

Genome-scale metabolic models



<http://sysmedicine.com>

Dr. Cheng Zhang
KTH-Royal Institute of Technology

Genome-scale metabolic models (GEMs)

1. Concept of GEMs
2. Flux balance analysis
3. GEMs vs metabolic networks
4. Reconstruction of GEMs
5. Applications of GEMs

What is a GEM

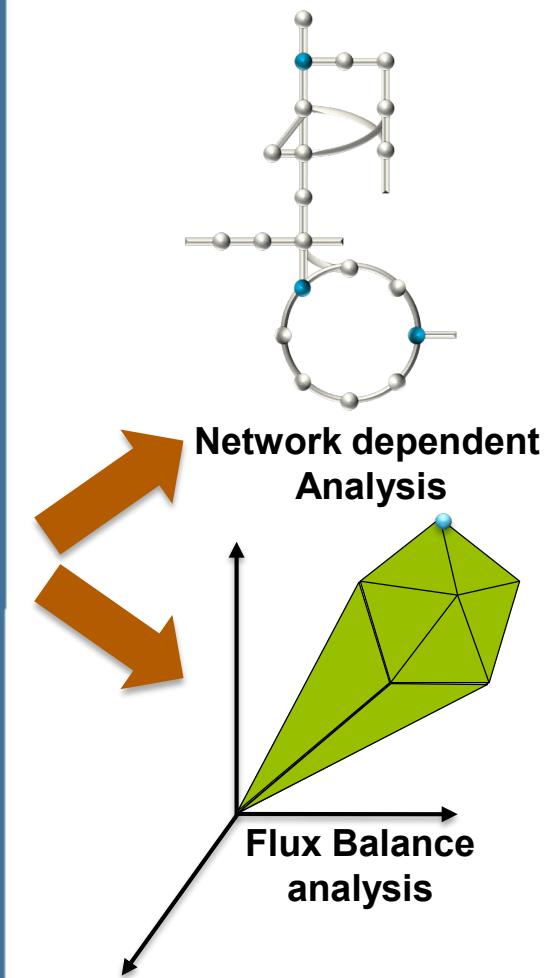
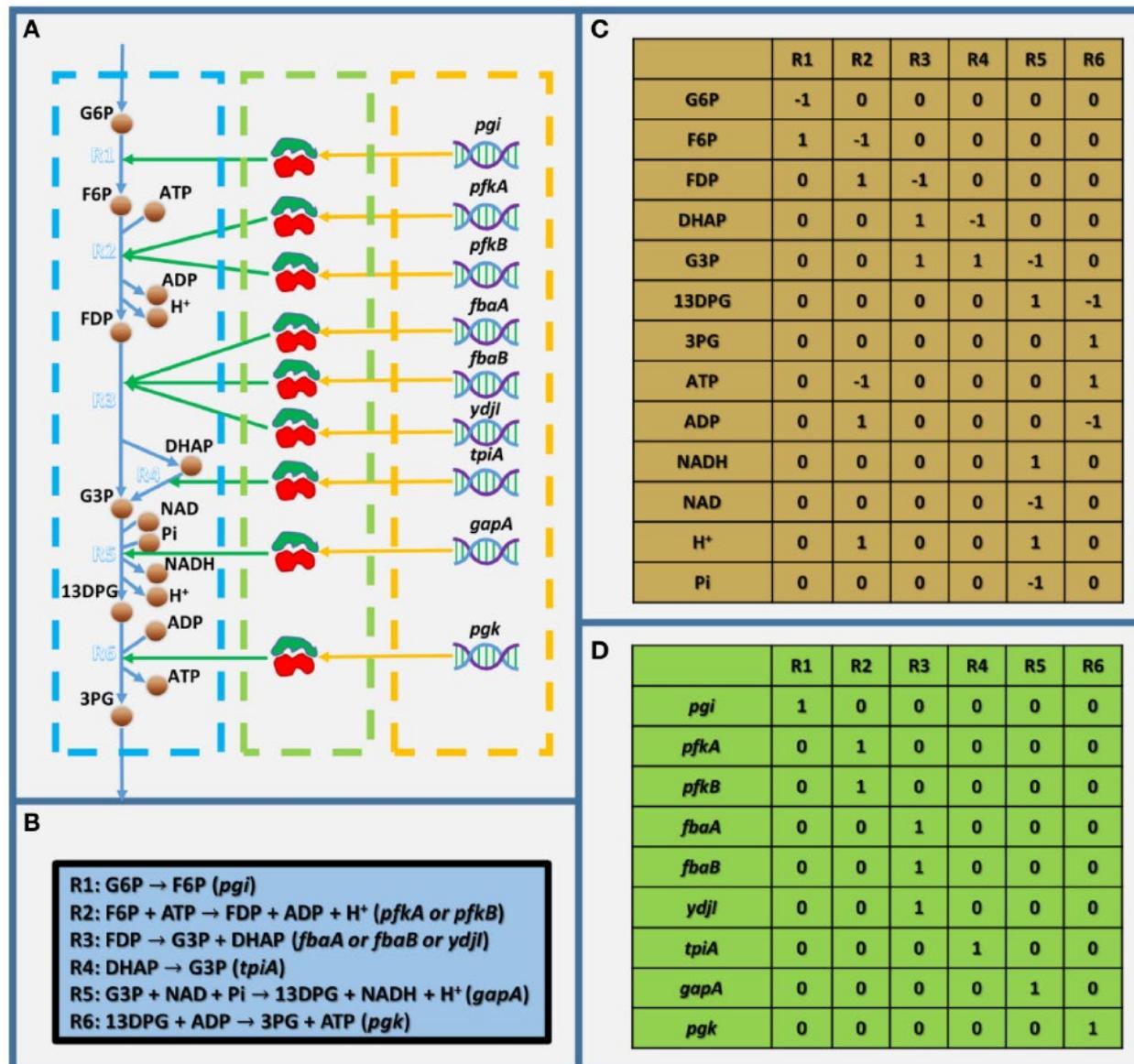
Genome-scale metabolic models (GEMs) are mathematical reconstructions of the metabolic networks with all known metabolic reactions of many kinds of cells, including those of microorganisms, plants, and mammals.

In some cases, GEMs could represent the whole tissue or body of a multicellular organism.

In these metabolic networks, the gene-protein-reaction (GPR) relationships are annotated.

In addition, all the reactions in GEMs are mass-balanced, ensuring stoichiometric balance.

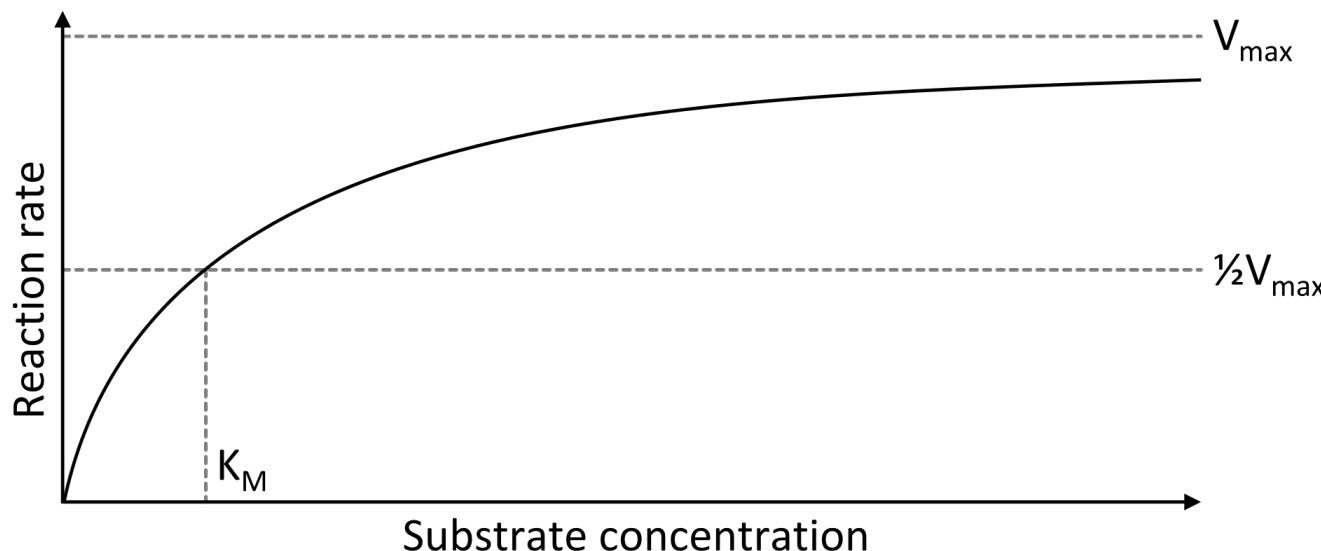
Structure of GEMs



Flux describes any effect that appears to pass or travel (whether it actually moves or not) through a surface or substance.

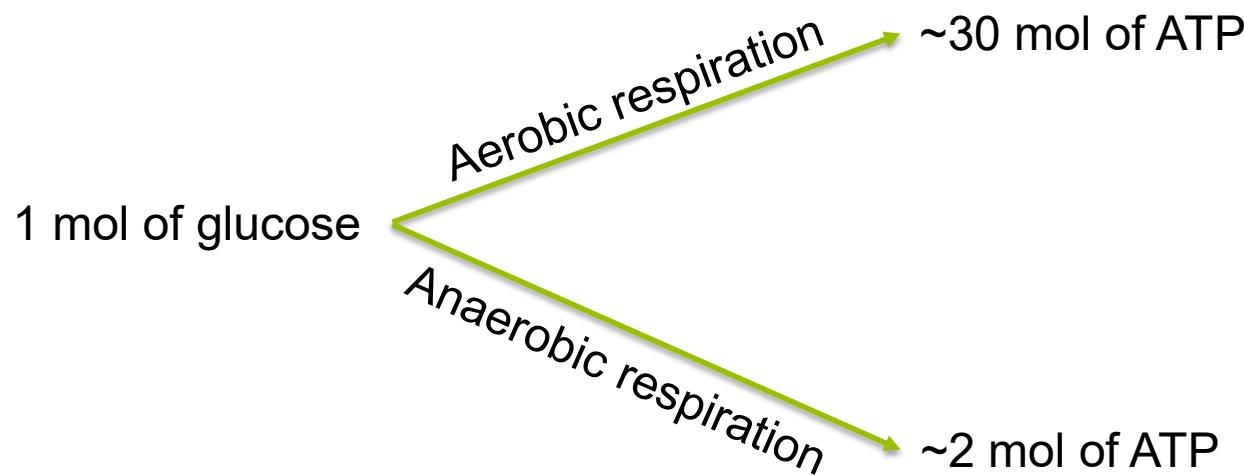
In flux balance analysis, flux is the representative of the metabolic flow of a metabolic reaction. One very good example is Michaelis–Menten kinetics, where the flux was defined as the speed of the enzymatic reaction. The most common unit used in genome-scale metabolic models (GEMs) in this kind of scenario is mmol/(gDCW*h).

$$v = \frac{d[P]}{dt} = V_{\max} \frac{[S]}{K_M + [S]} = k_{\text{cat}} [E]_0 \frac{[S]}{K_M + [S]}.$$



What is flux

There is also common scenario, that we use flux to represent the maximum (theoretical) yield.

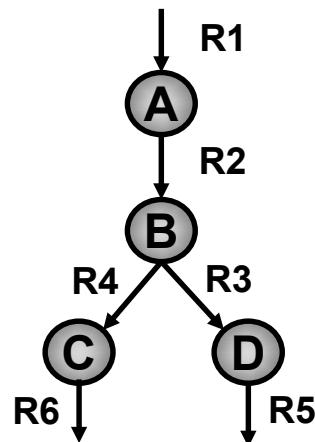


Both the yields of aerobic and anaerobic respiration that catabolize glucose and generate ATP can be seen as proxy of fluxes.

Why flux balance analysis?

“Flux balance analysis (FBA) is a mathematical method for simulating metabolism in genome-scale reconstructions of metabolic networks. In comparison to traditional methods of modeling, FBA is less intensive in terms of the input data required for constructing the model.”

$$v = \frac{d[P]}{dt} = V_{\max} \frac{[S]}{K_M + [S]} = k_{\text{cat}} [E]_0 \frac{[S]}{K_M + [S]}.$$



How many parameters we need to measure in this case to calculate flux in this case?

Flux balance analysis (FBA)

Mathematical formulation of FBA:

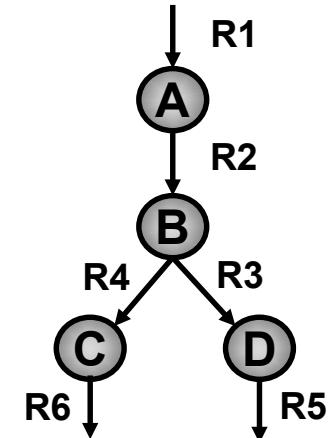
$$S^*v = b$$
$$LB \leq v \leq UB$$

S: the stoichiometric matrix with rows and columns respectively represent metabolites and reactions

v: a column vector with each element represents a flux of a reaction

b: a column vector with each element represents the production/consumption of a metabolite

LB/UB: a column vector with each element represents the minimum/maximum flux in the corresponding reaction



$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} b_A \\ b_B \\ b_C \\ b_D \end{pmatrix}$$

$$\begin{array}{lcl} v_1 - v_2 & & = b_A \\ v_2 - v_3 - v_4 & & = b_B \\ v_4 & & - v_6 = b_C \\ v_3 & & - v_5 = b_D \end{array}$$

Pseudo-steady state assumption: No unexpected accumulation or consumption of intra-cellular metabolite.

Which, in the formulation, means,

$$S^*v = 0$$

This is equal to assuming that,

$$b_i = 0 \quad i \in \{A, B, C, D\}$$

Biological meaning? Advantage and disadvantage?

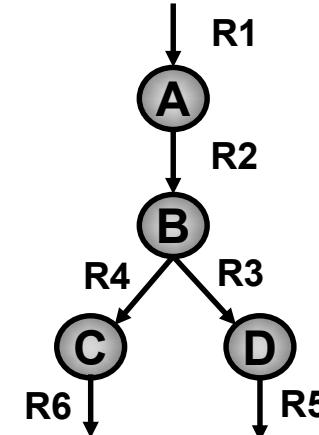
Pseudo-steady state

$$v_i \geq 0 \quad i = \{1, 2, 3, 4, 5, 6\}$$

$$v_1 = 5 \text{ & } v_5 = 1$$

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

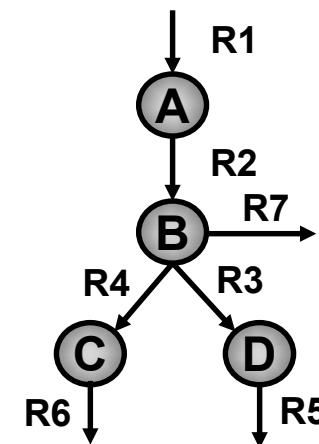


$$v_i \geq 0 \quad i = \{1, 2, 3, 4, 5, 6\}$$

$$v_1 = 5, v_5 = 1 \text{ & } v_7 = 1$$

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$



Pseudo-steady state assumption: No unexpected accumulation or consumption of intra-cellular metabolite.

Which, in the formulation, means,

$$S^*v = 0$$

This is equal to assuming that,

$$b_i = 0 \quad i \in \{A, B, C, D\}$$

Biological meaning? Advantage and disadvantage?

