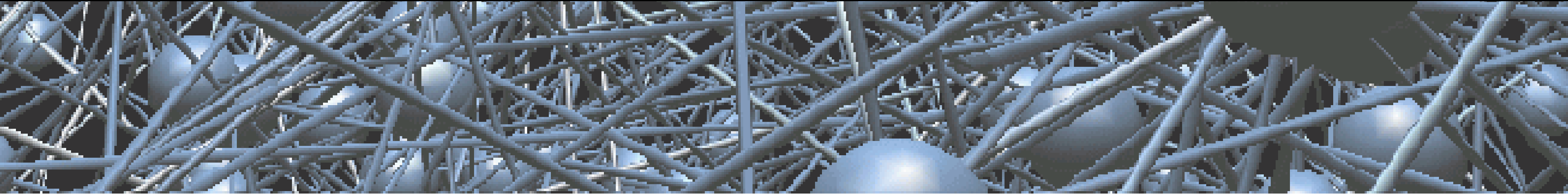
DATA INTEGRATION OF STOICHIOMETRIC MODELS



Principles and Applications of Systems Biology

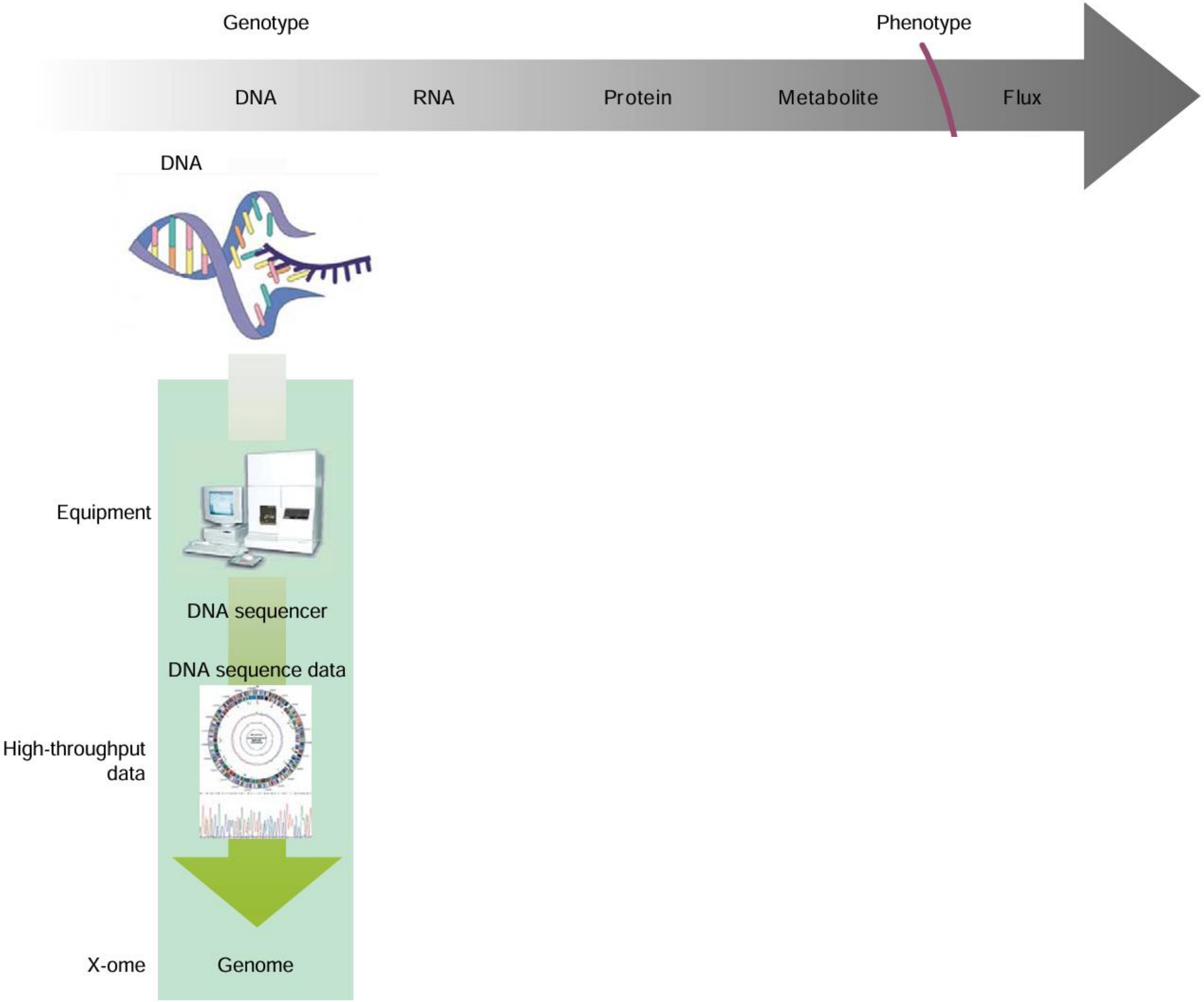
EPFL

Vassily Hatzimanikatis
October 2024

From central dogma to omics

Lee et al., Trends in Biotechnology, 2005

From central dogma to omics



Genotype

Phenotype

DNA

RNA

Protein

Metabolite

Flux

DNA

Equipment

DNA sequencer

DNA sequence data

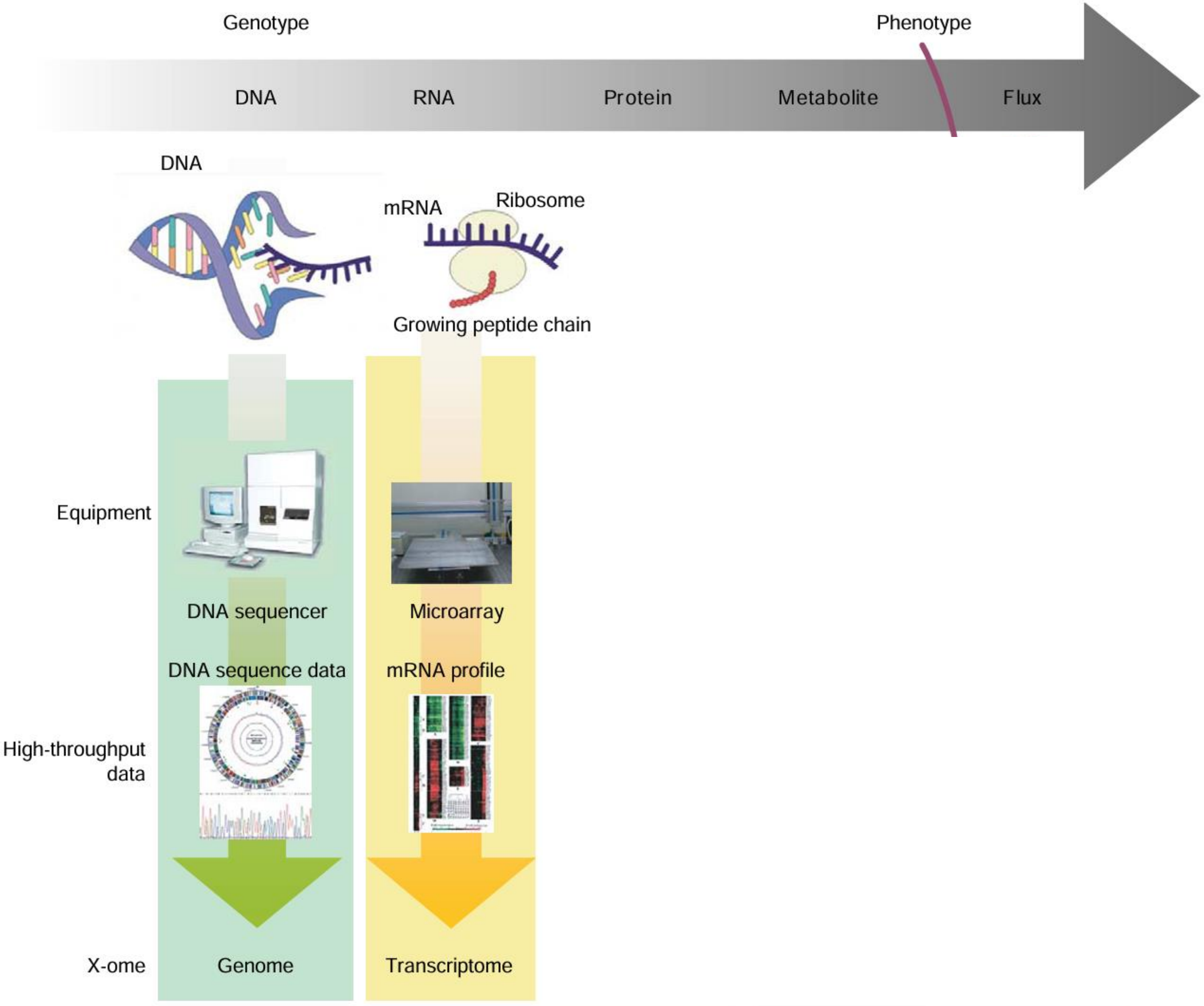
High-throughput data

X-ome

Genome

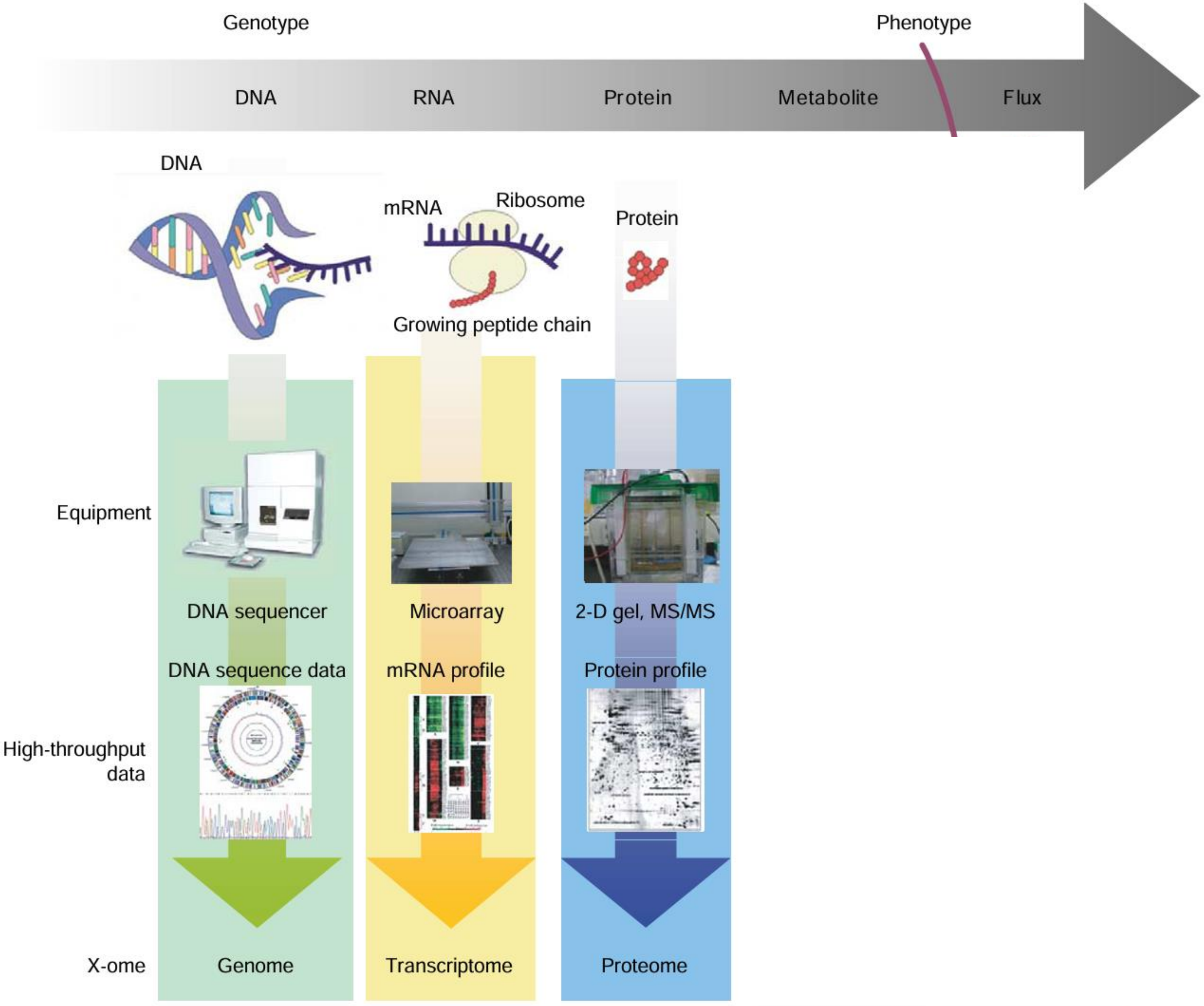
Lee et al., Trends in Biotechnology, 2005

From central dogma to omics



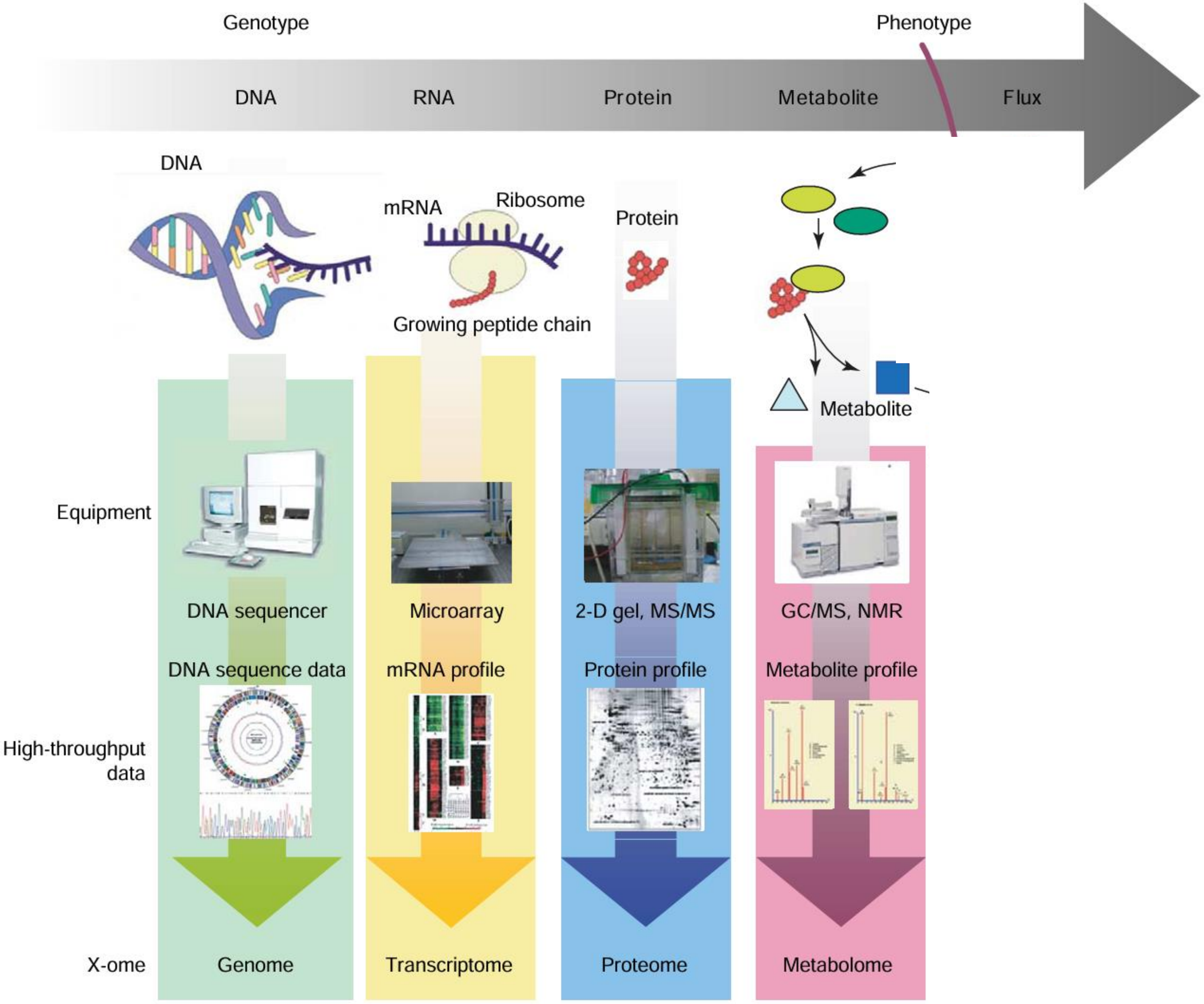
Lee et al., Trends in Biotechnology, 2005

From central dogma to omics



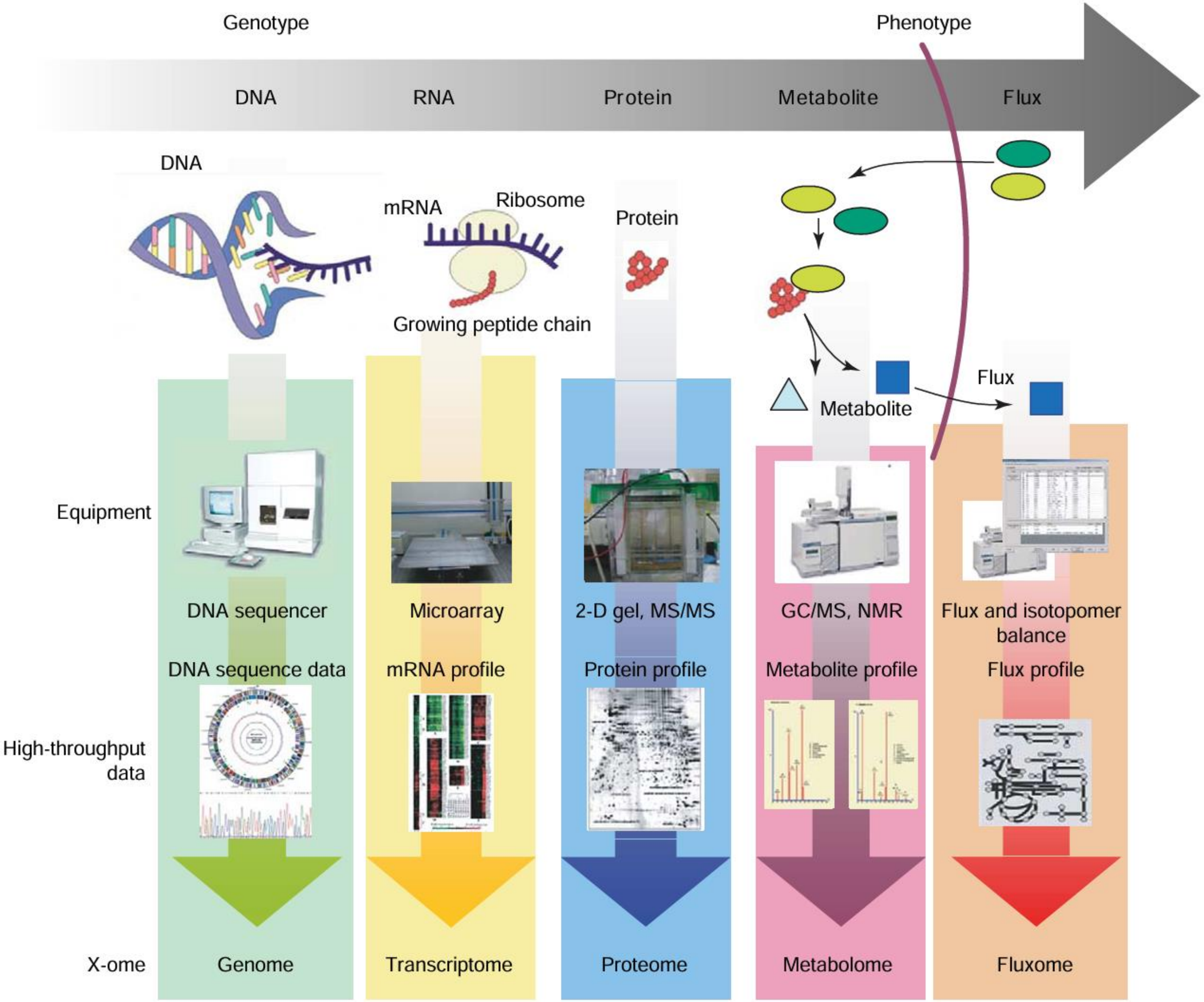
Lee et al., Trends in Biotechnology, 2005

