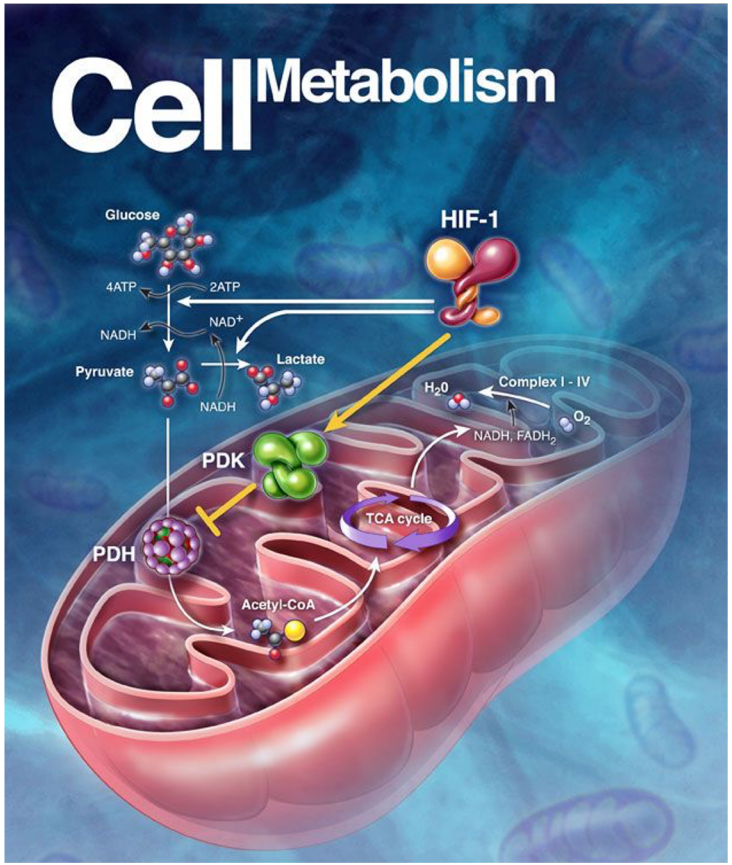
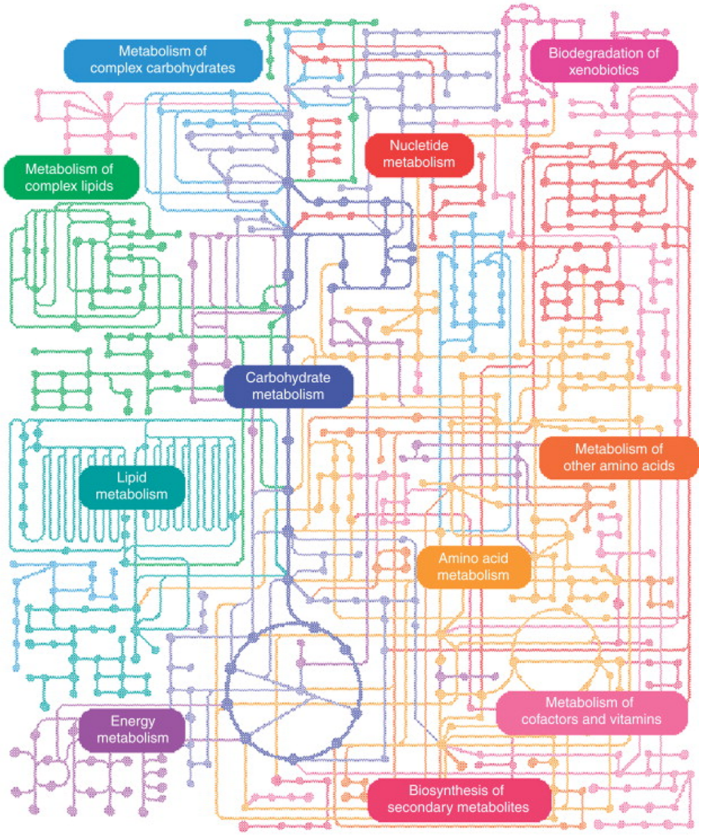
Inferring metabolic fluxes by maximizing information entropy conditioned on gene expression

Marcelo Rivas-Astroza

Departamento de Biotecnología
Universidad Tecnológica Metropolitana

Metabolism at the cellular level





T.Lengauer et al 2007

Inferring metabolic fluxes is useful but complicated

nature biotechnology

Explore content · About the journal · Publish with us

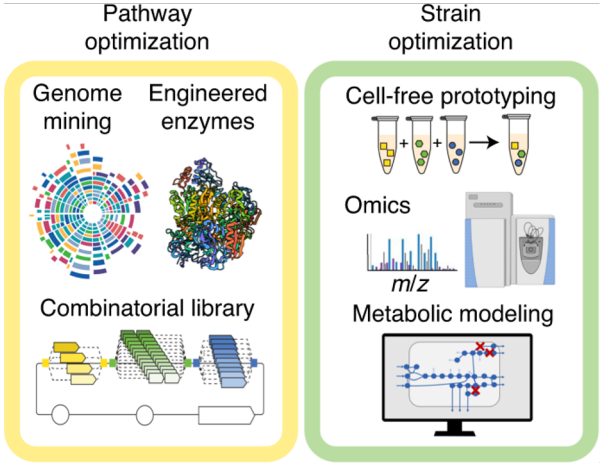
nature › nature biotechnology › articles › article

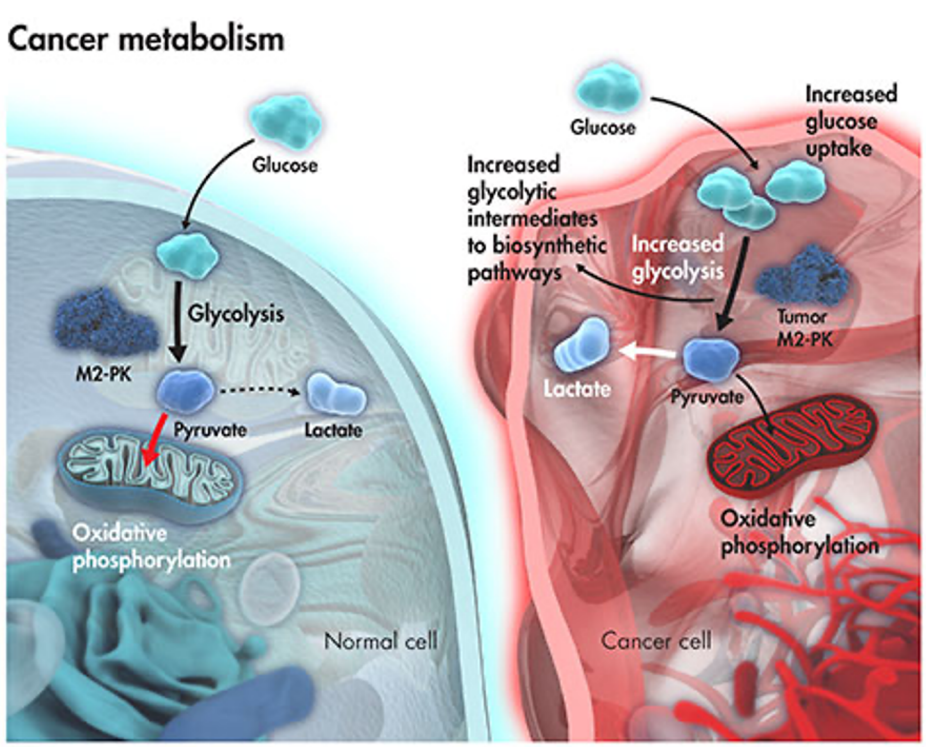
Article | Published: 21 February 2022

Carbon-negative production of acetone and isopropanol by gas fermentation at industrial pilot scale

Fungmin Eric Liew, Robert Nogle, ... Michael Köpke + Show authors

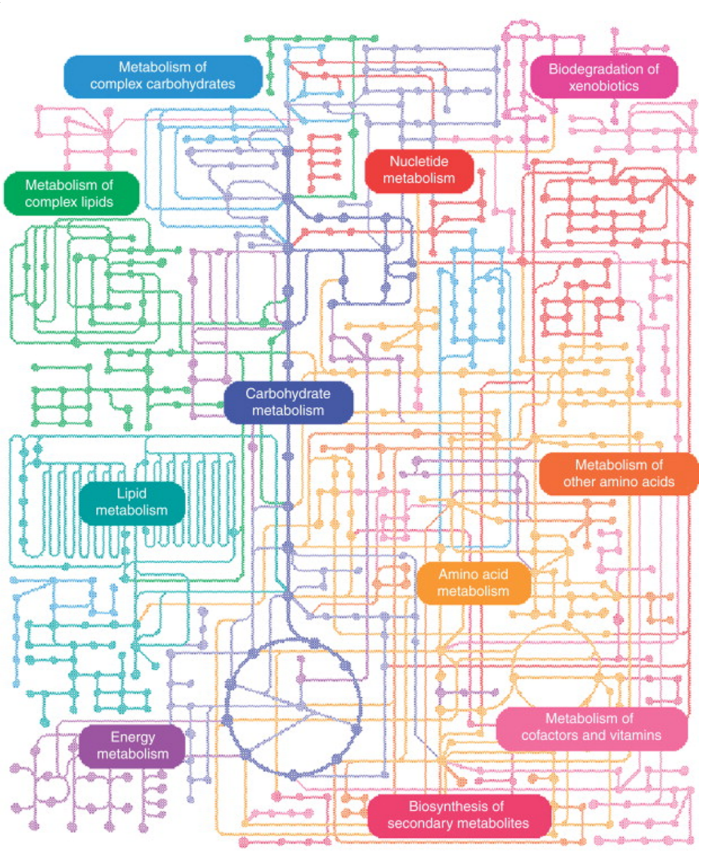
Nature Biotechnology (2022) | Cite this article