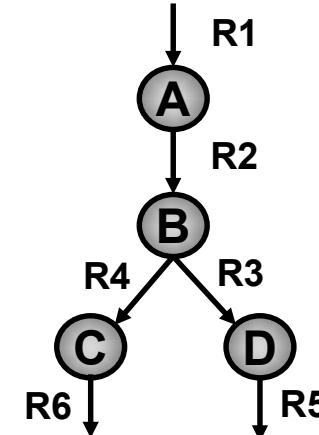
Pseudo-steady state

$$v_i \geq 0 \quad i = \{1, 2, 3, 4, 5, 6\}$$

$$v_1 = 5 \text{ & } v_5 = 1$$

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

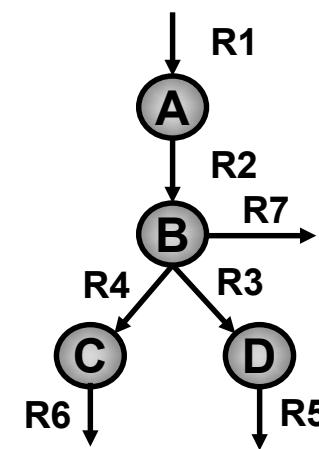


$$v_i \geq 0 \quad i = \{1, 2, 3, 4, 5, 6\}$$

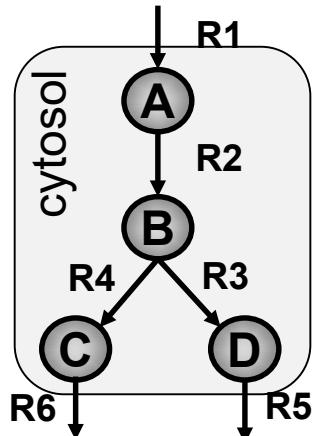
$$v_1 = 5, v_5 = 1 \text{ & } v_7 = 1$$

$$\begin{pmatrix} 1 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 & 0 & 0 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$



Exchange reactions & pool reactions

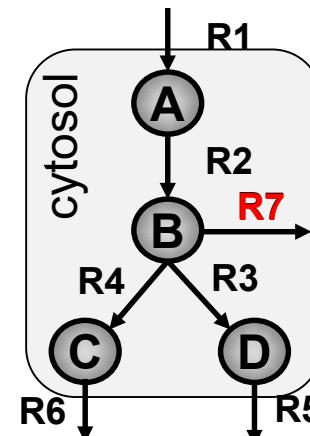


$$\begin{array}{c} \text{R1} \quad \text{R2} \quad \text{R3} \quad \text{R4} \quad \text{R5} \quad \text{R6} \\ \text{A} \left(\begin{array}{cccccc} 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & 0 & -1 & 0 \end{array} \right) \times \begin{array}{c} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_7 \end{array} = \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \end{array}$$

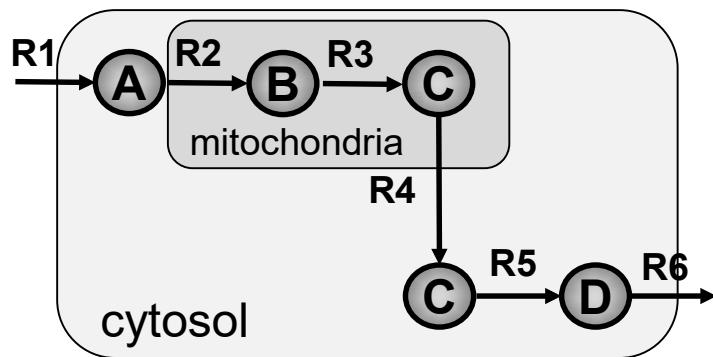
R1, R5 and R6 are exchange reactions, the others are intracellular reactions.

Why are exchange reaction important?

And what are pool reactions?



Compartmentalization



	R1	R2	R3	R4	R5	R6
A	1	-1	0	0	0	0
B	0	1	-1	0	0	0
C	0	0	1	?	-1	0
D	0	0	0	0	1	-1

In GEMs, metabolites present in different cellular compartment are treated as different metabolites even if they are actually the same molecule.

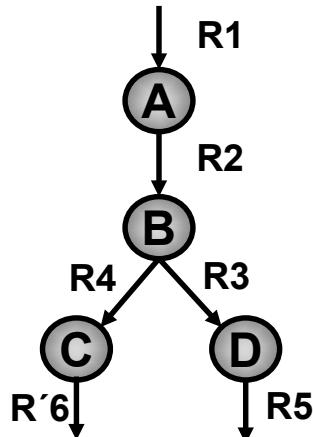
Why compartmentalization could be important?

	R1	R2	R3	R4	R5	R6
A[c]	1	-1	0	0	0	0
B[m]	0	1	-1	0	0	0
C[m]	0	0	1	-1	0	0
C[c]	0	0	0	1	-1	0
D[c]	0	0	0	0	1	-1

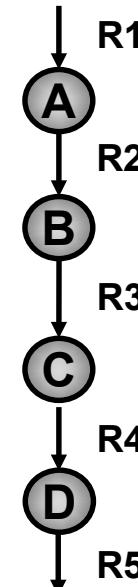
Degree of freedom in GEMs

In mathematics, a system of linear equations or a system of polynomial equations is considered underdetermined if there are fewer equations than unknowns (in contrast to an overdetermined system, where there are more equations than unknowns).

What is variable in GEMs? What is equation? Is it common that we have an underdetermined system? Could it be possible to have a GEM with overdetermined systems?



	R1	R2	R3	R4	R5	R6
M1	1	-1	0	0	0	0
M2	0	1	-1	-1	0	0
M3	0	0	0	1	0	-1
M4	0	0	1	0	-1	0



This is why constraint
are important!

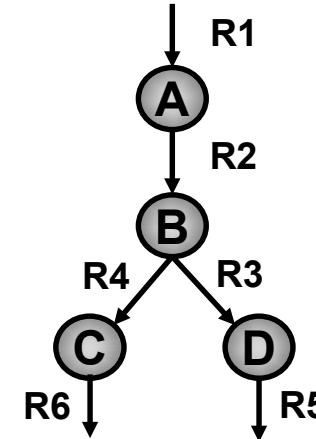
Objective function

Mathematical formulation of FBA with objective function:

$$\begin{array}{ll} \text{maximize} & Z = c^T v \\ \text{subject to} & S^* v = 0 \\ & l_b \leq v \leq u_b \end{array}$$

c: a column vector weighting the contribution of all fluxes to the objective function

$$\begin{array}{rl} v_1 - v_2 & = 0 \\ v_2 - v_3 - v_4 & = 0 \\ v_4 & - v_6 = 0 \\ v_3 & - v_5 = 0 \\ v_1 & = 1 \end{array}$$



This model is still an underdetermined system with defined input!

$$c^T = [0, 0, 0, 0, 0, 1]$$

Objective function $Z = c_1 v_1 + c_2 v_2 + c_3 v_3 + c_4 v_4 + c_5 v_5 + c_6 v_6 = v_6$

Objective function

Any differences between these three objective functions in this model if $v_1 = 1$?

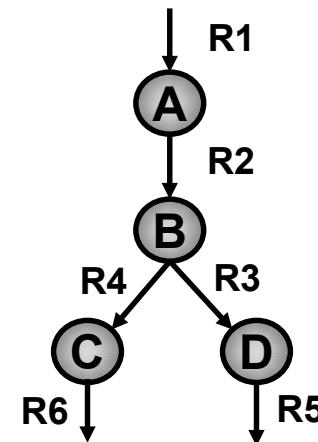
$$c^T = [0, 0, 0, 0, 0, 1] \quad Z = 1$$

$$c^T = [1, 1, 0, 0, 0, 1] \quad Z = 3$$

$$c^T = [0, 0, 0, 1, 0, 0] \quad Z = 1$$

$$c^T = [0, 0, 0, 0, 0.4, 0.6] \quad Z = 0.6$$

$$c^T = [0, 0, 0, 0, -1, 0] \quad Z = 0$$



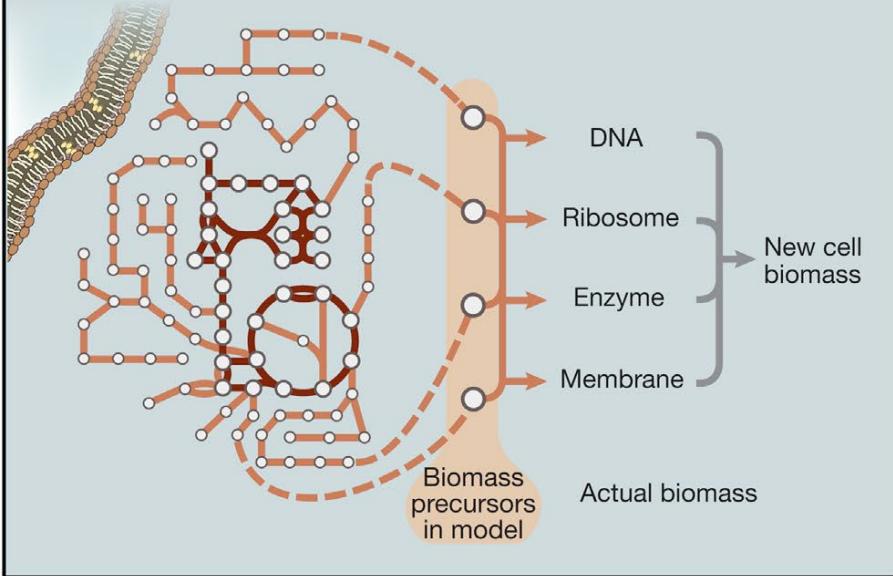
Even though the flux distribution will be identical, the value of the objective function can be different!

Objective function

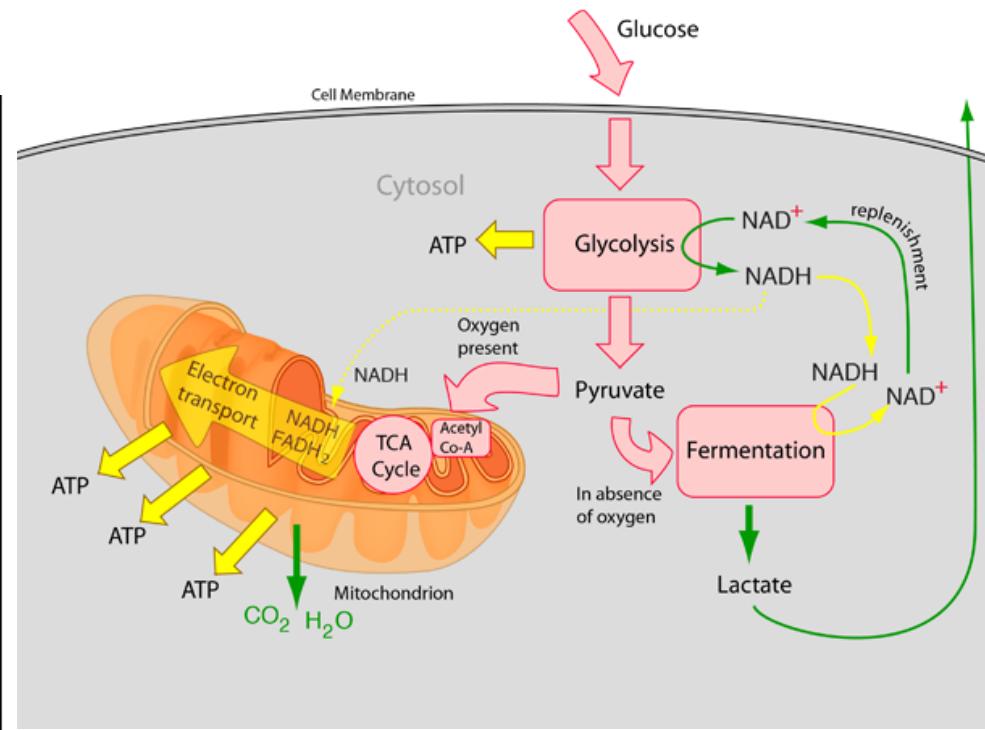
The common objective functions used

E

The biomass objective function



Growth



ATP generation

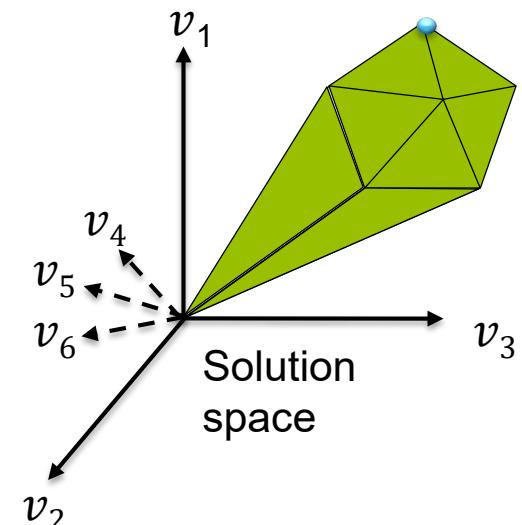
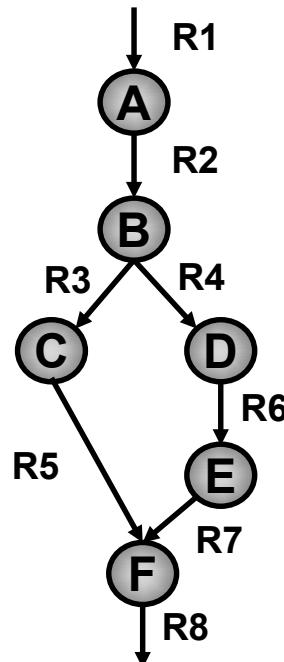
Why these two are good objective functions?

What if you optimize the glucose uptake?

Is objective function enough?

If $v_1 = 1$, and the objective function $Z = v_8$, can you obtain a fixed solution for this model?

$$v = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \text{ or } \begin{pmatrix} 1 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 1 \end{pmatrix} \text{ or } \dots \dots$$



Parsimonious FBA (pFBA)

Parsimonious FBA is a special kind of FBA which minimizes the sum of all fluxes in a GEM with defined constraints. This special kind of FBA assumes the cells always executes in the most metabolically economic way.

