OXFORD

popFBA: tackling intratumour heterogeneity with Flux Balance Analysis

Chiara Damiani[1,2,*,†], Marzia Di Filippo[1,3,†], Dario Pescini[1,4], Davide Maspero[3], Riccardo Colombo[1,2] and Giancarlo Mauri[1,2]

[1]SYSBIO Centre of Systems Biology, 20126 Milan, Italy, [2]Department of Informatics, Systems and Communication, [3]Department of Biotechnology and Biosciences and [4]Department of Statistics and Quantitative Methods, University Milano-Bicocca, 20126 Milan, Italy

*To whom correspondence should be addressed.

†The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

Abstract

Motivation: Intratumour heterogeneity poses many challenges to the treatment of cancer. Unfortunately, the transcriptional and metabolic information retrieved by currently available computational and experimental techniques portrays the average behaviour of intermixed and heterogeneous cell subpopulations within a given tumour. Emerging single-cell genomic analyses are nonetheless unable to characterize the interactions among cancer subpopulations. In this study, we propose *popFBA*, an extension to classic Flux Balance Analysis, to explore how metabolic heterogeneity and cooperation phenomena affect the overall growth of cancer cell populations.

Results: We show how clones of a metabolic network of human central carbon metabolism, sharing the same stoichiometry and capacity constraints, may follow several different metabolic paths and cooperate to maximize the growth of the total population. We also introduce a method to explore the space of possible interactions, given some constraints on plasma supply of nutrients. We illustrate how alternative nutrients in plasma supply and/or a dishomogeneous distribution of oxygen provision may affect the landscape of heterogeneous phenotypes. We finally provide a technique to identify the most proliferative cells within the heterogeneous population.

Availability and implementation: the popFBA MATLAB function and the SBML model are available at https://github.com/BIMIB-DISCo/popFBA.

Contact: chiara.damiani@unimib.it

1 Introduction

Populations of tumour cells display considerable phenotypic diversity both at the intertumour and intratumour level. Along with genetic and epigenetic factors, differential trophic supply and variations in the tumour microenvironment contribute in particular to intratumour metabolic heterogeneity and to the emergence of a complex cancer population architecture (Burrell *et al.*, 2013). Intratumour heterogeneity increases the repertoire of possible cellular responses to a drug and fosters the adaptive nature of cellular behaviours (Sun and Yu, 2015), compromising the efficacy of cancer therapies.

Although single cell-based technologies represent a promising approach for a more in-depth understanding of single cell behaviour within solid tumours, cancer populations are composed of both tumour and stromal cells that interact with each other by establishing a network of interactions that cannot be deciphered from the analysis of each of these individual components alone (Marusyk *et al.*, 2012).

Computational methodologies that allow to identify the possible cooperations that can be established to enhance the overall growth of the tumour mass and to investigate the mechanisms underlying them are therefore desired as they may facilitate the development of more effective cancer treatments. In particular, modelling efforts to integrate different sources of data and to determine the metabolic phenotype of cooperating cells would provide information that cannot be obtained directly from the genotype, transcriptome, proteome nor the metabolome alone (Holmes *et al.*, 2008).

Constraint-based modelling and especially Flux Balance Analysis (FBA) represents so far the most applied technique for studying metabolism and has effectively been exploited as a scaffold for 'omic' data integration (Cazzaniga *et al.*, 2014). In particular, many methods have been introduced for the integration of transcriptomic data into constraint-based models of metabolism (Machado and Herrgård, 2014). FBA is performed on a single metabolic network

and provides the (optimal) net flux distribution of possibly different metabolic populations, hindering the identification of possible metabolic interactions between subpopulations.

In Di Filippo et al. (2017), we preliminarily showed that clones of a constraint-based toy metabolic network can cooperate to maximize the ATP production of a total population. In this study, we propose a method to explore the space of possible interactions between heterogeneous cell populations within a putative tumour—given some constraints on plasma supply of nutrients—as well as a technique to identify the most proliferative sub-phenotypes.

2 Approach

Because the classic FBA approach on a single metabolic model predicts the optimal net flux distribution of a population (Di Filippo et al., 2017), the single metabolic model just represents a sort of black box of the investigated population, and on its own is unable to inform about the complex interactions occurring inside it. However, cooperation phenomena within some tumours have been reported, specifically between stromal and cancer cells (Fiaschi et al., 2012; Martinez-Outschoorn et al., 2011; Sanità et al., 2014; Whitaker-Menezes et al., 2011), pushing forward the need for computational approaches able to uncover these as well as other possible kind of interactions between tumour subpopulations. Therefore, we propose a new methodology aimed at investigating metabolic phenotypes of different subpopulations belonging to the same tumour mass, especially focusing on the relationships among them.

Assuming that a tumour mass may be composed of different types of cells (including stromal cells) and that, for reasons of spatial proximity, the communication with the plasma (in terms of nutrients exchange) of the different components may differ with respect to the communication with other cells within the populations, we separately modelled exchanges with plasma and exchanges with the tumour microenvironment.

In particular, the single metabolic model is used as a building block for constructing, in an automatic way, the population model characterized by multiple clones, all having identical stoichiometry and capacity constraints and sharing the plasma supply of nutrients. Exploiting linear programming optimization, we can investigate both the cooperation among different clones that is consistent with the achievement of the optimal growth rate of the entire tumour mass, and identify the strategies adopted by the most proliferative clones.

The proposed approach is schematically described in Figure 1 and formally defined in the following section.