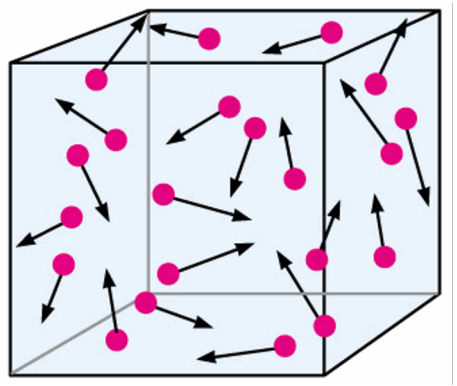
saboscIences

Inferring metabolic fluxes is useful but complicated



Framework of the problem

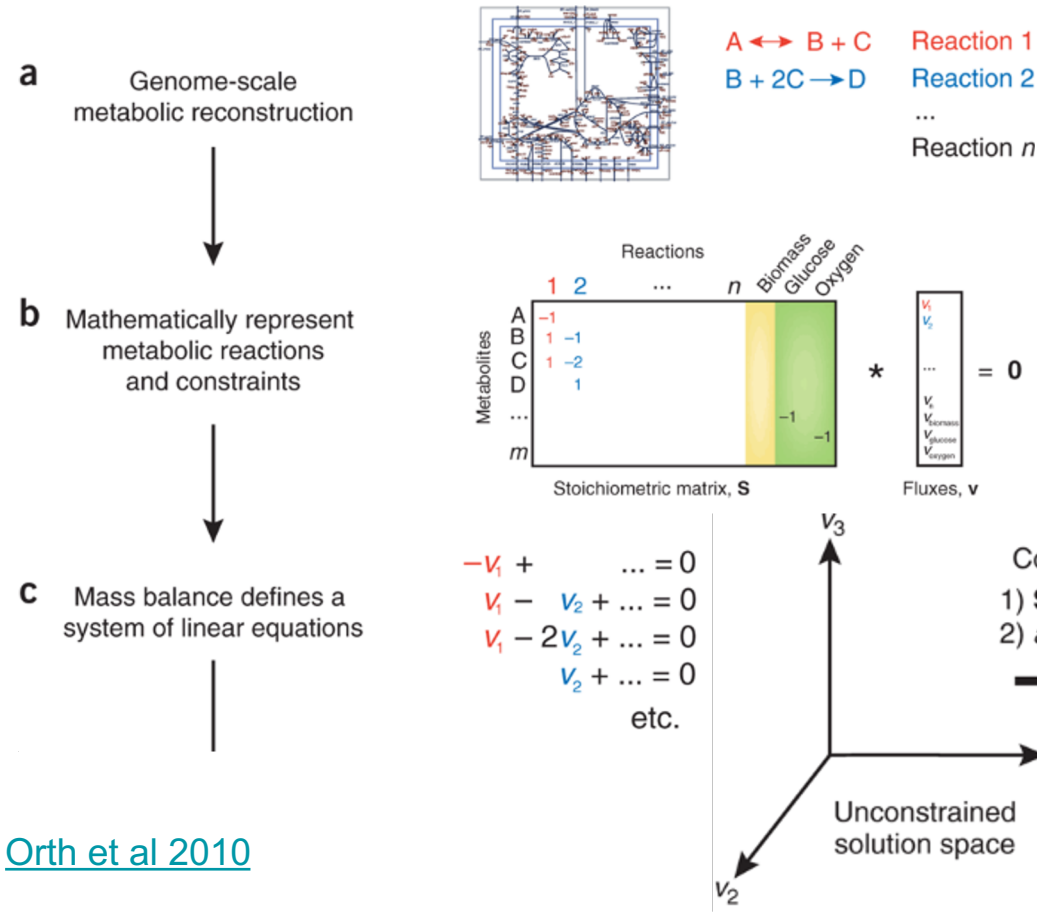
Simplifying assumptions



Evidence of improvement over alternative models

T.Lengauer et al 2007

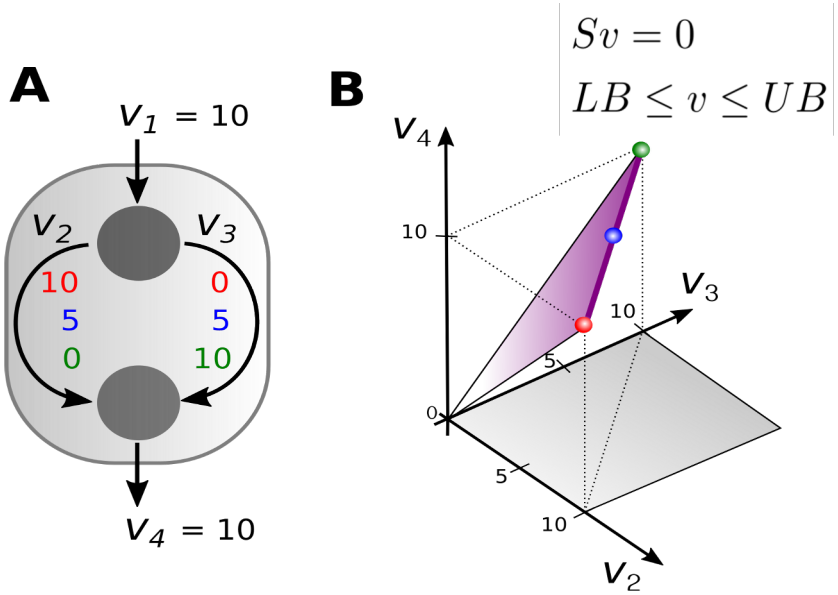
Framework



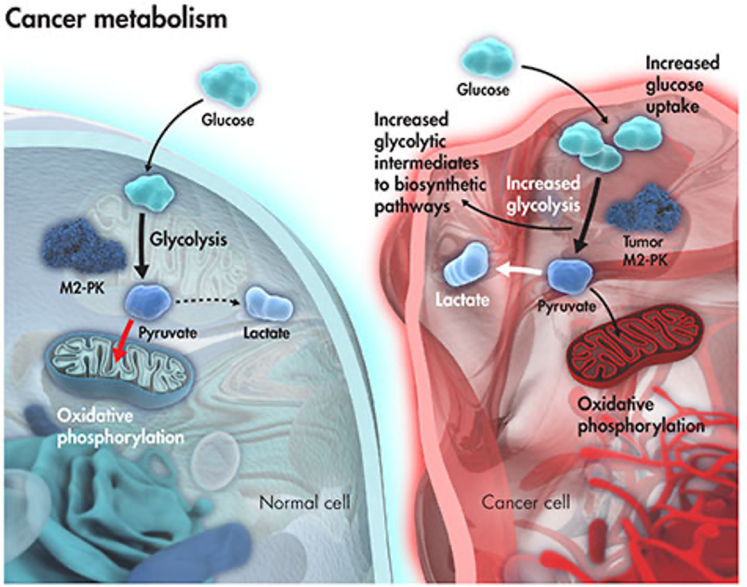
[Orth et al 2010]

Framework

Ambiguous inferences



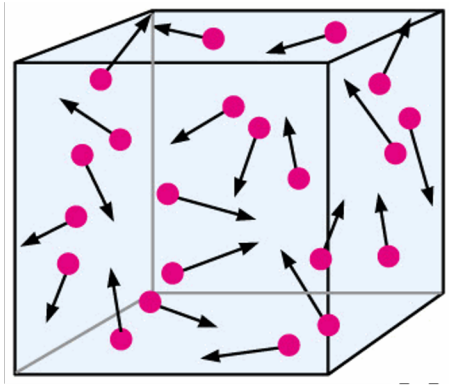
Not phenotype-specific



It is not always easy to derive the metabolic objective function

Framework

Ludwig Boltzmann



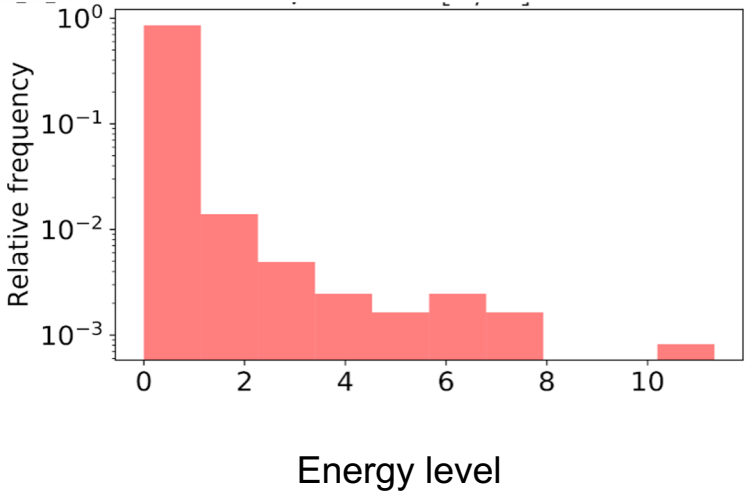
$$W = \frac{Q!}{\prod_i q_i}$$

$$max - \sum_i p_i Log(p_i)$$

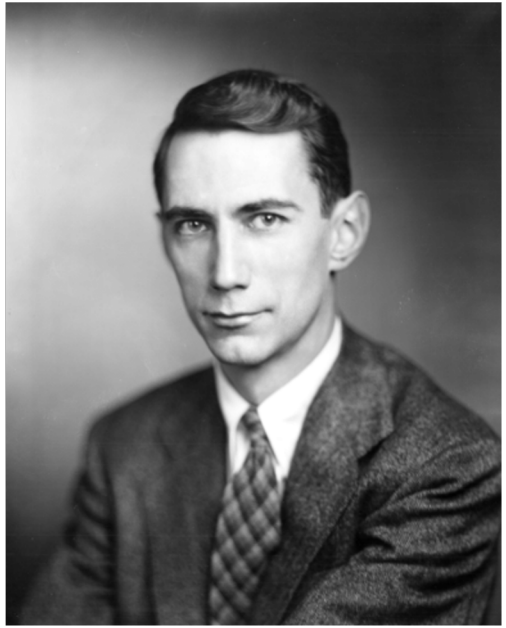
s. t.

