

Research



**Cite this article:** Wu W-H, Li F-Y, Shu Y-C, Lai J-M, Chang PM-H, Huang C-YF, 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. <http://dx.doi.org/10.1098/rsos.191241>

Received: 23 July 2019

Accepted: 26 February 2020

**Subject Category:**

Biochemistry, cellular and molecular biology

**Subject Areas:**

systems biology/computational biology/  
bioinformatics

**Keywords:**

cancer cell metabolism, flux balance analysis,  
head and neck squamous cell carcinoma,  
multiple-level optimization

**Author for correspondence:**

Feng-Sheng Wang

e-mail: [chmfs@ccu.edu.tw](mailto:chmfs@ccu.edu.tw)

Electronic supplementary material is available online at <https://doi.org/10.6084/m9.figshare.c.4887291>.

# Oncogene inference optimization using constraint-based modelling incorporated with protein expression in normal and tumour tissues

Wu-Hsiung Wu<sup>1</sup>, Fan-Yu Li<sup>1</sup>, Yi-Chen Shu<sup>1</sup>, Jin-Mei Lai<sup>2</sup>, Peter Mu-Hsin Chang<sup>3,4</sup>, Chi-Ying F. Huang<sup>5,6</sup> and Feng-Sheng Wang<sup>1</sup>

<sup>1</sup>Department of Chemical Engineering, National Chung Cheng University, Chiayi, Taiwan

<sup>2</sup>Department of Life Science, Fu Jen Catholic University, New Taipei City, Taiwan

<sup>3</sup>Department of Oncology, Taipei Veterans General Hospital, Taipei, Taiwan

<sup>4</sup>Faculty of Medicine, <sup>5</sup>Institute of Biopharmaceutical Sciences, and <sup>6</sup>Department of Biotechnology and Laboratory Science in Medicine, National Yang-Ming University, Taipei, Taiwan

F-SW, 0000-0001-5266-2346

Cancer cells are known to exhibit unusual metabolic activity, and yet few metabolic cancer driver genes are known. Genetic alterations and epigenetic modifications of cancer cells result in the abnormal regulation of cellular metabolic pathways that are different when compared with normal cells. Such a metabolic reprogramming can be simulated using constraint-based modelling approaches towards predicting oncogenes. We introduced the tri-level optimization problem to use the metabolic reprogramming towards inferring oncogenes. The algorithm incorporated Recon 2.2 network with the Human Protein Atlas to reconstruct genome-scale metabolic network models of the tissue-specific cells at normal and cancer states, respectively. Such reconstructed models were applied to build the templates of the metabolic reprogramming between normal and cancer cell metabolism. The inference optimization problem was formulated to use the templates as a measure towards predicting oncogenes. The nested hybrid differential evolution algorithm was applied to solve the problem to overcome solving difficulty for transferring the inner optimization problem into the single one. Head and neck squamous cells were applied as a case study to evaluate the

algorithm. We detected 13 of the top-ranked one-hit dysregulations and 17 of the top-ranked two-hit oncogenes with high similarity ratios to the templates. According to the literature survey, most inferred oncogenes are consistent with the observation in various tissues. Furthermore, the inferred oncogenes were highly connected with the TP53/AKT/IGF/MTOR signalling pathway through PTEN, which is one of the most frequently detected tumour suppressor genes in human cancer.

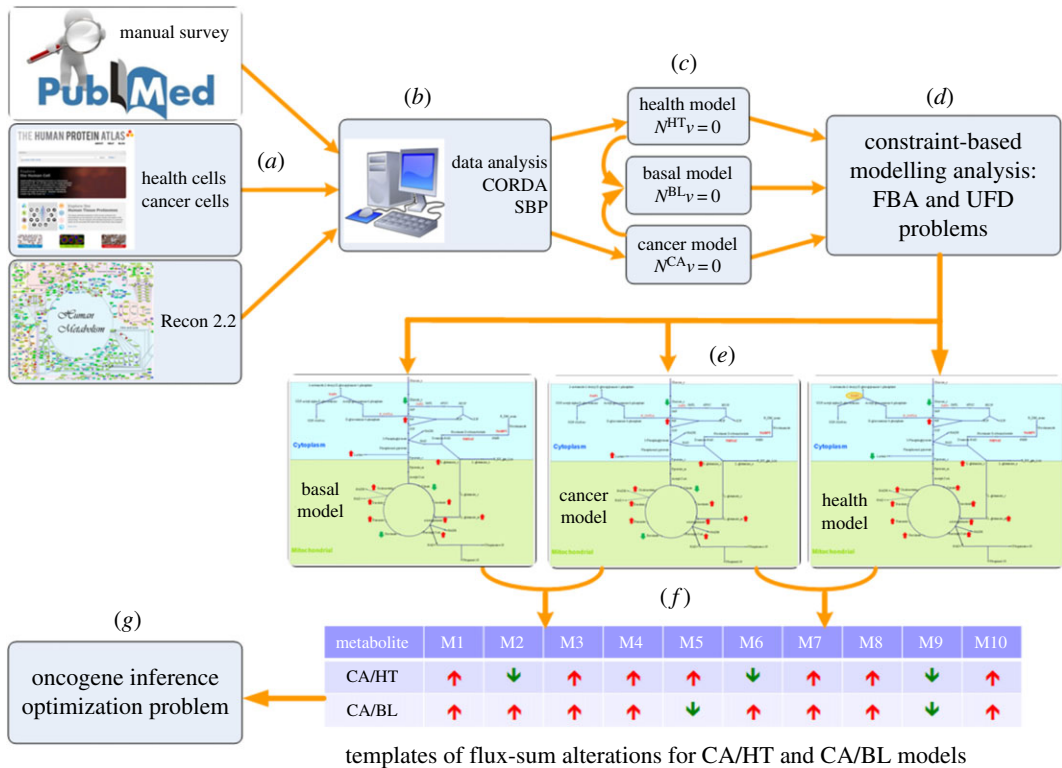
## 1. Introduction

Metabolic fluxes are altered in cancer cells and differ from the fluxes in the healthy situation. Normal cells primarily produce energy through mitochondrial oxidative phosphorylation. However, cancer cells undergo abnormal metabolism and use glucose differently than normal cells [1–4]. Cancer cells become habituated to certain fuel sources and metabolic pathways (metabolic reprogramming), profoundly changing how they consume and use nutrients such as glucose. Such flux reprogramming in cancer metabolism is characterized by high glucose consumption and lactate production, even under aerobic conditions, as well as increased glutamine catabolism and amino acid metabolism [1,2]. This is also termed as the Warburg effect [1–4]. Many recent studies on cancer metabolism have focused on the regulation of metabolic enzyme expression by oncogenes and tumour suppressors to identify multiple cancer-associated genes and pathways.

Human metabolism is complex and specialized in different tissue and cell types. Mapping out these tissue-specific metabolisms in genome-scale models will provide deeper insights into the metabolic basis of various physiological and pathological processes. Constraint-based model (CBM) approaches have been applied to predict oncogenes, essential enzymes and drug targets for developing novel medical treatments [5–11]. Flux-dependent and pruning methods are applied to reconstruct tissue-specific genome-scale metabolic models for predicting tissue-specific behaviours [6,10]. Schultz & Qutub [12] recently introduced cost optimization reaction dependency assessment (CORDA) to develop concise, but not minimalistic, tissue-specific metabolic models based on omics data and a generalized human metabolic reconstruction. The development of genome-scale human metabolic networks, such as Recon 1 [13], Recon 2 [14] and human metabolic reactions [8,15], has resulted in the emergence of network medicine. The recently updated and extended Recon 2 (i.e. Recon 2.2) and Recon 3D are the most comprehensive human genome-scale network reconstruction models [16,17]. Recon 3D [16] includes three-dimensional metabolite and protein structure data and enables the integrated analyses of metabolic functions in humans. Recon 2.2 is used to balance all reactions and improve the simulation of energy metabolism [17]. Recently, Richelle *et al.* [18] built 44 different genome-scale metabolic models from Recon 2.2 [17] and iHsa [19] using six model extraction methods with RNA-Seq data from NCI-60 cell line. Such an approach provides guidelines for the development of the next-generation of data contextualization methods. Ryu *et al.* [20] presented a systematic framework for the generation of gene-transcript-protein-reaction associations in the human metabolism and addition of new reactions from Recon 2.2 to build Recon 2M.2 that is biochemically consistent and transcript-level data compatible. Such gene-transcript-protein-reaction information enabled more accurate simulation of cancer metabolism and prediction of anticancer targets.

