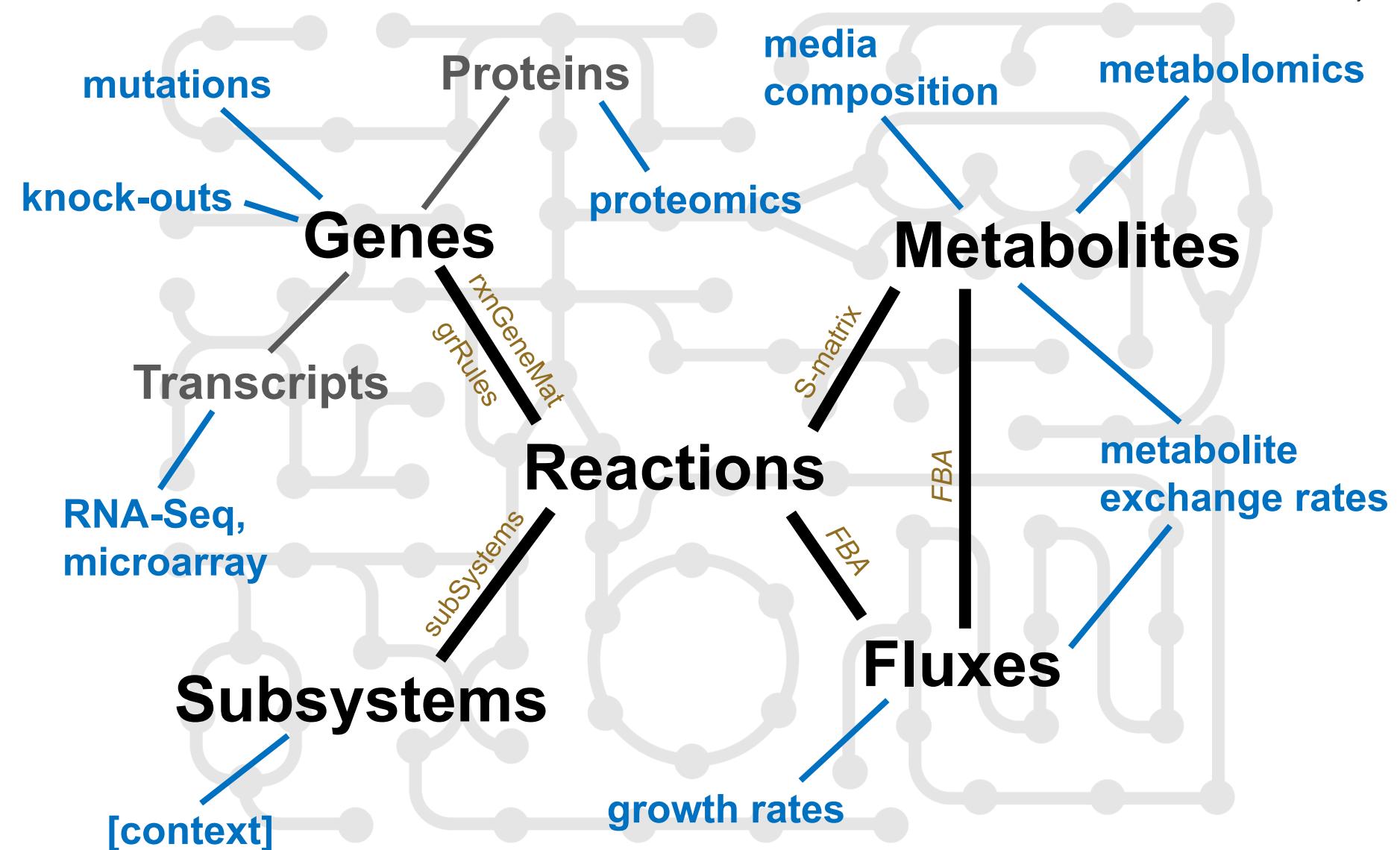


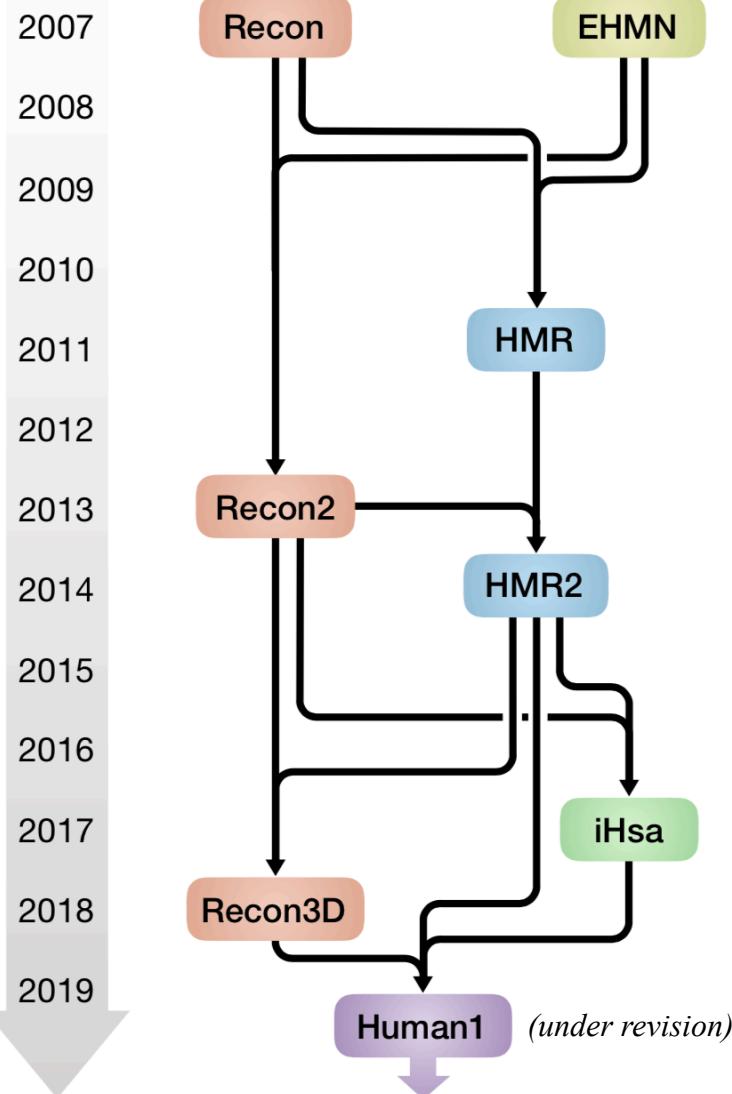
# Context specific analysis in metabolic modeling

Jonathan Robinson

Division of Systems and Synthetic Biology  
Department of Biology and Biological Engineering  
Chalmers University of Technology

Omics Integration and Systems Biology  
SciLifeLab, Stockholm, 2019-09-11





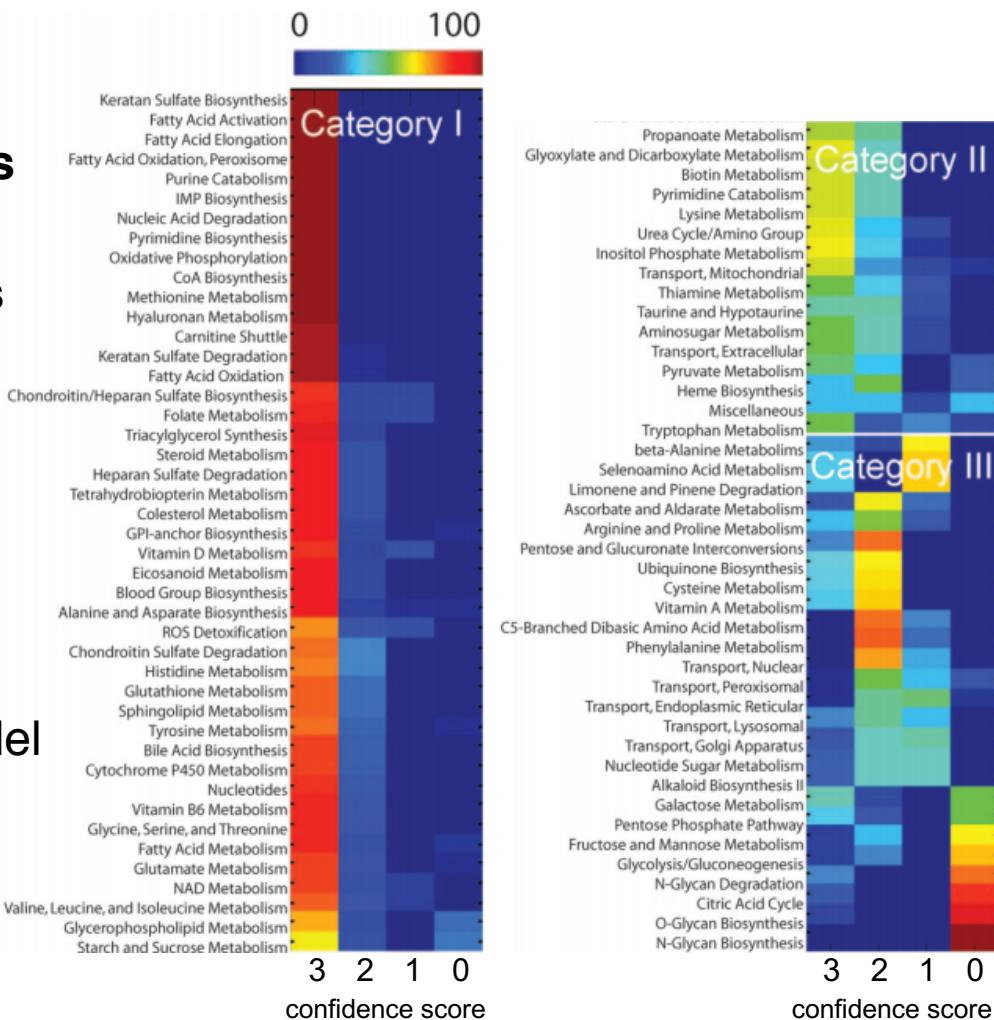
## Genome-scale models of human metabolism

- Began with Recon1 and EHMN (Edinburgh human metabolic network)
- Followed by the first generation of the Human Metabolic Reaction (HMR) model
- A few years later new versions Recon2 and HMR2 were published
- Most recent published model is Recon3D