$$p_i = \frac{q_i}{\sum_j q_j}$$

$$\sum_i p_i q_i = E$$



Framework



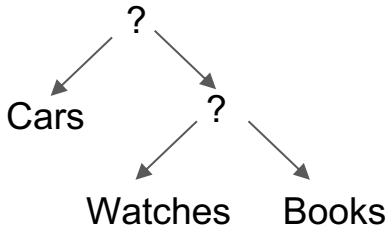
Claude Shannon

Florence:

Cars 50%

Watches 25%

Books 25%



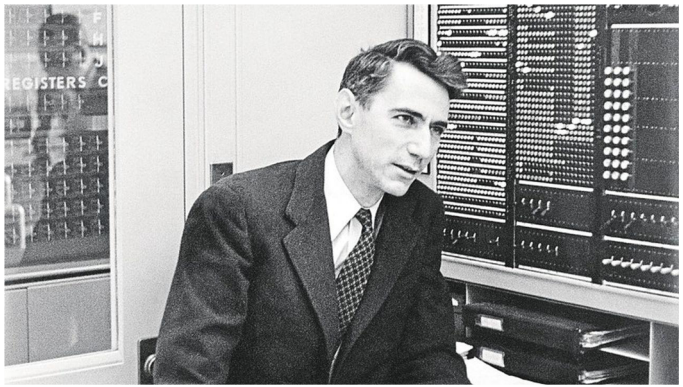
$$-\log(0.5) = 1$$

$$-\log(0.25) = 2$$

$$E[-\log(p_i)] = -\sum p_i \log(p_i)$$

$$= H$$

Framework: Using the principle of maximum entropy



E.T. Jaynes

$$\max_{v} H_v(X)$$

subject to:

$$Sv = 0$$

$$LB \leq v \leq UB$$

1. How do we define *H* in the context of the fluxome space?, and
2. How do we incorporate gene expression data into *H*?

Assumptions: MaxEnt

$$v_i = e_i f_i$$

For example, in Michaelis-Menten: $v = k\left(\frac{S}{K+S}\right)e$

$$P_i = \frac{e_i}{E}$$

$$P_i = \frac{v_i/f_i}{\sum_j v_j/f_j}$$

$$P_i = \frac{v_i}{V}$$

Rivas & Conejeros PLoS ONE 2018

$$H(v) = -\sum_{i=1}^{R} P_i \log P_i$$

$$= -\sum_{i=1}^{R} \frac{vi}{V} \log \frac{vi}{V}$$

We defined a constraint-based model, MaxEnt, as:

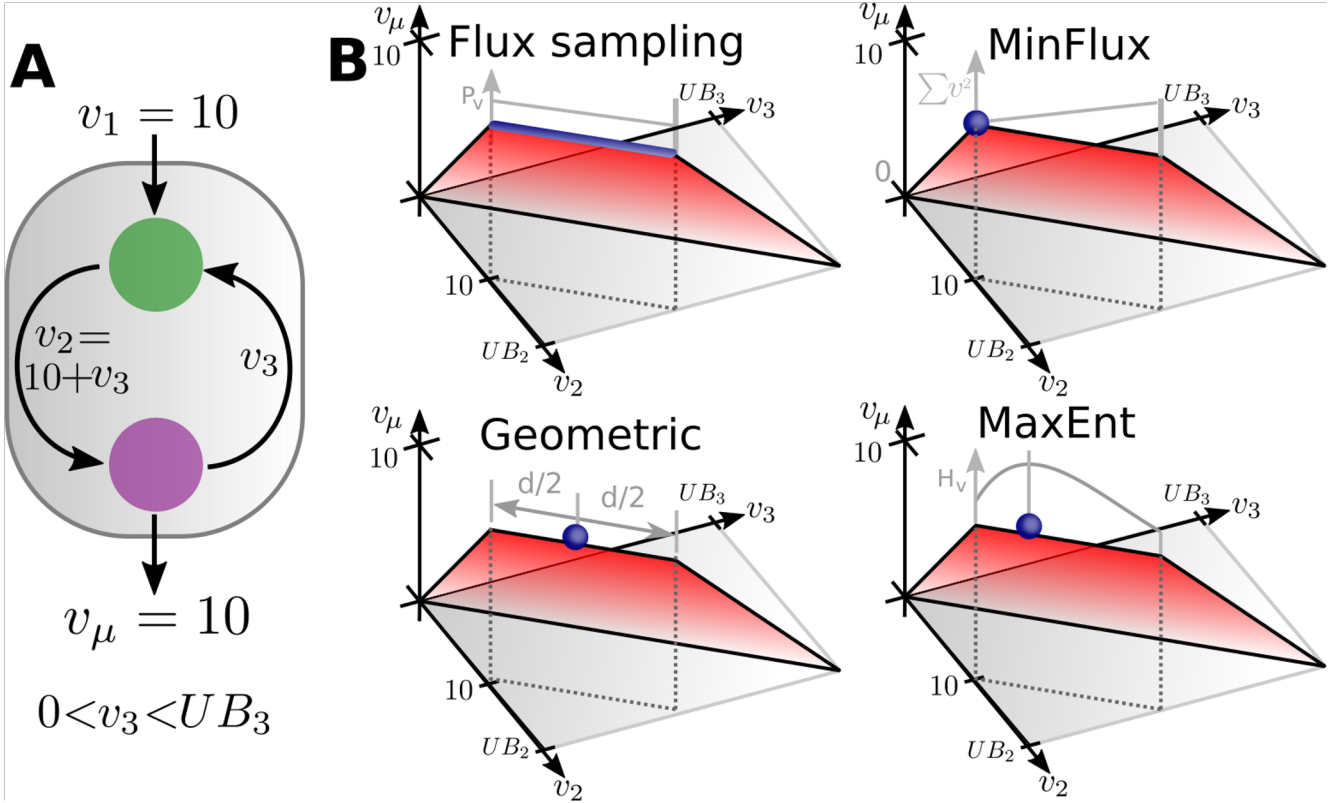
$$\max H(v)$$

$$\text{subject to:}$$

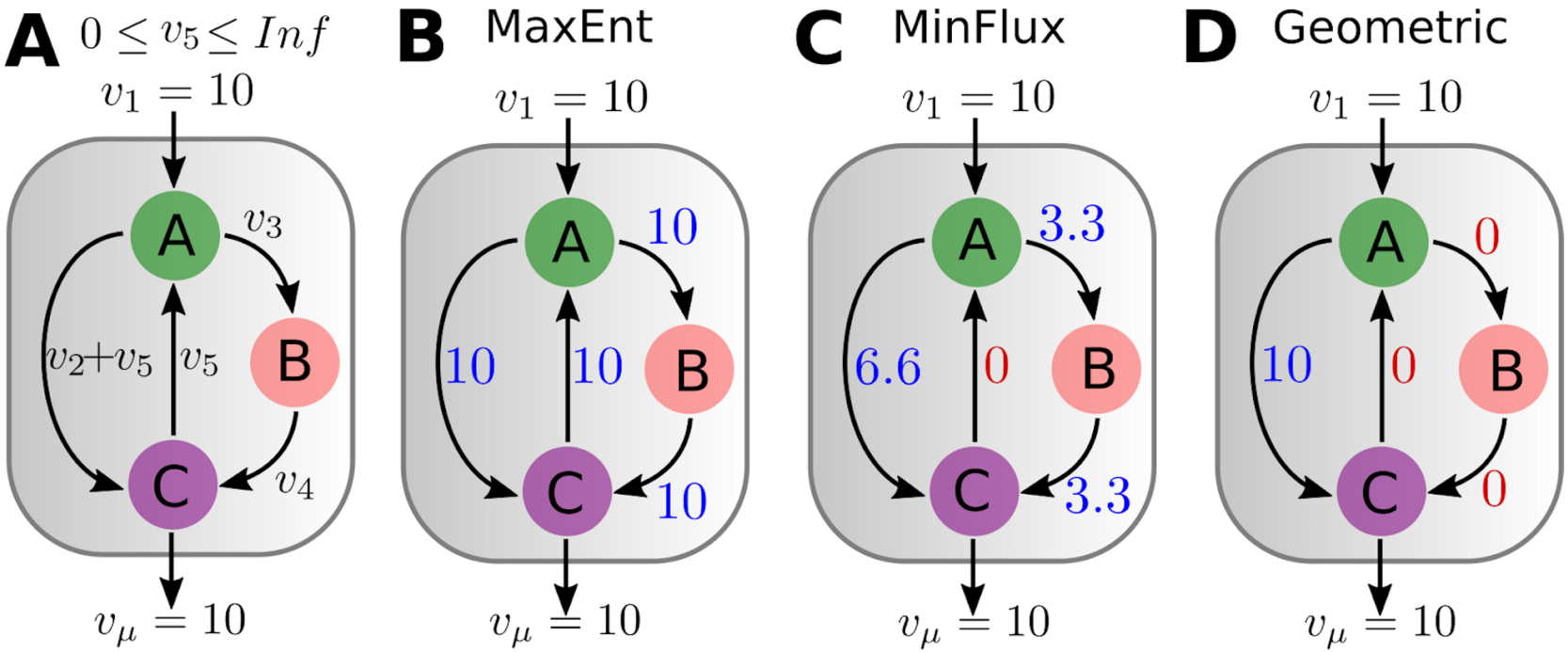
$$Sv = 0$$

$$LB \leq v \leq UB$$

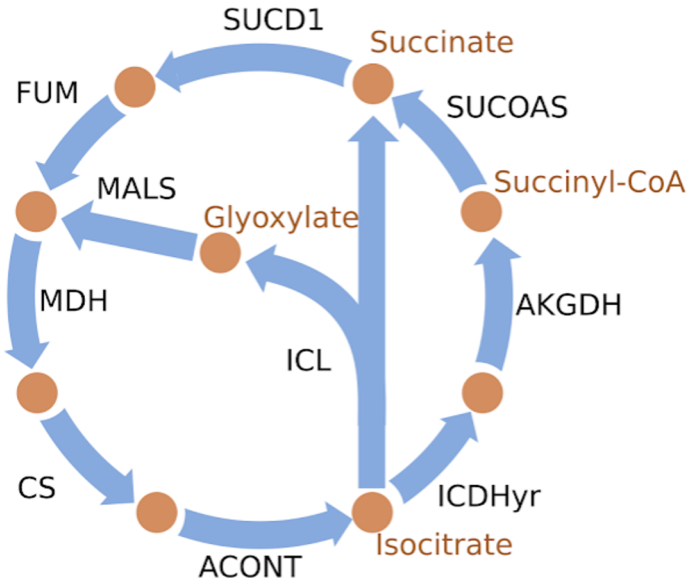
MaxEnt compared to alternative methods



MaxEnt does not eliminate flux loops nor produced fluxes reaching their bounds



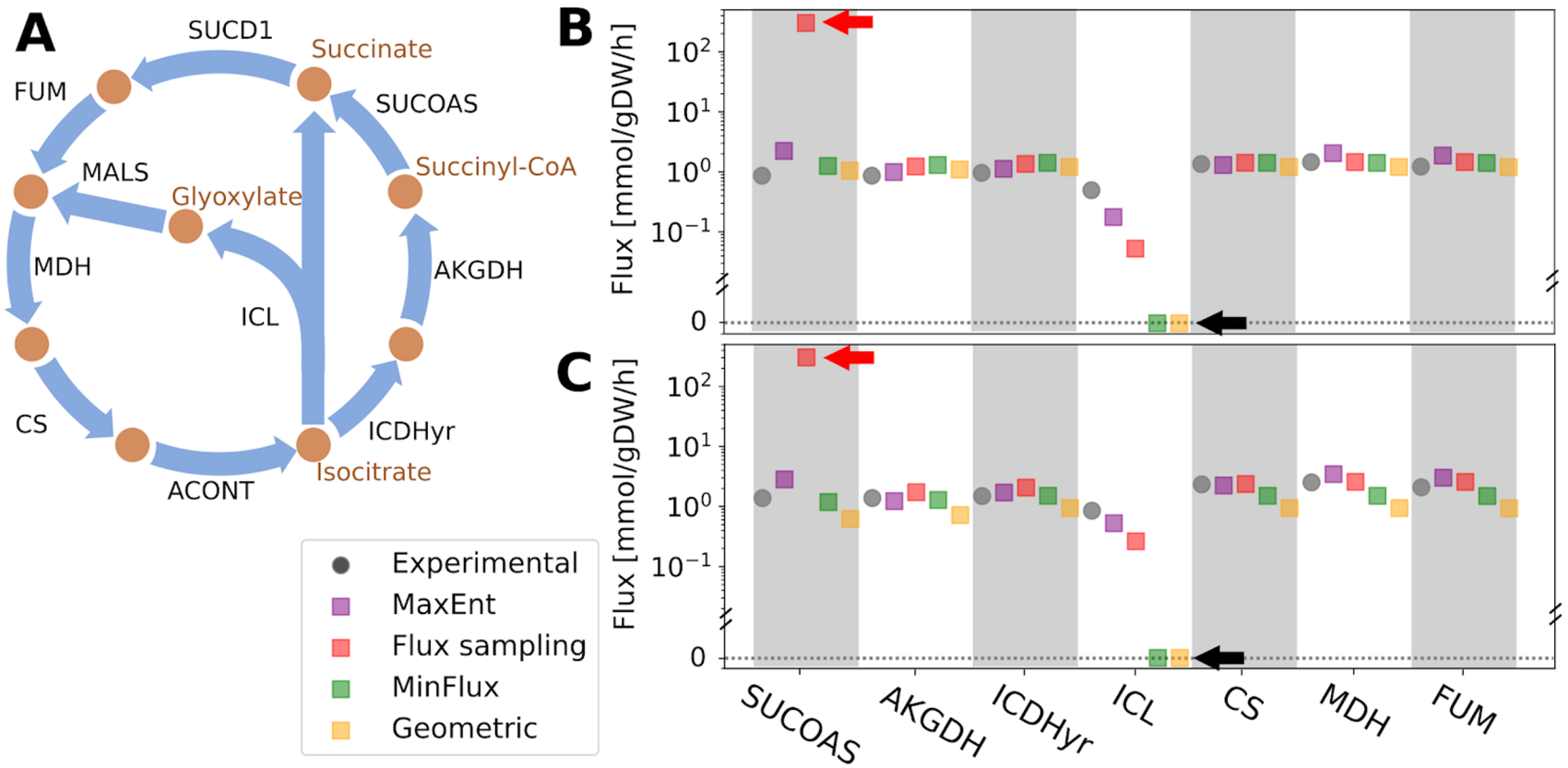
Some flux loops are thermodynamically feasible



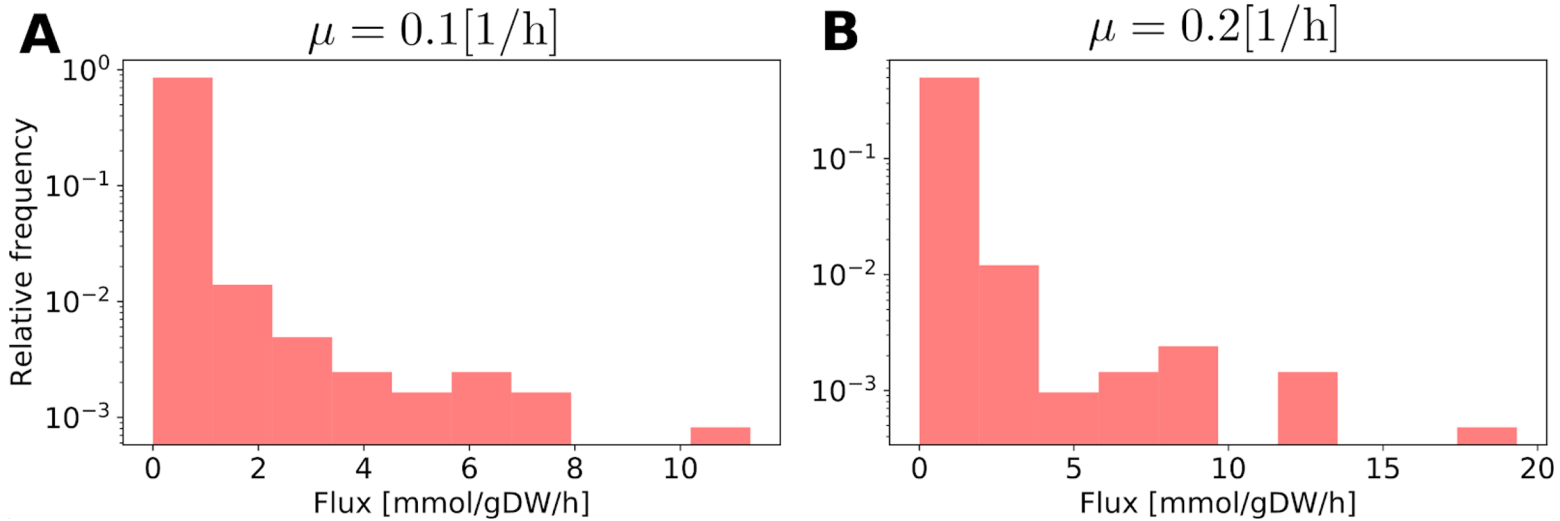
In *Escherichia coli*, it is know that the glyoxylate shunt carries flux.

Ishii et al 2007 Science

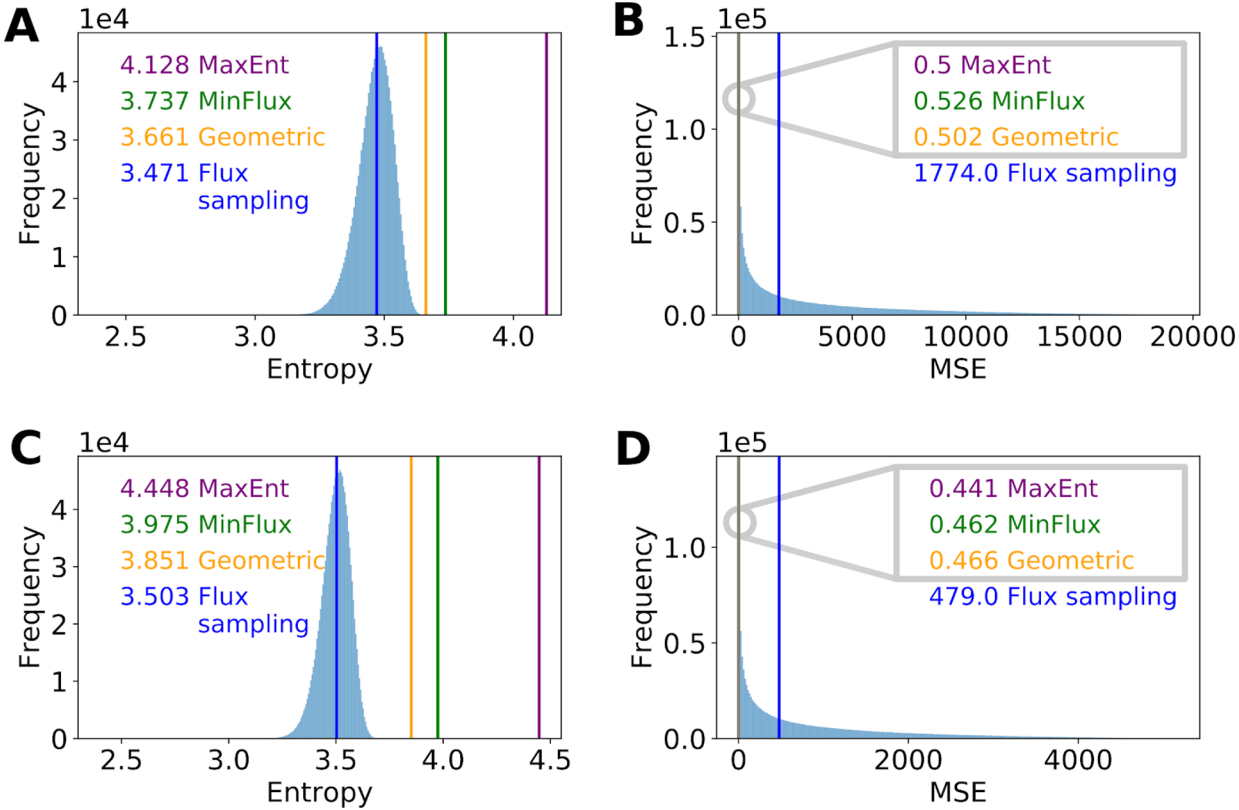
Results: MaxEnt does not eliminate flux loops nor produced fluxes reaching their bounds



