Supplementary Information: Metabolic coordination and phase transitions in spatially distributed multi-cellular systems

Krishnadev Narayanankutty, José Antonio Pereiro-Morejon, Arián Ferrero Fernández, Valentina Onesto, Stefania Forciniti, Loretta L. del Mercato, Roberto Mulet, Andrea De Martino, David S. Tourigny, Daniele De Martino

Contents

Supplementary Notes 2

1 Model parameters and constants 2

2 Multi-cellular diffusion constraints 2

3 Inter-cellular exchanges and overflow 3

Supplementary Methods 5

1 The partition function 5

2 The sampling algorithm 5

3 The mean-field approximation 6
- 3.1 Approximation of the partition function . . . 6
- 3.2 Analytical form of $Z(\beta_G, \beta_O, \phi)$ and its derivatives . . . 8
- 3.3 Numerical form of $Z_N(\beta_G, \beta_O)$ and the critical line . . . 9

4 Background lactate term 10

5 Limiting conditions 12
- 5.1 Limiting oxygen conditions . . . 12
- 5.2 Limiting glucose conditions . . . 12

6 Inter-cellular flux correlations 13
- 6.1 Correlations define emergent behavior in balanced phase . . . 13
- 6.2 Nearest neighbor correlations . . . 13
- 6.3 Effects of dimensionality on inter-cellular correlations . . . 14

7 The experimental dataset 15
- 7.1 Flux correction due weak acid reversible proton binding . . . 15
- 7.2 Conversion of lactic acid level to pH . . . 15
- 7.3 Background lactate accumulation . . . 17

8 Inverse modeling experimental data 18
- 8.1 Likelihood function . . . 18
- 8.2 The sampling method . . . 18
- 8.3 Training and test set split: setting the hyper-parameter γ . . . 19
- 8.4 Fitting the model without correlations . . . 19
- 8.5 Fitting the model with correlations . . . 19

9 Modeling single-cell oxygen dynamics 22

SUPPLEMENTARY NOTE 1

Model parameters and constants

In Supplementary Table 1, we provide the actual values of the parameters and constants used with the metabolic network model, including the inter-cellular diffusion constraints.

Name (ID)	Value/range	Unit	Reference
Cell diameter (R)	10	µm	This work
Linear size of culture (L)	1	cm	This work
Number of cells (N)	4×10^4	cells	This work
Mean field constant (K)	40	unit-less	This work**
Average cell dry weight (m)	1	ng	[2]
Maximum molecular oxygen import (U_{O})	3	mmol/g h	[1, 23]
Maximum glucose import (U_{G})	1	mmol/g h	[1, 21]
ATP maintenance demand (L_{M})	1	mmol/g h	[1, 21]
ATP produced by fermentation	1	ATP per half glucose	[15]
ATP produced by respiration	5	ATP per oxygen (O_2)	[15]
Glucose medium concentrations (c_{G})	25	mM	This work
Oxygen medium concentrations (c_{O})	0.25	mM	This work
Glucose diffusion constants (D_{G})	600	µm^2/s	[11]
Oxygen diffusion constants (D_{O})	2000	µm^2/s	[20]
Lactate diffusion constants (D_{L})	700	µm^2/s	[11]
Proton diffusion constants ($D_{\mathrm{H^+}}$)	7000	µm^2/s	[11]
Oxygen diffusion derived maximum intake ($u_{\mathrm{O}}^{\max}$)	230	mmol/g h	This work*
Glucose diffusion derived maximum intake ($u_{\mathrm{G}}^{\max}$)	6800	mmol/g h	This work*

(*) $u_{\mathrm{X}}^{\max} = 4\pi c_{\mathrm{X}} D_{\mathrm{X}} R/m$ (**) $K = NR/L$

Supplementary Table 1: Relevant constants and parameters of the multicellular metabolic network model.

SUPPLEMENTARY NOTE 2

Multi-cellular diffusion constraints

Consider a group of N spherical cells of identical radius R, with centers placed at positions $\mathbf{r}_i$ in a three-dimensional volume V ($i = 1, \ldots, N$), such that $|\mathbf{r}_i - \mathbf{r}_j| > 2R$ for any $i \neq j$. We assume that each cell is either a net absorber (flux $u_i > 0$) or a net emitter (flux $u_i < 0$) of a certain compound whose extracellular concentration is represented by a scalar field $c \equiv c(\mathbf{r}, t|\mathbf{r}_c, \mathbf{u})$, where $\mathbf{r}_c = \{\mathbf{r}_i\}$ and $\mathbf{u} = \{u_i\}$. Assuming that no material flow is present in the extracellular fluid and that neither the positions $\mathbf{r}_c$ of cells nor $\mathbf{u}$ change in time, such a field evolves due to (i) cellular emission and absorption, and (ii) random diffusion in V, leading to the diffusive problem

$$\frac{\partial c}{\partial t} = D\nabla^2 c \quad . \tag{S 1}$$

complemented by suited boundary conditions on the cells surfaces.

We are interested in characterizing its steady state, i.e. the solutions of the *Laplace equation*

$$\nabla^2 c = 0 \quad . \tag{S 2}$$

Basic mathematical properties like existence and uniqueness of solutions for different classes of boundary conditions are presented in great detail in classical physics textbooks such as [9], Chapters 2 and 3. We focus here on features that are specifically important for the present work.

To begin with, we recall that, for a perfect, isolated spherical absorber with $c = 0$ on the surface and $c = c_\infty$ for $|\mathbf{r}| \equiv r \to \infty$, the time-dependent diffusion equation (S 1) is solved by the radial function

$$c(r, t) = c_\infty \left[1 - \frac{R}{r}\left(1 - \operatorname{erf}\frac{r - R}{\sqrt{4Dt}}\right)\right] \quad . \tag{S 3}$$

Correspondingly, using the diffusion flux $\mathbf{j} = -D\nabla c$, the cellular intake rate is given by

$$u(t) = 4\pi R^2 |j(R,t)| = 4\pi D R c_\infty \left(1 + \frac{R}{\sqrt{\pi D t}}\right) \quad . \tag{S 4}$$

For $t \to \infty$ one gets

$$u \to u_s \equiv 4\pi D R c_\infty \quad \text{and} \quad c \to c_\infty \left(1 - \frac{R}{r}\right) = c_\infty - \frac{u_s}{4\pi D r} \quad . \tag{S 5}$$

(The formula for u_s is known as Smoluchowski's formula.) In the presence of multiple absorbers, shielding effects become relevant and u_s only represents an upper bound to the intake rate of an absorber. Because of the linearity of Laplace's equation, the steady state solution with N absorbers reads

$$c(r) = c_\infty - \sum_i \frac{u_i}{4\pi D |\mathbf{r} - \mathbf{r}_i|} \quad , \tag{S 6}$$

where u_i represents the uptake flux of cell i. For the sake of stability, at the position of cell j we must have

$$c_\infty \geq \sum_i \frac{u_i}{4\pi D |\mathbf{r}_j - \mathbf{r}_i|} \quad , \tag{S 7}$$

or, equivalently,

$$u_s \geq \sum_i A_{ji} u_i \qquad (\forall j) \quad , \tag{S 8}$$

where

$$A_{ji} = \delta_{ji} + (1 - \delta_{ji}) \frac{R}{|\mathbf{r}_j - \mathbf{r}_i|} \quad . \tag{S 9}$$

(We used Kronecker's δ-symbol: $\delta_{ij} = 1$ if $i = j$, and zero otherwise.) Eq. (S 8) can be seen as a global constraint that diffusion imposes on the feasible values of fluxes u_i. Notice that expression (S 6) holds in principle also in the presence of emitters with $u_i < 0$. Finally, if there is no reservoir of particles at infinity (i.e. if $c_\infty = 0$), then (S 8) becomes

$$\sum_i A_{ji} u_i \leq 0 \qquad (\forall j) \quad . \tag{S 10}$$

SUPPLEMENTARY NOTE 3

Inter-cellular exchanges and overflow

Consider a vector $\mathbf{u} = (u_1, ..., u_N) \in \mathbb{R}^N$ and invertible square matrix $\mathbf{A} \in \mathbb{R}^{N \times N}$. In the main text these are identified with the vector of lactate flux values (i.e. $u_i = -u_{\mathrm{L}}^{(i)}$ is the flux of lactate for the i^{th} cell, note the change in sign) and diffusion constraint matrix $\mathbf{A}$. Let $\mathbf{1} = (1, ..., 1)^T$ be the N-dimensional vector of all ones and $\mathbf{e}_i$ be the N-dimensional unit coordinate vector with one in the i^{th} position and zero everywhere else. Then, if we define the extended matrix

$$\mathbf{M} = \begin{pmatrix} \mathbf{A} & \mathbf{1} & -\mathbf{1} \end{pmatrix} \tag{S 11}$$

we have the following equivalence of constraints

$$\mathbf{u}^T \cdot \mathbf{M} \geq 0 \quad \iff \quad \mathbf{u}^T \cdot \mathbf{A} \geq 0, \quad \mathbf{u}^T \cdot \mathbf{1} = \sum_{i=1}^{N} u_i = 0. \tag{S 12}$$

In terms of the main text, this is the statement that a set of lactate flux values satisfies the diffusion constraints with the additional condition that there is no net lactate production. In particular, apart from certain contrived cases, a Flux Balance Analysis optimal yield flux pattern (where all cells are at point E in Figure 1b from the main text) must have this solution form, since otherwise it would be considered sub-optimal (i.e., there would always be another solution that could obtain a higher value of the objective function by decreasing net lactate production). We now derive a general condition on $\mathbf{A}$ that guarantees the only solution to these constraints is the trivial solution $u_1 = ... = u_N = 0$.

First note that, for any non-trivial solution $\mathbf{u}$, there exists some i such that $u_i = \mathbf{u}^T \cdot \mathbf{e}_i < 0$ without loss of generality (since from the constraint $\mathbf{u}^T \cdot \mathbf{1} = 0$ then we must also have some $j \neq i$ with $u_j > 0$). We can therefore write the conditions on a non-trivial solution as

$$\mathbf{u}^T \cdot \mathbf{M} \geq 0, \quad \mathbf{u}^T \cdot \mathbf{e}_i < 0, \tag{S 13}$$

and from Farkas' Lemma we can have no such $\mathbf{u}$ if we can find a solution to the system of equations

$$\mathbf{M} \cdot \mathbf{y} = \mathbf{e}_i, \quad \mathbf{y} \geq 0. \tag{S 14}$$

Here $\mathbf{y} \geq 0$ means that all components of $\mathbf{y}$ are non-negative. Explicitly, we have

$$\begin{pmatrix} \mathbf{A} & \mathbf{1} & -\mathbf{1} \end{pmatrix} \begin{pmatrix} \mathbf{x} \\ y_2 \\ y_3 \end{pmatrix} = \mathbf{A} \cdot \mathbf{x} + (y_2 - y_3)\mathbf{1} = \mathbf{e}_i, \tag{S 15}$$

where $\mathbf{x} \geq 0$ and from now on we identify $\lambda \equiv y_2 - y_3 \in \mathbb{R}$. Since $\mathbf{A}$ is invertible, we have

$$\mathbf{x} = \mathbf{A}^{-1} \cdot (\mathbf{e}_i - \lambda \mathbf{1}) \tag{S 16}$$

and

$$x_j = \mathbf{e}_j^T \cdot \mathbf{x} = (\mathbf{A}^{-1})_{ji} - \lambda \sum_{k=1}^{N} (\mathbf{A}^{-1})_{jk} \equiv a_{ji} - \lambda B_j \tag{S 17}$$

where a_{ji} is the $(i,j)^{\text{th}}$ component of the inverse matrix $\mathbf{A}^{-1}$ and B_j is the sum over elements in the j^{th} column of $\mathbf{A}^{-1}$. Provided B_j is non-zero, from the condition $x_j \geq 0$ for all j we get that we can always find $\lambda \in \mathbb{R}$ whenever

$$\min_{j:B_j>0} \frac{a_{ji}}{B_j} \geq \max_{j:B_j<0} \frac{a_{ji}}{B_j}. \tag{S 18}$$

It therefore follows that, provided this condition on $\mathbf{A}$ is satisfied for all i, then there will be no non-trivial solution to the conditions (S 12). It can be checked numerically that this condition is satisfied for any sensible choice of inter-cell distance matrix and geometrically corresponds to the intersection of all separating hyperplanes defined by the constraints meeting at just one point. Specifically, this implies that there will be no optimal solution to the maximum yield Flux Balance Analysis problem with lactate exchange between cells.

