Systems Biology Workflows for Data Analysis and Integration

Location: Monday, June 25, 8:30 am – 10:15 am in the Washington State Convention Center.

Presenter: Douglas McCloskey, Ph.D.

Outline

1. Multi-omic data integration and analysis architecture
2. Case studies
3. Model driven experimental design
4. Data processing pipelines
5. Biochemical model mapping
6. Biochemical modeling
7. Statistics
8. Visualization and sharing
9. Conclusion

Multi-omic data integration and analysis architecture

Introduction (1) Presenter background

Introduction (2) Systems biochemistry workflow

Introduction (1) Presenter background

- 2005-2010: BA (Biomedical Engineering) at University of California, Irvine
  - Tennis, Microfluidics, Software engineering (and beer pong)

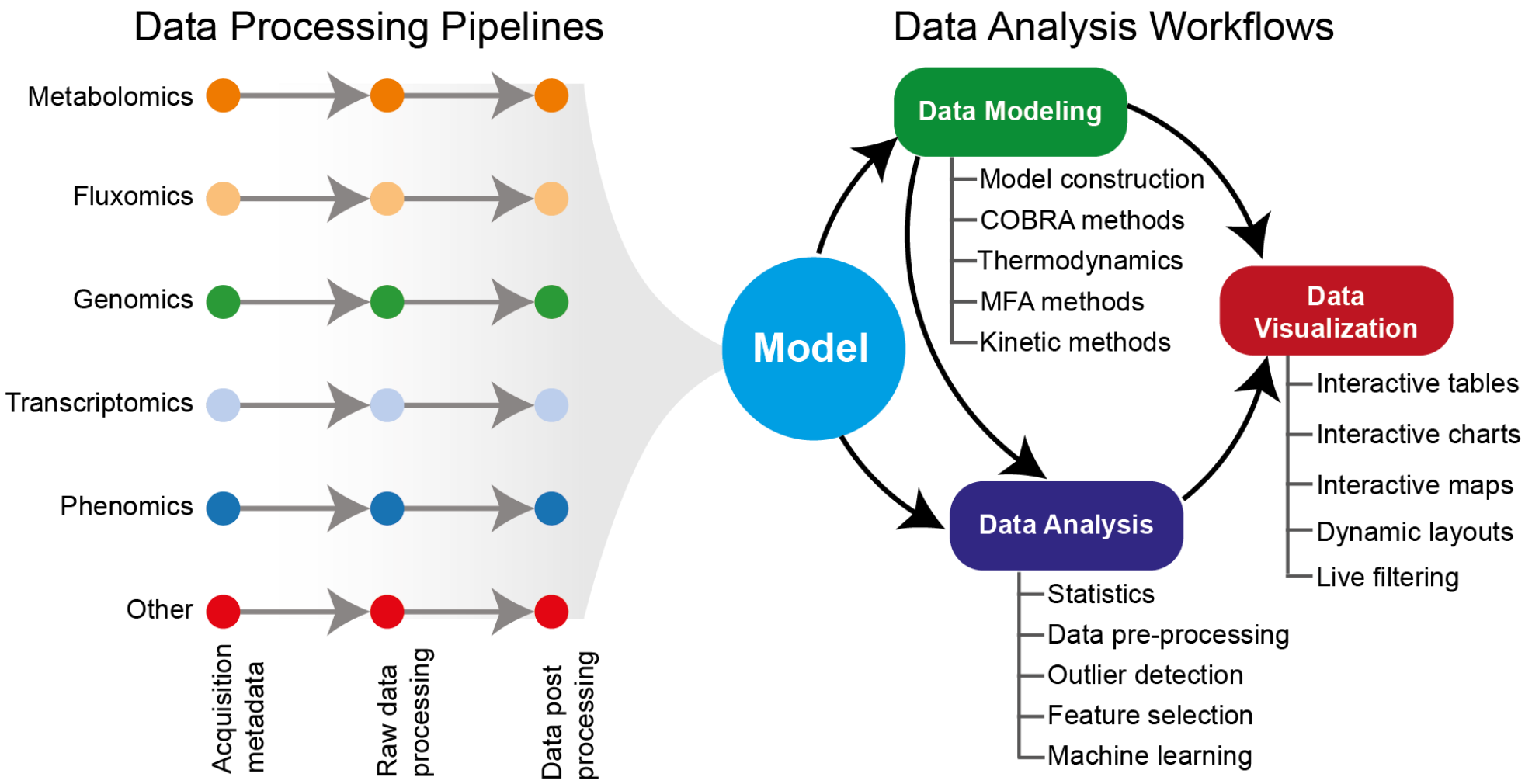
Introduction (1) Presenter background

- 2005-2010: BA (Biomedical Engineering) at University of California, Irvine
  - Tennis, Microfluidics, Software engineering (and beer pong)
- 2010-2017: PhD (Bioengineering Engineering) at University of California, San Diego
  - Metabolomics and fluxomics method development for Mass Spec
  - Multi-omics (i.e., Genomics, Transcriptomics, Phenomics, Metabolomics, and Fluxomics) data integration, analysis, and biochemical modeling
  - Applications in microbial evolution and human health

Introduction (1) Presenter background

- 2005-2010: BA (Biomedical Engineering) at University of California, Irvine
  - Tennis, Microfluidics, Software engineering (and beer pong)
- 2010-2017: PhD (Bioengineering Engineering) at University of California, San Diego
  - Metabolomics and fluxomics method development for Mass Spec
  - Multi-omics (i.e., Genomics, Transcriptomics, Phenomics, Metabolomics, and Fluxomics) data integration, analysis, and biochemical modeling
  - Applications in microbial evolution and human health
- 2017-Present: Group Leader at the Center for Biosustainability (CfB) at the Technical University of Denmark (DTU)
  - Automated and high throughput ”Big” metabolomics data generation, sample handling, acquisition, and data processing and analysis solutions using robotic liquid handlers and software solution
  - Application of machine learning and metabolomics (and other –omics) to human health

Introduction (2) Systems biochemistry workflow



Case studies

Case studies

Case study (1) Laboratory evolution of gene knockout strains of E. coli reveal regulatory architectures that are governed by metabolism

McCloskey D, et al. (2018) NCOMM accepted.
McCloskey D, et a. (2018) AEM, under revision
McCloskey D, et a. (2018) Frontiers, under revision
McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO. (2018) Metab Eng. doi: 10.1016/j.ymben.2018.06.005.
McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO. (2018) Metab Eng. doi: 10.1016/j.ymben.2018.05.012.

Case studies

Case study (1) Laboratory evolution of gene knockout strains of E. coli reveal regulatory architectures that are governed by metabolism

Case study (2) Personalized Whole-Cell Kinetic Models of Metabolism for Discovery in Genomics and Pharmacodynamics

McCloskey D, et al. (2018) NCOMM accepted.
McCloskey D, et a. (2018) AEM, under revision
McCloskey D, et a. (2018) Frontiers, under revision
McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO. (2018) Metab Eng. doi: 10.1016/j.ymben.2018.06.005.
McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO. (2018) Metab Eng. doi: 10.1016/j.ymben.2018.05.012.

Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

Case studies

Case study (1) Laboratory evolution of gene knockout strains of *E. coli* reveal regulatory architectures that are governed by metabolism

Case study (2) Personalized Whole-Cell Kinetic Models of Metabolism for Discovery in Genomics and Pharmacodynamics

Case study (3) RapidRIP quantifies the intracellular metabolome of 7 industrial strains of *E. coli*

McCloskey D, et al. (2018) NCOMM accepted.
McCloskey D, et a. (2018) AEM, under revision
McCloskey D, et a. (2018) Frontiers, under revision
McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO. (2018) Metab Eng. doi: 10.1016/j.ymben.2018.06.005.
McCloskey D, Xu S, Sandberg TE, Brunk E, Hefner Y, Szubin R, Feist AM, Palsson BO. (2018) Metab Eng. doi: 10.1016/j.ymben.2018.05.012.

Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

D McCloskey, J Xu, L Schrübbers, H Christensen, M Herrgård. RapidRIP quantifies the intracellular metabolome of 7 industrial strains of E. coli (2018). Metabolic Engineering, 47, 383-392

Model driven experimental design

Experimental design (1) Questions, hypotheses, or objectives

Experimental design (2) Model construction

Experimental design (3) Required data and analyses

Experimental design (1) Questions, hypotheses, or objectives

- What are my questions/hypotheses/objectives?
- The more clear your questions/hypotheses/objectives are the easier the downstream data integration and analysis will be

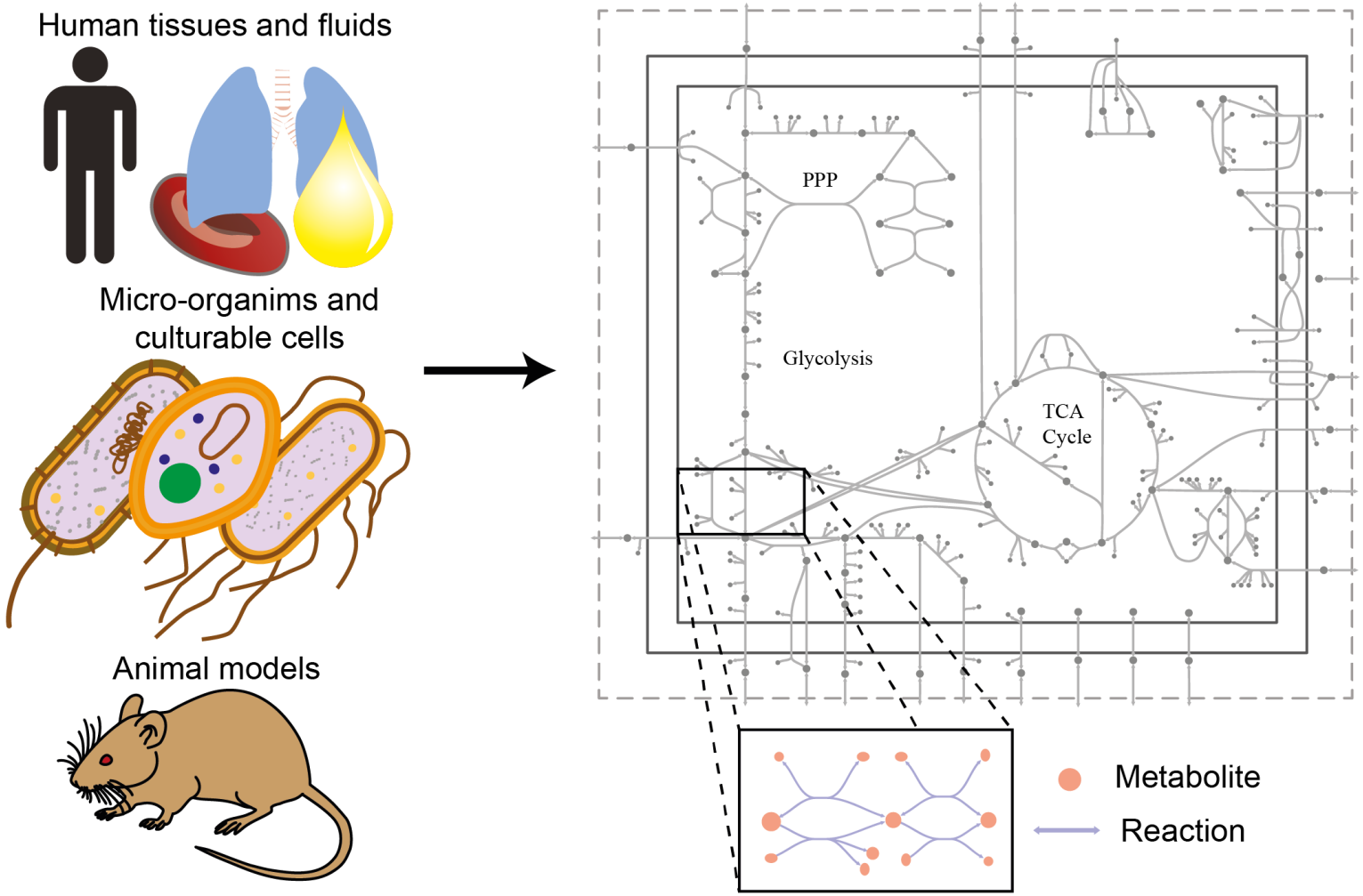
Experimental design (2) Model construction

- Model:
  - *Describes how all –omics data types relate to one another*

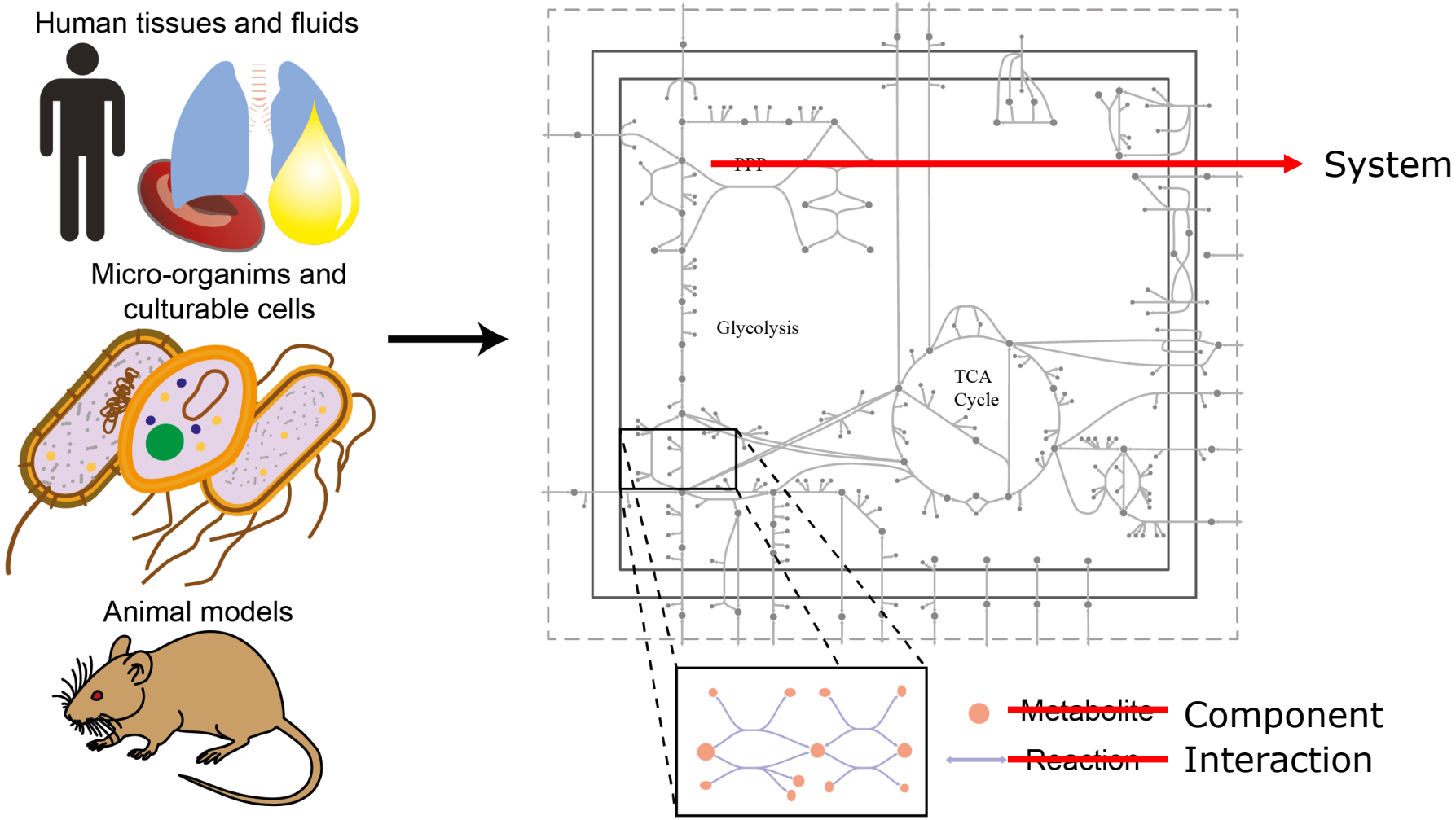
Experimental design (2) Model construction

- Model:
  - *Describes how all –omics data types relate to one another*
  - The biological system of study (e.g., Organism)
  - Scope (i.e., metabolism, cell signalling, gene expression, etc.)
  - Model resolution (e.g., reactions, atom mappings, kinetic parameters, etc.)

Experimental design (2) Model construction



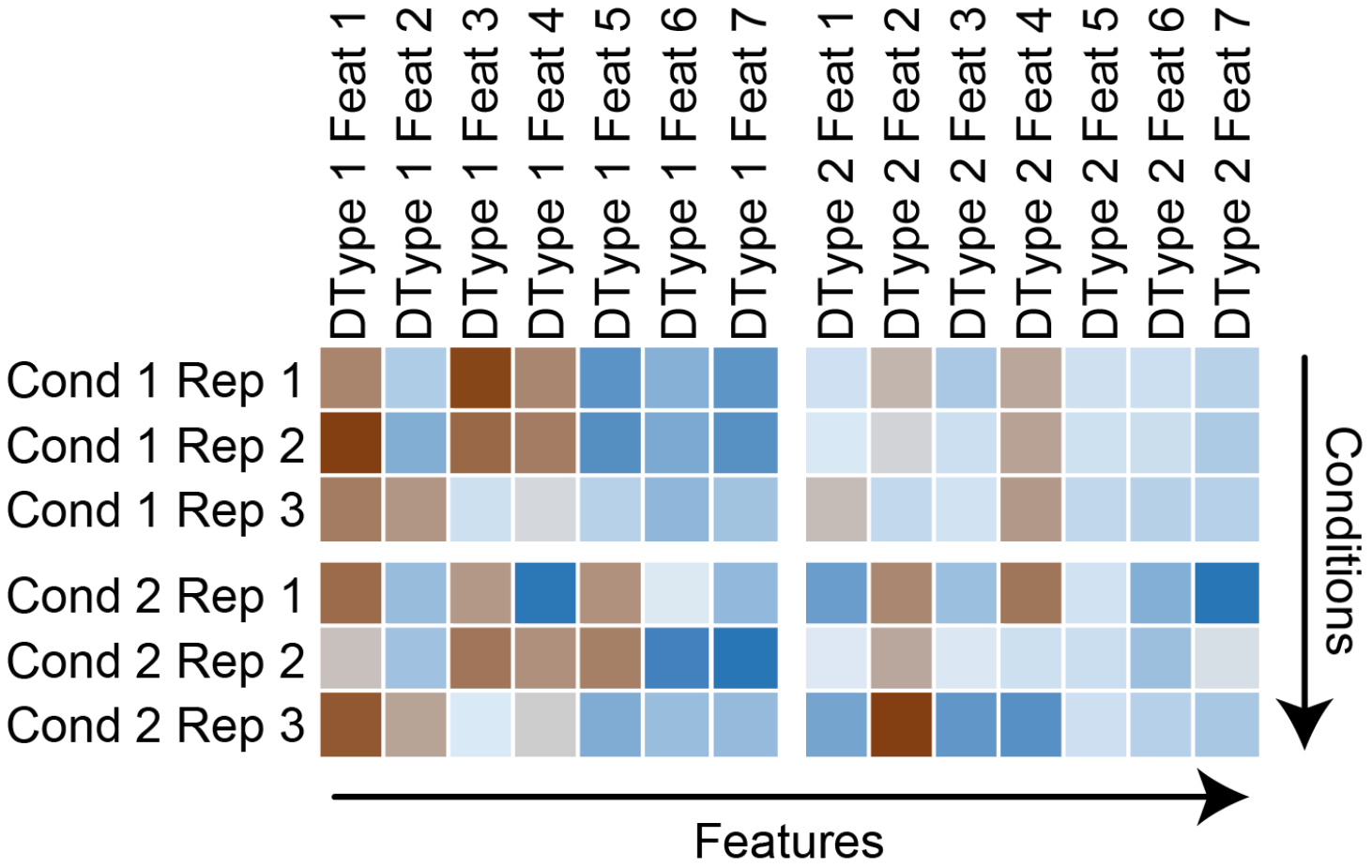
Experimental design (2) Model construction



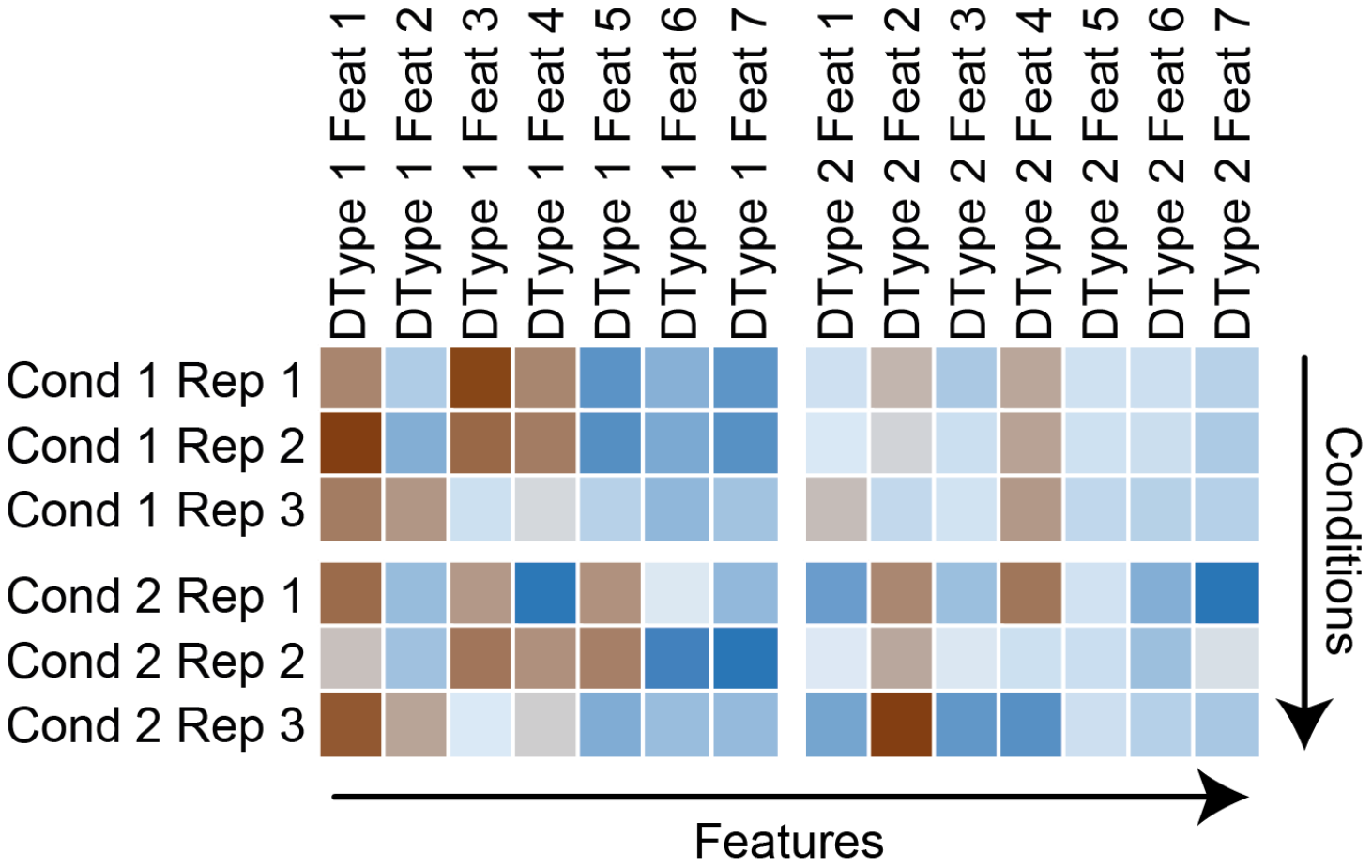
Experimental design (2) Model construction

- Model:
  - *Describes how all –omics data types relate to one another*
  - The biological system of study (e.g., Organism)
  - Scope (i.e., metabolism, cell signalling, gene expression, etc.)
  - Model resolution (e.g., reactions, atom mappings, kinetic parameters, etc.)
- Resources:
  - Model reconstruction: http://modelseed.org/
  - Published reconstructions: http://bigg.ucsd.edu/
  - Biochemical databases: https://biocyc.org/

Experimental design (3) Required data and analyses



Experimental design (3) Required data and analyses



- What are my experimental parameters?
  - Conditions and replicates
  - Controls
- What are my required analyses?
  - Hypothesis test
- What -Omics data are necessary?
  - Data types and number of features

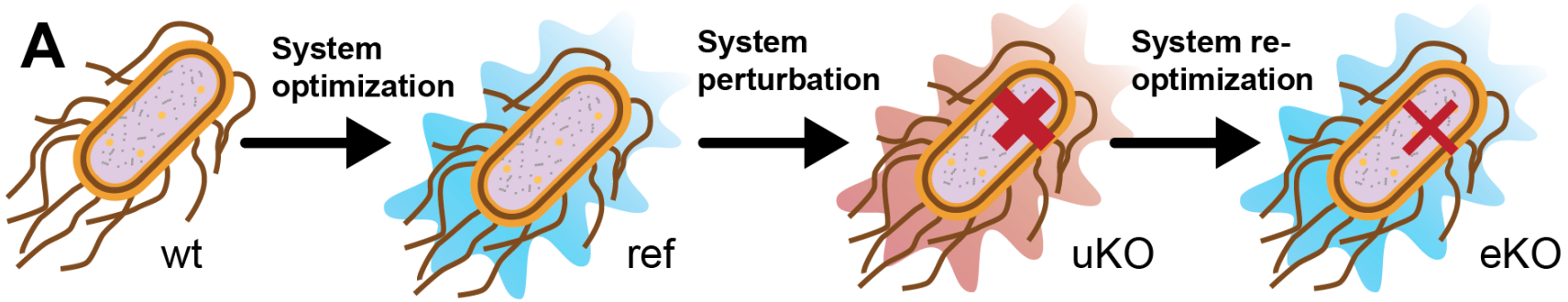
Case Study (1)

- Question: What are metabolic genes system’s level function?

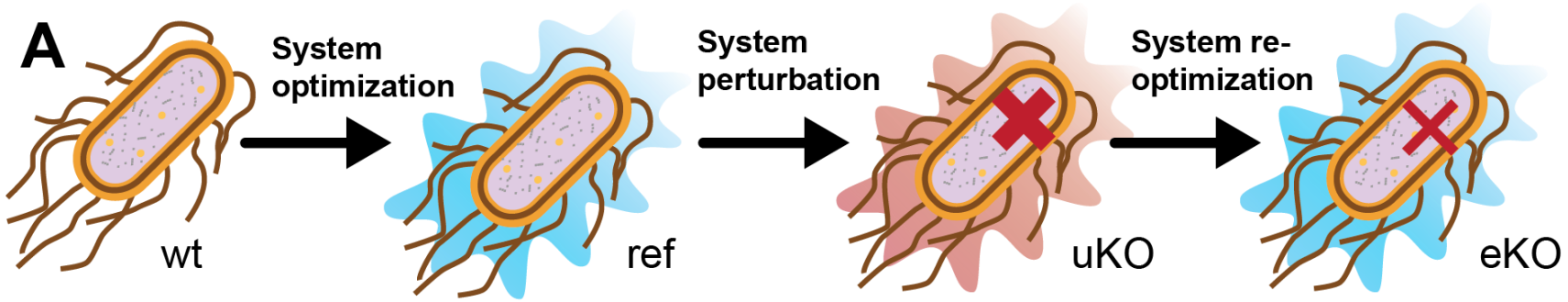
Case Study (1)

- Question: What are metabolic genes system's level function?
- Experiments: gene KOs, adaptive laboratory evolution (ALE)
  - Systems perturbation experiments in engineering: 1) perturb the system, 2) allow the system to recover, and 3) analyze how the system recovered

Case Study (1)

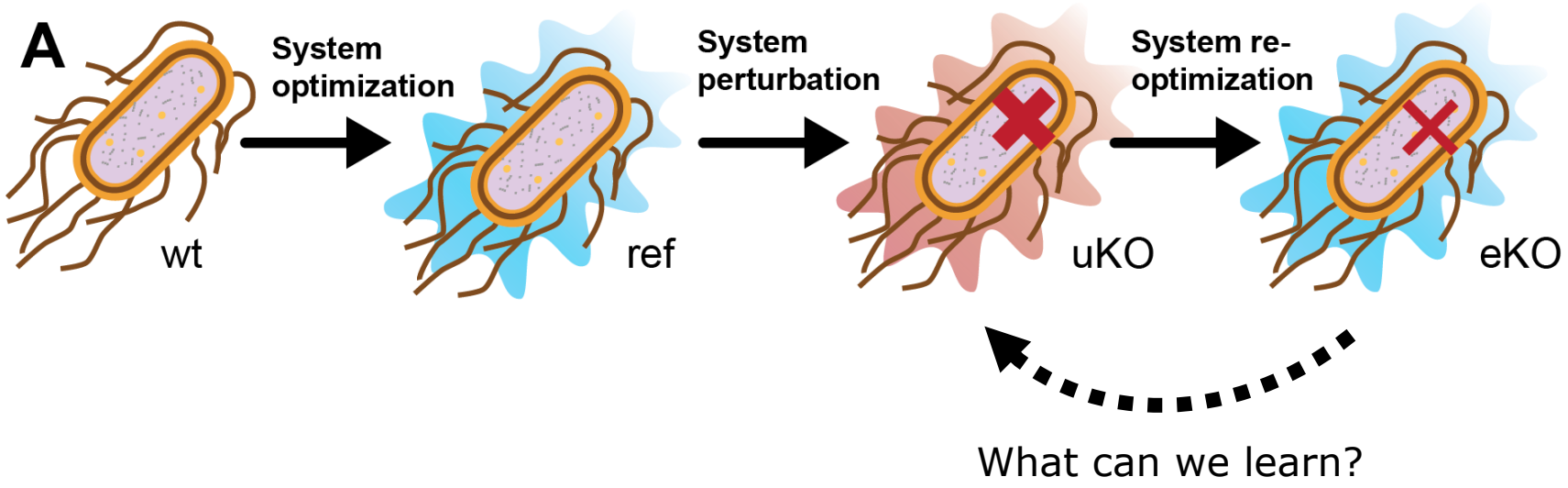

A
System optimization
System perturbation
System re-optimization
wt
ref
uKO
eKO

Case Study (1)



- Removed confounding variable of adaptation to the experimental conditions

Case Study (1)


A
System optimization
System perturbation
System re-optimization
wt
ref
uKO
eKO
What can we learn?

Case Study (1)

- Question: What are metabolic genes system's level function?
- Experiments: gene KOs, adaptive laboratory evolution (ALE)
  - Systems perturbation experiments in engineering: 1) perturb the system, 2) allow the system to recover, and 3) analyze how the system recovered
- Model: *E. coli* MG1655 K-12

Case Study (1)

- Question: What are metabolic genes system's level function?
- Experiments: gene KOs, adaptive laboratory evolution (ALE)
  - Systems perturbation experiments in engineering: 1) perturb the system, 2) allow the system to recover, and 3) analyze how the system recovered
- Model: *E. coli* MG1655 K-12
- Experimental parameters: Gene KO strains and control before and after ALE

Case Study (1)

Model-driven KO selection

Reaction
ATPS4rpp (atpIC)
PGK
GAPD
SUCDi
GLCptspp (ptsHIcrr)
TPI
FRD3
G6PDH2r (zwf)
GND
PGI
PFK
FBA
PPC

Case Study (1)

1. Sampled the model to rank order the highest flux reactions

Reaction	Sampled flux mean	Sampled flux St.D.
ATPS4rpp (atpIC)	56.37	0.27
PGK	19.05	0.09
GAPD	19.05	0.09
SUCDi	12.29	2.26
GLCptspp (ptsHIcrr)	11.64	0.19
TPI	8.60	0.09
FRD3	6.66	2.15
G6PDH2r (zwf)	5.86	0.19
GND	5.84	0.19
PGI	5.64	0.26
PFK	3.12	0.51
FBA	3.09	0.51
PPC	3.04	0.08

Case Study (1)

2. Gene essentiality analysis to determine genes that can be removed

Reaction	Sampled flux mean	Sampled flux St.D.	# of genes	Essential	Expressed
ATPS4rpp (atpIC)	56.37	0.27	9	No	Yes
PGK	19.05	0.09	1	Yes	Yes
GAPD	19.05	0.09	1	Yes	Yes
SUCDi	12.29	2.26	4	No	Yes
GLCptspp (ptsHIcrr)	11.64	0.19	8	No	Yes
TPI	8.60	0.09	1	No	Yes
FRD3	6.66	2.15	4	No	No
G6PDH2r (zwf)	5.86	0.19	1	No	Yes
GND	5.84	0.19	1	No	Yes
PGI	5.64	0.26	1	No	Yes
PFK	3.12	0.51	2	Yes	Yes
FBA	3.09	0.51	3	No	Yes
PPC	3.04	0.08	1	No	Yes

Case Study (1)

3. Experimental feasibility check

Reaction	Sampled flux mean	Sampled flux St.D.	# of genes	Essential	Expressed	Single KO
ATPS4rpp (atpIC)	56.37	0.27	9	No	Yes	Yes
PGK	19.05	0.09	1	Yes	Yes	Yes
GAPD	19.05	0.09	1	Yes	Yes	Yes
SUCDi	12.29	2.26	4	No	Yes	Yes
GLCptspp (ptsHIcrr)	11.64	0.19	8	No	Yes	Yes
TPI	8.60	0.09	1	No	Yes	Yes
FRD3	6.66	2.15	4	No	No	Yes
G6PDH2r (zwf)	5.86	0.19	1	No	Yes	Yes
GND	5.84	0.19	1	No	Yes	Yes
PGI	5.64	0.26	1	No	Yes	Yes
PFK	3.12	0.51	2	Yes	Yes	No
FBA	3.09	0.51	3	No	Yes	No
PPC	3.04	0.08	1	No	Yes	Yes

Case Study (1)

4. Coupling analysis to determine redundant KOs

Reaction	Sampled flux mean	Sampled flux St.D.	# of genes	Essential	Expressed	Single KO	Coupled
ATPS4rpp (atpIC)	56.37	0.27	9	No	Yes	Yes	
PGK	19.05	0.09	1	Yes	Yes	Yes	GAPD
GAPD	19.05	0.09	1	Yes	Yes	Yes	PGK
SUCDi	12.29	2.26	4	No	Yes	Yes	
GLCptspp (ptsHIcrr)	11.64	0.19	8	No	Yes	Yes	
TPI	8.60	0.09	1	No	Yes	Yes	
FRD3	6.66	2.15	4	No	No	Yes	
G6PDH2r (zwf)	5.86	0.19	1	No	Yes	Yes	PGL
GND	5.84	0.19	1	No	Yes	Yes	
PGI	5.64	0.26	1	No	Yes	Yes	
PFK	3.12	0.51	2	Yes	Yes	No	
FBA	3.09	0.51	3	No	Yes	No	
PPC	3.04	0.08	1	No	Yes	Yes	

Case Study (1)

5. Metabolomics coverage analysis to determine if nearby metabolites are possible to measure

Reaction	Sampled flux mean	Sampled flux St.D.	# of genes	Essential	Expressed	Single KO	Coupled	Mets up	Mets down
ATPS4rpp (atpIC)	56.37	0.27	9	No	Yes	Yes		adp	atp
PGK	19.05	0.09	1	Yes	Yes	Yes	GAPD	3pg, atp	adp
GAPD	19.05	0.09	1	Yes	Yes	Yes	PGK	nad, g3p	nadh
SUCDi	12.29	2.26	4	No	Yes	Yes		succ	fum
GLCptspp (ptsHIcrr)	11.64	0.19	8	No	Yes	Yes		pep	g6p, pyr
TPI	8.60	0.09	1	No	Yes	Yes		dhap	g3p
FRD3	6.66	2.15	4	No	No	Yes		fum	succ
G6PDH2r (zwf)	5.86	0.19	1	No	Yes	Yes	PGL	g6p, nadp	nadph
GND	5.84	0.19	1	No	Yes	Yes		nadp, 6pgc	Ru5p, nadph
PGI	5.64	0.26	1	No	Yes	Yes		g6p	f6p
PFK	3.12	0.51	2	Yes	Yes	No		f6p, atp	fdp, adp
FBA	3.09	0.51	3	No	Yes	No		fdp	g3p, dhap
PPC	3.04	0.08	1	No	Yes	Yes		pep	oaa

Case Study (1)

6. Implementation

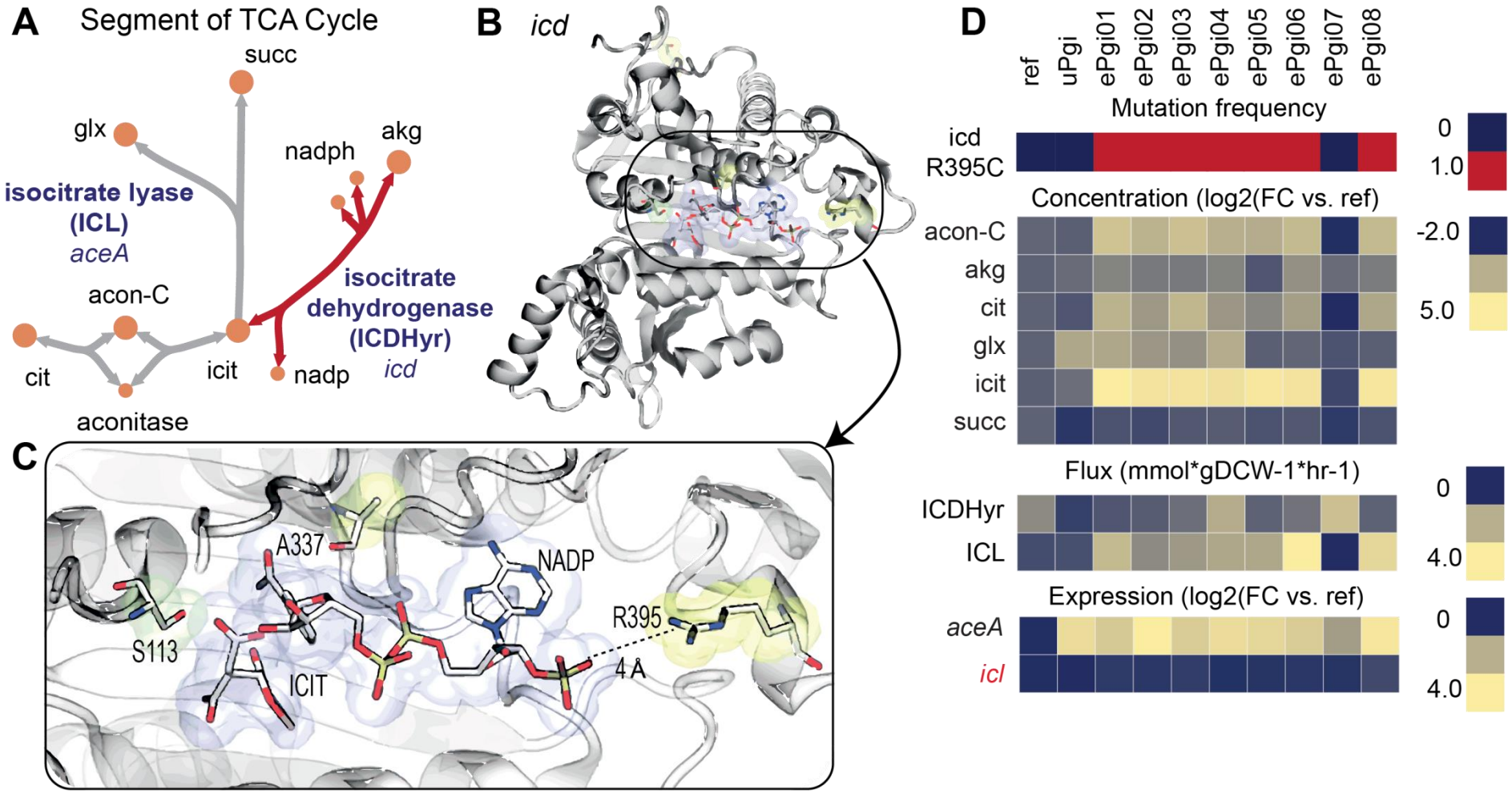
Reaction	Sampled flux mean	Sampled flux St.D.	# of genes	Essential	Expressed	Single KO	Coupled	Mets up	Mets down	Implemented
ATPS4rpp (atpIC)	56.37	0.27	9	No	Yes	Yes		adp	atp	No, failed using lambda red
PGK	19.05	0.09	1	Yes	Yes	Yes	GAPD	3pg, atp	adp	No
GAPD	19.05	0.09	1	Yes	Yes	Yes	PGK	nad, g3p	nadh	No
SUCDi	12.29	2.26	4	No	Yes	Yes		succ	fum	Yes
GLCptspp (ptsHIcrr)	11.64	0.19	8	No	Yes	Yes		pep	g6p, pyr	Yes
TPI	8.60	0.09	1	No	Yes	Yes		dhap	g3p	Yes
FRD3	6.66	2.15	4	No	No	Yes		fum	succ	No
G6PDH2r (zwf)	5.86	0.19	1	No	Yes	Yes	PGL	g6p, nadp	nadph	No, failed using lambda red
GND	5.84	0.19	1	No	Yes	Yes		nadp, 6pgc	Ru5p, nadph	Yes
PGI	5.64	0.26	1	No	Yes	Yes		g6p	f6p	Yes
PFK	3.12	0.51	2	Yes	Yes	No		f6p, atp	fdp, adp	No
FBA	3.09	0.51	3	No	Yes	No		fdp	g3p, dhap	No
PPC	3.04	0.08	1	No	Yes	Yes		pep	oaa	No, resulted in auxotrophy for asp-L

Case Study (1)

- Question: What are metabolic genes system's level function?
- Experiments: gene KOs, adaptive laboratory evolution (ALE)
  - Systems perturbation experiments in engineering: 1) perturb the system, 2) allow the system to recover, and 3) analyze how the system recovered
- Model: *E. coli* MG1655 K-12
- Experimental parameters: Gene KO strains and control before and after ALE
- Analyses: structural analysis, MFA, transcription factor perturbation analysis

Case Study (1)

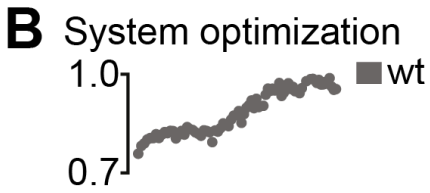
Phosphoglucose isomerase (*pgi*) KO



Case Study (1)

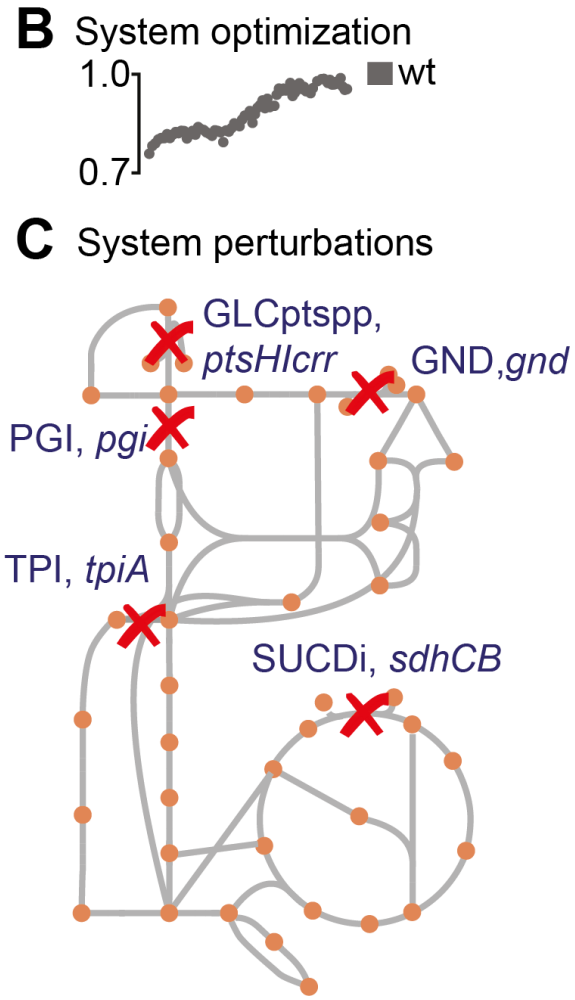
- Question: What are metabolic genes system's level function?
- Experiments: gene KOs, adaptive laboratory evolution (ALE)
  - Systems perturbation experiments in engineering: 1) perturb the system, 2) allow the system to recover, and 3) analyze how the system recovered
- Model: *E. coli* MG1655 K-12
- Experimental parameters: Gene KO strains and control before and after ALE
- Analyses: structural analysis, MFA, transcription factor perturbation analysis
- Data types: metabolomics, fluxomics, transcriptomics, genomics, and phenomics

Case Study (1)

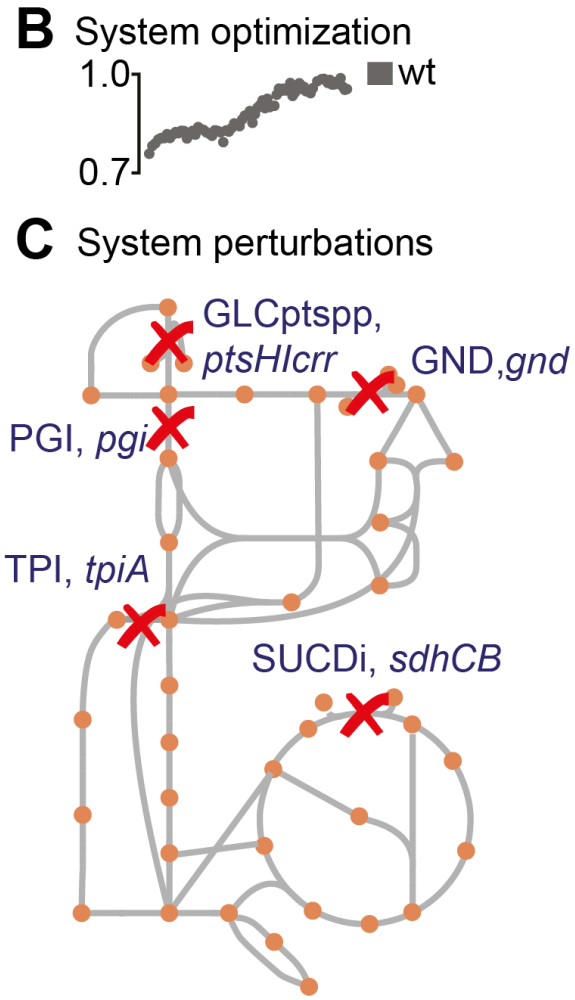


LaCroix et al., 2015 Appl. Environ. Microbiol. 81, 17–30.