## Recon1

N.C. Duarte, et al. *PNAS* 2007

- Included intracellular **compartments** and exchange
- References and **confidence scores** were provided for each model component
- Highlighted the large differences in characterization of each pathway
  - Category I, II, and III
- Integrated transcriptomic data from gastric bypass patients with the model
  - Gene fold-changes before/after surgery
  - Mapped to network and **visually** identified regions of coordinated expression change

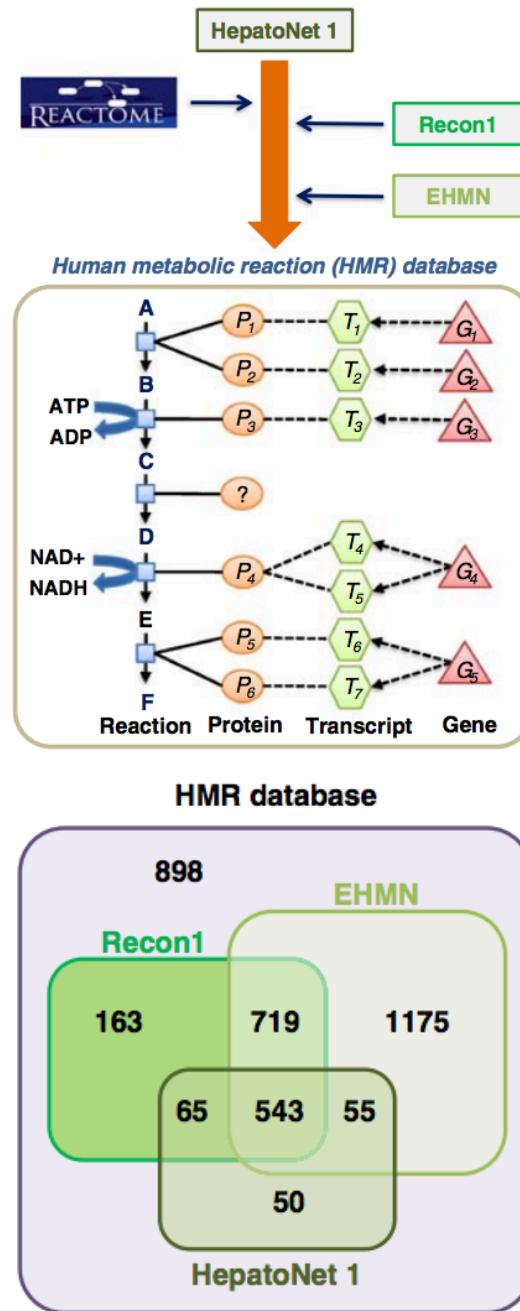


## HMR (Human Metabolic Reaction) model

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

A. Mardinoglu, et al. *Mol Syst Biol* 2013

- Initially formulated as more of a database than a model
- Merged Recon1 and EHMN with other databases (HumanCyc and KEGG)
- Focused on metabolites and reactions with standard identifiers (KEGG, InChI, etc.)
- HMR was integrated with healthy tissue and cancer proteomics and transcriptomics to generate tissue- and cancer-specific models
  - Developed the INIT algorithm to perform the omics data integration  
*(we will get to this later...)*



## Recon2

I. Thiele, et al. *Nat Biotechnol* 2013

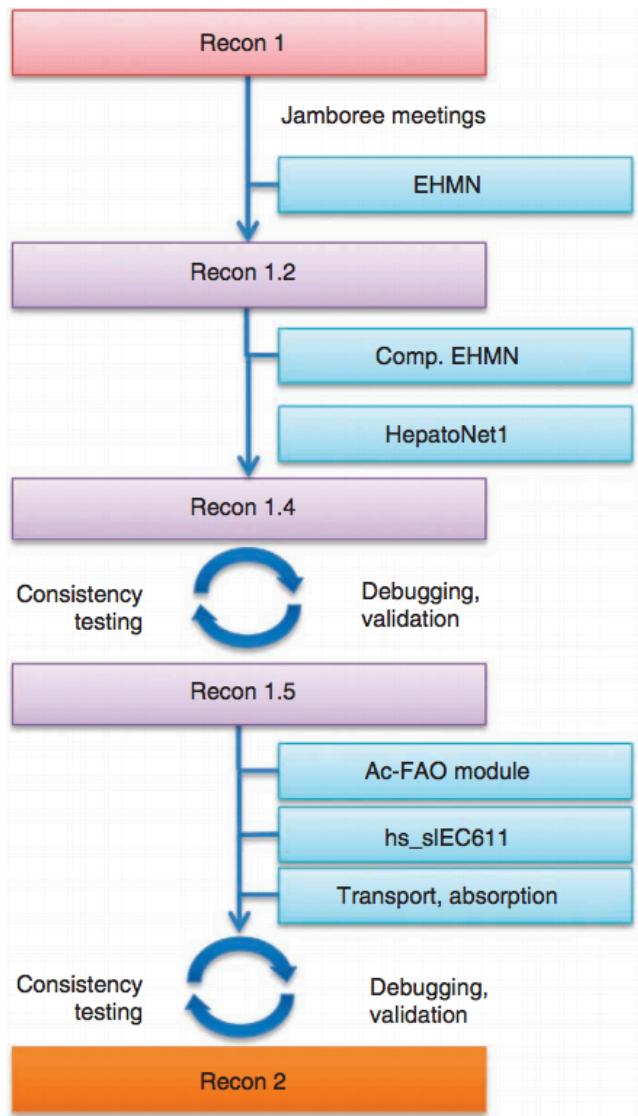
- Aimed to develop a consensus reconstruction, combining a few previous models
- Used the model to predict biomarkers for inborn errors of metabolism (IEM)
  - Constrained reactions catalyzed by affected enzymes and identified significantly altered exchange reaction fluxes
  - Recon2 outperformed Recon1
- Generated 65 cell type-specific GEMs using HPA expression data (with iMAT)
  - Compared structures (reaction content)
  - 25% of the models could generate biomass

<i>In silico</i>		
	Up	Down
Up	24	1
Down	16	5

Accuracy = 63%  
 $P = 0.054$

<i>In vivo</i>		
	Up	Down
Up	66	5
Down	18	10

Accuracy = 77%  
 $P = 7.9 \cdot 10^{-4}$



## HMR 2.0 database

## Literature based GEMs

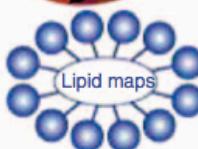
## Generic human GEMs

- iHuman1512
- Recon 1
- Edinburgh model (EHMN)

## Cell type specific GEM

- *iAdipocytes1809*
- HepatoNET 1

## Pathway / process databases



**HUMANCYC**  
A member of the bioCyc  
database collection



## HMR2 (Human Metabolic Reaction) model

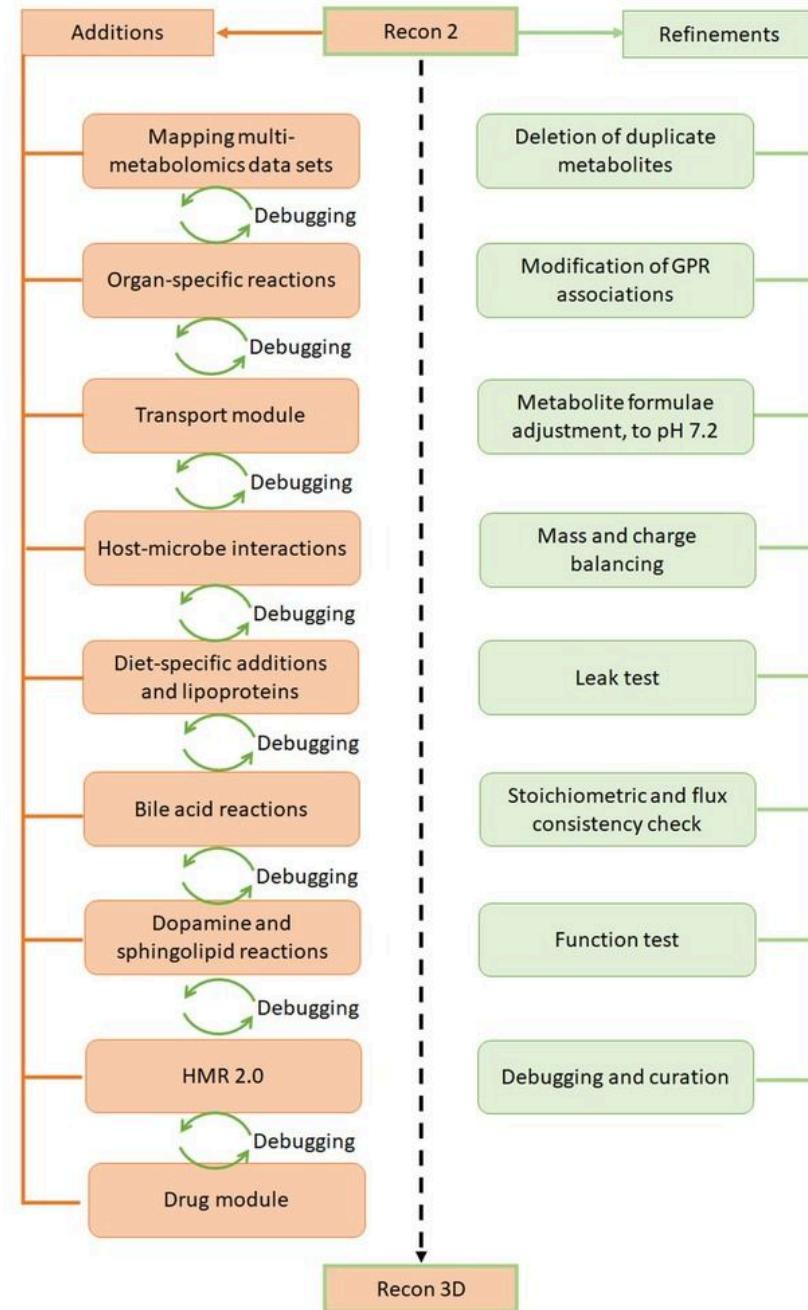
A. Mardinoglu, et al. *Nat Commun* 2014

- Incorporated extensive lipid metabolism
- Improved reaction-gene associations
  - However, all genes are still assumed to encode isozymes for their associated reactions
- HMDB, Lipid Map, KEGG, and ChEBI identifiers were assigned to metabolites
- KEGG IDs and EC numbers were assigned to reactions
- Also included genes and reactions in Recon2

## Recon3D

E. Brunk, et al. *Nat Biotechnol* 2018

- Expanded Recon2 by incorporating other models/networks
  - e.g., HMR2 and drug metabolism
- Curated and fixed many errors present in Recon2
- Added 3D metabolite and protein structure data
- A separate “database version” and “model version” exist
  - The database version contains all the reactions and information, but is not properly balanced.
  - The model version is suitable for simulation purposes (e.g., FBA).



