CA models and includes 2692 species, 4284 reactions, and 1539 genes. These three models were engaged in the FBA

and UFD problems to compute flux-sum distributions. The templates of flux-sum alterations for CA to BL and HT models were computed using the LFC_m defined in Equation (4).

Total 565 choke-point metabolites (metabolites involve in only one synthesis or degradation reaction) in the normal and cancerous metabolic networks were determined by topology analysis. Due to that different uptakes can lead to a bias of the computational results, same uptakes (Supplementary Materials, Table S1) for normal and cancerous metabolic networks were used in this study.

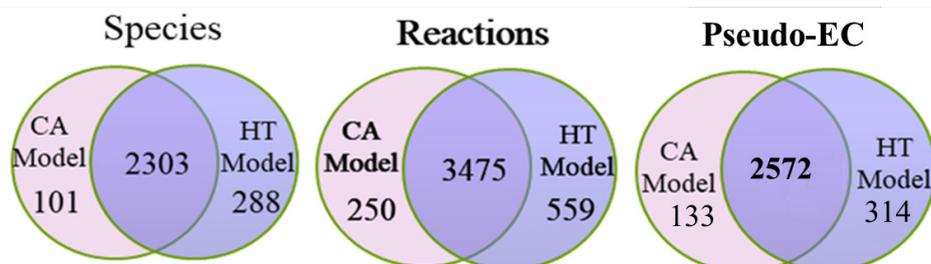


Figure 4. Statistics of cancer (CA) and healthy (HT) metabolomic models. Numbers of genes, species, and reactions for CA and HT models reconstructed by the CORDA algorithm taking the Recon 2.2 general model and HPA protein expression data as input. The basal (BL) model is composed of the union set of HT and CA models.

3.2. Inferred Oncogenes

The NHDE algorithm was applied to solve the TLOP, and the top 20 one-hit oncogenes were determined; their average similarity ratios are presented in Table 1. These genes participated in various metabolic pathways. The average similarity ratio (Ave. SR) for each mutant was greater than 98%, and the extent of similarity for flux change ratio (Ave. CR) was greater than 93%. The inferred results revealed that these mutants have a similar flux pattern to that of cancer cells. Surveys of PubMed and GeneCards indicated the inferred oncogenes are related to various carcinomas and diseases, as shown in Table 1, and among the results, seven oncogenes were observed in CRC. From the survey of GeneCards, we obtained 1360 and 852 genes associated with FAP and HNPCC, respectively, and there are 117 FAP-associated and 49 HNPCC-associated enzyme genes included in the reconstructed colorectal model (Supplementary Materials, Table S2). Three enzyme genes (*CAT*, *G6PD*, and *CDO1*) obtained from the NHDE algorithm are FAP-associated genes. Therefore, the results reveal that the inference optimization framework is a useful computer-aided oncogene detection system.

We also applied RNA-sequencing data of the healthy and cancerous colorectal tissues obtained from the TCGA database [24] on the analysis of the 20 inferred oncogenes. The RNA-sequencing data consist of 41 healthy samples and 478 cancer samples. A two-tailed *t*-test was used to analyze the RNA-sequencing data, and a *p*-value less than 0.05 (typically ≤ 0.05) is statistically significant. The result shows that 17 out of the 20 oncogenes are significant as shown in Table 1. The RNA-sequencing data obtained from the TCGA database can also be employed to build GSMNs for oncogene inference in the TLOP problem. The reconstructed GSMNs were used to evaluate the similarity ratio and the extent of similarity of the mutant flux patterns for the 20 oncogenes (Supplementary Materials, Table S3). We found that 19 out of the 20 oncogenes have high similarity level as those in Table 1. Moreover, GSMNs reconstructed by the CORDA and iMAT algorithms based on Recon 2.02 and Recon 3D general models were used to validate the 20 oncogenes. The computation results are shown in Supplementary Materials (Table S3) for comparison and reveals that these 20 oncogenes could also have high similarity ratios in different GSMNs.

Table 1. Top 20 one-hit oncogenes.