Case Study (1)


B System optimization
1.0
0.7
wt
C System perturbations
GLCptspp,
ptsHIcrr
GND,gnd
PGI, pgi
TPI, tpiA
SUCDi, sdhCB

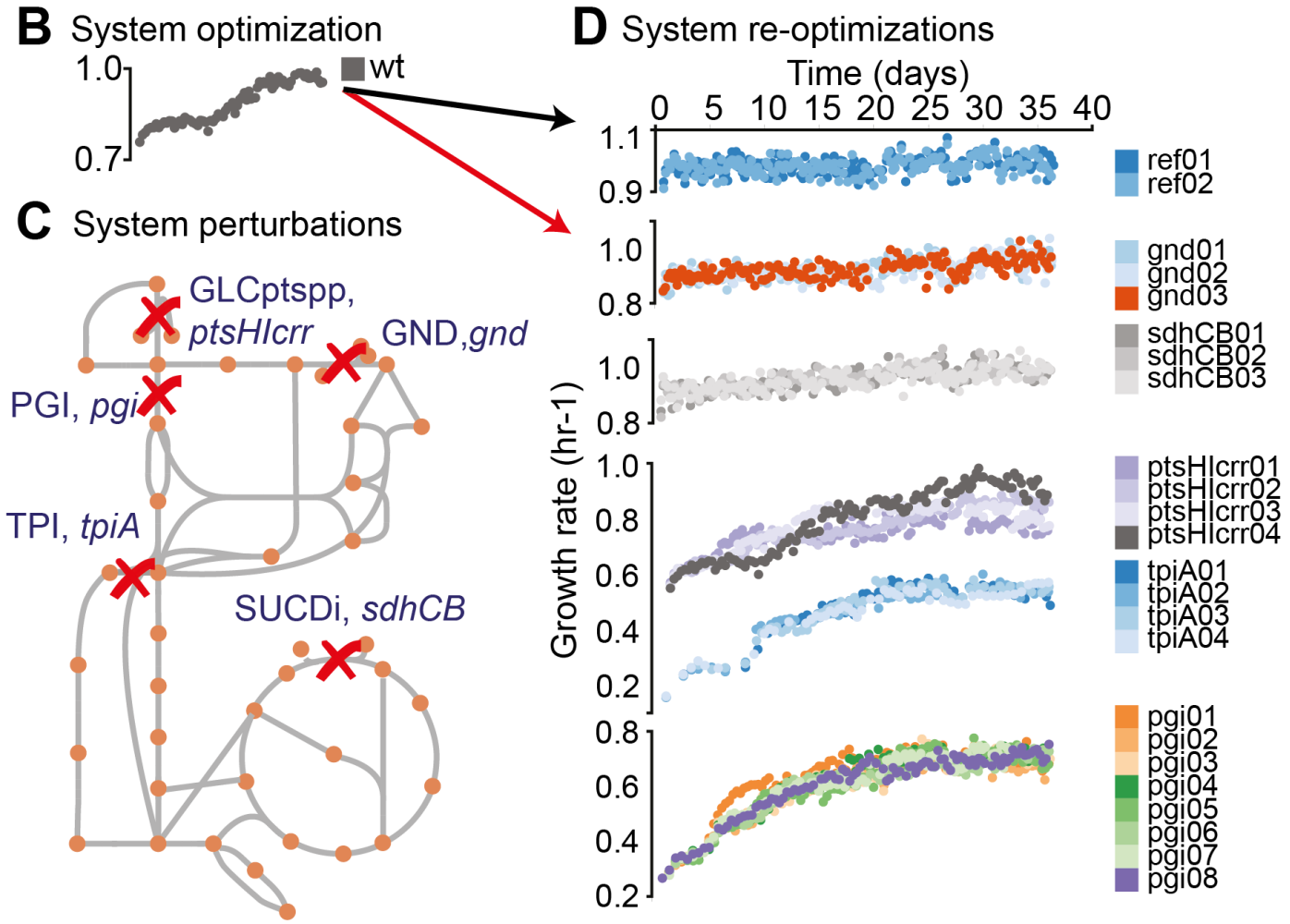
Case Study (1)



The most network perturbing knockouts were selected using a model-driven approach

Details to come…

Case Study (1)


B System optimization
1.0
0.7
wt
C System perturbations
GLCptspp,
ptsHIcrr
GND,gnd
PGI, pgi
TPI, tpiA
SUCDi, sdhCB
D System re-optimizations
Time (days)
0 5 10 15 20 25 30 35 40
Growth rate (hr-1)
1.1
0.9
1.0
0.8
1.0
0.8
1.0
0.8
0.6
0.4
0.2
0.8
0.6
0.4
0.2
ref01
ref02
gnd01
gnd02
gnd03
sdhCB01
sdhCB02
sdhCB03
ptsHIcrr01
ptsHIcrr02
ptsHIcrr03
ptsHIcrr04
tpiA01
tpiA02
tpiA03
tpiA04
pgi01
pgi02
pgi03
pgi04
pgi05
pgi06
pgi07
pgi08

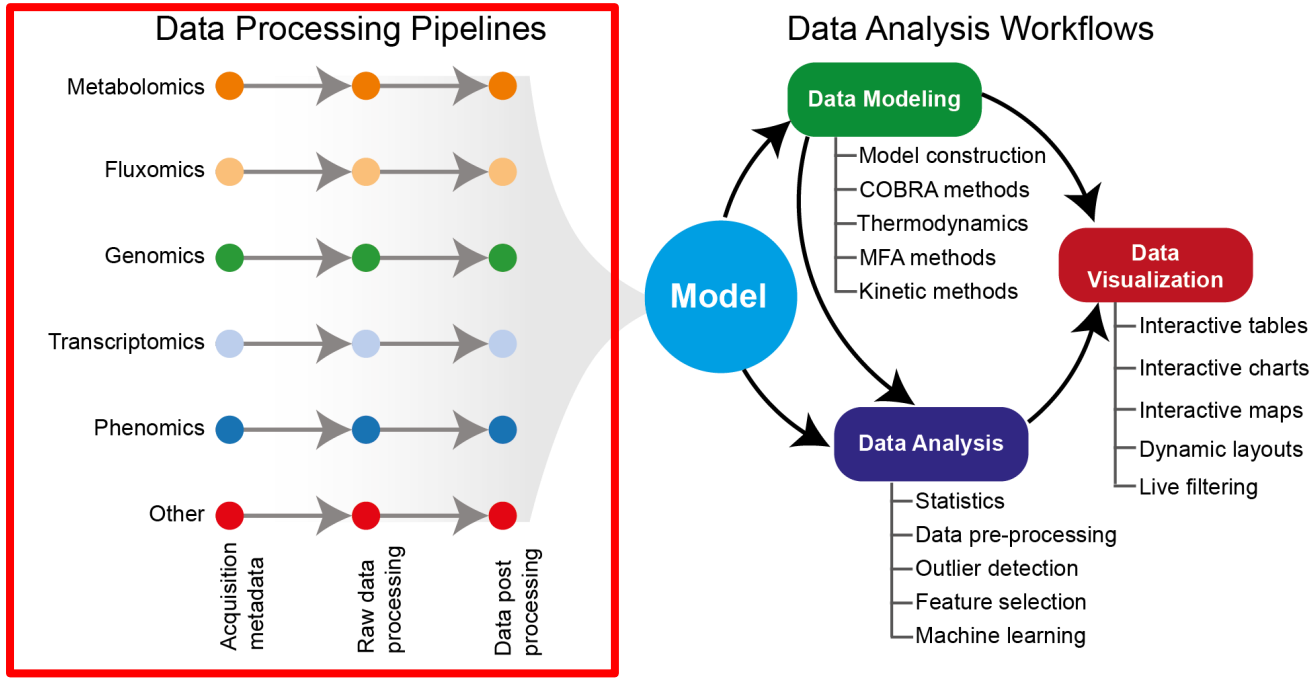
Case Study (1)

E System changes	Metabolomics		Transcriptomics		Fluxomics		Mutations	
	Ave % of total	STD % of total	Ave % of total	STD % of total	Ave % of total	STD % of total	Ave	STD
ref vs. eRefs	2.2	0.0	0.4	0.1	0.0	0.0	6.5	0.7
uGnd vs. eGnd	41.1	1.1	14.8	0.3	17.0	0.8	10.7	7.4
uPgi vs. ePgi	29.9	5.2	3.8	2.8	43.0	3.8	14.0	7.3
uPtsHIcrr vs. ePtsHIcrr	20.8	2.8	19.1	2.7	38.3	1.9	12.0	4.8
uSdhCB vs. eSdhCB	24.4	5.9	10.4	0.5	13.8	5.5	7.7	2.3
uTpiA vs. eTpiA	36.9	2.9	10.5	0.8	40.2	0.1	19.7	7.4

Confirmation of the Experimental Design

Data processing pipelines *From raw to processed data*

Data processing (1) endo-metabolomics
Data processing (2) exo-metabolomics
Data processing (3) isotope labeling
Data processing (4) proteomics
Data processing (5) transcriptomics
Data processing (6) genomics

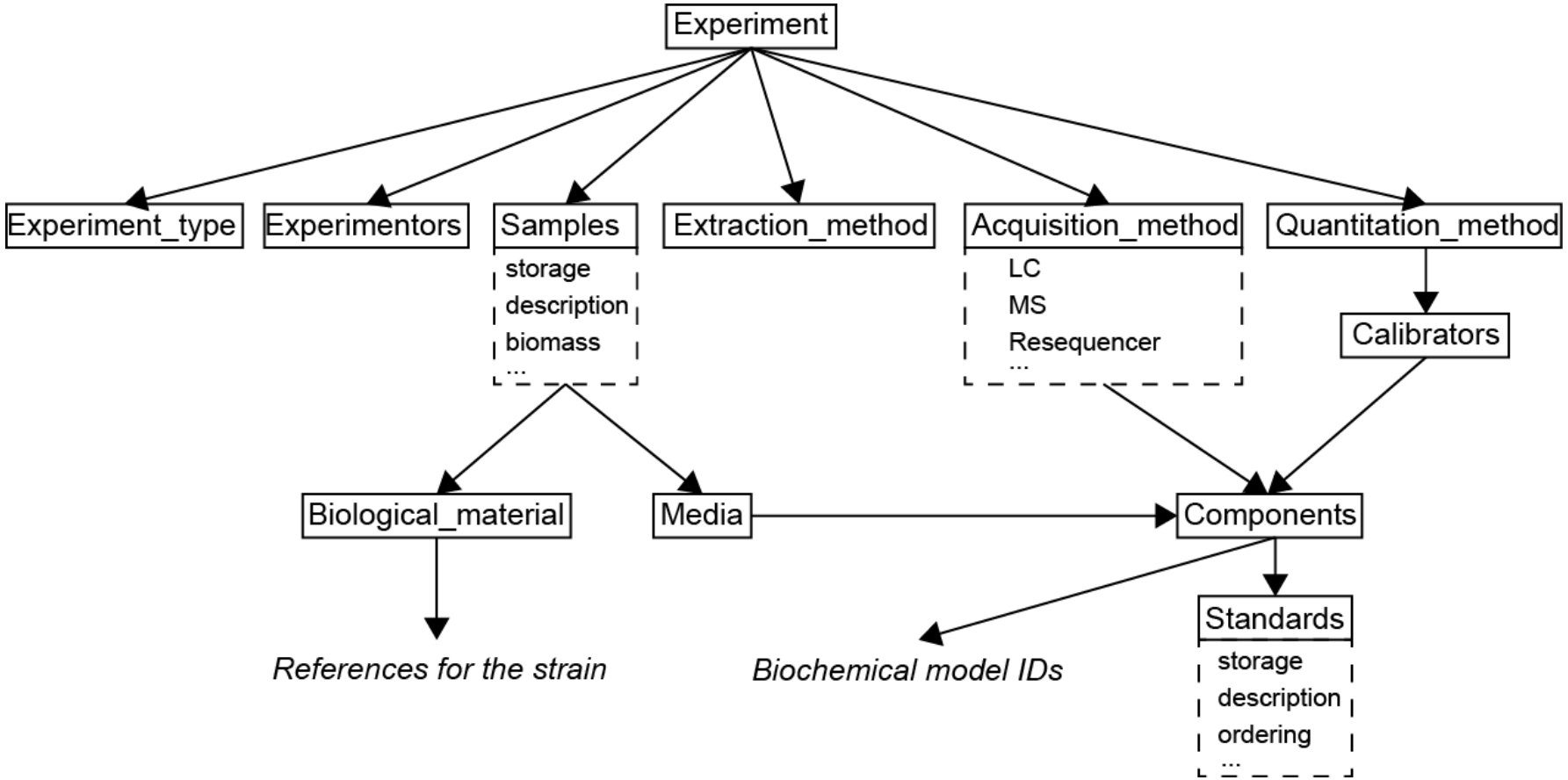


Data processing (0) LIMS

- Systems biology supported LIMS

Data processing (0) LIMS

- Systems biology supported LIMS



Data processing (0) LIMS

- Systems biology supported LIMS
- Controlled vocabulary (CV)

Data processing (0) LIMS

- Systems biology supported LIMS
- Controlled vocabulary (CV)
- General schema
  - Experiment (workflow) -> sample type
  - SOPs of the experiment capture important metadata in structure fields
  - script to pull out and propagate/accumulate metadata chain of experimental workflows used to derive the sample

Data processing (0) LIMS

- Systems biology supported LIMS
- Controlled vocabulary (CV)
- General schema
  - Experiment (workflow) -> sample type
  - SOPs of the experiment capture important metadata in structure fields
  - script to pull out and propagate/accumulate metadata chain of experimental workflows used to derive the sample
- Pushes all experimental data to a data lake for “Big data” analysis
  - Component_name (Feature): MetaData
  - Sample_name: MetaData
  - Value
  - Dimensions: time, location, etc.,
  - Units

Data processing (1/2) endo/exo-metabolomics

Pre-requisites: sampling and extraction, and labeled internal standards

Metabolomics (2015) 11:198–209
DOI 10.1007/s11306-014-0686-2

ORIGINAL ARTICLE

Fast Swinnex filtration (FSF): a fast and robust sampling and extraction method suitable for metabolomics analysis of cultures grown in complex media

Douglas McCloskey · Jose Utrilla · Robert K. Naviaux · Bernhard O. Palsson · Adam M. Feist

Data processing (1/2) endo/exo-metabolomics

Pre-requisites: LC-MS/MS method

Metabolomics (2015) 11:1338–1350
DOI 10.1007/s11306-015-0790-y

ORIGINAL ARTICLE

A pH and solvent optimized reverse-phase ion-paring-LC–MS/MS method that leverages multiple scan-types for targeted absolute quantification ofintracellular metabolites

Douglas McCloskey · Jon A. Gangoiti · Bernhard O. Palsson · Adam M. Feist

Data processing (1/2) endo/exo-metabolomics

Pre-requisites: LC-MS/MS method

Metabolic Engineering 47 (2018) 383–392

Contents lists available at ScienceDirect

Metabolic Engineering

journal homepage: www.elsevier.com/locate/meteng

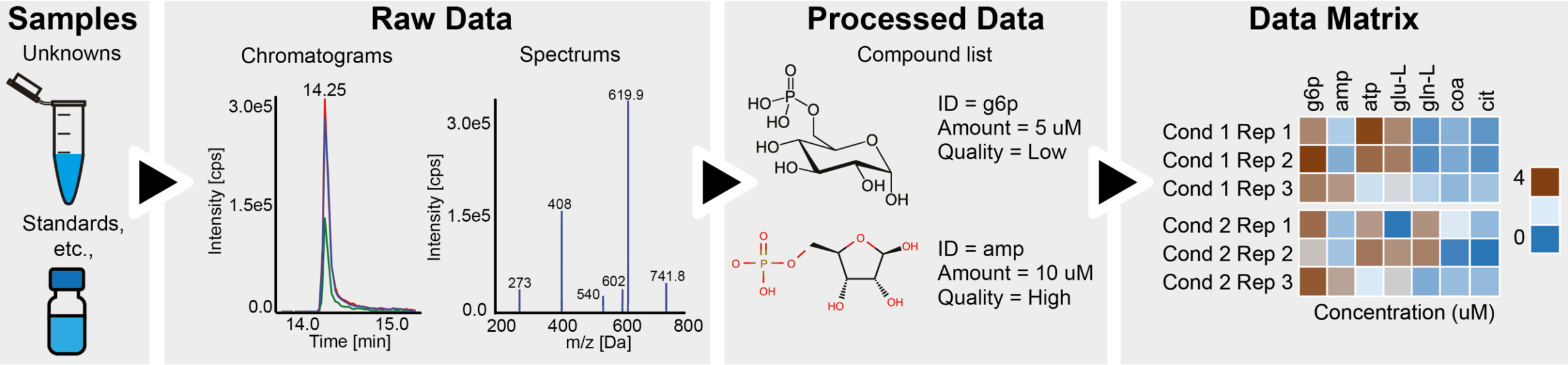
RapidRIP quantifies the intracellular metabolome of 7 industrial strains of E. coli

Douglas McCloskey [a], Julia Xu [b], Lars Schrübbers [a], Hanne B. Christensen [a], Markus J. Herrgård [a,*]

[a] Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark
[b] Department of Bioengineering, University of California - San Diego, La Jolla, CA 92093, USA

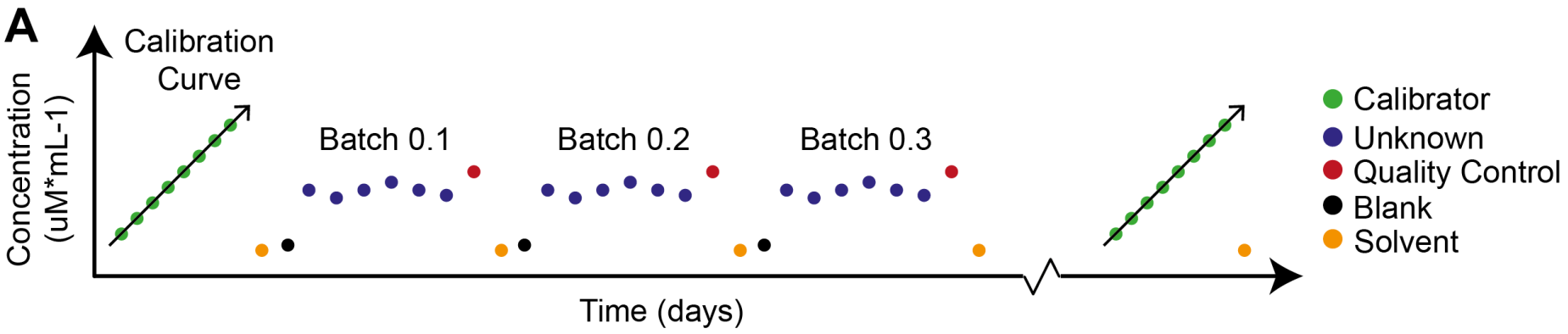
Data processing (1/2) endo/exo-metabolomics

- Chromatograms and spectrum to list of annotated metabolites with corresponding concentrations/abundances/fold changes



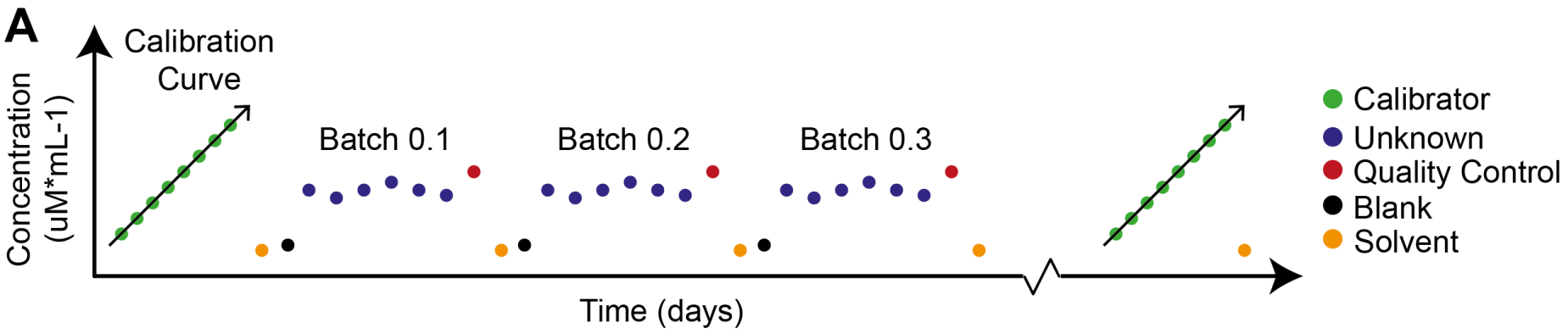
Data processing (1/2) endo/exo-metabolomics

Experiment design



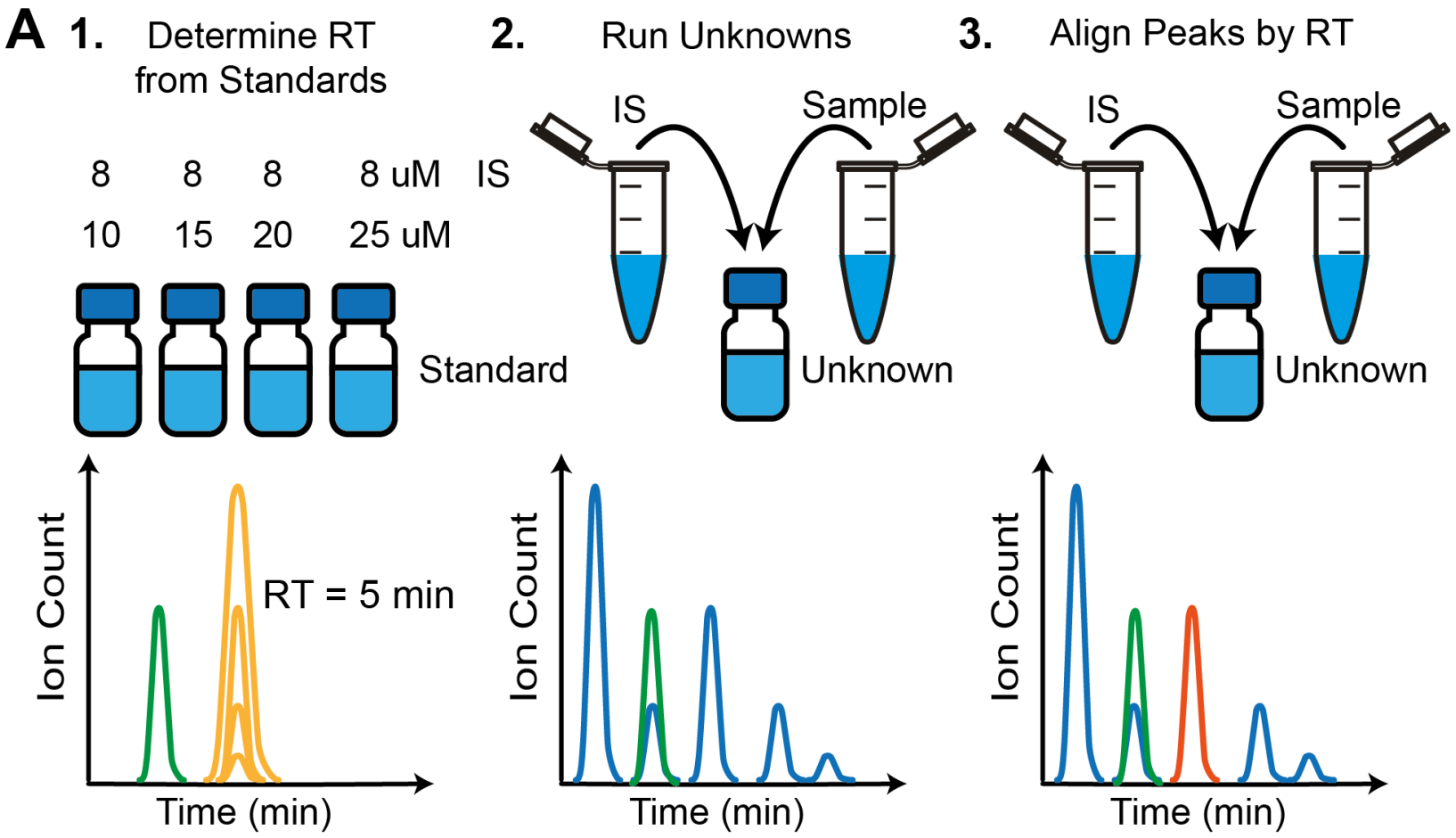
Data processing (1/2) endo/exo-metabolomics

Experiment design



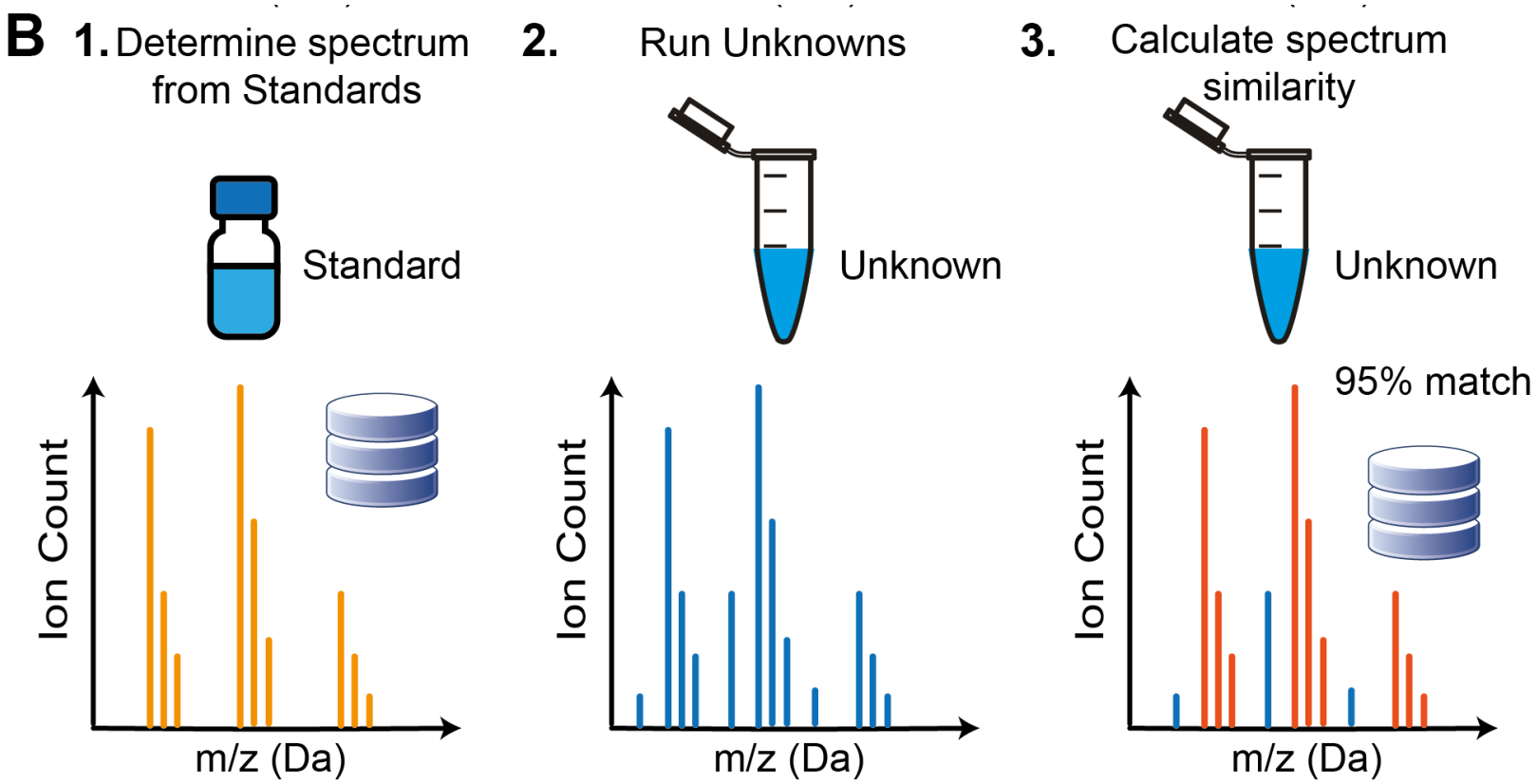
Extended batch design

Data processing (1/2) endo/exo-metabolomics



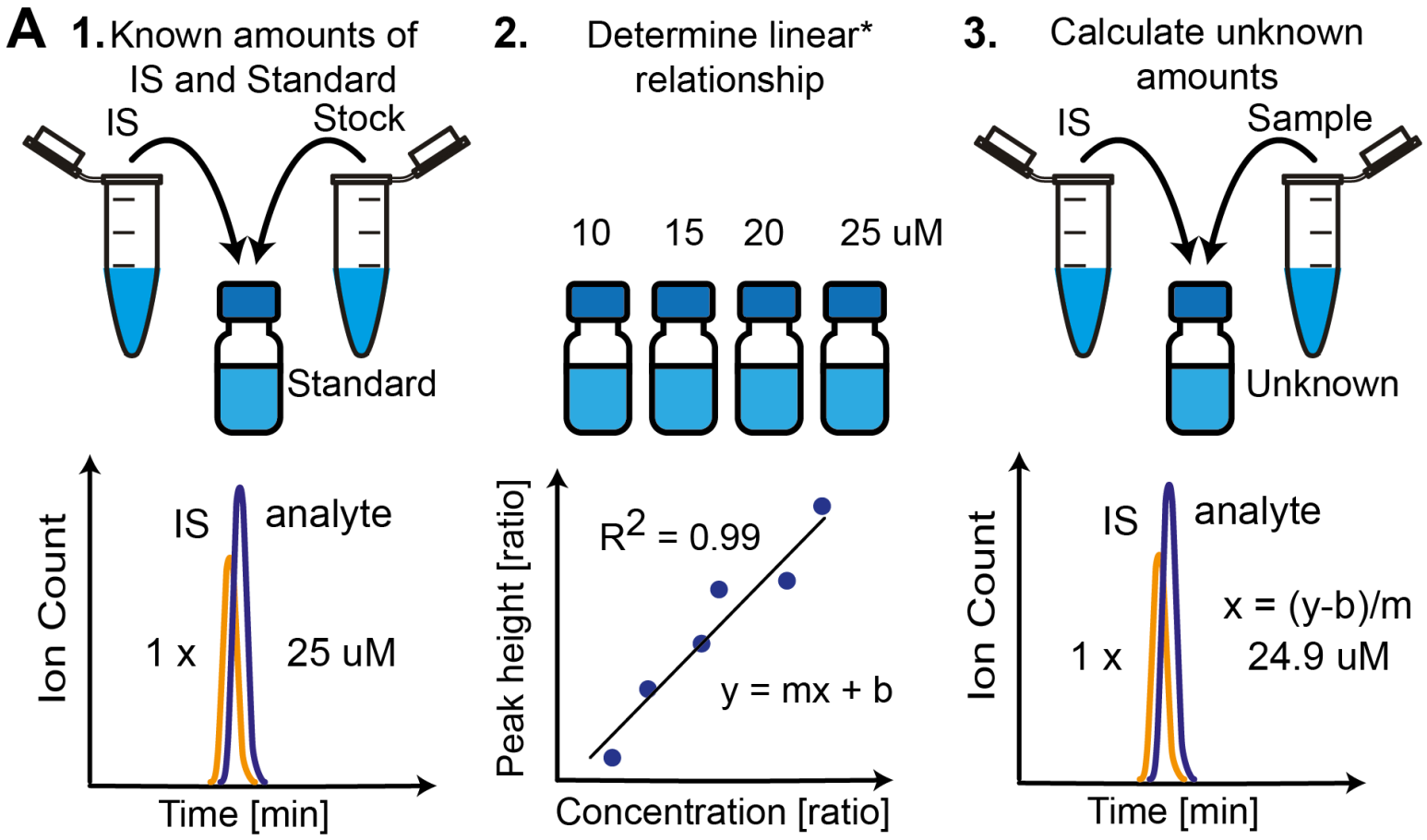
Retention time (RT) alignment

Data processing (1/2) endo/exo-metabolomics



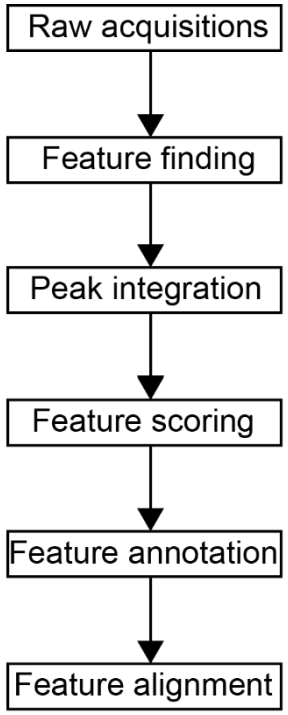
Met ID from Spectrum

Data processing (1/2) endo/exo-metabolomics



Quantitation

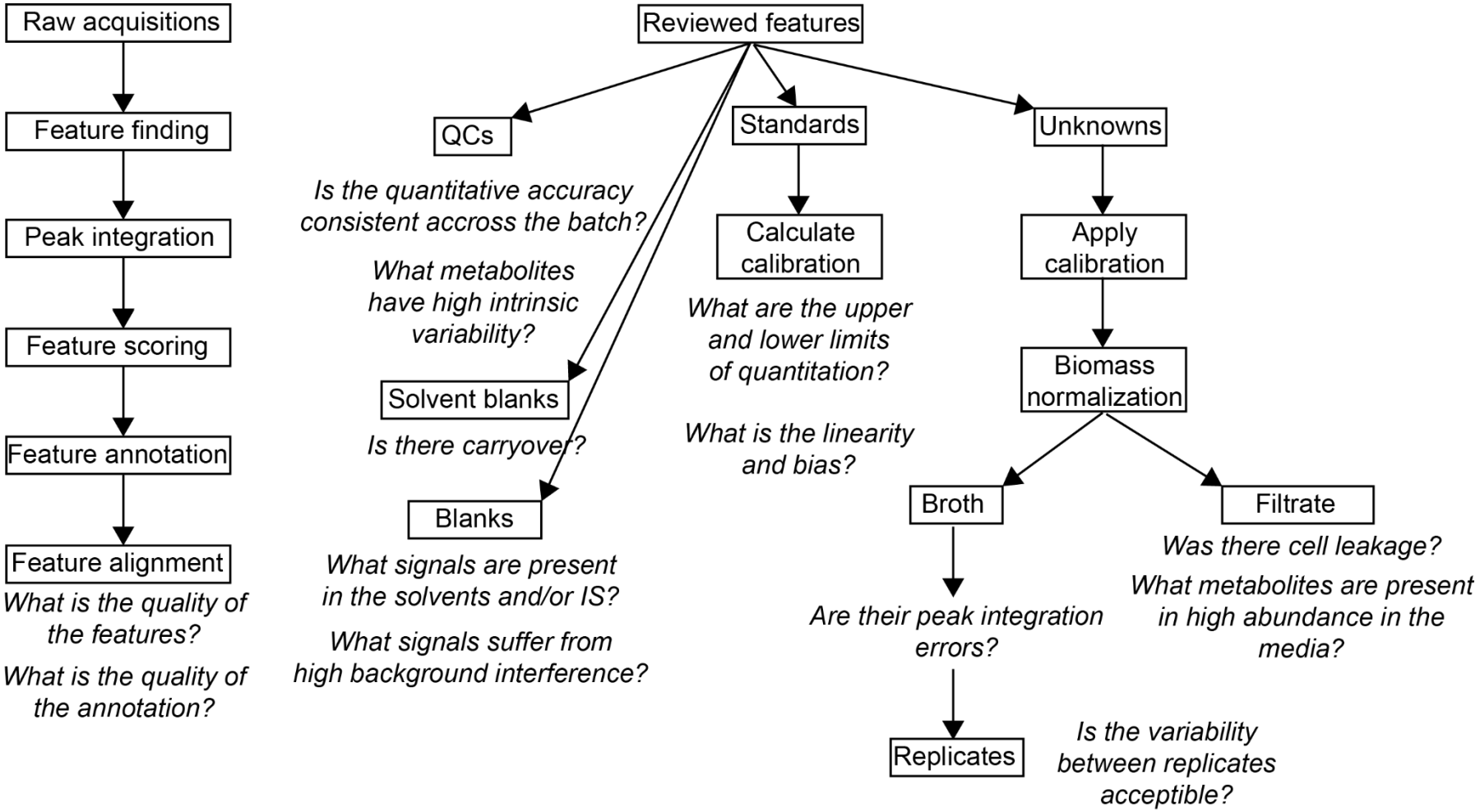
Data processing (1/2) endo/exo-metabolomics



What is the quality of the features?

What is the quality of the annotation?

Data processing (1/2) endo/exo-metabolomics



Data processing (3) isotope labeling

Pre-requisites: LC-MS/MS method

analytical chemistry

Article

pubs.acs.org/ac

MID Max: LC−MS/MS Method for Measuring the Precursor and Product Mass Isotopomer Distributions of Metabolic Intermediates and Cofactors for Metabolic Flux Analysis Applications

Douglas McCloskey[†], Jamey D. Young[§,⊠], Sibei Xu[†], Bernhard O. Palsson[†,‡], and Adam M. Feist[*,†,‡]

[†]Department of Bioengineering, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, United States

[‡]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark

[§]Department of Chemical and Biomolecular Engineering [⊠]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee 37235-1604, United States

Data processing (3) isotope labeling

Pre-requisites: Isotope deconvolution

analytical chemistry

Article

pubs.acs.org/ac

Modeling Method for Increased Precision and Scope of Directly Measurable Fluxes at a Genome-Scale

Douglas McCloskey[†], Jamey D. Young[§,⊠], Sibei Xu[†], Bernhard O. Palsson[†,‡], and Adam M. Feist[*,†,‡]

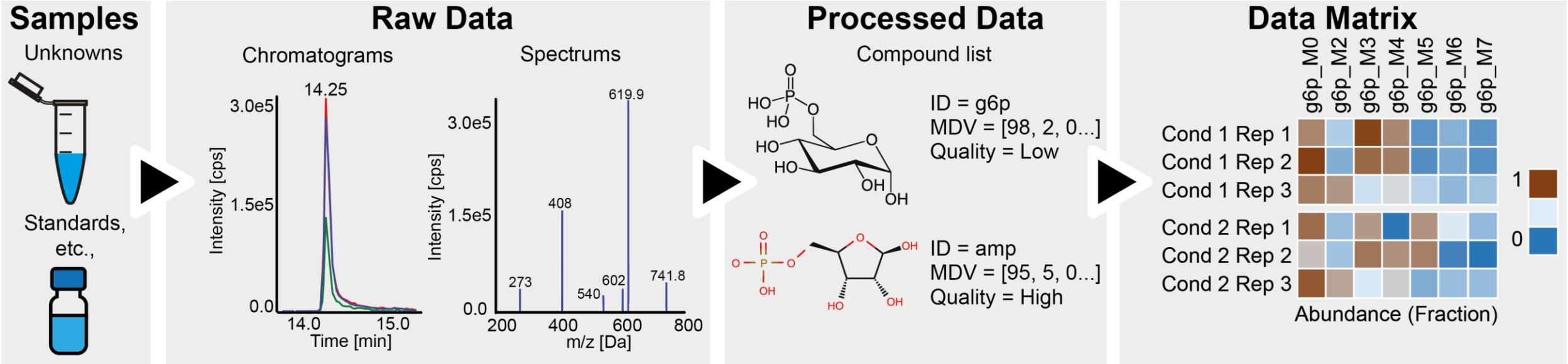
[†]Department of Bioengineering, University of California, San Diego, California 92093, United States

[‡]Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, 2800 Lyngby, Denmark

[§]Department of Chemical and Biomolecular Engineering, [⊠]Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee 37235-1604, United States

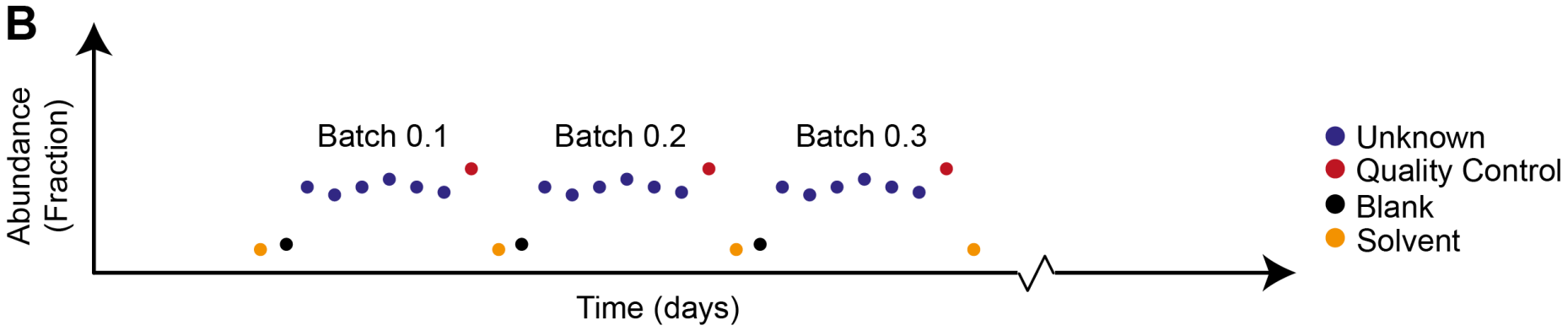
Data processing (3) isotope labeling

- Chromatograms and spectrum to list of annotated metabolites with mass distribution vectors (MDVs)



Data processing (1/2) endo/exo-metabolomics

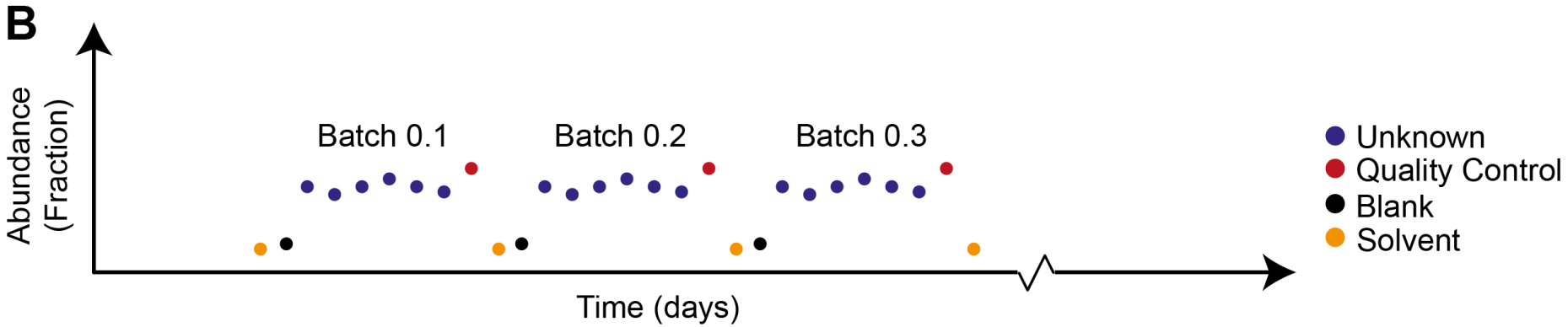
Experiment design



Extended batch design

Data processing (1/2) endo/exo-metabolomics

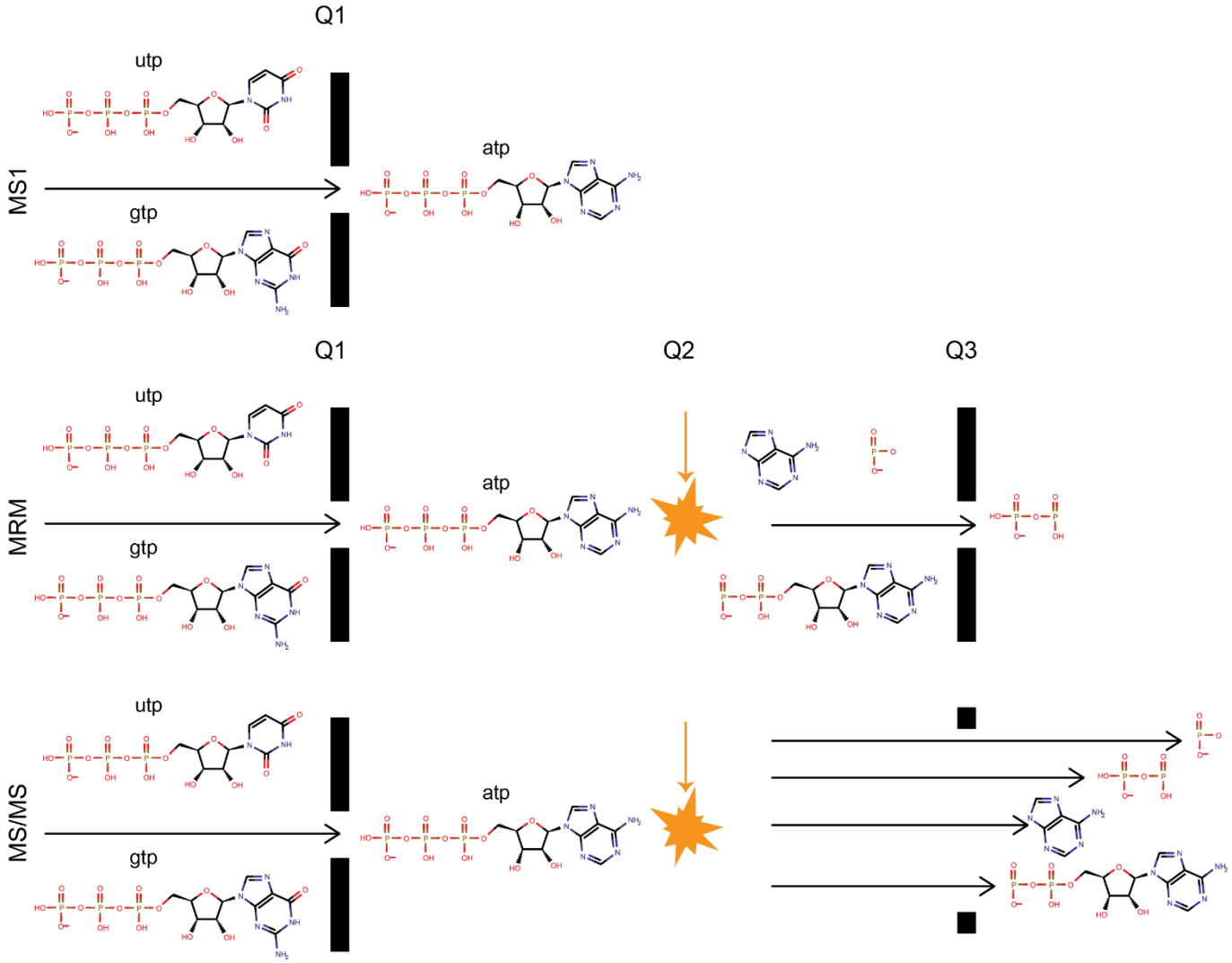
Experiment design



Extended batch design

- Unknown = biological samples grown *with* and *without* labeled media
- Quality control = pooled biological sample to test for compound variance
- Blank = sample matrix (without biological material) that has been through the entire sample workup
- Solvent = carryover checks