# On the size and complexity of GEMs

In early 1982, the Lisa software team was trying to buckle down for the big push to ship the software within the next six months. Some of the **managers decided to track the progress of each individual engineer in terms of the amount of code that they wrote.** They devised a form that each engineer was required to submit every Friday, which included a field for the number of lines of code written that week.

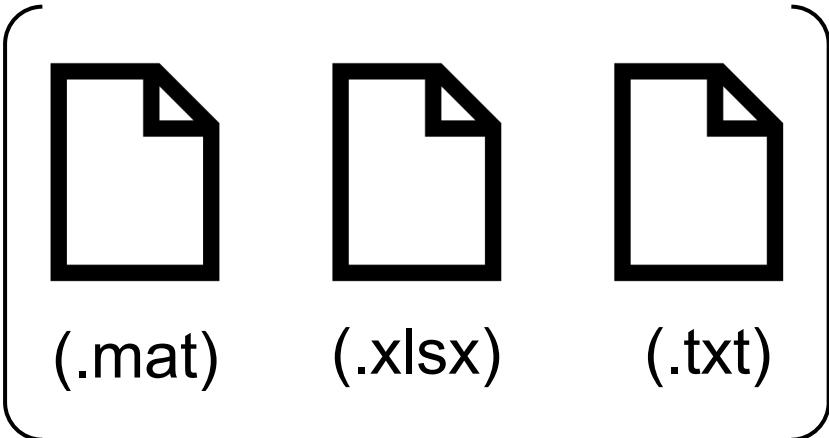


Steve Jobs and Bill Atkinson posing prior to the launch of the Macintosh. source: cnet.com

Bill Atkinson was working on optimizing Quickdraw's region calculation machinery, and had completely rewritten the region engine using a **simpler algorithm** which made operations almost **six times faster**. The rewrite also **saved around 2,000 lines** of code. When he got to the lines of code part, he thought about it for a second, and then wrote in the number: -2000.

They stopped asking Bill to fill out the form.

Model published  
and distributed as:



These are static files with no visible history or capacity for change

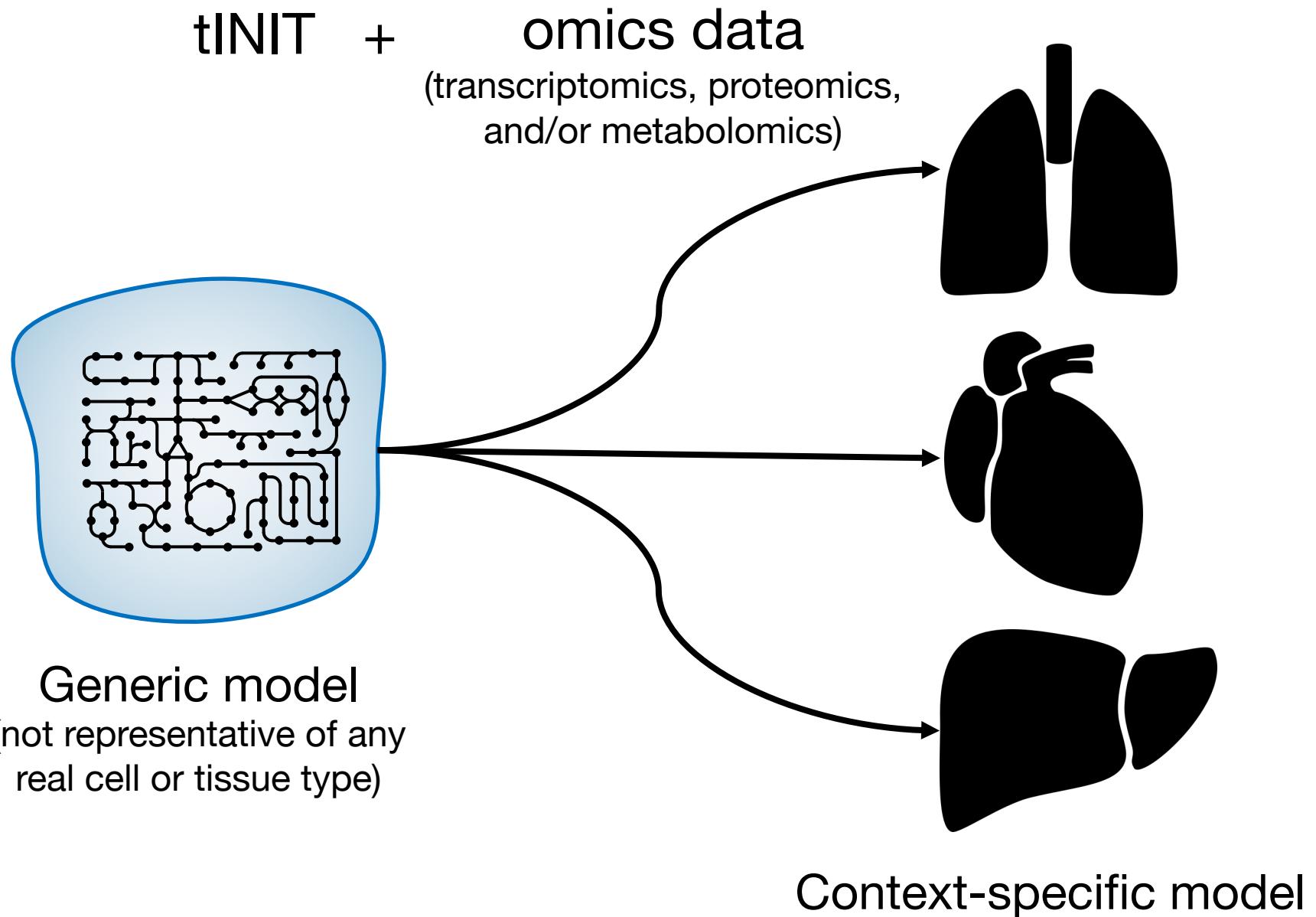
- What if I find an error?
- How do I determine when and why a change was made?
- How do I make suggested changes/improvements?



A repository has a transparent **history**, can **continue to develop**, and is designed to facilitate **collaboration**

<https://github.com/SysBioChalmers/Human-GEM>

Example: HMR\_5407: CO<sub>2</sub> + H<sub>2</sub>O <=> carbonate, in “Nitrogen metabolism” subsystem



# 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)

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

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

$$|v_i| \leq 1000y_i$$

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

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

$$b_j \leq 1000x_i$$

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

$$b_j \geq 0$$

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

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

---

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

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  
dTTP de novo synthesis

### Uptake of essential amino acids

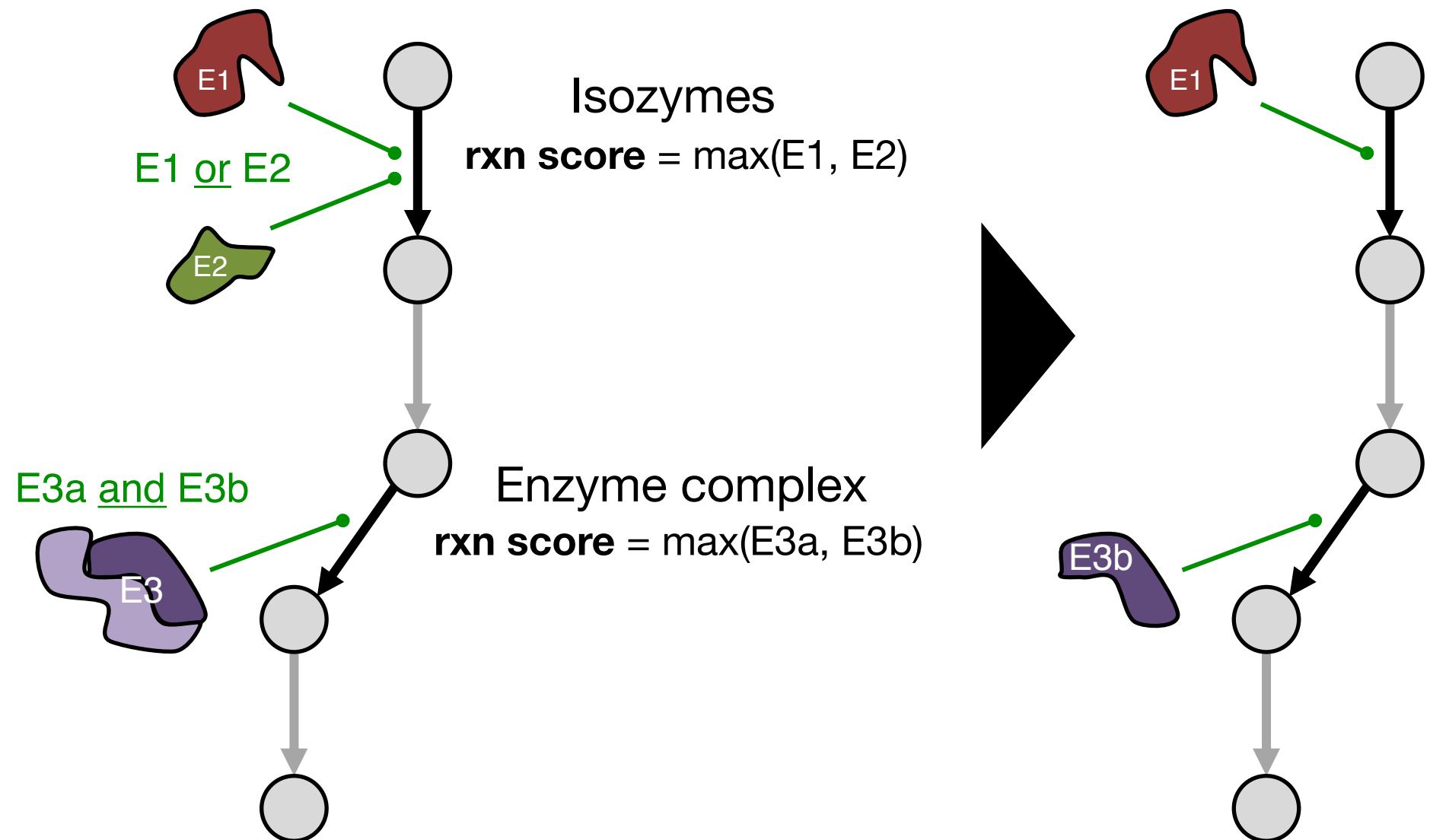
Histidine uptake  
Isoleucine uptake  
Leucine uptake  
Lysine uptake  
Methionine uptake  
Phenylalanine uptake  
Threonine uptake  
Tryptophan uptake  
Valine uptake