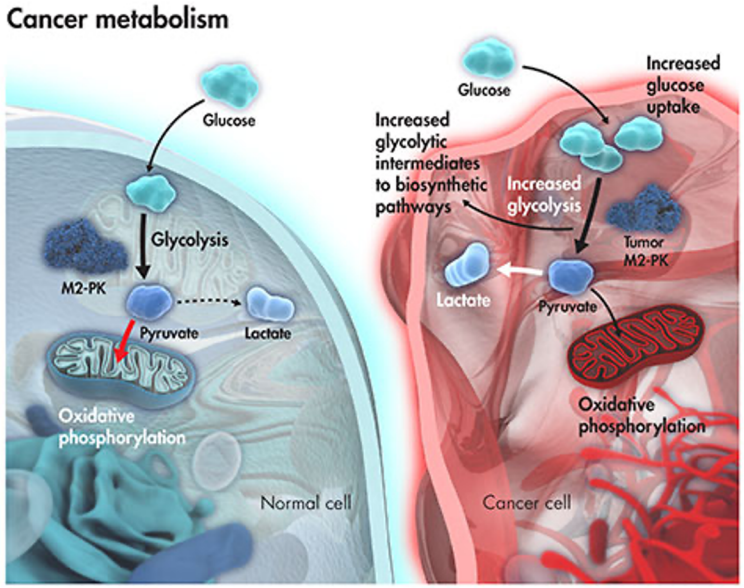
MaxEnt produces an structured distribution of fluxes (Almaas et al 2004 Nature)



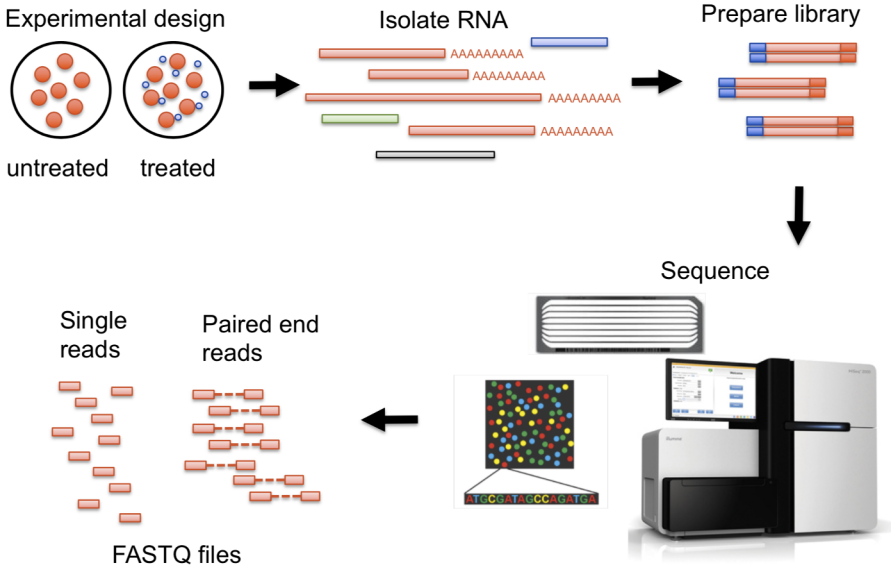
It is unlikely that flux sampling results in fluxomes with high entropy. On the other hand, MaxEnt produces better fluxome estimates than alternative methods.



Phenotype-specific estimations



Constraint-based models conditioned on phenotype-specific data



SPOT (Kim et al 2016)

$$\max \ \bar{v} \cdot \bar{g}$$

$$\text{subject to} \begin{cases} \bar{S}\bar{v} = 0 \\ 0 \leq \bar{v} \\ \|\bar{v}\|^2 \leq 1 \end{cases}$$

Defining *H* and adding gene expression (*g*) into it.

$$v_i = f_i e_i$$

$$v_i = f_i g_i$$

$$f_i = v_i / g_i$$

$$P_g(v_i) = \frac{v_i/g_i}{V}$$

$$\begin{aligned} H_g(v) &= -\sum_{i=1}^{R} \sum_{j=1}^{g_i} P_g(v_i) \log P_g(v_i) \\ &= -\sum_{i=1}^{R} g_i P_g(v_i) \log P_g(v_i) \\ &= -\sum_{i=1}^{R} g_i \frac{v_i/g_i}{V} \log \frac{v_i/g_i}{V} \\ &= -\sum_{i=1}^{R} \frac{v_i}{V} \log \frac{v_i/g_i}{V} \end{aligned}$$

We called our approach Pheflux

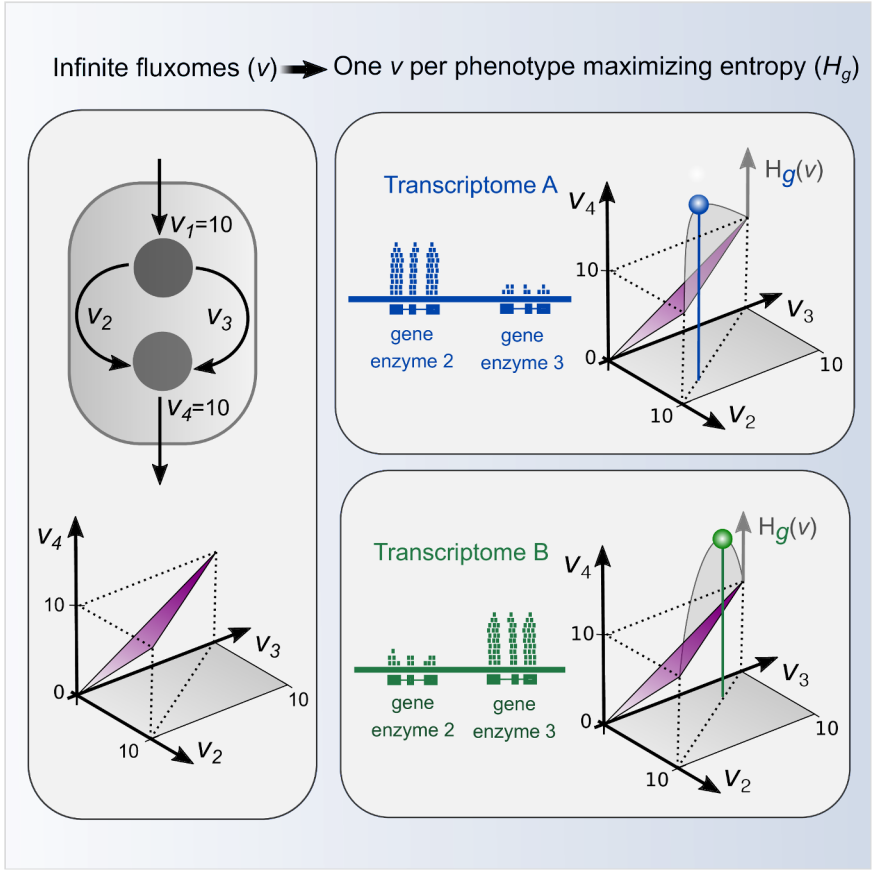
$$\max_{v} H_g(v)$$

subject to:

$$Sv = 0$$

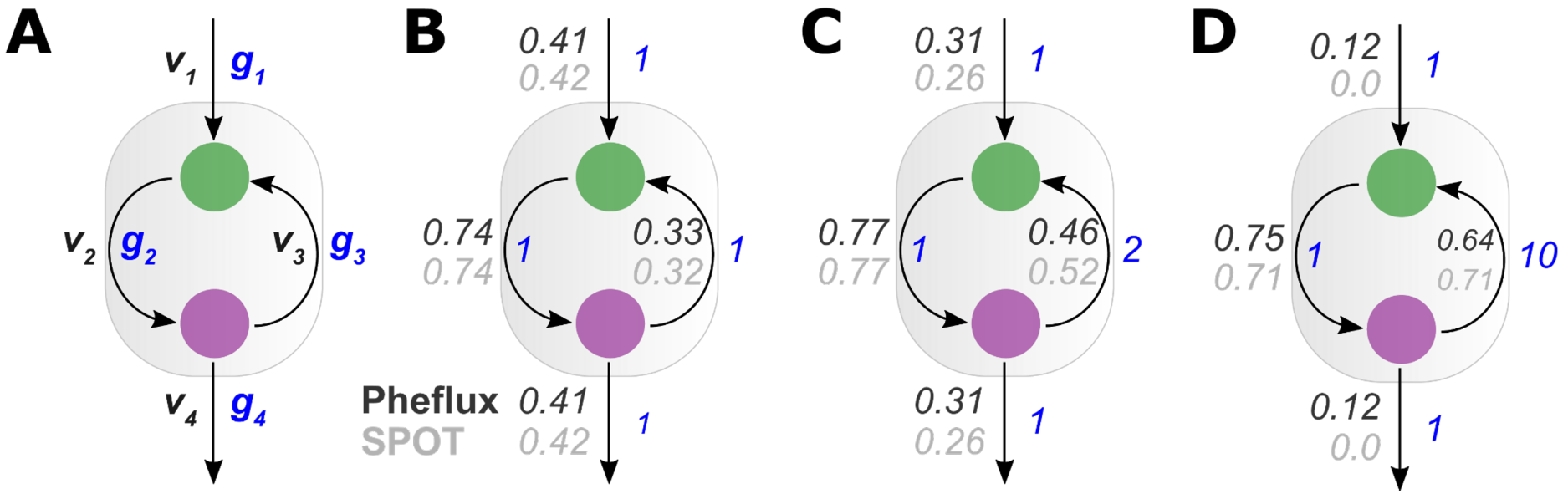
$$LB \leq v \leq UB$$

$$V = k$$



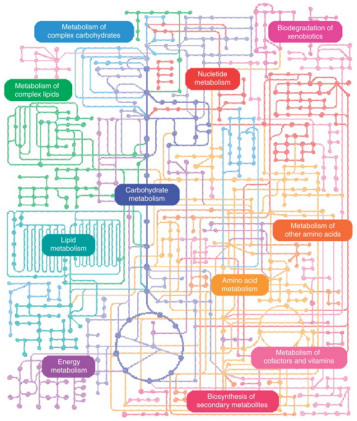
Gonzalez, Inostroza, Conejeros & Rivas
iScience 2023

How does Pheflux compare to SPOT?