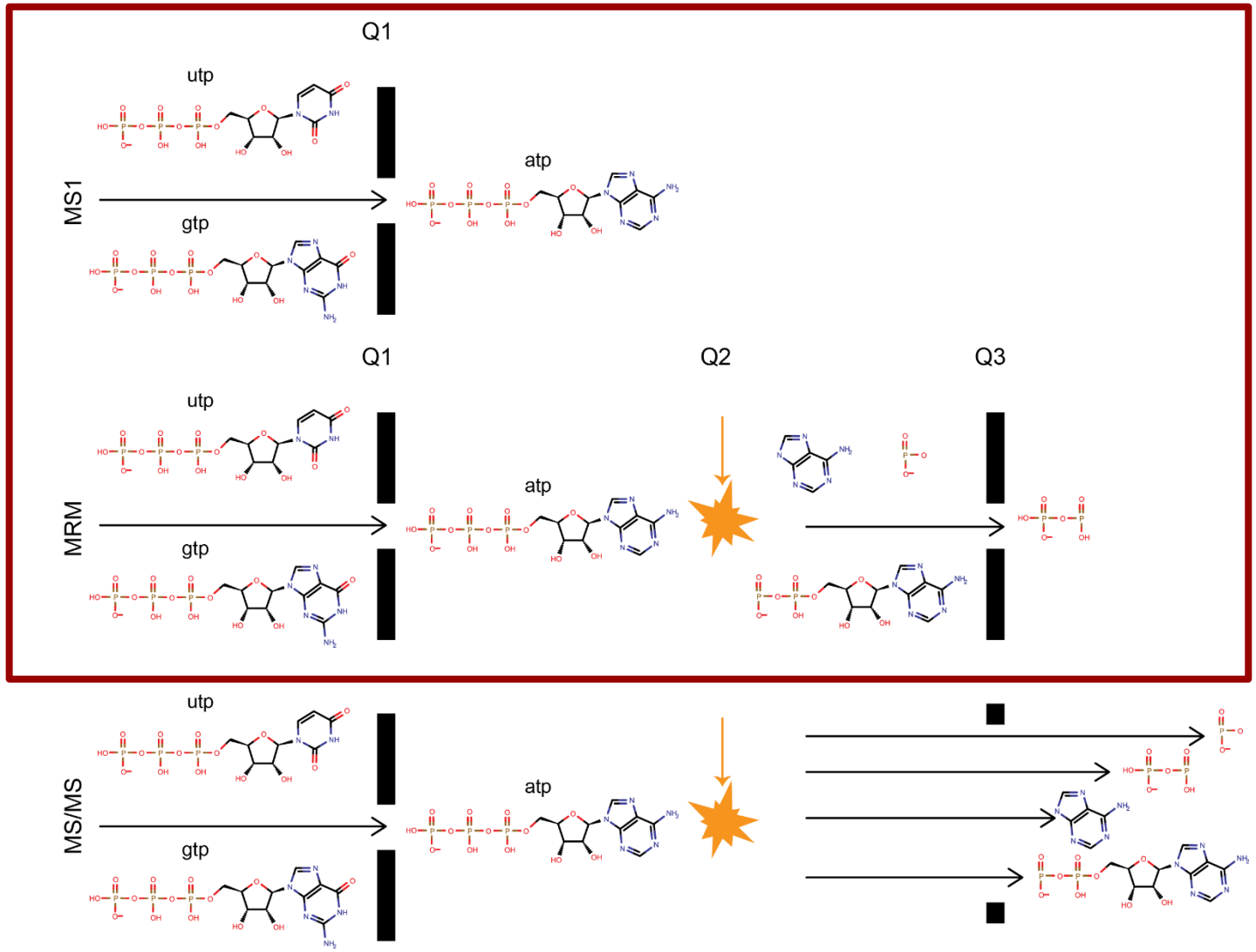
Data processing (3) isotope labeling



Scan types to capture precursor and product labeling information

10.1021/acs.analchem.5b03887

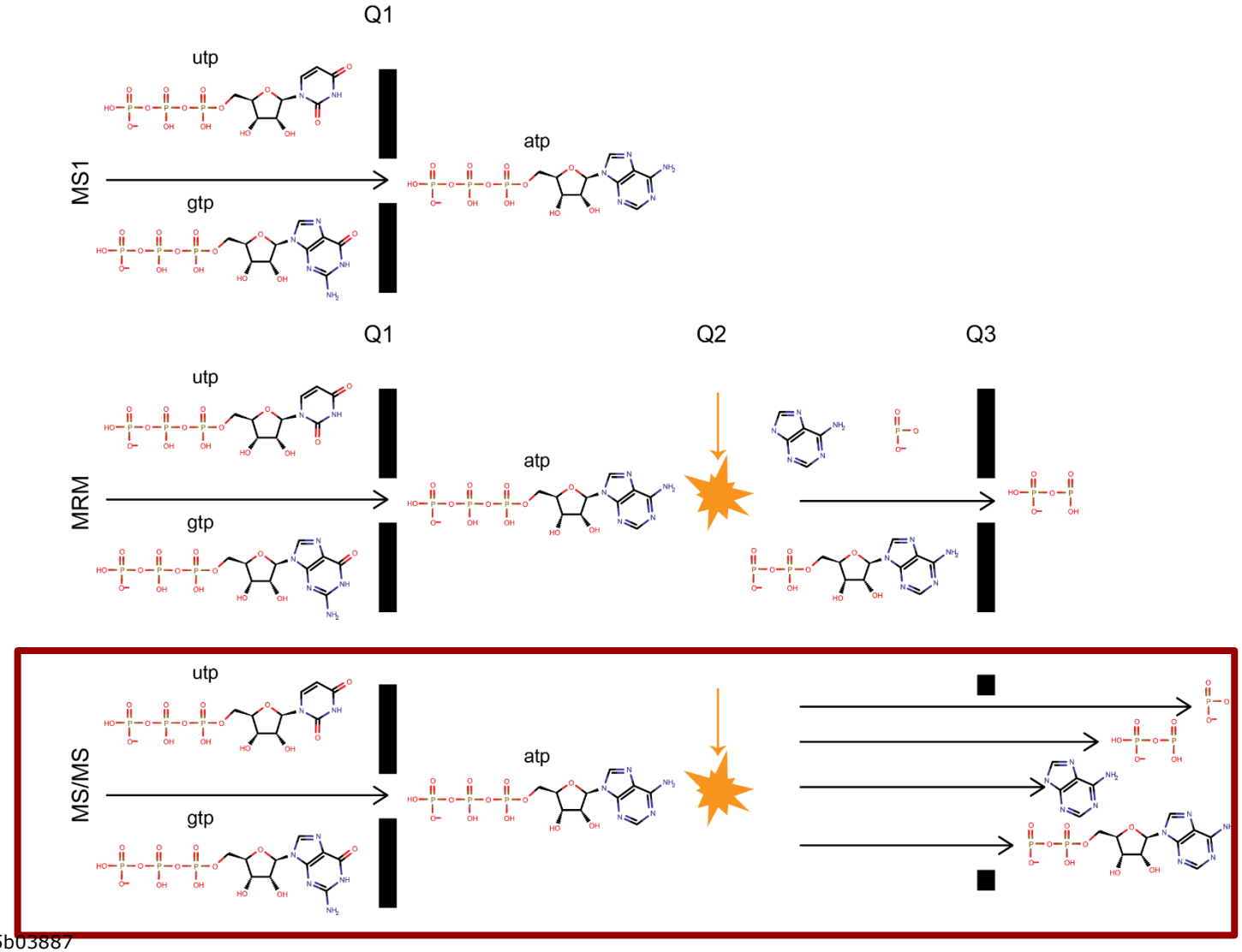
Data processing (3) isotope labeling



Scan types to capture **precursor** and product labeling information

10.1021/acs.analchem.5b03887

Data processing (3) isotope labeling

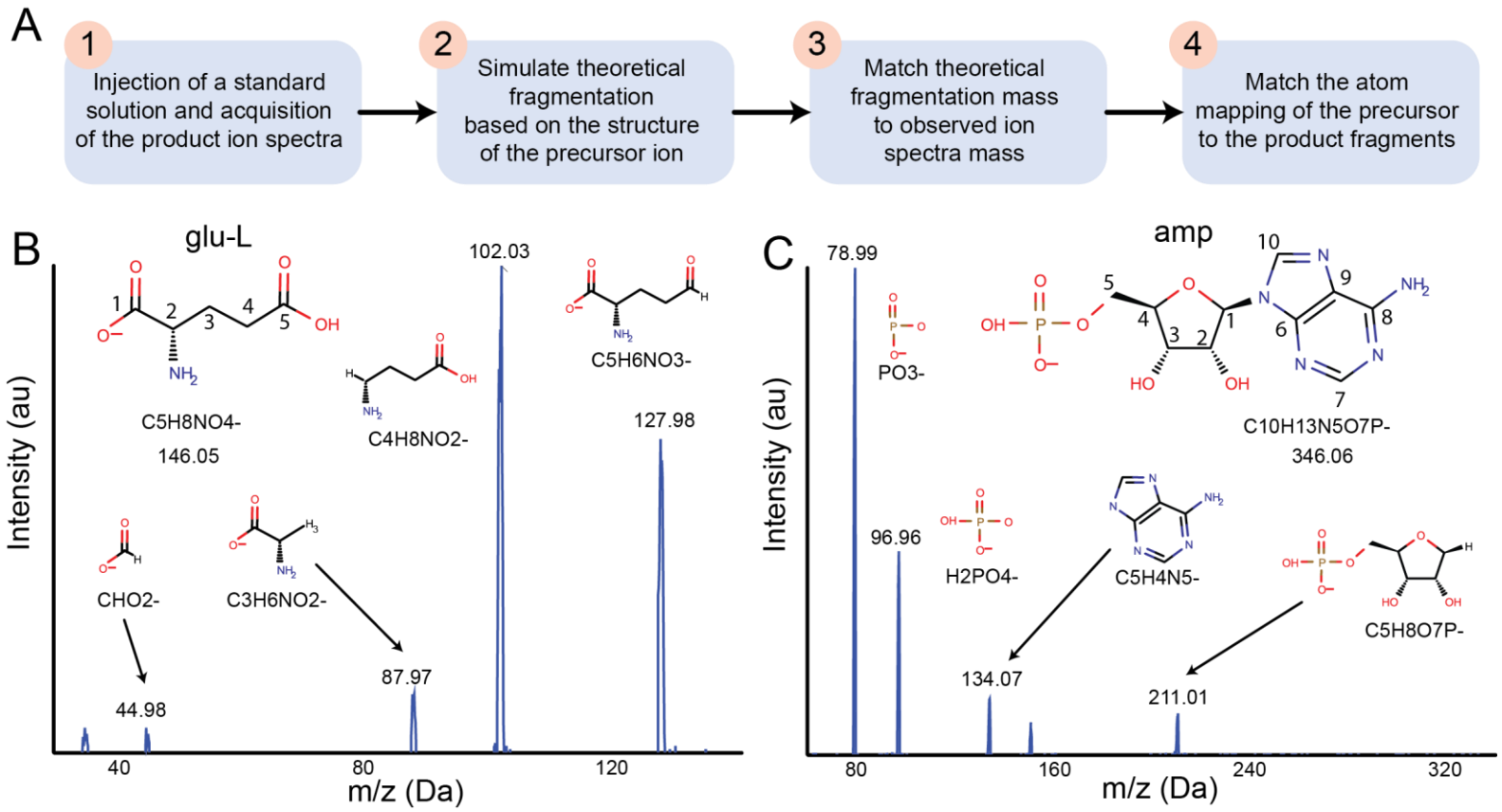


Scan types to capture precursor and **product** labeling information

10.1021/acs.analchem.5b03887

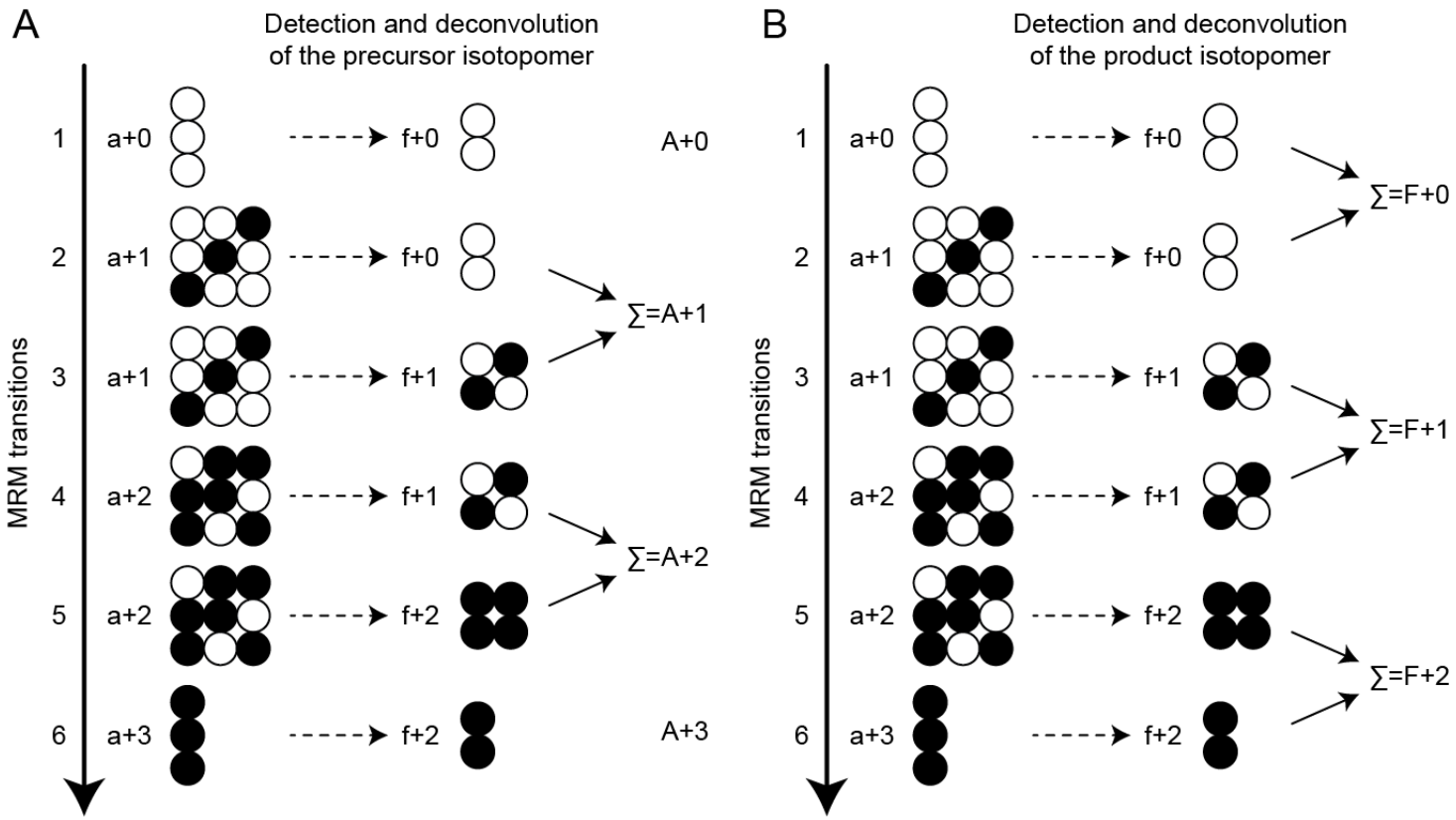
Data processing (3) isotope labeling

Atom mapping of measured fragments



10.1021/acs.analchem.5b03887

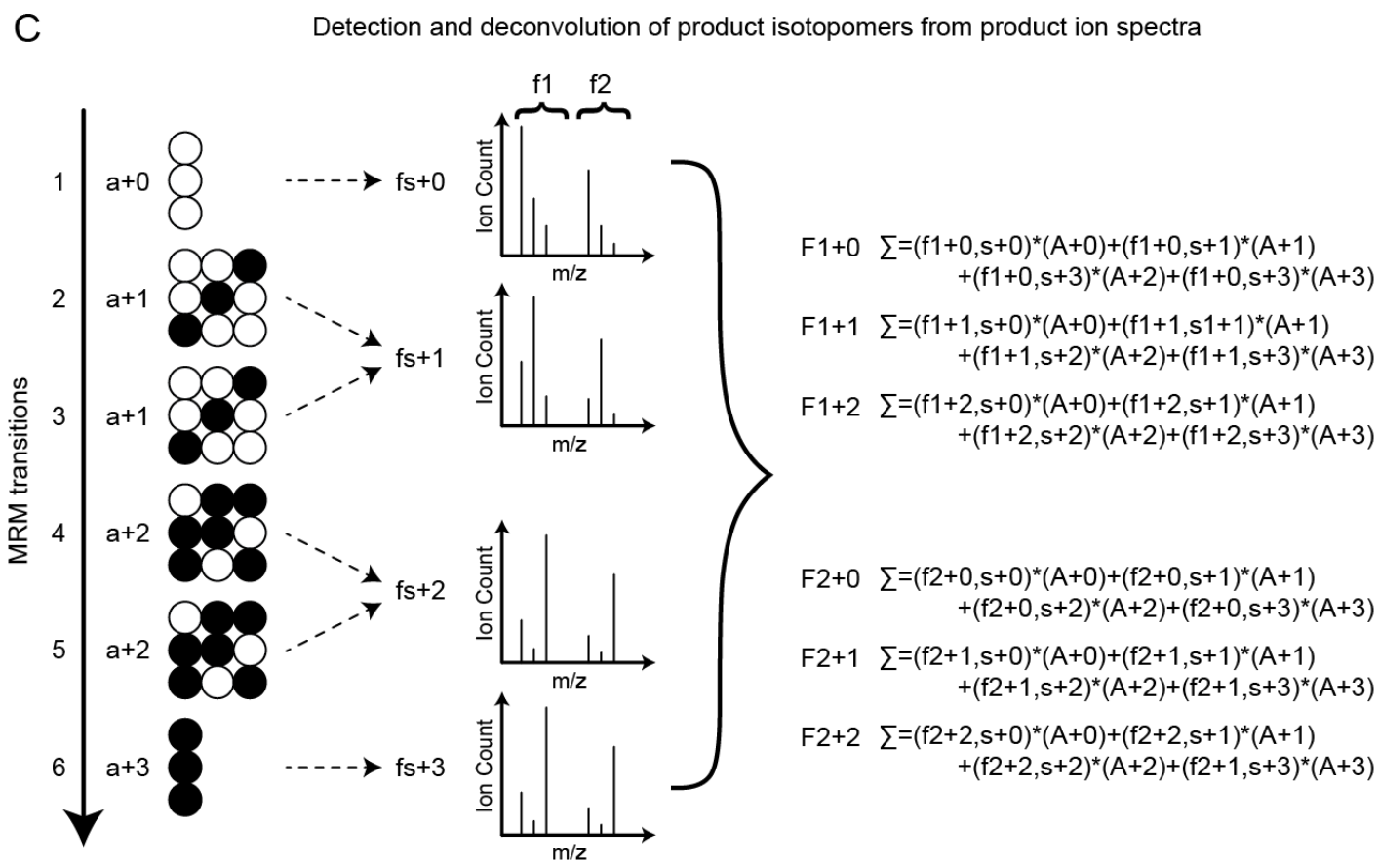
Data processing (3) isotope labeling



Chromatogram/Spectrum deconvolution

10.1021/acs.analchem.5b03887

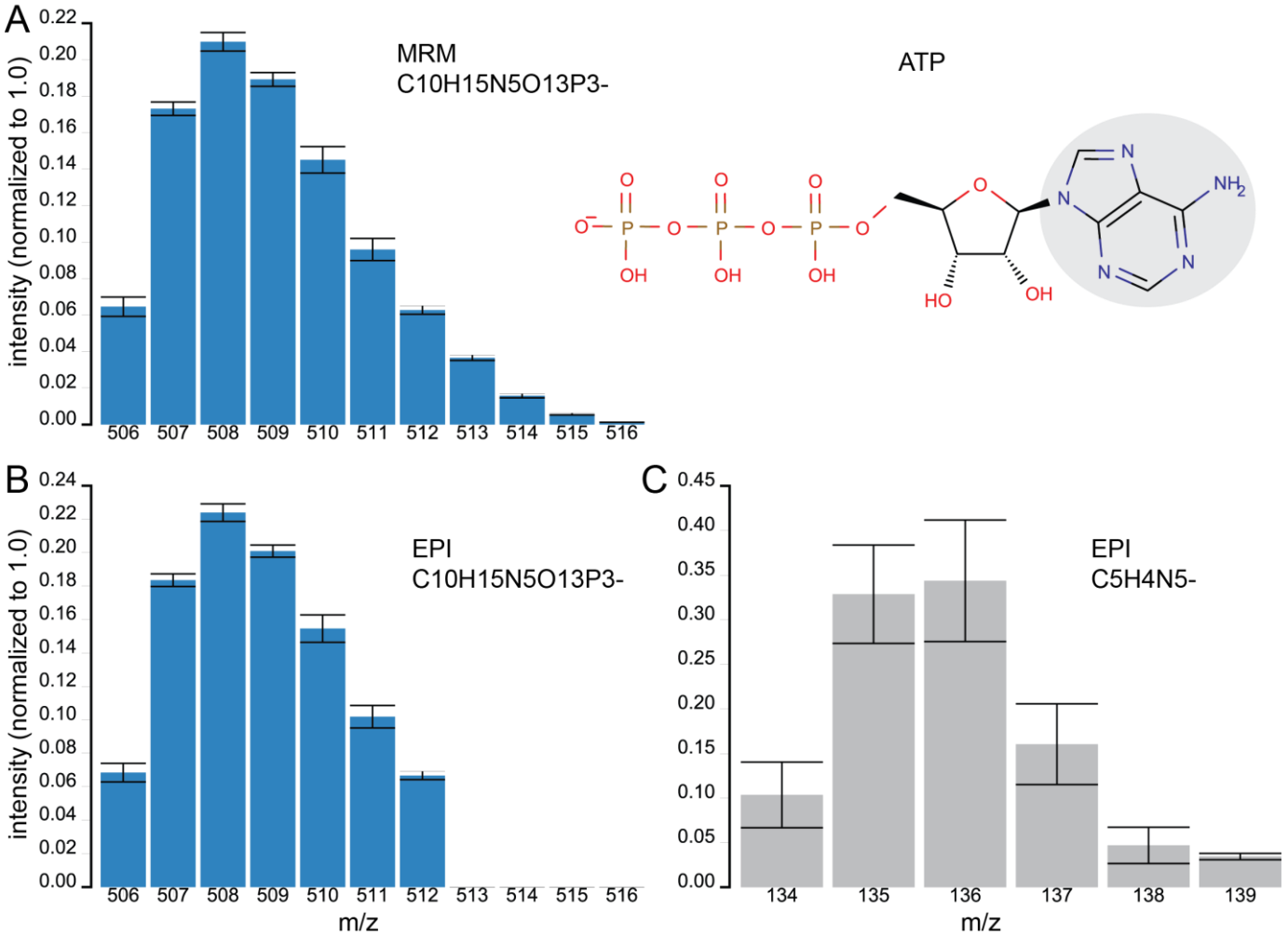
Data processing (3) isotope labeling



Chromatogram/Spectrum deconvolution

10.1021/acs.analchem.5b03887

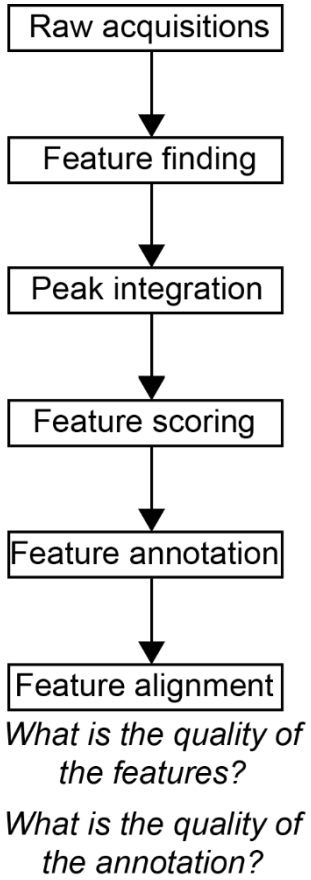
Data processing (3) isotope labeling



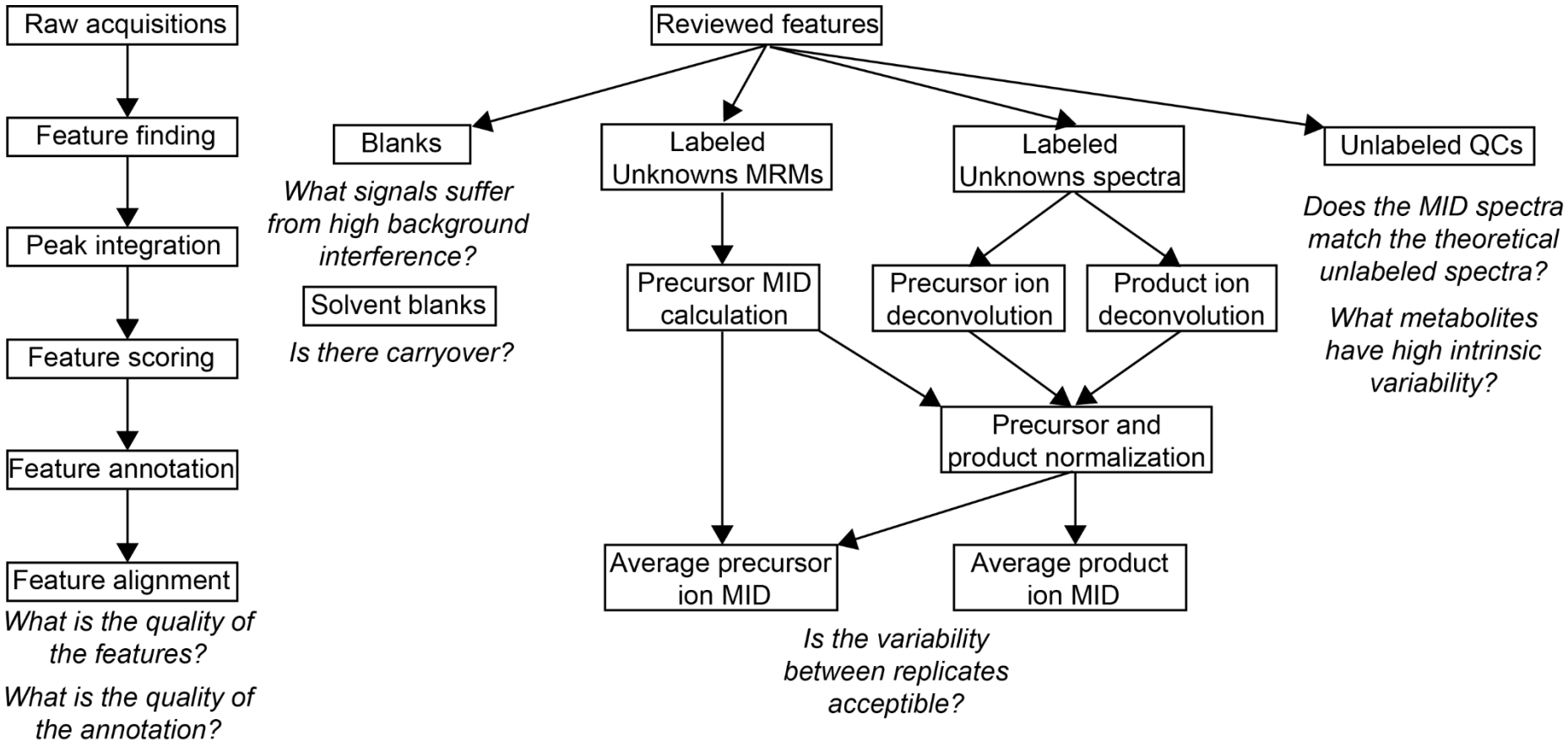
Conversion to MDVs

10.1021/acs.analchem.5b03887

Data processing (3) isotope labeling

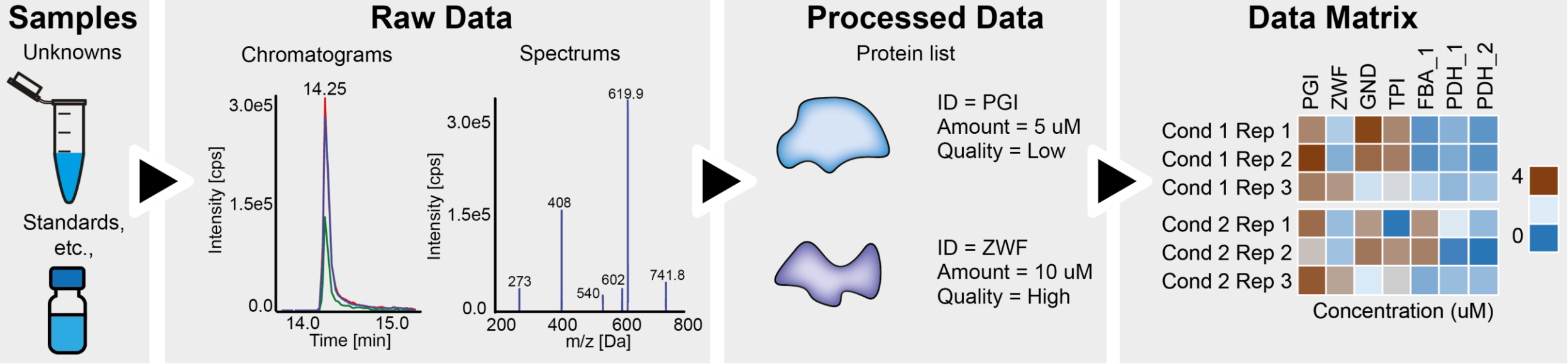


Data processing (3) isotope labeling



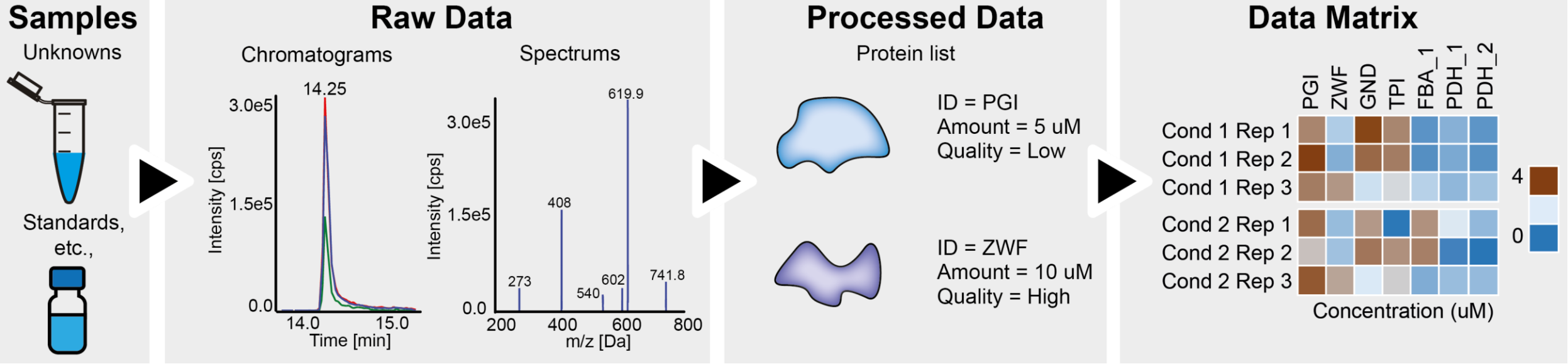
Data processing (4) proteomics

- Chromatograms and spectrum to list of annotated proteins with concentrations/abundances/fold changes



Data processing (4) proteomics

- Chromatograms and spectrum to list of annotated proteins with concentrations/abundances/fold changes

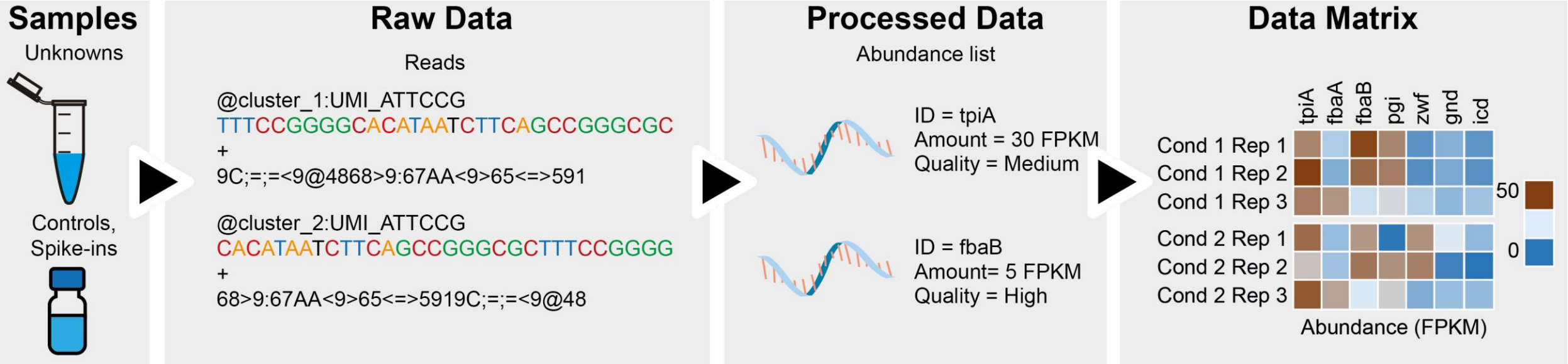


Details:

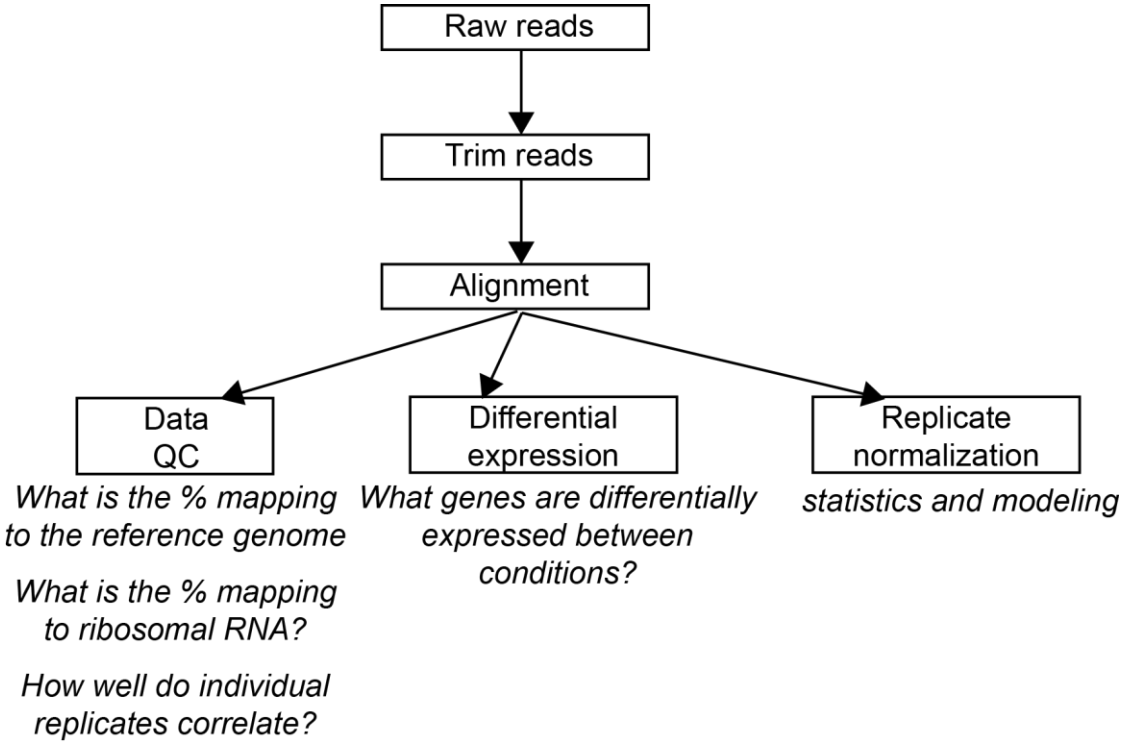
- Same platform used for metabolomics (More to come…)

Data processing (5) transcriptomics

- Reads to list of annotated transcripts with abundances/fold changes

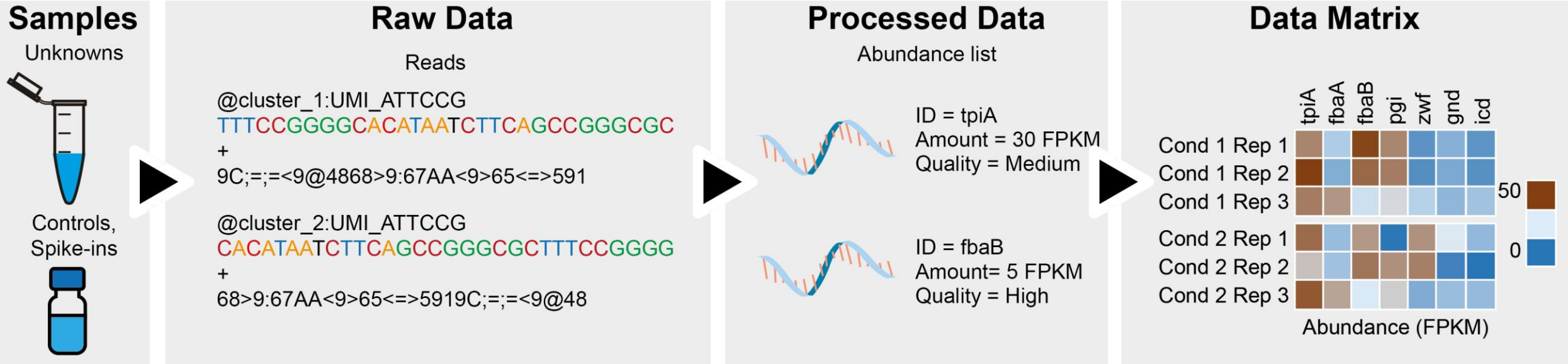


Data processing (5) transcriptomics



Data processing (5) transcriptomics

- Reads to list of annotated transcripts with abundances/fold changes

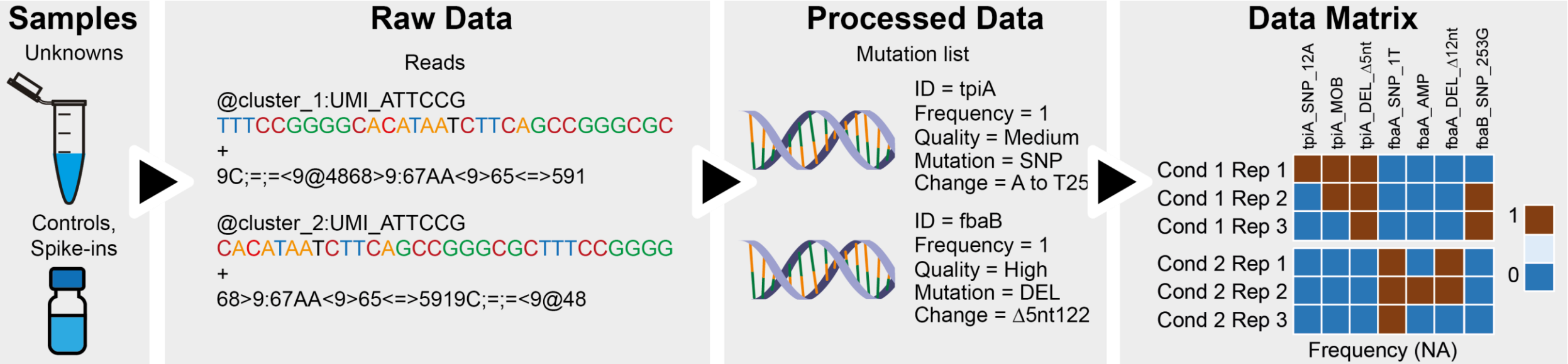


Details:

- Fastx for read trimming
- Bowtie for read alignment
- Samtools for sorting
- CuffLink, CuffDiff, and CuffNorm for differential expression calls and normalization, respectively
- Python/R for statistical and gene set enrichment analysis (details to come…)
- Deployed using Docker containers

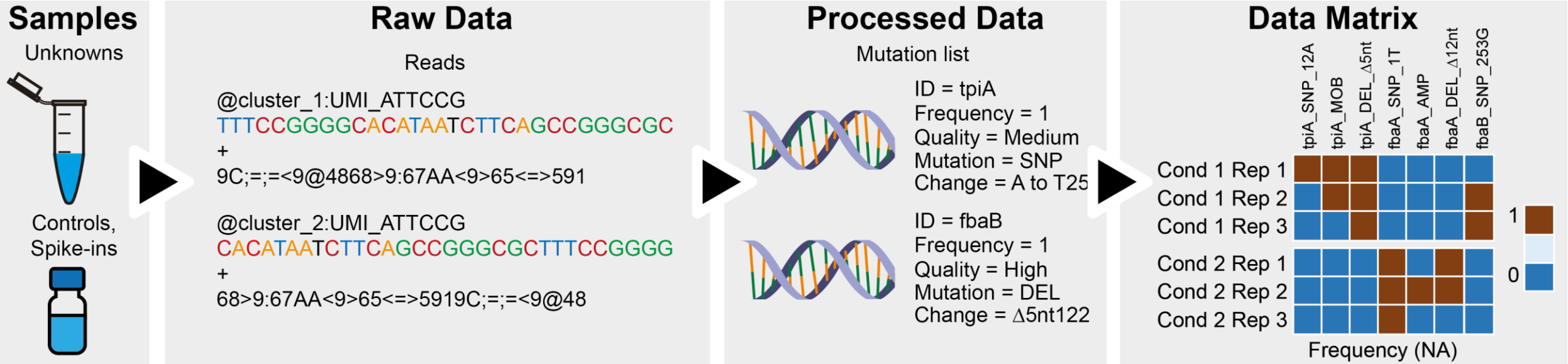
Data processing (6) genomics

- Reads to list of annotated genomic regions with variant calls



Data processing (6) genomics

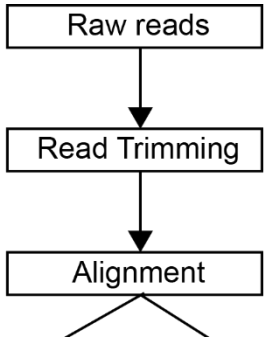
- Reads to list of annotated genomic regions with variant calls

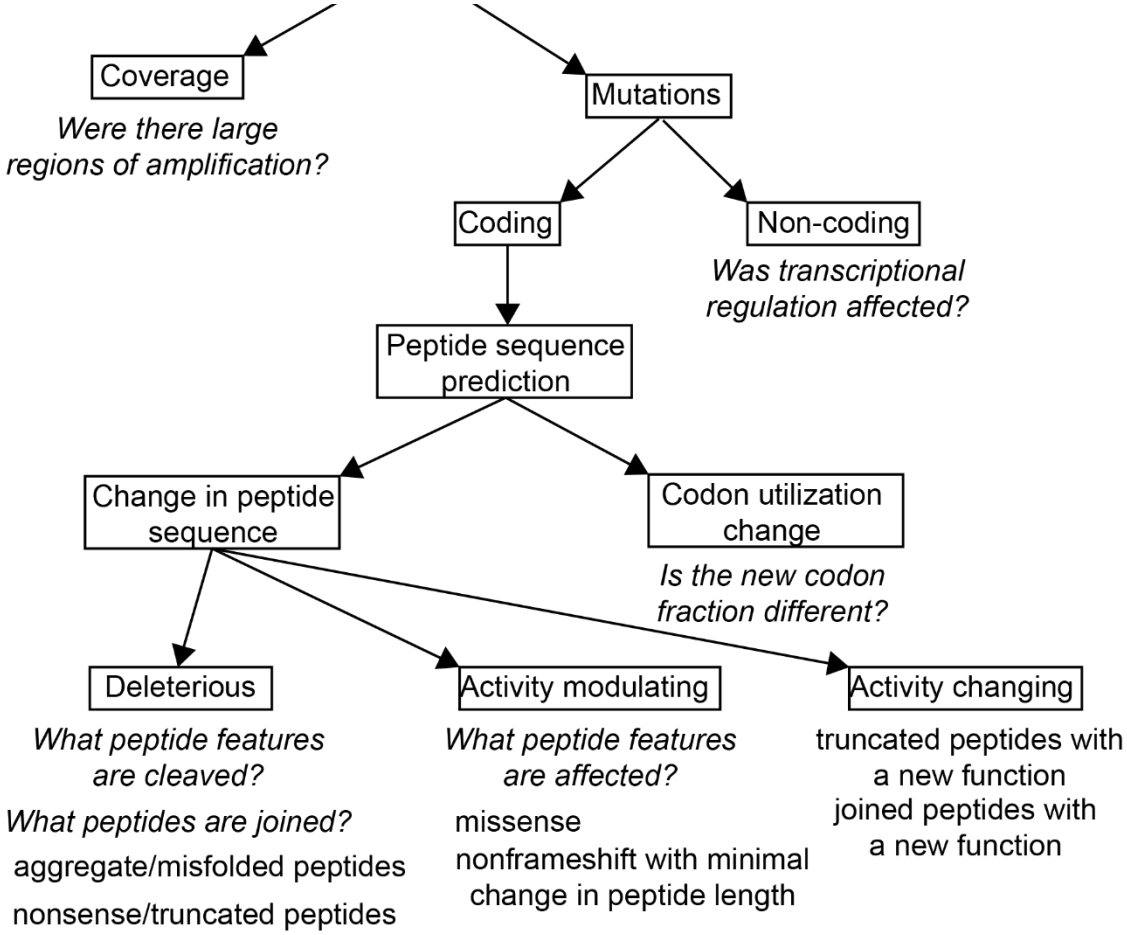


Caveats:

- Not genome assembly, but realignment to an existing genome

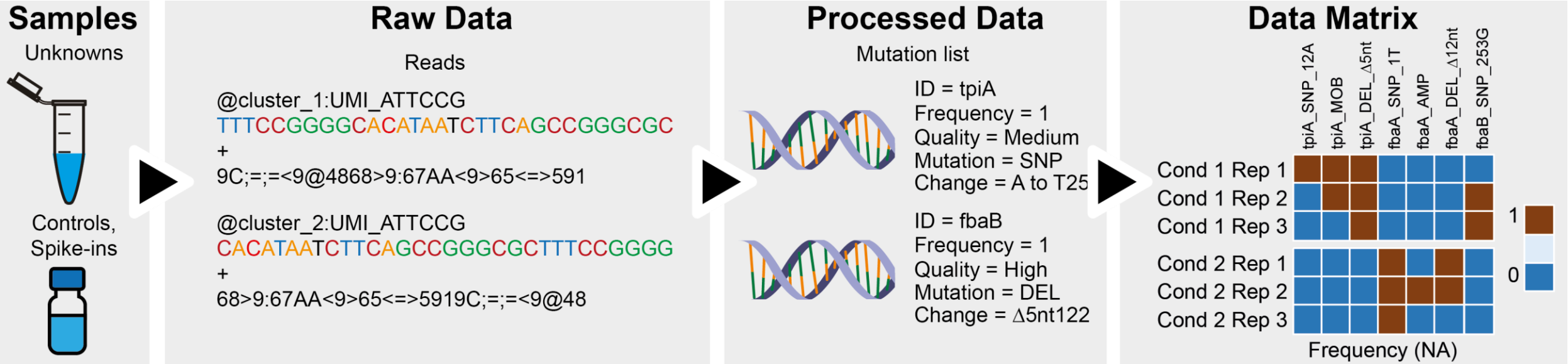
Data processing (6) genomics





Data processing (6) genomics

- Reads to list of annotated genomic regions with variant calls



Details:

- Fastx for read trimming
- Bowtie for read alignment
- Samtools for sorting
- Breseq for variance calling
- Python/R for statistical analyses and Biopython and MVD for structural analyses (details to come…)
- Deployed using Docker containers

Data processing Summary

Goals:

✓Fast

✓Accurate

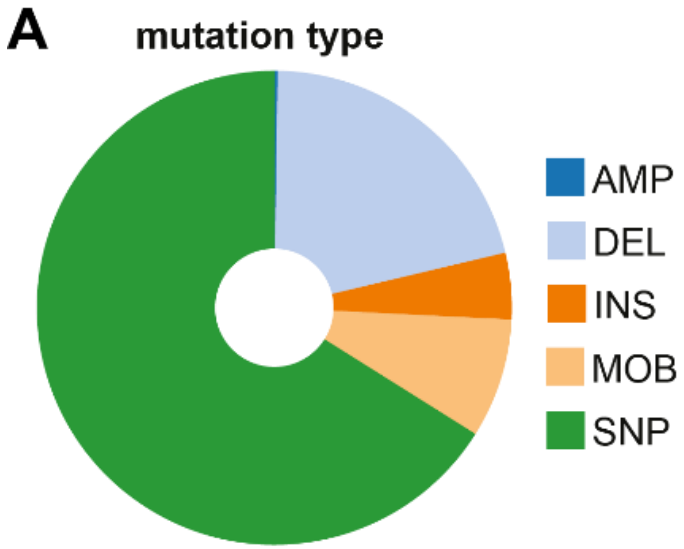
✓Reproducible

✓Push-button

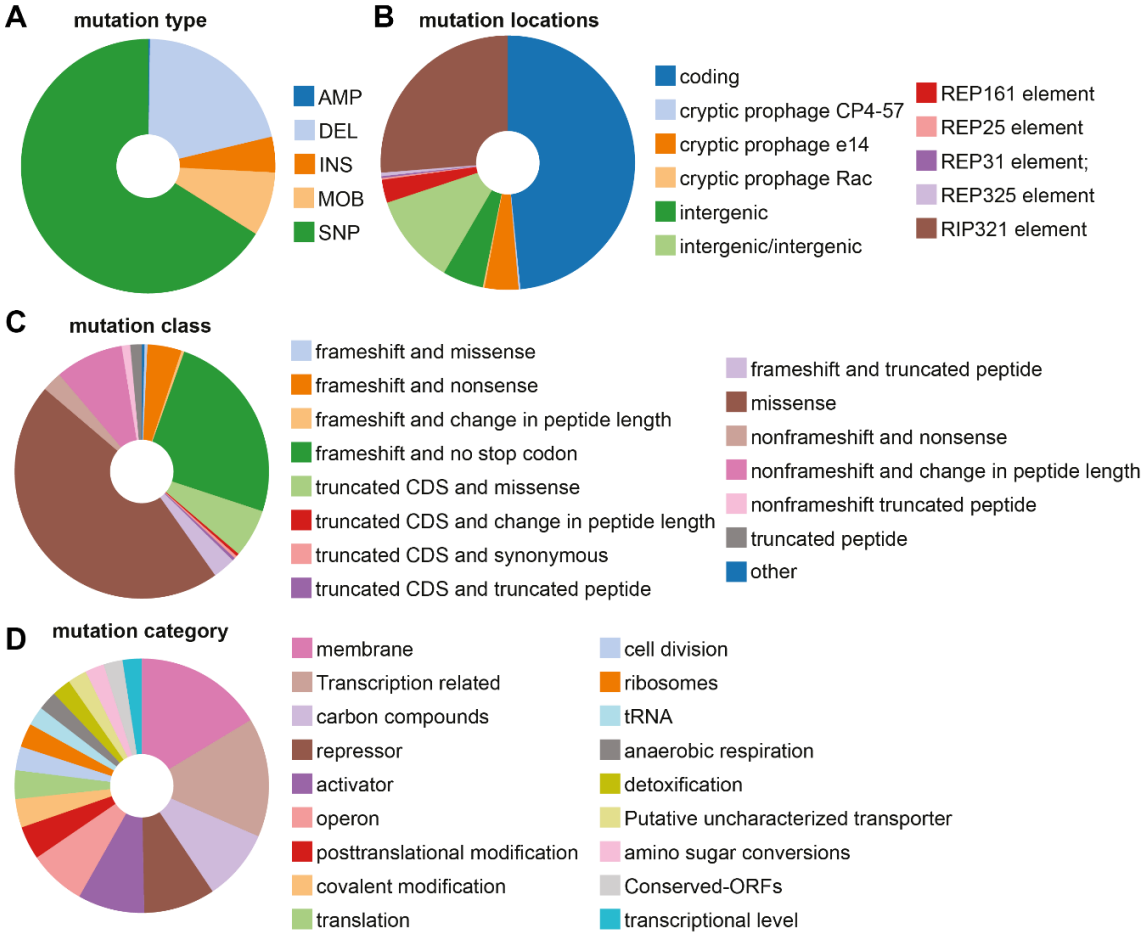
Case Study (1)

Collected **metabolomics, fluxomics, transtriptomics, genomics**, and **phenomics** on Ref, uKO, and eKO strains

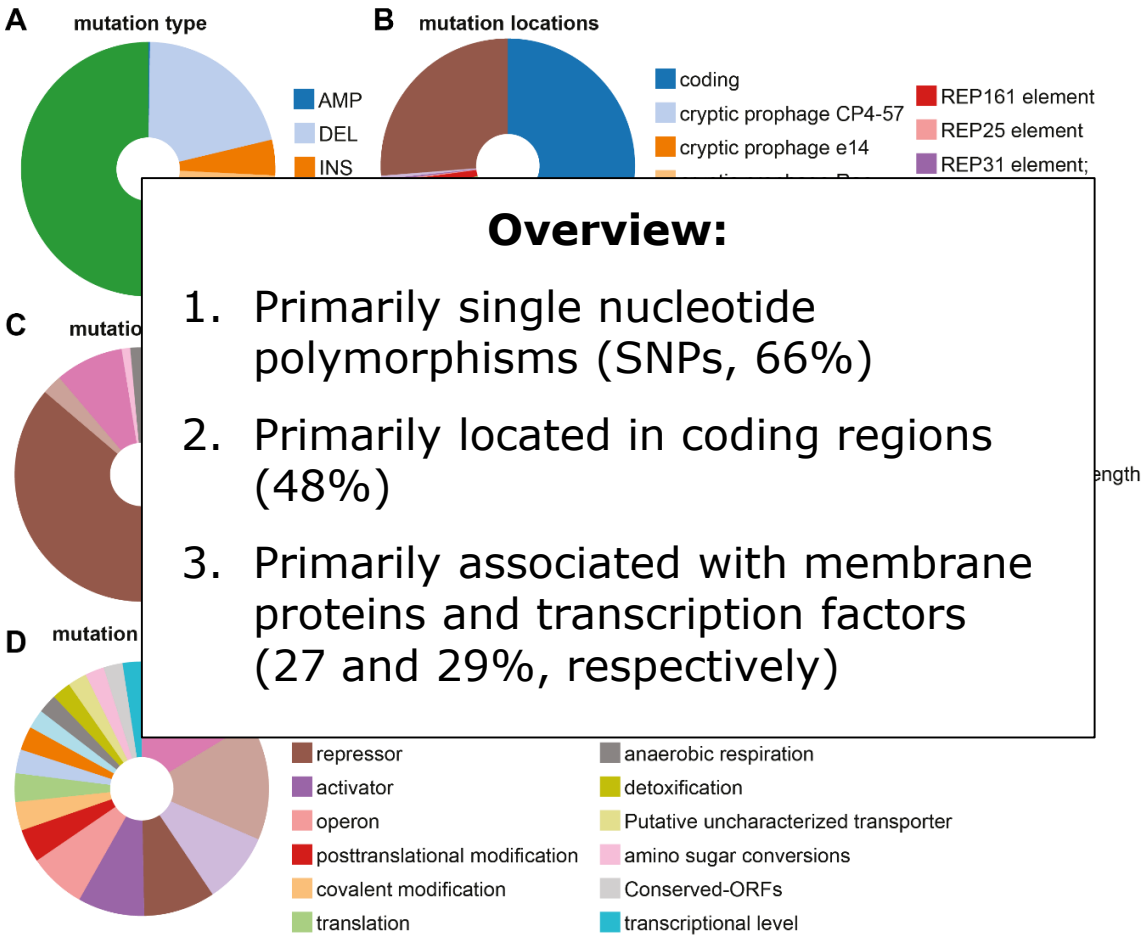
Case Study (1)



Case Study (1)


A
mutation type
AMP
DEL
INS
MOB
SNP
B
mutation locations
coding
cryptic prophage CP4-57
cryptic prophage e14
cryptic prophage Rac
intergenic
intergenic/intergenic
REP161 element
REP25 element
REP31 element;
REP325 element
RIP321 element
C
mutation class
frameshift and missense
frameshift and nonsense
frameshift and change in peptide length
frameshift and no stop codon
truncated CDS and missense
truncated CDS and change in peptide length
truncated CDS and synonymous
truncated CDS and truncated peptide
frameshift and truncated peptide
missense
nonframeshift and nonsense
nonframeshift and change in peptide length
nonframeshift truncated peptide
truncated peptide
other
D
mutation category
membrane
Transcription related
carbon compounds
repressor
activator
operon
posttranslational modification
covalent modification
translation
cell division
ribosomes
tRNA
anaerobic respiration
detoxification
Putative uncharacterized transporter
amino sugar conversions
Conserved-ORFs
transcriptional level

Case Study (1)



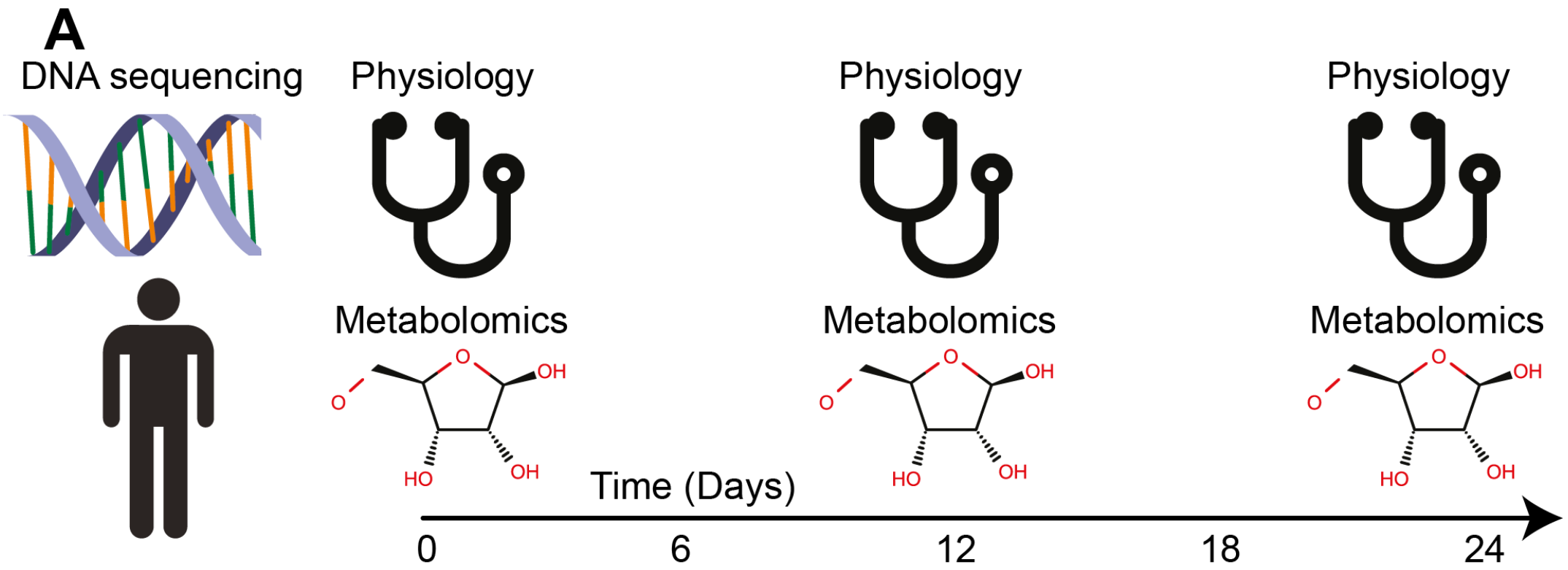
Overview:

1. Primarily single nucleotide polymorphisms (SNPs, 66%)
2. Primarily located in coding regions (48%)
3. Primarily associated with membrane proteins and transcription factors (27 and 29%, respectively)

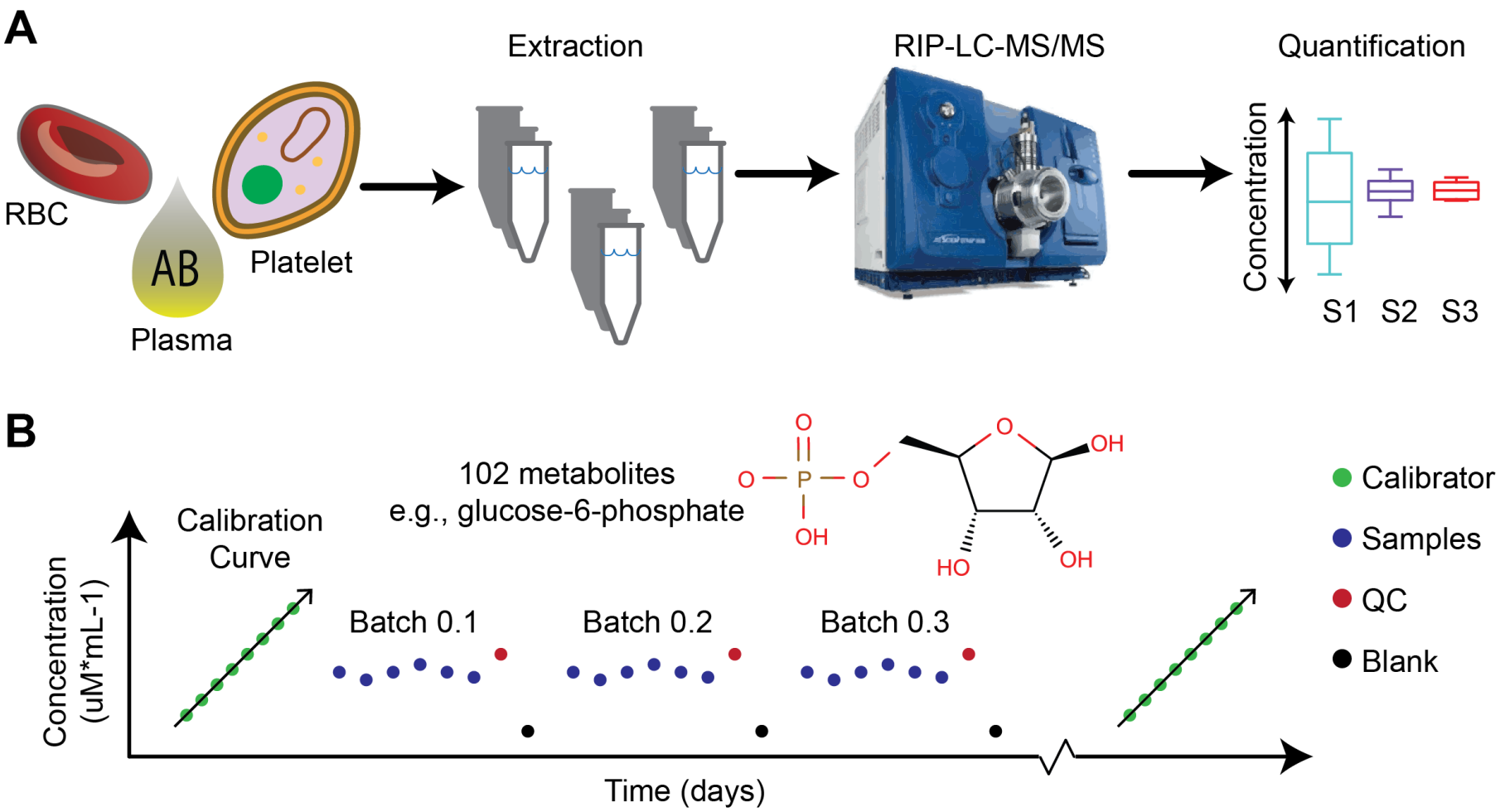
Case Study (2)

Collected **metabolomics** and **genomics**
30 healthy volunteers

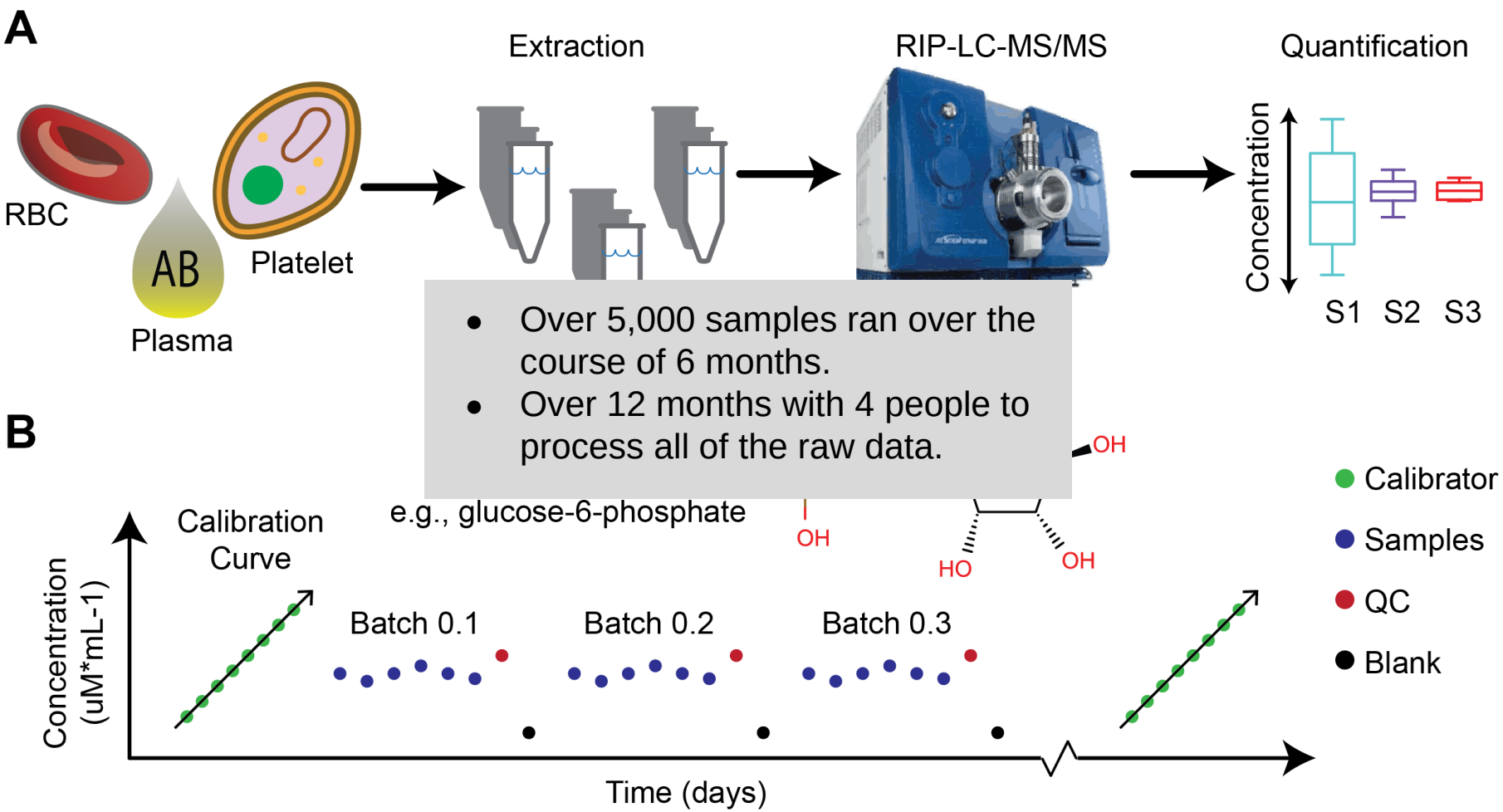
Case Study (2)


A
DNA sequencing
Physiology
Metabolomics
Physiology
Metabolomics
Physiology
Metabolomics
O
OH
HO
OH
Time (Days)
0
6
12
18
24

Case Study (2)

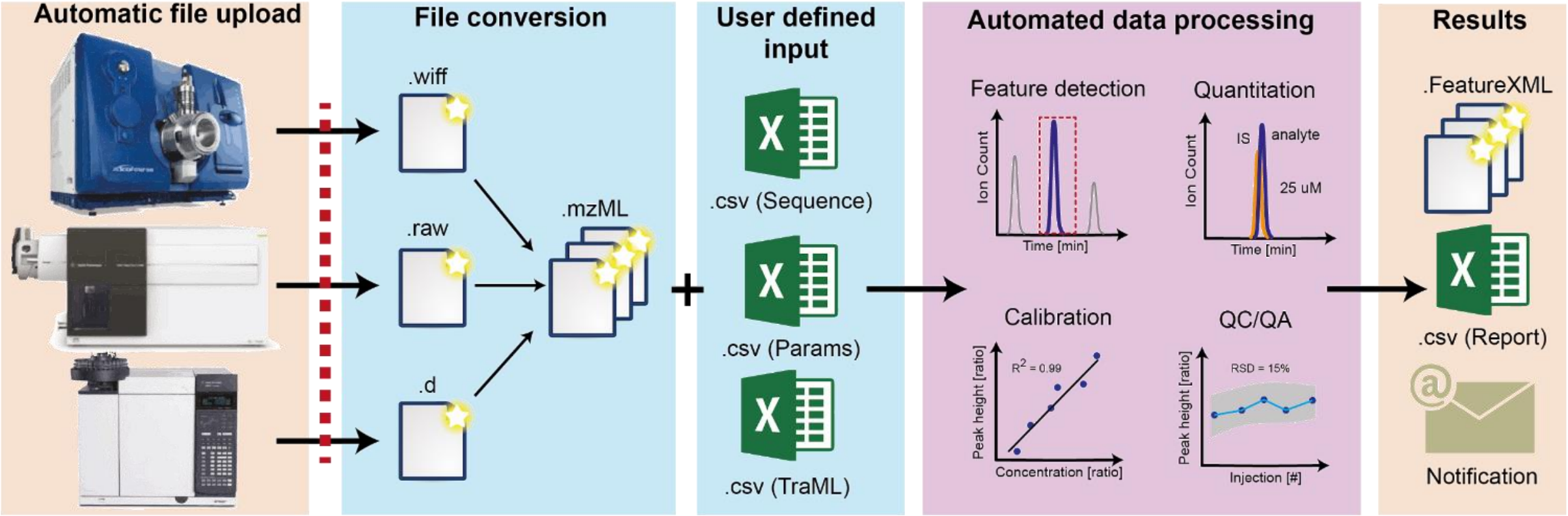

A
RBC
AB
Plasma
Platelet
Extraction
RIP-LC-MS/MS
Quantification
Concentration
S1
S2
S3
B
102 metabolites
e.g., glucose-6-phosphate
O
P
OH
HO
Calibration
Curve
Concentration
(uM*mL-1)
Batch 0.1
Batch 0.2
Batch 0.3
Time (days)
Calibrator
Samples
QC
Blank

Case Study (2)



- Over 5,000 samples ran over the course of 6 months.
- Over 12 months with 4 people to process all of the raw data.