Wu *et al.* [21] developed a CBM based on the published Recon 2 and the Warburg effects of mouse hepatocytes deficient in miR-122a, and inferred that *DDC* is an oncogene. A total of 38 metabolic profiles obtained using LC/MS from the liver tissues of 10 control mice and 10 *Mir122a*<sup>-/-</sup> mice were applied to CBM for evaluating similarity ratios between the deficient and normal states. However, genome-wide information is not used in this approach, and the changes in metabolite concentrations may not be equal to the flux-sum alteration. In the present study, we established an algorithm to incorporate the Recon 2.2 network [17] with the Human Protein Atlas (HPA) database [22] and used the CORDA to reconstruct genome-scale metabolic network (GSMN) models of head and neck squamous cells (HNSCs) at healthy and cancer states. The models were then applied to build templates of flux alterations between cancer and normal cells.

A tri-level inference optimization framework integrated the templates and CBM was developed to infer dysregulated enzymes that contribute to inducing head and neck squamous cell carcinoma (HNSCC). Such framework can also be used to mimic gene screening procedures in wet laboratory for detecting dysregulated oncogenes. A tri-level optimization problem (TLOP) integrating splice-isoform expression has been introduced to depict breast cancer metabolism [23]. This study



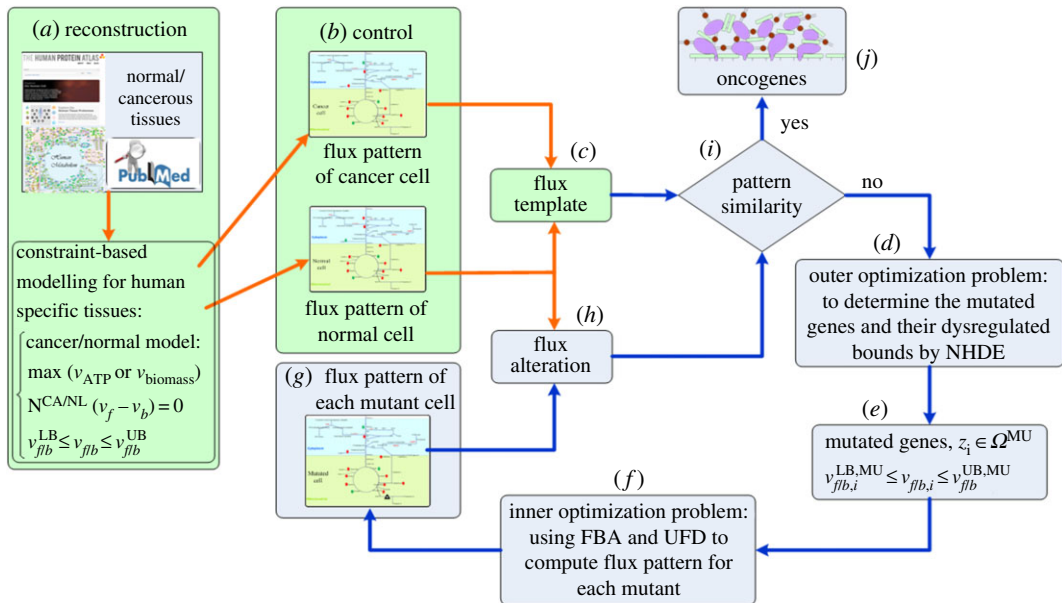
**Figure 1.** Templates of flux-sum alterations. (a) Procedure for survey and collection of information for head and neck squamous cells from PubMed, HPA and Recon 2.2. (b) Develop cancer (CA), healthy (HT) and basal (BL) specific models using CORDA and SBP algorithms. (c) The basal (BL) model, a union set of both models in (b), was used to investigate how healthy cells can smoothly detour the metabolic reprogramming to that of cancer cells. (d) Constraint-based methods were applied to the flux distributions for each model. (e) Flux-sum distributions were computed for each model. (f) Two templates of flux-sum alterations for CA related to HT and CA to BL models were generated. (g) Detect oncogenes by solving the oncogene inference optimization problem using the template similarity as the objective. The up-arrow/down-arrow notation indicates an increasing/decreasing flux-sum of the  $i$ th metabolite in cancer state.

introduced a similarity measure in the TLOP to decide mutated genes and their corresponding dysregulated bounds. The similarity between mutant flux patterns and templates of flux alterations was used as the objective. Duality theory is generally incapable of transforming inner problems into single-level problems. Solving the tri-level problem is difficult. We introduced a nested hybrid differential evolution (NHDE) algorithm for solving the TLOP to detect dysregulated oncogenes. The similarity measure was provided for NHDE to evolve new mutants for achieving higher ranked oncogenes.

## 2. Methods

### 2.1. Templates of flux-sum alterations for normal and cancer cells

This study introduced six steps (figure 1) to establish GSMN models for cancer (CA), healthy (HT) and basal (BL) models of head and neck squamous cells (HNSC) and to build their corresponding templates of flux-sum alterations. The BL model refers to the normal model, which represents normal situations of head and neck squamous tissues. PubMed literature survey, the HPA database [22] and the Recon 2.2 human metabolic network [17] were used to generate the specific information for HNSC (figure 1a,b). We first retrieve gene expression using the gene association of Recon 2.2 and the coding protein expression in normal and tumour tissues from the HPA database. Applying dependency assessment, the reactions having such expressions are identified into high, medium and negative confidence groups. The whole metabolic network of Recon 2.2 and the dependence reactions are provided as the input information for the CORDA algorithm [12] to reconstruct a



**Figure 2.** Oncogene inference problem incorporated with the templates. (a) Reconstruct the cancer and normal metabolic networks of colon cells using databases, such as HPA and Recon 2.2, and build the constraint-based model for the networks. (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 for the control of oncogene inference. (d–j) Blue arrows present a tri-level optimization problem to mimic mutant schemes in a wet laboratory for oncogene inference. (d) The outer optimization problem is employed to decide modulated genes by NHDE. (e) A set of mutated genes and their corresponding dysregulated bounds are obtained. (f) The inner optimization problem consisting of flux balance analysis (FBA) and uniform flux distribution (UFD) problems computes the flux distribution for each mutant. (g) Each mutant flux pattern is obtained from FBA and UFD. (h) Use the mutant flux patterns to evaluate flux alteration of each mutant. (i) Compare flux template and flux alteration until the pattern similarity satisfied; otherwise, repeat the computation. (j) Obtain the optimal oncogenes.

tissue-specific metabolic network (figure 1b), which consists of a thousand reactions and species. We have developed a systems biology program (SBP) tool to automatically build cancer and healthy stoichiometric models (figure 1c) in the GAMS programming language. The basal model (union set of CA and HT models) was used to investigate how the healthy cell could smoothly detour metabolic reprogramming to cancer. CBM approaches were then applied to analyse the structure of the models and to compute flux distributions for each model (figure 1c,d). Such flux distributions were applied to build the templates of flux-sum alterations for CA to HT models and CA to BL models (figure 1f). For example, flux-sum alteration of metabolite M1 increased between CA and HT model, but decreased for metabolite M2 as shown in figure 1f.

## 2.2. Tri-level optimization for inferring oncogenes

The oncogene inference framework (figure 2) integrated the templates and the TLOP was used to simulate a wet laboratory to detect dysregulated oncogenes in HNSC. The first, second and third procedures in figure 2 are discussed in figure 1 to build the flux template, which acts as a control in the oncogene inference problem. GSMN models of cancerous and normal cells were reconstructed as shown in figure 2a. Both models were then applied to compute the flux distribution patterns at the cancer and normal situations (figure 2b). The flux template was built according to the flux distributions of cancer and normal models (figure 2c). Blue arrows (figure 2d–j) present the TLOP to mimic mutant schemes in a wet laboratory for oncogene inference. The outer optimization problem was employed to decide mutation genes and their corresponding dysregulated bounds. The decision variables were used in the flux balance analysis (FBA) and uniform flux distribution (UFD) problems to hierarchically compute flux distribution for each mutant. The fluxes obtained from the decision variables in FBA and UFD problems were then applied to evaluate the similarity ratios in the outer problem. The evolutionary computation was repeated until

the optimal result was achieved. The TLOP was defined as follows:

$$\left\{ \begin{array}{l} \text{Outer optimization problem:} \\ \max_{\delta, z_i} (\text{SR}^{T, \text{HT}} + \text{SR}^{W, \text{HT}}) + \text{SR}^{T, \text{HT}} \text{SR}^{W, \text{HT}} \\ \max_{\delta, z_i} (\text{SR}^{T, \text{BL}} + \text{SR}^{W, \text{BL}}) + \text{SR}^{T, \text{BL}} \text{SR}^{W, \text{BL}} \\ \text{subject to the inner optimization problems,} \\ \text{FBA problem:} \\ \left\{ \begin{array}{l} \max_{\mathbf{v}_{f/b}} v_{\text{biomass}} \\ \text{subject to} \\ \mathbf{N}(\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\ v_{f/b,i}^{\text{LB}, \text{MU}} \leq v_{f/b,i} \leq v_{f/b,i}^{\text{UB}, \text{MU}}, z_i \in \Omega^{\text{MU}} \\ v_{f/b,j}^{\text{LB}} \leq v_{f/b,j} \leq v_{f/b,j}^{\text{UB}}, z_j \notin \Omega^{\text{MU}} \end{array} \right. \\ \text{UFD problem:} \\ \left\{ \begin{array}{l} \min_{\mathbf{v}_{f/b}} \sum_{k \in \Omega^{\text{int}}} (v_{f,k})^2 + (v_{b,k})^2 \\ \text{subject to} \\ \mathbf{N}(\mathbf{v}_f - \mathbf{v}_b) = \mathbf{0} \\ v_{f/b,i}^{\text{LB}, \text{MU}} \leq v_{f/b,i} \leq v_{f/b,i}^{\text{UB}, \text{MU}}, z_i \in \Omega^{\text{MU}} \\ v_{f/b,j}^{\text{LB}} \leq v_{f/b,j} \leq v_{f/b,j}^{\text{UB}}, z_j \notin \Omega^{\text{MU}} \\ v_{\text{biomass}} \geq v_{\text{biomass}}^* \end{array} \right. \end{array} \right. \quad (2.1)$$

where  $\mathbf{v}_f$  and  $\mathbf{v}_b$  are the forward and backward fluxes of reversible reactions, respectively;  $\mathbf{N}$  is an  $m \times n$  stoichiometric matrix, where  $m$  is the number of metabolites and  $n$  is the number of reactions;  $v_{f/b,j}^{\text{LB}}$  and  $v_{f/b,j}^{\text{UB}}$  are the positive lower and upper bounds of the  $j$ th forward/backward flux, respectively;  $v_{f/b,i}^{\text{LB}, \text{MU}}$  and  $v_{f/b,i}^{\text{UB}, \text{MU}}$  are the lower and upper bounds of the  $i$ th upregulation, downregulation, or knockout flux due to the  $i$ th enzyme dysregulation, which is determined using the outer optimization problem;  $v_{\text{biomass}}^*$  is the objective value obtained from FBA;  $\Omega^{\text{int}}$  is the set of internal reactions; and  $\Omega^{\text{MU}}$  is the set of mutated reactions. The objective of the FBA is the maximization of the biomass growth rate for the cancer and mutant cells. However, normal cells may have different objectives depending on growth signals [24]; thus, the maximization of ATP synthesis rate  $v_{\text{ATP}}$  was applied in this situation.

The reversible reactions were separated into forward and backward reactions,  $v_f$  and  $v_b$ , for building the constraint-based model of HNSC. This representation could directly compute flux-sum distributions for each metabolite, and evaluate the choke-point metabolites through the stoichiometric matrix. We define three categories of choke-point metabolites: a choke-point metabolite connected with a single-ingoing and multi-outgoing reaction, a choke-point metabolite connected with a multi-ingoing and single-outgoing reaction, or a choke-point metabolite connected with a single-ingoing and single-outgoing reaction (electronic supplementary material, S1).

The value for the dysregulated bounds is computed using the following equations:

Upregulation:

$$\left\{ \begin{array}{l} (1 - \delta)v_{f,i}^{\text{basal}} + \delta v_{f,i}^{\text{UB}} \leq v_{f,i} \leq v_{f,i}^{\text{UB}} \\ v_{b,i}^{\text{LB}} \leq v_{b,i} \leq (1 - \delta)v_{b,i}^{\text{basal}} + \delta v_{b,i}^{\text{LB}}, z_i \in \Omega^{\text{MU}} \end{array} \right.$$

Downregulation:

$$\left\{ \begin{array}{l} v_{f,i}^{\text{LB}} \leq v_{f,i} \leq (1 - \delta)v_{f,i}^{\text{basal}} + \delta v_{f,i}^{\text{LB}} \\ (1 - \delta)v_{b,i}^{\text{basal}} + \delta v_{b,i}^{\text{UB}} \leq v_{b,i} \leq v_{b,i}^{\text{UB}}, z_i \in \Omega^{\text{MU}} \end{array} \right. \quad (2.2)$$

Knockout:

$$v_{f,i} = v_{b,i} = 0, z_i \in \Omega^{\text{MU}},$$

where  $v_{f/b,i}^{\text{basal}}$  is the basal flux in the normal state;  $\delta$  determined in the outer optimization problem is the regulation strength parameter with a value within (0, 1]. The integer variable  $z_i$  is used to determine the dysregulated enzyme, so that the bounds of the modulated reactions are restricted to be backward or forward upregulation, backward or forward downregulation, or knockout reactions, as shown in equation (2.2).

The objective function in the outer optimization problem is a similarity measure that compares the templates of flux-sum alterations for CA with HT and BL models and with the Warburg hypothesis. Here,  $SR^{T,BL/HT}$  and  $SR^{W,BL/HT}$  are defined as the similarity ratios for the templates of flux-sum alterations compared CA with BL/HT models and the Warburg hypothesis obtained from the literature, respectively, and the similarity ratios (electronic supplementary material, S1) are evaluated as follows:

$$SR^{T/W} = \frac{\sum_{m=1}^{N_{T/W}} |\mu_m^{T/W}|}{N_{T/W}}, \quad (2.3)$$

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

$$\mu_m^{T/W} = \begin{cases} 1, & \text{if } LFC_m > \text{tol and } LFC_m^{T/W} > \text{tol} \\ -1, & \text{if } LFC_m < -\text{tol and } LFC_m^{T/W} < -\text{tol} \\ 0, & \text{otherwise,} \end{cases} \quad (2.4)$$

where the logarithmic fold changes in the  $m$ th metabolite,  $LFC_m^{T/W}$ , for the templates or the Warburg hypothesis are provided in advance. Here the tolerance is defined as  $\text{tol} = \log_2(1 + \varepsilon)$ , and  $\varepsilon$  is a percentage of flux alteration. The logarithmic fold change ( $LFC_m$ ) between the synthesis rates of the  $m$ th metabolite in cancer/dysregulated (denoted as deficient) and basal/healthy (denoted as normal) states is computed as follows:

$$LFC_m = \log_2 \left( \frac{r_{m,\text{deficient}}}{r_{m,\text{normal}}} \right), \quad (2.5)$$

where the overall synthesis rate ( $r_m$ ) of the  $m$ th intracellular compound at deficient and normal states is evaluated as follows:

$$r_m = \sum_{i \in \Omega^c} \left( \sum_{N_{ij} > 0, j} N_{ij} v_{f,j} - \sum_{N_{ij} < 0, j} N_{ij} v_{b,j} \right), \quad m \in \Omega^m. \quad (2.6)$$

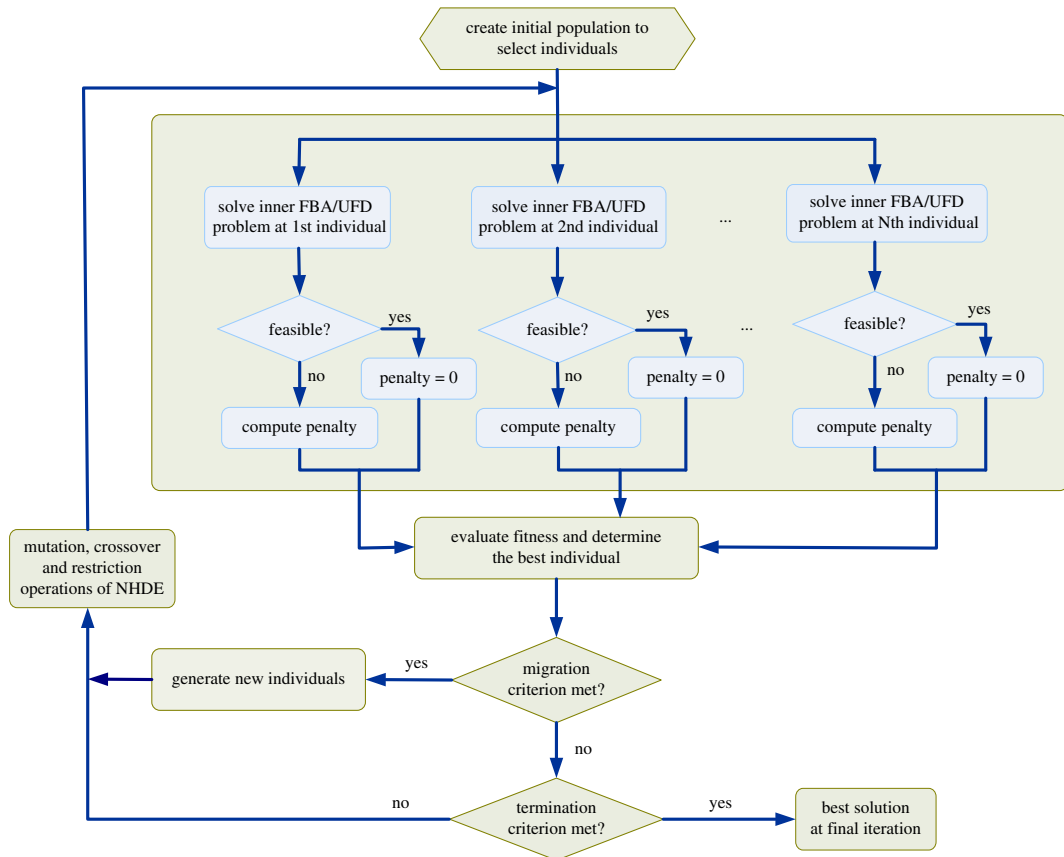
Here  $\Omega^c$  is the set of metabolites located in different compartments, and  $\Omega^m$  is the set of metabolites. The forward and backward fluxes ( $v_{f,j}$  and  $v_{b,j}$ ) are obtained from the UFD problem.

