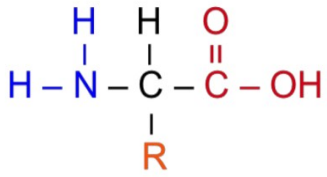
Syllabus and grading

#	Date	Short title	Lecturer	Subject
1	7/10/25	introduction	MR	Overview of Bioinformatics, sequence alignment
2	14/10/25	Linux/shell/ssh	AD	Introduction to Linux and the command line, bash scripting and ssh
3	21/10/25	QC+RNASeq	MR	Next generation sequencing: introduction, quality control and gene expression analysis for RNAseq
4	4/11/25	R (1)	AD	Introduction to the R programming language and Rstudio usage
5	11/11/25	R (2)	AD	Advances R subjects, introduction to Bioconductor
6	18/11/25	bedtools/vcftools/samtools fl	AD	Command line tool usage: bedtools, vcftools, samtools etc.
7	25/11/25	Denovo	MR	NGS for denovo genome and transciptome assembly
8	2/12/25	Exome/SNP calling	AD	Pipelines for SNP calling, especially for exome sequencing using the GATK pipeline
9	9/12/25	ChipSeq/chirp	MR	NGS analysis for molecular interactions (ChipSeq, (Par-)Clip, structural sequencing, chromosome conformation capture (3C))
10	16/12/25	metabolomics	MR	Pipelines for SNP calling, especially for exome sequencing using the GATK pipeline
11	13/1/26	presentations	MR+AD	Paper presentations by students
12	20/1/26	presentations	MR+AD	Genome-scale models of metabolism and macromolecular expression, Biological applications of Transformers
13	27/1/26	final projects support	MR+AD	Support for the final project

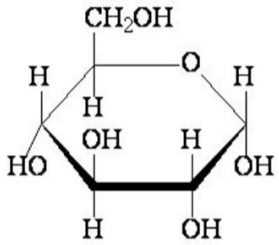
Grade	**100%**
Presentation	30%
Exercises	20%
Final Project	50%

Metabolomics: what are metabolites?

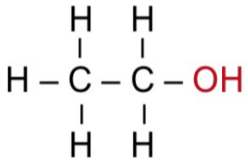
MetaboLights



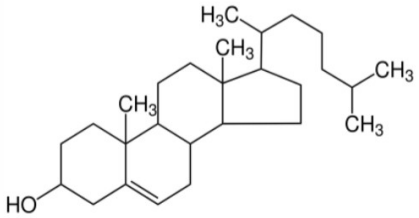
Amino acids



Sugars



Alcohols



Lipids

- small molecule
  - *MW <1500 Da*

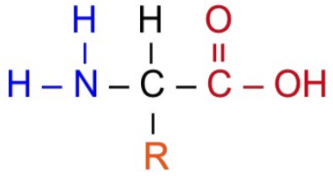
  eg. sugars, amino acids, phenolics, lipids, small peptides

- physically and chemically complex
  - *vast range of chemical structures & properties; polarity, solubility, volatility, stability*

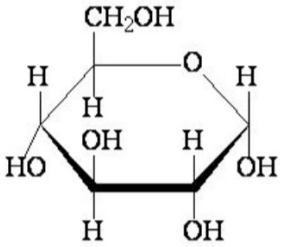
- *wide concentration range*
  - *mM to sub pM*
  - dynamic with constant synthesis, absorption, degradation, modification
  - *metabolite levels change with half life of mins or secs*

Metabolomics: what are metabolites?

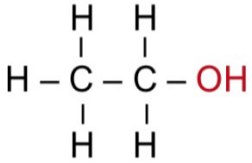
MetaboLights



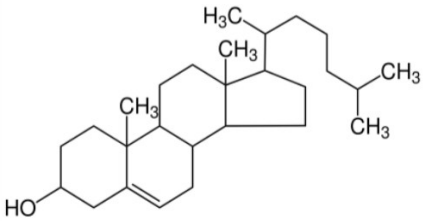
Amino acids



Sugars



Alcohols



Lipids

Primary Metabolites

directly involved in normal growth, development, and reproduction.

- *Glucose, lactic acid, vitamin B2*

Secondary Metabolites

not required for primary metabolic processes but have an important ecological role or other function.

- *defense against predators and other interspecies defenses (plants), toxins, antibiotics, flavorings*

An Introduction to Metabolomics

MetaboLights

metabolite

/mɪˈtabəlʌɪt/

noun BIOCHEMISTRY

noun: **metabolite**; plural noun: **metabolites**

a substance formed in or necessary for metabolism.

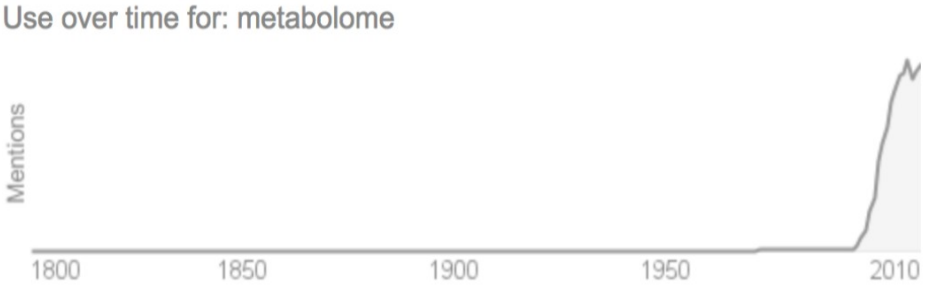
metabolome

/mɪˈtabələʊm/

noun BIOCHEMISTRY

noun: **metabolome**; plural noun: **metabolomes**

the total number of metabolites present within an organism, cell, or tissue.



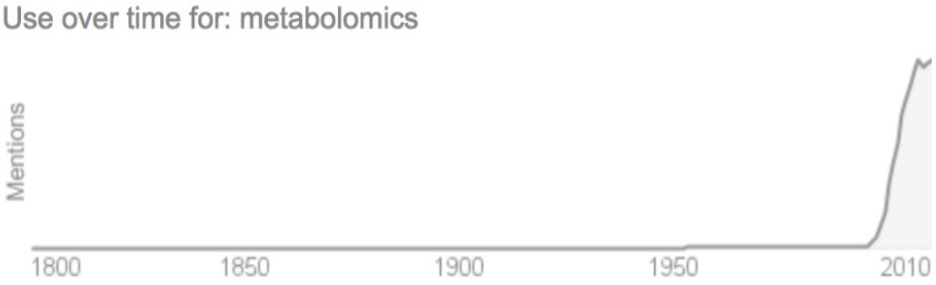
metabolomics

/mɪˈtabəlɒmɪks/

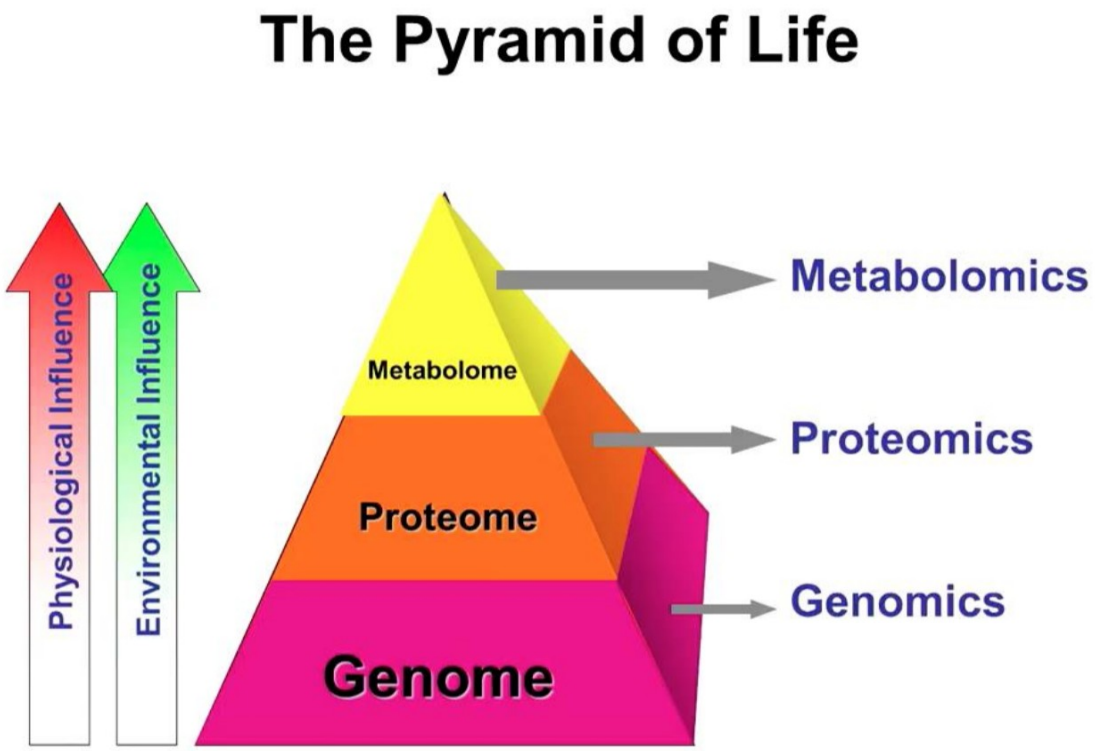
noun BIOCHEMISTRY

noun: **metabolomics**

the scientific study of the set of metabolites present within an organism, cell, or tissue.



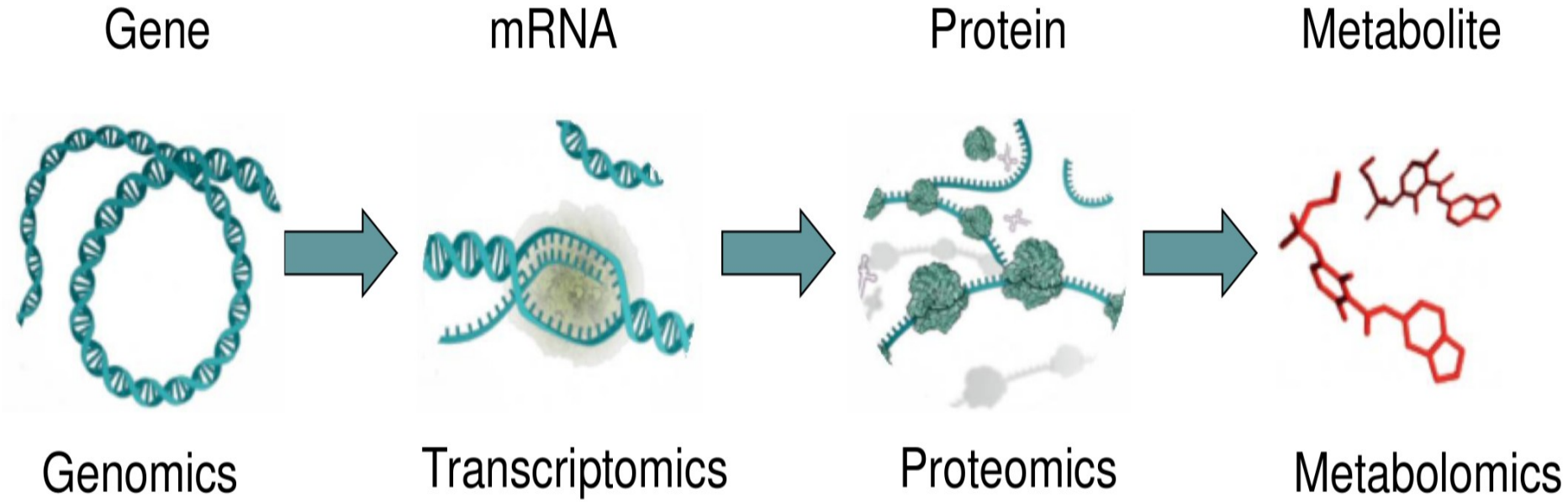
MetaboLights



D Wishart

Metabolomics: across the omics…

MetaboLights

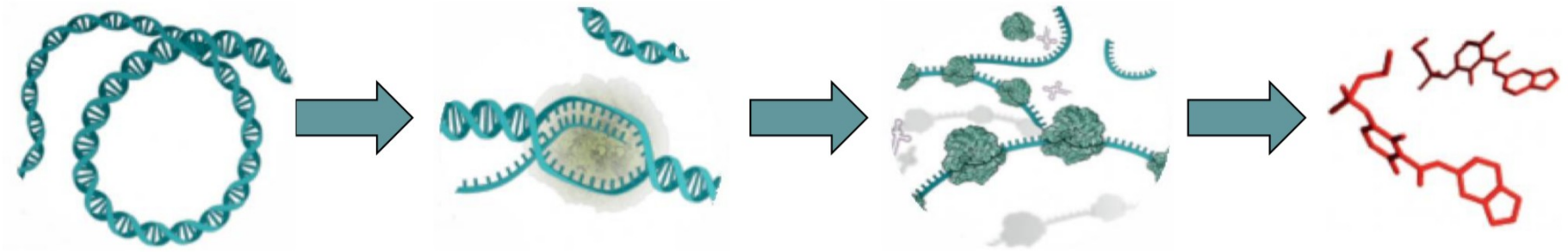


Metabolomics is complementary to genomic and proteomic studies

Metabolomics: across the omics…

MetaboLights

Gene → mRNA → Protein → Metabolite



Genomics → Transcriptomics → Proteomics → Metabolomics

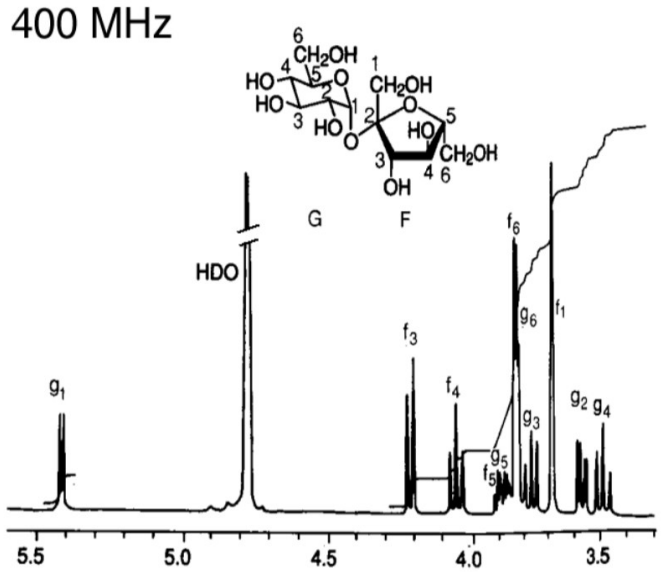
Genomics	Proteomics	Metabolomics
4 bases	20 amino acids	10^5+ chemicals
Single instrument	1 - 2 instruments	multiple instruments
22,000 genes	5000 – 10,000 proteins	200 is good!

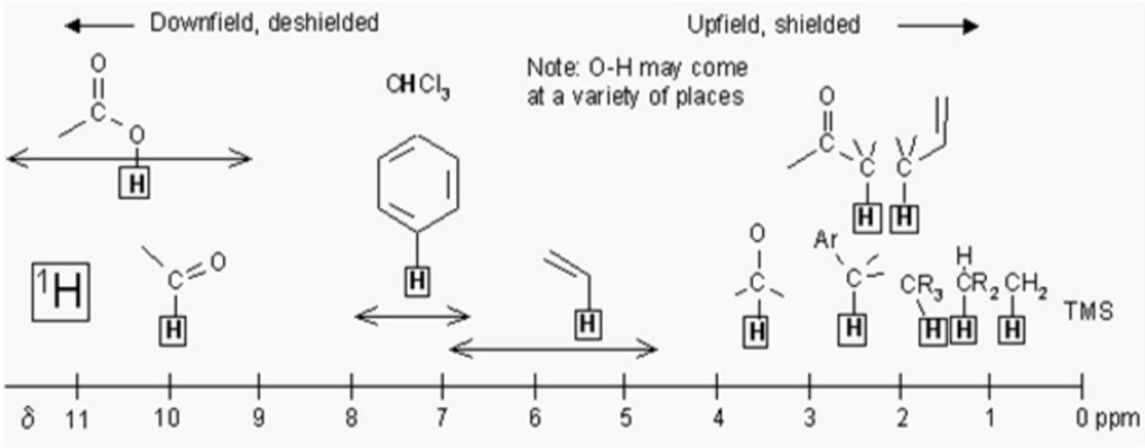
Relative abundance of metabolites requires sensitive instrumentation

Metabolomics: technology

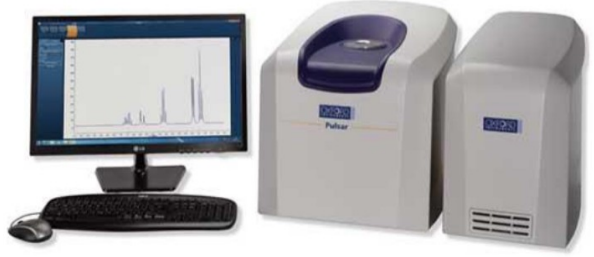
MetaboLights

Mass spectrometry (MS) and ***Nuclear Magnetic Resonance (NMR) spectroscopy*** *are the techniques most often used for metabolome profiling.*