Case Study (2)



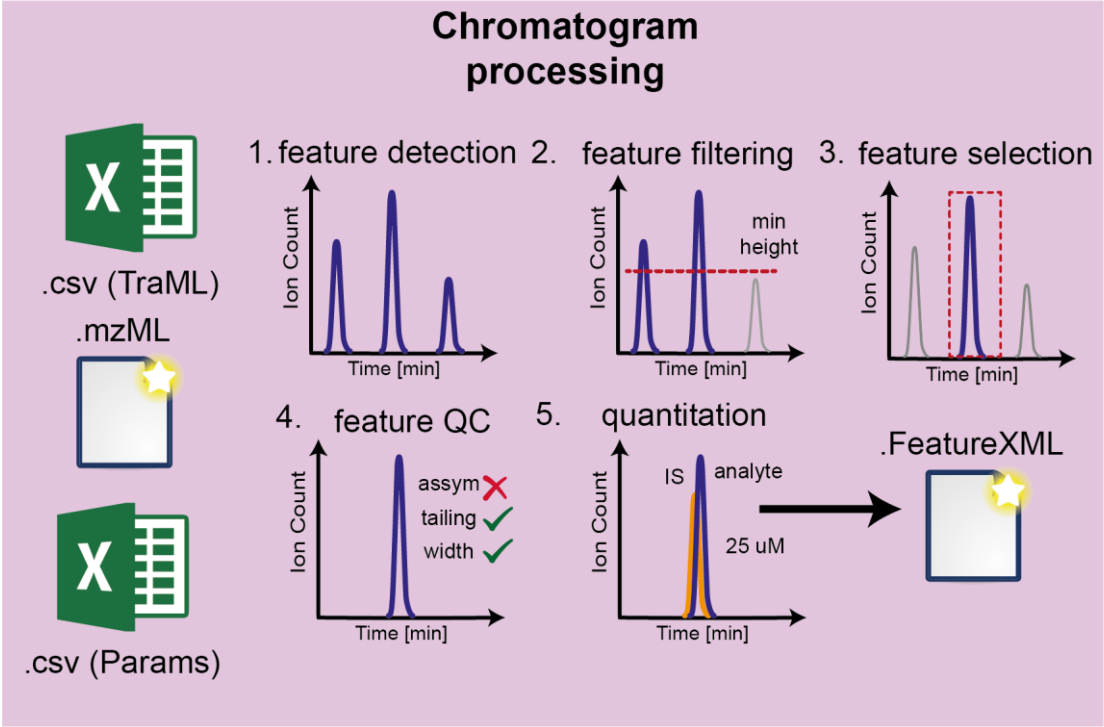
Fast, automated, and intelligent processing of GC- and LC-MS data, and HPLC data.

Data processing (1/2) endo/exo-metabolomics

Targeted (semi-quantitative to quantitative) and untargeted metabolomics pipeline

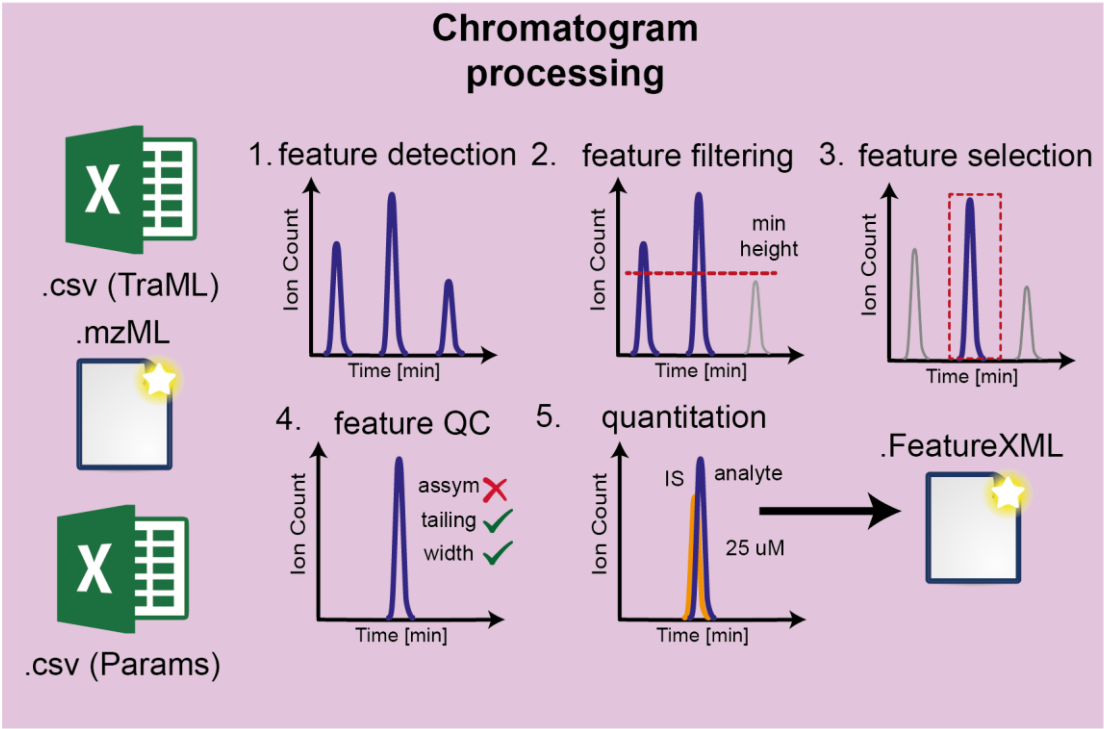
Data processing (1/2) endo/exo-metabolomics

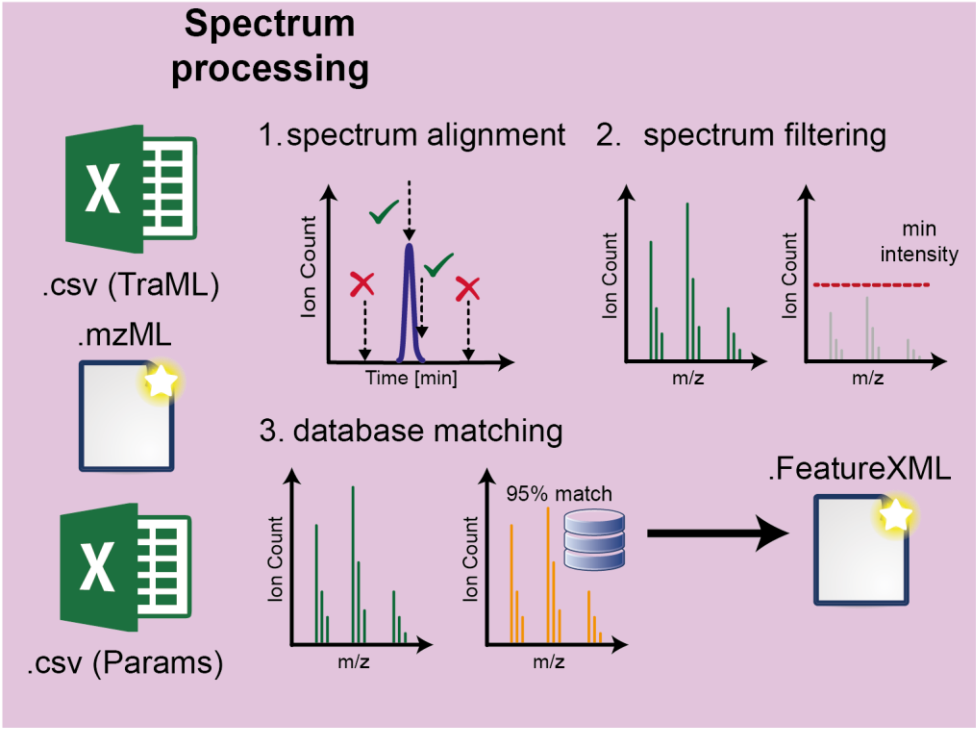
Targeted (semi-quantitative to quantitative) and untargeted metabolomics pipeline



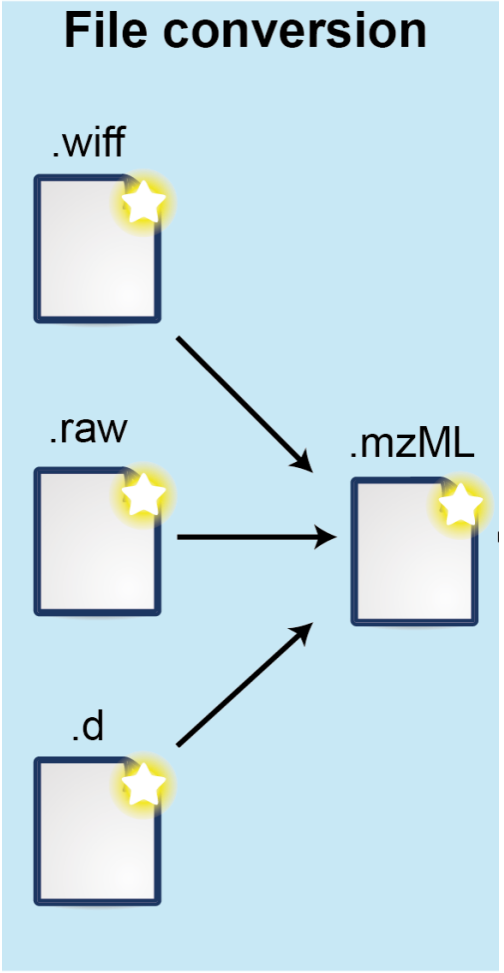
Data processing (1/2) endo/exo-metabolomics

Targeted (semi-quantitative to quantitative) and **untargeted** metabolomics pipeline





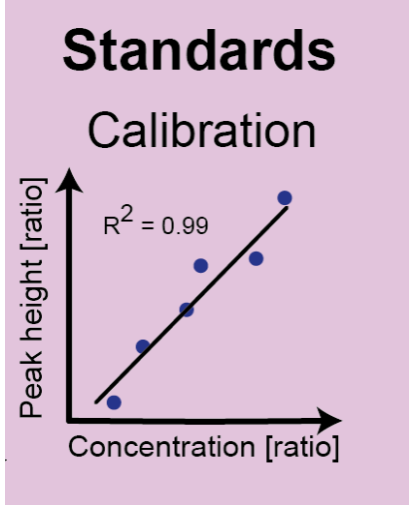
Features: File conversion

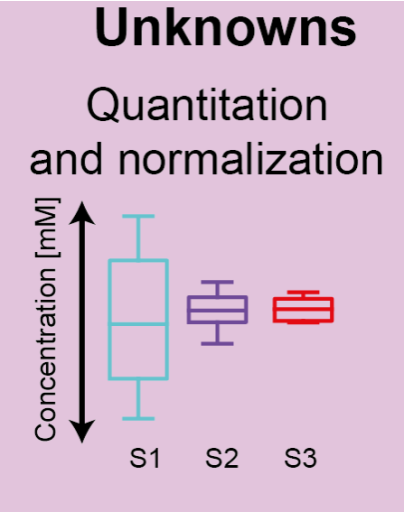


- ProteoWizard:
  - Supports open-source and most proprietary MS formats
  - Complete conversion of SRM, MRM, DIA (e.g., SWATH), DDA, and Full Scan (precursor and/or product)
- Custom:
  - HPLC text file conversion

McCloskey D, Collianni P, et al. in progress.

Features: Quantitation

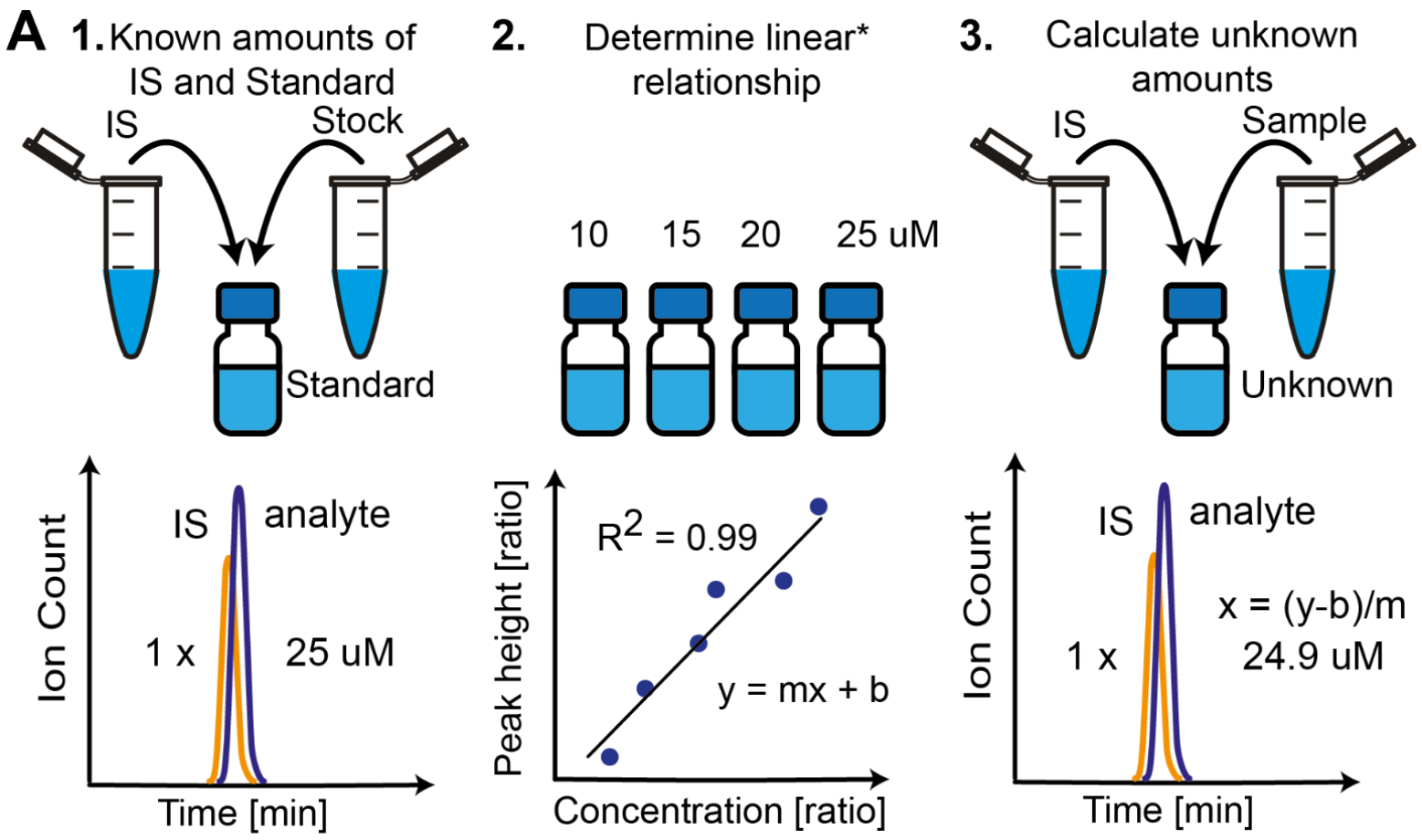




- Automated calibration curve fitting
  - Outlier detection with bias and R-squared optimization
  - User defined calibration model and acceptance criteria
- External calibration curve application
  - Back calculation of sample concentration with or without IS

Features: Quantitation

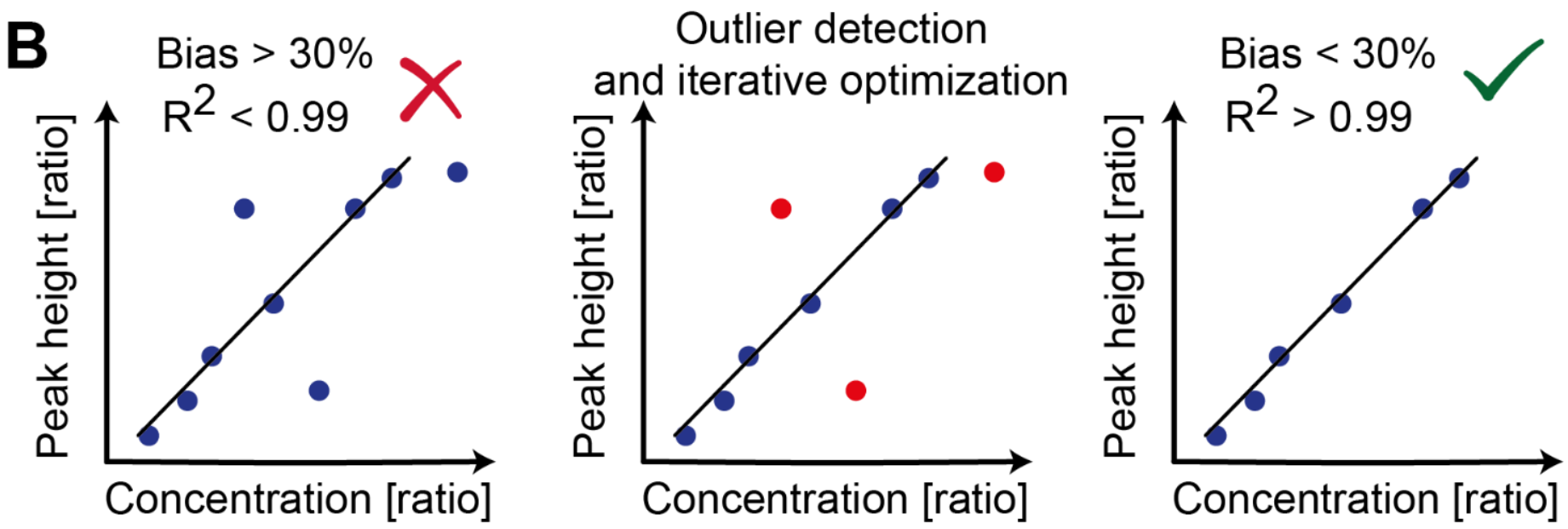
Isotope dilution mass spectrometry (IDMS)



* linear model with x and y weighting currently supported

Features: Quantitation

Automated calibration curve optimization



McCloskey D, Collianni P, et al. in progress.

Features: Quantitation

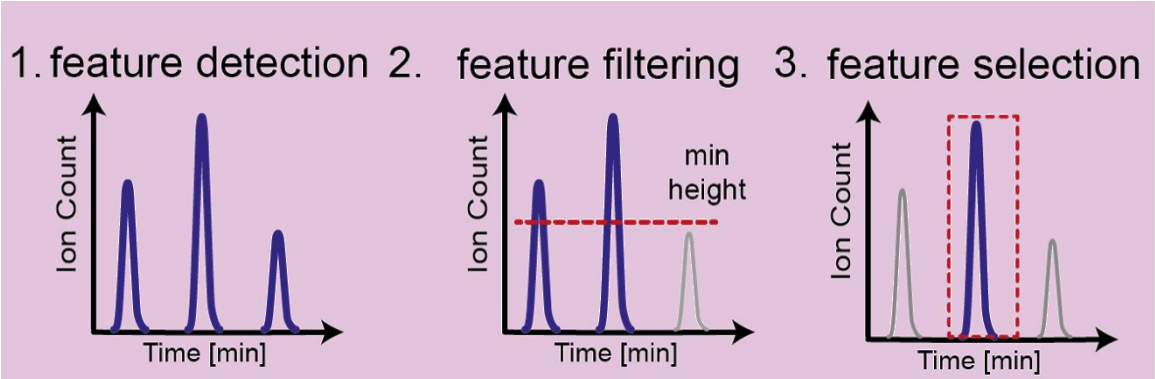
Automated calibration curve optimization

	Manual	Automated
Acceptable fit	0.97	0.97
Time (sec)	28,800	< 1.0

Table 2: Calibration curves for 116 compounds. Bias >= 30%, R >= 0.9, and minimum of 4 points.

McCloskey D, Collianni P, et al. in progress.

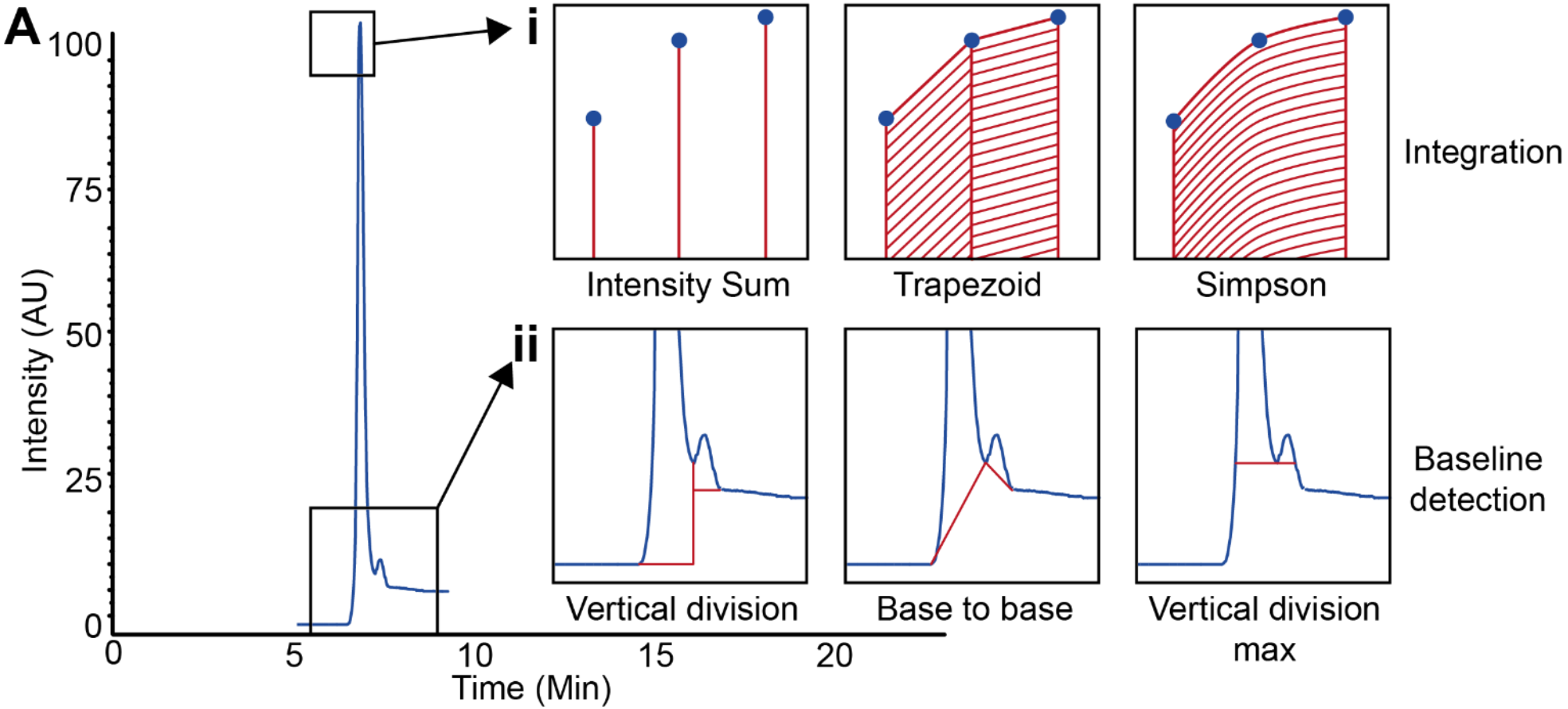
Features: Peak integration and selection



- Automated peak picking and peak selection
  - Advanced peak integration and baseline detection models
  - User defined thresholds for peak filtering
  - Multi-stage, adaptive optimization algorithm for peak selection

McCloskey D, Collianni P, et al. in progress.

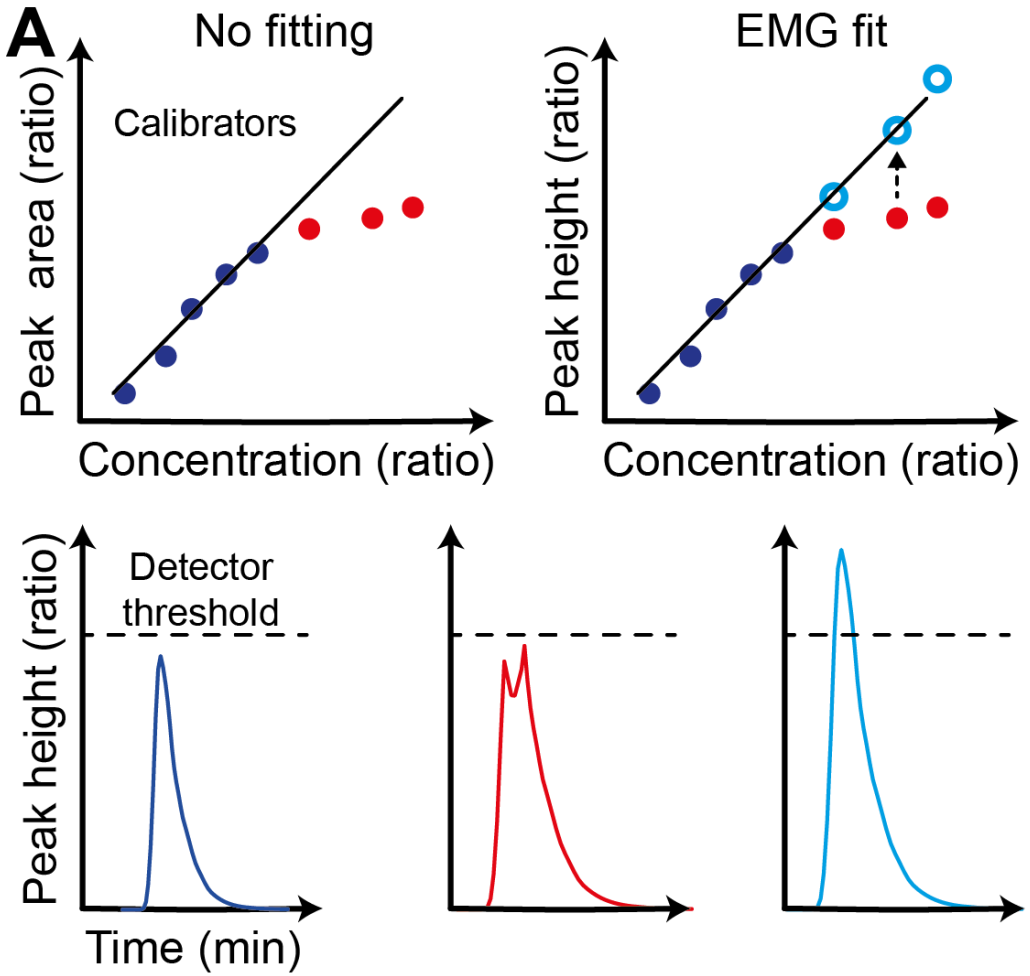
Features: Peak integration and selection



McCloskey D, Collianni P, et al. in progress.

Features: Peak integration and selection

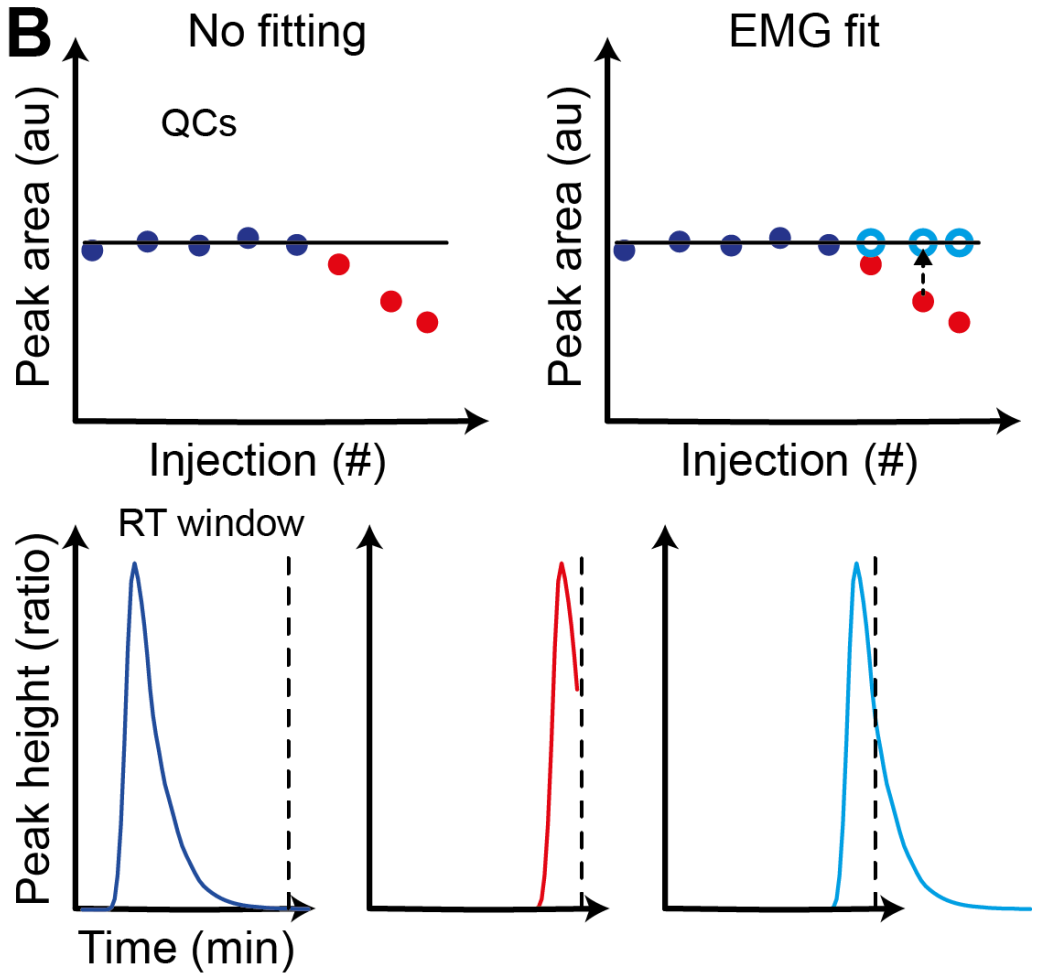
Saturated peak



McCloskey D, Collianni P, et al. in progress.

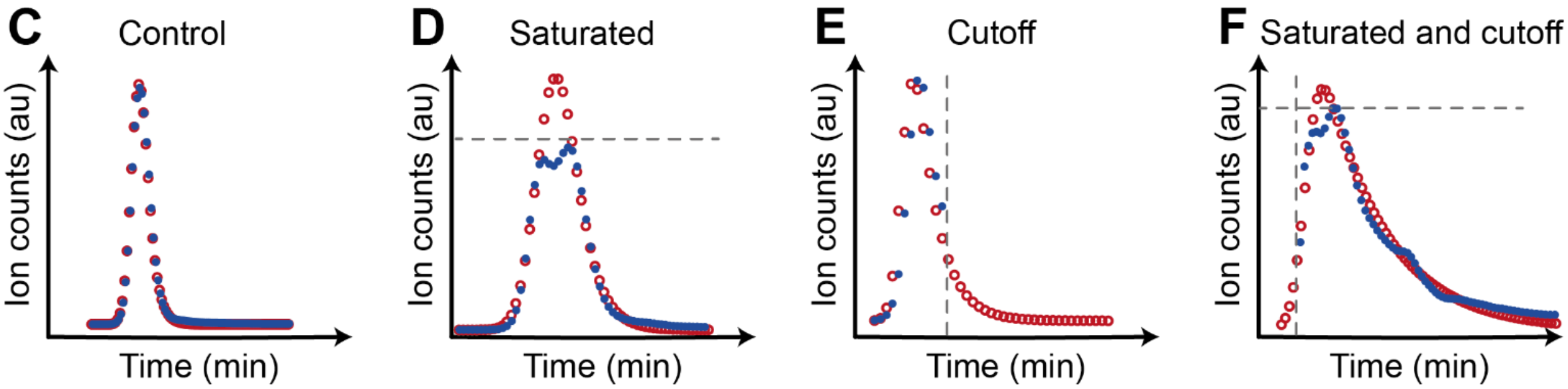
Features: Peak integration and selection

Cutoff peak



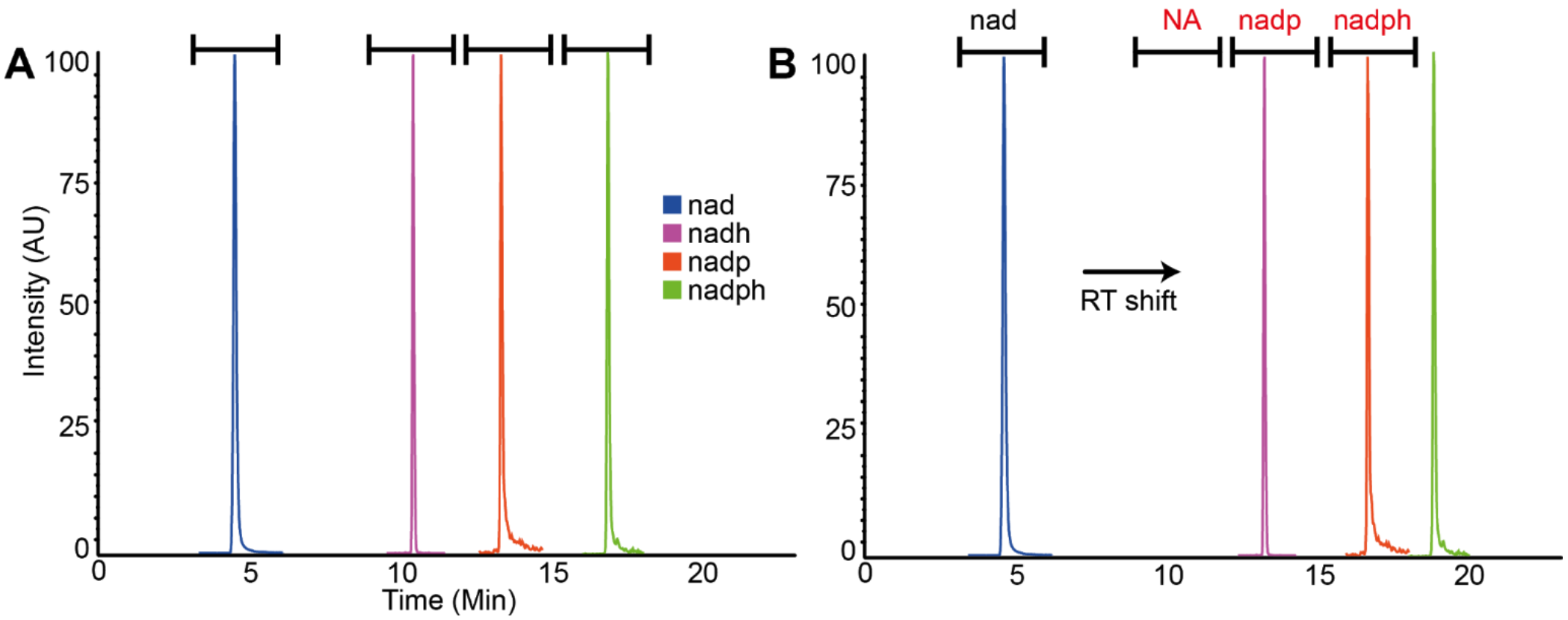
McCloskey D, Collianni P, et al. in progress.

Features: Peak integration and selection



McCloskey D, Collianni P, et al. in progress.

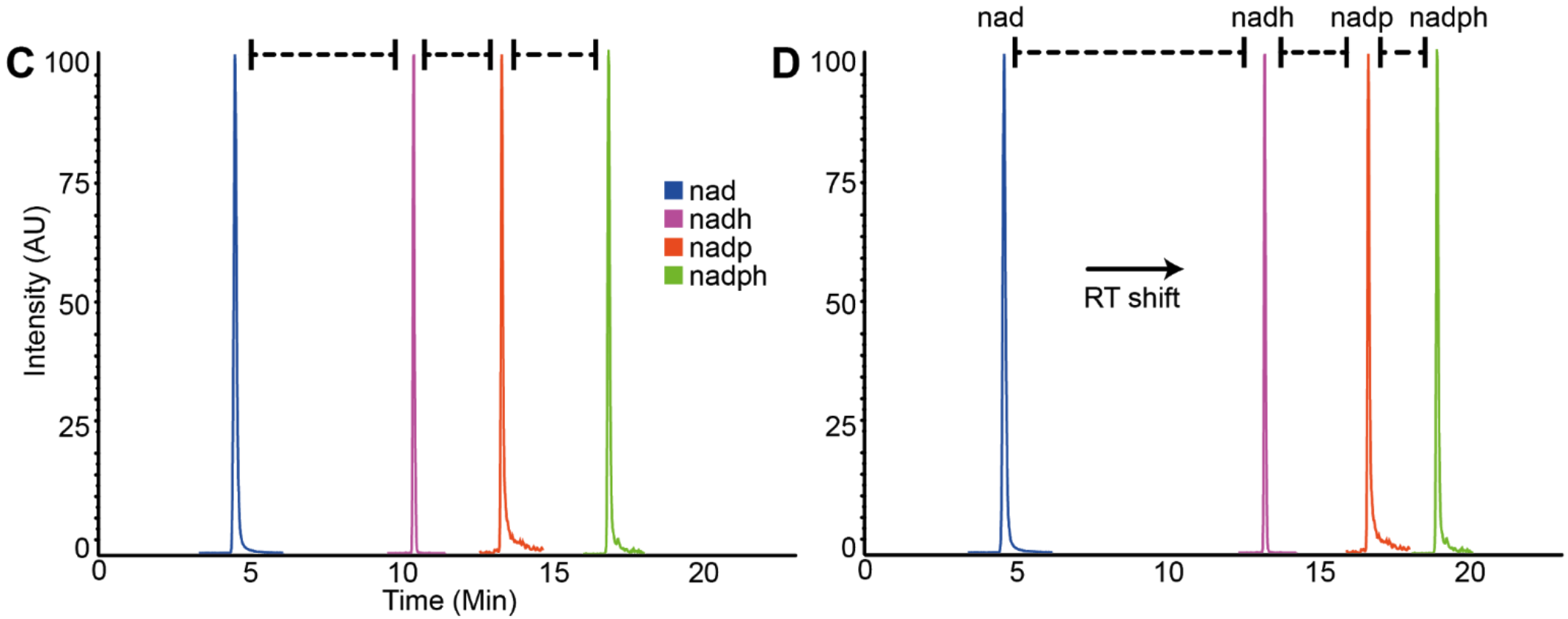
Features: Peak integration and selection



Absolute retention is susceptible to retention time shifts

LP to select the best peak based on signal intensity and absolute retention time

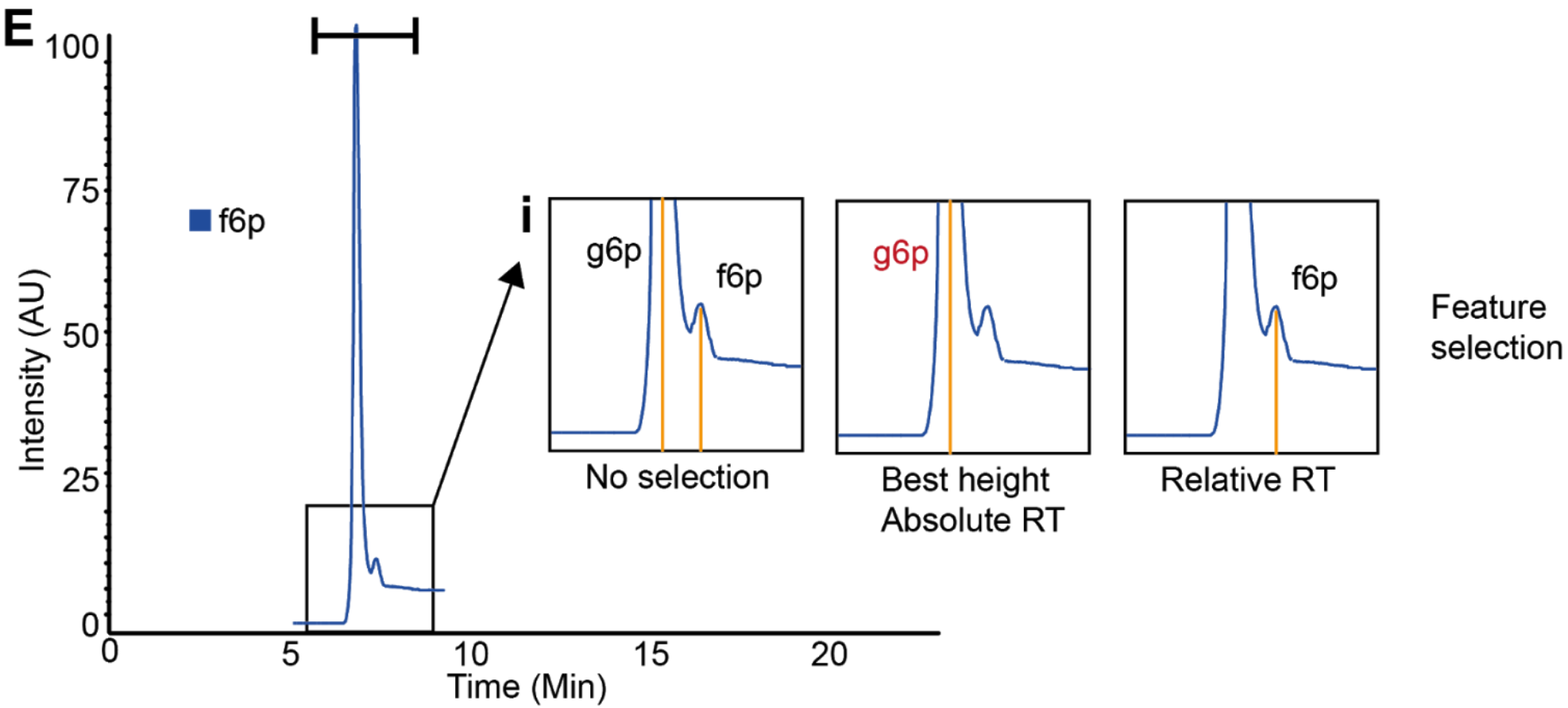
Features: Peak integration and selection



Retention difference is more robust to retention time shifts

QIP algorithm to "select" the most consistent set of peaks based on retention time difference

Features: Peak integration and selection



McCloskey D, Collianni P, et al. in progress.

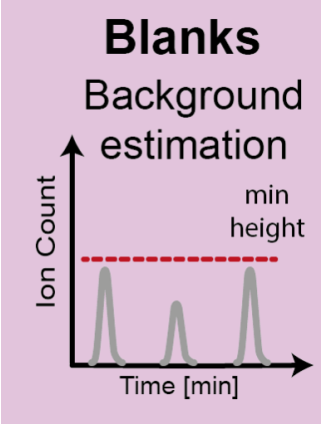
Features: Peak integration and selection

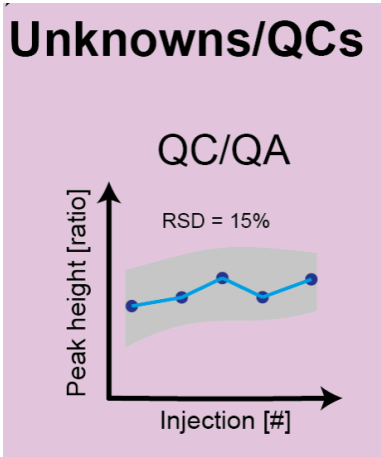
	Absolute RT	Relative RT
Average	0.87	0.95
StDev	0.07	0.03

Table 1: Peak Picking and Selection validation results. Tested on Human RBC, Plasma, and Platelet metabolomics (n=1297)

McCloskey D, Collianni P, et al. in progress.

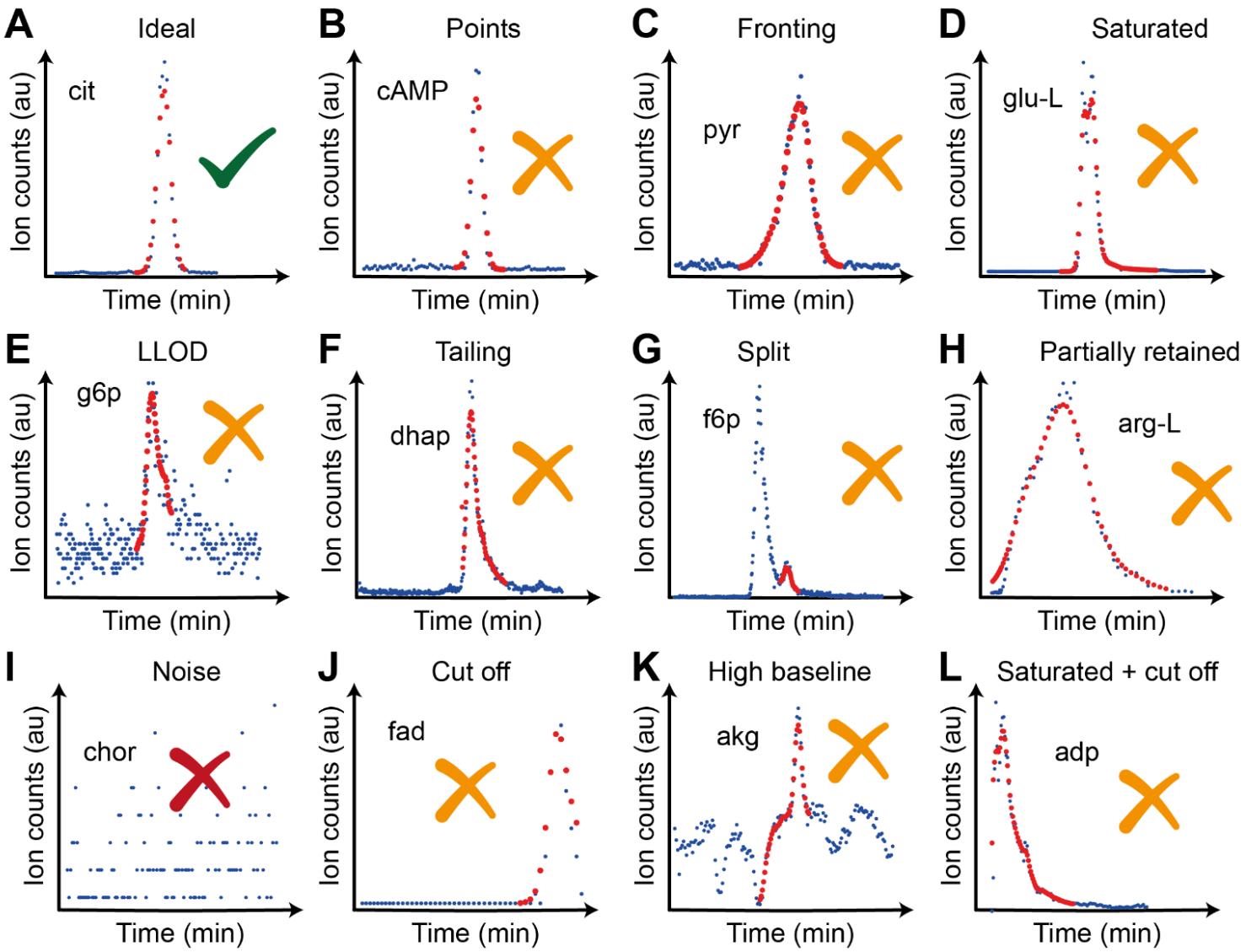
Features: Quality Control





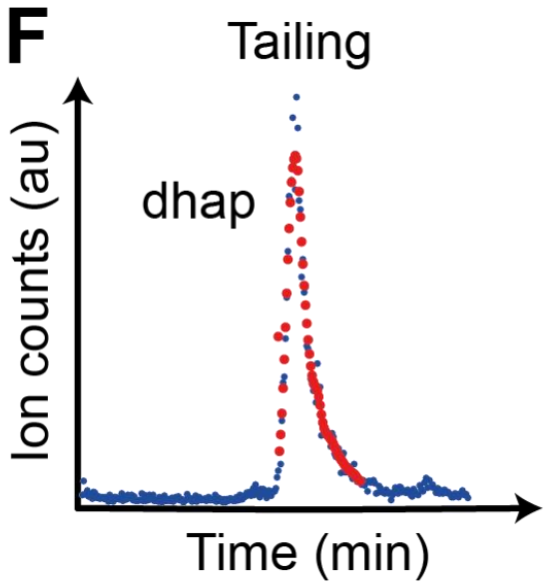
- Flexible, user-defined peak, run, or batch quality control reporting

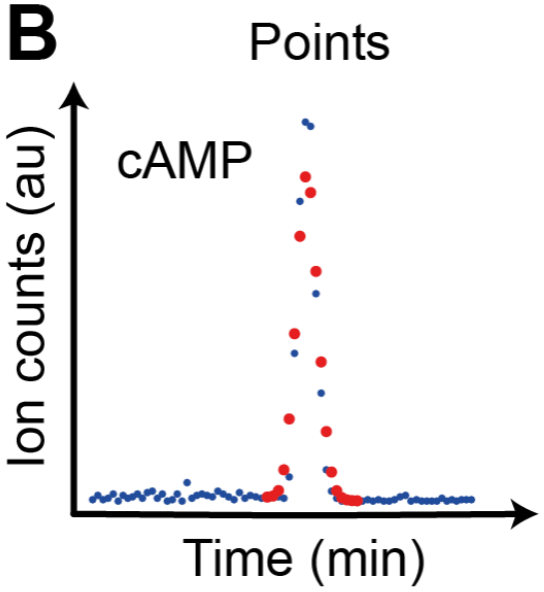
Feature: Quality Control



Feature: Quality Control

- Transition/peak level scoring system
  - Shape QC:
    - Total peak width, baseline delta to height
    - Tailing and asymmetry factors
  - Quality QC:
    - Retention time, intensity, quality, etc.,
    - Points across full width and half width
  - Quantitation QC:
    - LLOQ/ULOQ





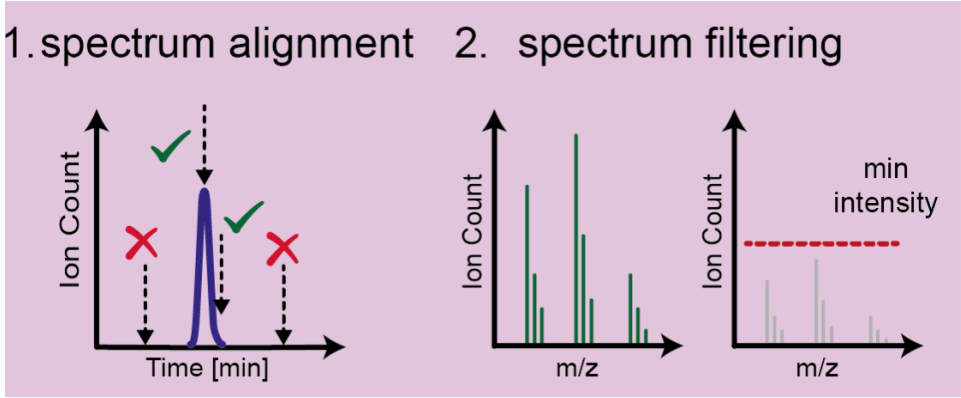
McCloskey D, Collianni P, et al. in progress.