From central dogma to omics



Lee et al., Trends in Biotechnology, 2005

From central dogma to omics



Lee et al., Trends in Biotechnology, 2005

From omics to GEMs

Genome

Transcriptome

Proteome

Metabolome

Fluxome

From omics to GEMs

Genome

Transcriptome

Proteome

Metabolome

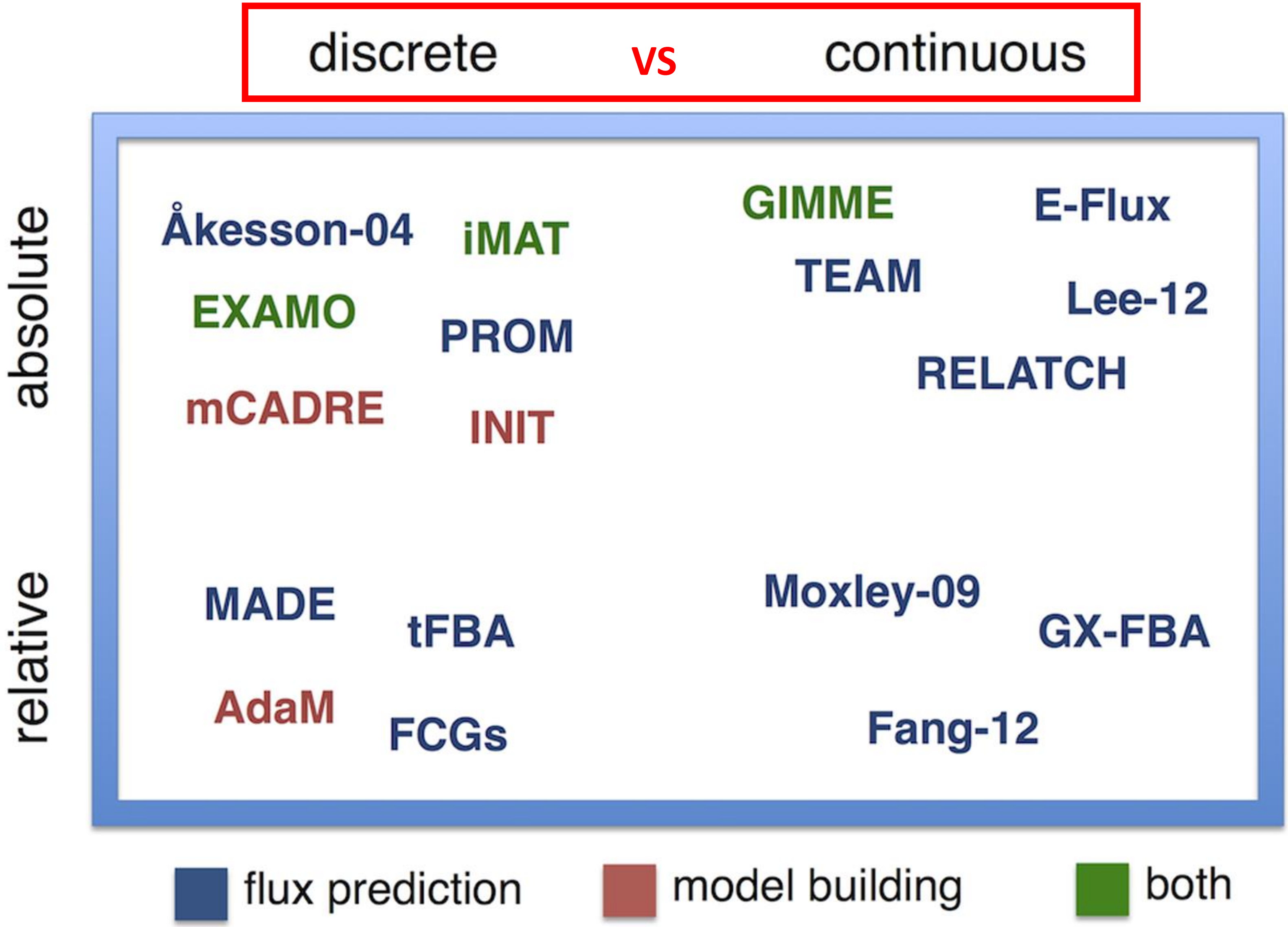
Fluxome

Build context specific models

&

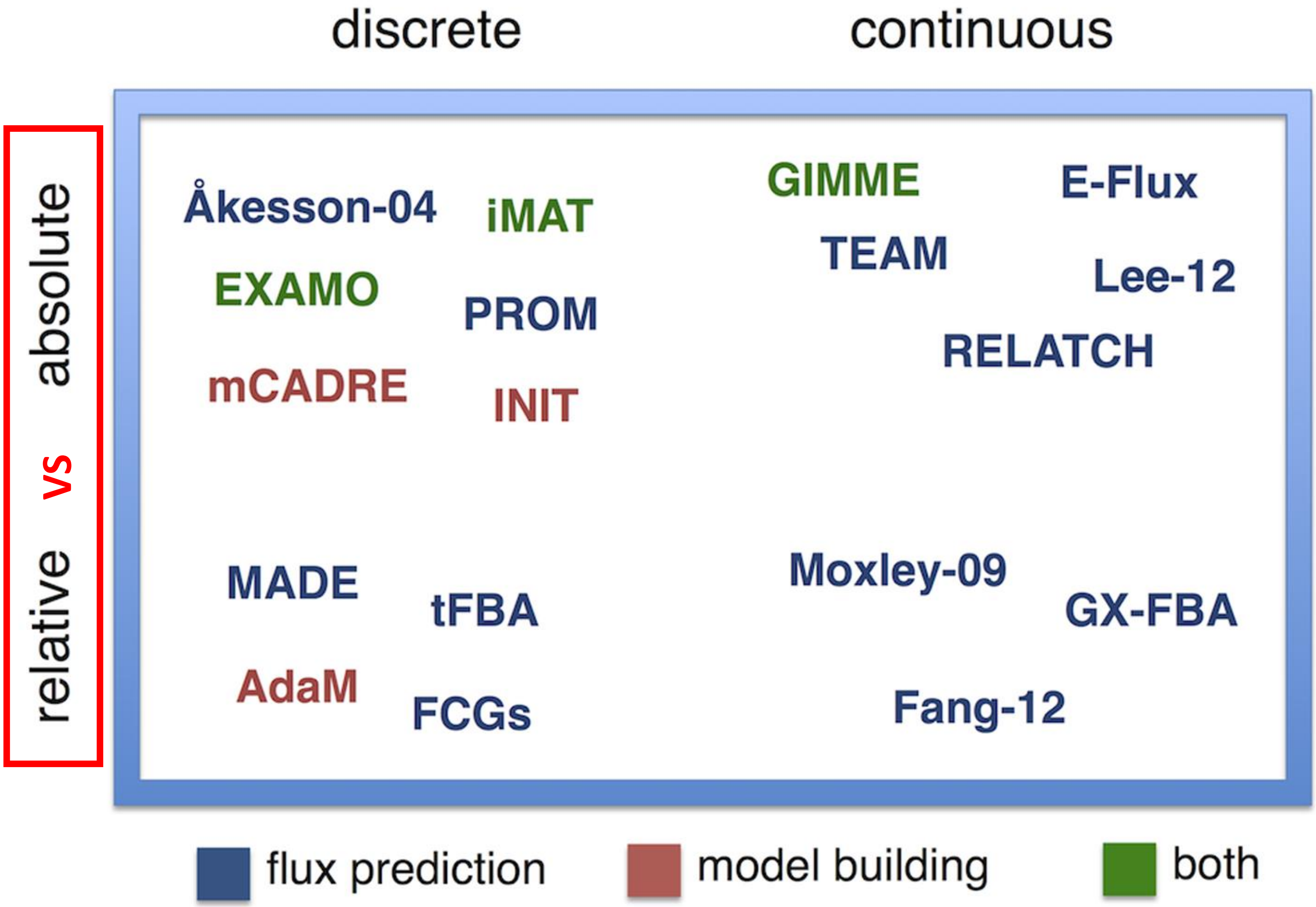
Improve flux prediction

Data integration into GEMs



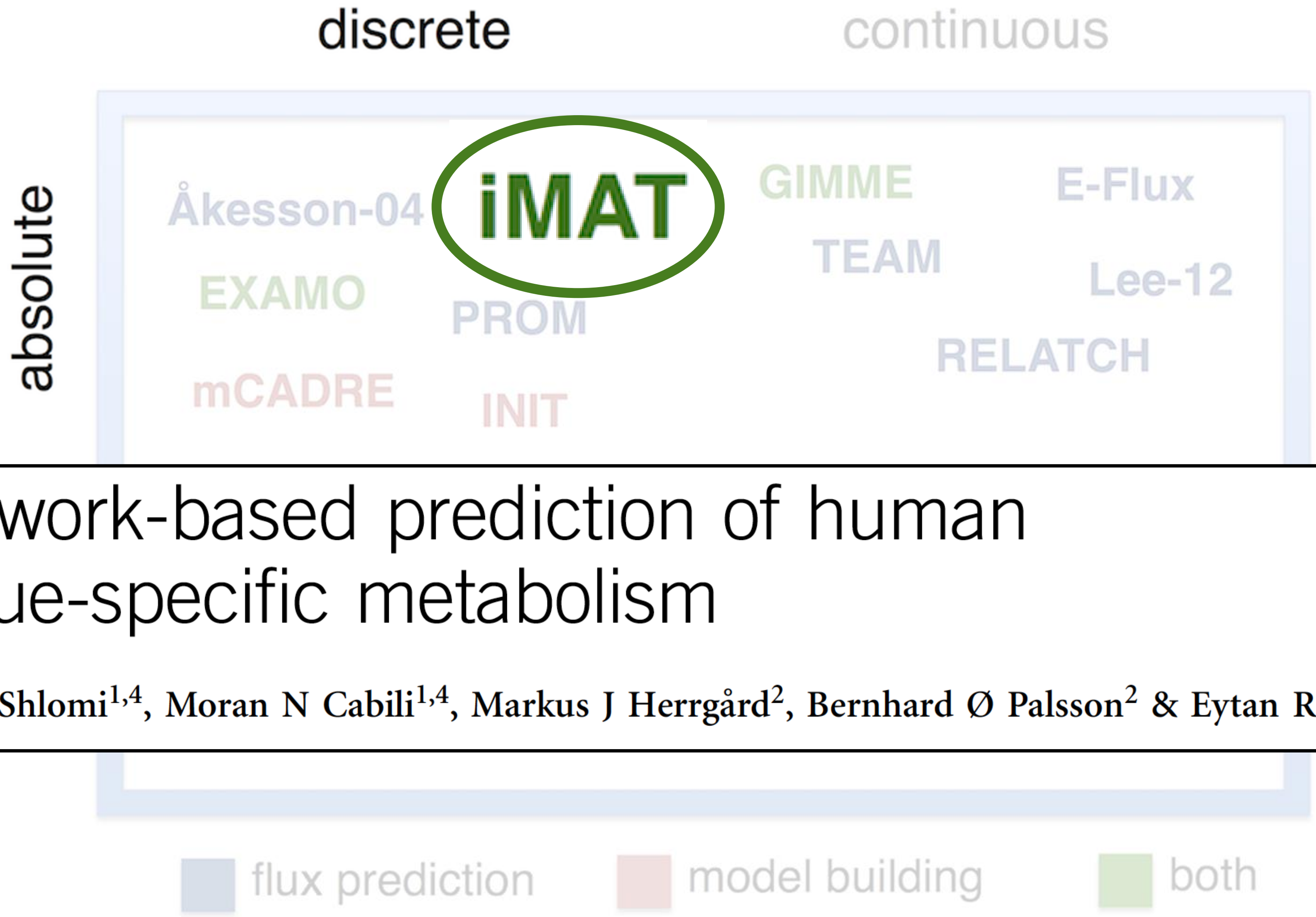
Machado and Herrgard, PLOS Computational Biology, 2014

Data integration into GEMs



Machado and Herrgard, PLOS Computational Biology, 2014

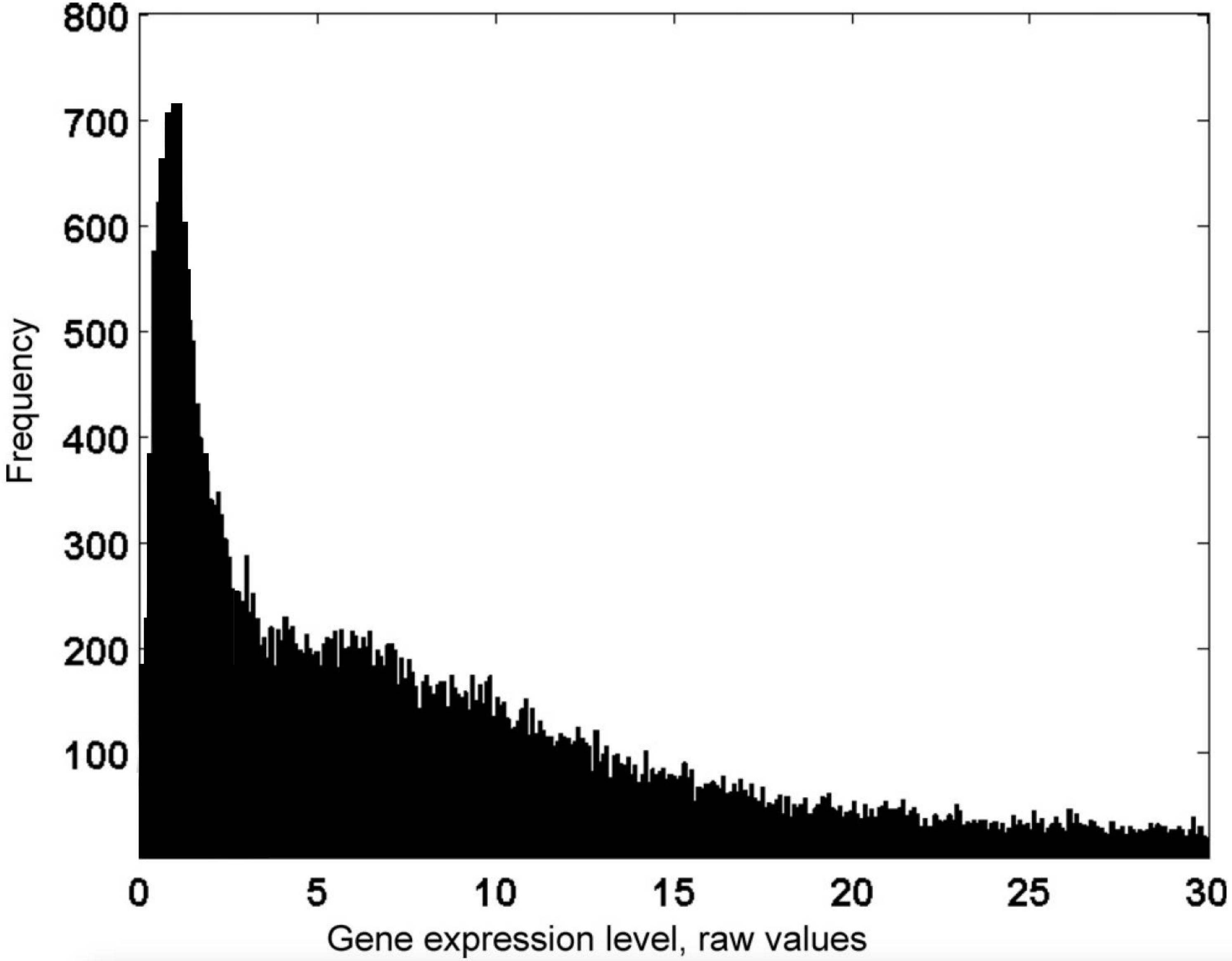
iMAT method



Network-based prediction of human tissue-specific metabolism

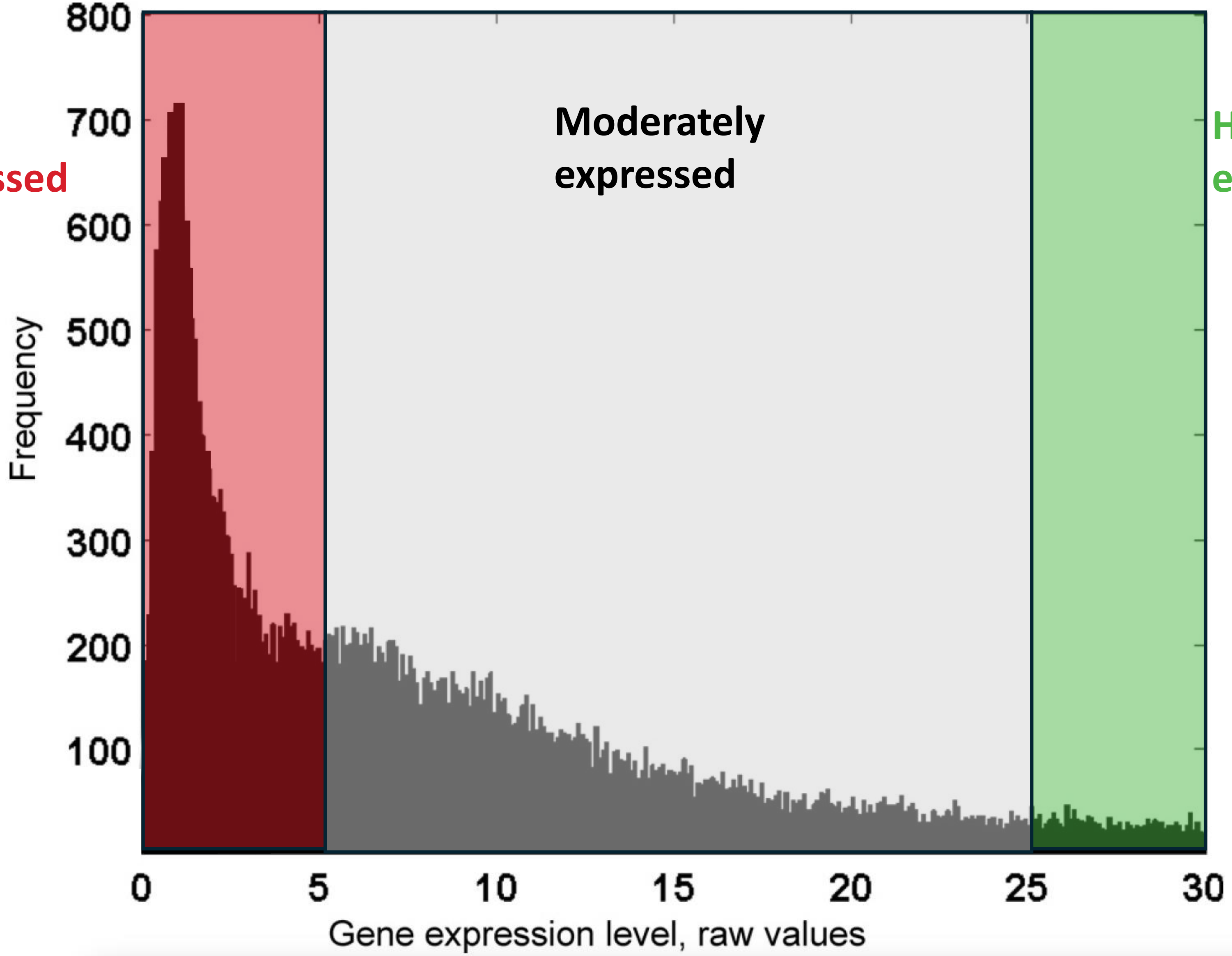
Tomer Shlomi[1,4], Moran N Cabili[1,4], Markus J Herrgård[2], Bernhard Ø Palsson[2] & Eytan Ruppin[1,3]

