

Genome-scale metabolic modeling for data integration and simulation

Rui Benfeitas

NBIS - National Bioinformatics Infrastructure Sweden
Science for Life Laboratory, Stockholm

rui.benfeitas@scilifelab.se

metabolic
ATLAS



NBIS

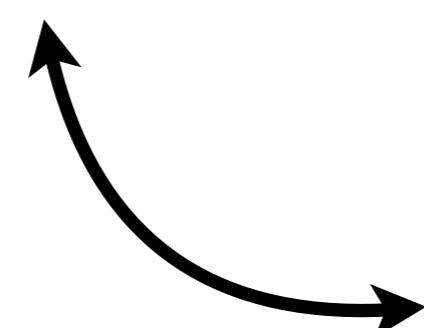
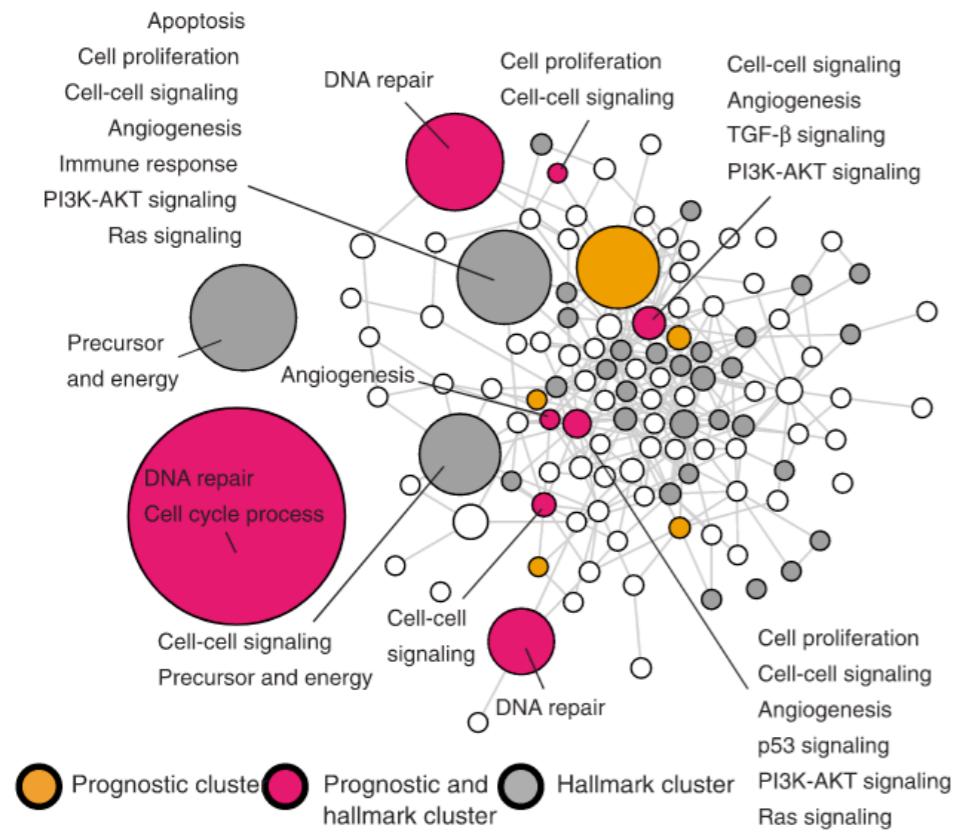


SciLifeLab

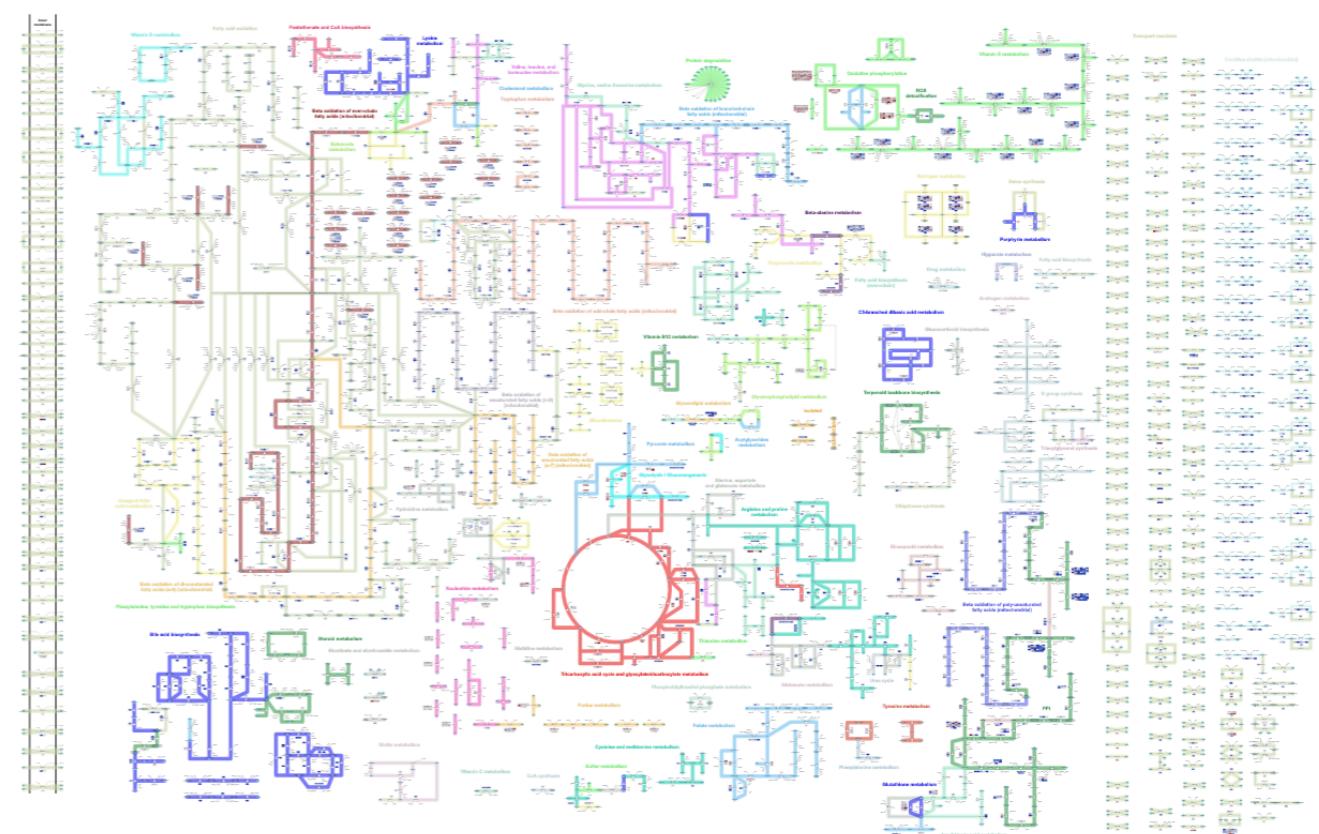


Frameworks for biological network analysis in health and disease

Introduction to application of graph analysis in disease



Genome-scale metabolic modeling for data integration and simulation



Uhlen 2017

<https://metabolicatlas.org/>

Companion resources

Pre-course information and installation instructions

Data pre-processing notebook: [link](#)

Introduction to metabolic modeling with cobrapy: [link](#)

Lab: introduction to metabolic modeling with cobrapy

To run the following notebook, retrieve it from [here](#) and include it in your local folder as `/workshop_omics_integration/session_gems/lab.ipynb`. Please use the jupyter container to run it. If you would like to further explore these techniques, we strongly suggest to [have a look here](#).

CZI Omics Integration

Rui Benfeitas, Scilifelab, NBIS National Bioinformatics Infrastructure Sweden

rui.benfeitas@scilifelab.se

Abstract

In this notebook we will perform some of the basic operations in working with a genome-scale metabolic model (GEM). The vast majority of software that has been developed surrounding GEMs has been done in MATLAB, likely because this form of modeling has origins in engineering (specifically chemical engineering). Although well-suited for metabolic modeling, MATLAB is not open-source and therefore limits the accessibility of such software. Fortunately, the modeling community has implemented the MATLAB COnstraint-Based Reconstruction and Analysis (**COBRA**) Toolbox in Python, as **COBRApy**.

COBRApy is still relatively new and therefore lacks some of the functionality of its MATLAB counterparts, but the core utilities are available and quickly expanding. Here, we will demonstrate some of the basic functions and classes of the **COBRApy** package, which should also familiarize the user with the fundamentals of GEM structure and simulation.

Most of the commands and material covered in this tutorial can be found in the [COBRApy Documentation](#), so we encourage you to reference the documentation if you encounter errors, warnings, or need further detail about something.

Contents

- 1 Global configuration object
- 2 Importing and inspecting models
- 3 Adding reactions to the model
- 4 Flux balance analysis (FBA)
- 5 Perform an in silico knockout

In [1]:

```
import cobra
import cobra.test
import os
```

Global configuration object

Before jumping right into things, it is always nice to see what sort of default settings are in place. **COBRApy** has organized such defaults into a **global configuration object**, which can be

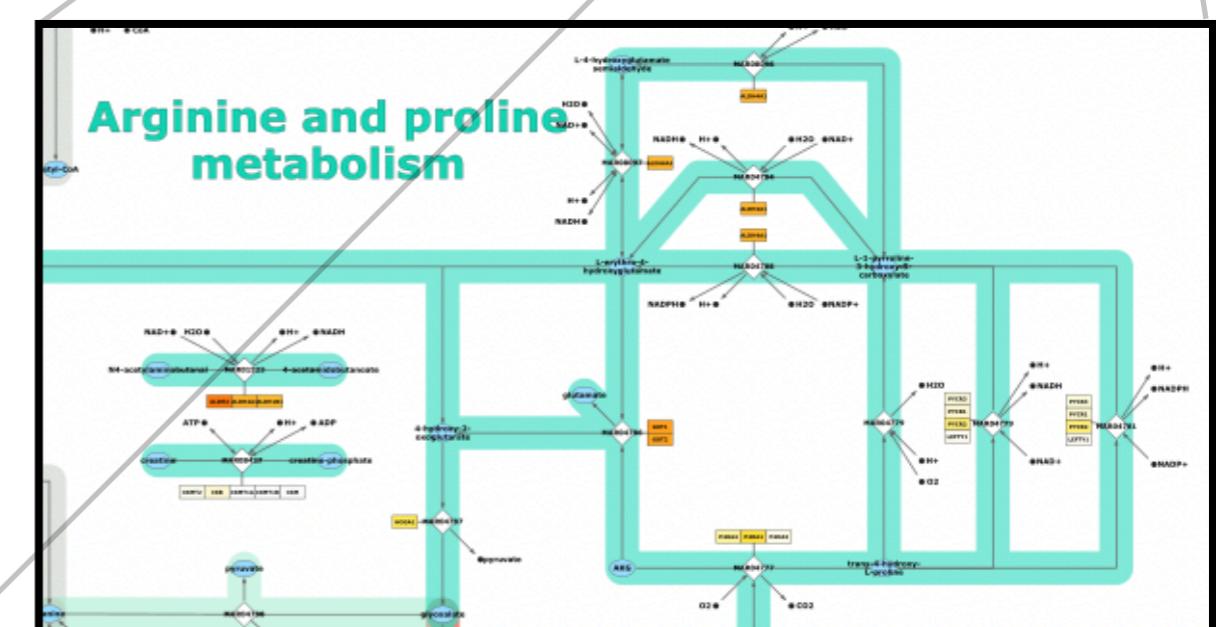
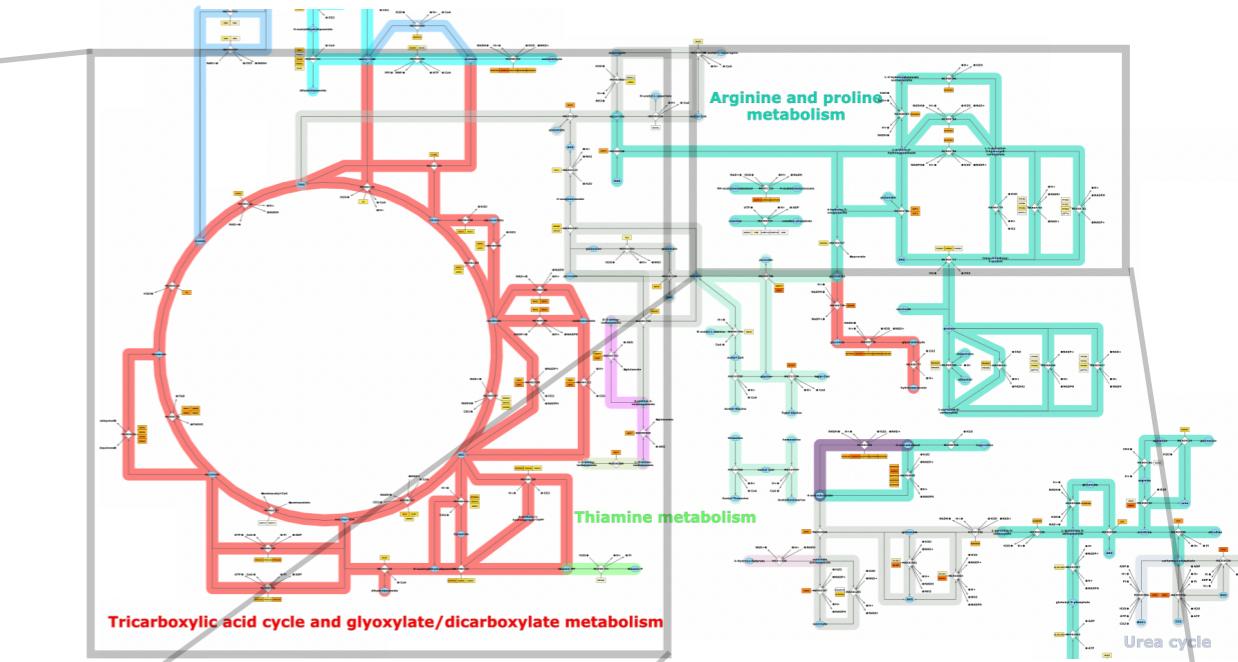
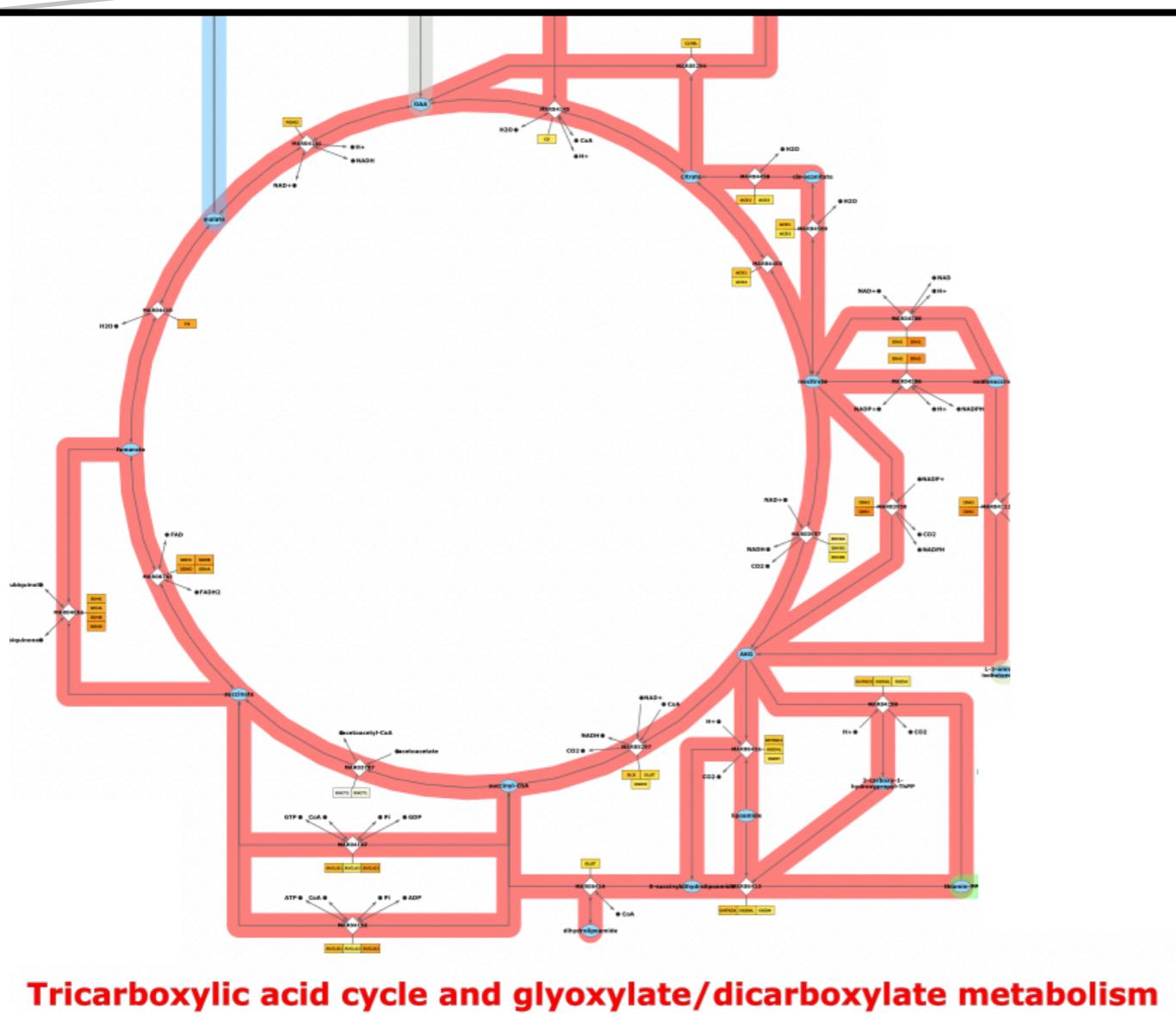
Overview

- 1. Challenge in systematically characterising metabolic disruption**
 2. Introduction to metabolic modelling and FBA
 3. Downstream analysis and combination with topology analysis

Original sources of images provided as reference and hyperlinks where applicable.