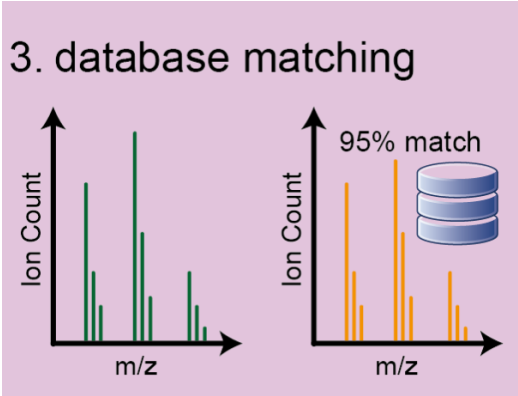
Feature: Quality Control

- Transition group level scoring system
  - Compound ID QC:
    - Ion ratios
  - # of transitions QC:
    - # of Light and Heavy transitions
    - # of quantifying and detection transitions
  - Quality QC:
    - Retention time, total intensity, total quality, etc.

McCloskey D, Collianni P, et al. in progress.

Features: DDA/DIA





McCloskey D, Collianni P, et al. in progress.

- Spectrum annotation
  - Spectrum filtering based on RT
  - Spectrum scoring based on resolution and quality
  - Spectrum annotation based on DB matching score
- DB matching
  - Internal or external (.msp format)
  - Contrast angle similarity score

Automated analytics data processing

- ✓ A single data processing solution for metabolomics, fluxomics, and proteomics applications by GC-MS(/MS), LC-MS(/MS), and HPLC-UV, RI, etc.

Automated analytics data processing

- ✓ A single data processing solution for metabolomics, fluxomics, and proteomics applications by GC-MS(/MS), LC-MS(/MS), and HPLC-UV, RI, etc.
- ✓ Improved accuracy and reduced of peak integration and peak selection

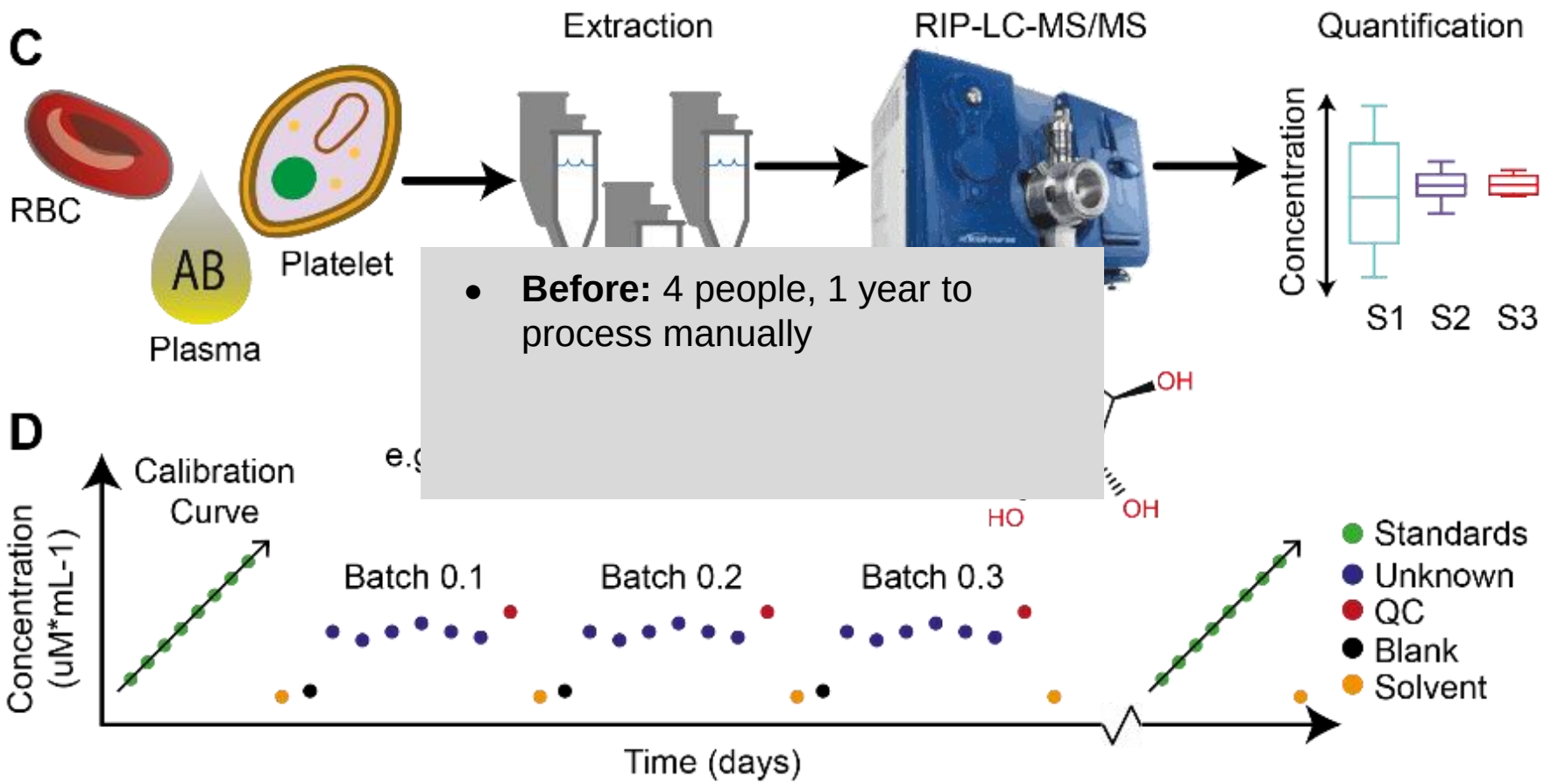
Automated analytics data processing

- ✓ A single data processing solution for metabolomics, fluxomics, and proteomics applications by GC-MS(/MS), LC-MS(/MS), and HPLC-UV, RI, etc.
- ✓ Improved accuracy and reduced of peak integration and peak selection
- ✓ Integration with open-source tools for seamless integration with bioinformatics pipelines

Automated analytics data processing

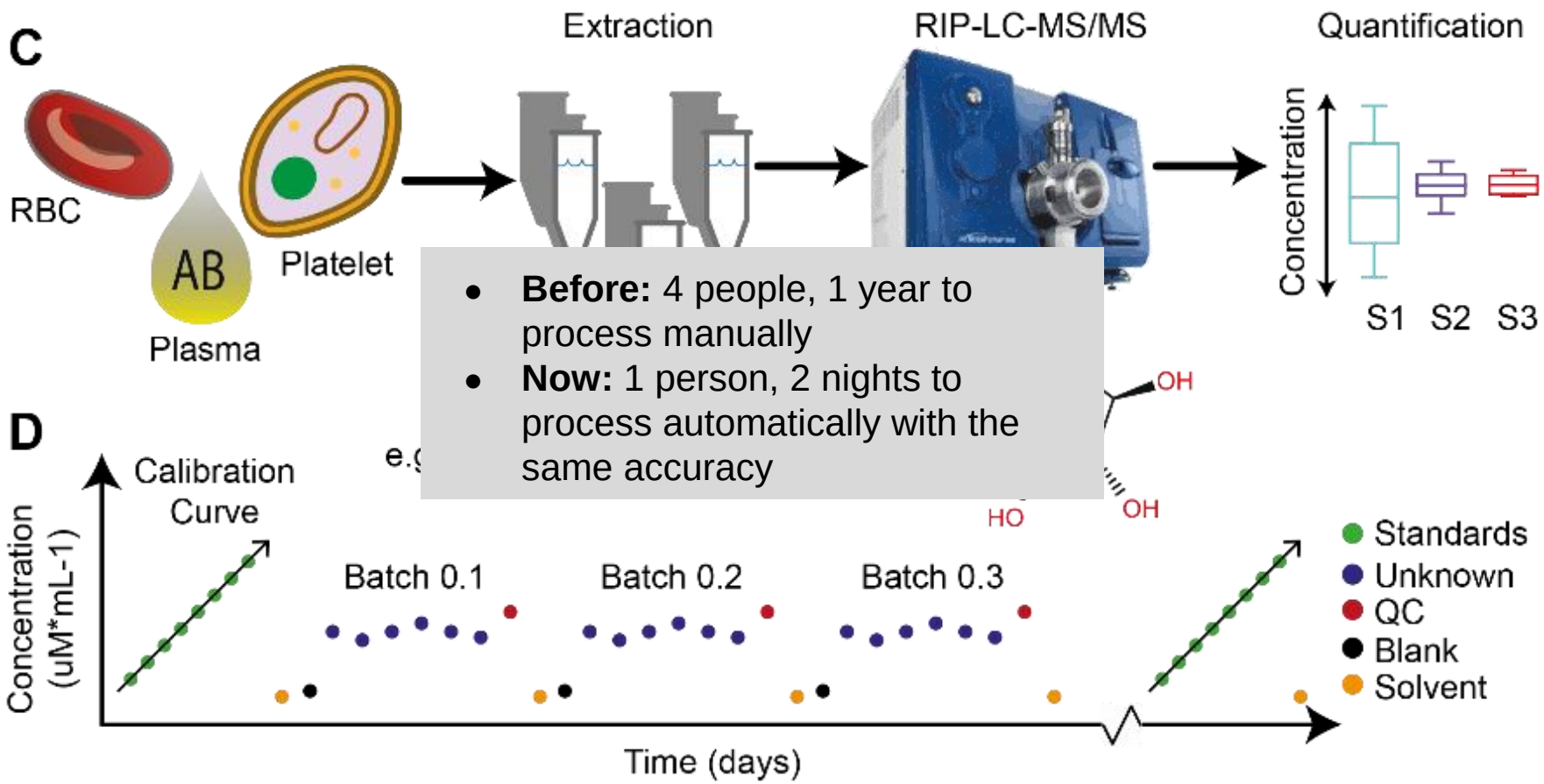
- ✓ A single data processing solution for metabolomics, fluxomics, and proteomics applications by GC-MS(/MS), LC-MS(/MS), and HPLC-UV, RI, etc.
- ✓ Improved accuracy and reduced of peak integration and peak selection
- ✓ Integration with open-source tools for seamless integration with bioinformatics pipelines
- ✓ Reduced time spent in lengthy and problematic peak integration and peak review sessions

Automated analytics data processing



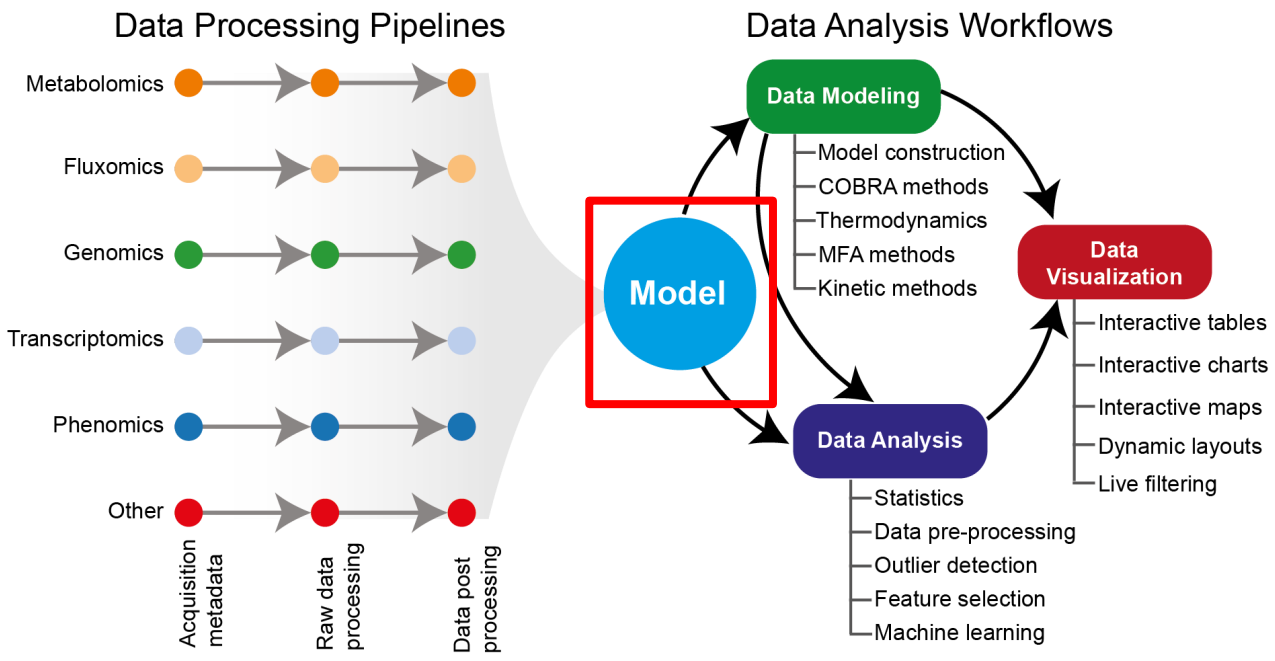
- **Before:** 4 people, 1 year to process manually

Automated analytics data processing



- **Before:** 4 people, 1 year to process manually
- **Now:** 1 person, 2 nights to process automatically with the same accuracy

Biochemical model mapping



Biochemical model mapping

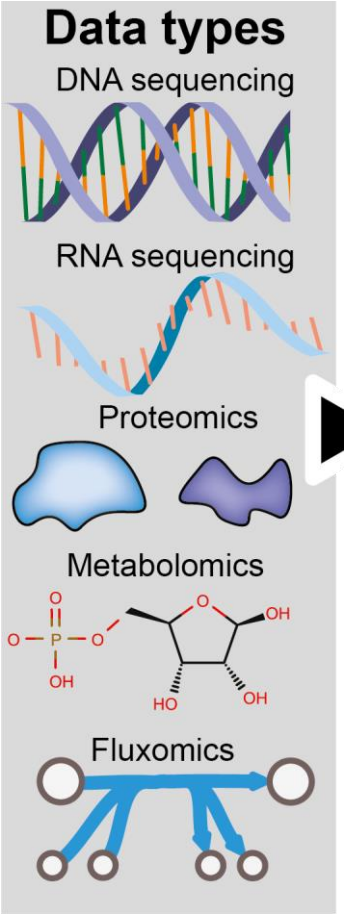
- How do I connect all –omics data features to one another?
- Resolution of the component connections
  - Organ to tissue to cell to organelle

Case Study (1)

Whole cell *E. coli* biochemical interaction network

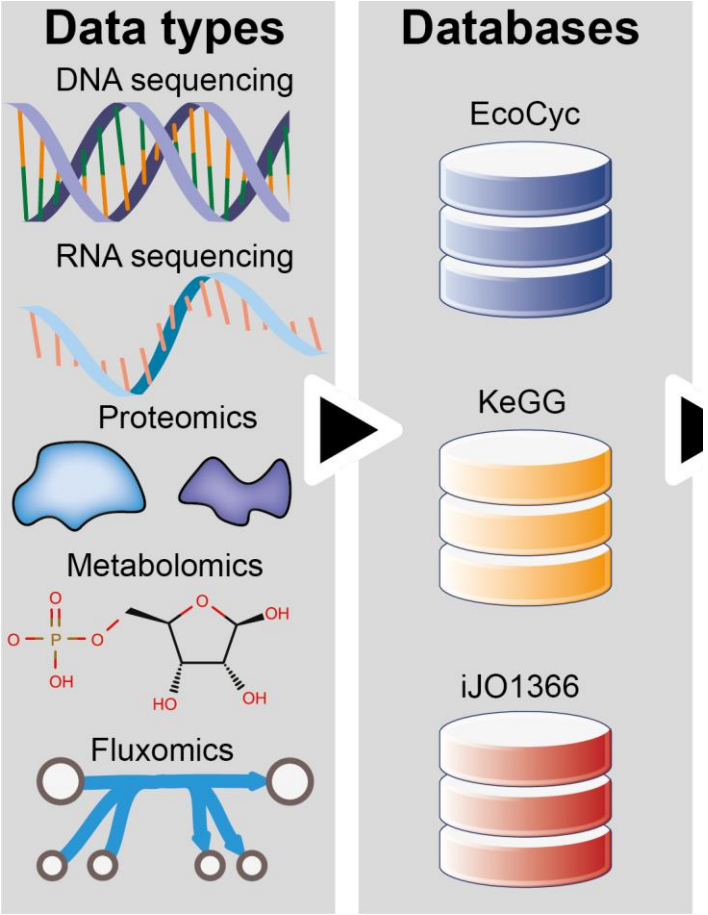
Case Study (1)

Whole cell *E. coli* biochemical interaction network



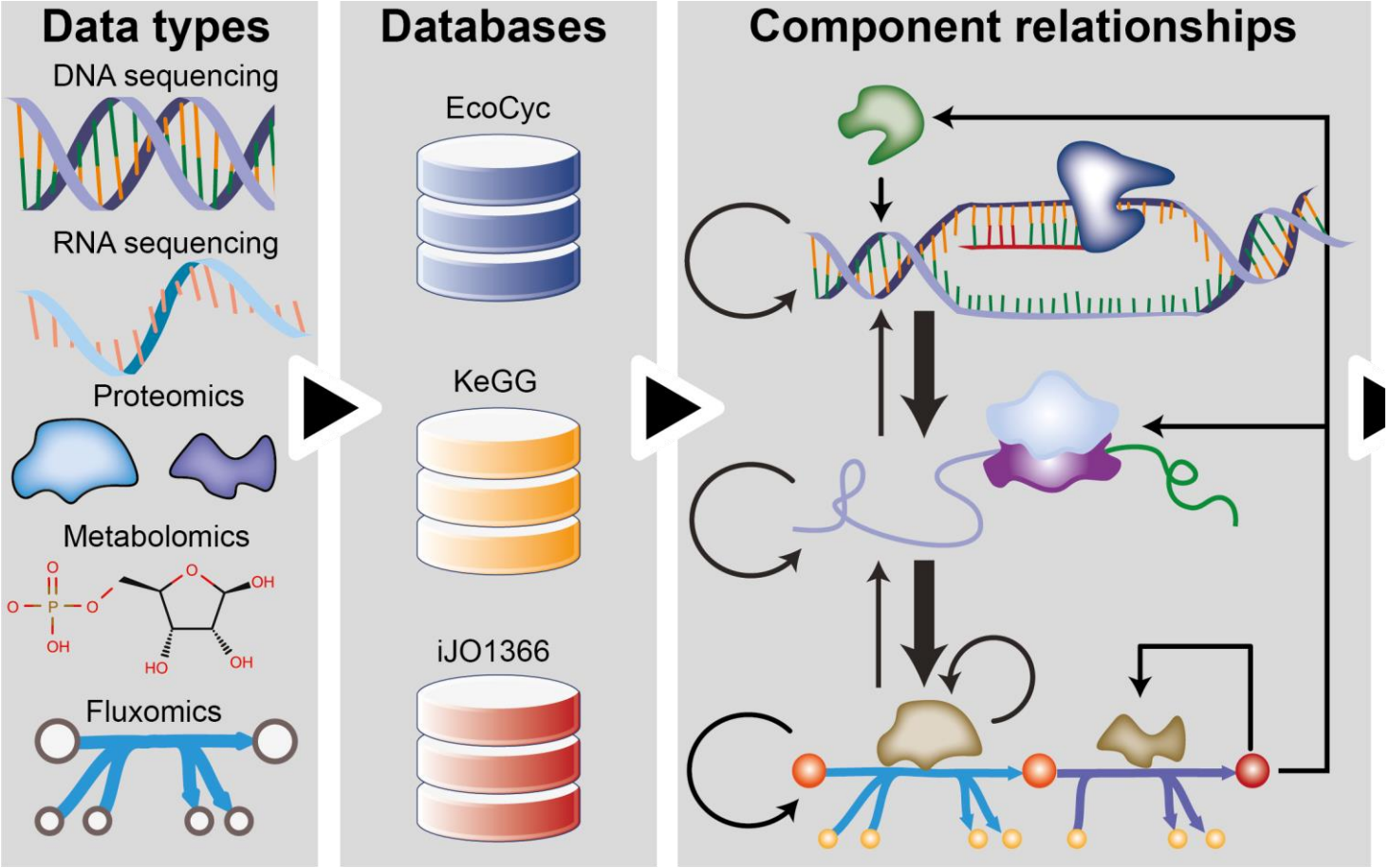
Case Study (1)

Whole cell *E. coli* biochemical interaction network



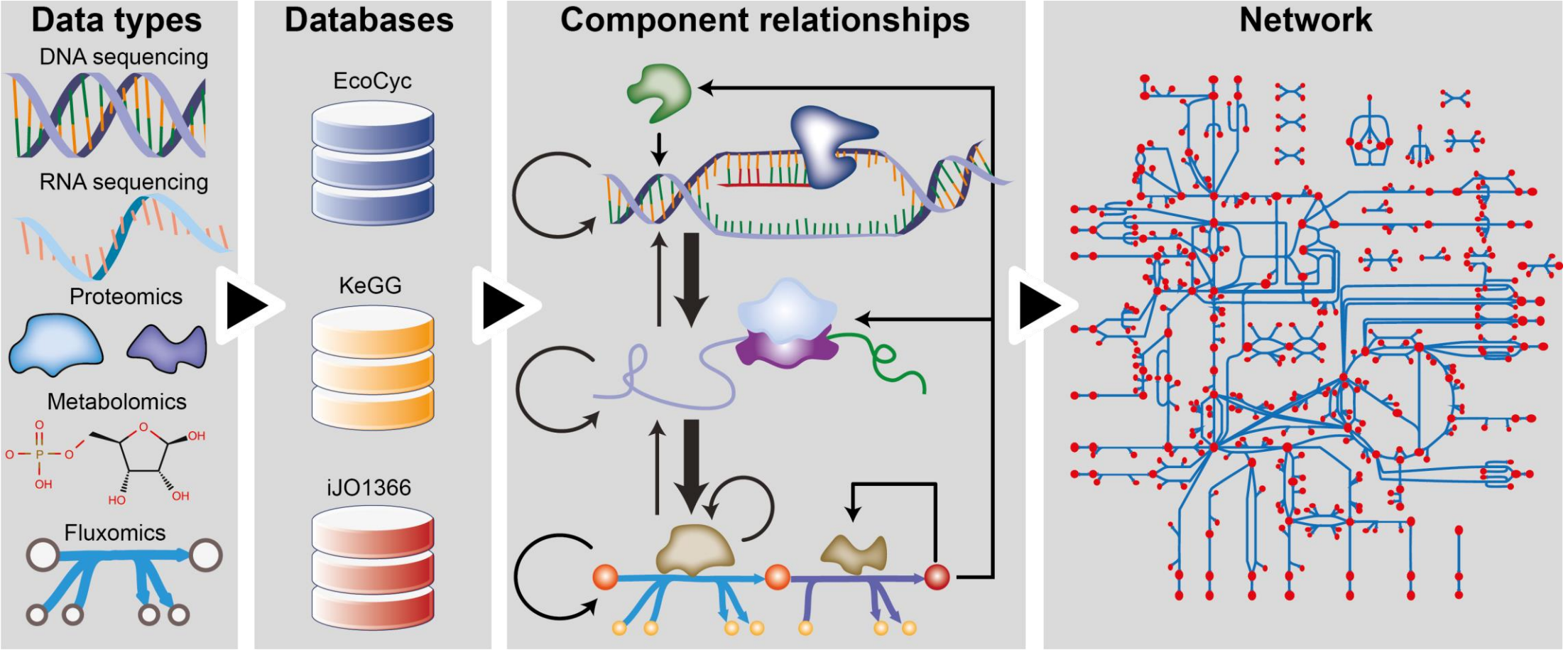
Case Study (1)

Whole cell *E. coli* biochemical interaction network



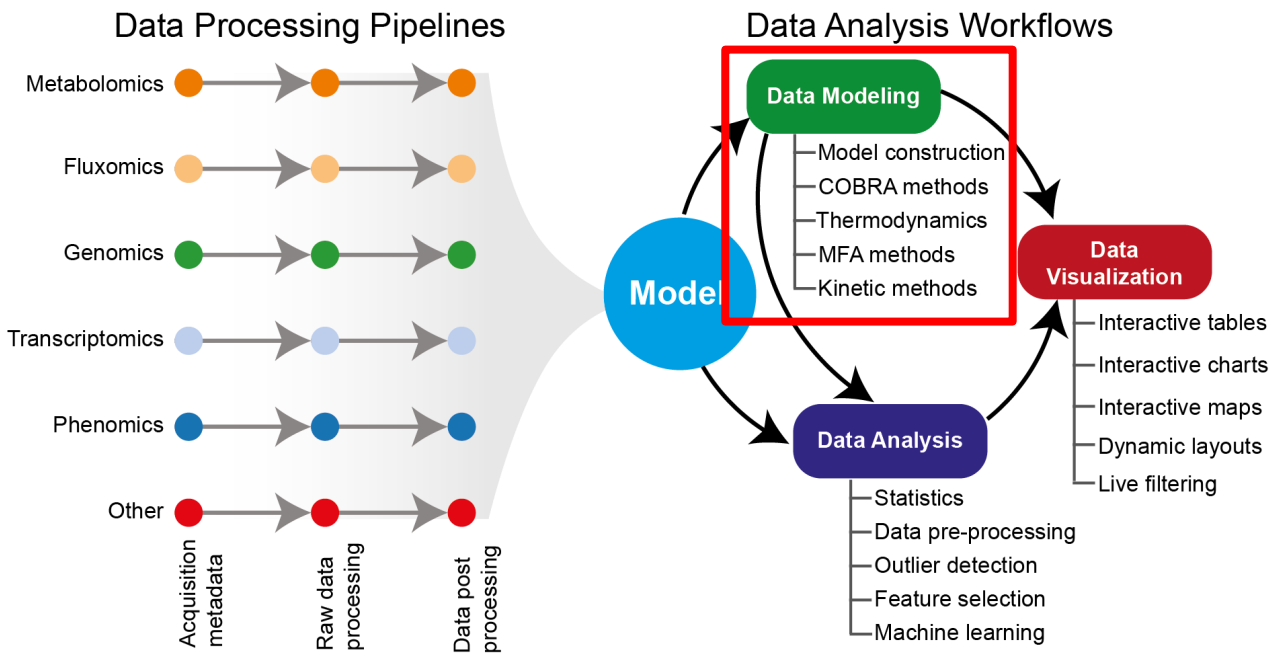
Case Study (1)

Whole cell *E. coli* biochemical interaction network



Biochemical modeling

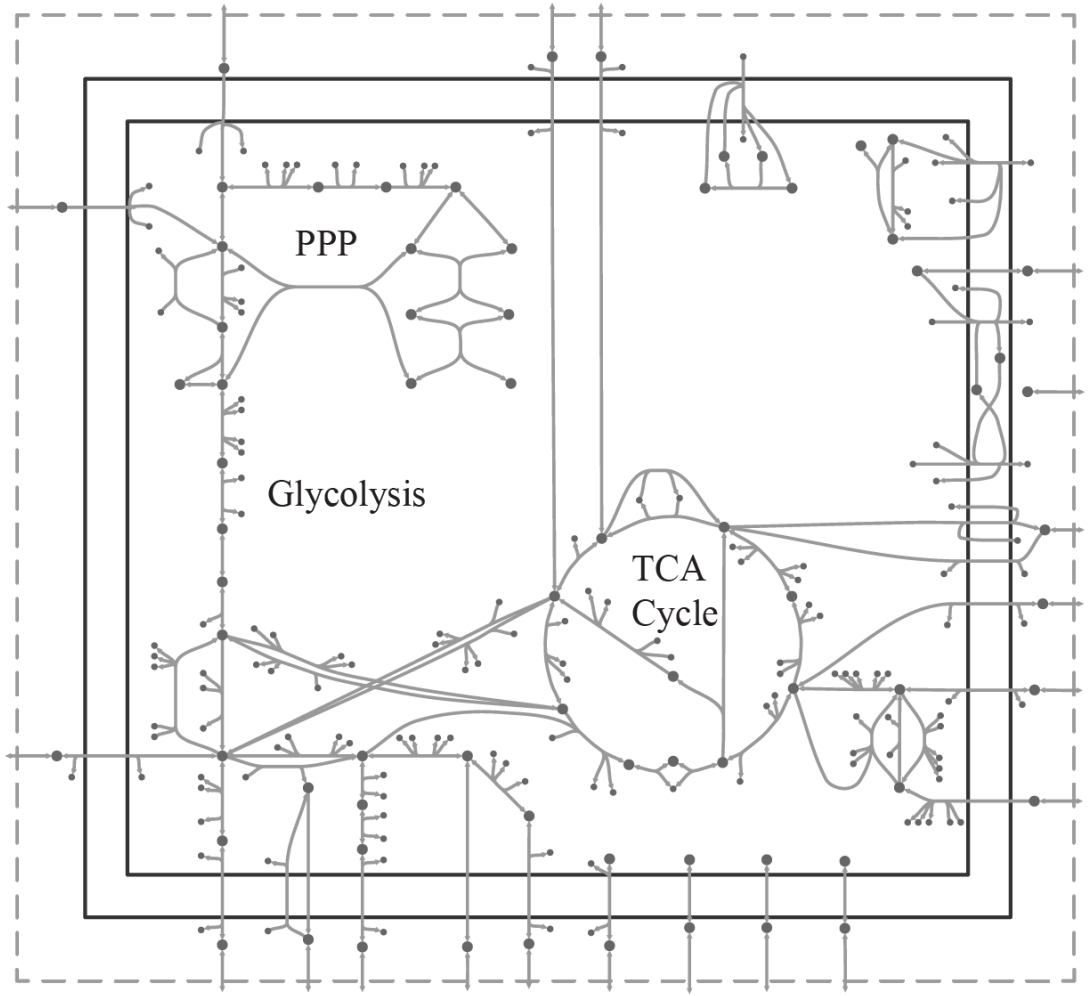
Modeling (1) constraint-based analysis
Modeling (2) thermodynamic analysis
Modeling (3) metabolic flux analysis (MFA)
Modeling (4) kinetic analysis



Modeling (1) constraint-based analysis

- Synopsis: mathematical approach to predict network fluxes (i.e., reaction rates) at steady-state based on mass balance and optimization
- Pros:
  - Can incorporate little to no –omics data to make predictions
  - Scalable from e.g., metabolism to whole cell models to multi-tissue models
  - Provides estimations of the optimal growth, uptake, and secretion rate given a systems external (i.e., environment, media, etc.,) and internal (i.e., genotype) state
- Cons:
  - Limited accuracy due to missing kinetic constraints
- Applications:
  - Media optimization in biotechnology
  - Genotype optimization in biotechnology
  - Experimental design

Modeling (1) constraint-based analysis



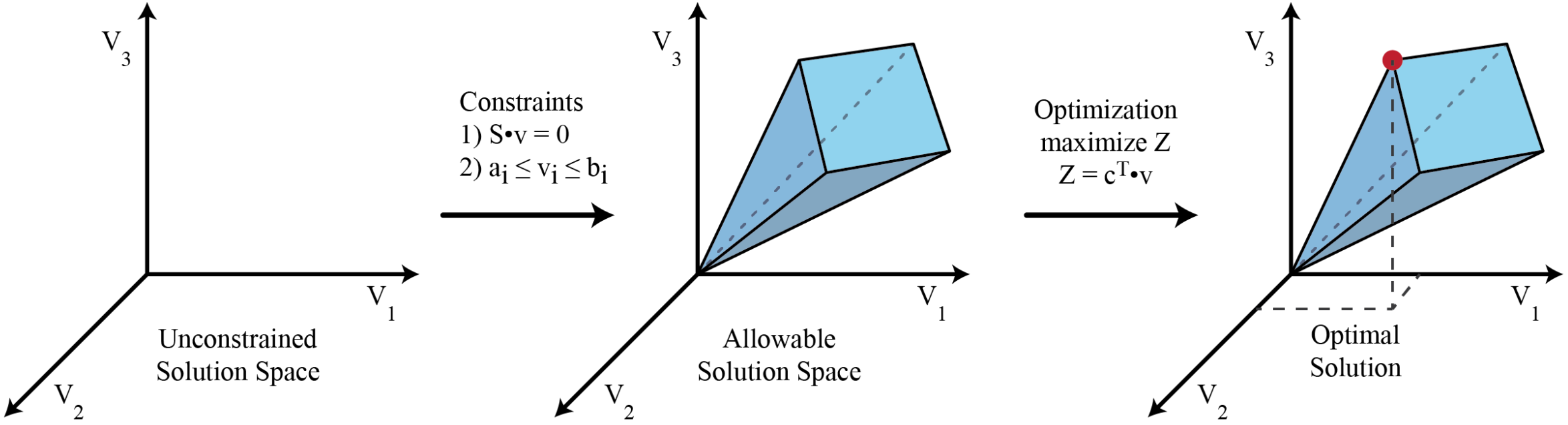
Stoichiometric Matrix

Reaction

Metabolite

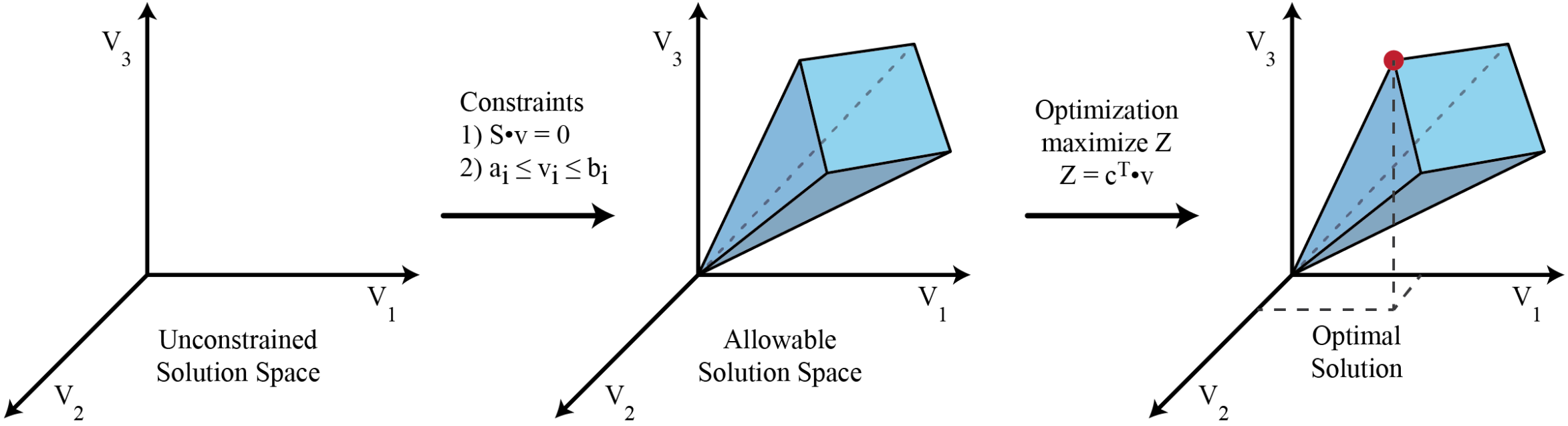
$$S = \begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Modeling (1) constraint-based analysis



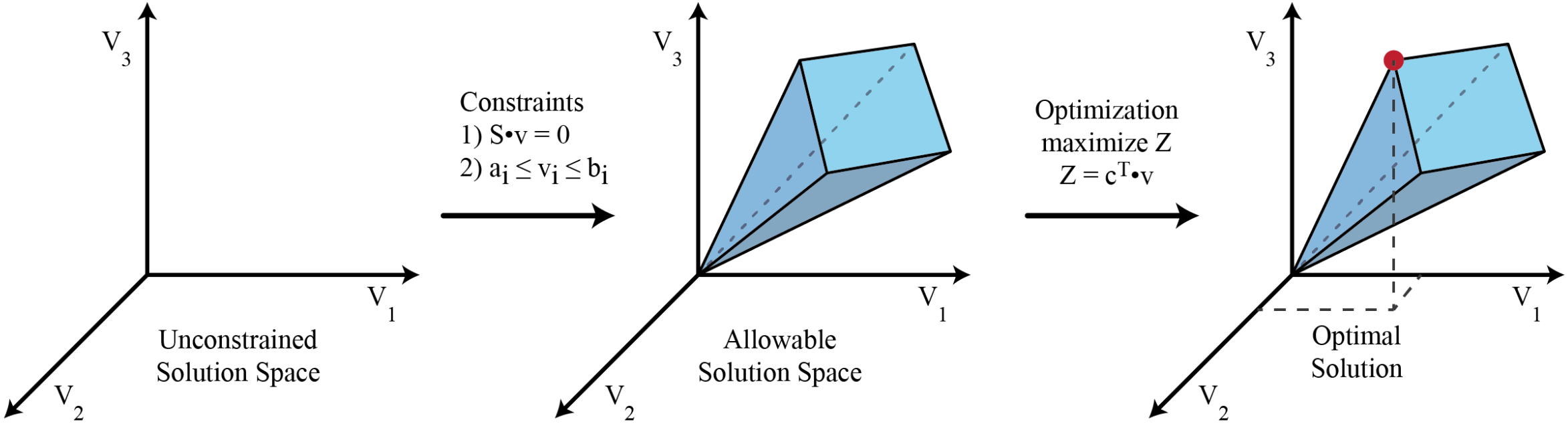
Modeling (1) constraint-based analysis

Objective: Find the optimal network fluxes (i.e., reaction rates) given set of reactions (V) and constraints



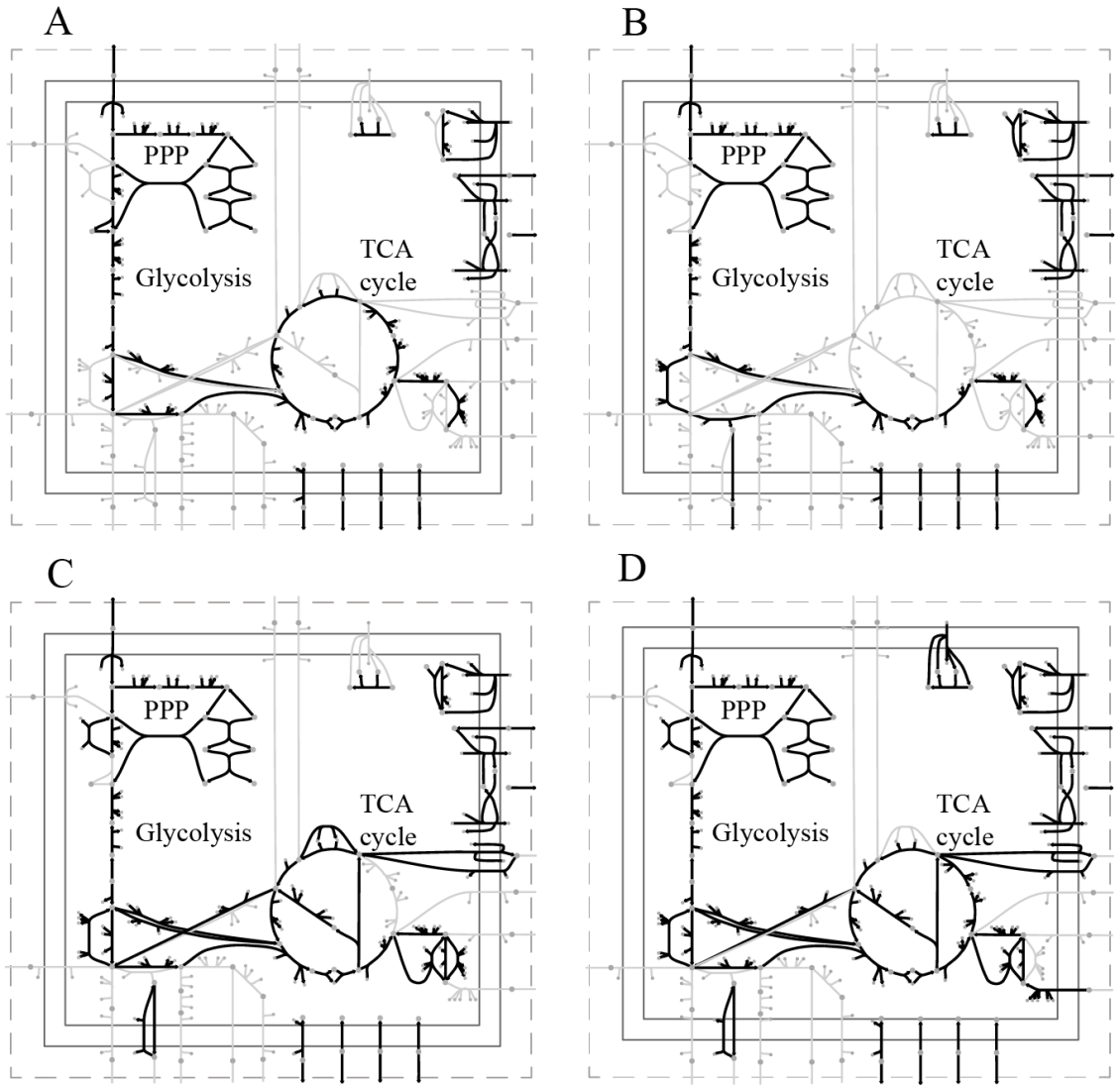
Modeling (1) constraint-based analysis

Objective: Find the optimal network fluxes (i.e., reaction rates) given set of reactions (V) and constraints



Optimization: Linear, Quadratic, Non-linear, or mixed integer programming

Modeling (1) constraint-based analysis



Simulate: Predicted flux under 4 different conditions

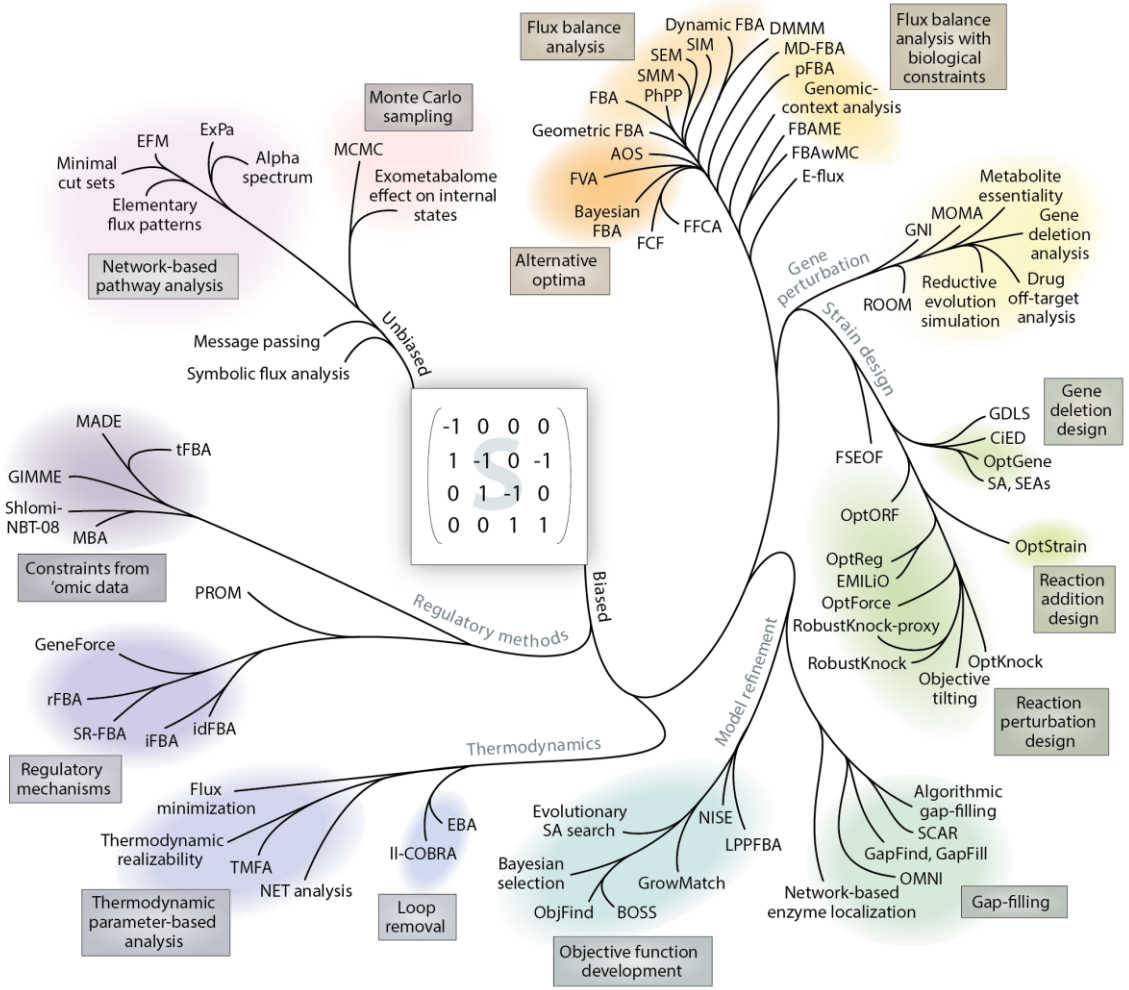
Modeling (1) constraint-based analysis

Constraining the metabolic genotype–phenotype relationship using a phylogeny of *in silico* methods

Nathan E. Lewis[1], Harish Nagarajan [2] *and Bernhard O. Palsson*[1]