How does Pheflux performance compares to alternative methods?

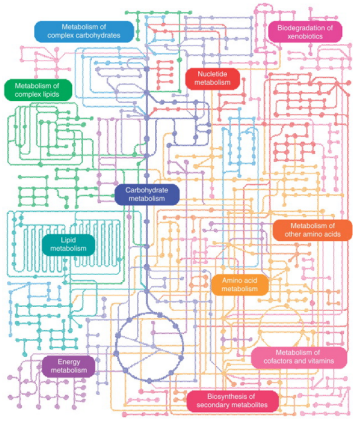
We used as benchmark C13 labeling (~20 fluxes)

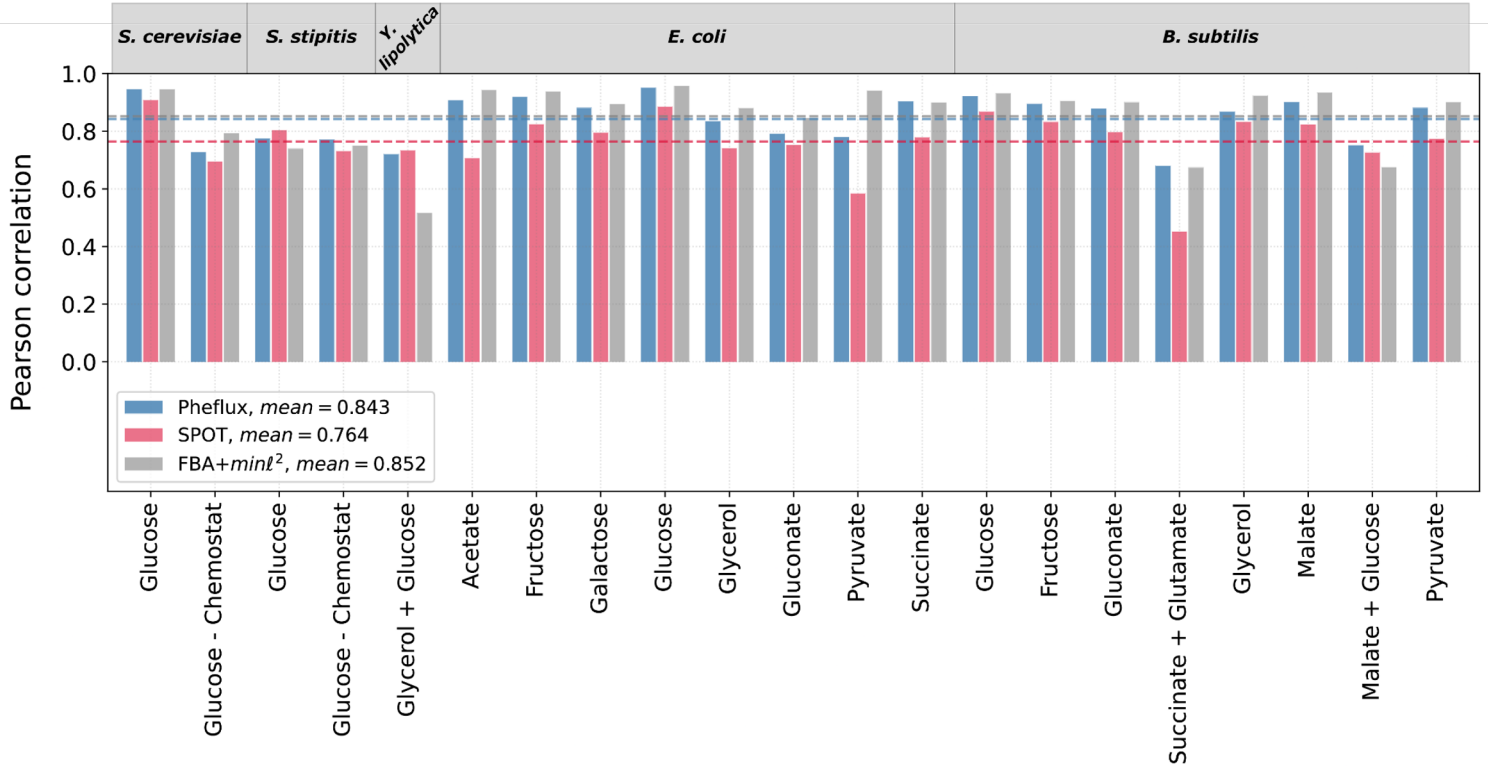


Organism & genome-scale metabolic model	Culture conditions	Transcriptomic Data	Fluxomic Data
S. cerevisiae iMM904 [1]	Two conditions —chemostat and batch— supplemented by glucose as carbon source.	Nookaew et al. (2012)[2]: Data measured using RNA-seq technology. Three replicates per condition. Normalized by FPKM.	Papini et al. (2012)[3]: Fluxes measured using ^{13}C labeled. No replicates.
S. stipitis iTL885 [4]	Two conditions —chemostat and batch— supplemented by glucose as carbon source.	Papini et al. (2012)[3]: Data measured using RNA-seq technology. Three replicates per condition. Normalized by FPKM.	Papini et al. (2012)[3]: Fluxes measured using ^{13}C labeled. No replicates.
Y. lipolytica iYali [5]	One condition —mixed culture— supplemented by glycerol and glucose as carbon source.	Sabra et al. (2017)[6]: Data measured using RNA-seq technology. Two replicates. Normalized by FPKM.	Sabra et al. (2017)[6]: Fluxes measured using ^{13}C labeled. No replicates.
E. coli iJO1366[7]	Eight conditions supplemented by glucose, gluconate, galactose, succinate, pyruvate, glycerol, succinate, acetate and fructose, respectively.	Gerosa et al. (2015)[8]: Data measured using microarray technology. Three replicates per condition. Normalized by quantile normalization.	Gerosa et al. (2015)[8]: Fluxes measured using ^{13}C labeled. No replicates.
B. subtilis iYO844 [9]	Eight conditions supplemented by glucose, fructose, gluconate, succinate + glutamate, glycerol, malate, malate + glucose and pyruvate, respectively.	Nicolas et al. (2012) [10]: Data measured using microarray technology. Three replicates per condition. Normalized by quantile normalization.	Chubukov et al. (2013) [11]: Fluxes measured using ^{13}C labeled. No replicates.

How does Pheflux performance compares to alternative methods?

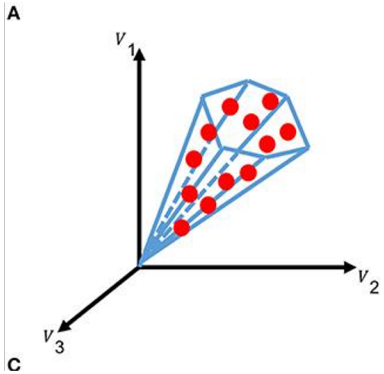
We used as benchmark C13 labeling (~20 fluxes)