3 Materials and methods

3.1 Metabolic network model

A metabolic network is formalized by specifying a set $\mathcal{X} = \{X_1, \ldots, X_M\}$ of metabolites, and the set $\mathcal{R} = \{R_1, \ldots, R_N\}$ of chemical transformations taking place among them. Reactions are defined as:

$$R_j : \sum_{i=1}^{M_j^s} \alpha_{ji} X_i \leftrightarrow \sum_{i=1}^{M_j^p} \beta_{ji} X_i, \qquad (1)$$

where M_j^s and M_j^p are, respectively, the number of reactants and products of reaction j—relatively to the case in which the reaction proceeds in the forward direction (from left to right)—and $\alpha_{ji}, \beta_{ji} \in N$ are stoichiometric coefficients associated, respectively, with the ith substrate and the ith product of the jth reaction.
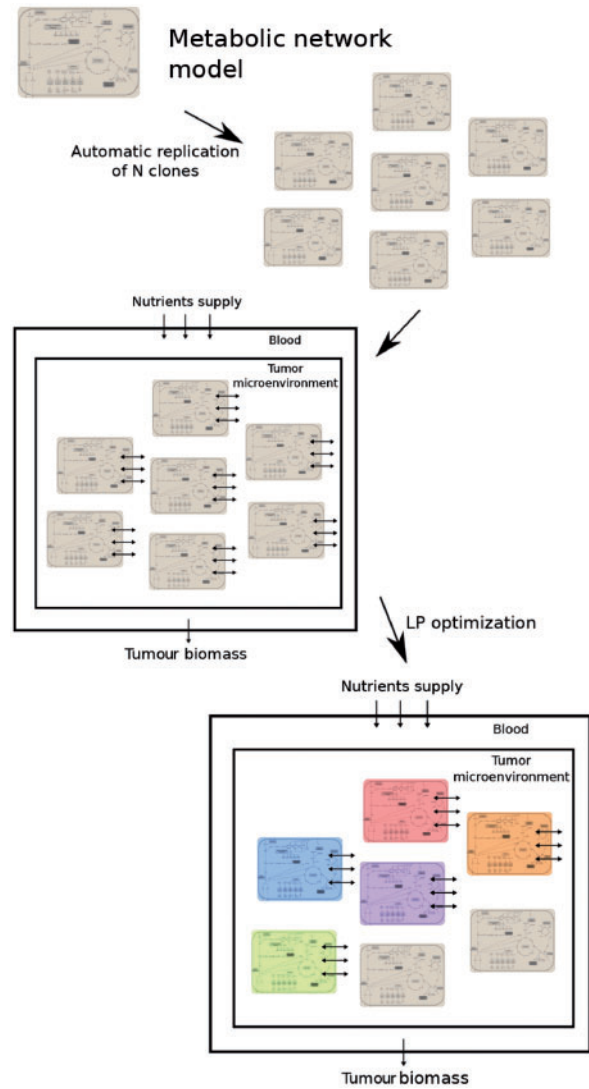

Fig. 1. Graphical representation of popFBA methodology

For the analyses reported in this study we used the model of central carbon metabolism, which we refer to as 'coreHMR', extracted from the HMR model (Mardinoglu et al., 2013) and introduced in Di Filippo et al. (2016), composed of 243 metabolites and 271 reactions. To adapt the model to popFBA analyses, we set the mitochondrial isocitrate dehydrogenase-catalyzed reactions as reversible; we introduced a consumption of ATP within the biomass reaction; we introduced a cell maintenance reaction ($ATP \rightarrow ADP$); and finally we structurally removed the thermodynamically infeasible loops detected with the algorithm developed in De Martino et al. (2013).

3.2 FBA

The assumption underlying FBA is that metabolic networks will reach a steady state: the concentration of each metabolite is assumed constant: $d[X_i]/dt = 0 \ \forall X_i \in \mathcal{X}$. The stoichiometric constraints lead to a bounded solution space of all feasible flux distributions, which can be further restricted by specifying maximum and minimum fluxes through any particular reaction. These bounds on admissible fluxes allow to define constraints on the reversibility of reactions, to impose experimentally measured flux ranges, and to set the extent of nutrients supply. The exchange of matter with the environment is

represented as a set $\mathcal{E} = \{E_1, \ldots, E_{N_{ext}}\}$ of N_{ext} unbalanced reactions (exchange reactions), enabling a predefined set of metabolites (including the pseudo-metabolite representing biomass) $\mathcal{Y} = \{Y_1, \ldots, Y_{N_{ext}}\} \subset \mathcal{X}$ to be inserted in or removed from the system, through reactions as E_j:

$$E_j : Y_j \leftrightarrow \varnothing \qquad (2)$$

The first step to perform FBA is the derivation of the stoichiometric matrix S, of size $M \times (N + N_{ext})$, whose element s_{ji} takes value $-\alpha_{ji}$ if the species X_i is a reactant of reaction R_j, $+\beta_{ji}$ if the species X_i is a product of reaction R_j and 0 otherwise.

FBA is then applied to determine the rate v_i at which each reaction in $\mathcal{R} \cup \mathcal{E}$ occurs, that is, the flux distribution $\vec{v} = (v_1, v_2, \ldots, v_{N+N_{ext}}) = (r_1, \ldots, r_N, e_1, \ldots, e_{N_{ext}})$ that maximizes or minimizes the objective function $Z = \sum_{i=1}^{N+N_{ext}} w_i v_i$, where w_i is the weight that quantifies the contribution of reaction i, while r_i and e_i are, respectively, the flux value associated to an "internal" or an exchange reaction.

In order to simulate tumour growth, in this study we maximize the flux $e_{biomass}$ through the biomass exchange reaction, thus $Z = e_{biomass}$ The optimization problem is postulated as a general Linear Programming (LP) formulation:

$$
\begin{aligned}
&\text{maximize or minimize } Z \\
&\text{subject to } S\vec{v} = \vec{0}, \\
&\vec{v}_L \leq \vec{v} \leq \vec{v}_B
\end{aligned}
\qquad (3)
$$

$\vec{v}_L$ and $\vec{v}_B$ are vectors specifying the lower and upper bound, respectively, for each flux v_i of $\vec{v}$. A negative lower bound indicates that flux is allowed in the backward reaction.