### De novo synthesis of key intermediates

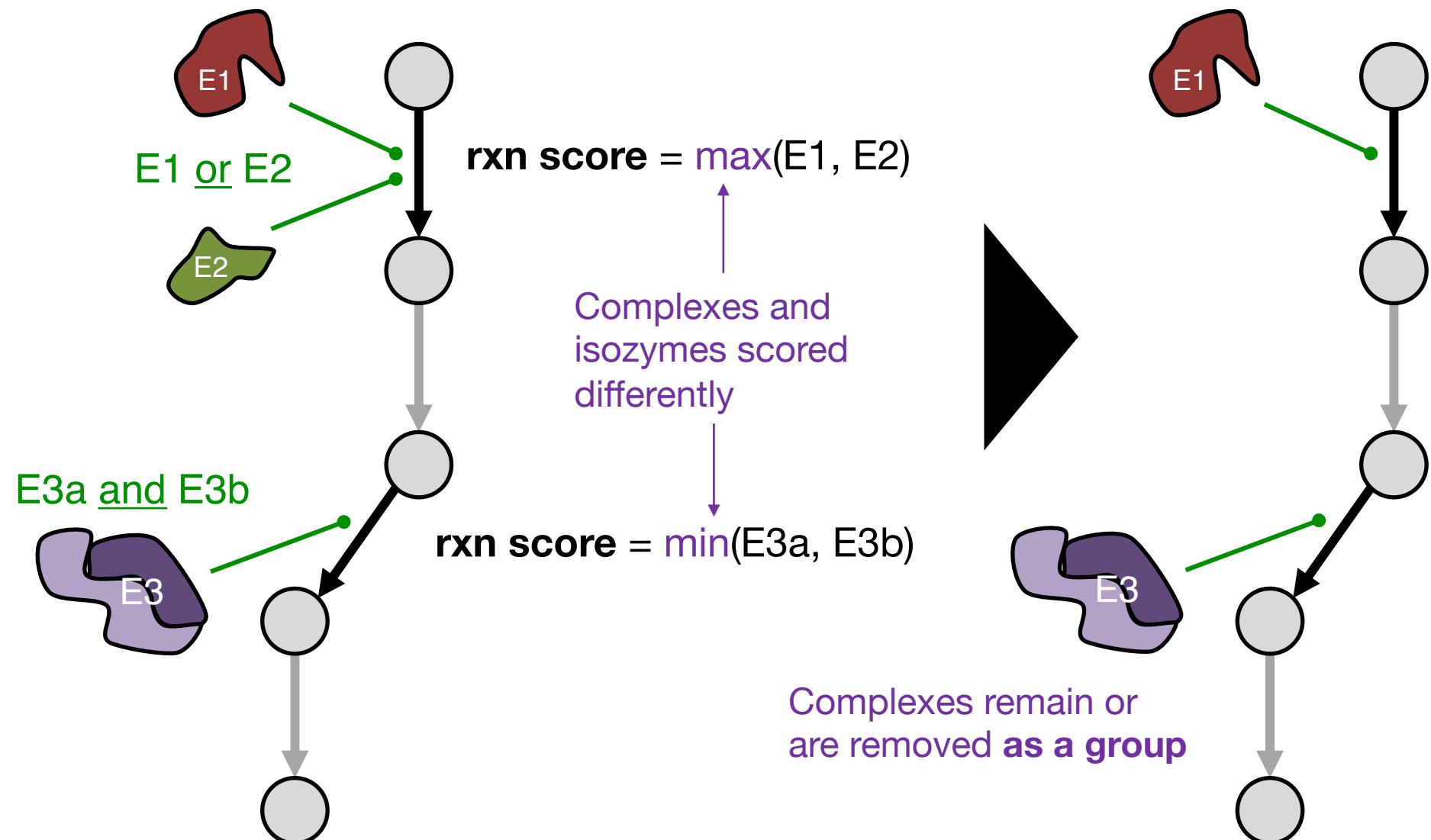
Glycerate 3-phosphate de novo synthesis  
Mitochondrial acetyl-CoA de novo synthesis  
Mitochondrial AKG de novo synthesis  
Erythrose 4-phosphate de novo synthesis  
Fructose 6-phosphate de novo synthesis

:

## Original approach



## New approach



## Original approach

**Ref. model rule:** (G1 and G2) or (G3 and G4)

Gene scores
G1: 3.2
G2: 1.0
<b>G3: -2.9</b>
G4: 6.7

$$\text{score} = \max(G1, G2, G3, G4) = 6.7$$

**tINIT model rule:** G1 or G2 or G4

New approach

**Ref. model rule:** (G1 and G2) or (G3 and G4)

Gene scores
G1: 3.2
G2: 1.0
G3: -2.9
G4: 6.7

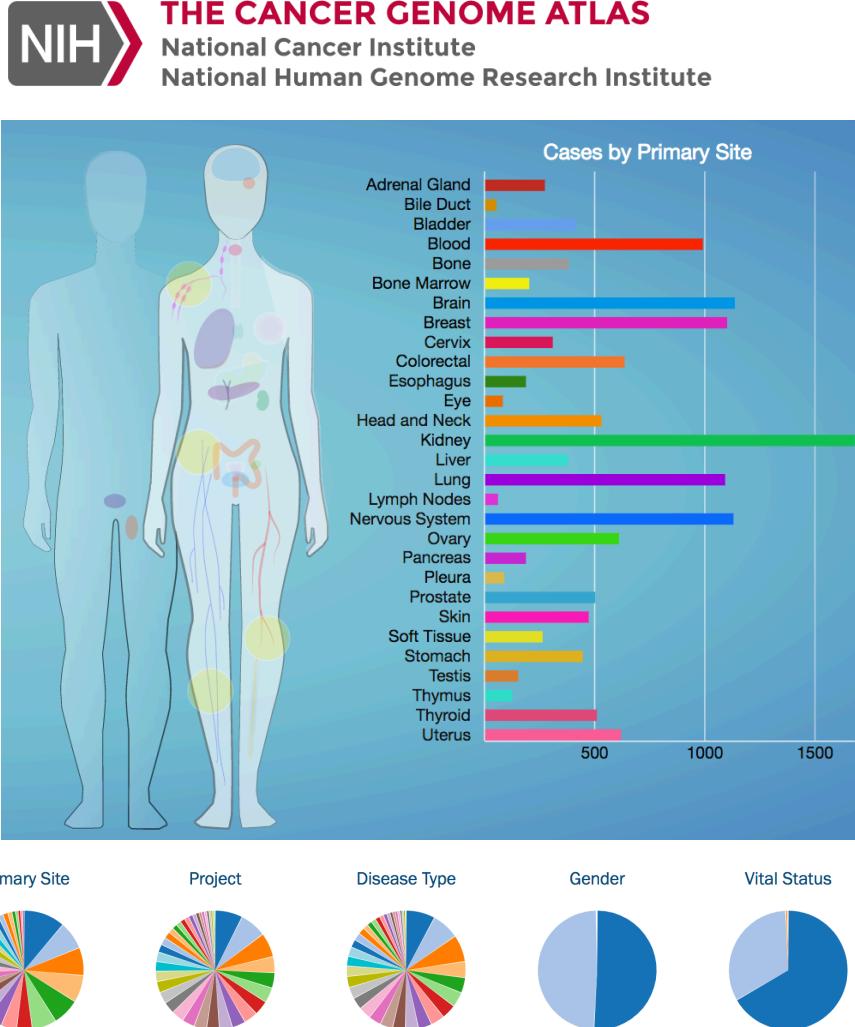
$$\text{score} = \max[ \min(G1, G2), \min(G3, G4) ] = 3.2$$