Gene expression level distribution



adapted from Dozmorov et al., BMC Infectious Diseases, 2009

Gene expression level distribution



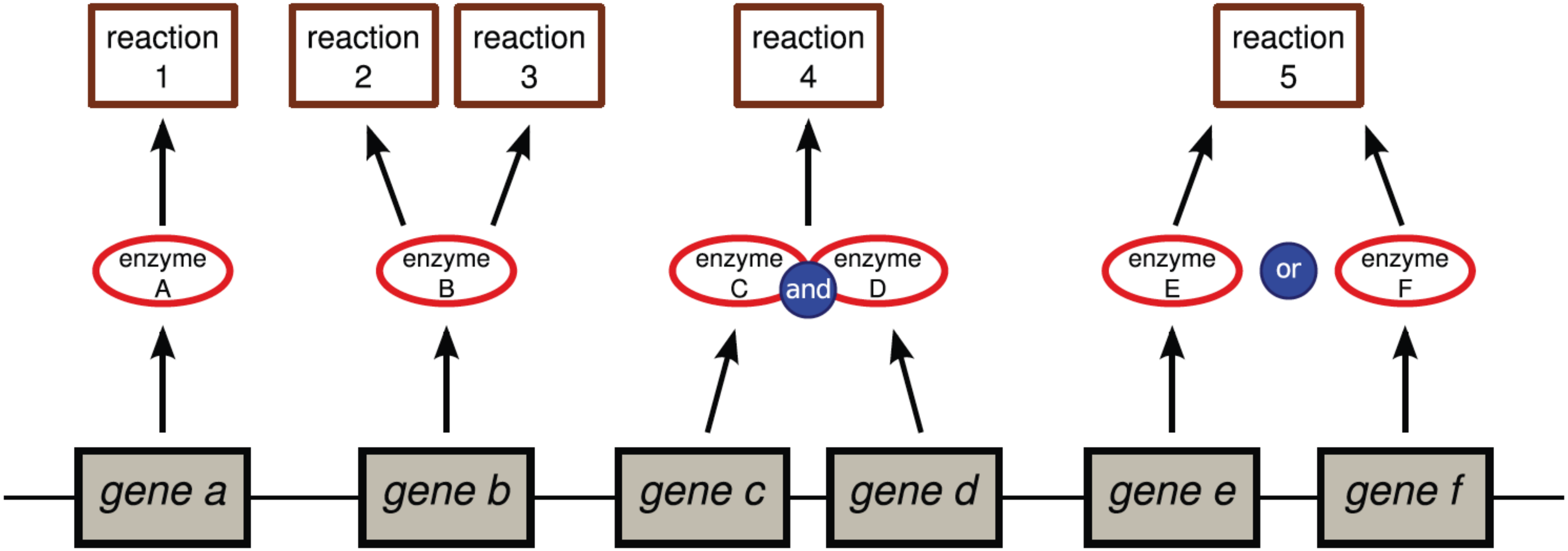
Lowly expressed

Moderately expressed

Highly expressed

adapted from Dozmorov et al., BMC Infectious Diseases, 2009

Gene-Protein-Reaction (GPR) rules

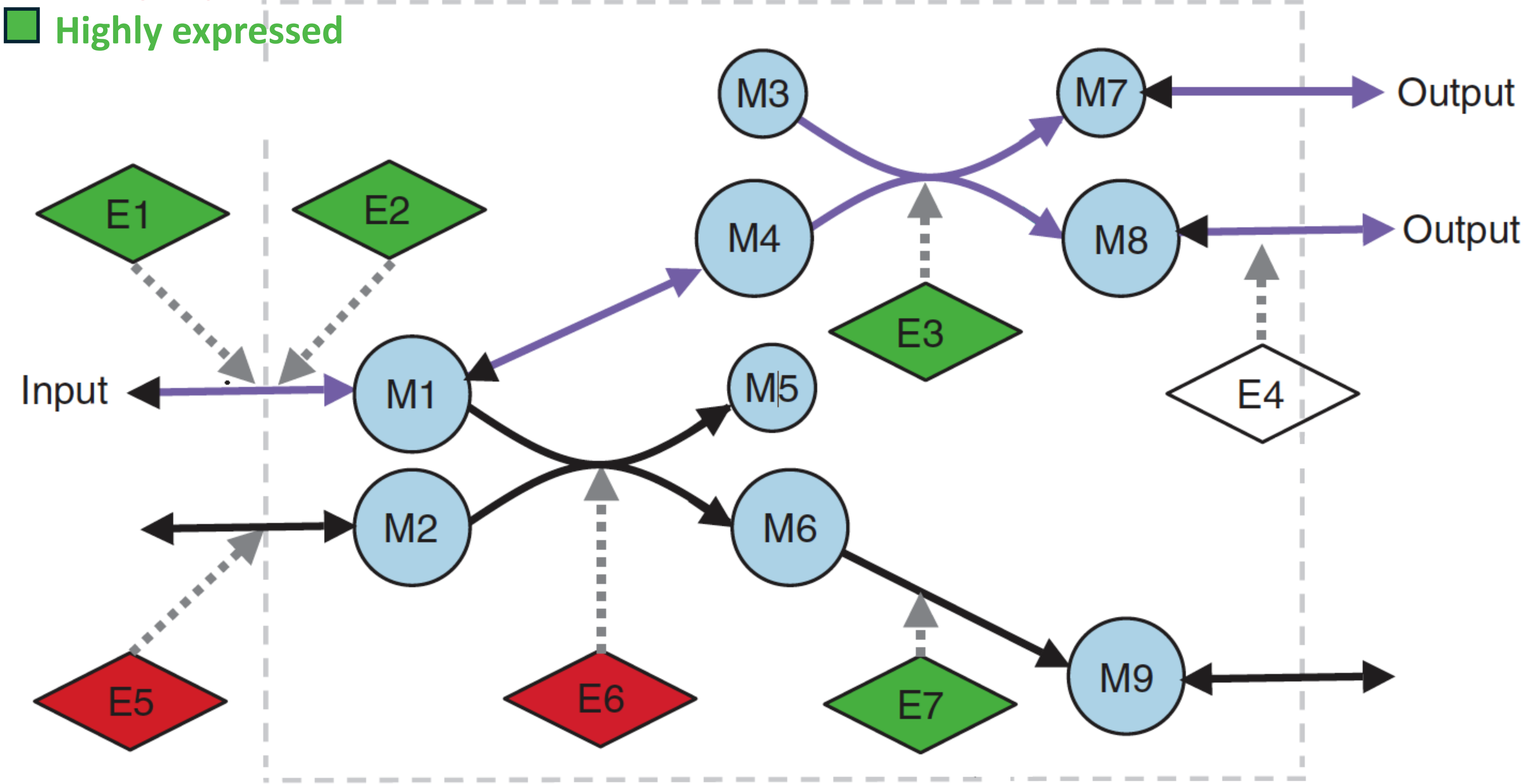


Jensen et al., BMC Systems Biology, 2011

Shlomi, Cabili et al., Nature Biotechnology, 2008

iMAT method (example)

☐ **Moderately expressed**
🟥 **Lowly expressed**
🟩 **Highly expressed**

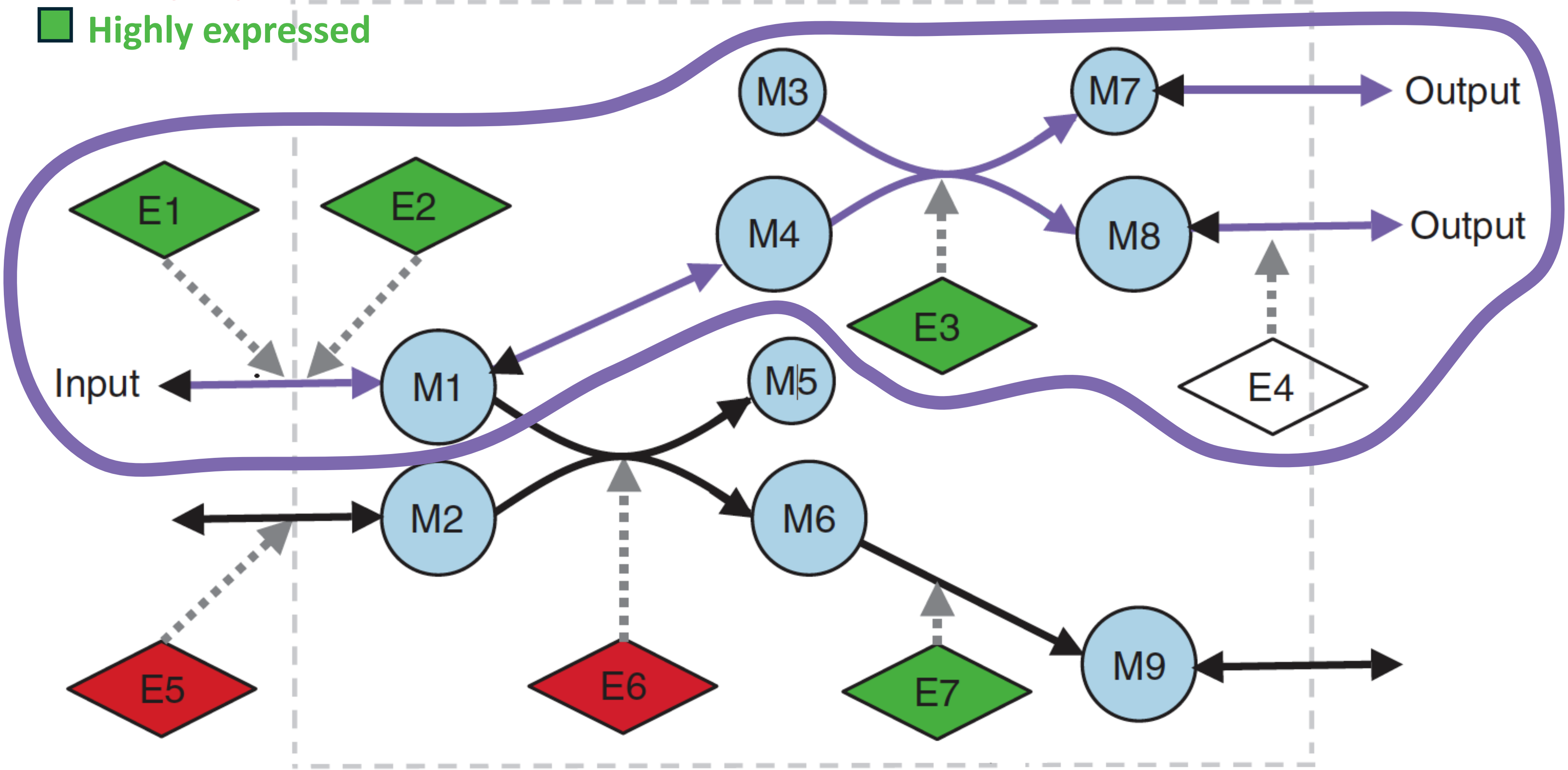


iMAT method (example)

☐ **Moderately expressed**
🟥 **Lowly expressed**
🟩 **Highly expressed**

Maximize data integration score

(E1, E2, E3, E5, E6, E7)

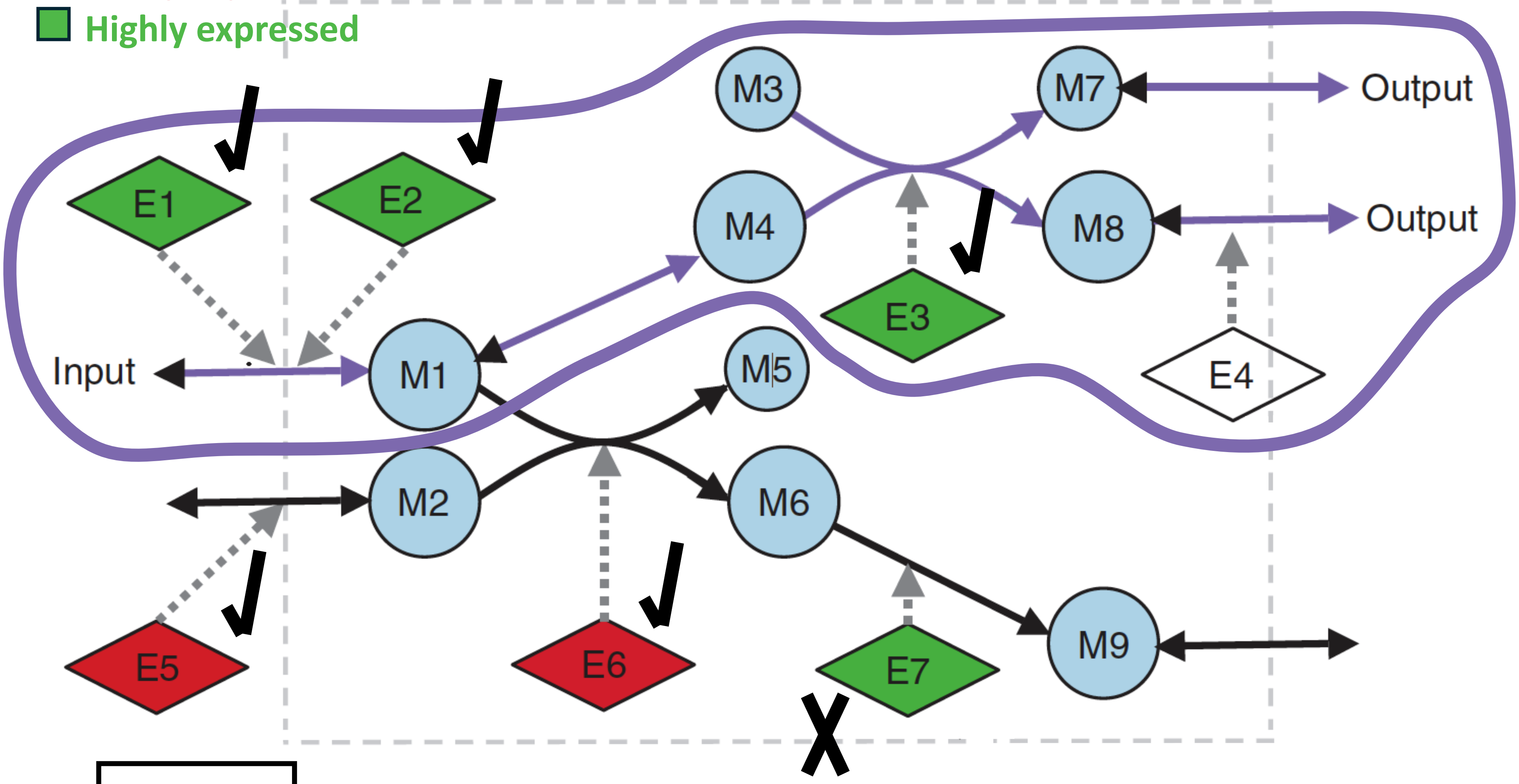


Shlomi, Cabili et al., Nature Biotechnology, 2008

iMAT method (example)

☐ **Moderately expressed**
🟥 **Lowly expressed**
🟩 **Highly expressed**

Maximize data integration score
(E1, E2, E3, E5, E6, E7)



5/6

Shlomi, Cabili et al., Nature Biotechnology, 2008

iMAT method (example)

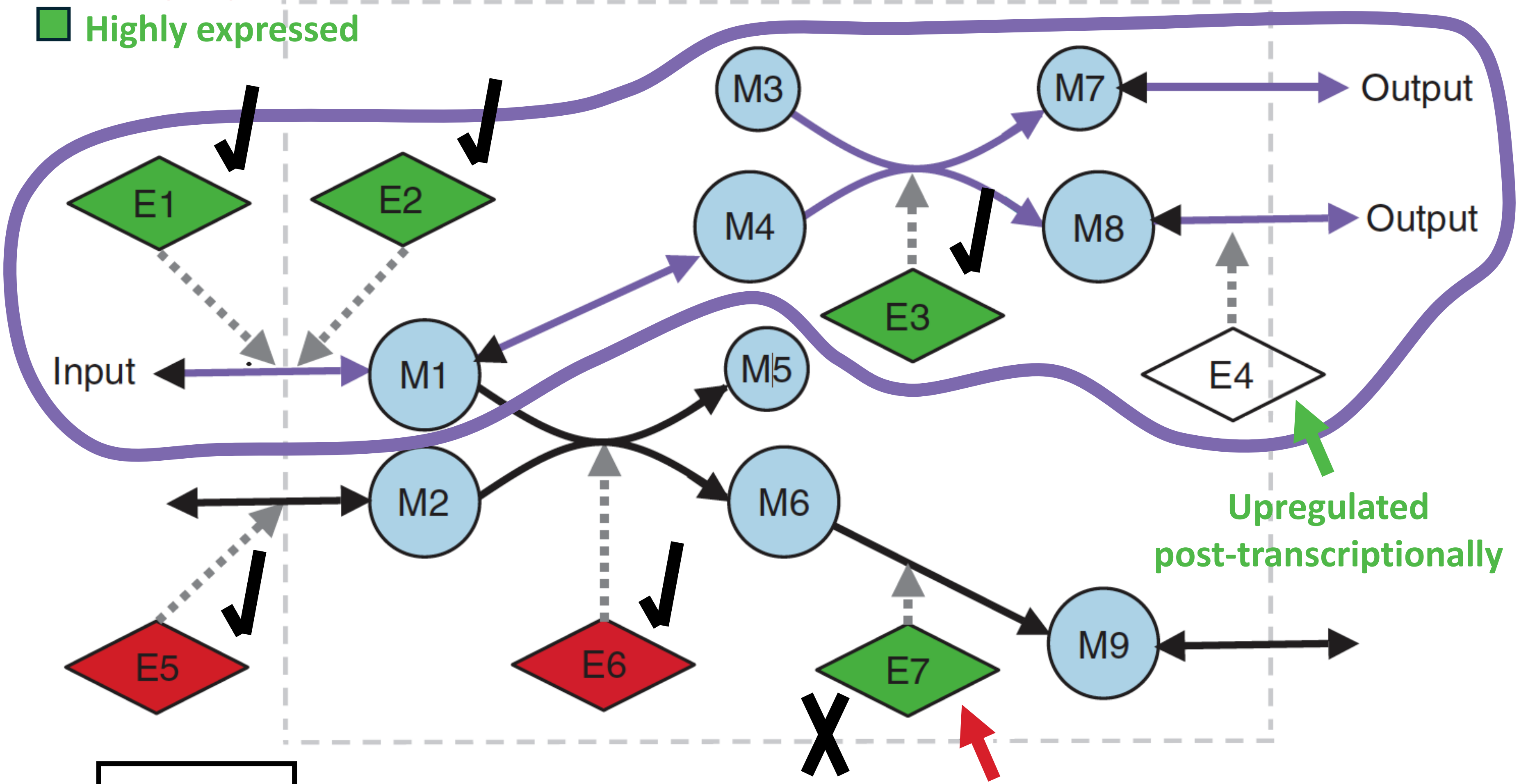
☐ **Moderately expressed**

🟥 **Lowly expressed**

🟩 **Highly expressed**

Maximize data integration score

(E1, E2, E3, E5, E6, E7)



