Identifying Essential Genes in Genome-Scale Metabolic Models of Consensus Molecular Subtypes of Colorectal Cancer

Chao-Ting Cheng¹, Jin-Mei Lai², Peter Mu-Hsin Chang³,⁴, Yi-Ren Hong⁵, Chi-Ying F. Huang⁶,⁷ and Feng-Sheng Wang¹*

¹ Department of Chemical Engineering, National Chung Cheng University, Chiayi, Taiwan; easonchao1008@gmail.com (C.-T.C.); chmfsw@ccu.edu.tw (F.-S.W.)
² Department of Life Science, College of Science and Engineering, Fu Jen Catholic University, New Taipei City 24205, Taiwan; jmlai@mail.fju.edu.tw
³ Department of Oncology, Taipei Veterans General Hospital, Taipei, Taiwan; ptchang@vghtpe.gov.tw
⁴ Faculty of Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan; ptchang@vghtpe.gov.tw
⁵ Department of Biochemistry and Graduate Institute of Medicine, Kaohsiung Medical University, Kaohsiung City, Taiwan; m835016@cc.kmu.edu.tw
⁶ Institute of Biopharmaceutical Sciences, National Yang Ming Chiao Tung University, Taipei, Taiwan; cyhuang5@ym.edu.tw
⁷ Department of Biotechnology and Laboratory Science in Medicine, National Yang Ming Chiao Tung University, Taipei, Taiwan; cyhuang5@ym.edu.tw

*Corresponding author: E-mail: chmfsw@ccu.edu.tw

Short title: Essential genes of Consensus Molecular Subtypes of Colorectal Cancer
Abstract

Identifying essential targets in genome-scale metabolic networks of cancer cells is a time-consuming process. This study proposed a fuzzy hierarchical optimization framework for identifying essential genes, metabolites and reactions. On the basis of four objectives, the framework can identify essential targets that lead to cancer cell death, and evaluate metabolic flux perturbations of normal cells due to treatment. Through fuzzy set theory, a multiobjective optimization problem was converted into a trilevel maximizing decision-making (MDM) problem. We applied nested hybrid differential evolution to solve the trilevel MDM problem to identify essential targets in the genome-scale metabolic models of five consensus molecular subtypes (CMSs) of colorectal cancers. We used various media to identify essential targets for each CMS, and discovered that most targets affected all five CMSs and that some genes belonged to a CMS-specific model. We used the experimental data for the lethality of cancer cell lines from the DepMap database to validate the identified essential genes. The results reveal that most of the identified essential genes were compatible to colorectal cancer cell lines from DepMap and that these genes could engender a high percentage of cell death when knocked out, except for EBP, LSS and SLC7A6. The identified essential genes were mostly involved in cholesterol biosynthesis, nucleotide metabolisms, and the glycerophospholipid biosynthetic pathway. The genes in the cholesterol biosynthetic pathway were also revealed to be determinable, if the medium used excluded a cholesterol uptake reaction. By contrast, the genes in the cholesterol biosynthetic pathway were non-essential, if a cholesterol uptake reaction was involved in the medium used. Furthermore, the essential gene CRLS1 was revealed as a medium-independent target for all CMSs irrespective of whether a medium involves a cholesterol uptake reaction.
Keywords: Cancer metabolism, genome-scale metabolic model, constraint-based model, multi-level optimization, fuzzy optimization, hybrid differential evolution

Author summary

Essential genes are indispensable genes for cells to grow and proliferate under certain physiological condition. Identifying essential genes in genome-scale metabolic networks of cancer cells is a time-consuming process. We develop an anticancer target discovery platform for identifying essential genes that conduct cell death when the genes of cancer cells are deleted. Meanwhile, the essential genes are also inactive on their healthy cells to maintain their cell viability and smaller metabolic alterations. We use fuzzy set theory to measure metabolic deviation of the perturbation of normal cells relative to healthy and cancer templates towards predicting side effects for treatment of each identified gene. The platform can identify essential genes, metabolites and reactions for treating five consensus molecular subtypes (CMS) of colorectal cancers with using various media. We discovered that most targets affected all five CMSs and that some genes belonged to a CMS-specific model. We found that the genes in the cholesterol biosynthetic pathway are nonessential for the cells that be compensated by a cholesterol uptake reaction from a medium. Furthermore, CRLS1 was revealed as an essential gene for all CMS colorectal cancer in a medium-independent manner that is unrelated to a cholesterol uptake reaction.
1. Introduction

Colorectal cancer (CRC) is a worldwide health burden, it is the third- and second-ranked cancer in terms of incidence and mortality, respectively [1], indicating that the world requires improved prognoses and treatment strategies. More than 1.9 million new CRC (including anal cancer) cases and 935000 deaths were estimated to have occurred in 2020; they represent approximately one in 10 cancer cases and deaths [1]. The classification of CRC plays a pivotal role in predicting a patient’s prognosis and determining treatment strategies. The tumor, node, and metastasis (commonly abbreviated as TNM) classification system is commonly used to determine the progression of CRC; however, an in-depth characterization is necessary to improve the assessment of treatment strategies and prognoses. The consensus molecular subtype (CMS) is an RNA expression-based classification system for CRC; it was developed from 18 CRC data sets and contains 4151 CRC samples as of 2015 [2]. CRC is classified into four subtypes that each exhibit distinct molecular and biological characteristics and pathological and genetic signatures. Molecular subtype-based therapies provide a new framework for implementing preferred and precise medical treatments. Several studies have used CMS classification to predict the prognosis of a patient with CRC and determine treatment strategies for the patient [3-9].

Tissue-specific genome-scale metabolic models (GSMM) are applied to identify anticancer targets and provide detailed insights into the metabolic bases of physiological and pathological processes [10-30]. The Cancer Genome Atlas (TCGA) [31] and Human Protein Atlas (HPA) [32] databases are incorporated into human metabolic networks such as Recon X [32-37] and Human-GEM [24, 38], to reconstruct tissue-specific GSMMs for investigating metabolic behaviors. Both the TCGA and HPA databases have not used to reconstruct CMS-based tissue-specific GSMMs.
Furthermore, a literature review revealed that no study has attempted to identify targets for CRC by integrating CMS classification with genome-scale metabolic modeling. The present study used the CRC data set from TCGA to classify patient samples into five CMSs and then incorporated with Recon3D to reconstruct the CMS-specific GSMMs of CRC. We applied these CMSs to identify essential genes, metabolites, and reactions and used a fuzzy decision-making method to evaluate cancer cell mortality, healthy cell viability, and metabolic perturbation effects due to the blockage of the corresponding fluxes.

The present study proposed a fuzzy hierarchical optimization framework for identifying essential genes for treating each CMS of CRC. This framework represents an extension of the Identifying AntiCancer Targets (IACT) framework [25, 26]; it accounts for RNA-sequence expression information in inner optimization problems to yield uniform flux patterns for treated and perturbed cells. Traditional cell culture media (e.g., such as Dulbecco’s Modified Eagle Medium [DMEM] and HAM’s medium [HAM]), were designed to ensure continuous cancer cell proliferation in vitro. However, their composition do not necessarily summarize the nutritional requirements of tumor cell cultures. Studies have demonstrated that the use of specific medium components can produce cell culture results that are highly relevant to tumor metabolism [39-41]. The present study also considered various media as uptake reactions to investigate relationships of tumor cell growth with nutritional components and essential targets for CMSs of CRC.

2. Materials and Methods

An essential gene is an indispensable element in an organism that is essential to the survival of an organism. Because an essential gene is required for a cancer cell to
grow, proliferate and survive, deleting it from a cancer cell leads to the death of this cell or causes a severe proliferation defect; by contrast, healthy cells are unlikely to be affected by its deletion. The present study developed an anticancer target discovery (ACTD) platform for identifying essential genes that can be deleted to cause the death of a cancer cell while allowing its healthy counterparts to survive with few side effects. Figure 1 illustrates the work flowchart of the ACTD platform, which mimics a wet-lab experiment that is designed for identifying essential genes.