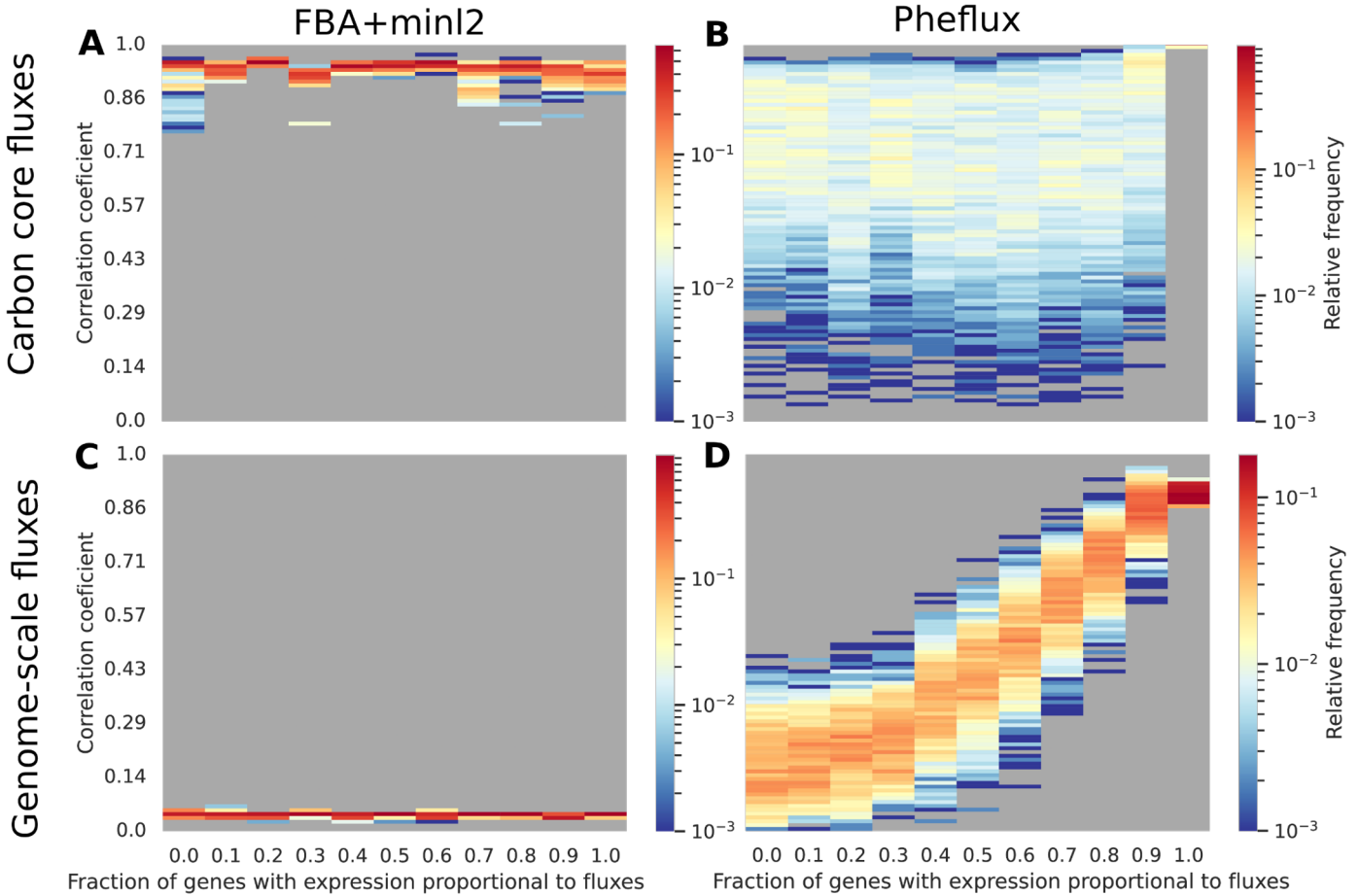




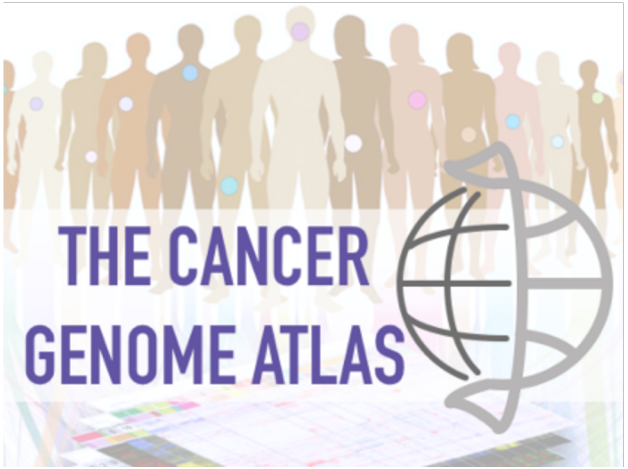
Does the situation changes at genome-wide scale?

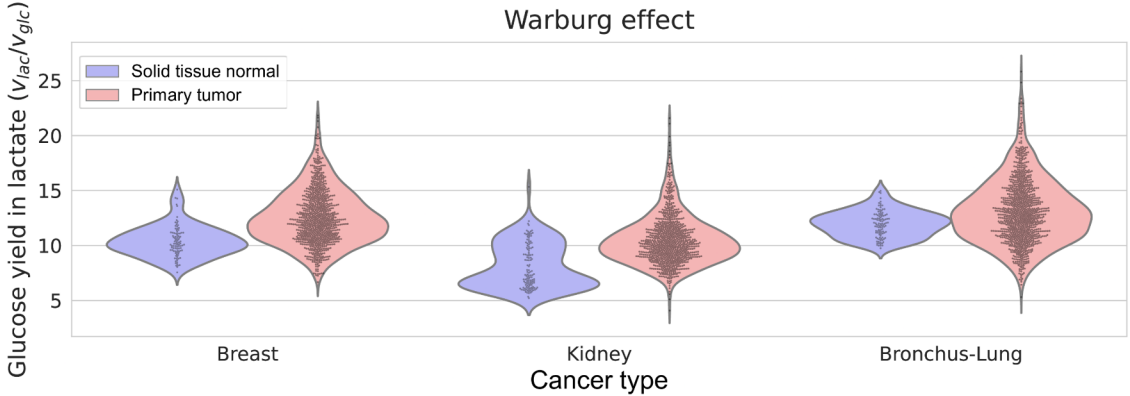
We simulated 1000 fluxomes using uniform sampling from the fluxome space.



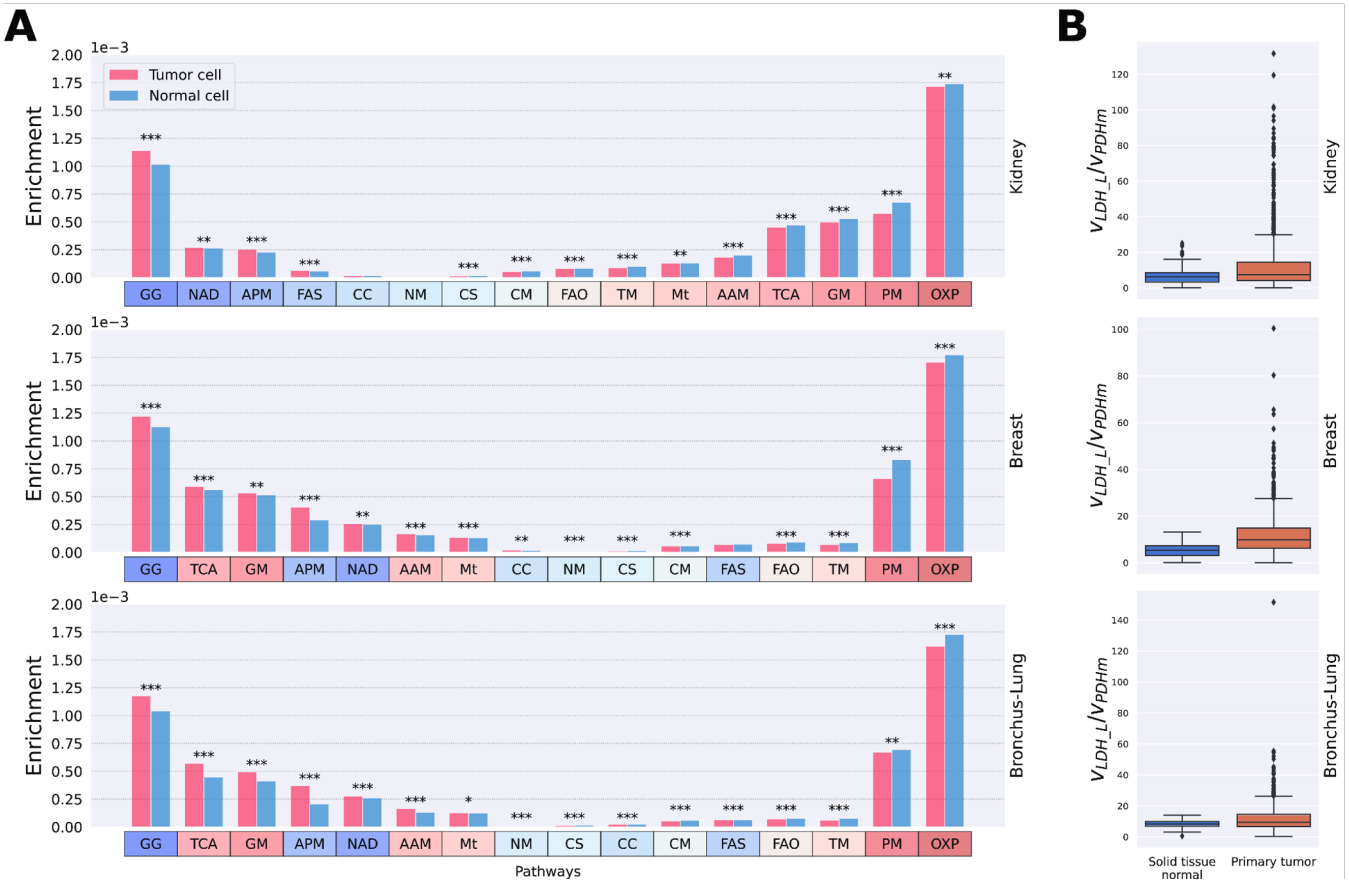


Does Pheflux recapitulates the Warburg effect?

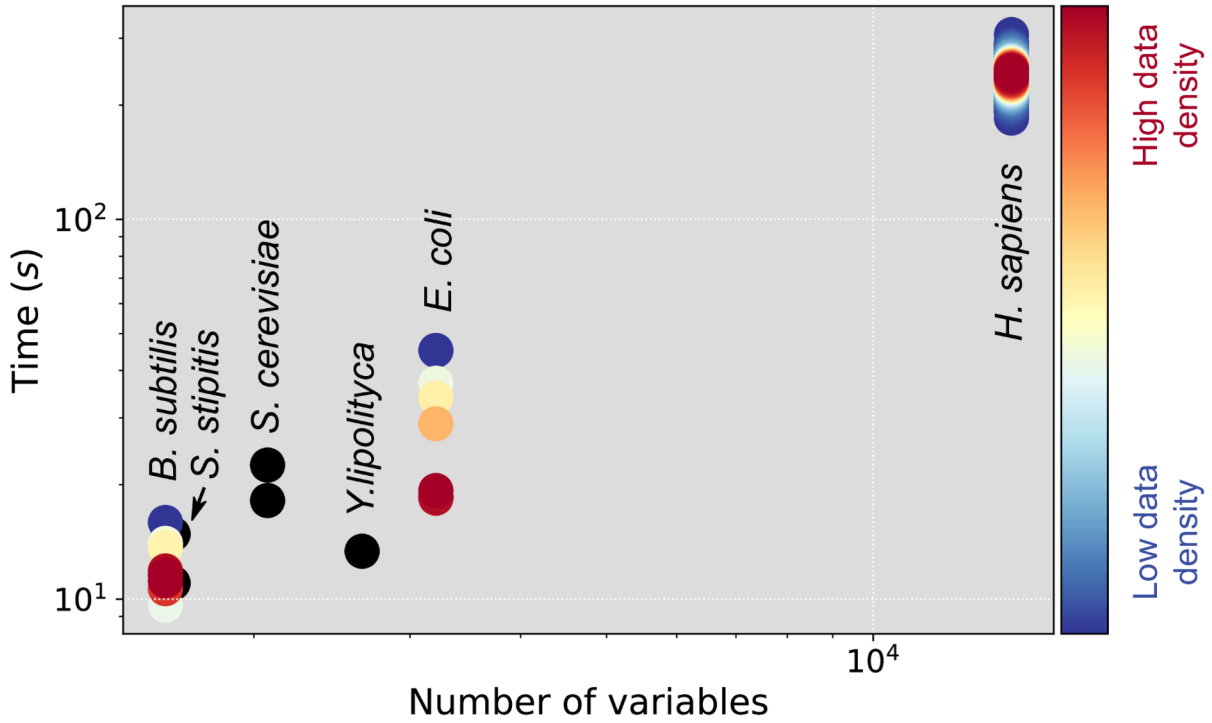




How Pheflux inferences add new insights into cancer metabolism?