SUPPLEMENTARY METHOD 1

The partition function

The partition function Z_N for a system of N cells is a function of Lagrange multipliers β_{G} and β_{O} given by the following $2N$-dimensional integral over the multi-cellular flux space $\mathcal{F}_N$:

$$Z_N\left(\beta_{\mathrm{G}},\beta_{\mathrm{O}}\right)=\int_{\mathcal{F}_N}\prod_{n=1}^{N}du_{\mathrm{G}}^{(n)}\,du_{\mathrm{O}}^{(n)}\,e^{\beta_{\mathrm{G}}u_{\mathrm{G}}^{(n)}}e^{\beta_{\mathrm{O}}u_{\mathrm{O}}^{(n)}}\ . \tag{S 19}$$

The partition function also serves as a moment-generating function for the expectation values and higher moments of net fluxes (normalised by the number of cells) defined with respect to the maximum entropy probability measure over multi-cellular flux space. Specifically, given the exponential form of the integrand in (S 19), we have that

$$\langle u_{\mathrm{G}}\rangle=\frac{1}{N}\frac{\partial}{\partial\beta_{\mathrm{G}}}\ln Z_N(\beta_{\mathrm{G}},\beta_{\mathrm{O}})\quad;\quad\langle u_{\mathrm{O}}\rangle=\frac{1}{N}\frac{\partial}{\partial\beta_{\mathrm{O}}}\ln Z_N(\beta_{\mathrm{G}},\beta_{\mathrm{O}}), \tag{S 20}$$

and, by stoichiometry,

$$\langle u_{\mathrm{L}}\rangle=\frac{\langle u_{\mathrm{O}}\rangle}{3}-2\,\langle u_{\mathrm{G}}\rangle\,. \tag{S 21}$$

Similarly, variances of these flux values are given by the second-order partial derivatives of Z_N

$$\sigma^2_{u_{\mathrm{G}}}\equiv\left\langle u_{\mathrm{G}}^2\right\rangle-\langle u_{\mathrm{G}}\rangle^2=\frac{1}{N^2}\frac{\partial^2}{\partial\beta_{\mathrm{G}}^2}\ln Z_N(\beta_{\mathrm{G}},\beta_{\mathrm{O}})\quad;\quad\sigma^2_{u_{\mathrm{O}}}\equiv\left\langle u_{\mathrm{O}}^2\right\rangle-\langle u_{\mathrm{O}}\rangle^2=\frac{1}{N^2}\frac{\partial^2}{\partial\beta_{\mathrm{O}}^2}\ln Z_N(\beta_{\mathrm{G}},\beta_{\mathrm{O}}). \tag{S 22}$$

For the variance of the net lactate flux, we use that

$$\left\langle u_{\mathrm{L}}^2\right\rangle=\left\langle\left(\frac{u_{\mathrm{O}}}{3}-2u_{\mathrm{G}}\right)^2\right\rangle=\frac{\left\langle u_{\mathrm{O}}^2\right\rangle}{9}+4\left\langle u_{\mathrm{G}}^2\right\rangle-\frac{4}{3}\left\langle u_{\mathrm{O}}\,u_{\mathrm{G}}\right\rangle \tag{S 23}$$

$$\langle u_{\mathrm{L}}\rangle^2=\left(\frac{\langle u_{\mathrm{O}}\rangle}{3}-2\,\langle u_{\mathrm{G}}\rangle\right)^2=\frac{\langle u_{\mathrm{O}}\rangle^2}{9}+4\,\langle u_{\mathrm{G}}\rangle^2-\frac{4}{3}\,\langle u_{\mathrm{O}}\rangle\,\langle u_{\mathrm{G}}\rangle \tag{S 24}$$

which gives

$$\sigma^2_{u_{\mathrm{L}}}=\left(\frac{\sigma_{u_{\mathrm{O}}}}{3}\right)^2+\left(-2\sigma_{u_{\mathrm{G}}}\right)^2-\frac{4}{3}\,\mathrm{Corr}\left[u_{\mathrm{O}},u_{\mathrm{G}}\right] \tag{S 25}$$

where

$$\mathrm{Corr}\left[u_{\mathrm{O}},u_{\mathrm{G}}\right]\equiv\left\langle u_{\mathrm{O}}\,u_{\mathrm{G}}\right\rangle-\langle u_{\mathrm{O}}\rangle\,\langle u_{\mathrm{G}}\rangle=\frac{1}{N^2}\frac{\partial^2}{\partial\beta_{\mathrm{G}}\partial\beta_{\mathrm{O}}}\ln Z_N(\beta_{\mathrm{G}},\beta_{\mathrm{O}}). \tag{S 26}$$

Thus, once $Z_N\left(\beta_{\mathrm{G}},\beta_{\mathrm{O}}\right)$ is known, one can obtain the expectation values and variances of any net flux value as a function of β_{G} and β_{O}.

We highlight that, in a more general setting, it is of course possible to introduce $2N$ Lagrange multipliers, one for each cell, that constrain expectation values of each single-cell flux value individually. In that case, the partition function would depend on $2N$ parameters and individual single-cell moments of the flux distribution obtained by differentiation in the same way as described for net flux values above. However, in this work we restrict our focus to a simple, two-parameter model with the goal of parsimoniously reproducing statistical features of existing experimental data and describing the collective behavior of multi-cellular flux distributions. More complex models of the same nature could be applied to experimental data that becomes available in the future.

SUPPLEMENTARY METHOD 2

The sampling algorithm

The multi-cellular metabolic model defines a high-dimensional convex polytope ($D=2N\approx300$, where N is the number of cells) and the computational task at hand is to characterize this space, in this case with flux distributions weighted by a Boltzmann factor defined by the maximum entropy constraint. This problem is connected to a class of NP-hard computational problems such as the computation of a matrix permanent, volume

of a high-dimensional convex body or the Ising partition function, which can be solved numerically in polynomial time using Markov chain Monte Carlo algorithms [19]. In particular, over-relaxed algorithms like hit-and-run [22] have been shown to work very well in the context of constraint-based genome-scale metabolic modeling, once the ill-conditioning problems connected to heterogeneous scales have been tackled by approximate ellipsoidal rounding [6]. In this work we employed an hit-and-run Markov chain to sample the space, outlined schematically by the following workflow:

0. Initial data: a point P_0 inside the polytope (can be found with a relaxation algorithm such as [13]).
1. Given point P_i, generate a direction/versor $\hat{n}$ uniformly at random (a point on a unit hyper-sphere, e.g. with the Marsaglia method [12]).
2. Find the intersections t_1, t_2 of the line $L(t) = P_i + t\hat{n}, \quad t \in \mathbb{R}$ with the boundary of the polytope.
3. Extract $t^* \in [t_1, t_2]$ by inverting the cumulative distribution function of the marginalized Boltzmann distribution over the segment [12]. Set $P_{i+1} = P_i + t^*\hat{n}$. Return to step 1 (or end the algorithm if you think you have enough points).

It is interesting to notice that, by comparison with the case of sampling steady states of bulk genome scale metabolic networks, the ill-conditioning problem is much less severe in our case. This is due the symmetric structure of the space, which is given by the product of N identical single-cell metabolic flux spaces (plus the diffusion constraints). A code implementation of the sampling algorithm is provided in the repository https://github.com/KrishnadevN/MulticellularMetabolicNetworks.

SUPPLEMENTARY METHOD 3

The mean-field approximation

3.1 Approximation of the partition function

The full partition function (S 19) can be written explicitly as

$$Z_N(\beta_{\mathrm{G}}, \beta_{\mathrm{O}}) = \int_0^{U_{\mathrm{G}}} \int_0^{U_{\mathrm{O}}} \cdots \int_0^{U_{\mathrm{G}}} \int_0^{U_{\mathrm{O}}} \prod_{n=1}^{N} du_{\mathrm{G}}^{(n)}\, du_{\mathrm{O}}^{(n)}\, e^{\beta_{\mathrm{G}} u_{\mathrm{G}}^{(n)} + \beta_{\mathrm{O}} u_{\mathrm{O}}^{(n)}}\, \theta\big(f_{\mathrm{ATP}}^{(n)} - L_{\mathrm{M}}\big)\, \theta\Big(-u_{\mathrm{L}}^{(n)} - \sum_{\substack{m=1 \\ m\neq n}}^{N} u_{\mathrm{L}}^{(m)} A_{mn}\Big)$$

where we have used the Heaviside step function

$$\theta(x) \equiv \begin{cases} 1, & x \geq 0 \\ 0, & x < 0, \end{cases}$$

to impose the constraints of minimal ATP demand and those given by the diffusion of lactate in the medium. Here, $u_{\mathrm{L}}^{(n)}$ and $f_{\mathrm{ATP}}^{(n)}$ are related to the independent variables $u_{\mathrm{G}}^{(n)}$ and $u_{\mathrm{O}}^{(n)}$ as

$$u_{\mathrm{L}}^{(n)} = \frac{u_{\mathrm{O}}^{(n)}}{3} - 2u_{\mathrm{G}}^{(n)} \quad \text{and} \quad f_{\mathrm{ATP}}^{(n)} = 2u_{\mathrm{G}}^{(n)} + \frac{14}{3} u_{\mathrm{O}}^{(n)}. \tag{S 27}$$

To obtain an exact expression for Z_N, we make the approximation that $A_{mn} \approx K/N$ for all cells $m \neq n$, which is equivalent to the mean-field approximation that all cells are fully connected and equally distanced, as discussed in the main text. It is important to remark that this approximation also assumes the system to be homogeneous, whereas in general the partition function could depend on the specific spatial positions of individual cells. This homogeneity assumption also means that the moments of the single-cell distribution do not depend on the cell index (n), i.e. the fluxes are identically distributed (albeit correlated) random variables. In particular, when $N \gg 1$ we have

$$\frac{1}{N} \sum_{\substack{m=1 \\ m\neq n}}^{N} u_{\mathrm{L}}^{(m)} \approx \frac{1}{N} \sum_{m=1}^{N} u_{\mathrm{L}}^{(m)} \equiv \overline{u_{\mathrm{L}}}, \tag{S 28}$$

i.e., the contribution of any one cell to the arithmetic mean across cells, $\overline{u_{\mathrm{L}}}$, becomes negligible. With this approximations, the inter-cellular constraint originating from the diffusion of lactate becomes

$$\theta\Big(-u_{\mathrm{L}}^{(n)} - \sum_{\substack{m=1 \\ m\neq n}}^{N} u_{\mathrm{L}}^{(m)} A_{mn}\Big) \approx \theta\Big(-u_{\mathrm{L}}^{(n)} - K\,\overline{u_{\mathrm{L}}}\Big) \tag{S 29}$$

so that

$$Z_N(\beta_{\rm G},\beta_{\rm O}) \approx \int_0^{U_{\rm G}} \int_0^{U_{\rm O}} \cdots \int_0^{U_{\rm G}} \int_0^{U_{\rm O}} \prod_{n=1}^{N} du_{\rm G}^{(n)}\, du_{\rm O}^{(n)}\, e^{\beta_{\rm G} u_{\rm G}^{(n)} + \beta_{\rm O} u_{\rm O}^{(n)}}\, \theta\big(f_{\rm ATP}^{(n)} - L_{\rm M}\big)\, \theta\big(-u_{\rm L}^{(n)} - K\overline{u_{\rm L}}\big). \tag{S 30}$$

To manipulate the above integral into a manageable form, we introduce ϕ and λ through the properties of the Dirac delta and its Fourier transform, to express unity as

$$\begin{aligned} 1 &= N \int_{-\infty}^{\infty} \delta\left(N\phi - \sum_{n=1}^{N} u_{\rm L}^{(n)}\right) d\phi && \text{(S 31)}\\ &= \frac{N}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left(\lambda\left[N\phi - \sum_{n=1}^{N} u_{\rm L}^{(n)}\right]\right) d\phi\, d\lambda && \text{(S 32)}\\ &= \frac{N}{2\pi} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\lambda\, e^{i\lambda N\phi} \prod_{n=1}^{N} e^{2i\lambda u_{\rm G}^{(n)} - i\lambda u_{\rm O}^{(n)}/3}. && \text{(S 33)} \end{aligned}$$

Inserting this into the integrand of (S 30), we obtain the identity

$$\prod_{n=1}^{N} e^{\beta_{\rm G} u_{\rm G}^{(n)} + \beta_{\rm O} u_{\rm O}^{(n)}}\, \theta\big(-u_{\rm L}^{(n)} - K\overline{u_{\rm L}}\big) = \frac{N}{2\pi} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\lambda\, e^{i\lambda N\phi} \prod_{n=1}^{N} e^{(\beta_{\rm G}+2i\lambda)u_{\rm G}^{(n)} + (\beta_{\rm O} - i\lambda/3)u_{\rm O}^{(n)}}\, \theta\Big(-u_{\rm L}^{(n)} - K\phi\Big),$$

which enables the approximation of the partition function (S 30) to be written as

$$\begin{aligned} Z_N\left(\beta_{\rm G},\beta_{\rm O}\right) &\approx \frac{N}{2\pi} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\lambda\, e^{i\lambda N\phi} [Z\big(\beta_{\rm G} + 2i\lambda,\, \beta_{\rm O} - i\lambda/3,\, \phi\big)]^N && \text{(S 34)}\\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} d\phi \int_{-\infty}^{\infty} d\lambda\, \exp\left[NF(\beta_{\rm G},\beta_{\rm O},\phi,i\lambda)\right] && \text{(S 35)} \end{aligned}$$

with

$$Z(\beta_{\rm G},\beta_{\rm O},\phi) = \int_0^{U_{\rm G}} \int_0^{U_{\rm O}} du_{\rm G} du_{\rm O}\, e^{\beta_{\rm G} u_{\rm G}} e^{\beta_{\rm O} u_{\rm O}}\, \theta\big(f_{\rm ATP} - L_{\rm M}\big)\, \theta\big(-u_{\rm L} - K\phi\big) \tag{S 36}$$

and

$$F(\beta_{\rm G},\beta_{\rm O},\phi,p) = p\phi + \log Z\big(\beta_{\rm G} + 2p,\, \beta_{\rm O} - p/3,\, \phi\big) + \frac{1}{N}\log N. \tag{S 37}$$

The introduction of ϕ and λ has thus allowed us to "decouple" the diffusion constraints to study the effective two-dimensional, single-cell flux space on which Z is defined.