**Figure 1.** Framework for identifying essential anticancer genes. (A) Retrieval of biological data such as RNA-seq expression for cancerous cells (CA) and healthy cells (HT), and human genome-scale metabolic networks. (B) Statistical analysis of accessed RNA-seq expression data to yield high, medium, low and not detect enzyme encoding genes. (C) Reconstruction of tissue-specific genome-scale metabolic models and gene-protein-reaction models for each CMS and HT, respectively. (D) Flux distribution patterns for each CMS are obtained from clinical data (if available); otherwise a CA template can be computed through flux balance analysis (FBA) and uniform flux distribution (UFD) problems without considering dysregulated restriction. (E) Flux distribution patterns of HT cells can be obtained from clinical data (if available); otherwise, a HT template can be computed through FBA and UFD problems without considering dysregulated restriction. (F) A set of candidate genes are identified by the nest hybrid differential evolution (NHDE) algorithm and used in FBA and UFD models of CMS to compute fluxes for treatment. (G) Flux distribution and metabolite flow rates for each candidate treatment are obtained. (H) Identical genes are used for
FBA and UFD models of HT cells during treatment. (I) Flux distribution and metabolite flow rates of perturbed HT cells for each candidate treatment are obtained. (J) Fuzzy membership functions for each fuzzy objective are defined to enable the evaluation of decision criterion. (K) A multiobjective objective optimization problem is converted into a maximizing decision-making problem to be solved by the NHDE algorithm. (L) Optimal essential genes are identified on the basis of decision criterion; otherwise, Steps (F) to (L) are repeated using the next set of candidate genes generated by NHDE algorithm.

2.1 Tissue-Specific Genome-Scale Metabolic Models

The present study combined a human metabolic network Recon3D [37] with RNA-seq expression data obtained from the TCGA database [31] to reconstruct CMS-based tissue-specific GSMMs for CRC and its healthy counterpart tissue (Figure 1A-1C). The RNA-seq expression data of 51 healthy colorectal samples (41 colon and 10 rectum samples) were downloaded from the TCGA database. A total 478 colon adenocarcinoma and 166 rectum adenocarcinoma samples were obtained as cancer samples. On the basis of CMS classification [2], the cancer samples were categorized into four subtypes that each exhibit distinct molecular and biological characteristics, namely CMS1 (microsatellite instability immune), CMS2 (canonical), CMS3 (metabolic), and CMS4 (mesenchymal). In addition, samples that could not be classified into one of the four aforementioned subtypes were categorized under CMS5 (unknown). Table 1 lists the number of accessed CRC samples that are categorized under each CMS group.

Table 1. Classification of colorectal cancer samples accessed from The Cancer Genome Atlas database by consensus molecular subtype (CMS). CMS classification is based on literature [2]. A sample that cannot be classified under one of the four subtypes is classified as CMS5.

|       | CMS1 | CMS2 | CMS3 | CMS4 | CMS5 |
|-------|------|------|------|------|------|
| No. Samples | 83   | 226  | 72   | 151  | 112  |
RNA-seq expressions (Table 1) were used to reconstruct CMS-specific GSMMs and healthy/basal (HT) tissue. Quantile normalization was applied to normalize the raw data from healthy and CMS samples to compute the mean, confidence interval, and coefficient of dispersion for each gene. The data were then used to evaluate supportive genes and to obtain a high differential expression of enzyme-encoding genes between each CMS and its healthy cell counterpart. Recon3D comprised 2247 enzyme-encoding genes, which were classified into four levels on the basis of their participation level (i.e., high, medium, low, and not detected). Four groups of confidence reactions (i.e., high, medium, negative, and others) were obtained through gene-protein-reaction association in Recon3D. These four groups of confidence reactions were used in the CORDA algorithm [42] to reconstruct each CMS-specific GSMM and its healthy cell counterpart.

Genome-scale reconstructions provide a mechanistic link between genotypes and phenotypes through stoichiometric models for metabolites and reactions. Gene-protein-reaction (GPR) associations are typically implemented as Boolean rules, which link the metabolic reactions in stoichiometric models to the gene-encoded enzymes in cells. A reaction can be catalyzed by one enzyme or isozymes. Moreover, reactions may be regulated by redundant enzymes, that is, more than two enzymes can catalyze a reaction. The redundant enzymes in a GSMM can be deleted to yield a reduced GPR association, thereby avoiding surplus computations in evolutionary optimization procedures. We developed a systems biology program that automatically builds stoichiometric and reduced GPR models in the files of the General Algebraic Modeling System (GAMS, https://www.gams.com/) for computation. The detailed procedures are discussed in another study [25].
2.2 Discovery of Essential Genes

The present study presents an ACTD framework to mimic a wet-lab experiment to identify essential genes (Figure 1D-1M). The framework was formulated as a fuzzy hierarchical optimization problem (Table 2).

Table 2. Optimization platform for anticancer target discovery design to evaluate performance of identified essential genes on basis of four fuzzy objectives.

| Objectives for the outer optimization problem |
|------------------------------------------------|
| 1. The first objective is to measure the mortality of cancer cells following gene knockout or deletion. |
| 2. The second objective is to maximize the cell viability of perturbed normal cells during gene knockout or deletion. |
| 3. The third objective is to measure the metabolic deviation of the perturbation of normal cells as a large dissimilarity relative to the cancer template. |
| 4. The fourth objective is to measure the metabolic deviation of perturbation as a close similarity relative to the normal template. |

subject to constraint-based models in inner optimization problems

1. FBA and UFD problems for treating cancer cells
2. FBA and UFD problems for perturbation of normal cells due to treatment

The ACTD platform was directly used to compute fluxes and metabolite flow rates for evaluating the fuzzy objectives. This method differs to that of the IACT framework, which examines logarithmic fold changes between dysregulated and template levels. An evaluation performed using the IACT framework may produce numerical inaccuracies if an evaluated value is close to zero. Furthermore, in the ACTD platform, the goal targeting cancerous (CA) cell mortality is equivalent to the minimization of both cell growth rate and adenosine triphosphate (ATP) production rate. The four goals are as follows:

The first goal is to evaluate fuzzy minimization of the cell growth rate and ATP production rate of treated CA cells as the following equations:
The second goal is to evaluate fuzzy minimization of the cell growth rate and fuzzy maximization of the ATP production rate of perturbed HT cells as the following equations:

\[
\begin{align*}
\min_z \theta_{b}^{TB} \\ \min_z \theta_{b}^{ATP} \\ \min_z \theta_{b}^{PB} \\ \max_z \theta_{b}^{ATP} \\ \max_z \theta_{b}^{PB} 
\end{align*}
\]  

The third goal is to measure fuzzy dissimilarity of fluxes and metabolite flow rates of perturbed HT cells relative to the CA template as the following equations:

\[
\begin{align*}
\text{dissimilarity}_{z} v_{b}^{PB} & \not\approx v_{b}^{CA} \\ \text{dissimilarity}_{z} r_{b}^{PB} & \not\approx r_{b}^{CA} 
\end{align*}
\]  

The fourth goal is to measure fuzzy similarity of fluxes and metabolite flow rates of perturbed HT cells relative to the HT template as the following equations:

\[
\begin{align*}
\text{similarity}_{z} v_{b}^{PB} & \approx v_{b}^{HT} \\ \text{similarity}_{z} r_{b}^{PB} & \approx r_{b}^{HT} 
\end{align*}
\]  

In the aforementioned equations, the decision variable $z$ represents the gene encoding enzymes as determined by a nest hybrid differential evolution (NHDE) algorithm for knockout. Fuzzy minimization (i.e., $\min$) is used to evaluate the minimum cell growth rate and ATP production rate of treated CA cells. By contrast, fuzzy maximization (i.e., $\max$) is used to measure the maximum ATP production rate of perturbed HT cells. Fuzzy dissimilarity (i.e., $\text{dissimilarity}$) is used to evaluate the disparity of the fluxes (i.e., $v_{b}^{PB}$) and metabolite flow rates (i.e., $r_{b}^{PB}$) for the perturbed HT cells relative to those of the CA template. A substantial disparity means that the flux changes of the
perturbed cells are considerably different from those of CA template, indicating that the perturbation of HT cells cannot lead to tumorigenesis during treatment. Fuzzy similarity (i.e., \( \text{similarity} \)) is used to evaluate the metabolic deviation between the perturbed HT cells and their template. The flow rate of the \( m \)th metabolite is computed using the following equations:

\[
r_m = \sum_{i \in \Omega} \left( \sum_{N_y > 0, j} N_y v_{f, j} - \sum_{N_y < 0, j} N_y v_{b, j} \right), \quad m \in \Omega^{a}
\]

where \( \Omega^{c} \) is the set of species located in various compartments of the cell, and \( N_y \) is a stoichiometric coefficient of the \( j \)th metabolite in the \( j \)th reaction of a GSMM. The forward flux \( v_{f, j} \) and backward flux \( v_{b, j} \) of the \( j \)th reaction are calculated by applying FBA and UFD models in the inner optimization problem as follows:

Treated CMS-specific model:

FBA problem:

\[
\begin{align*}
\max_{v_{f/b}} v_{\text{biomass}} \\
\text{subject to} \\
N^{CA} (v_f - v_b) = 0 \\
v_{f/b,d} = 0, \quad i \in \Omega^{KO} \\
v_{LB}^{f/b,j} \leq v_{f/b,j} \leq v_{UB}^{f/b,j}, \quad j \not\in \Omega^{KO} \\
\end{align*}
\]

UFD problem:

\[
\begin{align*}
\min_{v_{f/b}} \sum_{k \in \Omega^{CA}} C_k^{CA} \left( (v_{f,k})^2 + (v_{b,k})^2 \right) \\
\text{subject to} \\
N^{CA} (v_f - v_b) = 0 \\
v_{f/b,d} = 0, \quad i \in \Omega^{KO} \\
v_{LB}^{f/b,j} \leq v_{f/b,j} \leq v_{UB}^{f/b,j}, \quad j \not\in \Omega^{KO} \\
v_{\text{biomass}} \geq v_{\text{biomass, CA}}^* \\
\end{align*}
\]

Perturbed HT model:

FBA problem:

\[
\begin{align*}
\max_{v_{f/b}} v_{\text{ATP}} \\
\text{subject to} \\
N^{HT} (v_f - v_b) = 0 \\
v_{f/b,d} = 0, \quad i \in \Omega^{KO} \\
v_{LB}^{f/b,j} \leq v_{f/b,j} \leq v_{UB}^{f/b,j}, \quad j \not\in \Omega^{KO} \\
\end{align*}
\]

UFD problem:

\[
\begin{align*}
\min_{v_{f/b}} \sum_{k \in \Omega^{HT}} C_k^{HT} \left( (v_{f,k})^2 + (v_{b,k})^2 \right) \\
\text{subject to} \\
N^{HT} (v_f - v_b) = 0 \\
v_{f/b,d} = 0, \quad i \in \Omega^{KO} \\
v_{LB}^{f/b,j} \leq v_{f/b,j} \leq v_{UB}^{f/b,j}, \quad j \not\in \Omega^{KO} \\
v_{\text{ATP}} \geq v_{\text{ATP,HT}}^* \\
\end{align*}
\]
In the aforementioned equations, $N_{CA}$ and $N_{HT}$ are the stoichiometric matrices for each CMS-specific model and HT model, respectively, and they are reconstructed from the models presented in the previous subsection; $v^{LB}_{f/b,j}$ and $v^{UB}_{f/b,j}$ are the positive lower bound (LB) and positive upper bound (UB) of the $j^{th}$ forward flux and $j^{th}$ backward flux, respectively. The RNA-seq expressions for cancer and normal cells and GPR associations in Recon3D are not only used to reconstruct GSMMs but also to set the weighting factors $c^CA_k$ and $c^HT_k$ for UFD problems; the four groups of confidence reactions are assigned as follows:

$$c^{CA/HT}_k = \begin{cases} 
\frac{1}{4}, & k \in \text{high confidence} \\
\frac{1}{2}, & k \in \text{medium confidence} \\
\frac{3}{4}, & k \in \text{negative confidence} \\
1, & k \in \text{other confidence or non-gene-expression}
\end{cases}$$  \hspace{1cm} (7)

For a high confidence reaction, the smallest weighting factor is set to obtain a higher flux value from the UFD problem. Forward and backward fluxes are set to be zero if their corresponding gene encoding enzymes are knocked out. A reaction may be catalyzed by isozymes, indicating that it is still active. GPR association is used to set up the knockout reactions as follows:

$$v_{f,j} = v_{b,j} = 0; z_i \in \Omega^{KO}\setminus\Omega^{IZ}$$

$$\begin{cases} 
(1-\varepsilon)v_{f,j}^{basal} \leq v_{f,j} \leq (1+\varepsilon)v_{f,j}^{basal} \\
(1-\varepsilon)v_{b,j}^{basal} \leq v_{b,j} \leq (1+\varepsilon)v_{b,j}^{basal} ; z_i \in \Omega^{KO} \cap \Omega^{IZ}
\end{cases}$$  \hspace{1cm} (8)

where $\Omega^{IZ}$ is a set of reactions regulated by isozymes that are represented in the GPR model.

2.3 Maximizing Decision-Making Problem
The ACTD problem expressed in Eq. (1) to (8) is a hierarchical multiobjective optimization problem (MOOP). Numerous methods can be applied for solving hierarchical MOOPs and obtain a Pareto optimal solution [25, 43, 44]. These methods are classified into two categories, namely generating methods and preference-based methods, and each method has its advantages and disadvantages [43, 44]. Generating methods apply a scalarization approach to convert a hierarchical MOOP into a single-objective optimization problem with multiple weighting factors to identify a Pareto optimal solution. By contrast, preference-based methods require a decision maker to indicate their preferences in advance before they can identify a satisfactory solution. The present study used CA and HT templates (Figure 1D and 1E) as preferences to convert an ACTD problem into a maximizing decision-making (MDM) problem through the application of fuzzy set theory (Figure 1J and 1K). One-side linear membership functions are applied to attribute fuzzy minimization (red line in Figure 1J) and fuzzy maximization (brown line) by applying the equations as follows:

\[
\eta_{\text{min}} = \begin{cases} 
1, & \text{if } FV < LB \\
\frac{UB - FV}{UB - LB}, & \text{if } LB \leq FV \leq UB \\
0, & \text{if } FV > UB
\end{cases}
\]

\[
\eta_{\text{max}} = \begin{cases} 
0, & \text{if } FV < LB \\
\frac{FV - LB}{UB - LB}, & \text{if } LB \leq FV \leq UB \\
1, & \text{if } FV > UB
\end{cases}
\]

where \( FV \) represents the flux values computed using the treated CA or perturbed HT model. The lower bound \( (LB) \) and upper bound \( (UB) \) are obtained using the corresponding CA and HT templates, (i.e., \( LB = ST/4 \) and \( UB = 4ST \); \( ST \) is the standard value for CA or HT template) used in the present study. Two-sided linear membership functions are used to attribute fuzzy dissimilarity (blue line in Figure 1J) and fuzzy
similarity (green line). Fuzzy dissimilarity is a complement of fuzzy similarity; therefore, fuzzy similarity grade is derived using the equation as follows:

Left-hand side membership function:

\[
\eta_L^{PBST} = \begin{cases} 
0, & \text{if } FV < LB \\
\frac{FV - LB}{ST - LB}, & \text{if } LB \leq FV \leq ST \\
1, & \text{if } FV = ST 
\end{cases}
\]

Right-hand side membership function:

\[
\eta_R^{PBST} = \begin{cases} 
1, & \text{if } FV = ST \\
\frac{UB - FV}{UB - ST}, & \text{if } ST \leq FV \leq UB \\
0, & \text{if } FV > UB 
\end{cases}
\]

Therefore, fuzzy similarity grade is evaluated by applying the equation

\[
\eta_{SI} = \max \left\{ \min \left\{ \eta_L^{PBST}, \eta_R^{PBST} \right\}, 1, 0 \right\},
\]

and its complement of fuzzy dissimilarity is obtained by applying the equation \( \eta_{DS} = 1 - \eta_{SI} \).

