



# 19<sup>th</sup> European Conference on Computational Biology

## Planetary Health and Biodiversity

31<sup>st</sup> August 8<sup>th</sup> September 2020

## T05 Part 2: Introduction to COBRA

Miguel Ponce de León - BSC ([miguel.ponce@bsc.es](mailto:miguel.ponce@bsc.es))

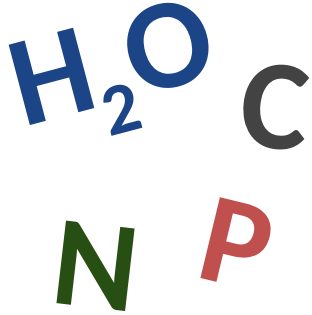
Marta Cascante - UB ([martacascante@ub.edu](mailto:martacascante@ub.edu))



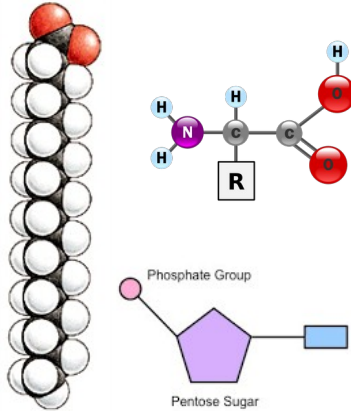
# What is a cell made of?

*Level (scale) of description*

## Chemistry



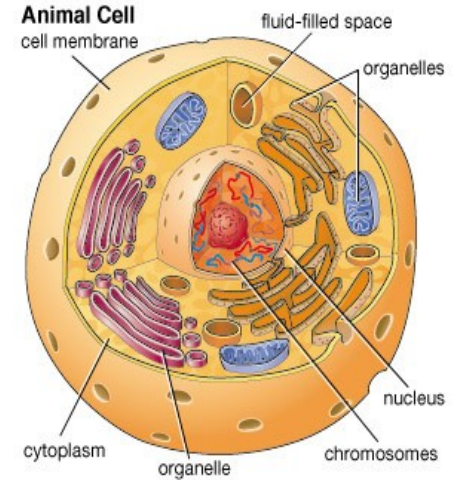
## Building blocks



## Macromolecules



## Cell



Cell's molecular factory: metabolism → What is metabolism?

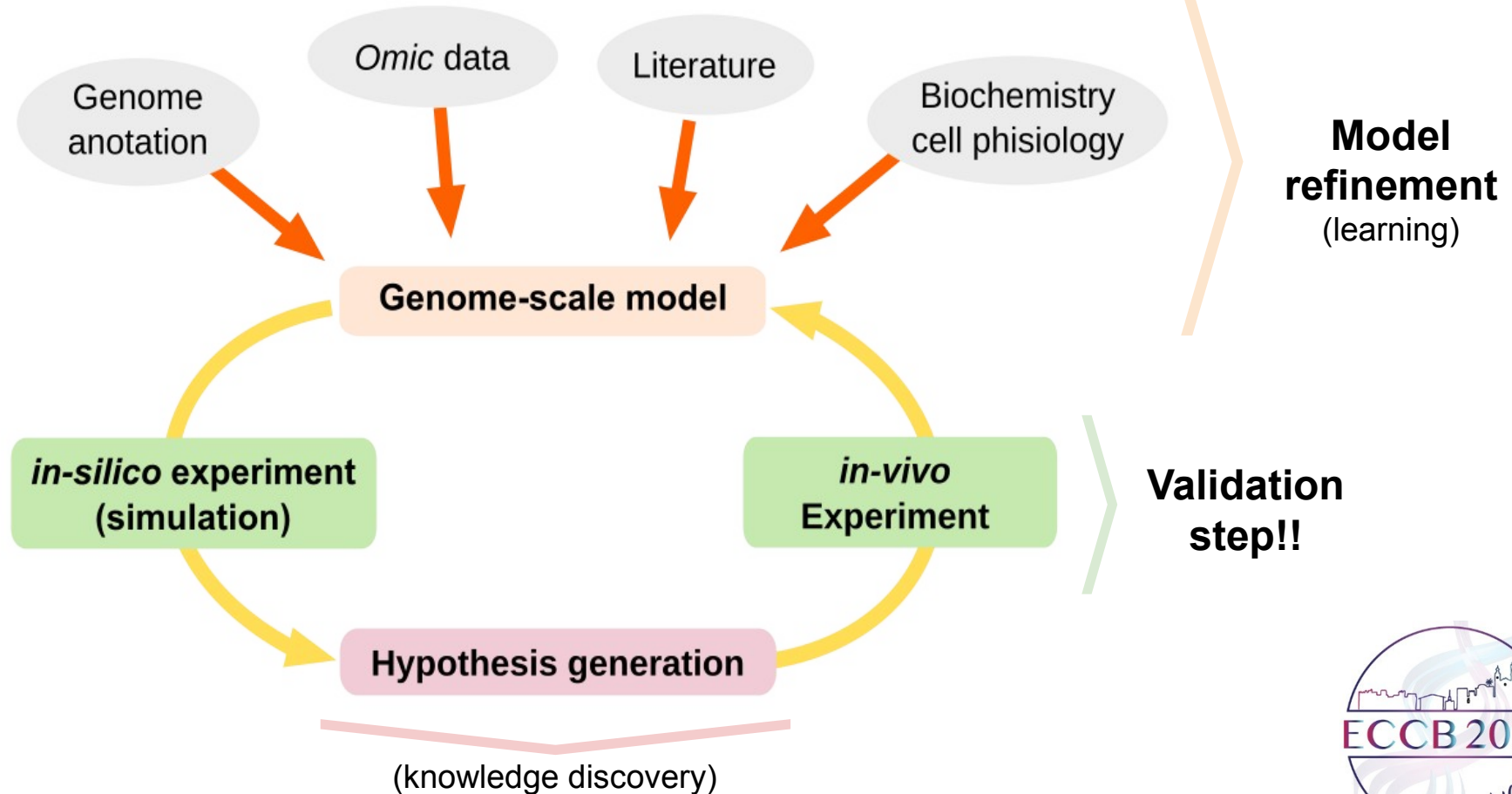
# What is a Genome-scale metabolic model?

Is a computational representation the metabolism of a cell

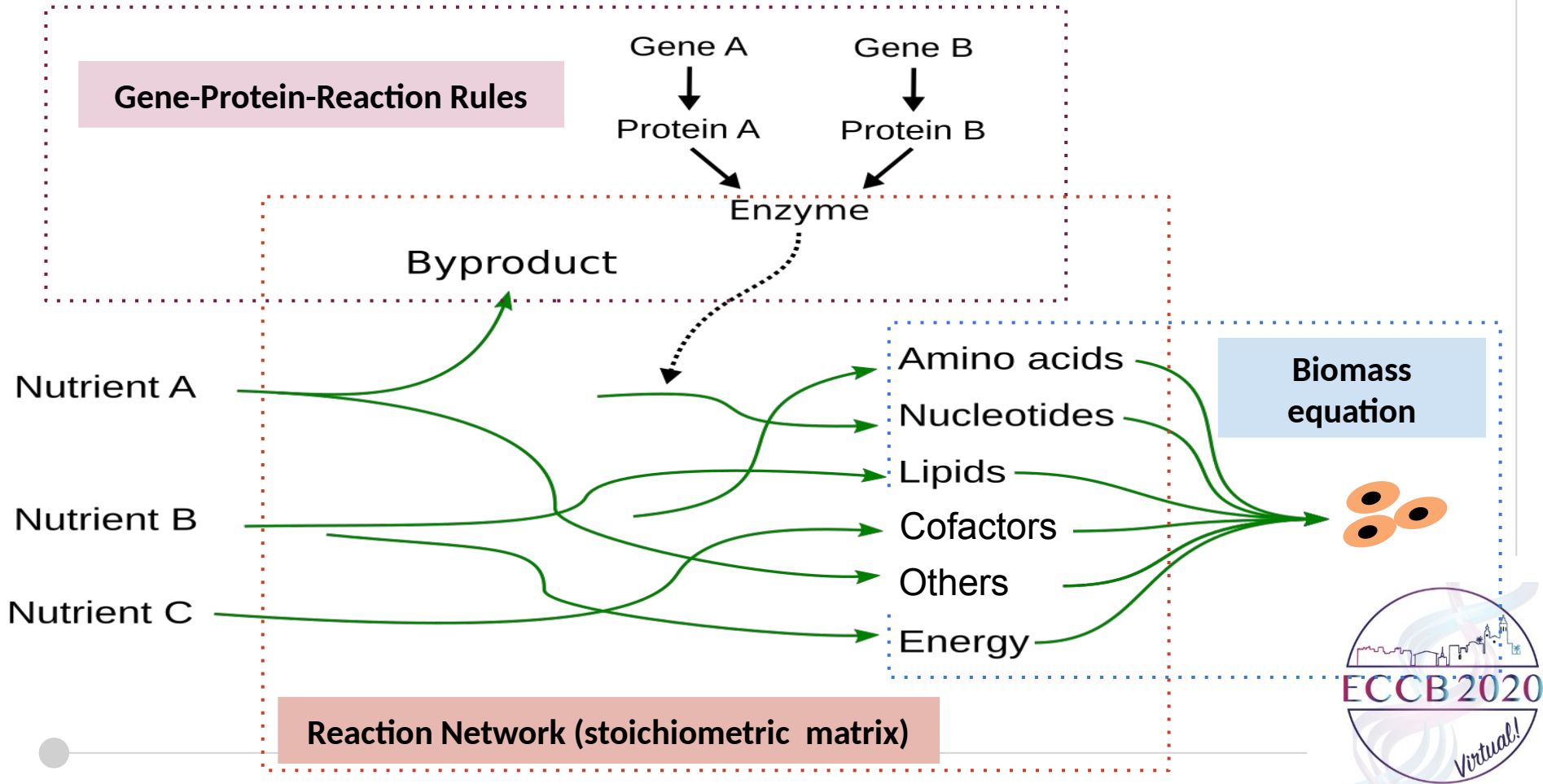
- **Includes:** genes and complexes, biochemical reactions, metabolites, transporters, cell compartments.
- **Uses:** Omic data integration, simulations, *in-silico* predictions.