Finally, to evaluate (S 35) in the limit $N \to \infty$, where the $\log(N)/N$ term in (S 37) can safely be ignored, we use the method of steepest descent (also called the saddle point approximation) [3]. For $N \gg 1$, we obtain

$$\frac{1}{N}\log Z_N\left(\beta_{\rm G},\beta_{\rm O}\right) \;\approx\; p_*\phi_* + \log Z\big(\beta_{\rm G} + 2p_*,\, \beta_{\rm O} - p_*/3,\, \phi_*\big) \tag{S 38}$$

where $(\phi_*, p_* \equiv i\lambda^*)$ is the stationary point given by

$$\left.\frac{\partial F}{\partial \phi}\right|_{(\phi_*,p_*)} = 0 \qquad \text{and} \qquad \left.\frac{\partial F}{\partial p}\right|_{(\phi_*,p_*)} = 0. \tag{S 39}$$

In particular, from (S 37), these saddle point equations correspond to the self-consistency equations

$$p^*(\beta_{\rm G},\beta_{\rm O}) = -\frac{\partial}{\partial\phi}\ln Z\big(\beta_{\rm G} + 2p_*,\, \beta_{\rm O} - p_*/3,\, \phi_*\big) \tag{S 40}$$

$$\phi^*(\beta_{\rm G},\beta_{\rm O}) = -\frac{\partial}{\partial p}\ln Z\big(\beta_{\rm G} + 2p_*,\, \beta_{\rm O} - p_*/3,\, \phi_*\big) \tag{S 41}$$

where the explicit dependence of p^* and ϕ^* on $\beta_{\rm G}$ and $\beta_{\rm O}$ is indicated. In fact, by using the chain rule we see that

$$\phi^*(\beta_{\rm G},\beta_{\rm O}) = \left(\frac{1}{3}\frac{\partial}{\partial\beta_{\rm O}} - 2\frac{\partial}{\partial\beta_{\rm G}}\right)\log Z\big(\beta_{\rm G} + 2p_*,\, \beta_{\rm O} - p_*/3,\, \phi_*\big) \equiv \langle\!\langle u_{\rm L}\rangle\!\rangle \tag{S 42}$$

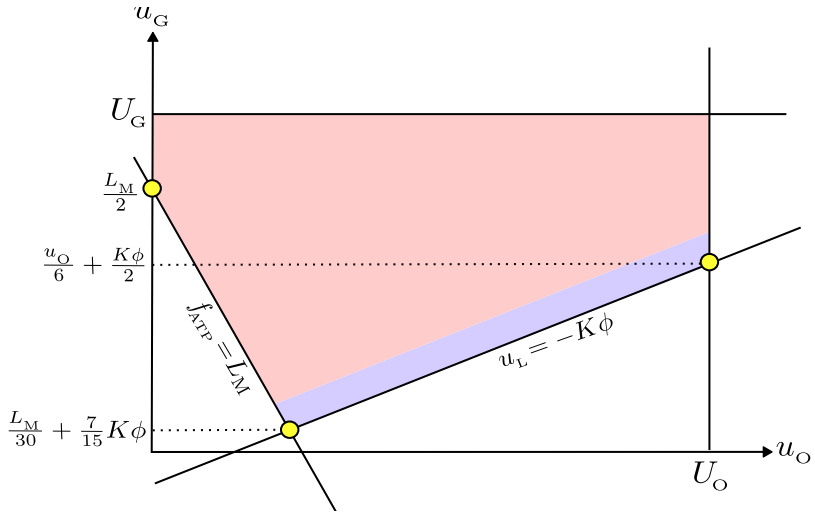
where $\langle\!\langle\cdots\rangle\!\rangle$ is the expectation value defined using the measure associated with Z on the effective single-cell space. Namely, from (S 36) and (S 42) with $\beta_{\rm G}^* \equiv \beta_{\rm G} + 2p^*$ and $\beta_{\rm O}^* = \beta_{\rm O} - p^*/3$ we have, self-consistently,

$$\langle\!\langle\cdots\rangle\!\rangle \equiv \frac{1}{Z(\beta_{\rm G}^*,\beta_{\rm O}^*,\langle\!\langle u_{\rm L}\rangle\!\rangle)} \int_0^{U_{\rm G}} \int_0^{U_{\rm O}} du_{\rm G} du_{\rm O}\, (\cdots) e^{\beta_{\rm G}^* u_{\rm G}} e^{\beta_{\rm O}^* u_{\rm O}}\, \theta\big(f_{\rm ATP} - L_{\rm M}\big)\, \theta\big(-u_{\rm L} - K\langle\!\langle u_{\rm L}\rangle\!\rangle\big). \tag{S 43}$$

The values ϕ^* and p^* therefore permit approximation of Z_N using (S 38) once Z and its derivatives are known. We provide exact analytical formulae for these in the next subsection.

3.2 Analytical form of $Z(\beta_G, \beta_O, \phi)$ and its derivatives

We now proceed to evaluate the integral defined in (S 36) to find an analytical formula for $Z(\beta_G, \beta_O, \phi)$ and its derivatives, which are required for computation of the full partition function Z_N. We recall that Z is defined on the effective single-cell space illustrated in Supplementary Figure 1, restricted to the domain $\phi \leq 0$ due to its evaluation at ϕ^*, as described in the previous subsection. However, we remind the reader that the self-consistent dependence of ϕ^* on β_G and β_O is only imposed after the saddle point approximation so that here ϕ is treated as an independent argument of the three-variable function $Z(\beta_G, \beta_O, \phi)$. The resulting domain of integration can be intuitively thought of as the single-cell flux space from the main text (Figure 1b) combined with an additional upper bound on the rate of lactate uptake, given by the constraint $u_L \leq -K\phi$.



Supplementary Figure 1: The effective single-cell flux space (in color) given by the limits of integration for $Z(\beta_G, \beta_O, \phi)$. Analogously to the single-cell flux space described in the main text, the pink region has the physical interpretation of lactate export and the purple region denotes lactate import.

In terms of variables u_G, u_O, these constraints give the following limits of integration for the integral in (S 36):

$$u_G^{\min} \leq u_G \leq U_G \quad ; \quad u_O^{\min} \leq u_O \leq u_O^{\max} \tag{S 44}$$

where,

$$\begin{aligned}
u_G^{\min} &= \max\left(0,\ \frac{L_M}{30} + \frac{7}{15}K\phi\right) \\
u_O^{\min} &= \begin{cases} \frac{3}{14}\left(L_M - 2u_G\right), & u_G^{\min} \leq u_G \leq u_G^a \\ 0, & u_G^a \leq u_G \leq U_G \end{cases}; \quad u_G^a \equiv \min\left(\frac{L_M}{2},\ U_G\right) \\
u_O^{\max} &= \begin{cases} 6u_G - 3K\phi, & u_G^{\min} \leq u_G \leq u_G^b \\ U_O, & u_G^b \leq u_G \leq U_G \end{cases}; \quad u_G^b \equiv \max\left(0,\ \frac{U_O}{6} + \frac{K\phi}{2}\right).
\end{aligned} \tag{S 45}$$

We evaluate

$$Z(\beta_G, \beta_O, \phi) = \int_{u_G^{\min}}^{U_G} du_G \int_{u_O^{\min}}^{u_O^{\max}} du_O\, e^{\beta_G u_G} e^{\beta_O u_O} \tag{S 46}$$

$$= \frac{1}{\beta_O}\int_{u_G^{\min}}^{U_G} du_G\, e^{\beta_G u_G} e^{\beta_O u_O^{\max}} - \frac{1}{\beta_O}\int_{u_G^{\min}}^{U_G} du_G\, e^{\beta_G u_G} e^{\beta_O u_O^{\min}}. \tag{S 47}$$

Since the dependence of $u_O^{\min}$ and $u_O^{\max}$ on u_G changes at u_G^a and u_G^b, respectively, we split each integral into two and obtain

$$\begin{aligned}
Z(\beta_G, \beta_O, \phi) \ = \ & \frac{1}{\beta_O}\int_{u_G^{\min}}^{u_G^b} du_G\, e^{(\beta_G + 6\beta_O)u_G} e^{-3K\phi\beta_O} + \frac{1}{\beta_O}\int_{u_G^b}^{U_G} du_G\, e^{\beta_G u_G} e^{\beta_O U_O} \\
& - \frac{1}{\beta_O}\int_{u_G^{\min}}^{u_G^a} du_G\, e^{(\beta_G - \frac{3}{7}\beta_O)u_G} e^{\frac{3}{14}L_M\beta_O} - \frac{1}{\beta_O}\int_{u_G^a}^{U_G} du_G\, e^{\beta_G u_G}
\end{aligned} \tag{S 48}$$

which finally gives

$$\begin{aligned} Z(\beta_{\rm G},\beta_{\rm O},\phi) \;=\; & \frac{e^{-3K\phi\beta_{\rm O}}}{\beta_{\rm O}}\left[\frac{e^{\beta_1 u_{\rm G}^b}}{\beta_1}-\frac{e^{\beta_1 u_{\rm G}^{\rm min}}}{\beta_1}\right]+\frac{e^{U_{\rm O}\beta_{\rm O}}}{\beta_{\rm O}}\left[\frac{e^{\beta_{\rm G}U_{\rm G}}}{\beta_{\rm G}}-\frac{e^{\beta_{\rm G}u_{\rm G}^b}}{\beta_{\rm G}}\right] \\ & -\frac{e^{\frac{3}{14}L_{\rm M}\beta_{\rm O}}}{\beta_{\rm O}}\left[\frac{e^{\beta_2 u_{\rm G}^a}}{\beta_2}-\frac{e^{\beta_2 u_{\rm G}^{\rm min}}}{\beta_2}\right]-\frac{1}{\beta_{\rm O}}\left[\frac{e^{\beta_{\rm G}U_{\rm G}}}{\beta_{\rm G}}-\frac{e^{\beta_{\rm G}u_{\rm G}^a}}{\beta_{\rm G}}\right] \end{aligned} \tag{S 49}$$

where

$$\beta_1 \equiv \beta_{\rm G}+6\beta_{\rm O} \quad \text{and} \quad \beta_2 \equiv \beta_{\rm G}-\frac{3}{7}\beta_{\rm O}. \tag{S 50}$$

For evaluating the derivatives, we first note that

$$\frac{\partial}{\partial\beta}\left(\frac{e^{\beta' u}}{\beta'}\right)=\left(u\frac{\partial\beta'}{\partial\beta}+\beta'\frac{\partial u}{\partial\beta}-\frac{1}{\beta'}\frac{\partial\beta'}{\partial\beta}\right)\frac{e^{\beta' v}}{\beta'}. \tag{S 51}$$

From (S 50), we have

$$\frac{\partial\beta_1}{\partial\beta_{\rm G}}=1 \quad ; \quad \frac{\partial\beta_1}{\partial\beta_{\rm O}}=6 \quad ; \quad \frac{\partial\beta_2}{\partial\beta_{\rm G}}=1 \quad ; \quad \frac{\partial\beta_2}{\partial\beta_{\rm O}}=\frac{-3}{7}. \tag{S 52}$$

For convenience of notation, we also define

$$A=-3K\phi-\frac{1}{\beta_{\rm O}}-\frac{6}{\beta_1} \quad ; \quad B=\frac{3}{14}L_{\rm M}-\frac{1}{\beta_{\rm O}}+\frac{3}{7}\frac{1}{\beta_2}. \tag{S 53}$$

Using the above, we differentiate (S 49) term-by-term to obtain

$$\begin{aligned} \frac{\partial Z}{\partial\beta_{\rm G}}(\beta_{\rm G},\beta_{\rm O},\phi) \;=\; & \frac{e^{-3K\phi\beta_{\rm O}}}{\beta_{\rm O}}\left[\left(u_{\rm G}^b-\frac{1}{\beta_1}\right)\frac{e^{\beta_1 u_{\rm G}^b}}{\beta_1}-\left(u_{\rm G}^{\rm min}-\frac{1}{\beta_1}\right)\frac{e^{\beta_1 u_{\rm G}^{\rm min}}}{\beta_1}\right] \\ & +\frac{e^{U_{\rm O}\beta_{\rm O}}}{\beta_{\rm O}}\left[\left(U_{\rm G}-\frac{1}{\beta_{\rm G}}\right)\frac{e^{\beta_{\rm G}U_{\rm G}}}{\beta_{\rm G}}-\left(u_{\rm G}^b-\frac{1}{\beta_{\rm G}}\right)\frac{e^{\beta_{\rm G}u_{\rm G}^b}}{\beta_{\rm G}}\right] \\ & -\frac{e^{\frac{3}{14}L_{\rm M}\beta_{\rm O}}}{\beta_{\rm O}}\left[\left(u_{\rm G}^a-\frac{1}{\beta_2}\right)\frac{e^{\beta_2 u_{\rm G}^a}}{\beta_2}-\left(u_{\rm G}^{\rm min}-\frac{1}{\beta_2}\right)\frac{e^{\beta_2 u_{\rm G}^{\rm min}}}{\beta_2}\right] \\ & -\frac{1}{\beta_{\rm O}}\left[\left(U_{\rm G}-\frac{1}{\beta_{\rm G}}\right)\frac{e^{\beta_{\rm G}U_{\rm G}}}{\beta_{\rm G}}-\left(u_{\rm G}^a-\frac{1}{\beta_{\rm G}}\right)\frac{e^{\beta_{\rm G}u_{\rm G}^a}}{\beta_{\rm G}}\right] \end{aligned} \tag{S 54}$$

and

$$\begin{aligned} \frac{\partial Z}{\partial\beta_{\rm O}}(\beta_{\rm G},\beta_{\rm O},\phi) \;=\; & \frac{e^{U_{\rm O}\beta_{\rm O}}}{\beta_{\rm O}}\left(U_{\rm O}-\frac{1}{\beta_{\rm O}}\right)\left[\frac{e^{\beta_{\rm G}U_{\rm G}}}{\beta_{\rm G}}-\frac{e^{\beta_{\rm G}u_{\rm G}^b}}{\beta_{\rm G}}\right]+\frac{1}{\beta_{\rm O}^2}\left[\frac{e^{\beta_{\rm G}U_{\rm G}}}{\beta_{\rm G}}-\frac{e^{\beta_{\rm G}u_{\rm G}^a}}{\beta_{\rm G}}\right] \\ & +\frac{e^{-3K\phi\beta_{\rm O}}}{\beta_{\rm O}}\left[\left(A+6u_{\rm G}^b\right)\frac{e^{\beta_1 u_{\rm G}^b}}{\beta_1}-\left(A+6u_{\rm G}^{\rm min}\right)\frac{e^{\beta_1 u_{\rm G}^{\rm min}}}{\beta_1}\right] \\ & -\frac{e^{\frac{3}{14}L_{\rm M}\beta_{\rm O}}}{\beta_{\rm O}}\left[\left(B-\frac{3}{7}u_{\rm G}^a\right)\frac{e^{\beta_2 u_{\rm G}^a}}{\beta_2}-\left(B-\frac{3}{7}u_{\rm G}^{\rm min}\right)\frac{e^{\beta_2 u_{\rm G}^{\rm min}}}{\beta_2}\right]. \end{aligned} \tag{S 55}$$