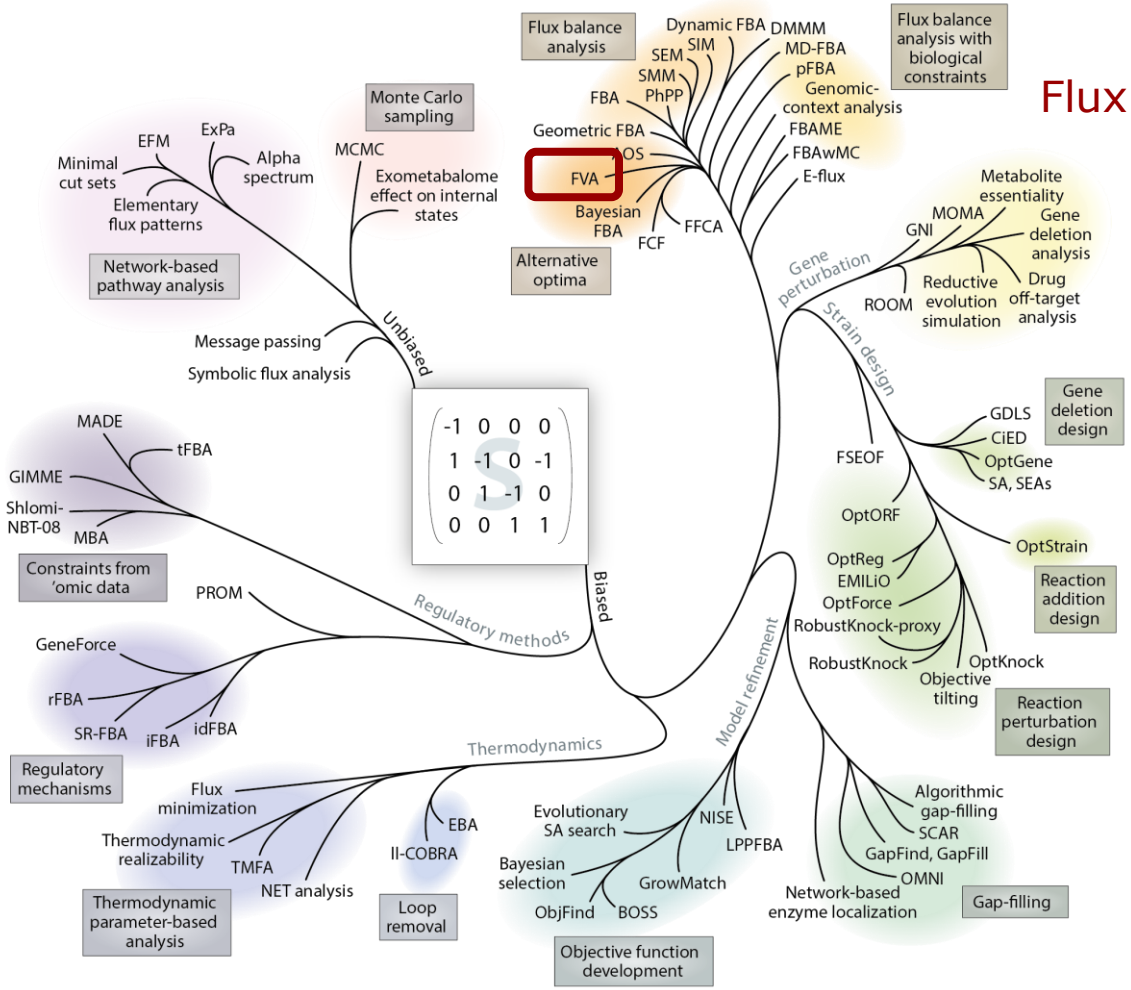
doi:10.1038/nrmicro2737

Modeling (1) constraint-based analysis



doi:10.1038/nrmicro2737

Modeling (1) constraint-based analysis



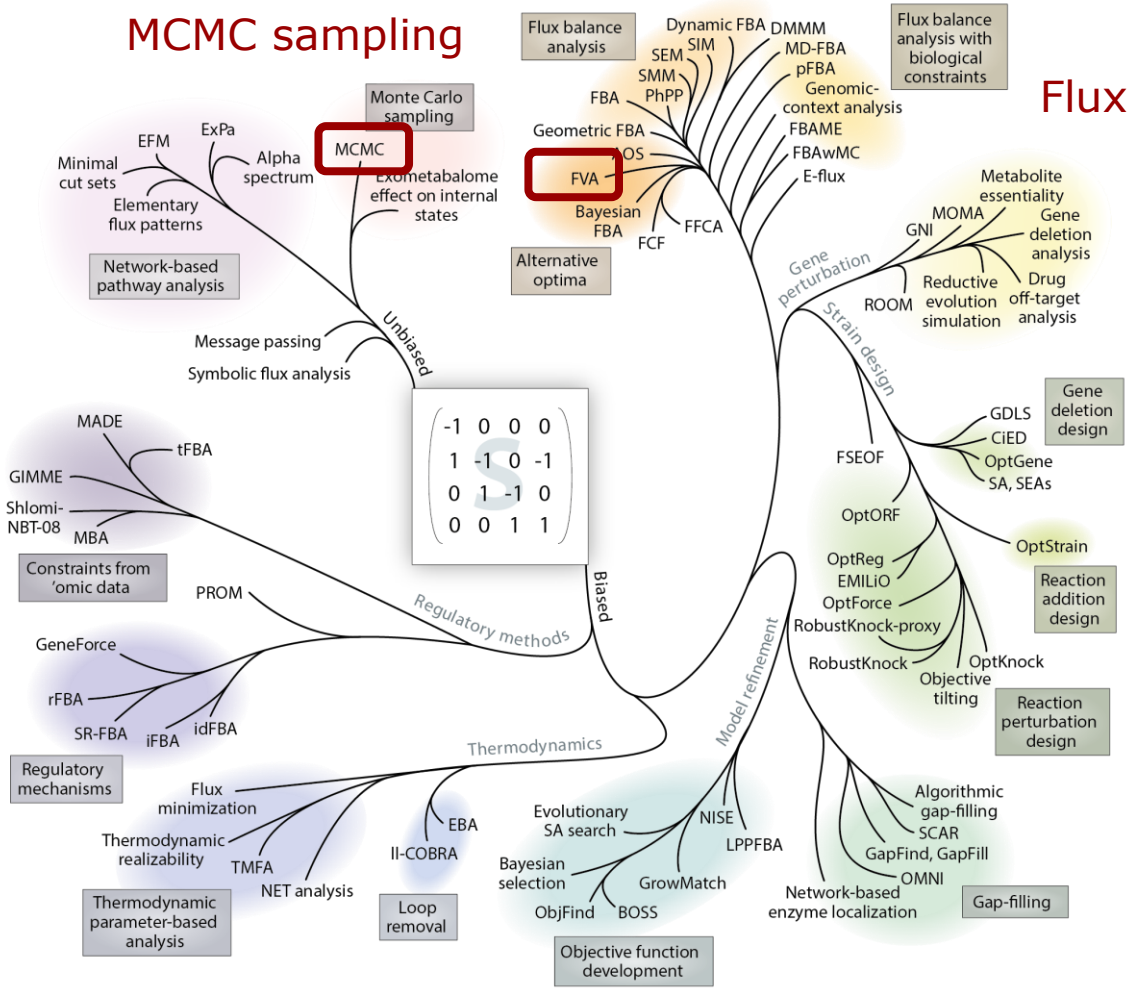
Flux variability analysis

doi:10.1038/nrmicro2737

Modeling (1) constraint-based analysis

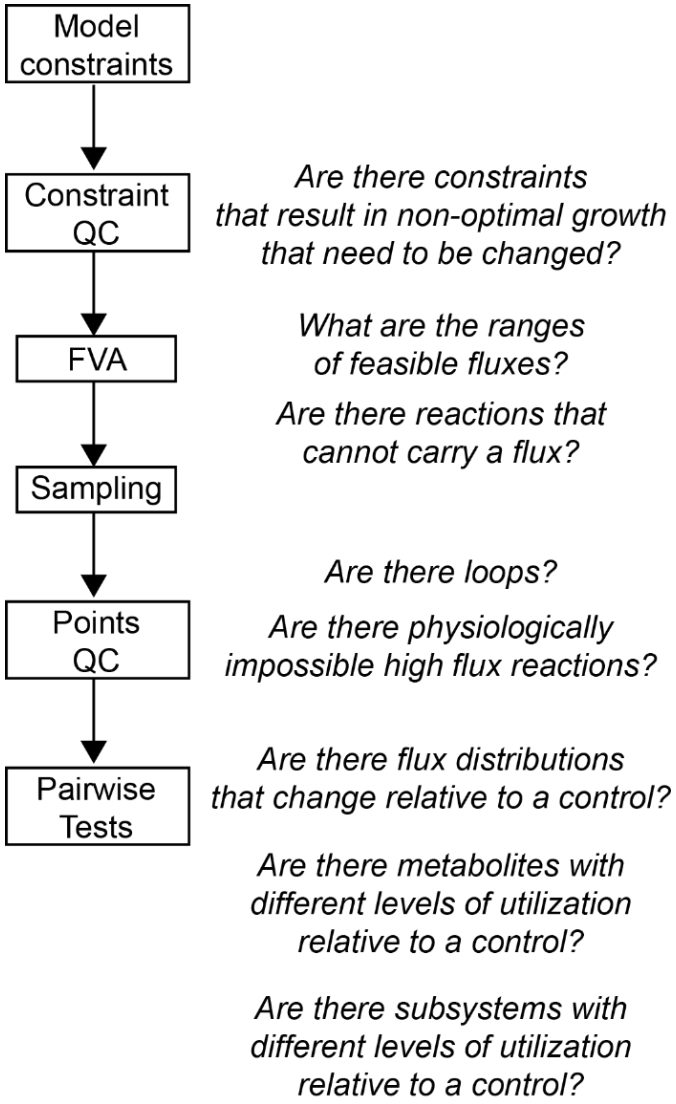
MCMC sampling

Flux variability analysis



doi:10.1038/nrmicro2737

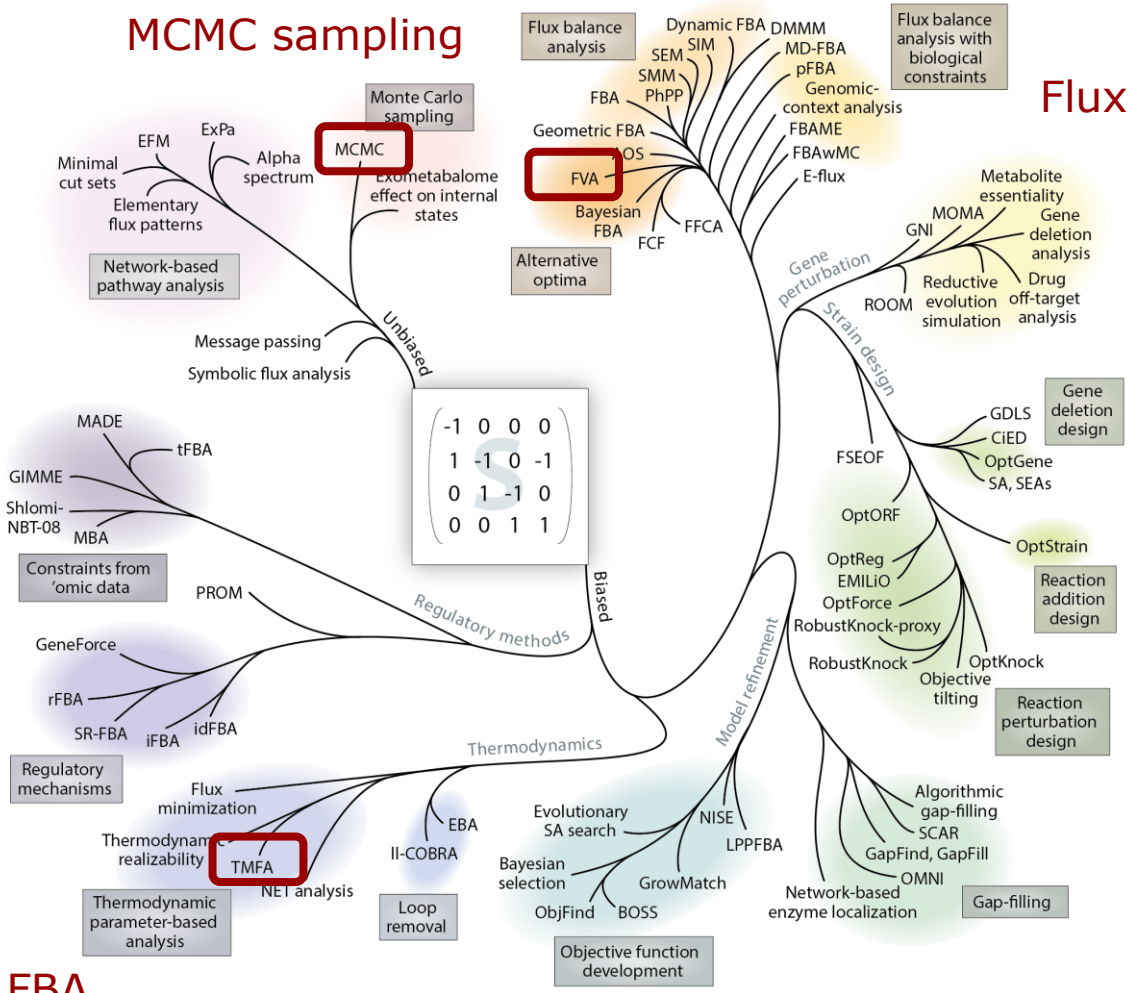
Modeling (1) constraint-based analysis



Modeling (1) constraint-based analysis

MCMC sampling

Flux variability analysis



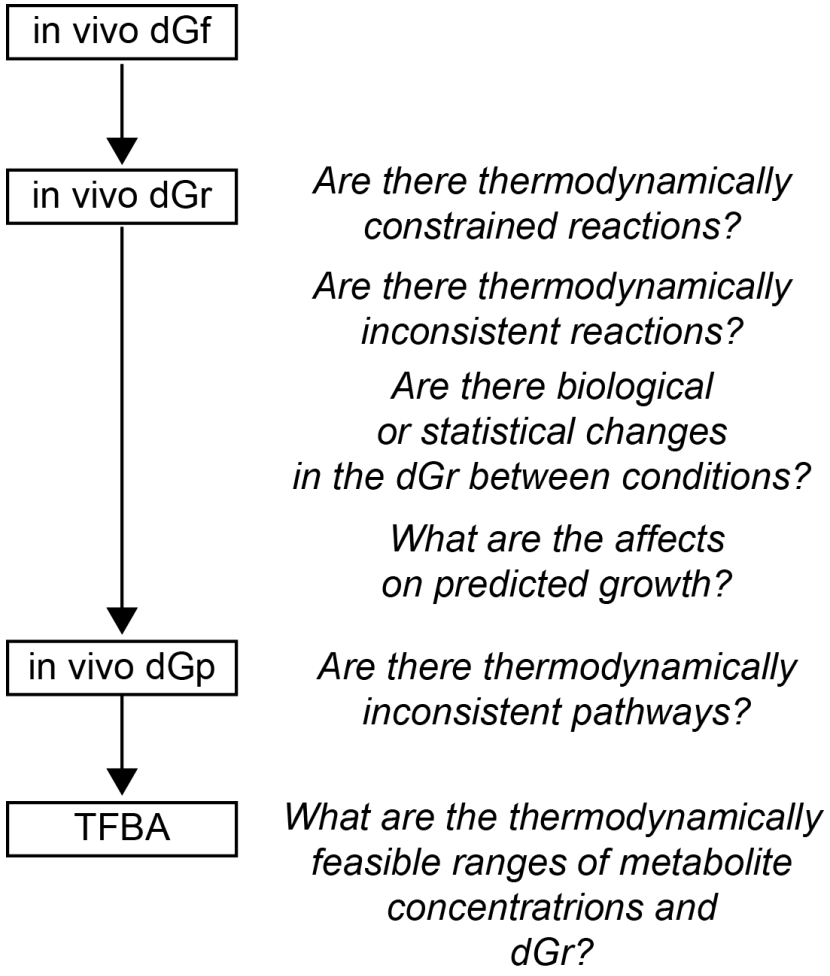
Thermodynamic FBA

doi:10.1038/nrmicro2737

Modeling (2) thermodynamic analysis

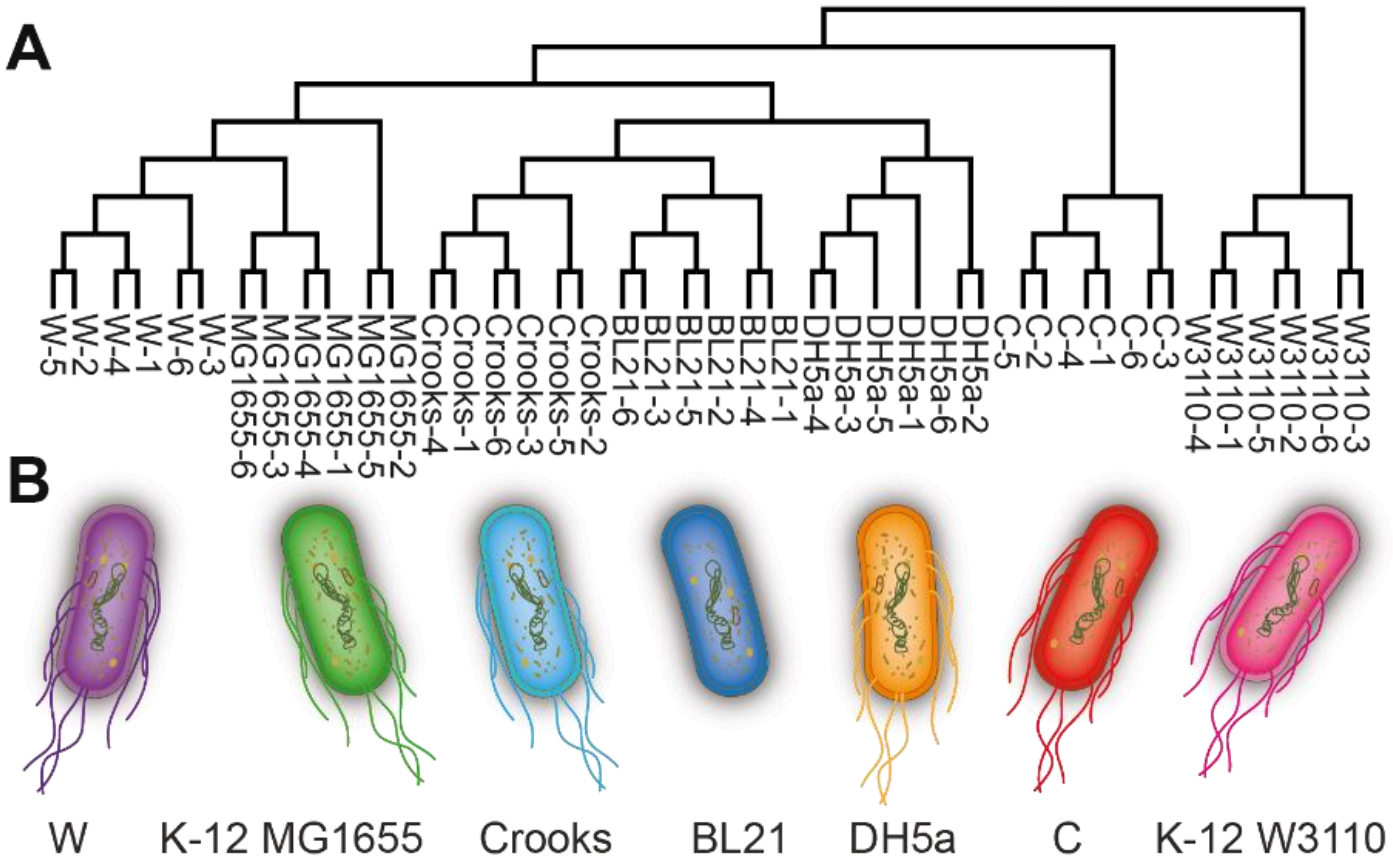
- Synopsis: addition of thermodynamic constraints using metabolomics data to constrain reaction directionality
- Pros: incorporates metabolomics data directly into the optimization problem
- Cons: requires compound heats of formation data
- Applications:
  - Quality control of metabolomics data to determine thermodynamic feasibility
  - Model minimization for MFA analysis
  - Sample missing metabolomics data that is thermodynamically feasible

Modeling (2) thermodynamic analysis



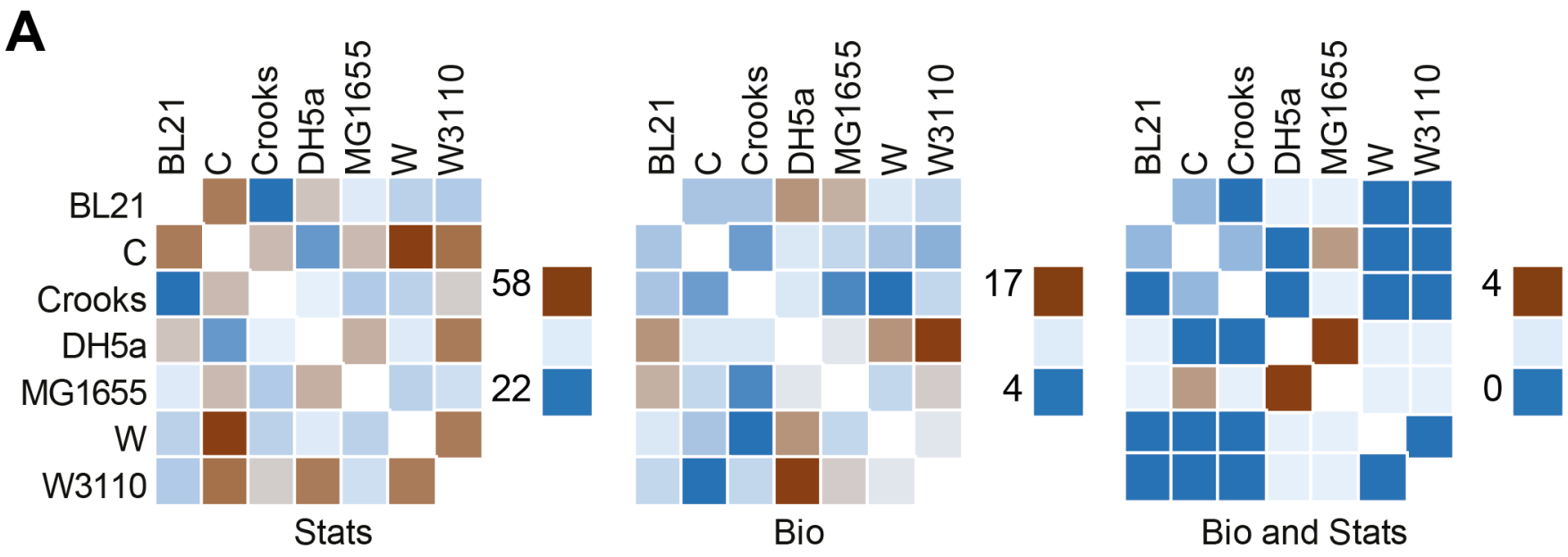
Case Study (3)

Case Study (3)



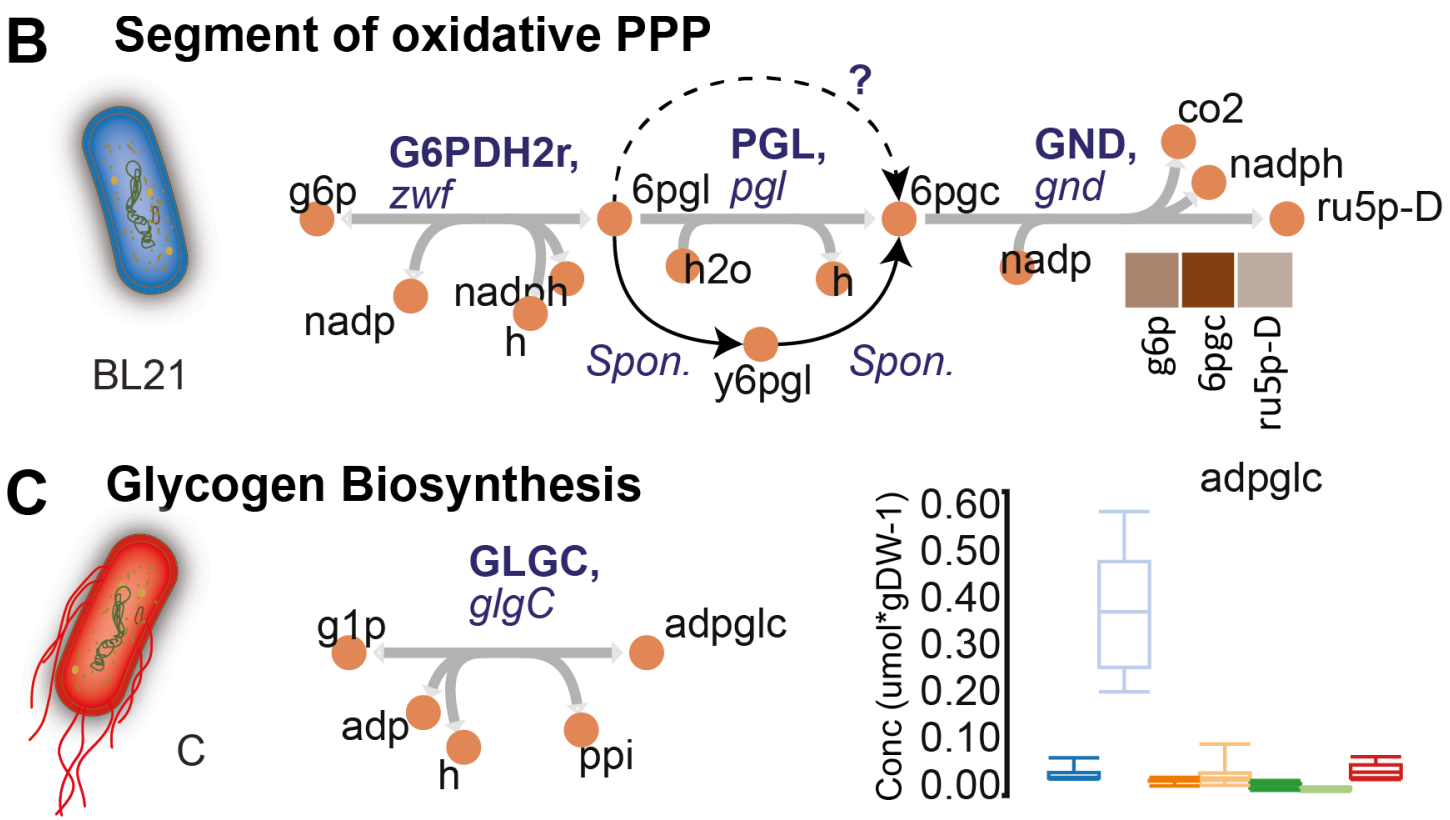
D McCloskey, J Xu, L Schrübbers, H Christensen, M Herrgård. RapidRIP quantifies the intracellular metabolome of 7 industrial strains of E. coli (2018). Metabolic Engineering, 47, 383-392

Case Study (3)



D McCloskey, J Xu, L Schrübbers, H Christensen, M Herrgård. RapidRIP quantifies the intracellular metabolome of 7 industrial strains of E. coli (2018). Metabolic Engineering, 47, 383-392

Case Study (3)



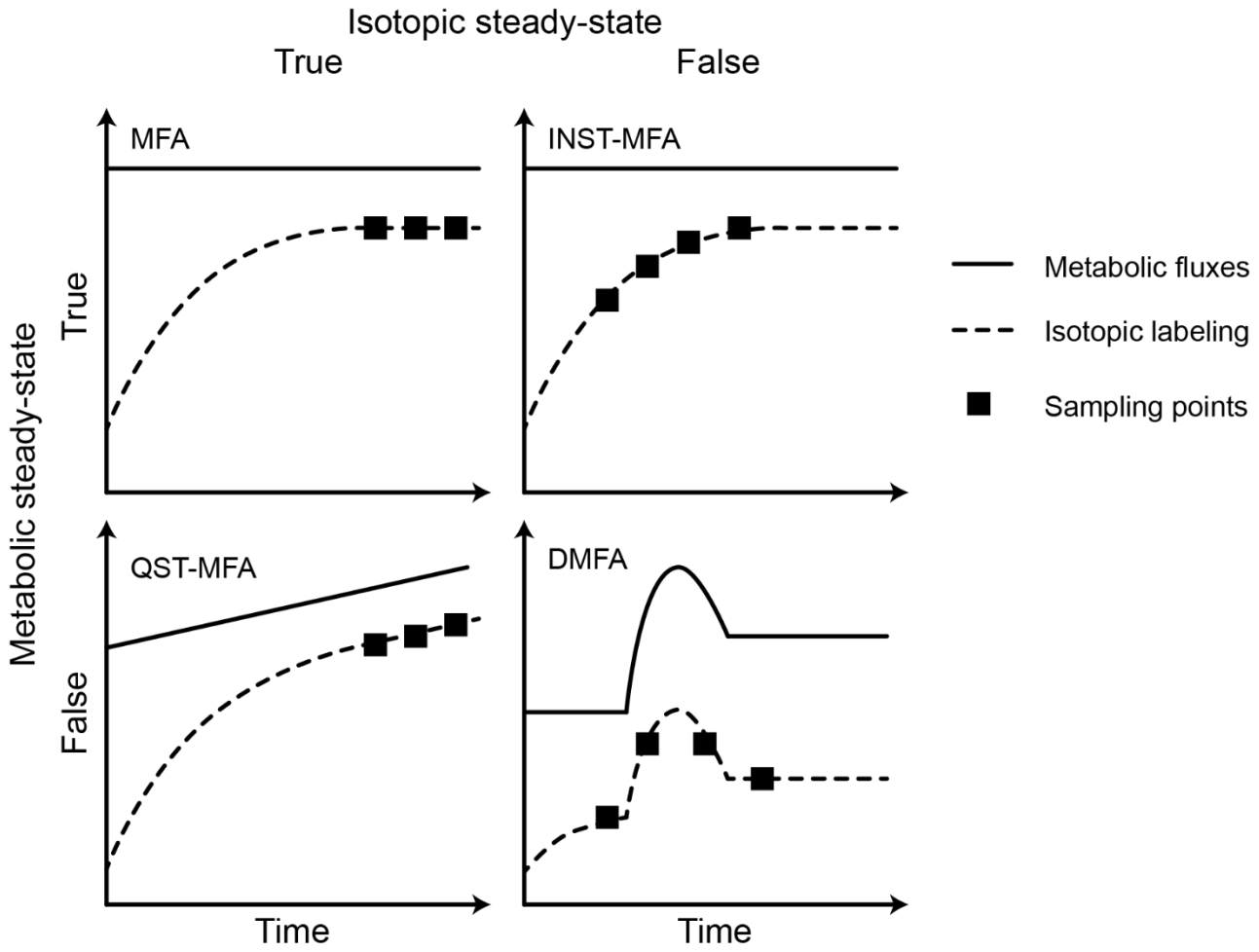
D McCloskey, J Xu, L Schrübbers, H Christensen, M Herrgård. RapidRIP quantifies the intracellular metabolome of 7 industrial strains of E. coli (2018). Metabolic Engineering, 47, 383-392

Modeling (3) metabolic flux analysis (MFA)

- Synopsis: utilizes isotope labeling information to back calculate the most consistent fluxes given a set of reactions
- Pros:
  - The most accurate way to calculate intracellular fluxes
- Cons:
  - Requires an atom mapping for all reactions
  - Often constrained to small reaction networks
  - Expensive both in terms of instrumentation and raw material required for the experiments
- Applications:
  - Optimization of carbon/nitrogen flow in biotechnology
  - Understanding the transfer of metabolites between compartments, cells, and tissues

Modeling (3) metabolic flux analysis (MFA)

MFA methodologies



Based on Fig. 2 from doi: 10.1007/978-1-62703-299-5_18

Modeling (3) metabolic flux analysis (MFA)

Choice of model scope

DOF = free fluxes - measured fluxes - measured fragment isotopes

Modeling (3) metabolic flux analysis (MFA)

Choice of model scope

Traditional/Core

- Accounts for the minimal amount of metabolism
- Fast simulation time
- Atom mapping is readily available
- Fewer reactions = greater DOF

DOF = free fluxes - measured fluxes - measured fragment isotopes

Modeling (3) metabolic flux analysis (MFA)

Choice of model scope

Traditional/Core

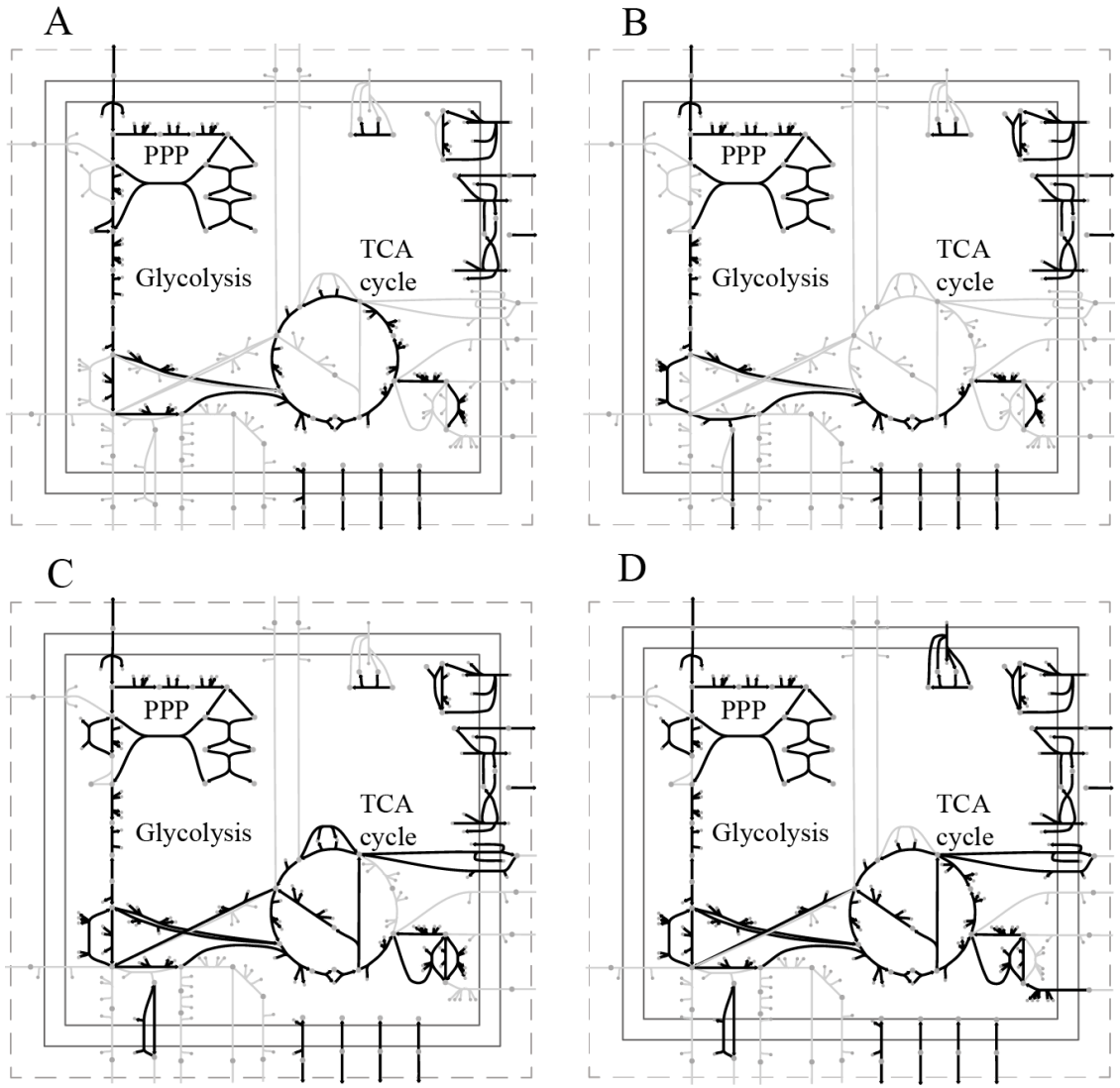
- Accounts for the minimal amount of metabolism
- Fast simulation time
- Atom mapping is readily available
- Fewer reactions = greater DOF

Genome-scale

- Accounts for cofactor usage, biosynthesis, and salvage
- Slow simulation time
- Atom mapping is not always available
- More reactions = less DOF
    - Requires additional measurements to generate a statistically significant fit

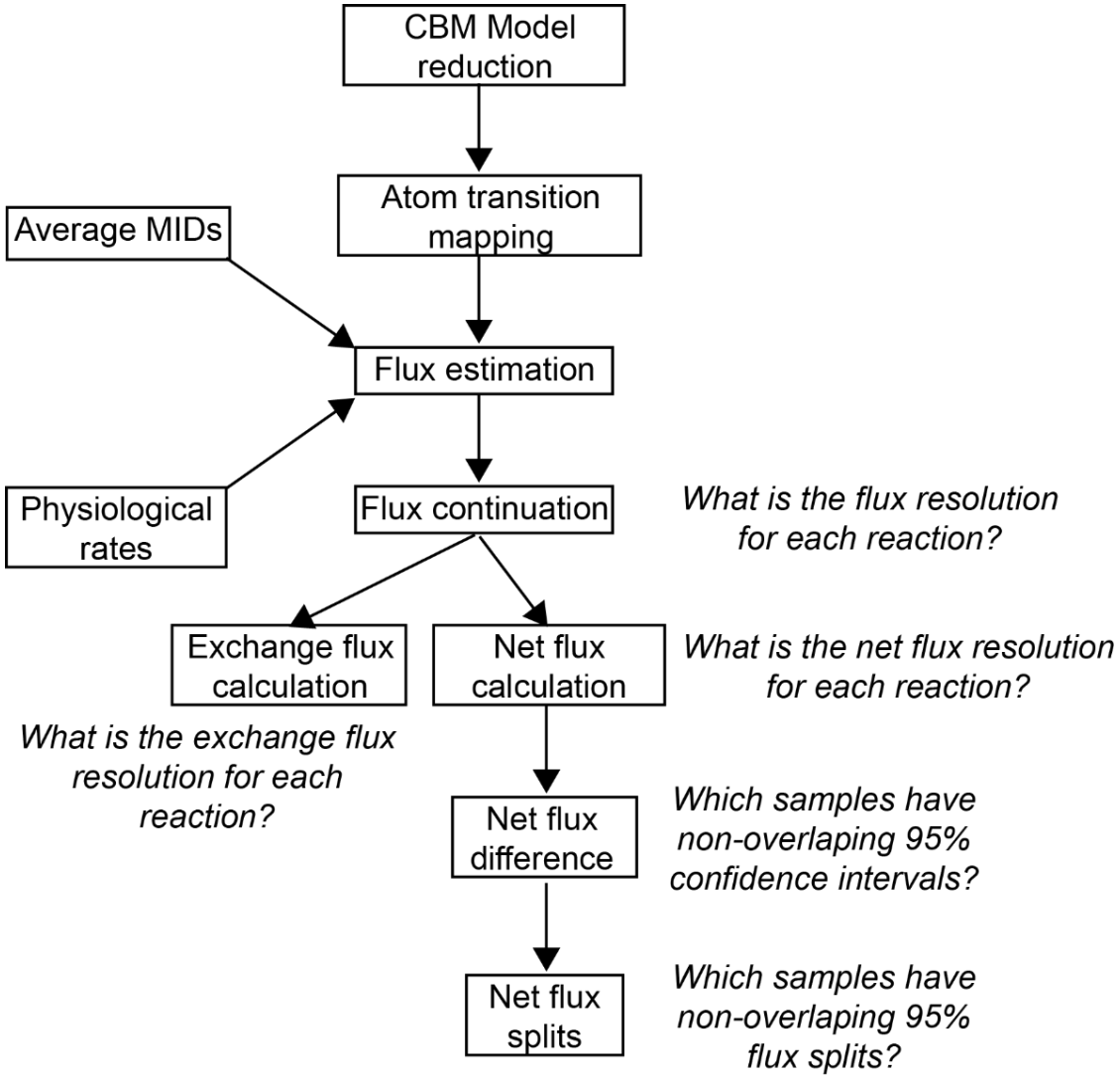
DOF = free fluxes - measured fluxes - measured fragment isotopes

Modeling (3) metabolic flux analysis (MFA)



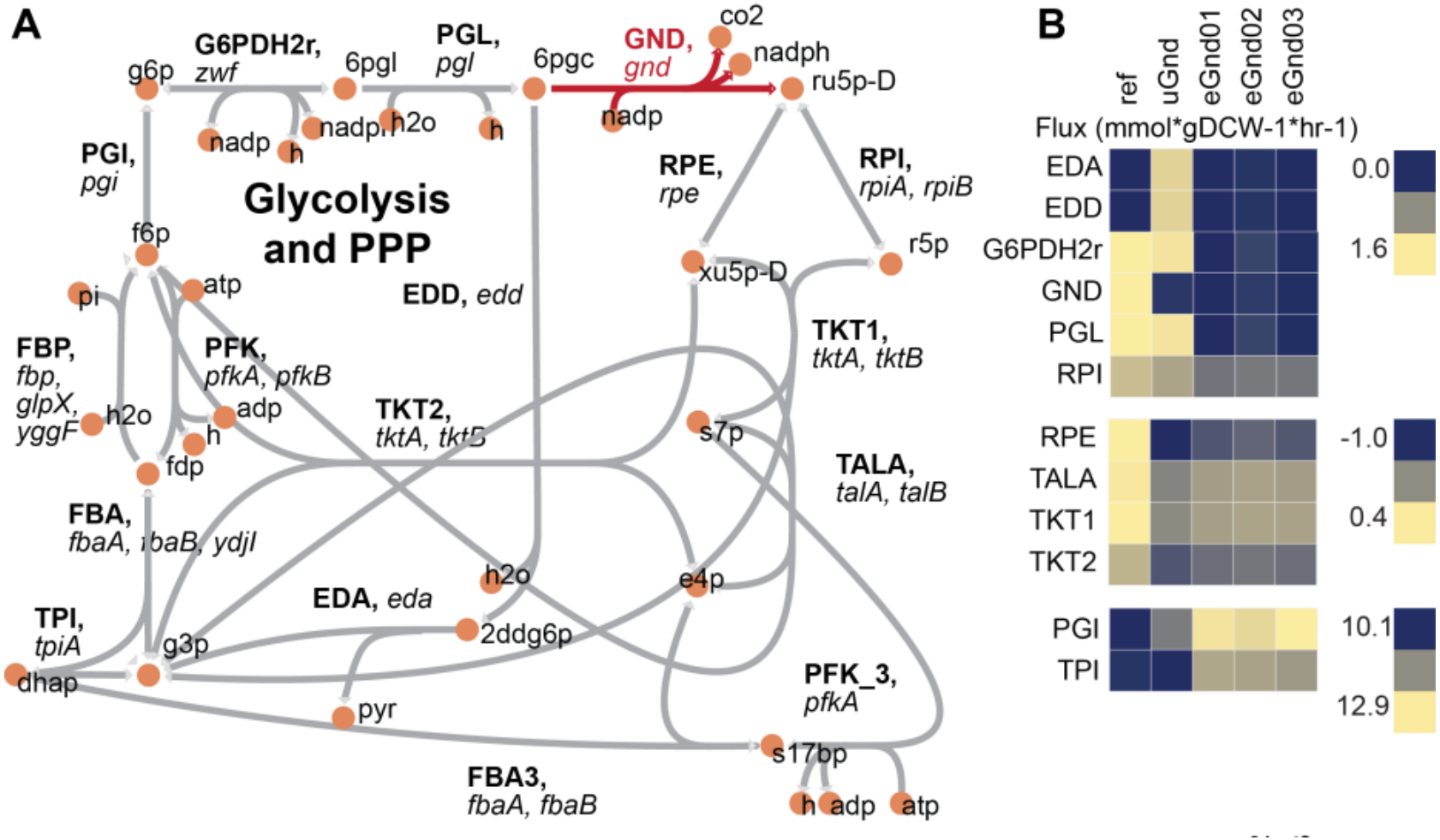
Calculate: Measured flux under 4 different conditions

Modeling (3) metabolic flux analysis (MFA)

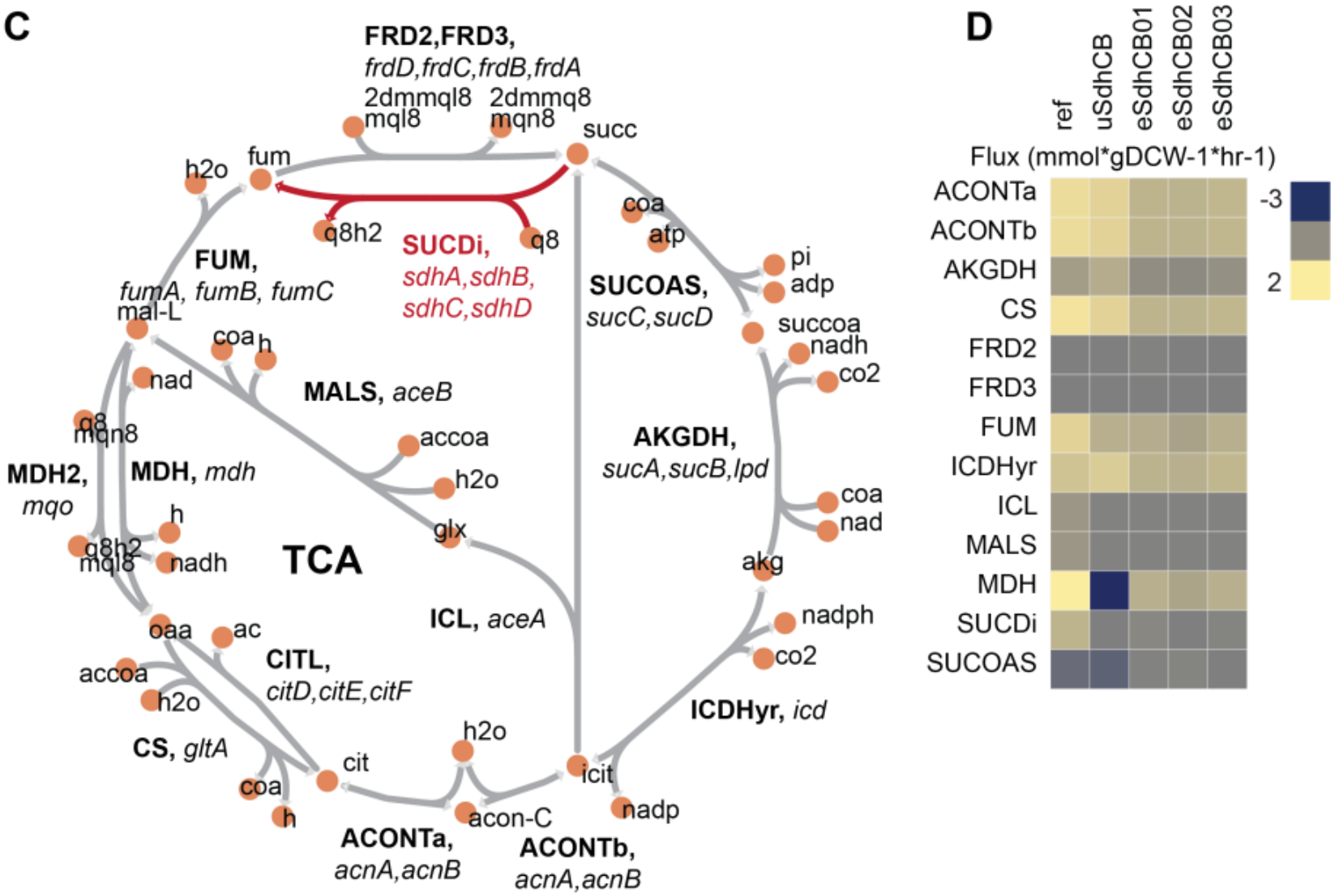


Case Study (1)

Case Study (1)


A
Glycolysis and PPP
G6PDH2r, zwf
PGL, pgl
GND, gnd
PGI, pgi
RPE, rpe
RPI, rpiA, rpiB
EDD, edd
FBP, fbp, glpX, yggF
PFK, pfkA, pfkB
TKT2, tktA, tktB
TKT1, tktA, tktB
TALA, talA, talB
FBA, fbaA, fbaB, ydjI
EDA, eda
TPI, tpiA
PFK_3, pfkA
FBA3, fbaA, fbaB
g6p
6pgl
6pgc
co2
nadph
ru5p-D
nadp
h
h2o
f6p
pi
atp
r5p
xu5p-D
s7p
fdp
adp
g3p
dhap
2ddg6p
e4p
pyr
s17bp
B
ref
uGnd
eGnd01
eGnd02
eGnd03
Flux (mmol*gDCW-1*hr-1)
EDA
EDD
G6PDH2r
GND
PGL
RPI
RPE
TALA
TKT1
TKT2
PGI
TPI
0.0
1.6
-1.0
0.4
10.1
12.9

Case Study (1)


C
FRD2,FRD3,
frdD,frdC,frdB,frdA
2dmmql8
mql8
2dmmq8
mqn8
succ
fum
h2o
q8h2
SUCDi,
sdhA,sdhB,
sdhC,sdhD
q8
coa
atp
pi
adp
FUM,
fumA, fumB, fumC
mal-L
SUCOAS,
sucC,sucD
succoa
nadh
co2
coa
h
nad
MALS, aceB
q8
mqn8
accoa
h2o
AKGDH,
sucA,sucB,lpd
MDH2,
mqo
MDH, mdh
h
q8h2
mql8
nadh
glx
TCA
coa
nad
akg
ICL, aceA
nadph
co2
oaa
ac
CITL,
citD,citE,citF
accoa
h2o
ICDHyr, icd
CS, gltA
h2o
cit
icit
coa
h
acon-C
nadp
ACONTa,
acnA,acnB
ACONTb,
acnA,acnB
D
ref
uSdhCB
eSdhCB01
eSdhCB02
eSdhCB03
Flux (mmol*gDCW-1*hr-1)
ACONTa
ACONTb
AKGDH
CS
FRD2
FRD3
FUM
ICDHyr
ICL
MALS
MDH
SUCDi
SUCOAS
-3
2

Modeling (4) kinetic analysis

- Synopsis: mechanistic model of biochemical interactions that provides information on system dynamics
- Pros:
  - As mechanistic as one can get in describing how biological entities interact
  - Predictions can be the most accurate
- Cons:
  - Requires multiple –omics data types
  - Requires knowledge of enzyme kinetic parameters including reaction mechanisms
  - Difficult to scale to large networks*
- Applications:
  - Flux and enzyme kinetics optimization in biotechnology
  - Cell signaling
  - Studies of enzyme kinetics between organisms
  - Affect of genomic variants on enzyme activity and regulation

Modeling (4) kinetic analysis

Biotechnology Advances 35 (2017) 981– 1003

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

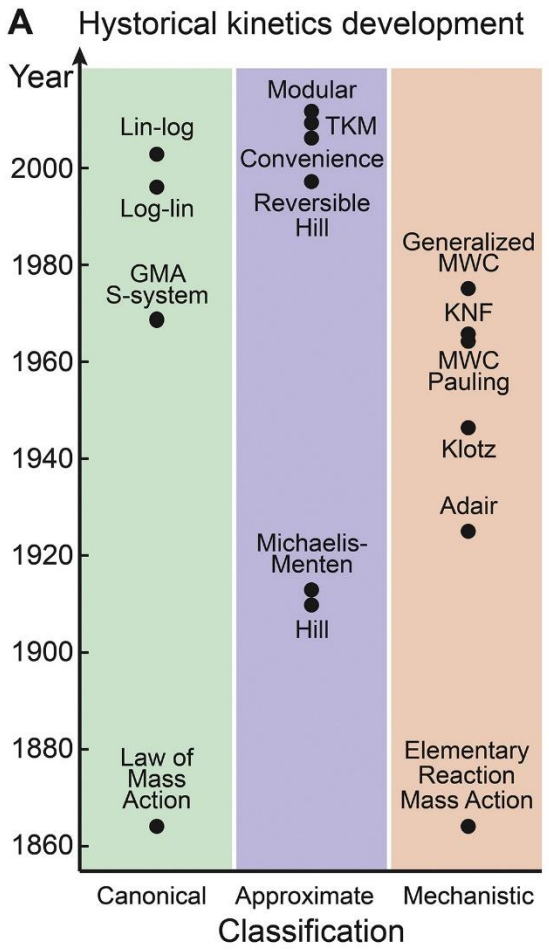
Formulation, construction and analysis of kinetic models of metabolism: A review of modelling frameworks

Pedro A. Saa [a], Lars K. Nielsen [a,b,⊠]

[a] Australian Institute for Bioengineering and Nanotechnology (AIBN), The University of Queensland, Brisbane, QLD 4072, Australia
[b] The Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark, Kemitorvet, Building 220, DK-2800 Kongens Lyngby, Denmark

DOI:10.1016/j.biotechadv.2017.09.005

Modeling (4) kinetic analysis

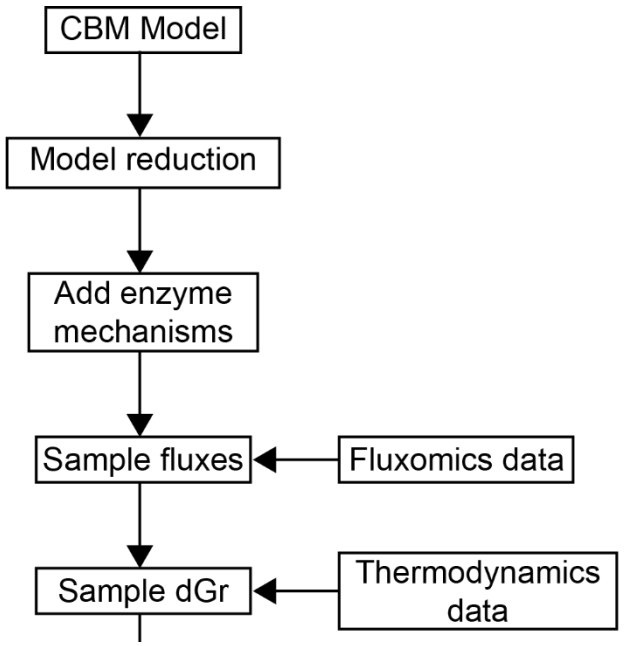


B Comparison of popular enzyme kinetics formalisms

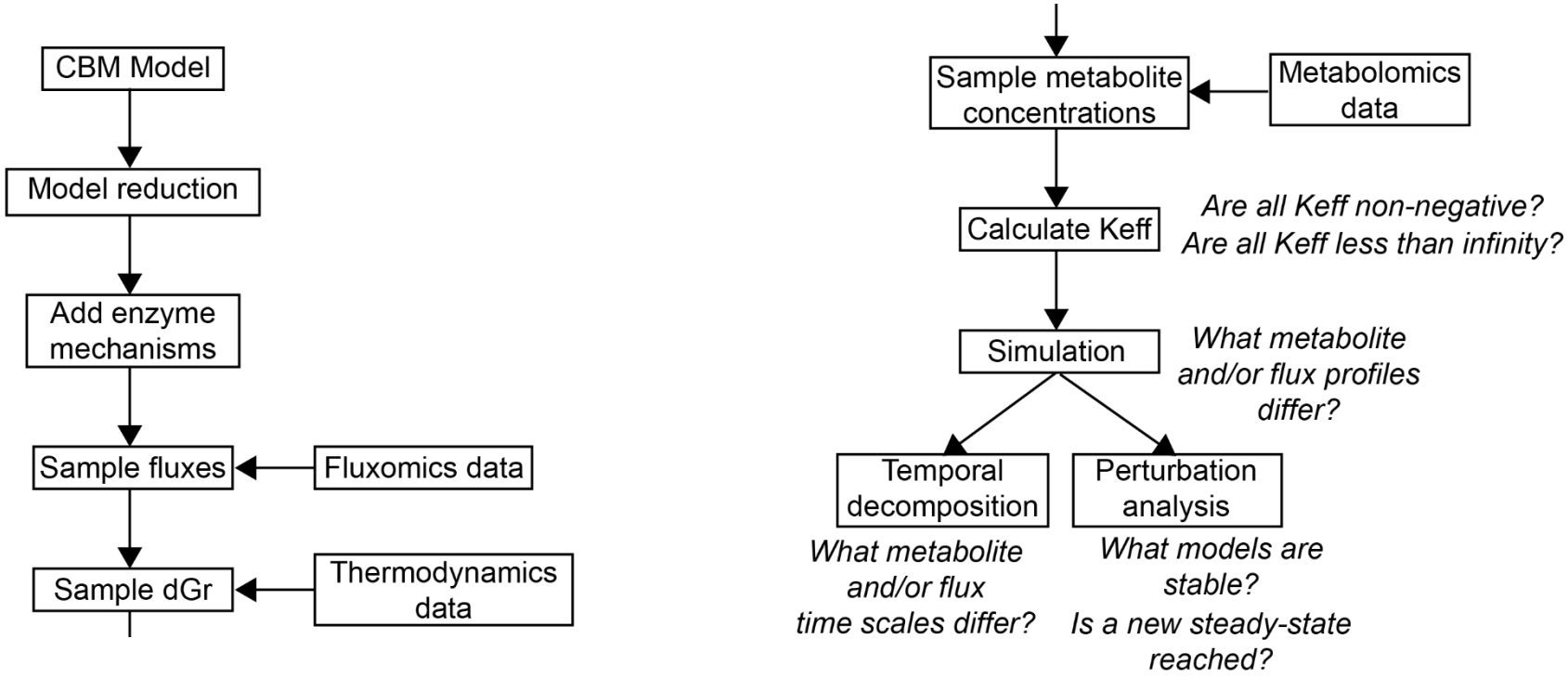
	Formalism	Kinetics supported	Parameterization
Canonical	LMA	Non-allosteric (only locally valid close to equilibrium)	rate constants (k^+,k^-) for the overall reaction
	GMA / S-system	General (only locally valid)	rate constants (k^+,k^-) for the overall reaction, kinetic orders (h_M,h_E)
	Log-lin / Lin-log	General (only locally valid)	reactants and effector elasticities (ε_E,ε_M)
Approximate	Michaelis-Menten	Non-allosteric	max catalytic rate V_{max}, Michaelis constant K_d
	Reversible Hill	General	catalytic rate constant k_{cat+}, equilibrium constant K_{eq}, dissociation constants K_M, Hill coefficient n_H
	Convenience and Modular kinetics	General	catalytic rates (k_{cat+},k_{cat-}), dissociation constants K_M, regulatory constants (k^A,k^I)
	TKM	General (depends on availaility of R)	thermokinetic potentials ξ_M, resistance $R(\xi_M)$
Mechanistic	ER-MA	Non-allosteric	rate constants (k^+,k^-) for each elementary step in the reaction mechanism
	MWC	Cooperative and allosteric binding	binding constants for active (R) and tense (T) conformational states (K_R,K_T), subunit number n, effector binding constants (K_E), allosteric constant L
	KNF	Cooperative and allosteric binding	protein-ligand binding constant K_S, subunit number n, allosteric constant L, subunit interaction energy σ
	Generalized MWC	General	catalytic rates (k_{cat+},k_{cat-}), reactant disociation constants for active (R) and tense (T) conformational states (K_R,K_T), subunit number n, effector binding constants (K_E), allosteric constant L