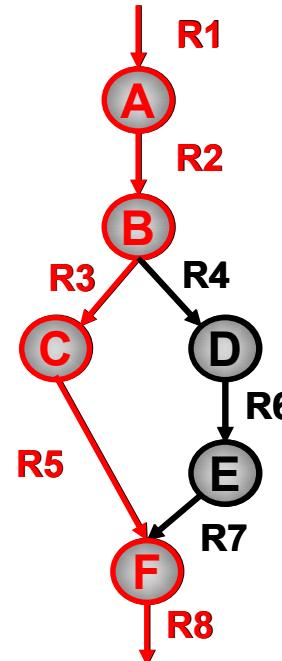
$$\text{minimize} \quad Z = \sum v_i$$

$$\text{s.t.} \quad S^*v = b$$

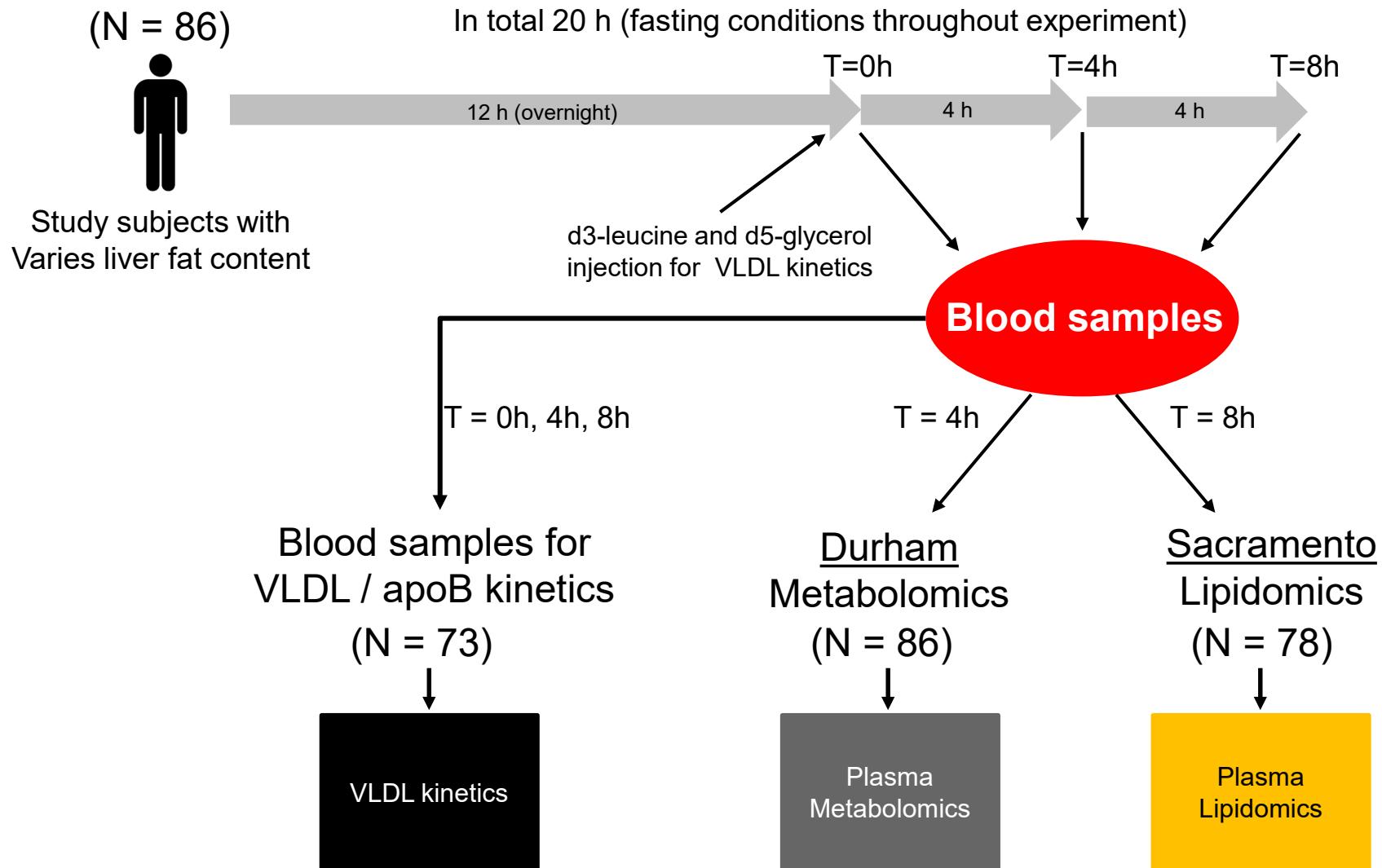
$$LB \leq v \leq UB$$

$$v = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad or \quad \begin{pmatrix} 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad or \quad \begin{pmatrix} 1 \\ 1 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 0.5 \\ 1 \end{pmatrix}$$

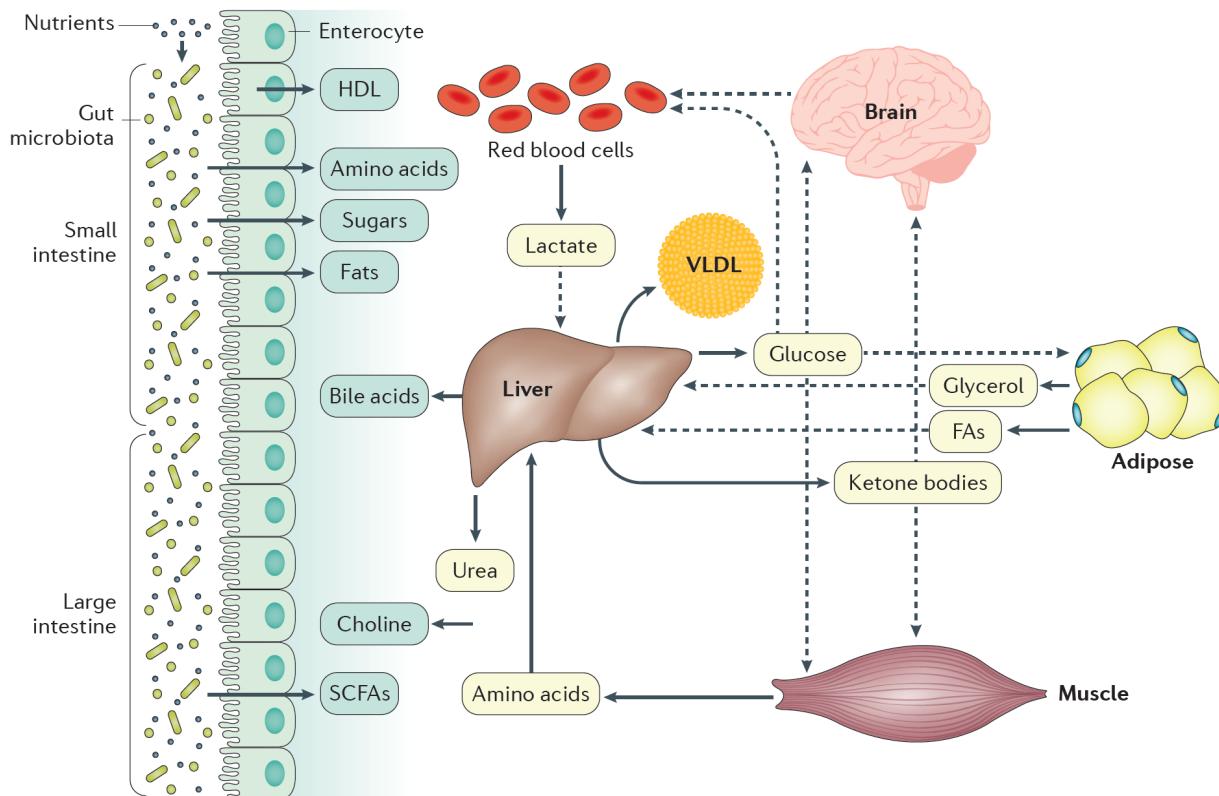
$$Z = 5 \quad Z = 6 \quad Z = 5.5$$



Subjects with NAFLD



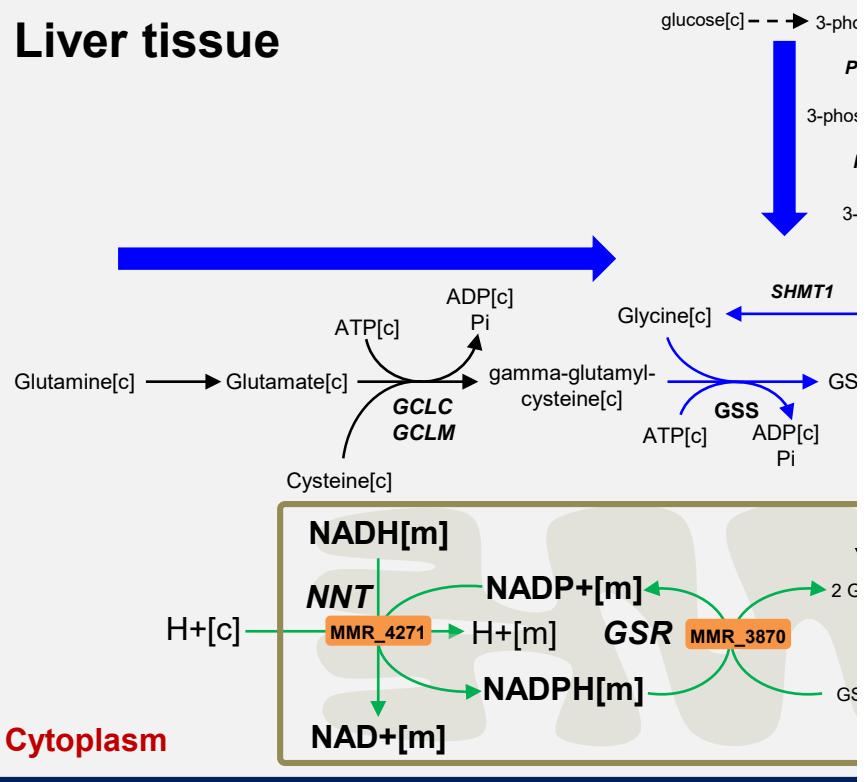
GEMs could be used to simulate the flux of different subject for comparison



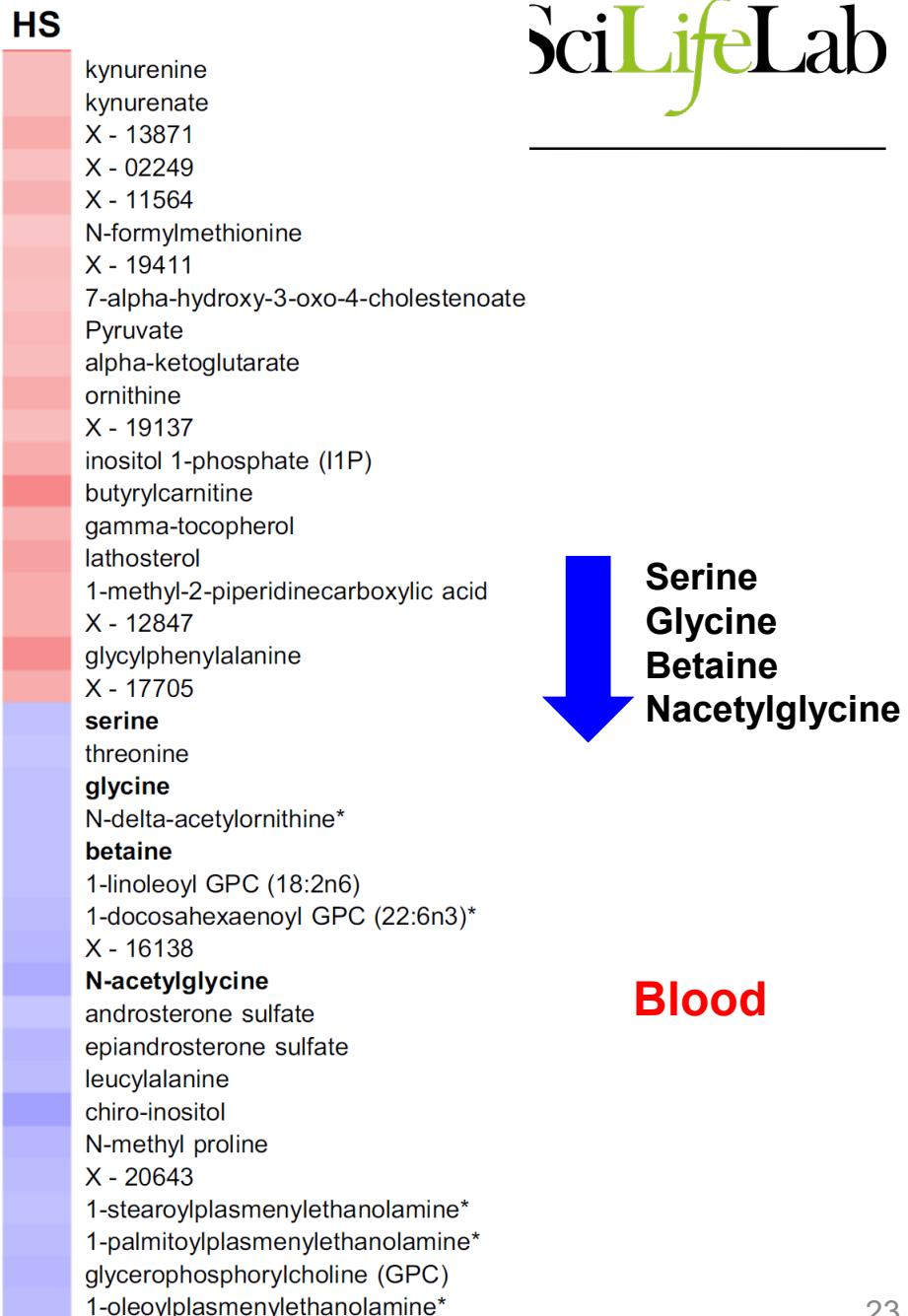
Nature Reviews | Gastroenterology & Hepatology

Correlation of intra with live

Liver tissue



Blue colour indicates down reg



Is pFBA enough?

If $v_1 = 1$, and the objective function $Z = \sum v_i$,
what will happen?

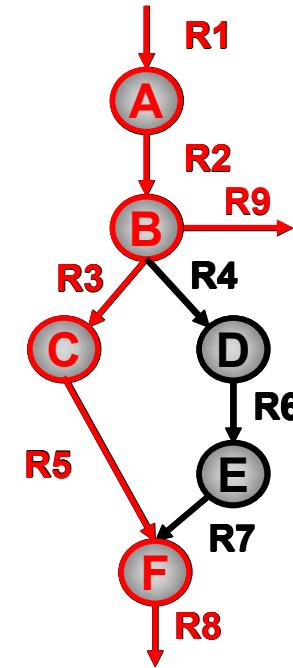
$$\text{minimize} \quad Z = \sum v_i$$

$$\text{s.t.} \quad S^*v = b$$

$$LB \leq v \leq UB$$

$$v_8 = v_8^{opt} = 1$$

$$v = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \quad v = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 0 \\ 1 \\ 0 \\ 0 \\ 1 \\ 0 \end{pmatrix}$$



Therefore, pFBA is often combined with biological objective function!

