Computers in Biology and Medicine 170 (2024) 108052

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Reconstruction of cell-specific models capturing the influence of metabolism on DNA methylation in cancer

Tânia Barata [a,b], Vítor Pereira [c,d], Ricardo Pires das Neves [a,b,e,*], Miguel Rocha [c,d,f,**]

[a] CNC - Center for Neuroscience and Cell Biology, University of Coimbra, 3004-517 Coimbra, Portugal
[b] CIBB - Centre for Innovative Biomedicine and Biotechnology, University of Coimbra, 3004-517 Coimbra, Portugal
[c] Centre of Biological Engineering, University of Minho - Campus de Gualtar, Braga, Portugal
[d] LABBELS - Associate Laboratory, Braga/Guimarães, Portugal
[e] IIIUC-Institute of Interdisciplinary Research, University of Coimbra, 3030-789 Coimbra, Portugal
[f] Department of Informatics, University of Minho, Portugal

ARTICLE INFO

Dataset link: https://doi.org/10.5281/zenodo.10031423, https://github.com/BioSystemsUM/epigen

Keywords:
Cancer
Genome-scale metabolic models (GSMMs)
DNA methylation
GSMMs enhanced with enzymatic constraints using kinetic and omics data (GECKOs)

ABSTRACT

The imbalance of epigenetic regulatory mechanisms such as DNA methylation, which can promote aberrant gene expression profiles without affecting the DNA sequence, may cause the deregulation of signaling, regulatory, and metabolic processes, contributing to a cancerous phenotype. Since some metabolites are substrates and cofactors of epigenetic regulators, their availability can be affected by characteristic cancer cell metabolic shifts, feeding cancer onset and progression through epigenetic deregulation. Hence, there is a need to study the influence of cancer metabolic reprogramming in DNA methylation to design new effective treatments. In this study, a generic Genome-Scale Metabolic Model (GSMM) of a human cell, integrating DNA methylation or demethylation reactions, was obtained and used for the reconstruction of Genome-Scale Metabolic Models enhanced with Enzymatic Constraints using Kinetic and Omics data (GECKOs) of 31 cancer cell lines. Furthermore, cell-line-specific DNA methylation levels were included in the models, as coefficients of a DNA composition pseudo-reaction, to depict the influence of metabolism over global DNA methylation in each of the cancer cell lines. Flux simulations demonstrated the ability of these models to provide simulated fluxes of exchange reactions similar to the equivalent experimentally measured uptake/secretion rates and to make good functional predictions. In addition, simulations found metabolic pathways, reactions and enzymes directly or inversely associated with the gene promoter methylation. Two potential candidates for targeted cancer epigenetic therapy were identified.

1. Introduction

In the past decades, there has been an increase in the incidence of early-onset cases of cancer [1]. Changes in lifestyle, environment, and diet, together with genetic susceptibilities, have contributed to genetic mutations that trigger an imbalance of cell differentiation, survival, and/or proliferation, promoting cancer onset and development [1,2]. In addition to genetic mutations, which directly affect the DNA sequence, the de-regulation of epigenetic mechanisms, which control the attachment of chemical groups to DNA, histones, and nucleosome-positioning protein complexes, can also induce cancerous phenotypes. In particular, the imbalance in epigenetic modifications may change the chromatin accessibility to transcriptional complexes, and subsequently, induce aberrant gene expression profiles without affecting the genomic sequence [2].

Another fundamental feature of cancer is its metabolic rewiring, as cancer cells are forced to adapt their metabolism to generate enough energy and elementary metabolites for the synthesis of new cellular membranes, proteins, or nucleic acids necessary for cell proliferation [3,4]. Furthermore, given that distinct metabolites are also substrates or cofactors of epigenetic regulators [5], alterations in their availability, as a consequence of metabolic reprogramming, can induce a cancerous phenotype through epigenetic deregulation [2,6], whereas, on the other end, epigenetic alterations on genes encoding metabolic enzymes may contribute to the metabolic shift characteristic of cancer. Therefore, there is an urge to investigate the cross-talk between cancer, epigenetics, and metabolism to develop new and efficient therapeutic strategies against the disease.

* Corresponding author.
** Corresponding author at: Centre of Biological Engineering, University of Minho - Campus de Gualtar, Braga, Portugal.
E-mail address: mrocha@di.uminho.pt (M. Rocha).

https://doi.org/10.1016/j.compbiomed.2024.108052
Received 6 June 2023; Received in revised form 18 January 2024; Accepted 26 January 2024
Available online 1 February 2024

Genome-Scale Metabolic Models (GSMMs) are mathematical representations of all metabolic reactions of a cell, where reactions catalyzed by enzymes are mapped to associated genes and/or proteins [7]. By assuming the steady state (that metabolite concentrations do not change over time), it has been possible to utilize these *in silico* constructs in the prediction of metabolic phenotypes. In detail, the product of a matrix with the stoichiometric coefficients (where columns and rows represent respectively the reactions and metabolites) and a vector of reaction fluxes (rates) is assumed to be zero upon the steady state. This results in a solvable system of linear equations, the solution of which comprises fluxes of all metabolic reactions represented in the system [8]. The development of methodologies for omics data acquisition over the years has favored the reconstruction of this type of model under specific biological contexts by integration of omics data of a particular tissue or cell type. In particular, a variety of healthy cells and morbidities [9–13], including cancer [14–16] have been modeled with context-specific GSMMs.

The main disadvantage of traditional constraint-based GSMMs is that, unless nutrient uptake fluxes are known, no finite flux distribution can be obtained. Unlike traditional GSMMs, Genome-Scale Metabolic Models enhanced with Enzymatic Constraints using Kinetic and Omics data (GECKOs) do not require nutrient-uptake rates to produce finite flux values during simulations, as they integrate both enzymatic kinetic information and concentration, that serve as additional constraints to the flux solution space. Specifically, enzymes are added as pseudo-metabolites that although represented as substrates, do not affect the mass balance of the reactions they catalyze, and pseudo-uptake reactions for each enzyme are included to guarantee enzyme mass balance. This results in an extended version of the abovementioned matrix of stoichiometric coefficients, where additional rows representing the enzyme mass balance and columns depicting enzyme usage reactions are introduced. The catalytic information is presented in the form of the inverse of turnover number (k_{cat}) values, as coefficients to the enzymes in metabolic reactions, whereas enzyme concentrations are used as the upper bounds of enzyme usage reactions. When no proteomics data is available to limit the flux of each enzyme usage reaction, an enzyme usage reaction of the pool of all enzymes is introduced instead and each enzyme is drawn from the enzyme pool [17].

Few studies have attempted to use GSMMs to model the interaction of metabolism and epigenetics. An old study from 2014 [18] integrated the decrease in gene expression observed upon mutation of histone tails, which are often mutated in cancer and are targeted by epigenetic marks, into a yeast model to represent the effect of those mutations on the rate of production/consumption of acetyl-CoA, a substrate for histone acetylation. In the following year, an analysis was published [19] where metabolic models reconstructed for different time points, based on time-course transcriptomics data, provided simulations that were compared with ChIP-seq data for a histone-acetylation mark, to capture the differentiation of primary human monocytes to macrophages. The authors observed that enhancers of metabolic genes under high regulatory load (close to histones with high levels of the acetylation mark) were mainly associated with transport reactions and other metabolic pathway entry points in comparison with other metabolic genes, suggesting that the former are critical epigenetic-regulatory control points for the metabolic reprogramming during monocyte to macrophage differentiation [19]. In another study from 2017, Chandrasekaran et al. [20] tried to predict in which of the two states murine pluripotent stem cells go through during embryonic development, preceding (naïve state) or succeeding (primed state) the implantation of the embryo in the uterus, was producing more S-Adenosyl-Methionine (SAM), a substrate for methylation. Using a semi-dynamic modeling approach, the authors suggested that histone methylation was more intense in the primed cells, which was experimentally verified afterward [20]. Most recently, Shen et al. successfully predicted the increase or decrease in protein acetylation levels in human cells in the presence of different nutrient sources. Furthermore, through the inclusion of one reaction representing the overall protein acetylation, cancer cell lines that were more sensitive to *vorinostat*, a deacetylase inhibitor used in cancer treatment, were estimated to have higher acetylation levels, suggesting that GSMMs could be used to identify cancer cells more responsive to treatments with deacetylase inhibitors [21].

Although those studies represent important steps toward the modulation of the interplay between metabolism and epigenetics, they all focus on histone modifications, particularly acetylation. The only study that addresses methylation dwells on histone methylation in murine cells and simply uses the flux of SAM as a surrogate for methylation. In the present work, we reconstruct models for 31 human cancer cell lines which included DNA methylation and demethylation reactions described in the literature, as well as DNA methylation levels estimated from experimental data. Furthermore, these models are GECKOs, which present the advantage of providing more accurate flux distributions than traditional GSMMs when experimental flux values are unavailable.

2. Results

In this study, GECKO models containing DNA methylation and demethylation reactions were reconstructed for different cancer cell lines. Those reactions, which included DNA containing modified cytosines were retrieved from literature, adapted for charge and mass balance, and were first introduced on the generic GSMM *Human1* [22] (more details on *Creation of the generic DNA methylation model* section of Materials and Methods). The complete list of reactions introduced is shown in 1xt and a simplified visual representation of how those reactions integrate with the model is presented in Fig. 1.

In a nutshell, the DNA methylation process starts when SAM is produced in the one-carbon cycle in the cytoplasm through reaction *MAR03875*, and, once inside the nucleus, it is used as a substrate of DNA methylation through reaction *MAR08641* (Fig. 1). DNA can then be demethylated using different pathways (Fig. 1). DNA-5-methylcytosines (DNA5mC) can be successively oxidized to DNA-5-hydroxymethylcytosines (DNA5hmC), DNA-5-formylcytosines (DNA-5fC), and DNA-5-carboxylcytosines (DNA5CaC) or converted to thymines, i.e. DNA-5-methyluracils (DNA5mU). DNA5fC and DNA5CaC can be transformed back to unmethylated cytosines in DNA (through the enzyme non-catalyzed reactions *consdirectDNA5fC* and *consdirectDNA5CaC*), or like DNA5mU, they can be replaced by unmethylated-cytosines through the cellular Base-Excision Repair (BER) mechanism. The BER starts with the excision of the modified cytosine (reactions *prodAPsite3* and *prodAPsite4*) or the mismatched thymine (*prodAPsite1* reaction) using DNA-glycosylases that cleave the bound between the base and the deoxyribose creating an apyrimidinic site (APsite). An endonuclease then cuts the phosphate backbone at the APsite, leaving a nick and a deoxyribo-5'-phosphate (dRP) to which the excised base was connected (in reaction *proddRPsite*). A new unmethylated cytosine is inserted afterward, while the dRP is still hung by its 3' side to the phosphate backbone (in *prodhangdRPsite* reaction). The dRP is excised by a dRP lyase, creating a nick in the DNA strand (in *prodDNAnick* reaction), which is then ligated by a DNA ligase, restoring the unmethylated DNA (in *ligate DNA* reaction) (see Fig. 1).

Since the ratio of DNA5mC, DNA5hmC, and DNA5fC in relation to unmethylated DNA can be estimated, a pseudo-reaction representing the composition of total DNA (DNAtot) in terms of those species was also introduced in the model (*prodDNAtot* reaction), and the original biomass reaction was replaced by an equivalent one (*adaptbiomass* reaction) where DNA was substituted by DNAtot (Fig. 1).

The final adapted generic model was able to produce biomass and none of the introduced (de)/methylation reactions were blocked. Context-specific GSMMs were then built for different NCI-60 cell lines through the integration of transcriptomics data, and a procedure from Robinson et al. [22] was used to convert those traditional GSMMs into GECKO models.