The ACTD problem in Eq. (1) to (8) can be transformed into an MDM problem. By applying the membership functions as follows:

\[
\max_{\eta_D} \left( \eta_{TR} + \min \{ \eta_{TR}, \eta_{CV}, \eta_{MD} \} \right)/2
\]

subject to inner optimization problems

1. FBA and UFD problems for treated CMS cells
2. FBA and UFD problems for perturbed BL cells

where the decision objective \( \eta_D \) is a hierarchical criterion, and the cell mortality grade \( \eta_{TR} \) is used to evaluate the first goal in Eq. (1) of the outer optimization problem and denoted as the first priority in the fuzzy decision-making problem. The cell viability grade \( \eta_{CV} \) is applied to measure the second goal in Eq. (2), and the metabolic deviation grade \( \eta_{MD} \) is used to measure the third and fourth goals to evaluate the mean-min level (i.e. \( \eta_{MD} = ((\eta_{DS} + \eta_{SI})/2 + \min\{\eta_{DS}, \eta_{SI}\})/2) \). The second priority of the decision objective is used to evaluate the lowest grade in the set \{ \eta_{TR}, \eta_{CV}, \eta_{MD} \} when the cell
viability or metabolic deviation grade is less than the cell mortality grade. The MDM problem in Eq. (11) is a mixed-integer optimization problem with linear and quadratic programming problems in its inner loop. It is a high-dimensional nondeterministic-polynomial-time-hard problem that cannot be solved by currently available commercial software. We employed the NHDE algorithm to solve the MDM problem. The NHDE algorithm is a parallel direct search procedure, which is an extended version based on hybrid differential evolution [45]. The detailed computational procedures are provided in the supporting information (see S1 Text). The source programs of the ACTD platform for identifying anticancer genes, metabolites and reactions of CMSs are available in http://doi.org/10.5281/zenodo.7136561.

3. Results

3.1 CMS-specific Metabolic Models

We used RNA-seq expressions for each CMS and its HT tissue to reconstruct the corresponding CMS-specific and healthy GSMMs (Figure 1A-1C). Figure 2 illustrates the numbers of species, reactions, genes, and feasible encoded enzymes for each reconstructed model. As indicated by the blue regions in Figure 2, five CMSs and HT models share numerous similarities in terms of their species, reactions, genes and enzymes. The orange regions indicate additional shared items in the five CMSs, and the grey regions illustrates the items that are specific to each CMS and HT model. Figure 3 illustrates the top 10 species and reaction classifications for each CMS and HT model. Most species in the fatty acyl groups are shared by the CMS and HT models. High percentages of organooxygen compounds, carboxylic acids, and steroids are also shared by the CMS and HT models (Figure 3A). The species that are specific to CMS2 to CMS5 comprise numerous steroid derivatives. Furthermore, extracellular transport
reactions accounted for the highest percentage of reactions that occur in CMS and HT models. More than 800 fatty acid oxidation reactions are shared by CMS and HT models (Figure 3B).

Figure 2. Number of species, reactions, genes, and feasible enzymes for each consensus molecular subtype (CMS) and its basal model.
Figure 3. Classification of species and reactions in metabolic models of consensus molecular subtypes and their healthy cells.
3.2 Identifying Single Essential Genes

We first used DMEM (see S1 Table) as an uptake nutrient, and 51 uptake reactions were used to set as reversible exchangeable reactions in the computation. By contrast, secretion reactions were set as irreversible reactions. The NHDE algorithm [25-29, 43] was used to perform a series of computations to solve the MDM problem and identify a set of essential genes for each CMS (Table 3). Through the computation, we determined 26 essential genes for CMSs (Table 3), of which 20 were shared by the CMSs and six were CMS-specific. We also used a brute-force enumerate method to determine essential genes one by one to validate our computations, and the results obtained were identical to those obtained through the NHDE algorithm. Through the use of STRING (https://string-db.org/) and GeneCards https://www.genecards.org/), the protein-protein interaction (PPI) networks encoded by the 26 genes were classified into five classes (Figure 4A). The first class comprised 12 genes that are involved in cholesterol biosynthesis. The second class comprised six genes that participate in nucleotide metabolism (specifically purine and pyrimidine metabolism) and one gene in the pentose phosphate pathway. In the third class, four genes are involved in the glycerophospholipid biosynthetic pathway, whereas the others are myo-inositol and ornithine transporters.
Table 3. Cell mortality grades and metabolic deviation grades for essential genes of each consensus molecular subtype, as obtained using DMEM medium. N/D is obtained by dividing number of cell deaths (N) by total number of cancerous cells (D) and used in DepMap-based test. “No. Drugs” denotes number of drugs retrieved from DrugBank that modulate each gene. “SE. Score” denotes side effect score, as calculated using average adverse events.

| Gene   | Cell Mortality Grade | Metabolic Deviation Grade | N/D | No. Drugs | SE. Score |
|--------|----------------------|---------------------------|-----|-----------|-----------|
|        | CMS1 | CMS2 | CMS3 | CMS4 | CMS5 | CMS1 | CMS2 | CMS3 | CMS4 | CMS5 |       |         |
| DHODH  | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.658 | 0.682 | 0.498 | 0.404 | 0.508 | 17/54 | 26 | 0.351   |
| CAD    | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.654 | 0.696 | 0.508 | 0.385 | 0.500 | 27/54 | 3 | --       |
| RPIA   | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.638 | 0.611 | 0.549 | 0.512 | 0.516 | 6/54 | 1 | --       |
| SQLE   | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.575 | 0.693 | 0.548 | 0.489 | 0.488 | 2/54 | 4 | 0.415   |
| ADSS2  | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.569 | 0.680 | 0.464 | 0.499 | 0.504 | 31/54 | 2 | --       |
| EBP    | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.568 | 0.668 | 0.450 | 0.566 | 0.504 | 0/54 | 1 | 0.412   |
| TM7SF2 | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.558 | 0.702 | 0.496 | 0.496 | 0.507 | 1/54 | -- | --       |
| MVD    | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.557 | 0.628 | 0.509 | 0.480 | 0.495 | 49/54 | -- | --       |
| MVK    | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.557 | 0.628 | 0.509 | 0.482 | 0.495 | 52/54 | 1 | --       |
| PMVK   | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.557 | 0.681 | 0.509 | 0.480 | 0.495 | 19/54 | -- | --       |
| LSS    | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.556 | 0.693 | 0.547 | 0.488 | 0.489 | 0/54 | 2 | --       |
| UMP    | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.554 | 0.547 | 0.546 | 0.470 | 0.532 | 29/54 | 2 | --       |
| FDFI   | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.553 | 0.617 | 0.494 | 0.496 | 0.503 | 7/54 | 1 | --       |
| HMGCR  | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.543 | 0.625 | 0.469 | 0.485 | 0.544 | 52/54 | 20 | 0.413   |
| CRLS1  | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.539 | 0.667 | 0.469 | 0.572 | 0.494 | 35/54 | -- | --       |
| SLC7A6 | 0.657 | 0.704 | 0.668 | 0.605 | 0.690 | 0.536 | 0.598 | 0.443 | 0.491 | 0.485 | 0/54 | 1 | --       |
| PG51   | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.488 | 0.596 | 0.482 | 0.498 | 0.501 | 52/54 | -- | --       |
| SC5D   | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.485 | 0.696 | 0.486 | 0.503 | 0.501 | 0/54 | -- | --       |
| AD5L   | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.470 | 0.701 | 0.469 | 0.502 | 0.517 | 51/54 | -- | --       |
| NSDHL  | 0.660 | 0.705 | 0.670 | 0.606 | 0.691 | 0.433 | 0.588 | 0.526 | 0.489 | 0.499 | 0/54 | 1 | --       |
| SLC2A13| 0.660 | --   | --   | 0.606 | --   | 0.448 | --   | 0.528 | --   | --   | 0/54 | -- | --       |
| CYP51A1| 0.660 | --   | --   | --   | 0.691 | 0.435 | --   | --   | --   | 0.495 | 0/54 | 5 | 0.359   |
| CT5S1  | --   | 0.705 | --   | --   | 0.691 | --   | 0.653 | --   | --   | 0.495 | 44/54 | 1 | --       |
| SLC5A3 | --   | --   | 0.670 | --   | --   | --   | 0.541 | --   | --   | --   | 2/54 | 1 | --       |
| PTPMT1 | --   | --   | --   | 0.606 | --   | --   | --   | 0.502 | --   | --   | 28/54 | -- | --       |
| PTDSS1 | --   | --   | --   | 0.606 | --   | --   | --   | 0.488 | --   | --   | 8/54 | 1 | --       |
Figure 4. Protein–protein interactions of identified essential genes for the union set of five consensus molecular subtypes under various nutrient uptakes. (A) DMEM: Dulbecco’s Modified Eagle Medium. (B) HAM: HAM’s medium. (C) VMH: The uptake reactions are obtained from the VMH database (https://www.vmh.life/#home). (D) DMEM+Cholesterol: DMEM medium included the cholesterol uptake. (E) HAM-Cholesterol: HAM medium excluded the cholesterol uptake. (F) VMH-Cholesterol: VMH medium excluded the cholesterol uptake.
For each essential gene, the cancer cells under all the CMS classifications died (cell growth rate \( \leq 10^{-6} \)), and the ATP production rate of these cells decreased by approximately 60% relative to their maximum levels. As a result, the cell mortality grades of cancer cells was more than 0.6 (Table 2), indicating that the first fuzzy goal in Eq.(1) was achieved a 60% satisfaction level. On the basis of this computation, we assumed that the genes in HT cells were also knocked out to investigate the cell viability and metabolic deviation of perturbed cells. We discovered that perturbed HT cells maintained a cell maintenance of \( 10^{-8} \) and their maximum ATP production rate such that the second fuzzy goal was achieved with a 100% satisfaction level. The metabolic deviation grade \( \eta_{MD} \) was used to perform mean-min calculations to evaluate the third and fourth goals in Eq.(1), that is, for measuring the perturbation of the fluxes and metabolite flow rates of the perturbed HT cells relative to CA and HT templates, respectively. For the CMS1 model, the highest metabolic deviation grade (\( \eta_{MD} \)) of 0.658 was achieved with the knockout of dihydroorotate dehydrogenase DHODH, and the lowest \( \eta_{MD} \) value of 0.433 was achieved with the knockout of sterol-4-alpha-carboxylate 3-dehydrogenase NSDHL. A higher grade indicates fewer predicted metabolic perturbation. For CMS2 to CMS5, the highest metabolic deviation grade achieved differed depending on gene that was knocked out (e.g., TM7SF2, RPIA, CRLS1 and HMGCR; Table 3).