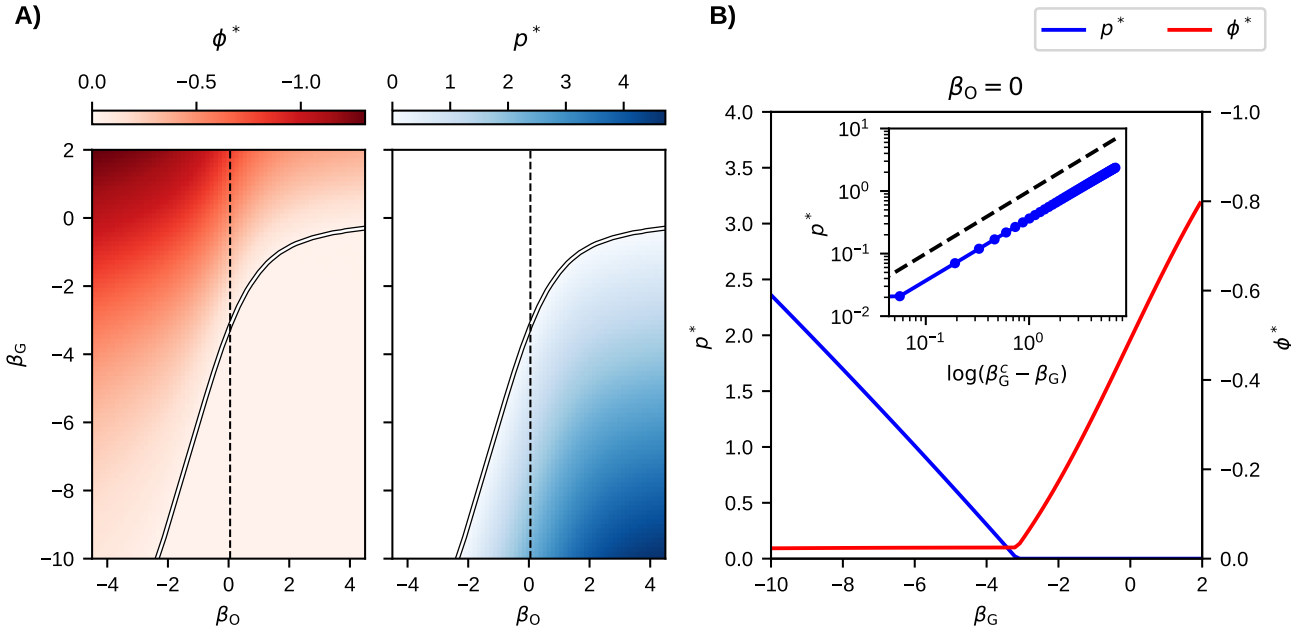
These analytical formulae form the basis of the self-consistency equations (S 41) that we solve to obtain Z_N and its derivatives numerically, as described in the next subsection.

3.3 Numerical form of $Z_N(\beta_{\rm G},\beta_{\rm O})$ and the critical line

Using the results of the previous two subsections, we are now in a position to calculate Z_N from (S 38) using ϕ^* and p^* as functions of $\beta_{\rm G}$ and $\beta_{\rm O}$, obtained numerically from the self-consistency equations (S 40) and (S 41). Python code implementing this procedure is provided in the repository https://github.com/KrishnadevN/MulticellularMetabolicNetworks. Importantly, first- and second-order derivatives of Z_N then give the expectation values and variances of net flux values, respectively, as described in Supplementary Method 1. These are compared with the values obtained from sampling multi-cellular flux space in Figure 3a-c from the main

text. We also confirmed that numerical values for the expectation value of the net fluxes overlap nearly identically with those calculated under the mean-field measure (S 43), e.g. $\langle u_{\mathrm{L}} \rangle \approx \langle\!\langle u_{\mathrm{L}} \rangle\!\rangle$, as might be expected in the $N \to \infty$ limit.

We remark that the analytical form of the self-consistency equations (S 40) and (S 41) reveals some important insights into the nature of the phase-transition described by the mean-field model. Inspecting the limits of integration (S 45), we find a critical value $\phi_c = -U_{\mathrm{O}}/3K$ such that the constraints remain constant for $\phi^* < \phi_c$. Thus, $p^* = 0$ in this region of $(\beta_{\mathrm{G}}, \beta_{\mathrm{O}})$ space, identified with the overflow phase and net lactate production as displayed in Supplementary Figure 2 and in Figure 3d from the main text. Conversely, in the regime $0 > \phi^* > \phi_c$ the limits of integration become dependent on the value of ϕ^* such that p^* steadily grows as ϕ^* slowly plateaus towards zero. This regime therefore corresponds to the balanced phase with minimal net lactate production. The overall effect is a critical line in phase space that delineates the separation between the balanced and overflow regimes (Supplementary Figure 2), defined by all critical values $(\beta_{\mathrm{G}}^c, \beta_{\mathrm{O}}^c)$ such that $\phi^*(\beta_{\mathrm{G}}^c, \beta_{\mathrm{O}}^c) = \phi_c$. Equivalently, this can be represented as a critical value of $\langle\!\langle u_{\mathrm{L}} \rangle\!\rangle = -U_{\mathrm{O}}/3K$, as displayed in Figure 4g from the main text.



Supplementary Figure 2: (A) Phase diagrams in the $(\beta_{\mathrm{G}}, \beta_{\mathrm{O}})$ plane: order parameters ϕ^* (left) and p^* (right). The phase diagram is divided by the critical line (continuous white). (B) The order parameters as a function of β_{G} (for $\beta_{\mathrm{O}} = 0$), corresponding to the dashed line in (A). Inset: p^* as a function of $(\beta_{\mathrm{G}}^c - \beta_{\mathrm{G}})$ in logarithmic scale showing that the critical exponent is $\alpha = 1$. Here, β_{G}^c is the critical value of β_{G} corresponding to $\beta_{\mathrm{O}} = 0$.

SUPPLEMENTARY METHOD 4

Background lactate term

While the model outlined in the previous sections forbids the net uptake of lactate (assuming none is exogenously supplied), any excess lactate produced by the culture will eventually accumulate in the growth medium to become available for net uptake at later time points. It is straightforward to extend the model to include a background lactate concentration term (and therefore the possibility of net uptake of lactate by the culture) in the inter-cellular diffusion constraints. In presence of a background term, these read

$$\sum_i A_{ij} u_{\mathrm{L}}^{(i)} \leq u_{\mathrm{L}}^{\max} \qquad (\forall j) \tag{S 56}$$

where the maximum flux limited by diffusion $u_{\mathrm{L}}^{\max}$ is given by the formula

$$u_{\mathrm{L}}^{\max} = 4\pi c_{\mathrm{L},b} D_{\mathrm{L}} R/m \tag{S 57}$$

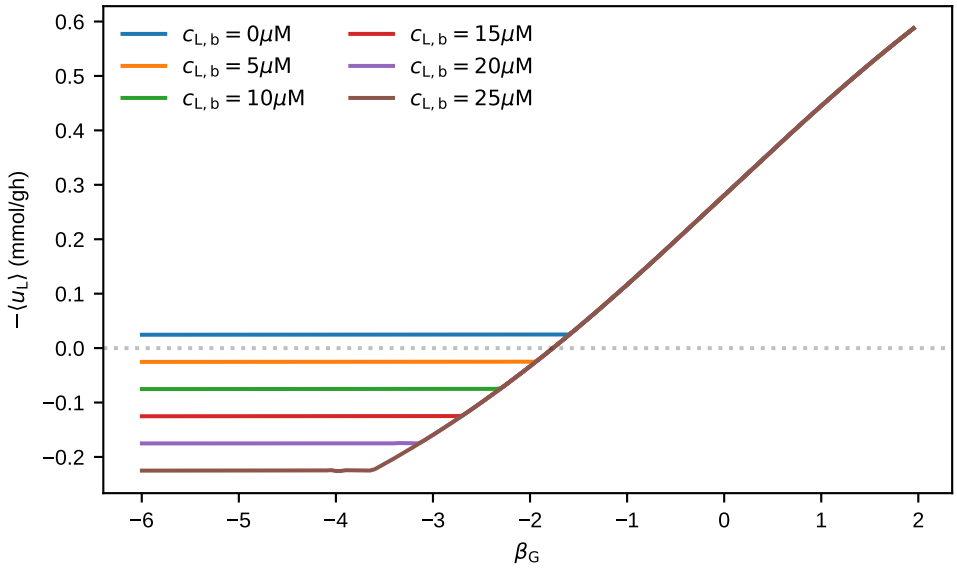
with $c_{\mathrm{L},b}$ the background concentration of lactate, D_{L} is the diffusion coefficient of lactate, R is the cell size and m is the cell mass. The mean-field equations in presence of a background lactate term are then modified by substituting

$$K\langle u_{\mathrm{L}} \rangle \to K\langle u_{\mathrm{L}} \rangle - u_{\mathrm{L}}^{\max}, \tag{S 58}$$

which results in a shift of the critical point to

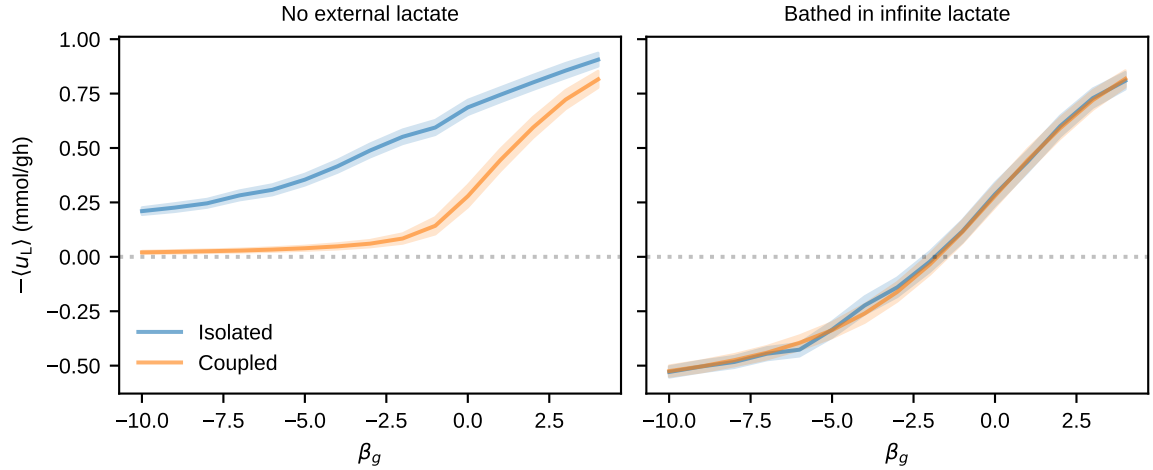
$$\phi_c = -U_{\mathrm{O}}/3K \to -U_{\mathrm{O}}/3K + u_{\mathrm{L}}^{\max}/K \tag{S 59}$$

and could potentially revert its sign. We illustrate this in Fig. 3 showing the overflow transition for a range of relevant $c_{\mathrm{L},b}$ values and parameter values taken from Table 1.



Supplementary Figure 3: Plot of the mean-field lactate flux as a function of β_{G} at fixed $\beta_{\mathrm{O}} = 1$ for a range of relevant $c_{\mathrm{L},b}$ concentrations (0-25 μM).

Eventually, in the limit $c_{\mathrm{L},b} \to \infty$, we find that ϕ_c becomes larger than the maximum uptake permitted by the oxidative capacity constraint, which causes the phase transition to disappear from the model. In this limit, the model is equivalent to a system with N isolated single cells (where the inter-cellular diffusion constraints are trivially satisfied). This is illustrated in Fig. 4 where we show results from simulations of the full spatial model and compare these to those from a model with isolated single cells.



Supplementary Figure 4: Numerical simulations of average lactate flux as a function of β_{G} at fixed $\beta_{\mathrm{O}} = 1$ for $N = 150$ isolated or coupled cells. Simulations were run in the extreme cases of zero background lactate (left, $c_{\mathrm{L},b} = 0$) or an infinite reservoir of lactate (right, $c_{\mathrm{L},b} \to \infty$). Shaded regions represent standard deviation on the mean.

SUPPLEMENTARY METHOD 5

Limiting conditions

5.1 Limiting oxygen conditions

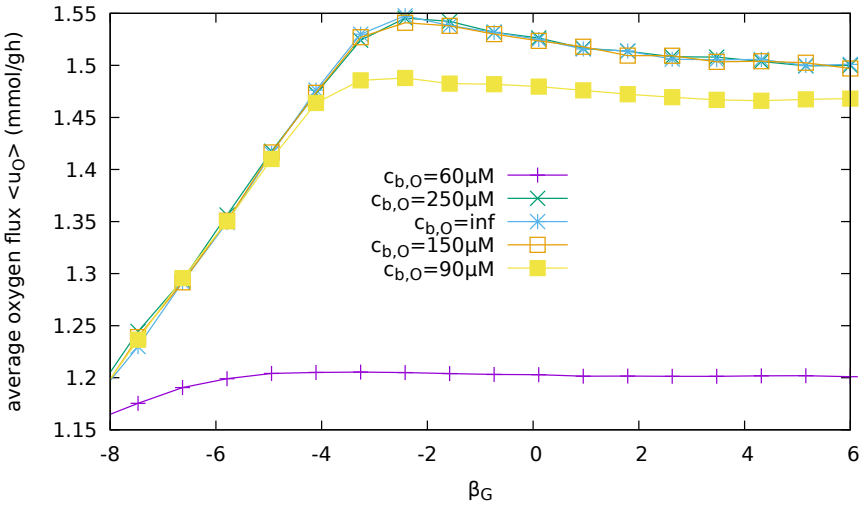
Using the same approach for inclusion of the background lactate term, we can establish conditions where oxygen concentrations are not limiting for the critical behavior predicted by our model. Inter-cellular diffusion constraints for oxygen read

$$\sum_i A_{ij} u_{\mathrm{O}}^{(i)} \leq u_{\mathrm{O}}^{\max} \qquad (\forall j) \tag{S 60}$$

where an important difference compared to the case with lactate is that oxygen can only be imported ($u_{\mathrm{O}}^{(i)} \geq 0$). From Table 1 we have

$$u_{\mathrm{O}}^{\max}/K \sim 5.75 \text{ mmol/g h}. \tag{S 61}$$

Since this quantity represents the average local oxygen concentration experienced by single cells and is larger than the maximum oxidative capacity ($U_{\mathrm{O}} = 3$ mmol/g h), we can rationalize that intercellular diffusion constraints play little or no role for our the experimental conditions outlined in Supplementary Method 7 (in particular in regard to cellular density). We demonstrate in Supplementary Figure 5 simulations of the spatial system at different background oxygen levels by plotting the average oxygen flux as a function of β_{G} (at fixed $\beta_{\mathrm{O}} = 1$) for $c_{O,b} \to \infty$ (ie not including the constraints), $c_{O,b} = 250\mu$M (room conditions as in experiments) and different levels of hypoxia $c_{O,b} = 150, 80, 60\mu$M . As expected we do not see changes for our experimental conditions while hypoxia starts to matter below $\sim 100\mu$M.