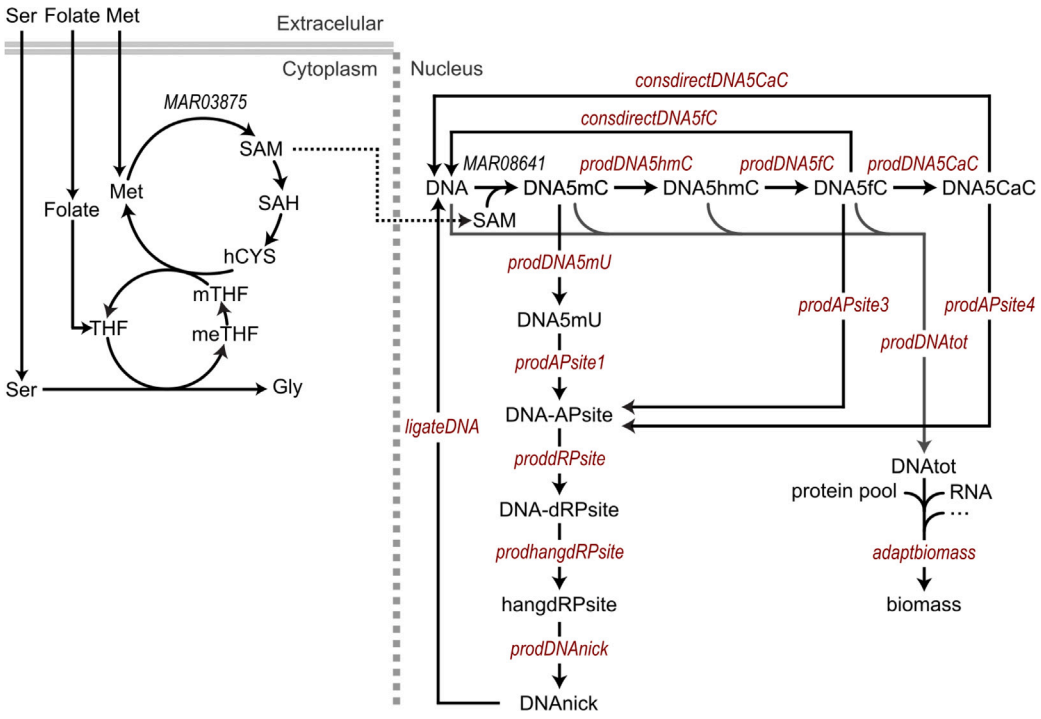

Fig. 1. Visual representation of reactions contributing to DNA methylation and demethylation. Note that this is just a simplified scheme, as it does not reflect the stoichiometric proportions and excludes many metabolites and transport reactions (for all complete reactions see Tab. A.1). Reaction identifiers are in italic. Those in red color were added to the original *Human1* (version 1.12) generic model. Gly: Glycine; Ser: Serine; Met: Methionine; THF: Tetrahydrofolate; mTHF: methyl-THF; meTHF: 5,10-methylene-THF; SAM: S-Adenosyl-Methionine; SAH: S-Adenosyl-Homocysteine; hCYS: Homocysteine; DNA5mC: DNA-5-methylcytosine (i.e. methylated DNA); DNA5hmC: DNA-5-hydroxymethylcytosine; DNA5fC: DNA-5-formylcytosine; DNA5CaC: DNA-5-carboxylcytosine; DNA5mU: DNA-5-methyluracil.

2.1. Reconstruction of cell-specific metabolic models

This study tested two strategies previously applied to build GSMMs of NCI-60 cell lines. One of those, introduced by Richelle et al. [23] in MATLAB, and here implemented in Python, is based on the inclusion of cell-type specific metabolic tasks. Initially, gene scores resulting from the preprocessing of transcriptomics data were converted to reaction scores for each cell line, using Gene-Protein-Reaction (GPR) rules. The highest reaction scores in a cell line were then attributed to all reactions necessary for each generic metabolic task, deeming a generic task as a metabolic task done by all cell types. Also, the same procedure was applied for reactions necessary for cell-type specific tasks, as long as those tasks are done by the specific cell type under consideration. Reactions necessary for a task were identified as the ones carrying flux after the implementation of the task-associated flux constraints on the generic model upon minimization of the sum of all fluxes. In order to determine whether a cell type performs a certain cell-specific task, a metabolic score was calculated for each task and cell type combination (see more details in *Reconstruction of cell line-specific traditional GSMMs* section of Materials and Methods). Afterward, the reconstruction algorithm FASTCORE was applied to build the cell-specific models, because, like other MBA-based methods and unlike iMAT-based methods (such as INIT), it preserves almost all tasks after reconstruction [23].

The second strategy consisted of directly using a version of the tINIT algorithm already implemented in MATLAB by Robinson et al. [22], which preserves the generic metabolic tasks (see more details in the *Reconstruction of cell line-specific traditional GSMMs* section of Materials and Methods).

For the selection of the best reconstruction and simulation strategies, models were initially built for only 40 to 42 of the NCI-60 cell lines, due to the lack of transcriptomics data, DNA methylation measurements, and metabolite uptake rates for some cell lines. The exact number varied with the reconstruction strategy applied depending on the number of infeasible models (1 or 2). However, after the selection of the best strategy, the number of models used in subsequent simulations was reduced to 31, due to the lack of another data type needed for model integration with the degree of DNA methylation, as will be explained further below.

2.2. Detection and validation of the best reconstruction and simulation pipelines

Since no flux distribution can be obtained from unconstrained traditional GSMMs, the uptake rates of metabolites in Ham's media were loosely constrained (from −1000 to 1000) and those of other metabolites were closed (set from 0 to 1000). Then, a parsimonious Flux Balance Analysis (pFBA) was applied and the resulting simulated fluxes of exchange reactions of 26 metabolites were compared with experimentally measured ones. Although there was a small correlation between simulated and measured fluxes, 0.33–0.51 and 0.34–0.49 of Pearson and Spearman correlation respectively (with a p-value of zero), most of the simulated values did not match measured ones, i.e. the logarithm of their absolute values was higher or lower than ± 1 of $\log_{10}(|\text{measured value}|)$ (most data points fell outside the pink area of the graphs in Fig. A.1). Note that the absolute values of the fluxes were logarithmized as the majority presented small values (close to zero) (Fig. A.1–E). Richelle's approach (Fig. A.1–A,B) showed a higher percentage of matching values than Robinson's strategy (Fig. A.1–C,D), and the integration of tasks was slightly detrimental to the correlation values in both reconstruction methodologies (Fig. A.1–A,C versus B,D). The predicted flux values of biomass were much higher than the measured ones and the relative errors of predicted growth rates were high (Fig. A.2). Note that the goal of this analysis is not to assess whether one reconstruction algorithm is better than the other (tINIT and FASTCORE), as that would require both algorithms to be run with the same programming language, but instead find a reconstruction approach that provides good quality models for the specific NCI-60 cell line dataset.

Given that constraints in the uptake/secretion rates of three metabolites (glucose, lactate, and threonine) with experimentally measured values had been previously reported as sufficient to generate small growth rate prediction errors for the models of eleven of the NCI-60 cell lines [22], we decided to test the effect of those constraints here. The absolute values of the fluxes were again logarithmized because many presented small values (close to zero) (Fig. A.3-E), and only 23 metabolites were taken into account, as the three metabolites whose fluxes were constrained were excluded from the analysis to prevent bias. Overall, there was an increase in the percentage of simulated fluxes whose values were similar to those measured (the $\log_{10}$(|simulated value|) was within $\log_{10}$(|measured value|) ± 1, as 65%–74% of data points are inside the pink area of the graphs in Fig. A.3) in relation to the loosely constrained models (Fig. A.1). This was coupled with an improvement in the correlation between simulated and measured values (the Pearson and Spearman correlations enhanced to 0.47-0.62 and 0.54-0.59). Furthermore, the correlations and the percentages of biomass flux values in close proximity to those measured greatly increased, while the relative errors of predicted growth rates reduced in comparison with the loosely constrained models (Fig. A.4 versus Fig. A.2). Unlike the loosely constrained models, the integration of cell-specific tasks provided a slight improvement when using Richelle's strategy (Fig. A.3-A versus B) and the overall best-performing reconstruction strategy was, in this case, Robinson's approach (Fig. A.3-A,B versus C,D).

With this data set, good simulations were obtained by limiting the fluxes of three exometabolites with experimental data. However, one of the purposes of this study is to present a pipeline that can be adopted in the future to different datasets, creating models that depict the interplay of metabolism and DNA methylation in other biological contexts, for most of which such experimentally measured metabolite uptake/secretion rates are unknown. Therefore, we assessed whether GECKO models without constraints on exchange metabolite uptake rates could be enough to make accurate predictions. Although an enzymatic pFBA with GECKO models in which the only constraint was the limitation of the protein pool uptake (with estimated cell-specific total protein concentrations) provided smaller correlations, it predicted more fluxes in close agreement with measured values (69%–77% in Fig. A.5) than both the traditional loosely constrained GSMMs (1%–3% in Fig. A.1) and those constrained with the three exometabolites uptake rates (65%–74% in Fig. A.3). Robinson's strategy was the reconstruction approach that gave the best results with enzyme-constrained models (Fig. A.5-A,B versus C,D) and the inclusion of tasks was detrimental (Fig. A.5-A,C versus B,D). Hence, subsequent simulations were performed with GECKO models reconstructed with Robinson's approach and excluding reactions necessary for tissue-specific tasks.

Even though 100% of $\log_{10}$(|biomass flux|) values predicted with GECKO models lay within ± 1 of $\log_{10}$(|measured value|) (fig. A.6) and the relative error in prediction of growth rates is in agreement with previously reported values for eleven of the NCI-60 cell lines [22], there is no significant correlation (p-value > 0.05) between simulated and real biomass flux values as the simulated values were underestimated (i.e. most data points are below the diagonal line in Fig. A.6-A–D).

One possible explanation for this is the assignment of default values to two parameters influencing the limitation given to the total protein uptake flux. Those parameters are σ, which accounts for the level of enzyme saturation *in vivo*, and f, the mass fraction of enzymes that are accounted for in the model out of all proteins present in the cell. These parameters can change with the cell type and are unknown for NCI-60 cell lines. Another factor that could have contributed to the underestimation of biomass flux is an incorrect assessment of the real value of total protein concentration.

In fact, when biomass fluxes together with the total protein concentration were constrained with experimental values and an FBA with minimization of total protein uptake reaction was performed on GECKOs reconstructed with the best strategy (Robinson's pipeline and without tasks), only two models were feasible, reinforcing that the aforementioned parameters or total protein concentrations are not correct. Therefore, we did a similar simulation where biomass fluxes were limited with experimentally measured flux rates, but without limiting the total enzyme pool uptake rate.

As expected, the limitation of the biomass flux with bounds determined from experimental values led the biomass simulated flux to be closer to the mean measured flux (Fig. 2-B versus Fig. A.6-C), improving the relative error of the growth rate (Fig. 2-D versus Fig. A.6-E). Regarding the fluxes of the 26 exometabolites, the restriction of growth rates gave similar results as without the constraints on biomass (Fig. 2-A versus Fig. A.5-C). Furthermore, the percentage of simulated flux values within proximity to those measured in GECKOs with a constraint on biomass (77% in Fig. 2) is higher than with traditional GSMMs with a constraint on biomass (71% in Fig. A.7). Hence, subsequent simulations were done with GECKO models reconstructed with Robinson's approach and limited by experimental growth rates, while minimizing the total enzyme usage.

2.3. Integration of models with cell line-specific DNA methylation levels and generic DNA methylation flux rules

The overall degree of protein acetylation of different human cell lines has been previously predicted in a study using traditional GSMMs, in which the simulated flux of a pseudo-reaction of global protein acetylation was shown to correlate with the amount of one type of histone acetylation mark that functions as an epigenetic regulator [21]. Therefore, in this study, we assessed whether an equivalent correlation could be observed between the simulated DNA methylation flux and the degree of DNA methylation estimated with experimental data (details on the estimation procedure in *Comparison of fluxes of reactions involved in DNA (de)/methylation and the degree of DNA methylation* section of Material and Methods).

The results presented in Fig. A.8-A demonstrate a lack of strong correlation between the actual global DNA methylation level (details of its estimation in *Calculation of the composition of total DNA* section of Materials and Methods) and the simulated flux of the DNA methylation reaction (*MAR08641*) as well as between the first and the flux of the reaction that produces the cytoplasmatic SAM (*MAR03875*), which is one of the substrates of DNA methylation. Note that the values in the scatter plots were logarithmized because many of the simulated flux values of reactions *MAR03875* and *MAR08641* were close to zero (Fig. A.8-B). Only a weak, but significant (p-value ≤ 0.05) correlation, with just the Spearman (not with the Pearson) method, was observed between the flux of each mentioned reaction and the global DNA methylation. The genomic region that gave the best significant correlations for the DNA methylation reaction (*MAR08641*), although still weak (0.36 and 0.52 of Pearson and Spearman coefficients respectively), was the one comprising 1000bp upstream of the genes' Transcription Start Sites (TSS), i.e. gene promoters (see reaction *MAR08641* in Fig. A.8-A). In addition, it was observed that the overall correlation values are slightly higher for the DNA methylation reaction (*MAR08641*) than the one producing SAM (*MAR03875*).