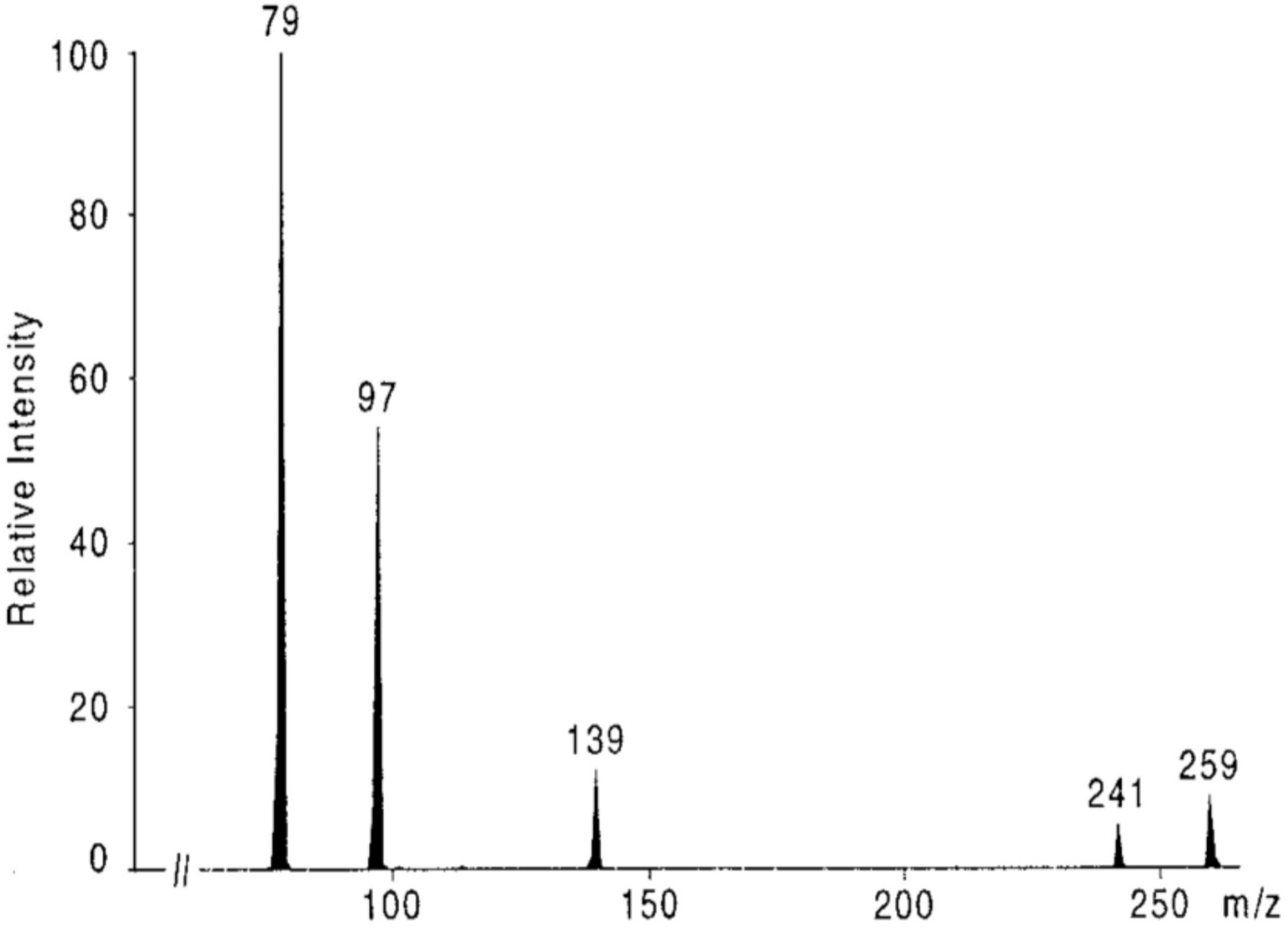


Nuclear Magnetic Resonanc

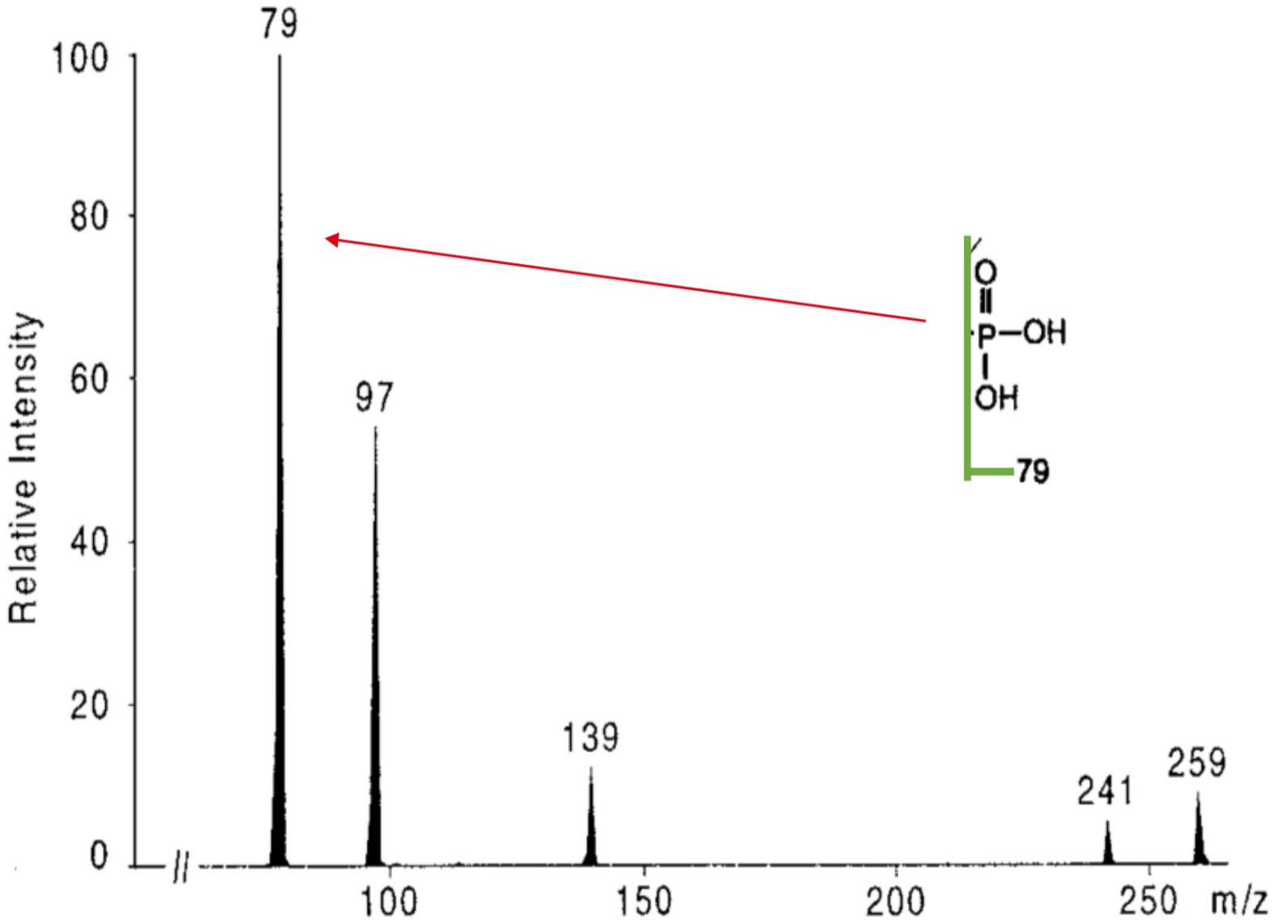


Metabolomics: technology

MetaboLights

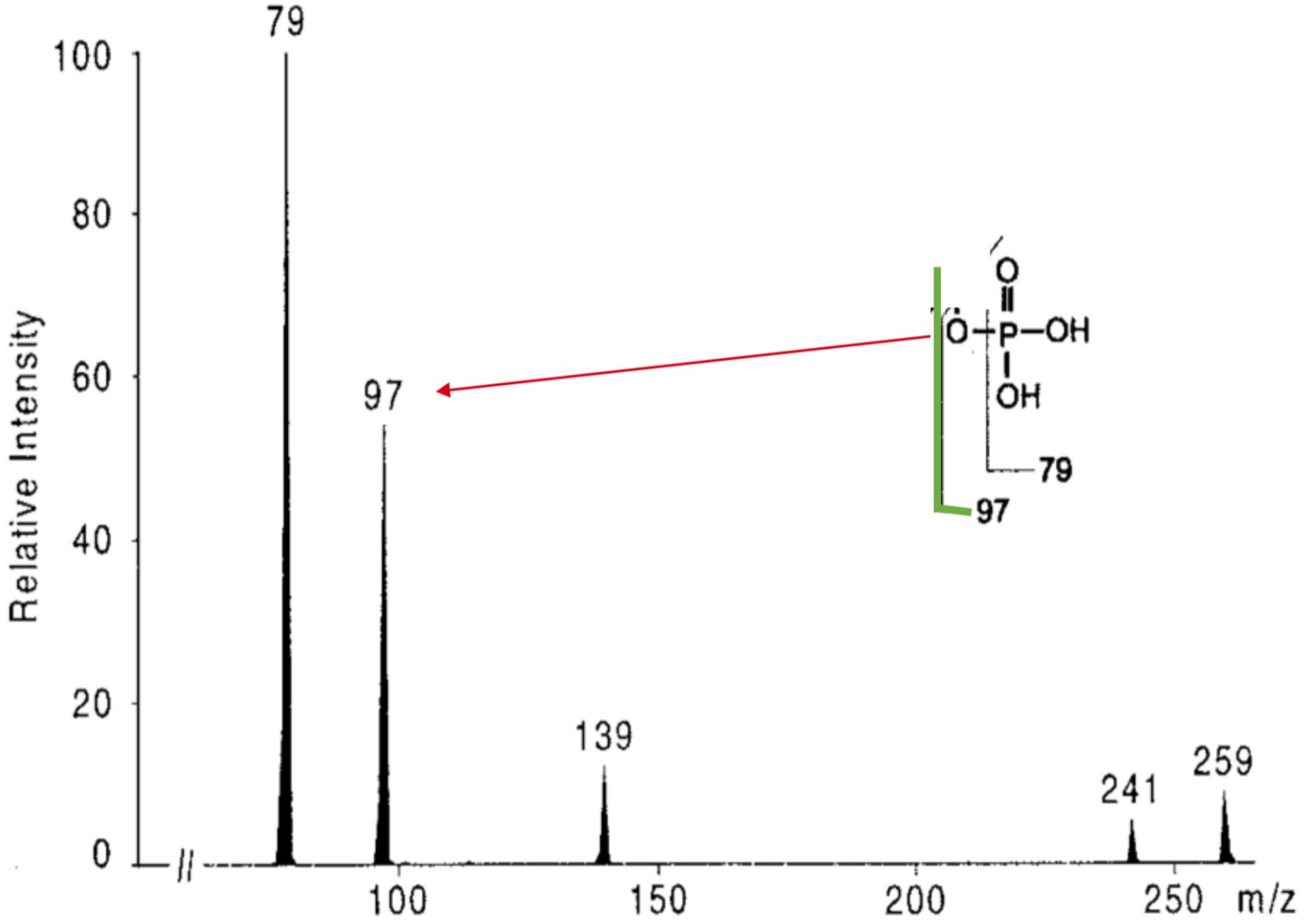


MetaboLights



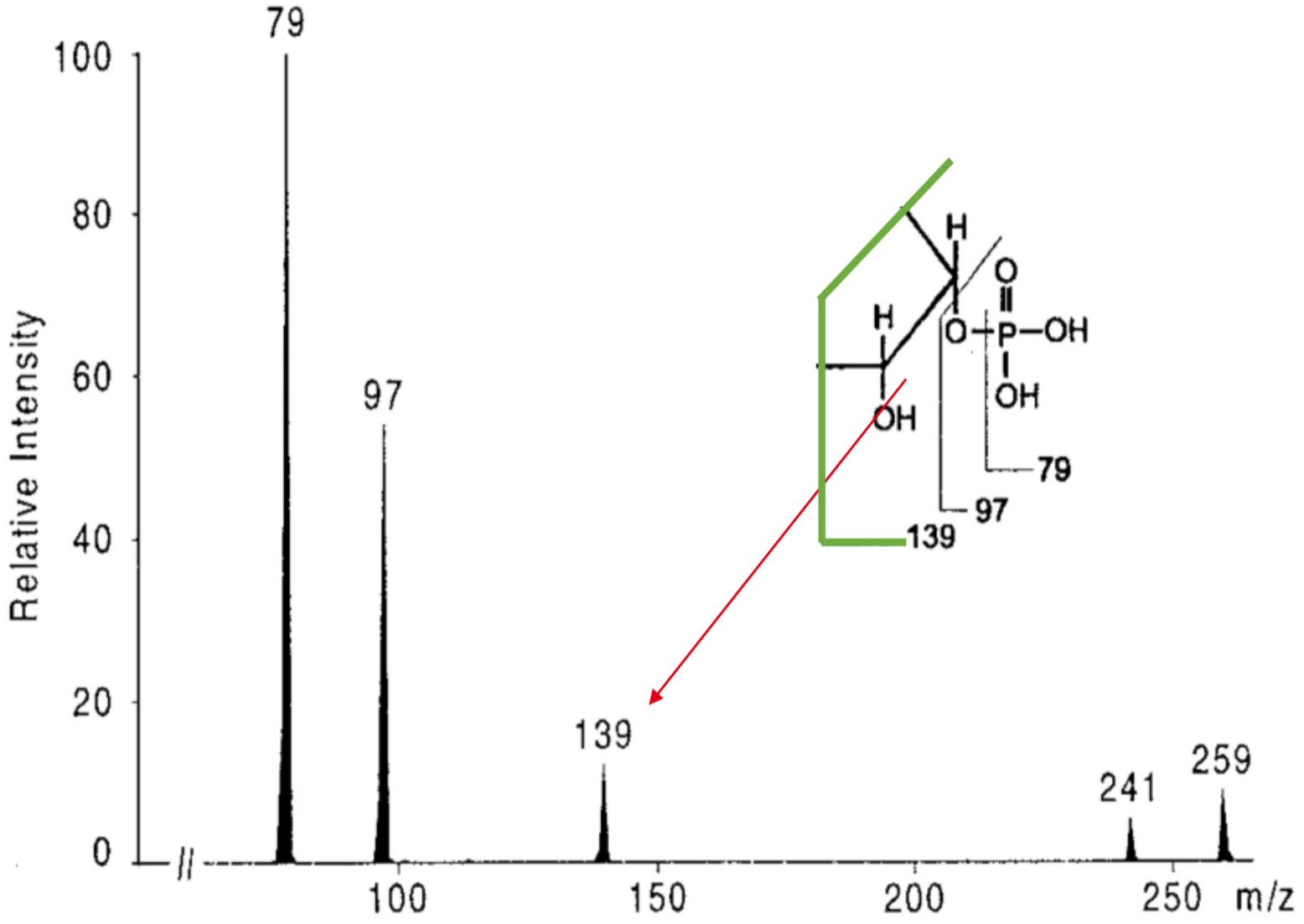
Metabolomics: technology

MetaboLights



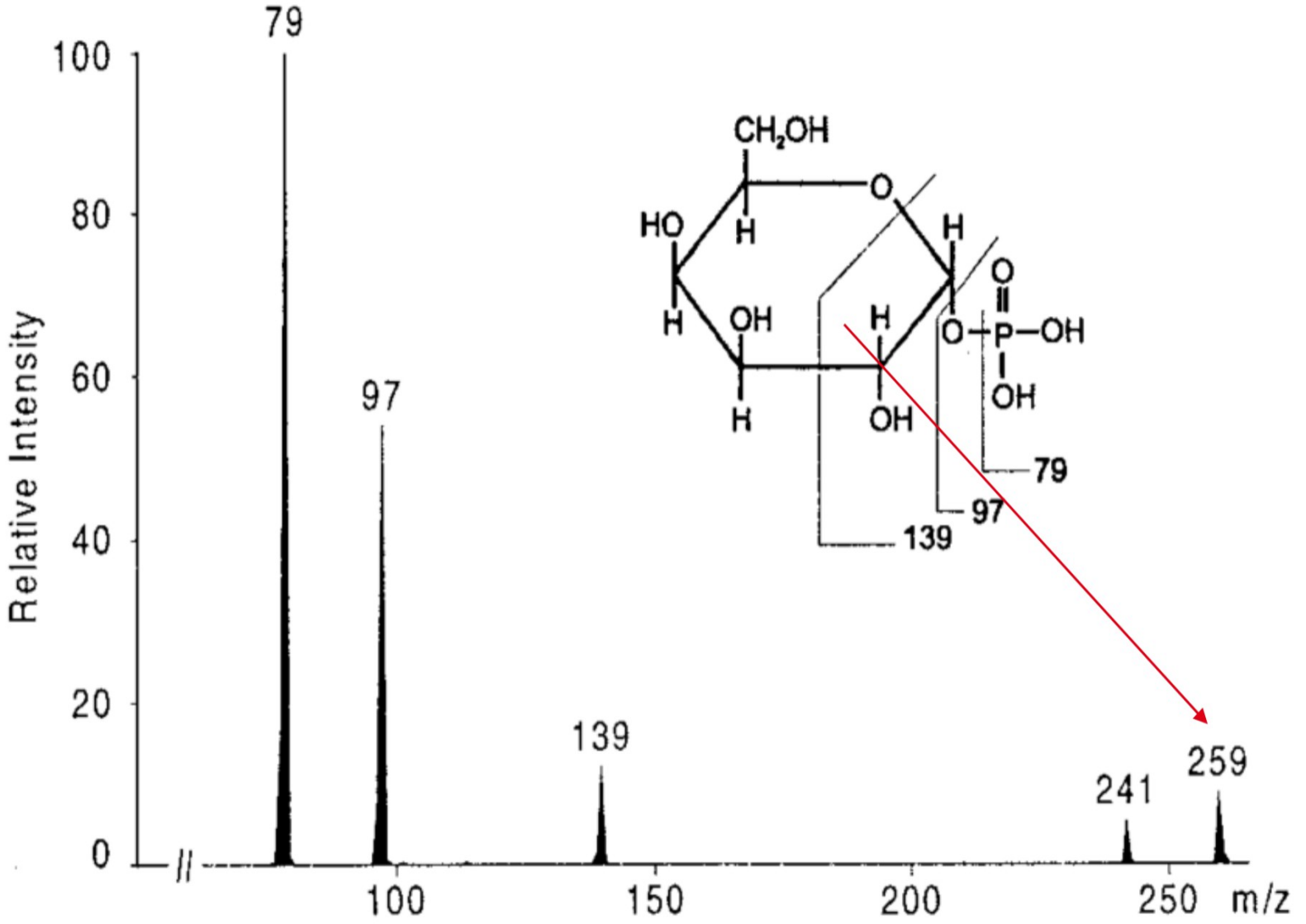
Metabolomics: technology

MetaboLights



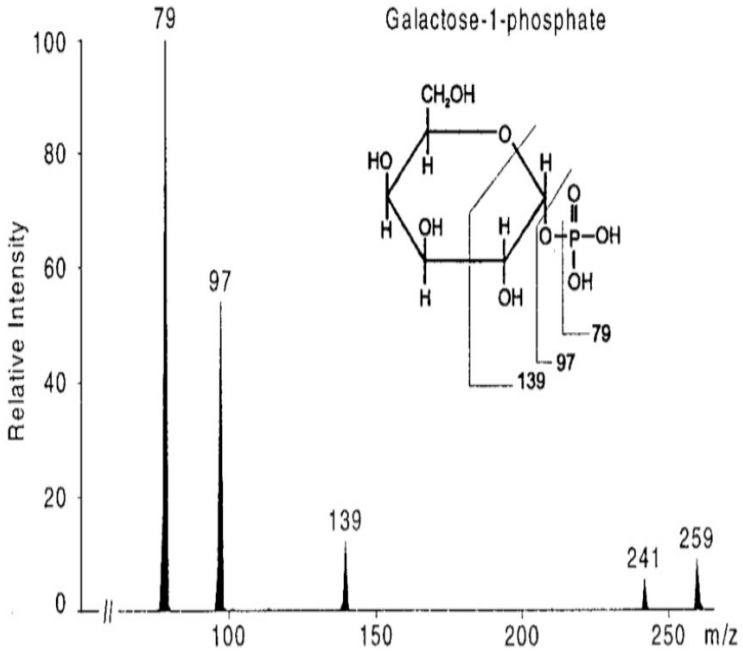
Metabolomics: technology

MetaboLights



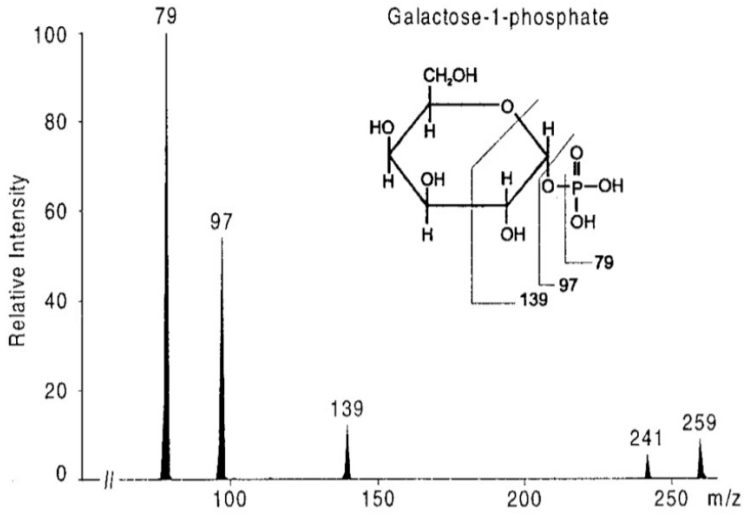
Metabolomics: technology

MetaboLights



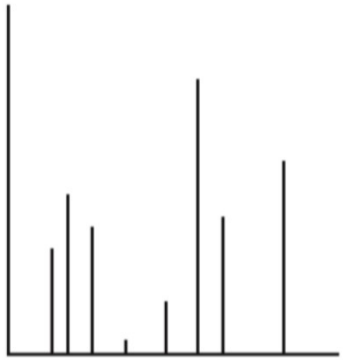
Metabolomics: technology

MetaboLights



<< >>

Spectral Libraries



^
^
v
v

Unknown sample 1

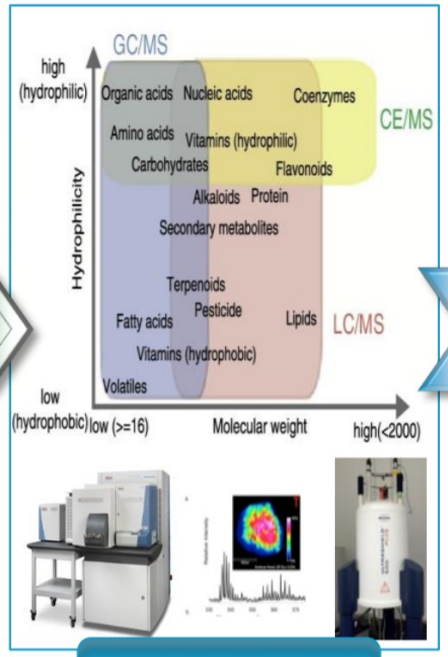
Reference compound

MetaboLights

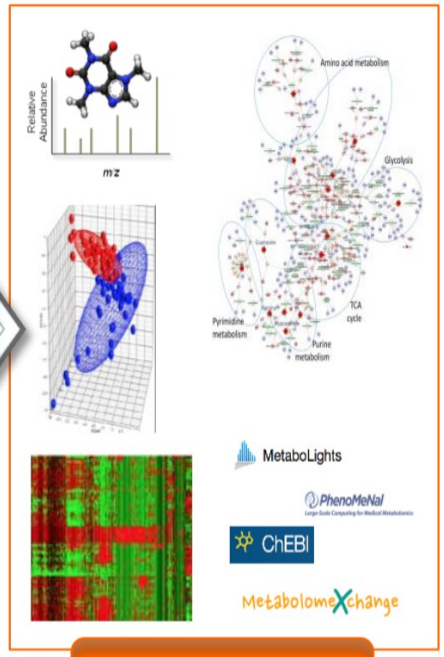
Metabolomics strategies

Sample Collection

Sample Preparation



Analytical Analysis



Data Interpretation

Intra-cellular metabolites

Extracellular metabolites
Biofluids (urine, plasma)
Metabolic footprint (secreted intracellular)

Tissues

Solvent extraction

Precipitation

No / minimal prep

Available equipment

Identification

Quantification

Analysis methods

Publication

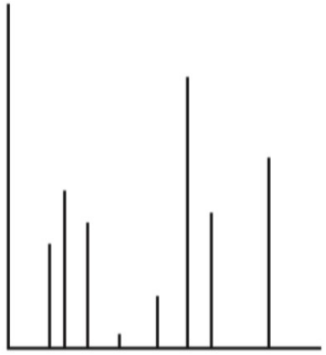
Intellectual Property

Data Sharing

Metabolomics Resources

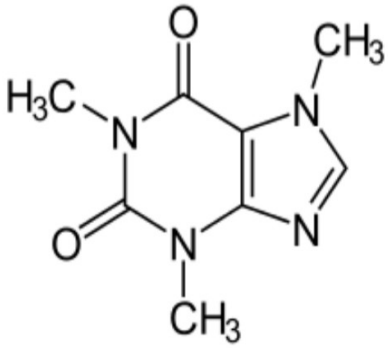
MetaboLights

Spectral Libraries



Analysis Tools

Metabolite Knowledgebases



Repositories

hmdb
The Human Metabolome Database

LIPID MAPS

METLIN

KEGG Kyoto Encyclopedia of Genes and Genomes

GNPS

MetaboLights

PhenoMeNal
Large-Scale Computing for Medical Metabolomics

4 Wm Workflow4metabolomics

MetaboAnalyst 4.0
Statistics
Pathway
Function
Biomarker

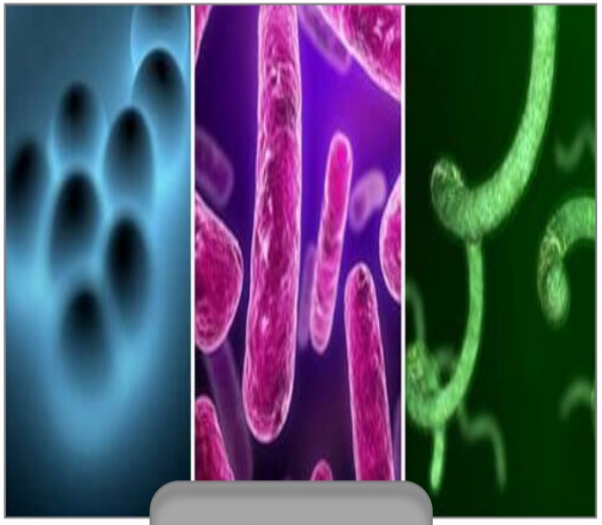
VMH
VIRTUAL METABOLIC HUMAN

Metabolomics: applied

MetaboLights

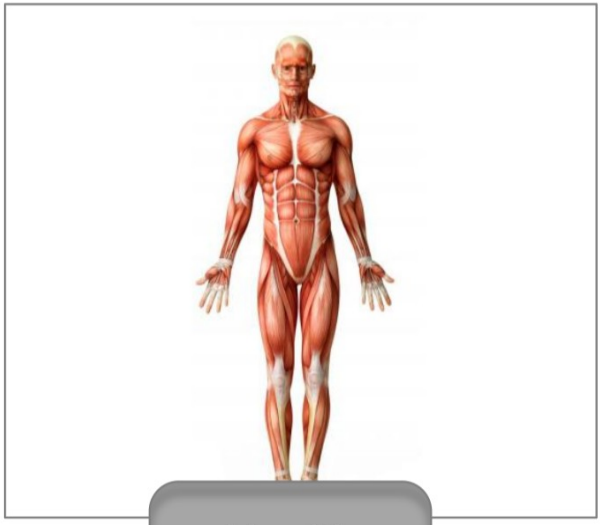
Agriculture R&D

Research in agriculture including crop protection and engineering
e.g. salt and drought tolerance; plant defence



Food Industry

Product and stress testing in food industries, e.g. water quality, control of pesticides and identification of potentially harmful bacterial strains



Health care

Medical diagnostics in healthcare, and future applications in personalised medicine
e.g. newborn screening

Metabolomics in oncology

Cell, Tumor, and Stem Cell Biology

Mass Spectrometry–Based Metabolic Profiling Reveals Different Metabolite Patterns in Invasive Ovarian Carcinomas and Ovarian Borderline Tumors

Carsten Denkert, Jan Budczies, Tobias Kind, Wilko Weichert, Peter Tablack, Jalid Sehouli, Silvia Niesporek, Dominique Könsgen, Manfred Dietel, and Oliver Fiehn

DOI: 10.1158/0008-5472.CAN-06-0755 Published November 2006

Letter

Metabolomic profiles delineate potential role for sarcosine in prostate cancer progression

Arun Sreekumar, Laila M. Poisson [...] Arul M. Chinnaiyan

Nature **457**, 910–914 (12 February 2009)
doi:10.1038/nature07762
Download Citation

Received: 09 October 2008
Accepted: 06 January 2009
Published: 12 February 2009
Corrigendum: 05 June 2013

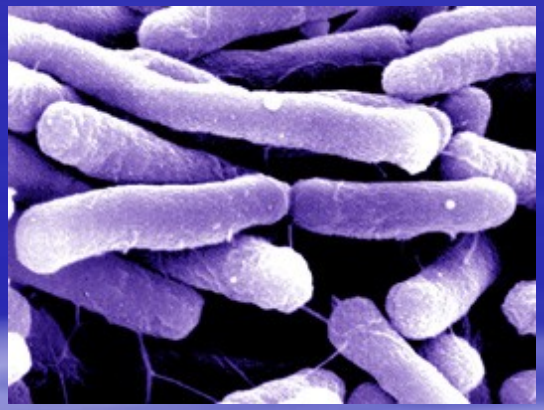
Oncometabolite*	Mechanism or role	Refs
2-hydroxyglutarate	• Inhibits ATP synthase and mTOR signalling • Inhibits 2-oxoglutarate-dependent oxygenases, which activate oncogenic hypoxia-induced factor pathways and alter DNA methylation patterns • Produced by gain-of-function mutations in the gene encoding isocitrate dehydrogenase • Elevated in gliomas and acute myeloid leukemia	63,144
Fumarate	• Inhibits 2-oxoglutarate-dependent oxygenases, which activate oncogenic hypoxia induced factor pathways and alter DNA methylation patterns • Leads to protein succination and disrupted metabolism • Produced by loss-of-function mutations in the gene encoding fumarate hydratase • Elevated in renal carcinoma	144,145
Succinate	• Inhibits 2-oxoglutarate-dependent oxygenases which activate oncogenic hypoxia induced fac pathways and alter DNA methylation • Produced by loss-of-function muta genes encoding succinate deh • Elevated in paraganglion and thyroid tumours	144,146
Sarcosine	• Activates mTOR g pathway • Elevated b glycine *N*-methyl transferase • Elevate metastatic prostate cancer	147,148
Glucose	source of carbon to support cancer cell abolism, TCA anaplerosis and aerobic glycolysis Activates hexokinase II • Activates glucose-regulated proteins that alter signalling, proliferation, invasion and apoptosis • Elevated in most cancers	149,150
Glutamine	• Essential source of nitrogen to support cancer cell anabolism and aerobic glycolysis • Essential source of carbon for TCA anaplerosis • Elevated in MYC-dependent cancers	151,152
Asparagine	• Essential source of nitrogen to support cancer cell anabolism and aerobic glycolysis • Anti-apoptotic agent • Elevated in acute lymphoblastic leukemia	153
Choline	• Serves as a methyl donor for DNA methylation which disrupts DNA repair and gene expression • Modifies lipid signalling • Essential source of carbon and nitrogen to support phospholipid synthesis in rapidly dividing cells • Elevated in breast, brain and prostate cancer	154
Lactate	• Lowers extracellular pH and induces metastasis • Induces local immunosuppression • Elevated in most cancers	155,156

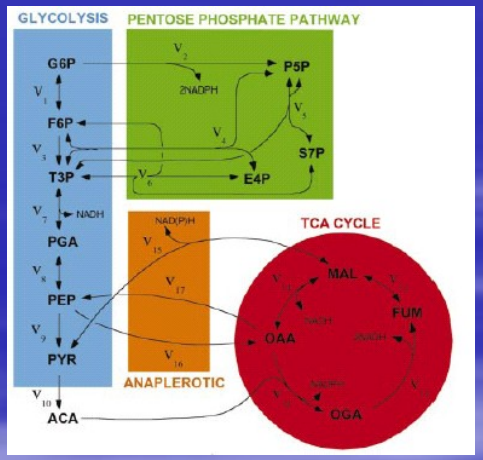
oncometabolites

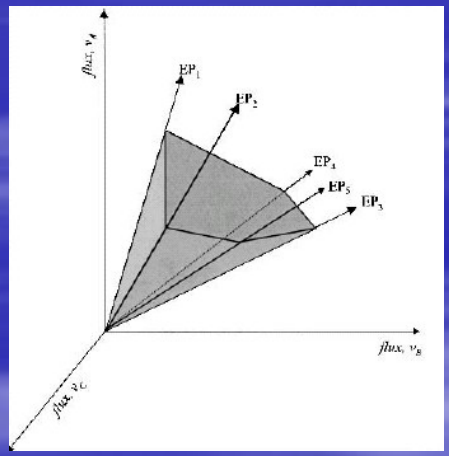
David S Wishart

NATURE REVIEWS | DRUG DISCOVERY

Flux balance analysis in metabolic networks







Lecture notes by Eran Eden, https://systemsbiology.usu.edu/documents/Flux%20Balance%20Analysis%20Overview%20-%202016.pdf

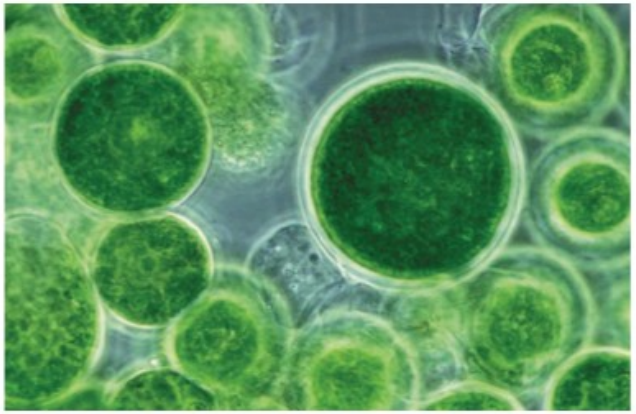
additional slides by B. Palsson from http://systemsbiology.ucsd.edu/sites/default/files/Attachments/Images/classes/taiwan_notes/6_slides _fba.pdf

and Jordan Nikkel https://www.youtube.com/watch?v=87OKtTpSRB8 Edit+supp. Mat. by Martin Reczko

Systems biology

Understand:

evolution
environment <==> genome



By:

systems biology
knowledge of cells <==> computer model

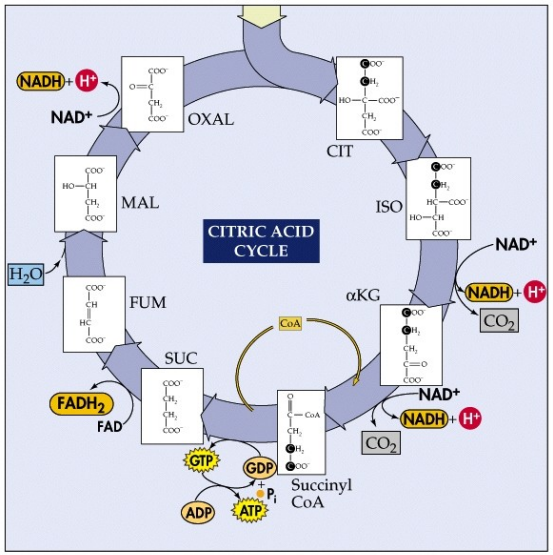
In order to:

alter our own human environment

Flux balance analysis in metabolic networks.

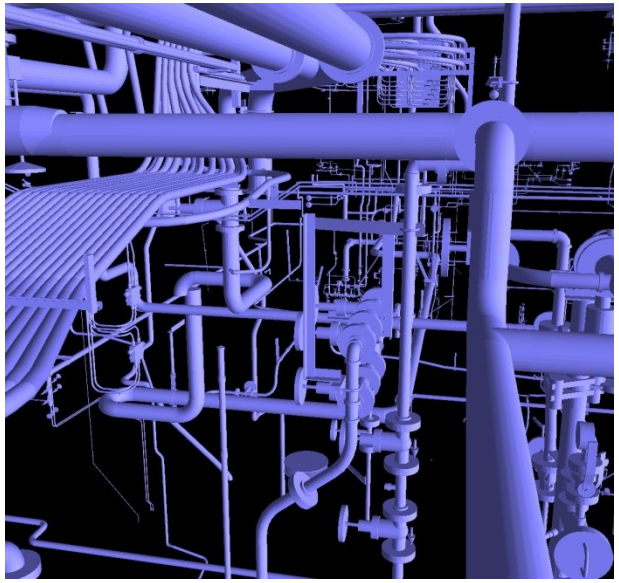
1. Metabolic networks

"Metabolism is the process involved in the maintenance of life. It is comprised of a vast repertoire of enzymatic reactions and transport processes used to convert thousands of organic compounds into the various molecules necessary to support cellular life" Kenneth et al. 2003



http://bill.srnr.arizona.edu/classes/182/CitricAcidCycle-LowRes.jpeg

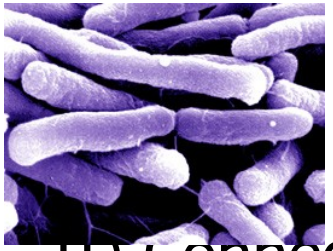
2. Flux Balance Analysis

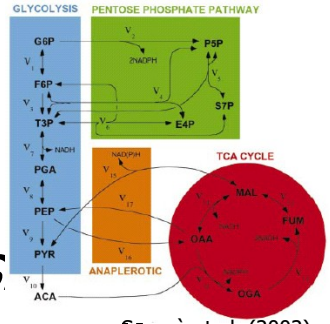


http://images.google.com/imgres?imgurl=http://www.cs.unc.edu/~walk/models/double_eagle/pieces/pipes.jpg&imgrefurl=http://www.cs.unc.edu/~walk/models/double_eagle/pieces/&h=1095&w=1156&sz=3078&tbnid=1C0069trx90J:&tbnh=142&tbnw=150&hl=en&start=1&prev=/images%3Fq%3D%2Bpipes%2B%26hl%3Den%26lr%3D

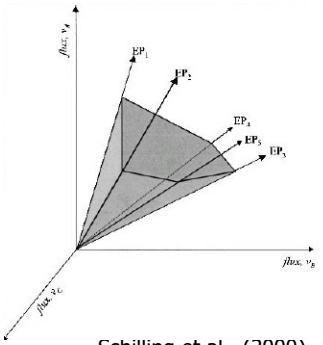
Plan

I) Creating an *in silico* model in order to describe an organisms metabolism in steady state



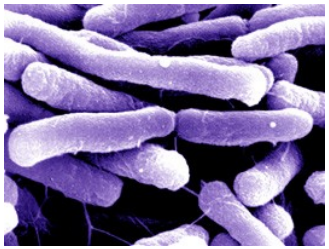


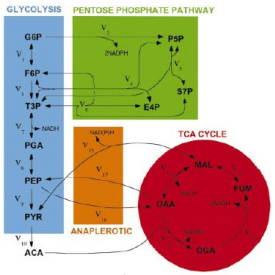
Segre` et al. (2002)

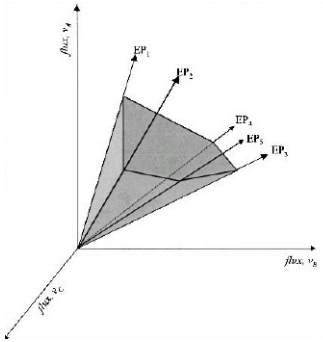


Schilling et al. (2000)

II) Connecting *in s* to *in-vivo* experiments in e. coli





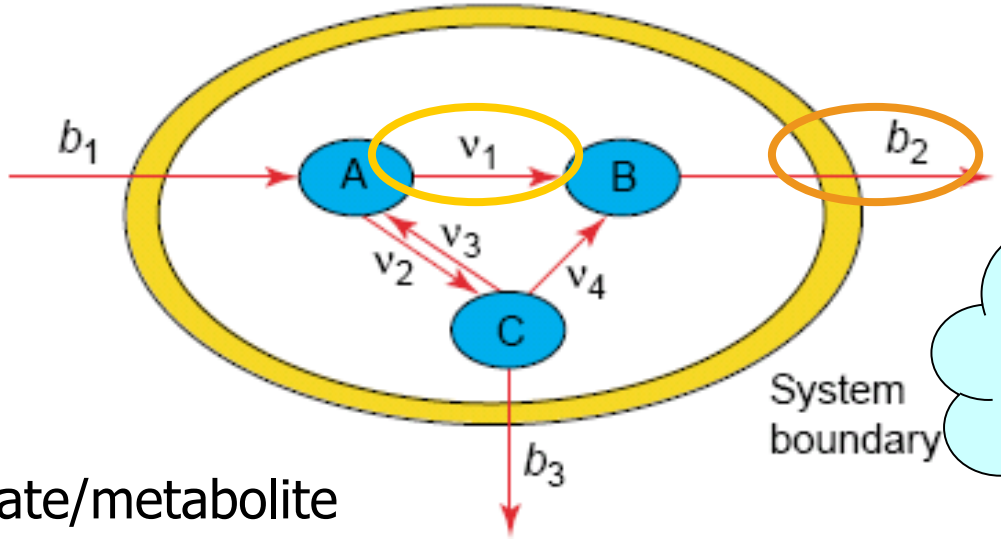


Motivation for studying metabolic pathways

- Better understanding of cellular physiology.
- Understanding vulnerabilities of unicellular metabolism.
- Manipulation, Bio-Engeneering

Constructing the model… Step (I) - Definitions

We begin with a very simple imaginary metabolic network represented as a directed graph:



How do we define a biologically significant system boundary?

Vertex - substrate/metabolite concentration.

Edge - flux (conversion mediated by enzymes of one substrate into the other)

Internal flux edge

External flux edge

(II) - Dynamic mass balance

Concentration vector | Stoichiometry Matrix | Flux vector

$$\frac{dx}{dt} = S \cdot v$$

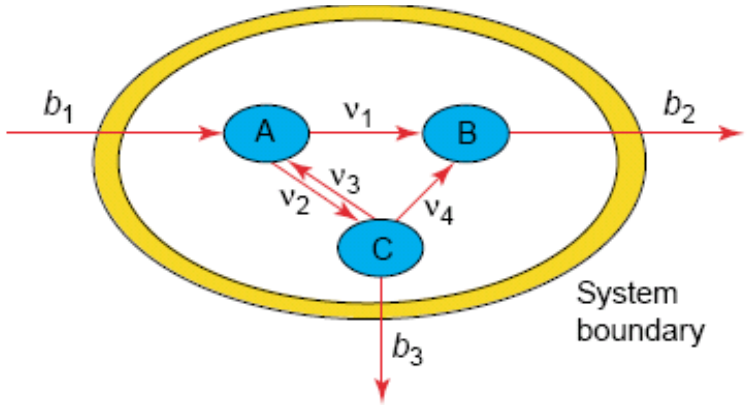
$$\frac{dA}{dt} = -v_1 - v_2 + v_3 + b_1$$

$$\frac{dB}{dt} = v_1 + v_4 - b_2$$

$$\frac{dC}{dt} = v_2 - v_3 - v_4 - b_3$$

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (columns) V (rows)



(II) - Dynamic mass balance

Concentration vector | Stoichiometry Matrix | Flux vector

$$\frac{dx}{dt} = S \cdot v$$

Problem …

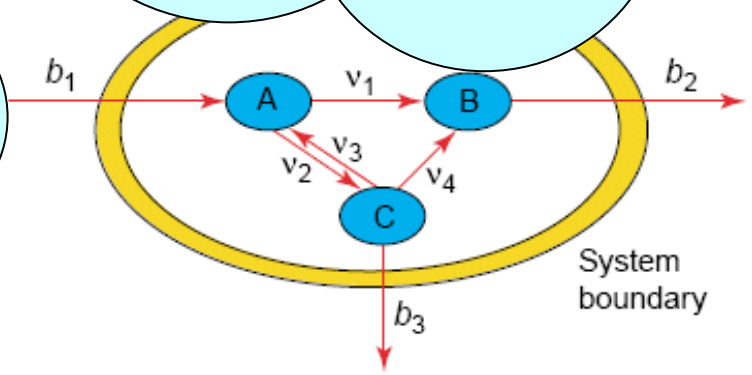
V=V(k1, k2,k3…) is actually a function of concentration as well as several kinetic parameters.

it is very difficult determine kinetic parameters experimentally.

Consequently there is not enough kinetic information in the literature to construct the model.

Solution !

In order to identify invariant characteristics of the network we assume the network is at **steady state**.



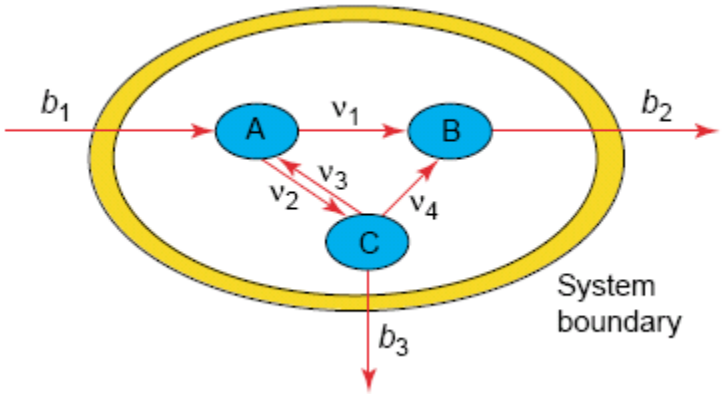
(III) - Dynamic mass balance at steady state

1. What does "steady state" mean?

2. Is it biologically justifiable to assume it?

3. Does it limit the predictive power of our model?

4. Most important question…



"The steady state approximation is generally valid because of fast equilibration of metabolite concentrations (**seconds**) with respect to the time scale of genetic regulation (**minutes**)" – Segre 2002

Yes…

4. Why does the steady state assumption help us solve our problem?

$$\frac{dx}{dt} = S \cdot v \longrightarrow 0 = S \cdot v$$

Steady state assumption

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (7 columns) × V (7 rows)

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{bmatrix}$$

S (7 columns) × V (7 rows)

(VI) adding constraints

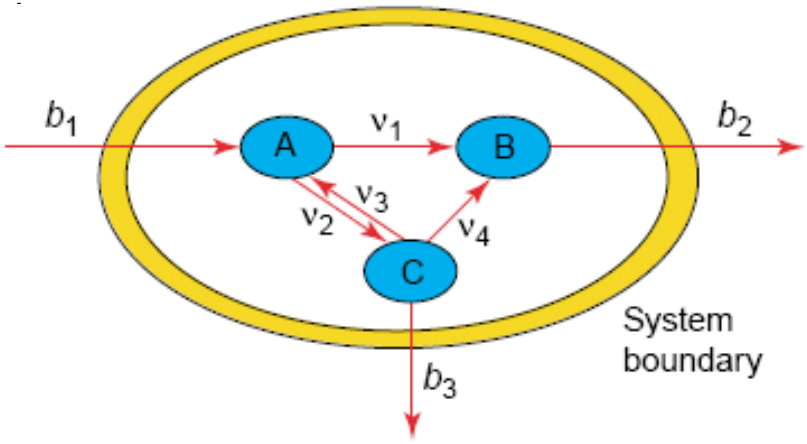
Constraints on internal fluxes: $v_i \geq 0, \forall$

Constraints on external fluxes:

Source $b_j \leq 0$

Sink $b_j \geq 0$

Sink/source b_j is unconstrained



In other words flux going into the system is negative while flux leaving the system is considered positive.

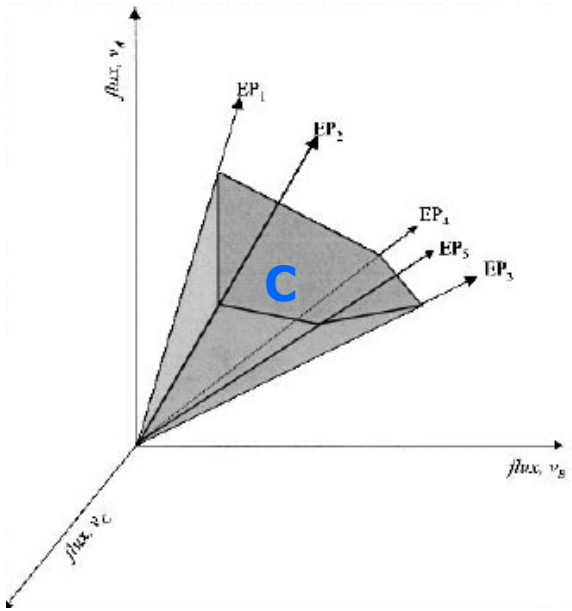
Remark: later on we will impose further constraints both on the internal flux as well as the external flux…

(V) Flux cone and metabolic capabilities

$$0 = S \cdot v$$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ b_1 \\ b_2 \\ b_3 \end{matrix} \quad \begin{matrix} \\ \\ \\ \\ \\ \\ \end{matrix} \quad v$$

S



Observation: the number of reactions considerably exceeds the number of metabolites

↓

The S matrix will have more columns than rows

↓

The null space of viable solutions to our linear set of equations contains an infinite number of solutions.

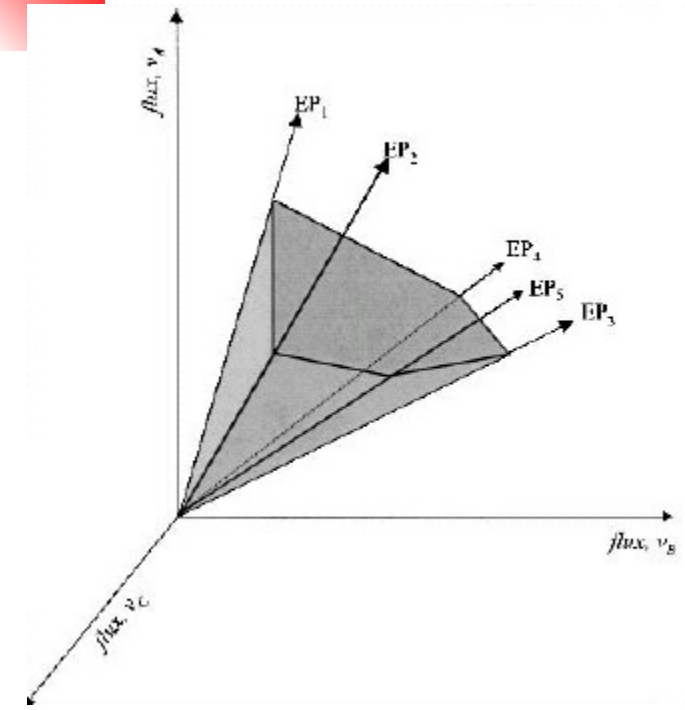
What about the constraints?

↓

"The solution space for any system of linear homogeneous equations and inequalities is a convex polyhedral cone." - Schilling 2000

Our flux cone contains all the points of the null space with non negative coordinates (besides exchange fluxes that are constrained to be negative or unconstrained)

(V) Flux cone and metabolic capabilities



What is the significance of the flux cone?

- It defines what the network can do and cannot do!
- Each point in this cone represents a flux distribution in which the system can operate at steady state.
- The answers to the following questions (and many more) are found within this cone:
  - what are the building blocks that the network can manufacture?
  - how efficient is energy conversion?
  - Where is the critical links in the system?

(VI) Navigating through the flux cone – using "Extreme pathways"

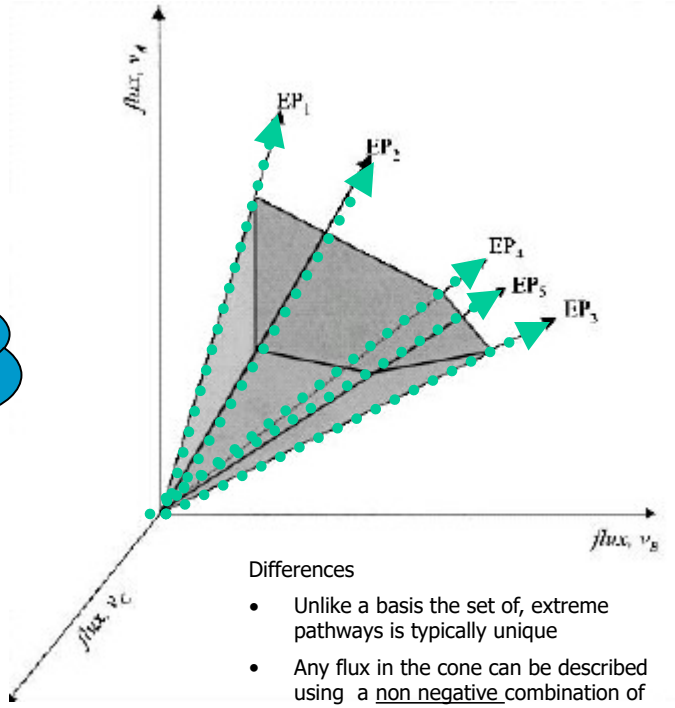
Next thing to do is develop a way to describe and interpret any location within this space.

- We will <u>not</u> use the traditional reaction/enzyme based perspective
- Instead we use a <u>pathway perspective:</u>

Extreme rays - "extreme rays correspond to edges of the cone. They are said to generate the cone and cannot be decomposed into non-trivial combinations of any other vector in the cone."- schilling 2000

What is the analogy in linear algebra?