Flux variability analysis (FVA)

Flux variability analysis (FVA) is the repetition of FBA which maximize and minimize each of the reaction with the defined constraints.

for every reaction i in the model:

minimize/maximize

$$v_i$$

s.t.

$$S^*v = b$$

$$LB \leq v \leq UB$$

$$v_8 = v_8^{opt} = 1$$

FVA results

$$v_1 = 1$$

$$v_2 = 1$$

$$0 \leq v_3 \leq 1$$

$$0 \leq v_4 \leq 1$$

$$0 \leq v_5 \leq 1$$

$$0 \leq v_6 \leq 1$$

$$0 \leq v_7 \leq 1$$

$$0 \leq v_8 \leq 1$$

$$0 \leq v_9 \leq 1$$

FVA results with
maximal OBJ

$$v_1 = 1$$

$$v_2 = 1$$

$$0 \leq v_3 \leq 1$$

$$0 \leq v_4 \leq 1$$

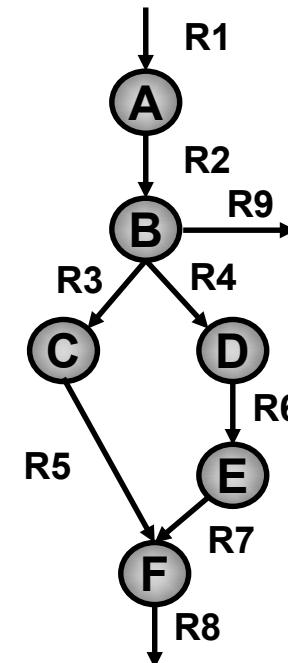
$$0 \leq v_5 \leq 1$$

$$0 \leq v_6 \leq 1$$

$$0 \leq v_7 \leq 1$$

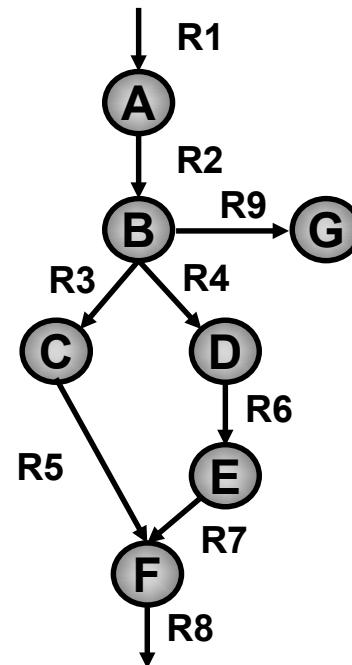
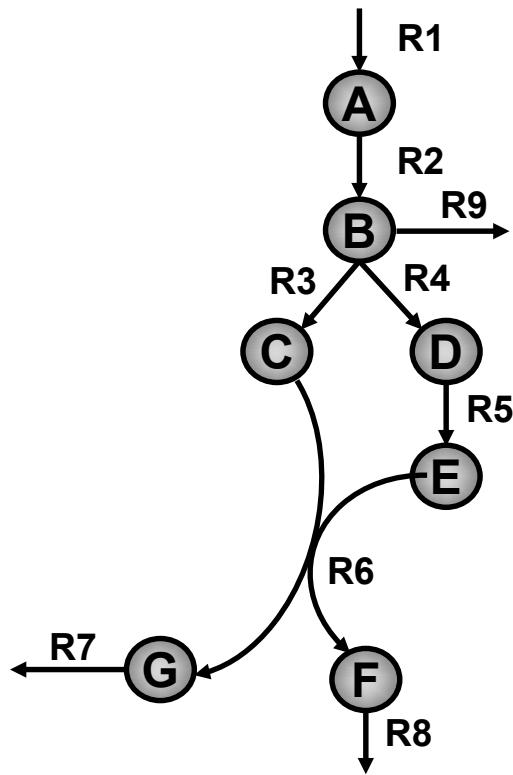
$$v_8 = 1$$

$$v_9 = 0$$



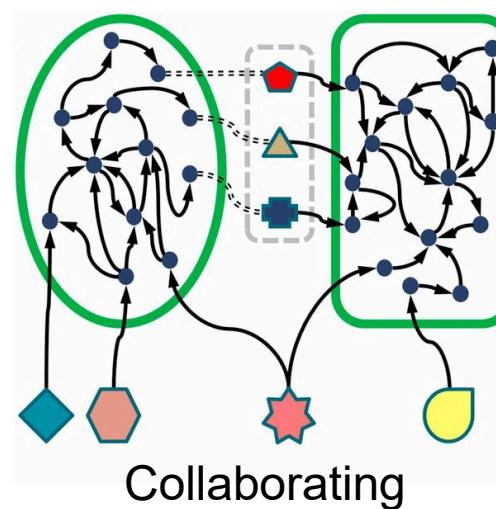
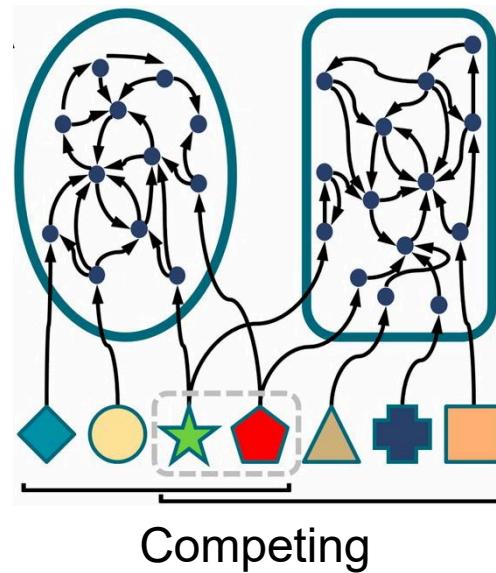
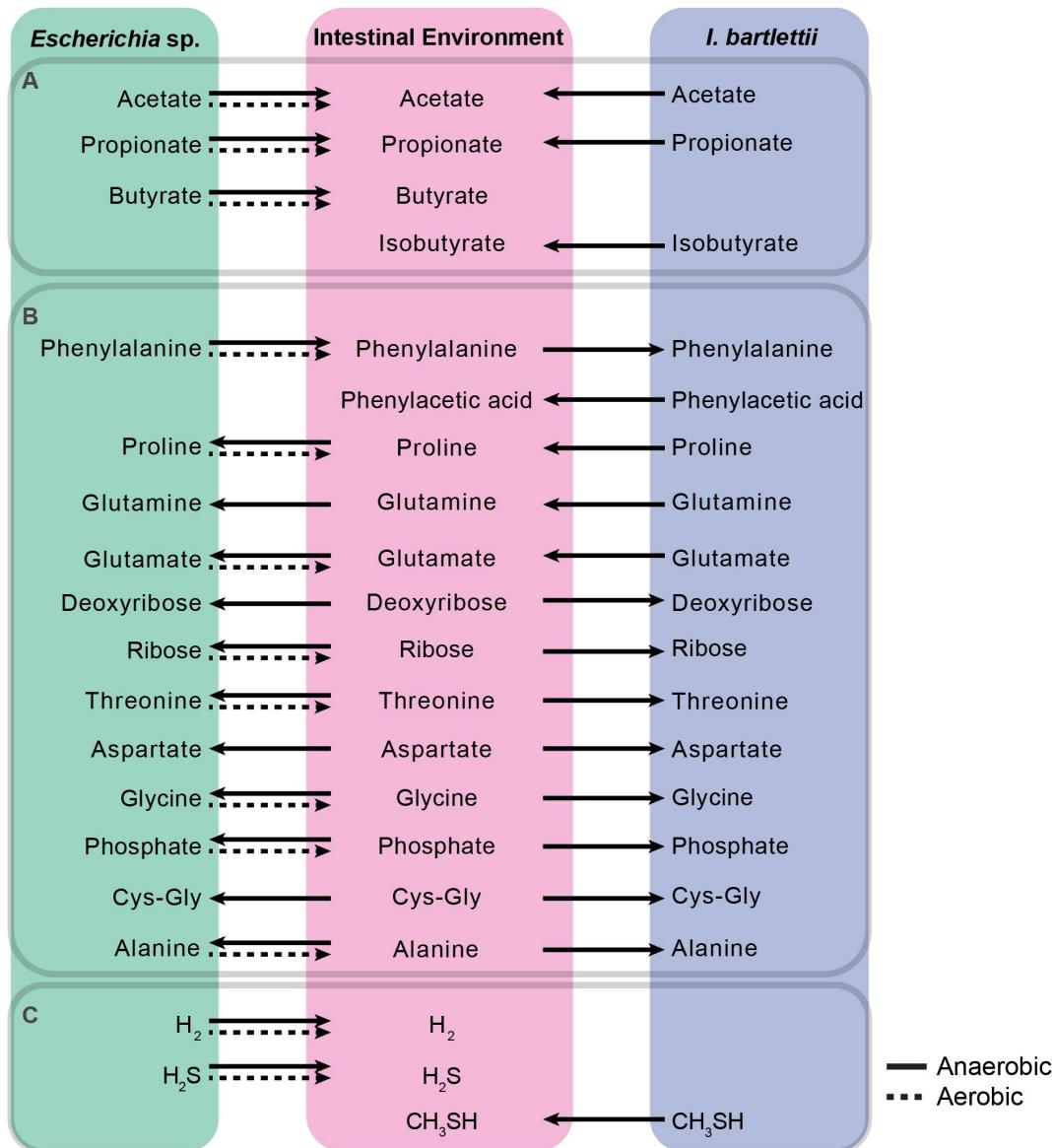
Flux variability analysis (FVA)

A few examples where FVA could be useful ($v_1 = 1$ & $Z = v_8$).

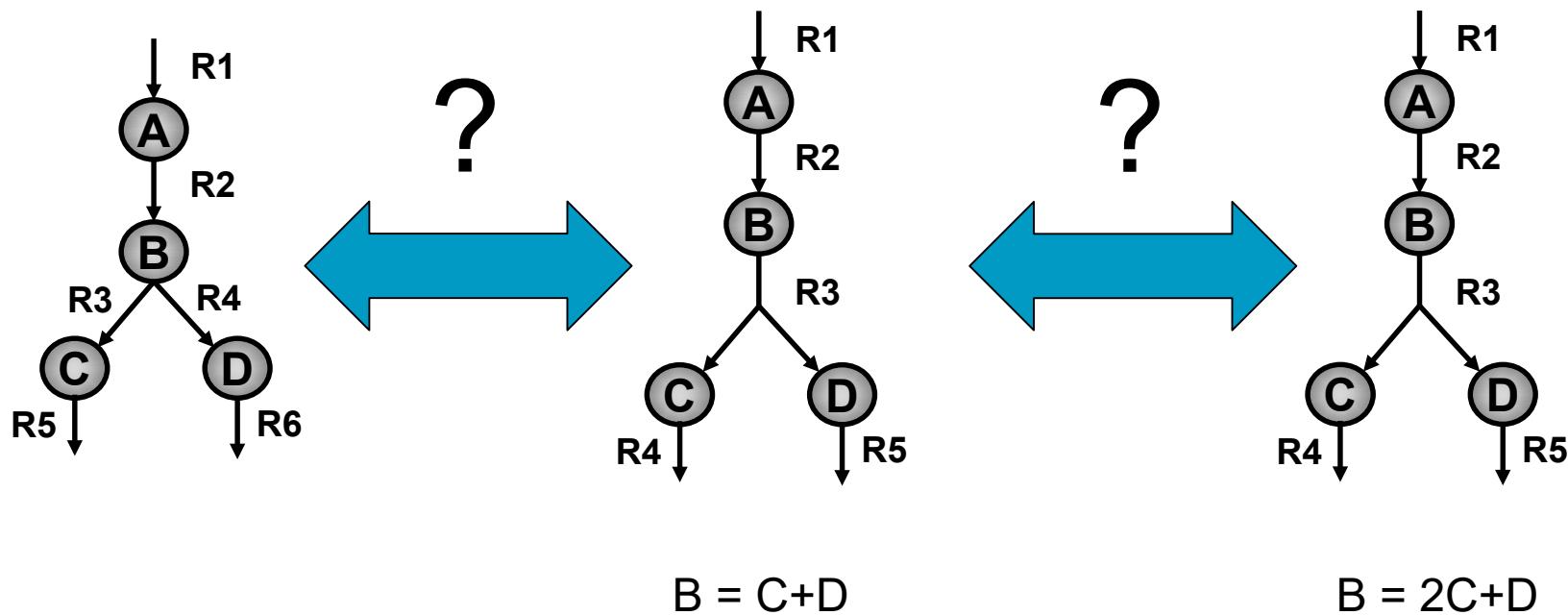


Essential reactions? Optional reactions? Blocked reactions?

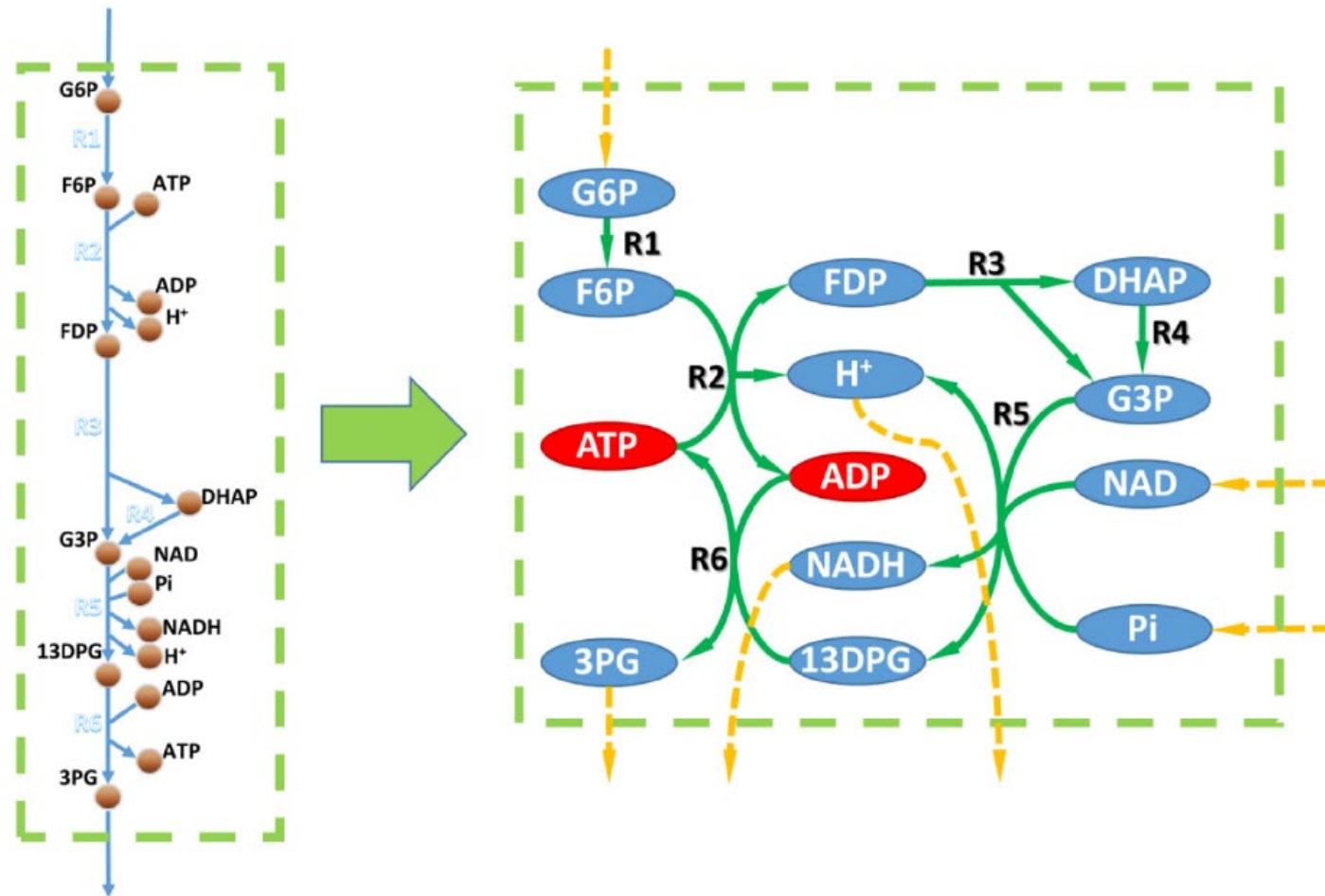
Flux variability analysis (FVA)



Difference between topology and stoichiometric matrix



GEMs vs metabolic networks



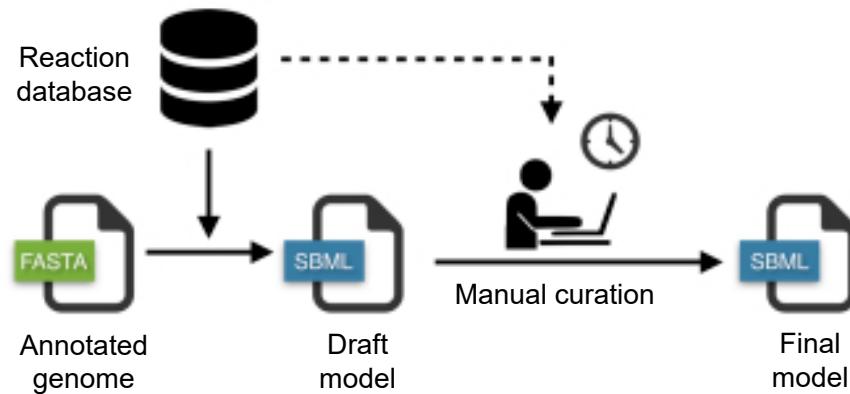
Metabolic networks with only topological info will miss ADP and ATP as required input and output, respectively

Result shows that model based flux balance analysis exhibited better prediction compared to network based analysis

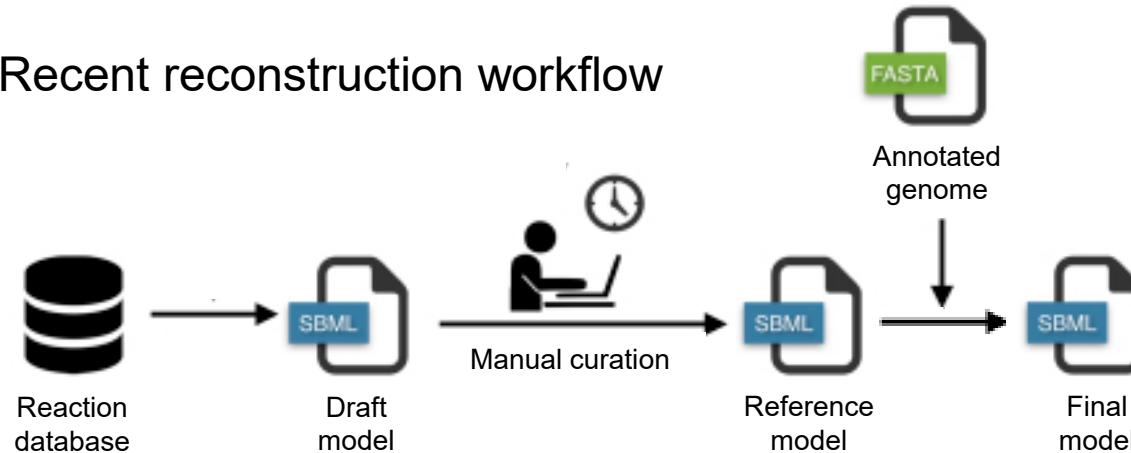
TABLE 2 | Spearman correlation between experimentally derived CERES scores and the *in silico* methods.