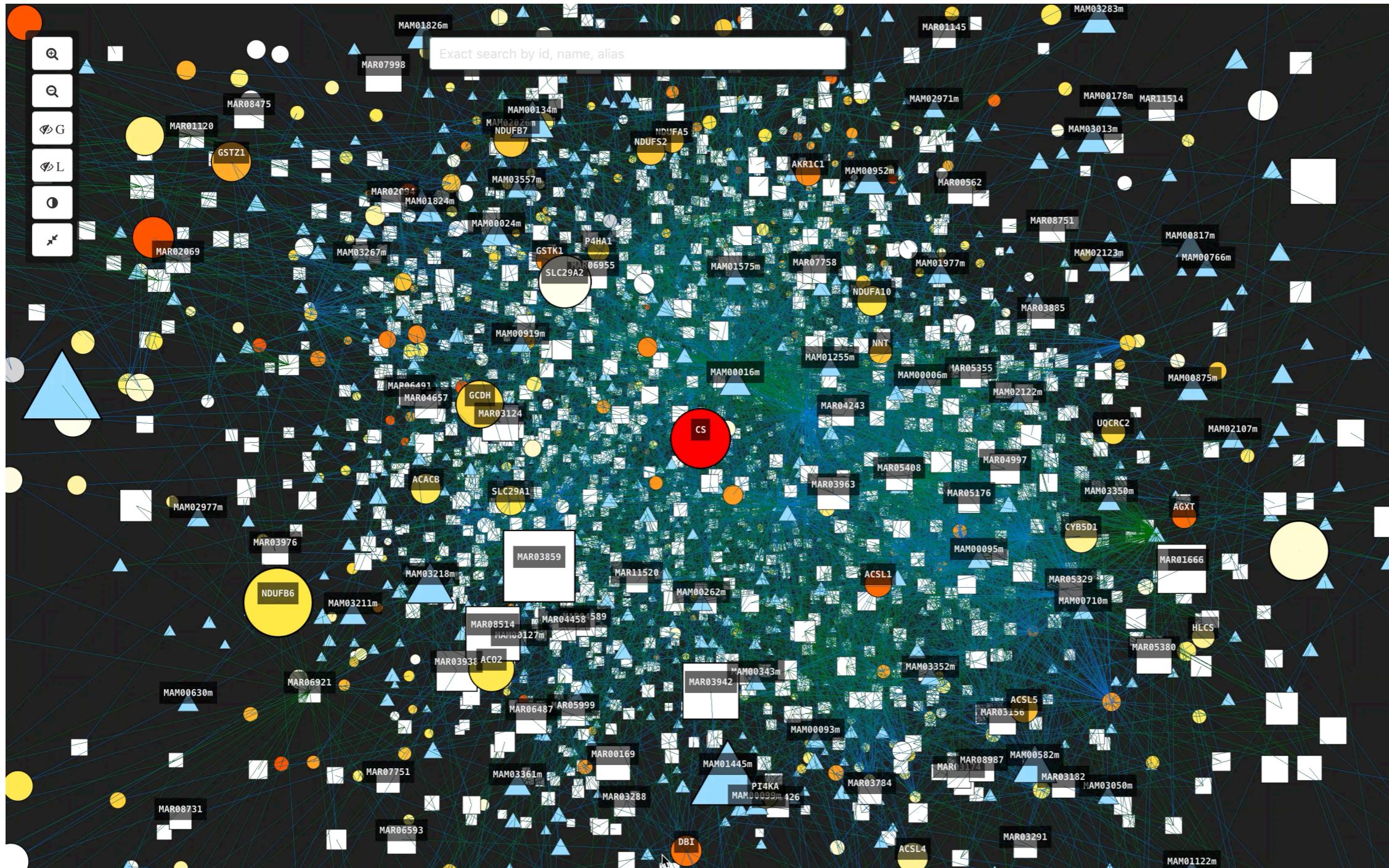
Characterizing metabolic disruption in disease is hindered by biological complexity

Human mitochondrial reactions



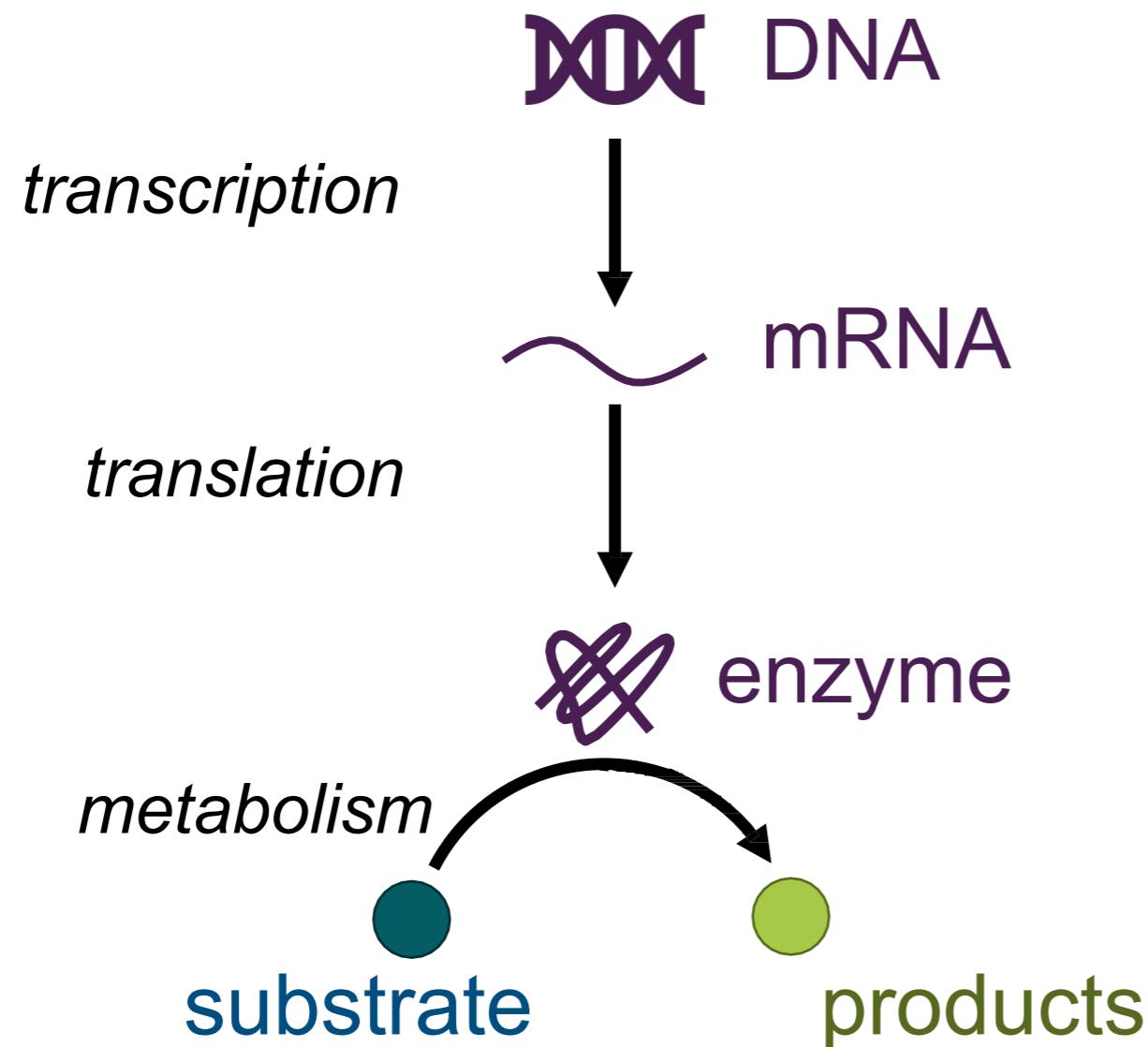
Characterizing metabolic disruption in disease is hindered by biological complexity

3D map of mitochondrial reactions in Human



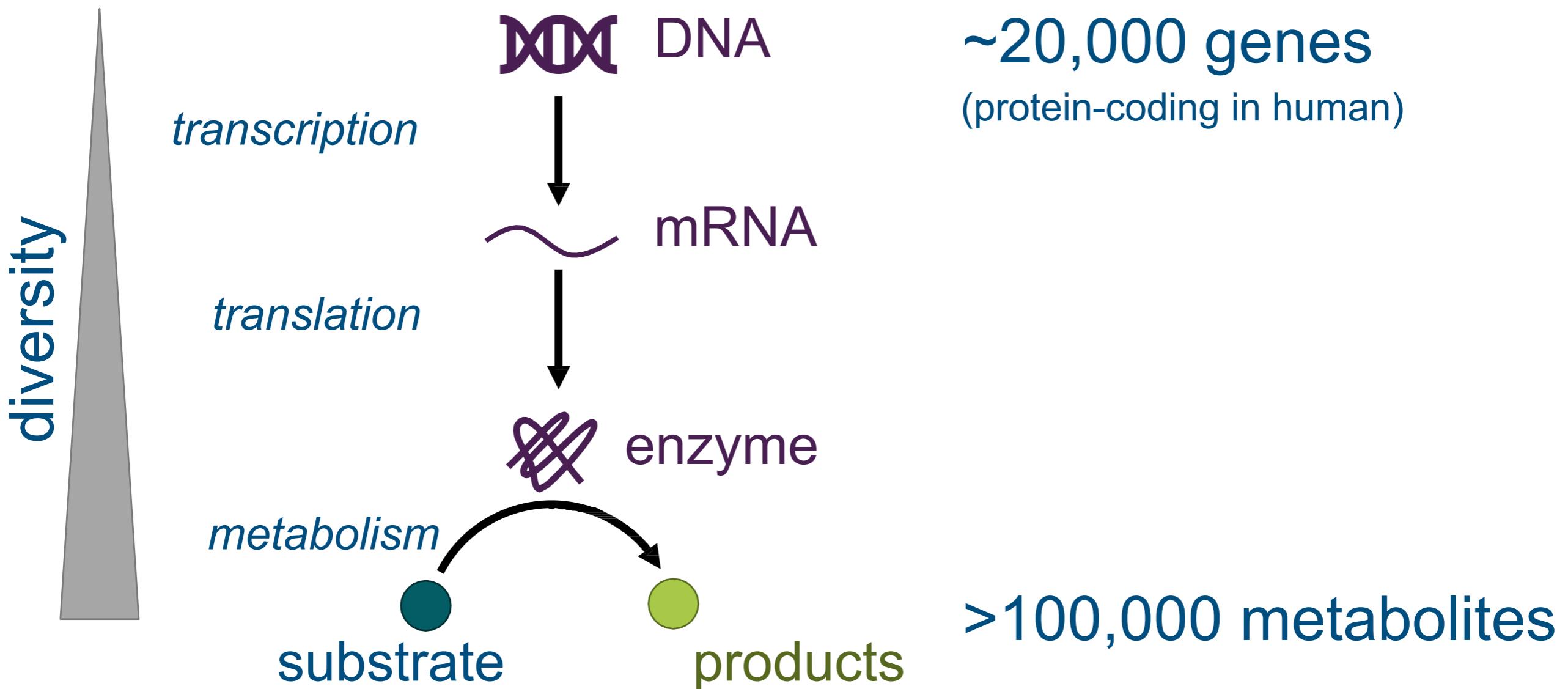
**How to systematically and integratively
examine metabolic disruption in disease?**

Moving from genetic to metabolic characterisations

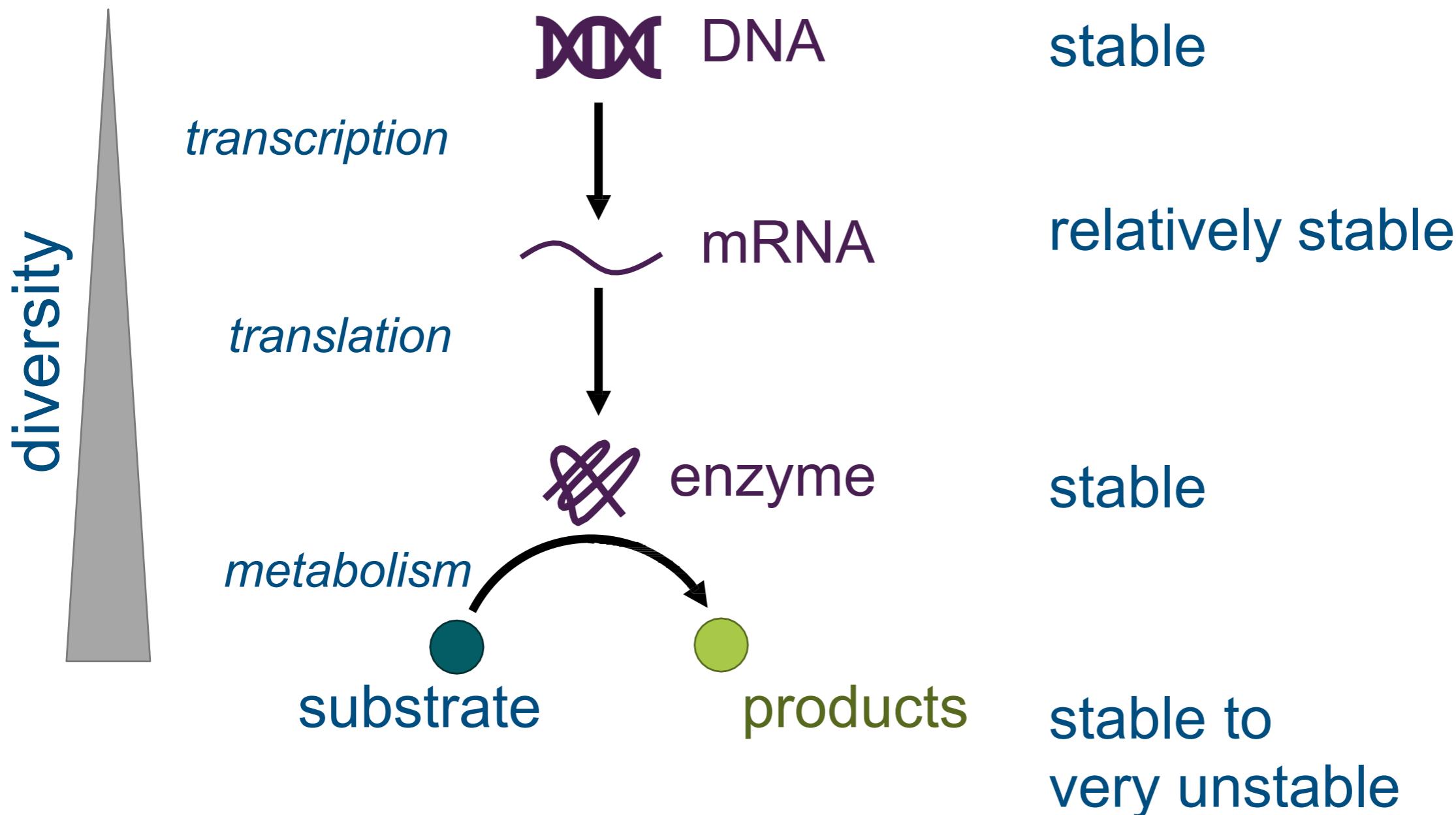


Metabolism provides the energy and building blocks necessary to sustain life.

Moving from genetic to metabolic characterisations



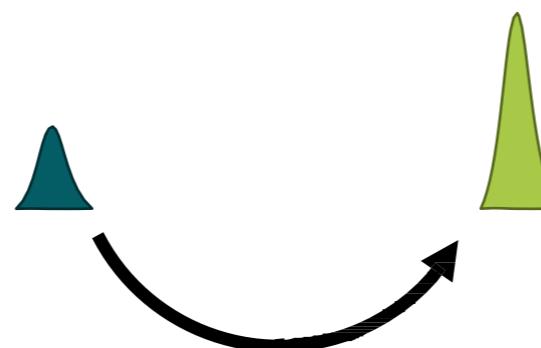
Moving from genetic to metabolic characterisations



Quantifying fluxes



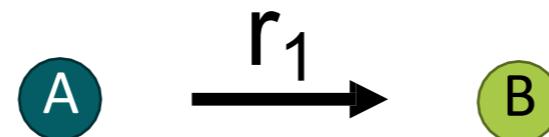
We can generally measure
metabolite concentrations



...but what is often important is
the flow or **flux** of metabolites
through the reactions.