We use the term **Extreme Pathways** when referring to Extreme rays of a convex polyhedral cone that represents metabolic fluxes



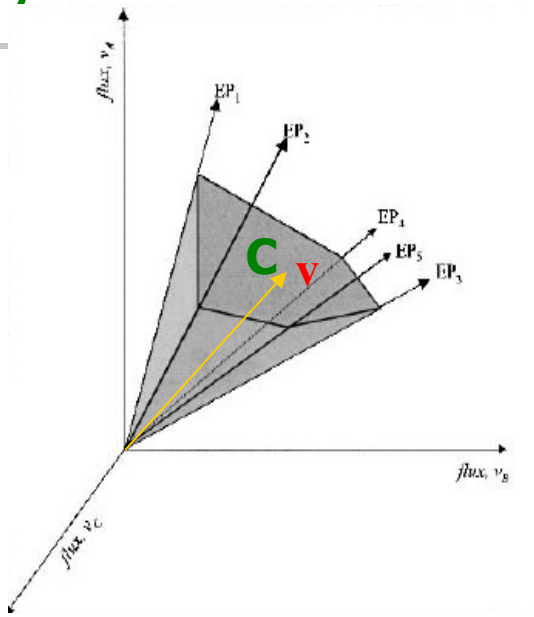
Differences

- Unlike a basis the set of, extreme pathways is typically unique
- Any flux in the cone can be described using a <u>non negative</u> combination of extreme rays.

(VI) Navigating through the flux cone – using "Extreme pathways"

- Extreme Pathways will be denoted by vector $\mathbf{EP_i}$ ($0 \leq i \leq k$)
- Every point within the cone can be written as a non-negative linear combination of the extreme pathways.

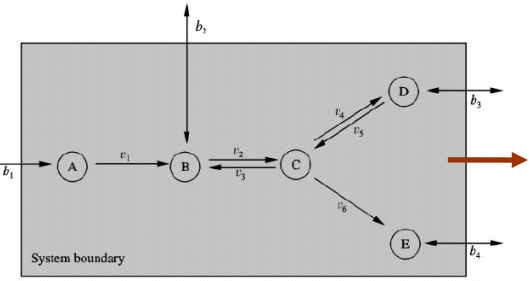
$$C = \{v \mid v = \sum_{i=1}^{k} w_i EP_i, \; w_i \geq 0 \; \forall i\}$$



In biological context this *means that :*

any steady state flux distribution can be represented by a non-negative linear combination of extreme pathways.

The entire process from a bird's eyes



$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

Internal flux constraints

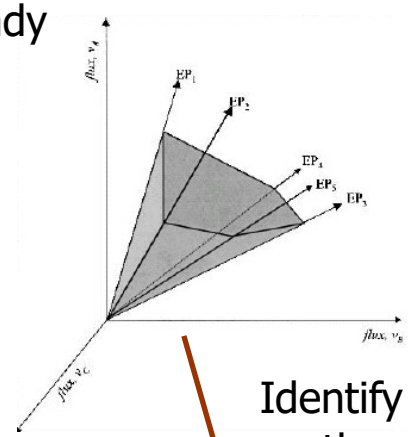
$v_j \geq 0, \quad j = 1, \ldots, 6$

Exchange flux constraints

$-\infty \leq b_j \leq +\infty, \quad j = 1, \ldots, 4$

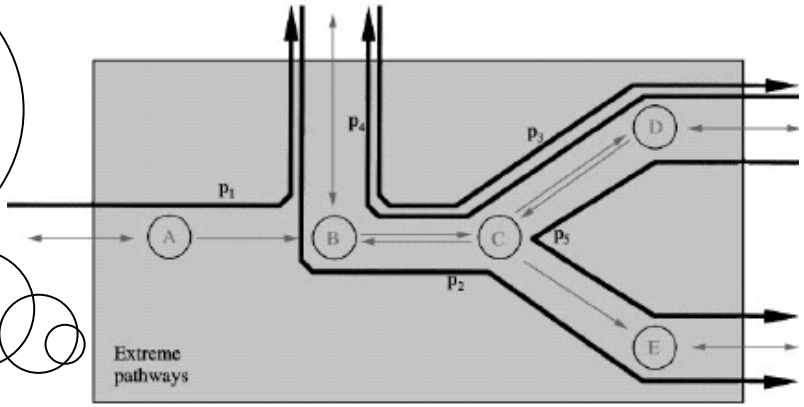
Compute steady state flux Convex

\+ constraints



Identify Extreme pathways (using algorithm presented in Schilling 2000)

$$\mathbf{P} = \begin{array}{c} \begin{matrix} \mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3 & \mathbf{p}_4 & \mathbf{p}_5 & \mathbf{p}_6 & \mathbf{p}_7 \end{matrix} \\ \left[\begin{array}{ccccccc} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ \hline -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & -1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 \end{array}\right] \begin{matrix} v_1 \\ v_2 \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ \hline b_1 \\ b_2 \\ b_3 \\ b_4 \end{matrix} \end{array}$$



Example

Lets look at a specific vector v' :

$$v = \begin{bmatrix} 4 \\ 2 \\ 0 \\ 1 \\ 0 \\ 1 \\ -4 \\ 2 \\ 1 \\ 1 \end{bmatrix}$$

Is v inside the flux cone?

Easy to check…

1. Does v fulfill constraints? ✓

2. Is v in the null space of Sv=0 ?

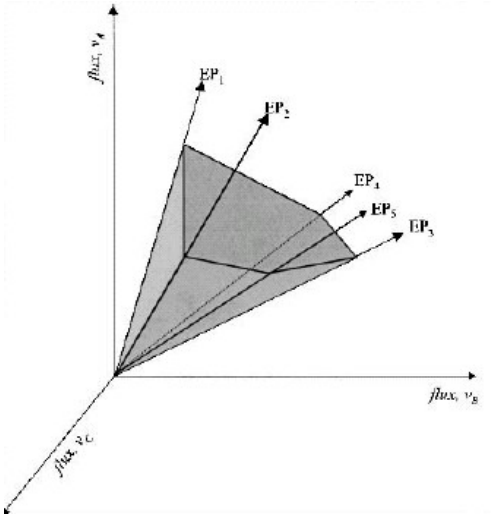
$$\begin{bmatrix} -1 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 \\ 1 & -1 & 1 & 0 & 0 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 4 \\ 2 \\ 0 \\ 1 \\ 0 \\ 1 \\ -4 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$ ✓

(I) Narrowing the steady state flux cone

- The steady state flux cone contains an infinite flux distributions!
- Only a small portion of them is physiologically feasible.
- More constraints on the external fluxes.

These depend on factors as:

- Organism
- Environment and accessibility substrates
- maximum rates of diffusion mediated transport
- Etc…



(II) Calculating optimal flux distribution

- The constrained flux cone in E.coli contains ~10^6 (Schilling 2001)
- How can we identify a "biologically meaningful" flux?

Assumption…

the metabolic network is optimized with respect to a certain objective function Z.

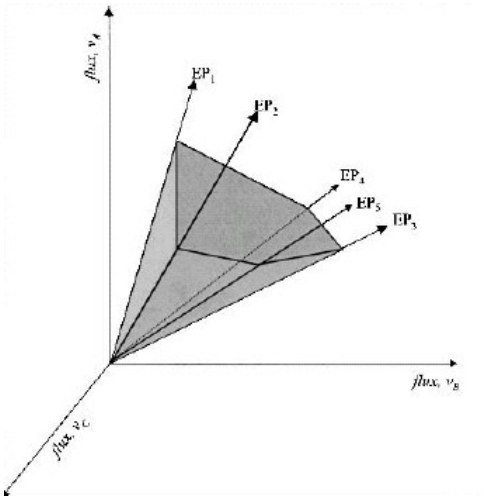
Z will be a linear function. Later, we will deal with how exactly to choose Z

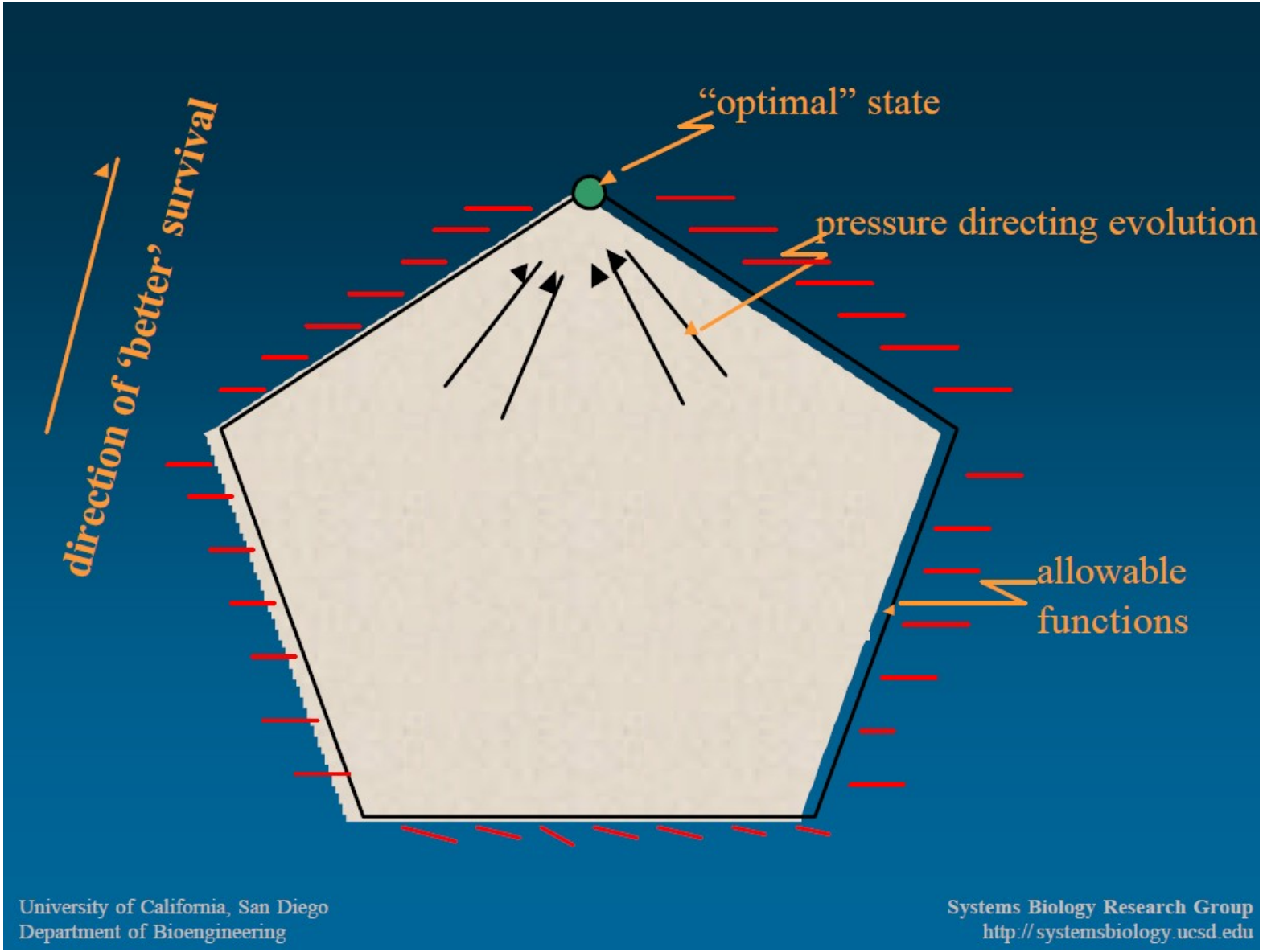
What we want to do is find the vector v in the flux cone which maximizes Z.

…this can be can formulated as an optimization problem:

Minimize/Maximize $Z = \sum_j c_j v_j$ such that $S \cdot v = 0$ + inequality constraints

This optimization problem is a classical linear programming **(LP)** problem that can be solved using the simplex algorithm. W. Wiechert . Journal of biotechnology(2002)



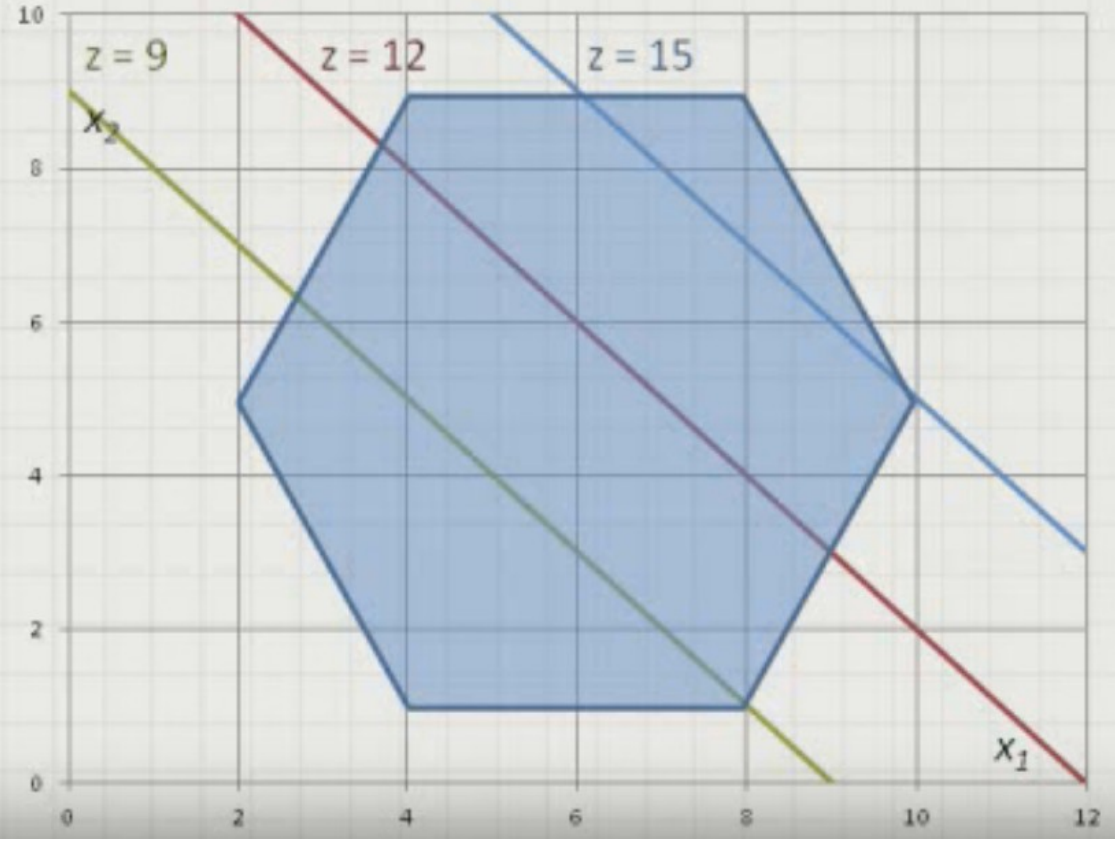

direction of 'better' survival
"optimal" state
pressure directing evolution
allowable functions
University of California, San Diego
Department of Bioengineering
Systems Biology Research Group
http:// systemsbiology.ucsd.edu

Geometrical Interpretation of Linear Programs

Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$
$$x_1 \geq 0$$
$$x_2 \geq 0$$

Maximize $z = x_1 + x_2$



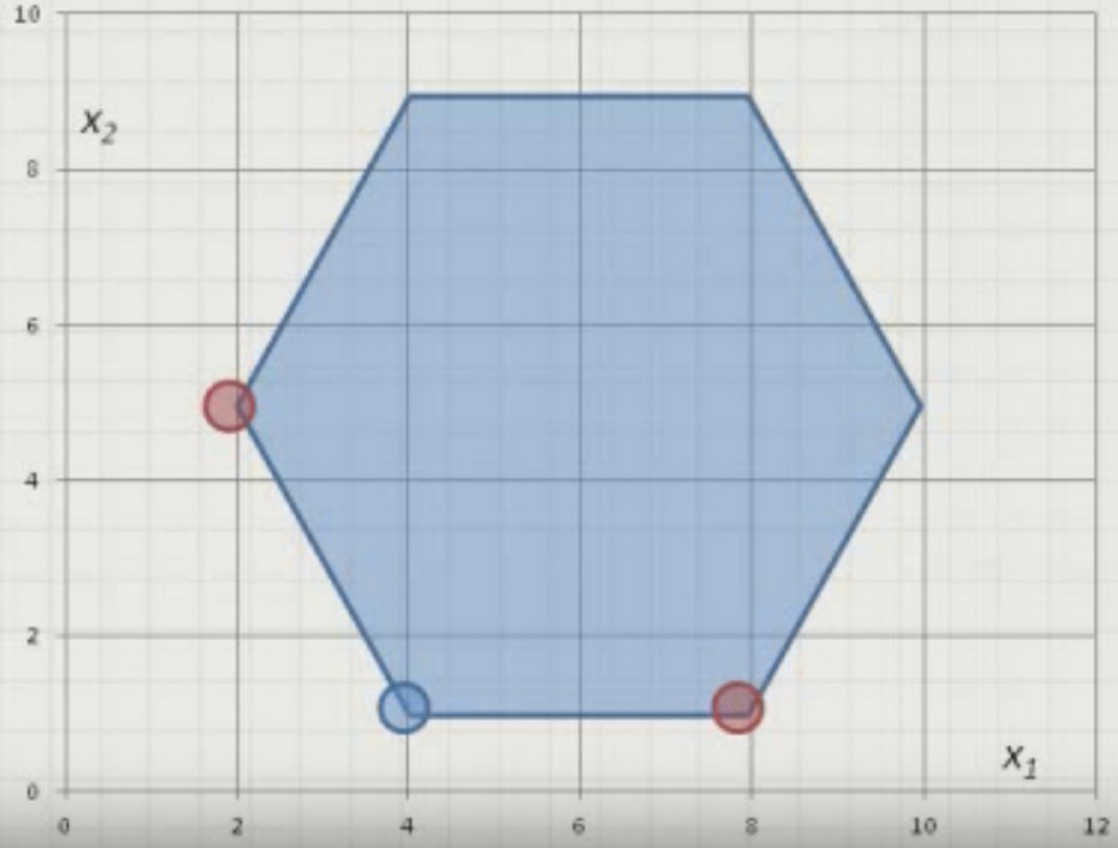
The Simplex Method

Maximize $z = x_1 + x_2$

Neighbors: (2,5) and (8,1)

Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$



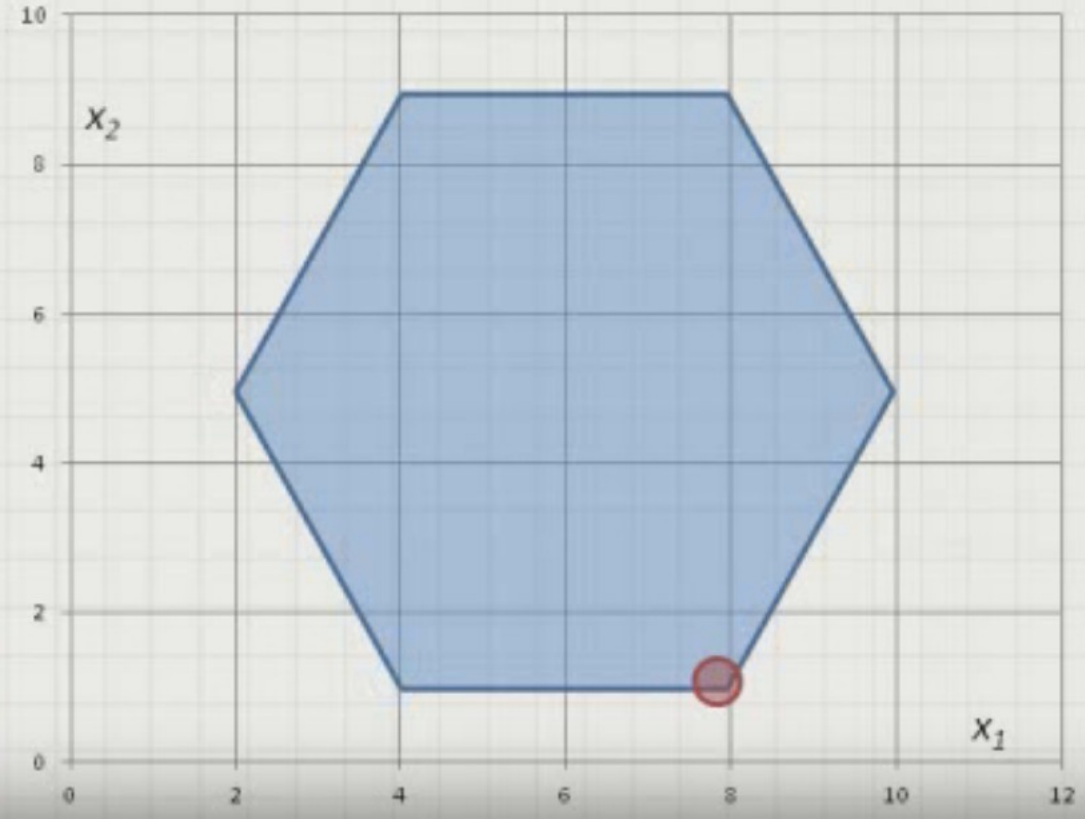
The Simplex Method

Maximize $z = x_1 + x_2$

Current point: $(x_1, x_2) = (8, 1)$

Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$



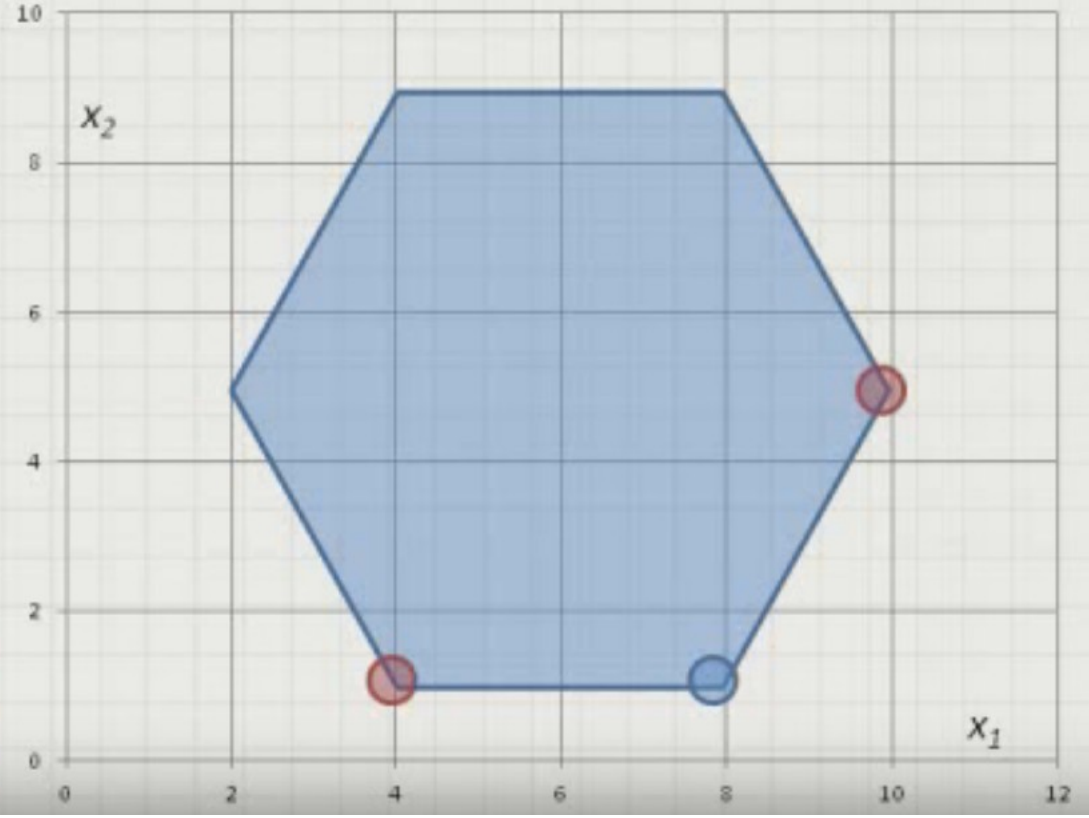
The Simplex Method

Maximize $z = x_1 + x_2$

Neighbors: (4,1) and (10,5)

Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$



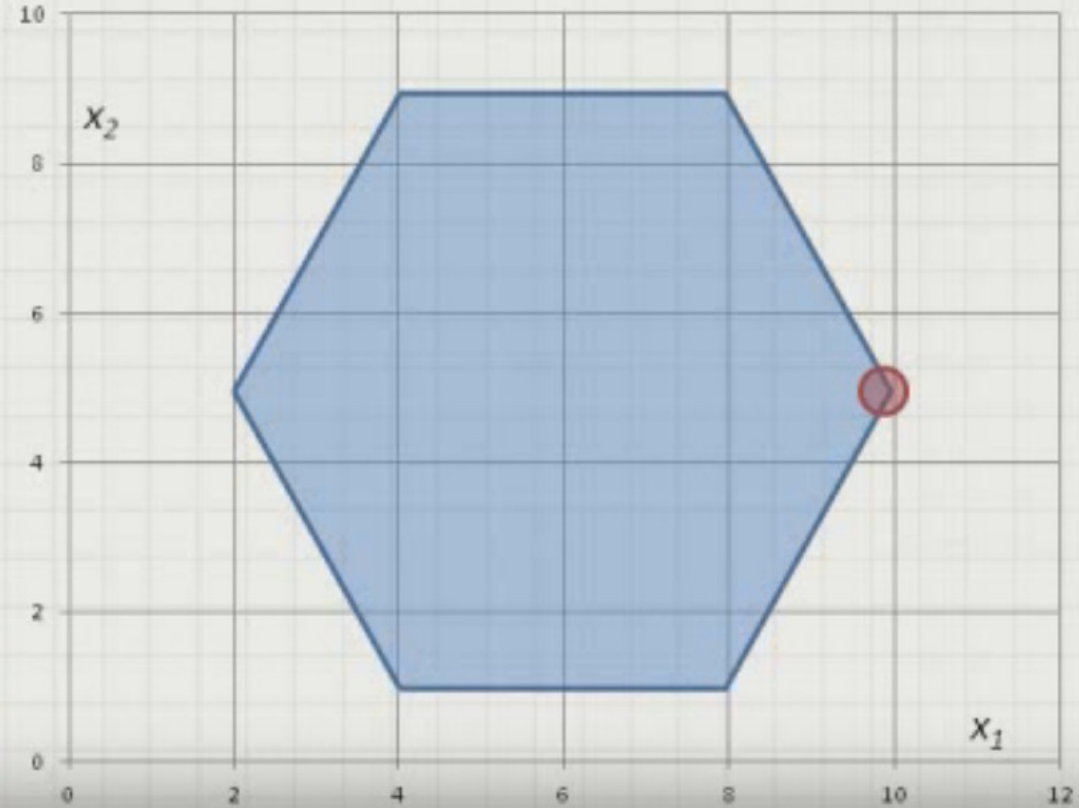
The Simplex Method

Maximize $z = x_1 + x_2$

Current point: $(x_1, x_2) = (10, 5)$

Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$



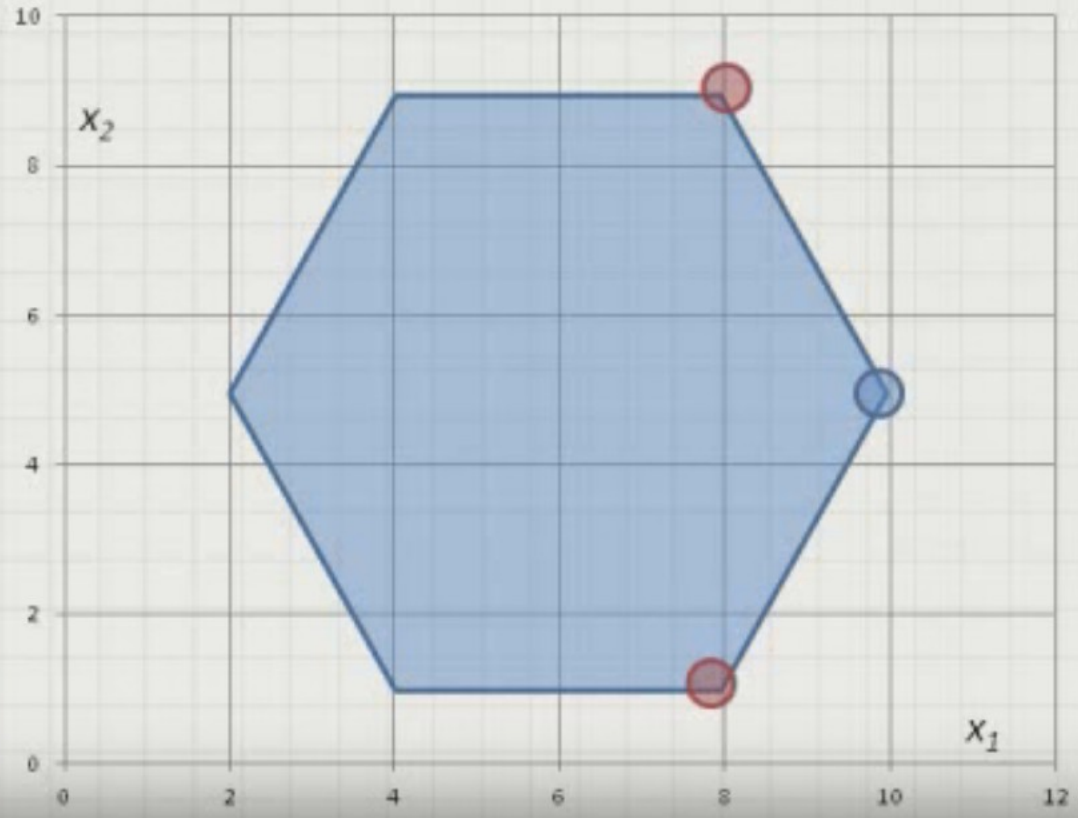
The Simplex Method

Maximize $z = x_1 + x_2$

Neighbors: (8,1) and (8,9)

Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$



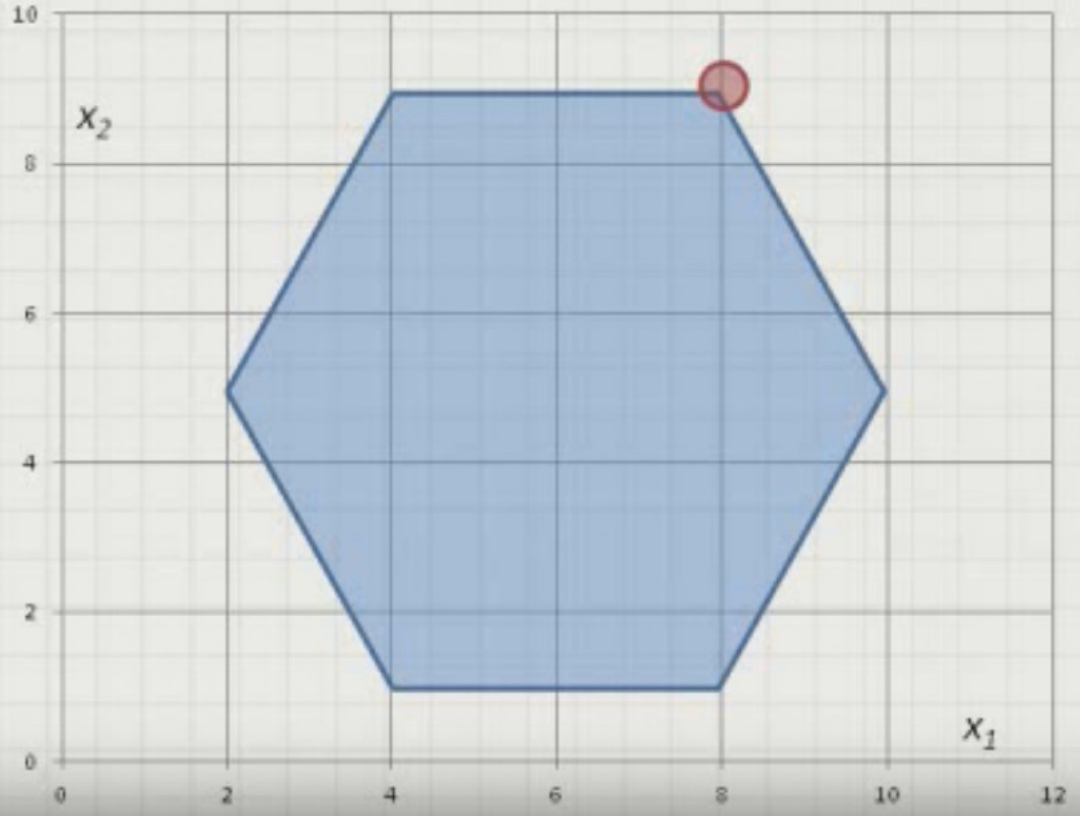
The Simplex Method

Maximize $z = x_1 + x_2$

Current point: $(x_1, x_2) = (8, 9)$

Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$



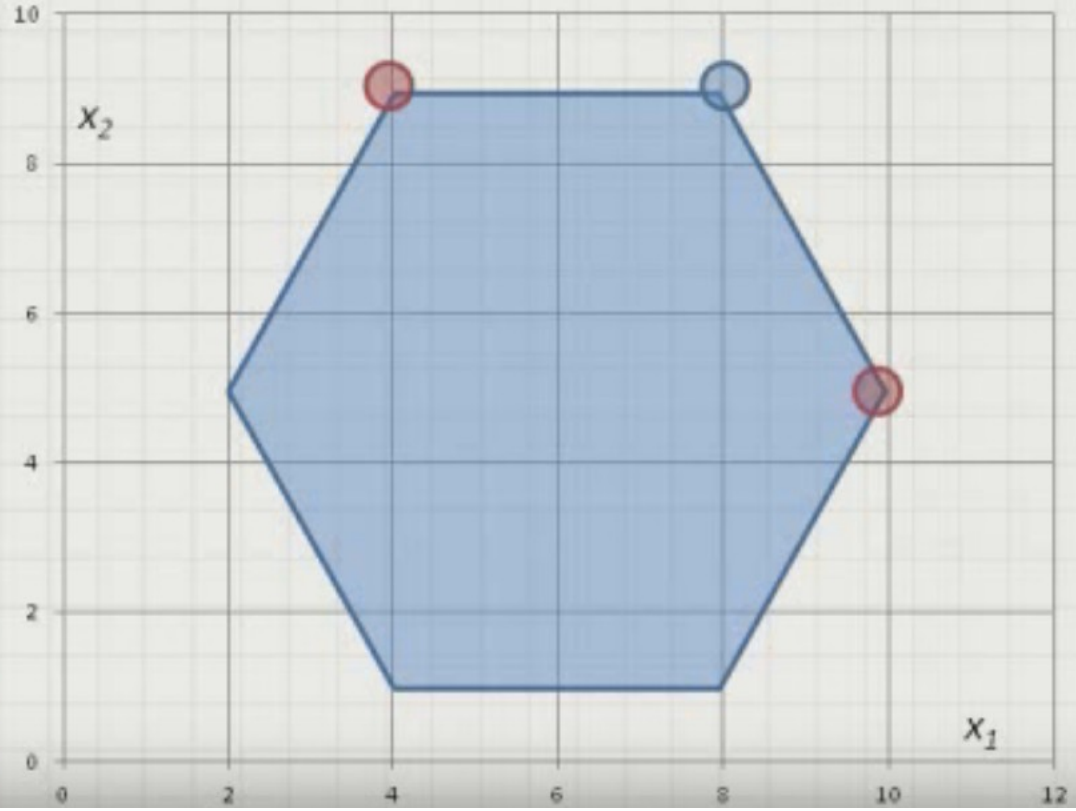
The Simplex Method

Maximize $z = x_1 + x_2$

Neighbors: (10,5) and (4,9)

Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$



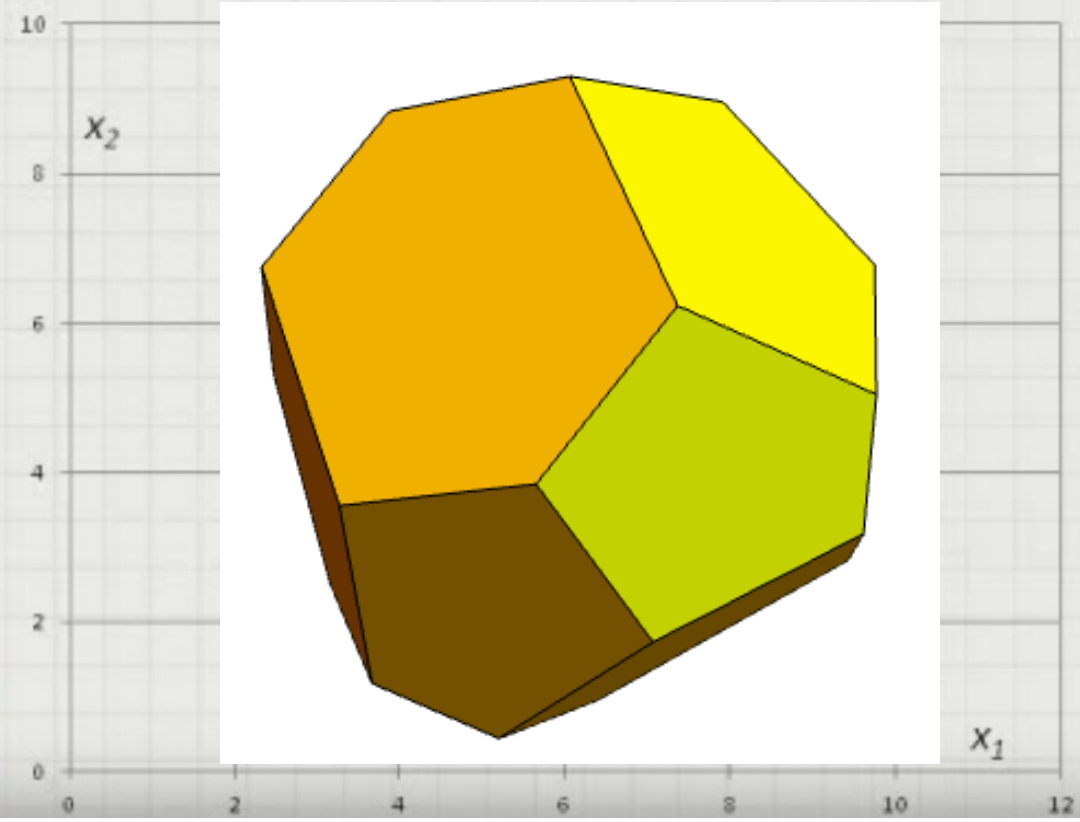
The Simplex Method

Maximize $z = x_1 + x_2$

Optimal solution: $(x_1, x_2) = (8, 9)$

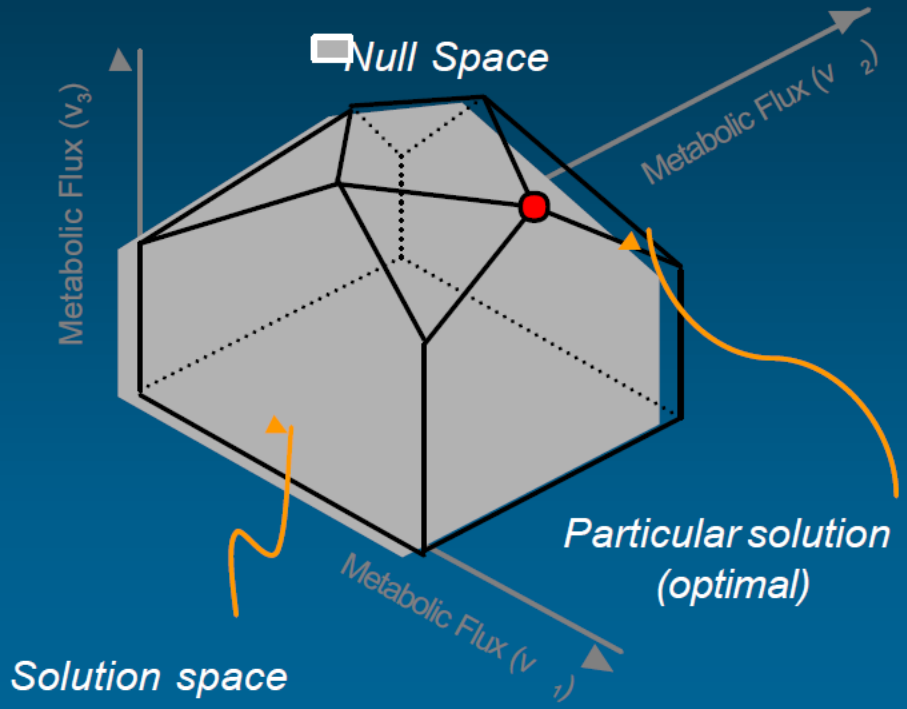
Subject to

$$x_2 \leq 9$$
$$-x_2 \leq -1$$
$$2x_1 + x_2 \leq 25$$
$$-2x_1 - x_2 \leq -9$$
$$-2x_1 + x_2 \leq 1$$
$$2x_1 - x_2 \leq 15$$



Linear Programming: What is it?

finding an optimal solution in a confined space



(III) How to choose the objective function Z

We want to choose a Z that is biologically meaningful.