Upregulated post-transcriptionally

Downregulated post-transcriptionally

5/6

iMAT method

MILP formulation

$$\max_{v,y^+,y^-}\left(\sum\nolimits_{i\in R_H}(y_i^+ + y_i^-) + \sum\nolimits_{i\in R_L} y_i^+\right)$$

$$s.t$$

$$S\cdot v = 0$$

$$v_{\min} \le v \le v_{\max}$$

$$v_i + y_i^+\left(v_{\min,i} - \varepsilon\right) \ge v_{\min,i}, i \in \ R_H$$

$$v_i + y_i^-\left(v_{\max,i} + \varepsilon\right) \le v_{\max,i}; i \in \ R_H$$

$$v_{\min,i}(1 - y_i^+) \le v_i \le v_{\max,i}(1 - y_i^+), i \in \ R_L$$

$$v \in R^m$$

$$y_i^+,\ y_i^- \in \ [0,1]$$

R_H : highly expressed reactions

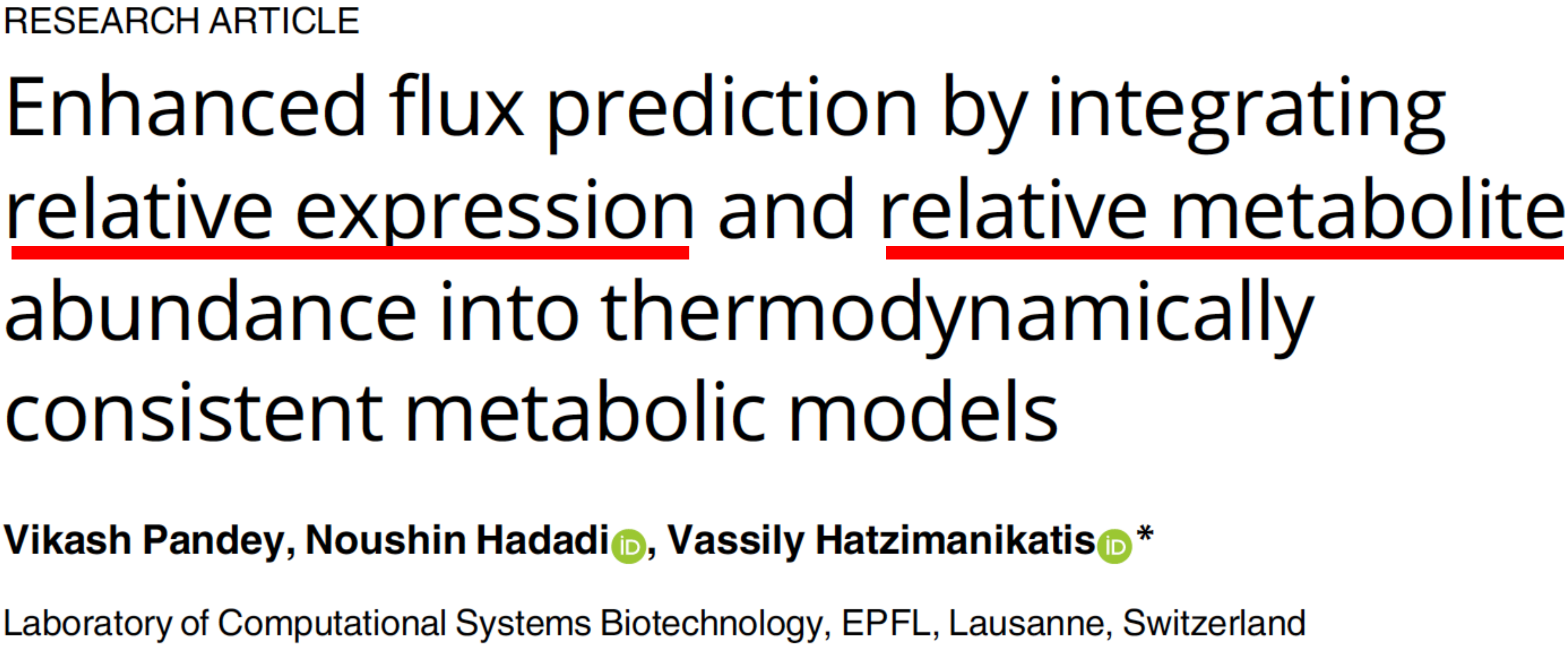
R_L : lowly expressed reactions

y_i^+ : reaction is active

y_i^- : reaction is not active

Shlomi, Cabili et al., Nature Biotechnology, 2008

REMI method



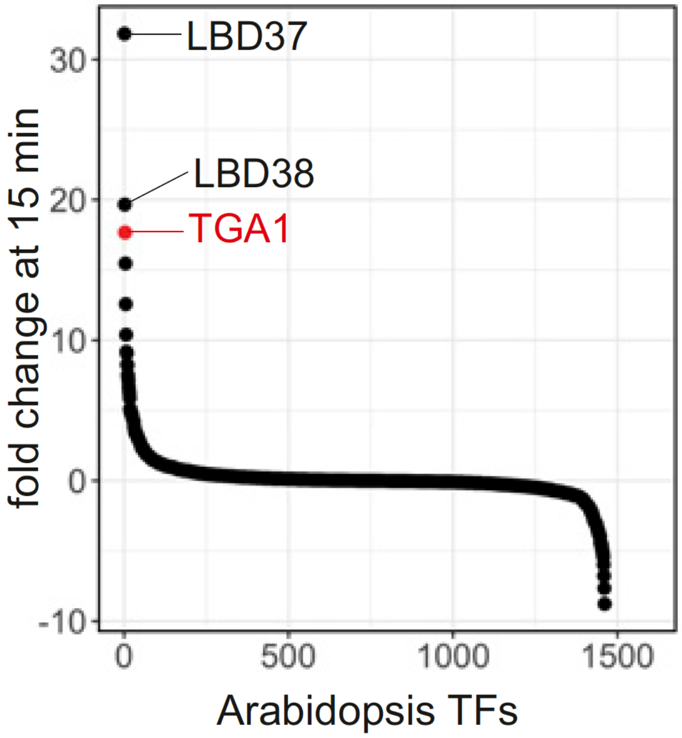
RESEARCH ARTICLE

Enhanced flux prediction by integrating relative expression and relative metabolite abundance into thermodynamically consistent metabolic models

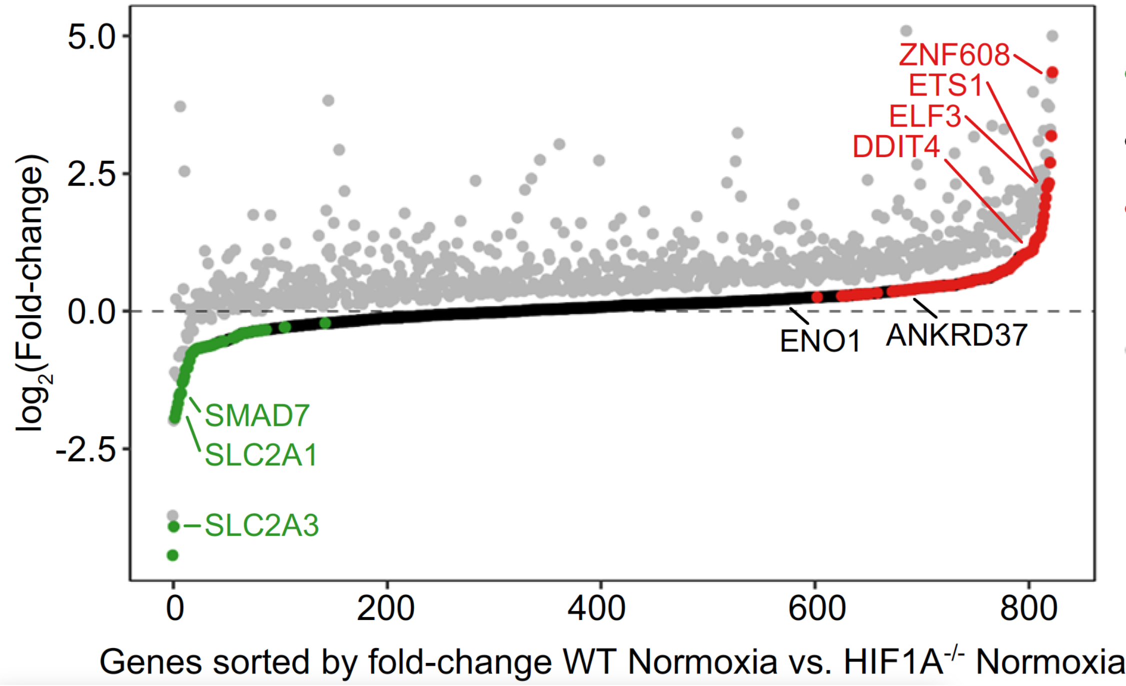
Vikash Pandey, Noushin Hadadi, Vassily Hatzimanikatis*