The TLOP in equation (2.1) consists of FBA and UFD problems in the inner optimization problem. The objective of FBA is to maximize the biomass formation rate for the cancer model. However, normal cells may have different objectives depending on growth signals [24]; thus, the maximum ATP synthesis rate is applied in this situation. The optimal flux distribution of the FBA problem is generally not unique; a large set of alternative flux distributions with an identical objective value exists. We minimize the squared sum of all internal fluxes for the UFD to ensure the efficient channelling of all fluxes through all pathways to eliminate the multiplicity of flux distributions in equation (2.1). The minimizing Euclidean norm problem is referred to as the UFD problem, which is a quadratic programming problem that has a unique solution.

### 2.3. Nested hybrid differential evolution

It is difficult to convert the inner optimization problems (FBA and UFD) of the TLOP into constraints by using duality transformation. In this study, we extended the NHDE algorithm to solve the TLOP (electronic supplementary material, S1). The computational concept of NHDE (figure 3) is based on hybrid differential evolution (HDE), which was extended from the original differential evolution (DE) algorithm [25–27]. The basic operations of the NHDE algorithm, except for coding representation, selection and evaluation operations, are similar to those of the DE and HDE algorithms. The NHDE algorithm has been applied to solve metabolic engineering [28] and biomedical problems [29,30]. The NHDE algorithm is used to identify integer variables in the outer optimization problem to determine which genes are selected to be regulated, 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 FBA problem is convergent, and the set of these individual solutions comprises a feasible solution to the TLOP.

In this study, the NHDE algorithm was implemented in the General Algebraic Modelling System (GAMS) environment, and its performance and solution quality depended on three setting factors: the tolerance ratio used in migration, population size and maximum number of iterations. The crossover factor and tolerance ratio were set to 0.5 and 0.05, respectively. A population size of 50 was used, and the maximum number of iterations was 100.



**Figure 3.** Flowchart of the NHDE algorithm. Hybrid differential evolution was applied to determine the outer decision variables and objective function. The inner optimization problem is a parallel computation in which each FBA problem was solved through linear programming to evaluate fitness in the NHDE algorithm.

### 3. Results and discussion

HNSCC, which accounts for more than 80% of head and neck malignancies, is one of the most common cancers worldwide. HNSCC is characterized by a high locoregional recurrence rate and includes cancers of the oral cavity, oropharynx, hypopharynx and larynx. The incidence and mortality rates of HNSCC vary by geographical location and site. A high incidence of oral cancer (excluding lip cancer) is found in South and Southeast Asia, parts of Western and Eastern Europe, parts of Latin America, and Caribbean and Pacific regions. Taiwan has one of the world's highest incidence rates of oral cancers. Improving the treatment outcome by identifying predictive targets for HNSCC treatment is an unmet clinical need. The aim of this study was to develop a computational strategy incorporated with expression profiles of proteins in head and neck tissue at normal and cancer states to reconstruct GSMN models.

#### 3.1. Templates of flux-sum alterations

Recon 2.2 is a curated version of Recon 2 and can be used to ensure full elemental balancing, in turn facilitating constraint-based analyses [17]. It comprises 5324 metabolites, 7785 reactions and 1675 associated genes. To date, Recon 2.2 has not been employed to reconstruct a GSMN model of a tissue-specific cell. The proposed procedures presented in figure 1 were used to reconstruct GSMN models of HNSCs at healthy and cancer states. Total 12865 gene expression profiles for human HNSCs and normal tissue were obtained from the HPA database. Based on the gene expression data and gene-protein-reaction (GPR) association, 384 and 411 high-confidence reactions at normal and cancer states were obtained, respectively, and applied to the CORDA algorithm to reconstruct the HNSC models. The reconstructed HNSC models comprised 2256 metabolites and 3427 reactions for the healthy (HT) model and 2147 metabolites and 3253 reactions for the cancer (CA) model. Both models identically had 82.2% and 75.7% of metabolites and reactions. The two models were combined to form a basal

**Table 1.** The inferred one-hit and two-hit oncogenes. AOX1, aldehyde oxidase; GFPT1, glutamine-fructose-6-phosphate aminotransferase [isomerizing] 1; GNPDA1, glucosamine-6-phosphate isomerase 1; IREB2, iron-responsive element-binding protein 2; OPLAH, 5-oxoprolinase; PGAM1, phosphoglycerate mutase 1; PI4KA, phosphatidylinositol 4-kinase alpha; PIK3CA, phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit alpha isoform; PIKFYVE, 1-phosphatidylinositol 3-phosphate 5-kinase; PIP5K1, phosphatidylinositol 4-phosphate 5-kinase type-1; RFK, riboflavin kinase; THTPA, thiamine-triphosphatase; TPK1, thiamin pyrophosphokinase 1.

gene	metabolic subsystem	Ave. SR <sup>a</sup>	remark <sup>b</sup>
<i>TPK1</i>	thiamine metabolism	0.843	vitamin B1/thiamine metabolism [31,32]
<i>G6PC</i>	glycolysis metabolism	0.841	metabolic regulator of glioblastoma [33]
<i>GNPDA1</i>	aminosugar metabolism	0.841	cell migration [34]
<i>PGAM1</i>	glycolysis metabolism	0.841	prognosis of OSCC/cell migration [35]
<i>GFPT1</i>	aminosugar metabolism	0.839	prognosis in patients with pancreatic cancer [36]
<i>RFK</i>	vitamin B2 metabolism	0.839	ROS production/necrotic cell death [37]
<i>THTPA</i>	thiamine metabolism	0.839	invasion/metastasis of cancers [38,39]
<i>OPLAH</i>	glutathione metabolism	0.838	potential marker in some human cancers [40]
<i>PTEN</i>	inositol phosphate metabolism	0.838	tumour suppressor [41,42]
<i>PI4KA</i>	inositol phosphate metabolism	0.837	therapy of HCC [43]
<i>PIKFYVE</i>	glycerophospholipid metabolism	0.837	cancer cell migration/invasion [44]
<i>PIP5K1</i>	glycerophospholipid metabolism	0.836	prognosis in prostate cancer [45]
<i>PIK3CA</i>	inositol phosphate metabolism	0.835	induce invasion in HNSCC [46]
( <i>IREB2, AOX1</i> )	citric acid cycle and vitamin B6 metabolism	0.871	biochemical recurrence of prostate cancer [47]
( <i>RFK, IREB2</i> )	vitamin B2 metabolism and citric acid cycle	0.867	see remark for RFK and IREB2
( <i>PTEN, IREB2</i> )	inositol phosphate metabolism and citric acid cycle	0.865	cellular proliferation [48]
( <i>G6PC, TPK1</i> )	glycolysis and thiamine metabolism	0.863	see remark for G6PC and TPK1
( <i>G6PC, PGAM1</i> )	glycolysis metabolism	0.857	see remark for G6PC and PGAM1

<sup>a</sup>Average similarity ratio of the mutant flux pattern to the template.