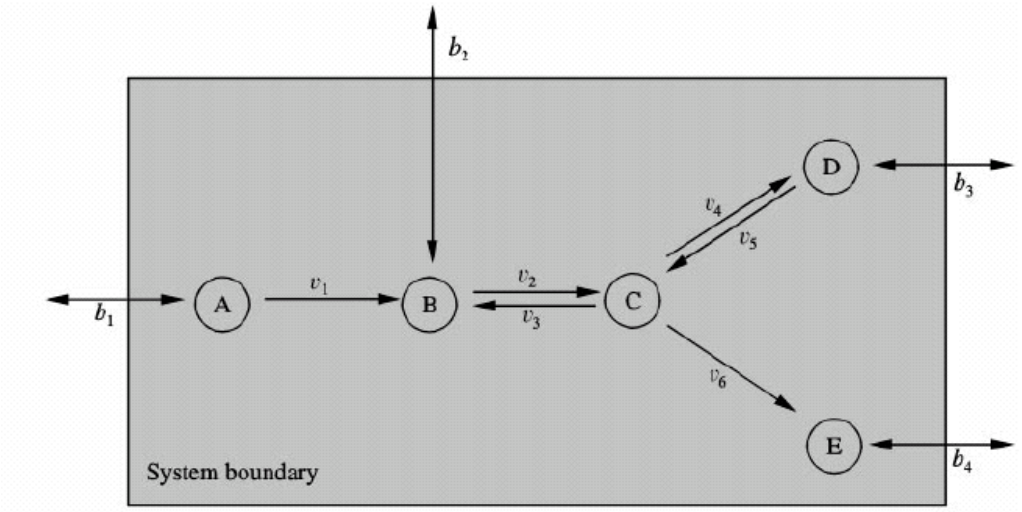
Reasonable options could be:

1. Z: Cellular growth (maximization)
2. Z: Particular metabolite engineering (maximization)
3. Z: Energy consumption (minimization)

Example:

cellular growth is correlated with the production of B,D and 2E.

We want a v that:
(A) Resides in side the cone.
(B) **maximizes Z=B+D+2E.**



(III) How to choose the objective function Z

1. "It has been shown that under rich growth conditions (i.e. no lack of phosphate and nitrogen), E. Coli grows in a stoichiometrically optimal manner." (Schilling 2001, Edwards 1994)

2. "It is reasonable to hypothesize that unicellular organisms have evolved toward maximal growth performance." (Segre, 2002.)

We shall use Z which reflects:

Cellular Growth

The growth requirements

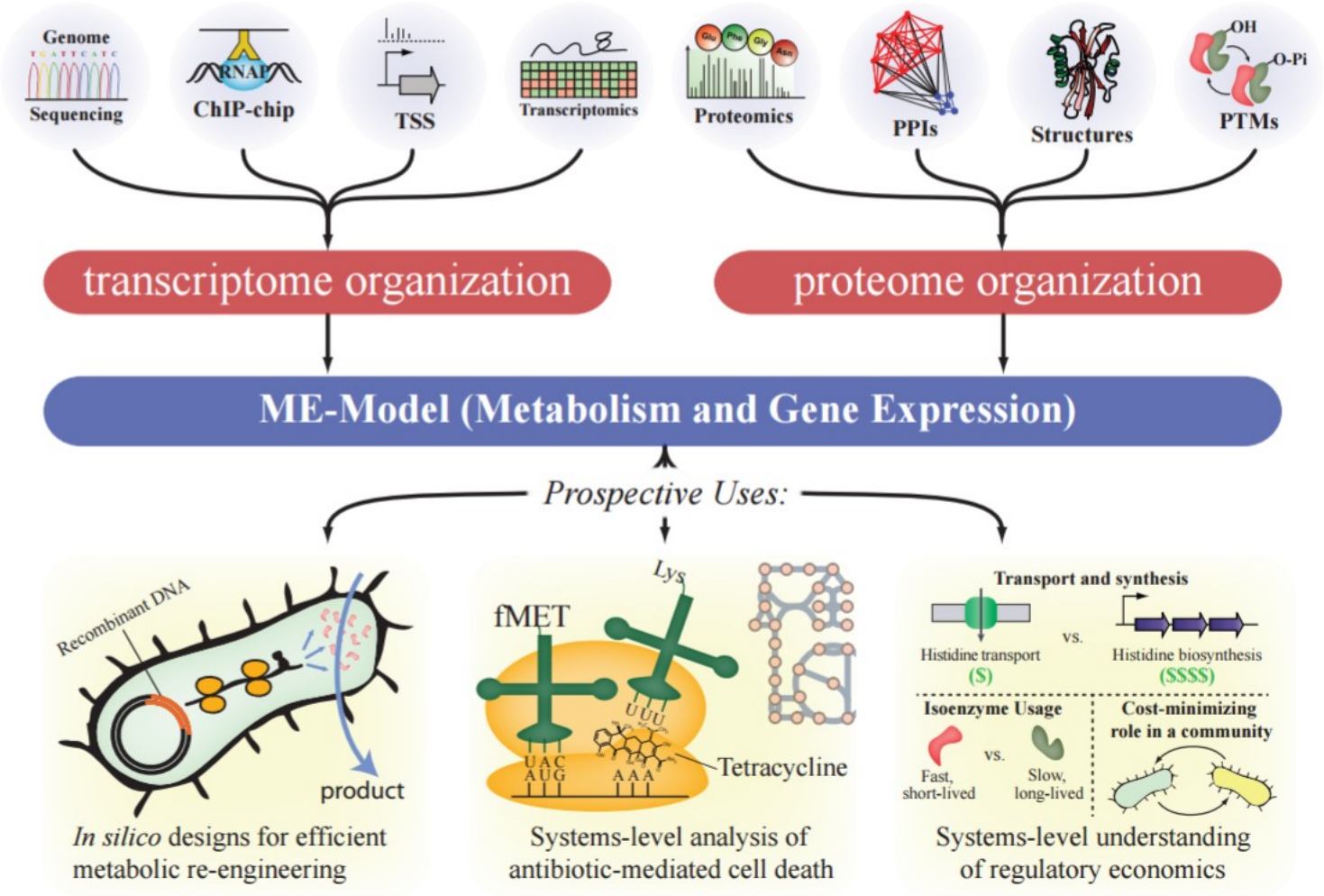
Metabolic demands of precursors and cofactors required for 1 g of biomass of *E. coli.*

These precursors are removed from the metabolic network in the corresponding ratios.

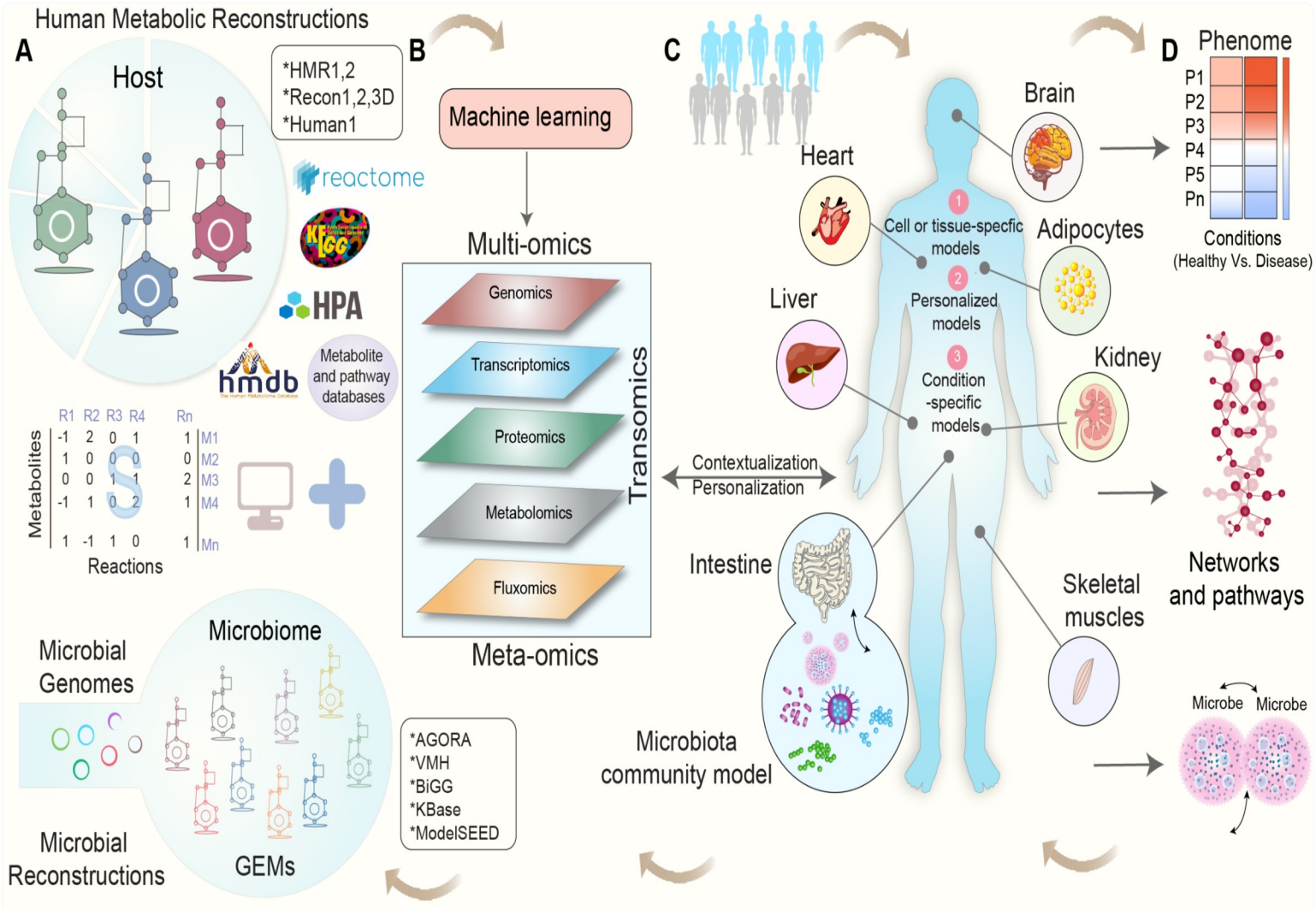
Thus, the objective function is:

$$Z = 41.2570\, v_{ATP} - 3.547 v_{NADH} + 18.225 v_{NADPH} + \ldots.$$

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


Genome
Sequencing
ChIP-chip
RNAP
TSS
Transcriptomics
Proteomics
PPIs
Structures
OH
O-Pi
PTMs
transcriptome organization
proteome organization
ME-Model (Metabolism and Gene Expression)
Prospective Uses:
Recombinant DNA
product
In silico designs for efficient metabolic re-engineering
Lys
fMET
UUU
UAC
AUG
AAA
Tetracycline
Systems-level analysis of antibiotic-mediated cell death
Transport and synthesis
Histidine transport
(S)
vs.
Histidine biosynthesis
($$$$)
Isoenzyme Usage
Fast, short-lived
vs.
Slow, long-lived
Cost-minimizing role in a community
Systems-level understanding of regulatory economics

Workflow for context-specific genome-scale metabolic reconstructions using multi-omics data



From: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10383060/pdf/metabolites-13-00855.pdf

OptFlux demo

- # 1st time installation
- mkdir ~/tools
- cd ~/tools
- cp /home/reczko/ecoli1633.xml .
- wget http://genomics-lab.fleming.gr/fleming/uoa/vm/OptFlux-3.3.0-linux-x64-installer.run

- chmod u+x OptFlux-3.3.0-linux-x64-installer.run

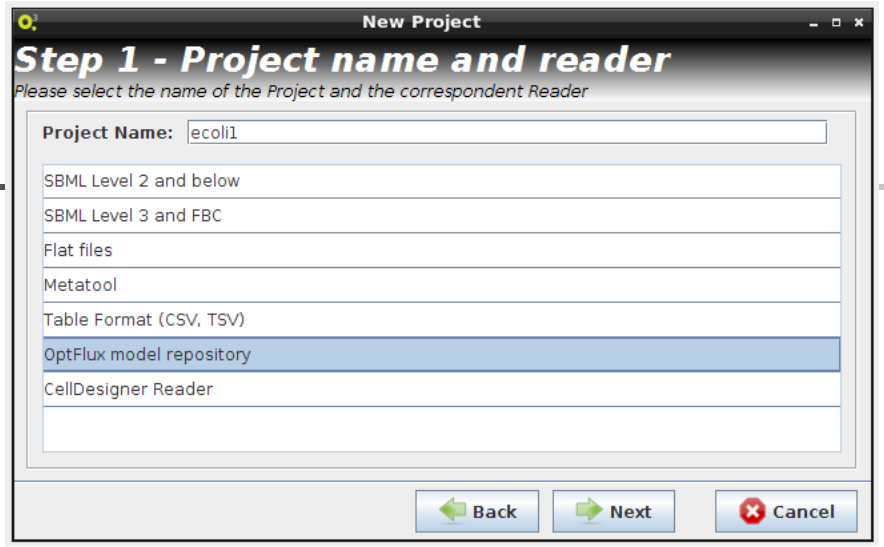
- ./OptFlux-3.3.0-linux-x64-installer.run
- # accept the license, click 'run after installation', don't accept the update options

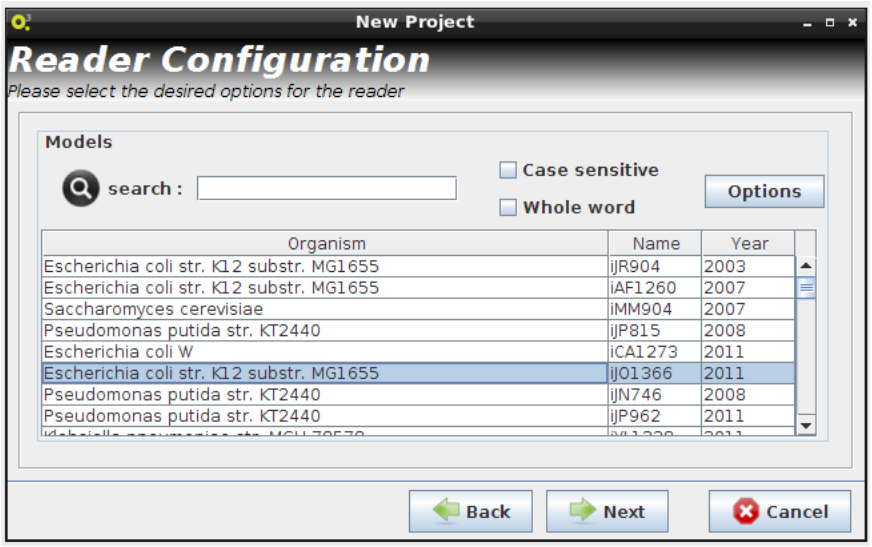
- # run again after installation

- cd ~/OptFlux-3.3.0
- ./optflux.sh

- # Full tutorial at: http://www.optflux.org/wiki/tutorial/TutorialOptFlux3.pdf
- Or at https://web.archive.org/web/20240619164025/http://www.optflux.org/wiki/tutorial/TutorialOptFlux3.pdf

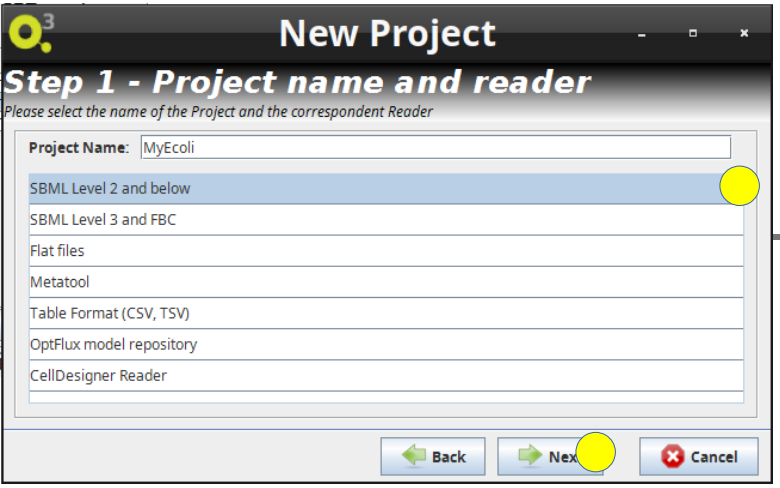
File → New Project → OptFlux model repository → Next

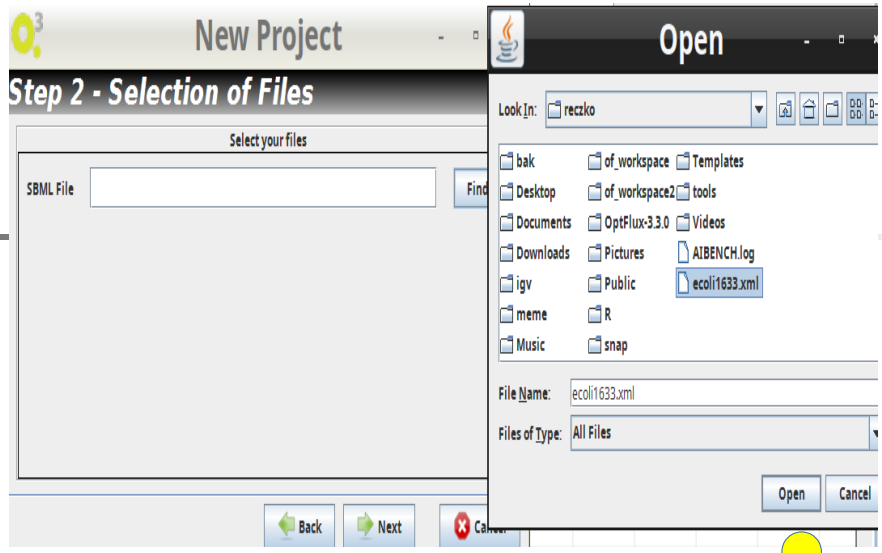


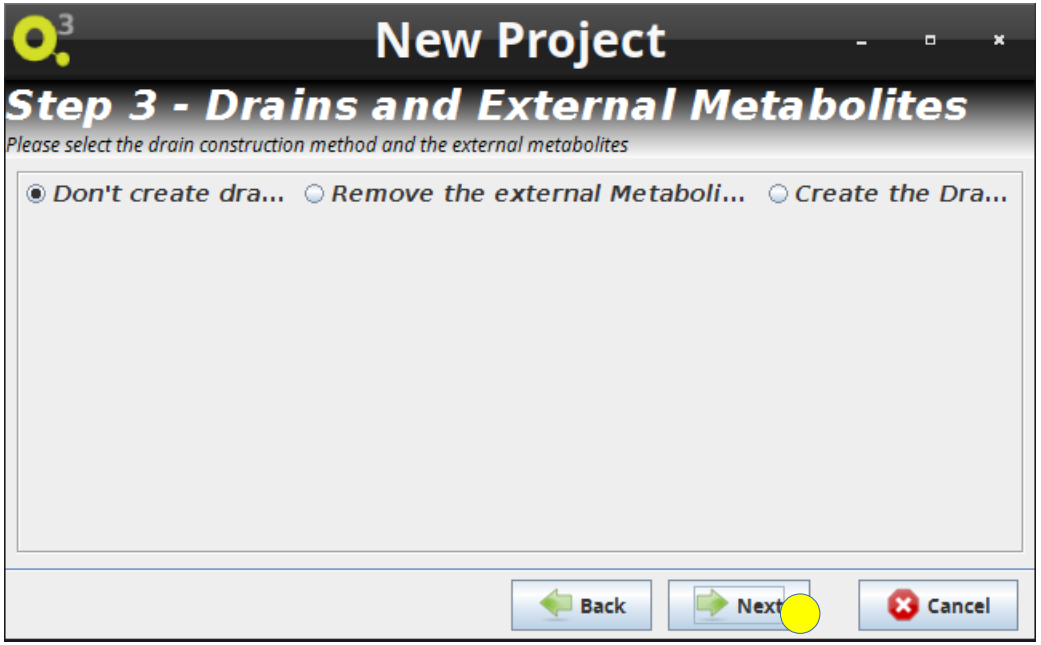


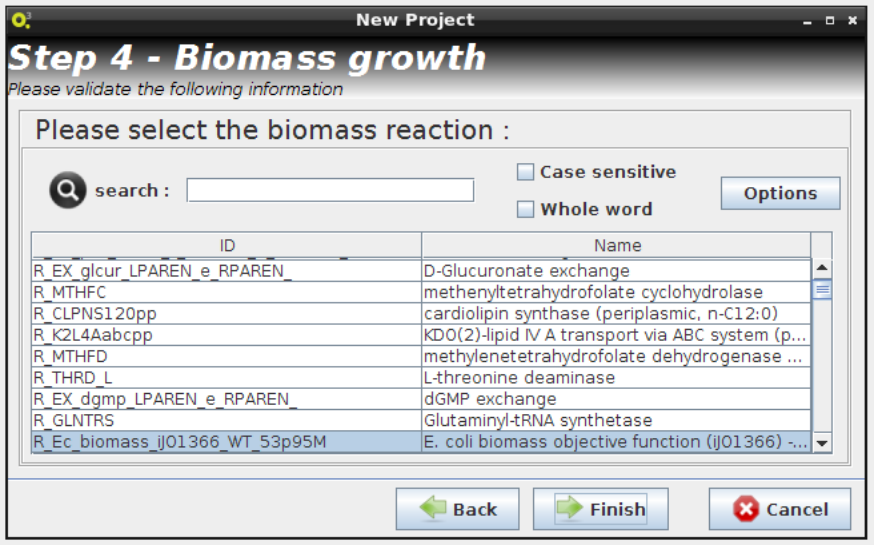
If repository server is unavailable, load sbml as on the next slide:

File → New Project → SBML Level 2 and below → Next

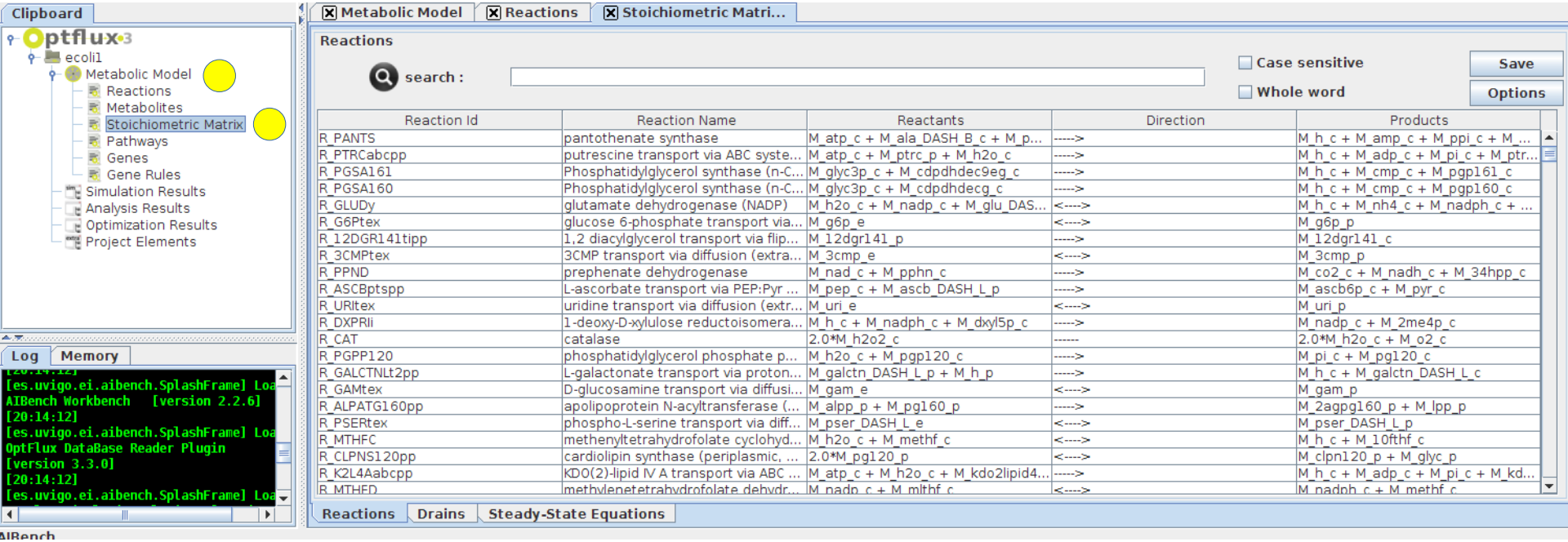


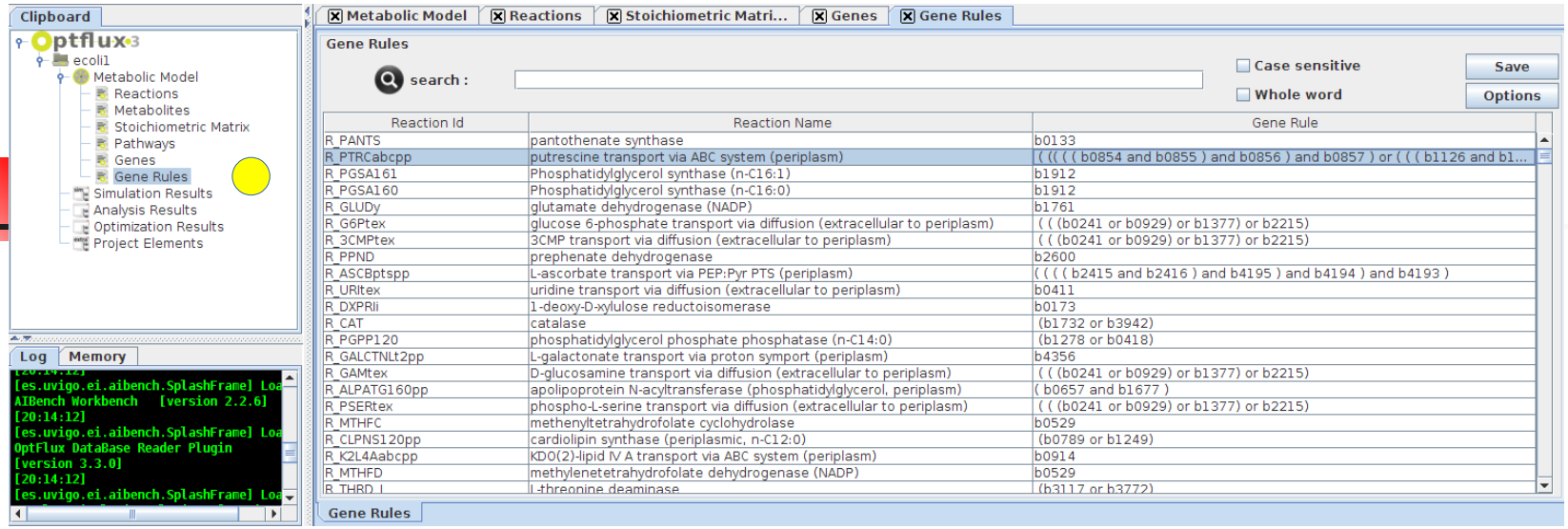




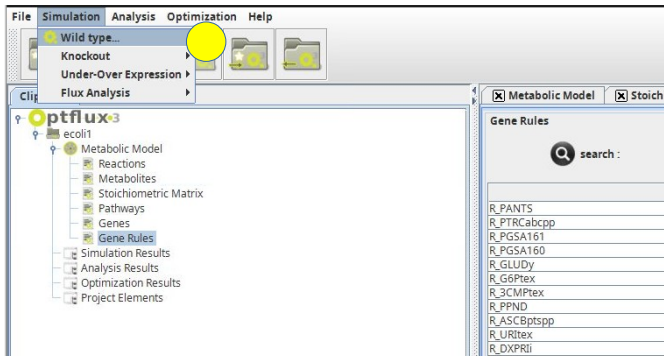


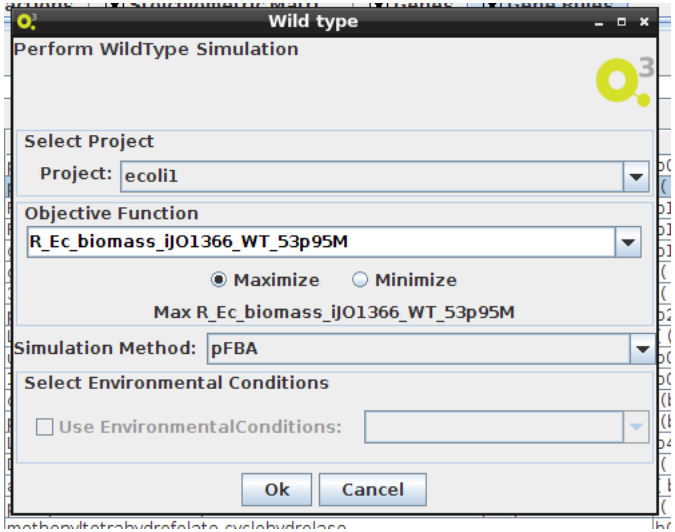
Browse model



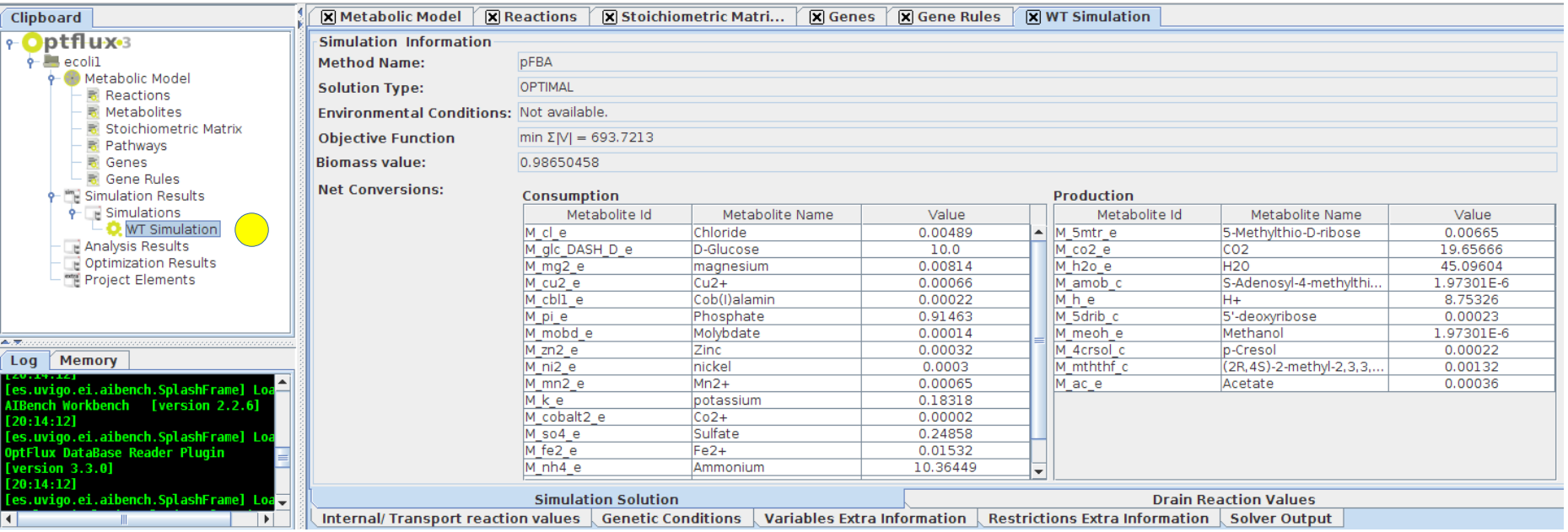


Simulation → Wild type

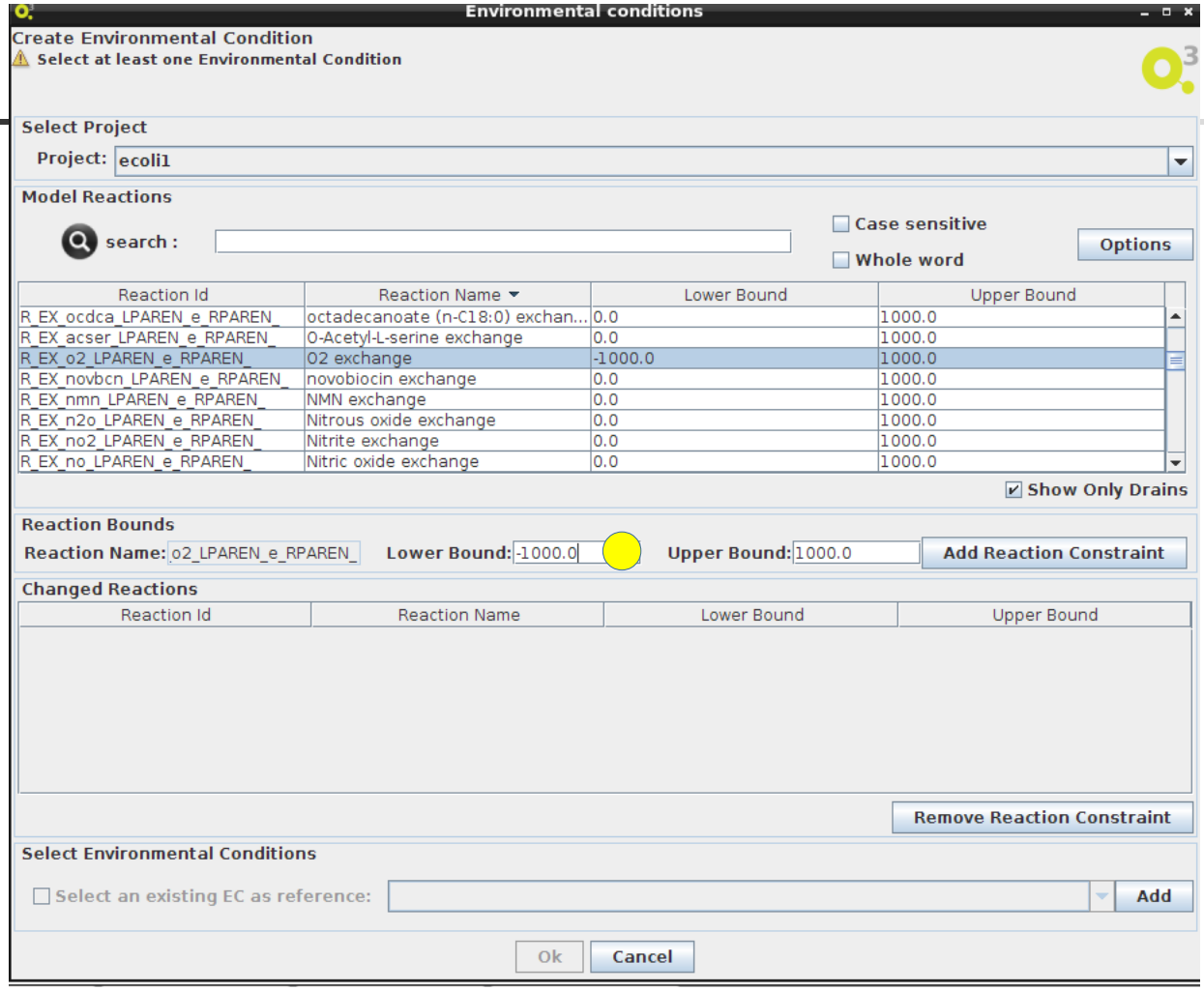




Results at Simulations → WT Simulation

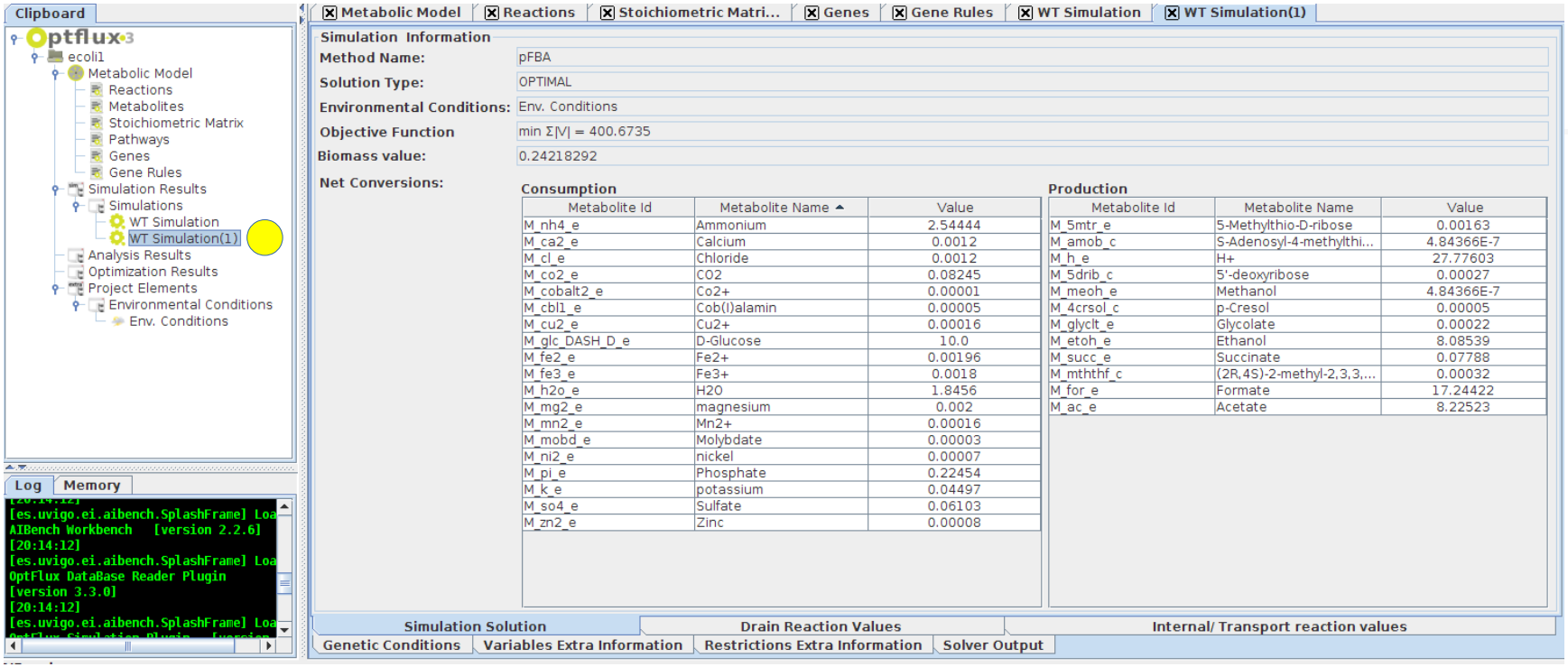


File → Create → Environmental conditions...



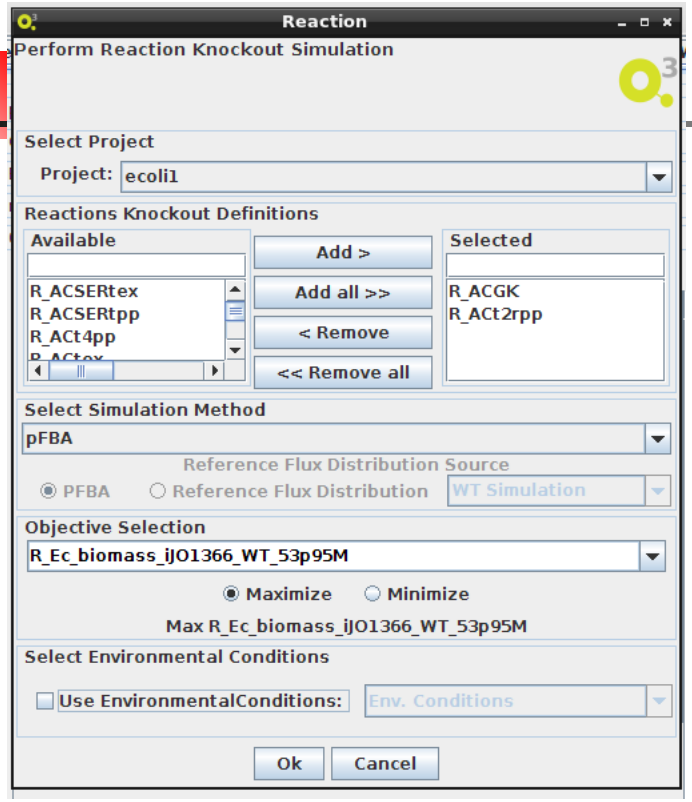
Lower Bound → 0.0 → Add Reaction Constraint → Ok

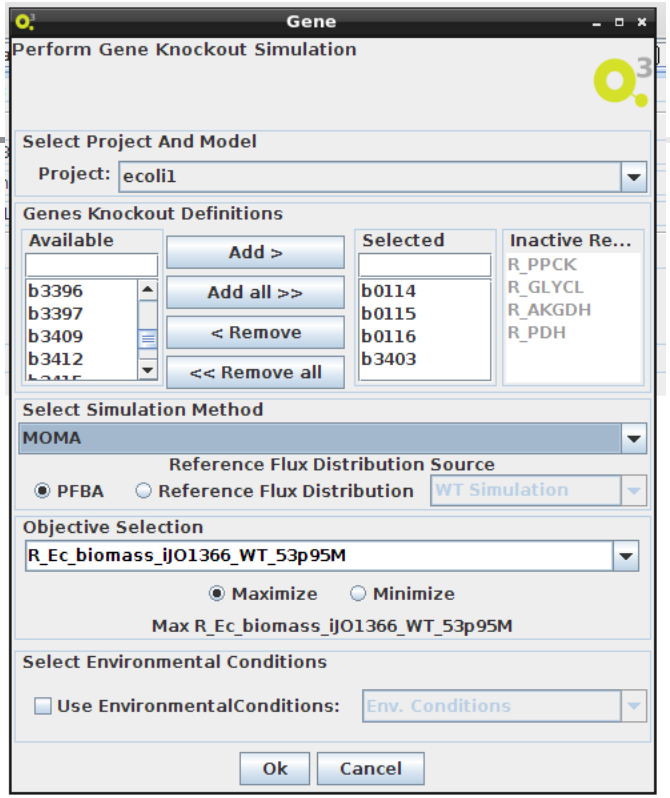
Simulation → Wild type → Check 'Use Environmental Conditions' → Ok



Compare Consumption,Production with aerobic simulation

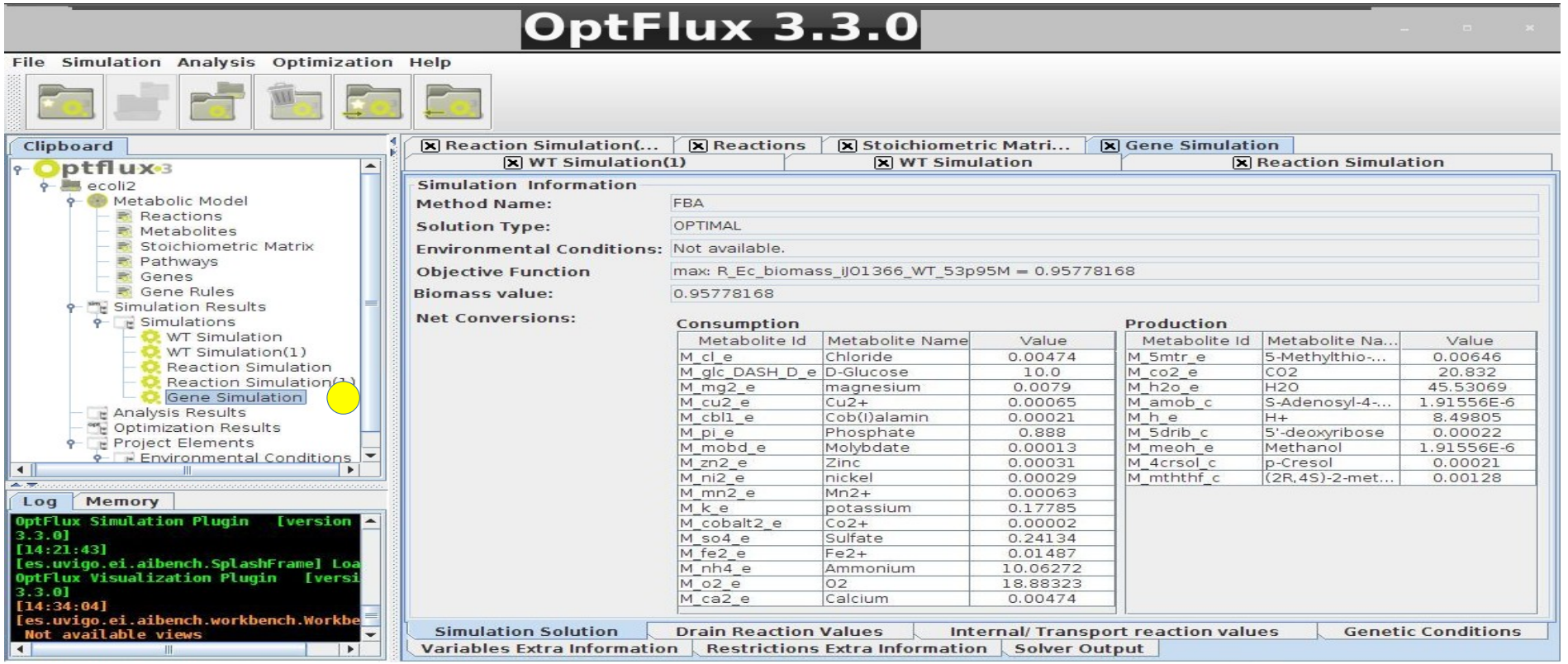
Simulation → Kockout → Reaction/Gene





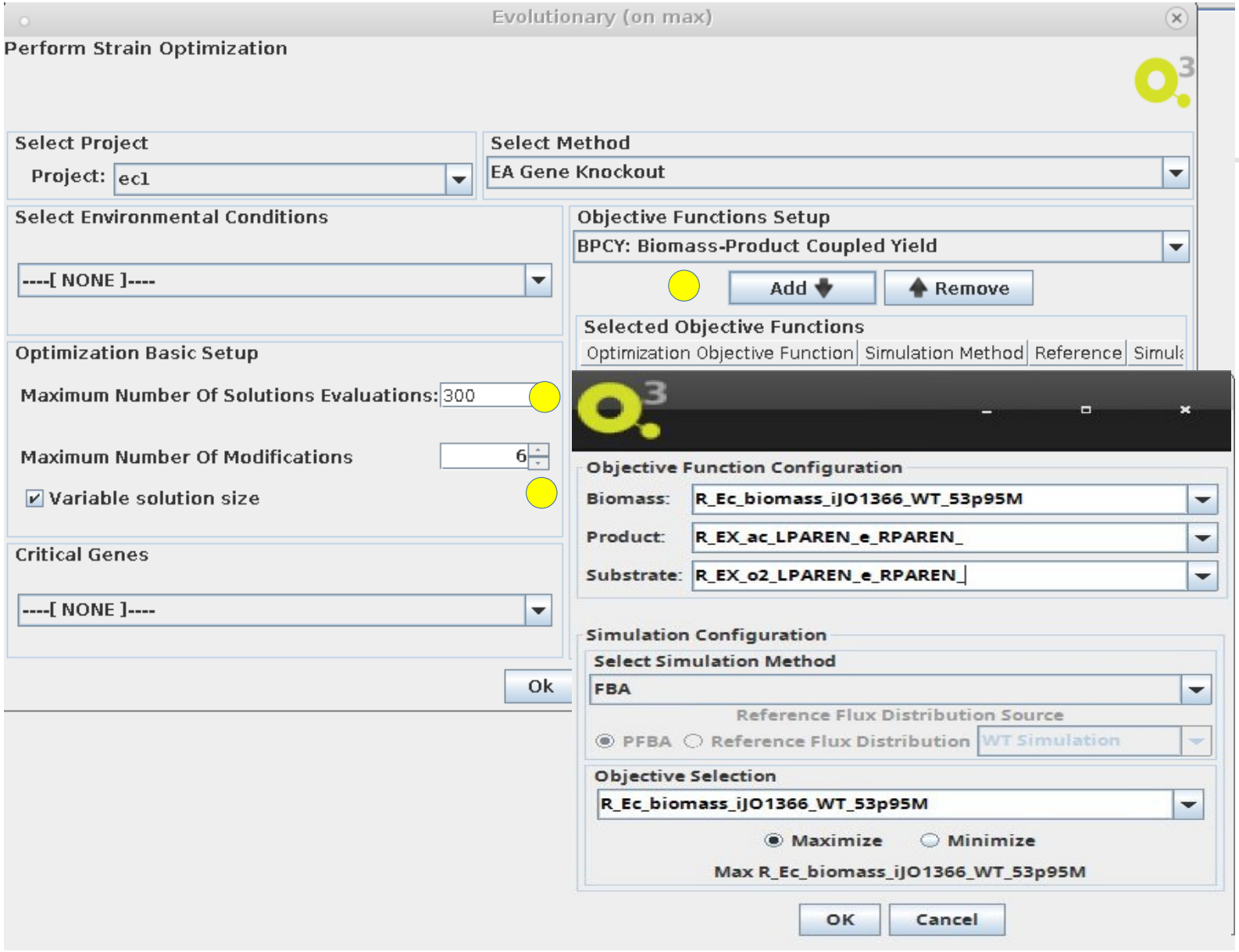
Knockout acetylglutamate_kinase,...

Knockout genes that knockout R_PPCK,...

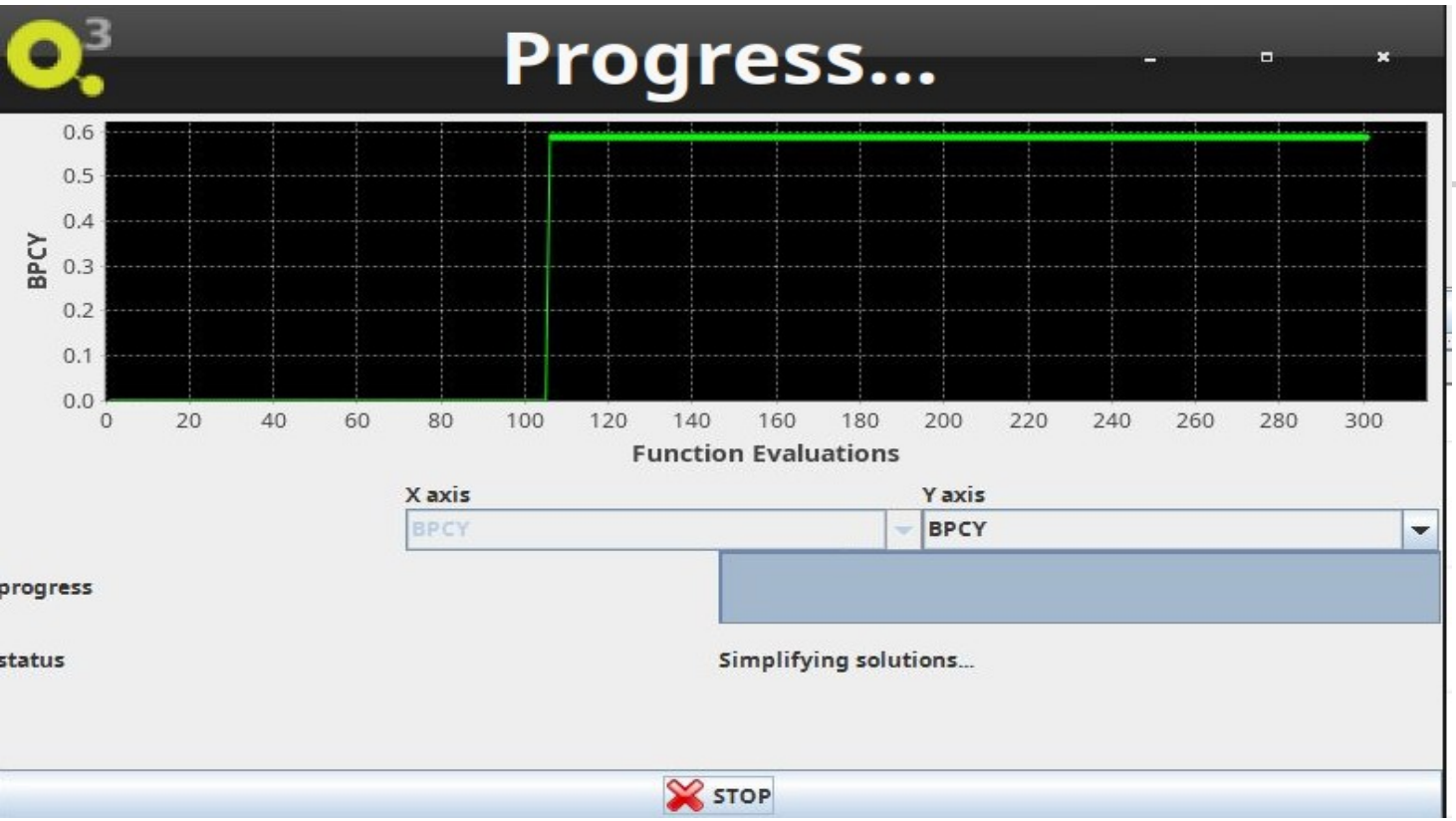


Compare Consumption,Production with WT Simulation:
no acetate production

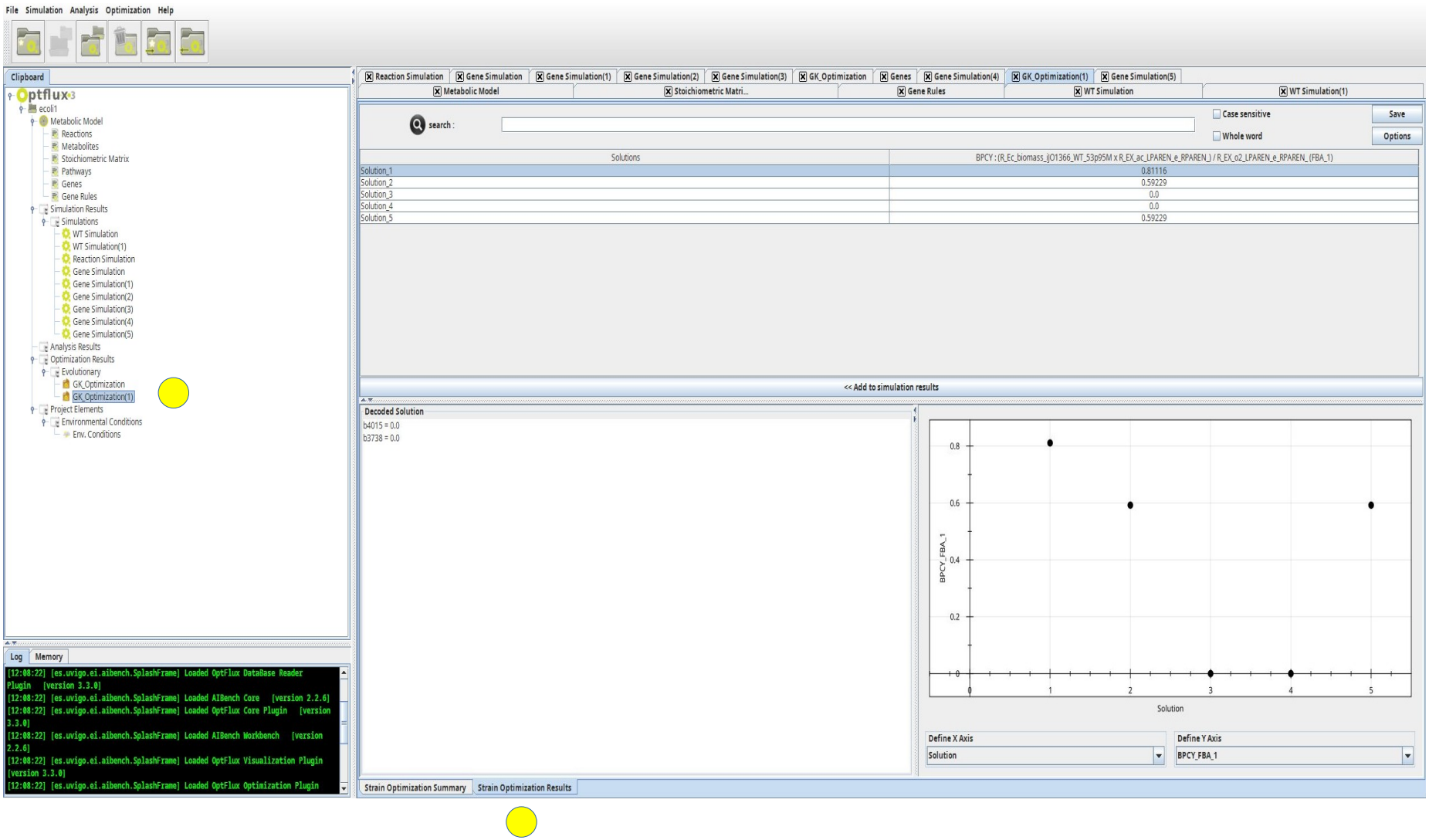
Find optimized strain: Optimization → Evolutionary



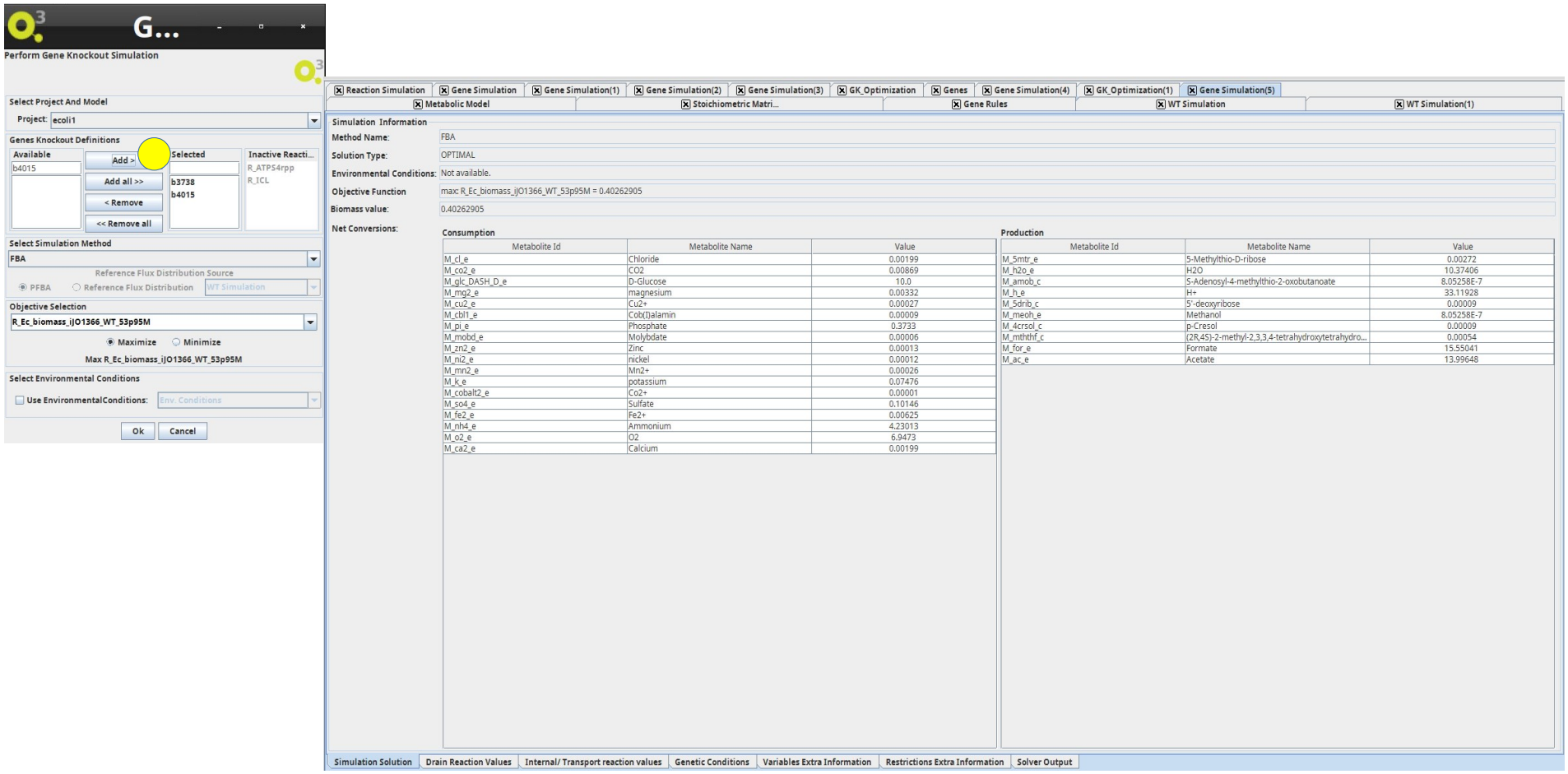
Optimization progress (not deterministic)



Optimization result



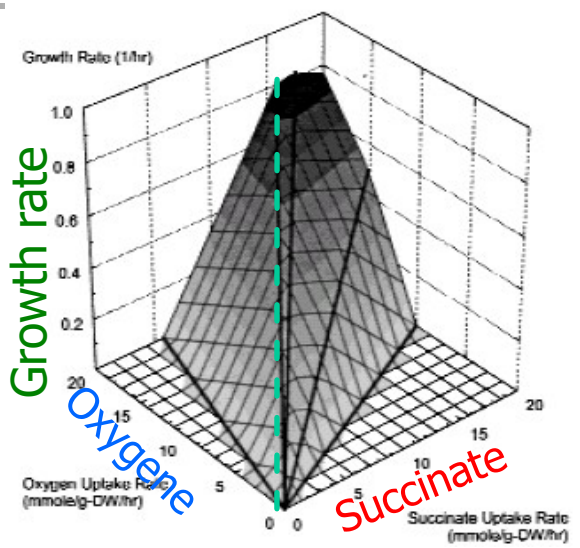
>38800 fold increased acetate production (M_ac_e):



(IV) Phenotype phase planes- PPP

Predicting cellular growth

Schilling 2001



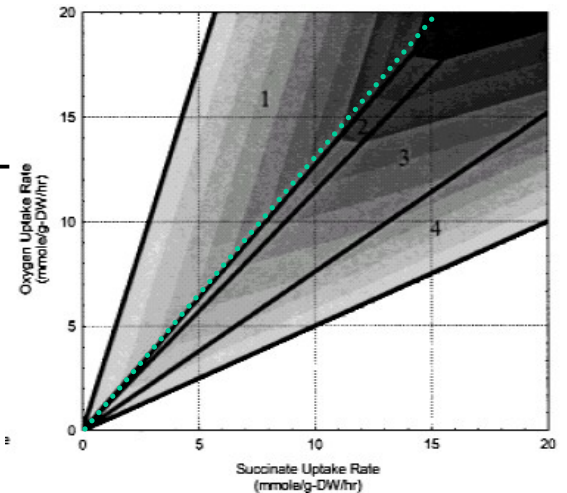
X axis – Succinate uptake rate

Y axis – Oxygene uptake rate

Z axis - Growth rate (maximal value of the objective function as function of succinate and oxygen uptake)

Observations:

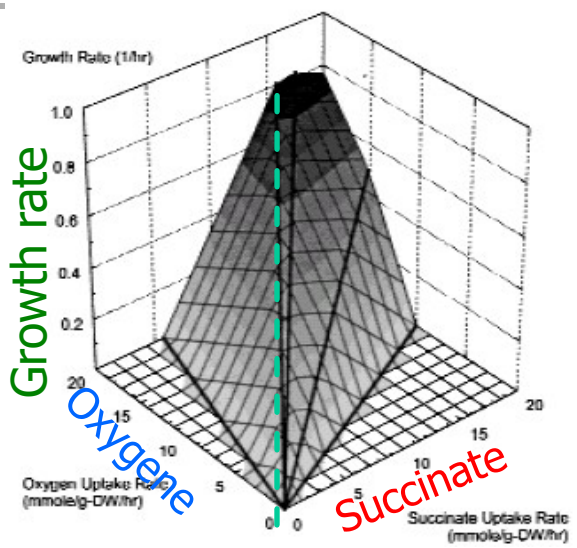
- Metabolic network is unable to utilize succinate as sole carbon source in anaerobic conditinos.
- **Region 1**: oxygen excess – this region is wasteful – (less carbon is available for biomass production since it is oxidized to eliminate the excess oxygen.)
- **Line of optimality**



(IV) Phenotype phase planes- PPP

Predicting cellular growth

Schilling 2001



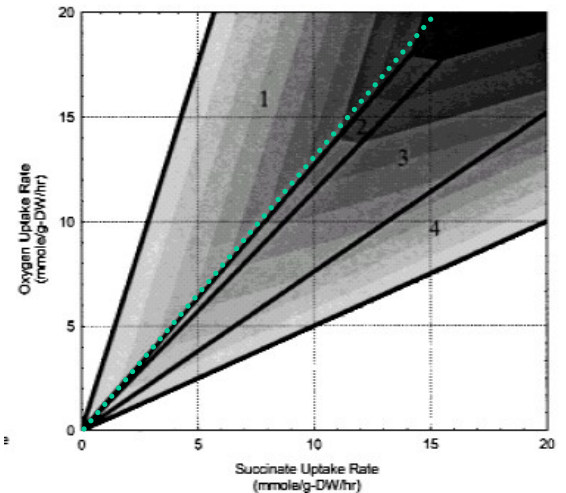
X axis – Succinate uptake rate

Y axis – Oxygene uptake rate

Z axis - Growth rate (maximal value of the objective function as function of succinate and oxygen uptake)

Observations:

- **Region 2** – limitation on both oxygen and succinate
- **Region 3-** the uptake of additional succinate has a negative effect. Cellular resources are required to eliminate excessive succinate.



Model vs. biological experiments

Does E. coli behave according to optimal behavior predictions?

- E. coli was grown with malate as sole carbon source.
- A range of substrate concentrations and temperatures were used in order to vary the malate uptake rate (MUR).
- Oxygen uptake rate (OUR) and growth rate were measured

.