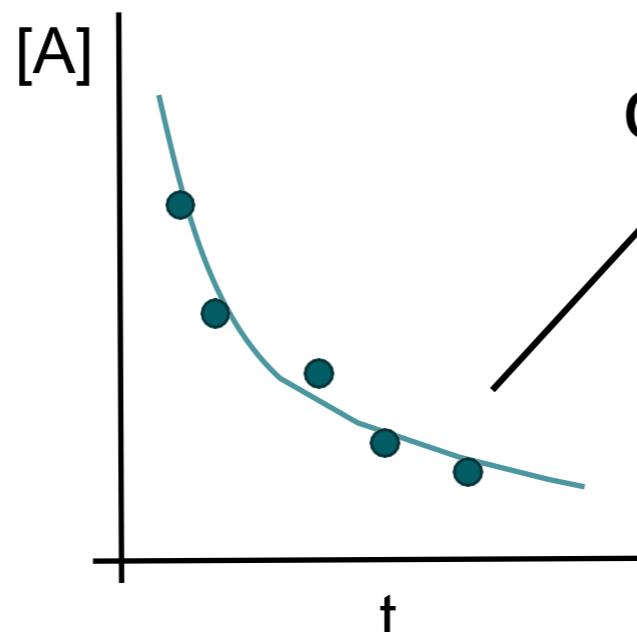
Quantifying fluxes



$$\text{flux} = v_1$$

$$\frac{d[A]}{dt} = -v_1$$

$$\frac{d[B]}{dt} = v_1$$

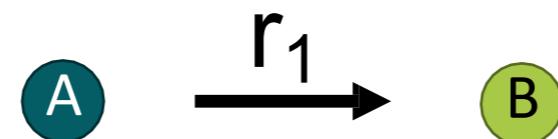


calculate v_1

v_1 = production rate of B

$$v : \frac{\text{mmol}}{g_{DCW} \cdot h}$$

Enzyme kinetics require knowledge of many kinetic parameters

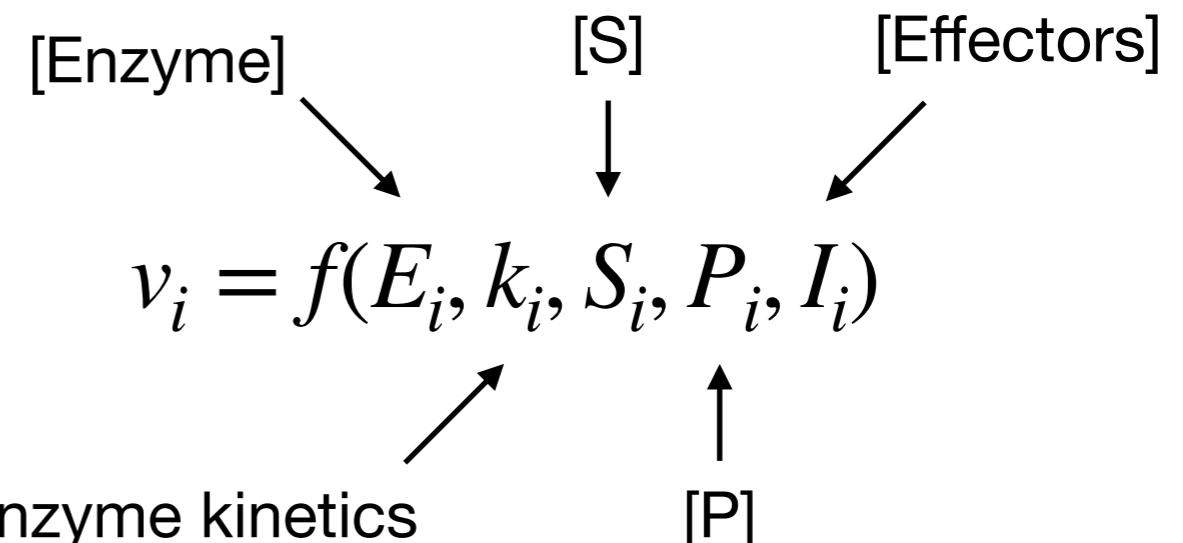


$$\text{flux} = v_1$$

Estimated experimentally

$$\frac{d[A]}{dt} = -v_1 = \underline{k_1} \times [A]$$

$$\frac{d[A]}{dt} = -v_1 = \underline{\frac{V_{max} \times [A]}{K_M + [A]}}$$

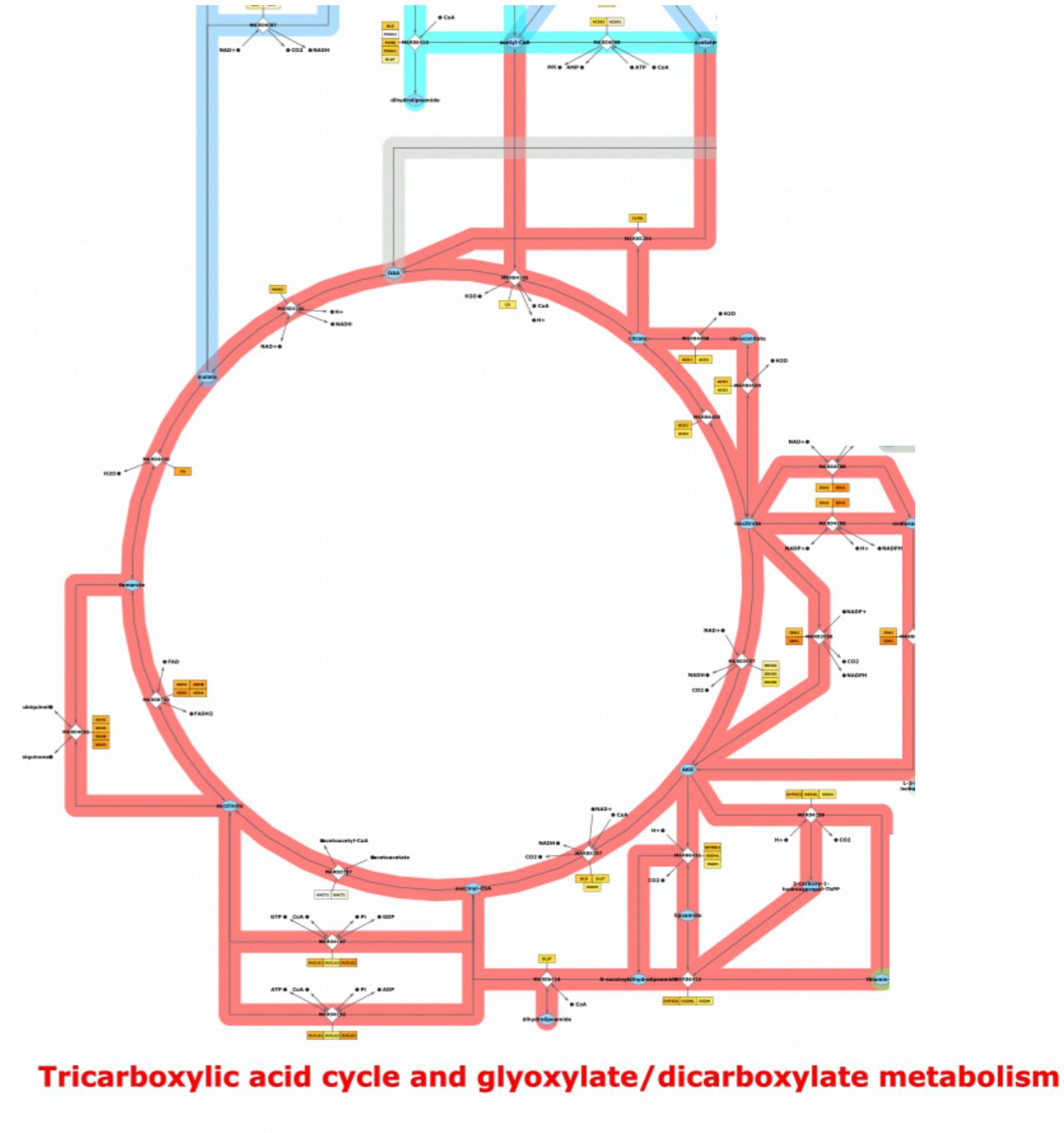


Expanding flux simulations globally

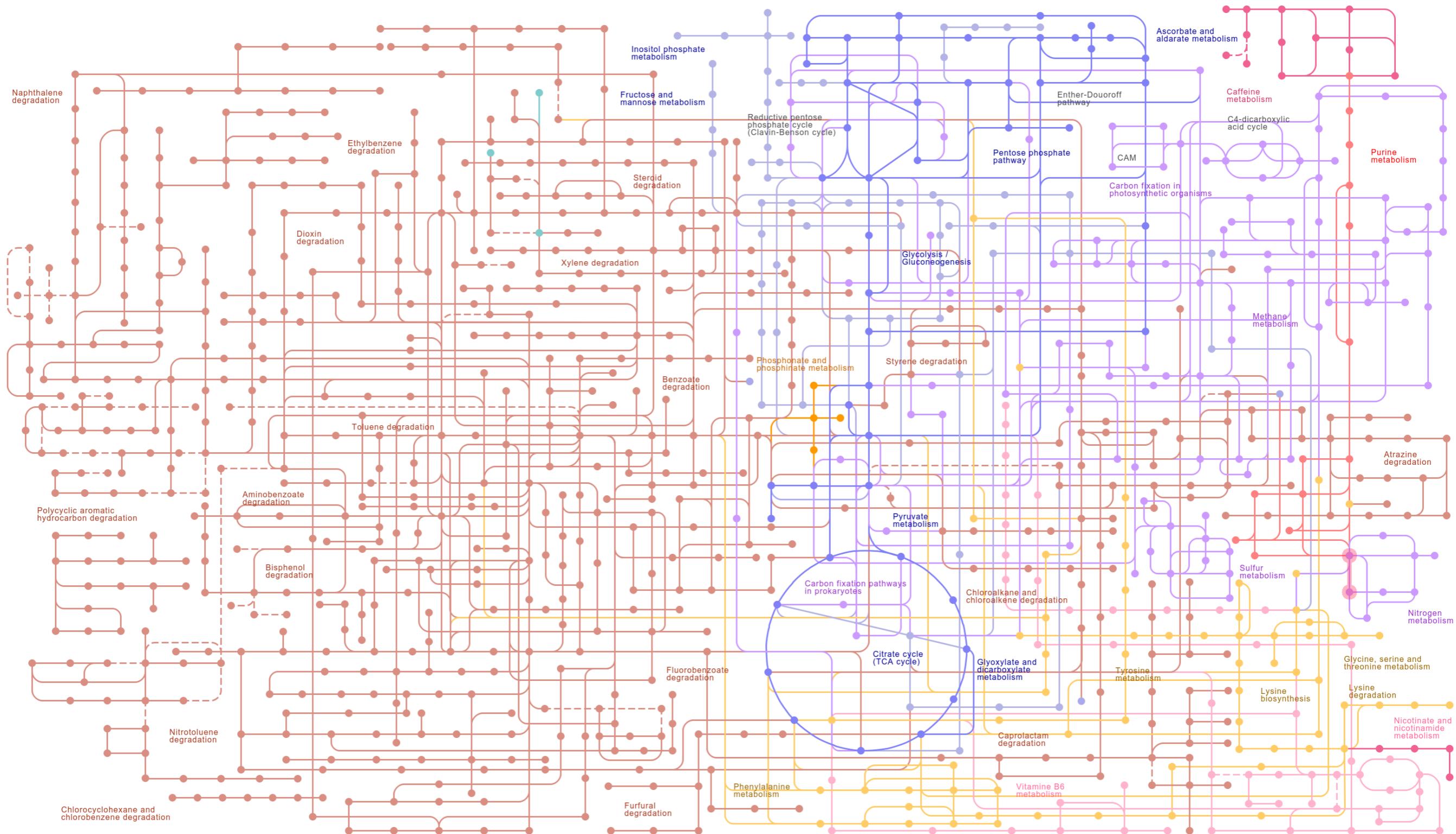
Detailed kinetics available for many elements of central metabolism

Many metabolite abundances are not possible to quantify

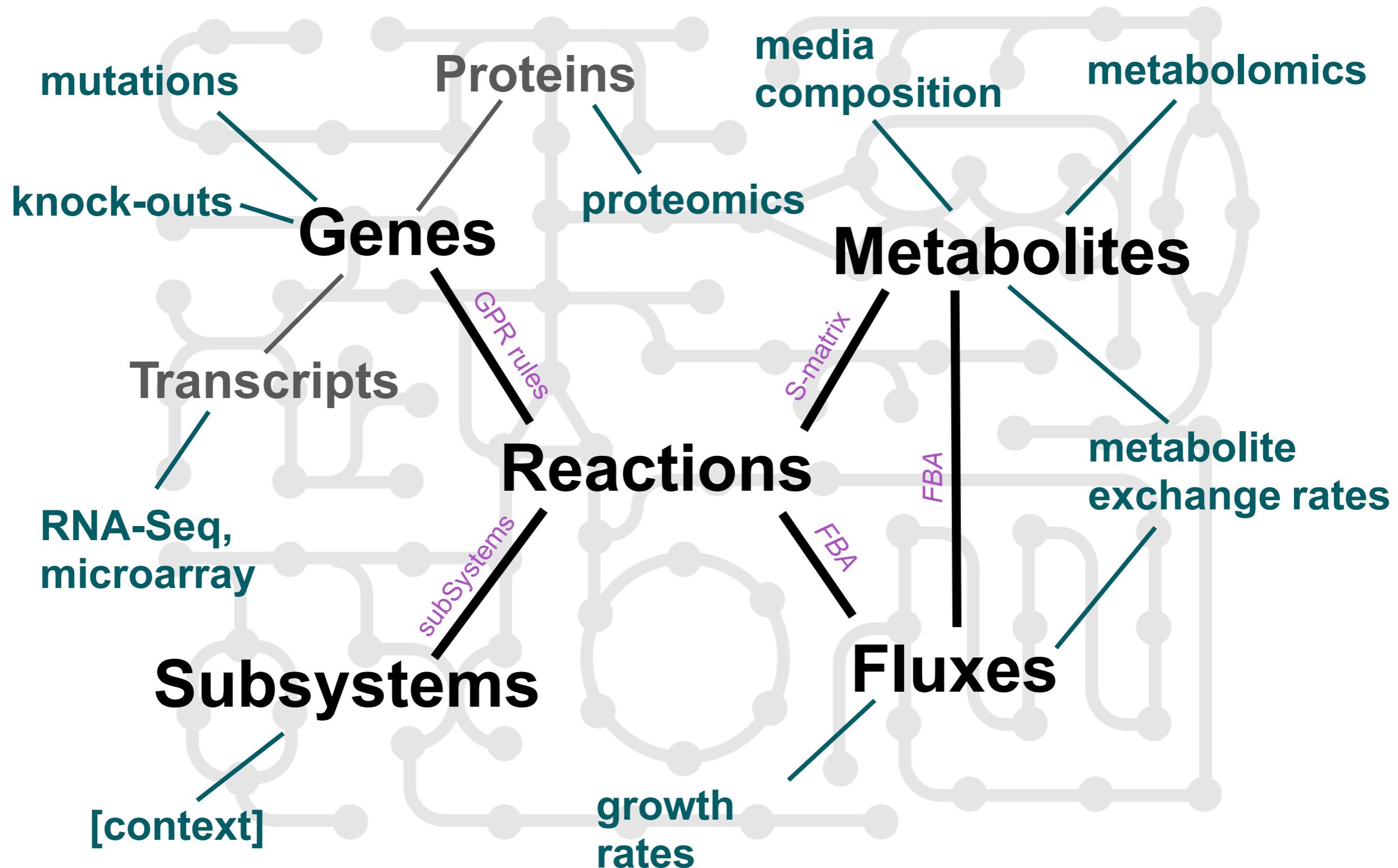
Multi-dimensional problem (genomic, metabolomic, proteomic, ...)



Expanding flux simulations globally



GEMs as an integrative tool



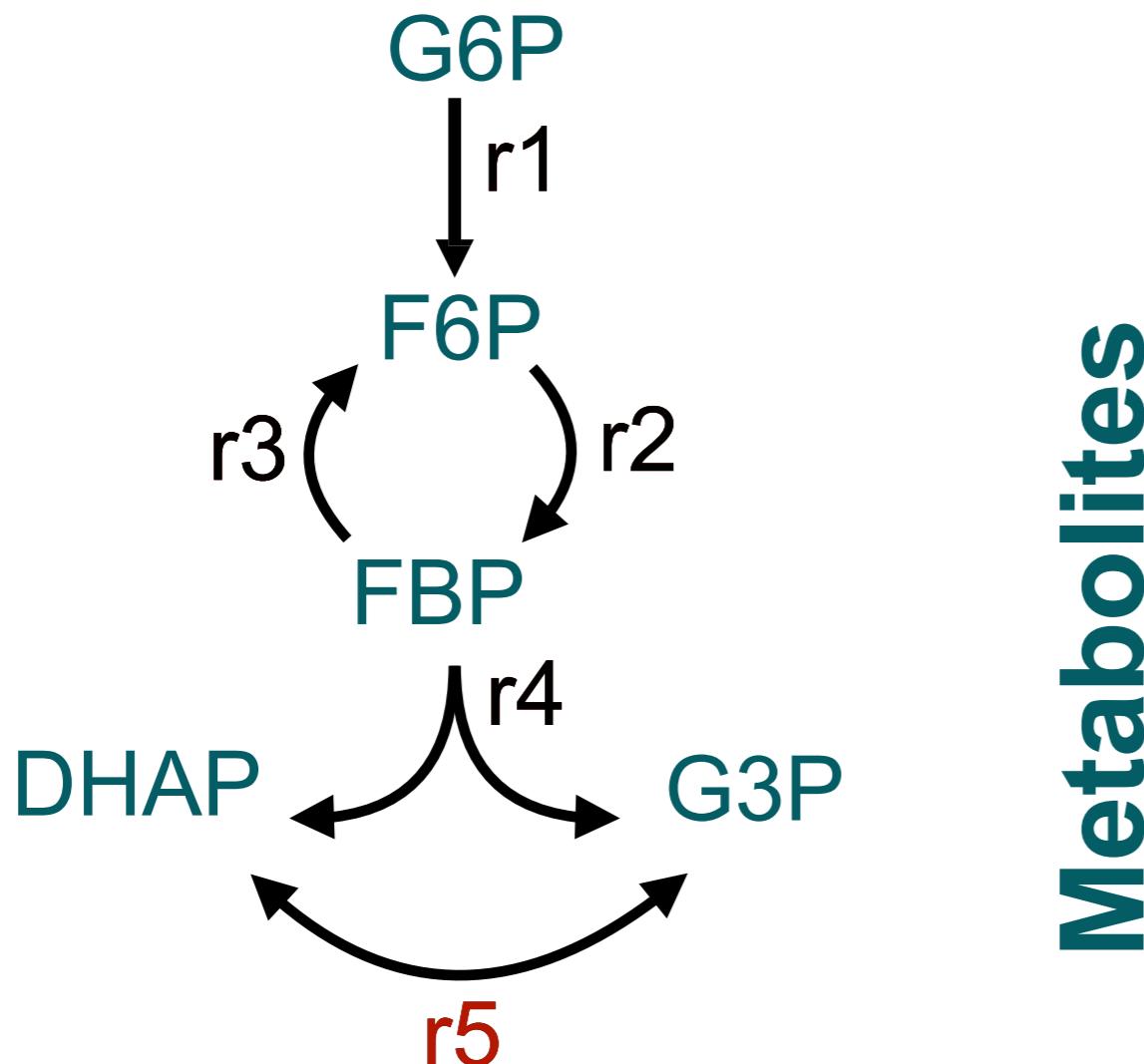
Overview

2. Introduction to metabolic modelling and FBA

3. Downstream analysis and combination with topology analysis

Original sources of images provided as reference and hyperlinks where applicable.