Methods	iIPC298		iNCIH1299	
	Correlations	P-values	Correlations	P-values
Essentiality scores	-0.194	0.014	-0.205	0.007
Betweenness centrality	-0.074	0.354	-0.138	0.070
Closeness centrality	-0.026	0.746	-0.065	0.396
Eccentricity centrality	-0.046	0.563	-0.011	0.883
Degree	0.070	0.376	-0.009	0.911

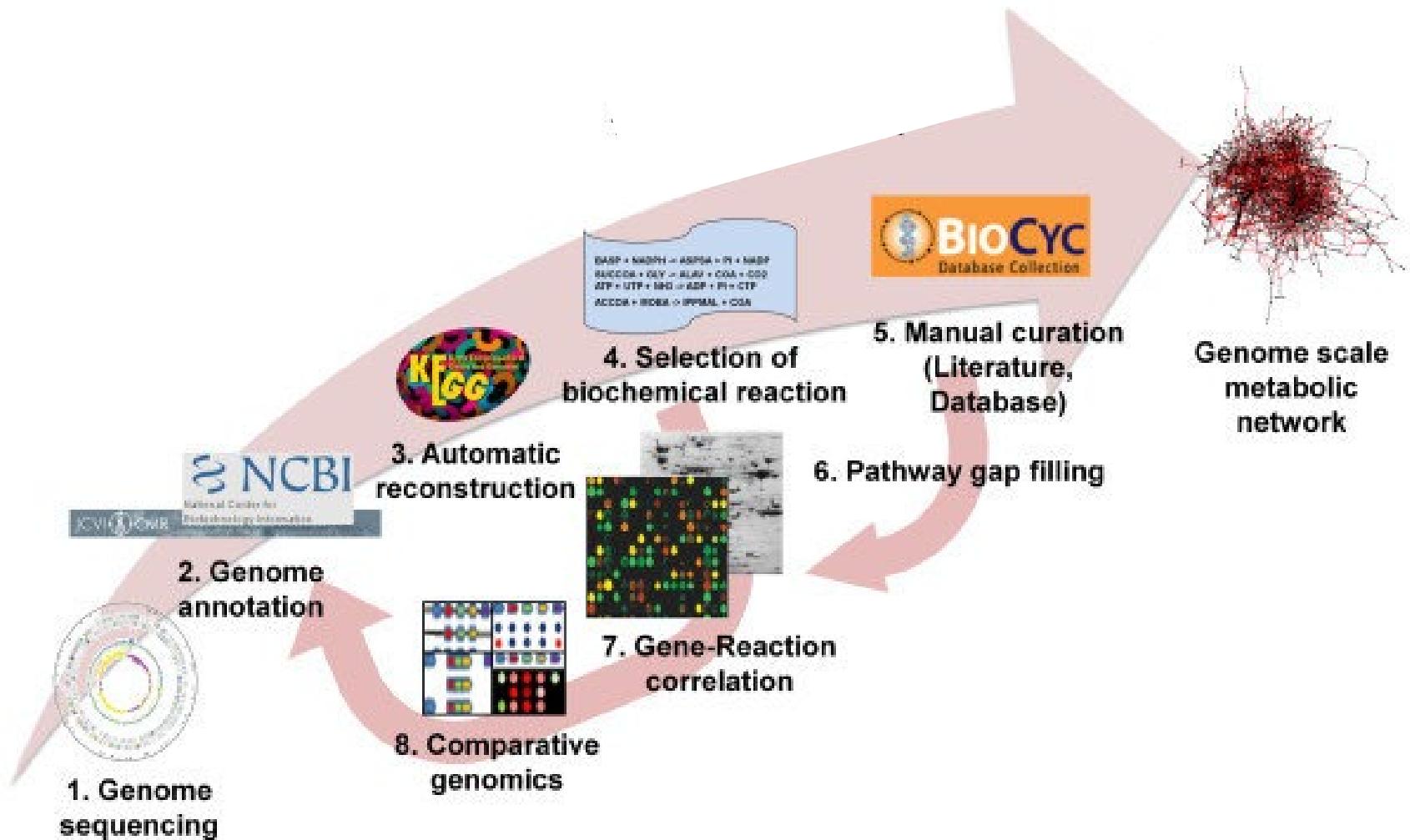
Classic reconstruction workflow



Recent reconstruction workflow

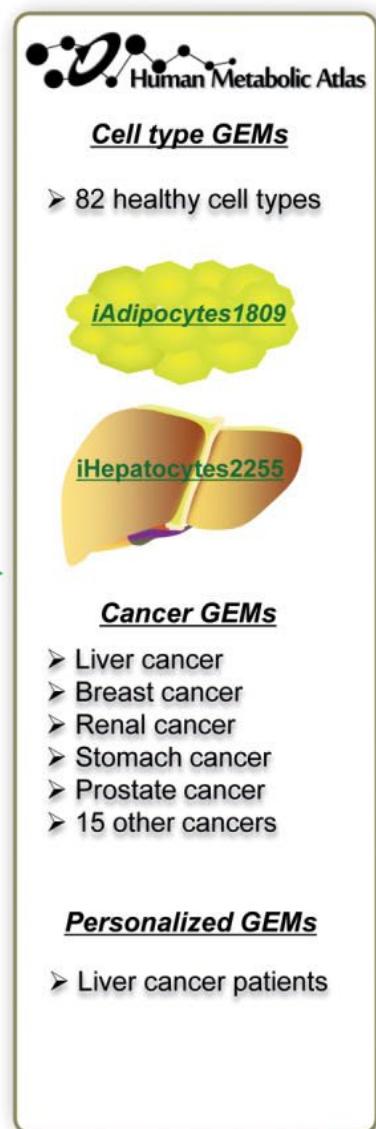
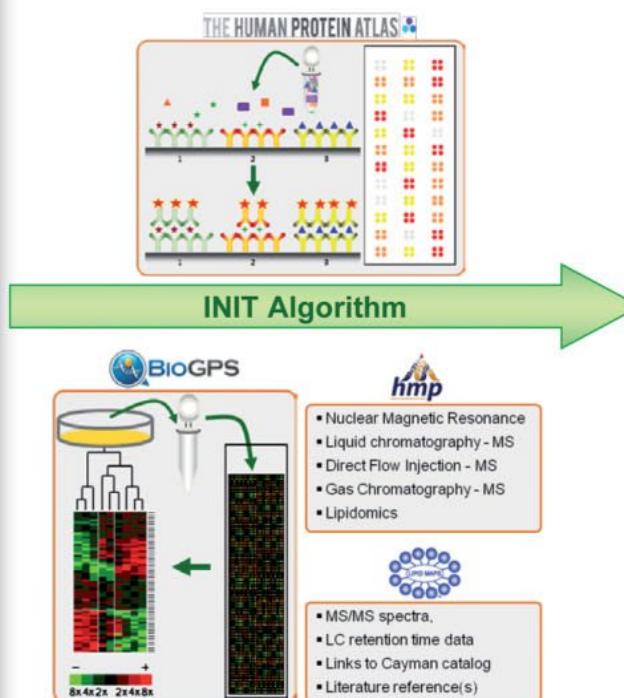
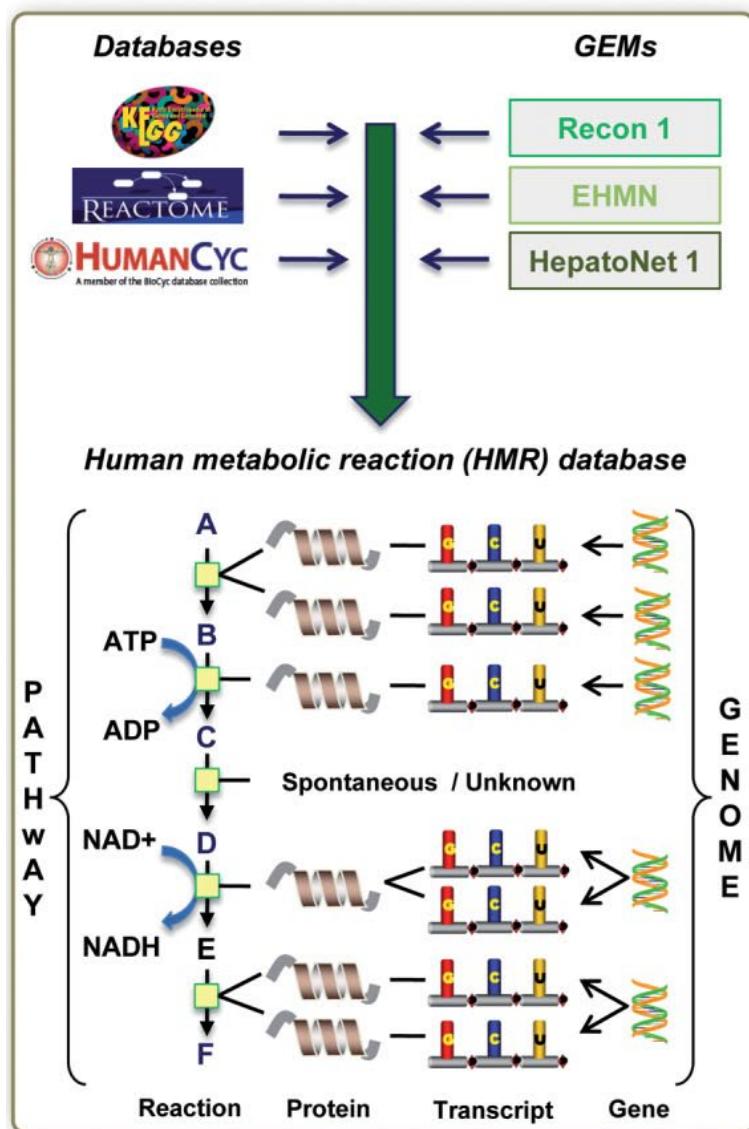


Classic reconstruction of GEMs



Context specific reconstruction of GEMs

SciLifeLab



INIT (Integrative Network Inference for Tissues)

R. Ågren, et al. *PLoS Comput Biol* 2012

- Uses proteomic, transcriptomic, and/or metabolomic data
- An optimization is performed to
 - maximize the number of high-confidence (high expression) reactions that are “on”
 - minimize the number of low-confidence (low-expression) reactions that are “on”
- **All reactions in the final model must be able to carry flux**
- **Metabolites are allowed to accumulate** during the optimization
 - An additional term in the algorithm maximizes the number of “present” metabolites that can be produced
 - Distinction of which metabolites should be “present” are based on literature or data (e.g., metabolomics)

$$\max \left(\sum_{i \in R} w_i y_i + \sum_{j \in M} x_j \right)$$

$$S\vec{v} = \vec{b}$$

$$|v_i| \leq 1000y_i$$

$$|v_i| + 1000(1 - y_i) \geq \varepsilon$$

$$v_i \geq 0, i \in \text{irreversible rxns}$$

$$b_j \leq 1000x_i$$

$$b_j + 1000(1 - x_i) \geq \varepsilon$$

$$b_j \geq 0$$

$$x_j = 1, j \in \text{present}$$

$$y_i, x_j \in \{0, 1\}$$

$$w_{i,j} = 5 \log \left(\frac{\text{Signal}_{i,j}}{\text{Average}_i} \right)$$

tINIT (Task-driven Integrative Network Inference for Tissues)

R. Ågren, et al. *Mol Syst Biol* 2014

- Identical formulation as INIT, with added steps
 - INIT does not necessarily yield simulation-ready models
- User defines a series of metabolic tasks that the model must perform
- Reactions that are required for these tasks are identified
 - A requirement that these reactions are active is included as an additional constraint in the optimization
- A follow-up evaluation of each task is performed
 - If a task fails, a gap-filling algorithm is used to enable task completion

Metabolic Tasks

Rephosphorylation of nucleoside triphosphates

Aerobic rephosphorylation of ATP from glucose
Aerobic rephosphorylation of GTP
Aerobic rephosphorylation of CTP
Aerobic rephosphorylation of UTP

De novo synthesis of nucleotides

ATP de novo synthesis
CTP de novo synthesis
GTP de novo synthesis
UTP de novo synthesis
dATP de novo synthesis
dCTP de novo synthesis
dGTP de novo synthesis
dTTP de novo synthesis

Uptake of essential amino acids

Histidine uptake
Isoleucine uptake
Leucine uptake
Lysine uptake
Methionine uptake
Phenylalanine uptake
Threonine uptake
Tryptophan uptake
Valine uptake

De novo synthesis of key intermediates

Glycerate 3-phosphate de novo synthesis
Mitochondrial acetyl-CoA de novo synthesis
Mitochondrial AKG de novo synthesis
Erythrose 4-phosphate de novo synthesis
Fructose 6-phosphate de novo synthesis

High throughput Reconstruction of gut microbiome GEMs

SciLifeLab

RESOURCE

nature
biotechnology

Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota

Stefanía Magnúsdóttir^{1,2}, Almut Heinken^{1,2}, Laura Kutt¹, Dmitry A Ravcheev¹, Eugen Bauer¹, Alberto Noronha¹, Kacy Greenhalgh¹, Christian Jäger¹, Joanna Baginska¹, Paul Wilmes¹, Ronan M T Fleming¹ & Ines Thiele¹

7542–7553 *Nucleic Acids Research*, 2018, Vol. 46, No. 15
doi: 10.1093/nar/gky537

Published online 21 June 2018

Fast automated reconstruction of genome-scale metabolic models for microbial species and communities

Daniel Machado, Sergej Andrejev, Melanie Tramontano and Kiran Raosaheb Patil*

European Molecular Biology Laboratory (EMBL), Meyerhofstrasse 1, 69117 Heidelberg, Germany

Received April 04, 2018; Revised May 17, 2018; Editorial Decision May 27, 2018; Accepted May 29, 2018

Essential genes are those genes of an organism that are thought to be critical for its survival. However, being essential is highly dependent on the circumstances in which an organism lives. Therefore, experimental assessment of gene-scale essentiality is a non-trivial effort.