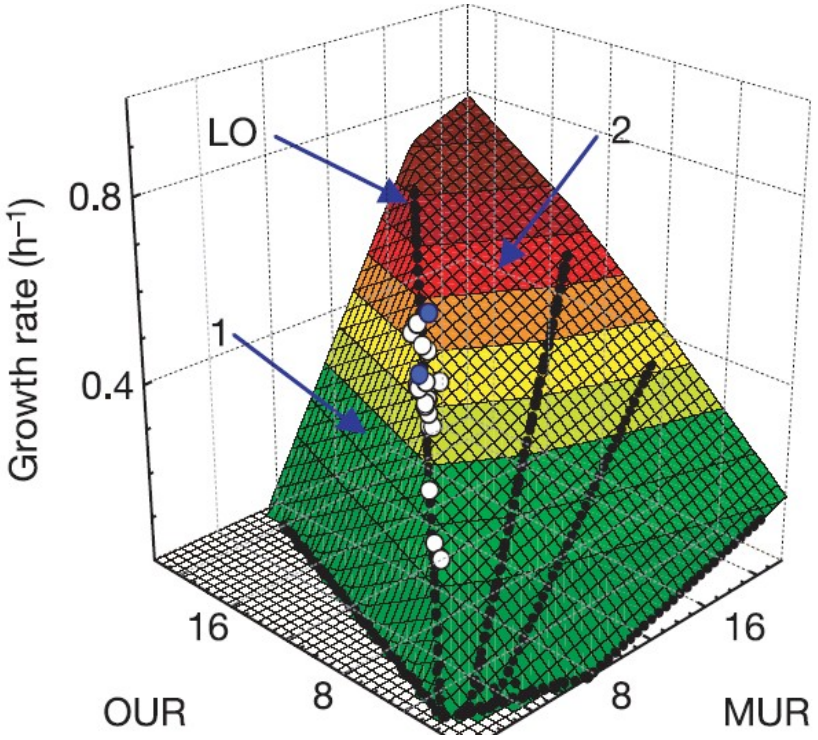
.

.

Does E. coli behave according to optimal behavior predictions?

1- The experimentally determined growth rate were on the line of optimality of the PPP !

Malate/oxygen PPP



Ibarra et al., Nature 2002

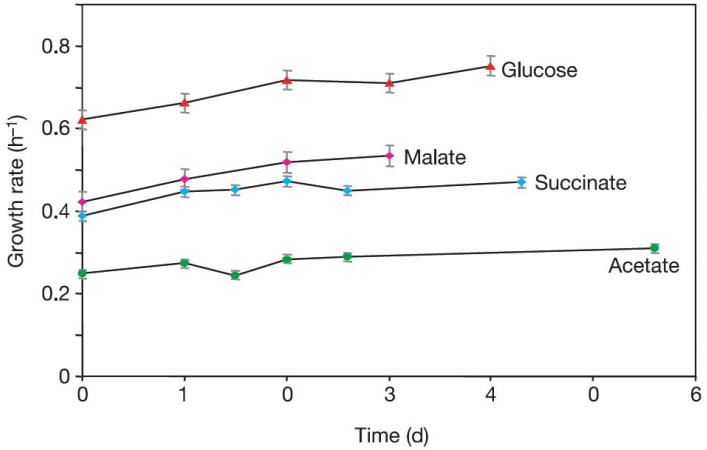
Does E. coli behave according to optimal behavior predictions?

Is the optimal performance on malate stable over prolonged periods of time?

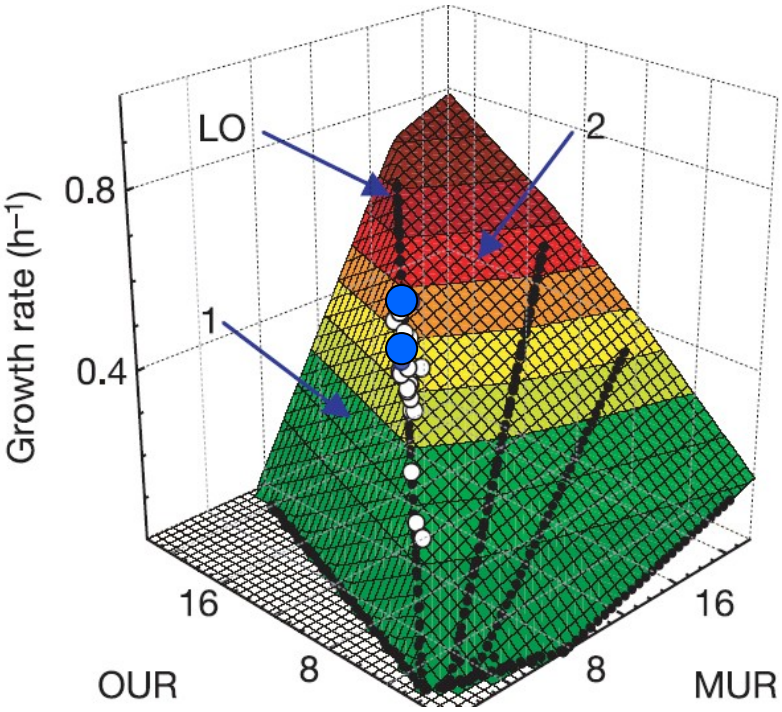
Evolution of E. coli on malate was studied for 500 generations in a single condition…

2- An adaptive evolution was observed with an increase of 19%in growth rate!

3- Same adaptive evolution was observed for succinate and Malate!



Malate/oxygen PPP



Ibarra et al., Nature 2002

Does E. coli behave according to optimal behavior predictions?

Same experiments were made using glycerol as sole carbon source

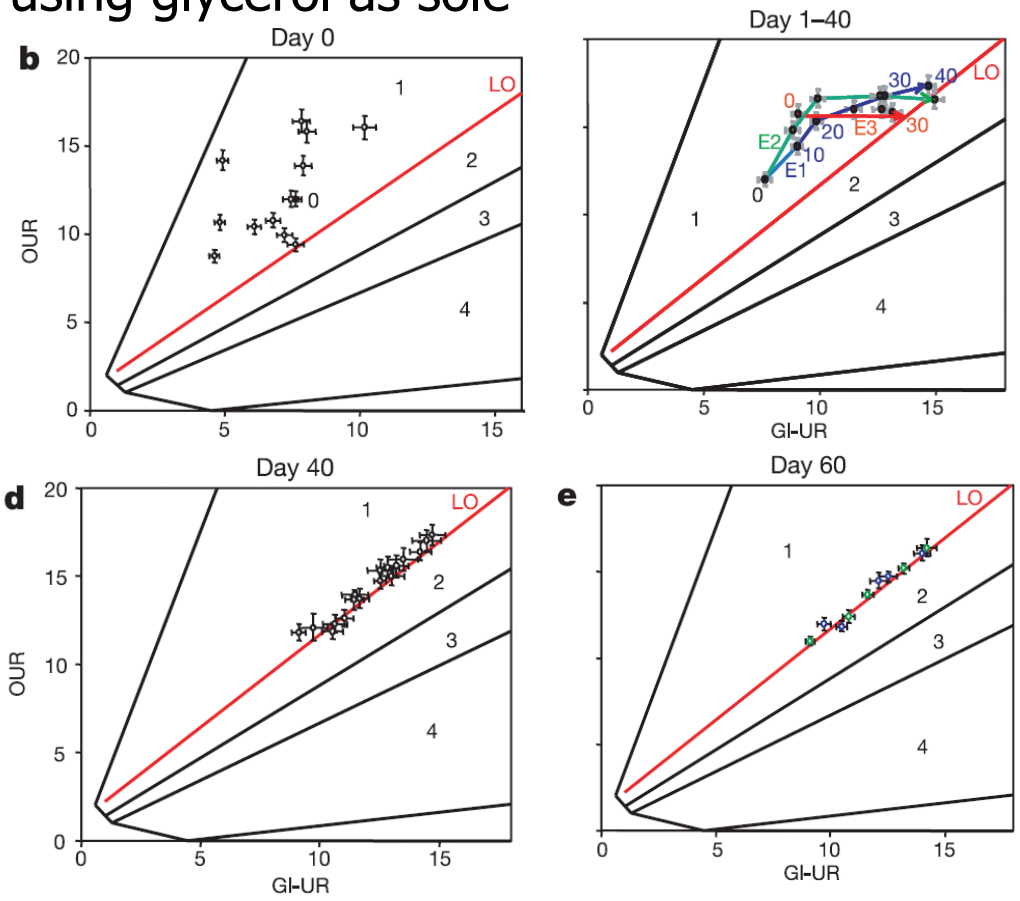
Day 0 – Sub optimal growth

Why?

Day 1-40 – evolution toward optimal growth

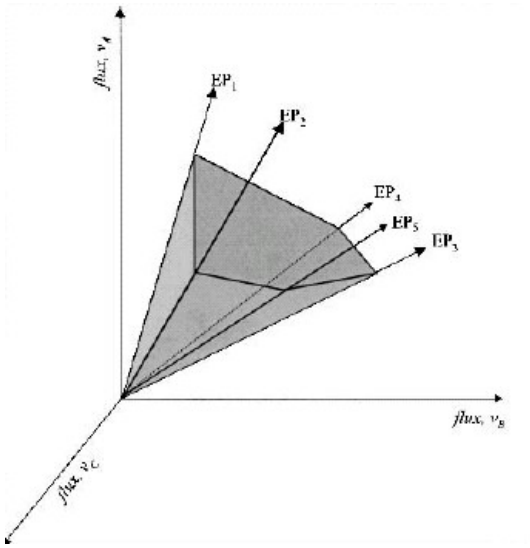
Day 40 –optimal growth

Day 60 –optimal growth (no change)



Considering instances where FBA predictions are inaccurate… **MOMA**

- What happens to the metabolism in the case of a mutation/genetically engineered bacteria?
- What happens in terms of the flux cone?



$$0 = S \cdot v$$

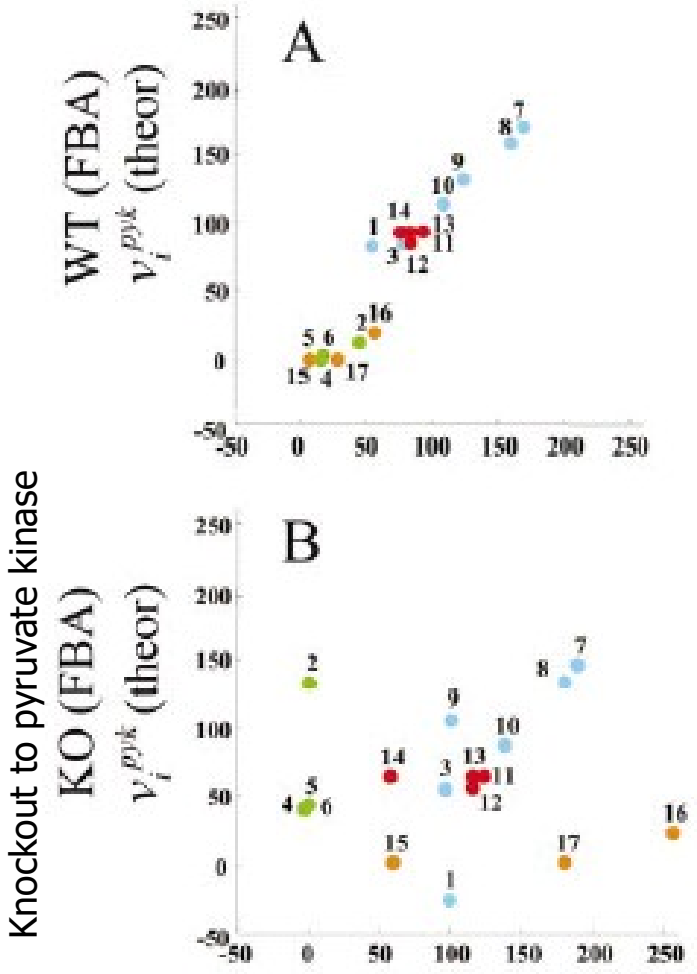
$$\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix} = \begin{bmatrix} -1 & -1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & -1 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} v_1 \\ 0 \\ v_3 \\ v_4 \\ 0 \\ b_2 \\ b_3 \end{bmatrix}$$

S

V

Considering instances where FBA predictions are inaccurate…

- FBA – assumes optimality of growth for wild type
- This assumption is not necessarily correct some instances…



Considering instances where FBA predictions are inaccurate… **MOMA**

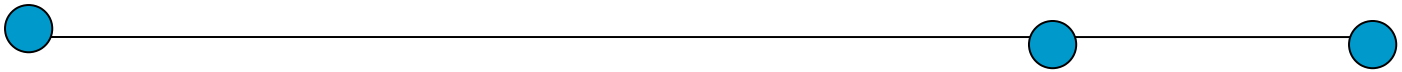
- Is there any other objective function Z that can capture the biological essence of these mutations?
- Perhaps another model…

MOMA - minimization of metabolic adjustments

[Segre, Vituk and Church 2002]

MOMA

- Uses the same steady state flux cone as FBA.
- Relaxes the assumption of maximal optimal growth.
- a mutant is likely to display a suboptimal flux distribution between wild-type optimum and mutant optimum.



Mutant optimal growth

Mutant growth actual

Wild-type growth

FBA

MOMA

FBA

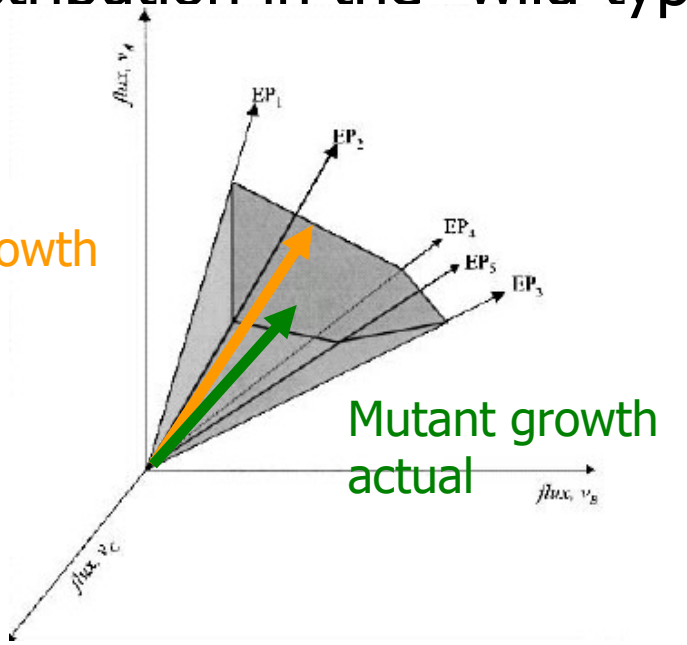
How does **MOMA** work**?**

- **Assumption – "Initially , the mutant remains as close as possible to the wild-type optimum in terms of flux values."**

How does **MOMA** work?

- In other words:

  MOMA searches for the flux distribution in the "mutant flux space" which is closest to the optimal flux distribution in the "wild-type flux space".



How does **MOMA** work?

Formally:

V^{wt} – the wild-type optimal growth vector.

V^{m} – a vector in mutant flux space.

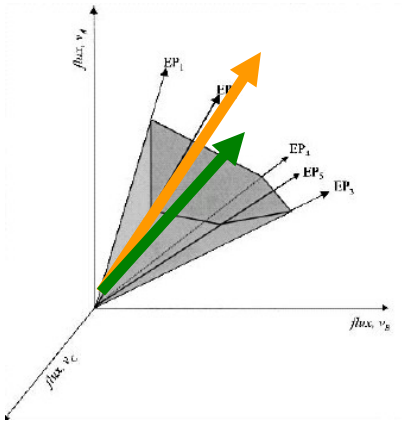
Find V^m which minimizes the Euclidian distance to V^{wt} :

$$D(V^m, V^{wt}) = \sqrt{\sum_{i=1}^{n} (V_i^m - V_i^{wt})^2}$$

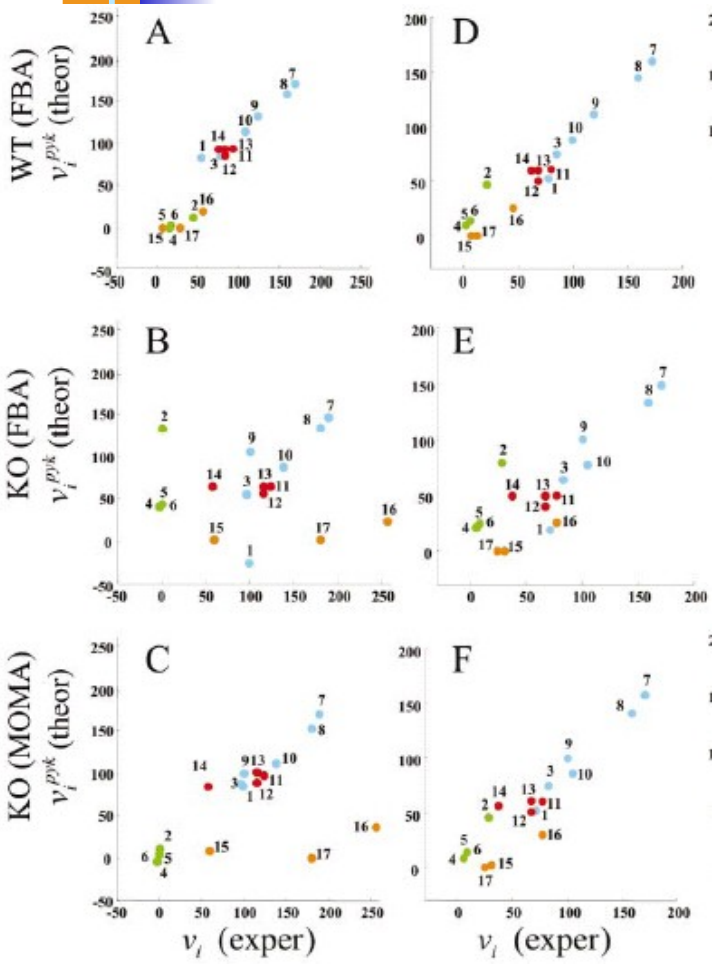
This can be stated as a QP problem. That is, minimize

$$f(x) = L \cdot x + \frac{1}{2} x^T Q x$$

Under a set of linear constraints.

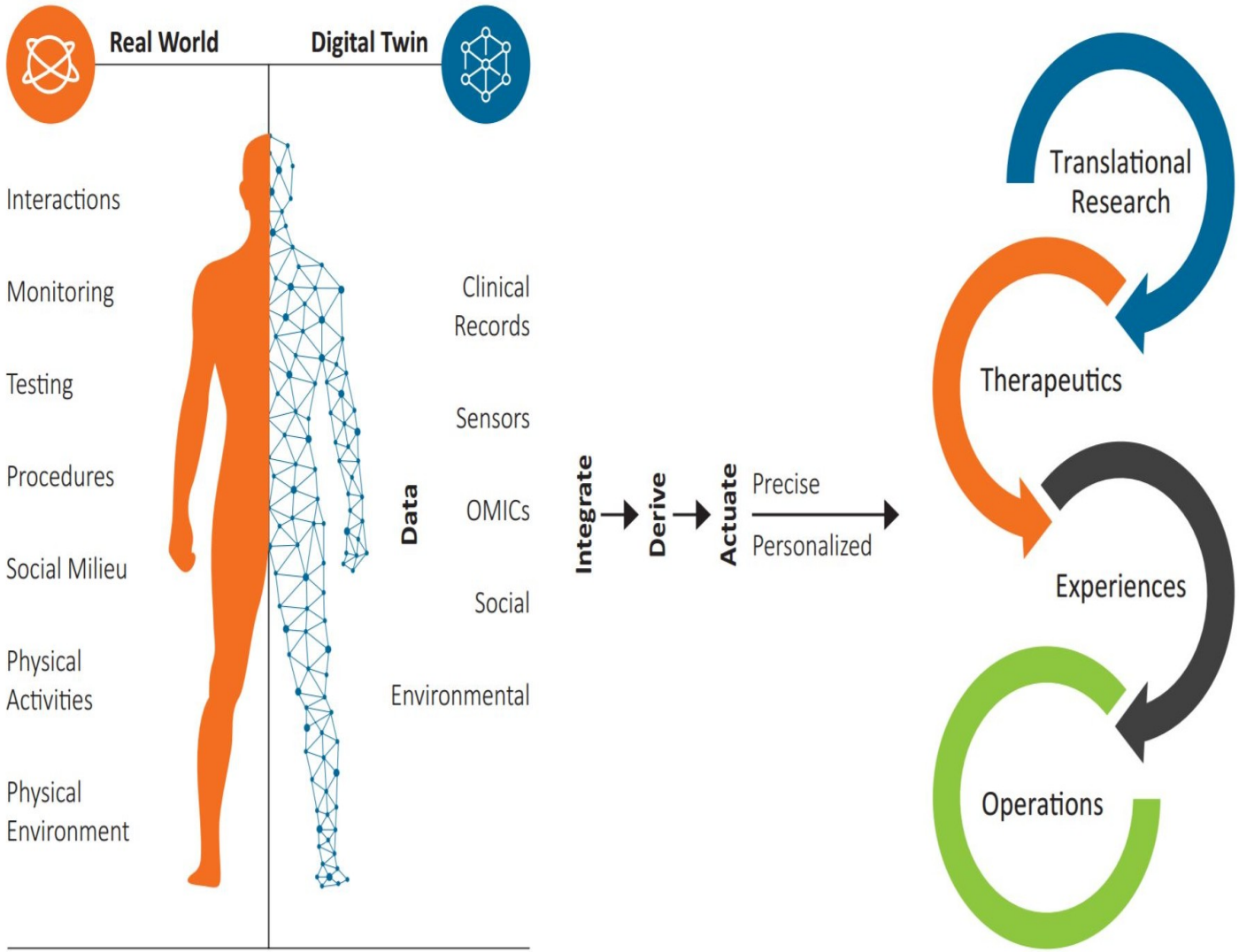


Comparing MOMA and FBA on mutant strains



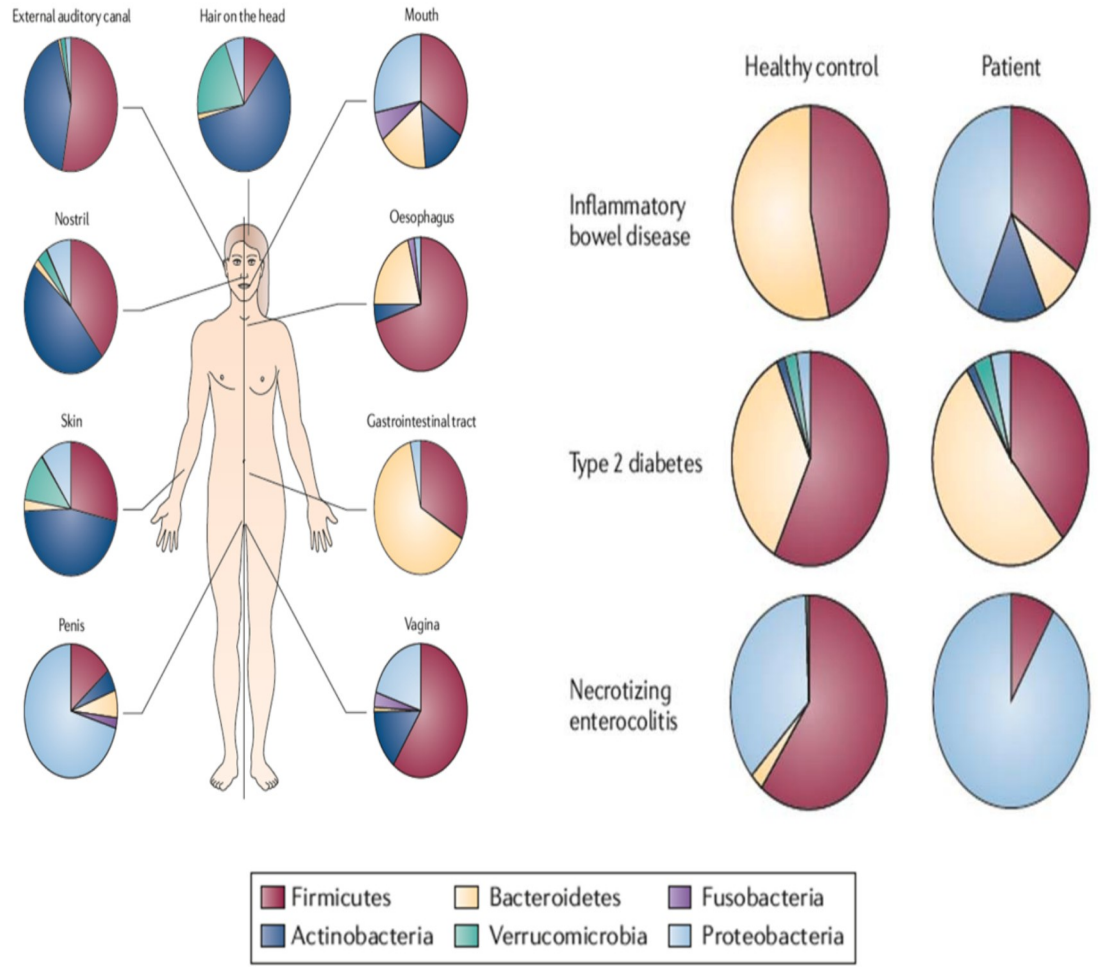
Condition	Method	Absolute ρ_1	Absolute *P* value (a)	Absolute *P* value (b)
C-0.08	WT	0.91	8.2×10^{-8}	
	KO (FBA)	−0.064	6.0×10^{-1}	3.3×10^{-3}
	KO (MOMA)	0.56	7.4×10^{-3}	
C-0.4	WT	0.97	8.1×10^{-12}	
	KO (FBA)	0.77	8.1×10^{-5}	2.5×10^{-3}
	KO (MOMA)	0.94	2.6×10^{-9}	

Digital Twin Powering The New Healthcare



https://www.persistent.com/wp-content/uploads/2019/02/digital-twins-whitepaper.pdf

Microbiome and diseases



http://www.fortunejournals.com/articles/the-human-microbiome-an-emerging-key-player-in-health-and-disease.pdf

Microbiome and diseases

- intestinal bowel disease
- diabetes
- obesity
- depression
- cardiovascular disease
- colorectal cancer
- ...

dynamic Flux balance analysis (dFBA):

$$\text{maximize} \quad \mu \tag{1}$$

$$\text{subject to} \quad Sv = 0 \tag{2}$$

$$v_{\min} \leq v \leq v_{\max} \tag{3}$$

$$b[t] = b[t - \delta t]e^{\mu \cdot \delta t} \tag{4}$$

$$exC[t + \delta t] = exC[t] - v_{ex}\frac{b[t]}{\mu}(1 - e^{\mu \cdot \delta t}) \tag{5}$$

$$v_{\min}^{ex} = -\frac{exC}{b \cdot \delta t} \tag{6}$$

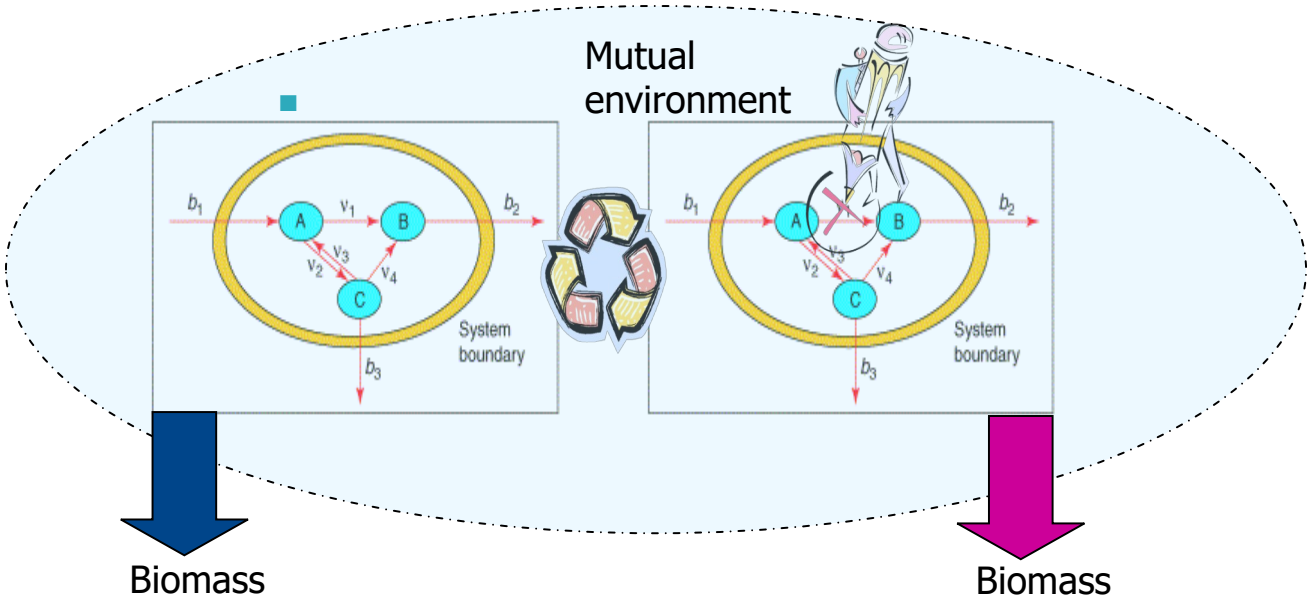
Our multispecies dFBA algorithm:

Tzamali *et al. BMC Systems Biology* 2011, **5**:167
http://www.biomedcentral.com/1752-0509/5/167

RESEARCH ARTICLE Open Access

A computational exploration of bacterial metabolic diversity identifying metabolic interactions and growth-efficient strain communities

Eleftheria Tzamali[1,2*], Panayiota Poirazi[3], Ioannis G Tollis [1,2] and Martin Reczko[4,5*]



$$B(t) = B_1(t) + B_2(t) = \frac{1}{2} B_0 e^{\mu_1(t)t} + \frac{1}{2} B_0 e^{\mu_2(t)t}$$

Our msdFBA algorithm:

$$b_i[t] = b_i[t - \delta t]e^{\mu_i \cdot \delta t} \quad (7)$$

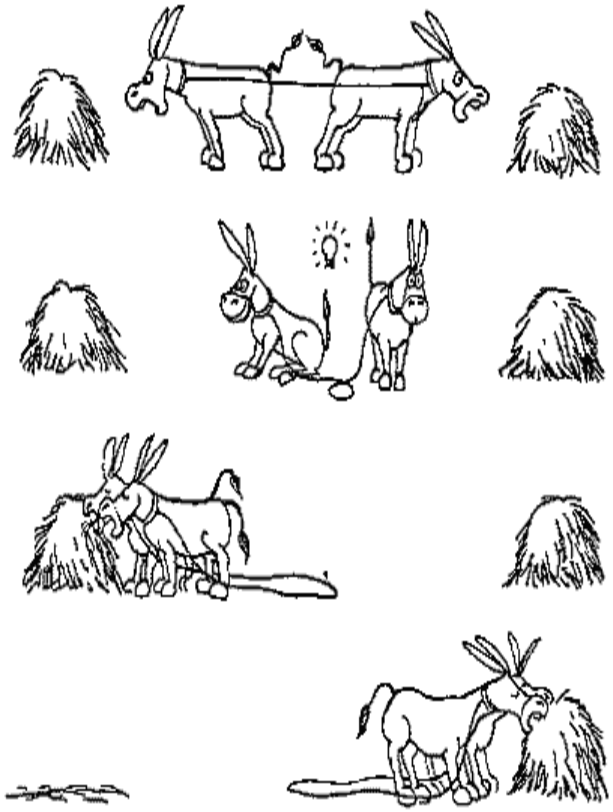
Biomass vector over time

$$exC[t + \delta t] = exC[t] - \sum_i v_{ex}^i \frac{b_i[t]}{\mu_i}(1 - e^{\mu_i \cdot \delta t}) \quad (8)$$

substrate vector over time

$$v_{\min}^{ex} = -\frac{exC}{\delta t \cdot \sum_i b_i}, \ \forall i \quad (9)$$

Our initial study: The benefit of cooperation



The AGORA collection

nature biotechnology

Generation of genome-scale metabolic reconstructions for 773 members of the human gut microbiota

Stefanía Magnúsdóttir[1,2], Almut Heinken[1,2], Laura Kutt[1], Dmitry A Ravcheev[1], Eugen Bauer[1], Alberto Noronha[1], Kacy Greenhalgh[1], Christian Jäger[1], Joanna Baginska[1], Paul Wilmes[1], Ronan M T Fleming[1] & Ines Thiele[1]

[1]Luxembourg Centre for Systems Biomedicine, University of Luxembourg, Esch-sur-Alzette, Luxembourg. [2]These authors contributed equally to this work. Correspondence should be addressed to I.T. (ines.thiele@uni.lu).