Gene	Pathway	Ave. CR ¶	Ave. SR §	p Value	Disease (Score) †	Remark ‡
CAT	Ethanol degradation	0.934	0.982	1.46×10^{-12}	Gonadoblastoma (1.42) Amelanotic Melanoma (1.39)	Related to ROS signaling pathway [38–40].
GPI	Pentose phosphate pathway	0.931	0.981	6.57×10^{-35}	Fibrosarcoma (1.08)	Gastric cancer [41].
PPA2	TRNA aminoacylation	0.935	0.982	0.4926	Sudden Cardiac Failure, Infantile (2.83)	Colorectal cancer [42] Prostate cancer [43].
HMGCL	Ketone body metabolism	0.935	0.982	3.8×10^{-29}	3-Hydroxy-3-Methylglutaryl-Coa Lyase Deficiency (2.83)	Nasopharyngeal carcinoma [44].
AGXT	Alanine and aspartate metabolism	0.933	0.982	0.0133	Hyperoxaluria, Primary, Type I (2.83)	Colorectal cancer [45].
GLRX2	PAK pathway	0.932	0.982	4.35×10^{-9}	NA	Oral squamous cell carcinoma [46].
GRHPR	Glyoxylate metabolism and glycine degradation	0.934	0.982	0.0018	Hyperoxaluria, Primary, Type Ii (2.83)	Hyperoxaluria [47].
G6PD	Methylene blue pathway	0.827	0.980	1.37×10^{-30}	Anemia (2.63) Glutathione Synthetase Deficiency (1.50)	Colorectal cancer [48] Obesity and diabetes [49].
H6PD	Pentose phosphate pathway	0.918	0.982	0.0018	Cortisone Reductase Deficiency 1 (2.83)	Cancer cell lines for colon, breast and lung [50,51].
G6PC3	Carbohydrate digestion and absorption	0.936	0.982	4.55×10^{-53}	Albinism, Oculocutaneous, Type Iv (1.26)	Breast cancer [52] Neutropenia [53].
SLC26A6	Mineral absorption	0.934	0.982	0.8577	Inflammatory Diarrhea (1.50)	Colorectal cancer cell lines [54] Pancreatic cancer cell [55].
SLC37A4	Carbohydrate digestion and absorption	0.930	0.982	0.4026	Glycogen Storage Disease (2.83) Pancreatic Ductal Adenocarcinoma (0.43)	Congenital hyperinsulinism of infancy [56].
SLC9A1	Osteoclast signaling	0.932	0.982	1.9×10^{-14}	Lichtenstein-Knorr Syndrome (2.83) Breast Cancer (0.38)	Colon cancer cells [57] Gliomas [58].
MLYCD	Peroxisomal lipid metabolism	0.933	0.982	1.84×10^{-8}	Malonyl-Coa Decarboxylase Deficiency (2.83) Pain-Chronic (1.43)	Proliferation of cancer cell lines [59].
PYCR3	Urea cycle and metabolism of amino groups	0.934	0.982	3.44×10^{-54}	Lung Cancer Susceptibility (0.42)	Related to metastasis of cancer cells [60].
PRODH2	Arginine and proline metabolism	0.933	0.981	4.2×10^{-5}	Primary Hyperoxaluria (1.34)	Hepatocellular carcinoma [61].
IMPDH1	Nucleotide metabolism	0.934	0.982	8.81×10^{-87}	Leber Congenital Amaurosis (2.83)	Small cell lung cancer [62].
CYBRD1	Mineral absorption	0.934	0.981	0.0013	Iron Metabolism Disease (1.36)	Breast and prostate cancer cells [63].
CDO1	Taurine and hypotaurine metabolism	0.934	0.982	1.08×10^{-5}	Small Intestine Cancer (1.31)	Colorectal cancer [64] Non-small cell lung cancer [65].
LIPC	Triacylglycerol degradation	0.940	0.981	0.0319	Hepatic Lipase Deficiency (2.83)	Colorectal cancer [66] Non-small cell lung carcinoma [67].

¶ Average similarity for flux change ratio; § Average similarity ratio of the mutant flux pattern to the template; † Disease is obtained from GeneCards database and score is accessed from GeneCards; ‡ Brief description of gene function and references from PubMed and cancer databases.

Catalase, encoded by *CAT*, is an essential antioxidant enzyme that catalyzes the decomposition of hydrogen peroxide (H_2O_2) to water and molecular oxygen. *CAT* catalyzes three decomposition reactions of H_2O_2 in the cytoplasm, peroxisome, and mitochondria of the GSMN model. Based on the computation results, we determined that a 78% downregulation of *CAT* could increase the production of reactive oxygen species (ROS) that exhibits carcinogenic effects [38,39]. We used the STRING database [68] to investigate how *CAT* is related to apoptosis, proliferation, and cell growth signaling pathways (Figure 5A). We observed that *CAT* was strongly related to the *TP53*, *mTOR*, *AKT1*, *MAPK1*, *EGFR*, *MYC*, *CDK8*, and *RAS* family; these are well-known carcinogenic genes. Notably, *CDK8* is a verified oncogene that induces aberrant activation of the canonical WNT/ β -catenin pathway occurring in almost all CRCs and contributes to cell growth, invasion, and survival [69]. We determined that *CAT* was also linked to *CDK8* through *TP53* and *MYC*. Finally, the 20 inferred genes in Table 1 were employed to investigate their protein-protein interactions (PPIs), as shown in Figure 5B. *CAT* showed all characteristics of a housekeeping gene, in that it acts as a hub to which most dysregulated genes are linked and then connected to the protein signaling pathways [40]. We determined from the STRING survey that *CYBRD1*, *CDO1*, and *LIPC* were disjointed in the PPI. The enzymes connected to *CAT* primarily participate in the glycolysis, pentose phosphate, amino acid, and lipid pathways. For example, *G6PD*, *H6PD*, *G6PC3*, *GPI*, *GRHPR*, and *SLC37A4* are associated with the glycolysis pathway. *G6PD* is a cytoplasmic enzyme that catalyzes the first step of the pentose phosphate pathway, which plays a key role in producing NADPH and the ribose required to synthesize DNA. *G6PD* deficiency is among the genetic factors hypothesized to protect against colorectal carcinogenesis [48].