GEM based essentiality analysis identified essential genes, reactions or metabolites that are critical for a specific function (e.g. growth) with computational prediction.

Essentiality analysis is one of the most accurate GEM based analysis, and it has been reported the accuracy of the prediction are around 90% in both *E. coli* and *S. cerevisiae* model.

Essentiality analysis

Recap of FBA

$$\text{maximize } Z = c^T v$$

$$\text{subject to } S^* v = 0$$

$$lb \leq v \leq ub$$

$$\text{maximize } Z = v_5$$

$$\text{subject to } v_1 - v_2 = 0$$

$$v_2 - v_3 - v_4 = 0$$

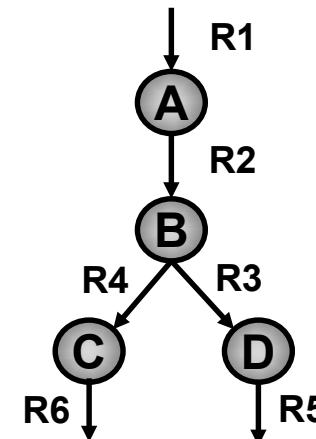
$$v_4 - v_6 = 0$$

$$v_3 - v_5 = 0$$

$$v_1 = 1$$

$$v_3 = 0$$

If $v_5^{KO} = 0$, R3 is essential.



How to mathematically check if R3 is essential for the objective function?

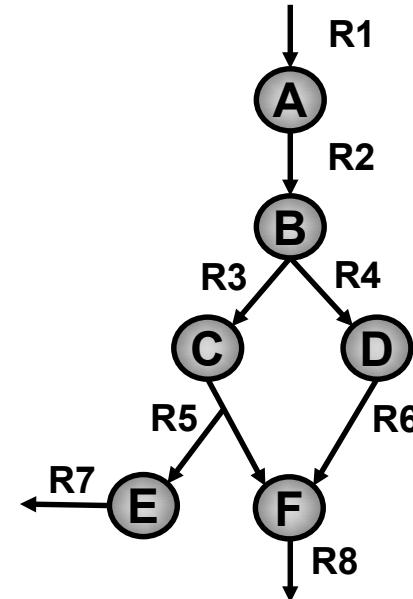
Essentiality analysis

If $v_1 = 1$ and objective function is $Z = v_8$

Which are the essential reactions?

R1, R2 (and R8 of course)

However, R4 and R6 are also important for the objective function. (Why?)



Note! R6: D = 2F

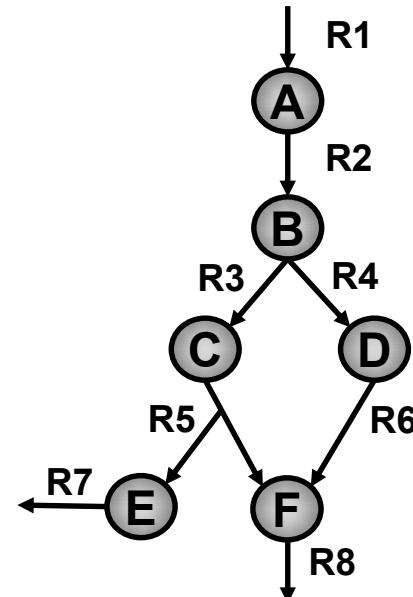
In order to also include reactions like R4 and R6, we could set a cutoff to modify the definition of essentiality

If $v_8^{KO} \leq \text{cutoff} * v_8^{\max}$ when R_i is knocked out, R_i is essential.

In this case, R4 and R6 are essential if cutoff = 0.6.

Gene essentiality analysis

	G1	G2	G3	G4	G5	G6
R1	1	0	0	0	0	0
R2	0	1	0	0	0	0
R3	0	0	1	1	0	0
R4	0	0	1	0	0	0
R5	0	0	0	0	1	1
R6	0	0	0	0	0	1
R7	0	0	0	0	0	0
R8	0	0	0	0	0	0



Which genes are essential if $v_1 = 1$
and objective function is $Z = v_8$

G6 is essential but G3 is not!

- R1: G1
- R2: G2
- R3: G3 or G4
- R4: G3
- R5: G5 and G6
- R6: G6
- R7: None
- R8: None

Gene essentiality analysis

SciLifeLab

GEM based gene essentiality analysis usually are more useful than reaction essentiality analysis, why? How we block a metabolic reaction in a cell in the lab?

It is useful for minimum genome design, drug target identification for CRISPR and siRNA, and many more.

Molecular Systems Biology 7; Article number 501; doi:10.1038/msb.2011.35
Citation: *Molecular Systems Biology* 7:501

© 2011 ELSIEVIER LTD. ALL RIGHTS RESERVED. 0874-5344/\$ - see front matter

WW



Contents lists available at ScienceDirect

Or

1

Sc

Ha

5

*

(

Fa

Re

Discovery of therapeutic agents for prostate cancer using genome-scale metabolic modeling and drug repositioning

Beste Turanli ^{a,b,c}, Cheng Zhang ^a, Woonghee Kim ^a, Rui Benfeitas ^a, Mathias Uhlen ^a, Kazim Yalcin Arga ^{b,*}, Adil Mardinoglu ^{a,d,e,*}

^a Science for Life Laboratory, KTH - Royal Institute of Technology, Stockholm SE-17121, Sweden

^b Department of Bioengineering, Marmara University, Istanbul, Turkey

^c Department of Bioengineering, Istanbul Medeniyet University, Istanbul, Turkey

^d Department of Biology and Biological Engineering, Chalmers University of Technology, Gothenburg SE-41296, Sweden

^e Centre for Host-Microbiome Interactions, Faculty of Dentistry, Oral & Craniofacial Sciences, King's College London, London, United Kingdom

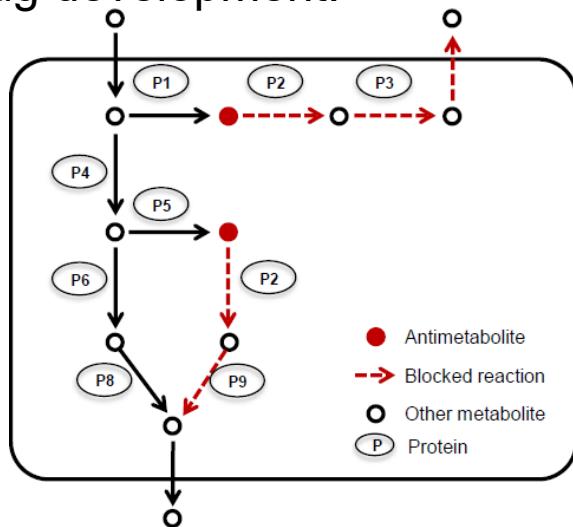
molecular
systems

EBioMedicine
Published by THE LANCET



Metabolite essentiality analysis

GEMs could also be used to predict the phenotype after metabolite removal, which could be useful in, for example, anti-metabolite type of cancer drug development.

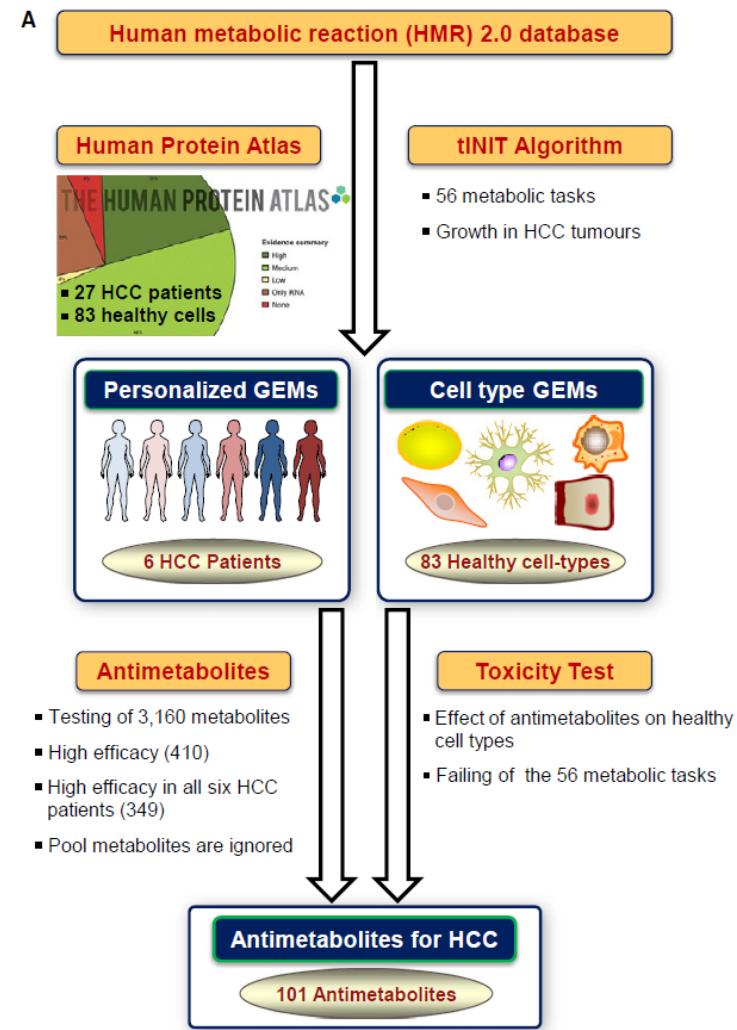


molecular
systems
biology

Article

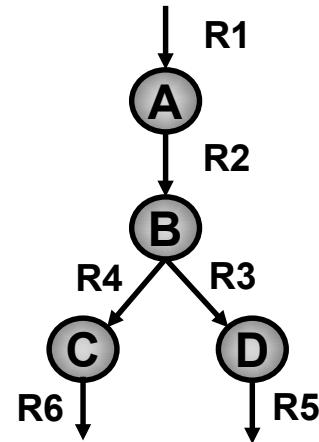
Identification of anticancer drugs for hepatocellular carcinoma through personalized genome-scale metabolic modeling

Rasmus Agren^{1,†}, Adil Mardinoglu^{1,†}, Anna Asplund², Caroline Kampf², Mathias Uhlen^{3,4} & Jens Nielsen^{1,3,*}



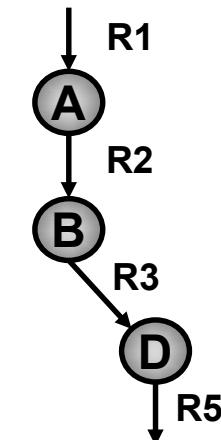
Metabolite essentiality analysis

$$\begin{array}{c} \text{R1} \quad \text{R2} \quad \text{R3} \quad \text{R4} \quad \text{R5} \quad \text{R6} \\ \text{A} \left(\begin{array}{cccccc} 1 & -1 & 0 & 0 & 0 & 0 \\ \text{B} & 0 & 1 & -1 & -1 & 0 & 0 \\ \text{C} & 0 & 0 & 0 & 1 & 0 & -1 \\ \text{D} & 0 & 0 & 1 & 0 & -1 & 0 \end{array} \right) \end{array}$$

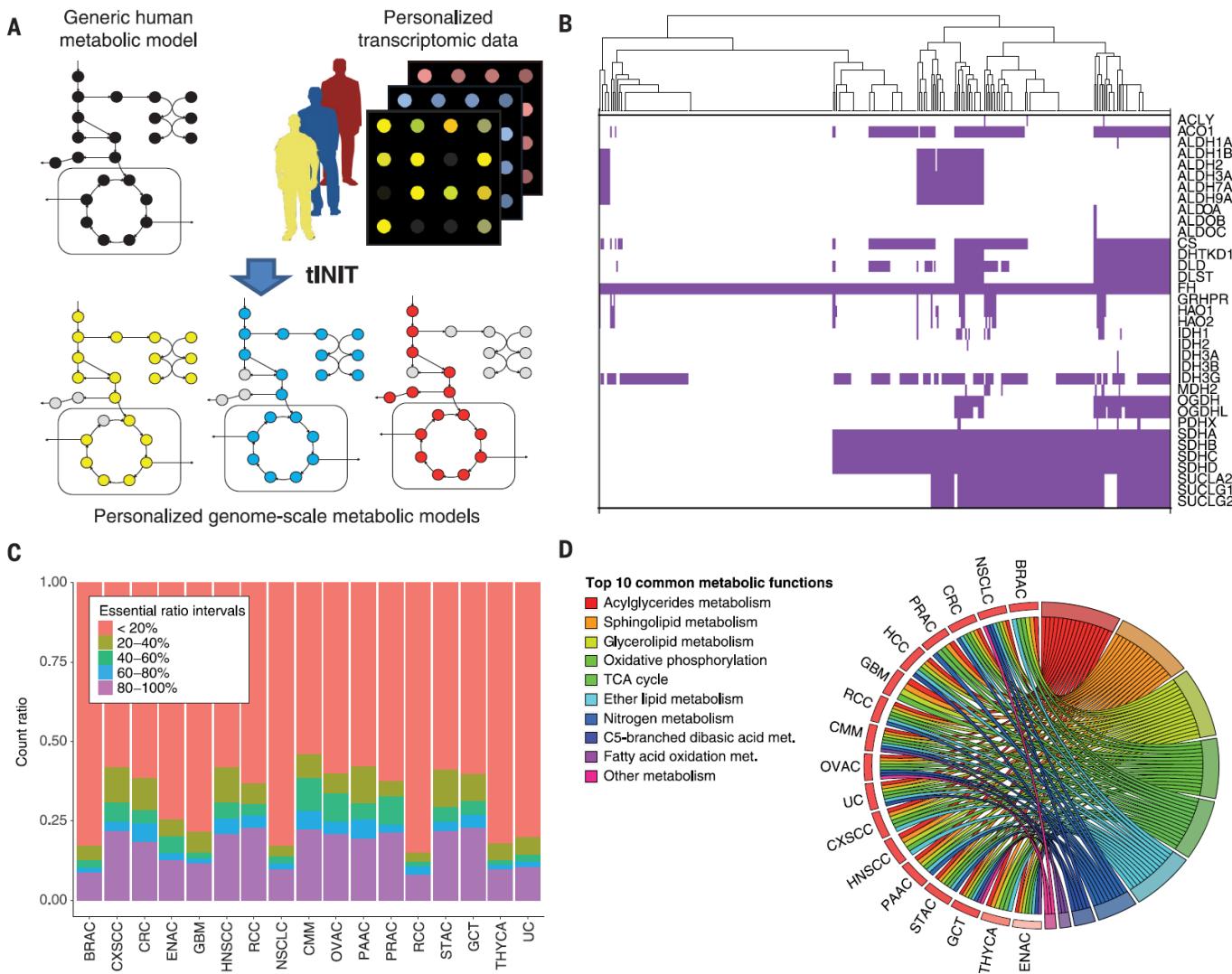


How to find all reactions blocked by removal of C through S matrix?

$$\begin{array}{c} \text{R1} \quad \text{R2} \quad \text{R3} \quad \text{R5} \\ \text{A} \left(\begin{array}{cccc} 1 & -1 & 0 & 0 \\ \text{B} & 0 & 1 & -1 & 0 \\ \text{D} & 0 & 0 & 1 & -1 \end{array} \right) \end{array}$$