Since the simulated methylation fluxes were not able to strongly predict the degree of DNA methylation, we switched our focus to understanding how metabolic mechanisms and metabolic shifts are related to the overall degree of DNA methylation in cancer, which is the ultimate goal of the present work. For that purpose, the degree of DNA methylation across the genome was integrated within the models. Specifically, the stoichiometric coefficients of the pseudo-reaction *prodDNAtot*, which represents the composition of total DNA in terms of DNA cytosine (de)/methylation marks, were modified based on published cell/tissue-specific DNA5mC and DNA5hmC datasets (more details on *Calculation of the composition of total DNA* section of Materials and Methods). Note that values of DNA5hmC levels were not available

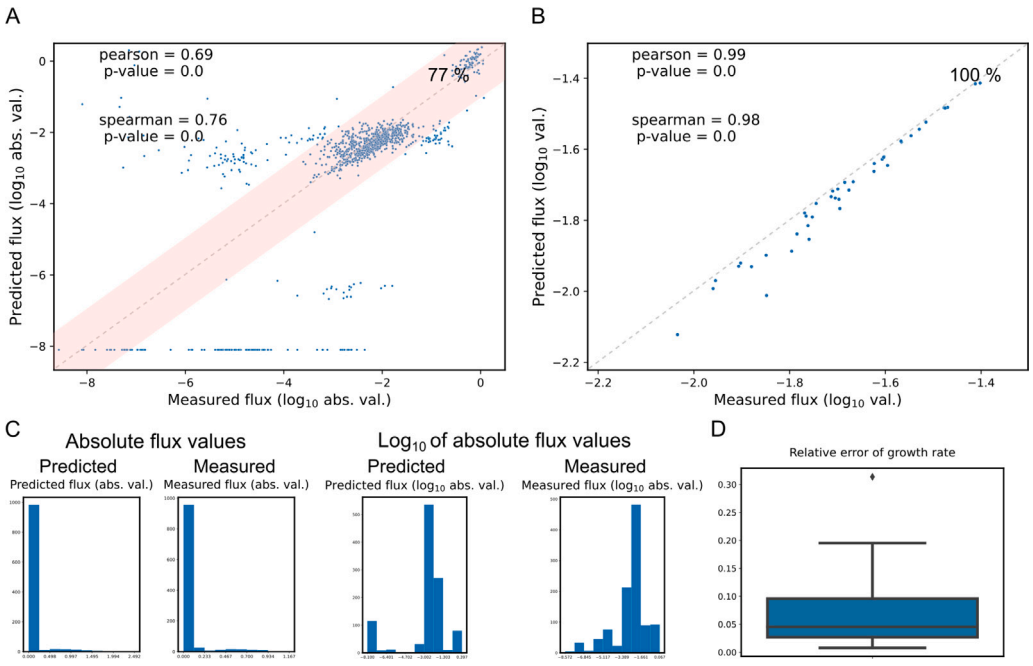

Fig. 2. Comparison of measured and simulated exchange fluxes produced by GECKO models constrained with measured growth rates. **A**: a scatter plot with $\log_{10}$ of values of simulated and measured fluxes of exchange reactions of 26 metabolites. Value at top right of **A** and **B** is the percentage of data points where $\log_{10}$(|predicted value |) is within $\log_{10}$(|measured value |) ± 1 (in **A**, it corresponds to the pink area). **B**: a scatter plot with $\log_{10}$ of values of simulated and measured growth rates. The correlation coefficients are not exactly one, because experimentally determined upper and lower bounds were used to constraint simulated biomass fluxes, while the value of the measured biomass in the graph is the average of those bounds. **C**: histograms with the distribution of absolute values of measured and simulated fluxes before and after logarithmization. **D**: Relative errors of predicted growth rates. Data points forming a line at the bottom of **A** correspond to metabolites with a predicted flux of zero, which are shown in the graphs as holding the lowest absolute measured value (besides zero), as the logarithm of zero is undefined. In **C**, the values of those metabolites fall in the lowest bin, creating an oddly tall bin at the beginning. Models used were reconstructed with Robinson's pipeline and without tissue-specific tasks.

for all cell lines. Hence, from then on simulations were made with models for only 31 of the 41 cell lines.

Another interesting observation from the previous simulations is that while the DNA methylation reaction always carried flux for any of the different cell lines, none of the DNA demethylation reactions included in the model was able to do the same, because the formation of unmethylated DNA necessary for the biomass production was directly obtained through the DNA polymerization reaction of the individual nucleotides instead of the DNA demethylation reactions. However, it is known that the extent to which the DNA is methylated depends on the balance between the rates of the reactions of methylation and demethylation, which produces a dynamic DNA methylation turnover steady-state [24]. In fact, variations in the proportion of those rates can originate methylation deregulation like the hypermethylation (i.e. silencing) of tumor suppressor genes and hypomethylation (i.e. activation) of pro-metastatic genes observed in cancer cells [24–26]. Hence, to guarantee that the simulations can reflect the dynamic DNA methylation turnover state, the flux of certain DNA demethylation reactions was forced to be positive in the subsequent simulations by first, including reactions necessary for DNA demethylation tasks when the task score was high in each specific cell line (above a threshold - details in Materials and Methods); second, by constraining the flux of DNA demethylation reactions in each model with reaction rate ratios previously described in the literature, as long as the imposed constraints would not produce an infeasible flux distribution. Those rate ratios are described in Tab. A.8 and will herein forth be called *methylation flux rules*.

As would be expected, there was an improvement in the correlation between the simulated fluxes of reactions *MAR08641* or *MAR03875* and the estimated degree of DNA methylation after adapting the composition of total DNA with cell-specific information (compare Fig. 3 with Fig. A.8). This time, the correlation between the DNA methylation flux (*MAR08641*) and the global DNA methylation was strong (0.62 and 0.73 of Pearson and Spearman coefficients respectively) and significant (p-value ≤ 0.05), and the *Upstream of TSS* was again the genomic feature that gave the best correlations (Fig. 3). Also, the integration of the abovementioned *methylation flux rules* for models where their inclusion provided feasible flux distributions enabled the activation of some DNA demethylation reactions in those models, whereas neither a positive nor negative correlation was observed between simulated fluxes of any DNA demethylation reaction and the degree of DNA methylation (Fig. A.9 and A.10). Note that, although the number of models used was lower than in previous simulations because the cell-specific methylation ratios could only be estimated for 31 cell lines, the correlations between measured and simulated fluxes and the percentage of simulated flux values in close proximity to the real ones for exchange reactions or biomass, as well as the relative errors of biomass were as good as without the cell-specific methylation ratios and the *methylation flux rules* (Fig. A.11 versus Fig. 2).

2.4. Analysis of active pathways and protein usage in cell-line-specific models

After model reconstruction and integration with methylation data, the results from simulated flux value distributions of each cell line were analyzed. From Fig. 4-A, it was possible to observe that among the central carbon metabolism pathways, *Glycolysis or Gluconeogenesis* is the one carrying the most flux for all cell lines, while the flux through the pathways directly or indirectly to DNA (de)/methylation (*Folate metabolism*, *Cysteine and methionine metabolism* and *DNA methylation or demethylation*) is very low. However, simulations suggest that these cell lines utilize a higher mass of enzymes to activate reactions associated with *Cysteine and methionine metabolism* than those involved in *Glycolysis or Gluconeogenesis* (Fig. 4-B). Furthermore, the top five pathways with the most flux were identified as *Glycolysis or Gluconeogenesis*, *Oxidative phosphorylation*, *Purine metabolism*, *Fatty-acid biosynthesis (even-chain)*, and *Aminoacyl-tRNA biosynthesis*. With respect to protein

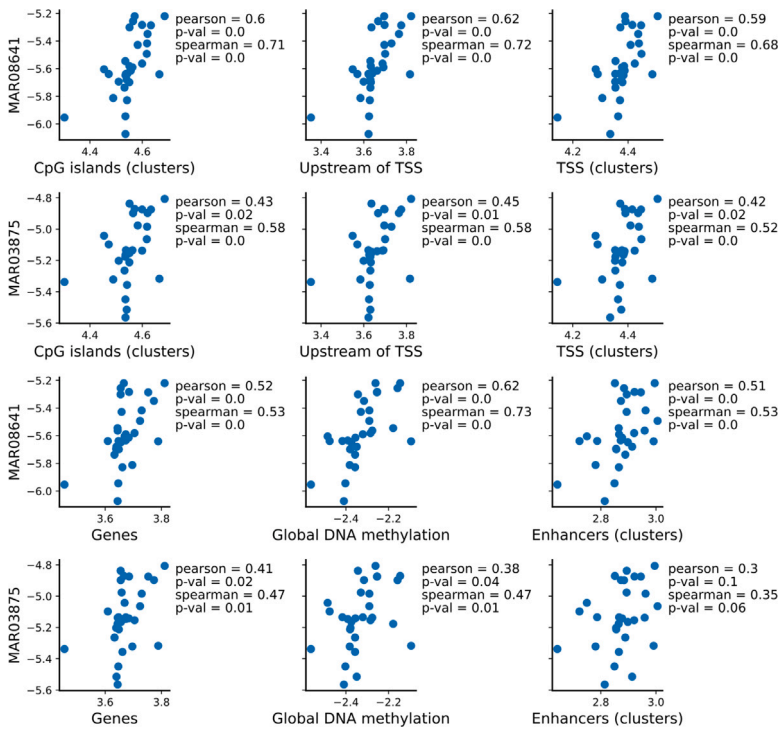

Fig. 3. Comparison of simulated fluxes of reactions involved in DNA methylation and the estimated degree of DNA methylation for models with *methylation flux rules*, as well as with cell-specific methylation ratios and demethylation tasks. Scatter plots with $\log_{10}$ values of simulated fluxes versus the experimentally estimated degree of methylation across all genome or in close proximity to different genomic features. The procedure and the description of datasets applied to estimate the level of DNA methylation is detailed in *Comparison of fluxes of reactions involved in DNA (de)/methylation and the degree of DNA methylation* section of *Materials and methods*. The vertical labels are identifiers of reactions shown in Fig. 1. Zero flux values were replaced with a very small value (1×10^{-15}) because $log_{10}(0)$ is undetermined. Only the 30 cell lines for which there was experimental data across all types of genomic intervals were used.

mass, the top five scoring pathways were *Cholesterol biosynthesis 2*, *Glycerophospholipid metabolism*, *Cholesterol biosynthesis 1 (Bloch pathway)*, *Acylglycerides metabolism*, and *Aminoacyl-tRNA biosynthesis* (Fig. 5).

2.5. Pathways and individual reactions/enzymes that influence or are influenced by overall DNA methylation

The pathways whose average flux and protein usage significantly correlated (p-value < 0.05) with overall DNA methylation levels independently of cell growth rate across the cell lines were identified to figure out which pathways could influence or be influenced by the overall DNA methylation. The pathways in which both average flux and protein usage are positively correlated with DNA methylation are *Dna (de)/methylation*, *Cysteine and methionine metabolism*, *Butanoate metabolism*, *Bile acid biosynthesis*, *Cholesterol metabolism* and *Cholesterol biosynthesis 3 (Kandustch-Russell pathway)*. The pathway in which both average flux and protein usage are negatively correlated with DNA methylation is *Beta oxidation of even-chain fatty acids (mitochondrial)*.

The specific reactions in which flux, and enzymes whose protein usage, significantly correlated (positively/negatively) with overall DNA methylation independently of growth rate were also detected and shown in tables Tab. A.9 and Tab. A.10. Most of them were involved in the metabolic pathways mentioned above. Some reactions and enzymes negatively correlated with methylation were also associated with the pathway *Fatty acid oxidation*. Among the reactions in which flux correlated positively with overall methylation, those with higher correlation coefficients and lowest p-values, apart from the DNA methylation reaction, are two transport reactions: one related to the transport of S-Adenosyl-Homocysteine (SAH), a product of DNA methylation, from the nucleus to the cytoplasm probably to be recycled by the methionine cycle, and the other is the transport of SAM from cytoplasm to the nucleus, to allow DNA methylation. The only reaction attributed to butanoate metabolism in which flux correlated positively with methylation (reaction id: MAR01434_REVNo1 in Tab. A.10) was the one converting acetyl-CoA to acetoacetyl-CoA. Also, one of the enzymes (Uniprot id: Q01581) whose protein usage is directly correlated with methylation is the one that catalyzes the condensation of acetyl-CoA with acetoacetyl-CoA to form HMG-CoA Tab. A.10, which can be converted to mevalonate, the precursor for cholesterol synthesis. This suggests that the only reason why flux and protein usage in *Butanoate metabolism* increases with methylation is that this allows the diversion of acetyl-CoA to cholesterol production.