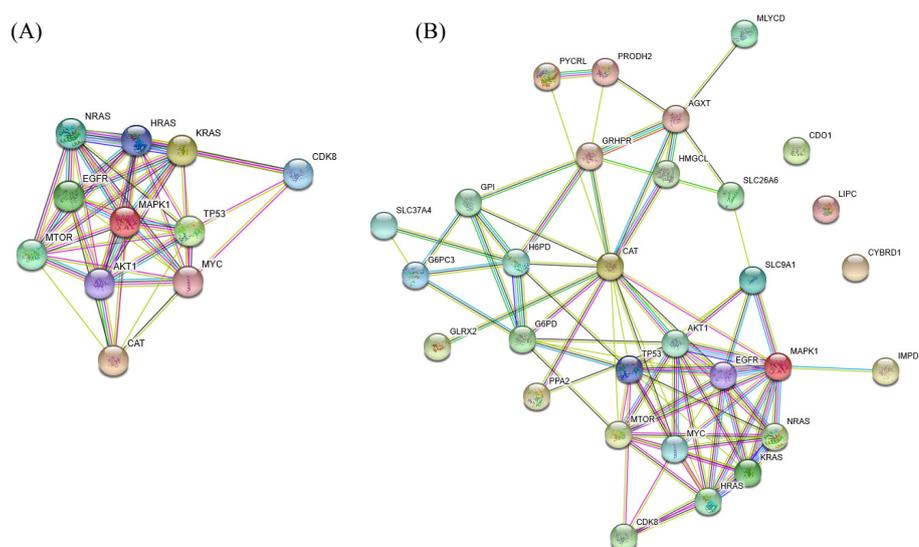


Figure 5. Protein-protein interactions (PPIs). (A) PPIs of the inferred oncogene *CAT*. *CAT* is strongly connected with the *TP53*, *mTOR*, *AKT1*, *MAPK1*, *EGFR*, *MYC*, *CDK8*, and *RAS* family. (B) *CAT* acts as a hub with most dysregulated genes linked to it.

3.3. Performance of Enzyme Pseudo-Coding

Furthermore, we applied a reaction-based approach to determine oncogenes in order to illustrate the performance of the proposed pseudo-enzyme strategy. Six top-ranked one-hit reactions are presented in Table 2. The reaction PGI is catalyzed by *GPI*, and the reaction r0161 is regulated by *AGXT*, such results are identical to that in Table 1. On the basis of the gene association of Recon 2.2, we determined that the gene *RPIA* catalyzes two reactions (r0249 and RPI). However, the reaction-based approach only determined r0249 with average similarity ratio of 0.981 and extent of flux changes of 0.935. The proposed pseudo-enzyme strategy was then applied to regulate both reactions. An average similarity ratio of 0.955 and average flux change ratio of 0.771 were obtained, which were smaller than

that obtained using the reaction-based approach. Following the posterior procedures, we observed that the gene *HMGCL* could obtain an identical result through the reaction-based approach, but it also catalyzed another reaction HMGLx and included the isozyme HMGCLL1. The gene *PRODH2* achieved the same result but catalyzed four reactions. Finally, *CAT* regulated three reactions (CATp, CATm, and r0010), but each reaction achieved different results.

Table 2. Top six one-hit reactions. The criteria may overestimate or underestimate compared with the results solved by the pseudo-enzyme strategy.

Reaction	Gene	Other Regulated Reactions	Isozyme	Ave. CR [§]	Ave. SR [†]	Remark [‡]
GPI	<i>GPI</i>	–	–	0.931	0.981	Gastric cancer [41].
r0161	<i>AGXT</i>	–	–	0.933	0.982	Colorectal cancer [45].
r0249	<i>RPIA</i>	RPI	–	0.935	0.981	Overestimated.
HMGLx	<i>HMGCL</i>	HMGLx	HMGCLL1	0.934	0.982	Nasopharyngeal carcinoma [44].
r0616	<i>PRODH2</i>	PROD2, r0615, PRO1x	–	0.934	0.982	Hepatocellular carcinoma [61].
CATp	<i>CAT</i>	CATp, r0010	–	0.932	0.982	Related to ROS signaling [38–40].
CATm	<i>CAT</i>	CATp, r0010	–	0.838	0.979	Underestimated, ROS signaling [38–40].
r0010	<i>CAT</i>	CATm, CATp	–	0.867	0.981	Underestimated, ROS signaling [38–40].

[§] Average similarity for flux change ratio; [†] Average similarity ratio of the mutant flux pattern to the template;

[‡] Brief description of gene function and references from PubMed and cancer databases.

3.4. Flux Variability Analysis

To avoid the solution bias of FBA used in the TLOP problem for evaluation of flux alterations, flux variability analysis (FVA) was applied to compute the minimum and maximum values of each metabolite for the normal model and the mutants. The flux ranges were compared between mutant and normal models, and each flux difference between mutant and normal models was classified into seven categories according to the definition presented in Supplementary Materials (Figure S4). The number of metabolites in different categories for the template and all mutants are shown in Figure 6. The result shows that about 71% of the synthesis rates of metabolites are complete decrease and about 5.8% complete increase. Such complete flux variances for all mutants are about 97% similar to the template and the average similarity ratios are about 82% as shown in Supplementary Materials (Figure S5). Furthermore, the choke-point metabolites in the normal and cancerous models were determined by topology analysis, and total 565 choke-point metabolites were determined. The flux variance patterns of the metabolites with gene association for the template and mutants are shown in Figure 7. Such choke-point metabolites can be used as essential candidates for discovering biomarkers, because the single flux of synthesis or degradation reaction for each choke-point metabolite is corresponding to the gene expression measurement. We observed that major deregulations occurred in lipids, nucleotides, carboxylic acids, organic acids, organic oxygen compounds, and organic heterocyclic compounds. Lipids consist of fatty acyls, glycerophospholipids, prenol lipids, sphingolipids, cholestane steroids, and steroidal glycosides. The fluxes of six choke-point metabolites (cardiolipin, 4beta-methylzymosterol-4alpha-carboxylic acid, 14-demethylanosterol, (S)-2,3-epoxysqualene, desmosterol, and 7-dehydrodesmosterol) among lipids are complete increase, and the others are decrease.