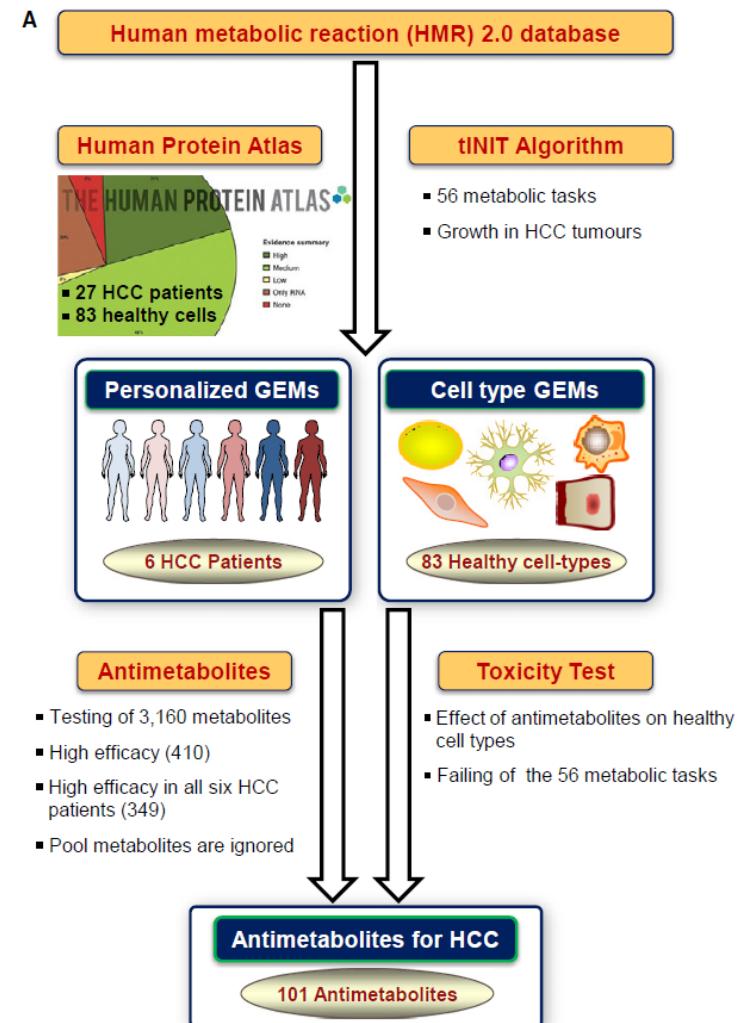
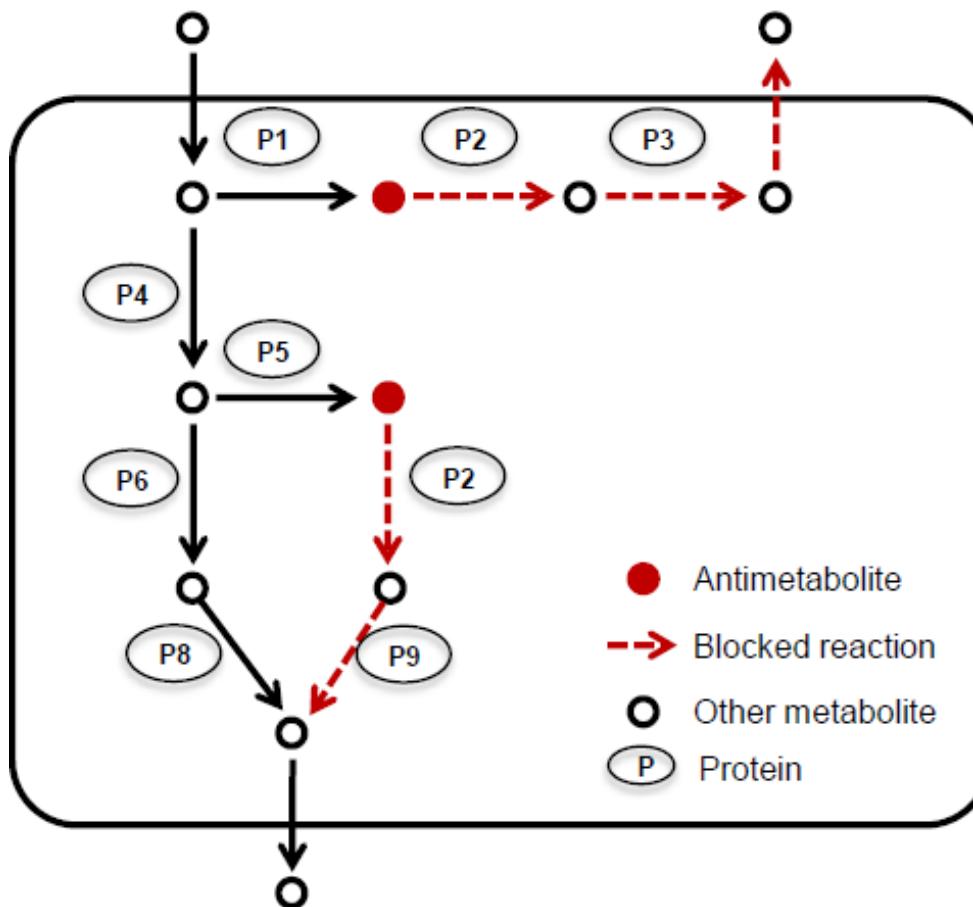


Essentiality analysis



Essentiality analysis

GEMs could also be used to predict the phenotype after metabolite removal, which could be potential anti-metabolite target for cancer treatment



In silico biological engineering is to use GEM to predict gene manipulation that could improve the production of desired bioproduct.

BIOTECHNOLOGY
and
BIOENGINEERING

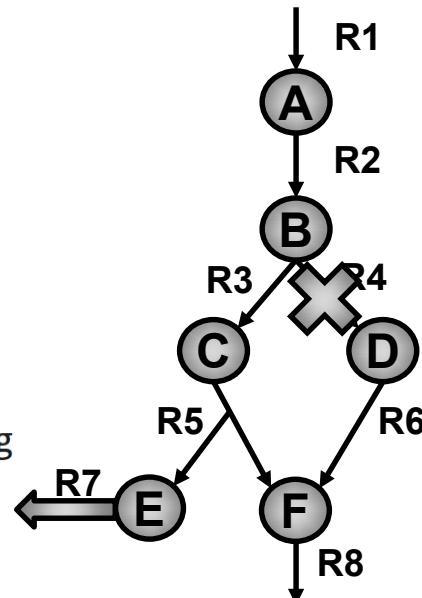
Regular Article

Optknock: A bilevel programming framework for identifying gene knockout strategies for microbial strain optimization

Anthony P. Burgard, Priti Pharkya, Costas D. Maranas 

First published: 07 October 2003 | <https://doi.org/10.1002/bit.10803> | Citations: 665

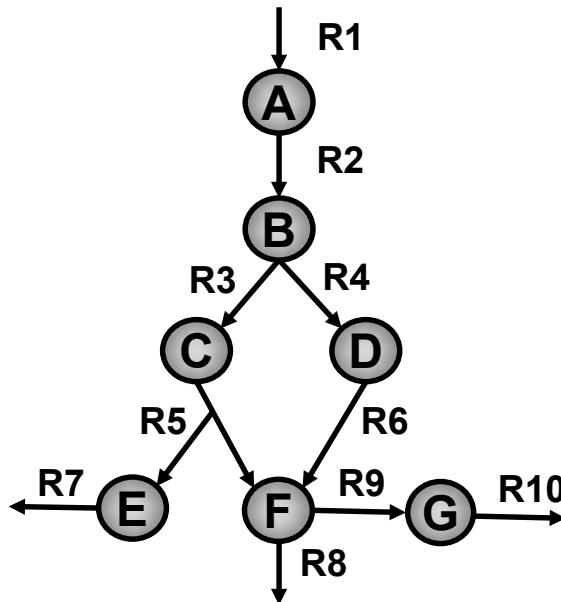
$$\begin{aligned}
 & \max \quad g'v \\
 \text{subject to} \quad & \sum_{l=1}^L y_l \leq C, \\
 & y_l \in \{0, 1\}, \quad l = 1, \dots, L, \\
 & Sv = 0, \\
 & (1 - y'G_j)a_j \leq v_j \leq (1 - y'G_j)b_j, \quad j = 1, \dots, n, \\
 & f'v = \sum_{j=1}^n v_j b_j - \mu_j a_j, \\
 & f_j - \sum_{i=1}^m \lambda_i S_{ij} - v_j + \mu_j - \xi_j = 0, \quad j = 1, \dots, n, \\
 & -Dy'G_j \leq \xi_j \leq Dy'G_j, \quad j = 1, \dots, n, \\
 & \mu, v \geq 0,
 \end{aligned}$$



- R1: G1
- R2: G2
- R3: G3 or G4
- R4: G3
- R5: G5 and G6
- R6: G6
- R7: None
- R8: None

maximize	bioengineering objective
(through gene knockouts)	
subject to	maximize cellular objective
(over fluxes)	
subject to	<ul style="list-style-type: none"> ◦ fixed substrate uptake ◦ network stoichiometry ◦ blocked reactions identified by outer problem
number of knockouts \leq limit	

We could also use GEMs to predict gene ‘knock-in’ that could lead to desired production.



Downloaded from genome.cshlp.org on November 27, 2019 - Published by Cold Spring Harbor Laboratory Press

Resource

OptStrain: A computational framework for redesign of microbial production systems

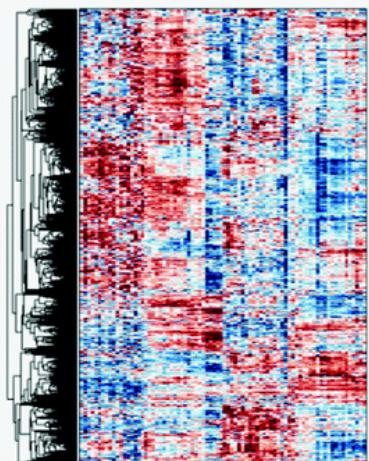
Priti Pharkya, Anthony P. Burgard, and Costas D. Maranas¹

Department of Chemical Engineering, The Pennsylvania State University, University Park, Pennsylvania 16802, USA

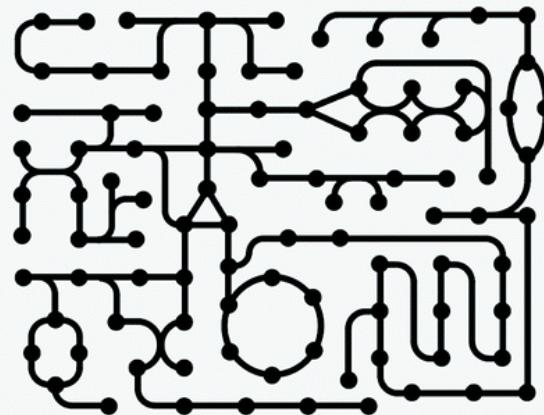
Integration of transcriptomic data

Integrating transcriptomic data into genome scale model has been a very appealing approach for improving the flux prediction and find key subnetwork within GEMs.

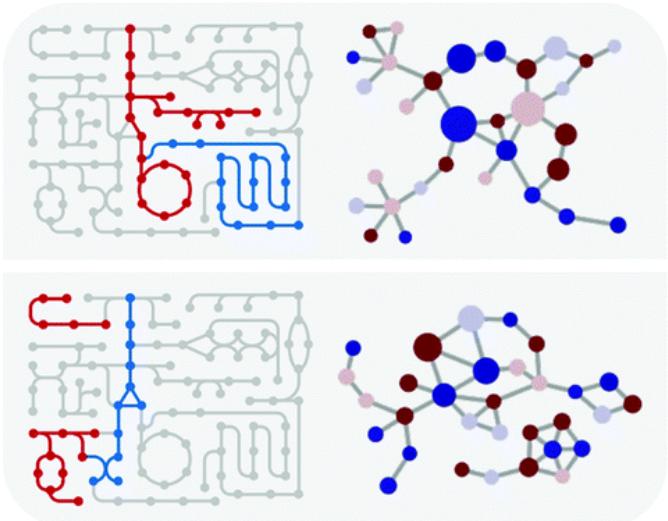
Transcriptomic



GEM

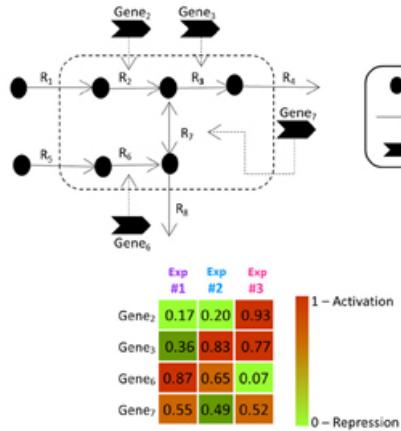


Flux distribution and subnetwork



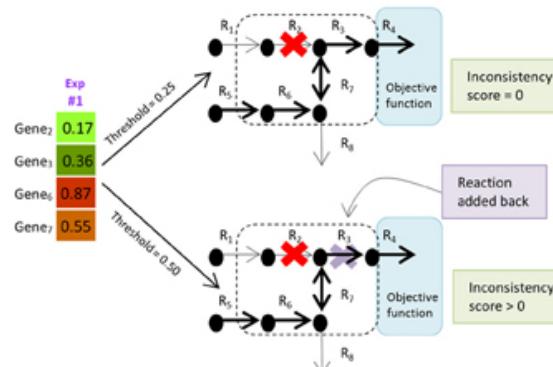
Integration of transcriptomic data

A Toy Network & Sample Expression Data



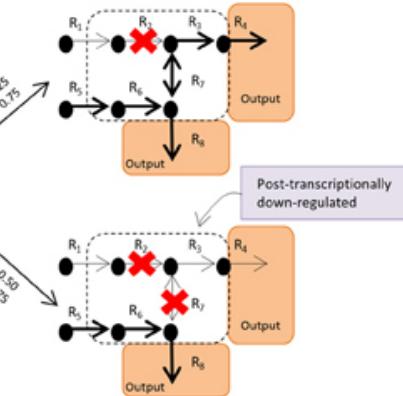
B GIMME

- +Requires only one set of expression data
- Requires thresholding



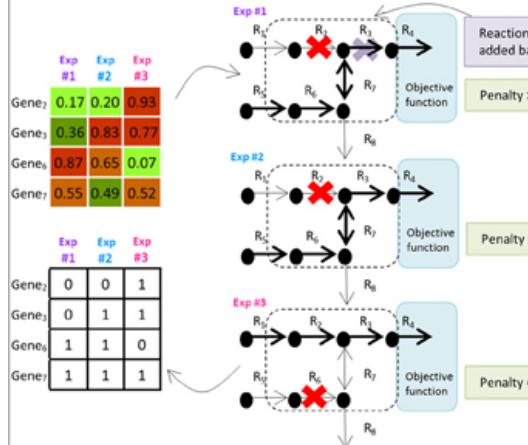
C iMAT

- +Does not require *a priori* knowledge of an objective function
- Requires discretization of data (lowly, moderately, highly expressed genes)