# Genome-Scale Modeling in Systems Biology



# Genome-scale metabolic model

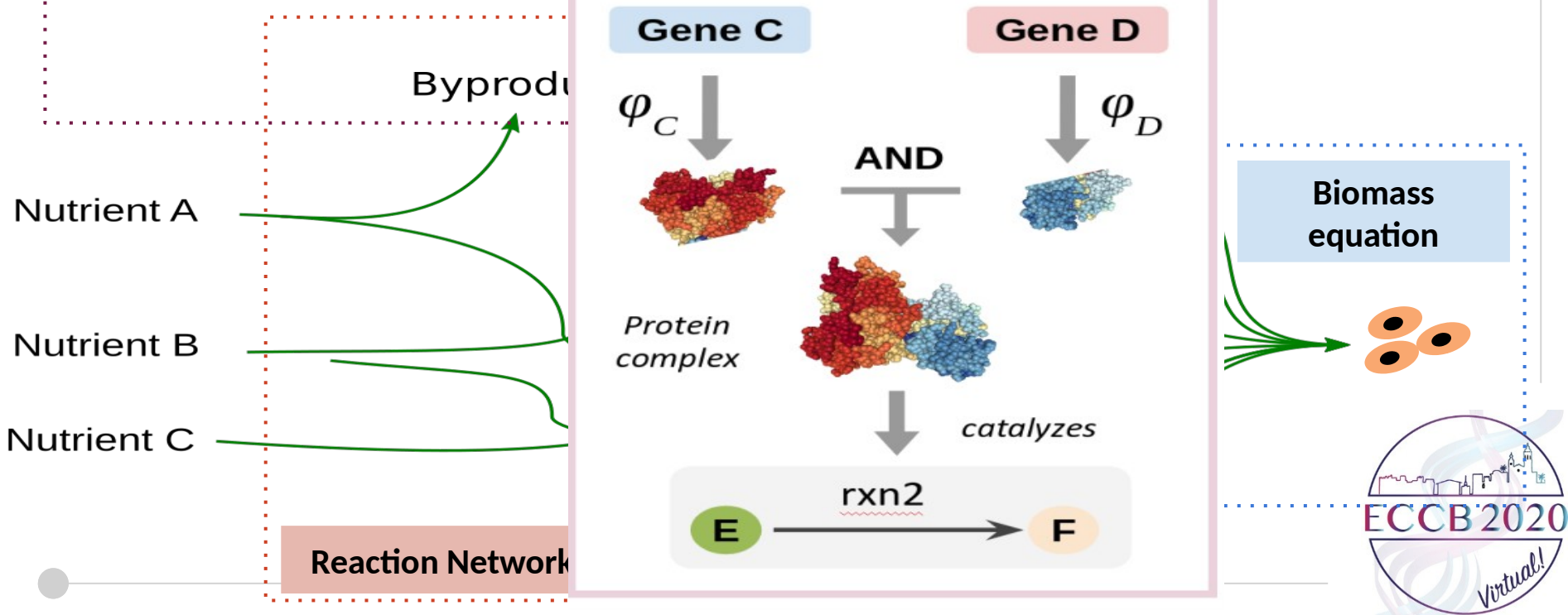


# Genome-scale metabolic model

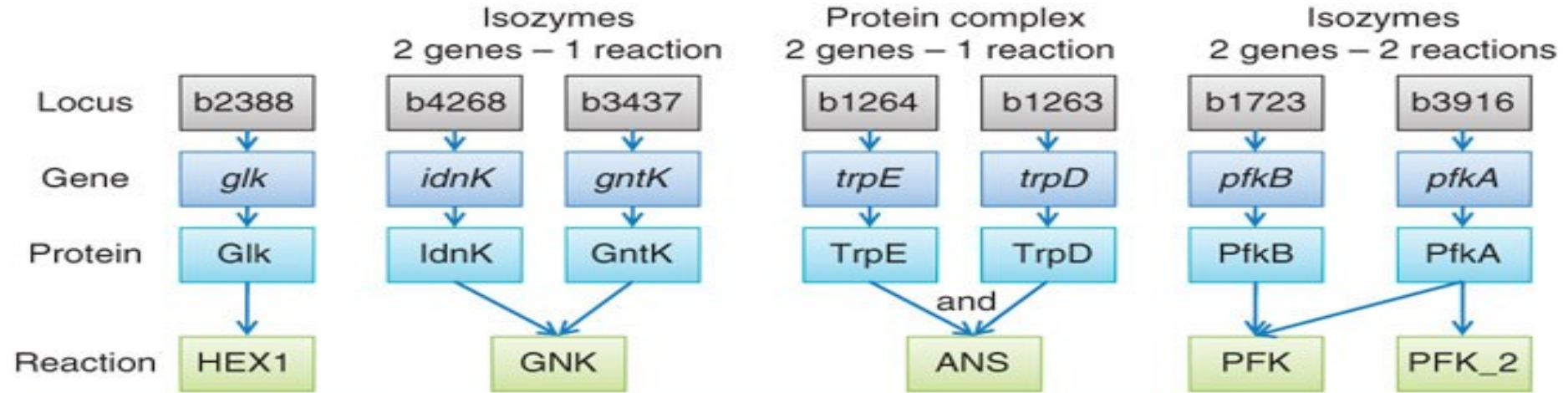
## Gene-Protein-Reaction Rules

## Gene-Protein Reaction Rules

*Logical relation between gene complexes and biochemical reactions*



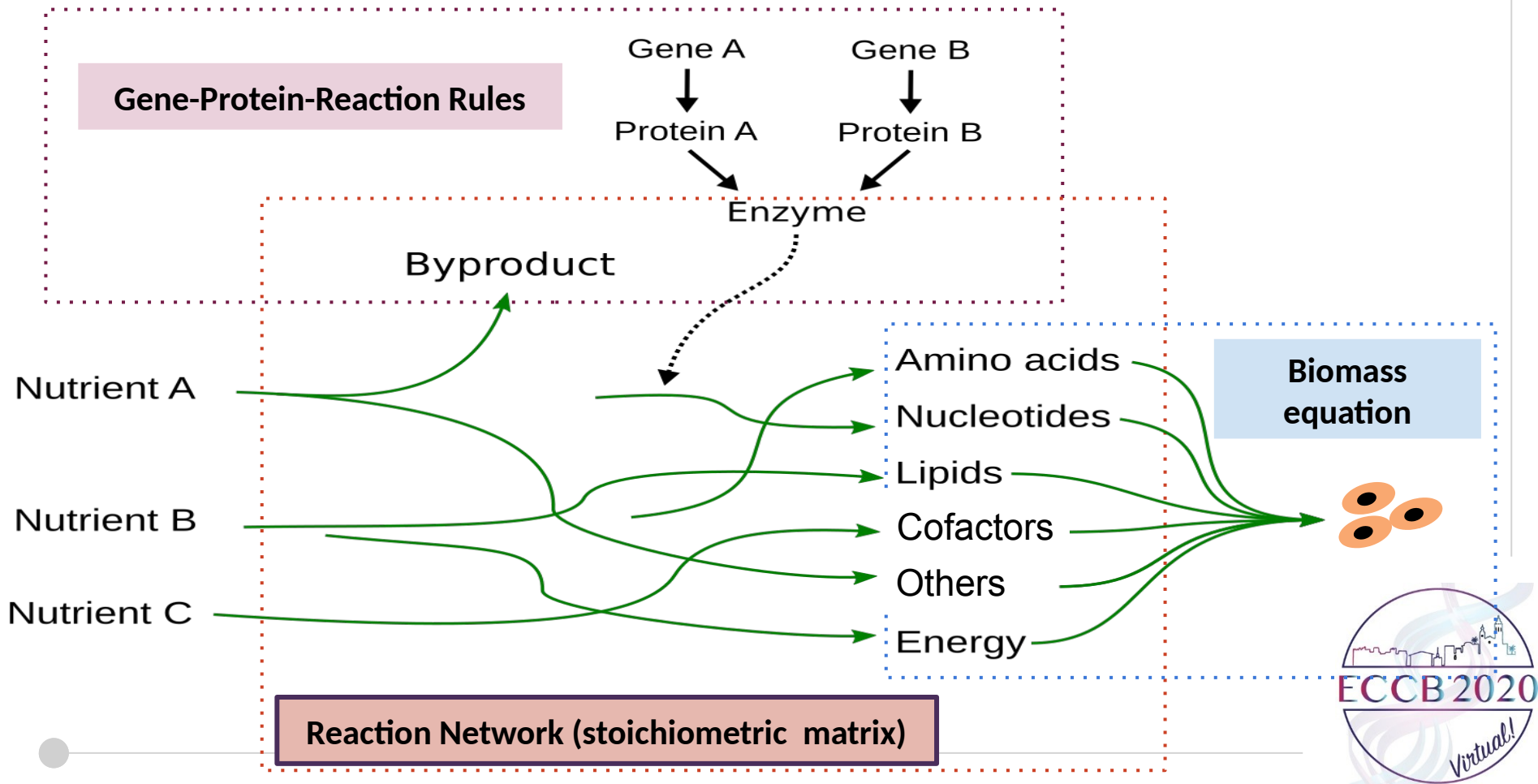
# Gene Protein Reaction rules: examples



Reaction abbreviation	Reaction name	E. C.number	GPR
HEX1	Hexokinase (D-glucose:ATP)	2.7.1.1	(b2388)
GNK	Gluconokinase	2.7.1.12	(b3437) or (b4268)
ANS	Anthranilate synthase	4.1.3.27	(b1264) and (b1263)
PFK	Phosphofructokinase	2.7.1.11	(b1723) or (b3916)
PFK_2	Phosphofructokinase (2)	2.7.1.11	(b3916)



# Genome-scale metabolic model





# Genome-scale metabolic model

Gene-Protein-Reaction Rules

Gene A

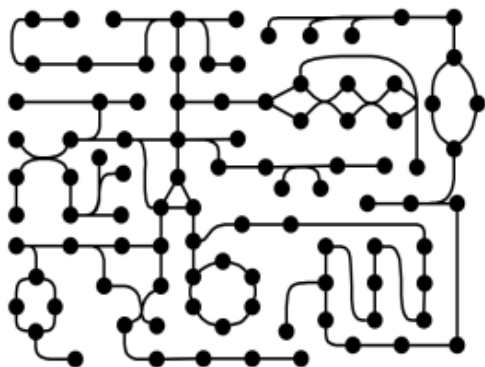


Gene B



## Metabolic Network: Human Metabolic model Recon 2.2.1

*Reactions, Transports, Metabolites & Cell compartments (-> stoichiometric matrix)*