Laboratory of Computational Systems Biotechnology, EPFL, Lausanne, Switzerland

Expression deregulation distribution

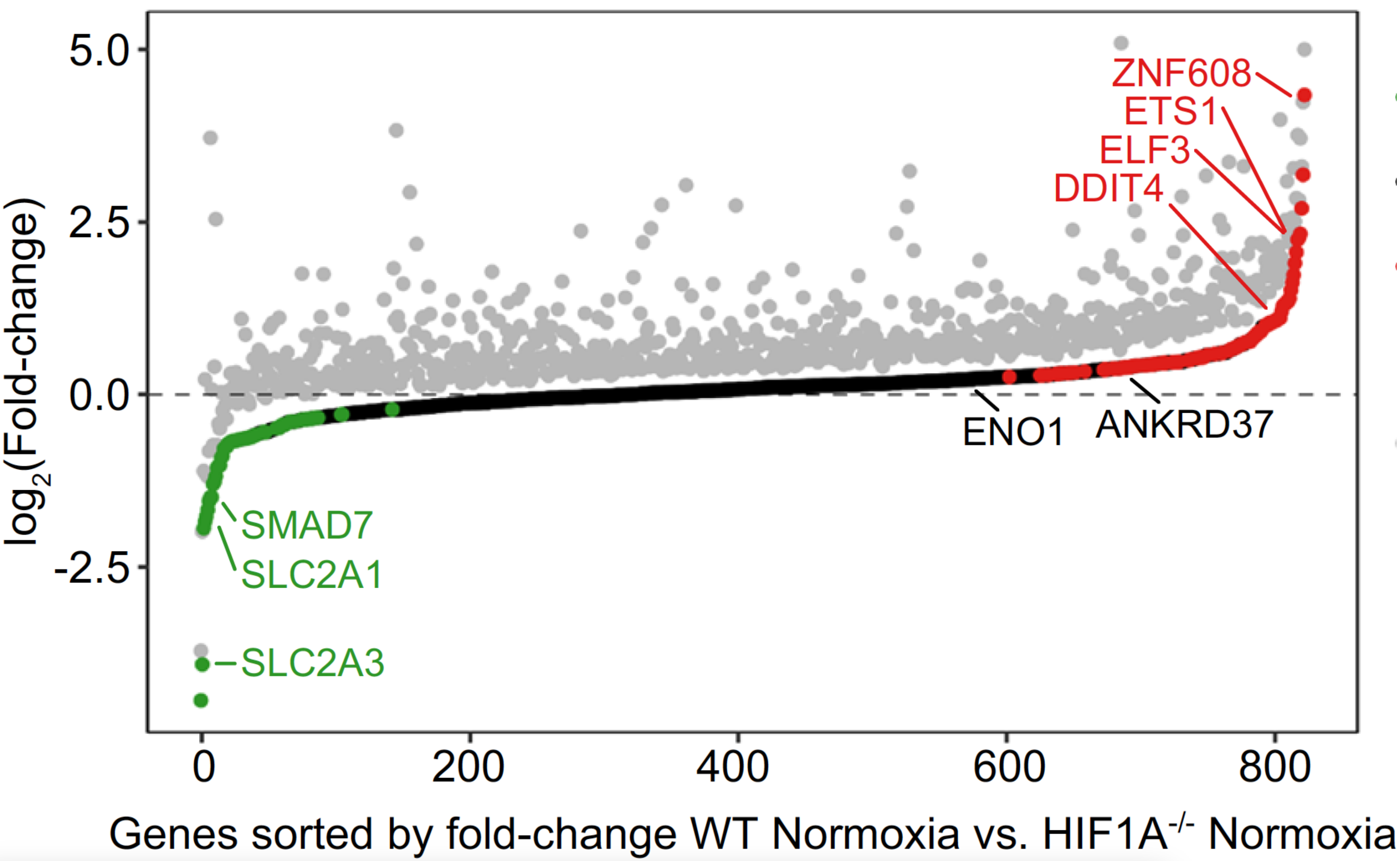


Andrysik et al., Nature Communications, 2021



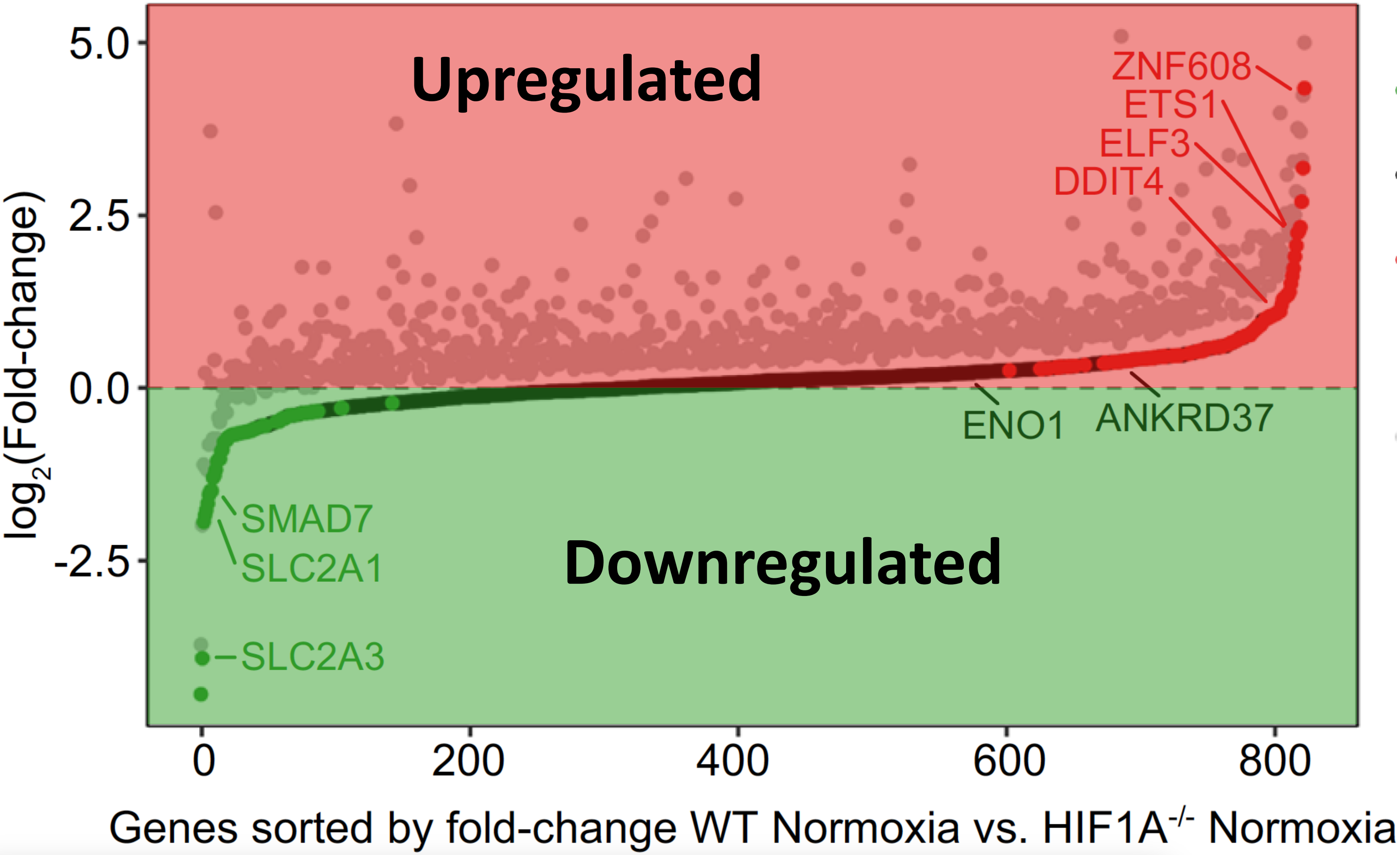
Swift, Alvarez et al., PNAS, 2020

Expression deregulation distribution



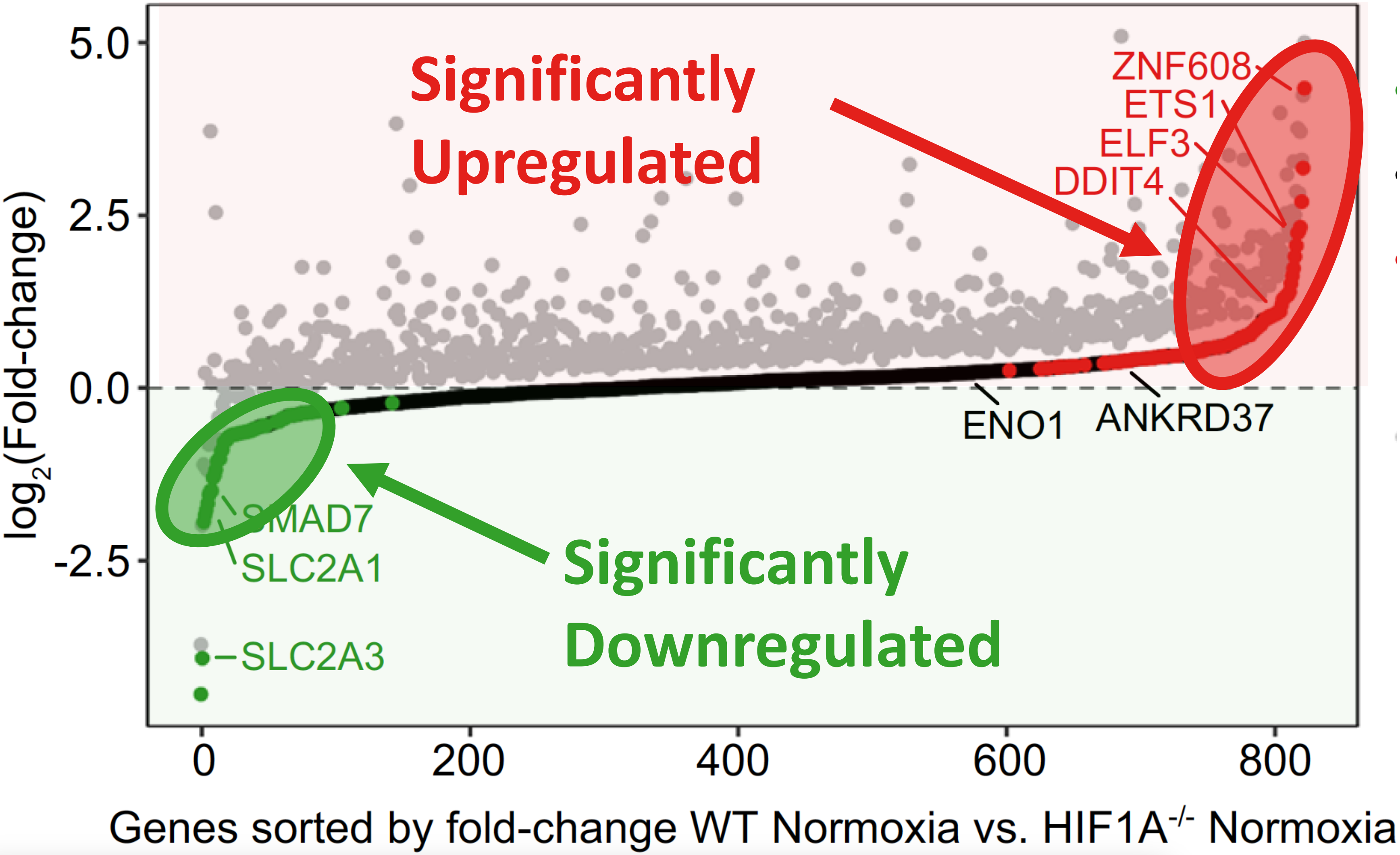
Swift, Alvarez et al., PNAS, 2020

Expression deregulation distribution



Swift, Alvarez et al., PNAS, 2020

Expression deregulation distribution

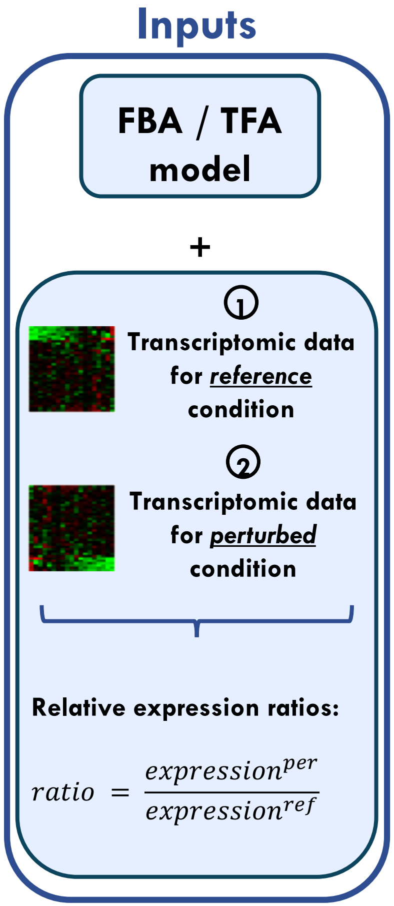


Significantly Upregulated

Significantly Downregulated

Swift, Alvarez et al., PNAS, 2020

REMI workflow for expression data integration



Inputs

FBA / TFA model

+

①

Transcriptomic data for *reference* condition

②

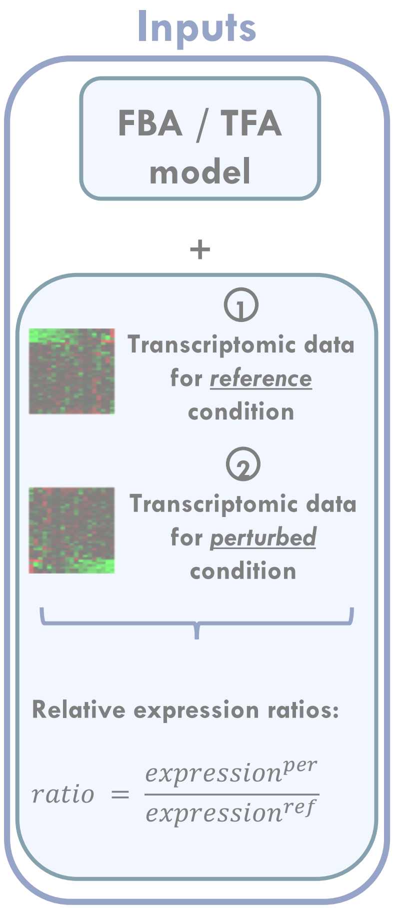
Transcriptomic data for *perturbed* condition

Relative expression ratios:

$$ratio = \frac{expression^{per}}{expression^{ref}}$$

Pandey et al., PLOS Computational Biology, 2019

REMI workflow for expression data integration



Inputs

FBA / TFA model

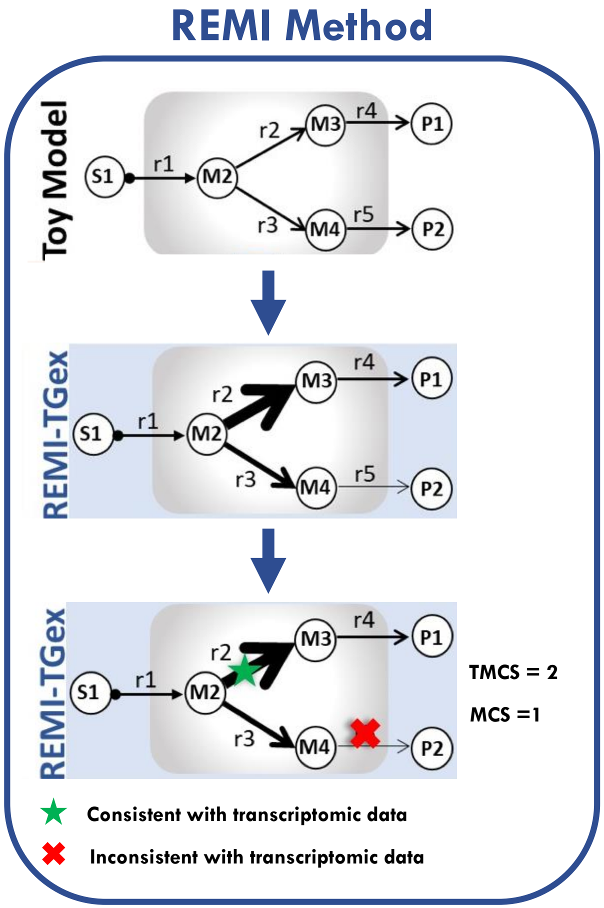
+

① Transcriptomic data for *reference* condition

② Transcriptomic data for *perturbed* condition

Relative expression ratios:

$$ratio = \frac{expression^{per}}{expression^{ref}}$$



REMI Method

TMCS = 2

MCS =1

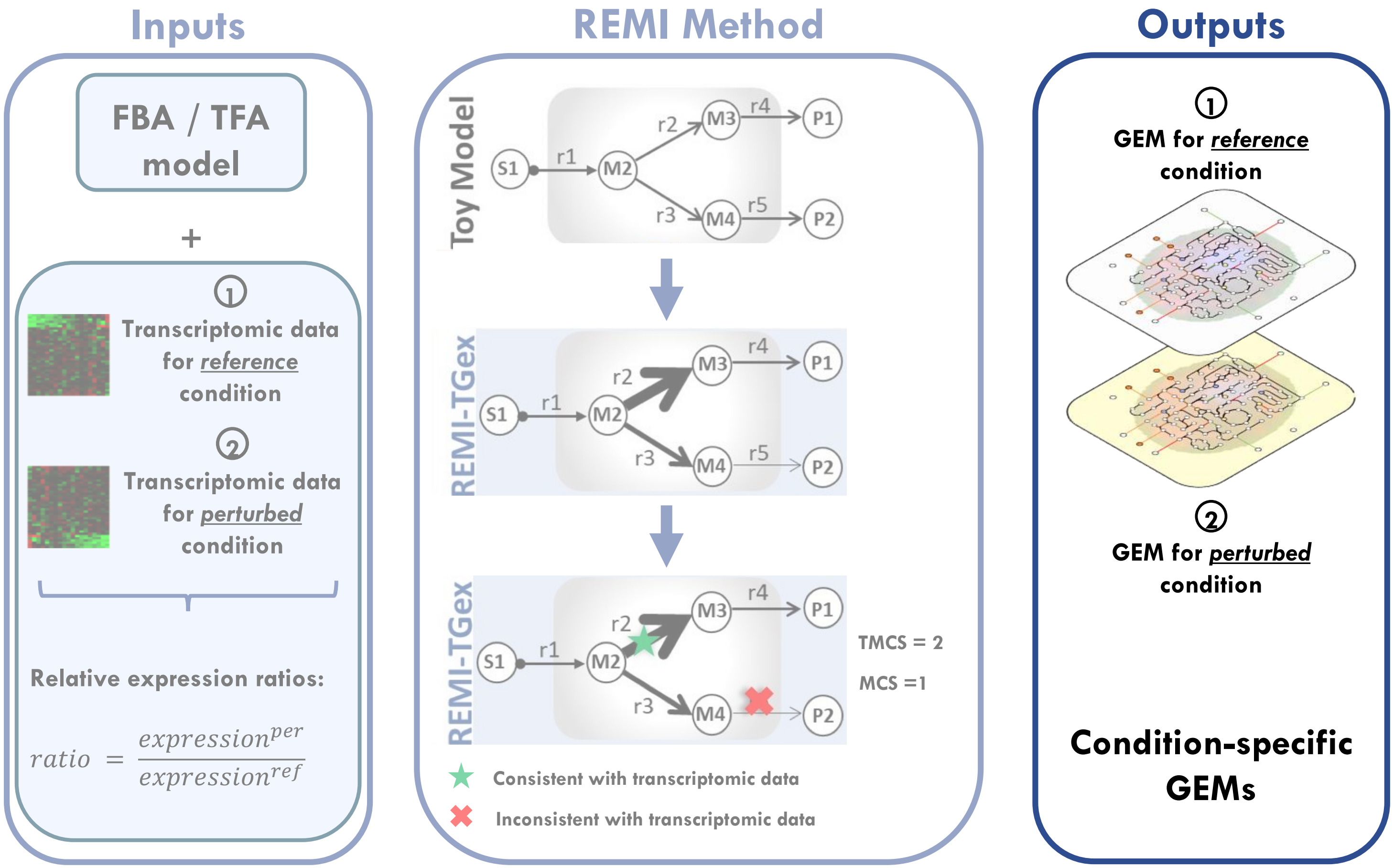
★ Consistent with transcriptomic data

✖ Inconsistent with transcriptomic data

* TMCS = Theoretical Maximum Consistency Score
MCS = Maximum Consistency Score

Pandey et al., PLOS Computational Biology, 2019

REMI workflow for expression data integration



Inputs

FBA / TFA model

\+

① Transcriptomic data for *reference* condition

② Transcriptomic data for *perturbed* condition

Relative expression ratios:

$$ratio = \frac{expression^{per}}{expression^{ref}}$$

REMI Method

TMCS = 2

MCS =1

★ Consistent with transcriptomic data

✖ Inconsistent with transcriptomic data

Outputs

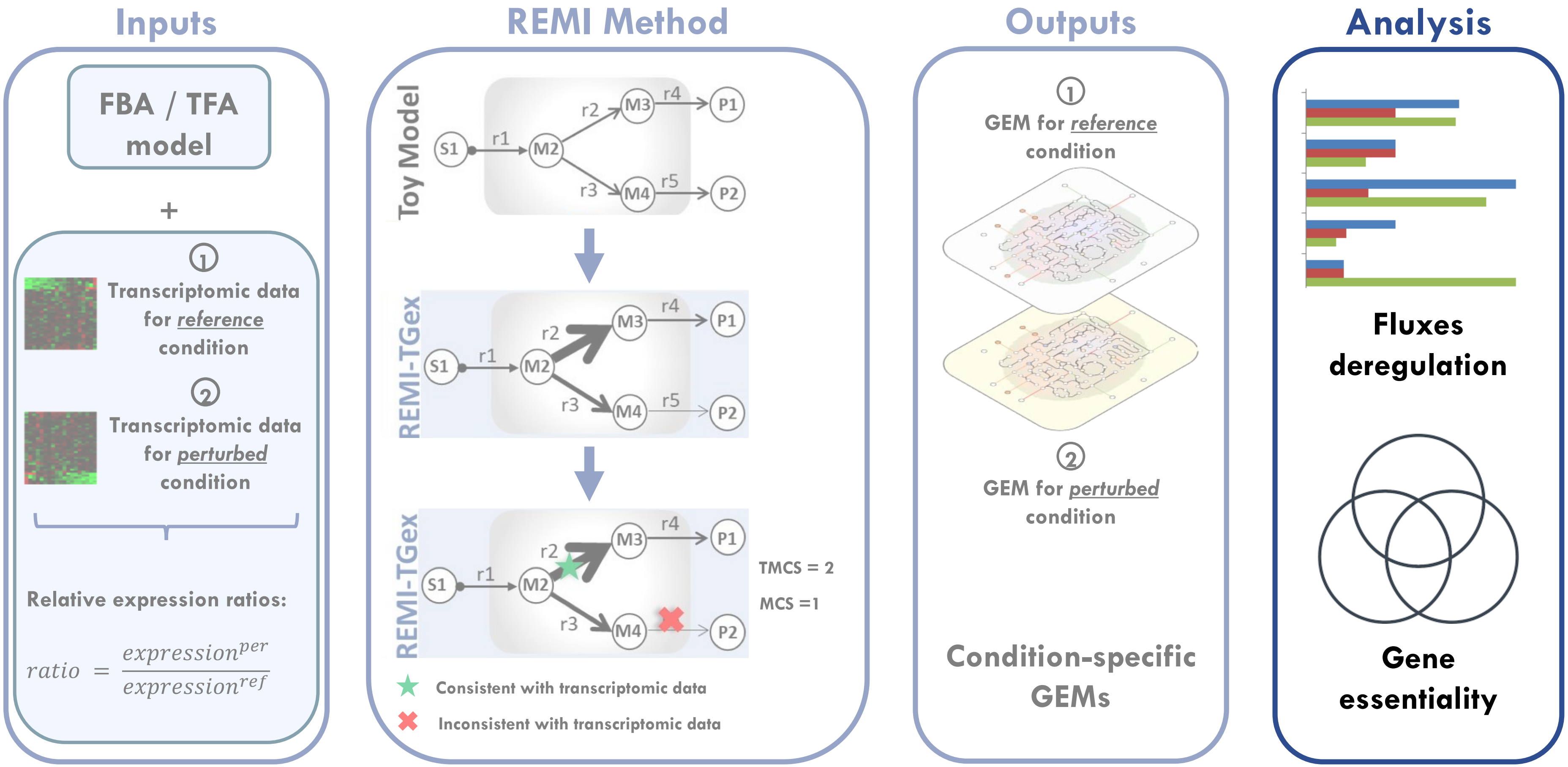
① GEM for *reference* condition

② GEM for *perturbed* condition

Condition-specific GEMs

Pandey et al., PLOS Computational Biology, 2019

REMI workflow for expression data integration



Inputs

FBA / TFA model

\+

① Transcriptomic data for *reference* condition

② Transcriptomic data for *perturbed* condition

Relative expression ratios:

$$ratio = \frac{expression^{per}}{expression^{ref}}$$

REMI Method

TMCS = 2

MCS =1

★ Consistent with transcriptomic data

✖ Inconsistent with transcriptomic data

Outputs

① GEM for *reference* condition

② GEM for *perturbed* condition

Condition-specific GEMs

Analysis

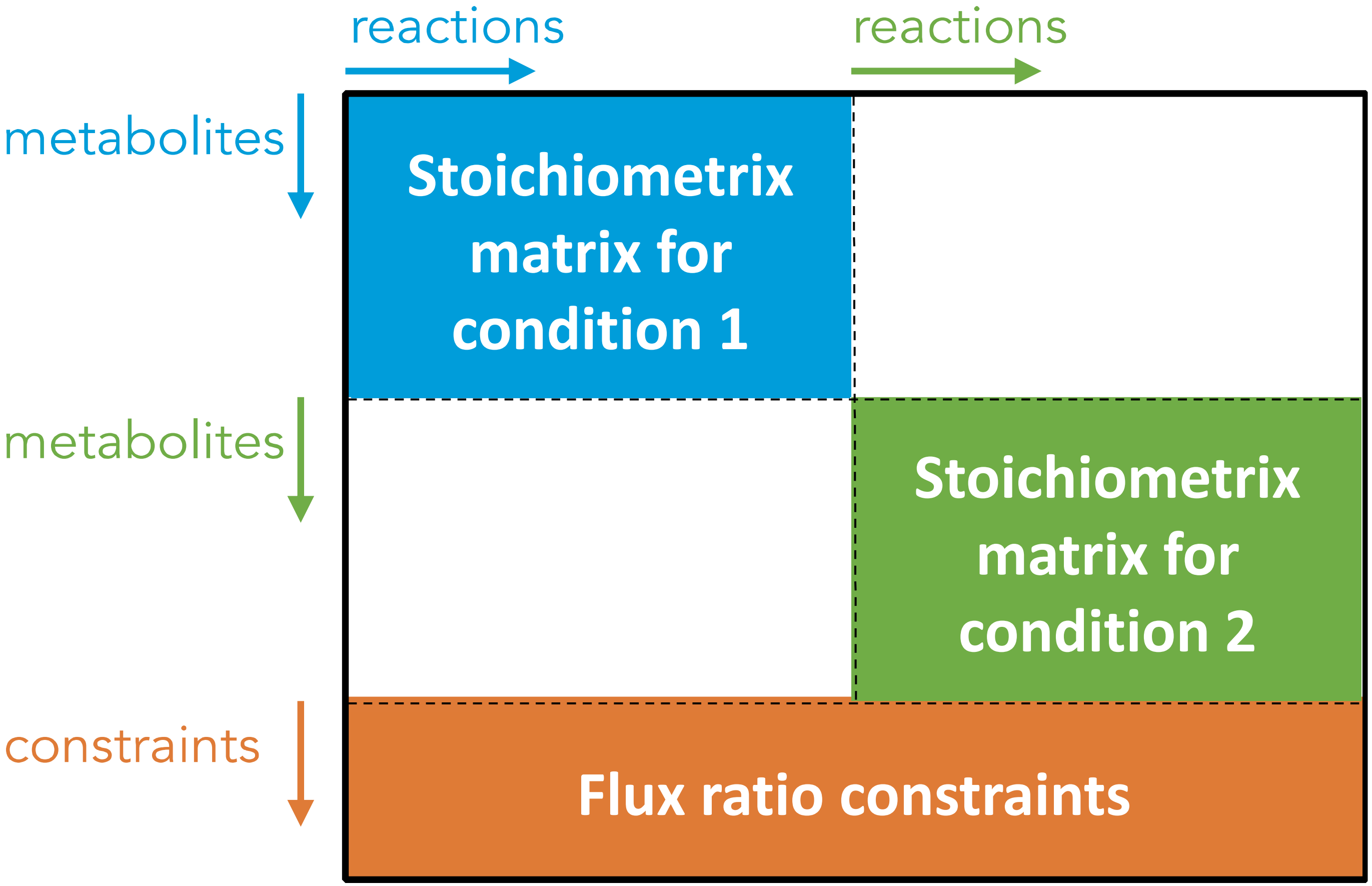
Fluxes deregulation

Gene essentiality

Pandey et al., PLOS Computational Biology, 2019

REMI method

Master S matrix



Stoichiometrix matrix for condition 1

Stoichiometrix matrix for condition 2

Flux ratio constraints

Pandey et al., PLOS Computational Biology, 2019

REMI method

MILP formulation

$$Consistency\ Score = Maximize \sum_{i=1}^{n} B_i$$

$$n = u + a$$

$$v_i^{mutant} \geq \varepsilon * B_i \quad [i = 1, .., n\ for\ both\ upregulated\ and\ downregulated\ reactions]$$

$$v_i^{wild} \geq \varepsilon * B_i \quad [i = 1, .., n\ for\ both\ upregulated\ and\ downregulated\ reactions]$$

$$v_i^{mutant} \geq p * v_r^{wild} - \sigma_i\ [i = 1, .., u\ for\ upregulated\ reactions]$$

$$v_i^{mutant} \leq p * v_r^{wild} + \sigma_i\ [i = 1, .., d\ for\ downregulated\ reactions]$$

$$\varepsilon * (1 - B_i) \leq \sigma_i \leq \varepsilon + (1 - B_i) * M' \quad i = 1, .., n$$