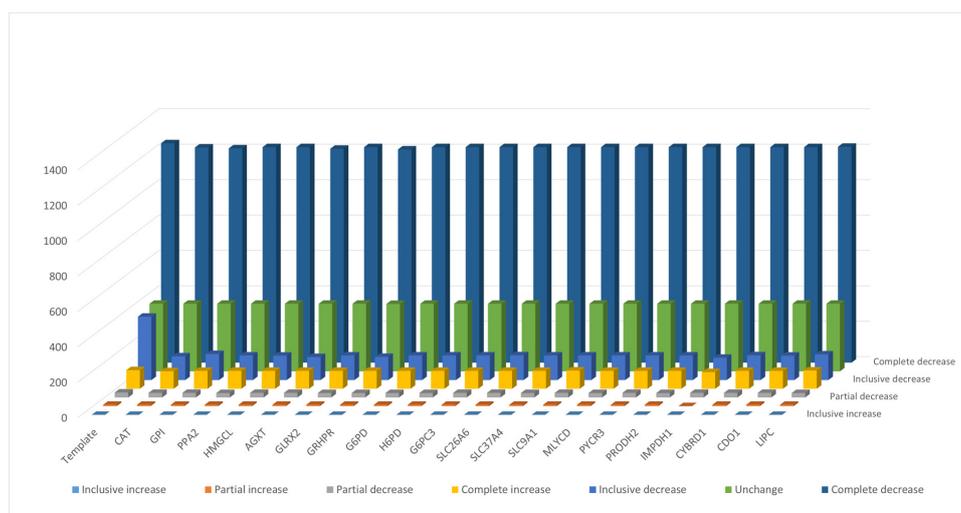


Figure 6. The number of metabolites in different categories for the template and 20 mutants. The definition of categories is presented in Supplementary Materials (Figure S4).

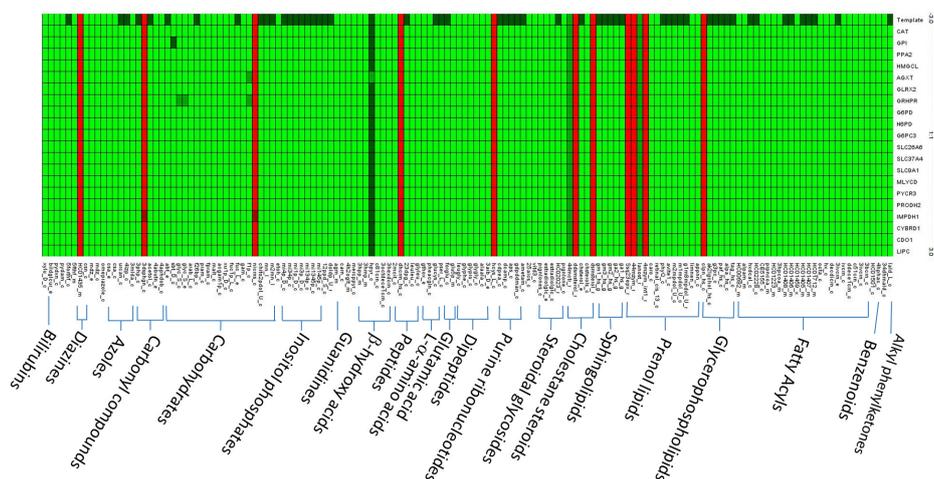


Figure 7. Flux variance patterns for the template and 20 mutants. Green indicates complete decrease, dark green means partial or inclusive decrease, and red denotes complete increase.

4. Conclusions

This study introduced a tri-level optimization framework that incorporated protein expressions of normal and tumor tissues for inferring oncogenes. The protein expression data or RAN-sequencing data, input of the optimization framework, was acquired from public databases such as HPA or TCGA. To enable provision of personalized medical treatment, the input of the optimization framework can be replaced with a patient's protein expression data. The proposed TLOP is an NP-hard problem, and no commercial software programs could be used to solve the problem. The NHDE algorithm with hierarchical objectives was applied to solve the problem. Moreover, this study introduced enzyme pseudo-coding numbers for expressing gene association to avoid the performing posterior decision-making that is necessary for the reaction-based method. CRC was applied as the case study, and the NHDE algorithm inferred 20 top-ranked oncogenes that cause various carcinomas and diseases based on surveys of PubMed and GeneCards. Notably, the evidence illustrated that the inference optimization framework enables oncogene prediction.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2218-1989/10/1/16/s1>, Figure S1: A numerical example to illustrate the computation of template, similarity ratio, and LFC, Figure S2: Flowchart of the NHDE algorithm for solving the TLOP, Figure S3: Definition of triangular membership function of fitness, Figure S4: Categories of flux variance between the normal and mutant model, Figure S5: Average similarity ratios computed by FVA, Table S1: Nutrients from RPMI-1640 and DMEM medium, Table S2: Genes associated to FAP and HNPCC from GeneCards database, Table S3: Average change ratios (CR) and similarity ratios (SR) of oncogenes in different GSMNs.