In order to test whether the metabolic shifts observed with the increase in overall DNA methylation independently of the growth rate could be predicted using only experimental data (i.e. without using the cell-specific models), we identified a list of the metabolic reactions associated with genes whose expression was positively and significantly correlated with the overall DNA methylation, together with the corresponding metabolic pathways. A hypergeometric test was done to assess the over-representation of each metabolic pathway in that list of reactions. Note that overall DNA methylation values were first divided by corresponding experimental cell growth rates to correct for the influence of cell growth. Tab. A.11 shows that, unlike the pathway whose flux and protein usage negatively correlated with overall methylation, none of those that positively correlated with it (independently of cell growth rates) could be predicted using only experimental (methylation and gene expression) data.

2.6. Genes whose changes in methylation may affect cancer cell growth rate

Next, genes that could be potentially targeted by DNA methylases or demethylases to inhibit cancer cell growth through gene expression

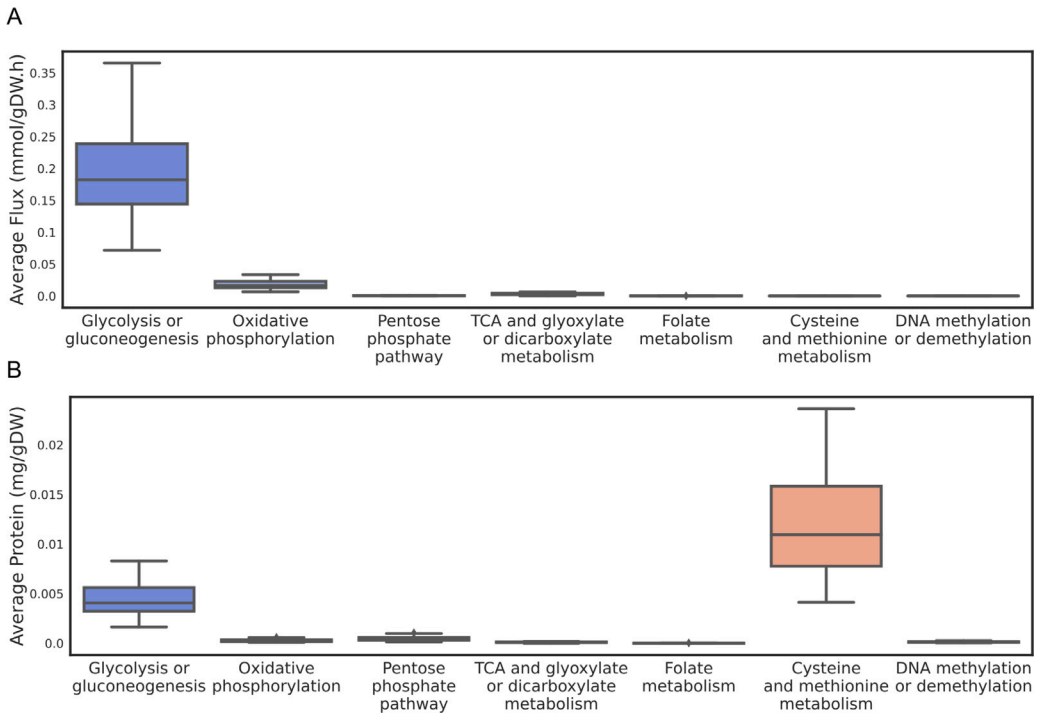

Fig. 4. Flux values and protein usage in pathways related with central carbon metabolism and DNA (de)/methylation. **A**: Boxplots show flux values across all cell lines. **B**: Boxplots show amount of protein spent across all cell lines.

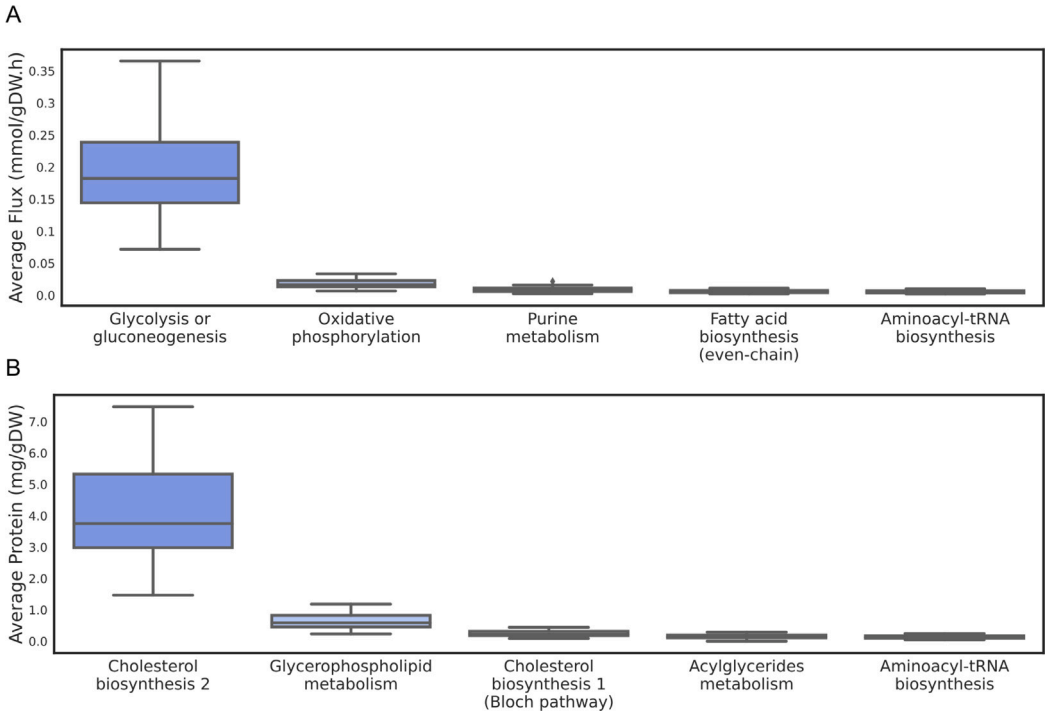

Fig. 5. Top five pathways with highest flux values or protein usage. **A**: Boxplots show flux values across all cell lines. **B**: Boxplots show amount of protein spent across all cell lines. Top five pathways were selected based on the median across all cell lines of average values of each pathway.

regulation were detected. For that, the list of genes whose promoter methylation significantly correlated (positively/negatively) with their transcription levels and with the cell growth rate across different cell lines was overlapped with the group of genes associated either with reactions whose flux or with enzymes whose protein usage was significantly (positively/negatively) correlated with cell growth rates. A diagram representing the different combinations of positive and negative correlations tested is shown in Tab. A.12. Since methylation is mostly (although not always) associated with gene expression silencing [2], and the gene promoter methylation moderately correlates with growth rate across the tested cell lines (Fig. A.12), it would be expected that genes in which promoter methylation increases would lead to a reduction in gene expression and consequently affect the flux and availability of some reactions/enzymes, indirectly enhancing the

cell growth rate. However, as the cell growth rate increased across the different cell lines, no genes could significantly enhance their promoters' methylation while simultaneously reducing both their expression and the flux or protein usage of associated metabolic reactions/enzymes (Tab. A.12 – Correlation group *B* in *Reaction flux* and *Enzyme usage* tabs).

In contrast, as the cell growth rate increased, two genes significantly enhanced their promoters' methylation while also increasing their expression and the flux of associated metabolic reactions, one of which encodes an enzyme whose protein usage kept up with the rise in growth rate: the TXNRD2 (Tab. A.12 – Correlation group *A*). In addition, one gene whose increase in promoters' methylation was inversely proportional to the rise of cell growth rate across the cell lines, had its expression and both the flux and protein usage of associated metabolic reactions/enzymes enhanced: the TPK1 (Tab. A.12 – Correlation group *C*). Interestingly, eighteen genes experienced an increase in methylation and in the flux of associated reactions with the rise of cell growth rate across the cell lines, while their expression was reduced (Tab. A.12 – Correlation group *G* in *Reaction flux* tab). Furthermore, from those, three also underwent a concerted increase in corresponding enzyme usage: GFPT2, UAP1, and SQLE (Tab. A.12 – Correlation group *G* in *Enzyme usage* tab). Note that each gene may be associated with more than one reaction and that is why the ACSF2 gene is in more than one group of genes in *Reaction flux* tab of Tab. A.12.

3. Discussion

To the best of our knowledge, this is the first time that models were built to simulate the influence of metabolism over DNA methylation. Past studies have tackled the interplay between metabolism and epigenetics, but most focused on histone acetylation [18,19,21]. Besides, the only study covering methylation [20] dealt with methylation of histones on murine cells and used the flux of the substrate for methylation (SAM) as a surrogate for methylation, instead of introducing protein methylation reactions.

In this study, we developed models for 31 human cancer cell lines that capture the influence of metabolism over DNA methylation, through the integration of reactions related to DNA methylation and demethylation, and cell-type specific DNA methylation levels.

Furthermore, past studies trying to mimic the relationship between metabolism and epigenetics utilized traditional GSMMs. When no metabolite uptake rates are known traditional constraint-based GSMMs cannot provide a finite flux distribution, because the solution is unbounded [22]. To improve the predictability of those models, it is important to constrain exchange fluxes, with real flux values [22], which are not always available. On the other hand, GECKO models do not require the definition of specific uptake rates to provide accurate flux distributions, because they introduce enzyme kinetics and concentration data that are able to reduce the flux solution space.

In the present study, even though uptake/production rates of some exometabolites were available and therefore could be used to make acceptable predictions with traditional GSMMs, the use of GECKO models provided better results than traditional GSMMs constrained with flux values of three metabolites (compare percentages between Fig. A.5 and Fig. A.3, and between Fig. 2 and Fig. A.7). Additionally, the use of cell-specific GECKO models is useful in presenting a proof-of-concept pipeline, that can be adopted in the future to other datasets for the production of models portraying the interplay of metabolism and DNA methylation in other biological contexts, for most of which such experimentally measured uptake/secretion rates may be unknown. Also, it is important to note that even though a constraint, the limitation of biomass flux with measured growth rates, had to be eventually applied to GECKO models, that process still requires gathering less experimental data than when applying measured fluxes of three different metabolites (as in constrained traditional models).

In this study, the stoichiometric coefficients of modified methylcytosines in the reaction representing the total DNA composition were adapted in accordance with experimentally measured DNA methylation levels specific to each cell line, so that the simulations could predict metabolic phenotypes associated with DNA methylation that were cell-type specific. Naturally, one of the consequences of this adaptation was the increase in correlation between fluxes of the reactions of DNA methylation and of the production of SAM and the experimentally observed DNA methylation levels. Notably, those correlations were higher in the region *Upstream of TSS* of genes (i.e. gene promoters) than in other genomic regions, including regions surrounding upstream and downstream of the TSS (the *TSS (clusters)*), suggesting that the variation in global DNA methylation across these cell lines can be in its majority attributed to alteration of methylation in gene promoters. Ghandi et al. [27] observed a negative correlation between gene expression and promoter methylation of many genes in these cell lines. Besides, another study with the same cell lines reported that the most significant correlation between DNA methylation and gene expression was an inverse correlation for epithelial and mesenchymal genes, more expressive for the former than the latter (−0.639 versus −0.525 correlation), and although to a smaller degree, with tumor suppressors as well [28]. This suggests that the increase in promoter methylation across different cell lines is mostly associated with endothelial-to-mesenchymal transition and silencing of tumor suppressor genes, and therefore, it would be expected that the proliferation ability of these cell lines to increase with the gene promoter methylation levels. To test this, we compared the growth rate of the cell lines and their promoter methylation levels and, in fact, a significant but moderate correlation could be observed in Fig. A.12.

Another important observation was that the inclusion of the aforementioned *methylation flux rules* based on reported values of ratios of DNA (de)/methylation reaction rates in the model simulations guaranteed flux through the DNA demethylation reactions at least in some models. This is relevant because it is known that there is a dynamic DNA methylation turnover steady-state, where both methylation and demethylation reactions are active [24]. Nevertheless, the *methylation flux rules* could only be applied to some of the cell lines without affecting the feasibility of the flux distribution, maybe because the metabolite pools (for e.g. protein pools) are not adapted to the cell type or because those flux rules are general rules that might not apply to all cell lines. In that sense, future *in silico* studies could try to integrate more cell-detailed experimentally determined flux rules as soon as they become available. The main reason the DNA demethylation reactions are not active without setting flux rules is that the model can produce demethylated DNA using less overall protein through the DNA polymerization reaction than by using the DNA demethylation reactions. And that might be happening just because there is a lack of information regarding reactions that happen inside the nucleus and that produce components of the biomass. In fact, the nucleus of the traditional generic model including the DNA (de)/methylation reactions is the organelle with less amount of reactions and metabolites: 218 and 142 respectively. Note that the model compartment *inner mitochondria* has less, but is part of the organelle mitochondria, that has much more. Besides, for some (de)/methylation reactions not to be blocked, there was a need to add reactions created based on assumptions (mainly transport reactions), which indicates that the biological knowledge is still missing information regarding reactions that happen in the nucleus.