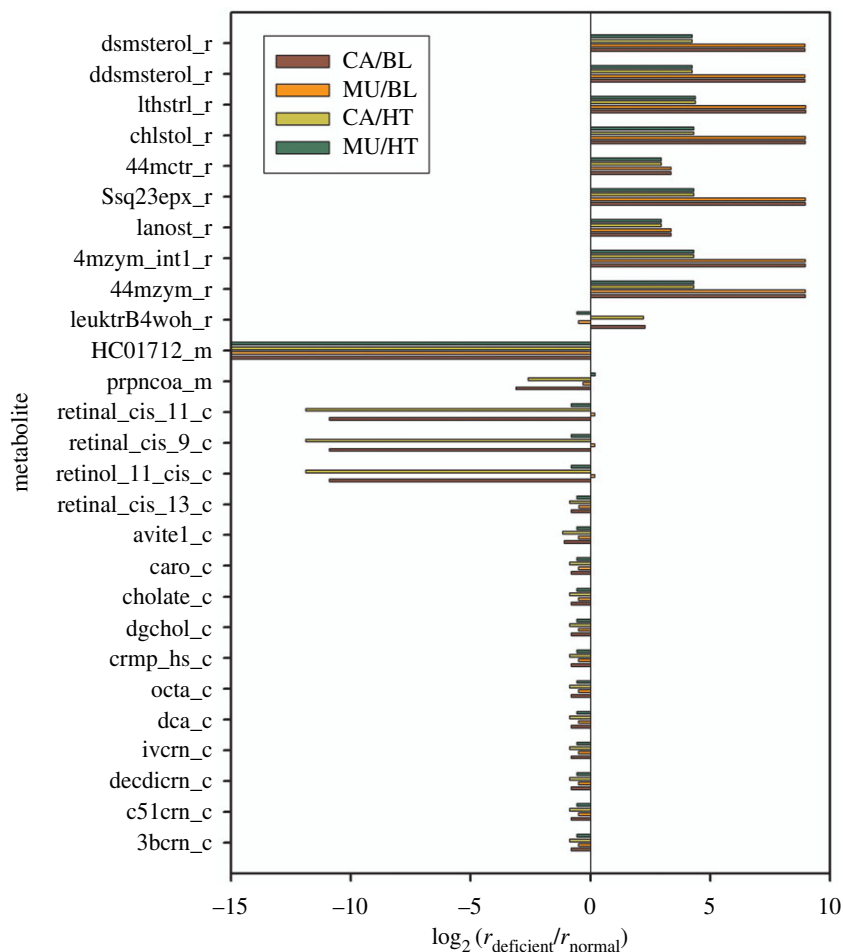
<sup>b</sup>Brief description of gene function and references.

(BL) model for investigating how a healthy cell can smoothly detour metabolic reprogramming to that of a cancer cell. The BL model was thus a union set of the HT and CA models and included 2417 metabolites and 3803 reactions. The three models were first employed in FBA and UFD problems to compute flux-sum distributions. We next used  $LFC_m$  to compute the templates of flux-sum alterations for CA to BL and HT models.

### 3.2. Detecting one-hit oncogene

The NHDE algorithm was applied to solve the TLOP, and 13 of the top-ranked one-hit oncogenes were determined; their similarity ratios are shown in table 1. These enzymes participated in seven metabolic subsystems, namely inositol phosphate, glycerophospholipid, glycolysis, thiamine, amino sugar, glutathione and vitamin B2 metabolisms. Four genes (*PTEN*, *PIKFYVE*, *PIK3CA* and *PI4KA*) were detected in inositol phosphate metabolism, wherein the participating molecules were involved in cellular growth and proliferation signalling processes. These genes were highly connected with AKT/TP53/mTOR signalling pathway. Phosphatidylinositol 3,4,5-trisphosphate 3-phosphatase (*PTEN*), encoded by *PTEN*, was dysregulated, which yielded the maximum similarity ratios of  $SR^{WBL} = 0.9286$  and  $SR^{WHT} = 0.8929$  (electronic supplementary material, S2) for the Warburg effect related to the

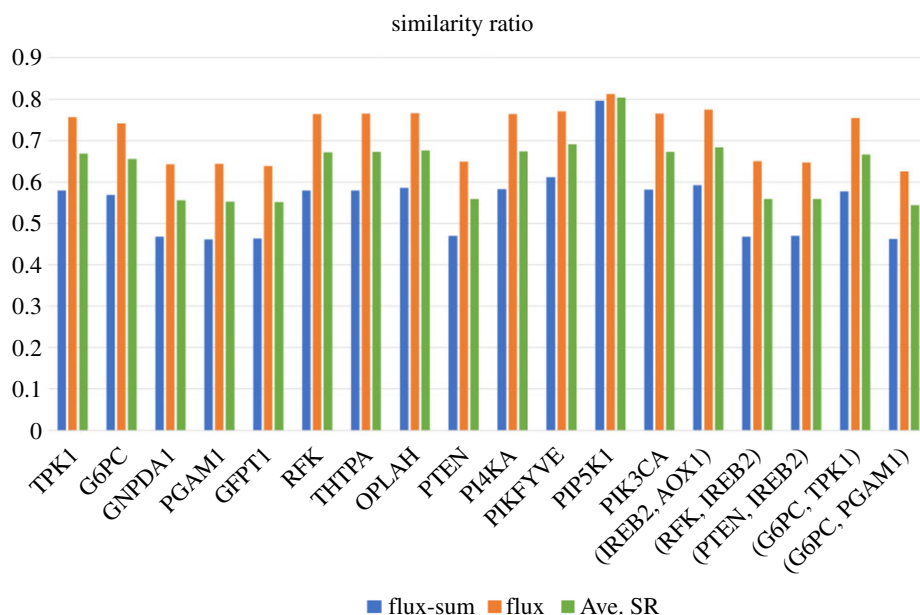




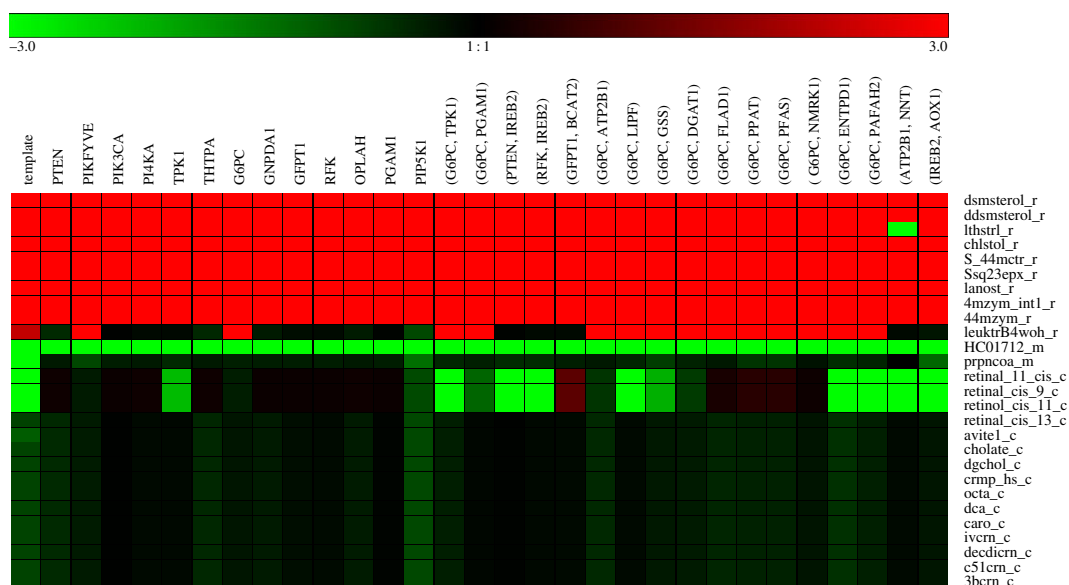
**Figure 4.** Logarithmic fold changes in the choke-point metabolites in lipid metabolism. In the endoplasmic reticulum, the levels of five cholesterols and their derivatives and four prenyl lipids increased. Fifteen choke-point lipids were located in the cytoplasm, and their synthesis rates decreased, except for three retinoids. The synthesis rates of the cancer template and *PTEN* dysregulation for S(8)-glutaryl dihydrolipoamide (HC01712\_m) in the mitochondria reduced to zero, thereby significantly reducing the LFC.

BL and HT models, respectively. *PTEN* dysregulation yielded the maximum ratios of  $SR^{T,BL} = 0.7478$  and  $SR^{T,HT} = 0.783$  compared with the templates of flux-sum alterations for the CA model to BL and HT models, respectively. The NHDE algorithm yielded the optimal average similarity (Ave. SR) ratio of 0.8381. *PTEN* is one of the most frequently observed tumour suppressor genes in human cancer. This biological cognition supports above computational results. Because altered AKT/PI3K/mTOR signalling has been frequently reported to contribute to human disease, researchers have made efforts to develop small-molecule inhibitors for AKT, PI3K and mTOR. Although the central role of *PTEN* in glucose metabolism has been well defined [31,43,46], little is known about how *PTEN* affects lipid metabolism [32,38]. The following results show the metabolic alterations by *PTEN* dysregulation.