To solve the above problem we exploited the GLPK solver within the COBRA Toolbox (Schellenberger *et al.*, 2011). For a more comprehensive description of FBA, the reader is referred to Orth *et al.* (2010).

3.3 popFBA

In order to investigate the role of cooperation within a population sharing a common environment, in this study we devised popFBA, an extension to FBA able to cope with the presence of several subpopulations exchanging a defined set of metabolites. Given a metabolic network A defined as $A = (\mathcal{X}, \mathcal{R}, \mathcal{E})$, popFBA maximizes the total biomass of N_{pops} clones A^c of A, which can cooperate by exchanging nutrients in the tumour microenvironment. For each clone A^c, let $\mathcal{X}^c = \{X_i^c\}$ be the set of its metabolites, $\mathcal{R}^c = \{R_j^c\}$ the set of its internal reactions, with $j = 1, \ldots, N$ and $c = 1, \ldots, N_{pops}$. To correct for the fact that in a population model a metabolite is not removed from the systems, but becomes a metabolite in the tumour microenvironment, each reaction $E_j \in \mathcal{E}$ of A is transformed into a *cooperation reaction* C_j^c with the form

$$C_j^c : Y_j^c \leftrightarrow Y_j' \qquad (4)$$
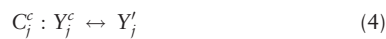
It is also necessary to define the new set of tumour microenvironment metabolites $\mathcal{Y}' = \{Y_i'\}$ with $i = 1, \ldots, N_{ext}$, together with a new set of N_{blood} exchange reactions $\mathcal{B} = \{B_1, \ldots, B_{N_{blood}}\}$ to allow a subset of metabolites $\mathcal{K} = \{K_1, \ldots, K_{N_{blood}}\} \subset \mathcal{Y}'$ to be exchanged with the blood supply:

$$B_j : K_j \leftrightarrow \varnothing \qquad (5)$$

The population model P is then defined by the union set of the metabolites $\mathcal{X}^P = \cup_c \mathcal{X}^c \cup \mathcal{Y}'$, of the internal reactions $\mathcal{R}^P = \cup_c \mathcal{R}^c$, of the cooperation reactions $\mathcal{C}^P = \cup_c \mathcal{C}^c$ and of the population exchange reactions $\mathcal{B}$.

A stoichiometric matrix S^P is then built for all reactions in $\mathcal{R}^P$, $\mathcal{C}^P$ and $\mathcal{B}$ and for all metabolites in $\mathcal{X}^P$.

The final size of matrix S^P is $(N_{pops} \cdot M + N_{blood}) \times (N_{pops} \cdot (N + N_{ext}) + N_{blood})$.

To obtain the matrix S^P, we implemented a MATLAB function that automatically replicates a number of times any (COBRA Toolbox compliant—Schellenberger *et al.*, 2011) SBML model to obtain the above defined population model, in a suitable form to then undergo constraint-based analyses. Linear programming is then applied as per Equation 3 to determine the flux distribution $\vec{v} = (v_1, \ldots, v_{N_{pops} \cdot (N+N_{ext})+N_{blood}}) = (r_1^1, \ldots, r_N^1, \ldots, r_1^{N_{pops}}, \ldots, r_N^{N_{pops}}, c_1^1, \ldots, c_{N_{ext}}^1, \ldots, c_1^{N_{pops}}, \ldots, c_{N_{ext}}^{N_{pops}}, b_1, \ldots, b_{N_{blood}})$ that maximizes the biomass exchange flux $b_{biomass}$, with v_i representing any flux i of the population model, and for each clone c, r_i^c representing the ith internal flux, c_i^c representing the ith a cooperation flux and b_i a plasma exchange flux.

3.4 Sampling in the region of optimal solutions

Linear programming only returns a single optimal solution. However, many alternative optimal flux distributions may exist. Flux Variability Analysis (Mahadevan and Schilling, 2003) has been efficiently exploited to identify the range of values that a flux can take across the complete set of optimal solutions. Nevertheless, in order to analyse the correlation between flux values and the proliferation rates of the model subpopulations, we need punctual solutions. Although methods have been proposed for enumerating alternative optimal solutions (Reed and Palsson, 2004), an exhaustive enumeration is not practicable for popFBA, due to the interchangeability of the flux distributions of the N_{pops} clones.

To cope with this problem, we set the bounds of the biomass exchange flux $b_{biomass}$ to the optimal value obtained with popFBA and we sampled the admissible solutions. The dominant algorithm of choice to uniformly sampling inside the region of allowed solutions is the so-called "Hit-and-Run" (HR) (Schellenberger and Palsson, 2009), according to which an initial valid point is moved repeatedly inside the space according to probabilistic rules.

In this study, we also exploited a recently proposed alternative approach (Bordel *et al.*, 2010; Damiani *et al.*, 2014): the simplex method with a random set of objective functions to be maximized. The maximization of each of these objective functions gives a corner in the space of solutions. In Bordel *et al.* (2010), random objective functions were generated by selecting random pairs of reactions. To maximize variability of sampled solutions, we instead let any number of reactions to take part in the objective function Z as in Damiani *et al.* (2014). The fraction τ of considered reactions is randomly drawn with uniform probability in $(0, 1]$. To any selected reaction is then assigned a random weight w_i uniformly tossed from the interval $(0, 1]$, where w_i takes value 0 with probability τ and a random value with uniform probability in $(0, 1]$ with probability $1 - \tau$.

For both methods, we controlled for repetitions in the sampled points.