The final models here created, which represented 31 different cell lines, provided a high percentage of simulated flux values in close proximity to corresponding experimental values (around 77% in Fig. A.11) for exchange reactions of 26 metabolites, which in itself serves as a good validation criteria. Furthermore, the flux and protein mass distribution across metabolic pathways agreed with the reported experimental evidence.

Glycolysis or Gluconeogenesis was the subsystem among the central carbon metabolism pathways to carry more flux, and it was the most active metabolic pathway. This is expected because aerobic glycolysis is one of the characteristics of cancer cells, as it allows them to quickly obtain energy and elementary metabolites for fast growth [29,30].

Although Glycolysis is more active than *Oxidative phosphorylation* across the models of all cell lines, as expected in cancer cells, the latter is still the second pathway with the most flux. A possible explanation for this is that, even though cancer cells prefer in general aerobic glycolysis, they utilize Oxidative Phosphorylation (OXPHOS) to produce at least some level of Reactive Oxygen Species (ROS). This is because a moderate amount of ROS is beneficial for tumorigenesis, resistance to chemotherapy and cancer progression [31]. Another putative reason could be the increased levels of citrate in cancer cells, due to the conversion of α-KG to citrate induced by oncogenes, which indirectly would allow the production of some energy through the *Oxidative phosphorylation* by feeding the Tricarboxylic acid cycle (TCA cycle) [30]. However, in that case, it would be expected that the TCA cycle was also among the most activated pathways, which is not the case.

In addition, the fact that *Purine metabolism*, *Fatty acid biosynthesis* and *Aminoacyl-tRNA biosynthesis* are the third, fourth and fifth pathways with the most flux in the models could also be anticipated because the first is necessary for the synthesis of nucleotides fundamental for the production of new DNA molecules, the second is essential for the formation of new cell membranes and the last allows protein production through mRNA translation, all of which are important factors for fast-dividing cells.

The first and third pathways to use most protein mass are related to *Cholesterol biosynthesis*, which is also in accordance with the literature. Cholesterol biosynthesis is often enhanced in cancer as on the one hand, it can activate mTORC1 signaling, which in turn promotes cell proliferation, invasion, and metastasis, while on the other hand it alters lipid rafts composition, promoting the loss of integrin-mediated cell adhesion, and consequently contributing to cancer aggressiveness [32,33]. The second and fourth pathways to use most protein mass, the *Glycerophospholipid metabolism* and *Acylglycerides metabolism*, are expected to be activated in cancer cells as well, because those metabolites are part of new cell membranes needed for intense cell proliferation.

With respect to pathways directly or indirectly related to DNA (de)/methylation, the flux is reduced. This is expected, since it is possible to anticipate that a small DNA methylation rate is enough to methylate less than 1% of the genome (the average percentage of methylation of the human genome). Nevertheless, the use of protein mass in one of those pathways, *Cysteine and methionine* metabolism is elevated, even more than in glycolysis, suggesting that although holding a small amount of flux, it is an important pathway.

Cysteine and methionine metabolism, *Bile acid biosynthesis*, *Butanoate metabolism* and subsystems related to cholesterol metabolism (including biosynthesis of cholesterol) were pathways predicted to carry more flux and use more protein on average, while *Beta oxidation of even-chain fatty acids (mitochondrial)* predicted to carry less flux and use less protein, as overall DNA methylation increases independently of cell growth rate. Besides, the flux of several individual reactions and protein usage of many individual enzymes of the *Fatty acid oxidation* pathway were inversely correlated with overall methylation independently of cell growth rate. Although the decrease in fatty acid oxidation could be anticipated without the use of models (based solely on experimental gene expression and overall DNA methylation), given the negative correlation of the gene expression in the pathways of fatty acid oxidation with the overall DNA methylation (independently of cell growth rates), the increase in activity of pathways positively correlated with methylation could only be predicted while utilizing the models, since the gene expression of such pathways did not correlate positively with overall methylation levels.

It is expected for *Cysteine and methionine metabolism* to increase with overall DNA methylation level, as the production of cysteine from serine and uptake of methionine can increase the activity of the one carbon-cycle (folate and methionine cycles) producing more SAM, that is used for methylation. In fact, even the observed increase of flux through the individual reactions transporting SAM and SAH respectively in and out of the nucleus could be anticipated. Moreover, Yokogami et al. previously reported that methionine depletion in glioma-initiating cells leads to a decrease in cholesterol biosynthesis by a decrease in the expression of SREBP2 and ACA43, while promoting fatty acid oxidation through an increase in the expression of methyltransferase METTL20 that methylates the electron carrier coupling mitochondrial fatty acid oxidation to respiration (flavoprotein subunit beta) and in that way serves as a positive regulator of beta-oxidation of fatty acids [34]. This suggests that an increase in methionine metabolism accompanying the overall increase in DNA methylation, as observed across the cell line-specific models, could enhance the cholesterol biosynthesis on one side and decrease beta-oxidation of fatty acids on the other. This is in accordance with our results as cholesterol metabolism and bile acid synthesis, which uses cholesterol as substrate, increase in activity, while *Beta oxidation of even-chain fatty acids (mitochondrial)* decrease with the rise in methionine metabolism (i.e. overall methylation) across the different cell lines. Nevertheless, the observed effect, in this case, does not seem to be mediated by changes in the expression of genes mentioned in the abovementioned publication (SREBF2, METTL20, METTL16) [34], since those were not in the list of genes which expression correlates with overall methylation independently of biomass. The involvement of ACA43 (i.e. SNORA17B) could not be tested as that gene was not present in the gene expression data.

As explained in the results section, the increase in flux and protein usage in Butanoate metabolism with overall methylation (independently of growth rate) is caused by the increase in the conversion of acetyl-CoA to acetoacetyl-CoA. Acetoacetyl-CoA is then used to produce HMG-CoA a precursor of the mevalonate pathway. Besides, acetoacetyl-CoA could not be produced instead from butyric acid/butanoate (a short-chain fatty acid resulting from gut bacteria fermentation and usually obtained by human cells [35,36]), as the models do not allow the uptake of compounds other than the pre-defined media components. Therefore, the observed increase in Butanoate metabolism is most probably a strategy to increase cholesterol synthesis through an increase in the consumption of acetyl-CoA.

This study identified five genes whose changes in methylation may potentially affect cancer cell growth rate. The methylation of TXNRD2 promoters significantly increased with its expression across the cell lines, concurrently with a raise in the flux and protein usage of associated reactions/enzymes and a boost in the cell growth rate. TXNRD2 encodes thioredoxin reductase 2, an antioxidant enzyme that controls the levels of ROS. It has been reported before that the treatment of triple-negative breast cancer cells with a cancer drug reduced TXNRD2 expression through an epigenetic mechanism (an increase in histone acetylation) resulting in ROS accumulation and cell senescence [37]. This suggests that expression of TXNRD2 in those cancer cells is needed to prevent cell death induced by ROS accumulation, and that could serve as an explanation for the TXNRD2 gene expression and enzyme usage/reaction flux association with higher cell growth rates in this analysis. There have been publications stating a negative correlation between gene expression and methylation of TXNRD2 in different biological contexts [38–40], and one article mentioning both the hypermethylation and hypomethylation of that gene in a cancer cell line respectively at 24 and 48 h after treatment with resveratrol [41]. Nonetheless, to the best of our knowledge, no association has been reported before between hypermethylation of TXNRD2 gene, its increase in expression, and the increase in cancer aggressiveness/progression (which can be understood as a surrogate for an increase in cancer cell growth rate). Such a relationship would therefore be an interesting research subject of experimental prospective studies.

TPK1 was the gene whose promoter's methylation significantly decreased with the rise in cell growth rate while its gene expression and the flux and protein usage of associated reactions/enzymes also increased. TPK1 is a gene that encodes the enzyme thiamine pyrophosphokinase 1, which catalyzes the phosphorylation of the essential metabolite thiamine (i.e. vitamin B1) into thiamine pyrophosphate (TPP). In turn, TPP acts as a cofactor of metabolic enzymes such as pyruvate dehydrogenase (producing acetyl-CoA from pyruvate and subsequently promoting the activation of TCA cycle), α-ketoglutarate dehydrogenase (catalyzing α-ketoglutarate to succinyl-CoA conversion in TCA cycle), and transketolase (an enzyme of the PPP pathway involved in the production of ribose-5-phosphate, a precursor of nucleotide synthesis) [42]. Besides its role in energy and nucleotide production, TPP may also act as an intracellular antioxidant, preventing the accumulation of ROS during stress conditions. Therefore, the increase in thiamine uptake and upregulation TPK1 during hypoxic and oxidative stress are characteristic of some cancer cells [42]. Furthermore, thiamine supplementation and TPP have been associated with cancer cell survival, proliferation, and resistance to chemotherapy [43]. This fact is in accordance with the correlation between the observed increase in cell growth rate and the rise in the expression of TPK1 and of the flux and protein usage. To the best of our knowledge, no connection was previously established between demethylation of TPK1 gene and higher cancer cell growth.

There are three genes whose promoter's methylation increases with the cell growth rate while their transcription reduces despite the rise of simulated flux and protein usage of corresponding reactions/enzyme. Different biological processes could explain an inconsistency between gene expression reduction and the increase of corresponding reaction flux and enzyme usage *in vivo*. One is that the enzymes could previously be produced and stored *in vivo*, and continue to be used afterward even if the transcription of the genes encoding those enzymes decreases. Another explanation could be that post-translation modifications (PTMs) are increasing the activity of those enzymes and accelerating the corresponding reactions, but as the models do not reflect changes in catalytic activity through PTMs the increase in activity translates into more protein usage in the models' simulations. This suggests that only the enzymes could be potential targets for cancer treatment and that epigenetic remodeling through targeted gene hypo/hypermethylation would not be effective.

The main disadvantage of these models is that they only mirror active DNA demethylation mechanisms, whereas passive DNA demethylation, i.e. the dilution of DNA methylation signal as a result of fast cellular division in the absence of functional DNA methylation maintenance machinery, is not taken into account [44].

Another limitation is the use of a generic biomass composition reaction, which can reduce the accuracy of predictions with cell line-specific models, since it is known that different cell types have distinct biomass compositions, which can even change according with environmental conditions [45,46]. Therefore, prospective investigations should integrate experimentally measured cell line-specific biomass composition values.

In addition, the models created in the present study can only model the impact of metabolism on global DNA methylation, while it is known that the DNA methylation also influences metabolism, directly through the regulation of transcription of genes encoding metabolic enzymes and indirectly by affecting the expression of genes associated with signaling proteins. Furthermore, DNA methylation is strongly swayed by other regulatory mechanisms, such as other epigenetic modifications [47]. Therefore, it will be important to integrate regulatory and signaling networks into future metabolic models of cancer cells with epigenetic machinery.

4. Conclusion

In this study, GECKO models that integrate the DNA methylation and demethylation machinery for 31 of the NCI-60 human cancer cell lines were built. Those models were validated by comparison of simulated fluxes of 26 metabolites with experimentally measured uptake/production rates. The flux and protein mass distribution observed across different metabolic pathways was in line with what was reported in the literature for cancer cells. Cell-line-specific DNA methylation levels estimated based on experimental data were introduced in the models to understand how metabolism affects DNA methylation. In fact, with these models it was possible to identify metabolic pathways and specific reactions/enzymes in which variation in flux or protein usage either follows or opposes the change in gene promoter methylation levels across the different cell lines. Furthermore, we were able to identify two genes whose changes in methylation seem to affect cancer cell growth rate: TXNRD2 and TPK1, and therefore might be good candidates to study their potential role as targets of cancer epigenetic drugs in an *in vitro* setting.