Using reaction stoichiometry to describe metabolism



Metabolites

Reactions

	r1	r2	r3	r4	r5
G6P	-1	0	0	0	0
F6P	1	-1	1	0	0
FBP	0	1	-1	-1	0
DHAP	0	0	0	1	-1
G3P	0	0	0	1	1

Genome-scale model (GEM)

Chemical formula

Charge

InChI code

Other external IDs

3

KEGG ID	Compartment	Name	Symbol	r1	r2	r3	r4	r5
C00668	cytosol [c]	glucose 6-phosphate	G6P	-1	0	0	0	0
C00085	cytosol [c]	fructose 6-phosphate	F6P	1	-1	1	0	0
C00354	cytosol [c]	fructose-1,6-bisphosphate	FBP	0	1	-1	-1	0
C00111	cytosol [c]	dihydroxyacetone phosphate	DHAP	0	0	0	1	-1
C00118	cytosol [c]	glyceraldehyde 3-phosphate	G3P	0	0	0	1	1

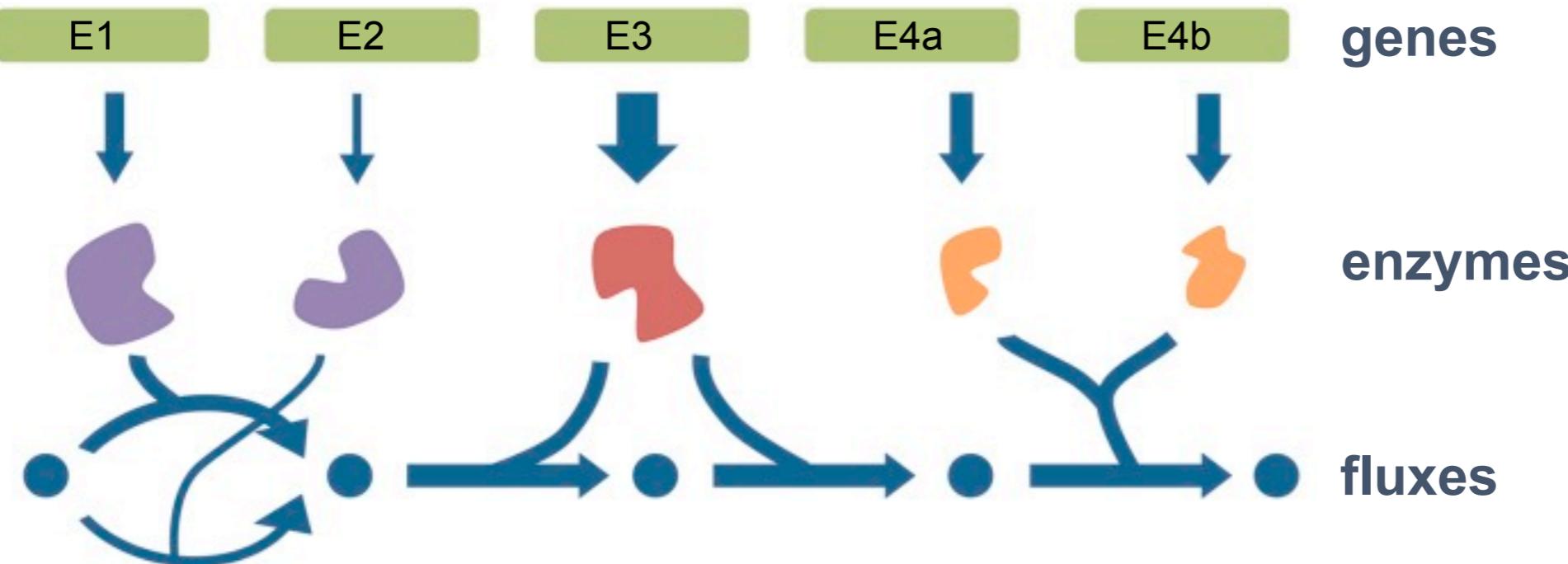
Genome-scale model (GEM)

		Genes (symbol)	Proteins (UniProt)	Transcript IDs	GO Terms	Orthologs
	r1	GPI	P06744			
	r2	n/a				
	r3	FBP1, FBP2	P09467, O00757			
	r4	ALDOA, ...	P04075, ...			
	r5	TPI1	P60174			
G6P	r1	-1	0	0	0	0
F6P	r2	1	-1	1	0	0
FBP	r3	0	1	-1	-1	0
DHAP	r4	0	0	0	1	-1
G3P	r5	0	0	0	1	1

Reactions linked to genes that encode the enzymes that catalyze the reaction

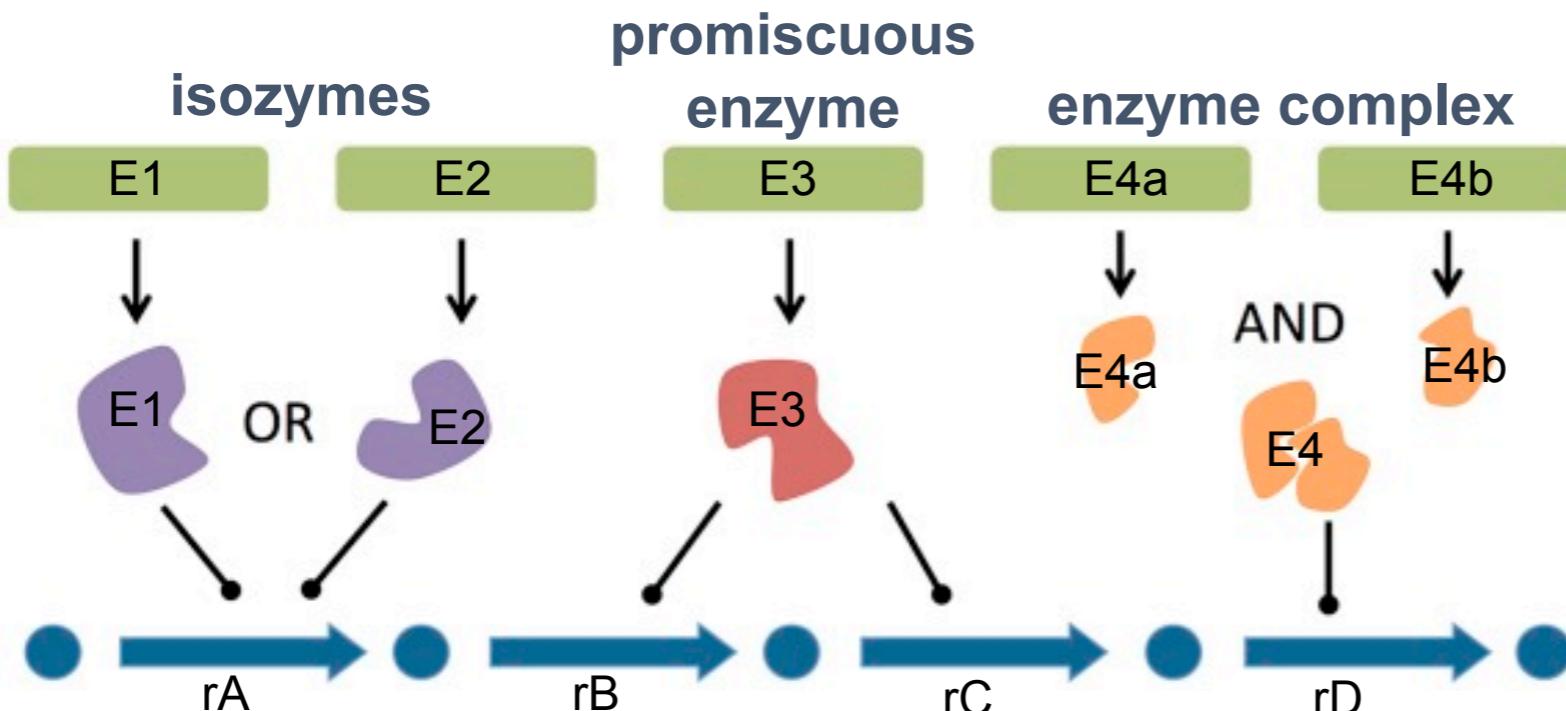
“gene-protein rules” (GPR rules)

GPR rules



GPR rules can be linked with gene expression

GPR rules

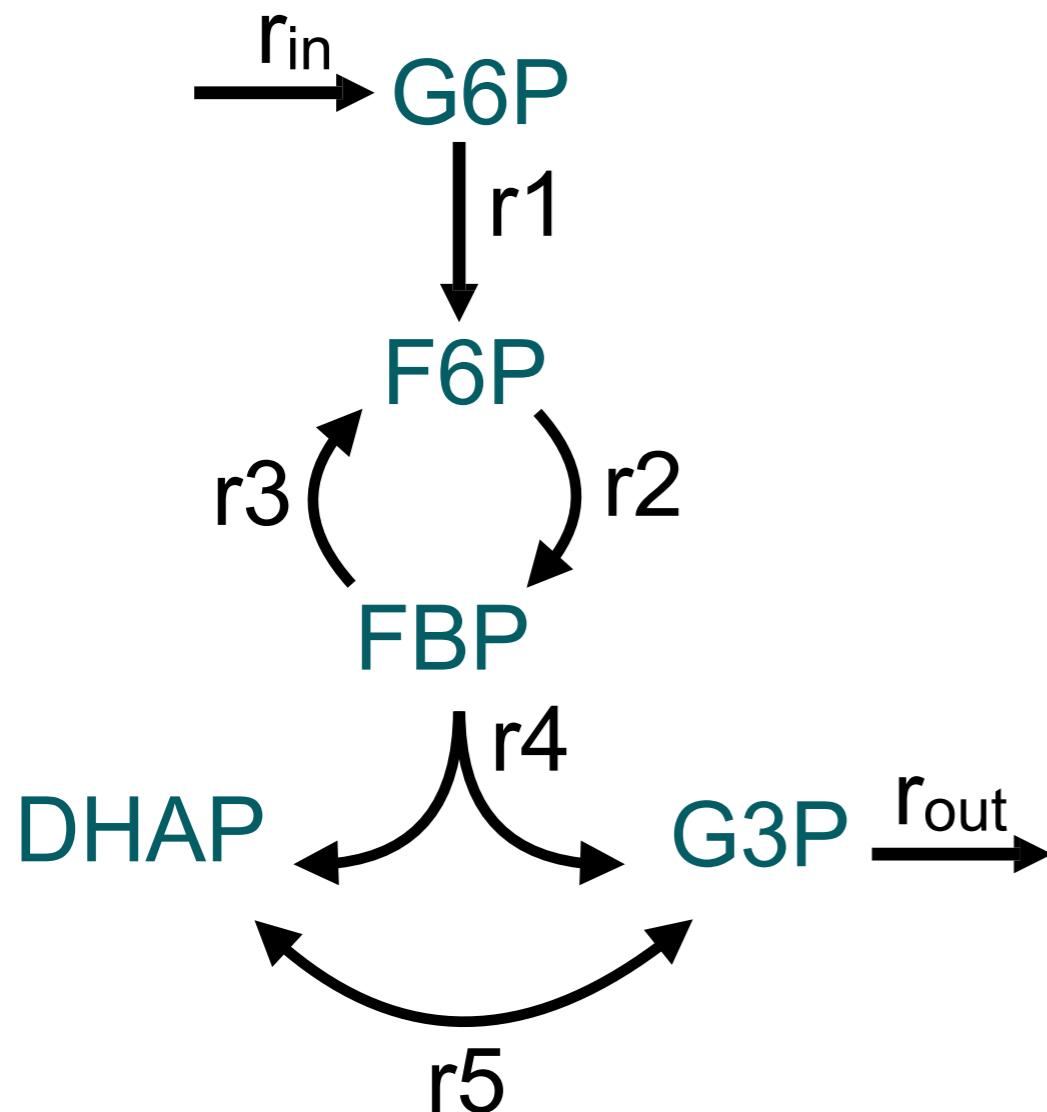


GPR Rules enable more accurate simulation of gene inactivation/knock-out

Knockout	Effect
E1	none
E2	none
E1 + E2	rA inactive
E3	rB rC inactive
E4a	rD inactive
E4b	rD inactive
E4a + E4b	rD inactive

Flux Balance Analysis (FBA)

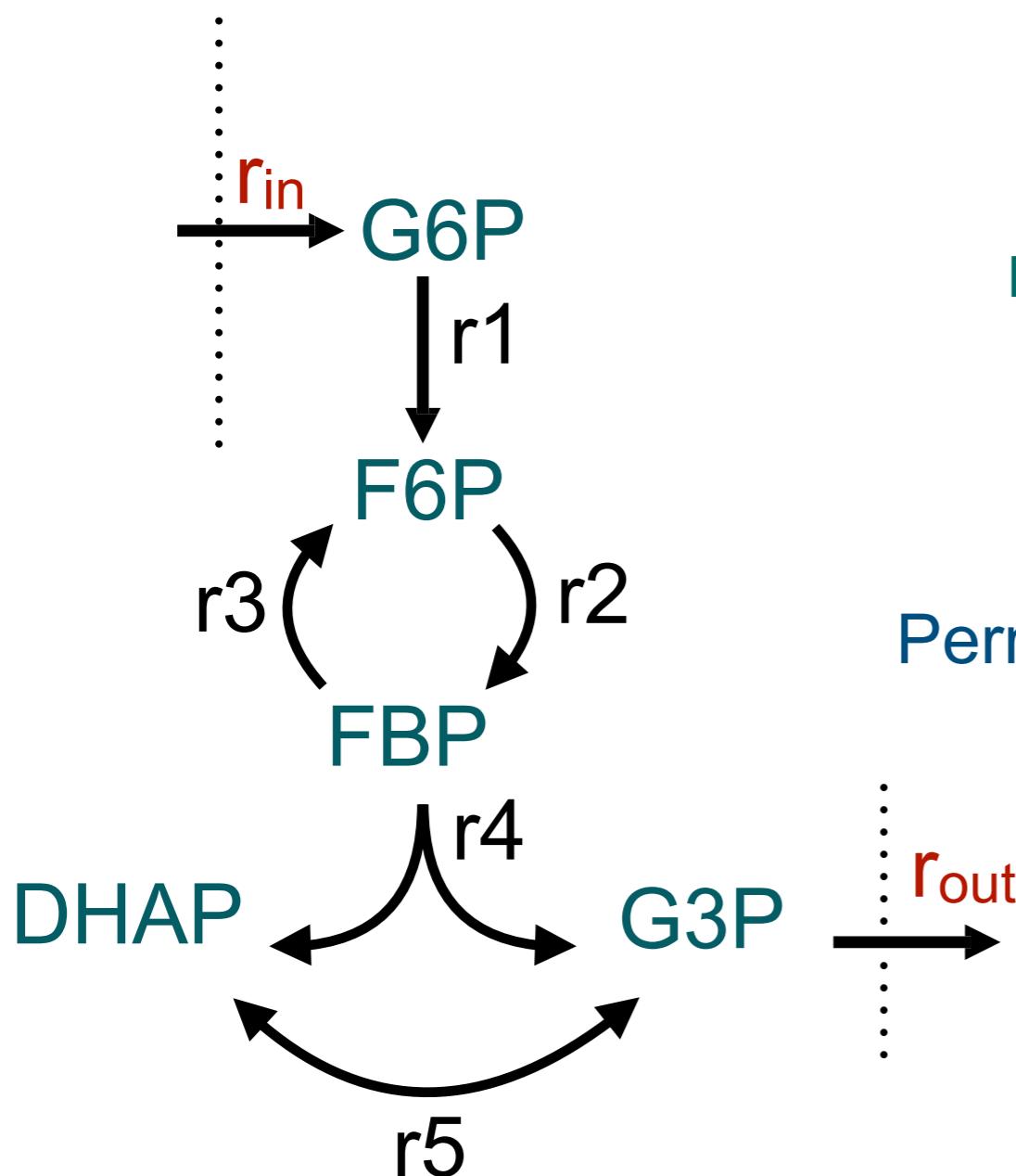
FBA seeks to calculate the reaction **fluxes (v)** of a network



Based on the **conservation of mass**: it cannot be created or destroyed

$$\frac{d[A]}{dt} = v_{prod} - v_{cons}$$

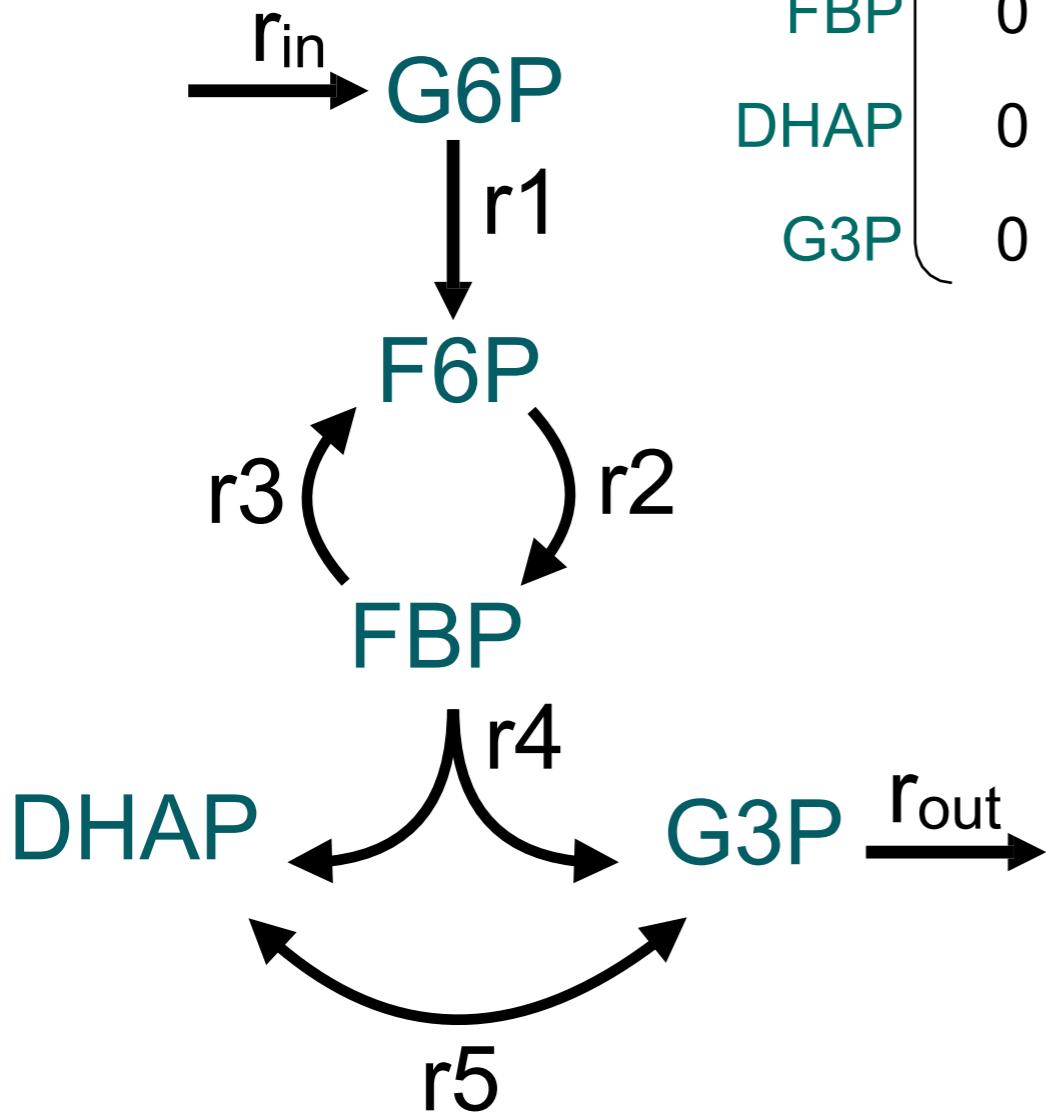
Flux Balance Analysis (FBA)



	r1	r2	r3	r4	r5	r _{in}	r _{out}
G6P	-1	0	0	0	0	1	0
F6P	1	-1	1	0	0	0	0
FBP	0	1	-1	-1	0	0	0
DHAP	0	0	0	1	-1	0	0
G3P	0	0	0	1	1	0	-1

Permits simulating contextual profiles

Flux Balance Analysis (FBA)