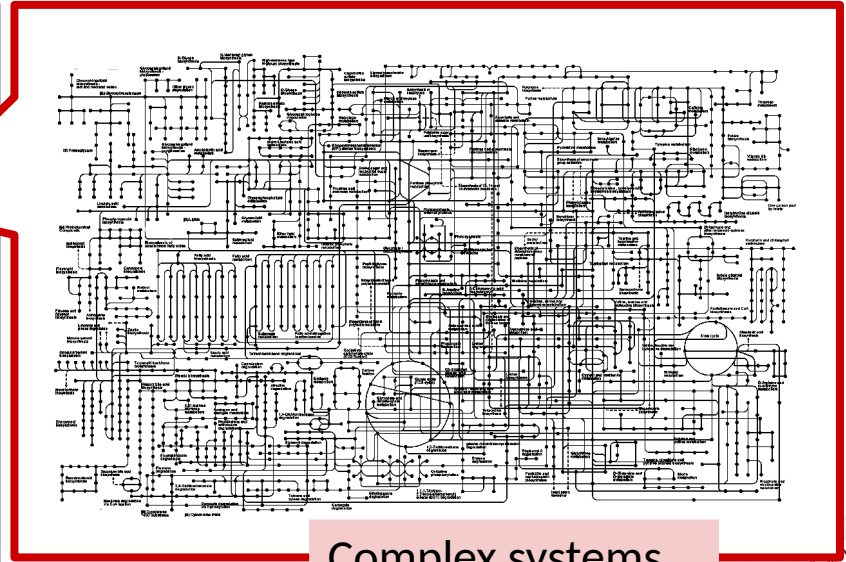
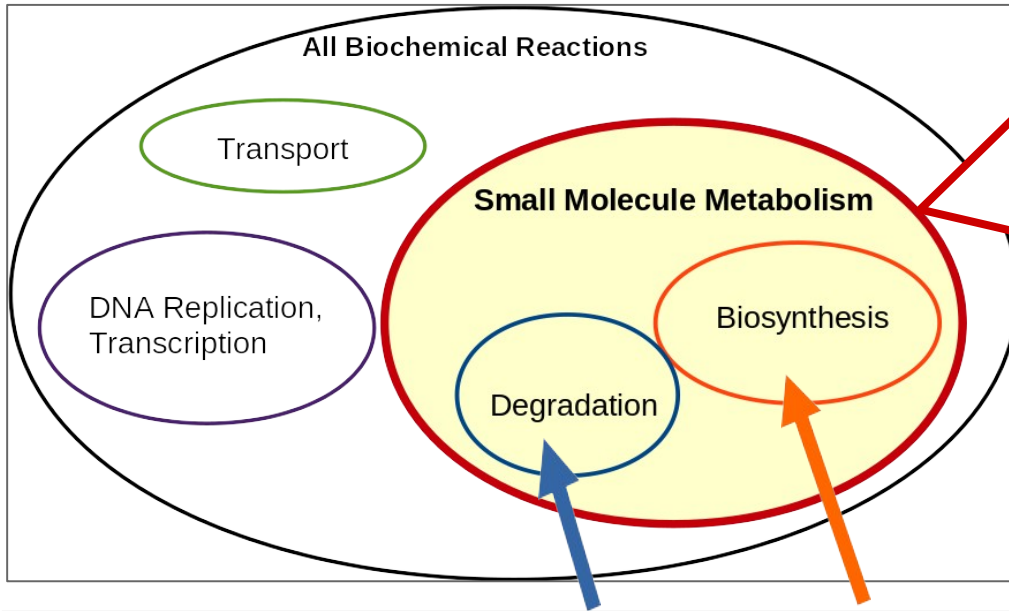
	...	rxn2
A		0
B		0
⋮		
E		-1
F		+1

	Total
Genes	1675
Metabolites	5324
Reactions	7785

Reaction Network (stoichiometric matrix)

# Metabolism: the molecular factory of the cell

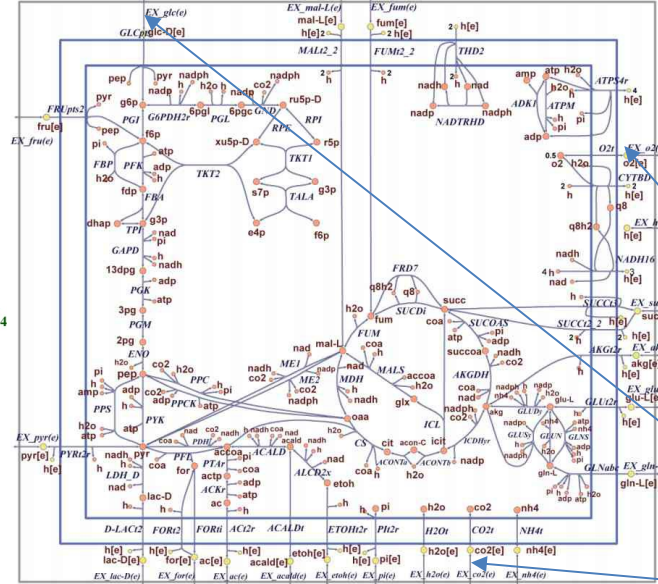
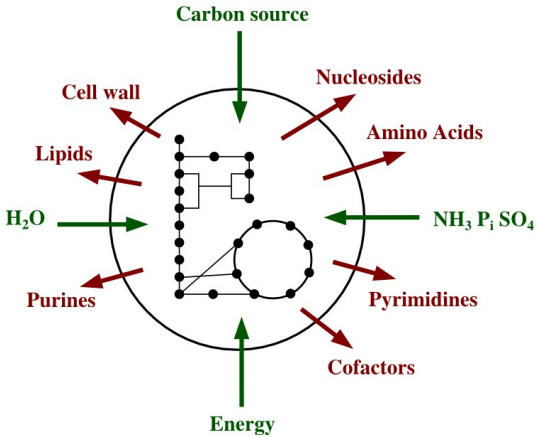
Is the **network** of biochemical reactions and transport processes that occur within a cell and allow **cell maintenance and growth**



Complex systems

- Generation of **energy (catabolism)** and **building block (anabolism)**
- Include the enzymatic reaction that act over small molecules

# Stoichiometric Matrix → Mass Balance Equations & Exchange fluxes (E.coli core)



## Mass Balance:

$$d[13\text{pg}(c)] / dt = \text{GAPD} - \text{PGK}$$

$$d[2\text{pg}(c)] / dt = \text{PGM} - \text{ENO}$$

$$d[6\text{pg}(c)] / dt = \text{PGL} - \text{GND}$$

$$d[6\text{pgl}(c)] / dt = \text{G6PDH2r} - \text{PGL}$$

$$d[\text{cit}(c)] / dt = \text{CS} - \text{ACONTa}$$

$$d[\text{dhap}(c)] / dt = \text{FBA} - \text{TPI}$$

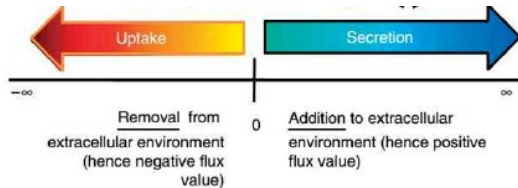
...(continues)

(c): cytosol

**Exchange fluxes:** variables that represent the exchanges of nutrients/by-products between the organism and its environment (uptake/secretion)

Lower bound:  
rate of nutrient uptake

Upper bound:  
rate of nutrient release



units: mmol/gDw/hr

	Min	Max
Glucose	-6	0
Oxygen	-20	0
CO <sub>2</sub>	0	Inf

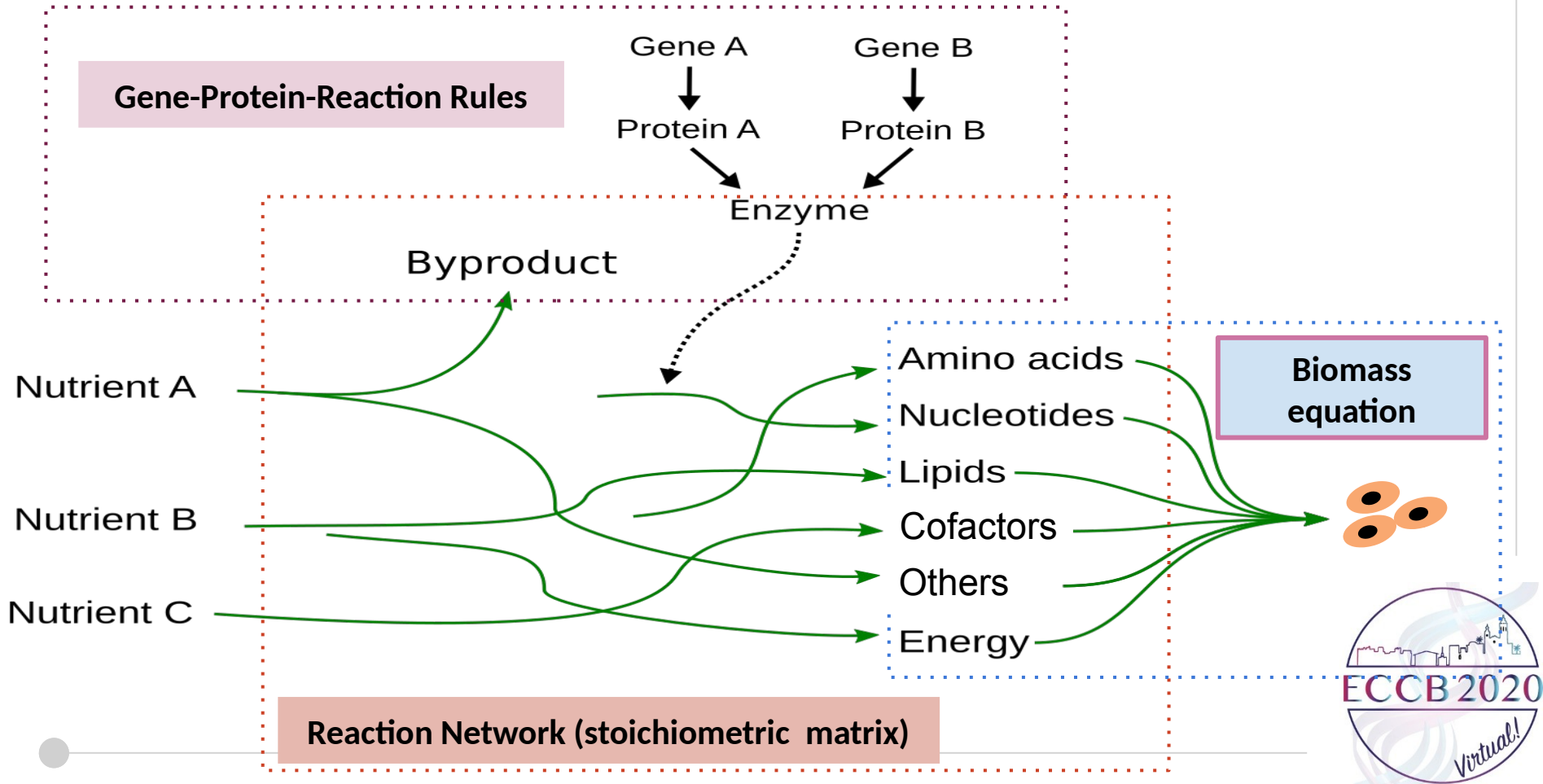
$\text{GLC}_{\text{Ex}}$   $\text{O}_{\text{Ex}}$   $\text{CO}_2_{\text{Ex}}$