Author Contributions: Conceptualization, methodology and writing original draft: F.-S.W., program coding and editing: W.-H.W., data analysis and database survey: W.-S.H., Y.-J.L. and K.-W.C. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and Technology of Taiwan (Grant MOST106-2221-E-194-049-MY3 and MOST107-2627-M-194-001).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AKT1	AKT Serine/Threonine Kinase 1
AGXT	Alanine–Glyoxylate And Serine–Pyruvate Aminotransferase
BL	Basal
CA	Cancer
CAT	Catalase
CDK8	Cyclin Dependent Kinase 8
CDO1	Cysteine Dioxygenase Type 1
CRC	Colorectal Cancer
CYBRD1	Cytochrome B Reductase 1
EGFR	Epidermal Growth Factor Receptor
FAP	Familial Adenomatous Polyposis
FBA	Flux Balance Analysis
FVA	Flux Variability Analysis
G6PC3	Glucose-6-Phosphatase Catalytic Subunit 3
G6PD	Glucose-6-Phosphate Dehydrogenase
GLRX2	Glutaredoxin 2
GPI	Glucose-6-Phosphate Isomerase
GRHPR	Glyoxylate And Hydroxypyruvate Reductase
GSM	Genome-Scale Metabolic Model
H6PD	Hexose-6-Phosphate Dehydrogenase/Glucose 1-Dehydrogenase
HMGCL	3-Hydroxy-3-Methylglutaryl-CoA Lyase
HNPCC	Hereditary Nonpolyposis Colon Cancer
HT	Healthy
IMPDH1	Inosine Monophosphate Dehydrogenase 1
LFCm	Logarithmic Fold Change Ratio
LIPC	Lipase C, Hepatic Type
MAPK1	Mitogen-Activated Protein Kinase 1
MLYCD	Malonyl-CoA Decarboxylase
mTOR	Mammalian Target Of Rapamycin
MYC	Myc Proto-Oncogene Protein
PPA2	Pyrophosphatase (Inorganic) 2
PPI	Protein-Protein Interaction
PRODH2	Proline Dehydrogenase 2
PYCR3	Pyrroline-5-Carboxylate Reductase 3
SLC26A6	Solute Carrier Family 26 Member 6
SLC37A4	Solute Carrier Family 37 Member 4

SLC9A1	Solute Carrier Family 9 Member A1
TLOP	Triple-Level Optimization Problem
TP53	Tumor Protein P53
UFD	Uniform Flux Distribution

References

- Morozova, O.; Marra, M.A. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **2008**, *92*, 255–264. [[CrossRef](#)] [[PubMed](#)]
- Barrett, T.; Wilhite, S.E.; Ledoux, P.; Evangelista, C.; Kim, I.F.; Tomashevsky, M.; Marshall, K.A.; Phillippy, K.H.; Sherman, P.M.; Holko, M.; et al. NCBI GEO: Archive for functional genomics data sets—update. *Nucleic Acids Res.* **2013**, *41*, D991–D995. [[CrossRef](#)] [[PubMed](#)]
- Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; et al. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **2010**, *28*, 1248. [[CrossRef](#)] [[PubMed](#)]
- Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [[CrossRef](#)] [[PubMed](#)]
- Winter, G.; Krömer, J.O. Fluxomics – connecting ‘omics analysis and phenotypes. *Environ. Microbiol.* **2013**, *15*, 1901–1916. [[CrossRef](#)] [[PubMed](#)]
- Asgari, Y.; Zabihinpour, Z.; Salehzadeh-Yazdi, A.; Schreiber, F.; Masoudi-Nejad, A. Alterations in cancer cell metabolism: The Warburg effect and metabolic adaptation. *Genomics* **2015**, *105*, 275–281. [[CrossRef](#)]
- Jalili, M.; Gebhardt, T.; Wolkenhauer, O.; Salehzadeh-Yazdi, A. Unveiling network-based functional features through integration of gene expression into protein networks. *Biochim. Biophys. Acta (BBA) Mol. Basis Dis.* **2018**, *1864*, 2349–2359. [[CrossRef](#)]
- Bordbar, A.; Monk, J.M.; King, Z.A.; Palsson, B.O. Constraint-based models predict metabolic and associated cellular functions. *Nat. Rev. Genet.* **2014**, *15*, 107. [[CrossRef](#)]
- Duarte, N.C.; Becker, S.A.; Jamshidi, N.; Thiele, I.; Mo, M.L.; Vo, T.D.; Srivas, R.; Palsson, B.Ø. Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl. Acad. Sci. USA* **2007**, *104*, 1777–1782. [[CrossRef](#)]
- Thiele, I.; Swainston, N.; Fleming, R.M.T.; Hoppe, A.; Sahoo, S.; Aurich, M.K.; Haraldsdottir, H.; Mo, M.L.; Rolfsson, O.; Stobbe, M.D.; et al. A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **2013**, *31*, 419. [[CrossRef](#)]
- Ma, H.; Sorokin, A.; Mazein, A.; Selkov, A.; Selkov, E.; Demin, O.; Goryanin, I. The Edinburgh human metabolic network reconstruction and its functional analysis. *Mol. Syst. Biol.* **2007**, *3*, 135. [[CrossRef](#)] [[PubMed](#)]
- Mardinoglu, A.; Agren, R.; Kampf, C.; Asplund, A.; Nookaew, I.; Jacobson, P.; Walley, A.J.; Froguel, P.; Carlsson, L.M.; Uhlen, M.; et al. Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol. Syst. Biol.* **2013**, *9*, 649. [[CrossRef](#)] [[PubMed](#)]
- Brunk, E.; Sahoo, S.; Zielinski, D.C.; Altunkaya, A.; Dräger, A.; Mih, N.; Gatto, F.; Nilsson, A.; Preciat Gonzalez, G.A.; Aurich, M.K.; et al. Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **2018**, *36*, 272–281. [[CrossRef](#)] [[PubMed](#)]
- Swainston, N.; Smallbone, K.; Hefzi, H.; Dobson, P.D.; Brewer, J.; Hanscho, M.; Zielinski, D.C.; Ang, K.S.; Gardiner, N.J.; Gutierrez, J.M.; et al. Recon 2.2: From reconstruction to model of human metabolism. *Metabolomics* **2016**, *12*, 109. [[CrossRef](#)]
- Agren, R.; Bordel, S.; Mardinoglu, A.; Pornputtapong, N.; Nookaew, I.; Nielsen, J. Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput. Biol.* **2012**, *8*, 1–9. [[CrossRef](#)]
- 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. [[CrossRef](#)]
- Jerby, L.; Shlomi, T.; Ruppin, E. Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism. *Mol. Syst. Biol.* **2010**, *6*, 401. [[CrossRef](#)]