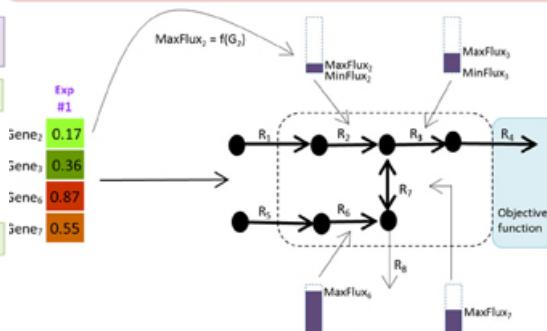
D MADE

- +Does not require thresholding of expression data
- Requires multiple data sets for differential expression



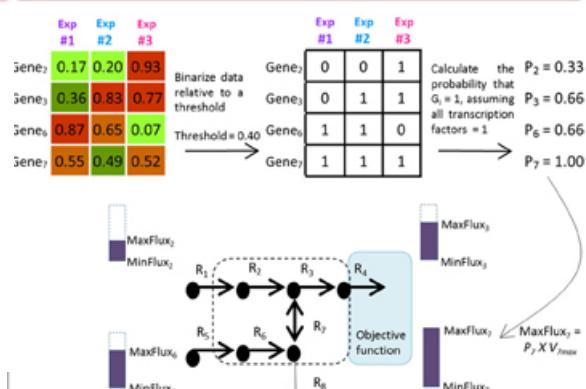
E E-Flux

- +Does not require binarization of expression data
- Requires a function to convert expression levels into an upper bound on fluxes

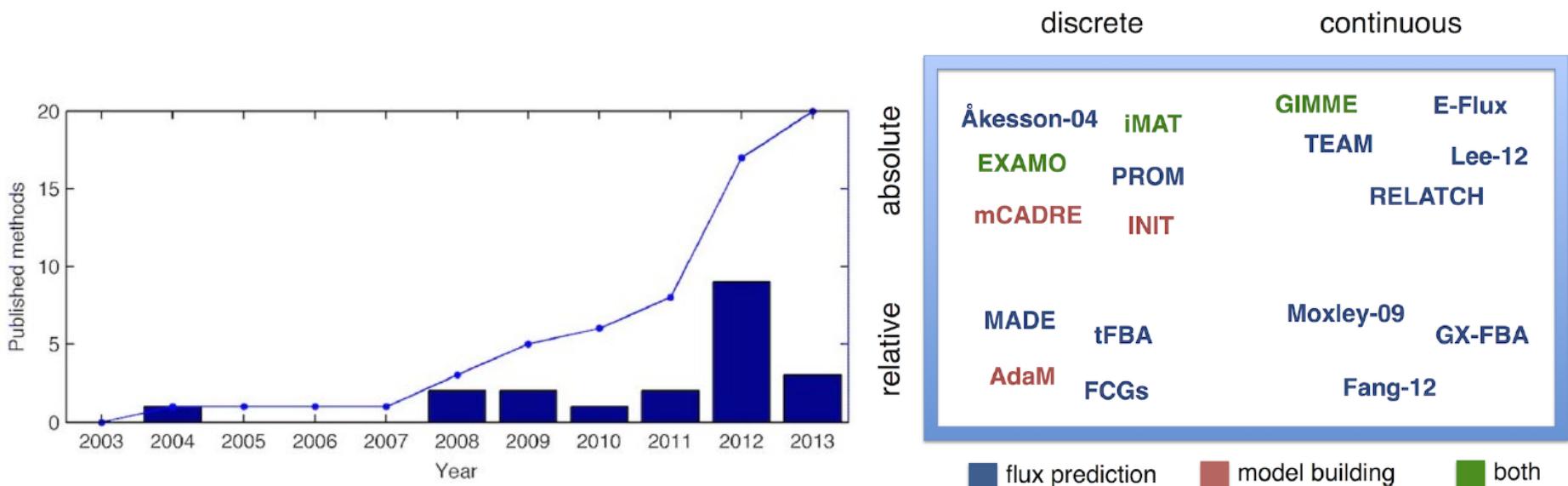


F PROM

- +Incorporates regulatory interactions without a mechanistic model
- Requires a large dataset for calculating transcription factor and target gene interactions



Integration of transcriptomic data



Many methods have been published for integration of transcriptomic data into GEMs.

OPEN ACCESS Freely available online

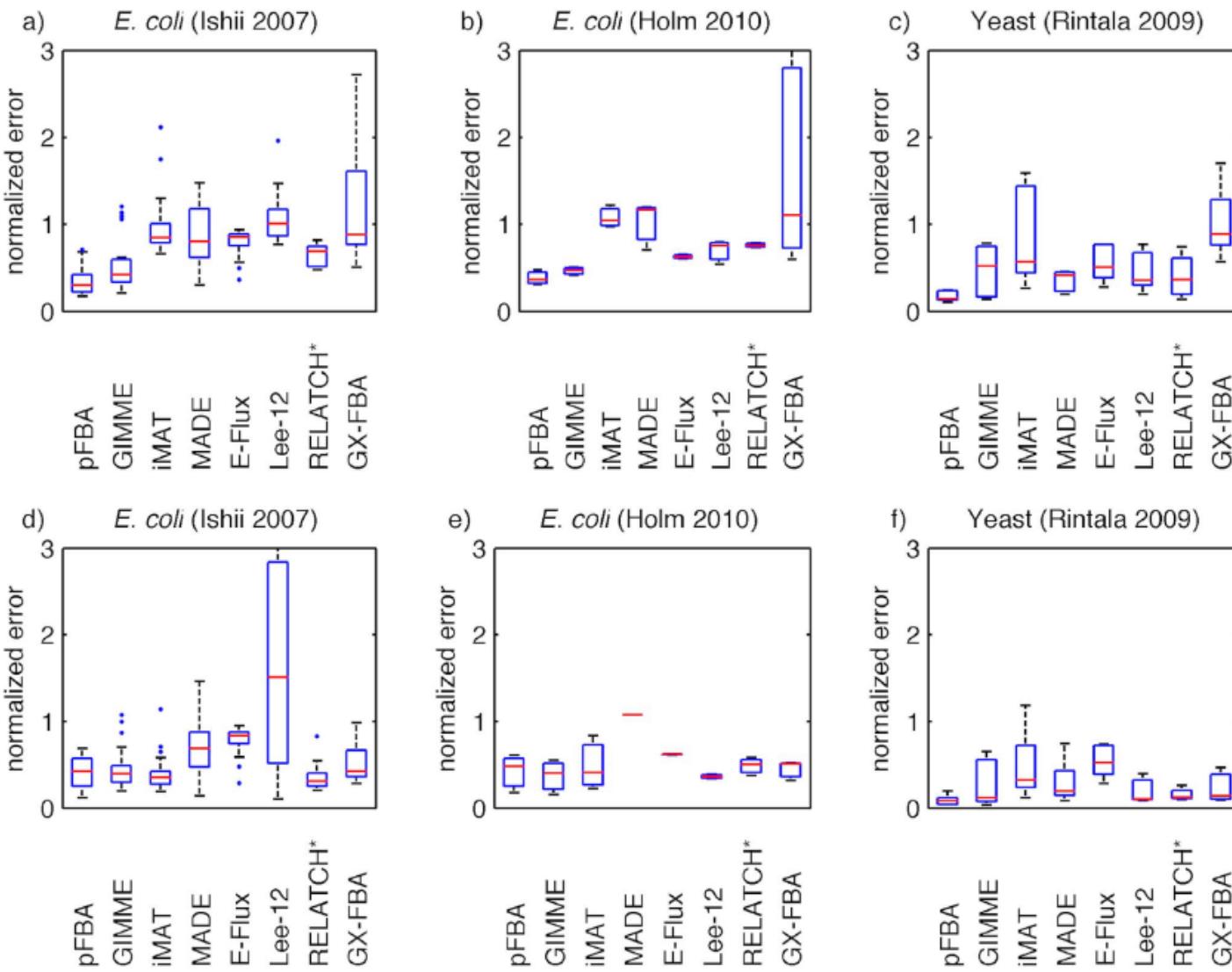
PLOS COMPUTATIONAL BIOLOGY

Systematic Evaluation of Methods for Integration of Transcriptomic Data into Constraint-Based Models of Metabolism

Daniel Machado, Markus Herrgård*

The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Horsholm, Denmark

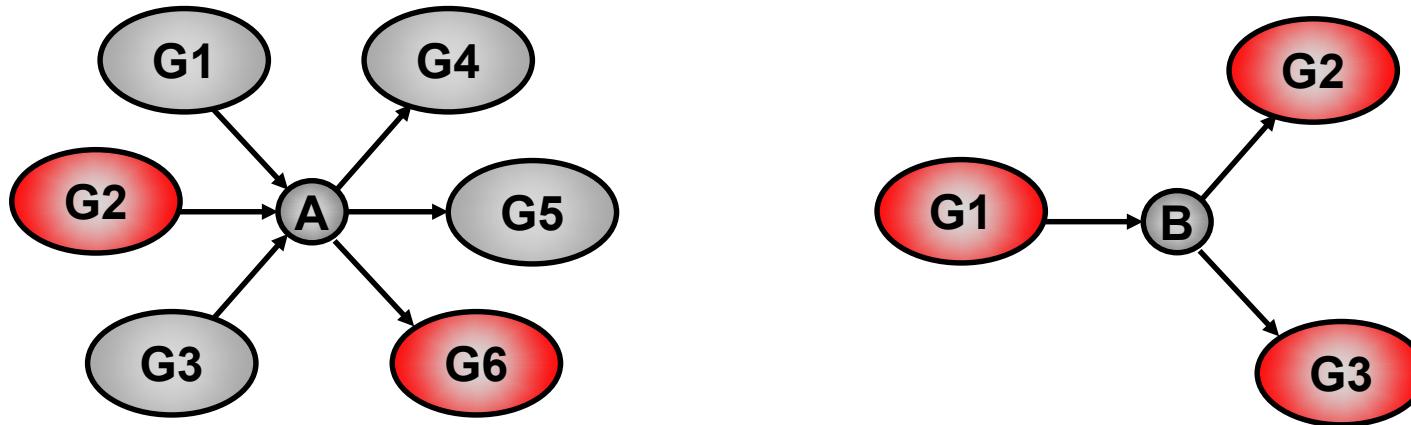
Integration of transcriptomic data



No major advantage has been found compared to pFBA in microorganism models. Why?

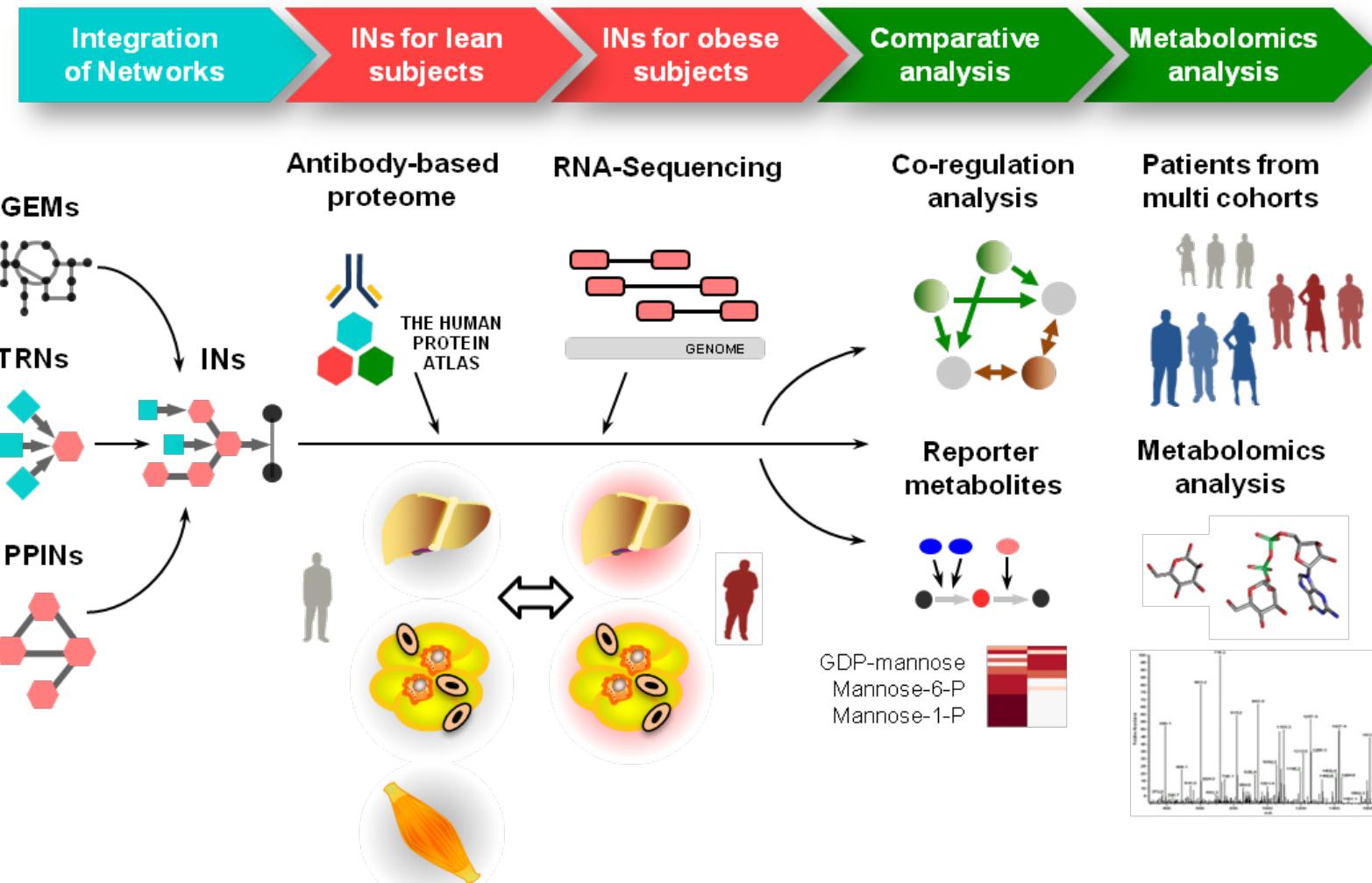
Are those method developed for integration of transcriptomic data valuable? Why do we still need them?

They can be good compliment for when there is no reliable objective function or input data, which is especially useful for tissue/cell specific GEMs of Mammalia.



In order to properly capture the connection between expression and network topology, we must determine whether the aggregated z score of a subnetwork is higher than expected relative to a random set of genes (drawn from the same expression data but independently of the network). We randomly sample gene sets of size k using a Monte Carlo approach, compute their aggregated scores, then use these to derive estimates for the mean score and standard deviation.

Reporter metabolites



Lee et. al (2016), Cell Metabolism, 24, 172-184

Association of mannose levels and incident T2D in 4,000 well phenotyped subjects

The EPIC-Norfolk STUDY

Model	N cases	N non-cases	HR	95% CI	p-value
Model 1: Study-specific covariates	673	830	1.68	1.54 - 1.82	1.94E-34
Model 1 + HbA1c	673	830	1.26	1.13 - 1.42	5.21E-05
Model 1 + HbA1c + random glucose	673	830	1.23	1.06 - 1.42	0.0065
Model 1 + HbA1c + random glucose + 1,5-anhydroglucitol	673	830	1.23	1.06 - 1.43	0.0058

METSIM STUDY

Model	N cases	N non-cases	HR	95% CI	p-value
Model 1: Study-specific covariates	96	2132	1.80	1.43 - 2.27	5.30E-07
Model 1 + HbA1c	96	2128	1.66	1.30 - 2.12	5.02E-05
Model 1 + HbA1c + FPG	96	2128	1.48	1.15 - 1.89	0.0022
Model 1 + HbA1c + FPG + 1,5-anhydroglucitol	96	2128	1.45	1.13 - 1.86	0.0036