A survey of a cancer dependency map (DepMap, https://depmap.org/portal/) revealed that most of the identified genes were compatible with the colon cancer cell lines obtained from DepMap and that the knockout of these genes (except for SLC7A6, SC5D, LSS, EBP, NSDHL and SLC2A13) resulted in a high percentage of cell death (Table 3). Some of the identified essential genes can be modulated by numerous approved drugs that are listed on DrugBank [46]. For example, 26 and 20 of the drugs
surveyed from DrugBank can act on DHODH and HMGCR, respectively [46]. To investigate the grades of adverse events (AEs), the aforementioned approved drugs were used in a SIDER survey (http://sideeffects.embl.de/) involving the use of the ADDReSS (http://www.bio-add.org/ADReCS/) databases. The National Cancer Institute Common Terminology Criteria for Adverse Events (CTCAE) provide precise clinical descriptions of AE severity (from mild AE to death related to AE) grades it on a scale from 1 to 5. We used this scale to calculate the average AE (Ave.AE) and convert it into a side effect score (i.e., $\text{SE.Score} = 1 - \text{Ave.AE}$) that correlates with corresponding metabolic deviation grade.

### 3.3 Nutrient Uptakes

We used HAM medium as the nutrient, and 63 uptake reactions (S1 Table) were used to identify the essential genes for each CMS. The computational results (S2 Table) revealed that the cell viability and metabolic deviation grades obtained through the HAM medium were almost identical to those obtained through DMEM. We discovered that five genes participate in purine and pyrimidine metabolism, RPIA participates in the pentose phosphate pathway, PGS1 and CRLS1 participate in the glycerophospholipid biosynthetic pathway, and SLC7A6 participates in ornithine transport; these are common essential genes for all CMSs. The union set of CMSs comprised 14 identified essential genes, of which nine were shared by CMSs and five were CMS-specific. The union set was used to classify PPIs into four classes (Figure 4B), and the results revealed that the identified essential genes are not involved in the cholesterol biosynthetic pathway.

The VMH database (https://www.vmh.life/#home) published 91 uptake reactions that serve as nutrient resource for human. We used these uptake reactions (referred to
as VMH medium) to investigate effects of nutrients on essential genes. Some of the uptake reactions do not exist in the reconstructed GSMMs of CMSs; thus, they were excluded in computations, and 83 uptake reactions (S1 Table) were used in computations. The NHDE algorithm revealed that 15 essential genes were shared by the CMSs and nine genes were CMS-specific (S2 Table). The union set of 24 essential genes for all CMSs was used to classify PPIs into four classes (Figure 4C), and the results revealed that the identified essential genes are not involved in the cholesterol biosynthetic pathway. The aforementioned findings indicate that these genes are nonessential on a HAM or VMH medium because the cancer cells under all CMSs will still survive if genes involved in cholesterol biosynthesis are knocked out.

Through a comparison of the uptake reactions in DMEM, the HAM medium, and the VMH mediums, the cholesterol uptake reaction was revealed to be absent from DMEM. We used three additional media to investigate the relationship of tumor cell growth with nutrient components and essential genes. DMEM was used as the first additional medium to involve the cholesterol uptake reaction (referred to as the DMEM+Cholesterol medium). The HAM and VMH media were used as the second and third media, respectively, to exclude the cholesterol uptake reaction (denoted as the HAM-Cholesterol and VMH-Cholesterol media). The ACTD platform used each medium to identify the essential genes for each CMS. Figure 4D to 4F illustrates PPI networks for the union set of the essential genes that were identified using the three additional media. The results are presented in S2 Table. When the DMEM+Cholesterol medium was used, the essential genes in the cholesterol biosynthetic pathway could not be determined (Figure 4D). However, the essential genes in the cholesterol biosynthetic pathway were identified using the HAM-Cholesterol and VMH-Cholesterol media. This finding reveals that the essential genes in the cholesterol biosynthetic pathway are
determinable if the medium that is used excludes the cholesterol uptake reaction (Figure 4A, 4E and 4F). By contrast, the genes in the cholesterol biosynthetic pathway were non-essential if the cholesterol uptake reaction was involved in a medium (Figure 4B, 4C and 4D). Our simulation results are consistent with those reported by other studies [38-40], that is, nutrient components can lead to cell growth that is highly relevant to tumor metabolism. Furthermore, we also set all exchange reactions (which comprised more than 782 reactions) for each CMS as reversible uptake reactions to identify essential genes; the results indicated that only CRLS1 was determinable for all CMSs, indicating that CRLS1 was a medium-independent essential gene.

The metabolite flow distributions for CMS1 that were achieved using DMEM and the DMEM+Cholesterol medium are illustrated in Figure 5, which explains the relationship between the cholesterol uptake reaction and essential genes. Normal cells primarily metabolized glucose to pyruvate with a flux of 13.649 mmol/gDW h when DMEM was used and a flux of 13.650 mmol/gDW h when the DMEM+Cholesterol medium was used (Figure 5). Pyruvate then entered the TCA cycle to generate ATP synthesis, which is required for cell survival. The Warburg hypothesis posits that cancer cells rewire their metabolism to promote growth, survival, proliferation, and long-term maintenance. Through the CMS1 model, we discovered that CA template metabolizes large amounts of glucose to pyruvate (23.059 for DMEM and 23.046 for DMEM+Cholesterol), and which was then mostly converted to lactic acid (7.625 for both media). Thereafter, glutamine was replenished in the cycle to meet the proliferation requirement. These results are consistent with the Warburg hypothesis [47-49]. We blocked HMGCR and other genes in the cholesterol biosynthetic pathway (Figure 5) to compute specific metabolite flow rates (i.e., rates listed in the third row of data boxes in Figure 5); the results revealed that the metabolite flow rate of mev-R was
0.0001 for both media and that a series reaction occurred to produce cholesterol at level of 0.0002. Cholesterol production involves intracellular biosynthesis (discussed in an earlier subsection) and an extracellular uptake. The extracellular cholesterol uptake was zero when DMEM was used (Figure 5) because this uptake reaction was excluded such that the genes in the cholesterol biosynthetic pathway became essential. By contrast, when the DMEM+Cholesterol medium was used, the deletion of HMGCR eliminated intracellular biosynthesis; however, the cholesterol was still supplied at a level of 0.001 through the extracellular uptake reaction of the medium. Consequently, HMGCR became a nonessential gene when the DMEM+Cholesterol medium was used.
Figure 5. Metabolite flow rates for CMS1, as obtained using DMEM (first column in matrix of data) and DMEM+Cholesterol (2nd column) media. First column in each data box presents the computational results, as obtained using DMEM; second column presents results, as obtained using DMEM+Cholesterol medium. First and second rows of each data box present CA and HT templates, respectively; third row presents knock-out of genes in cholesterol biosynthetic pathway (e.g., HMGCR and PMVK).
3.4 Combination of Essential Genes

An enumerative search is a time-consuming process for identifying two target combinations of essential genes because it requires more than 4,800,000 combinations for each CMS. The ACTD platform can be used to reduce the computational burden associated with performing evolutionary procedures. Accordingly, we applied the ACTD platform by using two groups of candidates with the NHDE algorithm to identify gene combinations. The first candidate group comprised the union set of essential genes shared by CMSs (Table 1 and S2 Table), and the second group comprised the other candidate genes from feasible encoding enzymes. This strategy substantially reduced the computational time and search space to approximately 86,000 possible combinations for the two candidate groups. We performed a series of computations to obtain a set comprising two identified target combinations for various nutrient uptakes (S3 Table). On the basis of our computational results, we discovered that the metabolic deviation grades for most of the two-target combinations were superior to those for their corresponding one-target essential genes and that each combination involved at least one essential gene. Moreover, on the one-target results presented in Figure 4 indicate that a medium involving cholesterol uptake cannot be used to identify a one-target essential gene in the cholesterol biosynthetic pathway; however, a one-target essential gene can be combined with another essential gene to improve their metabolic deviation grade.