DOI:10.1016/j.biotechadv.2017.09.005

Modeling (4) kinetic analysis

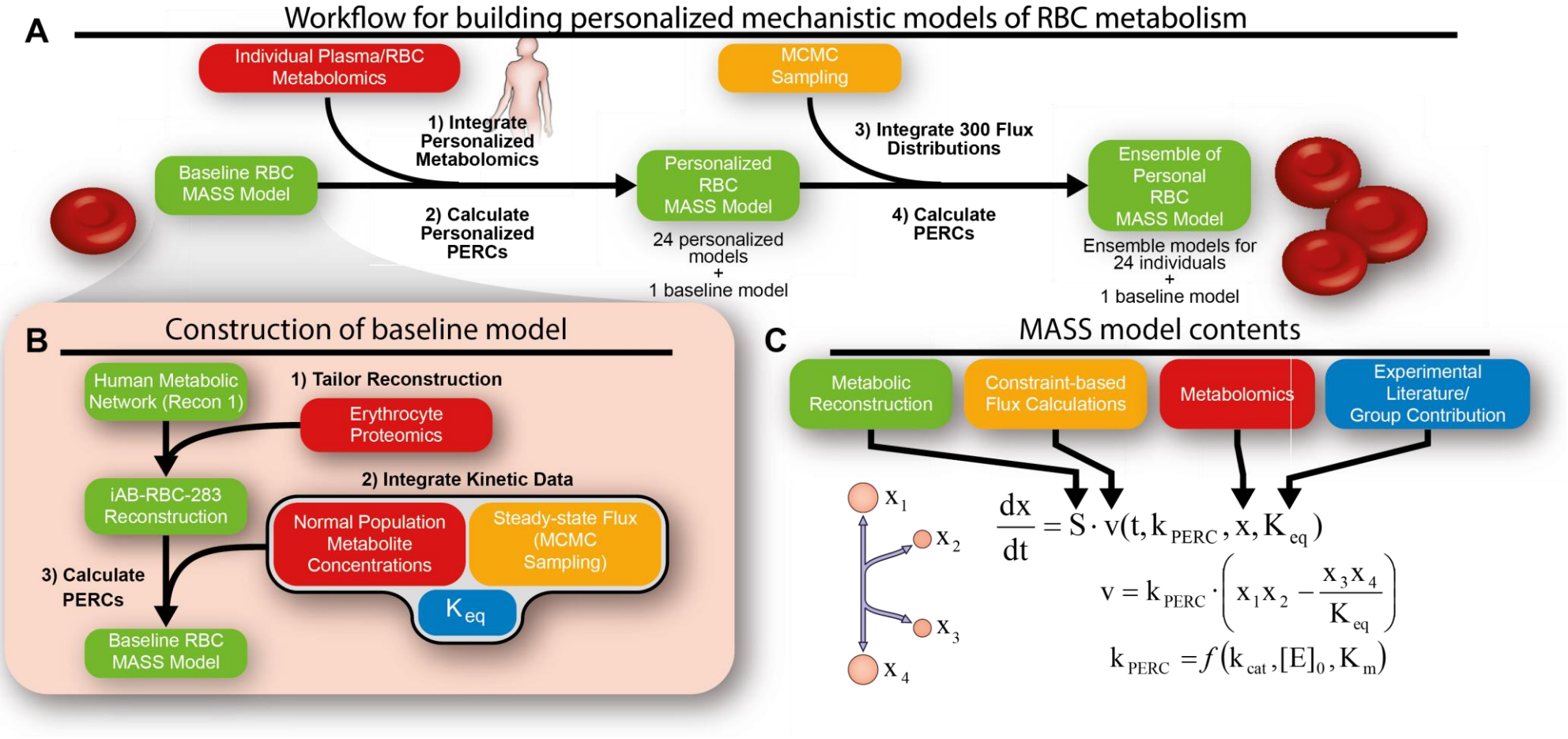


Modeling (4) kinetic analysis



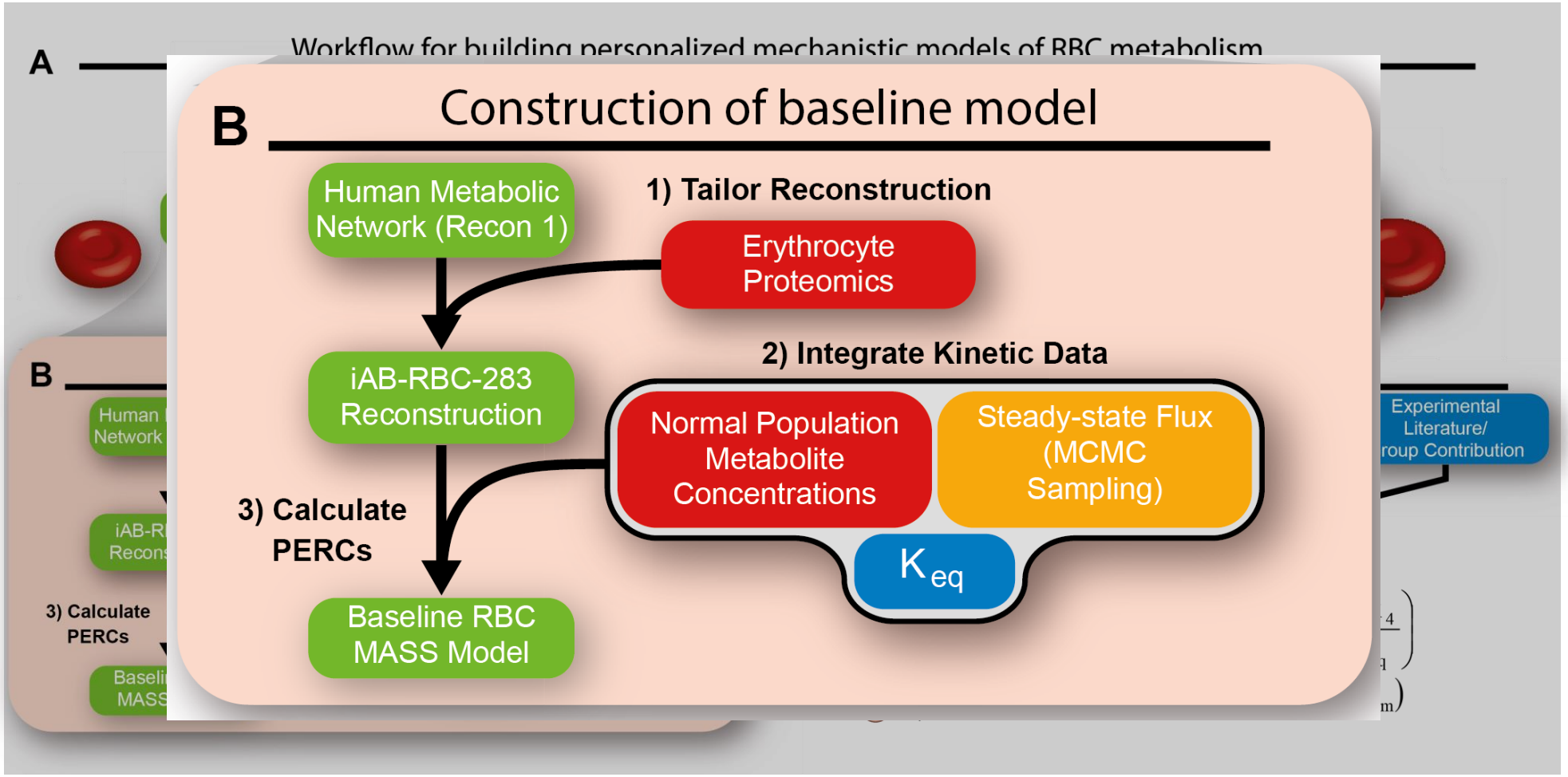
Case Study (2)

Case Study (2)



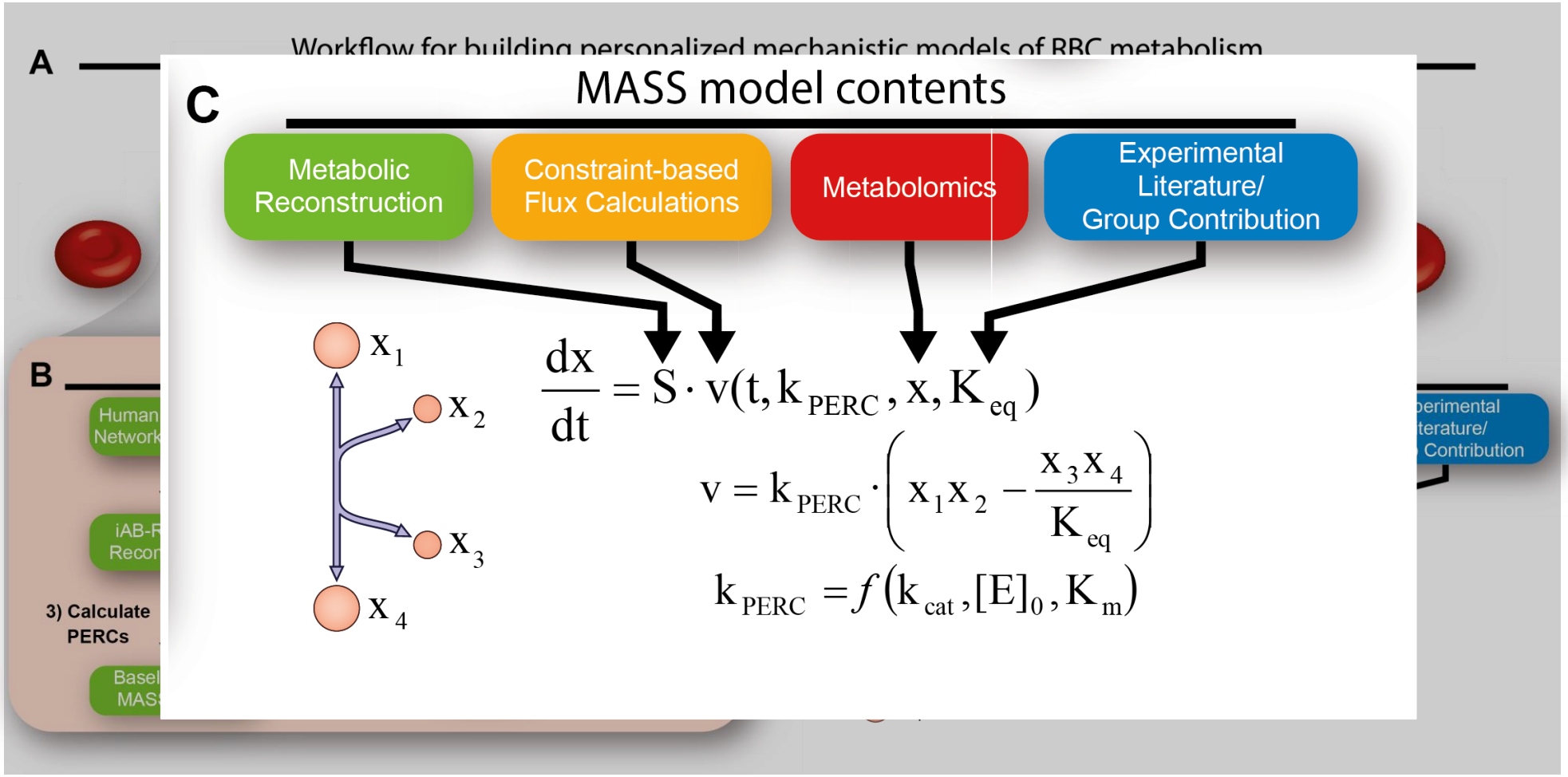
Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

Case Study (2)



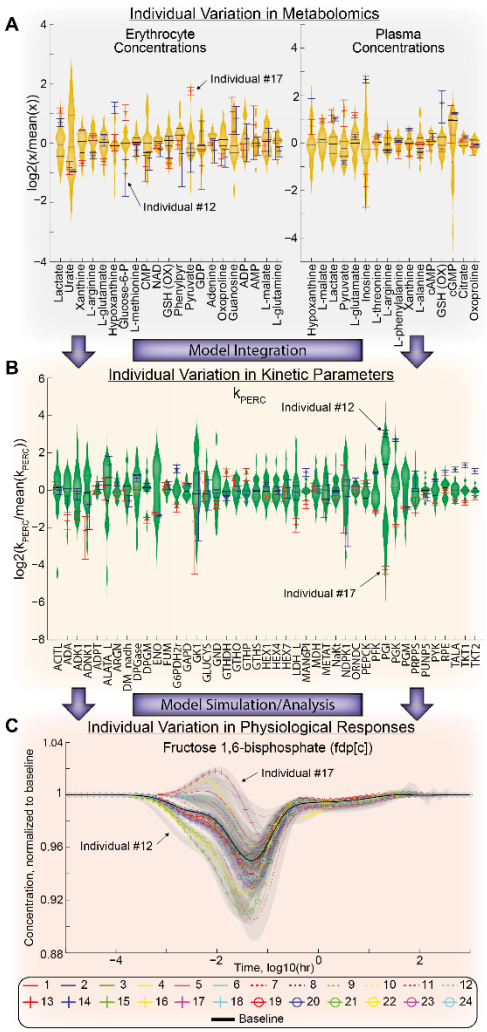
Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

Case Study (2)



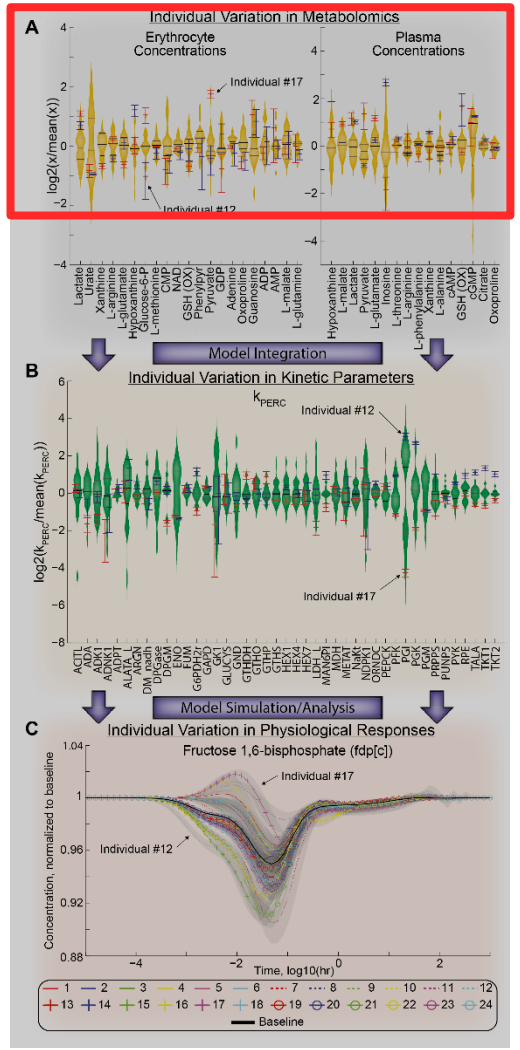
Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

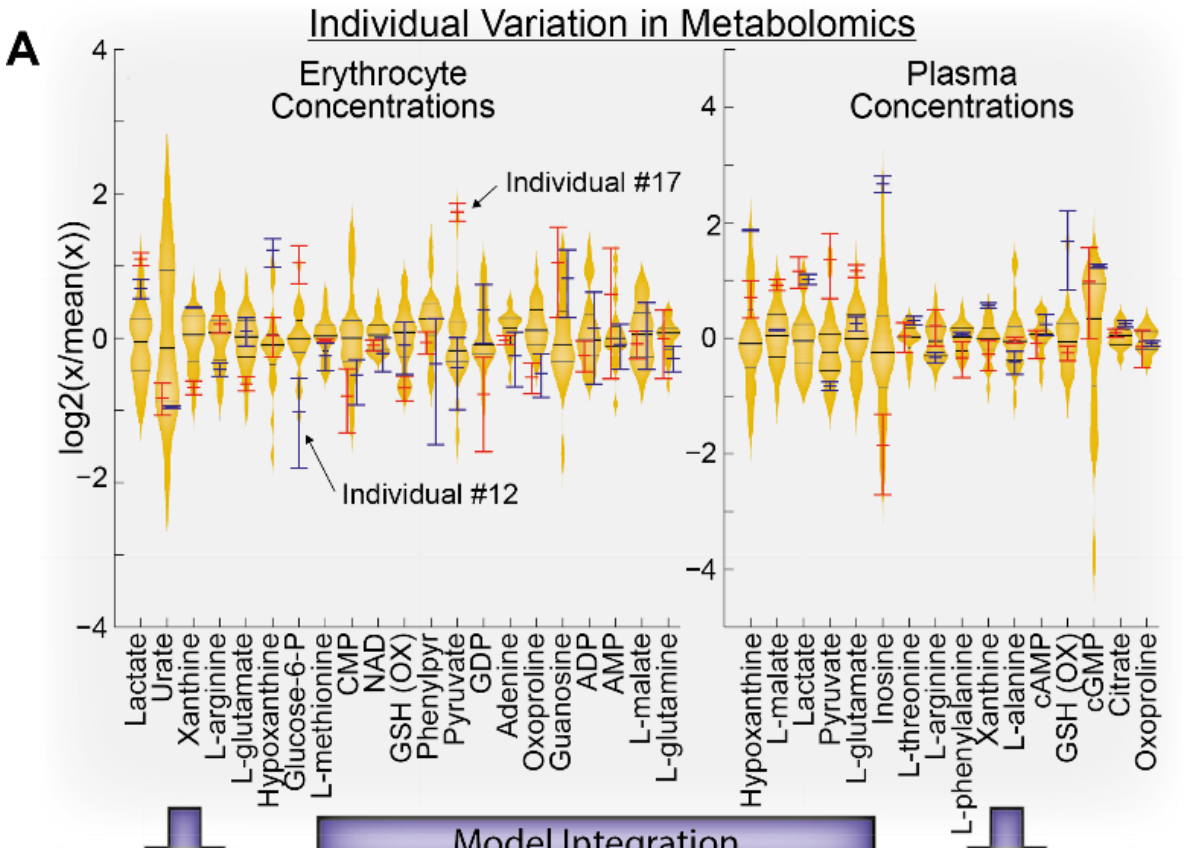
Case Study (2)



Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

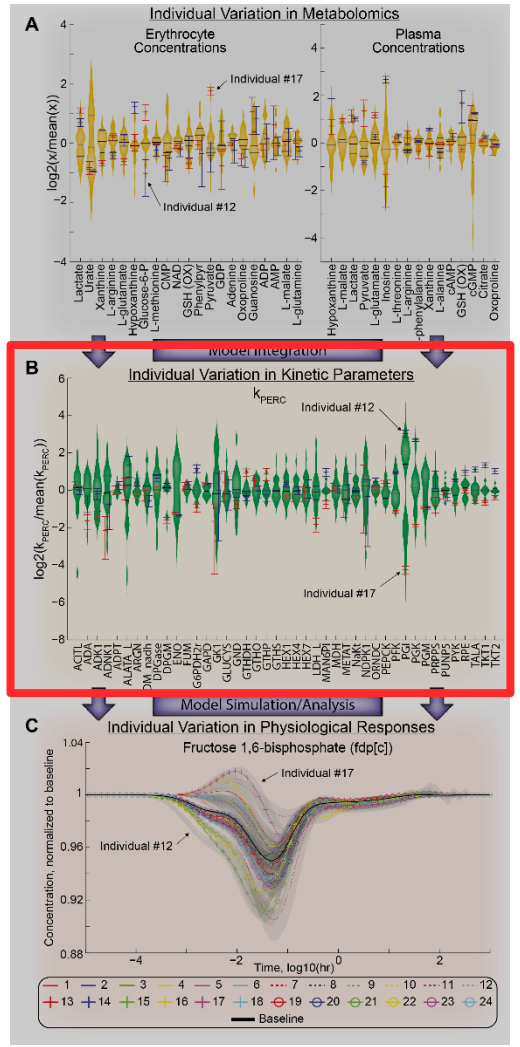
Case Study (2)

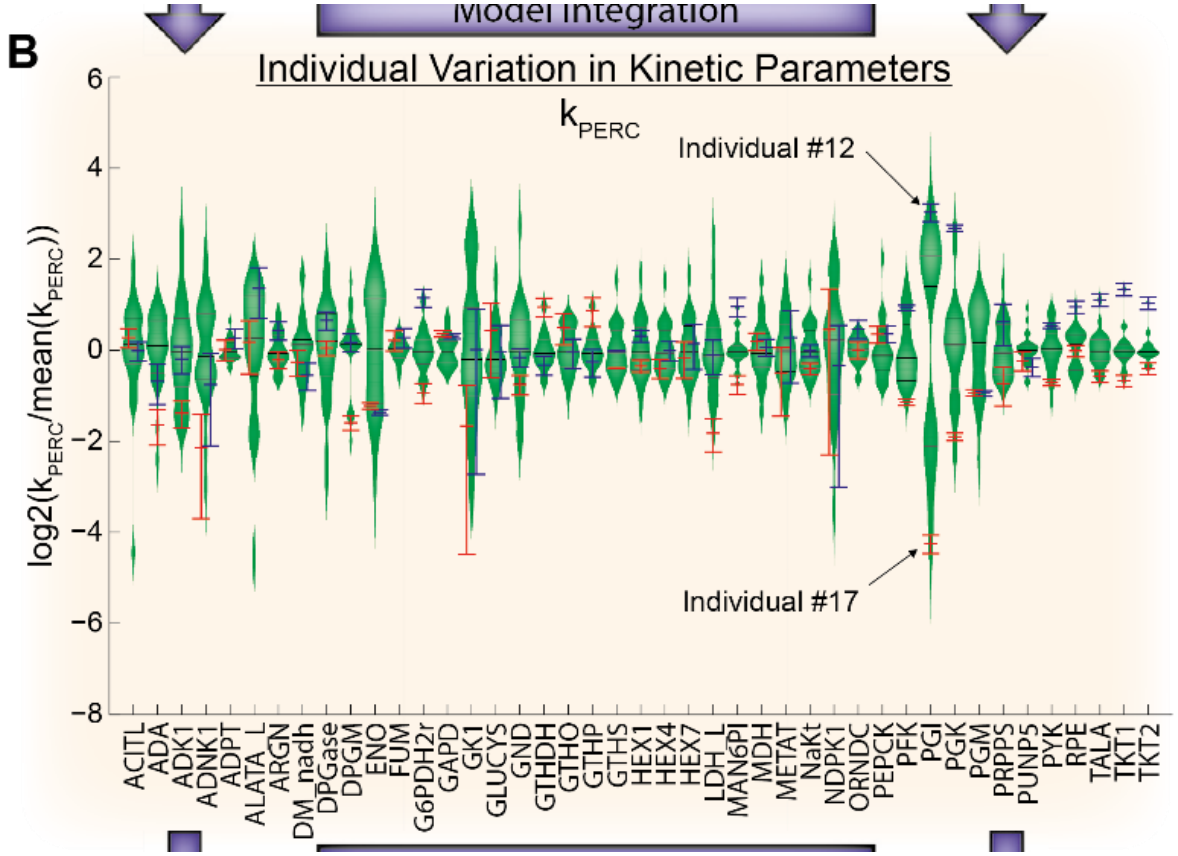




Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

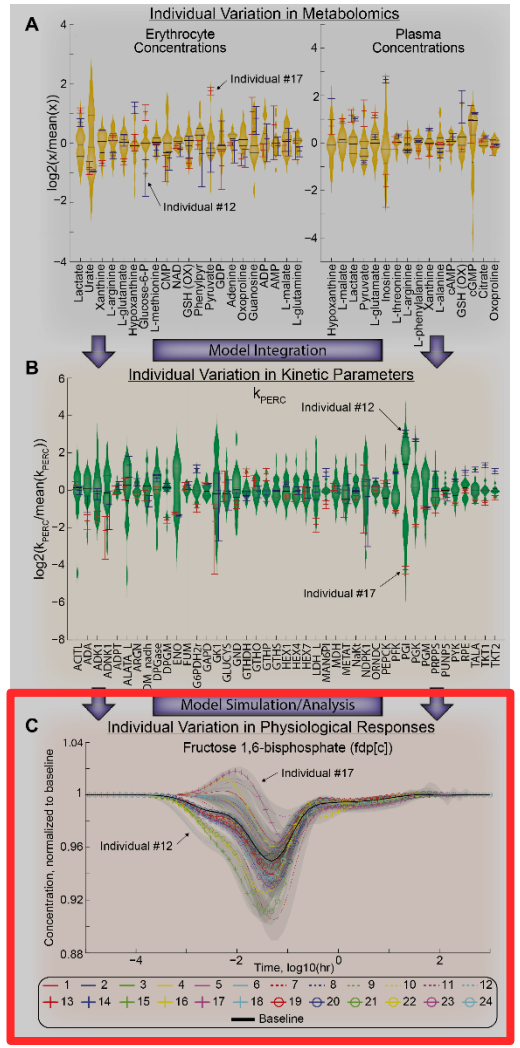
Case Study (2)

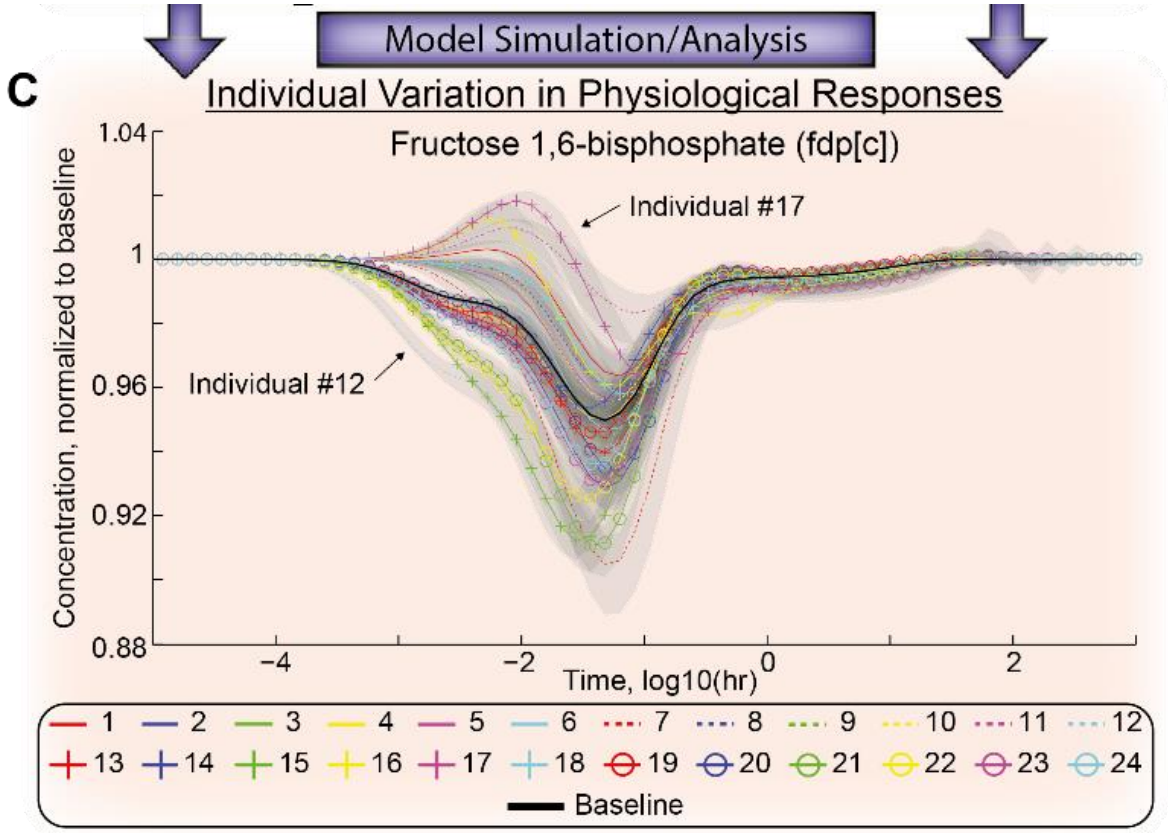




Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

Case Study (2)



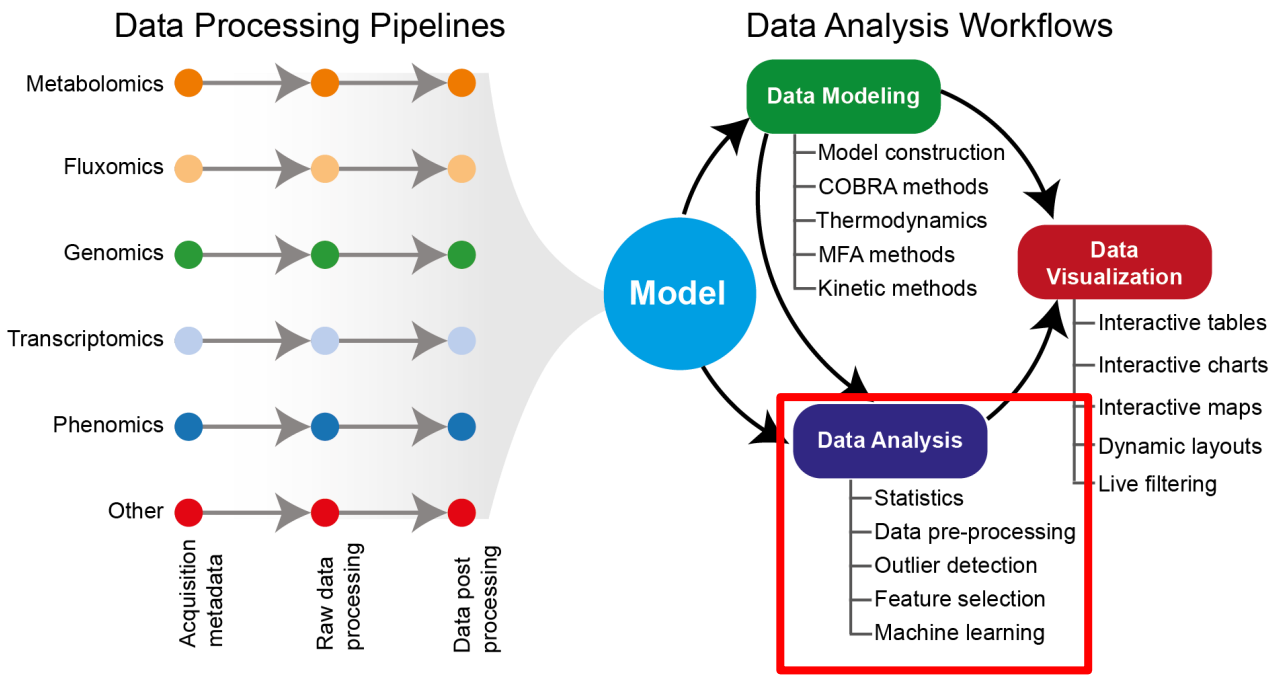


Bordbar A, McCloskey D, Zielinski D, Sonnenschein N, Jamshidi N, Palsson BO (2015). Cell Systems. http://dx.doi.org/10.1016/j.cels.2015.10.003

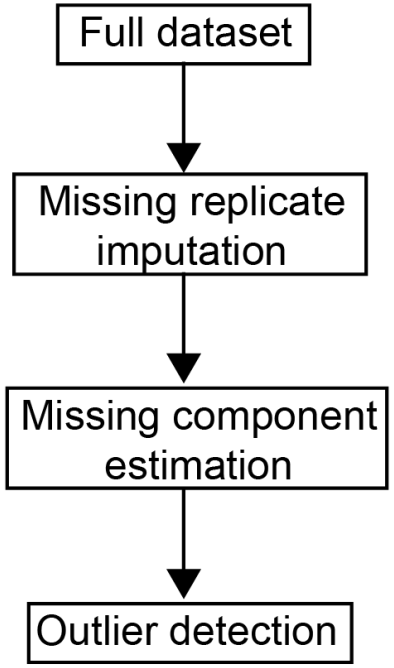
Statistics

Statistics (1) data imputation, normalization and exploration

Statistics (2) profiles and correlation

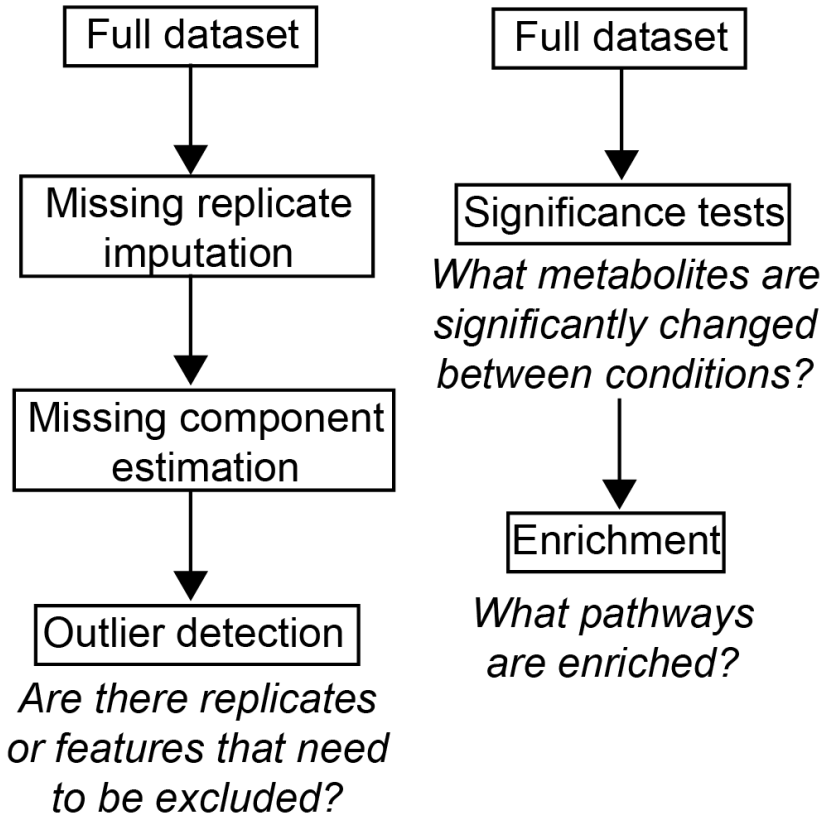


Statistics (1) data imputation, normalization and exploration

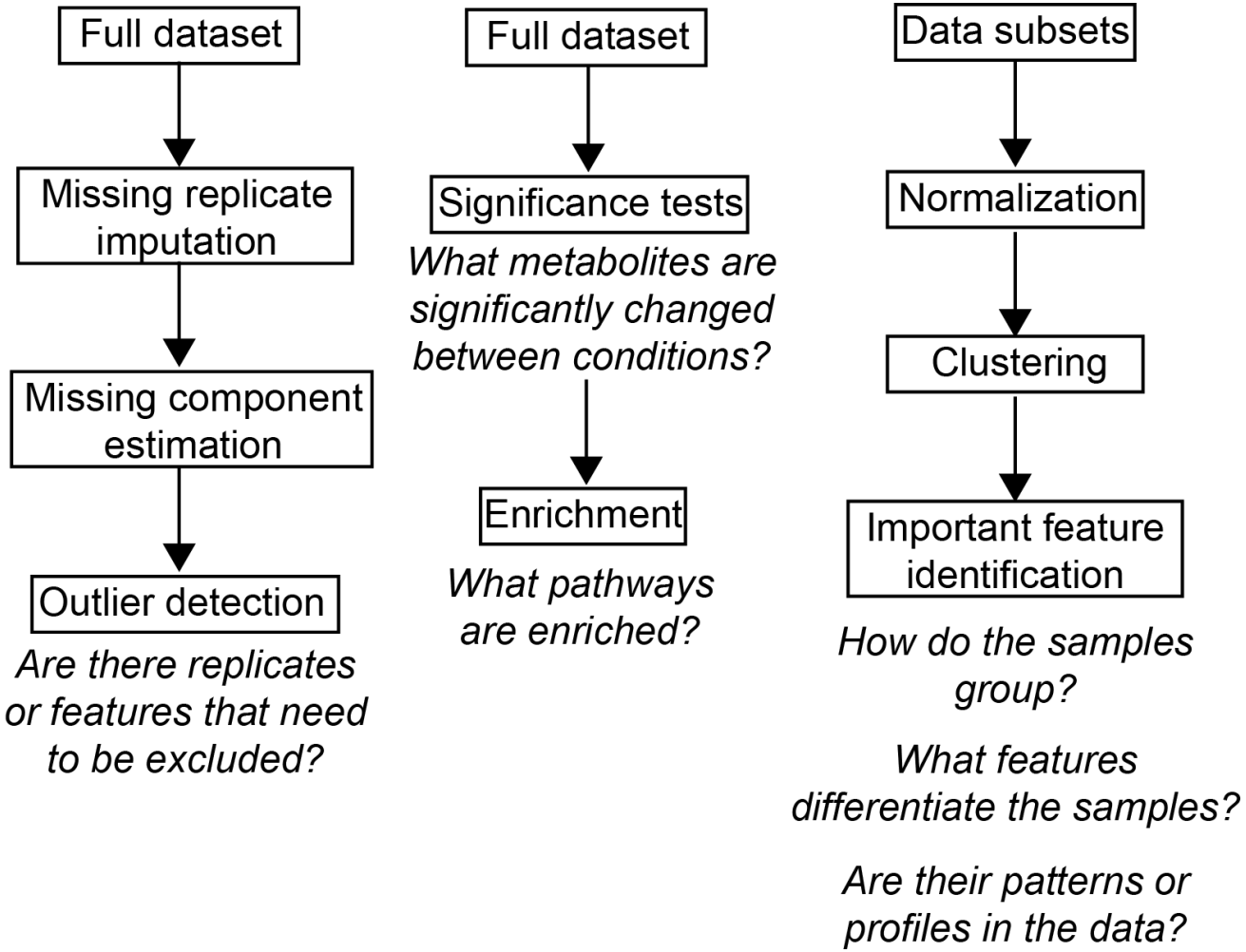


Are there replicates or features that need to be excluded?

Statistics (1) data imputation, normalization and exploration



Statistics (1) data imputation, normalization and exploration



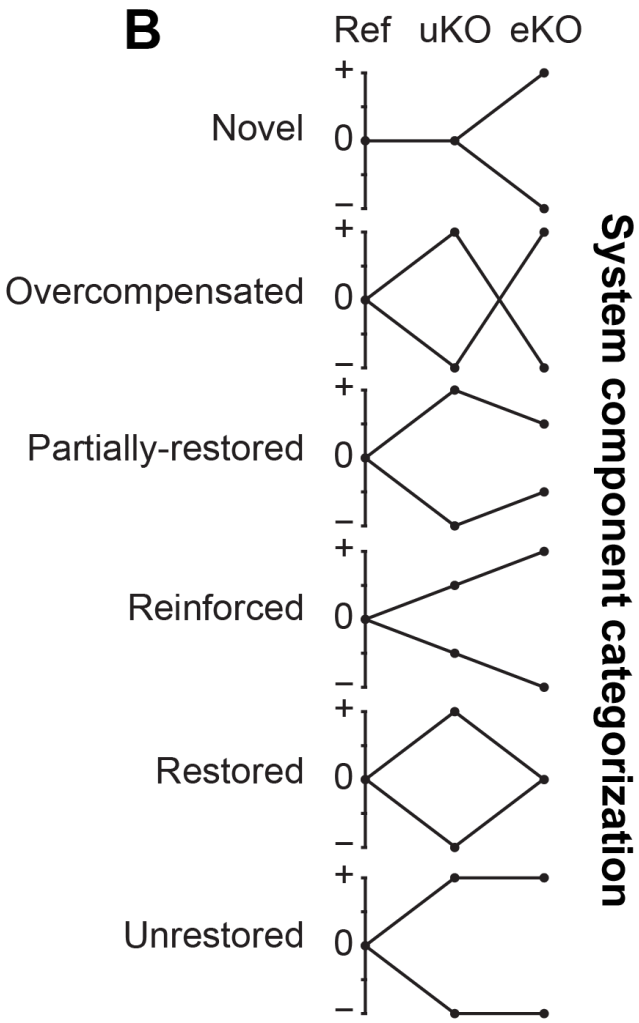
Statistics (2) profiles and correlation

- Provides an unbaised means to compare different –omics data across time or experimental conditions that can be independent of known biochemical mechanisms

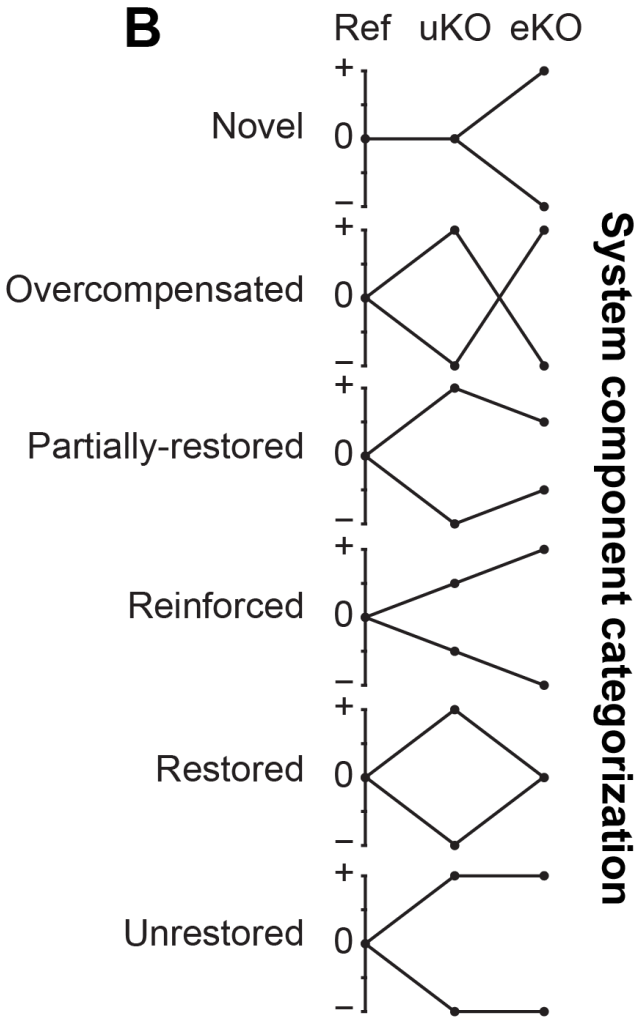
Statistics (2) profiles and correlation

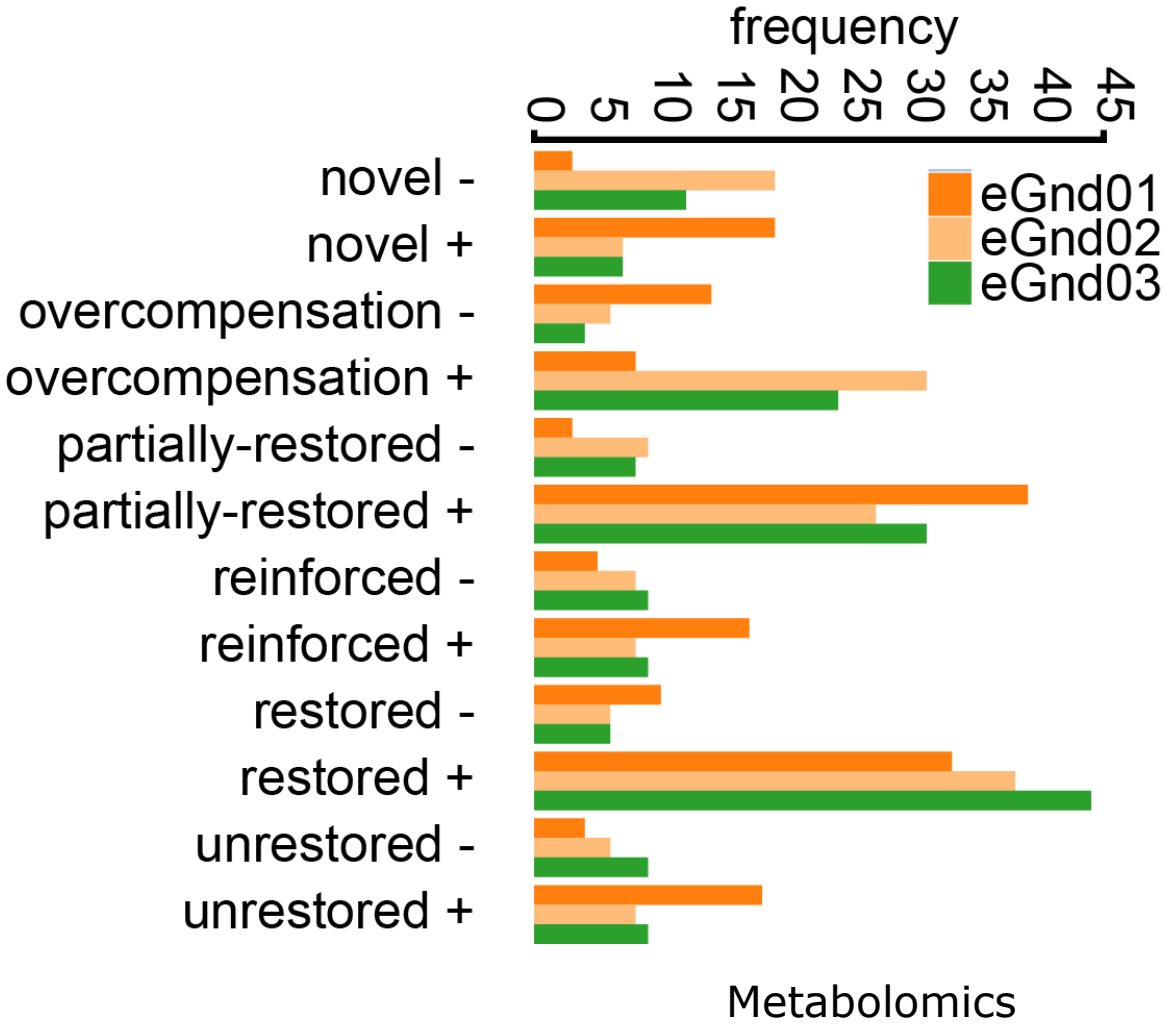
- Provides an unbaised means to compare different –omics data across time or experimental conditions that can be independent of known biochemical mechanisms
- First step towards greater mechanistic understanding of biology

Case Study (1)

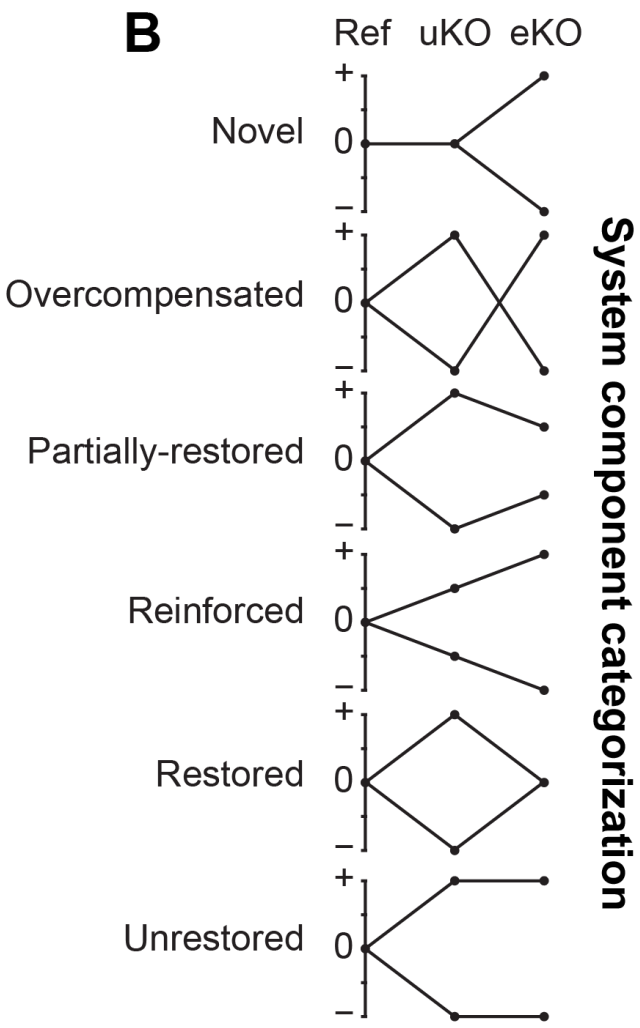

B
Ref uKO eKO
Novel
Overcompensated
Partially-restored
Reinforced
Restored
Unrestored
+
0
–
System component categorization

Case Study (1)


B
Ref uKO eKO
Novel
Overcompensated
Partially-restored
Reinforced
Restored
Unrestored
+
0
–
System component categorization

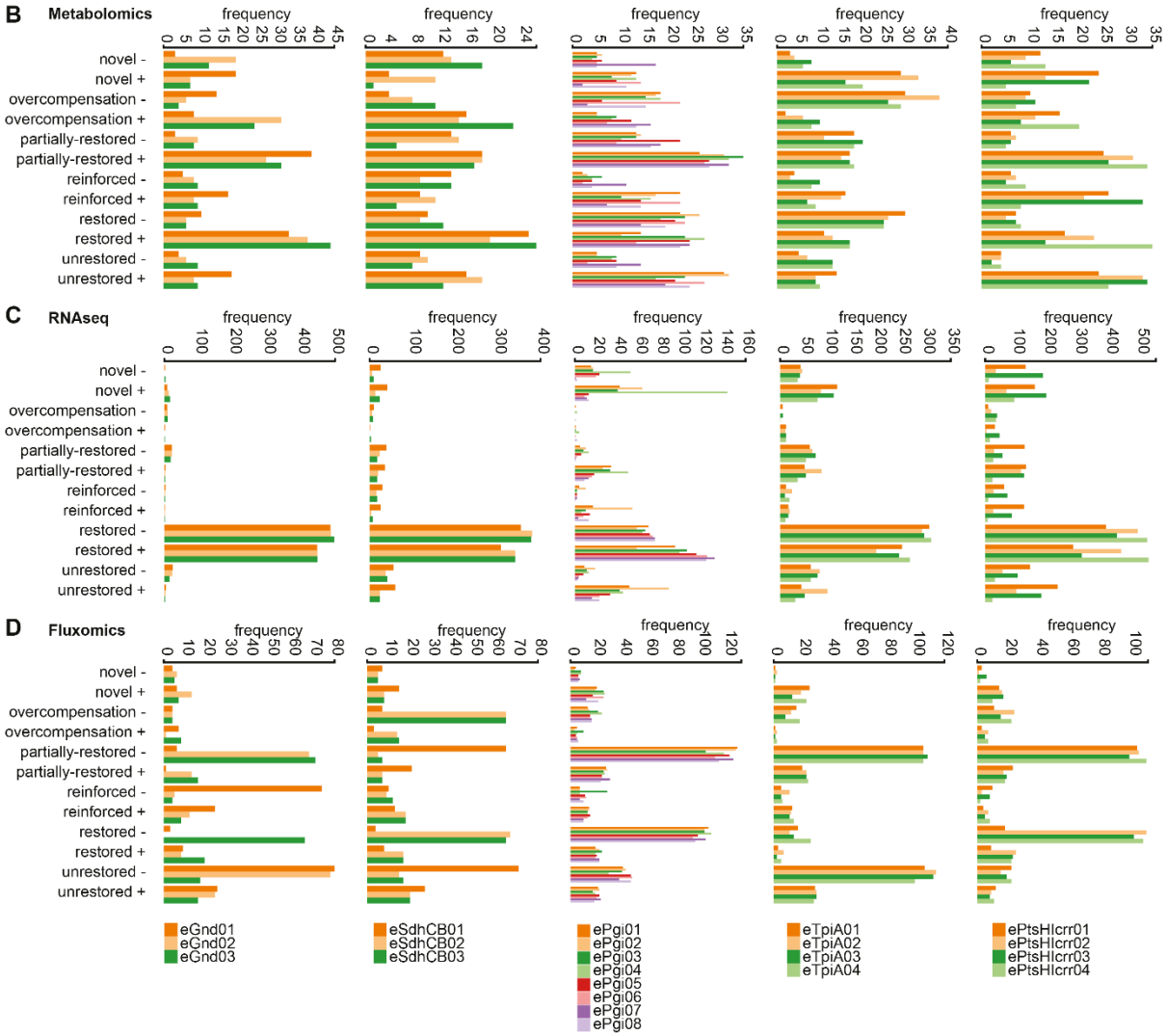

frequency
0 5 10 15 20 25 30 35 40 45
novel -
novel +
overcompensation -
overcompensation +
partially-restored -
partially-restored +
reinforced -
reinforced +
restored -
restored +
unrestored -
unrestored +
eGnd01
eGnd02
eGnd03
Metabolomics

Case Study (1)

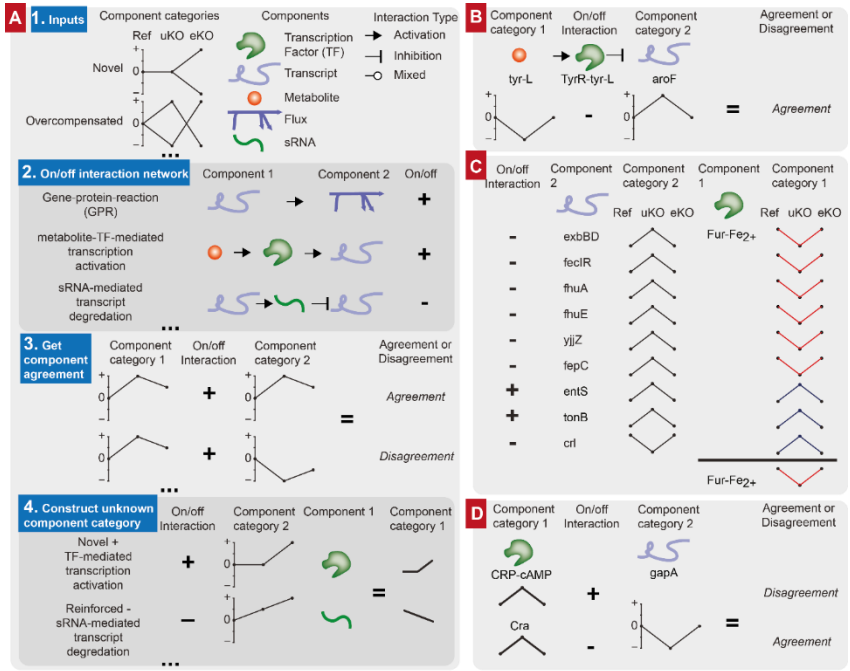

B
Ref uKO eKO
Novel
Overcompensated
Partially-restored
Reinforced
Restored
Unrestored
+
0
–
System component categorization

frequency
0
50
100
150
200
250
300
350
400
450
500
novel -
novel +
overcompensation -
overcompensation +
partially-restored -
partially-restored +
reinforced -
reinforced +
restored -
restored +
unrestored -
unrestored +
eGnd01
eGnd02
eGnd03
Transcriptomics

Case Study (1)

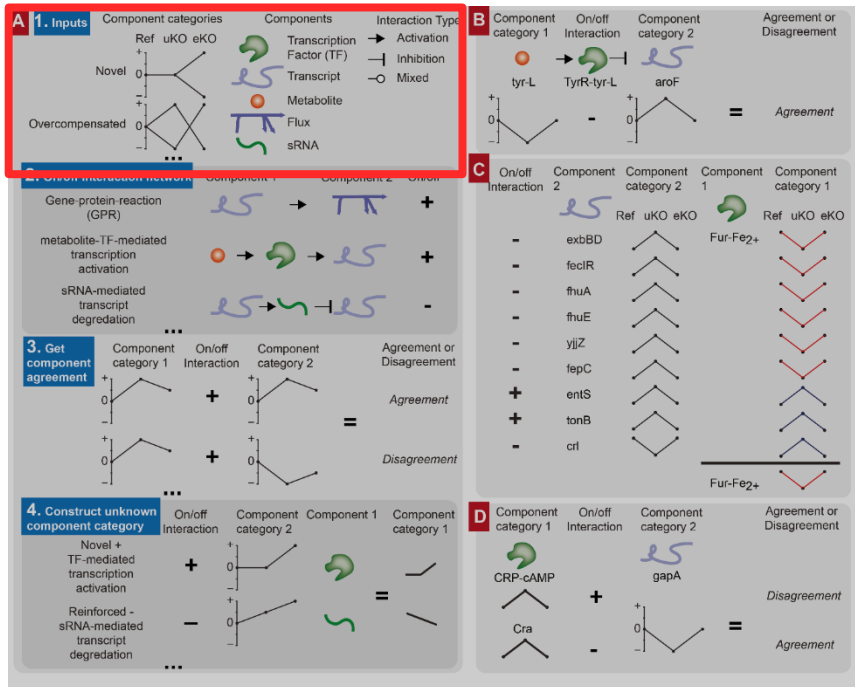

B Metabolomics
C RNAseq
D Fluxomics
frequency
novel -
novel +
overcompensation -
overcompensation +
partially-restored -
partially-restored +
reinforced -
reinforced +
restored -
restored +
unrestored -
unrestored +
eGnd01
eGnd02
eGnd03
eSdhCB01
eSdhCB02
eSdhCB03
ePgi01
ePgi02
ePgi03
ePgi04
ePgi05
ePgi06
ePgi07
ePgi08
eTpiA01
eTpiA02
eTpiA03
eTpiA04
ePtsHIcrr01
ePtsHIcrr02
ePtsHIcrr03
ePtsHIcrr04