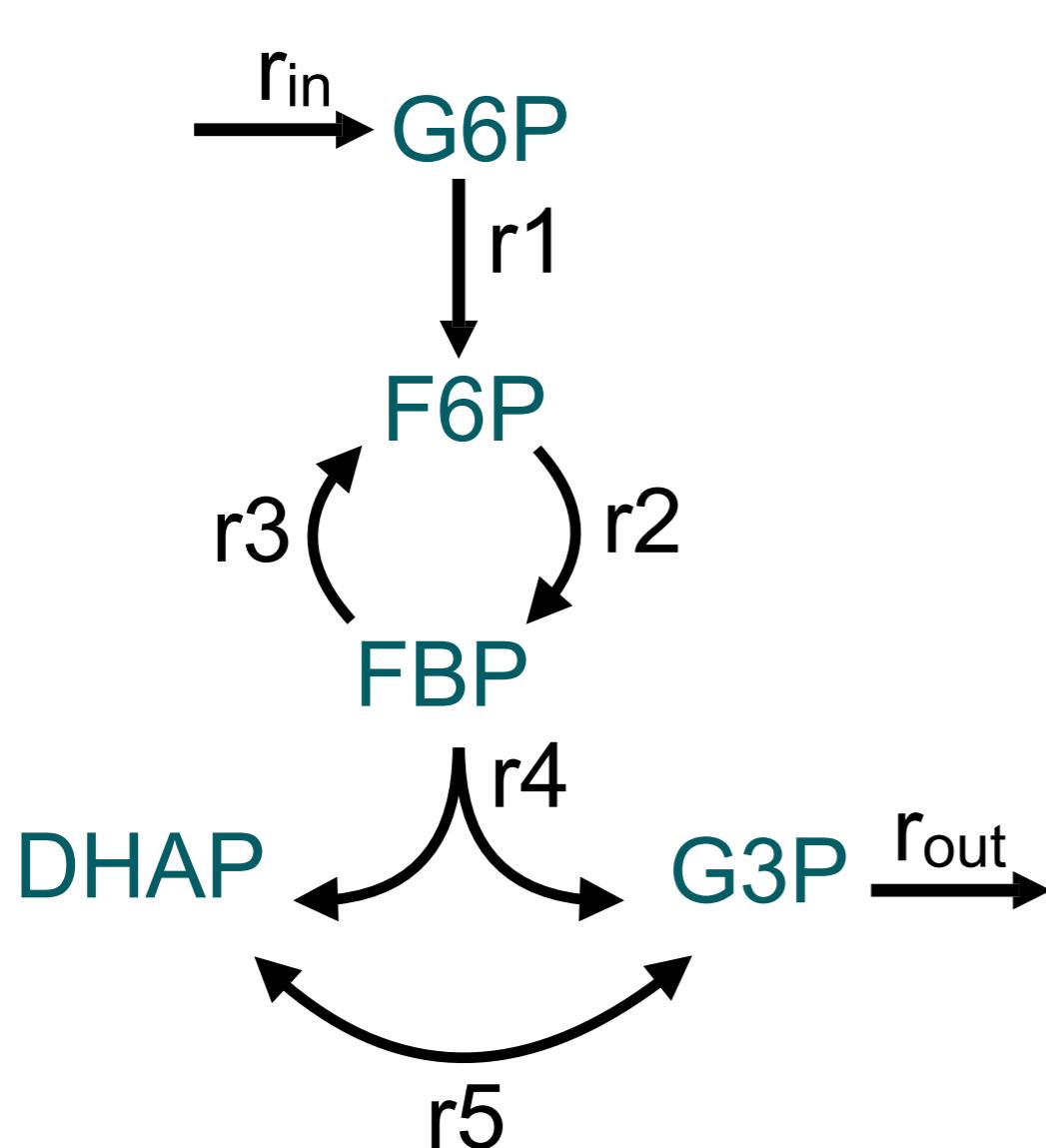
$$\begin{array}{c}
 \text{G6P} \\
 \text{F6P} \\
 \text{FBP} \\
 \text{DHAP} \\
 \text{G3P}
 \end{array}
 \begin{pmatrix}
 r_1 & r_2 & r_3 & r_4 & r_5 \\
 -1 & 0 & 0 & 0 & 0 \\
 1 & -1 & 1 & 0 & 0 \\
 0 & 1 & -1 & -1 & 0 \\
 0 & 0 & 0 & 1 & -1 \\
 0 & 0 & 0 & 1 & 1
 \end{pmatrix}
 \times
 \begin{pmatrix}
 r_{in} & r_{in} \\
 1 & 0 \\
 0 & 0 \\
 0 & 0 \\
 0 & -1
 \end{pmatrix}
 =
 \begin{pmatrix}
 \frac{dG6P}{dt} \\
 \frac{dF6P}{dt} \\
 \frac{dFBP}{dt} \\
 \frac{dDHAP}{dt} \\
 \frac{dG3P}{dt}
 \end{pmatrix}
 \quad \text{with } v_i \text{ representing } r_i$$

$$\frac{d[G6P]}{dt} = v_{in} - v_1$$

$$\frac{d[G3P]}{dt} = v_4 + v_5 - v_{out}$$

Flux Balance Analysis (FBA)

A key assumption to FBA is **steady state**: metabolite concentrations are **constant** through time



$$\frac{d[X]}{dt} = v_{prod} - v_{cons} = 0$$
$$\Rightarrow v_{prod} = v_{cons}$$

Flux Balance Analysis (FBA)

$$\begin{array}{c} \text{G6P} \\ \text{F6P} \\ \text{FBP} \\ \text{DHAP} \\ \text{G3P} \end{array} \begin{pmatrix} r_1 & r_2 & r_3 & r_4 & r_5 & r_{in} & r_{in} \\ -1 & 0 & 0 & 0 & 0 & 1 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & -1 \end{pmatrix} \times \begin{pmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_{in} \\ v_{out} \end{pmatrix} = \begin{pmatrix} d\text{G6P}/dt \\ d\text{F6P}/dt \\ d\text{FBP}/dt \\ d\text{DHAP}/dt \\ d\text{G3P}/dt \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

$$S \cdot v = 0$$

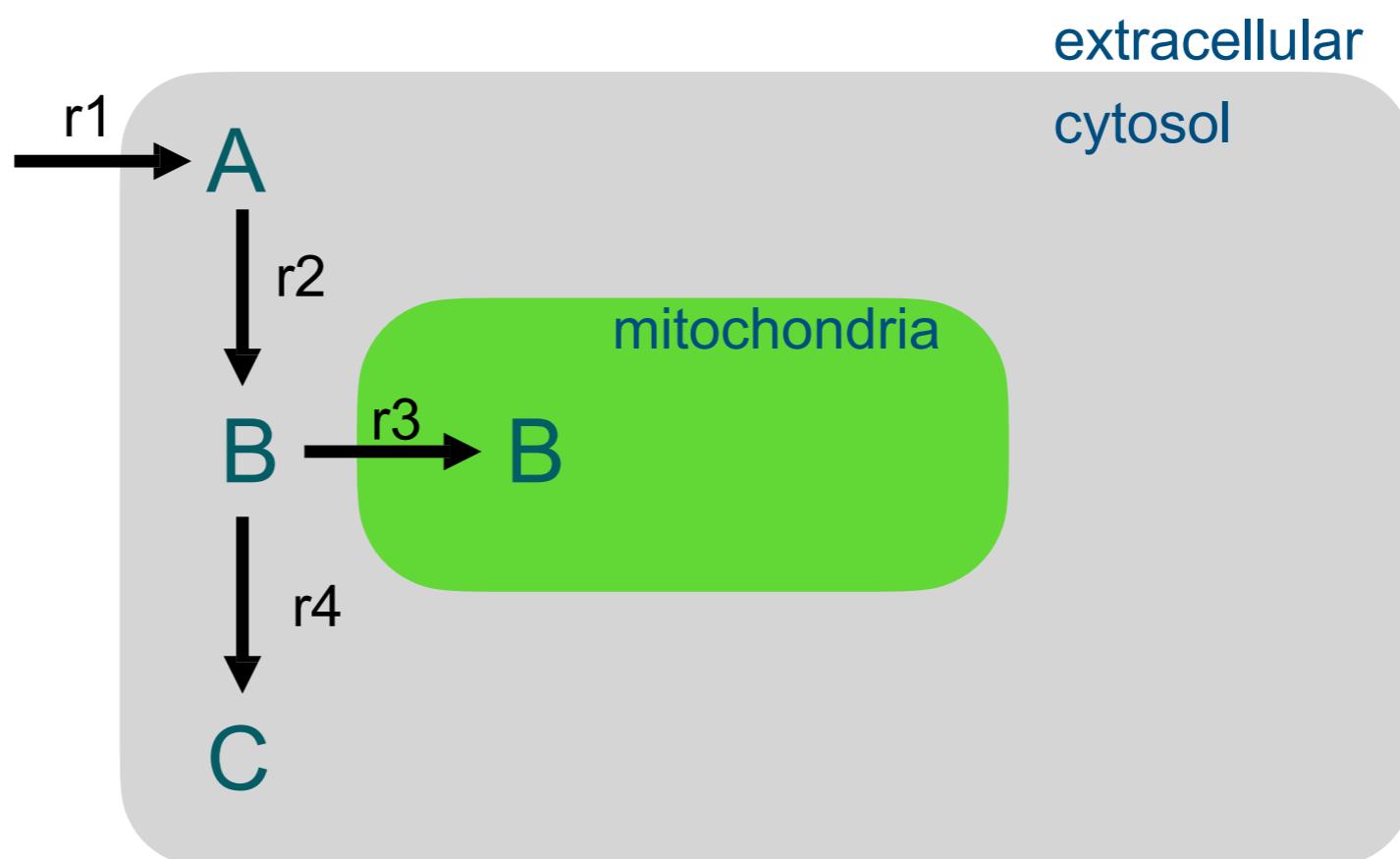
This assumption allows us to **ignore enzyme kinetics**, thus eliminating **many unknown parameters**

$$\frac{d[A]}{dt} = \frac{V_{max} \times [A]}{K_M + [A]}$$

Flux Balance Analysis (FBA)

Models account for compartments:

- Exchange reactions
- Intracellular compartments



Flux Balance Analysis (FBA)

We can further constrain the solution space by limiting reaction fluxes based on their reversibility:

Irreversible
reactions



$0 \leq v \leq \text{upper bound}$

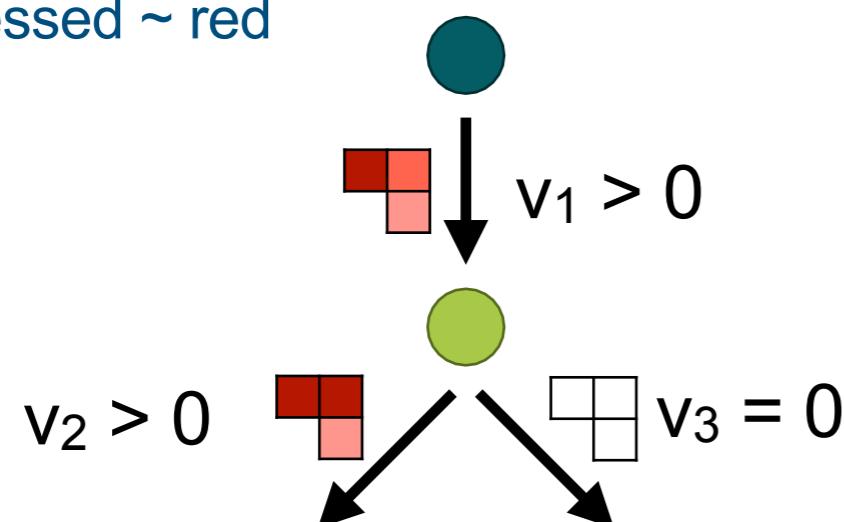
Reversible
reactions



lower bound $\leq v \leq \text{upper bound}$

Gene expression:

Expressed ~ red



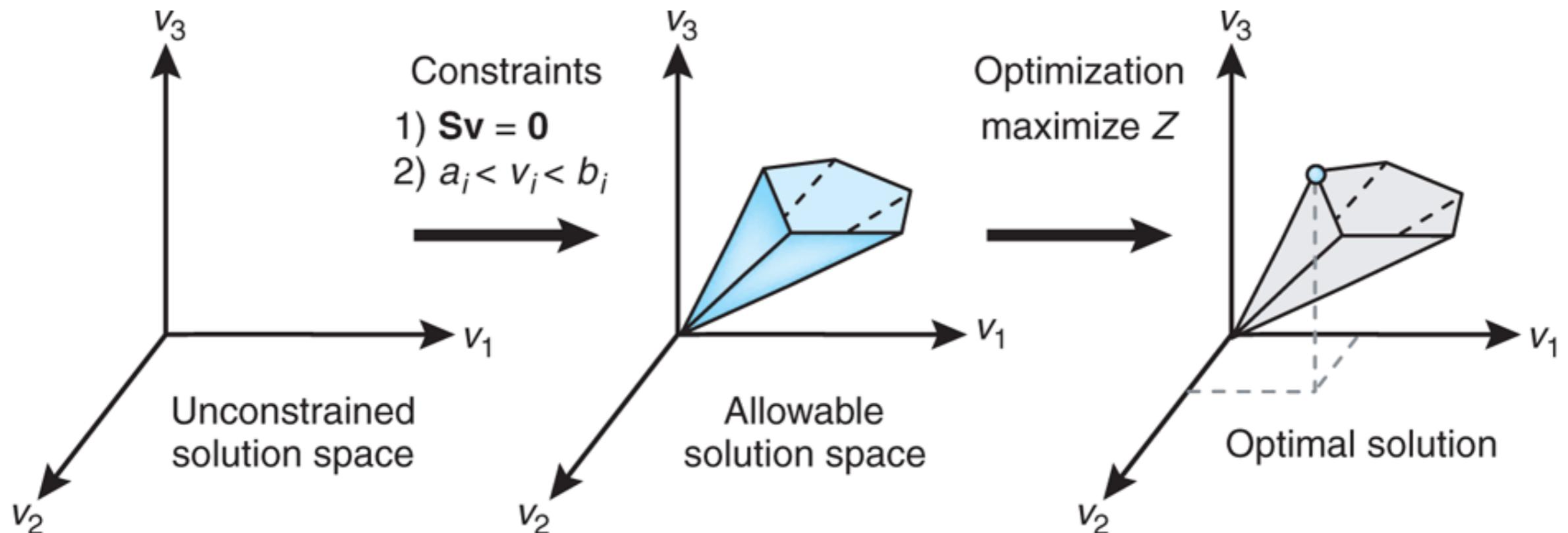
Others:

Enzyme kinetics
Thermodynamic constraints

Metabolic tasks
(~biological feasibility)

Flux Balance Analysis (FBA)

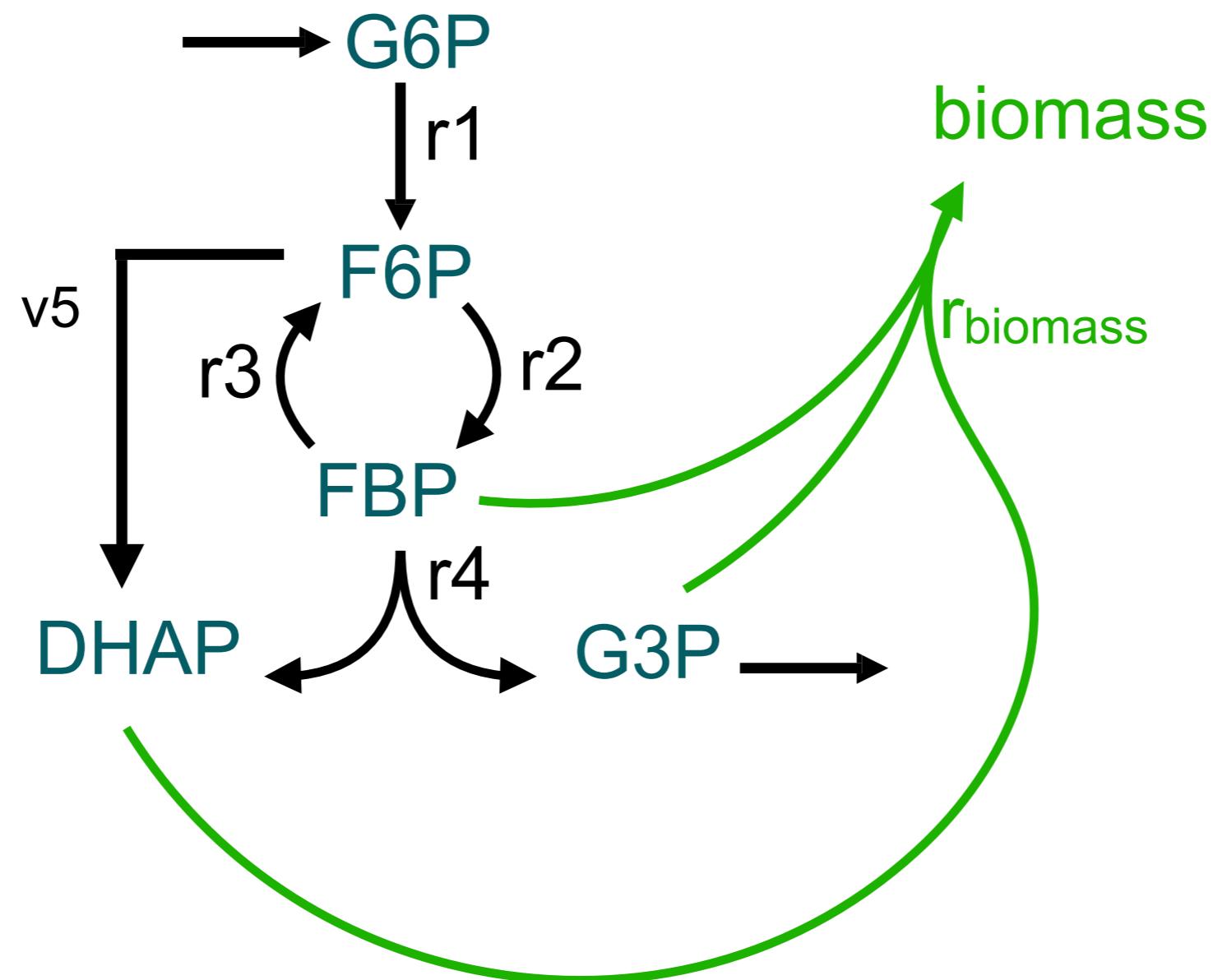
Since the problem is still **under-defined**, FBA uses linear **optimization** to identify a solution that maximizes (or minimizes) some **objective (Z)**



Orth, J., Thiele, I. & Palsson, B. *Nat Biotechnol* (2010).