-1

-1



# Genome-scale metabolic model



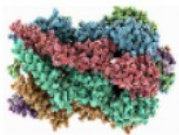
# Genome-scale metabolic model

## Biomass Equation

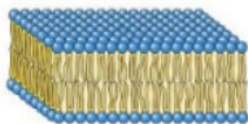
*Quantitative molecular composition of a cell*



DNA + RNA



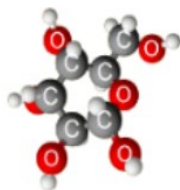
Proteins



Lipids



Vitamins/coenzymes



Carbohydrate

Biomass	Total
RNA components:	4
DNA components:	4
Protein components:	20
Carbohydrates:	1
Lipid components:	16
Cofactors and vitamins:	10

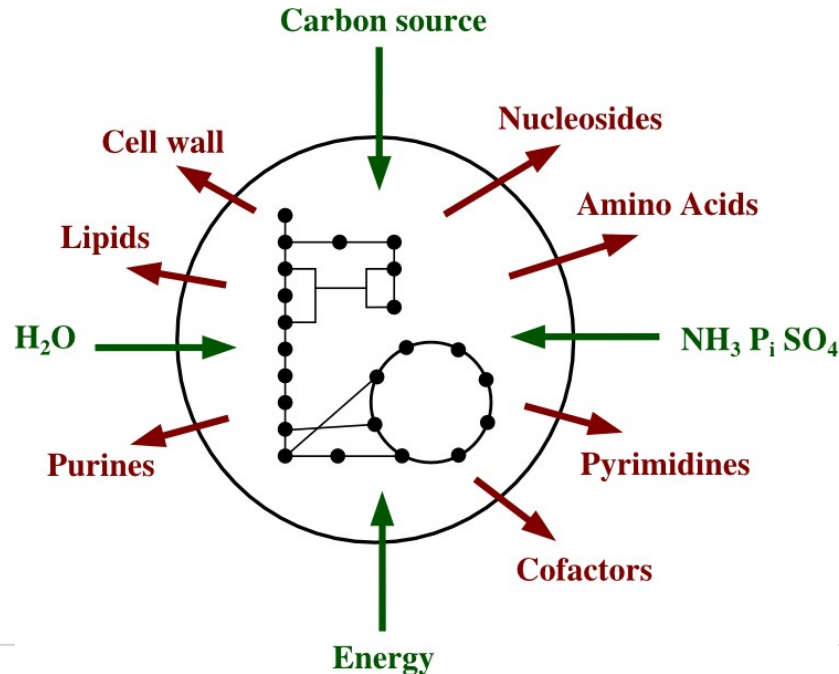
**Biomass  
equation**



Reaction Network (stoichiometric matrix)

# Biomass Equation of a cell

Description in stoichiometric terms of all components present in a gram (dry weight) of a cell.



Metabolite	Demand (mmol)
ATP	41.2570
NADH	-3.5470
NADPH	18.2250
G6P	0.2050
F6P	0.0709
R5P	0.8977
E4P	0.3610
T3P	0.1290
3PG	1.4960
PEP	0.5191
PYR	2.8328
AcCoA	3.7478
OAA	1.7867
AKG	1.0789

Neidhardt, et al. *Physiology of the Bacterial Cell* (1990)

# Modeling metabolic systems

## Kinetic Modelling (Differential equations)

- Time evolution of system variables (+)
- Sensitivity Analysis (+)
- Unknown Kinetic Parameter (-)
- Unknown Enzymatic Mechanisms (-)

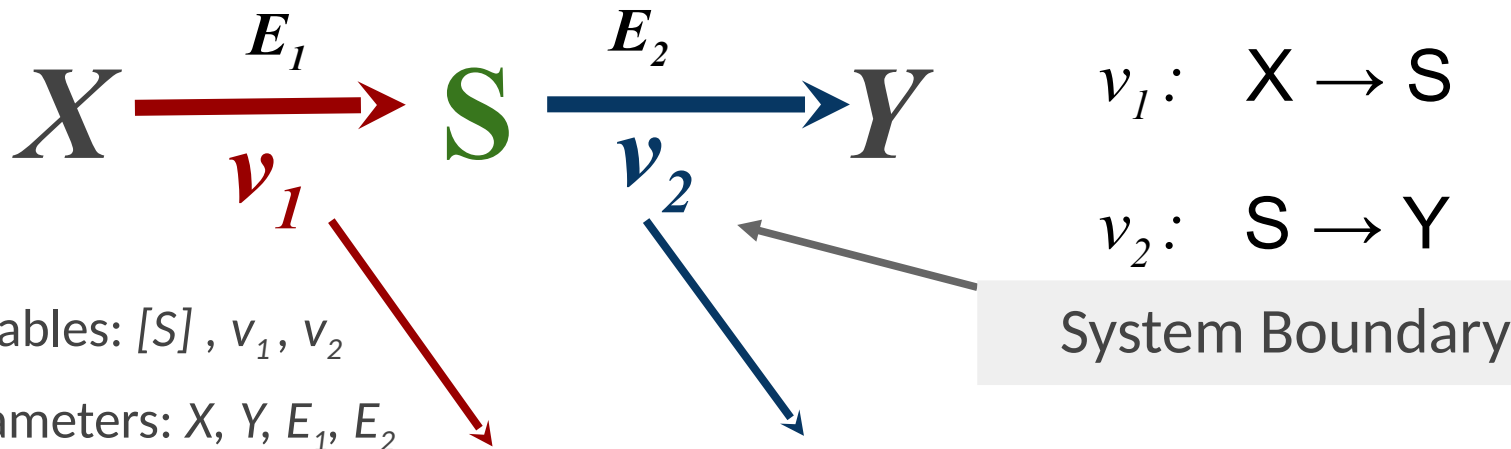
## Constraint-Based Modeling (CBM)

- Only need stoichiometry (+)
- Structure is an invariant property (+)
- Computationally tractable using **genome-scale models** (+)
- No information of metabolite concentrations (-)
- Only valid under steady-state (-)





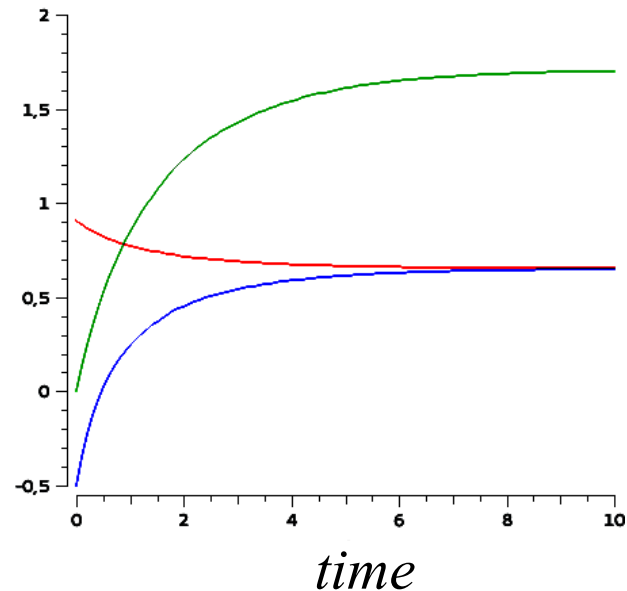
# Modeling metabolic systems



## Kinetic Mechanisms:

- Mass action
- Michaelis-Menten
- Others

# Kinetic description



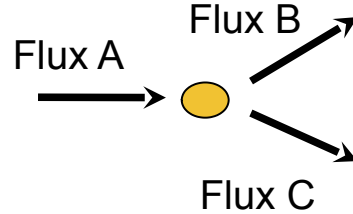
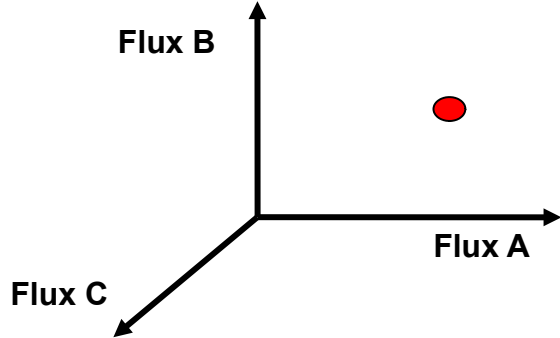
Many parameters!

$$\frac{d[S]}{dt} = \left( \frac{v_{f(v1)}[X] - v_{r(v1)}[S]}{K_{ms(v1)} + K_{mp(v1)} + 1 + \frac{[X]}{K_{ms(v1)}} + \frac{[S]}{K_{mp(v1)}}} \right) + \left( \frac{v_{f(v2)}[S] - v_{r(v2)}[Y]}{K_{ms(v2)} + K_{mp(v2)} + 1 + \frac{[S]}{K_{ms(v2)}} + \frac{[Y]}{K_{mp(v2)}}} \right)$$

# Constraint-based approach

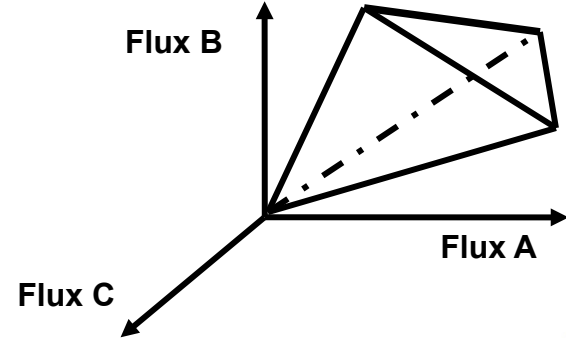
## Theory

- Full information
- Solution is a unique point (or limit cycle) in the system's phase space



## Genome-scale

- Incomplete information
- **Solution (flux) space**

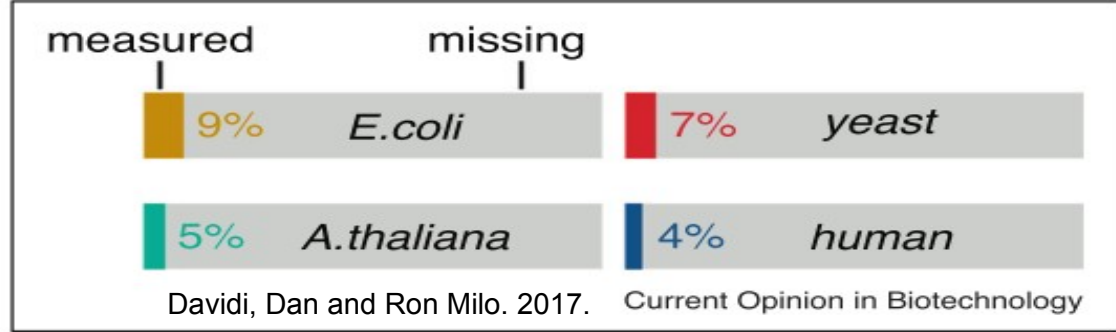


For genome-scale reconstructed metabolic network there are still not enough information to create the full kinetic description → too many unknown parameters!