Received 15 April; accepted 20 September; published online 28 November 2016; doi:10.1038/nbt.3703

AGORA, diets, human models from:
https://www.vmh.life

Nutrition

Search diets | enter search term | Search

Diet name	Total Energy (kcal)	Lipids % Energy	Carbohydrates %..	Proteins % Ener...	Alcohol % Energy
Vegetarian	1888.29	33.9803	51.3355	10.7402	0.464363
Vegan	1991.21	35.1134	47.6352	11.1888	0
EU average	2371.9	40.4589	41.1814	11.9428	3.23286
Mediterranean	2647.45	33.708	37.2041	17.4473	4.39819
DACH	2148.36	32.7024	41.3385	18.5617	0
High protein	1791.84	30.1023	37.3962	28.7401	0
Gluten free	2983.83	37.0946	42.1941	15.6876	0
High fiber	2435.67	29.6452	45.544	16.4689	0.306077
Type 2 diabetes	2084.57	34.4098	41.4647	17.3514	0
High fat, low carb	2215.48	69.8147	1.67909	24.7916	0
Unhealthy	3847.05	48.7122	30.1623	16.2132	2.50259

Page 1 of 1 | Displaying 1 - 11 of 11 | Change page size | OK | Download

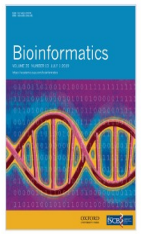
Feedback

Patient gut microbiomes from NIH Human Microbiome Project

The Human Microbiome Project expands the toolbox for studying host and microbiome interactions

https://commonfund.nih.gov/hmp

via:



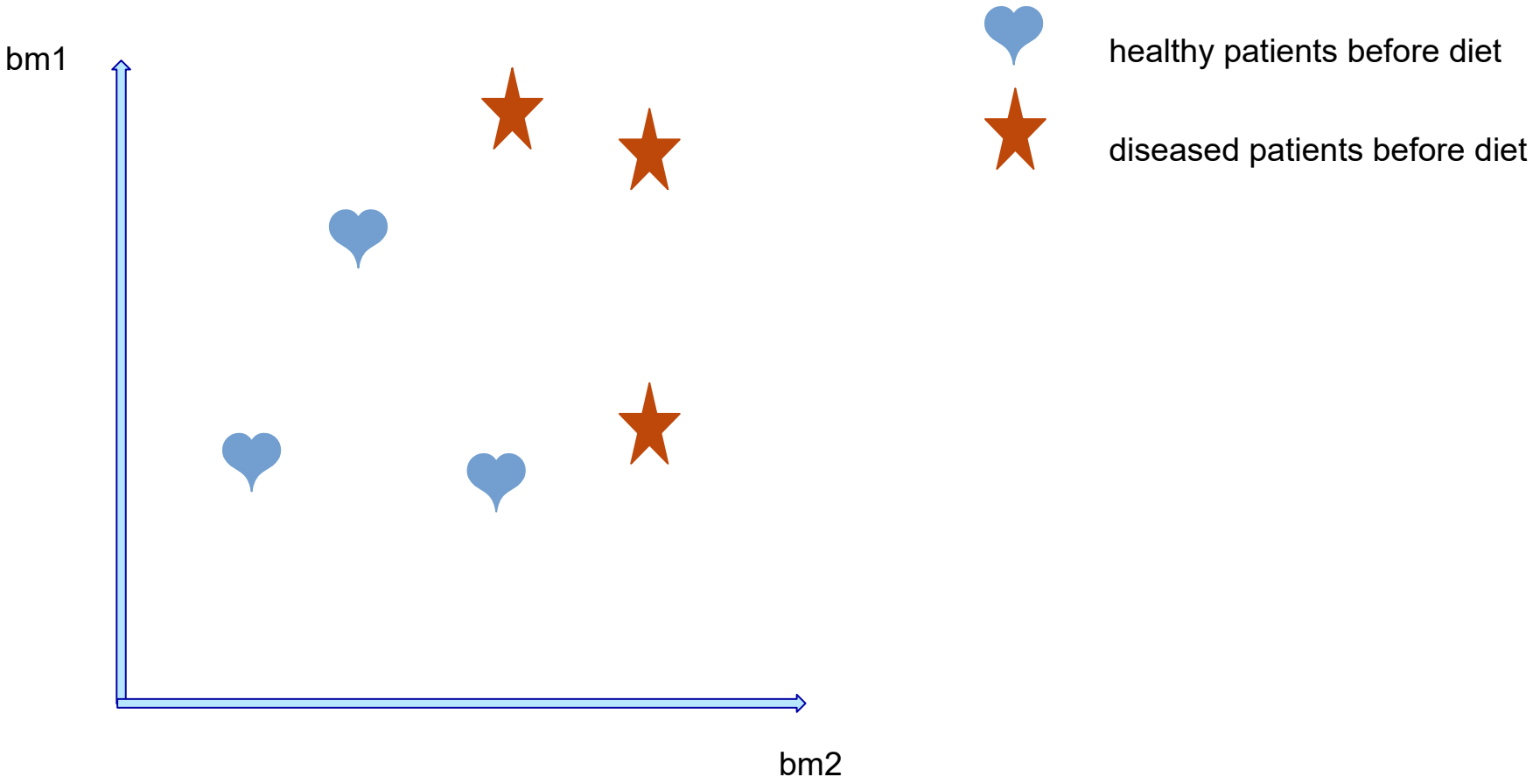
Volume 35, Issue 13
1 July 2019

The Microbiome Modeling Toolbox: from microbial interactions to personalized microbial communities

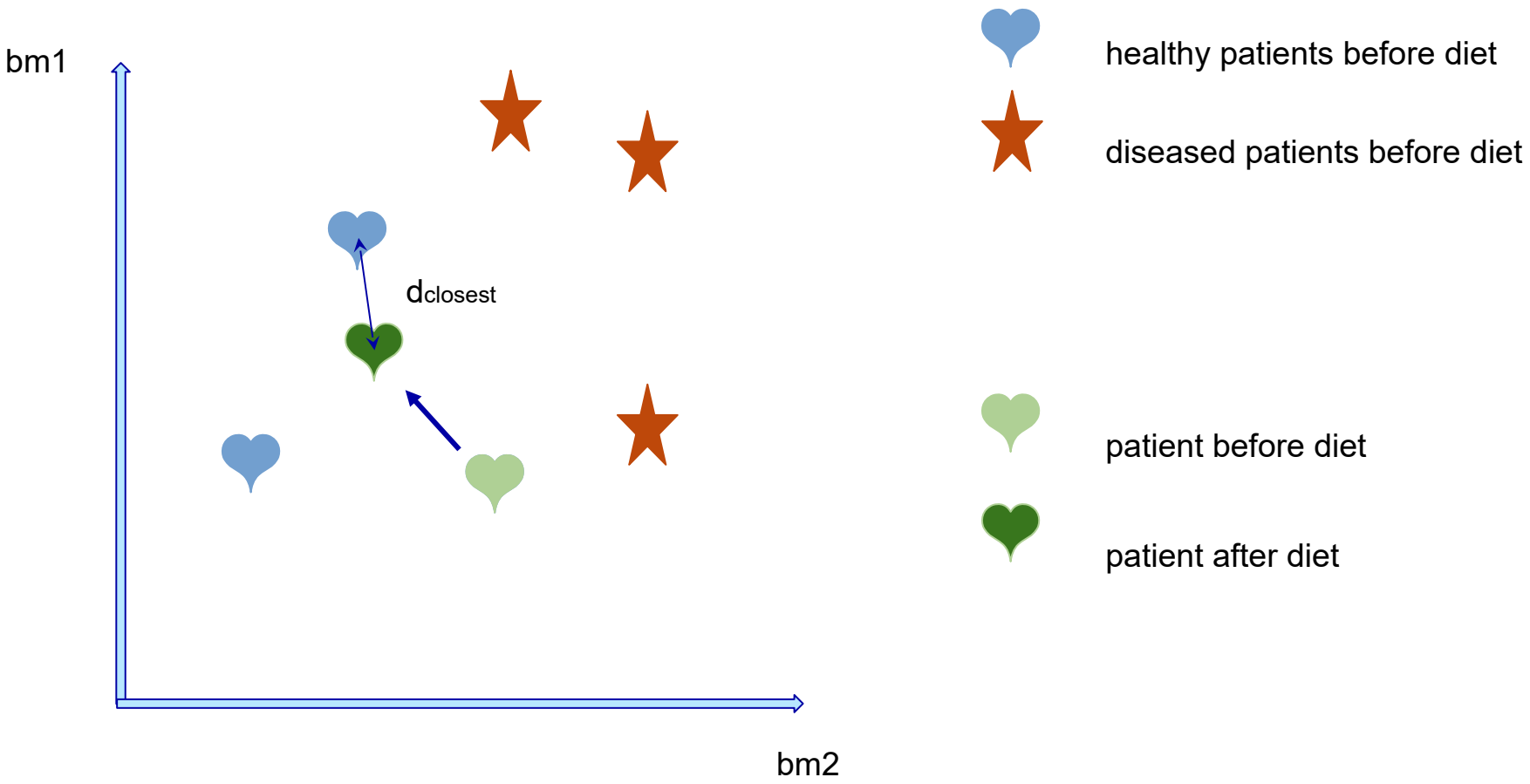
Federico Baldini, Almut Heinken, Laurent Heirendt, Stefania Magnusdottir, Ronan M T Fleming, Ines Thiele ✉

Bioinformatics, Volume 35, Issue 13, 1 July 2019, Pages 2332–2334,
https://doi.org/10.1093/bioinformatics/bty941

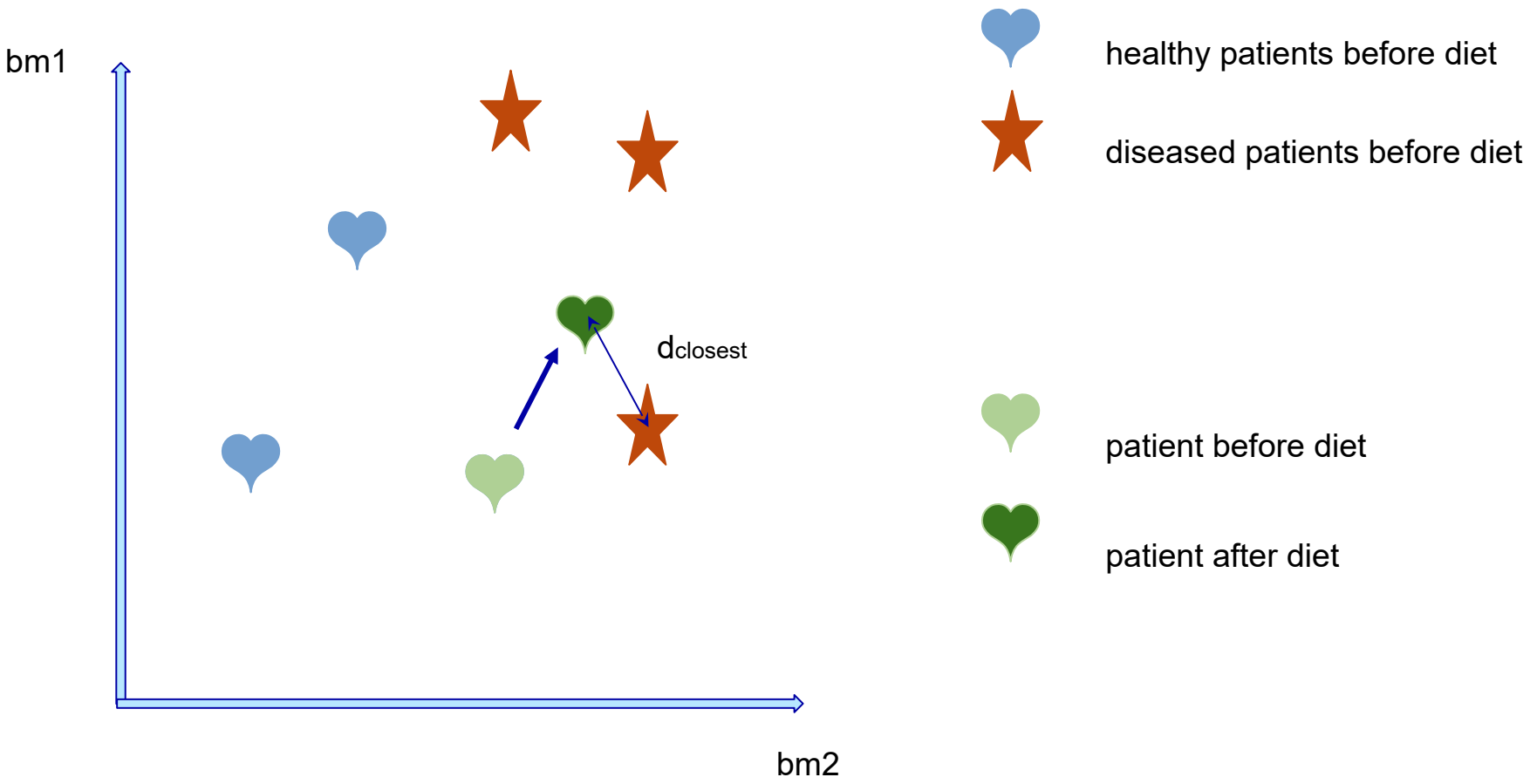
DietScore definition



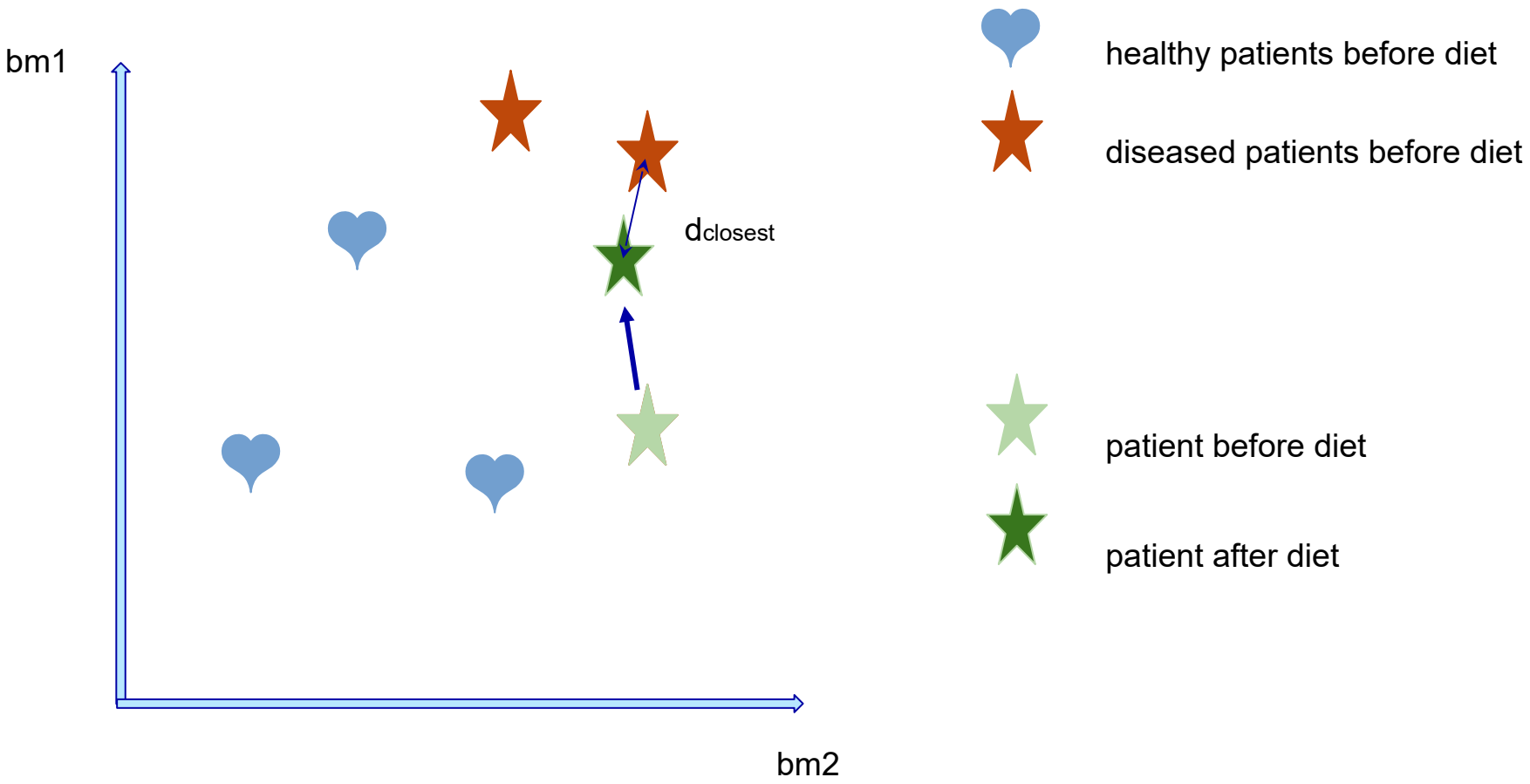
DietScore definition



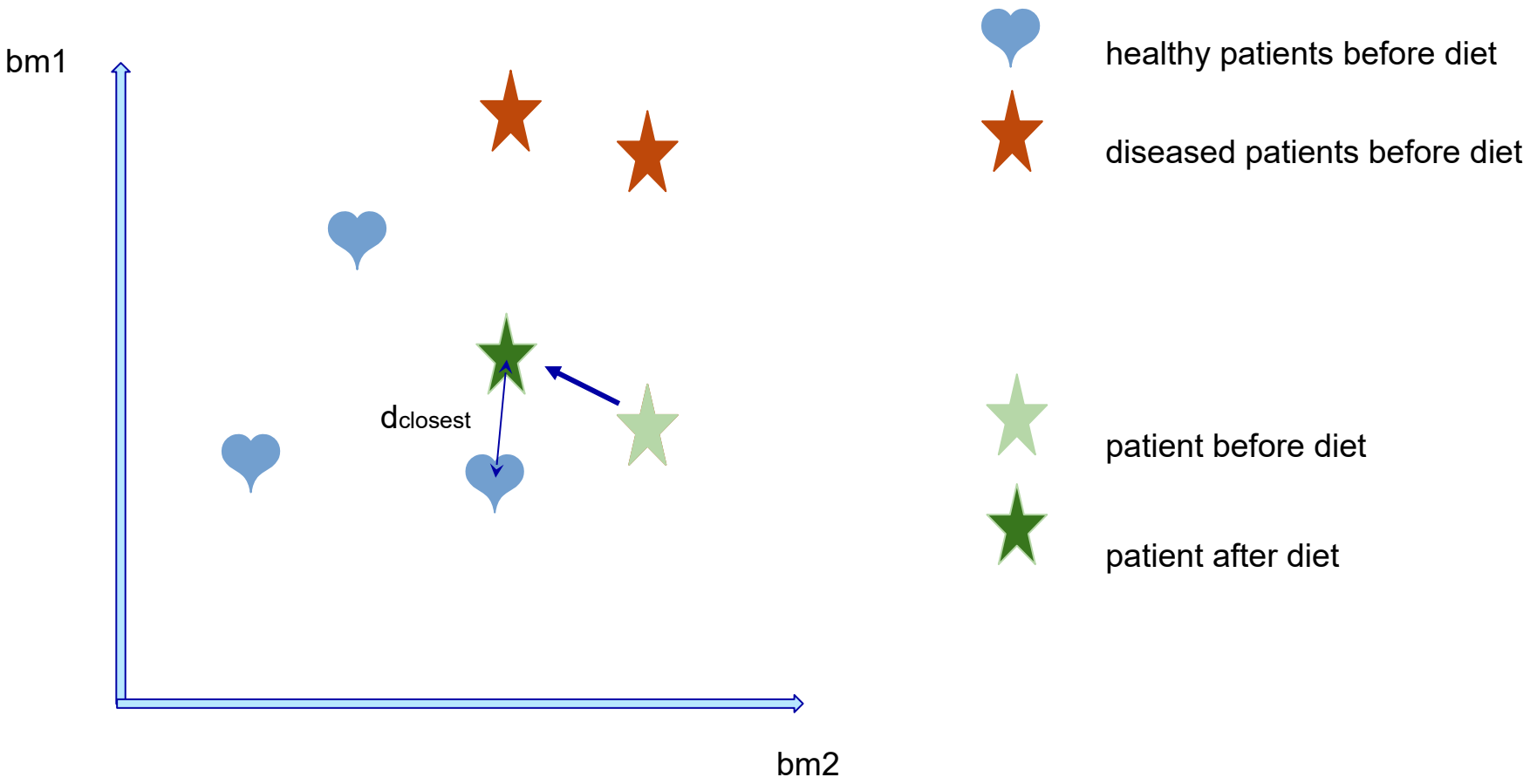
DietScore definition



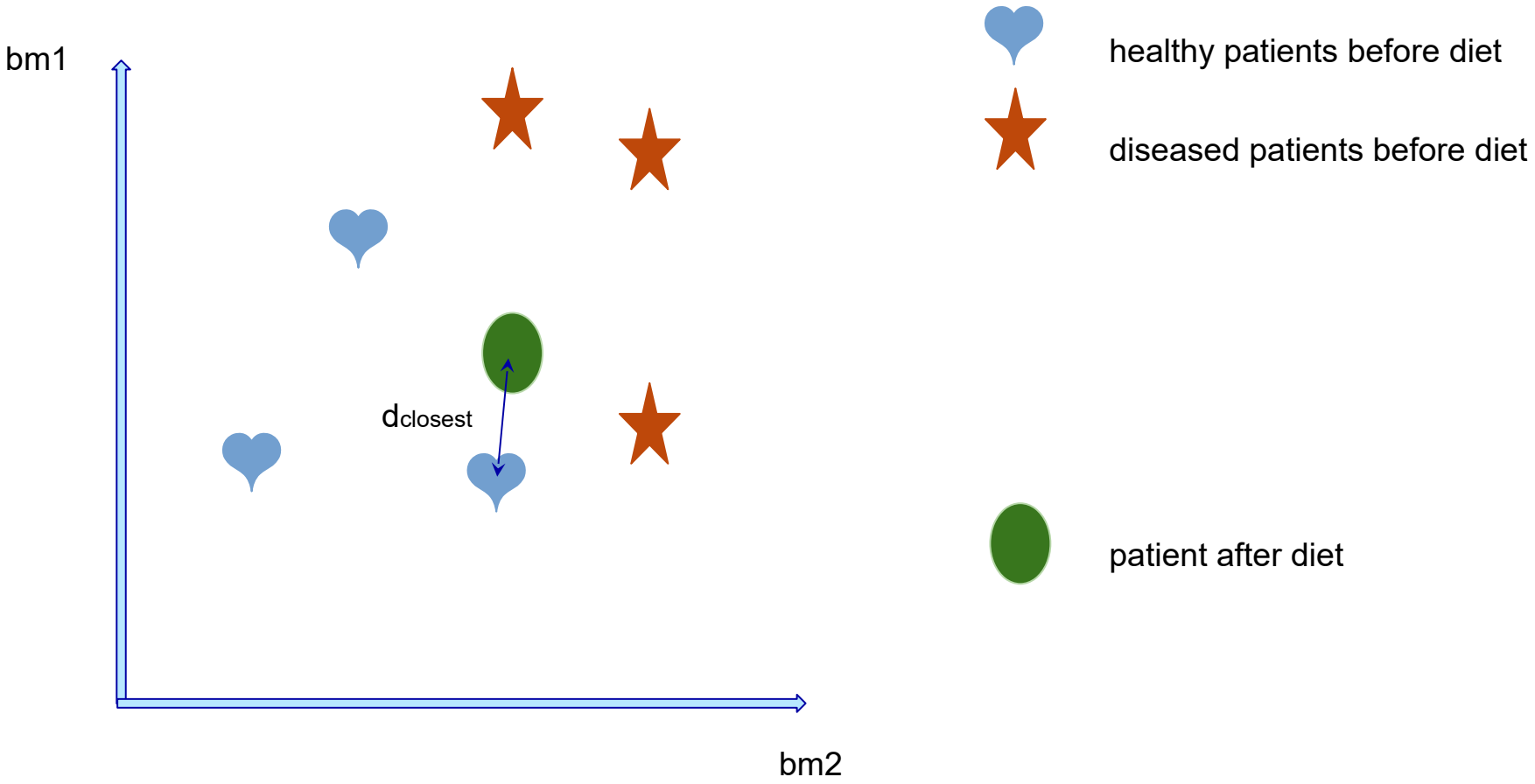
DietScore definition



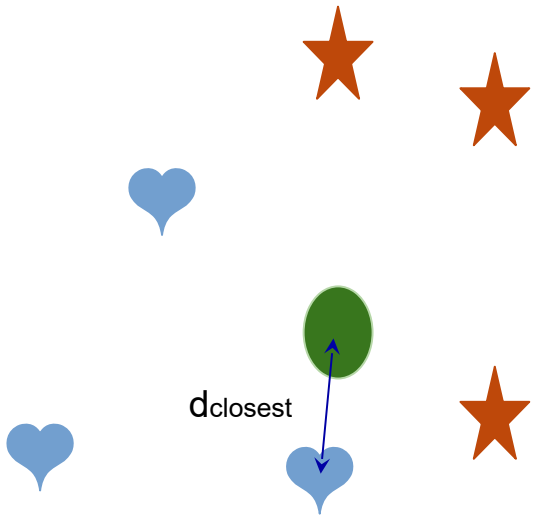
DietScore definition



DietScore definition



DietScore definition



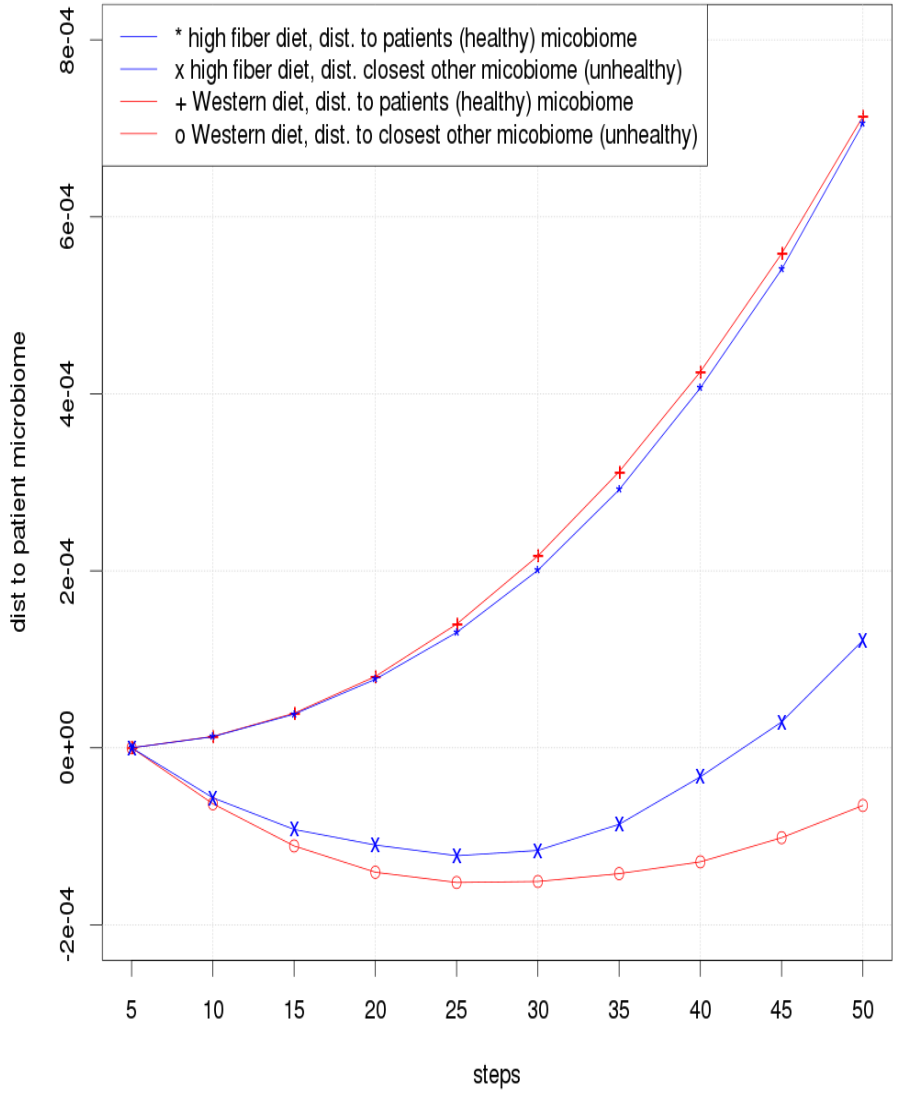
$$Diet_{score} = \frac{1}{n_{patients}} \sum_{patient} \frac{closest_{healthy}}{1 + d_{closest}} \quad (1)$$

where $d_{closest}$ is the closest biomass distance to any of the $n_{patients} = 149$ patients at the final time step of a msdFBA simulation and

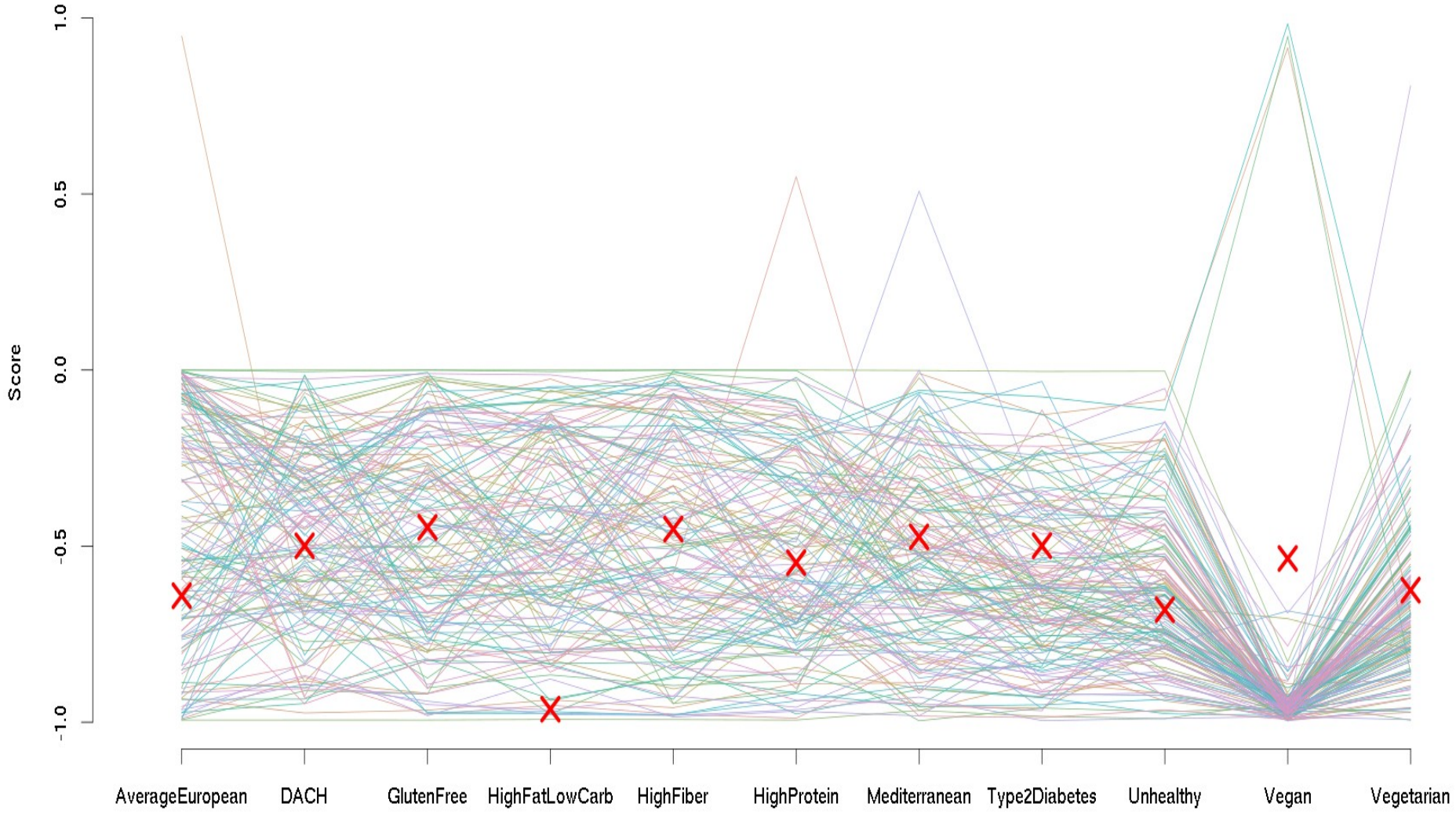
$$closest_{healthy} = \begin{cases} +1, if\ closest\ patient\ is\ healthy, \\ -1, else. \end{cases}$$

$Diet_{score}$ = +1.0 , if all patients under diet end at a healthy patient
$Diet_{score}$ = -1.0 , if all patients under diet end at a diseased patient

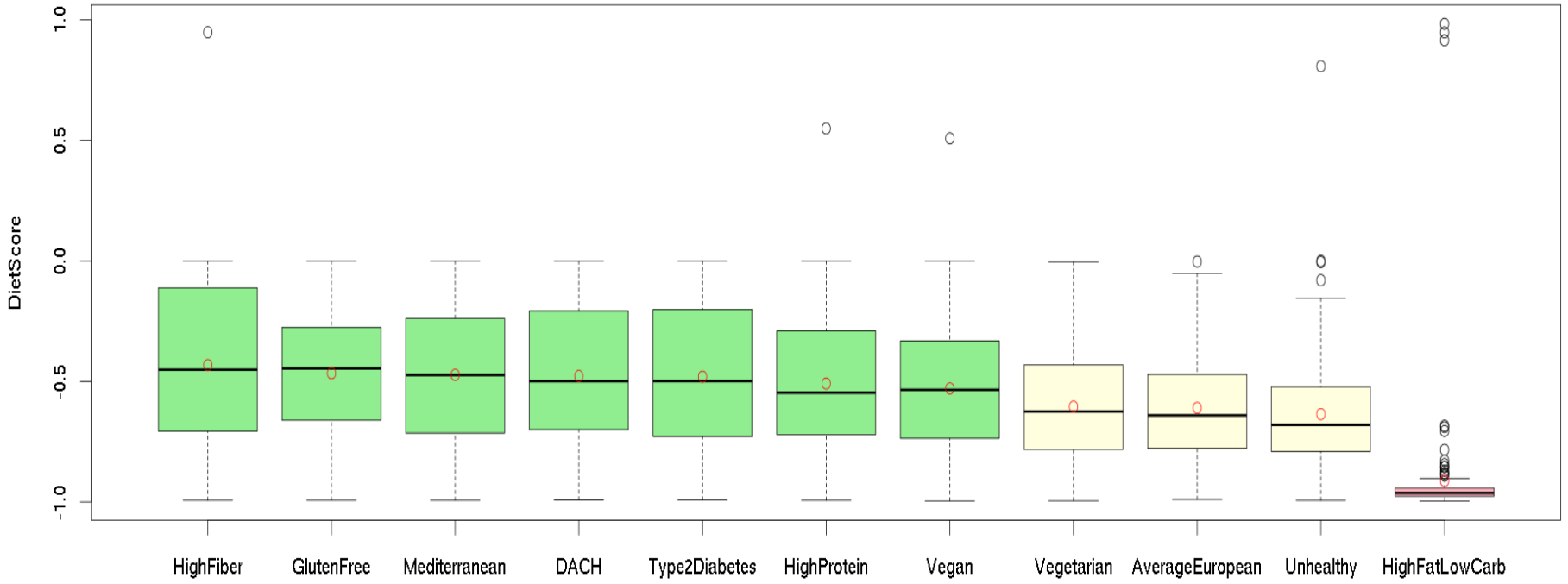
Time course of biomasses distances



DietScore profiles for each patient (IBD study)