Flux Balance Analysis (FBA)

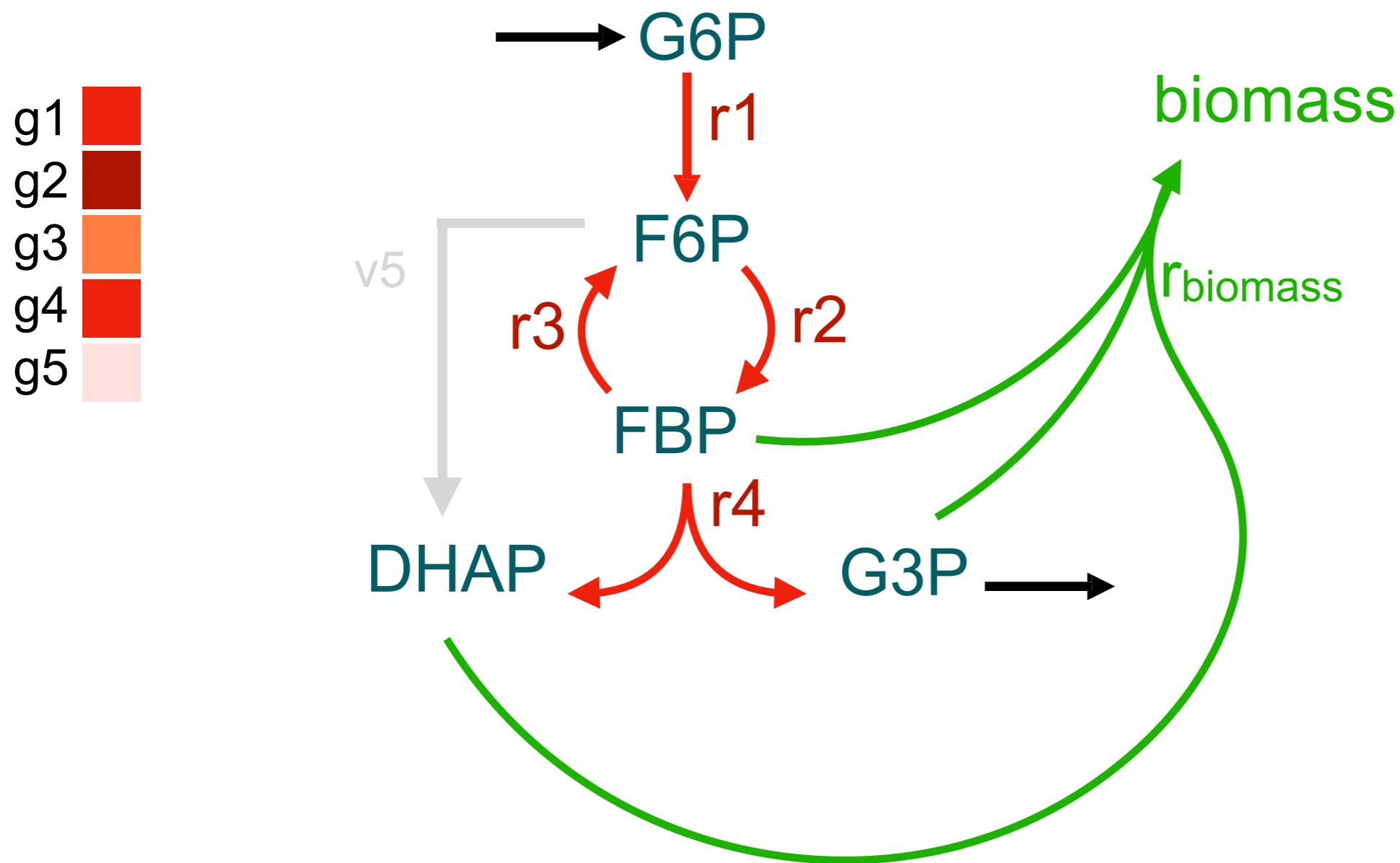
Objective function (i.e. optimisation objective) is often:
maximise an artificial “**biomass**” reaction or **ATP production**



Flux Balance Analysis (FBA)

Objective function (i.e. optimisation objective) is often:
maximise an artificial “**biomass**” reaction or **ATP production**

Optimisation is performed after model constraining

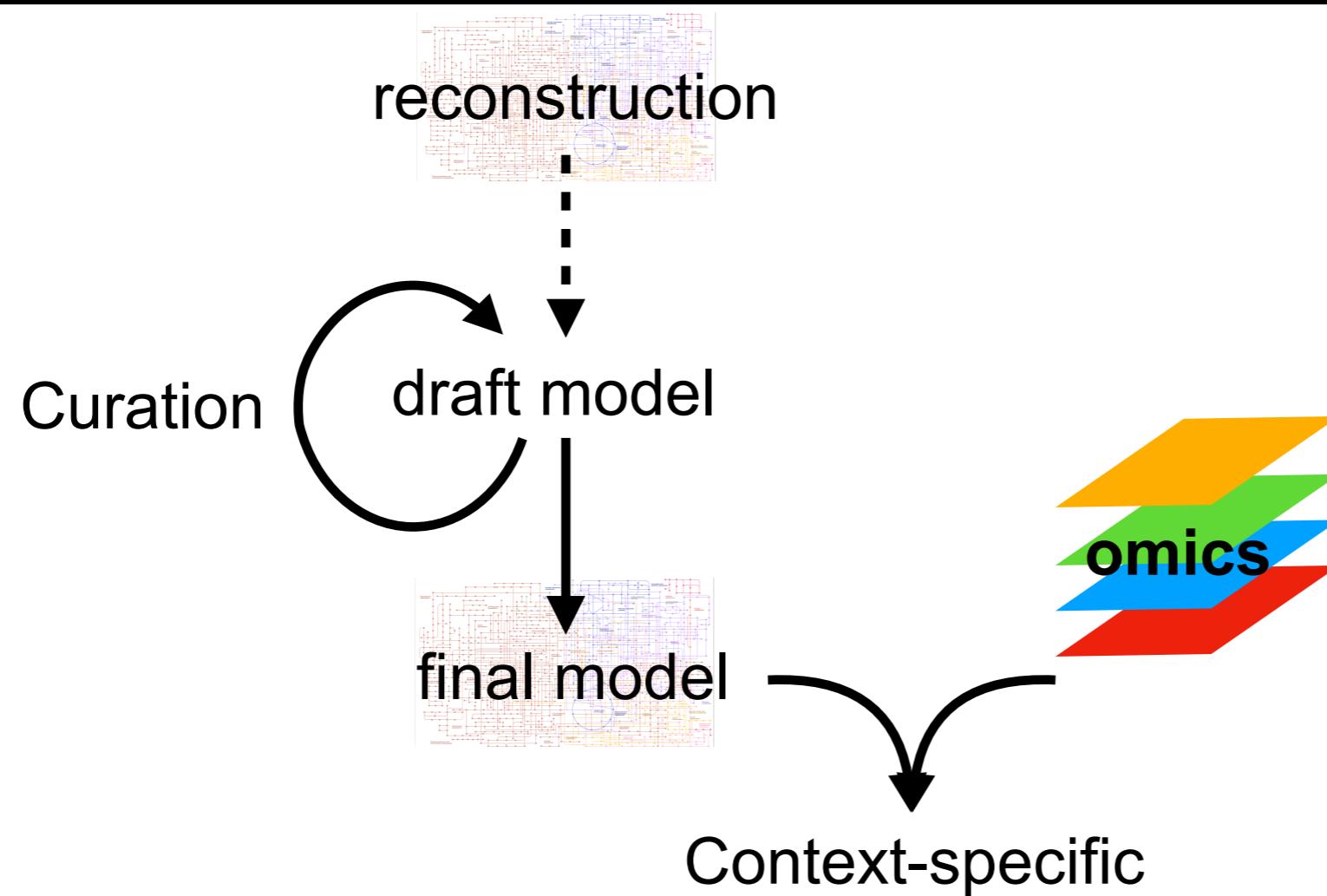


Overview

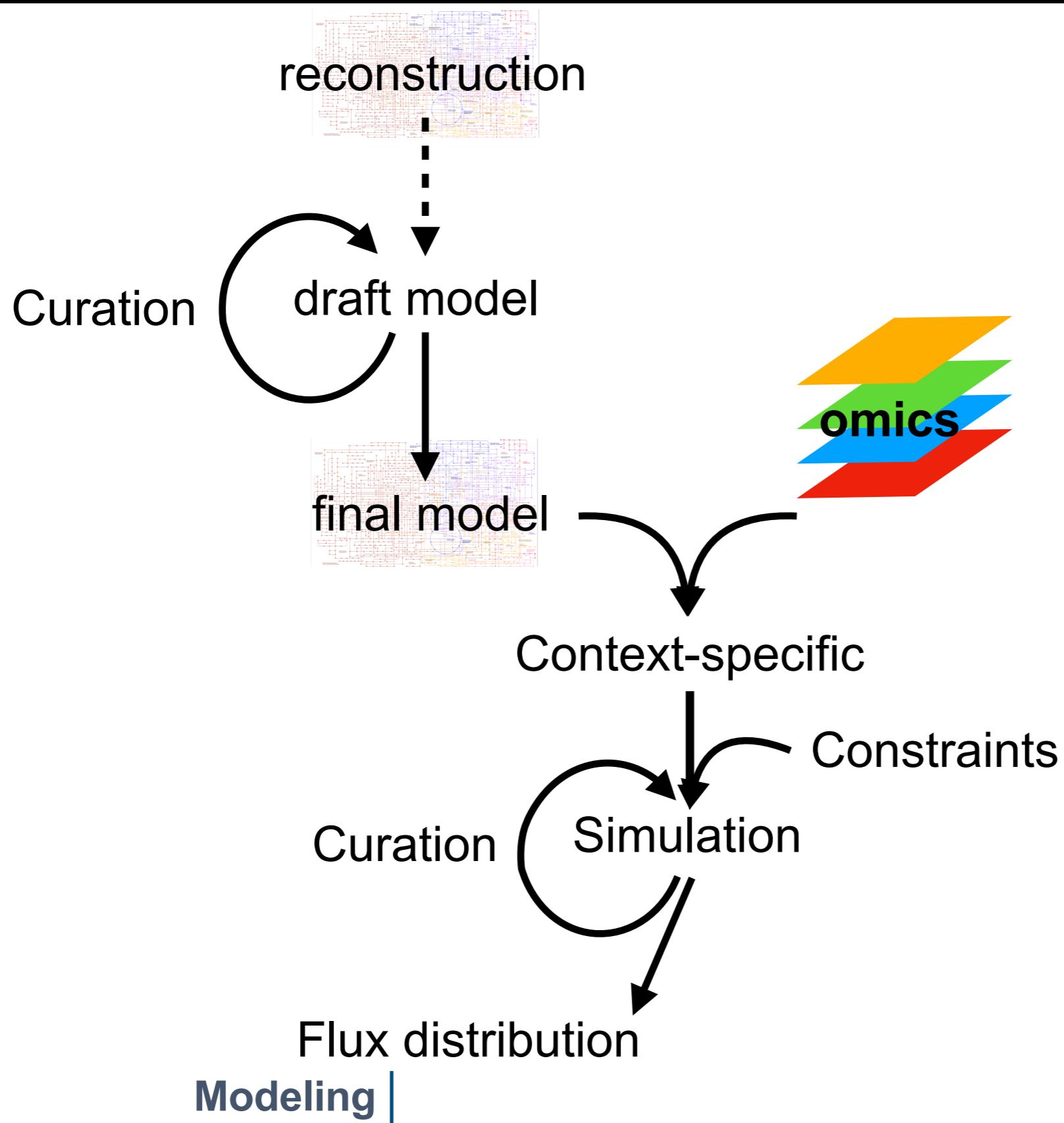
1. Challenge in systematically characterising metabolic imbalances
2. Introduction to metabolic modelling and FBA
- 3. Downstream analysis and combination with topology analysis**

Original sources of images provided as reference and hyperlinks where applicable.