Figure 4 shows the flux-sum alterations for 27 choke-point metabolites in lipid metabolism obtained using dysregulated *PTEN*. These choke-point metabolites were compartmented in the endoplasmic reticulum, mitochondrion and cytoplasm. In the endoplasmic reticulum, the flux-sum (LFC > 0) of five cholesterols and their derivatives and four prenyl lipids increased because of *PTEN* dysregulation. This observation indicated that the metabolite level should increase. The metabolite 20-hydroxy-leukotriene B4 (leuktrB4woh\_r) was classified as a fatty acyl in the endoplasmic reticulum; its synthesis rates in the templates increased but was inconsistent in dysregulated cases, as shown in figure 4. The synthesis rates of S(8)-glutaryl dihydrolipoamide (HC01712\_m) in the mitochondria in the cancer template and *PTEN* dysregulated cases reduced to zero; thus, the LFC was significantly reduced. The LFC of the template for acryloyl-CoA (prpncoa\_m) was inconsistent with the dysregulation. Fifteen choke-point lipids were located in the cytoplasm, and their synthesis rates decreased, except for three retinoids that exhibited increased flux rates for dysregulation. *N*-acylsphingosine 1-phosphate (crmp\_hs\_c) was classified as a sphingolipid in the cytoplasm, and its synthesis rate in the dysregulated cases was consistent with that in the template. Furthermore, the



**Figure 5.** Similarity ratios of flux and flux-sum synthesis rates evaluated by flux variability analysis in a posterior inspection. The similarity ratio for each mutant is the ratio of the number of metabolites with same flux/flux-sum trends in the flux interval of the mutant to those of the template.

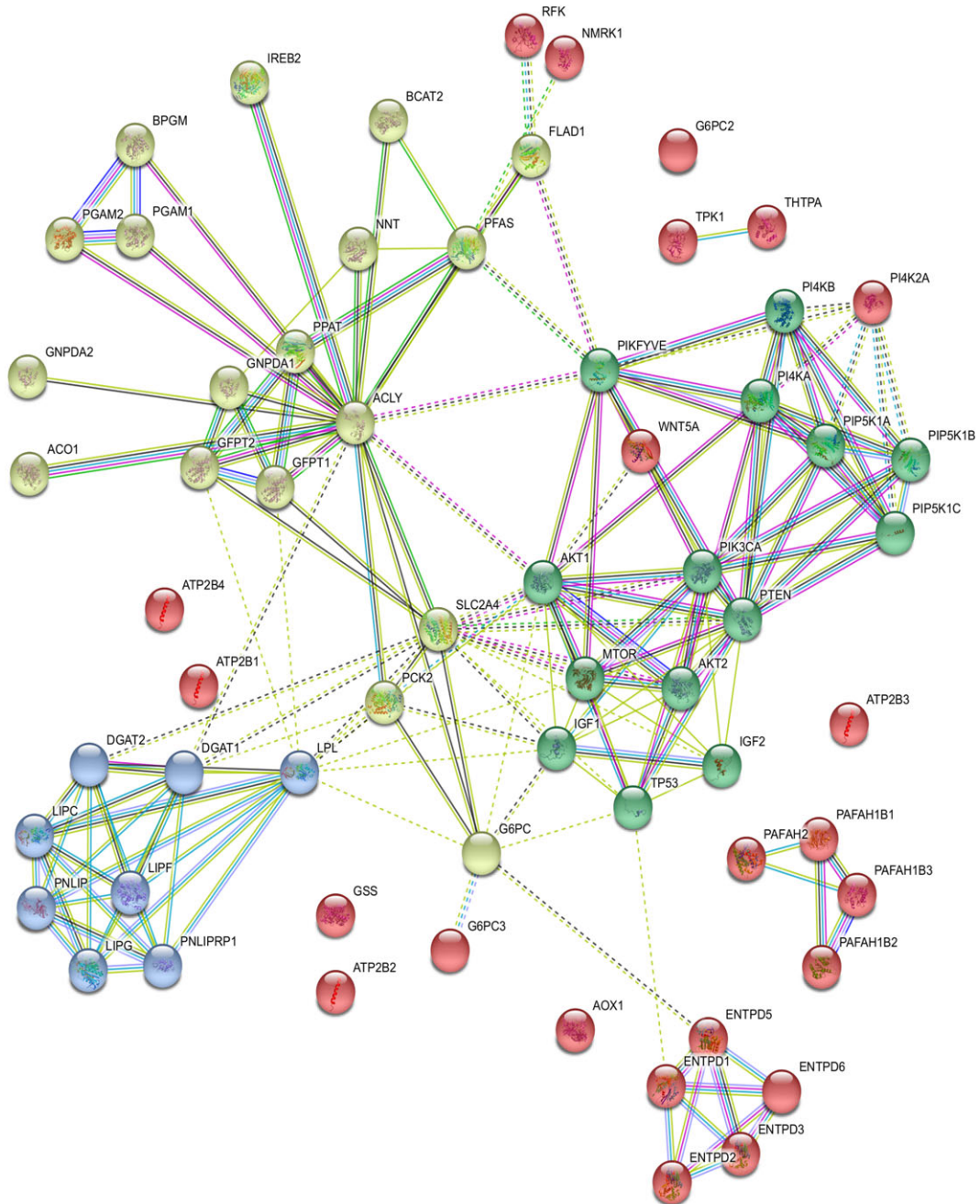


**Figure 6.** Hierarchical clustering of logarithmic fold changes in the choke-point metabolites. Hierarchical clustering of logarithmic fold changes,  $\log_2(r_{\text{deficient}}/r_{\text{normal}})$ , in the choke-point metabolites in lipid metabolism for the template and all one-hit and two-hit deficiencies. Red colour indicates an increase, and green indicates a decrease.

computational results that revealed increased cholesterol levels and decreased sphingolipid levels are consistent with the results for various carcinomas in the literature [49–51].

Genes from the *PIKFYVE*, *PIK3CA* and *PI4KA* families were also detected in inositol phosphate metabolism. In this study, the maximum similarity ratios of  $SR^{WBL} = 0.9286$ ,  $SR^{WHT} = 0.8929$ ,  $SR^{TBL} = 0.7424$  and  $SR^{THT} = 0.783$  were yielded through computation in the dysregulated cases. *PIKFYVE* has been reported to mediate EGFR trafficking to the nucleus in human bladder cancer tissue [52] and regulate lung carcinoma cell migration and invasion through RAC1 activation [44]. The similarity ratios for *PIK3CA* and *PI4KA* were nearly equal to those for *PTEN*. Dysregulation of *PIK3CA* and *PI4KA* is commonly found in HNSCC [46] and hepatocellular carcinoma [43], respectively.

Nine of the top-ranked one-hit oncogenes (table 1) had identical similarity ratios for Warburg effects, except for glucose-6-phosphatase (*G6PC*), which had a slightly higher ratio ( $SR^{WHT} = 0.8408$ ) than the



**Figure 7.** Protein–protein interactions of inferred oncogenes. Protein–protein interactions (PPIs) of the inferred oncogenes with eight additional signalling proteins and their families, namely ACLY, TP53, AKT, IGF, GSK, GLUT, PEPCK and MTOR, *k*-means clustering function in STRING database was applied to classify the PPI network into four classes denoted by four colours. The proteins in the first cluster (green balls) participated in inositol phosphate and glycerophospholipid metabolism, related to PTEN. The proteins in the second cluster (emerald green balls) were involved in triacylglycerol synthesis. The proteins in the third cluster (purple balls) were mainly involved in gluconeogenesis and amino sugar and citric acid cycle metabolisms. The proteins in the fourth cluster (red balls) were involved in other diversified metabolisms.

ratios of the other oncogenes. However, the similarity ratios of the templates were slightly different from those of PTEN. According to a survey of PubMed and cancer databases, each dysregulated enzyme could cause various cancers and diseases, as shown in table 1 and electronic supplementary material, S2.

Evidence indicates that cancer is the result of accumulated genetic mutations, also known as the Knudson hypothesis [53]. Following the same procedures (described in above section), the NHDE algorithm was applied to detect two-hit enzyme deficiencies. A total of 17 of the top-ranked two-hit enzymes yielded higher similarity ratios than those of one-hit enzymes (table 1; electronic supplementary

material, S2). They could achieve the optimal Ave. SRs about 0.86, which were approximately 3% higher than those for the one-hit dysregulation. A survival analysis for the oncogenes can be applied to investigate the clinical significance of the metabolic alterations. However, such investigations need several clinical trials. In this study, we surveyed survival analysis from the HPA to explain survival significance of the inferred oncogenes, and the detailed results were also shown in electronic supplementary material, S3. We observed that five genes (*GNPDA1*, *PGAM1*, *GFPT1*, *RFK* and *PTEN*) are significant for survival.