3.5 Assessing subpopulations heterogeneity

To assess the heterogeneity of the metabolism of the N_{pops} clones within a given optimal solution, we compared their flux distributions both quantitatively and qualitatively. To avoid taking into account clones that carry no flux, we disregarded the clones that are

not cooperating for tumour growth, by filtering out clones that do not have a least one non-zero flux through any of the cooperation reactions.

To count the number of quantitatively different subpopulations we considered clones a and b to be different if they differ by at least the value of one flux (rounded at the fourth digit): if $\exists i : v_i^a \neq v_i^b$. To count the number of qualitatively different subpopulations, the counting process is performed, rather than on $\vec{v'}$, on a discretized version $\vec{d'}$ of vector $\vec{v'}$, whose component d_i' is obtained as follows:

$$d_i' = \begin{cases} 1 & v_i > 0 \\ -1 & v_i < 0 \\ 0 & v_i = 0 \end{cases} \quad (6)$$

In this way two clones are considered as different *iff* they follow at least one different metabolic path, that is, a different route or a different flux direction.

4 Results

4.1 popFBA reveals the existence of cooperation and metabolic heterogeneity within cancer population models

We applied popFBA to 10 clones ($N_{pops} = 10$) of the coreHMR model. The value 10 was arbitrarily chosen to allow up to 10 possibly different subpopulations to be detected. In a first experiment we simulated a plasma supply of glucose, glutamine and oxygen, the three nutrients provided with the same order of molar magnitude, in accordance with the magnitude found when scanning data in literature. Although the molar concentration of glutamine may be lower as compared with that of glucose and oxygen, in our system glutamine represents the unique nitrogen source, so it is reasonable to increase its uptake flux to account also for other nitrogen sources that are generally available in the plasma. As a first approximation, we assumed equal bounds for the reactions of the 10 clones, thus disregarding spatial diffusion phenomena. The 10 clones can cooperate, via cooperation reactions described in the Section 3, by exchanging lactate, glutamine, glutamate and ammonia. Although these metabolites can be secreted in the microenvironment compartment, they cannot be disposed (as waste products) in the human plasma. This experimental setting, which we refer to as the reference condition, is better detailed in Figure 2A. In this condition, we sampled 2×10^4 different optimal solutions (10^4 points with the HR sampling methods and 10^4 with the CB sampling method), that are compatible with the same optimal tumour biomass.

We assessed the quantitative heterogeneity of the clones in each of the sampled solutions. Remarkably, in all sampled solutions, we observed that all 10 clones behave differently. This diversity results in a different biomass synthesis flux value (the growth rate) for distinct clones. This heterogeneity may partially be the result of slightly different phenotypes, following the very same metabolic paths but at different rates. We wanted therefore to assess the number of clones that follow different metabolic paths, that is, that differ in the set of reactions that is active and/or in the direction of the flux through such set of reactions. The distribution of the number of qualitatively different subpopulations is reported in Figure 2B and it shows that, although subpopulations of quantitatively different clones may overlap from a qualitative point of view, for both sampling methods, the clones typically all follow different metabolic routes.
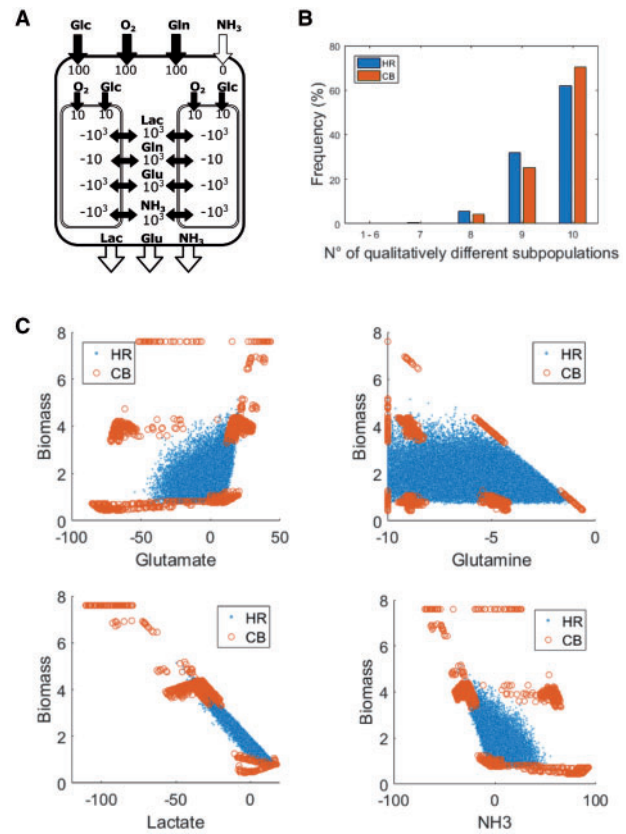

Fig. 2. Reference condition. (**A**) Experimental setting of the bounds imposed on the release/consumption of metabolites in/by the plasma, and on the cooperative reactions. Black or white filled arrows, respectively, indicate allowed and blocked reactions (bounds set to 0). (**B**) Histogram relative to the number of qualitatively different subpopulations obtained with HR (blue bars) and CB (orange bars) sampling methods. (**C**) Scatter plots obtained with HR (blue points) and CB (orange circles) sampling methods relative, in clockwise order, to the correlation between glutamate exchange and biomass synthesis, between glutamine exchange and biomass synthesis, between NH_3 exchange and biomass synthesis and between lactate exchange and biomass synthesis. Abbreviations: Glc, Glucose; O_2, Oxygen; Gln, glutamine; NH_3, ammonia; Lac, lactate; Glu, glutamate

Once we established that popFBA is able to highlight the possible heterogeneity of subpopulations of cells belonging to the same tumour population, we shifted the attention toward a more in-depth investigation of which types of interactions are compatible with the achievement of the maximum tumour biomass. The observed heterogeneity results indeed form cooperative behaviours among different subpopulations. In fact, as it can be observed in the scatter plots in Figure 2C, a secretion (negative values) or consumption (positive values) of metabolites in the tumour microenvironment, namely of lactate, glutamate and ammonia emerged among the different subpopulations. Among the four allowed cooperations, glutamine is the only metabolite that is always just consumed and it is not exchanged with other subpopulations.