18. Machado, D.; Herrgård, M. Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol.* **2014**, *10*, e1003580. [[CrossRef](#)]
19. Opdam, S.; Richelle, A.; Kellman, B.; Li, S.; Zielinski, D.C.; Lewis, N.E. A systematic evaluation of methods for tailoring genome-scale metabolic models. *Cell Syst.* **2017**, *4*, 318–329. [[CrossRef](#)]
20. Vlassis, N.; Pacheco, M.P.; Sauter, T. Fast reconstruction of compact context-specific metabolic network models. *PLoS Comput. Biol.* **2014**, *10*, e1003424. [[CrossRef](#)]
21. Wang, Y.; Eddy, J.A.; Price, N.D. Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* **2012**, *6*, 153. [[CrossRef](#)] [[PubMed](#)]
22. Wu, M.; Chan, C. Human metabolic network: Reconstruction, simulation, and applications in systems biology. *Metabolites* **2012**, *2*, 242–253. [[CrossRef](#)] [[PubMed](#)]
23. Schultz, A.; Qutub, A.A. Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Comput. Biol.* **2016**, *12*, 1–33. [[CrossRef](#)] [[PubMed](#)]
24. The Cancer Genome Atlas Program. National Cancer Institute of U.S. Department of Health and Human Services. 2019. Available online: <https://www.cancer.gov/tcga> (accessed on 1 November 2019).
25. Yizhak, K.; Chaneton, B.; Gottlieb, E.; Ruppin, E. Modeling cancer metabolism on a genome scale. *Mol. Syst. Biol.* **2015**, *11*, 817. [[CrossRef](#)] [[PubMed](#)]
26. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2018. *CA Cancer J. Clin.* **2019**, *68*, 7–30. [[CrossRef](#)]
27. Taiwan Health Promotion Administration. Taiwan Cancer Registry, Cancer Statistics: Incidence and Mortality Rates for the Top 10 Cancer in Taiwan, 2008–2014. 2018. Available online: <http://tcr.cph.ntu.edu.tw/main.php?Page=N2> (accessed on 24 November 2018).
28. Nam, H.; Campodonico, M.; Bordbar, A.; Hyduke, D.R.; Kim, S.; Zielinski, D.C.; Palsson, B.O. A systems approach to predict oncometabolites via context-specific genome-scale metabolic networks. *PLoS Comput. Biol.* **2014**, *10*, e1003837. [[CrossRef](#)]
29. Burgard, A.P.; Pharkya, P.; Maranas, C.D. OptKnock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization. *Biotechnol. Bioeng.* **2003**, *84*, 647–657. [[CrossRef](#)]
30. Pharkya, P.; Burgard, A.P.; Maranas, C.D. OptStrain: A computational framework for redesign of microbial production systems. *Genome Res.* **2004**, *14*, 2367–2376. [[CrossRef](#)]
31. Pharkya, P.; Maranas, C.D. An optimization framework for identifying reaction activation/inhibition or elimination candidates for overproduction in microbial systems. *Metab. Eng.* **2006**, *8*, 1–13. [[CrossRef](#)]
32. Ranganathan, S.; Suthers, P.F.; Maranas, C.D. OptForce: An optimization procedure for identifying all genetic manipulations leading to targeted overproductions. *PLoS Comput. Biol.* **2010**, *6*, e1000744. [[CrossRef](#)]
33. Wang, F.S.; Wu, W.H. Optimal design of growth-coupled production strains using nested hybrid differential evolution. *J. Taiwan Inst. Chem. Eng.* **2015**, *54*, 57–63. [[CrossRef](#)]
34. Kim, J.; Reed, J. OptORF: Optimal metabolic and regulatory perturbations for metabolic engineering of microbial strains. *BMC Syst. Biol.* **2010**, *4*, 53. [[CrossRef](#)]
35. Pozo, C.; Miró, A.; Guillén-Gosálbez, G.; Sorribas, A.; Alves, R.; Jiménez, L. PGlobal optimization of hybrid kinetic/FBA models via outer-approximation. *Comput. Chem. Eng.* **2015**, *72*, 325–333. [[CrossRef](#)]
36. Wu, W.H.; Chien, C.Y.; Wu, Y.H.; Wu, H.H.; Lai, J.M.; Chang, P.M.H.; Huang, C.Y.F.; Wang, F.S. Inferring oncoenzymes in a genome-scale metabolic network for hepatocytes using bilevel optimization framework. *J. Taiwan Inst. Chem. Eng.* **2018**, *91*, 97–104. [[CrossRef](#)]
37. Hsu, K.C.; Wang, F.S. Detection of minimum biomarker features via bi-level optimization framework by nested hybrid differential evolution. *J. Taiwan Inst. Chem. Eng.* **2017**, *81*, 31–39. [[CrossRef](#)]
38. Bauer, G. Tumor cell-protective catalase as a novel target for rational therapeutic approaches based on specific intercellular ROS signaling. *Anticancer Res.* **2012**, *32*, 2599–2624. [[PubMed](#)]
39. Glorieux, C.; Calderon, P.B. Catalase, a remarkable enzyme: Targeting the oldest antioxidant enzyme to find a new cancer treatment approach. *Biol. Chem.* **2017**, *398*, 1095–1108. [[CrossRef](#)]
40. Glorieux, C.; Zamocky, M.; Sandoval, J.M.; Verrax, J.; Calderon, P.B. Regulation of catalase expression in healthy and cancerous cells. *Free Radical Biol. Med.* **2015**, *87*, 84–97. [[CrossRef](#)]
41. Ma, Y.T.; Xing, X.F.; Dong, B.; Cheng, X.J.; Guo, T.; Du, H.; Wen, X.Z.; Ji, J.F. Higher autocrine motility factor/glucose-6-phosphate isomerase expression is associated with tumorigenesis and poorer prognosis in gastric cancer. *Cancer Manag. Res.* **2018**, *10*, 4969–4980. [[CrossRef](#)]