**tINIT model rule:** G1 and G2

# **Application:**

## GEM-based comparison of transcriptomes

# GEM-based comparison of transcriptomes



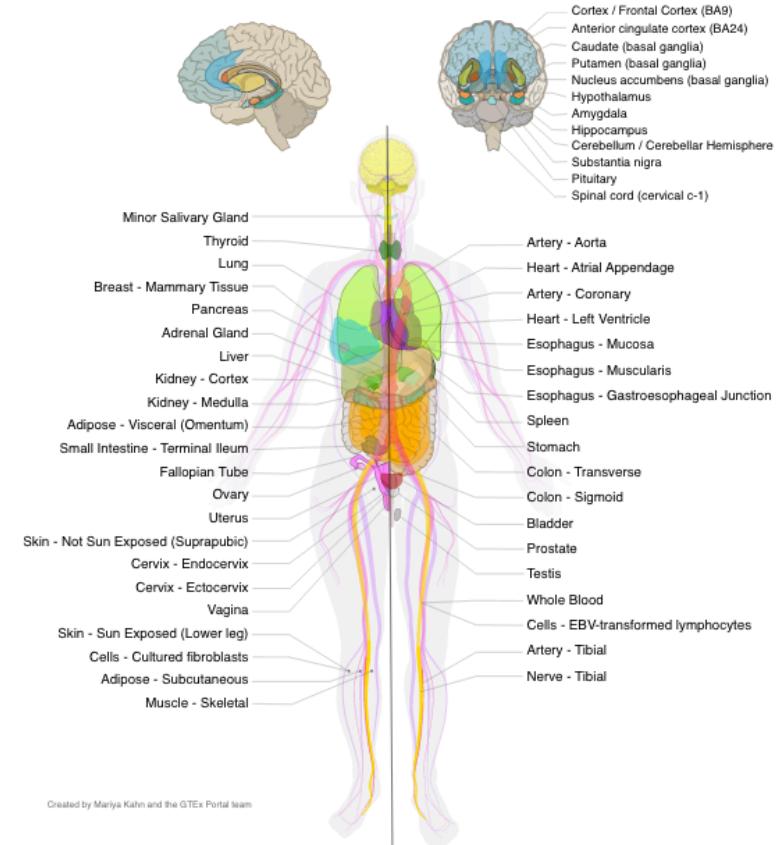
## GTEx Portal

Home Datasets Expression QTLs & Browsers Sample Data Do

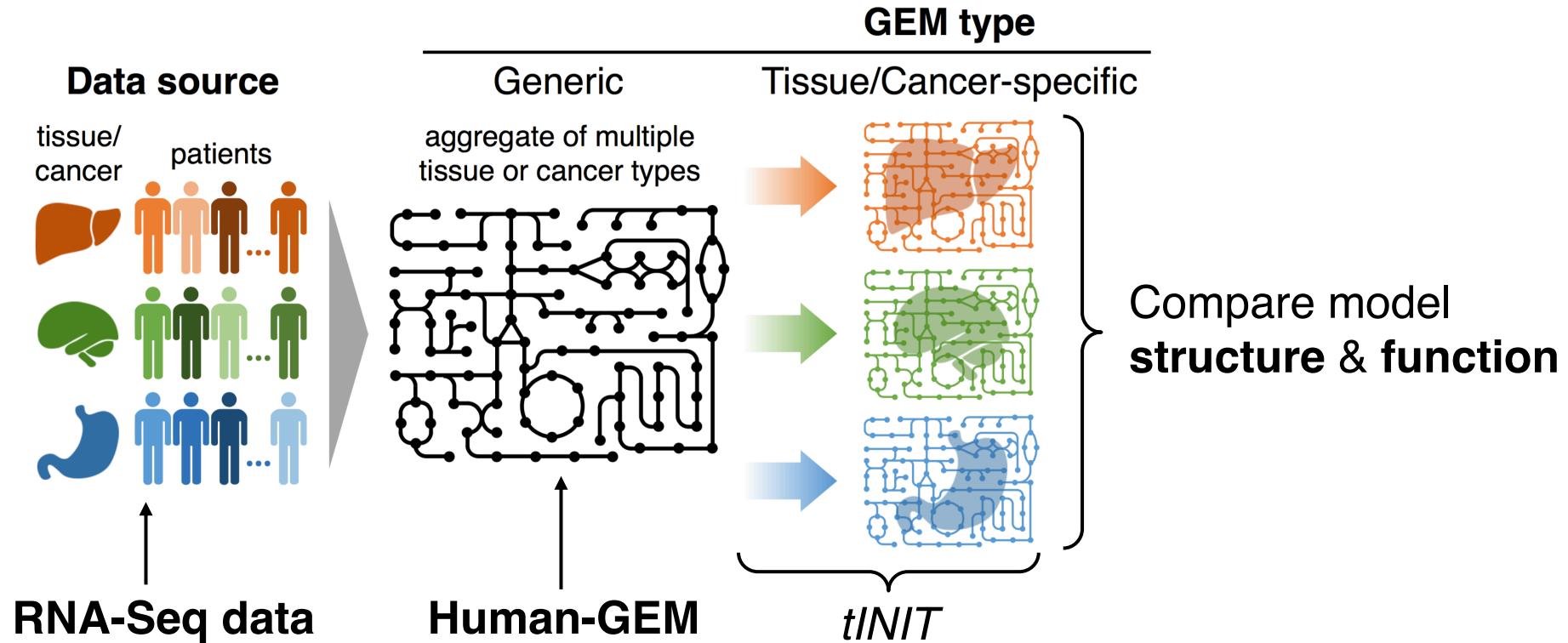
### Tissue Sampling Sites

This page provides a visual representation of the biospecimen source sites (BSSs) for the collection of tissue from postmortem/organ procurement cases for the Genotype-Tissue Expression (GTEx) project.

The full documentation on tissue collection procedures can be found on the [GTEx Tissue Harvesting Work Instruction](#).



**Objective:** To investigate healthy and tumor tissue transcriptomic differences *in the context of metabolism*



## RNA-Seq data

- TCGA
  - 33 tumor types
  - 23 paired-normal tissue types
- GTEx
  - 30 healthy tissue types

Comparison of model **structures**

Reaction	Lung Tumor	Lung Paired	Lung Healthy	Brain Tumor	Brain Paired
rxn1	1	0	1	1	1
rxn2	0	1	1	1	1
rxn3	0	0	0	0	0
rxn4	0	1	0	1	0
rxn5	1	1	0	1	1
rxn6	1	0	0	1	0
rxn7	0	0	1	1	0

...