### 3.3. Flux variability analysis

The FBA in the TLOP problem was employed to evaluate flux alterations for normal and cancer model. However, it is a bias method to compute the flux pattern. Flux variability analysis (FVA) can be applied in a posterior inspection to overcome such a drawback, but required a lot of computation time. We applied the FVA to compute the minimum and maximum quantities of each metabolite for the normal model and the mutants, respectively. The flux intervals of each deficient case were compared with the normal intervals, and classified into seven categories by using the definition presented in the electronic supplementary material, S1. We evaluated the trends of flux and flux-sum synthesis rate for each metabolite of the mutants and those of the template, and then counted the number of metabolites with same flux/flux-sum trends in the flux interval for each mutant and the template to obtain similarity ratios as shown in figure 5. We found that the flux-sum similarity ratios and flux similarity ratios of all mutants are greater than 47% and 65%, respectively, so that the average ratio for each mutant is higher than 56%.

### 3.4. Connection with signal pathway

We also investigated behaviours of lipid metabolism due to one-hit and two-hit deficiencies obtained from table 1. The LFCs of the choke-point metabolites for the templates and mutants are shown in figure 6 and electronic supplementary material, S4. For all dysregulations, we observed that the levels of five cholesterol and four prenol lipids in the endoplasmic reticulum increased, except for 5 $\alpha$ -cholest-7-en-3 $\beta$ -ol (*lthstr1\_r*) dysregulated by the (ATP2B1, NNT) pair. Furthermore, sphingolipid (*crmp\_hs\_c*) and HC01712\_m levels decreased; the same results were obtained for PTEN. In addition, the synthesis rates of all dysregulations of HC01712\_m reduced to zero, which were consistent with those for the templates. The choke-point metabolites could be used as potential biomarker candidates for detecting the metabolic programming between cancer and normal states.

We used the STRING database [54] to investigate protein–protein interactions (PPIs) for the inferred oncogenes obtained from table 1. Eight signalling proteins and their families, namely ACLY, TP53, AKT, IGF, GSK, GLUT, PEPCK and MTOR, were included in the survey and were related to cellular quiescence, proliferation, cancer and longevity in an intracellular signalling pathway. Almost all the proteins were strongly connected, and only a few others independently existed in the PPI network. We used *k*-means clustering to classify the network into four clusters (see figure 7 and electronic supplementary material S5 for details). The proteins in the first cluster (green balls) participated in inositol phosphate and glycerophospholipid metabolisms and were linked with the TP53/AKT/IGF/MTOR signalling pathway through PTEN. The proteins in the second cluster (emerald green balls) were involved in triacylglycerol synthesis and linked with PIKFYVE and AKT1 in the first cluster through ACLY. The proteins in the third cluster (purple balls) were mainly involved in gluconeogenesis and amino sugar and citric acid cycle metabolisms. It also connected with the second cluster. The proteins in the last cluster (red balls) were involved in other diversified metabolisms.

## 4. Conclusion

A tri-level inference optimization framework was applied to detect 13 one-hit and 17 two-hit deficient enzymes that contribute to inducing HNSCC. According to the PubMed survey, each detected one-hit deficiency participated in carcinogenesis in various tissues. The inferred oncogene *PTEN* is one of the most frequently mutated tumour suppressor genes in human cancer. In recent years, the function of PTEN as a metabolic regulator has attracted considerable research attention. This study also supported this function. From the literature survey, we found that *TPK1* is an oncogene for colon cancer, and three genes (*PTEN*, *GNPDA1* and *PIK3CA*) appear in head and neck cancer. Deficiency of G6PC leads to accelerated hepatic carcinogenesis in glycogen storage disease. In addition, G6PC was involved in 11 pairs of two-hit oncogenes, as revealed by the computation in this study. Moreover, we

will conduct wet-laboratory experiment to verify the inferred results in order to discover potential cancer cell metabolic targets.

**Data accessibility.** The datasets supporting this article have been uploaded as part of the electronic supplementary material. The optimization code used during this study is available via the Dryad Digital Repository <https://doi.org/10.5061/dryad.364vk23> (doi:10.5061/dryad.364vk23) [55].

**Authors' contributions.** W.-H.W. wrote the modelling software and revised the manuscript; F.-Y.L. and Y.-C.S. collected field data and carried out the analyses and paper survey; J.-M.L., P.M.-H.C. and C.-Y.F.H. participated in the design of the study and data analysis; F.-S.W. conceived of the study, designed the study, coordinated the study and drafted the manuscript. All authors gave final approval for publication and agree to be held accountable for the work performed therein.

**Competing interests.** The authors declare that they have no competing interests.

**Funding.** This work was supported by Ministry of Science and Technology of Taiwan (MOST) (grant nos. MOST106-2221-E-194-049-MY3 and MOST107-2627-M-194-001 to F.-S.W., MOST107-2627-M-030-001 to J.-M.L., MOST107-2627-M-075-001 to P.M.-H.C., MOST107-2320-B-010-040-MY3 to C.-Y.F.H.).

## References

- Cairns RA, Harris IS, Mak TW. 2011 Regulation of cancer cell metabolism. *Nat. Rev. Cancer* **11**, 85–95. (doi:10.1038/nrc2981)
- Hsu PP, Sabatini DM. 2008 Cancer cell metabolism: Warburg and beyond. *Cell* **134**, 703–707. (doi:10.1016/j.cell.2008.08.021)
- Tomita M, Kami K. 2012 Systems biology, metabolomics, and cancer metabolism. *Science* **336**, 990–991. (doi:10.1126/science.1223066)
- Ward PS, Thompson CB. 2012 Metabolic reprogramming: a cancer hallmark even Warburg did not anticipate. *Cancer Cell* **21**, 297–308. (doi:10.1016/j.ccr.2012.02.014)
- Eyassu F, Angione C. 2017 Modelling pyruvate dehydrogenase under hypoxia and its role in cancer metabolism. *R. Soc. open sci.* **4**, 170360–170360. (doi:10.1098/rsos.170360)
- Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T. 2011 Predicting selective drug targets in cancer through metabolic networks. *Mol. Syst. Biol.* **7**, 501. (doi:10.1038/msb.2011.35)
- Gottstein W, Olivier BG, Bruggeman FJ, Teusink B. 2016 Constraint-based stoichiometric modelling from single organisms to microbial communities. *J. R. Soc. Interface* **13**, 20160627. (doi:10.1098/rsif.2016.0627)
- Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J. 2014 Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease. *Nat. Commun.* **5**, 3083. (doi:10.1038/ncomms4083)
- Sean A, Andrew S, Mitali ST, Soyer Orkun S. 2019 Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses. *J. R. Soc. Interface* **15**, 20180125. (doi:10.1098/rsif.2018.0125)
- Wang Y, Eddy JA, Price ND. 2012 Reconstruction of genome-scale metabolic models for 126 human tissues using mCADRE. *BMC Syst. Biol.* **6**, 153. (doi:10.1186/1752-0509-6-153)
- Yizhak K, Chaneton B, Gottlieb E, Ruppin E. 2015 Modeling cancer metabolism on a genome scale. *Mol. Syst. Biol.* **11**, 817. (doi:10.15252/msb.20145307)
- Schultz A, Qutub AA. 2016 Reconstruction of tissue-specific metabolic networks using CORDA. *PLoS Comput. Biol.* **12**, 1–33. (doi:10.1371/journal.pcbi.1004808)
- Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, Vo TD, Srivas R, Palsson BO. 2007 Global reconstruction of the human metabolic network based on genomic and bibliomic data. *Proc. Natl Acad. Sci. USA* **104**, 1777–1782. (doi:10.1073/pnas.0610772104)
- Thiele I *et al.* 2013 A community-driven global reconstruction of human metabolism. *Nat. Biotechnol.* **31**, 419–425. (doi:10.1038/nbt.2488)
- Mardinoglu A *et al.* 2013 Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol. Syst. Biol.* **9**, 649. (doi:10.1038/msb.2013.5)
- Brunk E *et al.* 2018 Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* **36**, 272–281. (doi:10.1038/nbt.4072)
- Swainston N *et al.* 2016 Recon 2.2: from reconstruction to model of human metabolism. *Metabolomics* **12**, 109. (doi:10.1007/s11306-016-1051-4)
- Richelle A, Chiang AWT, Kuo CC, Lewis NE. 2019 Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions. *PLoS Comput. Biol.* **15**, e1006867. (doi:10.1371/journal.pcbi.1006867)
- Blais EM, Rawls KD, Dougherty BV, Li ZI, Kolling GL, Ye P, Wallqvist A, Papin JA. 2017 Reconciled rat and human metabolic networks for comparative toxicogenomics and biomarker predictions. *Nat. Commun.* **8**, 14250. (doi:10.1038/ncomms14250)
- Ryu JY, Kim HU, 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. (doi:10.1073/pnas.1713050114)
- Wu HQ *et al.* 2017 Flux balance analysis predicts Warburg-like effects of mouse hepatocyte deficient in miR-122a. *PLoS Comput. Biol.* **13**, 1–22. (doi:10.1371/journal.pcbi.1005618)
- Pontén F, Jirstrom K, Uhlen M. 2008 The human protein atlas—a tool for pathology. *J. Pathol.* **216**, 387–393. (doi:10.1002/path.2440)
- Angione C. 2017 Integrating splice-isoform expression into genome-scale models characterizes breast cancer metabolism. *Bioinformatics* **34**, 494–501. (doi:10.1093/bioinformatics/btx562)
- Nam H, Campodonico M, Bordbar A, Hyduke DR, Kim S, Zielinski DC, Palsson BO. 2014 A systems approach to predict oncometabolites via context-specific genome-scale metabolic networks. *PLoS Comput. Biol.* **10**, 1–13. (doi:10.1371/journal.pcbi.1003837)
- Chiou JP, Wang FS. 1999 Hybrid method of evolutionary algorithms for static and dynamic optimization problems with application to a fed-batch fermentation process. *Comput. Chem. Eng.* **23**, 1277–1291. (doi:10.1016/S0098-1354(99)00290-2)
- Storn R, Price K. 1997 Differential evolution – a simple and efficient heuristic for global optimization over continuous spaces. *J. Global Optim.* **11**, 341–359. (doi:10.1023/A:1008202821328)
- Wang FS, Chiou JP. 1997 Optimal control and optimal time location problems of differential-algebraic systems by differential evolution. *Ind. Eng. Chem. Res.* **36**, 5348–5357. (doi:10.1021/ie9702486)
- Wang FS, Wu WH. 2015 Optimal design of growth-coupled production strains using nested hybrid differential evolution. *J. Taiwan Inst. Chem. Eng.* **54**, 57–63. (doi:10.1016/j.jtice.2015.03.015)
- Hsu KC, Wang FS. 2017 Detection of minimum biomarker features via bi-level optimization framework by nested hybrid differential evolution. *J. Taiwan Inst. Chem. Eng.* **81**, 31–39. (doi:10.1016/j.jtice.2017.10.015)
- Wu WH, Chien CY, Wu YH, Wu HH, Lai JM, Chang PMH, Huang CY, Wang FS. 2018 Inferring oncoenzymes in a genome-scale metabolic network for hepatocytes using bilevel optimization framework. *J. Taiwan Inst. Chem. Eng.* **91**, 97–104. (doi:10.1016/j.jtice.2018.06.002)
- Kim S *et al.* 2015 Bioinformatic and metabolomic analysis reveals miR-155 regulates thiamine level in breast cancer. *Cancer Lett.* **357**, 488–497. (doi:10.1016/j.canlet.2014.11.058)