42. Uozie, A.C.; Selevsek, N.; Wahlander, A.; Nanni, P.; Grossmann, J.; Weber, A.; Buffoli, F.; Marra, G. Targeted proteomics for multiplexed verification of markers of colorectal tumorigenesis. *Mol. Cell. Proteom. MCP* **2017**, *16*, 407–427. [[CrossRef](#)]
43. Pang, J.; Liu, W.P.; Liu, X.P.; Li, L.Y.; Fang, Y.Q.; Sun, Q.P.; Liu, S.J.; Li, M.T.; Su, Z.L.; Gao, X. Profiling Protein markers associated with lymph node metastasis in prostate cancer by DIGE-based proteomics analysis. *J. Proteome Res.* **2010**, *9*, 216–226. [[CrossRef](#)] [[PubMed](#)]
44. Luo, W.; Qin, L.; Li, B.; Liao, Z.; Liang, J.; Xiao, X.; Xiao, X.; Mo, Y.; Huang, G.; Zhang, Z.; et al. Inactivation of HMGCL promotes proliferation and metastasis of nasopharyngeal carcinoma by suppressing oxidative stress. *Sci. Rep.* **2017**, *7*, 11954. [[CrossRef](#)] [[PubMed](#)]
45. Kjersem, J.B.; Thomsen, M.; Guren, T.; Hamfjord, J.; Carlsson, G.; Gustavsson, B.; Ikdahl, T.; Indrebø, G.; Pfeiffer, P.; Lingjærde, O.; et al. AGXT and ERCC2 polymorphisms are associated with clinical outcome in metastatic colorectal cancer patients treated with 5-FU/oxaliplatin. *Pharmacogenom. J.* **2015**, *16*, 272–279. [[CrossRef](#)] [[PubMed](#)]
46. Pedro, N.F.; Biselli, J.M.; Maniglia, J.V.; Santi-Neto, D.D.; Pavarino, É.C.; Goloni-Bertollo, E.M.; Biselli-Chicote, P.M. Candidate biomarkers for oral squamous cell carcinoma: Differential expression of oxidative stress-related genes. *Asian Pac. J. Cancer Prev. APJCP* **2018**, *19*, 1343–1349.
47. Cramer, S.D.; Ferree, P.M.; Lin, K.; Milliner, D.S.; Holmes, R.P. The gene encoding hydroxypyruvate reductase (GRHPR) is mutated in patients with primary hyperoxaluria type II. *Hum. Mol. Genet.* **1999**, *8*, 2063–2069. [[CrossRef](#)]
48. Ju, H.Q.; Lu, Y.X.; Wu, Q.N.; Liu, J.; Zeng, Z.L.; Mo, H.Y.; Chen, Y.; Tian, T.; Wang, Y.; Kang, T.B.; et al. Disrupting G6PD-mediated Redox homeostasis enhances chemosensitivity in colorectal cancer. *Oncogene* **2017**, *36*, 6282–6292. [[CrossRef](#)]
49. Buj, R.; Aird, K.M. Deoxyribonucleotide triphosphate metabolism in cancer and metabolic disease. *Front. Endocrinol.* **2018**, *9*, 177. [[CrossRef](#)]
50. Marini, C.; Ravera, S.; Buschiazzo, A.; Bianchi, G.; Orengo, A.M.; Bruno, S.; Bottoni, G.; Emionite, L.; Pastorino, F.; Monteverde, E.; et al. Discovery of a novel glucose metabolism in cancer: The role of endoplasmic reticulum beyond glycolysis and pentose phosphate shunt. *Sci. Rep.* **2016**, *6*, 25092. [[CrossRef](#)]
51. Tsachaki, M.; Mladenovic, N.; Štambergová, H.; Birk, J.; Odermatt, A. Hexose-6-phosphate dehydrogenase controls cancer cell proliferation and migration through pleiotropic effects on the unfolded-protein response, calcium homeostasis, and redox balance. *FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol.* **2018**, *32*, 2690–2705. [[CrossRef](#)]
52. Harami-Papp, H.; Pongor, L.S.; Munkácsy, G.; Horváth, G.; Nagy, Á.M.; Ambrus, A.; Hauser, P.; Szabó, A.; Tretter, L.; Györffy, B. TP53 mutation hits energy metabolism and increases glycolysis in breast cancer. *Oncotarget* **2016**, *7*, 67183–67195. [[CrossRef](#)]
53. Yeshayahu, Y.; Asaf, R.; Dubnov-Raz, G.; Schiby, G.; Simon, A.J.; Lev, A.; Somech, R. Testicular failure in a patient with G6PC3 deficiency. *Pediatr. Res.* **2014**, *76*, 197–201. [[CrossRef](#)] [[PubMed](#)]
54. Donnard, E.; Asprino, P.F.; Correa, B.R.; Bettoni, F.; Koyama, F.C.; Navarro, F.C.P.; Perez, R.O.; Mariadason, J.; Sieber, O.M.; Strausberg, R.L.; et al. Mutational analysis of genes coding for cell surface proteins in colorectal cancer cell lines reveal novel altered pathways, druggable mutations and mutated epitopes for targeted therapy. *Oncotarget* **2014**, *5*, 9199–9213. [[CrossRef](#)] [[PubMed](#)]
55. Aust, S.; Brucker, B.; Graf, J.; Klimpfner, M.; Thalhammer, T. Melatonin modulates acid/base transport in human pancreatic carcinoma cells. *Cell. Physiol. Biochem.* **2006**, *18*, 91–102 [[CrossRef](#)]
56. Cappello, A.R.; Curcio, R.; Lappano, R.; Maggiolini, M.; Dolce, V. The physiopathological role of the exchangers belonging to the SLC37 family. *Front. Chem.* **2018**, *6*, 122. [[CrossRef](#)] [[PubMed](#)]
57. Parks, S.K.; Cormerais, Y.; Durivault, J.; Pouyssegur, J. Genetic disruption of the pH_i-regulating proteins Na⁺/H⁺ exchanger 1 (SLC9A1) and carbonic anhydrase 9 severely reduces growth of colon cancer cells. *Oncotarget* **2017**, *8*, 10225–10237. [[CrossRef](#)] [[PubMed](#)]
58. Guan, X.; Luo, L.; Begum, G.; Kohanbash, G.; Song, Q.; Rao, A.; Amankulor, N.; Sun, B.; Sun, D.; Jia, W. Elevated Na/H exchanger 1 (SLC9A1) emerges as a marker for tumorigenesis and prognosis in gliomas. *J. Exp. Clin. Cancer Res.* **2018**, *37*, 255. [[CrossRef](#)]
59. Yizhak, K.; Gaude, E.; Le Dévédec, S.; Waldman, Y.Y.; Stein, G.Y.; van de Water, B.; Frezza, C.; Ruppin, E. Phenotype-based cell-specific metabolic modeling reveals metabolic liabilities of cancer. *eLife* **2014**, *3*, e03641. [[CrossRef](#)]