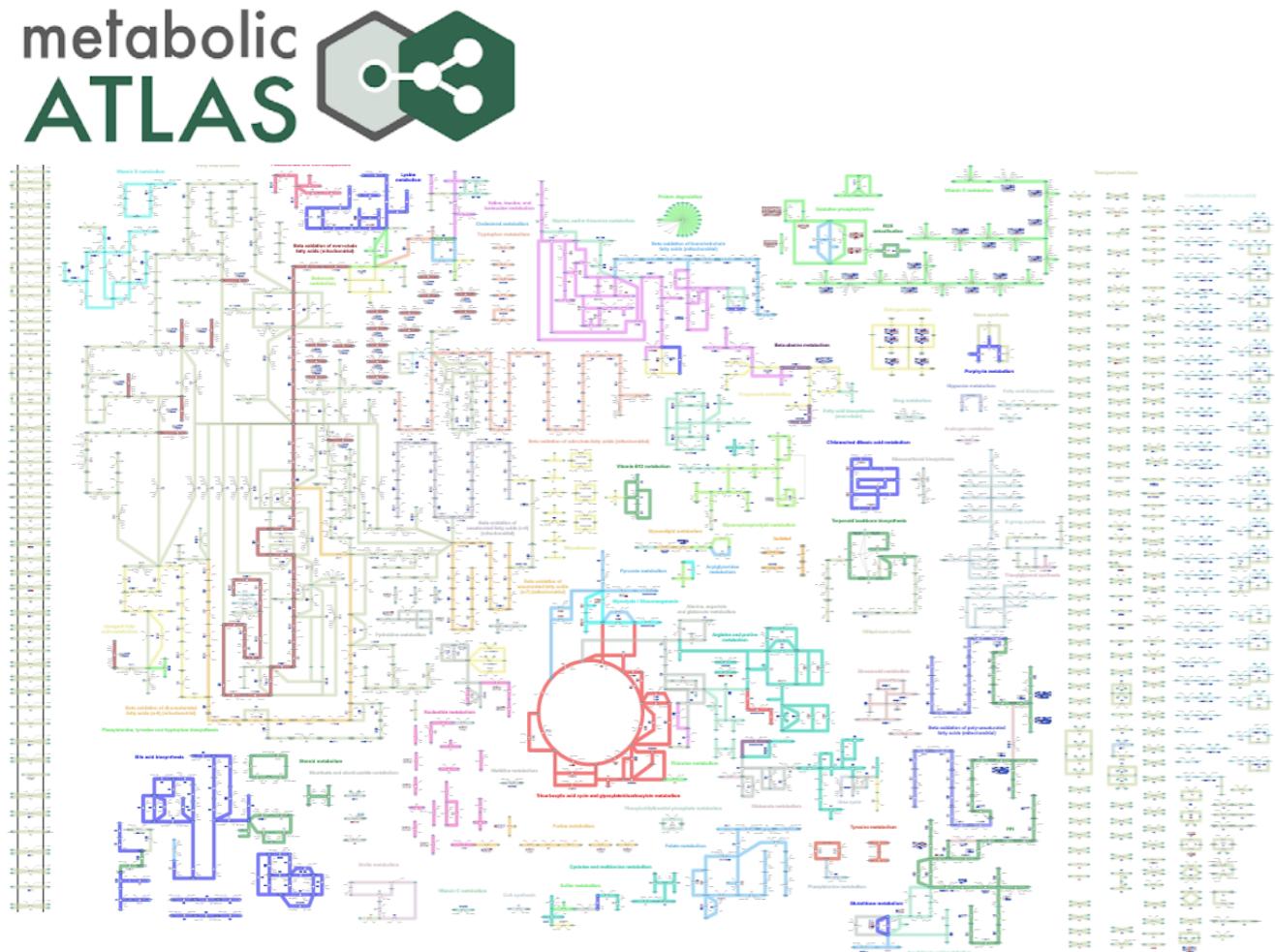
Analysis workflow



Analysis workflow



Genome-scale metabolic models as integrative networks



Simulate flux distributions and pathway activity

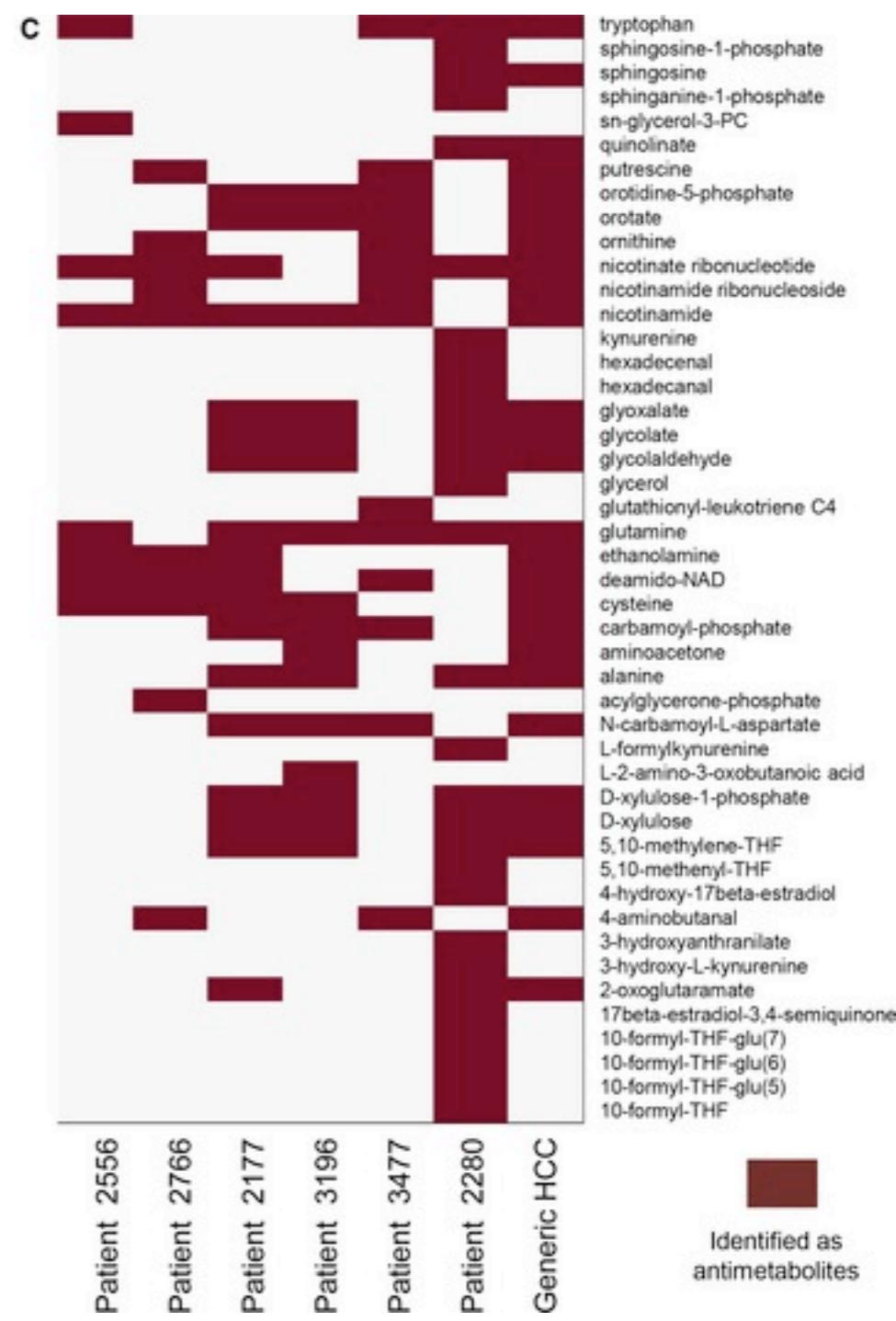
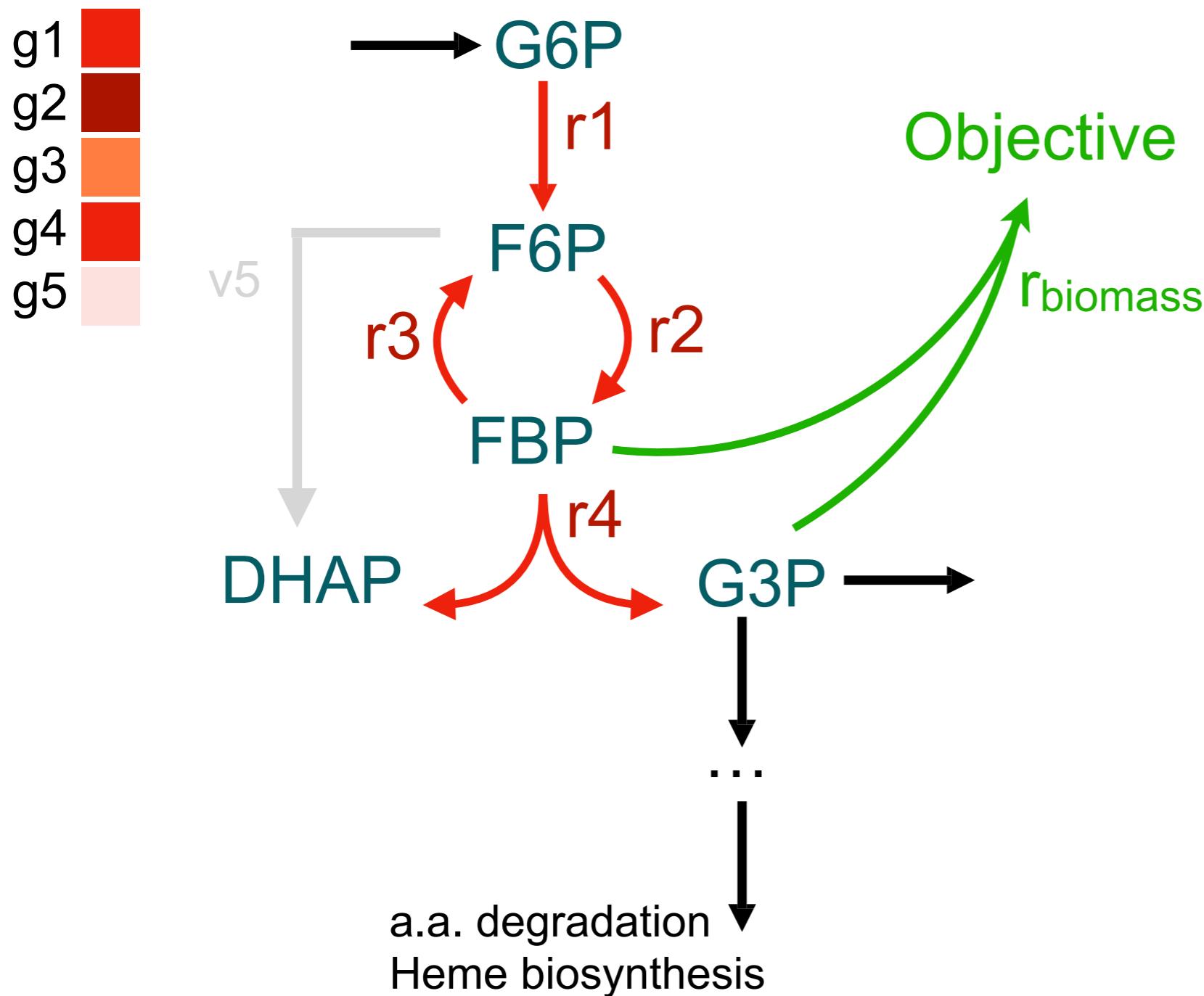
Dysregulated pathways

Essential genes & metabolites

Reporter metabolites

Essential genes, metabolites and metabolic tasks

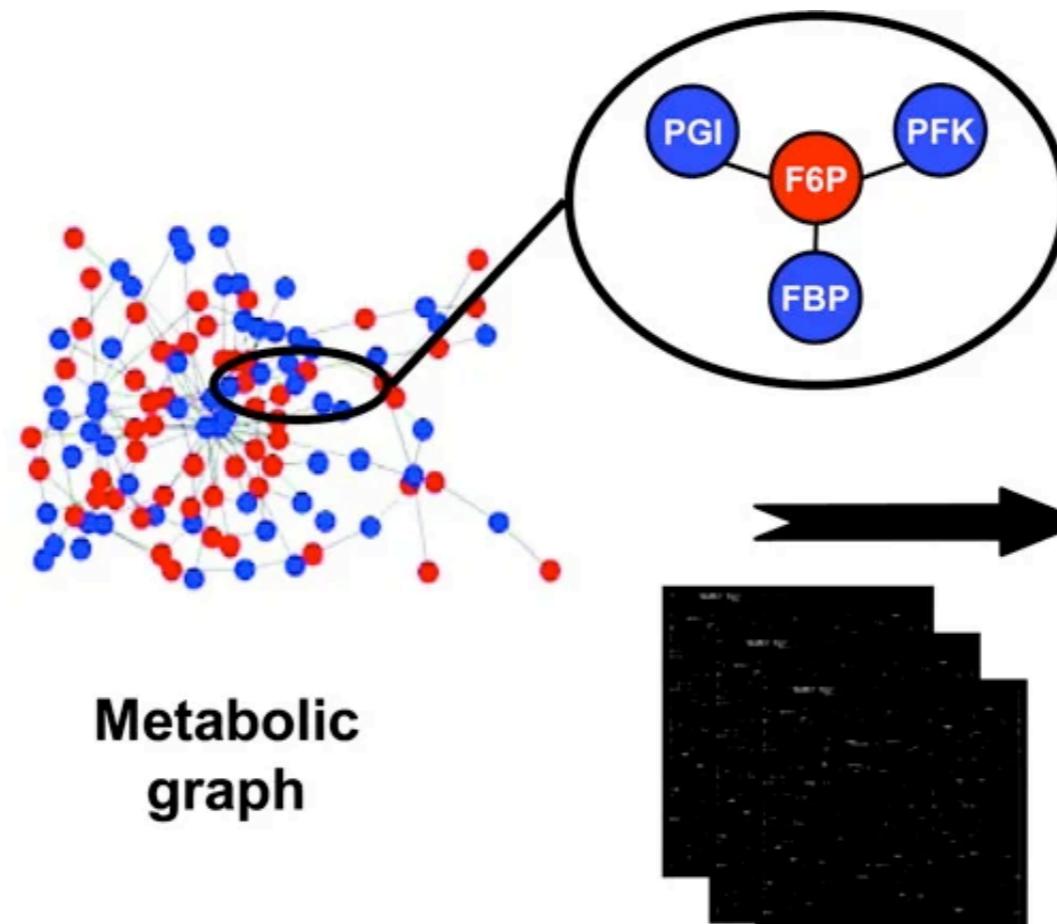
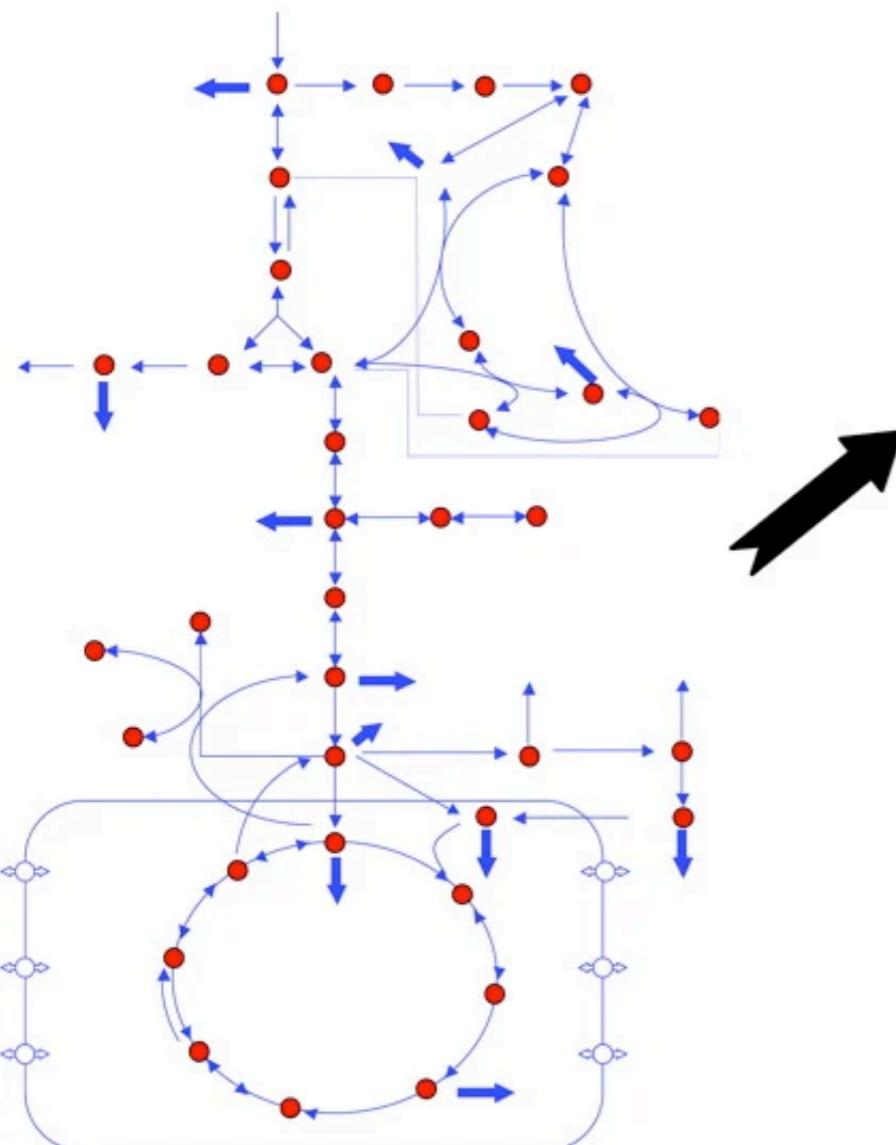
Essential metabolites and genes can be identified based on fulfilment of certain biological tasks



Agren 2014

Reporter metabolites are useful tools for identifying key metabolites linked with transcriptional changes

Metabolic pathways



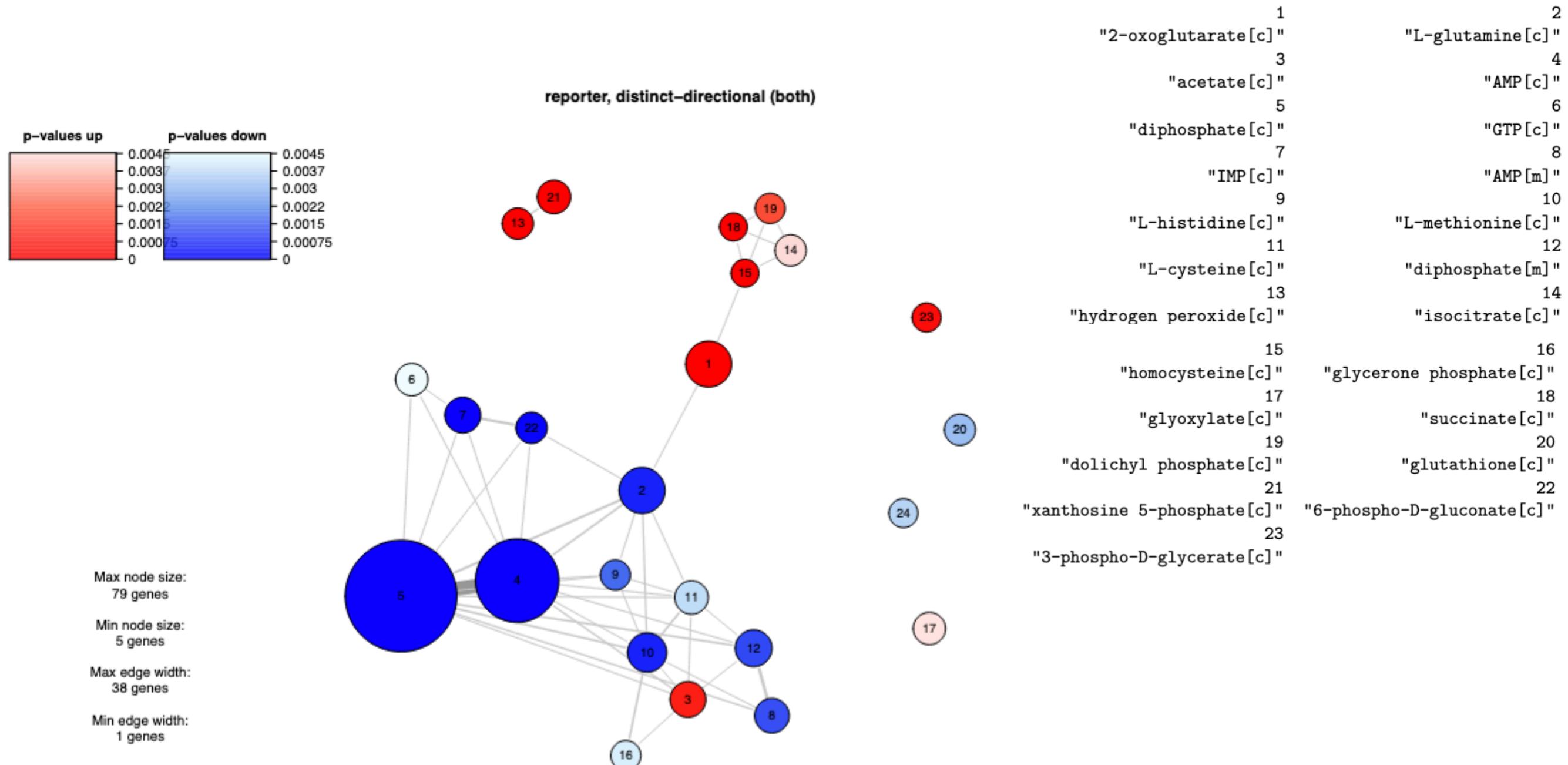
Fructose 6P
NH3
Glucose 6P
GABAxt
NH3xt
GABAxt

Reporter metabolites

Using GEMs as frameworks

Highlight metabolites around which substantial genetic changes occur

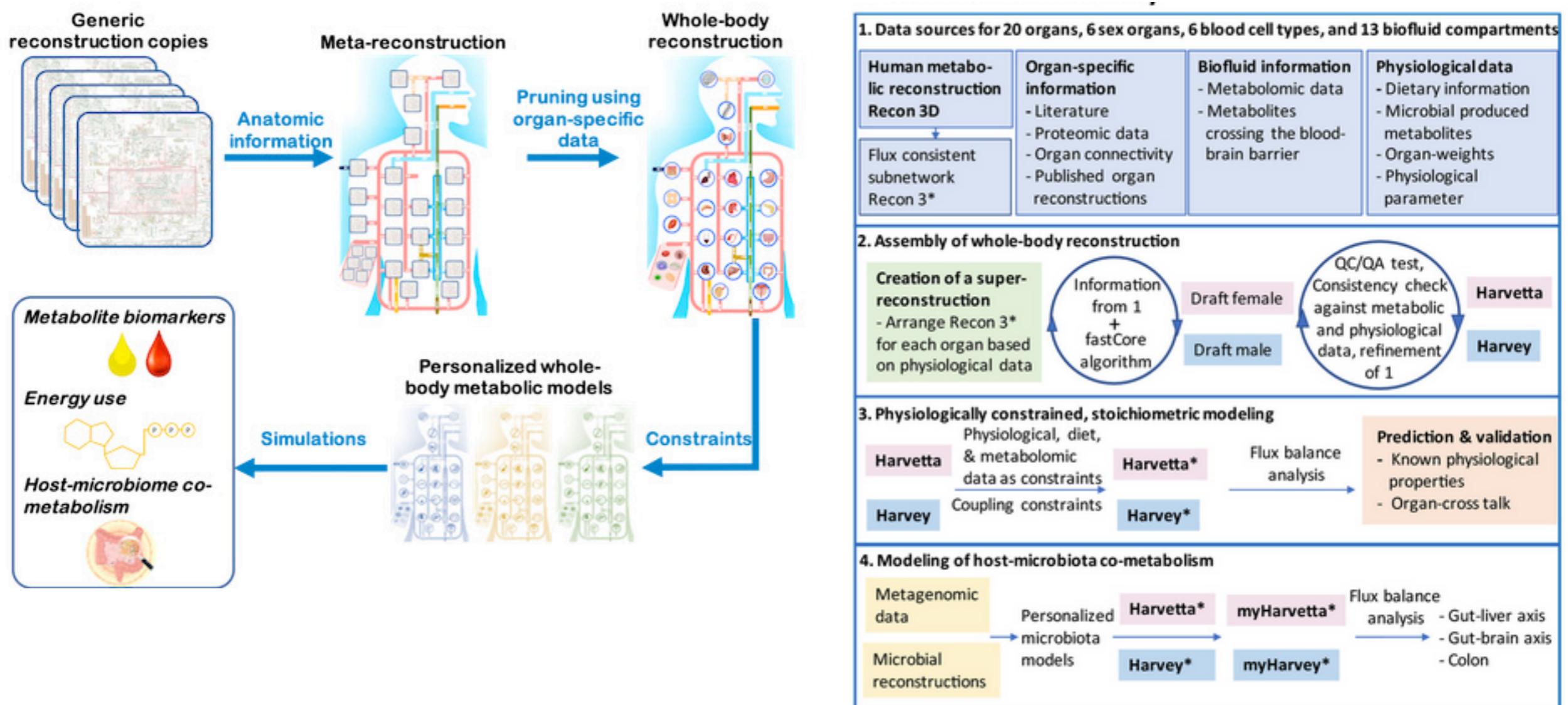
Reporter metabolites are useful tools for identifying key metabolites linked with transcriptional changes



Varemo 2013

GEMs in healthy and disease profiling

Personalised whole-body models of host + gut microbiome 26 organs
6 blood cell types
80,000 biochemical reactions