3.5 Essential Metabolites and Reactions

Essential metabolites and reactions denote that cancer cells that will die and the normal cells that will survive if all synthesis reactions to these metabolites are blocked. We also applied metabolite- and reaction-centric approaches on the ACTD platform to
identify essential metabolites and reactions, respectively, for various media (see S4 Table and S5 Table). We compared our computational results by examining various media with or without the cholesterol uptake reaction, and we discovered that most of the essential metabolites were shared by CMSs irrespective of the presence or absence of the cholesterol uptake reaction. The computational results also revealed that the essential metabolites, farnesyl diphosphate (abbreviated as frdp in Figure 5) and formate, can be identified on a medium without the cholesterol uptake reaction, and the only exception was the CMS5 model that used VMH-Cholesterol medium. Farnesyl diphosphate is classified as a chemical compound in prenol lipids that participate in the mevalonate metabolic pathway, and it is catalyzed by FDFT1 to form squalene (sql) and to progressively synthesize cholesterol (Figure 5). Formate is an intermediate metabolite in one-carbon metabolism, and it acts as the sources and sinks in the context of mammalian cell metabolism to link various biosynthetic pathways [50]. For example, formate is released during the synthesis of cholesterol and downstream of the lanosterol step. In the present study, the deletion of formate production reduced intracellular cholesterol synthesis such that the growth of cancer cells was eliminated.

The essential reactions for all CMSs as identified using various media as listed in S5 Table. The computational results revealed that an essential reaction in the cholesterol biosynthetic pathway could not be identified irrespective of whether a medium involved or excluded the cholesterol uptake reaction. This finding differs from that obtained through a gene-centric approach (Table 3), because each essential gene that is involved in the cholesterol biosynthetic pathway through GPR association can regulate at least two reactions. We used two reaction combinations catalyzed by HMGCR to compute cell mortality and metabolic deviation grades, and the results were consistent with those obtained through a HMGCR knockout. Table 4 reveals that the reactions R_DHORD9,
R_ADSD1 and R_ADSS were regulated by their corresponding genes, namely DHODH, ADSL, and ADSS1, respectively; consequently, the block of each reaction was identical to the corresponding gene knockout (Table 4). The essential gene CAD (Table 3) catalyzed three sequential reactions, namely R_CBPS, R_ASPCTr, and R_DHORTS. Therefore, the block of each reaction could eliminate the cell growth and proliferation of each CMS and achieve a satisfactory metabolic deviation grade (Table 4). The essential gene UMPS regulated two sequential reactions, namely R_OMPDC and R_ORPT, in the pyrimidine metabolic pathway such that similar results were yielded.

4. Discussion

CRC is a major disease burden, and the world requires improved prognoses and treatment strategies. Molecular subtype-based therapies provide a new potential framework for preferred and precise medical treatments. CMS classification uses RNA expressions of CRC to categorize patient samples into five subtypes. Several studies have used CMS classification to predict a patient's prognosis and determine treatment strategies for CRC. By contrast, few studies have discussed the use of CMS classification to reconstruct CMS-specific GSMMs and to analyze the metabolic characteristics of these GSMMs. The present study used the RNA-seq expressions of CRC that were obtained from the TCGA database to reconstruct five CMS-specific GSMMs. We discovered that the five CMSs and HT models shared numerous similarities in terms of species, reactions, genes and enzymes. Most species in the fatty acyl groups and more than 800 fatty acid oxidation reactions were shared by the CMS and HT models.

The proposed fuzzy hierarchical optimization framework can identify essential genes, metabolites and reactions for treating each CMS of CRC. The optimization
framework can identify essential targets that lead to cancer cell death and evaluate metabolic flux perturbations of normal cells due to treatment. Metabolic deviation grades and two-side fuzzy membership functions were used to measure the perturbations of fluxes and metabolite flow rates in the perturbed HT cells relative to HT and CA templates, respectively. A smaller metabolic deviation indicates fewer adverse effects. We used various media to identify essential targets for each CMS, and we discovered that most targets were shared by the five CMSs and that some genes belonged to a CMS-specific model. Furthermore, the essential genes in the cholesterol biosynthetic pathway were determinable if the medium used excluded the cholesterol uptake reaction. By contrast, the genes in the cholesterol biosynthetic pathway were nonessential if the medium used involved the cholesterol uptake reaction.
Table 4. Cell mortality grades and metabolic deviation grades for essential reactions of each consensus molecular subtype, as obtained using DMEM. Setting flux value of an essential reaction to zero leads to termination of cancer cell growth and proliferation.

| ID       | Cell Mortality Grade | Metabolic Deviation Grade | Name                                           |
|----------|----------------------|---------------------------|------------------------------------------------|
|          | CMS1  CMS2  CMS3  CMS4  CMS5 | CMS1  CMS2  CMS3  CMS4  CMS5                  |
| R_ASPCTr | 0.660 0.705 0.670 0.606 0.691 | 0.663 0.632 0.520 0.504 0.532 | Aspartate Carbamoyltransferase               |
| R_DHORD9 | 0.660 0.705 0.670 0.606 0.691 | 0.658 0.682 0.498 0.404 0.508 | Dihydroorotic Acid Dehydrogenase             |
| R_OMPDC  | 0.660 0.705 0.670 0.606 0.691 | 0.655 0.555 0.530 0.499 0.526 | Orotidine-5'-Phosphate Decarboxylase          |
| R_ORPT   | 0.660 0.705 0.670 0.606 0.691 | 0.618 0.687 0.524 0.496 0.532 | Orotate Phosphoribosyltransferase             |
| R_CBPS   | 0.660 0.705 0.670 0.606 0.691 | 0.594 0.527 0.422 0.513 0.500 | Carbamoyl-Phosphate Synthase                 |
| R_ADSS   | 0.660 0.705 0.670 0.606 0.691 | 0.569 0.680 0.464 0.499 0.504 | Adenylosuccinate Lyase                       |
| R_ADSSL1 | 0.660 0.705 0.670 0.606 0.691 | 0.569 0.680 0.464 0.499 0.504 | Adenylosuccinate Synthase                    |
| R_DHORTS | 0.660 0.705 0.670 0.606 0.691 | 0.552 0.551 0.497 0.513 0.520 | Dihydroorotate                               |
| R_DATPtn | 0.660 0.705 0.670 0.606 0.691 | 0.504 0.681 0.498 0.528 0.547 | DATP Diffusion in Nucleus                    |
| R_DGTPtn | 0.660 0.705 0.670 0.606 0.691 | 0.459 0.702 0.532 0.574 0.493 | DGTP Diffusion in Nucleus                    |
| R_DSAT   | 0.660 0.705 0.670 0.606 0.691 | 0.449 0.677 0.452 0.503 0.546 | Dihydrosphingosine N-Acyltransferase         |
| R_INST1tr| 0.660 -- -- 0.606 -- | 0.448 -- -- 0.528 -- | Transport of Inositol via Proton Symport     |
| R_TRDR   | 0.660 -- 0.670 -- -- | 0.510 -- 0.545 -- -- | Thioredoxin Reductase (NADPH)                |
| R_INST4  | -- -- 0.670 -- -- | -- -- 0.534 -- -- | Transport of Inositol via Sodium Symport     |
| R_PET    | -- -- 0.510 -- 0.560 | -- -- 0.512 -- 0.477 | Phosphatidylethanolamine Transport           |
| R_PGPP_hs| -- -- -- 0.606 -- | -- -- -- 0.502 -- | Phosphatidylglycerol Phosphate               |
| R_PSSA1 hs| -- -- -- 0.606 -- | -- -- -- 0.488 -- | Phosphatidylserine Synthase                  |
Financial Disclosure