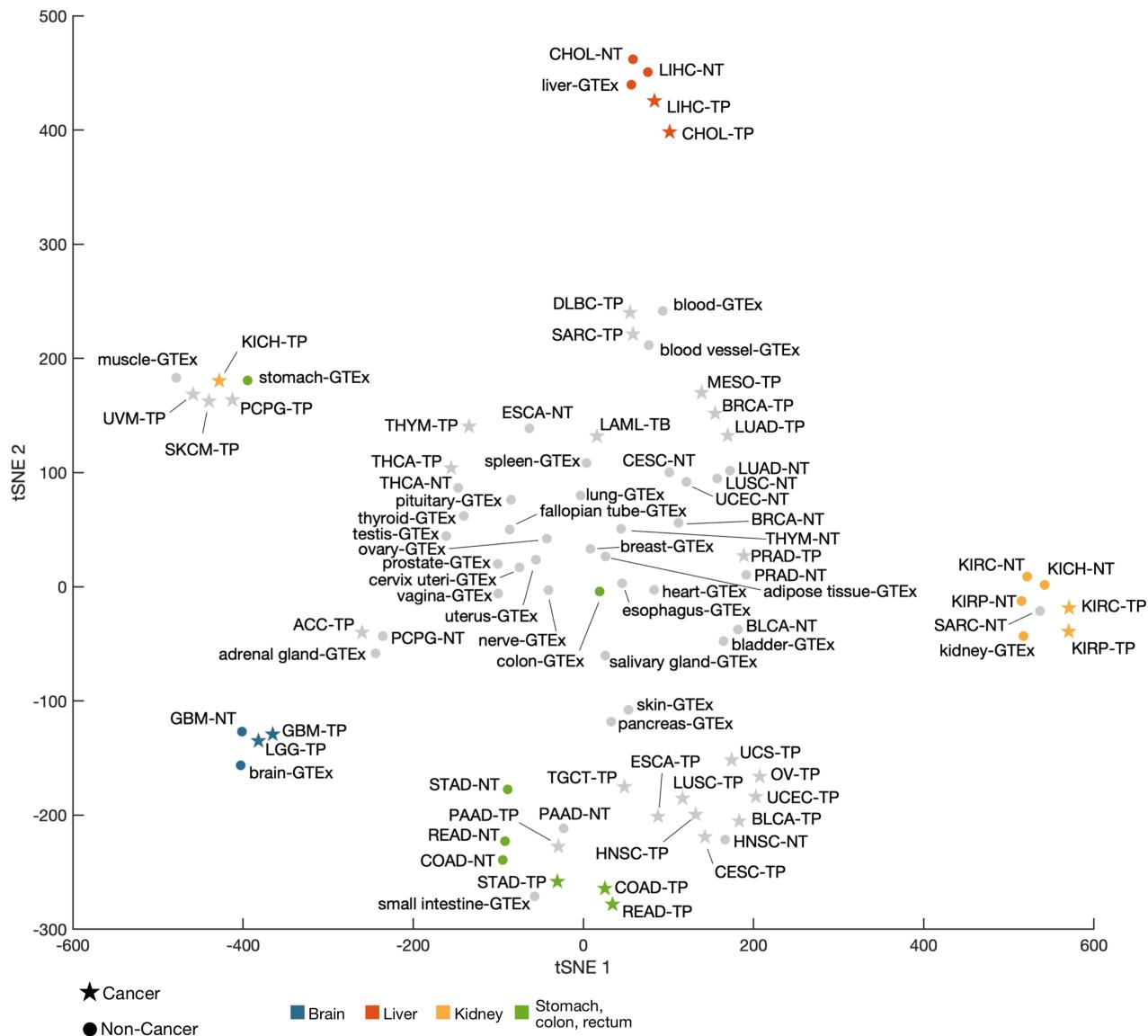
Model  
**contains**  
reaction

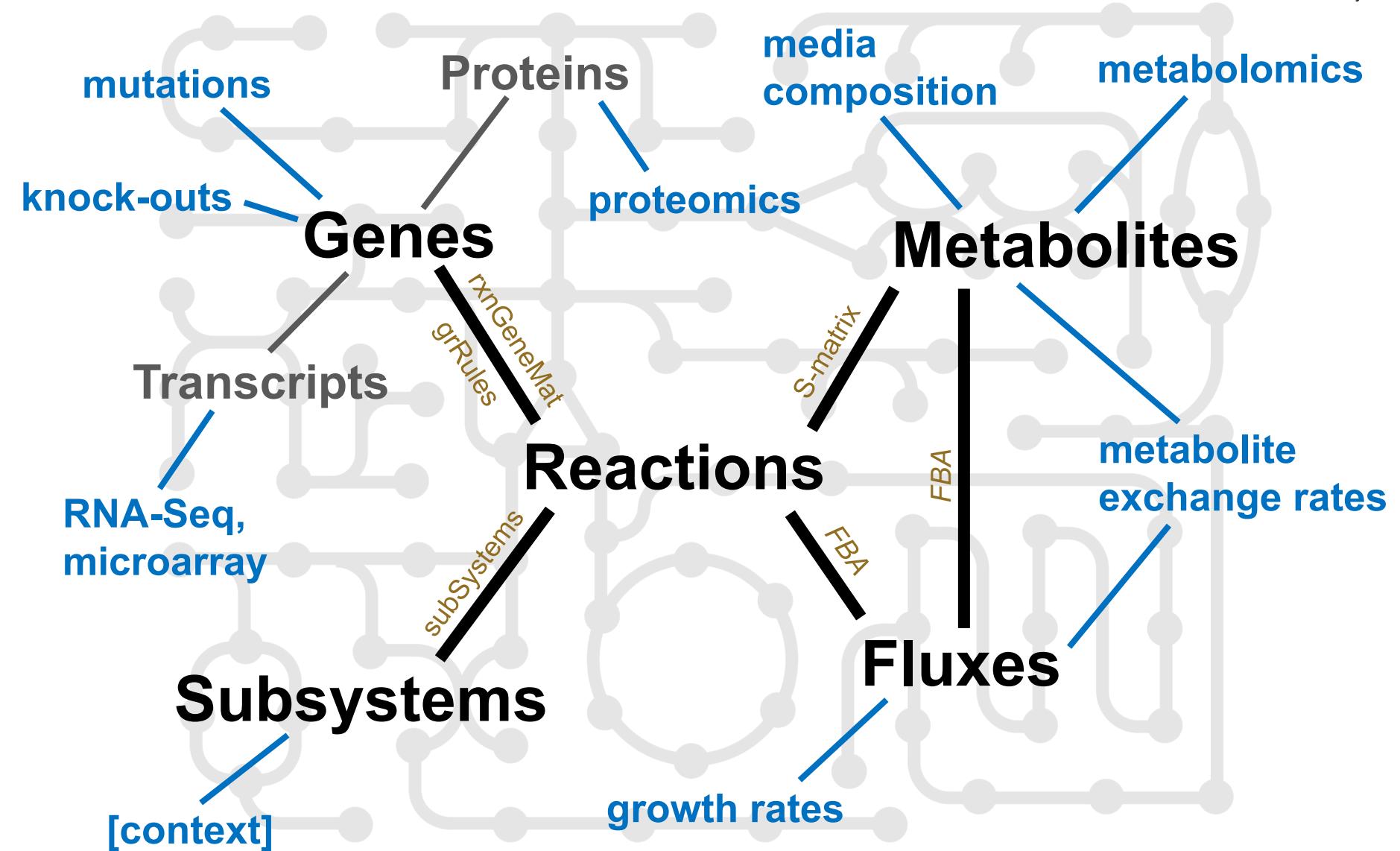
Model  
**missing**  
reaction

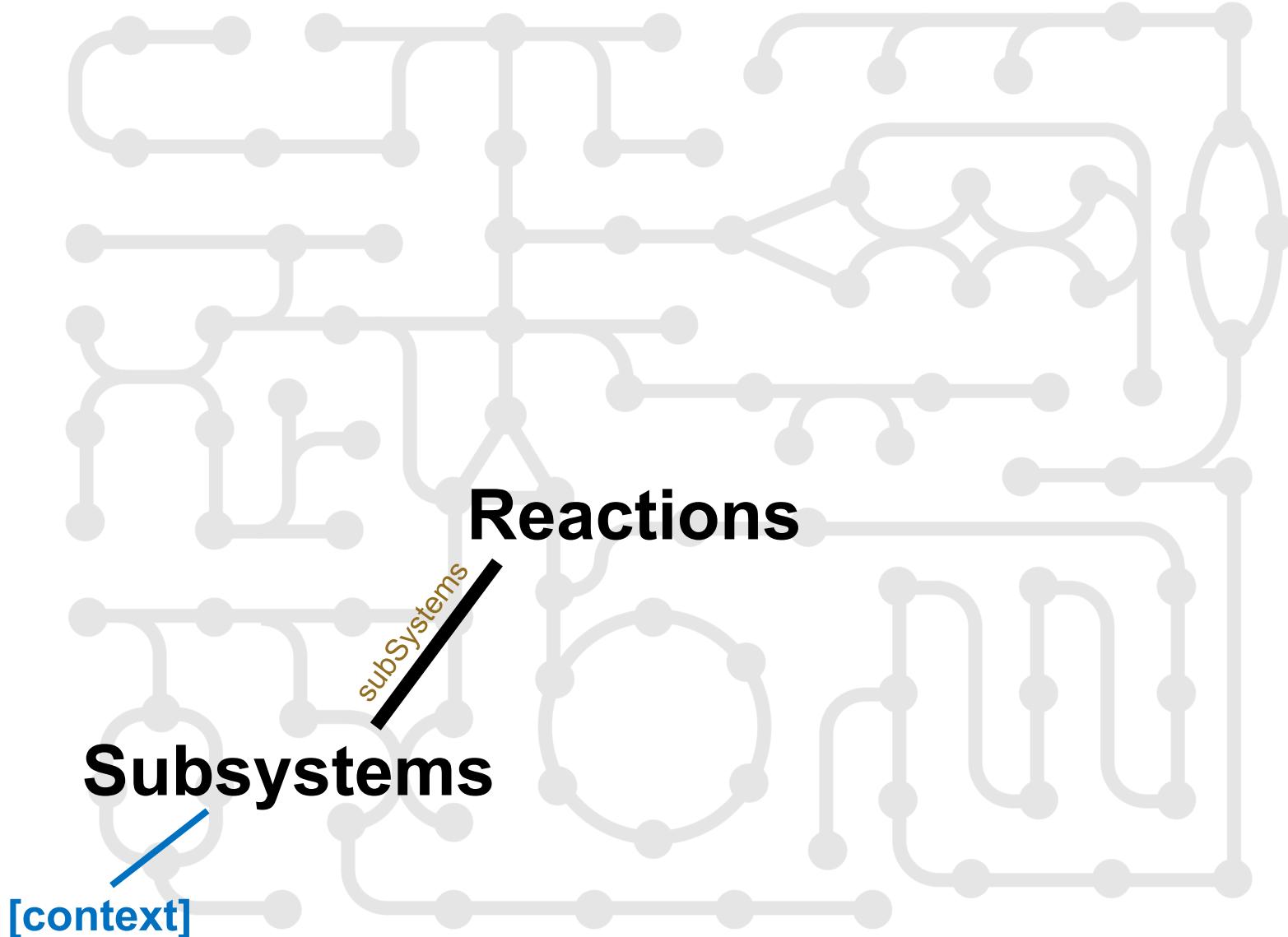
⋮

⋮

## tSNE of model reaction content matrix







## Comparison of model **structures**

Subsystem	Reaction	Lung Tumor	Lung Paired	Lung Healthy	Brain Tumor	Brain Paired	...
TCA cycle	rxn1	1	0	1	1	1	
TCA cycle	rxn2	0	1	1	1	1	
Glycolysis	rxn3	0	0	0	0	0	
TCA cycle	rxn4	0	1	0	1	0	
Fatty acid oxidation	rxn5	1	1	0	1	1	
Carnitine shuttle	rxn6	1	0	0	1	0	
Glycolysis	rxn7	0	0	1	1	0	