# Kinetic constants: the state of the art

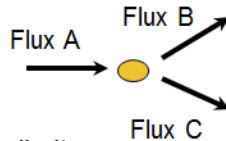
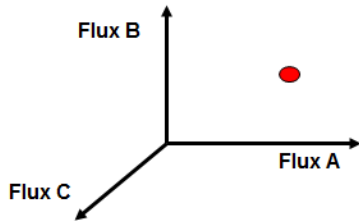
## N° of reactions from GEMs:

- *E. coli* (iJO1266): 2251
- Budding yeast (iIND750): 1149
- Arabidopsis (--): 1363
- Human (Recon1): 7785



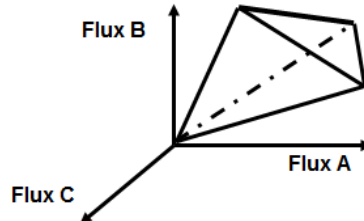
## Theory

- Full information
- Solution is a unique point (or limit cycle) in the system's phase space



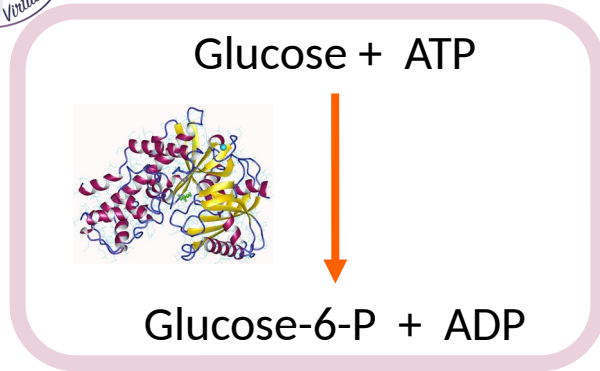
## Genome-scale

- Incomplete information
- Solution (flux) space



For **genome-scale** reconstructed metabolic network there are still not enough information to create the full **kinetic description** → **too many unknown parameters!**

# Constraint-based modeling



Glucokinase (single reaction)

Glucokinase

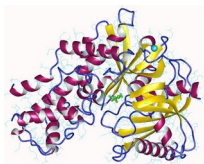
Glucose	-1	$\rightarrow N$
ATP	-1	
G-6-P	+1	
ADP	+1	

Stoichiometric matrix  $N$  (metabolic network)

## The Constraints

# Constraint-based modeling

Glucose + ATP



Glucose-6-P + ADP

Glucokinase (single reaction)

Glucokinase

Glucose  
ATP

G-6-P  
ADP

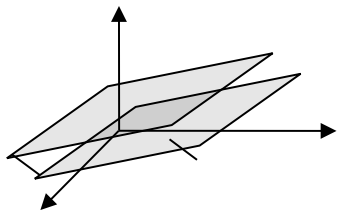
$$\begin{bmatrix} & -1 & \\ & -1 & \\ & +1 & \\ & +1 & \end{bmatrix} \rightarrow N$$

Stoichiometric matrix  $N$  (metabolic network)

## The Constraints

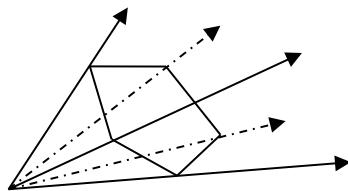
Mass Balance

$$N \cdot v = 0$$



Thermodynamics

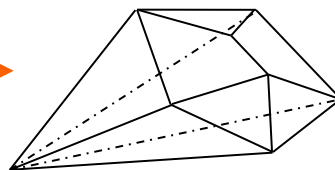
$$v_i > 0$$



Capacities (bounds)

$$v_i < v_{max}$$

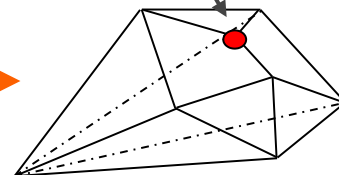
Flux Space



Cell Objective

Flux Balance Analysis

$$\text{Max } c^T \cdot v$$



How does it look (under the hood)?





# Genome-scale metabolic model naked

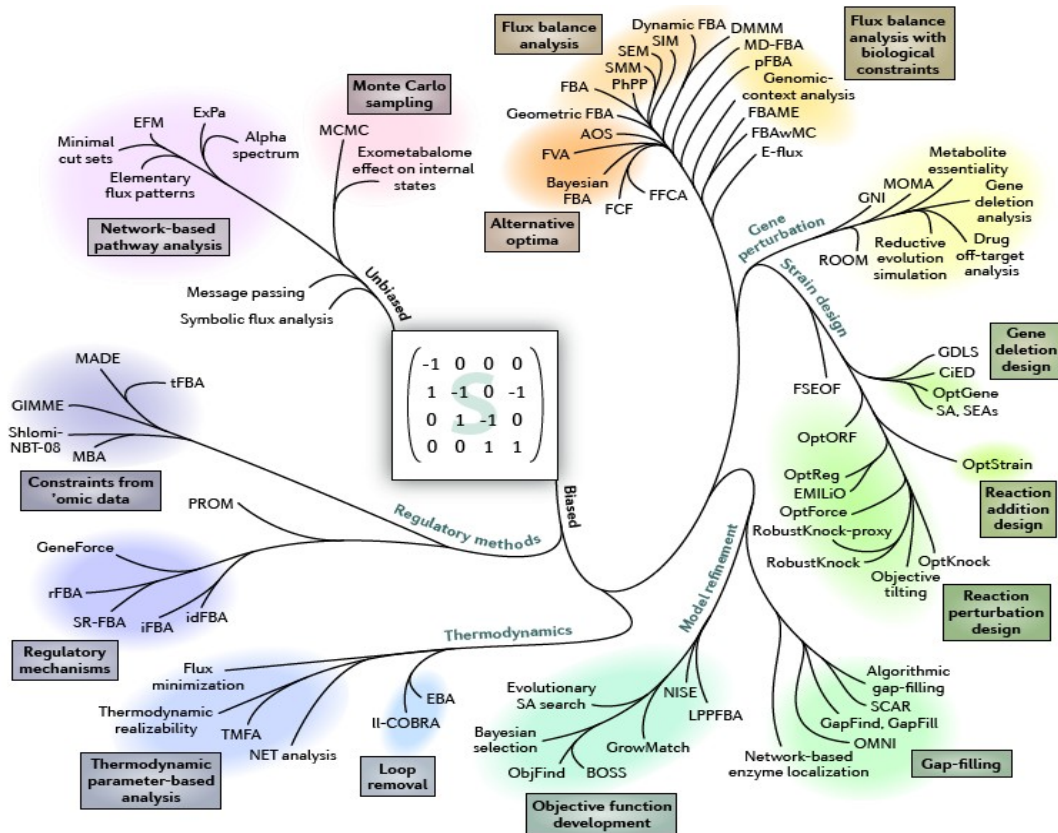


```
<model id="Recon2.2.1">
  <listOfUnitDefinitions>
    <unitDefinition id="mmol_per_gDW_per_hr"> ... </listOfUnitDefinitions>
  <listOfCompartments>
    <compartment id="g" name="Golgi apparatus" size="1"/>
    <compartment id="c" name="cytoplasm" size="1"/>
    ...
  </listOfCompartments>
  <listOfSpecies>
    <species id="M_10fthf5glu_c" name="10-formyltetrahydrofolate-[Glu](5)" compartment="c" charge="-6">
      <notes>FORMULA: C40H45N11O19</notes>
    </species>
    ...
  <reaction id="R_ENO" name="enolase" reversible="true">
    <notes>
      <p>GENE ASSOCIATION: HGNC:3350 or HGNC:3354 or HGNC:3353</p>
      <p>CONFIDENCE LEVEL: 5</p>
      <p>SUBSYSTEM: Glycolysis/gluconeogenesis</p>
    </notes>
    <listOfReactants>
      <speciesReference species="M_2pg_c" stoichiometry="1"/>
    </listOfReactants>
    <listOfProducts>
      <speciesReference species="M_h2o_c" stoichiometry="1"/>
      <speciesReference species="M_pep_c" stoichiometry="1"/>
    </listOfProducts>
    <kineticLaw>
      <listOfParameters>
        <parameter id="UPPER_BOUND" value="1000" units="mmol_per_gDW_per_hr"/>
        <parameter id="FLUX_VALUE" value="0" units="mmol_per_gDW_per_hr"/>
        <parameter id="OBJECTIVE_COEFFICIENT" value="0" units="dimensionless"/>
        <parameter id="LOWER_BOUND" value="-1000" units="mmol_per_gDW_per_hr"/>
      </listOfParameters>
    </kineticLaw>
  </reaction>
```