Pandey et al., PLOS Computational Biology, 2019

REMI method

MILP formulation

$$Consistency\ Score = Maximize \sum_{i=1}^{n} B_i$$

$$n = u + a$$

$$v_i^{mutant} \geq \varepsilon * B_i \quad [i = 1, .., n\ for\ both\ upregulated\ and\ downregulated\ reactions]$$

$$v_i^{wild} \geq \varepsilon * B_i \quad [i = 1, .., n\ for\ both\ upregulated\ and\ downregulated\ reactions]$$

Fold change ratios

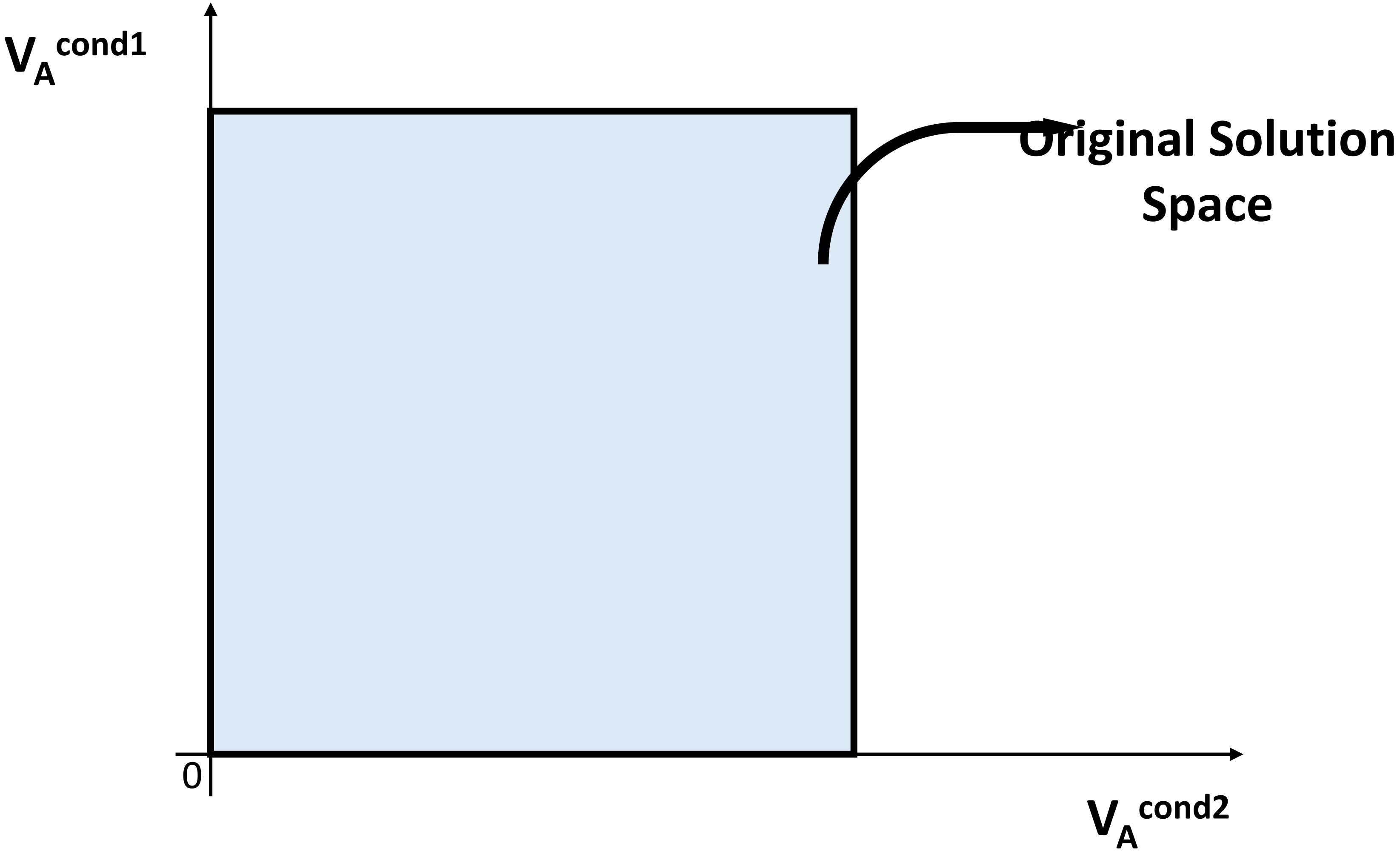
$$v_i^{mutant} \geq p * v_r^{wild} - \sigma_i \ [i = 1, .., u\ for\ upregulated\ reactions]$$

$$v_i^{mutant} \leq p * v_r^{wild} + \sigma_i \ [i = 1, .., d\ for\ downregulated\ reactions]$$

$$\varepsilon * (1 - B_i) \leq \sigma_i \leq \varepsilon + (1 - B_i) * M' \quad i = 1, .., n$$

Pandey et al., PLOS Computational Biology, 2019

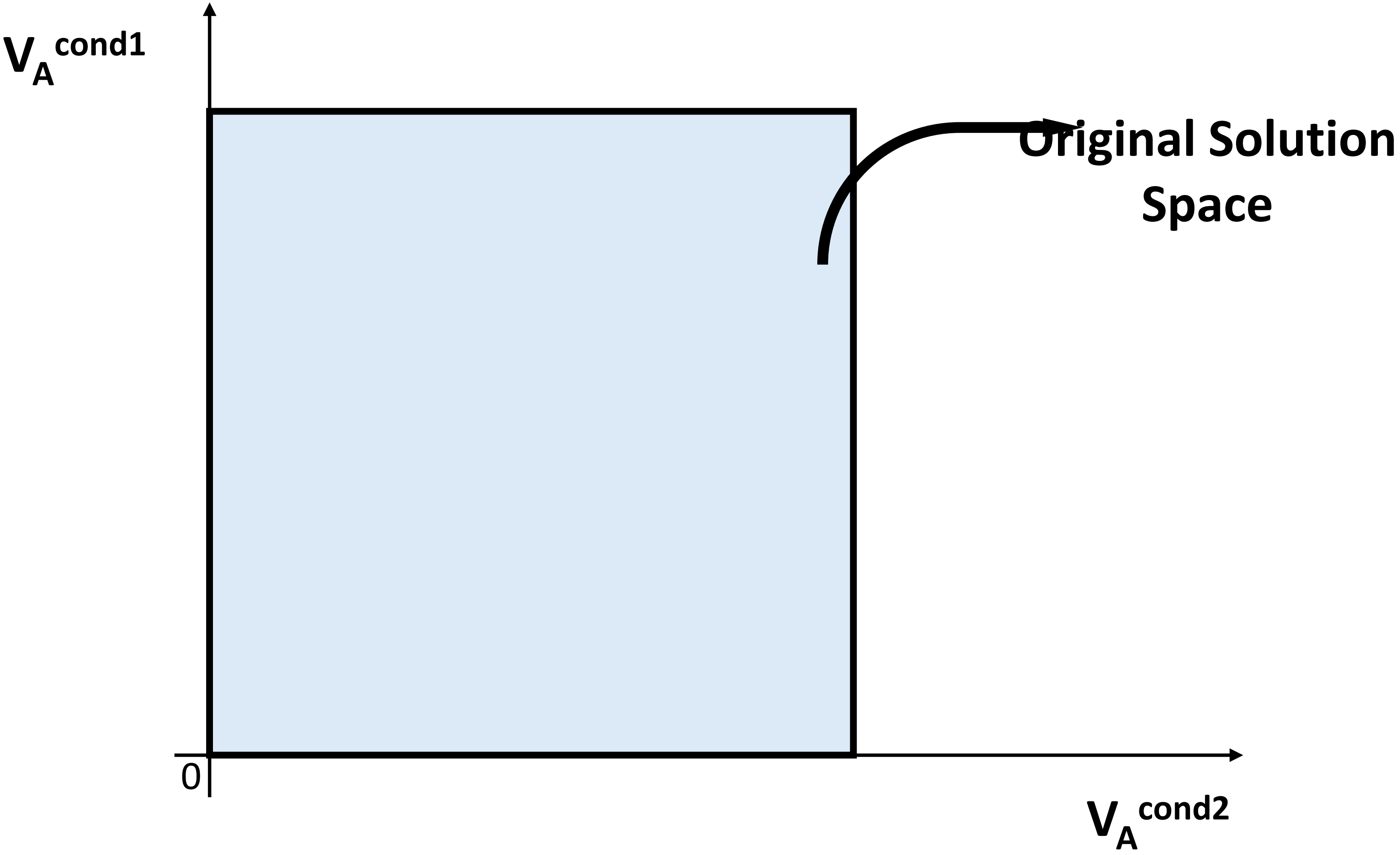
REMI method (example)



REMI method (example)

Flux ratio constraint:

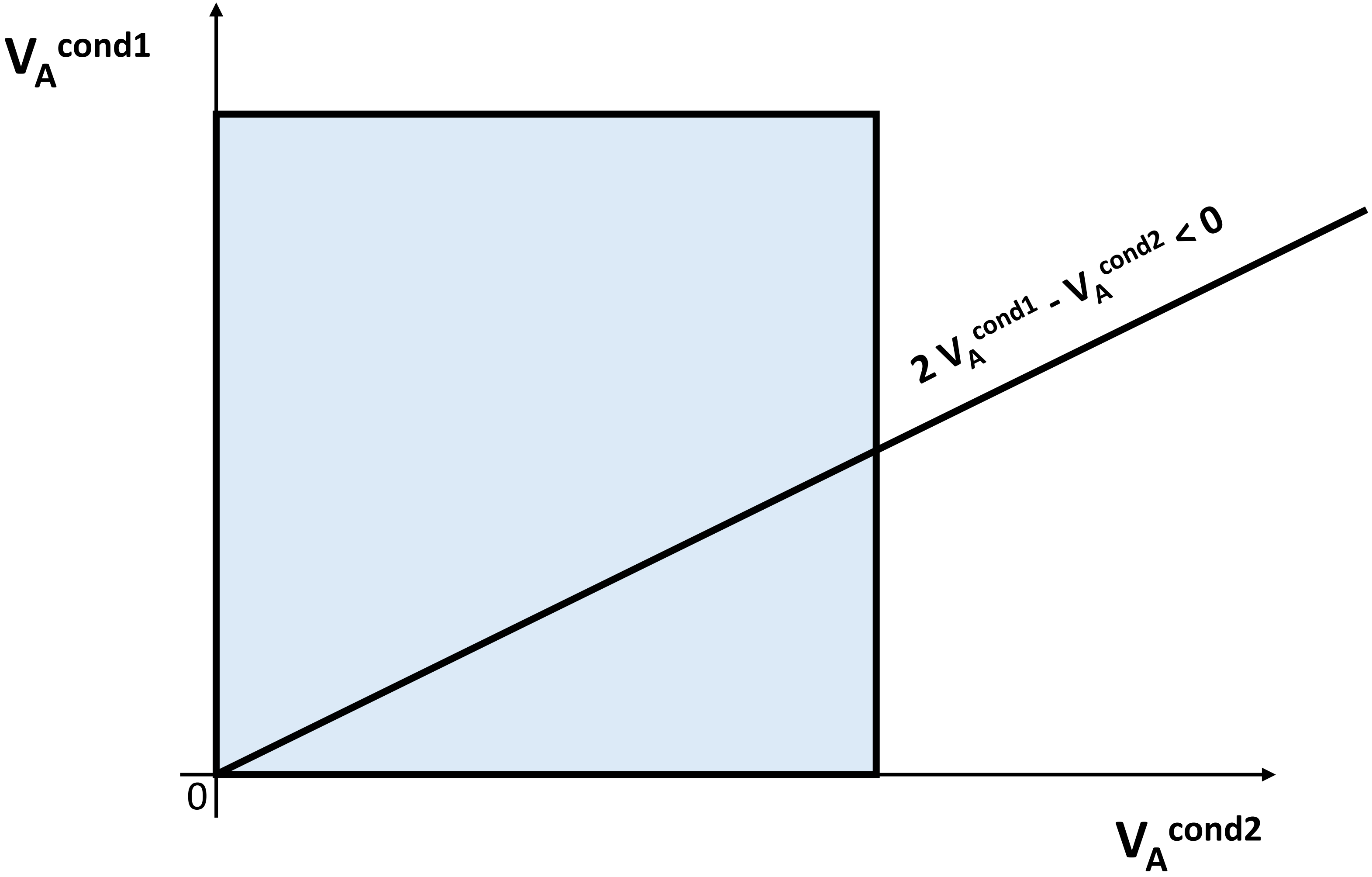
$$V_A^{cond2} > 2\,V_A^{cond1} \quad \Rightarrow \quad \boxed{2\,V_A^{cond1} - V_A^{cond2} < 0}$$



REMI method (example)

Flux ratio constraint:

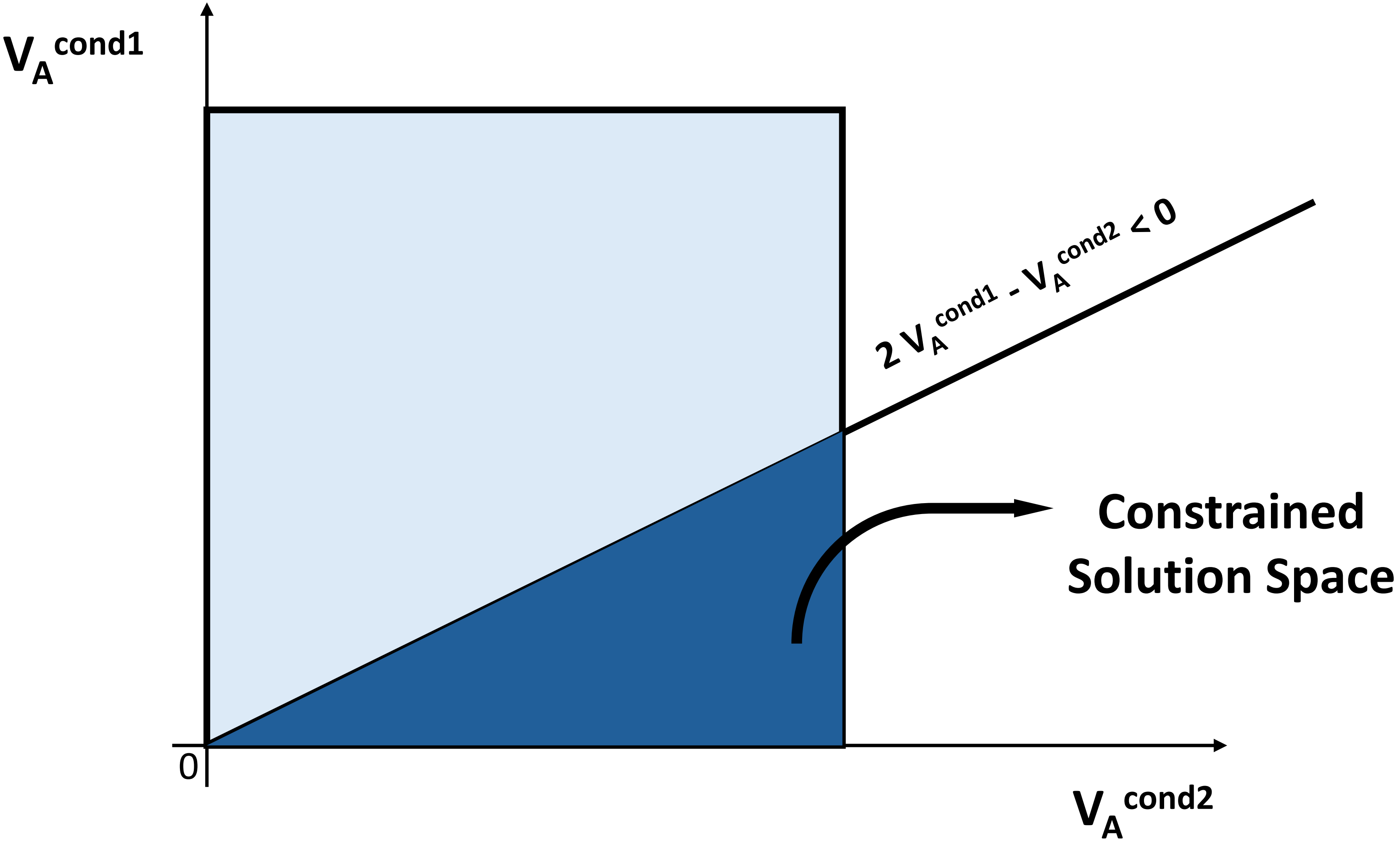
$$V_A^{cond2} > 2\,V_A^{cond1} \quad \Rightarrow \quad 2\,V_A^{cond1} - V_A^{cond2} < 0$$