⋮

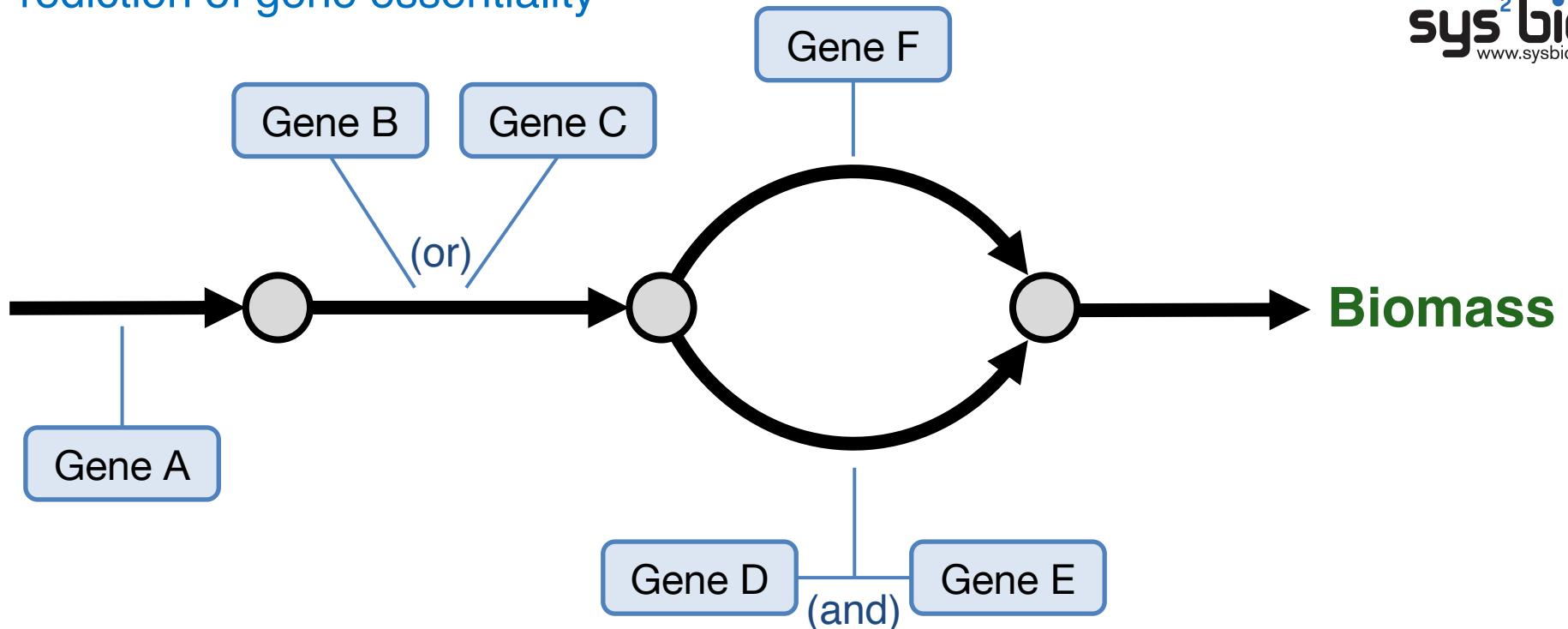
⋮



# **Application / Validation:**

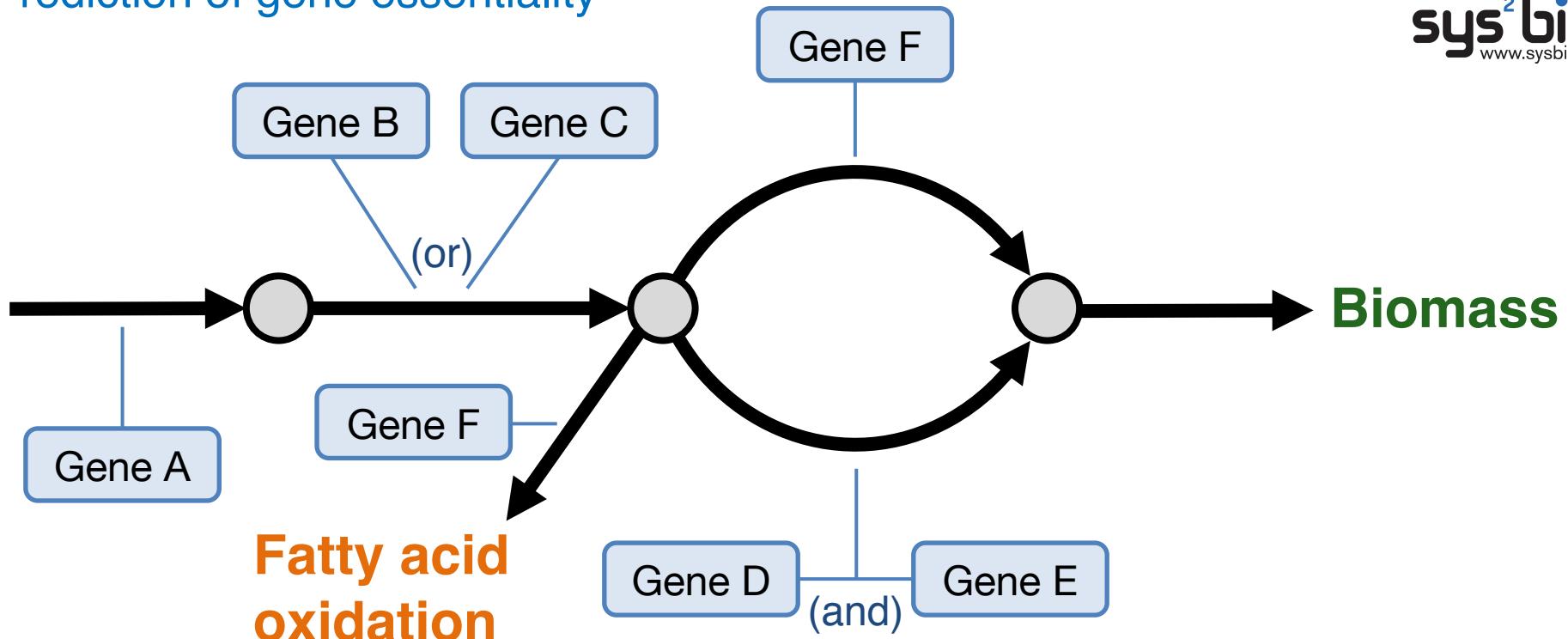
## Prediction of gene essentiality

# Prediction of gene essentiality



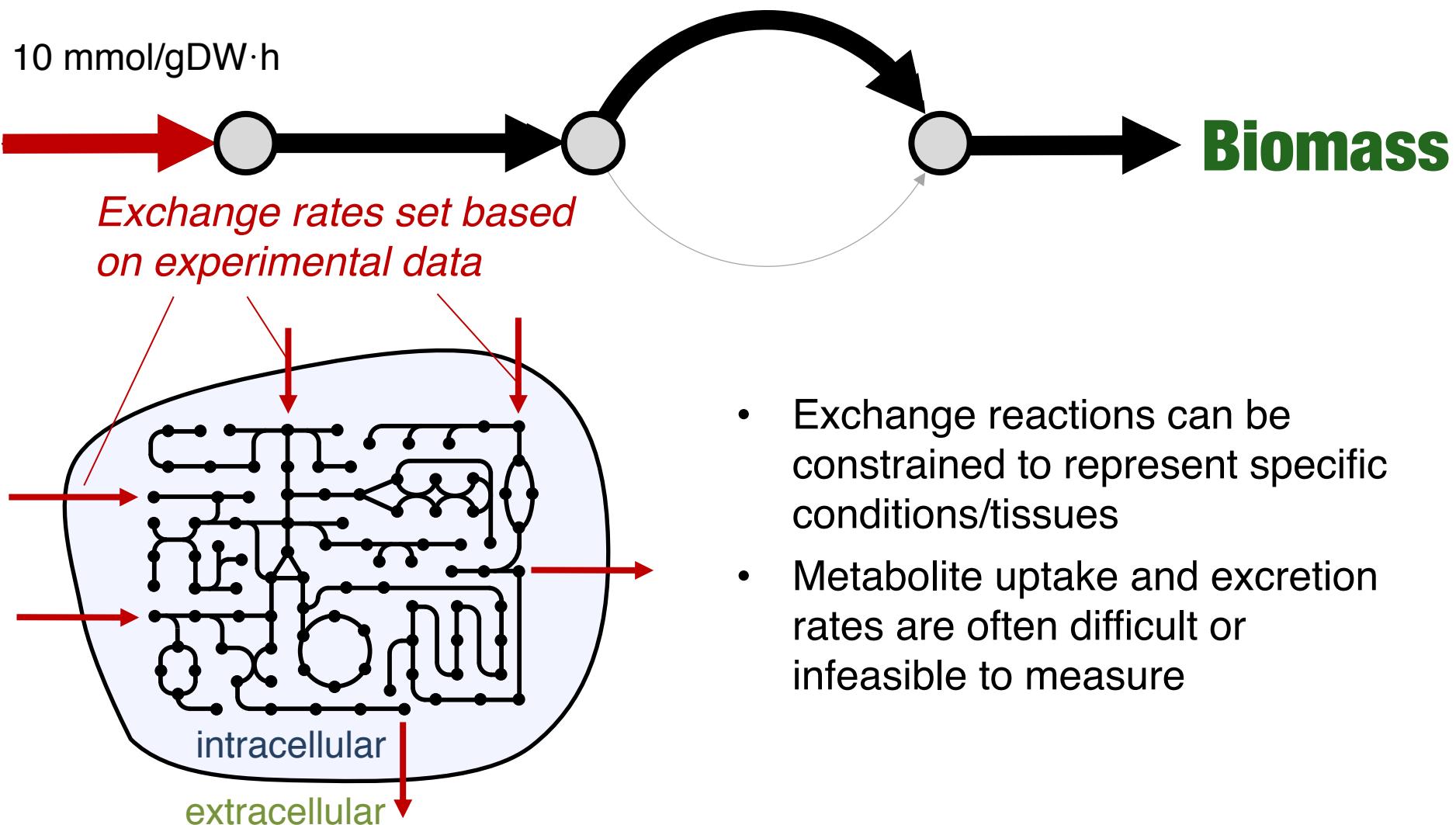
	Required for biomass production	Essential
Gene A	1	1
Gene B	0	0
Gene C	0	0
Gene D	0	0
Gene E	0	0
Gene F	0	0

# Prediction of gene essentiality



	Required for biomass production	Required for fatty acid oxidation	Essential
Gene A	1	1	1
Gene B	0	0	0
Gene C	0	0	0
Gene D	0	0	0
Gene E	0	0	0
Gene F	0	1	1

How do we set the flux bounds for our model?

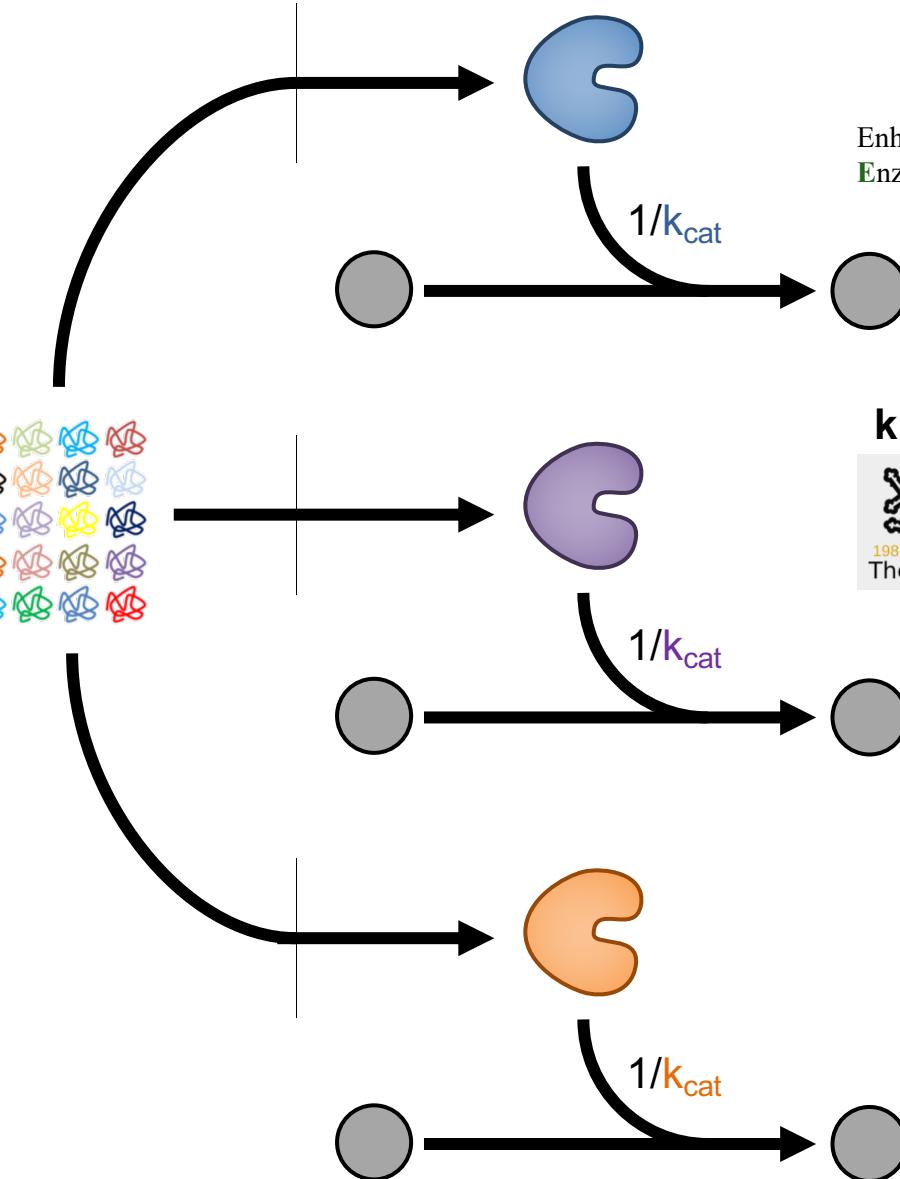
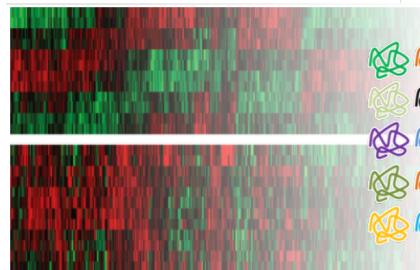