# A phylogeny of **CO**nstraint-Based **Re**construction and **A**nalysis (**COBRA**) methods for GSMMs

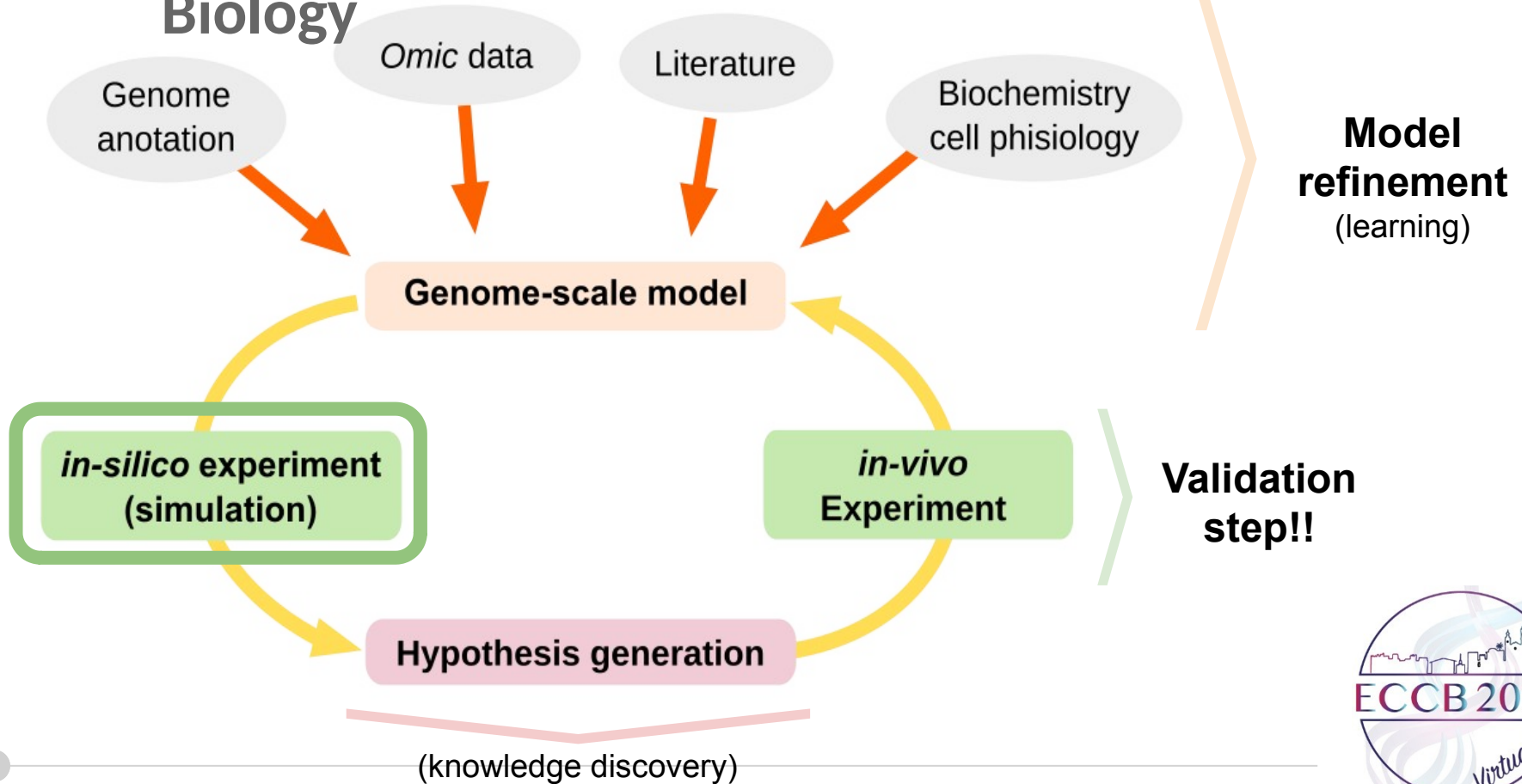
Because of the versatility and scalability, more than 100 COBRA methods have been developed for constraint-based modeling and analysis. Many of them implemented in software packages: <http://opencobra.github.io/>

All are based on the analysis of the underlying metabolic network structure (i.e., the **stoichiometric matrix**).



The phylogenetic tree depict similarities between applications of the COBRA methods, and the underlying algorithms (Lewis et al. 2012)

# Genome-Scale Modeling in Systems Biology

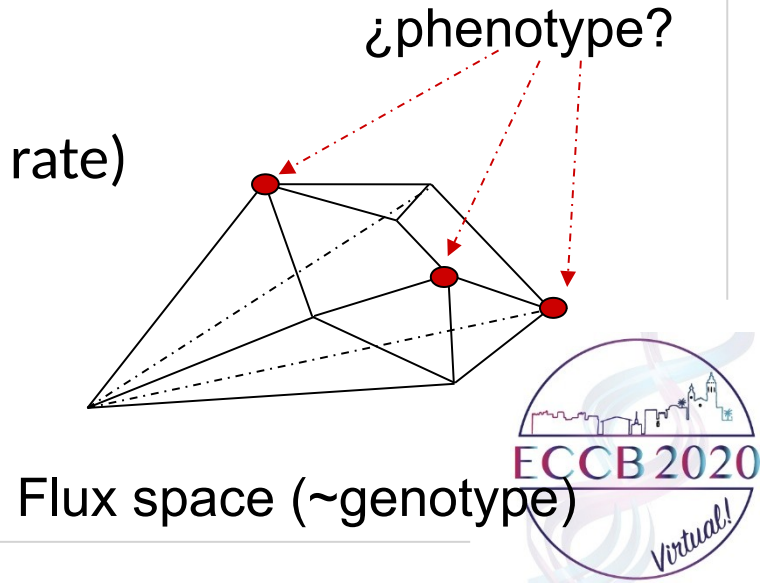


# Predicting plausible physiological states (phenotypes)

¿How to identify a particular flux distribution?

## Using optimization principles

- Adjusting with experimental data
- Maximize de Biomass production (growth rate)
- Maximize ATP production
- Minimize metabolic cost
- Multiple criteria



# Do we have Strategies to found a unique flux distribution?

## Flux Balance Analysis (FBA)

**Definition:** computational strategy that uses a set of constraints (e.g. mass balance, thermodynamics, etc) and linear optimization to determine the steady-state reaction flux distribution in a metabolic network by maximizing an objective function (**growth rate**)

### Constraints:

$$N \cdot \vec{v} = 0$$

Mass balance

$$v_j \geq 0 \quad \forall j \in R_{irrev}$$

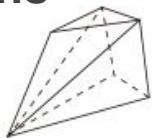
Thermodynamics

$$\alpha_j \leq v_j \leq \beta_j, \quad \forall j \in R$$

Enzyme and transport capacities

Feasible space  
(flux space)

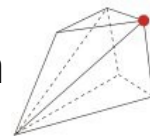
### Optimizations



Feasible space (Genotype)

Maximize an objective function

**FBA:** Max growth rate



Feasible space (Genotype)

Flux distribution (phenotype)