32. Tiwana GS *et al.* 2015 Identification of vitamin B1 metabolism as a tumor-specific radiosensitizing pathway using a high-throughput colony formation screen. *Oncotarget* **6**, 5978–5989. (doi:10.18632/oncotarget.3468)
33. Cho JH, Kim GY, Mansfield BC, Chou JY. 2018 Hepatic glucose-6-phosphatase- $\alpha$  deficiency leads to metabolic reprogramming in glycogen storage disease type Ia. *Biochem. Biophys. Res. Commun.* **498**, 925–931. (doi:10.1016/j.bbrc.2018.03.083)
34. Oikari S, Makkonen K, Deen AJ, Tyni I, Karna R, Tammi RH, Tammi MI. 2016 Hexosamine biosynthesis in keratinocytes: roles of GFAT and GNPDA enzymes in the maintenance of UDP-GlcNAc content and hyaluronan synthesis. *Glycobiology* **26**, 710–722. (doi:10.1093/glycob/cww019)
35. Shen S *et al.* 2017 A seven-gene prognostic signature for rapid determination of head and neck squamous cell carcinoma survival. *Oncol. Rep.* **38**, 3403–3411. (doi:10.3892/or.2017.6057)
36. Yang C *et al.* 2016 High expression of GFAT1 predicts poor prognosis in patients with pancreatic cancer. *Sci. Rep.* **6**, 39044. (doi:10.1038/srep39044)
37. Yazdanpanah B *et al.* 2009 Riboflavin kinase couples TNF receptor 1 to NADPH oxidase. *Nature* **460**, 1159–1163. (doi:10.1038/nature08206)
38. Kovacevic Z, Fu D, Richardson DR. 2008 The iron-regulated metastasis suppressor, Ndrq-1: identification of novel molecular targets. *Biochim. Biophys. Acta (BBA) – Mol. Cell Res.* **1783**, 1981–1992. (doi:10.1016/j.bbamcr.2008.05.016)
39. Fang BA, Kovačević Ž, Park KC, Kalinowski DS, Jansson PJ, Lane DJ, Sahni S, Richardson DR. 2014 Molecular functions of the iron-regulated metastasis suppressor, NDRG1, and its potential as a molecular target for cancer therapy. *Biochim. Biophys. Acta* **1845**, 1–19. (doi:10.1016/j.bbcan.2013.11.002)
40. Liu Y, Hyde AS, Simpson MA, Barycki JJ. 2014 Emerging regulatory paradigms in glutathione metabolism. *Adv. Cancer Res.* **122**, 69–101. (doi:10.1016/B978-0-12-420117-0.00002-5)
41. Li J *et al.* 1997 PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science* **275**, 1943–1947. (doi:10.1126/science.275.5308.1943)
42. Steck PA *et al.* 1997 Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nat. Genet.* **15**, 356–362. (doi:10.1038/ng0497-356)
43. Ilboudo A, Nault JC, Dubois-Pot-Schneider H, Corlu A, Zucman-Rossi J, Samson M, Le Seyec J. 2014 Overexpression of phosphatidylinositol 4-kinase type III $\alpha$  is associated with undifferentiated status and poor prognosis of human hepatocellular carcinoma. *BMC Cancer* **14**, 7. (doi:10.1186/1471-2407-14-7)
44. Oppelt A, Haugsten EM, Zech T, Danielsen HE, Sveen A, Lobert VH, Skotheim RI, Wesche J. 2014 PIKfyve, MTMR3 and their product PtdIns5P regulate cancer cell migration and invasion through activation of Rac1. *Biochem. J.* **461**, 383–390. (doi:10.1042/BJ20140132)
45. Semenas J *et al.* 2014 The role of PI3K/AKT-related PIP5K1 $\alpha$  and the discovery of its selective inhibitor for treatment of advanced prostate cancer. *Proc. Natl Acad. Sci. USA* **111**, E3689–E3698. (doi:10.1073/pnas.1405801111)
46. Kidacki M, Lehman HL, Green MV, Warrick JJ, Stairs DB. 2017 p120-catenin downregulation and PIK3CA mutations cooperate to induce invasion through MMP1 in HNSCC. *Mol. Cancer Res.* **15**, 1398–1409. (doi:10.1158/1541-7786.MCR-17-0108)
47. Møller M *et al.* 2017 Heterogeneous patterns of DNA methylation-based field effects in histologically normal prostate tissue from cancer patients. *Sci. Rep.* **7**, 40636. (doi:10.1038/srep40636)
48. Khuroya H *et al.* 2017 IRP2 as a potential modulator of cell proliferation, apoptosis and prognosis in nonsmall cell lung cancer. *Eur. Respir. J.* **49**, 1600711. (doi:10.1183/13993003.00711-2016)
49. Coant N, Sakamoto W, Mao C, Hannun YA. 2017 Ceramidases, roles in sphingolipid metabolism and in health and disease. *Adv. Biol. Regul.* **63**, 122–131. (doi:10.1016/j.jbior.2016.10.002)
50. Kuzu OF, Noory MA, Robertson GP. 2016 The role of cholesterol in cancer. *Cancer Res.* **76**, 2063–2070. (doi:10.1158/0008-5472.CAN-15-2613)
51. Lee WK, Kolesnick RN. 2017 Sphingolipid abnormalities in cancer multidrug resistance: chicken or egg? *Cell Signalling* **38**, 134–145. (doi:10.1016/j.cellsig.2017.06.017)
52. Kim J, Jahng WJ, Di Vizio D, Lee JS, Jhaveri R, Rubin MA, Shisheva A, Freeman MR. 2007 The phosphoinositide kinase PIKfyve mediates epidermal growth factor receptor trafficking to the nucleus. *Cancer Res.* **67**, 9229–9237. (doi:10.1158/0008-5472.CAN-07-1333)
53. Knudson AG. 2001 Two genetic hits (more or less) to cancer. *Nat. Rev. Cancer* **1**, 157–162. (doi:10.1038/35101031)
54. Szklarczyk D *et al.* 2017 The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. *Nucleic Acids Res.* **45**, D362–D368. (doi:10.1093/nar/gkw937)
55. Wu W-H, Li F-Y, Shu Y-C, Lai J-M, Chang PM-H, Huang C-YF, Wang F-S. 2020 Data from: Oncogene inference optimization using constraint-based modelling incorporated with protein expression in normal and tumour tissues. Dryad Digital Repository. (doi:10.5061/dryad.364vk23)