Furthermore, the same generic model and reconstruction pipeline validated in this study could be applied to other datasets. For example, to understand the interplay between metabolism and DNA methylation of cancer cells exposed to epigenetic remodeling compounds, or even to study the effect of metabolic diseases, like obesity, insulin resistance, or dyslipidemia in DNA methylation [48]. The models created present certain limitations, some of which might be solved in the future when information regarding cell line-specific methylation flux rules or the expansion of knowledge of the metabolic network in the cell nucleus becomes available.

5. Materials and methods

5.1. Creation of the generic DNA methylation model

Since, to the best of our knowledge, no human generic model available contained the reactions needed to simulate both the DNA methylation and demethylation, alterations were made to the model Human1 (version 1.12) to create a generic model depicting those processes, which was made available to the community.

Overall, the reaction of DNA methylation in the cytoplasm was removed (as there is no DNA in the cytoplasm), and the gene rule of the equivalent reaction in the nucleus was updated (explanation in Tab. A.4). Reactions and corresponding GPR rules involved in DNA (de)/methylation were obtained by literature curation (see explanation in Tab. A.1 and Tab. A.4). Some transport reactions were added, and two reactions that occur when cytosine is inside the DNA were assumed to take place also when it is in its monomeric form (*consdirect5fC* and *consdirect5CaC* reactions in Tab. A.1), to guarantee flux through the remaining DNA (de)/methylation reactions (i.e. to prevent their blockage). Metabolites taking part in the added reactions are described in Tab. A.5.

A pseudo-reaction representing the average total human DNA composition in the nucleus (*prodDNAtot*) in terms of DNA cytosine (de)/methylation marks, like DNA-5-methylcytosine (DNA5mC), DNA-5-hydroxymethylcytosine (DNA5hmC) and DNA-5-formylcytosine (DNA5fC) was introduced (Tab. A.2 and Tab. A.3 show how the composition was determined), so that subsequently generated context-specific models could reflect cell line-specific DNA (de)/methylation levels.

Moreover, the generic biomass reaction was replaced by a similar reaction (*adaptbiomass* reaction) where the DNA was changed into the pseudo-metabolite (DNAtot) representing the total DNA harboring all DNA methylation and demethylation marks (in Tab. A.1).

All introduced reactions were corrected for charge and mass balance (Tab. A.6). All blocked reactions and associated genes and metabolites were removed.

5.2. Reconstruction of cell line-specific traditional GSMMs

Cell line-specific GSMMs were built for different NCI-60 cell lines through the integration of transcriptomics data from Cancer Cell Line Encyclopedia (CCLE) deposited in DepMap repository in 2019. This version of the transcriptomics dataset was chosen because it was produced by the same study, Ghandi et al. [27], which the Reduced-Representation Bisulfite Sequencing (RRBS) dataset used in the present work was retrieved from. The approach presented in Richelle et al. [23] to build cell type-specific models was implemented in Python and made available to the public. Gene scores were determined from cell-specific gene expression data using the following expression:

$$\text{gene score} = 5 * log\left(1 + \frac{expression\ level}{threshold}\right) \quad (1)$$

where the threshold is the mean value of gene expression over all samples unless it is lower than the 25th or higher than the 75th percentiles of the gene expression value distribution, in which cases the threshold value is considered to be the same as the mentioned percentiles.

Reaction scores were subsequently calculated from gene scores taking the GPR rules into account so that: the score of a reaction catalyzed by an enzyme complex was the minimum score of all genes associated with the complex (*AND* rule) and that of a reaction catalyzed by isozymes was the maximum score of all genes encoding the isozymes (*OR* rule). The highest reaction scores of a cell line were attributed to the reactions considered necessary for the generic metabolic tasks (a.k.a. essential metabolic tasks), while the same procedure was applied to necessary reactions of other tasks if those are performed in that specific cell type. This was done to give the reconstruction algorithm a higher probability of building a cell type-specific model that can pass all the generic tasks and tasks specific to that cell type. A reaction was considered necessary for a task if it was carrying flux upon the inclusion of the task-associated flux constraints on the generic model followed by a minimization of the sum of all fluxes. In order to determine whether a task was done in a certain cell type, a metabolic score consisting of the average of the scores of the reactions previously identified as required for that task was calculated. When the task metabolic score was higher than $5 * log(2)$ (the gene/reaction score to which the expression level is equal to the aforementioned threshold) the task was considered to be done in that specific cell type. Then, the FASTCORE algorithm from Troppo package [49] was run to obtain the cell type-specific models. Note that three DNA demethylation tasks (each one corresponding to a distinct demethylation pathway) were created and added to the original list of tissue-specific tasks so that each final cell-line model could have all reactions necessary for the DNA demethylation pathway done by that specific cell type. For comparison purposes, an equivalent analysis was done without including the reactions necessary for cell type-specific tasks. In both types of analyses, with or without cell-line-specific tasks, we also included two DNA demethylation-associated reactions (*consdirectDNA5fC* and *consdirectDNA5CaC* in Fig. 1) that were non-necessary for enzyme-catalyzed DNA demethylation to occur (demethylation could happen in the generic model without them) because they were not associated with any gene (were not catalyzed by an enzyme), and therefore were always excluded from the reconstructed models (due to the lack of associated reactions scores) although they could happen without the presence of enzymes.

The second approach to building context-specific models consisted of the application of a version of the tINIT algorithm run in MATLAB that had already been implemented by Robinson et al. [22]. That version tries to include reactions with scores above a threshold while removing those below the threshold and keeping the connectivity of the model (i.e. making sure all reactions carry flux). The gene-to-reaction scores conversion strategy applied was again the minimum and maximum of gene scores for complexes and isozymes, respectively [22]. The final models also kept the ability to perform the generic metabolic tasks, which consisted of the previously reported 57 essential metabolic tasks [22]. The two non-catalyzed DNA demethylation-associated reactions *consdirectDNA5fC* and *consdirectDNA5CaC* were also introduced in all models due to the reasons explained above for Richelle's approach. When testing the inclusion of cell-line-specific tasks with this approach, the reactions needed for cell-specific tasks were included after reconstruction if the task score was above the threshold for a particular cell line. Note that in one of the analyses, to guarantee that demethylation could happen, the inclusion of reactions needed only for cell-specific DNA demethylation tasks together with *methylation flux rules* but without including the other cell-specific tasks, was also assessed.

5.3. Generation of cell line-specific GECKO models from traditional GSMMs

GECKO models were created from traditional GSMMs using a MATLAB script produced by Robinson et al. [22], whose pipeline was first described in Sanchez et al. [17]. In that pipeline, enzymes are introduced as pseudo-substrates in the reactions they catalyze and the stoichiometric coefficients are the inverse of the turnover numbers of the corresponding enzyme-metabolic substrate pairs. Reversible reactions are split into two irreversible reactions in opposite directions, and isozymes are separated into different reactions, each catalyzed by one of the isozymes. Furthermore, for each original un-split isozyme-associated reaction, a new pseudo-reaction, named *arm* reaction, is added. The only product of an *arm* reaction is an intermediary pseudo-metabolite which is used as a substrate by each of the isozyme-split reactions so that the flux bounds of each original un-split reaction can still be applied. For reactions catalyzed by complexes, each enzyme of the complex is introduced as a substrate and the stoichiometric coefficient is in that case the product of the inverse of its turnover number and the stoichiometric coefficient of the enzyme inside the complex.

Finally, supply reactions for each enzyme known as protein *draw* reactions, are added, where each reaction consumes a proportion (based on the enzyme molecular weight) of a total protein pool, which in turn is supplied by another included boundary reaction called *protein pool exchange* reaction [17,22].

5.4. Detection and validation of the best reconstruction and simulation pipelines

For the assessment of the best type of models (traditional GSMMs or GECKOs) and the selection of the most suitable model reconstruction and simulation strategies, values of simulated uptake/secretion rates of 26 metabolites across different cell lines were compared with corresponding experimentally measured ones originally obtained from Jain et al. [50]. The 26 metabolites chosen were the ones previously utilized to validate the reconstruction of eleven NCI-60 cell lines in Robinson et al. [22] article. The same comparison was made between the simulated fluxes of biomass reaction and measured rates of cell growth retrieved from Zielinski et al. [51]. The strategies and model types giving the best percentage of simulated flux values in close proximity to measured ones were identified. For simulations with traditional GSMMs, parsimonious Flux Balance Analyses (pFBAs) were carried out whereby the minimization of the sum of all fluxes took place after constraining the biomass flux with either, the objective value of an FBA whose metabolic objective was maximization of biomass flux or with experimentally measured growth rates. For GECKO models, the simulations were accomplished through either an enzymatic pFBA where the minimization of the flux of the total protein uptake reaction followed the maximization of the flux of biomass reaction, or by an FBA whose metabolic objective was to minimize the total protein uptake upon limitation of the flux of biomass reaction with measured growth rates. The last strategy was selected to be applied to all subsequent simulations, as it was the one to give the best results. Model manipulation and simulation were done with the *MEWpy* [52] and *COBRApy* [53] python modules.

5.5. Calculation of the composition of total DNA

The overall composition of the total DNA in terms of modified-cytosine species involved in methylation and demethylation was initially estimated for a generic human cell based on general knowledge of the human genome (see Tab. A.2), and the average level of DNA5hmC sites across different healthy human tissues (Tab. A.3) obtained from a Chemical-assistant C-to-T conversion of 5hmC sequencing (hmC-CATCH) experiment, whose values were kindly provided by the authors of He et al. [54]. Calculations made were similar to the ones of stoichiometric coefficients of reactions for synthesis of macromolecules (e.g. DNA, proteins) that are consumed in lumped-biomass reactions [55].

Those estimations were initially integrated into the total DNA composition reaction, the *prodDNAtot* reaction, of the generic model before the reconstruction of the context-specific models. However, simulations with the cell-line-specific models have later been performed with a cell-line-specific *prodDNAtot* reaction. The stoichiometric coefficient of the DNA5hmC in *prodDNAtot* reaction of a specific cell line was the estimated ratio of DNA with 5hmCs of the healthy tissue to which that cell line corresponds (calculation shown in Tab. A.3). The estimation of the stoichiometric coefficients of the remaining cytosine species was grounded on the results of a Reduced-Representation Bisulfite Sequencing (RRBS) experiment obtained from the same study that produced the transcriptomics data used in the model reconstruction [27]. The output of bisulfite sequencing is the proportion of all cytosines that have remained unconverted (i.e. were not converted to Uracil) during the bisulfite treatment, and it is generally used as a proxy for the ratio of cytosines that are methylated. However, in reality, not only DNA5mCs but also DNA5hmCs are not converted to Uracil [56] upon treatment with bisulfite, while aside from the fully unmethylated cytosines also the DNA5fCs are converted to Uracil [57] during the process. So, the assumption that the bisulfite sequencing signal is the ratio of DNA that is methylated could lead to imprecise estimations, as in reality, it represents the ratio (DNA5mC + DNA5hmC)/(unmethylated DNA-5-cytosine + DNA5fC). Fortunately, in this case, the ratio of DNA with DNA5hmCs and DNA5fCs could be calculated (from the hmC-CATCH results and literature, respectively), and therefore, there was no need to use the bisulfite sequencing signal directly as a proxy (calculations shown in formulas of excel Tab. A.7).

5.6. Comparison of fluxes of reactions involved in DNA methylation and the degree of DNA methylation

The correlation between simulated fluxes of important reactions involved in DNA methylation and demethylation and the overall level of DNA methylation across different cell lines (represented by the stoichiometric coefficients of DNA5mC calculated above) was assessed in this study. In addition to the global methylation state, the comparison to the methylation levels of CpGs at specific genomic regions was also analyzed. For this, the RRBS signal of CpGs within 1000 bp-length genomic intervals *Upstream of the TSS* of genes (gene promoters), retrieved from the same abovementioned RRBS study [27], was added across all genomic intervals for each cell line. Note that each genomic interval had at least one cell line for which the value was not available (i.e. with NA). So, instead of excluding all genomic intervals with at least one non-available value, the non-available values were replaced by the median of the genomic interval across all cell lines. The above-mentioned sum of the signal was then directly employed as a proxy for the DNA methylation level, in contrast to the procedure applied above in the calculation of stoichiometric coefficients, because in this case, the number of CpGs containing DNA5hmC and DNA5fC within the particular genomic regions was unknown.

This same analysis strategy was further applied to other datasets of the same study where the level of methylation in methylation clusters (i.e. regions where CpG sites have similar methylation changes across different cell lines) was retrieved for genomic intervals centered around TSS (from 3000 bp upstream to 2000 bp downstream), CpG islands, and enhancers (from 2000 bp upstream to 2000 bp downstream).