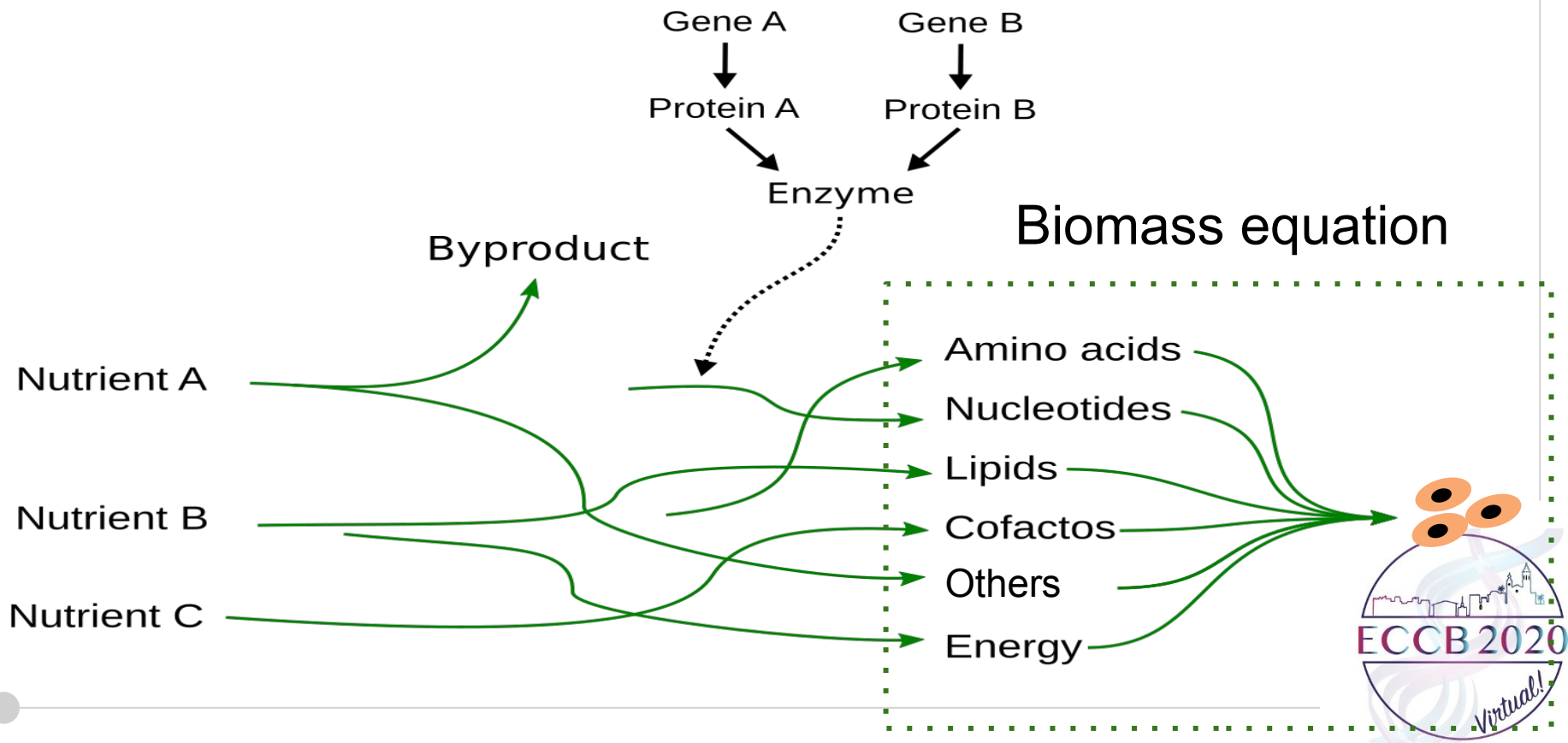
# Applications

## Gene *knockouts* predictions





## Simulations → *in-silico* predictions of gene KO effect

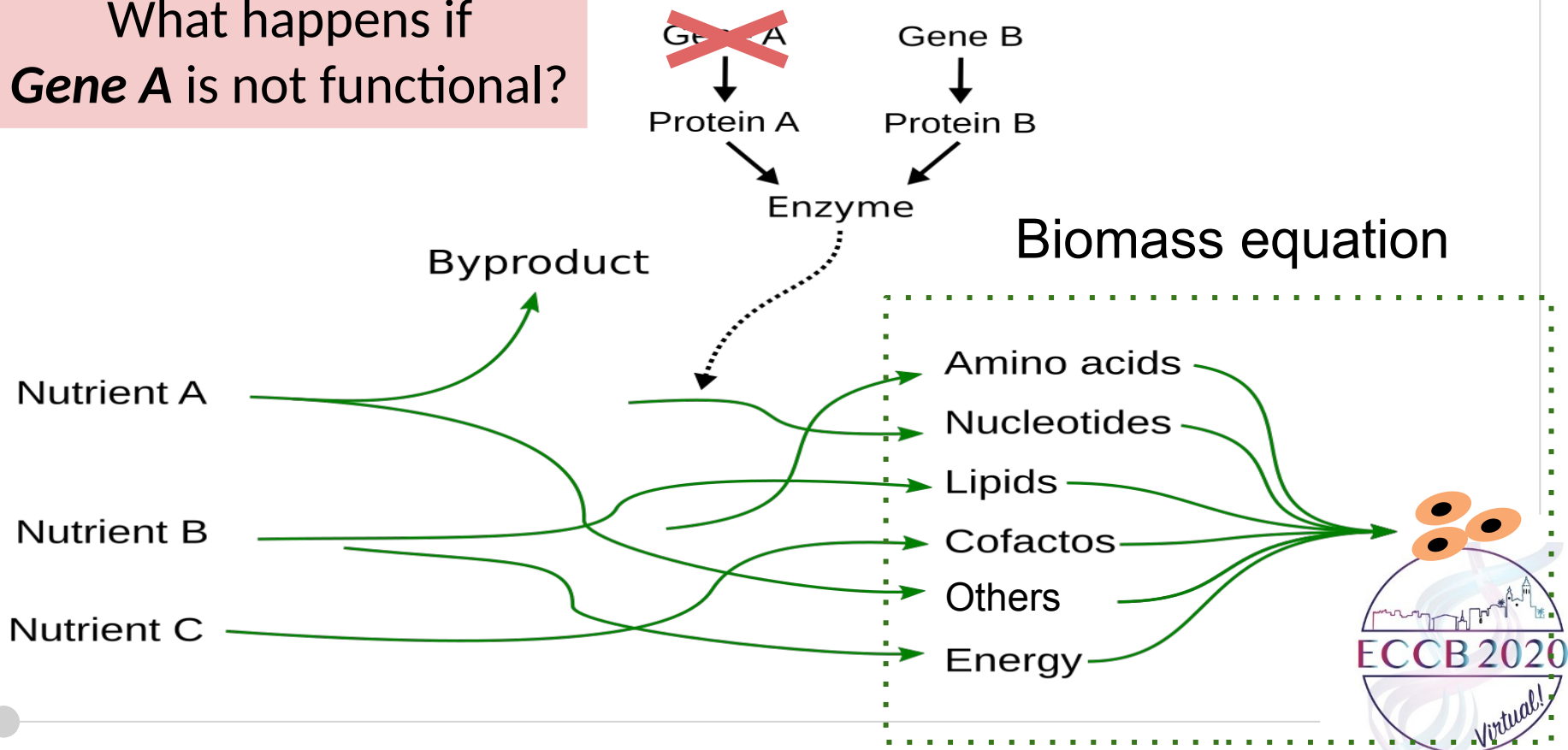






## Simulations → *in-silico* predictions of gene KO effect

What happens if  
**Gene A** is not functional?



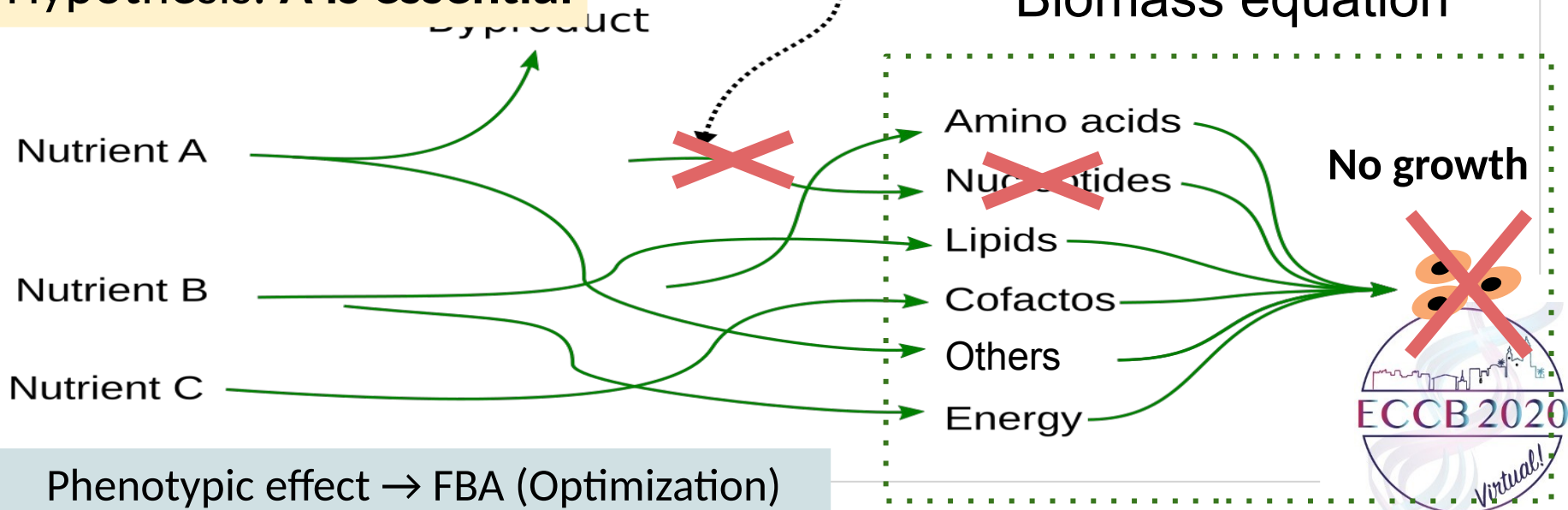


## Simulations → *in-silico* predictions of gene KO effect

What happens if  
**Gene A** is not functional?

Simulation → no growth

Hypothesis: **A is essential**



# **Applications**

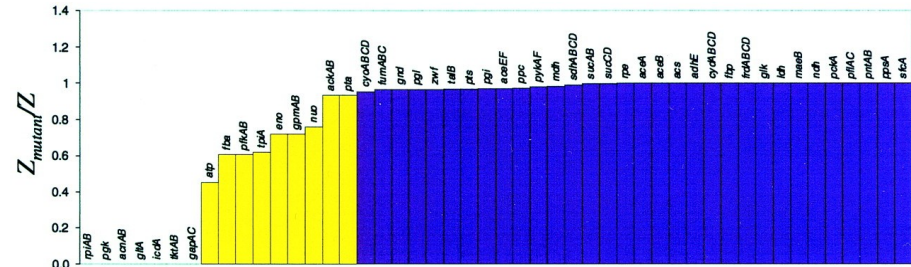
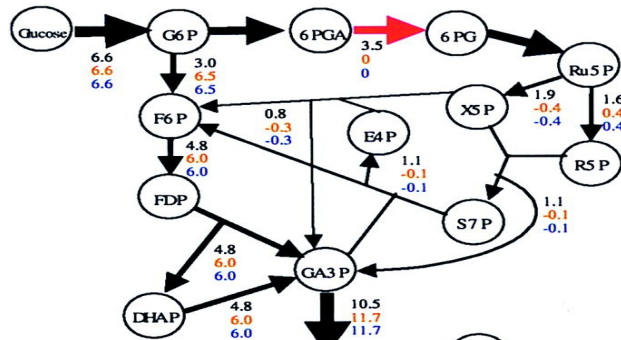
## **Gene *knockouts* predictions**

### **Does it really work?**



# Knockout predictions: original study

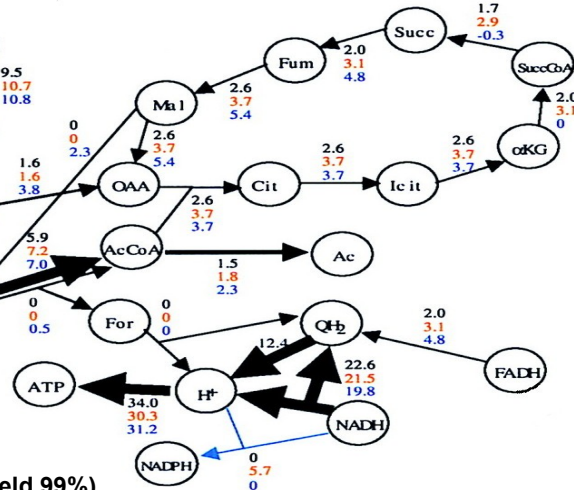
*In-silico* knockout predictions; maximal biomass



**86%** (68 of 79 cases) of the *in-silico* predictions were consistent with the experimental data.

## Rerouting of fluxes

- Wild type
- zwf* mutant (Biomass yield 99%)
- zwf pnt* double mutant (Biomass yield 99%)

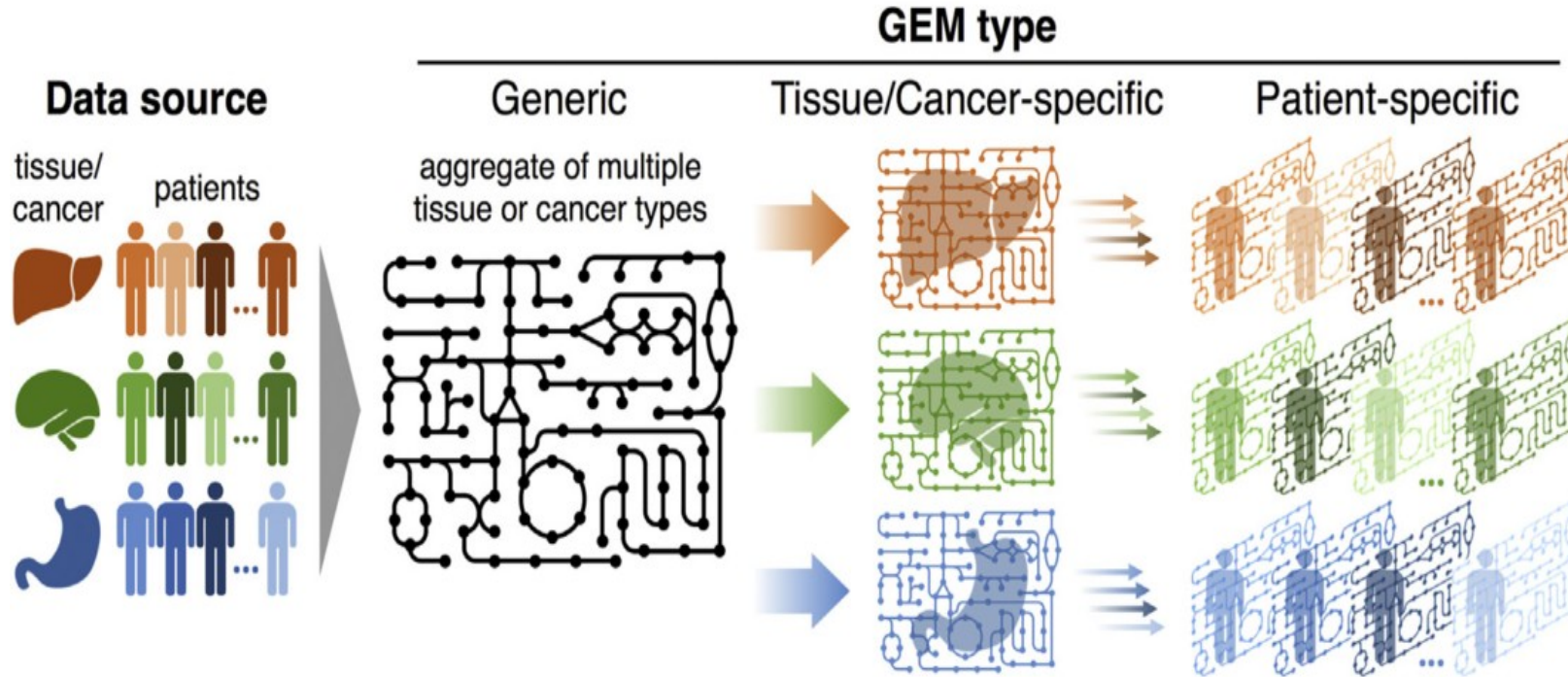


Edwards JS, Palsson BO. PNAS (2000)