REMI method (example)

Flux ratio constraint:

$$V_A^{cond2} > 2\,V_A^{cond1} \quad \Rightarrow \quad \boxed{2\,V_A^{cond1} - V_A^{cond2} < 0}$$



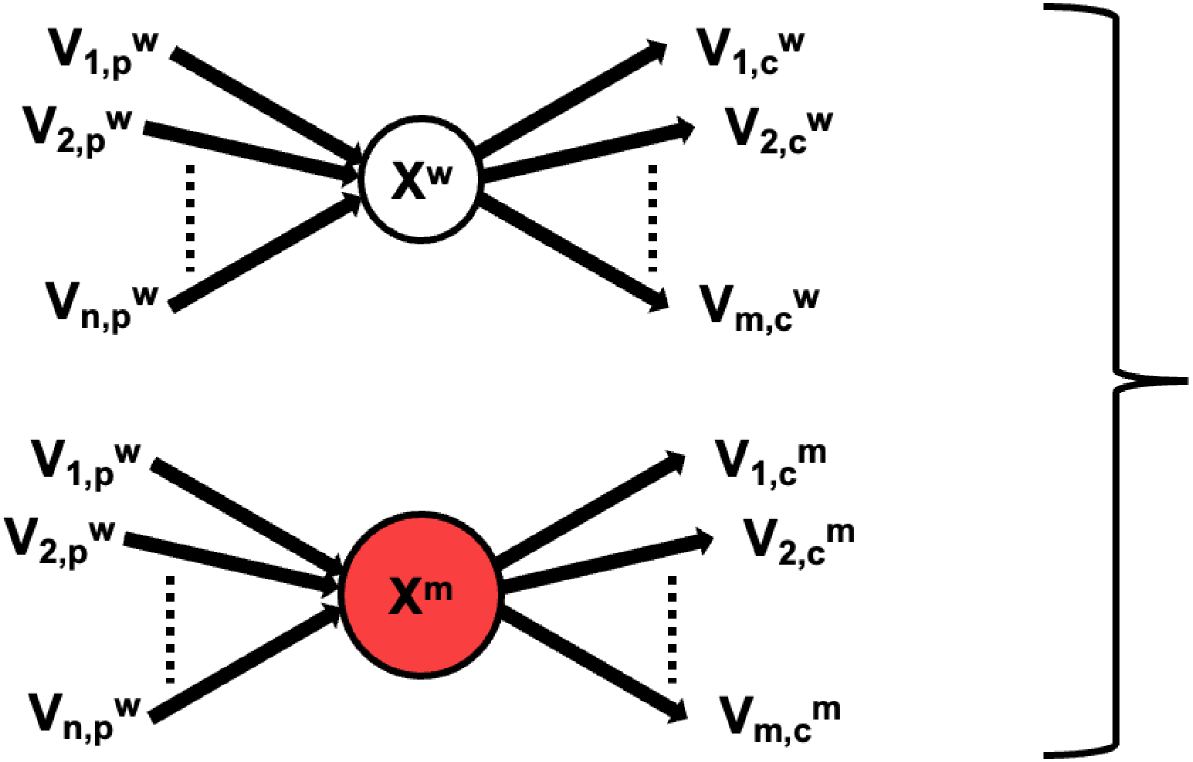
REMI method – Metabolomics

From metabolomics to fluxes

REMI method – Metabolomics

From metabolomics to fluxes

- Upregulated metabolite



$$\sum_{i=1}^{n} V_{i,p}{}^{m} \geq \frac{X^{m}}{X^{w}} \sum_{i=1}^{n} V_{i,p}{}^{w}$$

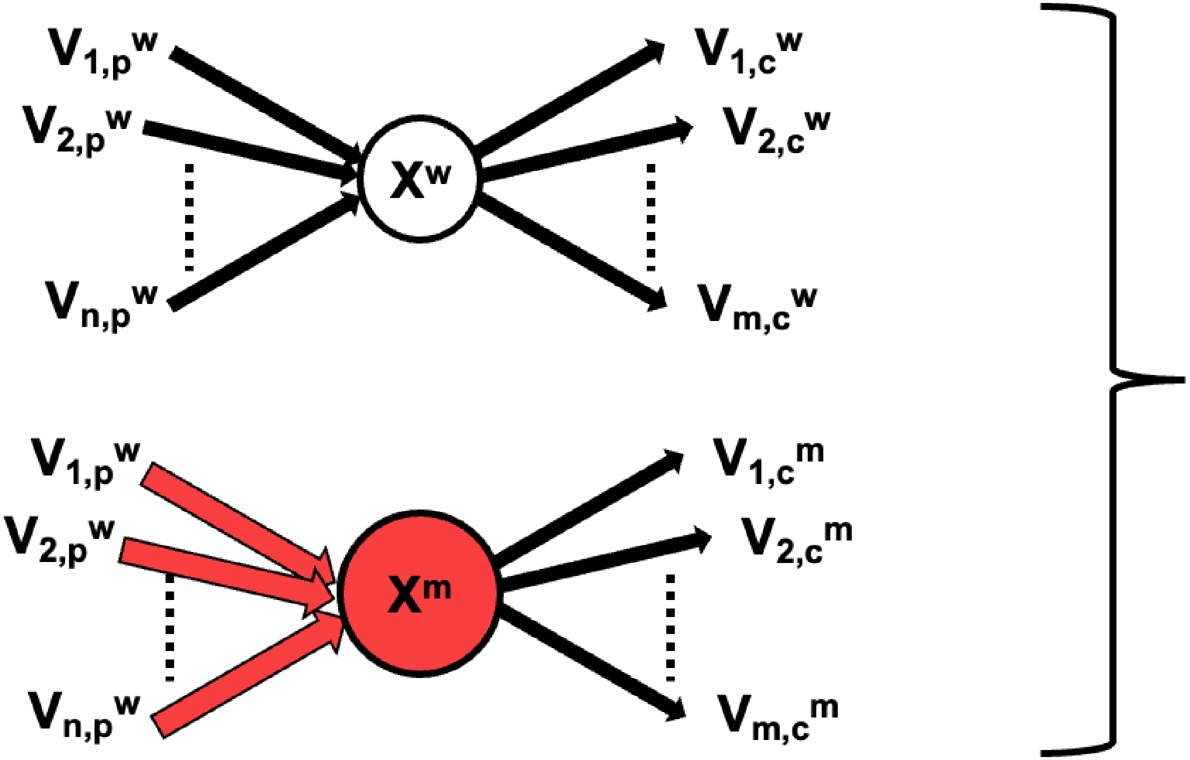
or

$$\sum_{j=1}^{m} V_{j,c}{}^{m} \leq \frac{X^{w}}{X^{m}} \sum_{j=1}^{m} V_{j,c}{}^{w}$$

REMI method – Metabolomics

From metabolomics to fluxes

- Upregulated metabolite (Increased Production)



$$\boxed{\sum_{i=1}^{n} V_{i,p}{}^{m} \geq \frac{X^{m}}{X^{w}} \sum_{i=1}^{n} V_{i,p}{}^{w}}$$

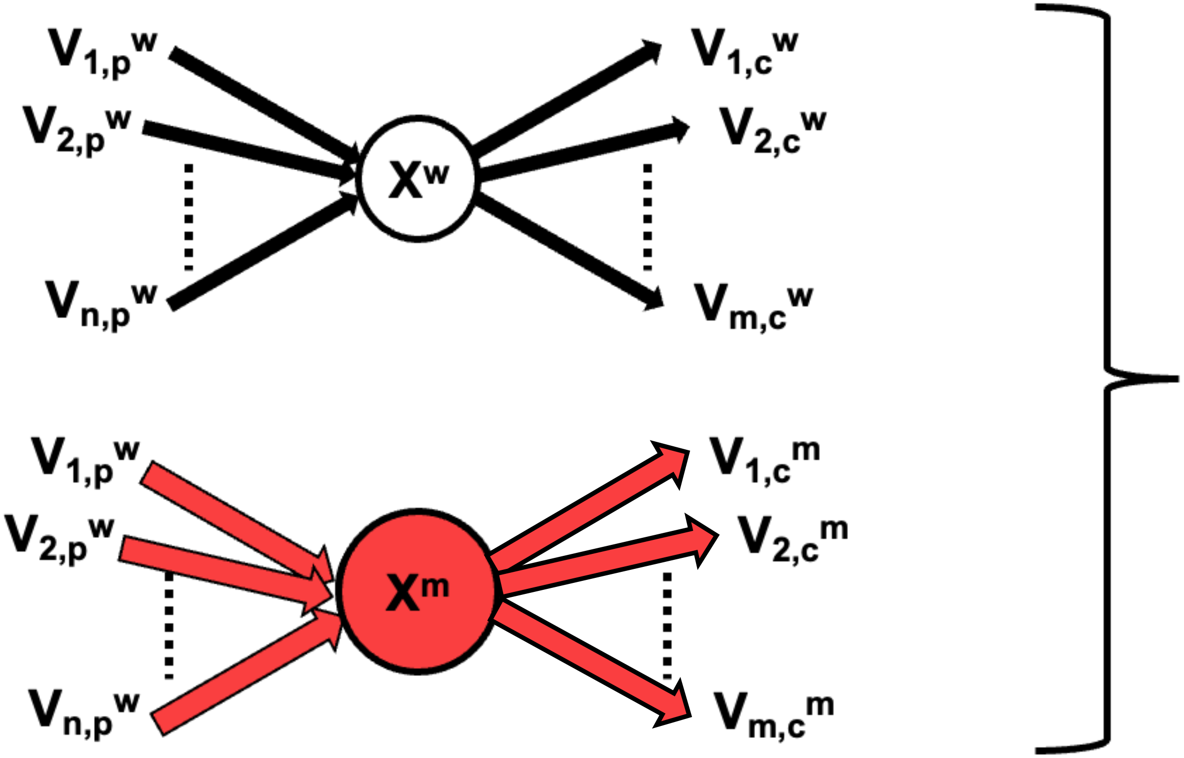
or

$$\sum_{j=1}^{m} V_{j,c}{}^{m} \leq \frac{X^{w}}{X^{m}} \sum_{j=1}^{m} V_{j,c}{}^{w}$$

REMI method – Metabolomics

From metabolomics to fluxes

- Upregulated metabolite (Increased Production)



$$\boxed{\sum_{i=1}^{n} V_{i,p}{}^{m} \geq \frac{X^{m}}{X^{w}} \sum_{i=1}^{n} V_{i,p}{}^{w}}$$

or

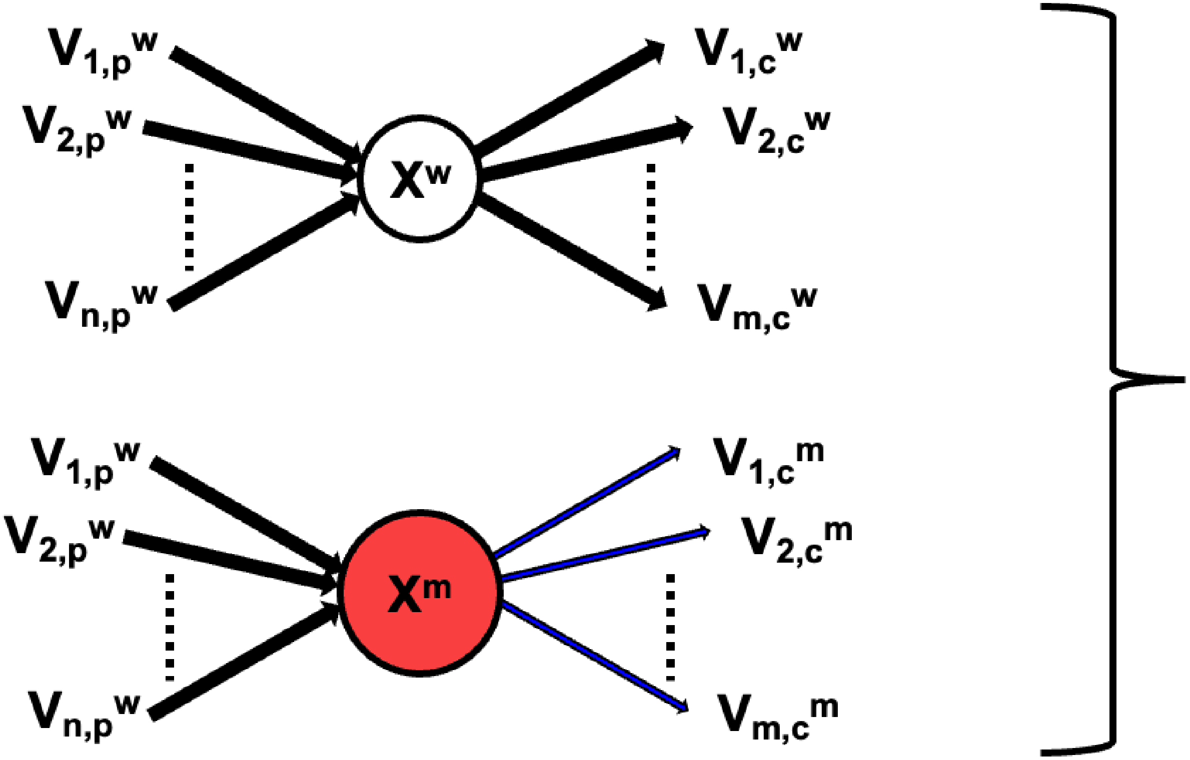
$$\sum_{j=1}^{m} V_{j,c}{}^{m} \leq \frac{X^{w}}{X^{m}} \sum_{j=1}^{m} V_{j,c}{}^{w}$$

@ S.S.

REMI method – Metabolomics

From metabolomics to fluxes

- Upregulated metabolite (Decreased Consumption)



$$\sum_{i=1}^{n} V_{i,p}{}^{m} \geq \frac{X^m}{X^w}\sum_{i=1}^{n} V_{i,p}{}^{w}$$

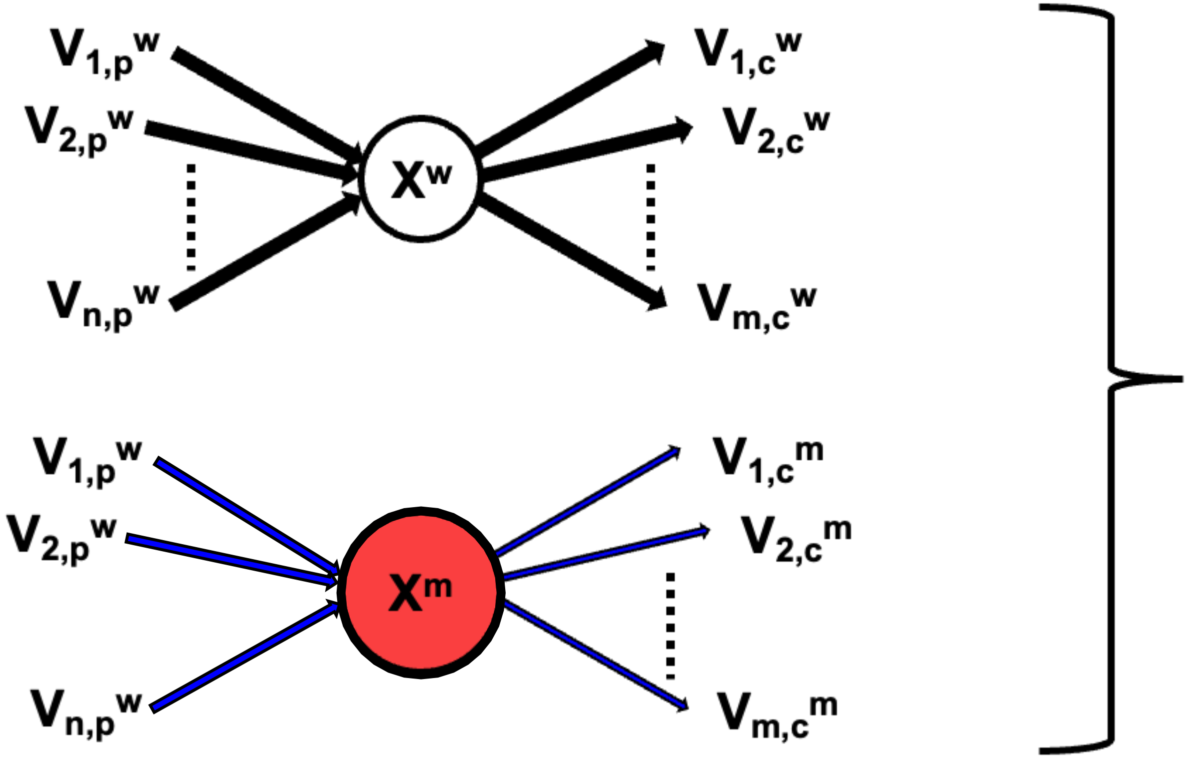
or

$$\boxed{\sum_{j=1}^{m} V_{j,c}{}^{m} \leq \frac{X^w}{X^m}\sum_{j=1}^{m} V_{j,c}{}^{w}}$$

REMI method – Metabolomics

From metabolomics to fluxes

- Upregulated metabolite (Decreased Consumption)



$$\sum_{i=1}^{n} V_{i,p}^{\ m} \geq \frac{X^m}{X^w}\sum_{i=1}^{n} V_{i,p}^{\ w}$$

or

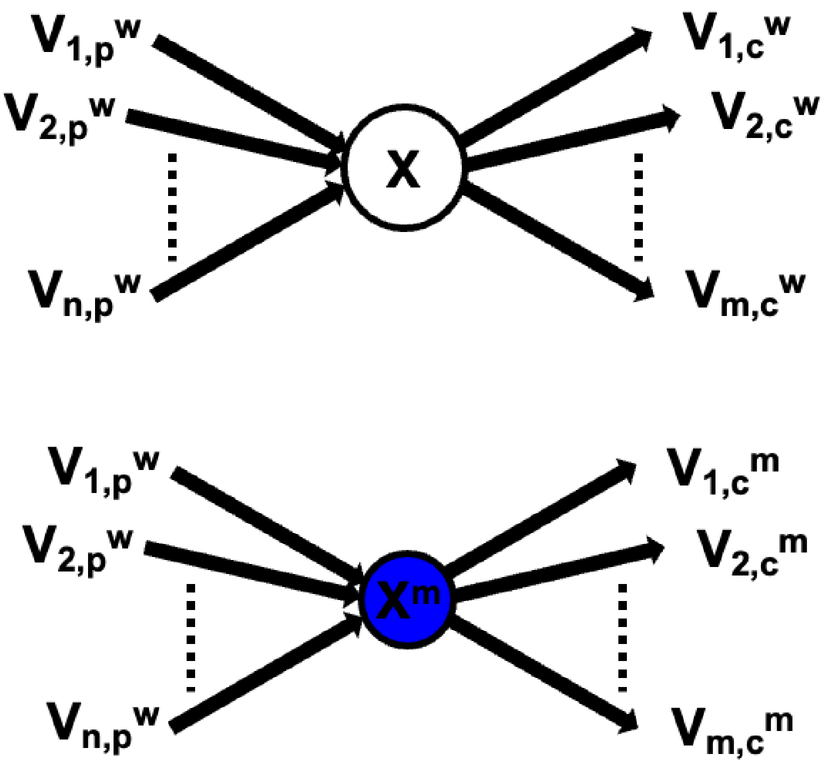
$$\boxed{\sum_{j=1}^{m} V_{j,c}^{\ m} \leq \frac{X^w}{X^m}\sum_{j=1}^{m} V_{j,c}^{\ w}}$$

@ S.S.