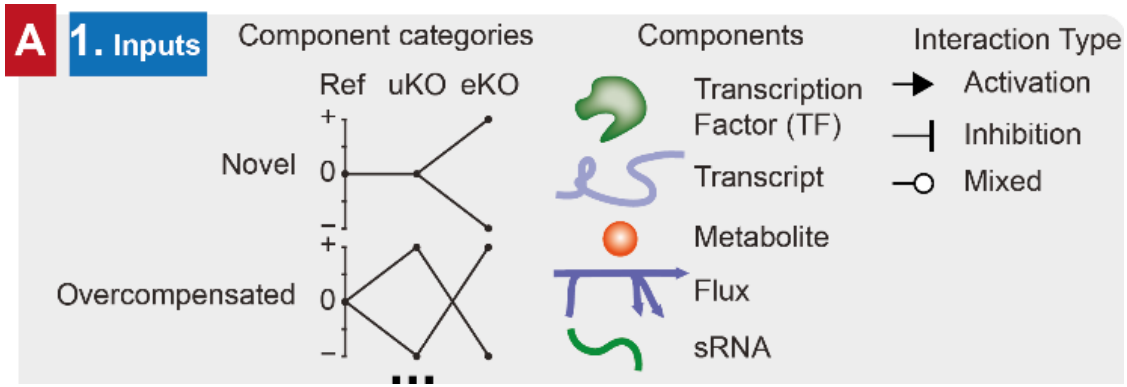
Case study (1)


A
1. Inputs
Component categories
Components
Interaction Type
Ref uKO eKO
Novel
Overcompensated
Transcription Factor (TF)
Transcript
Metabolite
Flux
sRNA
Activation
Inhibition
Mixed
2. On/off interaction network
Component 1
Component 2
On/off
Gene-protein-reaction (GPR)
metabolite-TF-mediated transcription activation
sRNA-mediated transcript degredation
3. Get component agreement
Component category 1
On/off Interaction
Component category 2
Agreement or Disagreement
Agreement
Disagreement
4. Construct unknown component category
On/off Interaction
Component category 2
Component 1
Component category 1
Novel + TF-mediated transcription activation
Reinforced - sRNA-mediated transcript degredation
B
Component category 1
On/off Interaction
Component category 2
Agreement or Disagreement
tyr-L
TyrR-tyr-L
aroF
Agreement
C
On/off Interaction
Component 2
Component category 2
Component 1
Component category 1
Ref uKO eKO
exbBD
fecIR
fhuA
fhuE
yjiZ
fepC
entS
tonB
crl
Fur-Fe2+
D
Component category 1
On/off Interaction
Component category 2
Agreement or Disagreement
CRP-cAMP
gapA
Cra
Disagreement
Agreement

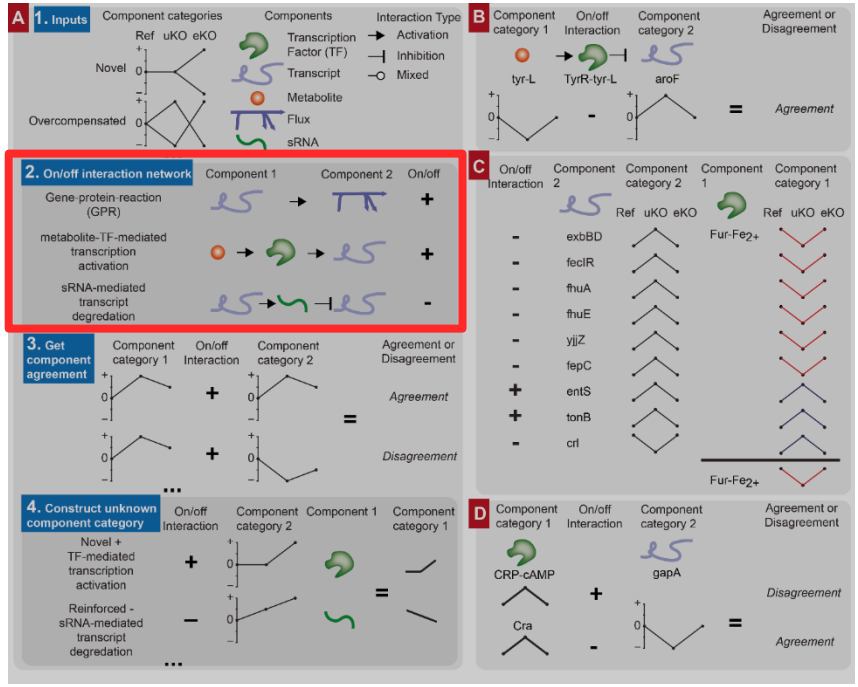
Case study (1)

1. Inputs

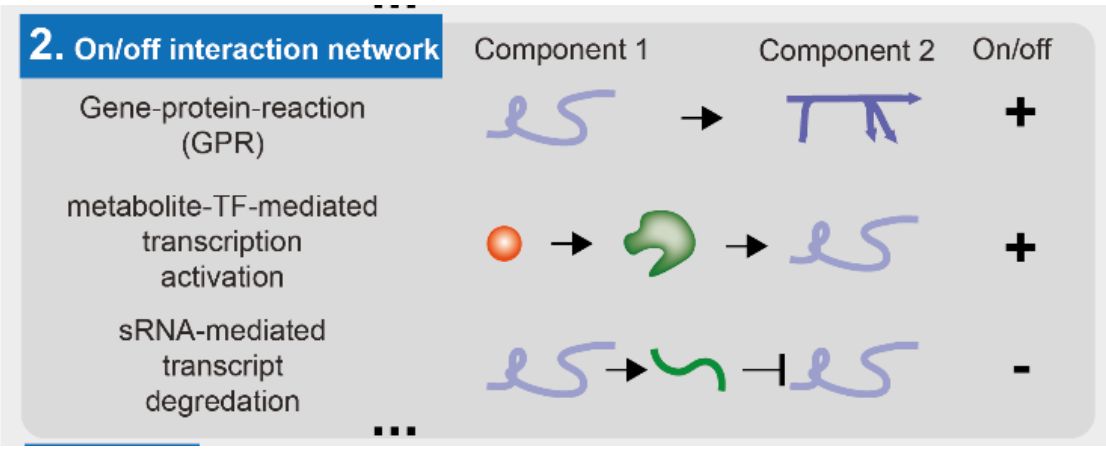




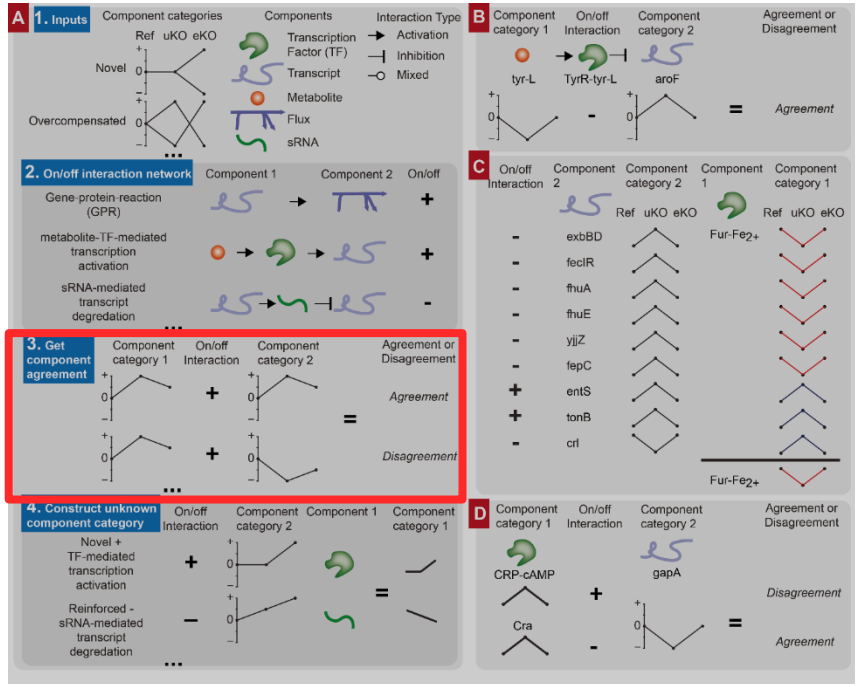
Case study (1)



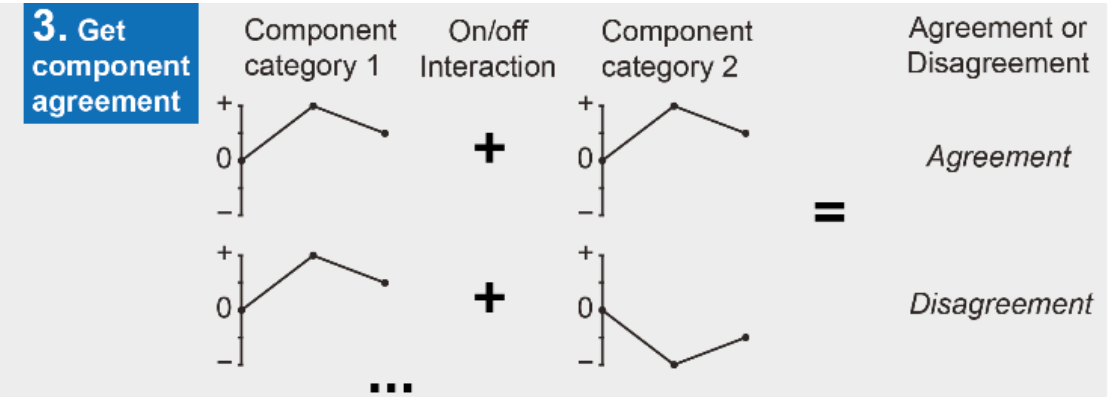
2. Create an on/off interaction network between all metabolic and regulatory components



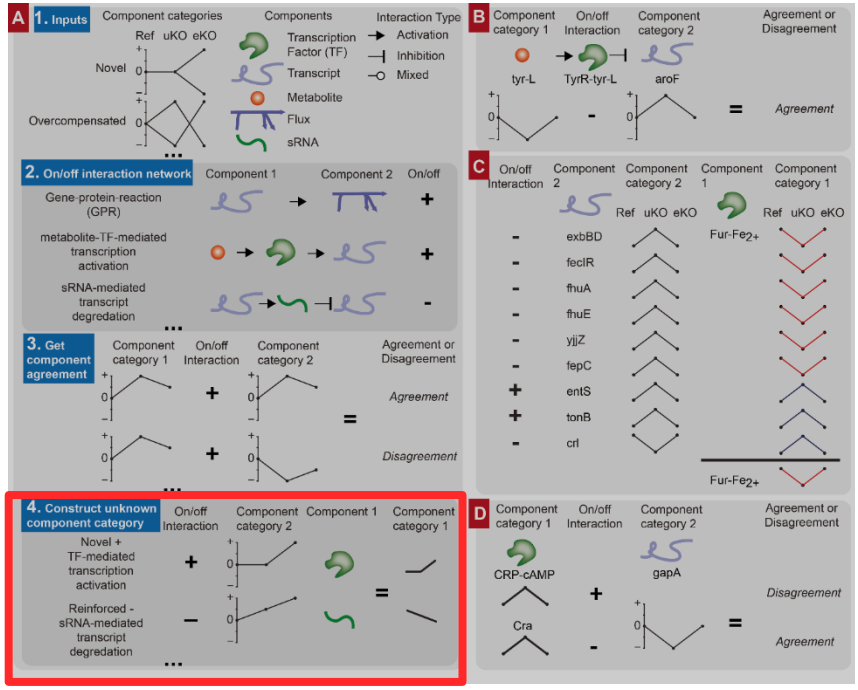
Case study (1)



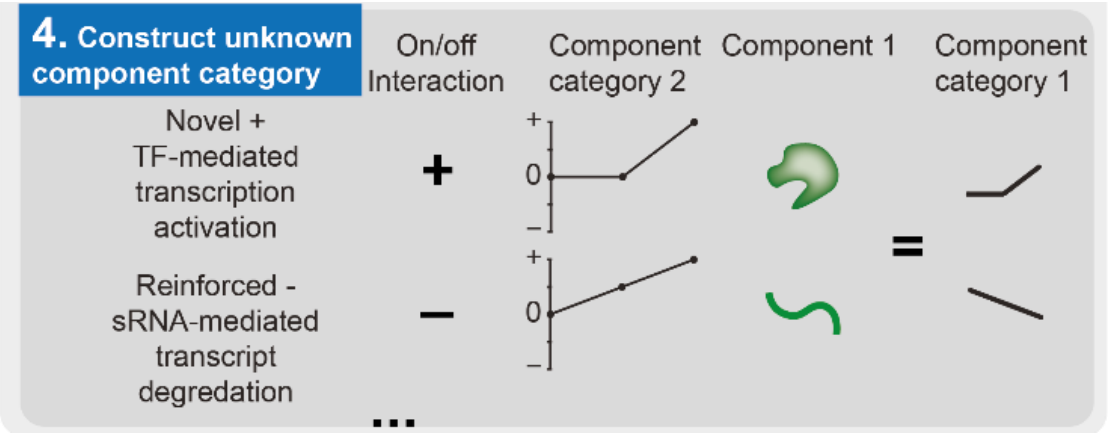
3. Determine the agreement between component profiles and known regulatory/metabolic interactions



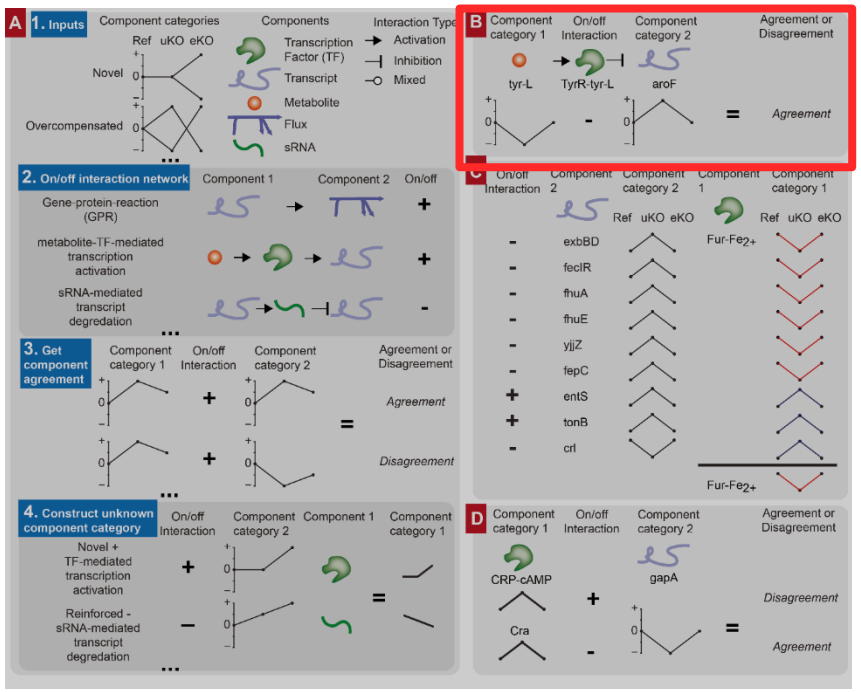
Case study (1)



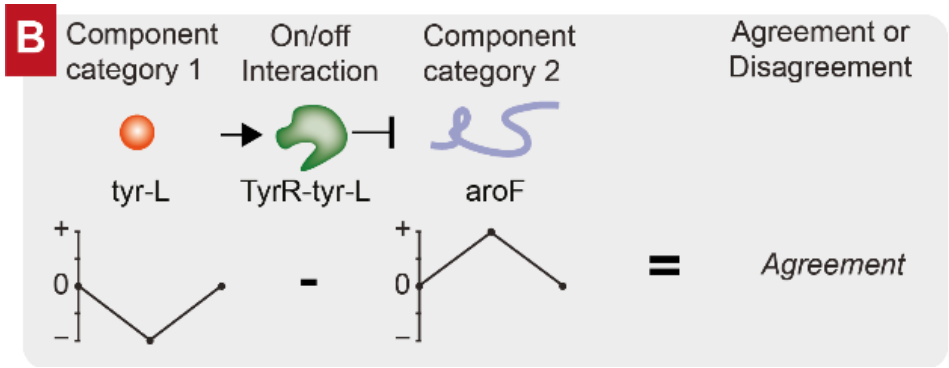
4. Construct unknown component category from known component categories



Case study (1)

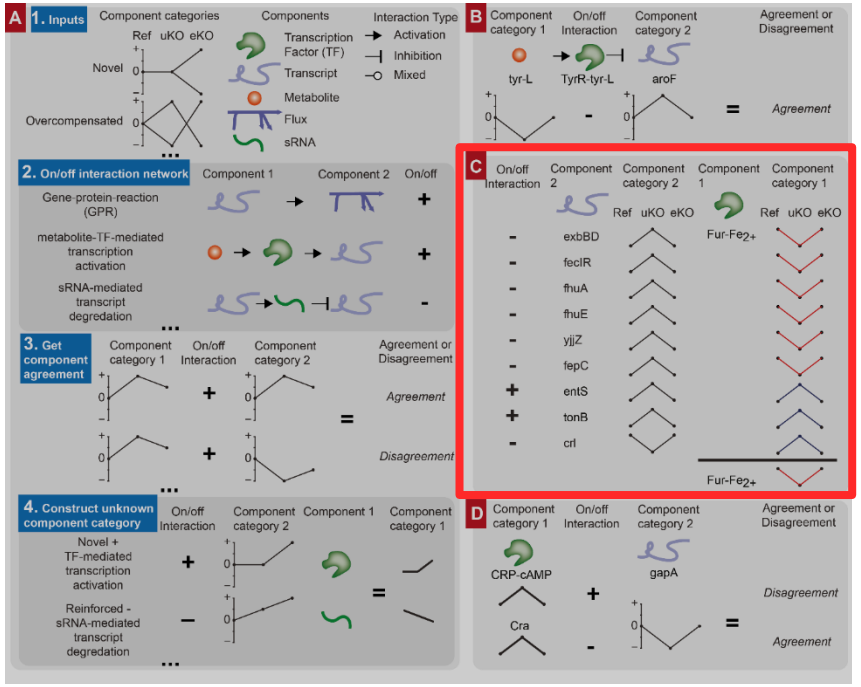


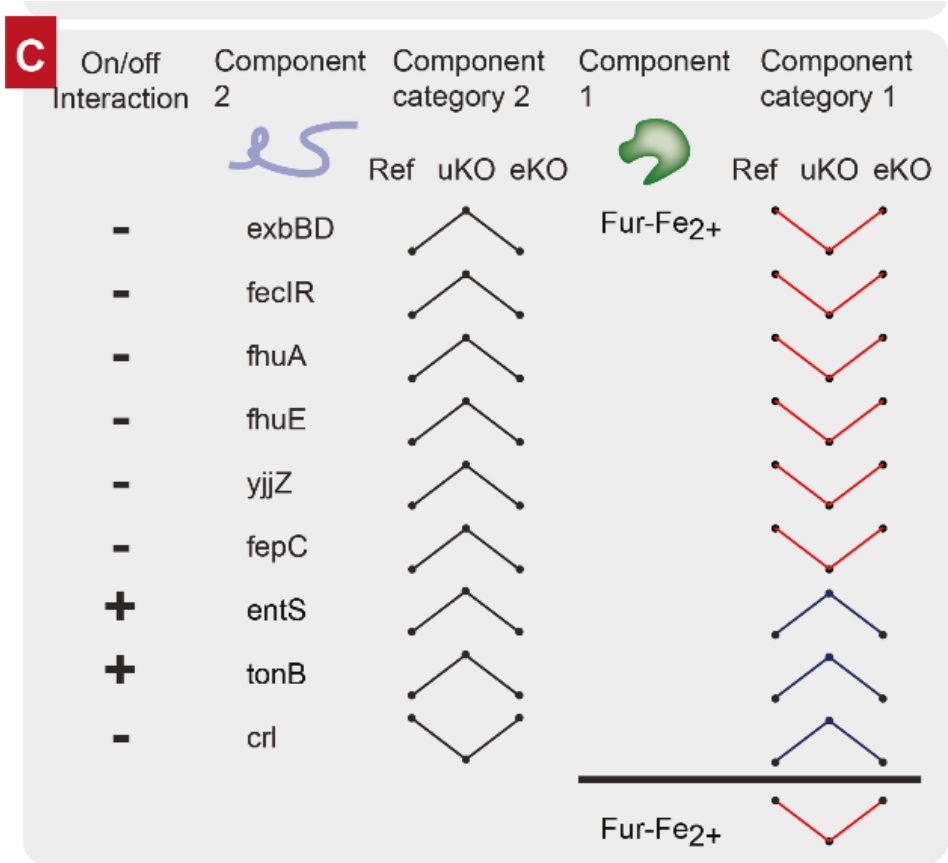
Calculate the activity of unmeasured components (i.e., transcription factor activation/deactivation status) based on expression of target operons and genes



Case study (1)

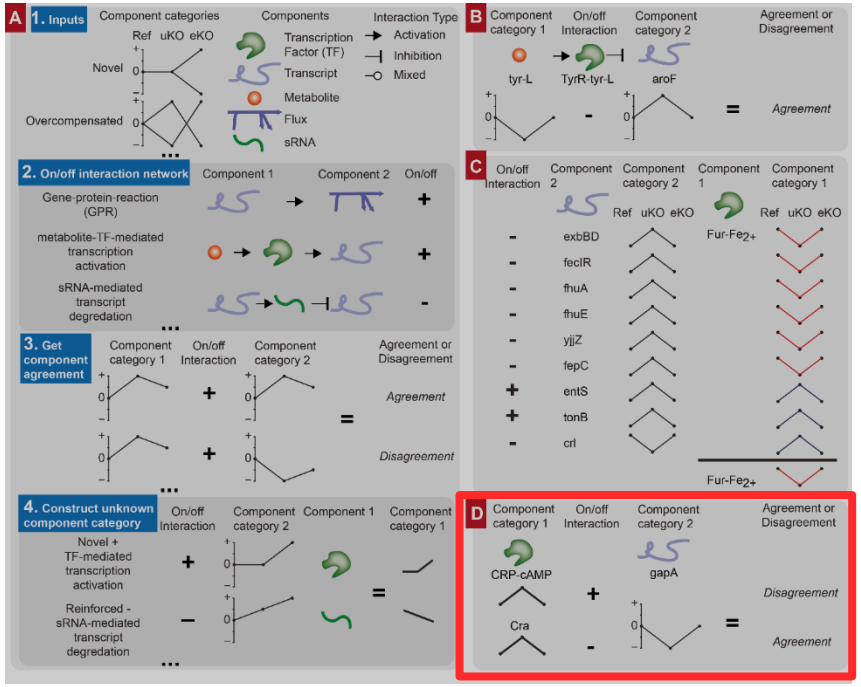
What is the consensus activation profile for a given regulator based on target gene expression profiles?

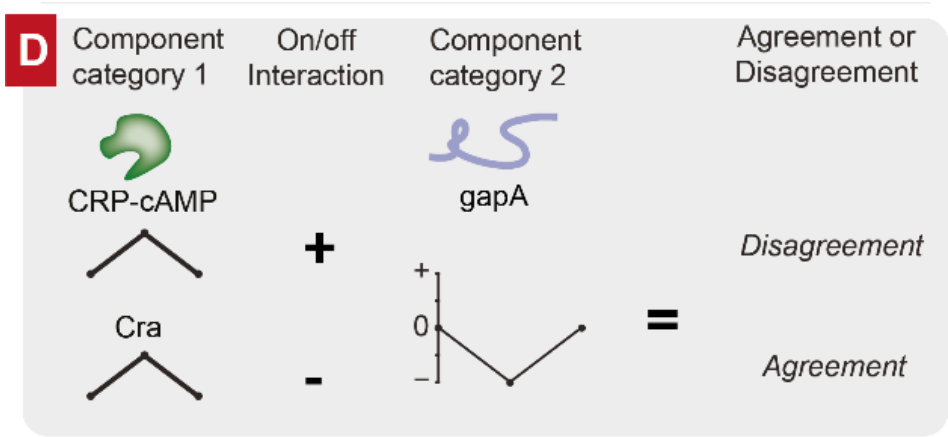




Case study (1)

What is the agreement between measured data and literature derived biochemical interaction networks?

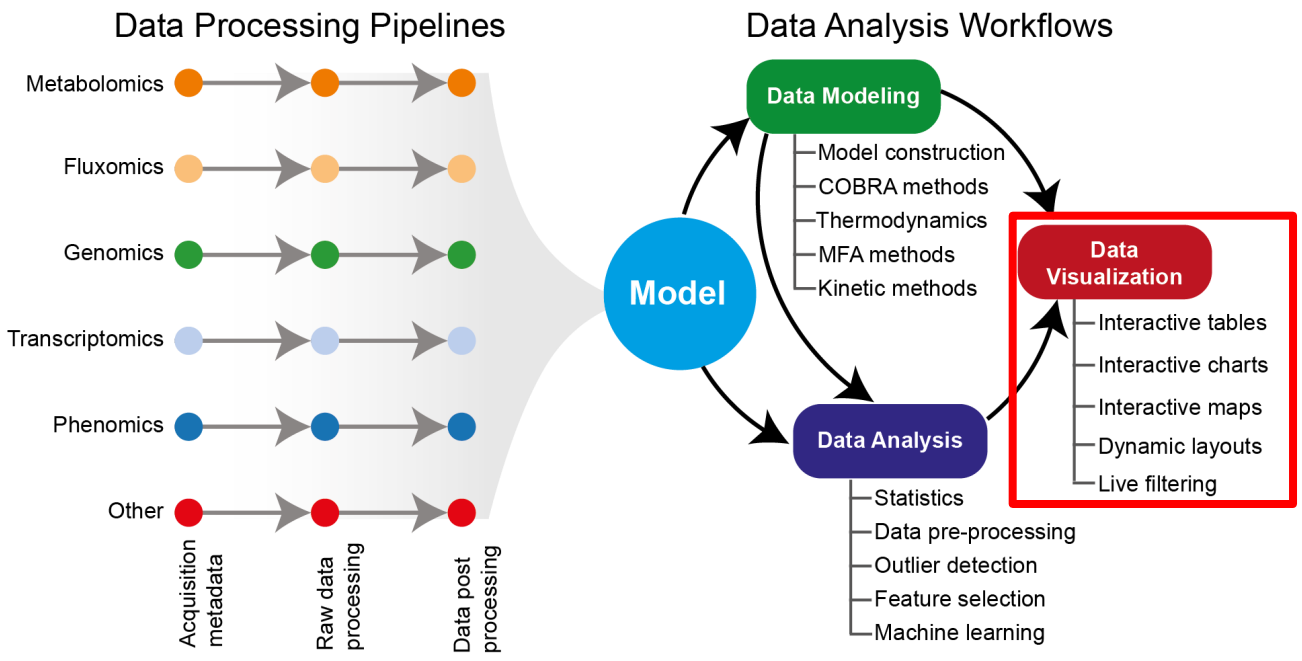




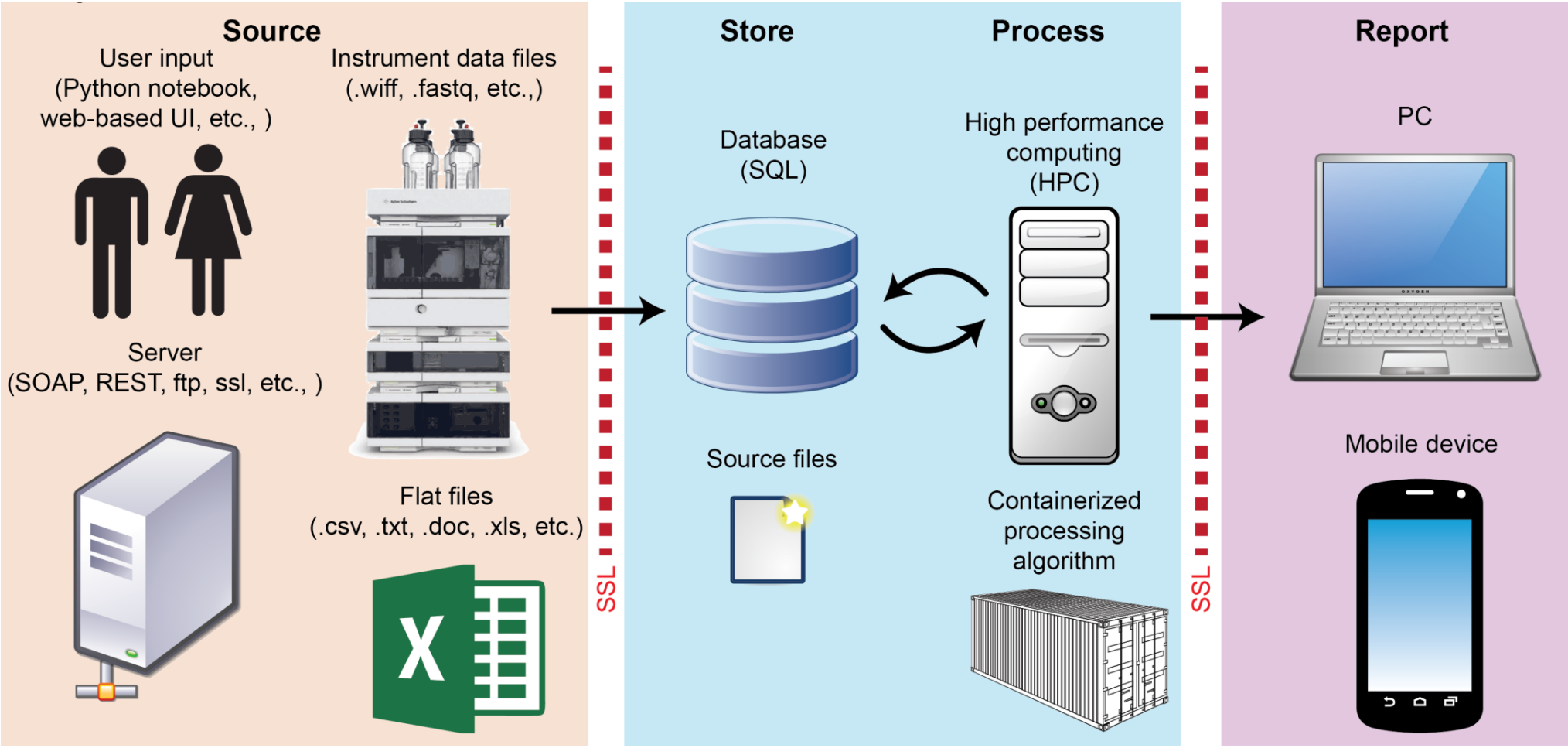
Visualization and sharing

Visualization (1) from static figures to dynamic dashboards

Visualization (2) view sharing



Visualization (1) from static figures to dynamic dashboards



Visualization (1) from static figures to dynamic dashboards

- Deployment
- Database and application
- Webserver
- Browser-based GUI

Visualization (1) from static figures to dynamic dashboards

- Software requirements
  - Database
  - Web server
  - Browser based rendering

Visualization (1) from static figures to dynamic dashboards

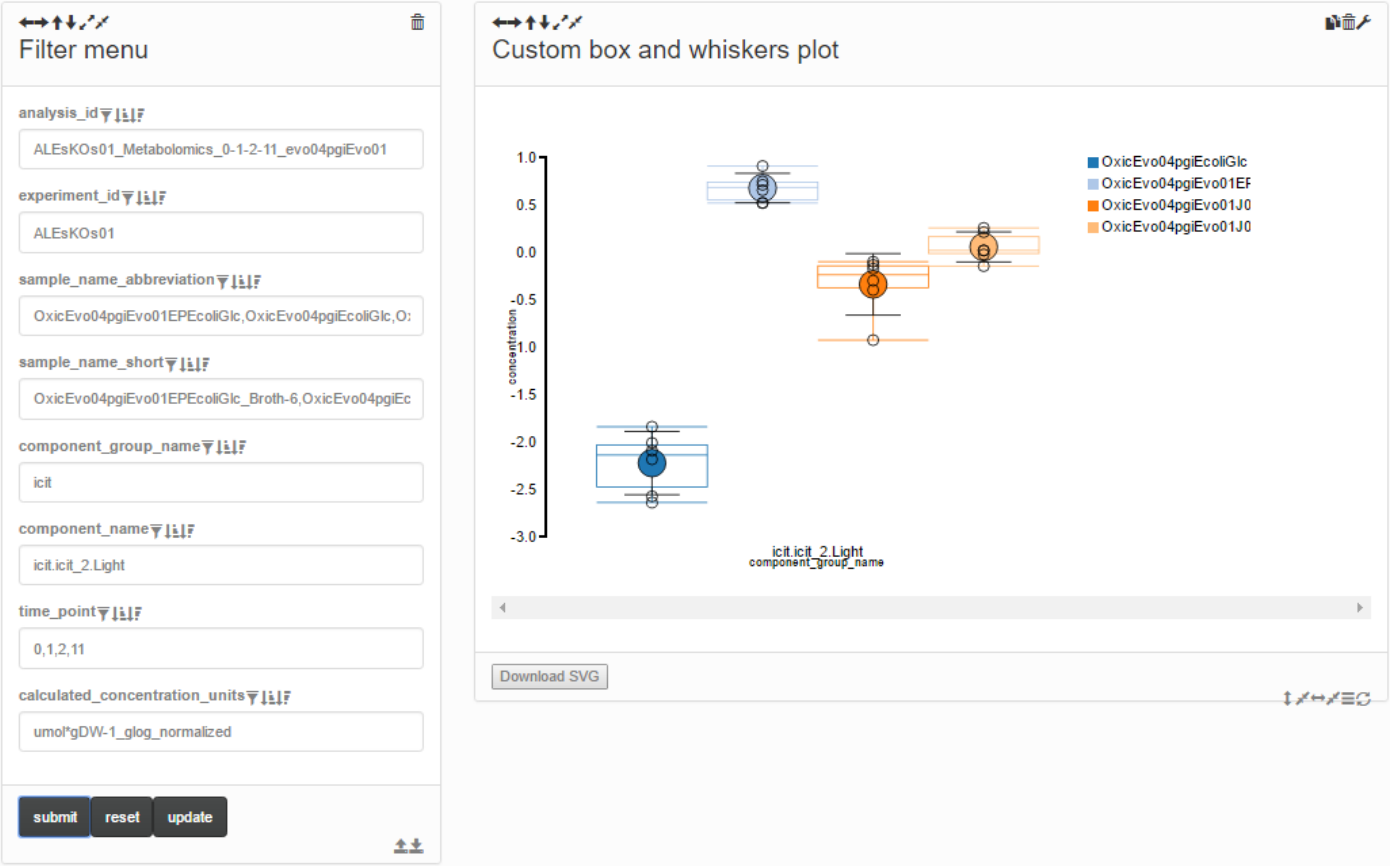
- Software requirements
  - Database
  - Web server
  - Browser based rendering
- Dashboard requirements
  - Plotting library
  - View and modify data in table views
  - Filtering and sorting capabilities
  - Real-time plot (and board) updates
  - Resizing and moving of dashboard elements

Visualization (2) view sharing

- Software requirements
  - Data format that describes the dashboard layout and applied filters
  - Encryption
  - Support for static link sharing or file sharing

Case Study (1/2/3)

- Browser-based dashboards and UIs based on a dynamic “tile” layout



- ✓ Filtering
- ✓ Moving tiles
- ✓ Messaging
- ✓ Sharing

Conclusion

Conclusion (1) Future trends

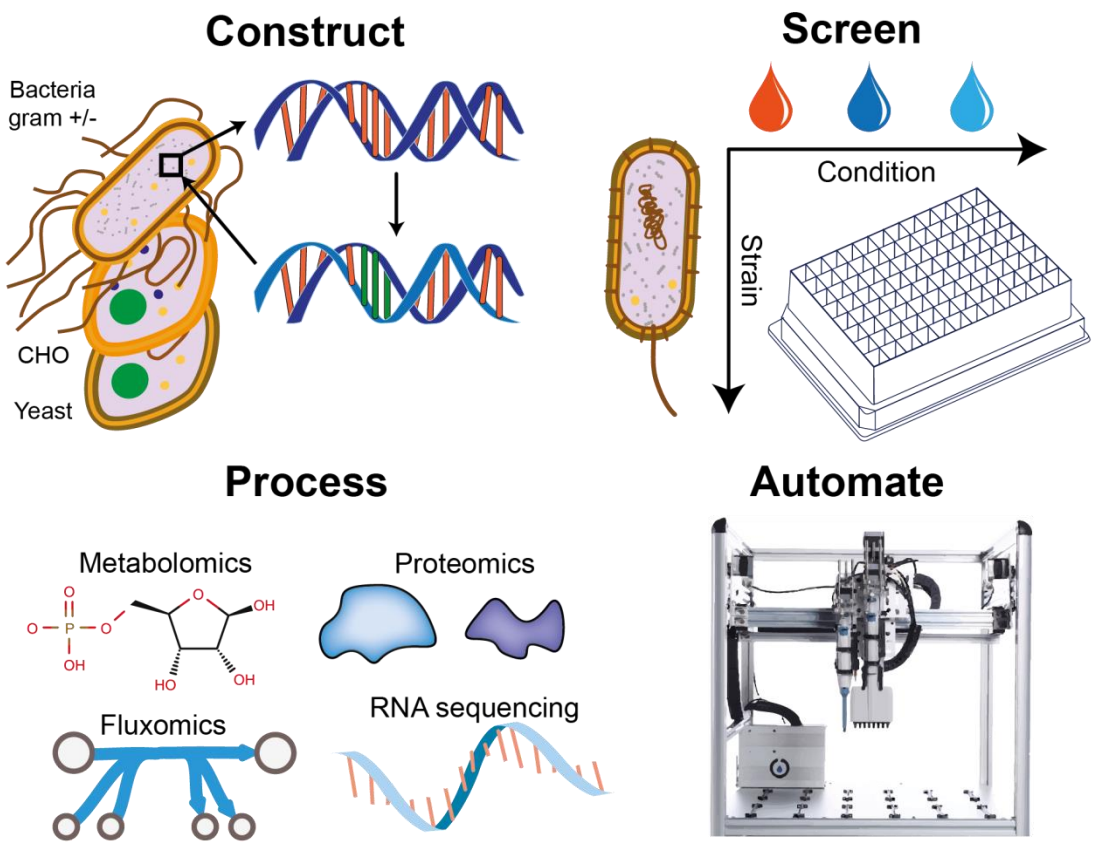
Conclusion (2) Re-cap

Conclusion (1) Future trends

- Automation
  - Sample handling
  - Data processing

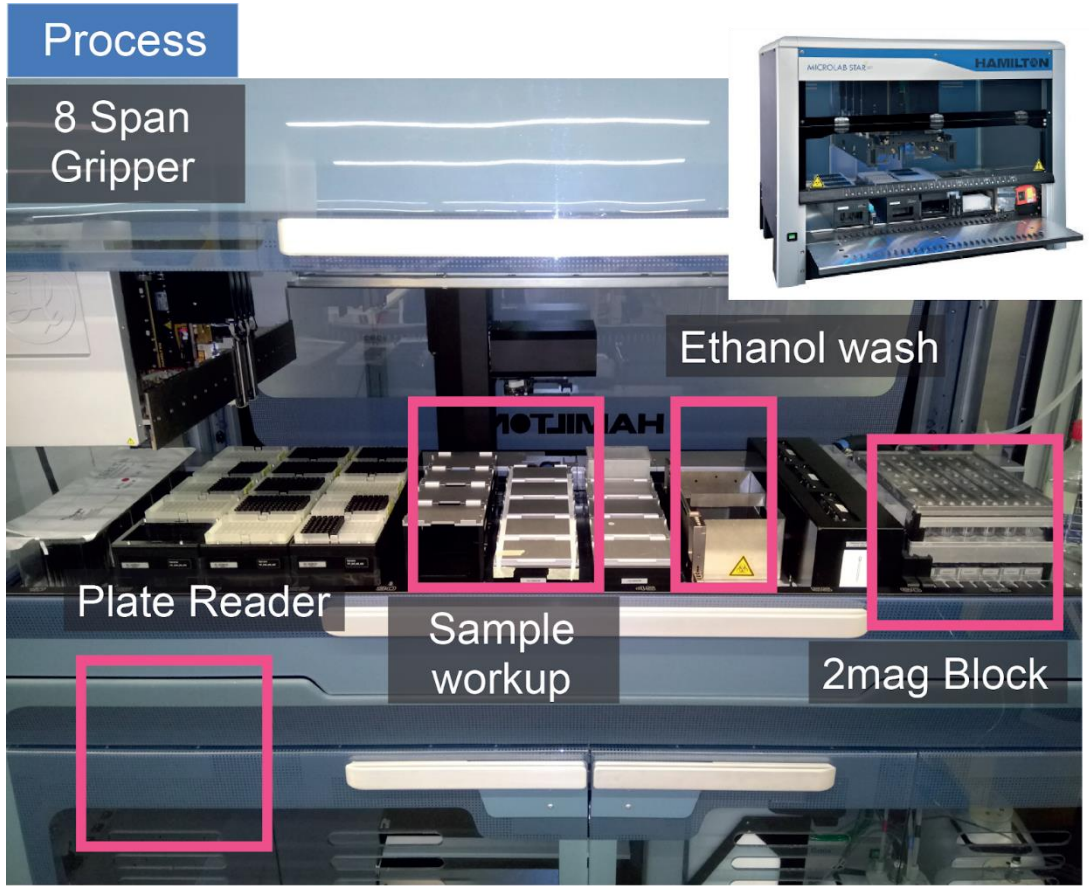
Conclusion (1) Future trends

- Automation
  - Sample handling
  - Data processing



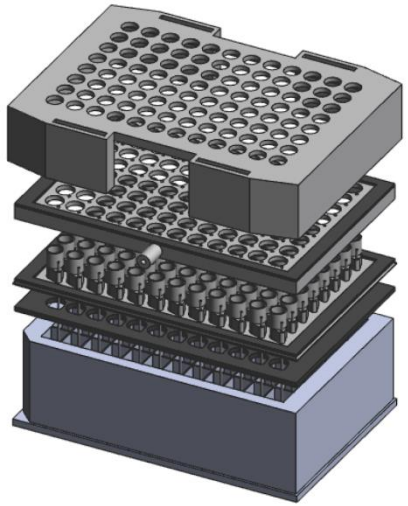
Conclusion (1) Future trends

- Automation
  - Sample handling
  - Data processing



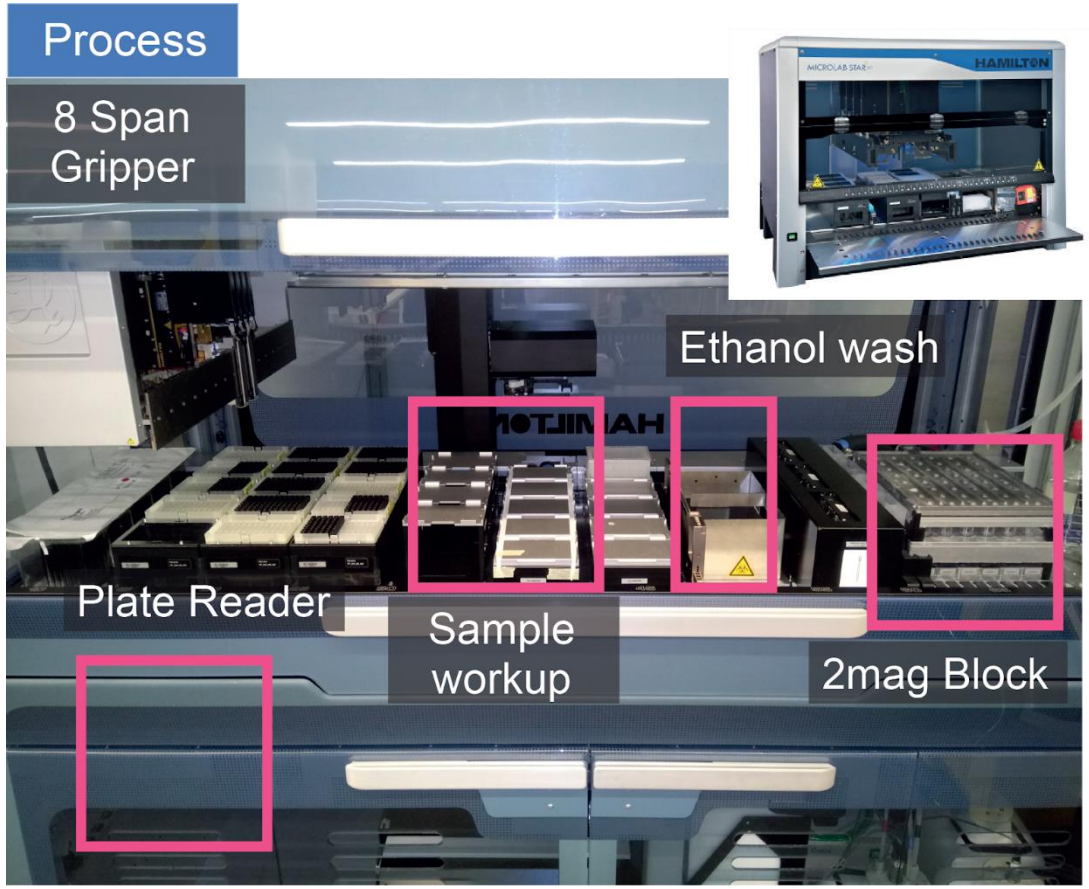
Hamilton Microlab Vantage Liquid Handling System

96 well HT cultivation



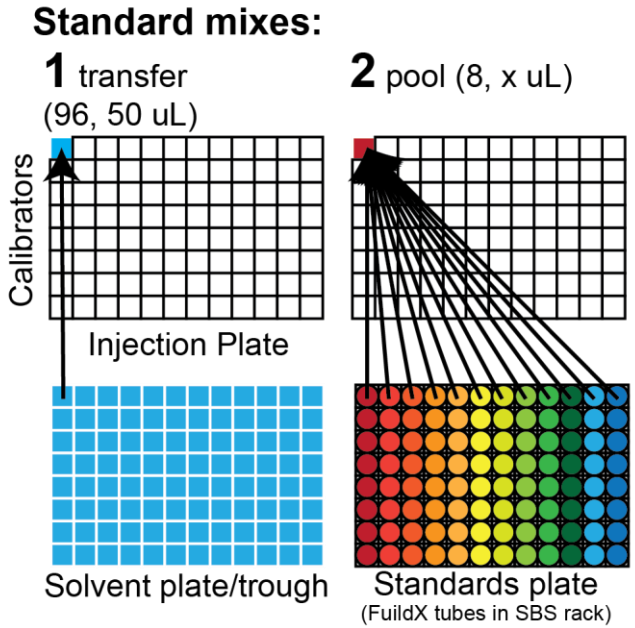
Conclusion (1) Future trends

- Automation
  - Sample handling
  - Data processing



Hamilton Microlab Vantage Liquid Handling System

-omics sample processing

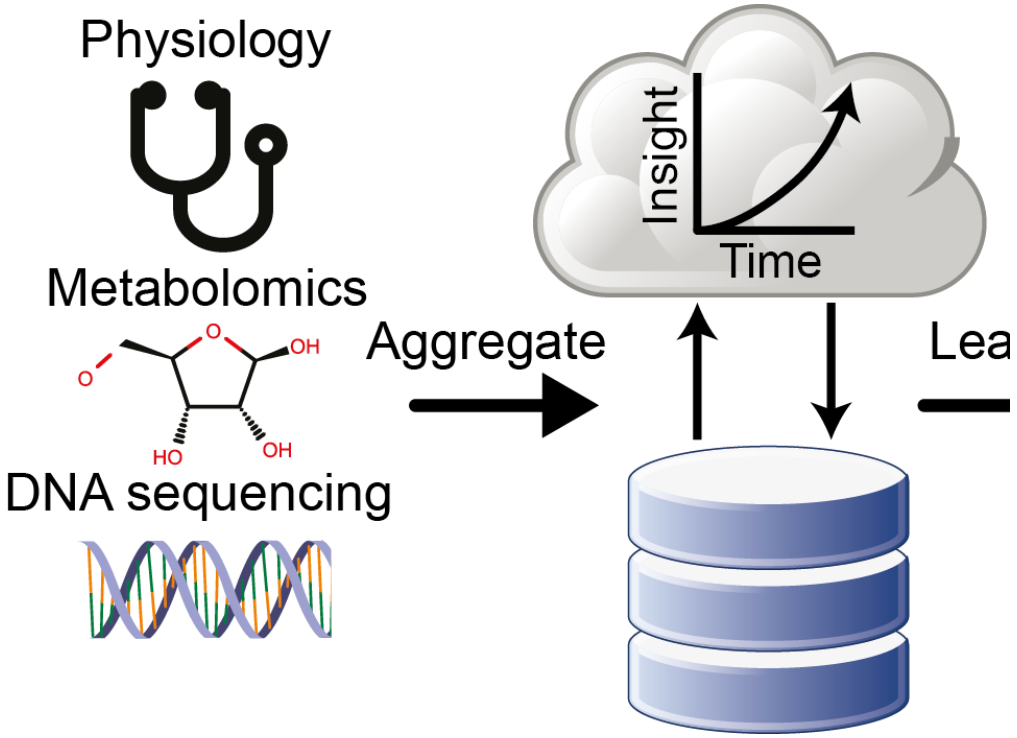


Conclusion (1) Future trends

- Automation
  - Sample handling
  - Data processing
- “Big Data”
  - Data aggregation
  - LIMS
  - High throughput computing (HTC)

Conclusion (1) Future trends

- Automation
  - Sample handling
  - Data processing
- "Big Data"
  - Data aggregation
  - LIMS
  - High throughput computing (HTC)

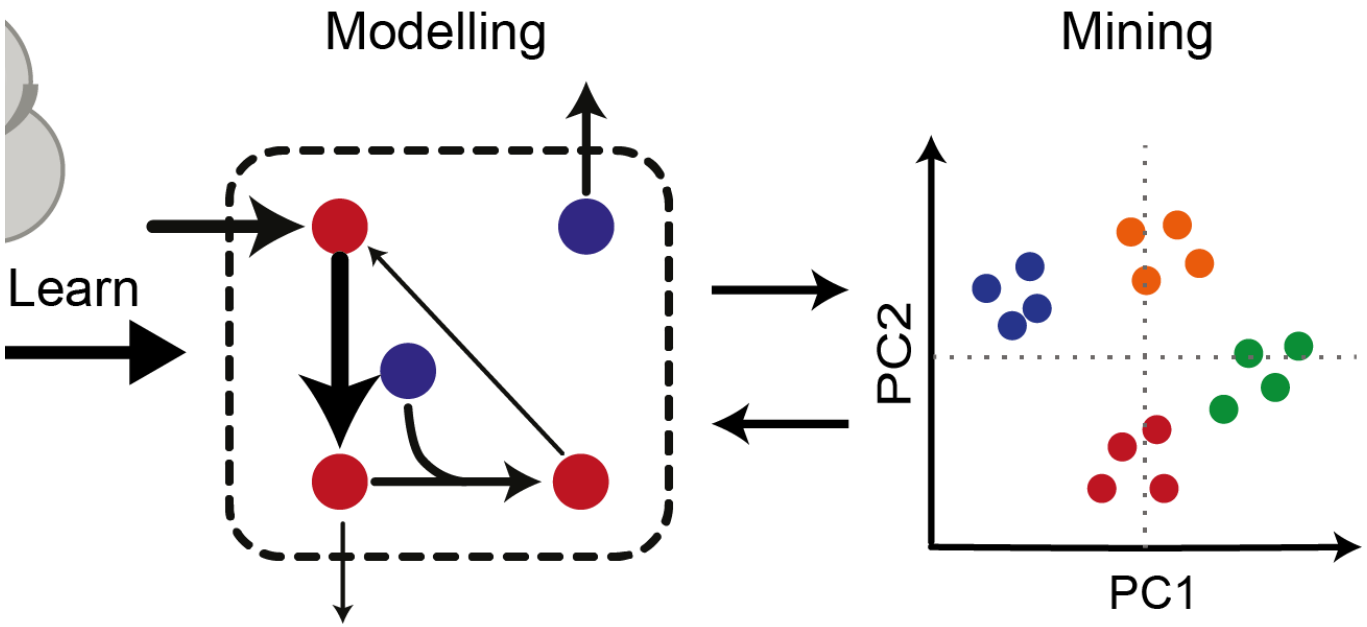


Conclusion (1) Future trends

- Automation
  - Sample handling
  - Data processing
- “Big Data”
  - Data aggregation
  - LIMS
  - High throughput computing (HTC)

Conclusion (1) Future trends

- Automation
  - Sample handling
  - Data processing
- "Big Data"
  - Data aggregation
  - LIMS
  - High throughput computing (HTC)
- Data analysis
  - Modeling
  - Machine Learning
  - Dynamic visualization



Conclusion (2) Re-cap

1. Multi-omic data integration and analysis architecture
2. Case studies
3. Model driven experimental design
4. Data processing pipelines
5. Biochemical model mapping
6. Biochemical modeling
7. Statistics
8. Visualization and sharing
9. Conclusion

Acknowledgements

UCSD:
- Bernhard Ø. Palsson
- Neema Jamshidi
- Andreas Dräger
- Julia Xu
- Edith Hernandez
- Richard Xie
- Wayne Yang

OpenMS:
- Oliver Kohlbacher
- Timo Sachsenberg
- Hannes Röst

DTU:
- Markus Herrgård
- Hanne Bjerre Christensen
- René Hanke
- Lars Schrübbers
- Suresh Sudarsan
- Jérôme Maury
- Simo Jacobsen
- Olivier Bitterlin
- Pasquale Colaianni

Upcoming events

Clinical Metabolomics Copenhagen 2018

Dates: Oct 25-26th, 2018

Location: Copenhagen, DK

Will be announced on www.eventbrite.com by mid July

Contact Information

Douglas McCloskey

Email (consulting): dmccloskey87@gmail.com

Email (academic): domccl@biosustain.dtu.dk

Phone: +45 93 51 19 57