# Metabolic modeling in humans



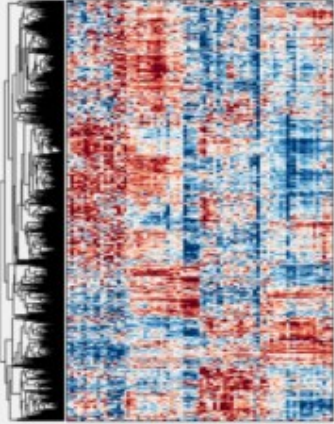
# Metabolic modeling in humans





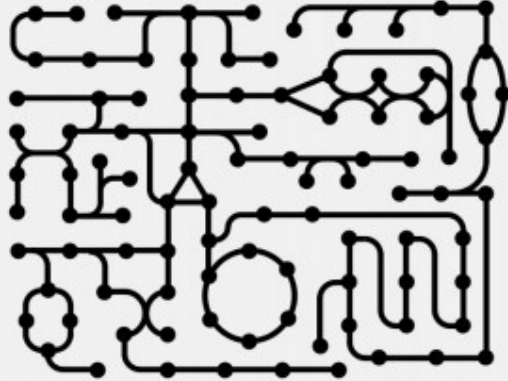
# Context-Specific Metabolic Modeling (CSM)

collect omics data



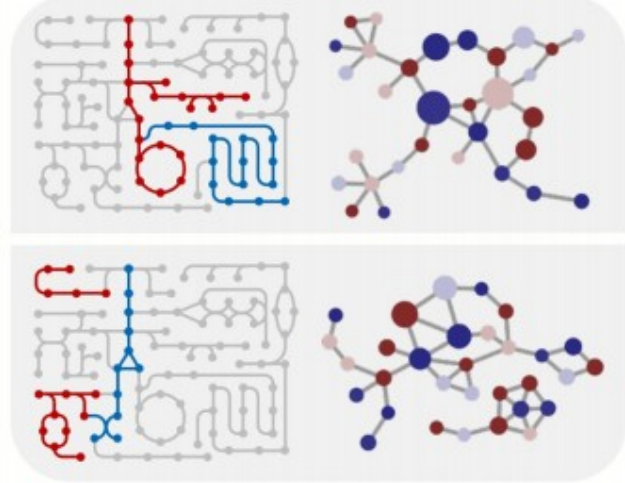
Cell Context (omics)

map to GEM



Universal Human metabolism

identify subnetworks, reporters



Cell-type specific metabolic models



# Genome-Scale Model of Human Metabolism

**Table 1**

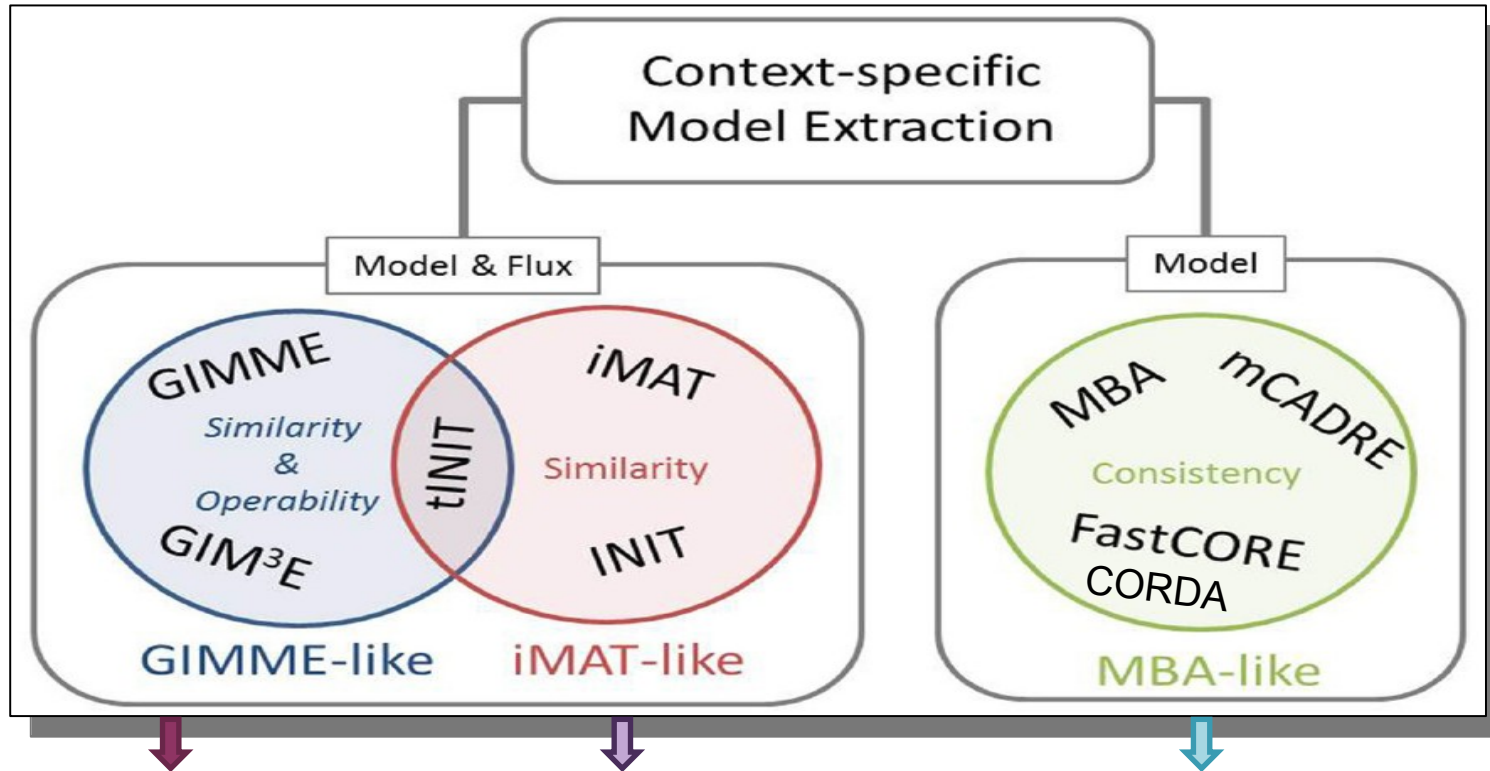
**Statistics of currently published generic human GEMs.**

Generic GEMs	Genes	Metabolites <sup>a</sup>	Reactions <sup>a</sup>	Features
RECON1	1496	1509	3744	Manually reconstructed from bibliomics data
EHMN	2322	2671	2823	Manually reconstructed from bibliomics data
RECON2	1789	2626	7440	Merging EHMN and HepatoNet1 with RECON1
RECON 2.2	1675	5324	7785	Reconstructed by integrating previous versions, with emphasis on mass and charge balance
HMR1.0	1512	3397	4144	Reconstructed based on RECON1, EHMN, HumanCyc and KEGG
HMR2.0	3765	3160	8181	Reconstructed based on HMR1, with additional emphasis on lipid metabolism by integrating iAdipocytes1809, KEGG, Lipidomics Gateway
Recon3D	2248	5835	10600	Reconstructed based on RECON2 and includes mapping to 3D structure of proteins through PDB ids

Swainston, N., et al (2016). *Metabolomics*, 12(7), 109.

- Several options available ( all derived from RECON1)
- Recon3D is most recent version

# Classes of Model Extraction Methods (MEMs)

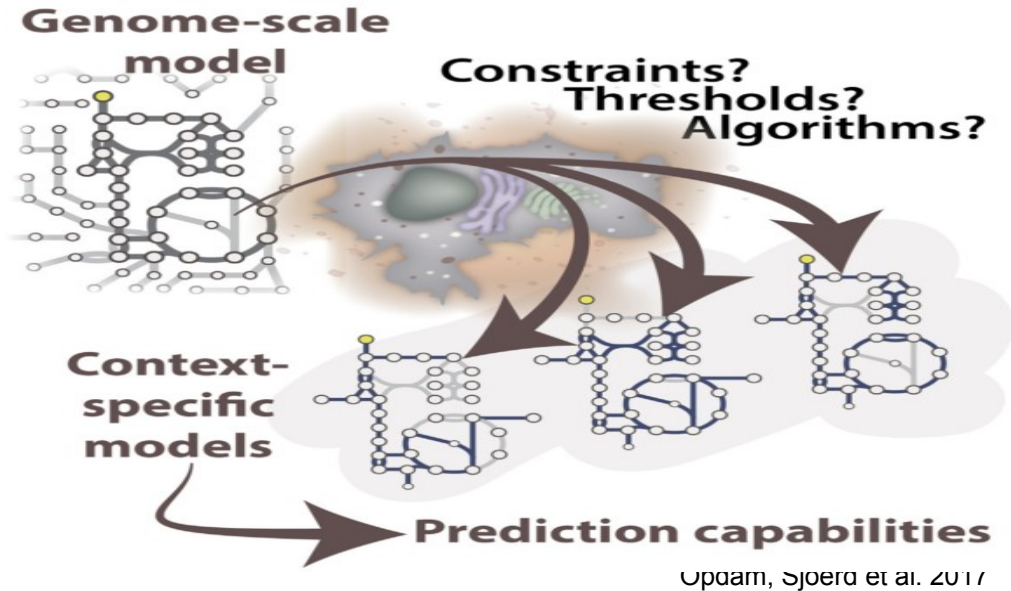


Minimizes flux through reactions associated with low gene expression

Finds an optimal trade-off between removing reactions associated with low gene expression and keeping reactions with high gene expression

The algorithms use sets of core reactions that should be retained and active while removing other reactions if possible

# Context-Specific Metabolic Modeling (CSM) depend on key decisions on methodology and data processing



- No strong evidence that a Model extraction Method gives the most accurate models
- Each method has different underlying assumptions that affect the resulting model
- **Gene expression discretization** seems to be the most determinant decision

# Hands on!

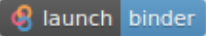
<https://github.com/migp11/ECCB2020-T05>

## ECCB2020-T05


### Computational modelling systems

This repo contains all the material for the

### Use with binder


 launch binder

1


 jupyter


Files Running Clusters


Select items to perform actions on them.

☐ 0 ▾  / ECCB2020-T05


☐ ..


☐  attractors


☐  cobrapy


☐  maboss


2

 data

 img

 out

 Tutorial.ipynb

 requirements.txt

3



ECCB 2020  
Virtual!