Supplementary Figure 5: Average oxygen flux as a function of β_{G} at fixed $\beta_{\mathrm{O}} = 1$ for several $c_{O,b}$ (in μM) from simulations.

5.2 Limiting glucose conditions

Analogously to oxygen, the inter-cellular diffusion constraints for glucose are

$$\sum_i A_{ij} u_{\mathrm{G}}^{(i)} \leq u_{\mathrm{G}}^{\max} \qquad (\forall j) \tag{S 62}$$

where once again glucose can only be imported ($u_{\mathrm{G}}^{(i)} \geq 0$). From Table 1 we have

$$u_{\mathrm{G}}^{\max}/K \sim 170 \text{ mmol/g h} \tag{S 63}$$

This is far higher than $U_{\mathrm{G}} = 1$ mmol/g h and we can therefore safely assume these constraints play no role in the model, provided glucose is not substantially depleted over time (a simple estimate indicate that within the 6h of experiments described in Supplementary Method 7 at most 10% is consumed). We note that such an abundance of glucose is a particular feature of standard cultivation media for the cell line being analyzed.

SUPPLEMENTARY METHOD 6

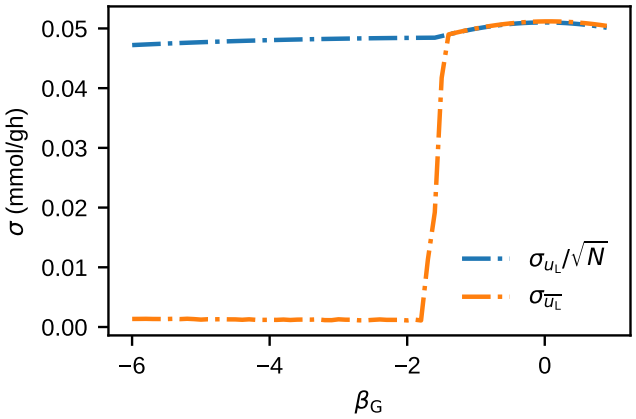
Inter-cellular flux correlations

6.1 Correlations define emergent behavior in balanced phase

In Figure 3 of the main text, we show that the mean-field model quantitatively reproduces the trends of the average fluxes as a function of the control parameters $(\beta_{\mathrm{O}}, \beta_{\mathrm{G}})$. The mean-field approximation also highlighted that the two phases (overflow and balanced) differ by the level of cell-cell lactate flux correlations, which become negative in the balanced phase. This can be seen in Supplementary Figure 6 that displays, as a function of β_{G} (at fixed $\beta_{\mathrm{O}} = 1$) the standard deviation of the lactate flux for a single cell, $\sigma_{u_{\mathrm{L}}}$ (rescaled by $1/\sqrt{N}$), compared to the standard deviation of the average across cells, $\sigma_{\bar{u}_{\mathrm{L}}}$. These two quantities are related by the formula

$$\sigma^2_{\bar{u}_{\mathrm{L}}} = \sigma^2_{u_{\mathrm{L}}}/N + \frac{1}{N^2}\sum_{i\neq j}\langle (u_{\mathrm{L}}^{(i)} - \langle u_{\mathrm{L}}\rangle)(u_{\mathrm{L}}^{(j)} - \langle u_{\mathrm{L}}\rangle)\rangle, \tag{S 64}$$

where the second term on the right-hand side represents the total inter-cellular correlations. For this setup,



Supplementary Figure 6: Standard deviation of the lactate flux for a single cell $\sigma_{u_{\mathrm{L}}}$, rescaled by $1/\sqrt{N}$, and standard deviation of the mean across cells $\sigma_{\bar{u}_{\mathrm{L}}}$ as function of β_{G} at fixed $\beta_{\mathrm{O}} = 1$ for mean field calculations. From mean field calculations.

above the point of overflow transition ($\beta_{\mathrm{G}} \sim -1.5$) we have that $\sigma_{\bar{u}_{\mathrm{L}}} \sim \sigma_{u_{\mathrm{L}}}/\sqrt{N}$, implying that correlations are zero in the overflow phase. Below it, the departure of $\sigma_{\bar{u}_{\mathrm{L}}}$ from $\sigma_{u_{\mathrm{L}}}/\sqrt{N}$ implies there are negative correlations among cells in the balanced phase.

This characterizes the difference between the balanced and overflow phases, where the former is defined as a "coordinated" state in terms of negative inter-cellular lactate exchange flux correlations, providing a concrete definition of an emergent phenomenon (a property that single cells do not have on their own, and emerges only when they interact in a wider whole). On the other hand, by assuming cells are fully connected and equally distanced, the mean-field approximation neglects any specific spatial structure. We therefore introduce nearest neighbor correlations to study the impact of spatial orientation on coordinated behavior, as these will ultimately enable a comparison with experimental data.

6.2 Nearest neighbor correlations

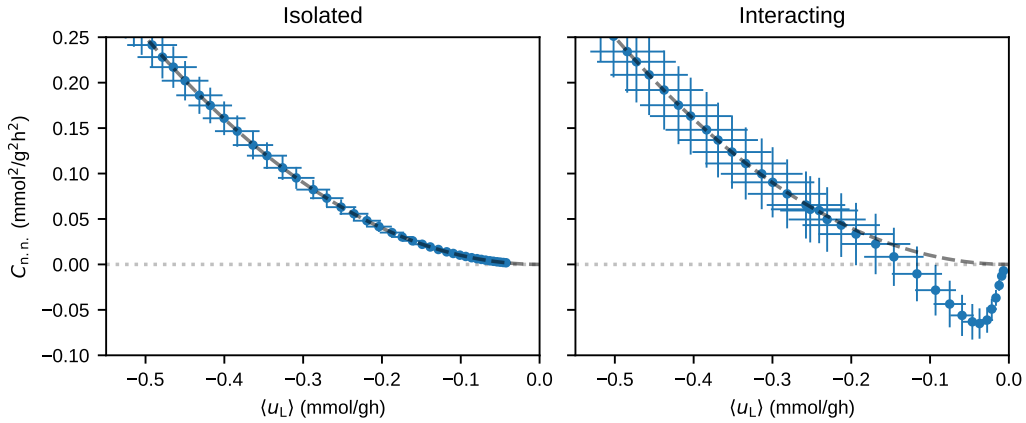
For N cells in a given random spatial configuration (from experiments or simulations) we define average nearest neighbor inter-cellular lactate flux correlations by the expectation value of

$$C_{\mathrm{n.n.}} = \frac{1}{N}\sum_i u_{\mathrm{L}}^{(i)} u_{\mathrm{L}}^{(n(i))} \tag{S 65}$$

where $n(i)$ is the index of the first neighbor of the cell i, based on absolute distance. In the absence of pairwise correlations, the expectation value of $C_{\mathrm{n.n.}}$ can be identified with the square of the average lactate flux since

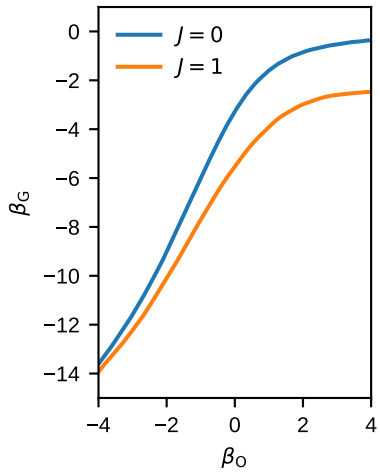
$$\langle C_{\mathrm{n.n.}}\rangle_{unc} = \frac{1}{N}\sum_i \left\langle u_{\mathrm{L}}^{(i)} u_{\mathrm{L}}^{(n(i))}\right\rangle = \frac{1}{N}\sum_i \left\langle u_{\mathrm{L}}^{(i)}\right\rangle \left\langle u_{\mathrm{L}}^{(n(i))}\right\rangle = \langle u_{\mathrm{L}}\rangle^2 . \tag{S 66}$$

As shown in Supplementary Figure 7, left, this is reproduced by simulations of isolated single cells without exogenous lactate, as considered in Supplementary Method 4.



Supplementary Figure 7: Left: Nearest neighbor correlations vs average flux for a model of isolated single cells verifies a simple scaling relation. Right: correlations vs the average flux for the coupled model violate the simple scaling relation. Simulations of the spatial model with $N = 150$ cells were performed via Monte-Carlo sampling, for a sweep in β_{G} with $\beta_{\mathrm{O}} = 1$. Error bars denote the standard error on the quantities.

On the other hand, the coupled model displays negative nearest neighbor correlations in the balanced phase, as displayed in Supplementary Figure 7, right.



Supplementary Figure 8: Critical line for the overflow threshold in the $(\beta_{\mathrm{O}}, \beta_{\mathrm{G}})$ plane with $(J = 1)$ and without $(J = 0)$ the phenomenological maximum entropy interaction term. Obtained numerically from simulations of the spatial model with $N = 150$ cells performed via Monte-Carlo sampling on a 20×20 grid.

Beyond the nearest neighbor correlations that emerge as a consequence of the spatial coupling (diffusion constraints), when fitting the model to experimental data we also consider appending it as an additional term in the exponent of the Boltzmann distribution. We weight this by small phenomenological constant J and augment the function h in the exponent according to

$$h \to h + JC_{\mathrm{n.n.}}. \tag{S 67}$$

Quadratic terms of this kind in metabolic network modeling have been studied elsewhere [7] and we checked that for the value inferred in this work this term only slightly perturb the behavior of the system. We show in Fig 8 the critical line in the $(\beta_{\mathrm{O}}, \beta_{\mathrm{G}})$ plane with $(J = 1)$ and without $(J = 0)$ this augmented form of the exponent, which has little effect on the critical behavior.

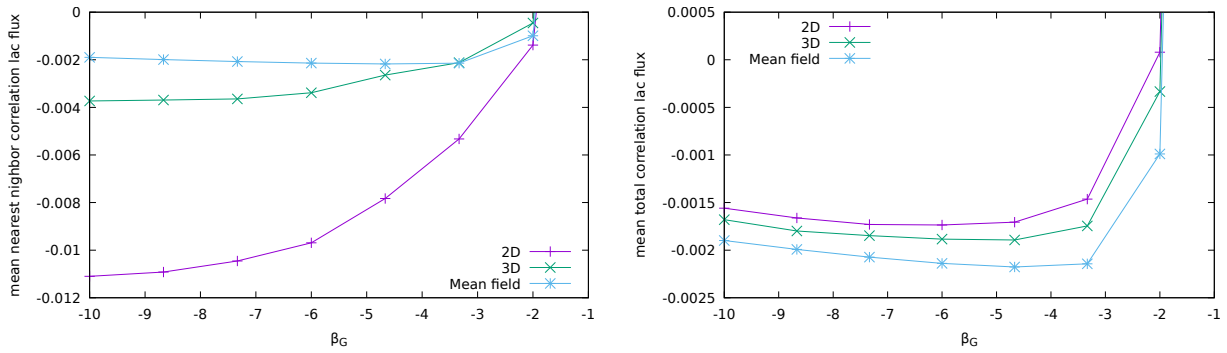
6.3 Effects of dimensionality on inter-cellular correlations

It is well known from statistical mechanics that mean field approximations improve upon increasing the dimension of the system, till becoming exact for many quantities of interest above a certain critical dimension [17]. Thus, as we move from 2D to 3D systems we expect both nearest neighbor and total correlations correlations to become better approximated using the mean-field model. This hypothesis is verified in Supplementary Figure 9, where for both types of correlations 3D cultures are found to lie between those from 2D cultures and the

mean-field model. Here the total density ρ_{3d} of the 3D system was chosen to match the magnitude of left-hand side of the inter-cellular diffusion constraint for 2D cultures by imposing

$$\rho_{2d} = \rho_{3d} L \tag{S 68}$$

where $L = 1cm$ is the total system size. By comparison with nearest neighbor correlations, which are most pronounced for 2D cultures, total correlations for both 2D and 3D cultures do not reach levels observed for the mean-field model (where both nearest neighbor and total correlations are the same by construction), which explains the small discrepancy between predictions for $\sigma_{\bar{u}_L}$ in the balanced phase from the mean-field model versus simulations reported in Figure 3 of the main text.



Supplementary Figure 9: Average lactate flux nearest neighbor (left) and total (right) correlations as a function of β_G at fixed $\beta_o = 1$ for systems of different dimensionality (2D, 3D and fully connected mean-field approximation).