4.2 Identification of the most proliferative subpopulations

To identify the most proliferative phenotypes among the heterogeneous subpopulations identified above, we computed the Pearson Correlation Coefficient (ρ) between the flux of each of the four cooperation reactions (exchange of glutamate, glutamine, lactate and

ammonia via the tumour microenvironment) and the biomass production rate of the corresponding clones, consistently with the achievement of the optimal tumour biomass.

We found that the correlation with the biomass synthesis flux, obtained with HR (and CB), is -0.32 (-0.4) for glutamine; $+0.35$ ($+0.44$) for glutamate; -0.55 (-0.65) for ammonia; and -0.96 (-0.99) for lactate. Notice that a negative ρ implies a positive correlation between nutrient consumption and biomass, while a positive ρ implies a positive correlation between nutrient secretion and biomass.

The ρ values, along with the corresponding scatter plots (Fig. 2C), clearly indicate that the most proliferative subpopulations are those consuming the lactate available in the tumour microenvironment. To evaluate whether lactate consuming subpopulations are oxidising this carbon source, we also analysed the correlation between the lactate exchange and oxygen consumption fluxes, obtaining a ρ of $+0.98$ ($+0.99$), confirming that the most proliferative subpopulations are consuming high levels of oxygen, which is plausibly exploited to oxidize the consumed lactate.

4.3 The set of nutrients exchanged with plasma affects possible cooperations

The observation that the glutamine supplied with plasma is consumed by all 10 popFBA clones, with no clone producing and exchanging it with the tumour microenvironment, was expected. Glutamine represents indeed a unique source of nitrogen supplied by plasma in the above simulations, and nitrogen is mandatory for the synthesis of amino acids and thus of biomass. The correlation between glutamine and biomass is however modest, because the clones are able to extract the nitrogen from glutamine and to exchange it in the form of NH_3 or glutamate via the tumour microenvironment. Unexpectedly, the correlation with biomass is indeed negative for NH_3, but it is positive for glutamate.
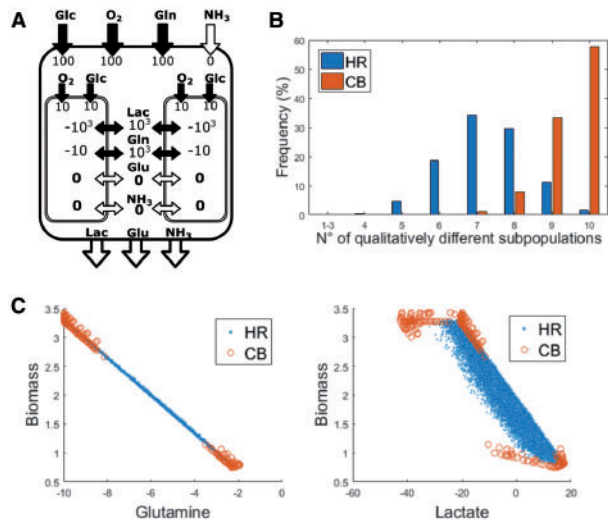
We observed that, when the cooperation flux for both NH_3 and glutamate is prevented (experimental setting in Fig. 3A and results in Fig. 3B and C), the correlation between glutamine and biomass becomes indeed close to -1 (-0.99 for both the HR and the CB method). In this situation, the negative correlation between lactate and biomass is preserved (-0.97 for HR and -0.96 for CB).

We also wanted to investigate whether the source of nitrogen in the plasma supply may affect the possible cooperative behaviours. We tested the situation in which the nitrogen source provided by the plasma to the tumour mass is not represented by glutamine, but by the ammonia (Fig. 4A), which may account for the nitrogen deriving from other amino acids in real cells. In this situation, as shown in Figure 4C, the exchange of glutamine among different subpopulations becomes possible. On the contrary, the exchange of NH_3 becomes not possible, as it is now exclusively consumed by the 10 clones (data not shown). Notably, we still observed a high correlation between lactate consumption and biomass formation: $\rho = -0.89$ (-0.98), as per Figure 4C.

Once the effect of an alternative nitrogen source on the internal transport reactions was investigated, we also analysed the possibility to have an outflow of either glutamate or ammonia or lactate from the tumour microenvironment compartment toward the plasma, maintaining the reference experimental setting. We observed that a slight increase in the tumour biomass value is caused by both glutamate (2%) and lactate secretion (2%) in the plasma.

Interestingly, we observed that the exit of glutamate from the tumour microenvironment towards the plasma does not prevent an exchange of glutamate among different subpopulations. On the contrary, as compared with the reference condition in Figure 2, this situation results in an enhancement of the correlations between glutamate exchange and biomass synthesis rate—from $+0.35$ (0.44) to $+0.47$ ($+0.61$); between glutamate exchange and lactate exchange—from -0.29 (-0.39) to -0.47 (-0.61); and between