Furthermore, the methylation level of genes was also roughly determined by adding the methylation values of all genes in each cell line from a dataset deposited at CellMiner [58] database which resulted from a DNA methylation array experiment [28]. In that case, the average gene methylation values (i.e. average of beta values) were given by the ratio of the intensity of the probes for methylated DNA and the intensity of all probes (those detecting methylated and unmethylated DNA) annotated to that gene.

5.7. Analysis of active pathways and protein usage

To compare the simulated flux distributions across the different metabolic pathways and cell lines, a generic GECKO model was first created from the generic traditional GSMM, where each reaction was associated with a metabolic subsystem. The fluxes of all reactions of the same metabolic subsystem in each cell-line-specific model were added and divided by the number of reactions attributed to that subsystem in the generic GECKO model, to correct for the bias that subsystems with more reactions have a higher chance to have more active reactions (and therefore carry more flux). Since each isozyme-associated reaction of a traditional GSMM is split into different reactions in a GECKO model (each associated with one of the isozymes) that consume the same pseudo-metabolite of an *arm* reaction (mentioned above), the flux of the *arm* reaction is the sum of the fluxes of the other split-reactions. Therefore, for reactions associated with isozymes only the *arm* reactions were considered in the analysis. Also, the abovementioned protein *draw* reactions and the *protein pool exchange* reaction were naturally excluded, as they were not associated with any subsystem.

For the estimation of protein usage in each metabolic subsystem, instead of directly quantifying the amount of enzyme spent in each *draw* reaction, the amount of protein used in each enzyme-reaction combination was calculated instead, because the same enzyme can participate in different reactions of distinct metabolic subsystems. The flux of each reaction (in $mmol.gDW^{-1}.h^{-1}$) was divided by the k_{cat} of each enzyme-reaction combination (in h^{-1}), and then multiplied by the molecular weight of the enzyme (in KDa, i.e. 1 $g.mmol^{-1}$) and 1000, to obtain the amount of the enzyme used in the reaction (in mg/gDW). All reactions that do not use any enzyme as a pseudo-substrate (the *arm* reactions and non-catalyzed reactions) were excluded. Then, the sum of all protein usage values of each metabolic subsystem was divided by the number of enzyme-reaction combinations attributed to that subsystem in the generic GECKO model to correct for the bias that subsystems with more reactions and with reactions containing more enzymes have the tendency to use more protein.

5.8. Identify pathways and individual reactions/enzymes that influence or are influenced by overall DNA methylation

In order to obtain the metabolic pathways that correlated with DNA methylation independently of cell growth rate across the different cell lines, each reaction's flux or protein usage was first divided by the simulated cell growth rate and the sum of flux or protein usage values of all reactions of each subsystem was divided by the number of reactions of that subsystem in the generic model. This average flux or protein usage value was calculated for each metabolic subsystem of each cell line. Then, those subsystems were identified in which flux or protein usage significantly correlated (p-value < 0.05) with the global DNA methylation level across the different cell lines. This was done with both the Spearman and Pearson correlation. The subsystems that correlated positively with global DNA methylation levels (with correlation coefficients > 0) in both types of correlation coefficients were selected and sorted based on their Pearson correlation coefficient.

The same was done with those negatively correlated (correlation coefficients < 0). The individual reactions whose flux and the enzymes whose protein usage significantly and positively or negatively correlated with overall DNA methylation level were also identified.

The genes whose transcription was positively or negatively and significantly (p-value < 0.05) correlated with the ratio between overall DNA methylation level and experimental cell growth rate were identified in this study. Then, the metabolic reactions and corresponding pathways/subsystems associated with those genes were listed using the generic model's gene-protein-reaction rules. A hypergeometric test was applied to determine the over-representation of each metabolic pathway in that list of reactions, followed by a multiple test correction of Benjamin-Hoersch. The pathways/subsystems that correlated significantly in both types of correlation coefficients were selected and sorted based on the adjusted p-value of the hypergeometric test. This was done to assess whether the metabolic shifts observed with the increase in overall DNA methylation independently of cell growth could be predicted only by utilizing experimental data instead of the metabolic models.

5.9. Identify genes whose changes in methylation may affect cancer cell growth rate

To identify genes whose changes in methylation levels may affect cancer growth rate through changes in gene expression regulation, a list of genes whose promoter methylation (1000 bp upstream of TSS) significantly correlated (either positively/negatively) with its transcription and with the cell growth rate across the different cell lines was intersected with either the group of genes associated with reactions whose flux, or the genes associated with enzymes whose protein usage, significantly correlated (either positively/negatively) with the cell growth rate. Entities (genes/reactions/enzymes) were considered significantly correlated when the p-value was below 0.5 with both the Spearman and Pearson correlation methods.

CRediT authorship contribution statement

Tânia Barata: Conceptualization, Formal analysis, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Vítor Pereira:** Methodology, Supervision. **Ricardo Pires das Neves:** Conceptualization, Supervision, Writing – review & editing. **Miguel Rocha:** Conceptualization, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare no conflict of interests.

Data availability

Data, graphs produced and adapted MatLab code are deposited at https://doi.org/10.5281/zenodo.10031423. Python code and analysis description is at https://github.com/BioSystemsUM/epigen.

Acknowledgments

This work was supported by *FCT – Fundação para a Ciência e a Tecnologia*, Portugal under projects *UIDB/04539/2020* and *UIDP/04539/2020* (Strategic Plan CIBB). We also acknowledge the Ph.D. studentship with reference *SFRH/BD/ 123028/2016* funded by *FCT*, Portugal.

Appendix A. Supplementary material

Supplementary materials can be found online at https://doi.org/10.5281/zenodo.1003142.

References

[1] T. Ugai, N. Sasamoto, H.-Y. Lee, M. Ando, M. Song, R.M. Tamimi, I. Kawachi, P.T. Campbell, E.L. Giovannucci, E. Weiderpass, T.R. Rebbeck, S. Ogino, Is early-onset cancer an emerging global epidemic? Current evidence and future implications, Nat. Rev. Clin. Oncol. 19 (2022) 656–673, http://dx.doi.org/10.1038/s41571-022-00672-8.

[2] P. Saggese, A. Sellitto, C.A. Martinez, G. Giurato, G. Nassa, F. Rizzo, R. Tarallo, C. Scafoglio, Metabolic regulation of epigenetic modifications and cell differentiation in cancer, Cancers 12 (2020) 1–24, http://dx.doi.org/10.3390/cancers12123788.

[3] H.A. Coller, Is cancer a metabolic disease? Am. J. Pathol. 184 (2014) 4–17, http://dx.doi.org/10.1016/j.ajpath.2013.07.035.

[4] J. Peixoto, J. Lima, Metabolic traits of cancer stem cells, Dis. Models Mech. 11 (2018) 1–13, http://dx.doi.org/10.1242/dmm.033464.

[5] S. Fernández-Arroyo, E. Cuyàs, J. Bosch-Barrera, T. Alarcón, J. Joven, J.A. Menendez, Activation of the methylation cycle in cells reprogrammed into a stem cell-like state, Oncoscience 2 (2016) 958–967, http://dx.doi.org/10.18632/oncoscience.280.

[6] A.M. Intlekofer, L.W.S. Finley, Metabolic signatures of cancer cells and stem cells, Nat. Metabol. 1 (2019) 177–188, http://dx.doi.org/10.1038/s42255-019-0032-0.

[7] C. Gu, G.B. Kim, W.J. Kim, H.U. Kim, S.Y. Lee, Current status and applications of genome-scale metabolic models, Genome Biol. 20 (2019) 1–18, http://dx.doi.org/10.1186/s13059-019-1730-3.

[8] S. Volkova, M.R. Matos, M. Mattanovich, I.M. de Mas, Metabolic modelling as a framework for metabolomics data integration and analysis, Metabolites 10 (2020) 1–27, http://dx.doi.org/10.3390/metabo10080303.

[9] L. Väremo, C. Scheele, C. Broholm, A. Mardinoglu, C. Kampf, A. Asplund, I. Nookaew, M. Uhlén, B.K. Pedersen, J. Nielsen, Proteome- and transcriptome-driven reconstruction of the human myocyte metabolic network and its use for identification of markers for diabetes, Cell Rep. 11 (2015) 921–933, http://dx.doi.org/10.1016/j.celrep.2015.04.010.

[10] A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, I. Nookaew, P. Jacobson, A.J. Walley, P. Froguel, L.M. Carlsson, M. Uhlen, J. Nielsen, Integration of clinical data with a genome-scale metabolic model of the human adipocyte, Mol. Syst. Biol. 9 (2013) 1–16, http://dx.doi.org/10.1038/msb.2013.5.

[11] S. Aller, A. Scott, M. Sarkar-Tyson, O.S. Soyer, Integrated human-virus metabolic stoichiometric modelling predicts host-based antiviral targets against Chikungunya, Dengue and Zika viruses, J. R. Soc. Interface 15 (2018) 1–12, http://dx.doi.org/10.1098/rsif.2018.0125.

[12] A. Mardinoglu, R. Agren, C. Kampf, A. Asplund, M. Uhlen, J. Nielsen, Genome-scale metabolic modelling of hepatocytes reveals serine deficiency in patients with non-alcoholic fatty liver disease, Nature Commun. 5 (2014) 1–11, http://dx.doi.org/10.1038/ncomms4083.

[13] S. McGarrity, Ó. Anuforo, H. Halldórsson, A. Bergmann, S. Halldórsson, S. Palsson, H.H. Henriksen, P.I. Johansson, Ó. Rolfsson, Metabolic systems analysis of LPS induced endothelial dysfunction applied to sepsis patient stratification, Sci. Rep. 8 (2018) 1–14, http://dx.doi.org/10.1038/s41598-018-25015-5.

[14] R. Agren, S. Bordel, A. Mardinoglu, N. Pornputtapong, I. Nookaew, J. Nielsen, Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT, PLoS Comput. Biol. 8 (2012) 1–9, http://dx.doi.org/10.1371/journal.pcbi.1002518.

[15] V. Vieira, J. Ferreira, M. Rocha, A pipeline for the reconstruction and evaluation of context-specific human metabolic models at a large-scale, PLoS Comput. Biol. 18 (2022) http://dx.doi.org/10.1371/journal.pcbi.1009294.

[16] T. Barata, V. Vieira, R. Rodrigues, R.P. das Neves, M. Rocha, Reconstruction of tissue-specific genome-scale metabolic models for human cancer stem cells, Comput. Biol. Med. 142 (2022) 1–12, http://dx.doi.org/10.1016/J.COMPBIOMED.2021.105177.

[17] B.J. Sánchez, C. Zhang, A. Nilsson, P.-J. Lahtvee, E.J. Kerkhoven, J. Nielsen, Improving the phenotype predictions of a yeast genome-scale metabolic model by incorporating enzymatic constraints, Mol. Syst. Biol. 13 (2017) 935, http://dx.doi.org/10.15252/msb.20167411.

[18] A. Salehzadeh-Yazdi, Y. Asgari, A.A. Saboury, A. Masoudi-Nejad, Computational analysis of reciprocal association of metabolism and epigenetics in the budding yeast: A Genome-Scale Metabolic Model (GSMM) approach, in: J. Vera (Ed.), PLoS One 9 (2014) e111686, http://dx.doi.org/10.1371/journal.pone.0111686.

[19] M.P. Pacheco, E. John, T. Kaoma, M. Heinäniemi, N. Nicot, L. Vallar, J.-L. Bueb, L. Sinkkonen, T. Sauter, Integrated metabolic modelling reveals cell-type specific epigenetic control points of the macrophage metabolic network, BMC Genom. 16 (2015) 809, http://dx.doi.org/10.1186/s12864-015-1984-4.

[20] S. Chandrasekaran, J. Zhang, Z. Sun, L. Zhang, C.A. Ross, Y.-C. Huang, J.M. Asara, H. Li, G.Q. Daley, J.J. Collins, Comprehensive mapping of pluripotent stem cell metabolism using dynamic genome-scale network modeling, Cell Rep. 21 (2017) 2965–2977, http://dx.doi.org/10.1016/j.celrep.2017.07.048.

[21] F. Shen, L. Boccuto, R. Pauly, S. Srikanth, S. Chandrasekaran, Genome-scale network model of metabolism and histone acetylation reveals metabolic dependencies of histone deacetylase inhibitors, Genome Biol. 20 (2019) 1–15, http://dx.doi.org/10.1186/s13059-019-1661-z.