This research was supported by grants from the National Science and Technology Council, Taiwan (https://www.nstc.gov.tw/)

MOST111-2320-B-194-003 and MOST111-2221-E-194-004 to FSW
MOST111-2320-B-030-008 to JML
MOST111-2320-B-075-009 to PMHC
MOST111-2320-B-037-028;KMU-DK(A)111006 to YRH
MOST111-2320-B-A49-036 to CYFH

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data and code availability

The source programs of anticancer target discovery platform and the tissue-specific CMS genome-scale metabolic models are coded by the General Algebraic Modeling System (GAMS, https://www.gams.com/), and are available in http://doi.org/10.5281/zenodo.7136561

Supplementary tables in Microsoft Excel format are available along with this article.

Author contributions

Conceptualization: Jin-Mei Lai, Peter Mu-Hsin Chang, Yi-Ren Hong, Chi-Ying F. Huang, Feng-Sheng Wang

Data curation: Chao-Ting Cheng, Feng-Sheng Wang

Formal analysis: Chi-Ying F. Huang, Feng-Sheng Wang

Funding acquisition: Jin-Mei Lai, Peter Mu-Hsin Chang, Yi-Ren Hong, Chi-Ying F.
Huang, Feng-Sheng Wang

Investigation: Jin-Mei Lai, Peter Mu-Hsin Chang, Yi-Ren Hong, Chi-Ying F. Huang,
Feng-Sheng Wang

Methodology: Chao-Ting Cheng, Feng-Sheng Wang

Project administration: Feng-Sheng Wang.

Resources: Chao-Ting Cheng, Feng-Sheng Wang

Software: Chao-Ting Cheng, Feng-Sheng Wang

Supervision: Jin-Mei Lai, Peter Mu-Hsin Chang, Yi-Ren Hong, Chi-Ying F. Huang,
Feng-Sheng Wang

Validation: Jin-Mei Lai, Peter Mu-Hsin Chang, Yi-Ren Hong, Chi-Ying F. Huang,
Feng-Sheng Wang

Visualization: Chao-Ting Cheng

Writing – original draft: Jin-Mei Lai, Peter Mu-Hsin Chang, Yi-Ren Hong, Chi-Ying F.
Huang, Feng-Sheng Wang

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

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, et.al., Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries, CA: A Cancer Journal for Clinicians. 2021; 71, 209-249. http://doi:10.3322/caac.21660 PMID: 33538338

2. Guinney J, Dienstmann R, Wang X, De Reynies A, Schlicker A, Soneson C, Angelino P, The consensus molecular subtypes of colorectal cancer, Nature Medicine. 2015; 21, 1350-6. http://doi:10.1038/nm.3967 PMID: 26457759

3. Eide PW, Bruun J, Lothe RA, Sveen A, CM Scaller: an R package for consensus molecular subtyping of colorectal cancer pre-clinical models, Scientific Reports. 2017; 7, 1:16618. http://doi:10.1038/s41598-017-16747-x PMID: 29968252

4. Komor MA, Bosch LJW, Bounova G, Bolijn AS, Delis-van Diemen PM, C. Rausch C, Jenster G, Consensus molecular subtype classification of colorectal adenomas. The Journal of Pathology. 2018; 246, 266-276. http://doi:10.1002/path.5129 PMID: 29968252

5. Valenzuela G, Canepa J, Simonetti C, de Zaldívar LS, Marcelain K, González-Montero J, Consensus molecular subtypes of colorectal cancer in clinical practice: A translational approach. World J. of Clinical Oncology. 2021; 12(11), 1000-1008. http://doi: 10.5306/wjco.v12.i11.1000 PMID: 34909395

6. Singh MP, Rai S, Pandey A, Singh NK, Srivastava S, Molecular subtypes of colorectal cancer: An emerging therapeutic opportunity for personalized medicine, Genes & Diseases. 2019; 8, 133-145. http://doi:10.1016/j.gendis.2019.10.013 PMID: 33997160

7. Hoorn ST, de Back TR, Sommeijer DW, Vermeulen L, Clinical value of consensus molecular subtypes in colorectal cancer: a systematic review and meta-analysis, Journal of the National Cancer Institute. 2022; 114, 503-516. http://doi:10.1093/jnci/djab106 PMID: 34077519

8. Mouillet-Richard S, Laurent-Puig P, YAP/TAZ signalling in colorectal cancer: Lessons from consensus molecular subtypes, Cancers. 2020; 12, 3160. http://doi:10.3390/cancers12113160 PMID: 33126419

9. Sawayama H, Miyamoto Y, Ogawa K, Yoshida N, Baba H, Investigation of colorectal cancer in accordance with consensus molecular subtype classification, Annals of Gastroenterological Surgery. 2020; 4, 528-539.
10. Agren R, Mardinoglu A, Asplund A, Kampf C, Uhlen M, Nielsen J, Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling, Molecular Systems Biology. 2014; 10, 721. http://doi:10.1002/msb.145122 PMID: 24646661

11. Folger O, Jerby L, Frezza C, Gottlieb E, Ruppin E, Shlomi T, Predicting selective drug targets in cancer through metabolic networks, Molecular Systems Biology. 2022; 7, 501. http://doi:10.1038/msb.2011.35 PMID: 21694718

12. Ghaffari P, Mardinoglu A, Nielsen J, Cancer metabolism: a modeling perspective, Frontiers in Physiology. 2015; 6, 382. http://doi:10.3389/fphys.2015.00382 PMID: 26733270

13. Jerby L, Ruppin E, Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. Clinical Cancer Research. 2012; 18, 5572-84. http://doi:10.1158/1078-0432.CCR-12-1856 PMID: 23071359

14. Lewis NE, Abdel-Haleem AM, The evolution of genome-scale models of cancer metabolism. Frontiers in Physiology. 2013; 4, 237. http://doi:10.3389/fphys.2013.00237 PMID: 24027532

15. Nilsson A, Nielsen J, Genome scale metabolic modeling of cancer. Metabolic Engineering. 2017; 43, 103-112. http://doi:10.1016/j.ymben.2016.10.022 PMID: 27825806

16. Robinson JL, Nielsen J, Anticancer drug discovery through genome-scale metabolic modeling, Current Opinion in Systems Biology. 2017; 4, 1-8. http://doi.org/10.1016/j.coisb.2017.05.007

17. Gatto F, Miess H, Schulze A, Nielsen J, Flux balance analysis predicts essential genes in clear cell renal cell carcinoma metabolism, Scientific Reports. 2015; 5, 10738, http://doi:10.1038/srep10738 PMID: 26040780

18. Bintener T, Pacheco MP, Sauter T, Towards the routine use of in silico screenings for drug discovery using metabolic modelling, Biochemical Society Transactions. 2020; 48, 955-969. http://doi:10.1042/BST20190867 PMID: 32369553

19. Joly JH, Chew BTL, Graham NA, The landscape of metabolic pathway dependencies in cancer cell lines, PLoS Computational Biology. 2021; 17(4): e1008942. https://doi.org/10.1371/journal.pcbi.1008942 PMID: 33872312

20. Turanli B, Zhang C, Kim W, Benfeitas R, Uhlen M, Arga KY, Mardinoglu A, Discovery of therapeutic agents for prostate cancer using genome-scale metabolic
modeling and drug repositioning, eBioMedicine. 2019; 42, 386-396. http://doi:10.1016/j.ebiom.2019.03.009 PMID: 30905848

21. Jerby L, Shlomi T, Ruppin E, Computational reconstruction of tissue-specific metabolic models: application to human liver metabolism, Molecular systems biology. 2010; 6, 401. http://doi:10.1038/msb.2010.56 PMID: 20823844

22. Ravi S, Gunawan R, ΔFBA—Predicting metabolic flux alterations using genome-scale metabolic models and differential transcriptomic data, PLoS Computational Biology. 2021, 17(11):e1009589. https://doi.org/10.1371/journal.pcbi.1009589 PMID: 34758020