Fig. 3. Cooperation reactions variation condition. (**A**) Experimental setting of the bounds imposed on the release/consumption of metabolites in/by the plasma, and on the cooperative reactions. Black or white filled arrows, respectively, indicate allowed and blocked reactions (bounds set to 0). (**B**) Histogram relative to the number of qualitatively different subpopulations obtained with HR (blue bars) and CB (orange bars) sampling methods. (**C**) Scatter plots obtained with HR (blue points) and CB (orange circles) sampling methods relative to the correlation between glutamine exchange and biomass synthesis (on the left), and between lactate exchange and biomass synthesis (on the right)
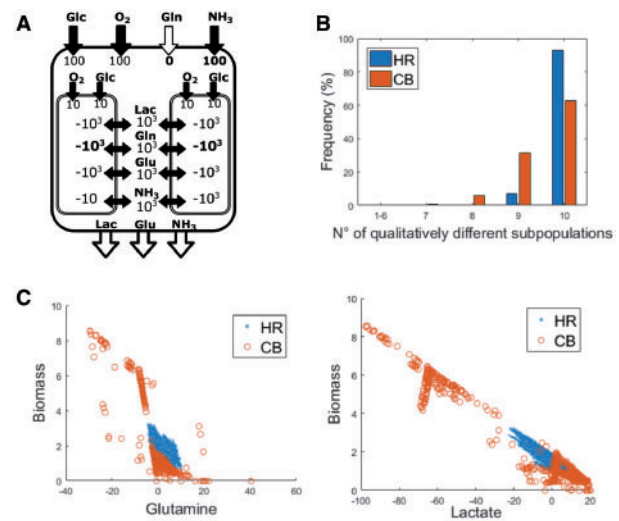

Fig. 4. Changing nitrogen source condition. (**A**) Experimental setting of the bounds imposed on the release/consumption of metabolites in/by the plasma, and on the cooperative reactions. Black or white filled arrows, respectively, indicate allowed and blocked reactions (bounds set to 0). (**B**) Histogram relative to the number of qualitatively different subpopulations obtained with HR (blue bars) and CB (orange bars) sampling methods. (**C**) Scatter plots obtained with HR (blue points) and CB (orange circles) sampling methods relative to the correlation between glutamine exchange and biomass synthesis (on the left), and between lactate exchange and biomass synthesis (on the right)

glutamate and oxygen—from −0.26 (−0.36) to −0.44 (−0.59). These results indicate that the subpopulations that are responsible for a glutamate secretion in the tumour microenvironment are characterized by a consumption of both lactate and oxygen.

4.4 Comparison of HR and CB sampling methods

In the scatter diagrams in Figures 2C, 3C and 4C it is evident how the two sampling methods (HR and CB) differ in the way they explore the space of possible cooperations among metabolic clones. HR uniformly samples within the space of optimal solutions, returning points that are close to one another inside the space of allowed solution, whereas, as already pointed out in Bordel *et al.* (2010), the CB method returns solutions corresponding to the corners in the region of allowed flux distributions. This difference in the two methods does not particularly affect the correlation coefficients between the release/consumption of metabolites within the tumour microenvironment and the growth rate of a given sub-population. It can be indeed observed in Figure 5C that the ρs obtained with the two methods are very similar. However, the HR method may in some cases underestimate the number of qualitatively different subpopulations. Although the distribution of the number of qualitatively different subpopulations of the two methods is very similar in the experiments presented in Figures 2B and 4B, it substantially diverges in the experiment in Figure 3B. A possible explanation of this phenomenon is that the experiment reported in Figure 3B refers to a case in which the clones have less possibilities to cooperate (glutamate and ammonia cooperation is prevented) and thus tend to be more similar to one another (indeed also the CB method finds more cases with a lower number of different subpopulations as compared with experiments in Figs. 2B and 4B), and that the HR method amplifies this effect.

As expected, the computation time of the two methods grows linearly with the sample size (Fig. 5D) and it is slightly higher for
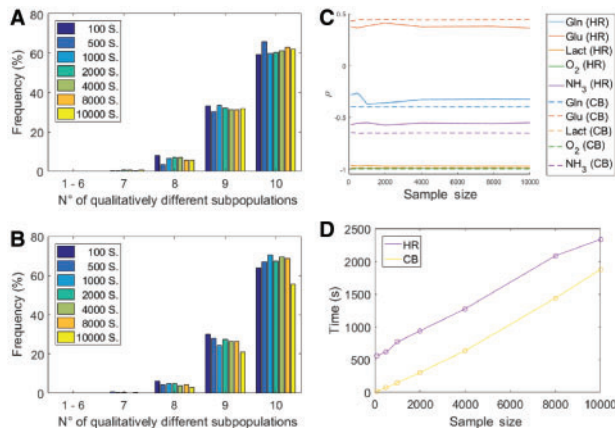
HR because of the fixed initial time to create the warm up points. In both cases, the computed ρs stabilize for a sample size greater than $\sim 4 \times 10^3$. Also the shape of the distribution of the number of different subpopulations (Fig. 5A and B for HR and CB, respectively) is not particularly affected by sample size, although some oscillations are possible especially for the CB case, while the frequency of any number of different subpopulations becomes stable for samples greater than $\sim 8 \times 10^3$. All in all these considerations confirm that the size of the sample chosen for our analysis ($10^4 + 10^4$) was reasonable.

4.5 Simulation of spatial diffusion phenomena with popFBA

The simulations presented so far assumed equal boundaries for the reactions of the 10 clones. There could be however cases where different biological conditions impose different uptake/secretion capabilities on the different subpopulations. Our approach could still be exploited to represent biologically distinct subpopulations, when information is available to constrain the exchange reactions differently. To prove the viability of popFBA to take into account spatial diffusion phenomena, we performed an experiment in which, differently from the experimental settings presented above, the 10 subnetworks are not identical in terms of access capability to the plasma supply.

To mimic a simplified oxygen gradient, we set the bounds of oxygen uptake of the 10 subnetworks according to a linear decreasing function (as illustrated in Fig. 6A). Assuming that subpopulations are radially stratified, subnetworks with a given availability of oxygen may represent subpopulations of cells at an equal distance from the blood vessel.

The scatterplots in Figure 6B clearly show that the subpopulations that consume less oxygen produce more lactate, which is then consumed by fast growing subpopulations with high oxygen consumption rates.