REMI method – Metabolomics

From metabolomics to fluxes

- Downregulated metabolite



$$\sum_{i=1}^{n} V_{i,p}{}^{m} \leq \frac{X^{m}}{X^{w}} \sum_{i=1}^{n} V_{i,p}{}^{w}$$

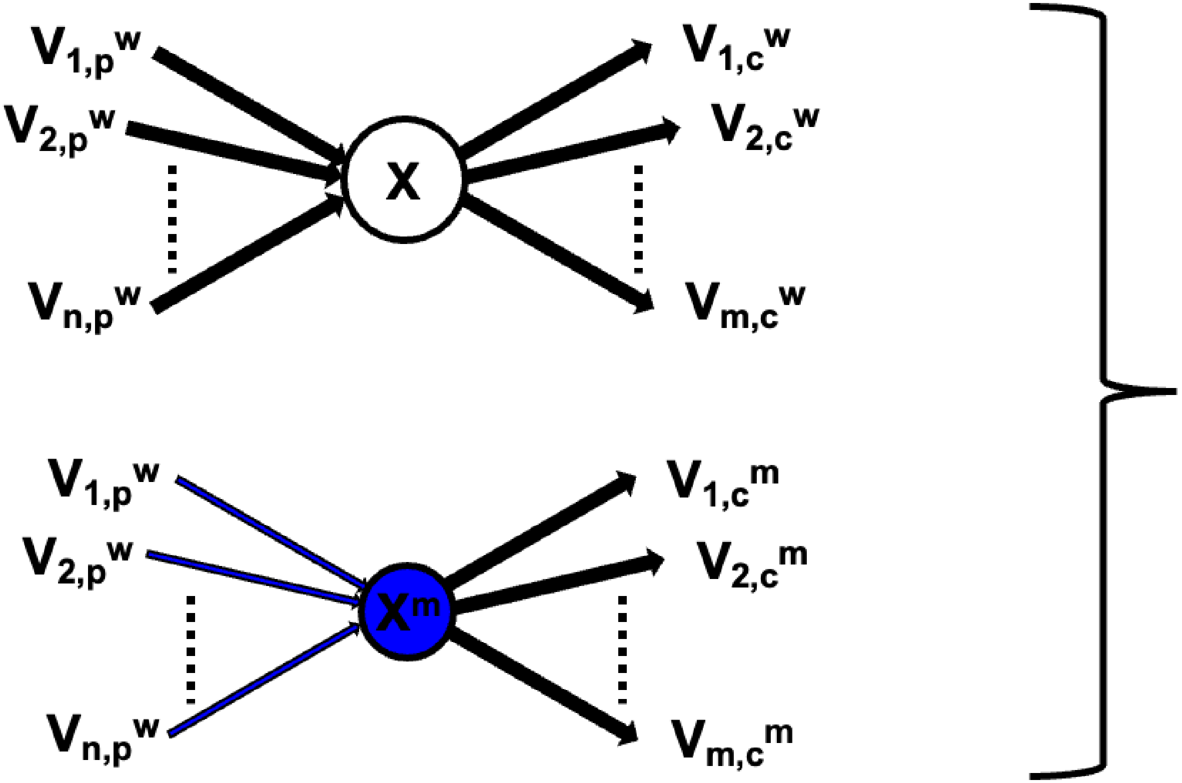
or

$$\sum_{j=1}^{m} V_{j,c}{}^{m} \geq \frac{X^{w}}{X^{m}} \sum_{j=1}^{m} V_{j,c}{}^{w}$$

REMI method – Metabolomics

From metabolomics to fluxes

- Downregulated metabolite (Decreased Production)



$$\boxed{\sum_{i=1}^{n} V_{i,p}{}^{m} \le \frac{X^{m}}{X^{w}} \sum_{i=1}^{n} V_{i,p}{}^{w}}$$

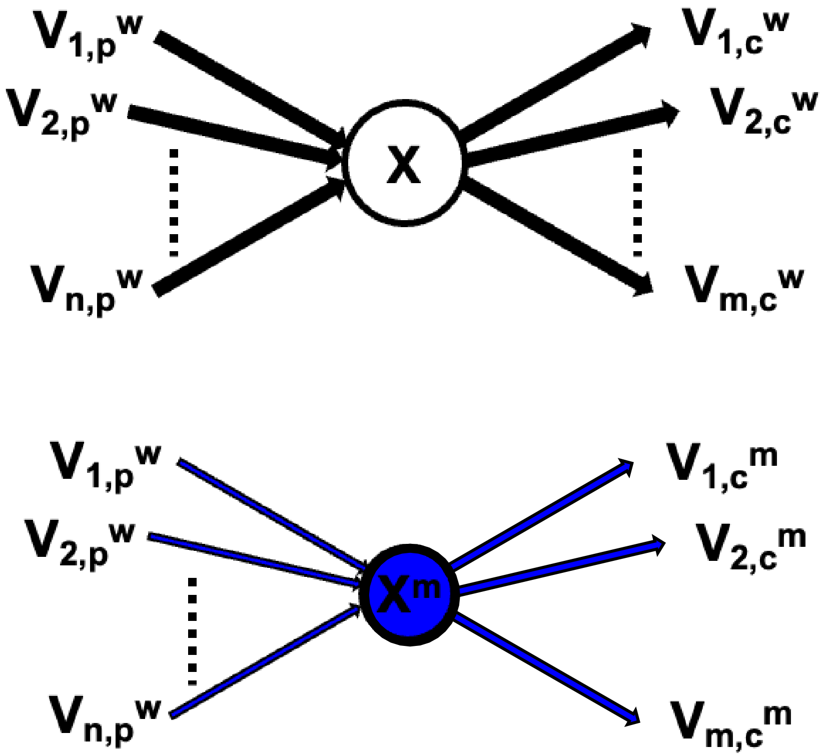
or

$$\sum_{j=1}^{m} V_{j,c}{}^{m} \ge \frac{X^{w}}{X^{m}} \sum_{j=1}^{m} V_{j,c}{}^{w}$$

REMI method – Metabolomics

From metabolomics to fluxes

- Downregulated metabolite (Decreased Production)



$$\boxed{\sum_{i=1}^{n} V_{i,p}{}^{m} \le \frac{X^{m}}{X^{w}} \sum_{i=1}^{n} V_{i,p}{}^{w}}$$

or

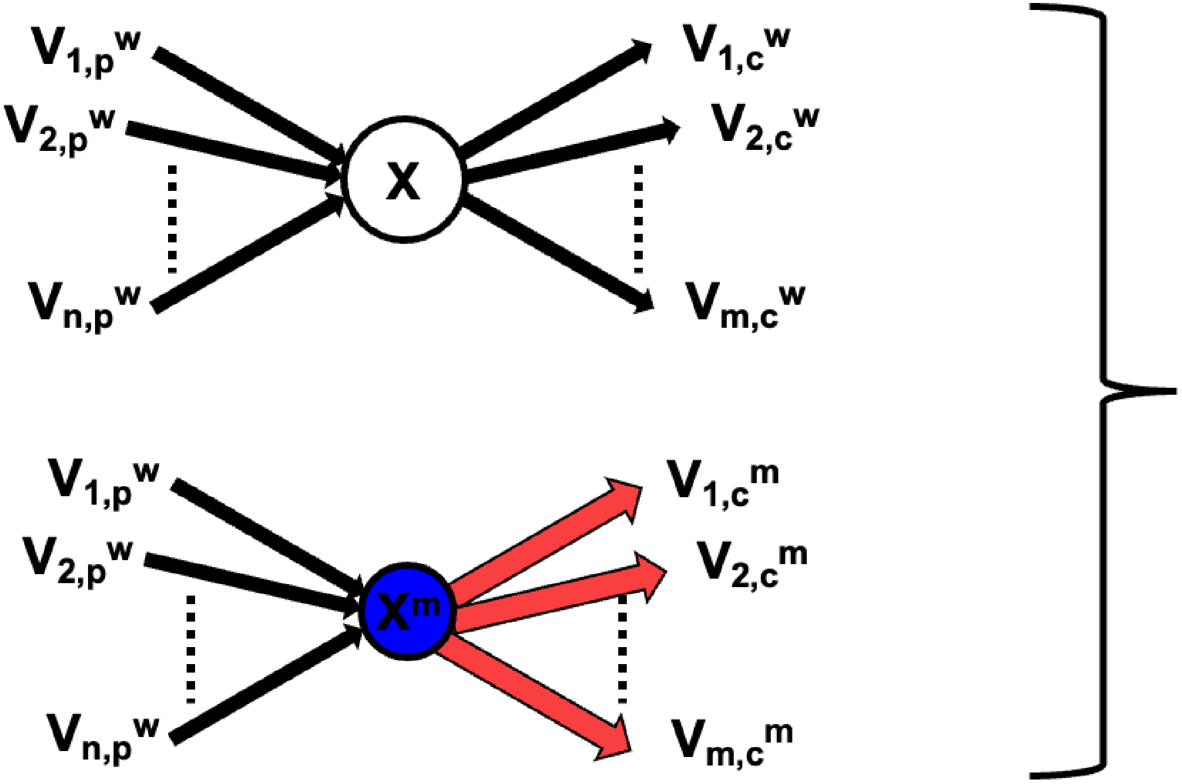
$$\sum_{j=1}^{m} V_{j,c}{}^{m} \ge \frac{X^{w}}{X^{m}} \sum_{j=1}^{m} V_{j,c}{}^{w}$$

@ S.S.

REMI method – Metabolomics

From metabolomics to fluxes

- Downregulated metabolite (Increased Consumption)



$$\sum_{i=1}^{n} V_{i,p}{}^{m} \leq \frac{X^{m}}{X^{w}} \sum_{i=1}^{n} V_{i,p}{}^{w}$$

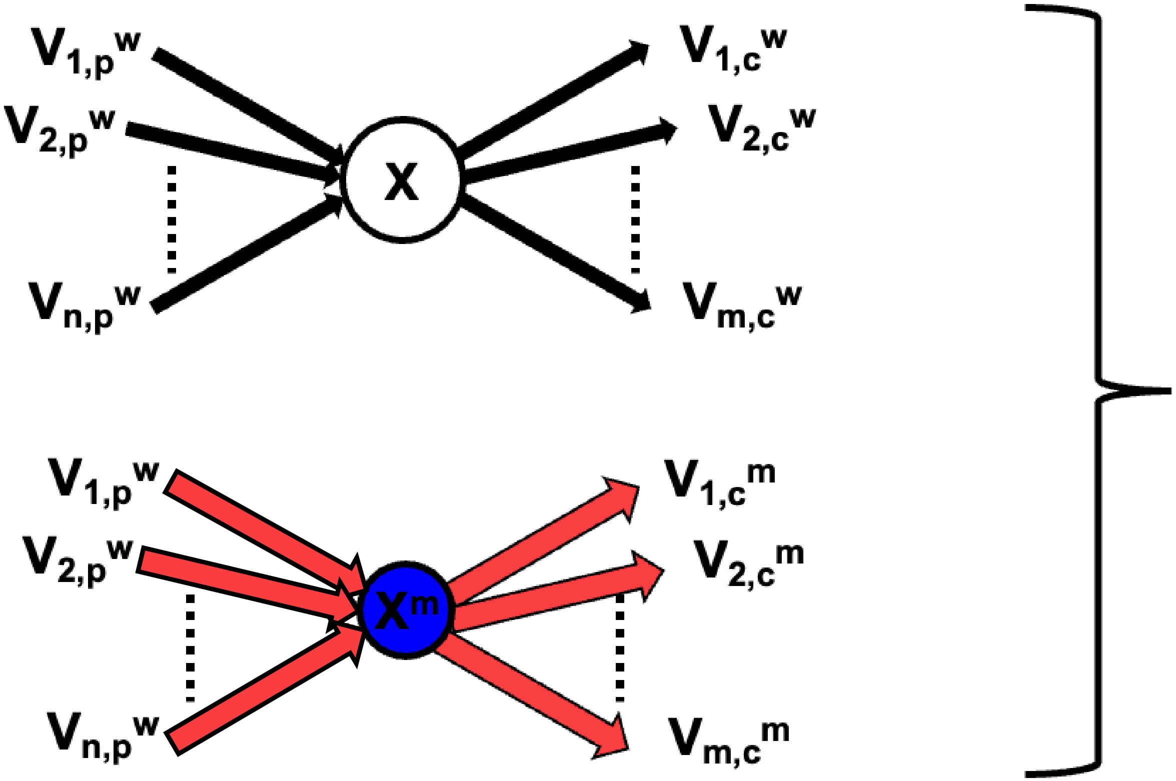
or

$$\boxed{\sum_{j=1}^{m} V_{j,c}{}^{m} \geq \frac{X^{w}}{X^{m}} \sum_{j=1}^{m} V_{j,c}{}^{w}}$$

REMI method – Metabolomics

From metabolomics to fluxes

- Downregulated metabolite (Increased Consumption)



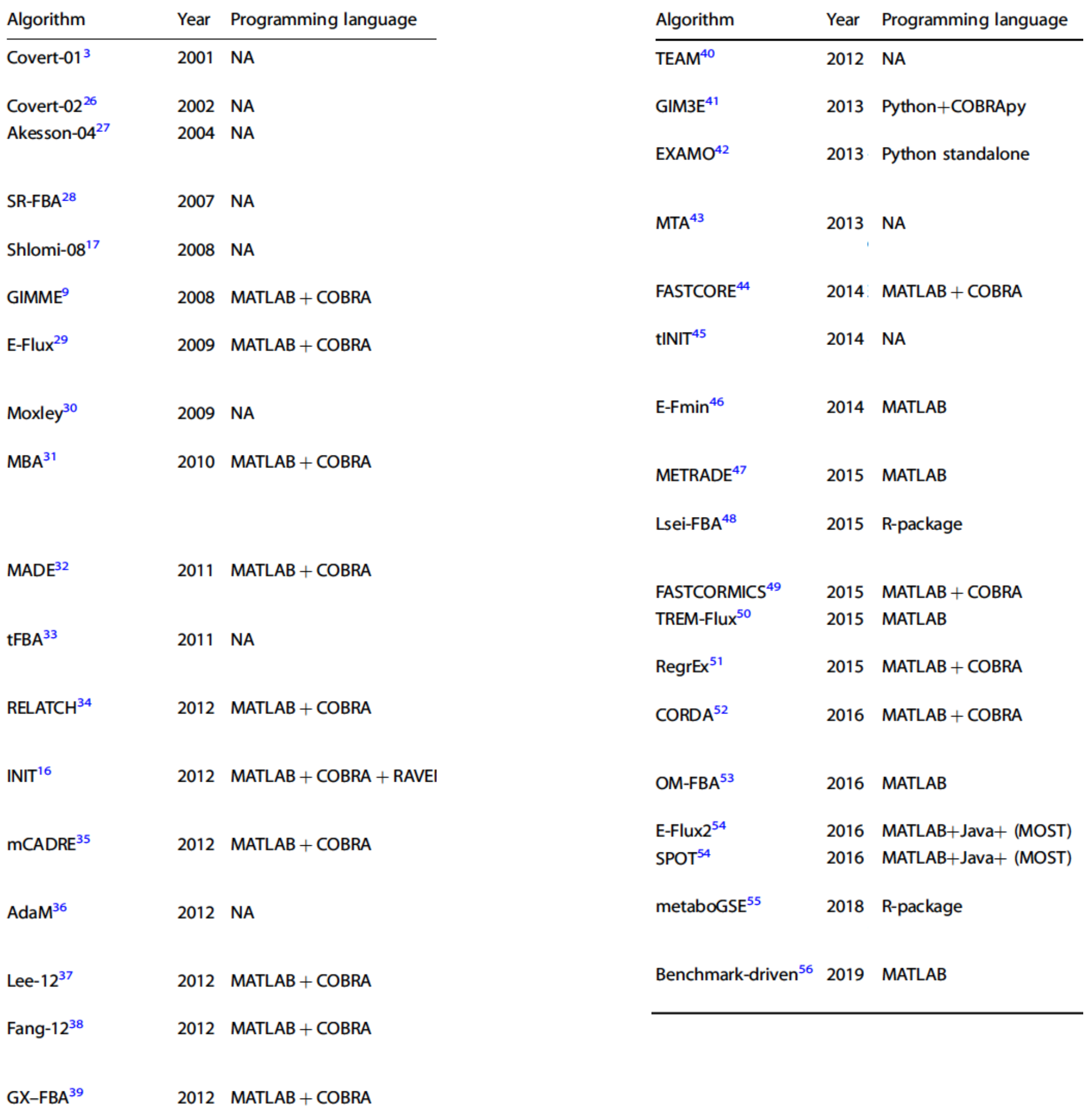
$$\sum_{i=1}^{n} V_{i,p}{}^{m} \leq \frac{X^{m}}{X^{w}} \sum_{i=1}^{n} V_{i,p}{}^{w}$$

or

$$\boxed{\sum_{j=1}^{m} V_{j,c}{}^{m} \geq \frac{X^{w}}{X^{m}} \sum_{j=1}^{m} V_{j,c}{}^{w}}$$

@ S.S.

And many more methods…



Algorithm	Year	Programming language
Covert-01[3]	2001	NA
Covert-02[26]	2002	NA
Akesson-04[27]	2004	NA
SR-FBA[28]	2007	NA
Shlomi-08[17]	2008	NA
GIMME[9]	2008	MATLAB + COBRA
E-Flux[29]	2009	MATLAB + COBRA
Moxley[30]	2009	NA
MBA[31]	2010	MATLAB + COBRA
MADE[32]	2011	MATLAB + COBRA
tFBA[33]	2011	NA
RELATCH[34]	2012	MATLAB + COBRA
INIT[16]	2012	MATLAB + COBRA + RAVE
mCADRE[35]	2012	MATLAB + COBRA
AdaM[36]	2012	NA
Lee-12[37]	2012	MATLAB + COBRA
Fang-12[38]	2012	MATLAB + COBRA
GX–FBA[39]	2012	MATLAB + COBRA
TEAM[40]	2012	NA
GIM3E[41]	2013	Python+COBRApy
EXAMO[42]	2013	Python standalone
MTA[43]	2013	NA
FASTCORE[44]	2014	MATLAB + COBRA
tINIT[45]	2014	NA
E-Fmin[46]	2014	MATLAB
METRADE[47]	2015	MATLAB
Lsei-FBA[48]	2015	R-package
FASTCORMICS[49]	2015	MATLAB + COBRA
TREM-Flux[50]	2015	MATLAB
RegrEx[51]	2015	MATLAB + COBRA
CORDA[52]	2016	MATLAB + COBRA
OM-FBA[53]	2016	MATLAB
E-Flux2[54]	2016	MATLAB+Java+ (MOST)
SPOT[54]	2016	MATLAB+Java+ (MOST)
metaboGSE[55]	2018	R-package
Benchmark-driven[56]	2019	MATLAB

Jalili et al., npj Systems Biology and Applications, 2023