SUPPLEMENTARY METHOD 7

The experimental dataset

This work includes experimental data extracted from [16]. Briefly, pH-sensing hybrid nanofibers were fabricated through electrospinning technique starting from polycaprolactone (PCL) polymer dissolved at a concentration of 10% (w/v) in chloroform and dimethyl sulfoxide (DMSO). The PCL solution was mixed with 36 mg/mL of ratiometric pH-sensing microparticles based on the pH-indicator dye fluorescein 5(6)-isothiocyanate (FITC) and the reference dye rhodamine B isothiocyanate (RBITC) synthesized as described previously [5]. For cell culture experiments, the pH-sensing nanofibers were placed into a μ-slide 4 well chamber slide with 4×10^4 cells/well of human pancreatic cancer cell line AsPC-1 and cancer-associated fibroblasts (CAFs)in the ratio 70% CAFs and 30% AsPC-1 tumor cells at 37 °C in a humidified 5% CO_2 incubator. The fluorescence response of the pH-sensing nanofibers during cell culture was monitored via time-lapse confocal laser scanning microscopy for 6 hours with a time interval of 10 minutes, maintaining a temperature of 37 °C. Before the experiment, 200 µL of Leibovitz L15 medium were added to the samples and the calibration of the whole system (pancreatic tumor and pancreatic stromal cells seeded on the pH-sensing nanofibers) was performed. Single-cell fermentation fluxes were inferred by inverse modeling the scalar pH field under the approximation of steady state diffusion from spherical sources and sinks. For further details see [16].

7.1 Flux correction due weak acid reversible proton binding

Experimental fluxes from the dataset [16] have been re-scaled here by a factor 7, which is approximately the ratio between the diffusion constant of protons and lactate, respectively (see Supplementary Table 1). The output of experiments in [16] in fact gives more precisely the ratio between the cellular uptake flux and the diffusion constant, the latter assumed to be the one of protons. This assumption, equivalent to a complete dissociation of the acid is incorrect for the case of a weak acid like lactate we are considering here.

7.2 Conversion of lactic acid level to pH

We will examine the simple case of a uniform dilute solution of acid. The situation can be treated as if the acid HA dissolves first in molecular form and ionises until equilibrium is reached. Let this initial concentration of acid be c. The two contributions to H_3O^+ in solution are (i) the ionisation of HA and (ii) the self-ionisation of

water. Let these concentrations be h_a and h_w respectively. We have,

$$[\mathrm{H_3O^+}] = [\mathrm{A^-}] + [\mathrm{OH^-}] = h_a + h_w,$$
$$[\mathrm{HA}] = c - h_a.$$

Using the expressions for K_a and K_w, we can write

$$K_\mathrm{a} = \frac{[\mathrm{H_3O^+}][\mathrm{A^-}]}{[\mathrm{HA}]} = \frac{(h_a + h_w)h_a}{c - h_a}, \tag{S 69}$$
$$K_\mathrm{w} = [\mathrm{H_3O^+}][\mathrm{OH^-}] = (h_a + h_w)h_w. \tag{S 70}$$

We need to solve the pair of equations for h_a and h_w, given K_a, K_w and c. It is convenient to substitute for $(h_a + h_w)$ and h_w in (S 70) using (S 69). Substituting

$$(h_a + h_w) = \frac{K_\mathrm{a}(c - h_a)}{h_a} \quad \text{and} \quad h_w = \frac{K_\mathrm{a}(c - h_a)}{h_a} - h_a \tag{S 71}$$

into (S 70) gives

$$K_\mathrm{w} = \frac{K_\mathrm{a}(c - h_a)}{h_a}\left[\frac{K_\mathrm{a}(c - h_a)}{h_a} - h_a\right], \tag{S 72}$$

which can be rearranged to get an implicit equation for h_a

$$h_a^2 = \frac{K_\mathrm{a}^2(c - h_a)^2}{K_\mathrm{w} + K_\mathrm{a}(c - h_a)}. \tag{S 73}$$

This equation can be solved using successive approximations, first assuming $h_a \ll c$, which is valid for a weak acid with low degree of ionisation, to obtain

$$h_a \approx \frac{K_\mathrm{a}c}{\sqrt{K_\mathrm{w} + K_\mathrm{a}c}} \tag{S 74}$$

and substituting the value back in (S 73) for an improved estimate. The approximation $c - h_a \approx c$ works best when $c/K_\mathrm{a} > 100$ or greater. Further, in the case of lactic acid, $K_\mathrm{a} \approx 1.38 \times 10^{-4}$ and $K_\mathrm{w} \approx 10^{-14}$ so that, unless the acid solution is very dilute (i. e. if $K_\mathrm{a}(c - h_a) \approx K_\mathrm{a}c \gg K_\mathrm{w}$), (S 73) can be approximated as

$$h_a^2 \approx K_\mathrm{a}(c - h_a) \quad \text{or} \quad h_a \approx \sqrt{K_\mathrm{a}c}. \tag{S 75}$$

To obtain a first approximation of $[\mathrm{H_3O^+}]$ and hence the pH of the medium, (S 74) could be substituted in (S 71) to obtain

$$[\mathrm{H_3O^+}] = (h_a + h_w) = \frac{K_\mathrm{a}c}{h_a} - K_\mathrm{a} \approx \sqrt{K_\mathrm{w} + K_\mathrm{a}c} - K_\mathrm{a}. \tag{S 76}$$

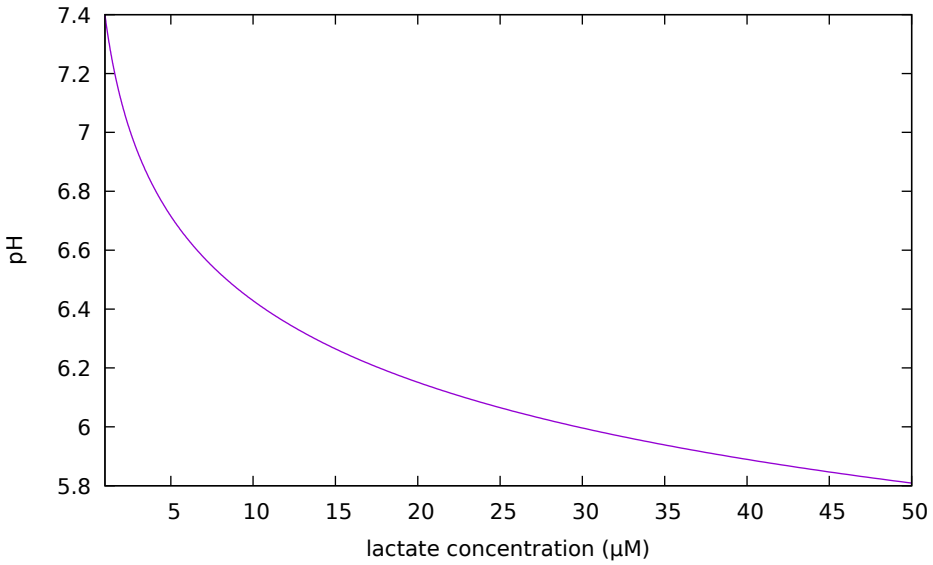
It is to be noted that (S 71) cannot be used directly when $c \to 0$, as both the numerator and denominator tend to 0. Instead, when $h_a = 0$, $[\mathrm{H_3O^+}] = h_w = \sqrt{K_\mathrm{w}}$ from (S 70). It is possible to solve (S 73) numerically and substitute the value in (S 71) to obtain $[\mathrm{H_3O^+}]$ as a function of c.

In our experimental setup, we do not have direct measurements of the concentration of acids. Instead, we have measurements of pH and hence of $[\mathrm{H_3O^+}]$. So in (S 69) and (S 70), the known quantity is $h_a + h_w$ and the unknown is c. We seek an expression for c in terms of $h = h_a + h_w$, given K_a and K_w. We can rewrite (S 69) and (S 70) in terms of h and substitute for h_a in (S 69) using (S 70) as

$$K_\mathrm{w} = hh_w = h(h - h_a) \implies h_a = \left(h - \frac{K_\mathrm{w}}{h}\right) \tag{S 77}$$

$$K_\mathrm{a} = \frac{hh_a}{c - h_a} \implies \frac{h}{K_\mathrm{a}} = \frac{c}{h_a} - 1$$
$$\implies c = h_a\left(1 + \frac{h}{K_\mathrm{a}}\right)$$
$$\implies c = \left(h - \frac{K_\mathrm{w}}{h}\right)\left(1 + \frac{h}{K_\mathrm{a}}\right) \tag{S 78}$$

It is easily verified that when $h = h_w = \sqrt{K_\mathrm{w}}$, (S 78) correctly gives $c = 0$. This function $c(h)$ can be numerically inverted to obtain a lookup table for the inverse function $h(c)$, from which we calculate pH as a function of lactate concentration in the medium. We further add an offset term to this formula to match the experimental pH, phenomenologically modeling the buffering of the medium. The resulting conversion curve is shown in Supplementary Figure 8.

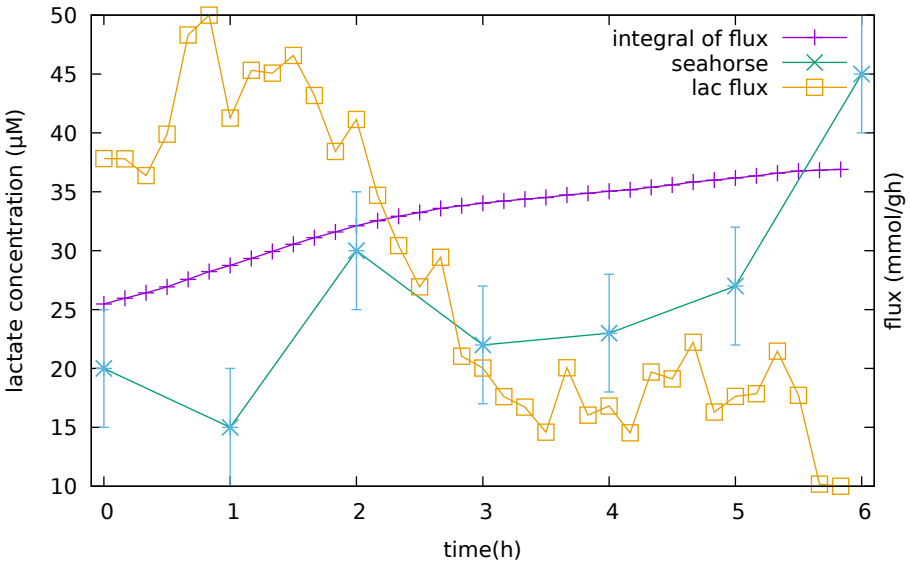


Supplementary Figure 10: pH as a function of the lactate concentration.

7.3 Background lactate accumulation

As described previously, although no exogenous source of lactate is supplied to the growth medium, lactate produced by cells can accumulate to become available for net uptake at later time points. Supplementary Figure 11 displays the experimental time course of total lactate concentrations as measured independently using a Seahorse assay with lactate dehydrogenase (see [16] for further details) and compared with the integral of the net lactate flux. Both measurements confirm that background concentrations of lactate are increasing in time as it is being produced by cell cultures, which motivates inclusion of the background lactate term described in Supplementary Method 4 when fitting the model to data. The implementation of the background term in the inverse modeling of experimental data has been performed by matching the experimentally measured lactate level with the sum of the of the average lactate level in the observed frame due to flux from cells and the boundary constant term according to

$$c_{\mathrm{L},exp} = \langle c_{\mathrm{L}} \rangle + c_{\mathrm{L},b}. \tag{S 79}$$



Supplementary Figure 11: Experimental lactate concentration as function of time, obtained from a lactate dehydrogenase assay (stars) and by integrating over average cell lactate flux (crosses). Also shown is the average lactate flux (squares). Data from [16]

SUPPLEMENTARY METHOD 8

Inverse modeling experimental data

8.1 Likelihood function

The maximum entropy model we are considering is determined by two parameters, which are the Lagrange multipliers $\beta_{\rm G}$ and $\beta_{\rm O}$ that fix the average net glucose and oxygen fluxes, respectively, across the cell population. Equivalently, one could consider any other linearly independent combinations of these Lagrange multipliers, such as those fixing the average net fluxes for ATP and lactate production. As we remarked at the end of Supplementary Method 1, the setting could be modified beyond the homogeneity assumption and allowing for $2N$ parameters to be inferred. This would make the problem computationally more difficult but still feasible with the help of expectation propagation techniques [4, 14, 18].