Results: Can Pheflux be run in a reasonable time?



Conclusions

Pheflux:

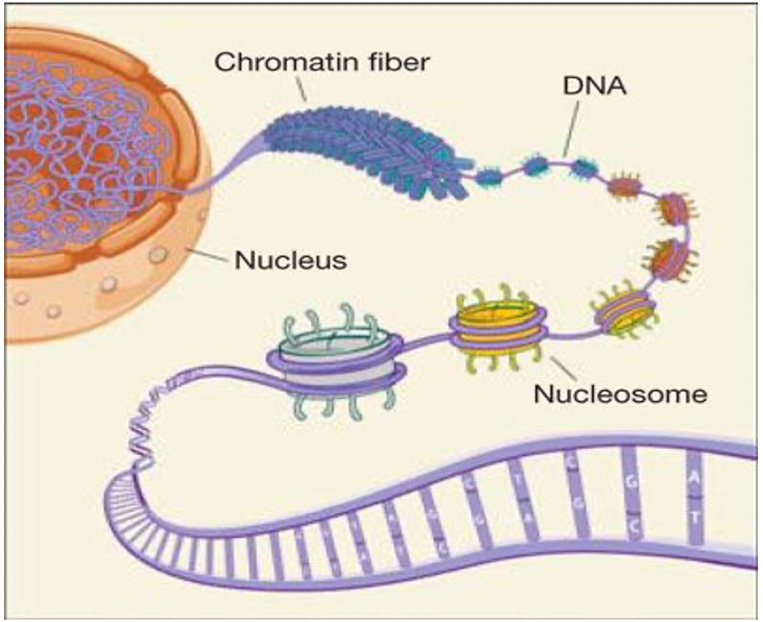
1. outperforms alternative CBMs (SPOT and FBA at genome-wide scale),
2. it produces phenotype-specific predictions that matches the literature (Warburg effect),
3. it may inform therapeutic targets (experimental validation needed), and
4. it can be run to model genome-scale models.

Room for improvement:

1. Pheflux does not prevent thermodynamically infeasible fluxes (M. Farias & N Améstica),
2. Using proteomic data, rather than gene expression data, should improved predictions.

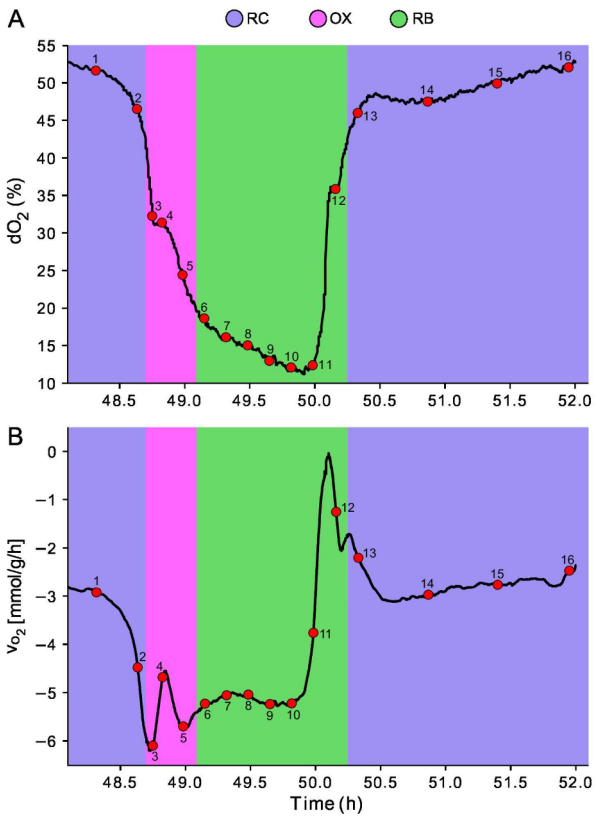
Work in progress: Epigenetics

Bárbara Guzmán
Universidad Tecnológica
Metropolitana

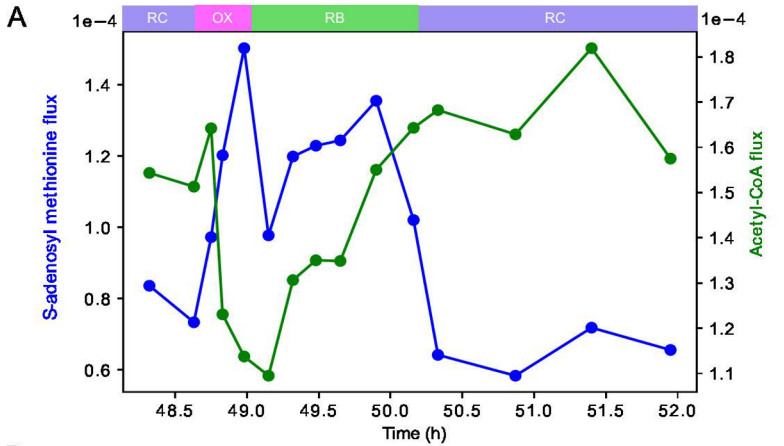


Baker 2011. Nature Methods

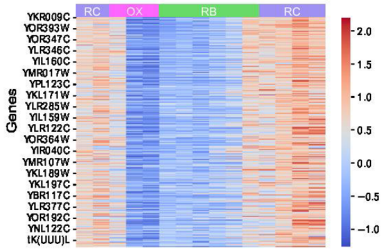
Work in progress:Epigenetics



Kuang et al Nat Struct Mol Biol 2014

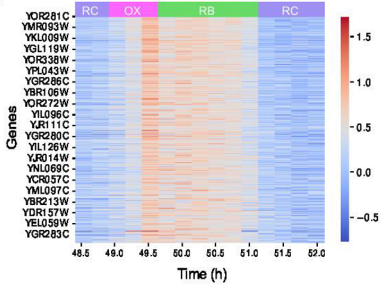


B



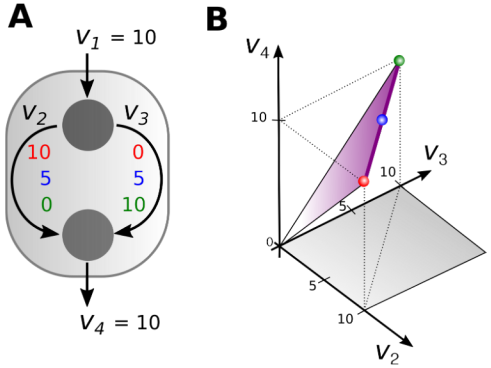
Function	P-Value	Benjamini
Peroxisome	2,50E-13	5,50E-12
Metabolic pathways	2,30E-09	8,70E-08
Carbon metabolism	2,90E-08	7,20E-07
Fatty acid beta-oxidation	1,00E-09	5,10E-07
Oxidoreductase activity	7,80E-10	2,30E-07
Fatty acid metabolic process	2,70E-08	6,50E-06
Biosynthesis of secondary metabolites	3,60E-07	5,40E-06
Fatty acid degradation	1,70E-07	3,10E-06
Peroxisomal matrix	1,20E-06	8,50E-05
NAD	3,10E-06	7,80E-05
Oxidoreductase	2,80E-07	1,10E-05
Fatty acid metabolism	1,00E-05	6,60E-04

C

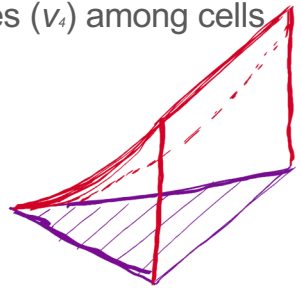


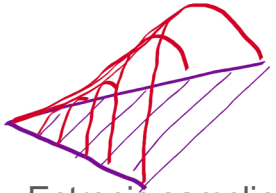
Function	P-Value	Benjamini
Ribosome biogenesis	2,30E-70	2,10E-67
rRNA processing	2,20E-58	9,90E-56
Nucleolus	2,50E-56	6,80E-54
Nucleoplasm	1,70E-33	2,30E-31
Small-subunit processome	2,50E-33	2,40E-31
Nucleus	3,10E-32	2,10E-30
90S preribosome	1,70E-29	9,50E-28
Ribonucleoprotein complex	1,20E-22	5,10E-21
RNA binding	3,00E-15	1,50E-12
Cytoplasmic translation	6,90E-14	8,90E-12
mRNA binding	2,80E-09	7,00E-07
RNA helicase activity	1,70E-07	2,20E-05

Work in progress: Entropic sampling



Entropic sampling of growth rates (v_4) among cells





Entropic sampling of fluxomes among cells

Selection of the most entropic fluxome (v) given v_1

$$H(v) = -\sum_{i=1}^{N} \frac{v_i}{V} Log(\frac{v_i}{V})$$

Andrea De Martino
Politecnico di Torino

Daniele De Martino
Biofisika Bizkaia Fundazioa

Raúl Conejeros
Pontificia Universidad Católica de Valparaíso

Thanks for your attention

This project was supported by the

- UTEM Internal Research Projects Grant (code LPR20-07), and
- The System of High Performance Computing PIDi-UTEM (SCC-PIDi-UTEM - CONICYT - FONDEQUIP - EQM180180).