23. Barata T, Vieira V, Rodrigues R, das Neves RP, Rocha M, Reconstruction of tissue-specific genome-scale metabolic models for human cancer stem cells, Computers in Biology and Medicine. 2022; 142, 105177. http://doi:10.1016/j.compbiomed.2021.105177 PMID: 35026576

24. Mardinoglu A, Agren R, Kampf C, Asplund A, Uhlen M, Nielsen J, Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease, Nature Communications. 2014; 5, 3083. http://doi:10.1038/ncomms4083 PMID: 24419221

25. Wang FS, Wang TY, Wu WH, Fuzzy multiobjective hierarchical optimization with application to identify antienzymes of colon cancer cells, Journal of the Taiwan Institute of Chemical Engineers. 2022; 132, 104121. http://doi:org/10.1016/j.jtice.2021.10.021

26. Cheng CT, Wang TY, Chen PR, Wu WH, Lai JM, et.al., Computer-aided design for identifying anticancer targets in genome-scale metabolic models of colon cancer, Biology. 2021; 10, 1115. http://doi:10.3390/biology10111115 PMID: 34827109

27. Wang YT, Lin MR, Chen WC, Wu WH, Wang FS, Optimization of a modeling platform to predict oncogenes from genome-scale metabolic networks of non-small-cell lung cancers, FEBS OpenBio. 2021; 11(8), 2078-2094. http://doi:10.1002/2211-5463.13231 PMID: 34137202

28. Wu WH, Li FY, Shu YC, Lai JM, Chang PMH. et.al., Oncogene inference optimization using constraint-based modelling incorporated with protein expression in normal and tumour tissues, Royal Society Open Science. 2020; 7, 191241. http://doi:10.1098/rsos.191241 PMID: 32269785

29. Wang FS, Wu WH, Hsiu WS, Liu YJ, Chuang KW, Genome-scale metabolic
modeling with protein expressions of normal and cancerous colorectal tissues for oncogene inference, Metabolites. 2019; 10, 16. [https://doi.org/10.3390/metabo10010016] PMID: 31881674

30. Wu HQ, Cheng ML, Lai JM, Wu HH, Chen MC, et.al., Flux balance analysis predicts Warburg-like effects of mouse hepatocyte deficient in miR-122a, PLoS Computational Biology. 2017; 13(7): e1005618. [https://doi.org/10.1371/journal.pcbi.1005618] PMID: 28686599

31. National Cancer Institute of U.S., The Cancer Genome Atlas Program, Department of Health and Human Services, 2020. [https://portal.gdc.cancer.gov/]

32. Uhlen M, Oksvold P, Fagerberg L, Lundberg E, Jonasson K, et.al., Towards a knowledge-based human protein atlas, Nature Biotechnology. 2010; 28(12), 1248-50. [http://doi:10.1038/nbt1210-1248] PMID: 21139605

33. Duarte NC, Becker SA, Jamshidi N, Thiele I, Mo ML, et.al., Global reconstruction of the human metabolic network based on genomic and bibiomic data, Proceedings of the National Academy of Sciences. 2007; 104(6), 1777-82. [http://doi:10.1073/pnas.0610772104] PMID: 17267599

34. Thiele I, Swainston N, Fleming RM, Hoppe A, Sahoo S, et.al., A community-driven global reconstruction of human metabolism, Nature Biotechnology. 2013; 31(5), 419-25. [http://doi:10.1038/nbt.2488] PMID: 23455439

35. Swainston N, Smallbone K, Hofsi H, Dobson PD, Brewer J, et.al., Recon 2.2: from reconstruction to model of human metabolism, Metabolomics. 2016; 12: 109. [http://doi:10.1007/s11306-016-1051-4] PMID: 27358602

36. Ryu JY, Kim HU, Lee SY, Framework and resource for more than 11,000 gene-transcript-protein-reaction associations in human metabolism, Proceedings of the National Academy of Sciences. 2017; 114(45), E9740-E9749. [http://doi:10.1073/pnas.1713050114] PMID: 29078384

37. Brunk E, Sahoo S, Zielinski DC, Altunkaya A, Dräger A, et.al., Recon3D enables a three-dimensional view of gene variation in human metabolism, Nature Biotechnology. 2018; 36(3), 272-281. [http://doi:10.1038/nbt.4072] PMID: 29457794

38. Robinson JL, Kocabas P, Wang H, Cholley PE, Cook D, et.al., An atlas of human metabolism. Science Signaling; 2020; 13(624), eaaz1482. [http://doi:10.1126/scisignal.aaz1482] PMID: 32209698

39. Cantor JR, Abu-Remaileh M, Kanarek N, Freinkman E, Gao X, et.al., Physiologic
medium rewires cellular metabolism and reveals uric acid as an endogenous inhibitor of UMP synthase, Cell. 2017; 169(2), 258-272. 
http://doi:10.1016/j.cell.2017.03.023 PMID: 28388410

40. Vande Voorde J, Ackermann T, Pfetzer N, Sumpton D, Mackay G, et.al., Improving the metabolic fidelity of cancer models with a physiological cell culture medium, Science Advances. 2019; 5(1): eaau7314. 
http://doi:10.1126/sciadv.aau7314 PMID: 30613774

41. Ackermann T, Tardito S, Cell culture medium formulation and its implications in cancer metabolism, Trends in Cancer. 2019; 5(6), 329-332. 
http://doi:10.1016/j.trecan.2019.05.004 PMID: 31208694

42. Schultz A, Qutub AA, Reconstruction of tissue-specific metabolic networks using CORDA. PLoS Computational Biology. 2016; 12(3): e1004808. 
http://doi:10.1371/journal.pcbi.1004808 PMID: 26942765

43. Wang FS, Nested differential evolution for mixed-integer bi-level optimization for genome-scale metabolic networks, in Differential evolution in chemical engineering Edited by W. H. Wu, Fuzzy G. P. Rangaiah, (2017), World Scientific, ISBN: 978-981-3207-51-6 https://doi.org/10.1142/10379

44. Sakawa M, Fuzzy sets and interactive multiobjective optimization, (1993), Springer Science & Business Media. ISBN: 978-1-4899-1635-8

45. Chiou JP, Wang FS, Hybrid method of evolutionary algorithms for static and dynamic optimization problems with application to a fed-batch fermentation process, Comput. Chem. Eng. 1999; 23(9), 1277-1291. 
http://doi.org/10.1016/S0098-1354(99)00290-2

46. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, et. Al., DrugBank 5.0: a major update to the DrugBank database for 2018, Nucleic Acids Research. 2018; 46(D1), D1074-D1082. 
http://doi:10.1093/nar/gkx1037 PMID: 29126136

47. Liberti MV, Locasale JW, The Warburg effect: how does it benefit cancer cells? Trends in Biochemical Sciences. 2016; 41(3), 211-218. 
http://doi:10.1016/j.tibs.2015.12.001 PMID: 26778478

48. Martínez-Reyes I, Chandel NS, Cancer metabolism: looking forward, Nature Reviews Cancer. 2021; 21(10), 669-680. 
http://doi:10.1038/s41568-021-00378-6 PMID: 34272515

49. Pavlova NN, Zhu J, Thompson CB, The hallmarks of cancer metabolism: Still emerging, Cell Metabolism. 2022; 34(3), 355-377.
50. Pietzke M, Meiser J, Vazquez A, Formate metabolism in health and disease, Molecular Metabolism. 2020; 33, 23-37. http://doi:10.1016/j.molmet.2019.05.012 PMID: 31402327

Supporting Information

S1 Table. Uptake reactions for DMEM, HAM medium and VMH medium

S2 Table. Cell mortality grades and metabolic deviation grades for essential genes of each consensus molecular subtype, as obtained using various media.

S3 Table. Cell mortality grades and metabolic deviation grades for combination of essential genes of each consensus molecular subtype, as obtained using various media.

S4 Table. Cell mortality grades and metabolic deviation grades for essential metabolites of each consensus molecular subtype, as obtained using various media.

S5 Table. Cell mortality grades and metabolic deviation grades for essential reactions of each consensus molecular subtype, as obtained using various media.

S1 Text. Description of the nested hybrid differential evolution algorithm. The source programs for identifying essential targets are coded by the GAMS software language.