Thiele 2020

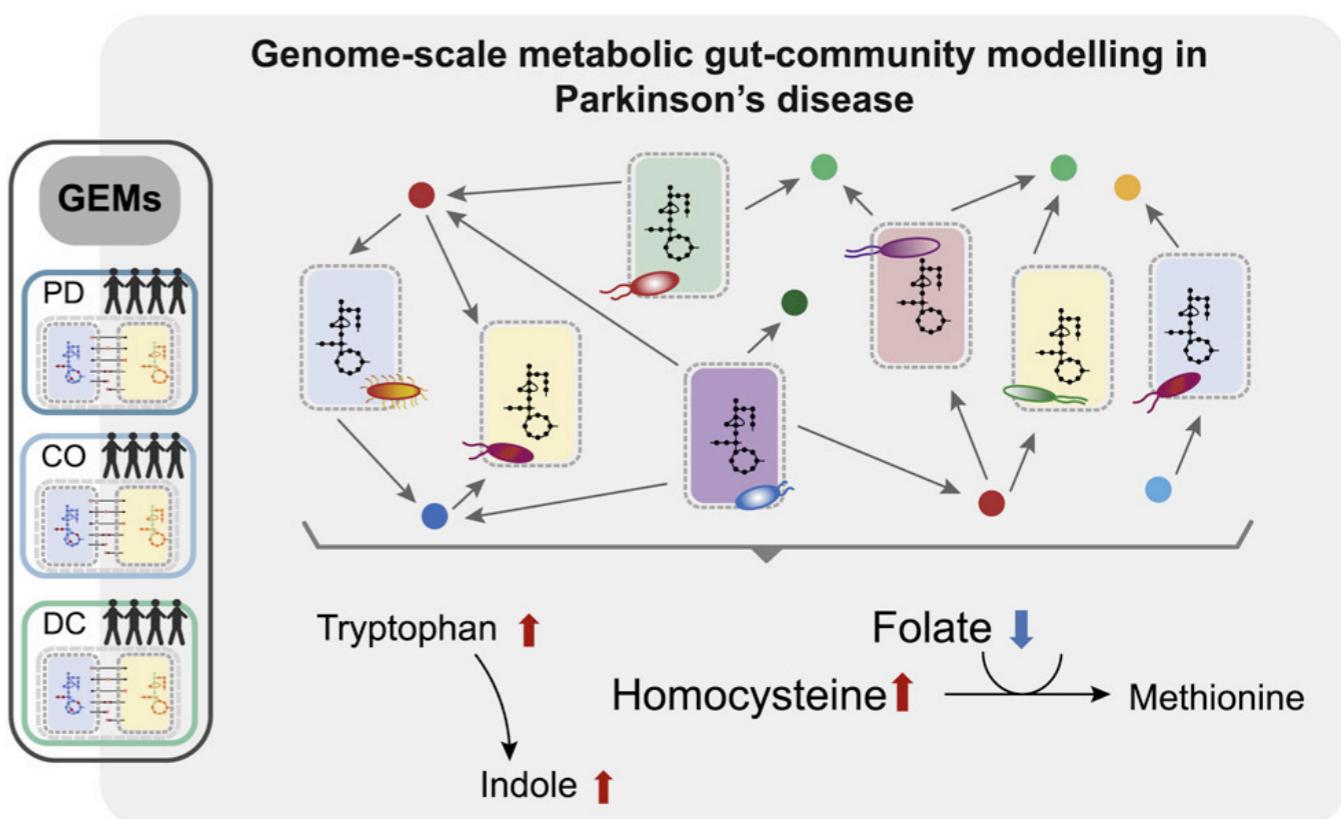
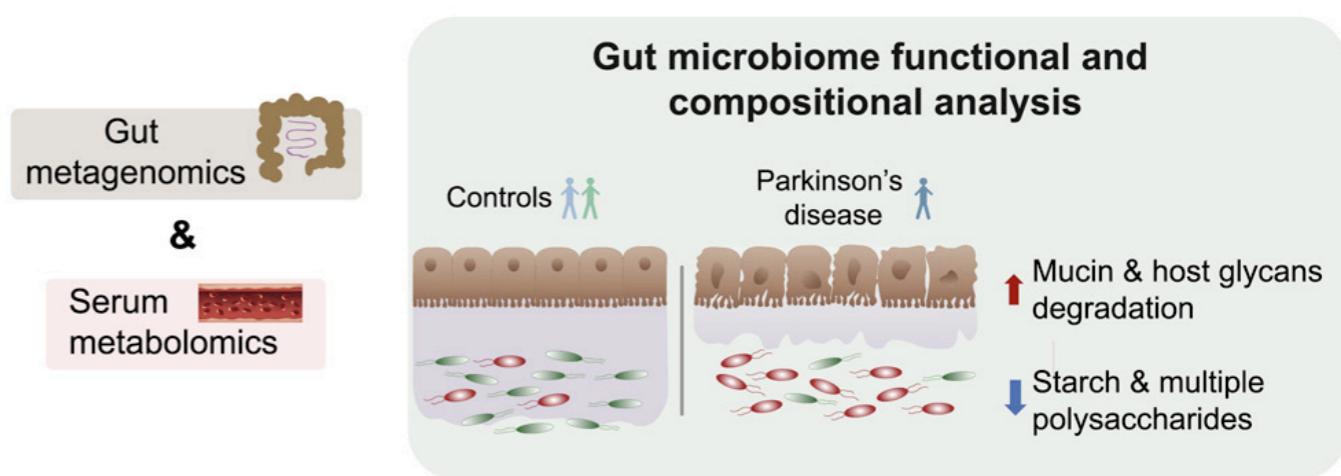
Integrative host + gut GEMs in Parkinson's disease

Cohort with PD, healthy controls (CO) and diseased controls (DC)

Microbial GEMs + Intestinal lumen

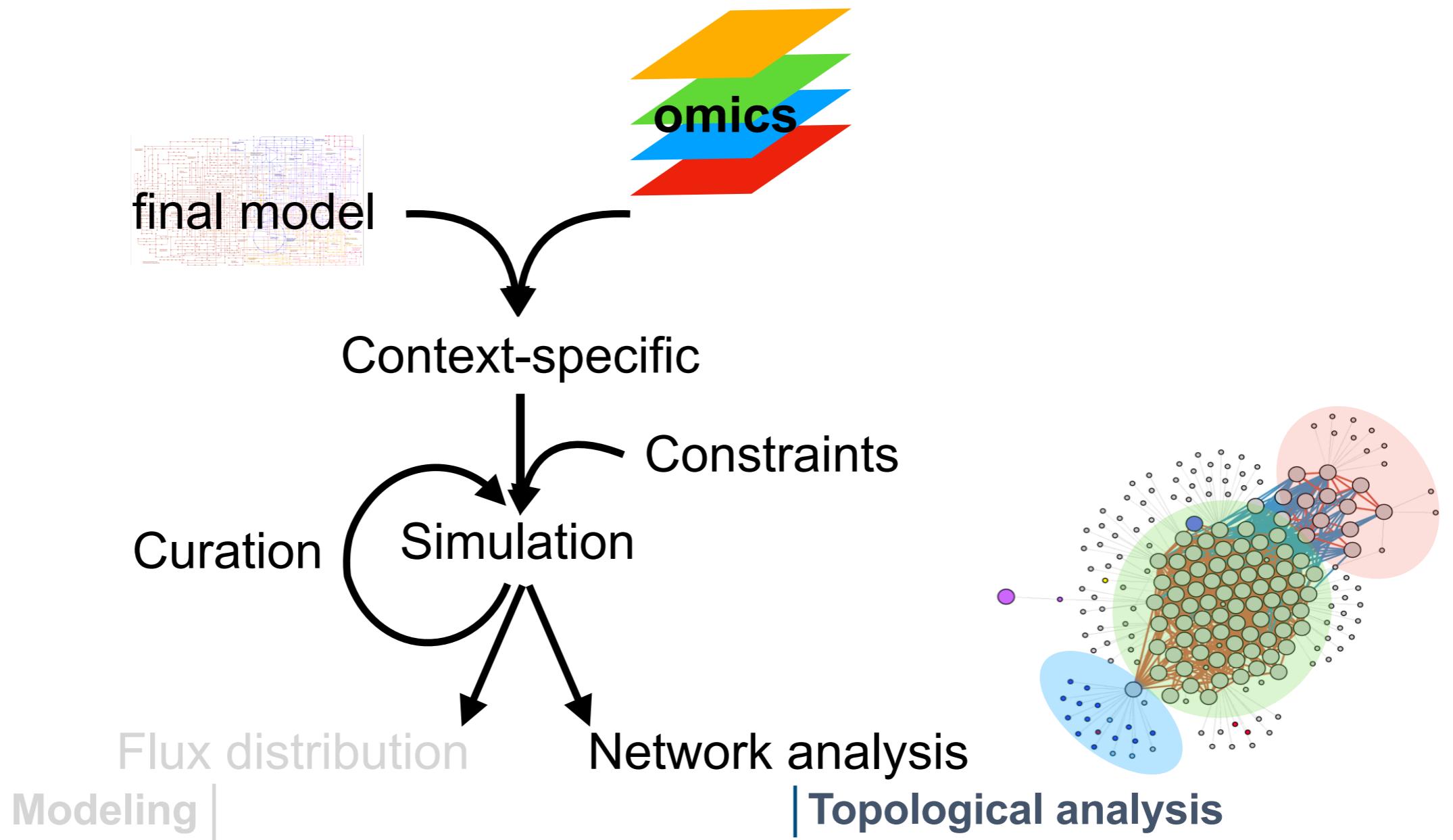
Personalised community GEMs are modelled and analysed

Substantial imbalances in bacterial mucin and host catabolic enzymes associated with PD



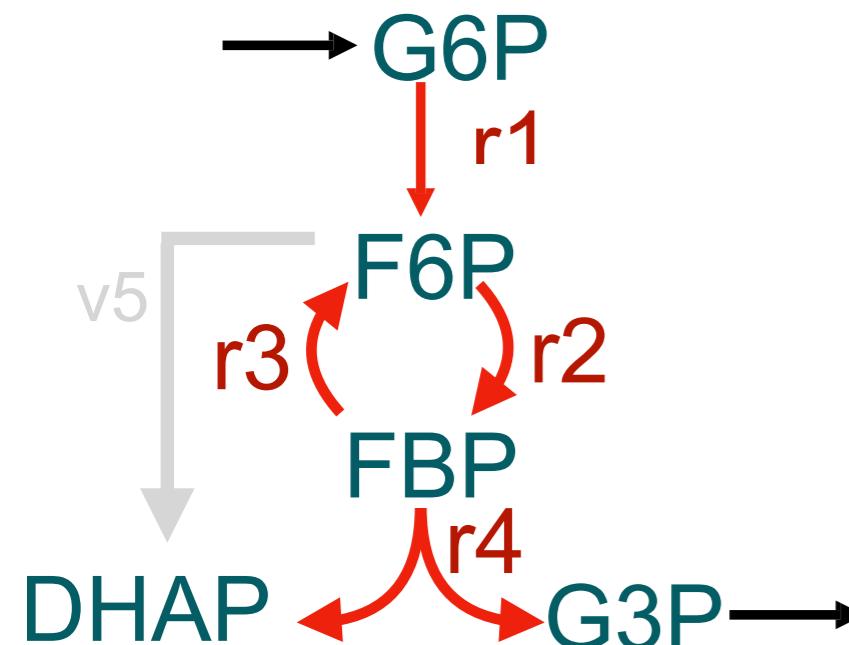
Rosario 2021

Approach for analysis

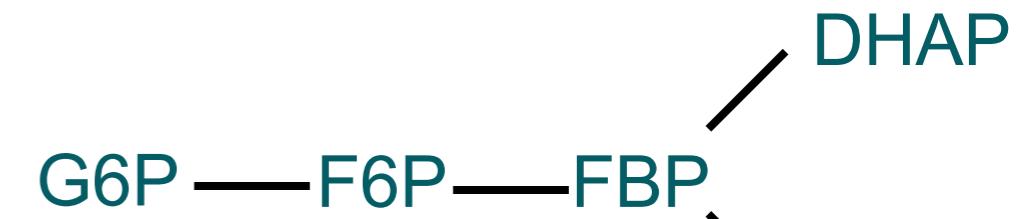


Approach for analysis

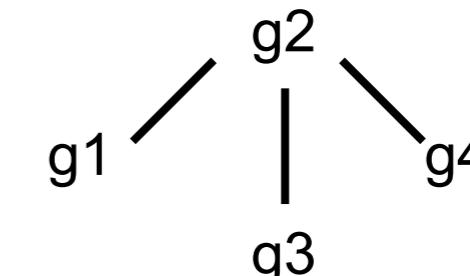
Networks from GEMs



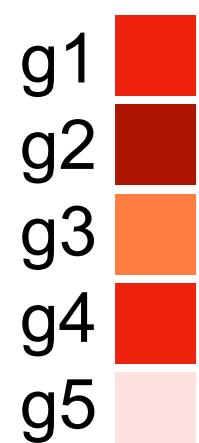
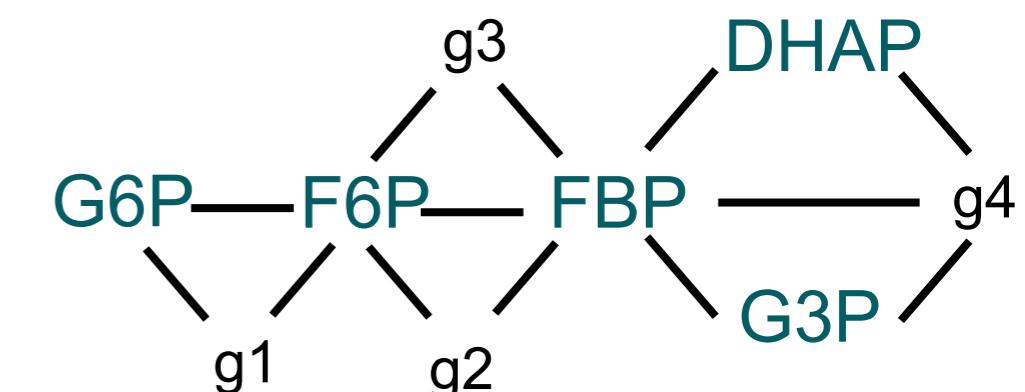
met-met



gene-gene



gene-met



Important to exclude uninformative metabolites (e.g. cofactors, H₂O)

Personalised GEMs + Topology analysis highlight disruption of central metabolism associated with SARS-CoV-2 severity

RNAseq + Metabolomics +

...

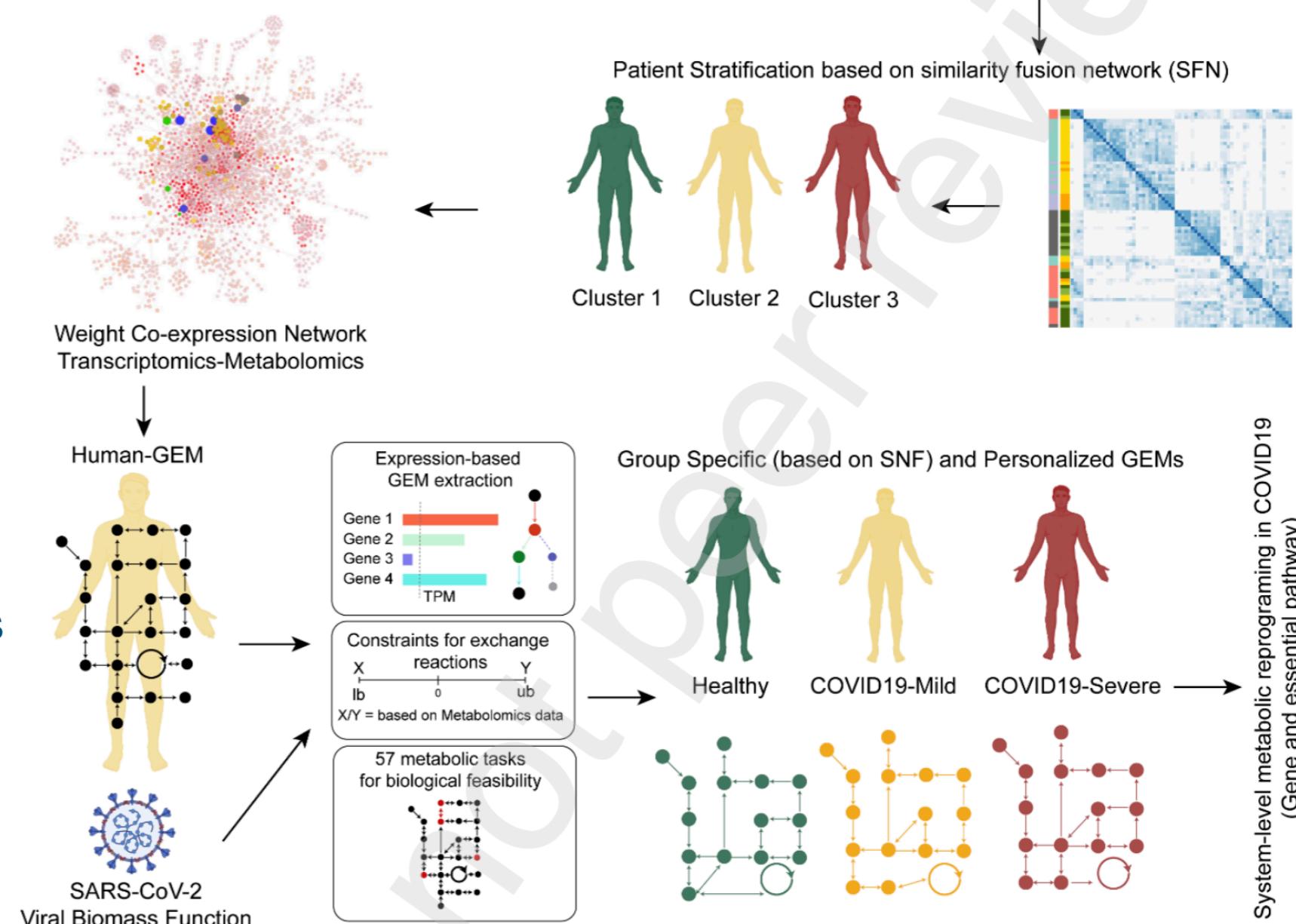
Data-driven patient stratification (SFN)

Integrated network analysis for characterisation and feature prioritisation

Human+viral personalised and subgroup-specific GEMs

GEM -> Topology analysis show altered metabolism in key transporters and central metabolites

Healthy + Mild + Severe + Convalescent subjects



System-level metabolic reprogramming in COVID19 (Gene and essential pathway)

Ambikan 2022 (pre-print)

Conclusions

Genome-scale metabolic models are solid frameworks for integration of multi-omic data, for single individuals or multiple species

GEMs may be specifically constructed for patient subgroups or personalised levels

Elucidate mechanistic alterations and key metabolic effectors associated with pathology

A combination of metabolic modelling and topology analysis provides systematic clues not easily elucidated by the individual approaches

Special thanks to Jonathan Robinson for some of the slides

metabolic
ATLAS



NBS



SciLifeLab