Ideally, experimental data should provide information about independent sets of net metabolic fluxes. In this respect, one caveat of our experimental data is that we have access only to the set of single cell lactate fluxes at each experimental time point. However, we can instead exploit the single-cell resolution nature of these data, to fit the moments of the observed single-cell experimental lactate flux distribution, which are predicted as functions of $\beta_{\rm G}$ and $\beta_{\rm O}$ in our model. This leads to a predictive model for the observables: average net lactate flux, fluctuations (as measured by the standard deviation) and nearest neighbor correlations, which can be fitted by maximizing the likelihood of the experimental data. Assuming that sources of noise are Gaussian and independently distributed, the log-likelihood $\mathcal{L}_\gamma(\vec{\beta}_{\rm O}, \vec{\beta}_{\rm G})$ can be written as the sum of two terms. The first is the residue and/or distance between the experimental observables and the corresponding predictions the model, while the second is a regularizer term that prevents over-fitting by ensuing continuity of the parameters across adjacent time points, controlled by the hyper-parameter γ:

$$\mathcal{L}_\gamma(\vec{\beta}_{\rm O}, \vec{\beta}_{\rm G}) = \sum_t -\Bigg[\frac{(\overline{u_{\rm L}}_{{\rm exp},t} - \langle u_{\rm L}\rangle_{{\rm mod},t})^2}{2\sigma^2_{{\rm mean},t}} + \frac{(\sigma_{u_{\rm L},{\rm exp},t} - \langle\sigma_{u_{\rm L}}\rangle_{{\rm mod},t})^2}{2\sigma^2_{{\rm fluc},t}} + \frac{(C_{\rm n.n.,exp,t} - \langle C_{\rm n.n.}\rangle_{{\rm mod},t})^2}{2\sigma^2_{{\rm corr},t}} - \gamma\left[(\beta_{\rm O}(t) - \beta_{\rm O}(t+1))^2 + (\beta_{\rm G}(t) - \beta_{\rm G}(t+1))^2\right]\Bigg]. \quad \text{(S 80)}$$

Here the sum is taken over experimental time points $t = 0, 1, ..., T$ (with $\vec{\beta}_{\rm O} \equiv (\beta_{\rm O}(0), \beta_{\rm O}(1), \beta_{\rm O}(2), ..., \beta_{\rm O}(T))$ and same for $\vec{\beta}_{\rm G}$) and experimental errors were estimated via jackknife resampling. The most probable values of the parameters are retrieved by maximizing $\mathcal{L}$ as described in the next subsection, while errors on the inferred parameters can be estimated by considering that knowledge of $\mathcal{L}$ leads to their posterior probability distribution approximated (up to to a constant) by

$$\text{prob}(\vec{\beta}_{\rm O}, \vec{\beta}_{\rm G}) \propto e^{\mathcal{L}}. \quad \text{(S 81)}$$

8.2 The sampling method

The parameters have been sampled according to the posterior probability distribution via a Metropolis Monte-Carlo algorithm according to the following steps:

0. Initial data: a set of grids in the $(\beta_{\rm O}, \beta_{\rm G})$ plane that have been calculated from the model with the values of the observables of interest (average, single cell fluctuations and nearest neighbor correlations of the lactate flux), one for each experimental time point, with increasing level of background lactate. This is done using the method described in Supplementary Method 2. Define the starting value for the Markov chain $(\vec{\beta}^n_{\rm O}, \vec{\beta}^n_{\rm G})$, $n = 0, 1, 2, \ldots$ with $(\vec{\beta}^0_{\rm O}, \vec{\beta}^0_{\rm G}) = (\vec{0}, \vec{0})$.
1. Choose uniformly at random an experimental time point $t^* \in [0, T]$ and uniformly at random from the grid corresponding to that time point a value $(\beta^*_{\rm O}, \beta^*_{\rm G})$. Construct a proposal by substituting in the current vector the component corresponding to that time point:
$$(\vec{\beta}^{\rm prop}_{\rm O}, \vec{\beta}^{\rm prop}_{\rm G}) = ((\beta^n_{\rm O}(t=0), \beta^n_{\rm G}(t=0)), \ldots, (\beta^*_{\rm O}(t^*), \beta^*_{\rm G}(t^*)), \ldots, (\beta^n_{\rm O}(t=T), \beta^n_{\rm G}(t=T))).$$
2. Calculate the difference $\Delta\mathcal{L} = \mathcal{L}(\vec{\beta}^{\rm prop}_{\rm O}, \vec{\beta}^{\rm prop}_{\rm G}) - \mathcal{L}(\vec{\beta}^n_{\rm O}, \vec{\beta}^n_{\rm G})$.
3. (Metropolis) Accept the proposal if $\Delta\mathcal{L} \geq 0$ or, if $\Delta\mathcal{L} < 0$, extract a random number $r \in [0, 1]$ at uniform and accept the proposal if $r < e^{\Delta\mathcal{L}}$. If proposal accepted go to 4, otherwise go to 5.
4. Set
$$(\vec{\beta}^{n+1}_{\rm O}, \vec{\beta}^{n+1}_{\rm G}) = (\vec{\beta}^{\rm prop}_{\rm O}, \vec{\beta}^{\rm prop}_{\rm G})$$
and go to 1.

5. Set

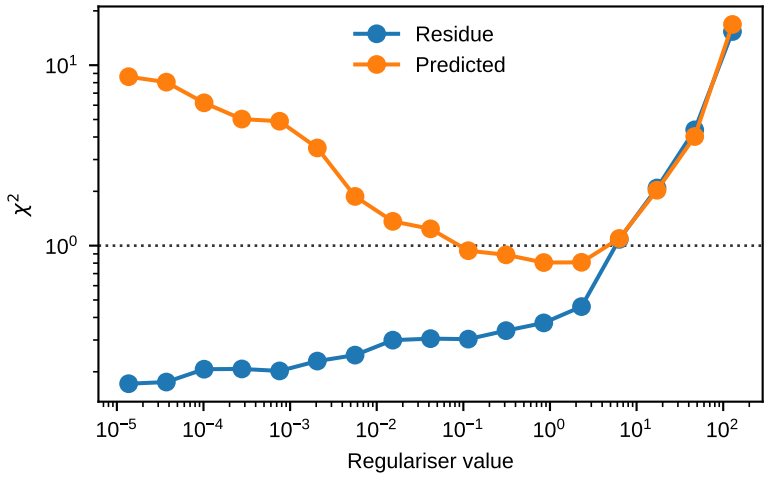
$$(\vec{\beta}_{\mathrm{O}}^{n+1}, \vec{\beta}_{\mathrm{G}}^{n+1}) = (\vec{\beta}_{\mathrm{O}}^{n}, \vec{\beta}_{\mathrm{G}}^{n})$$

and go to 1.

After disregarding an initial transient (around 10^5 points), we collected $\sim 10^6$ samples from which we evaluated the statistical quantities of interest.

8.3 Training and test set split: setting the hyper-parameter γ

We tested the predictive capabilities of the model with standard methods of statistical inference. We split the $M = 36 \times 3 = 118$ observables in two sets: a training set, comprising 90% of the observables taken uniformly at random, over which we perform the fit (i.e., values of the observables are included in the definition of $\mathcal{L}$), and a test set with the remaining 10% values used to evaluate model predictions. This procedure was repeated 10^4 times for different realization of the random split and it has been evaluated as a function of the hyper-parameter. The reduced χ^2 has been calculated with respect to the training and test set as a function of the hyper-parameter γ. The trend is monotonously increasing for the training set (starting from $\chi^2 =$ at $\gamma = 0$) while for the test set it decreases sharply up to $\gamma = 1$ where it assumes approximately the value of the of the training set (see fig 12). The trade-off is thus optimal around $\gamma = 2$, the value we finally set.



Supplementary Figure 12: Reduced χ^2 (average square residue) for the fitted and predicted data as a function of γ

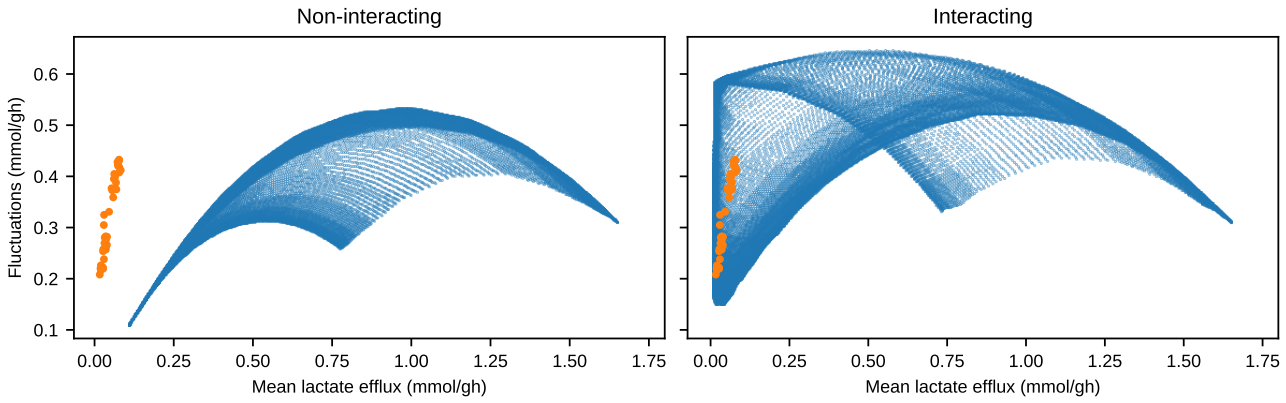
8.4 Fitting the model without correlations

In the first instance, we trialed fitting the two model parameters $\beta_{\mathrm{G}}, \beta_{\mathrm{O}}$ to two observables, and chose the average and standard deviation (fluctuations) of lactate flux $\langle u_{\mathrm{L}} \rangle$ and σ_{L}, respectively, by excluding the nearest neighbor correlations term from the likelihood function. In this scenario, we see by comparing the grids of model predictions (Supplementary Figure 13) that a model with isolated single cells (non-interacting) will never be able to reproduce these experimental data. On the other hand, the model involving inter-cellular diffusion constraints (interacting model) is successful in producing a set of $\langle u_{\mathrm{L}} \rangle$ and σ_{L} values that match the experimental observations (compare Supplementary Figure 14 with 15). However, the nearest neighbor correlations predicted by the interacting model fail to reproduce those observed experimentally (Supplementary Figure 14) and we find inferred $(\beta_{\mathrm{G}}, \beta_{\mathrm{O}})$ points have large errors bars extending deeply into the balanced phase that seem to self-organize around the critical lines (Supplementary Figure 16). The latter seems to be an indication of over-fitting since similar biases has been noted in inverse modeling neural systems.

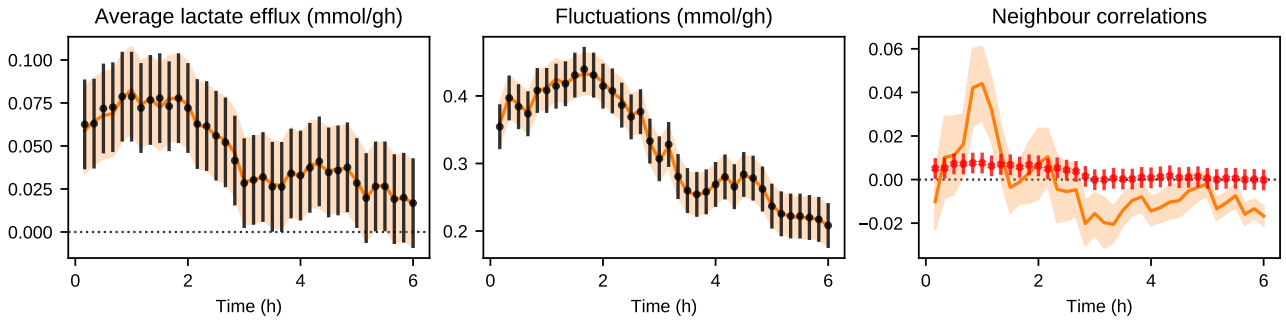
8.5 Fitting the model with correlations

In an effort to improve the fit of the interacting model, we thus included the term with nearest neighbor correlations in the likelihood. We note that a non-interacting model with isolated single cells cannot reproduce these correlations by construction. In the absence of a background lactate term, the interacting model then presented a reasonable fit of these experimental data, reproducing the negative nearest neighbor correlations observed in the balanced phase (Supplementary Figure 17).

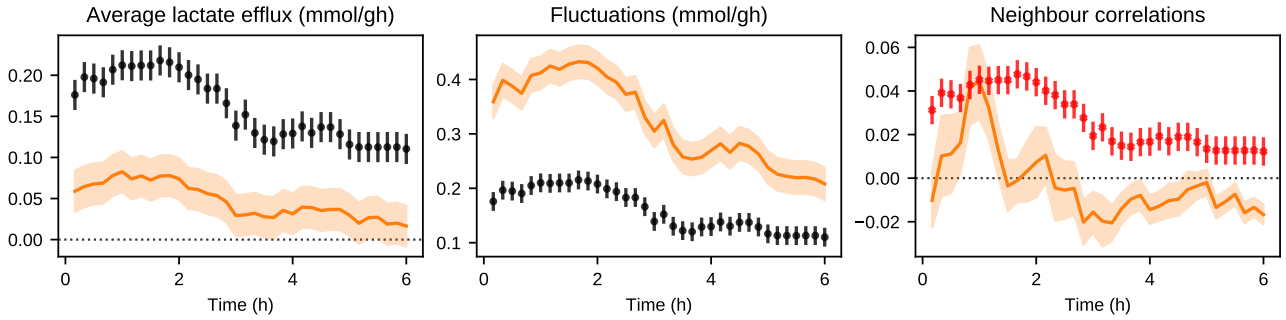
However, since the background concentration of lactate is both predicted and observed experimentally to increase over time, we next assessed a fit of the same data with the background lactate term estimated as described



Supplementary Figure 13: Grid of model predictions (blue) compared to experimental data (orange) for models with isolated single cells (left, non-interacting model) and cells coupled by inter-cellular diffusion constraints (right, interacting model).

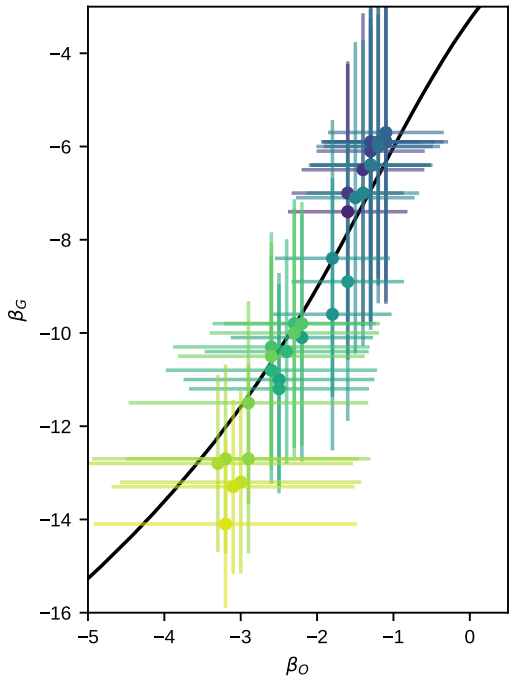


Supplementary Figure 14: Fit of the interacting model (black dots) to averages and fluctuations of the experimental lactate flux (orange lines). Nearest neighbor correlations (red) were excluded from the fit. See Supplementary Method 8.1 for details on error estimation.