# Enzyme-constrained Human-GEM



Enhancing a Genome-scale model to account for Enzyme Constraints, using Kinetics and Omics

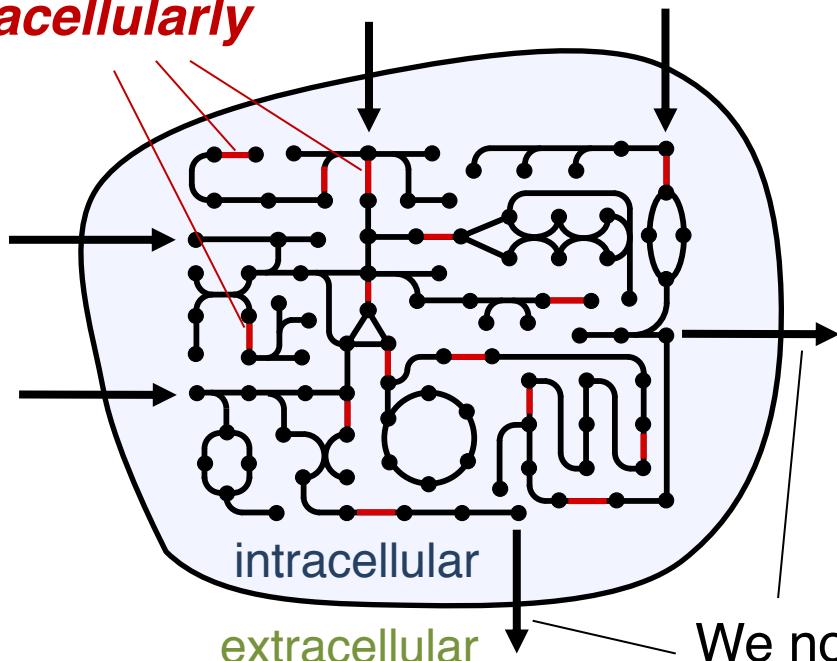
proteomic data



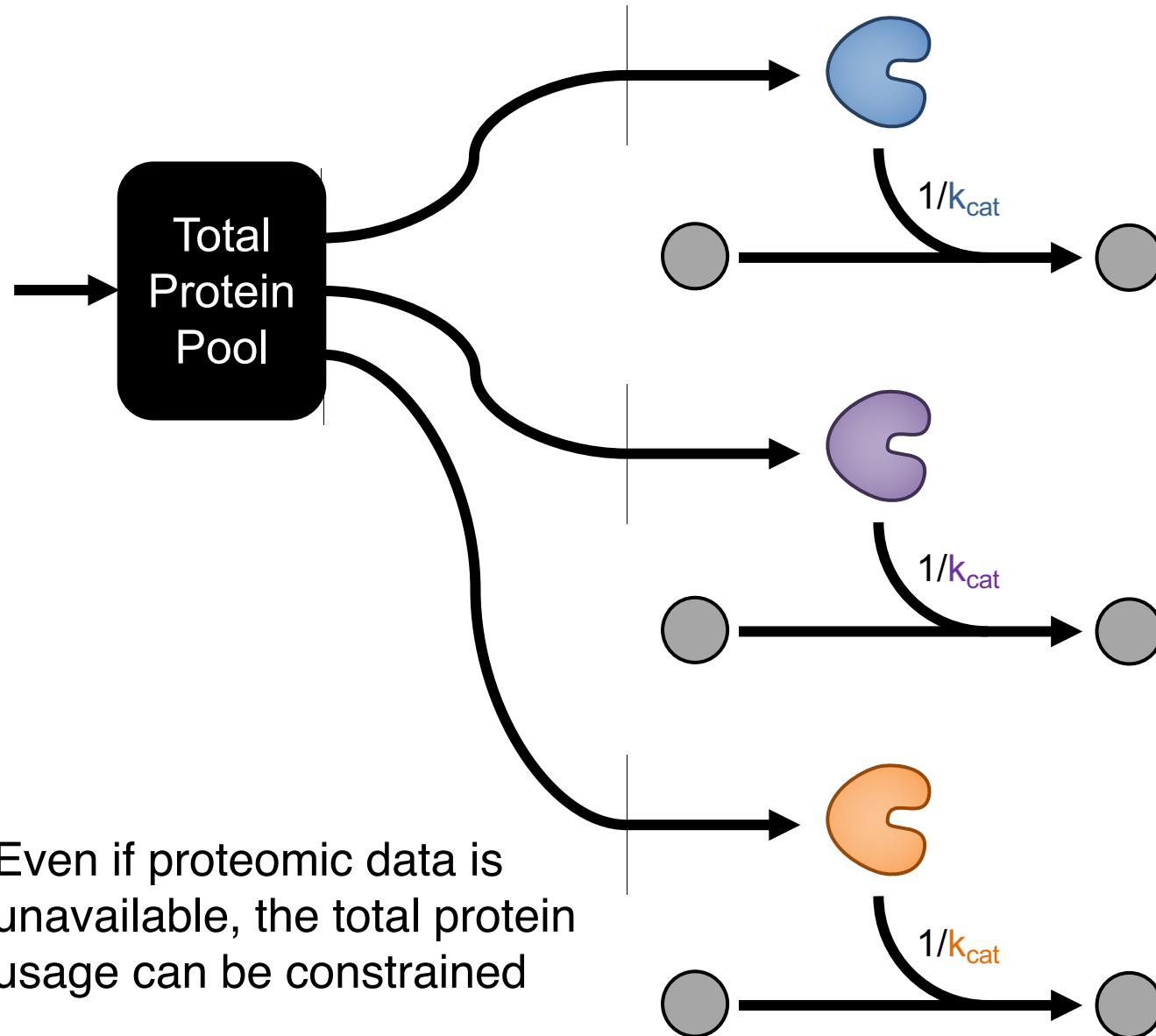
kinetic data ( $k_{cat}$  values)



*Model fluxes are now constrained **intracellularly***



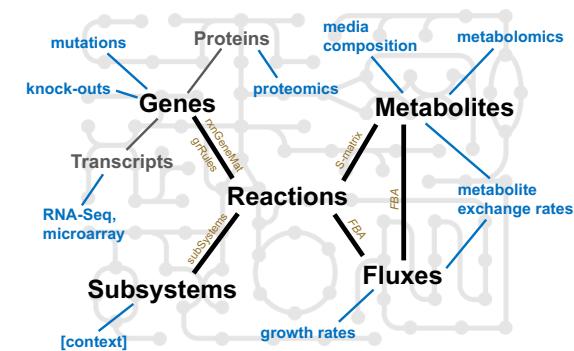
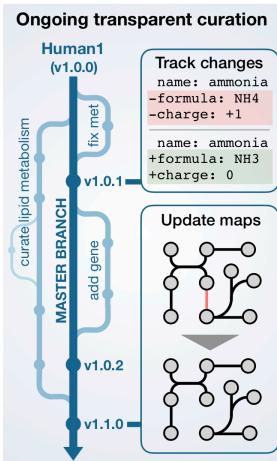
We now require little or no information about metabolite exchange rates



Even if proteomic data is unavailable, the total protein usage can be constrained

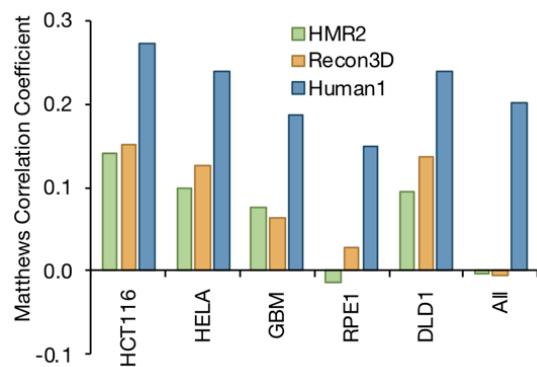
# Summary

GEMs can serve as a scaffold for integrating & studying diverse types of (omics) data

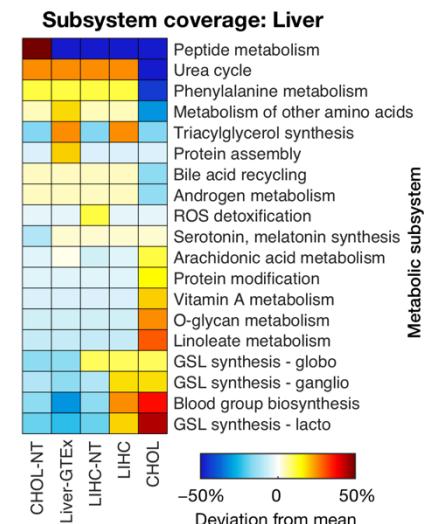


Transparent, git-based GEM curation addresses the need for quick and collaborative updates/extensions of the model

Human-GEM enables the analysis of transcriptomic data in the context of metabolism



Model curation efforts have increased the accuracy of gene essentiality predictions



The incorporation of enzyme constraints enables the simulation of physiologically relevant flux distributions using little or no metabolite exchange rate data



# Acknowledgements

Jens Nielsen

Hao Wang  
Pinar Kocabas  
Pierre Cholley  
Petre Mihail Anton  
Daniel Cook  
Avlant Nilsson  
Iván Domenzain  
Raphael Ferreira  
Johan Gustafsson  
Virinchi Billa  
Angelo Limeta  
Alex Hedin  
Eduard Kerkhoven  
Thomas Svensson  
Bernhard Palsson  
Adil Mardinoglu  
Lena Hansson  
Mathias Uhlén



*Knut och Alice  
Wallenbergs  
Stiftelse*

NBS

NIH NATIONAL CANCER INSTITUTE