[22] J.L. Robinson, P. nar Kocabaş, H. Wang, P. etienne Cholley, D. Cook, A. Nilsson, M. Anton, R. Ferreira, I. Domenzain, V. Billa, A. Limeta, A. Hedin, J. Gustafsson, E.J. Kerkhoven, L.T. Svensson, B.O. Palsson, A. Mardinoglu, L. Hansson, M. Uhlén, J. Nielsen, An Atlas of human metabolism, Sci. Signal. 13 (2020) 1–11, http://dx.doi.org/10.1126/scisignal.aaz1482.

[23] A. Richelle, A.W.T. Chiang, C. chung Kuo, N.E. Lewis, Increasing consensus of context-specific metabolic models by integrating data-inferred cell functions, in: C.A. Ouzounis (Ed.), PLoS Comput. Biol. 15 (2019) 1–19, http://dx.doi.org/10.1371/journal.pcbi.1006867.

[24] M. Turpin, G. Salbert, 5-methylcytosine turnover: Mechanisms and therapeutic implications in cancer, Front. Mol. Biosci. 9 (2022) http://dx.doi.org/10.3389/fmolb.2022.976862.

[25] D. Cheishvili, L. Boureau, M. Szyf, DNA demethylation and invasive cancer: Implications for therapeutics, British J. Pharmacol. 172 (2015) 2705–2715, http://dx.doi.org/10.1111/bph.12885.

[26] M. Ehrlich, DNA methylation in cancer: Too much, but also too little, Oncogene 21 (2002) 5400–5413, http://dx.doi.org/10.1038/sj.onc.1205651.

[27] M. Ghandi, F.W. Huang, J. Jané-Valbuena, G.V. Kryukov, C.C. Lo, E.R. McDonald, J. Barretina, E.T. Gelfand, C.M. Bielski, H. Li, K. Hu, A.Y. Andreev-Drakhlin, J. Kim, J.M. Hess, B.J. Haas, F. Aguet, B.A. Weir, M.V. Rothberg, B.R. Paolella, M.S. Lawrence, R. Akbani, Y. Lu, H.L. Tiv, P.C. Gokhale, A. de Weck, A.A. Mansour, C. Oh, J. Shih, K. Hadi, Y. Rosen, J. Bistline, K. Venkatesan, A. Reddy, D. Sonkin, M. Liu, J. Lehar, J.M. Korn, D.A. Porter, M.D. Jones, J. Golji, G. Caponigro, J.E. Taylor, C.M. Dunning, A.L. Creech, A.C. Warren, J.M. McFarland, M. Zamanighomi, A. Kauffmann, N. Stransky, M. Imielinski, Y.E. Maruvka, A.D. Cherniack, A. Tsherniak, F. Vazquez, J.D. Jaffe, A.A. Lane, D.M. Weinstock, C.M. Johannessen, M.P. Morrissey, F. Stegmeier, R. Schlegel, W.C. Hahn, G. Getz, G.B. Mills, J.S. Boehm, T.R. Golub, L.A. Garraway, W.R. Sellers, Next-generation characterization of the cancer cell line encyclopedia, Nature 569 (2019) 503–508, http://dx.doi.org/10.1038/s41586-019-1186-3.

[28] W.C. Reinhold, S. Varma, M. Sunshine, V. Rajapakse, A. Luna, K.W. Kohn, H. Stevenson, Y. Wang, H. Heyn, V. Nogales, S. Moran, D.J. Goldstein, J.H. Doroshow, P.S. Meltzer, M. Esteller, Y. Pommier, The NCI-60 methylome and its integration into CellMiner, Cancer Res. 77 (2017) 601–612, http://dx.doi.org/10.1158/0008-5472.CAN-16-0655.

[29] S. Agnihotri, G. Zadeh, Metabolic reprogramming in glioblastoma: the influence of cancer metabolism on epigenetics and unanswered questions, Neuro-Oncol. 18 (2016) 160–172, http://dx.doi.org/10.1093/neuonc/nov125.

[30] N.N. Pavlova, T.C. B., The emerging hallmarks of cancer metabolism, Cell Metabol. 23 (2016) 27–47, http://dx.doi.org/10.1016/j.cmet.2015.12.006.

[31] Z.G. Movahed, M. Rastegari-Pouyani, M. hossein Mohammadi, Cancer cells change their glucose metabolism to overcome increased ROS: One step from cancer cell to cancer stem cell? Biomed. Pharmacother. 112 (2019) 1–15, http://dx.doi.org/10.1016/j.biopha.2019.108690.

[32] X. Ding, W. Zhang, S. Li, H. Yang, The role of cholesterol metabolism in cancer, Am. J. Cancer Res. 9 (2019) 219–227.

[33] T. Murai, The role of lipid rafts in cancer cell adhesion and migration, Int. J. Cell Biol. 2012 (2012) 1–6, http://dx.doi.org/10.1155/2012/763283.

[34] K. Yokogami, T. Kikuchi, T. Watanabe, Y. Nakatake, S. Yamashita, A. Mizuguchi, H. Takeshima, Methionine regulates self-renewal, pluripotency, and cell death of GIC through cholesterol—rRNA axis, BMC Cancer 22 (2022) 1351, http://dx.doi.org/10.1186/s12885-022-10280-5, URL https://bmccancer.biomedcentral.com/articles/10.1186/s12885-022-10280-5.

[35] L. Wang, A.A.F. Shannar, R. Wu, P. Chou, M.S. Sarwar, H. chen Kuo, R.M. Peter, Y. Wang, X. Su, A.N. Kong, Butyrate drives metabolic rewiring and epigenetic reprogramming in human colon cancer cells, Mol. Nutr. Food Res. 66 (2022) http://dx.doi.org/10.1002/MNFR.202200028, URL https://pubmed.ncbi.nlm.nih.gov/35429118/.

[36] E.K. Mallott, K.R. Amato, Butyrate production pathway abundances are similar in human and nonhuman primate gut microbiomes, Mol. Biol. Evol. 39 (2022) http://dx.doi.org/10.1093/molbev/msab279.

[37] D. Yang, Q. Guo, Y. Liang, Y. Zhao, X. Tian, Y. Ye, J. Tian, T. Wu, N. Lu, Wogonin induces cellular senescence in breast cancer via suppressing TXNRD2 expression, Arch. Toxicol. 94 (2020) 3433–3447, http://dx.doi.org/10.1007/s00204-020-02842-y.

[38] X. Zhu, D. Li, Y. Du, W. He, Y. Lu, DNA hypermethylation-mediated downregulation of antioxidant genes contributes to the early onset of cataracts in highly myopic eyes, Redox Biol. 19 (2018) 179–189, http://dx.doi.org/10.1016/j.redox.2018.08.012.

[39] J. Nanduri, Y.-J. Peng, N. Wang, S.A. Khan, G.L. Semenza, N.R. Prabhakar, DNA methylation in the central and efferent limbs of the chemoreflex requires carotid body neural activity, J. Physiol. 596 (2018) 3087–3100, http://dx.doi.org/10.1113/JP274833.

[40] T.M. Barrow, N.W. Doo, R.L. Milne, G.G. Giles, E. Willmore, G. Strathdee, H.-M. Byun, Analysis of retrotransposon subfamily DNA methylation reveals novel early epigenetic changes in chronic lymphocytic leukemia, Haematologica 106 (2020) 98–110, http://dx.doi.org/10.3324/haematol.2019.228478.

[41] R. Medina-Aguilar, C. Pérez-Plasencia, L.A. Marchat, P. Gariglio, J.G. Mena, S.R. guez Cuevas, E. Ruíz-García, H.A. de la Vega, J.H. Juárez, A. Flores-Pérez, C. López-Camarillo, Methylation landscape of human breast cancer cells in response to dietary compound resveratrol, PLoS One 11 (2016) e0157866, http://dx.doi.org/10.1371/journal.pone.0157866.

[42] H.C. Jonus, B.S. Hanberry, S. Khatu, J. Kim, H. Luesch, L.H. Dang, M.G. Bartlett, J.A. Zastre, The adaptive regulation of thiamine pyrophosphokinase-1 facilitates malignant growth during supplemental thiamine conditions, Oncotarget 9 (2018) 35422–35438, http://dx.doi.org/10.18632/oncotarget.26259.

[43] K.V.Q. Luong, L.T.H. Nguyen, The role of thiamine in cancer: Possible genetic and cellular signaling mechanisms, Cancer Genom. Proteomics 10 (2013) 169–185.

[44] R.M. Kohli, Y. Zhang, TET enzymes, TDG and the dynamics of DNA demethylation, Nature 502 (2013) 472–479, http://dx.doi.org/10.1038/nature12750.

[45] C. Schulz, T. Kumelj, E. Karlsen, E. Almaas, Genome-scale metabolic modelling when changes in environmental conditions affect biomass composition, PLoS Comput. Biol. 17 (2021) 1–22, http://dx.doi.org/10.1371/journal.pcbi.1008528.

[46] D. Dikicioglu, B. Kırdar, S.G. Oliver, Biomass composition: the "elephant in the room" of metabolic modelling, Metabolomics 11 (2015) 1690–1701, http://dx.doi.org/10.1007/s11306-015-0819-2.

[47] H.T. Lee, S. Oh, D.H. Ro, H. Yoo, Y.W. Kwon, The key role of DNA methylation and histone acetylation in epigenetics of atherosclerosis, J. Lipid Atherosclerosis 9 (2020) 419, http://dx.doi.org/10.12997/JLA.2020.9.3.419.

[48] C. Carson, H.A. Lawson, Epigenetics of metabolic syndrome, Physiol. Genomics 50 (2018) 947–955, http://dx.doi.org/10.1152/physiolgenomics.00072.2018.

[49] J. Ferreira, V. Vieira, J. Gomes, S. Correia, M. Rocha, Troppo - A Python framework for the reconstruction of context-specific metabolic models, Adv. Intell. Syst. Comput. 1005 (2020) 146–153, http://dx.doi.org/10.1007/978-3-030-23873-5_18.

[50] M. Jain, R. Nilsson, S. Sharma, N. Madhusudhan, T. Kitami, A.L. Souza, R. Kafri, M.W. Kirschner, C.B. Clish, V.K. Mootha, Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation, Science 336 (2012) 1040–1044, http://dx.doi.org/10.1126/science.1218595.

[51] D.C. Zielinski, N. Jamshidi, A.J. Corbett, A. Bordbar, A. Thomas, B.O. Palsson, Systems biology analysis of drivers underlying hallmarks of cancer cell metabolism, Sci. Rep. 7 (2017) 1–14, http://dx.doi.org/10.1038/srep41241.

[52] V. Pereira, F. Cruz, M. Rocha, MEWpy: A computational strain optimization workbench in Python, Bioinformatics 37 (2021) 2494–2496, http://dx.doi.org/10.1093/BIOINFORMATICS/BTAB013.

[53] A. Ebrahim, J.A. Lerman, B.O. Palsson, D.R. Hyduke, COBRApy: Constraints-based reconstruction and analysis for Python, BMC Syst. Biol. 7 (2013) 1–6, http://dx.doi.org/10.1186/1752-0509-7-74.

[54] B. He, C. Zhang, X. Zhang, Y. Fan, H. Zeng, J. Liu, H. Meng, D. Bai, J. Peng, Q. Zhang, W. Tao, C. Yi, Tissue-specific 5-hydroxymethylcytosine landscape of the human genome, Nature Commun. 12 (2021) 1–12, http://dx.doi.org/10.1038/s41467-021-24425-w.

[55] S. Santos, I. Rocha, Estimation of biomass composition from genomic and transcriptomic information, J. Integr. Bioinform. 13 (2016) 285, http://dx.doi.org/10.2390/biecoll-jib-2016-285.

[56] Y. Kawasaki, Y. Kuroda, I. Suetake, S. Tajima, F. Ishino, T. Kohda, A novel method for the simultaneous identification of methylcytosine and hydroxymethylcytosine at a single base resolution, Nucl. Acids Res. 45 (2017) e24, http://dx.doi.org/10.1093/NAR/GKW994.

[57] M.J. Booth, S. Balasubramanian, Reduced bisulfite sequencing: Quantitative base-resolution sequencing of 5-formylcytosine, Methods Mol. Biol. (Clifton, N.J.) 2272 (2021) 3–12, http://dx.doi.org/10.1007/978-1-0716-1294-1_1.

[58] U.T. Shankavaram, S. Varma, D. Kane, M. Sunshine, K.K. Chary, W.C. Reinhold, Y. Pommier, J.N. Weinstein, CellMiner: A relational database and query tool for the NCI-60 cancer cell lines, BMC Genom. 10 (2009) 1–10, http://dx.doi.org/10.1186/1471-2164-10-277/FIGURES/4.