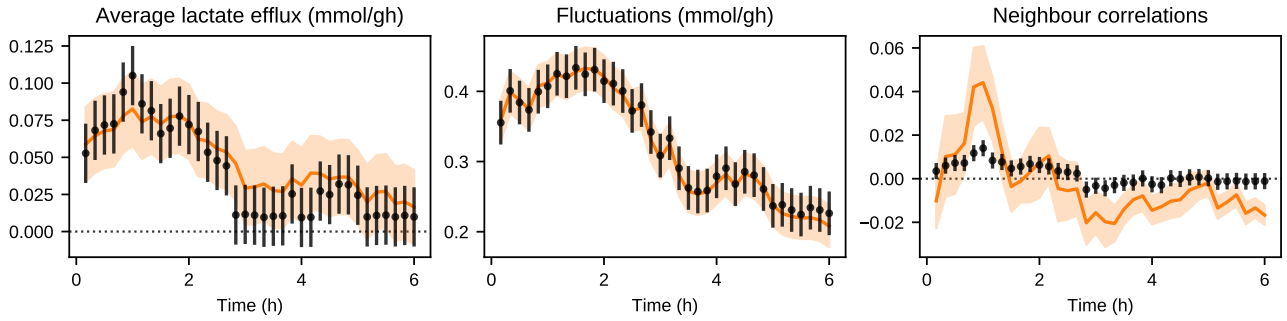


Supplementary Figure 15: Fit of the non-interacting model (black dots) to averages and fluctuations of the experimental lactate flux (orange lines). Nearest neighbor correlations (red) were excluded from the fit. See Supplementary Method 8.1 for details on error estimation.

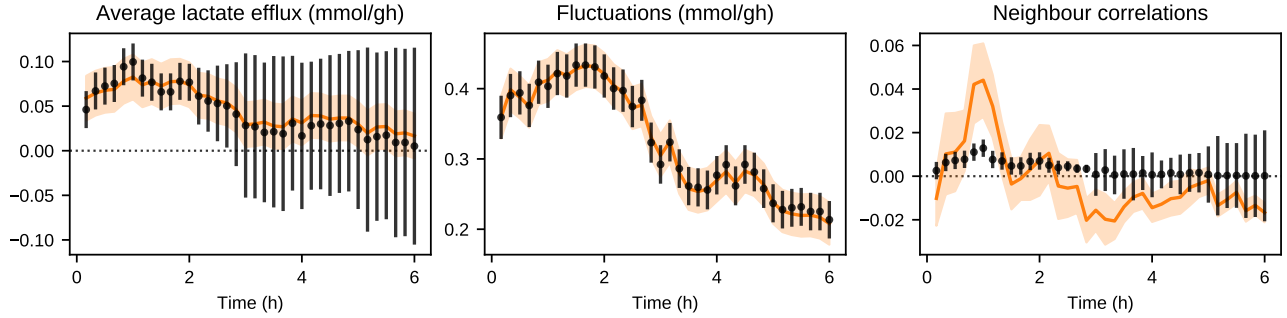
previously (Supplementary Methods 4 and 7.3). This lead to a mutual inconsistency between the three observables predicted by the model. As shown in Supplementary Figure 18, based on diffusion constraints alone, the interacting model cannot explain the negative nearest neighbor correlations observed experimentally when the background concentrations of lactate are high. This motivated an inclusion of an additional phenomenological term in the Boltzmann exponent as discussed in Supplementary Method 6 and main text. The final value of this phenomenological term has been fixed upon performing the inference for a range of six different values (Supplementary Figure 19).



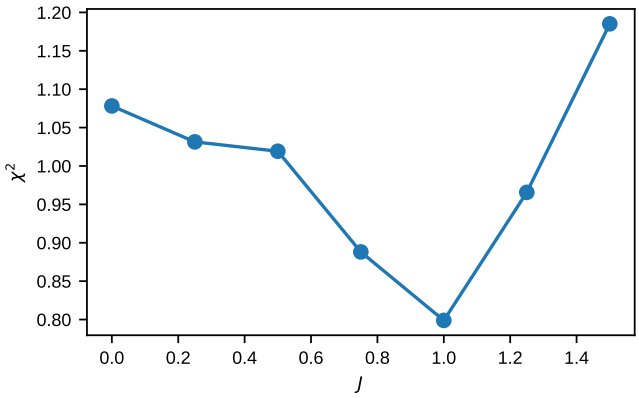
Supplementary Figure 16: Inferred points in the $(\beta_{\mathrm{O}}, \beta_{\mathrm{G}})$ phase space obtained from fitting the model to averages and standard deviations of the lactate flux, excluding correlations. The critical line obtained from the mean-field model without exogenous lactate is shown in black.



Supplementary Figure 17: Fit of the interacting model (black dots) to the average, fluctuations and nearest neighbor correlations of the experimental lactate flux (orange lines), neglecting background lactate accumulation. See Supplementary Method 8.1 for details on error estimation.



Supplementary Figure 18: Fit of the interacting model (black dots) to the average, fluctuations and nearest neighbor correlations of the experimental lactate flux (orange lines), including background lactate accumulation. See Supplementary Method 8.1 for details on error estimation.



Supplementary Figure 19: Reduced χ^2 as function of J

SUPPLEMENTARY METHOD 9

Modeling single-cell oxygen dynamics

From the inverse modeling, we have an estimate of the average oxygen flux $\langle u_{\rm O}\rangle_{\rm ME}$ that we can complement with the single cell measured value $u_{\rm L}^{(n)}$ of the lactate flux to obtain single cell estimates of the oxygen fluxes, once again, using the maximum entropy method. The lower bound $\mathcal{L}_{\rm O}^{(n)}$ and upper bound $\mathcal{U}_{\rm O}^{(n)}$ for oxygen flux $u_{\rm O}^{(n)}$ is calculated using the measured lactate flux $u_{\rm L}^{(n)}$ as

$$\begin{aligned}\mathcal{L}_{\rm O}^{(n)} &= \max\left(0,\ 3\,u_{\rm L}^{(n)},\ \frac{M+u_{\rm L}^{(n)}}{5}\right)\\ \mathcal{U}_{\rm O}^{(n)} &= \min\left(U_{\rm O},\ 6\,U_{\rm G}+3\,u_{\rm L}^{(n)}\right).\end{aligned} \tag{S 82}$$

For cells with $u_{\rm L}^{(n)} < -2\,U_{\rm G}$, we enforce $u_{\rm O}^{(n)} = 0$. We then enforce a maximum entropy constraint on the average oxygen flux with a Lagrange multiplier ξ that is independent of the cell index. To obtain the average oxygen flux for the cell n we integrate the marginal Boltzmann probability distribution over the segment $[\mathcal{L}_{\rm O}^{(n)}\mathcal{U}_{\rm O}^{(n)}]$

$$\left\langle u_{\rm O}^{(n)}\right\rangle = \frac{\int_{\mathcal{L}_{\rm O}^{(n)}}^{\mathcal{U}_{\rm O}^{(n)}} u e^{\xi u}\,du}{\int_{\mathcal{L}_{\rm O}^{(n)}}^{\mathcal{U}_{\rm O}^{(n)}} e^{\xi u}\,du} \tag{S 83}$$

that gives

$$\left\langle u_{\rm O}^{(n)}\right\rangle = \frac{\mathcal{U}_{\rm O}^{(n)} e^{\xi\,\mathcal{U}_{\rm O}^{(n)}} - \mathcal{L}_{\rm O}^{(n)} e^{\xi\,\mathcal{L}_{\rm O}^{(n)}}}{e^{\xi\,\mathcal{U}_{\rm O}^{(n)}} - e^{\xi\,\mathcal{L}_{\rm O}^{(n)}}} - \frac{1}{\xi}. \tag{S 84}$$

We then match the average $\left\langle u_{\rm O}^{(n)}\right\rangle$ with $\langle u_{\rm O}\rangle_{\rm ME}$

$$\frac{1}{N}\sum_{i=1}^{N}\left\langle u_{\rm O}^{(n)}\right\rangle = \langle u_{\rm O}\rangle_{\rm ME} \tag{S 85}$$

which gives

$$\frac{1}{\xi} = \frac{1}{N}\sum_{n=1}^{N}\frac{\mathcal{U}_{\rm O}^{(n)} e^{\xi\,\mathcal{U}_{\rm O}^{(n)}} - \mathcal{L}_{\rm O}^{(n)} e^{\xi\,\mathcal{L}_{\rm O}^{(n)}}}{e^{\xi\,\mathcal{U}_{\rm O}^{(n)}} - e^{\xi\,\mathcal{L}_{\rm O}^{(n)}}} - \langle u_{\rm O}\rangle_{\rm ME} \tag{S 86}$$

The value of ξ obtained solving numerically equation (S 86) can then be used to estimate $\left\langle u_{\rm O}^{(n)}\right\rangle$ from (S 84).

This procedure, with our data, does give large errors, i.e. there are many equivalent models with inferred single cell oxygen flux reproducing the lactate data that are compatible with the constraints. Our inferred model can be thus regarded as a "sloppy" one in the sense of [10] and does not lead to real single-cell measurements of the oxygen flux. This is a task that could be finally be achieved with the help of nanometric sensing [8].

Supplementary References

[1] Jorge Fernandez-de-Cossio-Diaz , Kalet Leon, and Roberto Mulet. Characterizing steady states of genome-scale metabolic networks in continuous cell cultures. PLoS Computational Biology, 13(11):1–22, 2017.

[2] Makarieva Am, Gorshkov Vg, Li Bl, Chown Sl, Reich Pb, and Gavrilov Vm. Mean mass-specific metabolic rates are strikingly similar across life's major domains: Evidence for life's metabolic optimum. Proceedings of the National Academy of Sciences of the United States of America, 105(44), 2008.

[3] Carl M Bender and Steven A Orszag. Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory. Springer Science & Business Media, 2013.

[4] Alfredo Braunstein, Anna Paola Muntoni, and Andrea Pagnani. An analytic approximation of the feasible space of metabolic networks. Nature communications, 8(1):14915, 2017.

[5] Anil Chandra, Saumya Prasad, Francesco Alemanno, Maria De Luca, Riccardo Rizzo, Roberta Romano, Giuseppe Gigli, Cecilia Bucci, Adriano Barra, and Loretta L del Mercato. Fully automated computational approach for precisely measuring organelle acidification with optical ph sensors. ACS Applied Materials & Interfaces, 14(16):18133–18149, 2022.

[6] Daniele De Martino, Matteo Mori, and Valerio Parisi. Uniform sampling of steady states in metabolic networks: heterogeneous scales and rounding. PloS one, 10(4):e0122670, 2015.

[7] Jorge Fernandez-de Cossio-Diaz and Roberto Mulet. Statistical mechanics of interacting metabolic networks. Physical Review E, 101(4):042401, 2020.

[8] Giuliana Grasso, Valentina Onesto, Stefania Forciniti, Eliana D'Amone, Francesco Colella, Lara Pierantoni, Valeria Famà, Giuseppe Gigli, Rui L Reis, Joaquim M Oliveira, et al. Highly sensitive ratiometric fluorescent fiber matrices for oxygen sensing with micrometer spatial resolution. Bio-Design and Manufacturing, pages 1–15, 2024.

[9] David J Griffiths. Introduction to Electrodynamics. Cambridge University Press, 2021.

[10] Ryan N Gutenkunst, Joshua J Waterfall, Fergal P Casey, Kevin S Brown, Christopher R Myers, and James P Sethna. Universally sloppy parameter sensitivities in systems biology models. PLoS computational biology, 3(10):e189, 2007.

[11] Rudolf Hober, David I. Hitchcock, J. B. Bateman, David R. Goddard, and Wallace O. Fenn. Physical Chemistry of Cells and Tissues. The Journal of Physical Chemistry, 50(4):386–387, 1946.

[12] Donald E Knuth. The Art of Computer Programming: Seminumerical Algorithms, Volume 2. Addison-Wesley Professional, 2014.

[13] Werner Krauth and Marc Mézard. Learning algorithms with optimal stability in neural networks. Journal of Physics A: Mathematical and General, 20(11):L745, 1987.

[14] Anna Paola Muntoni, Alfredo Braunstein, Andrea Pagnani, Daniele De Martino, and Andrea De Martino. Relationship between fitness and heterogeneity in exponentially growing microbial populations. Biophysical Journal, 121(10):1919–1930, 2022.

[15] David L Nelson, Michael M Cox, and Albert L Lehninger. Lehninger Principles of Biochemistry. W.H. Freeman and Company, 2013.

[16] Valentina Onesto, Stefania Forciniti, Francesco Alemanno, Krishnadev Narayanankutty, Anil Chandra, Saumya Prasad, Amalia Azzariti, Giuseppe Gigli, Adriano Barra, Andrea De Martino, et al. Probing single-cell fermentation fluxes and exchange networks via ph-sensing hybrid nanofibers. ACS nano, 17(4):3313–3323, 2022.

[17] Giorgio Parisi. Statistical field theory (advanced book classics). Perseus Pr, 1998.

[18] José Antonio Pereiro-Morejón, Jorge Fernandez-de Cossio-Diaz, and Roberto Mulet. Inference of metabolic fluxes in nutrient-limited continuous cultures: A maximum entropy approach with the minimum information. Iscience, 25(12), 2022.

[19] Miklós Simonovits. How to compute the volume in high dimension? Mathematical programming, 97:337–374, 2003.

[20] W.K. Subczynski and J.S. Hyde. Diffusion of oxygen in water and hydrocarbons using an electron spin resonance spin-label technique. Biophysical Journal, 45(4):743–748, 1984.

[21] Diana Széliová, Jerneja Štor, Isabella Thiel, Marcus Weinguny, Michael Hanscho, Gabriele Lhota, Nicole Borth, Jürgen Zanghellini, David E. Ruckerbauer, and Isabel Rocha. Inclusion of maintenance energy improves the intracellular flux predictions of CHO. PLOS Computational Biology, 17(6):e1009022, jun 2021.

[22] VF Turcin. On the computation of multidimensional integrals according to the monte carlo method. Teor. Verojatnost. i Primenen, 16:738–743, 1971.

[23] Brett A. Wagner, Sujatha Venkataraman, and Garry R. Buettner. The rate of oxygen utilization by cells. Free Radical Biology & Medicine, 51(3):700–712, aug 2011.