5 Conclusions

We introduced popFBA, an extension of FBA to take into account intratumour heterogeneity and interactions among different cell
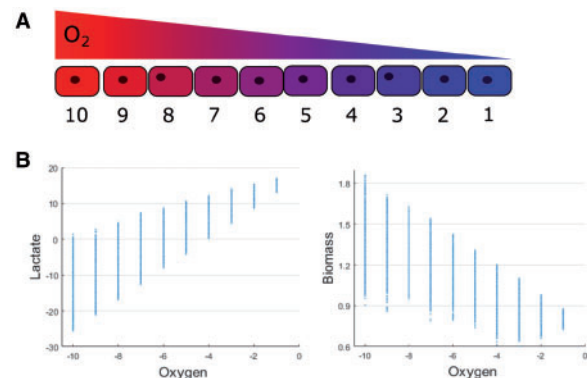

Fig. 5. (**A**) Distribution of the number of qualitatively different subpopulations obtained for seven independent samples of different size S ($S \in \{100, 500, 1000, 2000, 4000, 8000, 10\,000\}$) obtained with the HR method. Parameter nStepsPerPoint of the sampleCbModel MATLAB function was set to default value (200); parameter nWarmupPoints was set to 6000; parameters nFiles and nPointsPerFile were set in a way to maintain the ratio between their default values. (**B**) Distribution of the number of qualitatively different subpopulations obtained for seven independent samples of different size $\{100, 500, 1000, 2000, 4000, 8000, 10\,000\}$ obtained with the CB method. (**C**) Values of ρ between biomass and each of the exchange fluxes with the intratumour microenvironment (Gln, Glu, Lact, O$_2$, NH$_3$), obtained with HR (solid lines) and CB (dashed lines) as a function of sample size. (**D**) Computation time as a function of samples size for HR and CB

Fig. 6. Simulation of oxygen spatial diffusion. (**A**) Schematic representation of the performed experiment. The red-to-blue chromatic scale is used to mimic an oxygen gradient through the tumour population, respectively, from an aerobic to a hypoxic environment. The values on the bottom of the figure refer to maximum uptake flux value of oxygen imposed on each clone. (**B**) Scatter plots obtained with HR sampling methods relative to the correlation between lactate exchange and oxygen uptake (on the left), and between biomass synthesis and oxygen uptake (on the right)

populations within the same tumour. We applied popFBA to a model of 10 clones of the metabolic network of human central carbon metabolism, simulating a plasma supply of glucose, glutamine and oxygen, assuming equal bounds for the reactions of the 10 clones and an internal exchange of lactate, glutamine, glutamate and ammonia. We sampled different optimal solutions, by using both the HR sampling method and the CB sampling method, that are compatible with the same optimal tumour biomass. We observed that popFBA reveals the existence of metabolic heterogeneity and cooperation within the population model. Indeed, by assessing the quantitative heterogeneity of the clones in each of the sampled solutions, we observed that all 10 clones behave differently and are characterized by a different growth rate. Moreover, by assessing the distribution of the number of qualitatively different subpopulations, we found that the subpopulations are following different metabolic routes. This observed heterogeneity is the result of a cooperative behaviour among subpopulations: a consumption or secretion of lactate, glutamine, glutamate and ammonia (the four exchanged metabolites) emerged indeed among the different subpopulations.

Following the observation that popFBA approach is able to point out, if it is present, a cooperative behaviour among multiple subpopulations of the same population, we investigated which types of interactions are compatible with the achievement of the optimal tumour biomass.

In order to characterize the metabolism of the most proliferative subpopulations we assessed the correlation between the exchanges of metabolites within the tumour microenvironment and the growth rate of a given subpopulation. In this regard, we showed that, although the two sampling methods (HR and CB) differ in the way they explore the space of possible solutions, this difference has no particular effect on the computed correlation coefficients. Therefore, the two methods are equally effective when the goal is to determine the subpopulation with a propensity for growth.

Remarkably, in all of the investigated scenarios, we observed that the lactate exchange within tumour microenvironment is negatively correlated with respect to the biomass synthesis flux. This means that the most proliferative subpopulations consume the lactate that is secreted in the tumour microenvironment by less proliferative subpopulations, by using it as energy source. Because a high positive correlation between the lactate exchange and oxygen consumption fluxes emerged, we deduced that the most proliferative subpopulations are oxidising the consumed lactate.

A wide array of studies on clonal cancer populations indicated that cancer cells are characterized by high secretion of lactate in the medium and that this effect may be functional for growth (Cantor and Sabatini, 2012; Ward and Thompson, 2012). Conversely, experimental evidence (Fiaschi *et al.*, 2012; Martinez-Outschoorn *et al.*, 2011; Sanità *et al.*, 2014; Whitaker-Menezes *et al.*, 2011) of the existence of a stromal-cancer lactate shuttle in human tumours, a phenomenon named as "reverse Warburg effect" due to the fact that tumour stromal cells undergo aerobic glycolysis producing lactate that is used as energy source by the adjacent high proliferative cancer cells has been reported. By simulating scenarios in which this energy source is not fully released in the environment but must be taken up by other subpopulations, our approach effectively captured the side effects of lactate production on the overall growth of a heterogeneous tumour mass displaying the reverse Warburg effect. Under this assumption, our approach well describes the complex ("symbiotic" but also "parasitic") metabolic relationship between cancer and stromal cells in mixed cancer populations.

The agreement of our results with the above-mentioned experimental data supports the reliability of our approach and further confirms the need for computational and experimental approaches able to take into account the specificity of the subpopulations within a tumour rather than observing the average behaviour.

We have shown how popFBA may also be applied to simulate scenarios in which different biological conditions impose different uptake/secretion capabilities on the different subpopulations, as well as to investigate the metabolic plasticity of the tumour mass with respect to the adaptation of its components (i.e. different subpopulations belonging to the tumour) to changing external but also internal scenarios.