60. Elia, I.; Broekaert, D.; Christen, S.; Boon, R.; Radaelli, E.; Orth, M.F.; Verfaillie, C.; Grünewald, T.G.P.; Fendt, S.M. Proline metabolism supports metastasis formation and could be inhibited to selectively target metastasizing cancer cells. *Nat. Commun.* **2017**, *8*, 15267. [[CrossRef](#)]
61. Tang, L.; Zeng, J.; Geng, P.; Fang, C.; Wang, Y.; Sun, M.; Wang, C.; Wang, J.; Yin, P.; Hu, C.; et al. Global metabolic profiling identifies a pivotal role of proline and hydroxyproline metabolism in supporting hypoxic response in hepatocellular carcinoma. *Clin. Cancer Res.* **2018**, *24*, 474–485. [[CrossRef](#)]
62. Huang, F.; Ni, M.; Chalisehar, M.D.; Huffman, K.E.; Kim, J.; Cai, L.; Shi, X.; Cai, F.; Zacharias, L.G.; Ireland, A.S.; et al. Inosine monophosphate dehydrogenase dependence in a subset of small cell lung cancers. *Cell Metab.* **2018**, *28*, 369–382.e5. [[CrossRef](#)]
63. Rychtarcikova, Z.; Lettlova, S.; Tomkova, V.; Korenkova, V.; Langerova, L.; Simonova, E.; Zjablovskaja, P.; Alberich-Jorda, M.; Neuzil, J.; Truksa, J. Tumor-initiating cells of breast and prostate origin show alterations in the expression of genes related to iron metabolism. *Oncotarget* **2017**, *8*, 6376–6398. [[CrossRef](#)] [[PubMed](#)]
64. Vedeld, H.M.; Andresen, K.; Eilertsen, I.A.; Nesbakken, A.; Seruca, R.; Gladhaug, I.P.; Thiis-Evensen, E.; Rognum, T.O.; Boberg, K.M.; Lind, G.E. The novel colorectal cancer biomarkers CDO1, ZSCAN18 and ZNF331 are frequently methylated across gastrointestinal cancers. *Int. J. Cancer* **2015**, *136*, 844–853. [[CrossRef](#)] [[PubMed](#)]
65. Ooki, A.; Maleki, Z.; Tsay, J.C.J.; Goparaju, C.; Brait, M.; Turaga, N.; Nam, H.S.; Rom, W.N.; Pass, H.I.; Sidransky, D.; et al. A panel of novel detection and prognostic methylated DNA markers in primary non-small cell lung cancer and serum DNA. *Clin. Cancer Res.* **2017**, *23*, 7141–7152. [[CrossRef](#)] [[PubMed](#)]
66. Weidle, U.H.; Birzele, F.; Kruger, A. Molecular targets and pathways involved in liver metastasis of colorectal cancer. *Clin. Exp. Metastasis* **2015**, *32*, 623–635. [[CrossRef](#)]
67. Galluzzi, L.; Goubar, A.; Olaussen, K.A.; Vitale, I.; Senovilla, L.; Michels, J.; Robin, A.; Dorvault, N.; Besse, B.; Validire, P.; et al. Prognostic value of LIPC in non-small cell lung carcinoma. *Cell Cycle* **2013**, *12*, 647–654. [[CrossRef](#)]
68. Szklarczyk, D.; Morris, J.H.; Cook, H.; Kuhn, M.; Wyder, S.; Simonovic, M.; Santos, A.; Doncheva, N.T.; Roth, A.; Bork, P.; et al. The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **2017**, *45*, D362–D368. [[CrossRef](#)]
69. Firestein, R.; Bass, A.J.; Kim, S.Y.; Dunn, I.F.; Silver, S.J.; Guney, I.; Freed, E.; Ligon, A.H.; Vena, N.; Ogino, S.; et al. CDK8 is a colorectal cancer oncogene that regulates β -catenin activity. *Nature* **2008**, *455*, 547–551. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).