As it has already been pointed out (Resendis-Antonio *et al.*, 2015), the Warburg and reverse Warburg effect may represent only two paradigmatic examples of metabolic interconnection within cancers. Other scenarios may be simulated with popFBA, by exploiting experimental information to specifically constrain the distinct subpopulations, with particular regard to the information on single-cell transcriptome now enabled by single-cell RNA sequencing (scRNA-seq).

Taking inspiration from the plethora of existing methods to integrate transcriptomic data into classic FBA (Machado and Herrgård, 2014), we plan in the next future to define a method to integrate single cells transcriptomic data into our multiscale model, in order to pave the way to the integration of the increasing availability of scRNA-seq data into computational models. Single-cell flux distributions will be computed as a function of the transcriptome at the cell level, while assuring biomass formation at the population level.

Linear programming is computationally inexpensive, making a single popFBA computation applicable to populations composed of hundreds of clones and scalable to genome-scale metabolic networks. The computation time will increase linearly with the overall number of reactions. Conversely, the proper size of the set of sampled solutions should be assessed on a case-by-case basis, as it is likely to depend on the specific features of the metabolic network model, as well as on the set of cooperating metabolites. However, random sampling methods involve independent computations and may thus easily benefit from distributed implementations.

Funding

Conflict of Interest: none declared.

References

Bordel,S. *et al.* (2010) Sampling the solution space in genome-scale metabolic networks reveals transcriptional regulation in key enzymes. *PLoS Comput. Biol.*, **6**, e1000859.

Burrell,R.A. *et al.* (2013) The causes and consequences of genetic heterogeneity in cancer evolution. *Nature*, **501**, 338–345.

Cantor,J.R., and Sabatini,D.M. (2012) Cancer cell metabolism: one hallmark, many faces. *Cancer Disc.*, **2**, 881–898.

Cazzaniga,P. *et al.* (2014) Computational strategies for a system-level understanding of metabolism. *Metabolites*, **4**, 1034–1087.

Damiani,C. *et al.* (2014) An ensemble evolutionary constraint-based approach to understand the emergence of metabolic phenotypes. *Nat. Comput.*, **13**, 321–331.

De Martino,D. *et al.* (2013) Counting and correcting thermodynamically infeasible flux cycles in genome-scale metabolic networks. *Metabolites*, **3**, 946–966.

Di Filippo,M. *et al.* (2016) Zooming-in on cancer metabolic rewiring with tissue specific constraint-based models. *Comput. Biol. Chem.*, **62**, 60–69.

Di Filippo,M. *et al*. (2017) Constraint-Based Modeling and Simulation of Cell Populations. In: Rossi,F. *et al*. (eds) *Advances in Artificial Life, Evolutionary Computation, and Systems Chemistry*. WIVACE 2016. Communications in Computer and Information Science, vol 708. Springer, Cham.

Fiaschi,T. *et al*. (2012) Reciprocal metabolic reprogramming through lactate shuttle coordinately influences tumor-stroma interplay. *Cancer Res*., **72**, 5130–5140.

Holmes,E. *et al*. (2008) Metabolic phenotyping in health and disease. *Cell*, **134**, 714–717.

Machado,D., and Herrgård,M. (2014) Systematic evaluation of methods for integration of transcriptomic data into constraint-based models of metabolism. *PLoS Comput. Biol*., **10**, e1003580.

Mahadevan,R., and Schilling,C. (2003) The effects of alternate optimal solutions in constraint-based genome-scale metabolic models. *Metab. Eng*., **5**, 264–276.

Mardinoglu,A. *et al*. (2013) Integration of clinical data with a genome-scale metabolic model of the human adipocyte. *Mol. Syst. Biol*., **9**, 649.

Martinez-Outschoorn,U.E. *et al*. (2011) Cancer cells metabolically "fertilize" the tumor microenvironment with hydrogen peroxide, driving the Warburg effect: implications for PET imaging of human tumors. *Cell Cycle*, **10**, 2504–2520.

Marusyk,A. *et al*. (2012) Intra-tumour heterogeneity: a looking glass for cancer? *Nat. Rev. Cancer*, **12**, 323–334.

Orth,J.D. *et al*. (2010) What is flux balance analysis?. *Nat. Biotechnol*., **28**, 245–248.

Reed,J.L., and Palsson,B. (2004) Genome-scale in silico models of *E. coli* have multiple equivalent phenotypic states: assessment of correlated reaction subsets that comprise network states. *Genome Res*., **14**, 1797–1805.

Resendis-Antonio,O. *et al*. (2015). Modeling metabolism: a window toward a comprehensive interpretation of networks in cancer. *Semin. Cancer Biol*., **30**, 79–87.

Sanità,P. *et al*. (2014) Tumor-stroma metabolic relationship based on lactate shuttle can sustain prostate cancer progression. *BMC Cancer*, **14**, 154.

Schellenberger,J., and Palsson,B. (2009) Use of randomized sampling for analysis of metabolic networks. *J. Biol. Chem*., **284**, 5457–5461.

Schellenberger,J. *et al*. (2011) Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc*., **6**, 1290–1307.

Sun,X-X., and Yu,Q. (2015) Intra-tumor heterogeneity of cancer cells and its implications for cancer treatment. *Acta Pharmacol. Sin*., **36**, 1219–1227.

Ward,P.S., and Thompson,C.B. (2012) Metabolic reprogramming: a cancer hallmark even warburg did not anticipate. *Cancer Cell*, **21**, 297–308.

Whitaker-Menezes,D. *et al*. (2011) Evidence for a stromal-epithelial "lactate shuttle" in human tumors: MCT4 is a marker of oxidative stress in cancer-associated fibroblasts. *Cell Cycle*, **10**, 1772–1783.