DietScore distributions for IBD



DietScore distribution distances for IBD

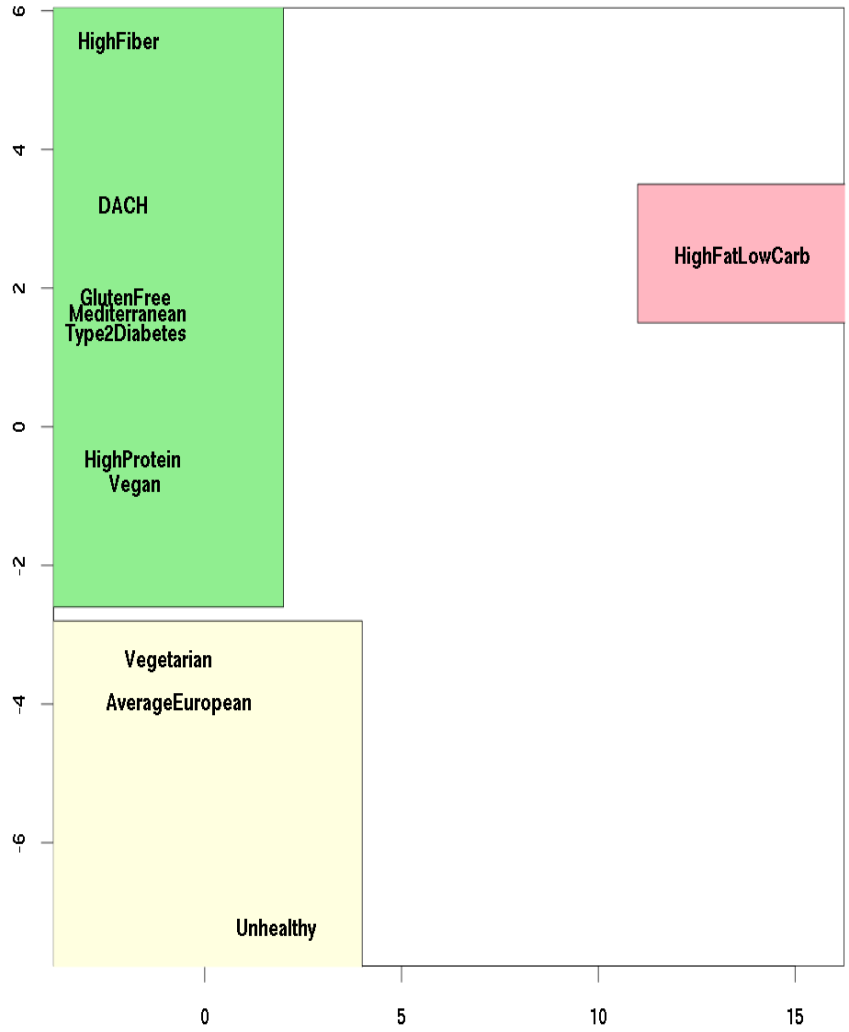
Diet1	Diet2	q-value
Mediterranean	GlutenFree	1.000000e+00
Type2Diabetes	GlutenFree	1.000000e+00
Type2Diabetes	Mediterranean	1.000000e+00
Vegetarian	AverageEuropean	9.999999e-01
Mediterranean	DACH	9.999906e-01
GlutenFree	DACH	9.999823e-01
Vegan	HighProtein	9.999762e-01
HighFiber	DACH	9.999452e-01
Type2Diabetes	DACH	9.994632e-01
Mediterranean	HighFiber	9.898656e-01
HighFiber	GlutenFree	9.871803e-01
Unhealthy	AverageEuropean	9.801268e-01
Type2Diabetes	HighProtein	9.756104e-01
Type2Diabetes	HighFiber	9.494194e-01
HighProtein	GlutenFree	9.197859e-01
Vegetarian	Unhealthy	9.075000e-01
Mediterranean	HighProtein	9.075000e-01
Vegan	Type2Diabetes	7.706096e-01
HighProtein	DACH	6.181543e-01
Vegan	GlutenFree	6.053679e-01
Vegan	Mediterranean	5.797649e-01
Vegan	DACH	2.476459e-01
HighProtein	HighFiber	2.214467e-01
Vegetarian	Vegan	8.463075e-02
Vegan	HighFiber	5.097940e-02

similar diet pairs

Diet1	Diet2	q-value
Vegan	AverageEuropean	3.311465e-02
Vegetarian	HighProtein	1.400984e-02
HighProtein	AverageEuropean	4.169973e-03
Vegan	Unhealthy	2.347788e-04
Vegetarian	Type2Diabetes	7.943436e-05
Vegetarian	GlutenFree	2.016660e-05
Unhealthy	HighProtein	1.973406e-05
Vegetarian	Mediterranean	1.973406e-05
Type2Diabetes	AverageEuropean	1.090171e-05
GlutenFree	AverageEuropean	2.461366e-06
Mediterranean	AverageEuropean	2.435656e-06
Vegetarian	DACH	6.631672e-07
DACH	AverageEuropean	6.433142e-08
Vegetarian	HighFiber	6.081344e-09
Unhealthy	Mediterranean	2.153218e-09
Unhealthy	Type2Diabetes	1.705444e-09
HighFiber	AverageEuropean	8.125022e-10
Unhealthy	GlutenFree	4.419831e-10
Unhealthy	DACH	1.242662e-11
Unhealthy	HighFiber	5.817569e-14
Unhealthy	HighFatLowCarb	1.887379e-15
HighFatLowCarb	AverageEuropean	2.220446e-16
HighFatLowCarb	DACH	2.220446e-16
HighFatLowCarb	GlutenFree	2.220446e-16
HighFiber	HighFatLowCarb	2.220446e-16
HighProtein	HighFatLowCarb	2.220446e-16
Mediterranean	HighFatLowCarb	2.220446e-16
Type2Diabetes	HighFatLowCarb	2.220446e-16
Vegan	HighFatLowCarb	2.220446e-16
Vegetarian	HighFatLowCarb	2.220446e-16

different diet pairs

Diets projected using DietScore distances for IBD



Patient microbiotas for colorectal cancer

Article | 28 November 2014 | OPEN ACCESS

SOURCE DATA

Potential of fecal microbiota for early-stage detection of colorectal cancer

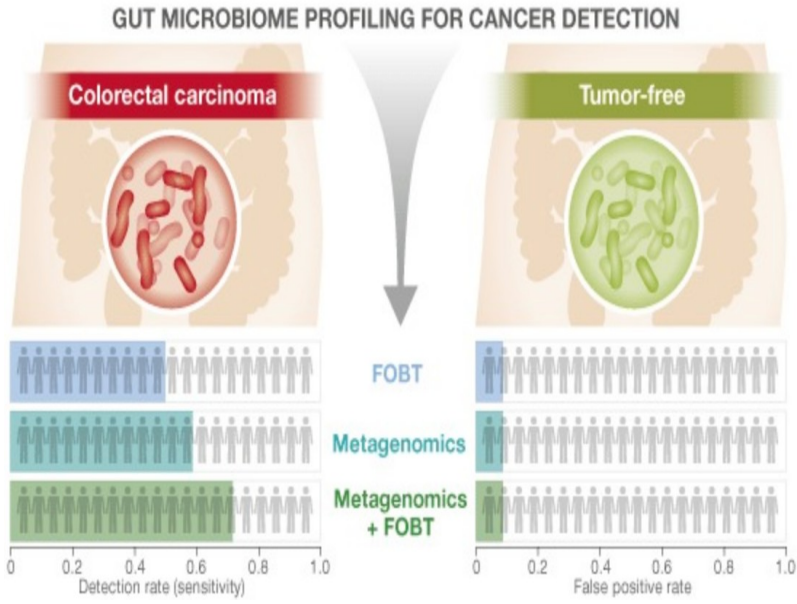
Georg Zeller, Julien Tap, Anita Y Voigt, Shinichi Sunagawa, Jens Roat Kultima, Paul I Costea, Aurélien Amiot, Jürgen Böhm, Francesco Brunetti, Nina Habermann, Rajna Hercog, Moritz Koch, Alain Luciani, Daniel R Mende, Martin A Schneider, Petra Schrotz-King, Christophe Tournigand, Jeanne Tran Van Nhieu, Takuji Yamada, Jürgen Zimmermann, Vladimir Benes, Matthias Kloor, Cornelia M Ulrich, Magnus von Knebel Doeberitz, Iradj Sobhani✉, Peer Bork✉

Author Information

Mol Syst Biol (2014) 10: 766 | https://doi.org/10.15252/msb.20145645

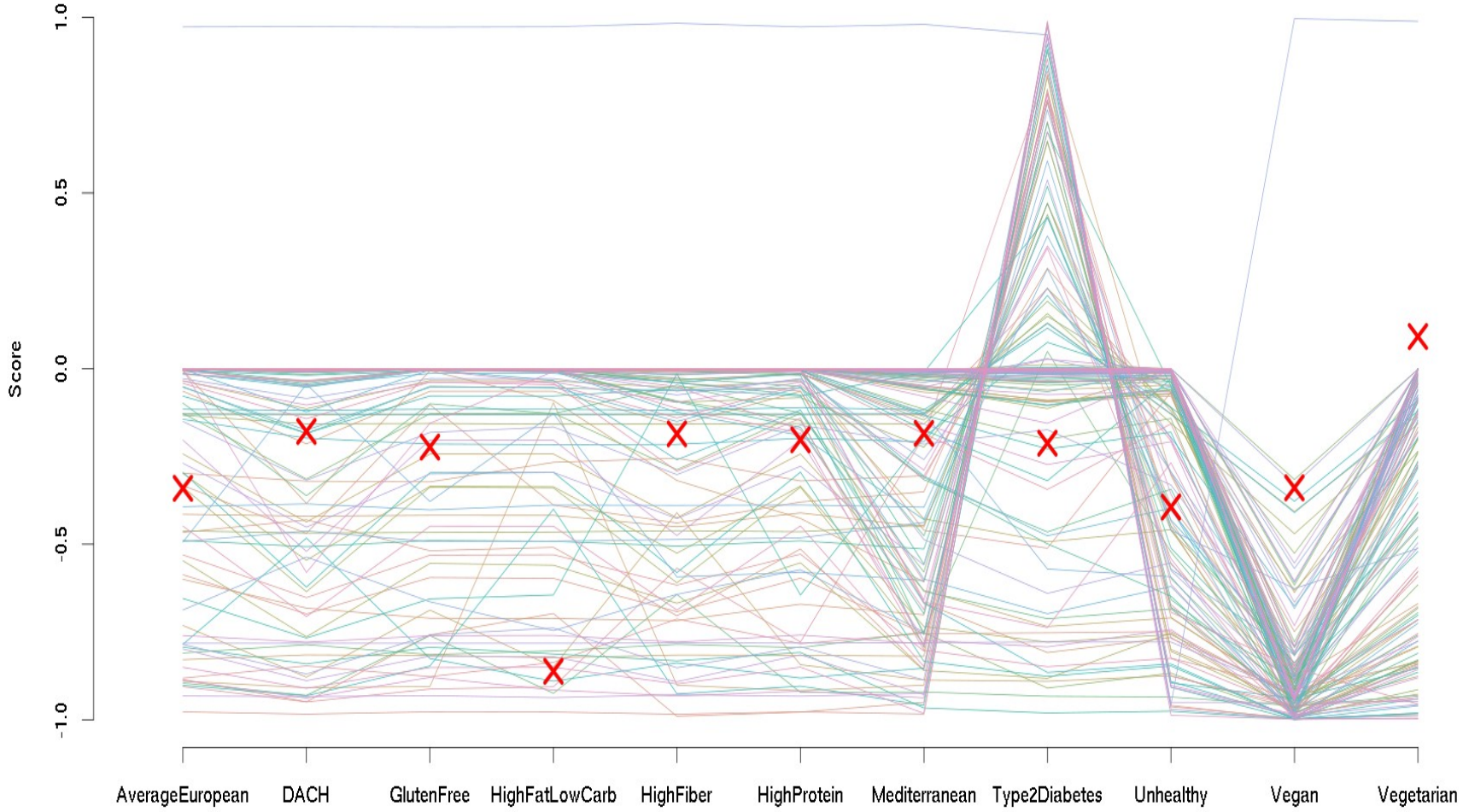
Volume 10
Issue 11
1 November 2014

IN THIS ISSUE

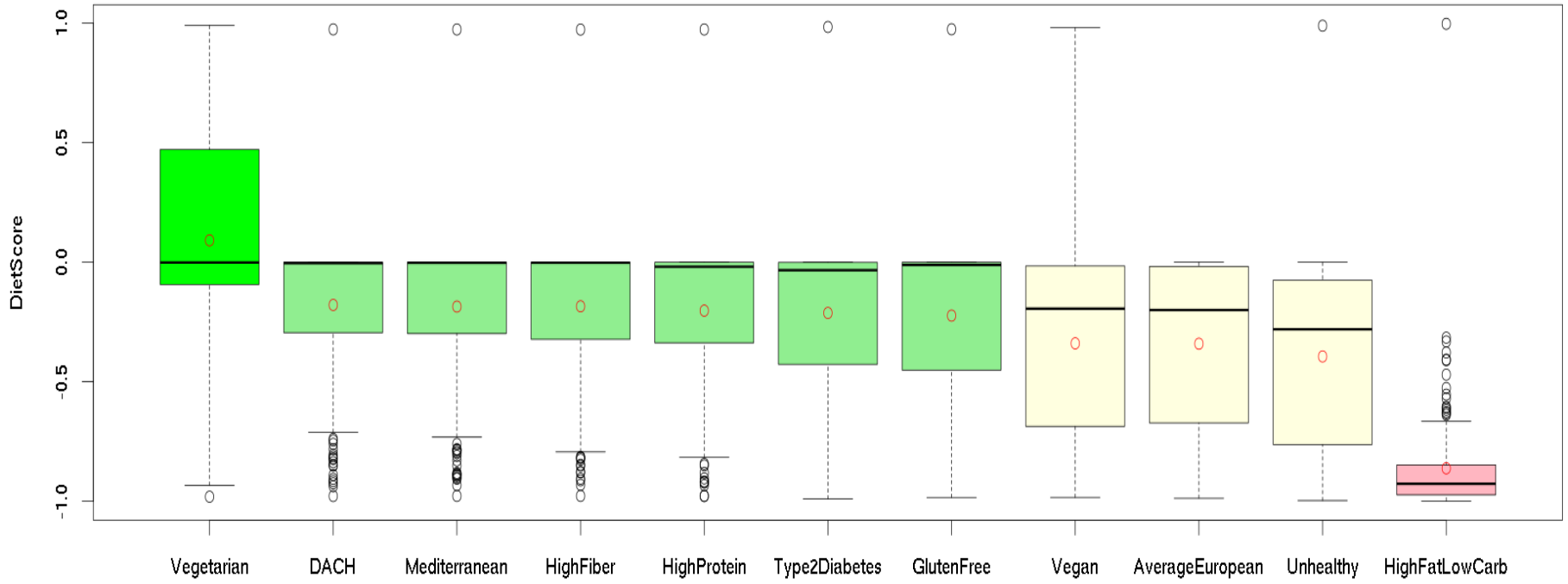


=> 141 patient samples mapped to 167 AGORA species

DietScore profiles for each patient (CRC study)



Host-microbiota metabolic interactions



DietScore distribution distances for CRC

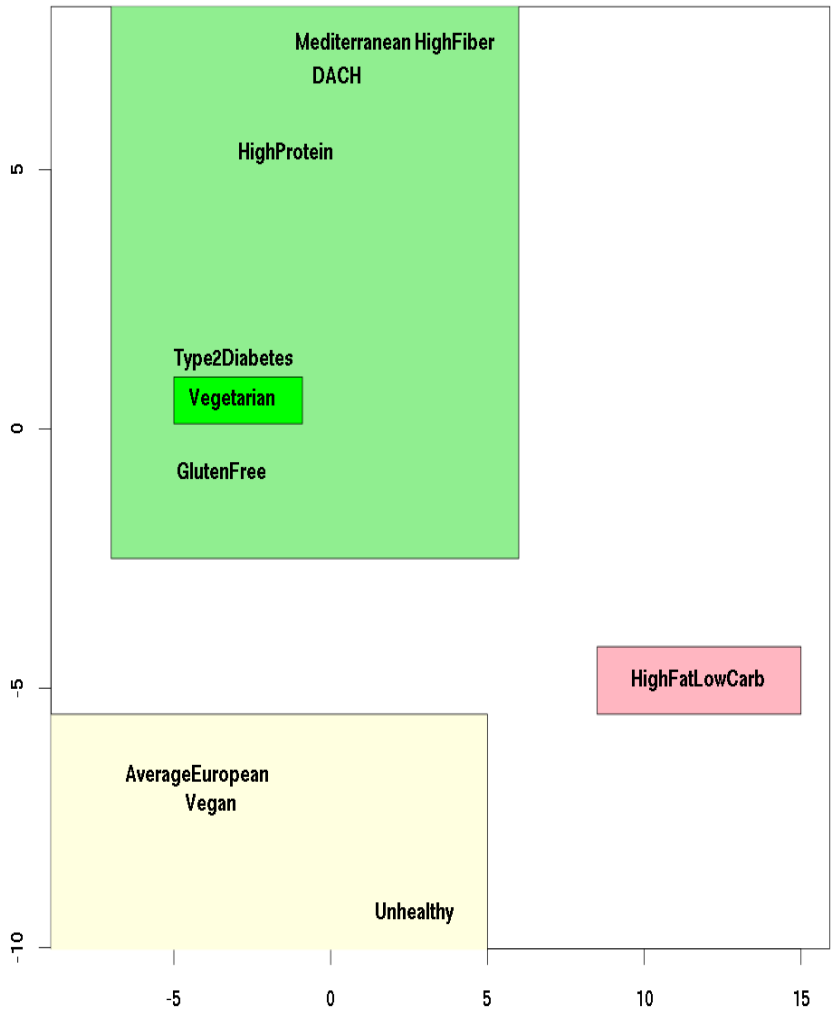
Diet1	Diet2	q-value
Vegan	AverageEuropean	1.000000e+00
Vegetarian	Type2Diabetes	9.999999e-01
Mediterranean	HighFiber	9.999970e-01
Vegetarian	GlutenFree	9.999692e-01
Type2Diabetes	GlutenFree	9.985670e-01
HighProtein	DACH	9.985670e-01
Vegan	Unhealthy	5.377574e-01
Type2Diabetes	HighProtein	4.110011e-01
Unhealthy	AverageEuropean	3.391787e-01
Mediterranean	DACH	2.865635e-01
Vegetarian	HighProtein	2.285077e-01
HighFiber	DACH	9.856680e-02
Type2Diabetes	DACH	5.104035e-02
HighProtein	GlutenFree	5.104035e-02

similar diet pairs

Diet1	Diet2	q-value
Mediterranean	HighProtein	2.683671e-02
Vegetarian	DACH	1.901759e-02
HighProtein	HighFiber	5.319388e-03
GlutenFree	AverageEuropean	4.302050e-03
GlutenFree	DACH	2.111175e-03
Vegan	GlutenFree	1.411617e-03
Vegetarian	AverageEuropean	4.003242e-04
Vegetarian	Vegan	1.155049e-04
Type2Diabetes	AverageEuropean	7.296435e-05
Vegan	Type2Diabetes	2.224698e-05
Type2Diabetes	Mediterranean	1.652751e-06
Type2Diabetes	HighFiber	3.429058e-08
Unhealthy	GlutenFree	2.055316e-08
Vegetarian	Mediterranean	4.437701e-08
Vegetarian	HighFiber	9.484192e-09
Unhealthy	HighFatLowCarb	5.485842e-09
Mediterranean	GlutenFree	5.207108e-10
Vegetarian	Unhealthy	1.150431e-10
HighProtein	AverageEuropean	4.622391e-11
HighFiber	GlutenFree	2.349831e-11
Unhealthy	Type2Diabetes	3.979927e-12
Vegan	HighProtein	1.246003e-12
DACH	AverageEuropean	1.276756e-14
Vegan	DACH	1.332268e-15
HighFatLowCarb	AverageEuropean	2.220446e-16
HighFiber	AverageEuropean	2.220446e-16
Mediterranean	AverageEuropean	2.220446e-16
HighFatLowCarb	DACH	2.220446e-16
Unhealthy	DACH	2.220446e-16
HighFatLowCarb	GlutenFree	2.220446e-16
HighFiber	HighFatLowCarb	2.220446e-16
HighProtein	HighFatLowCarb	2.220446e-16
Mediterranean	HighFatLowCarb	2.220446e-16
Type2Diabetes	HighFatLowCarb	2.220446e-16
Vegan	HighFatLowCarb	4.440892e-16
Vegetarian	HighFatLowCarb	2.220446e-16
Unhealthy	HighFiber	2.220446e-16
Vegan	HighFiber	2.220446e-16
Unhealthy	HighProtein	2.220446e-16
Unhealthy	Mediterranean	2.220446e-16
Vegan	Mediterranean	2.220446e-16

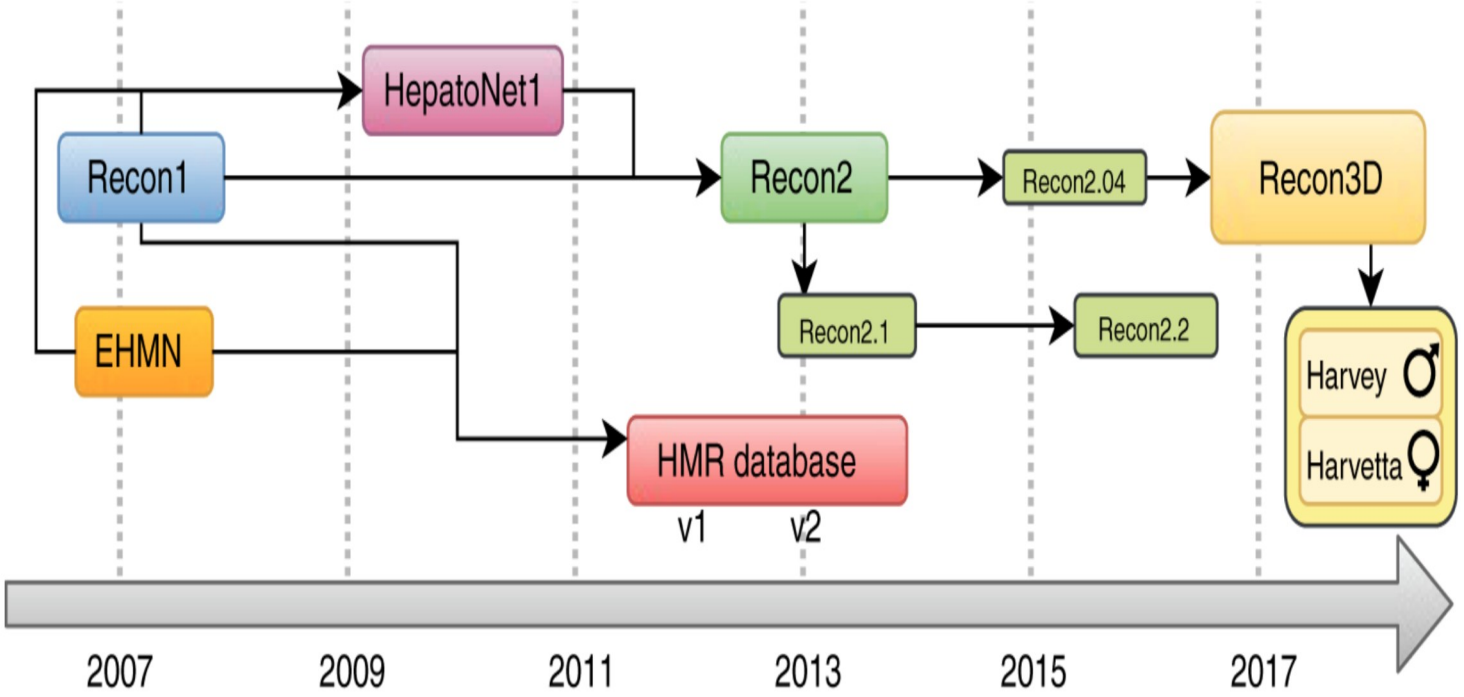
different diet pairs

Diets projected using DietScore distances for CRC



Microbiome simulation extensions

- add other microbiome related diseases (obesity, depression, liver cancer,…)
- online tool to generate patient msDFBA profiles
- extend to whole body metabolic models: Harvey & Harvetta

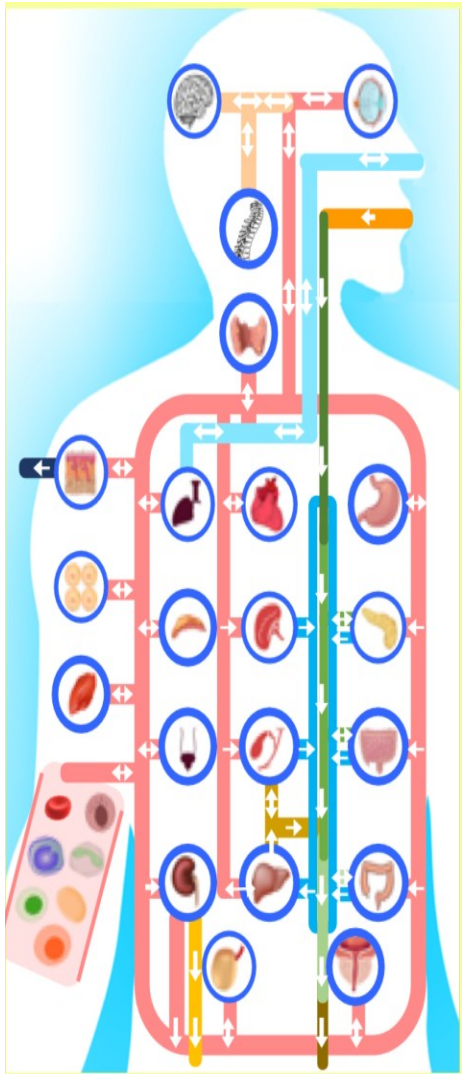


Recon4imd

“infant” Harvey& Harvetta

Figure from: https://www.degruyter.com/view/j/jib.2019.16.issue-1/jib-2018-0068/jib-2018-006

When metabolism meets physiology: Harvey and Harvetta

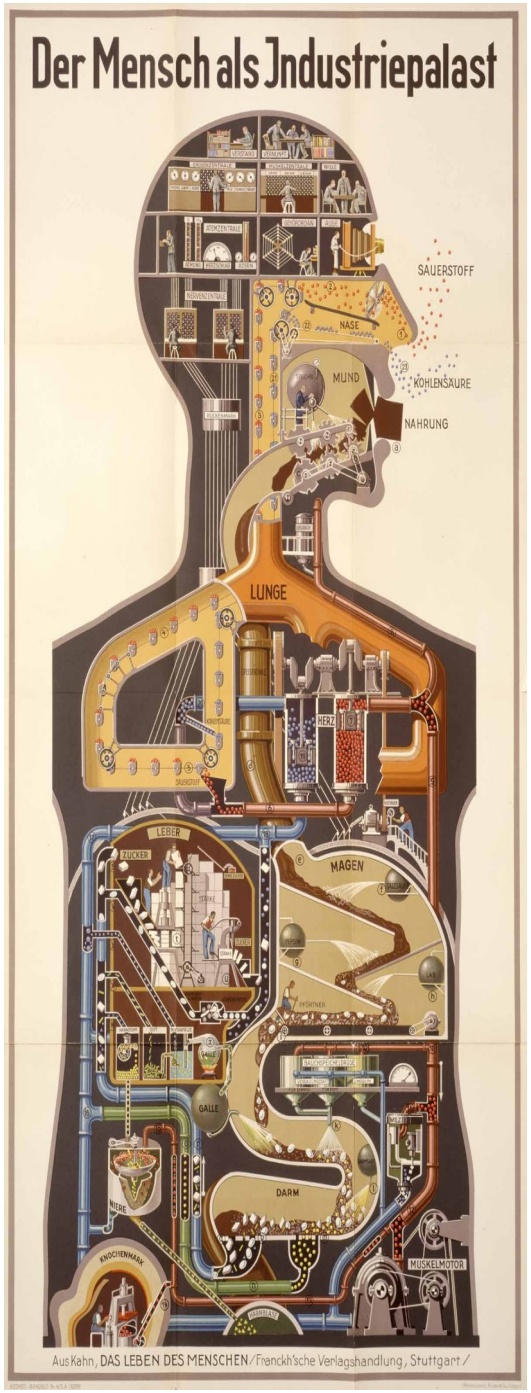


ORGANS	
Adipose tissue	Pancreas
Adrenal gland	Parathyroid gland
Brain	Retina
Colon	Skin
Gallbladder	Small intestine
Heart	Spinal cord
Kidney	Spleen
Liver	Stomach
Lung	Thyroid gland
Muscle	Urinary bladder

SEX SPECIFIC	
♀	♂
Breast	Testis
Cervix	Prostate
Ovary	
Uterus	

BLOOD CELLS
B-cells
CD4+ T-cells
Monocytes
Natural killer cells
Platelet
Red blood cell

,
Ines Thiele, Swagatika Sahoo, Almut Heinken, Laurent Heirendt, Maike K. Aurich, Alberto Noronha, Ronan M. T. Fleming, doi: https://doi.org/10.1101/255885



Visualizations by Fritz Kahn from 1922 coming to life

https://www.fritz-kahn.com/about/

elixir GREECE

Conclusions

selection pressure results in optimal performance through evolutionary process.

This optimal performance can be predicted using *in-silico* modeling.

Unicellular evolution can be thought in terms of an iterative optimization procedure whose objective function maximizes the organisms ability to survive and proliferate. If given enough time (iterations) a local maxima is struck….

Bibliography

[1] Daniel Segre` , Dennis Vitkup, and George M. Church. Analysis of optimality in natural and perturbed metabolic networks. PNAS, vol. 99, 2002.

[2] C. H. Schilling, D. Letscher and Bernhard Palsson. Theory for the Systemic Definition of Metabolic Pathways and their use in Interpreting Metabolic Function from a Pathway-Oriented Perspective. *J. theor. Biol.* (2000)

[3] Schillling et. Al Combining pathway analysis with flux balance analysis for the comprehensive study of metabolic systems. Biotechnology and bioengineering, 2001.

[4] Edwards et al. 2002. Characterizing the metabolic phenotype" A phenotype phase plan. Biotechnology and bioengineering

[5] Kenethh et al. Advances in flux balance analysis. Current Opinion in Biotechnology.

[6] Ibarra et al. Escherichia coli k-12 undergoes adaptive evolution to achiev in silico predicted optimal growth. Nature 2002.

[7] W. Wiechert . Modeling and simulation: tools for metabolic engineering. Journal of biotechnology(2002)

[8] Cornish-Bowden. From genome to cellular phenotype- a role for meatbolic flux analysis? Nature biotechnology, vol 18, 2000.

[9] Schuster et al. Detection of elelmtary flux modes in biochemical networks: a promising tool for pathway analysis and metabolic engineering. TIBTECH 1999

[10] J. Papin, Nathan D Price, B. Palsson. Extreme pathway lengths and reaction participation in genome scale metabolic networks. Genome research, 2002.

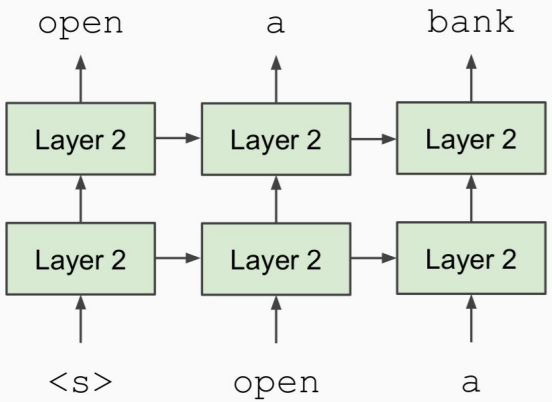
[11] Stelling eta l. Metabolic netwrok structure determines key aspects of functionality and regulation. Nature 2002.

[12] A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks.

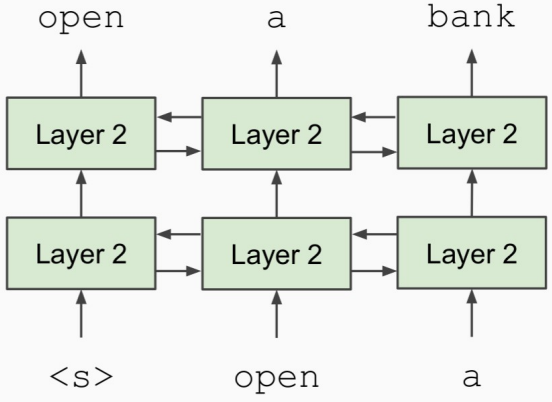

BERT
李宏毅
Hung-yi Lee

Unidirectional vs. Bidirectional Models

Unidirectional context
Build representation incrementally



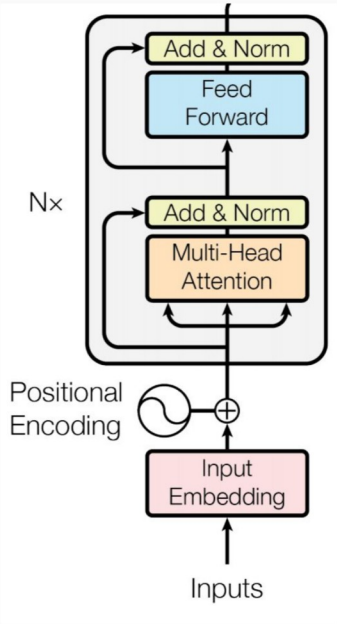
Bidirectional context
Words can “see themselves”

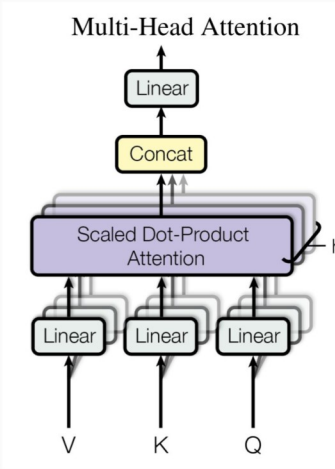


Model Architecture

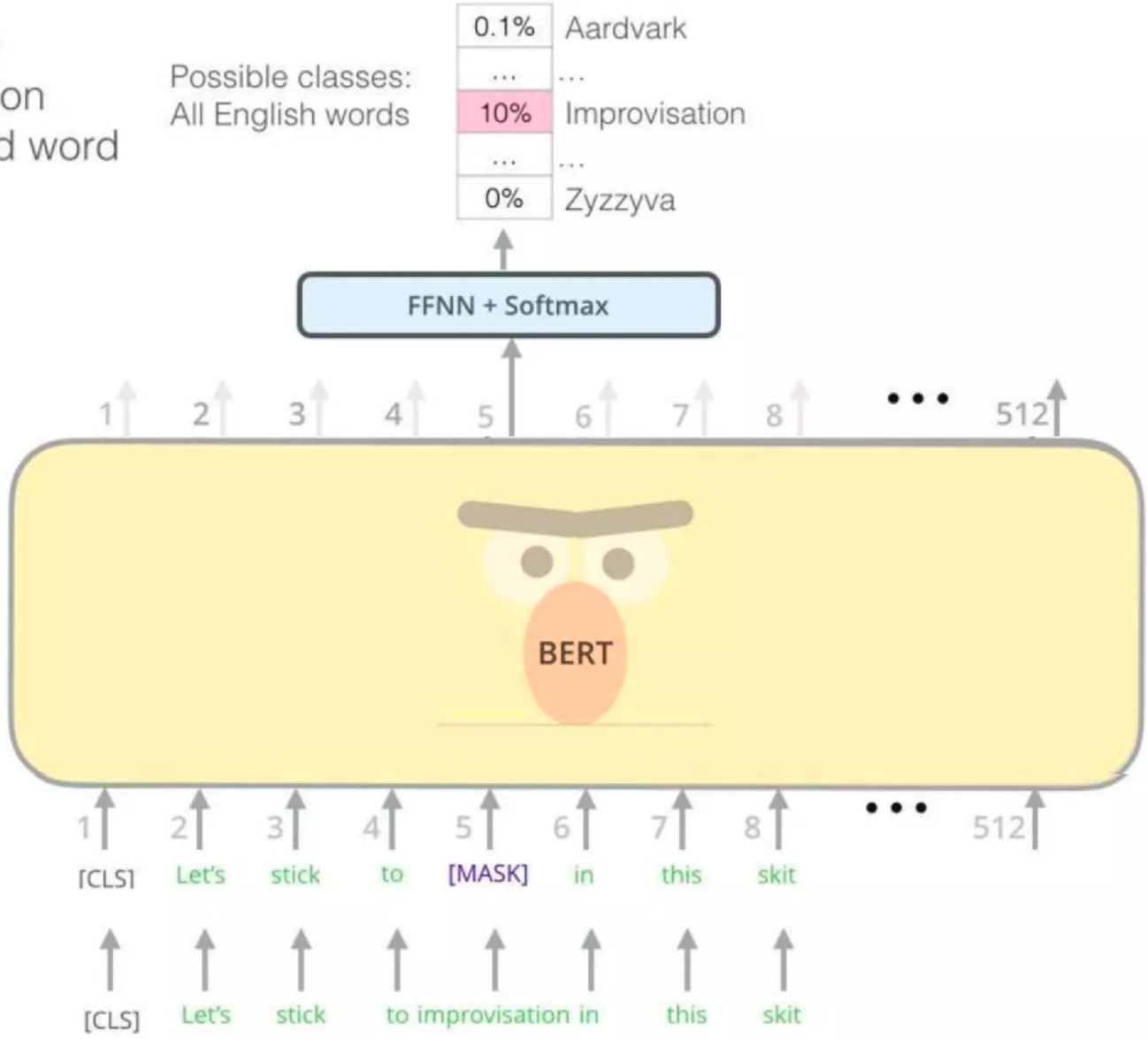
Transformer encoder

- Multi-headed self attention
  - Models context
- Feed-forward layers
  - Computes non-linear hierarchical features
- Layer norm and residuals
  - Makes training deep networks healthy
- Positional embeddings
  - Allows model to learn relative positioning





Use the output of the masked word's position to predict the masked word



Randomly mask 15% of tokens

Input

1 - Semi-supervised training on large amounts of text (books, wikipedia..etc).

The model is trained on a certain task that enables it to grasp patterns in language. By the end of the training process, BERT has language-processing abilities capable of empowering many models we later need to build and train in a supervised way.

Semi-supervised Learning Step

Model:

Dataset:

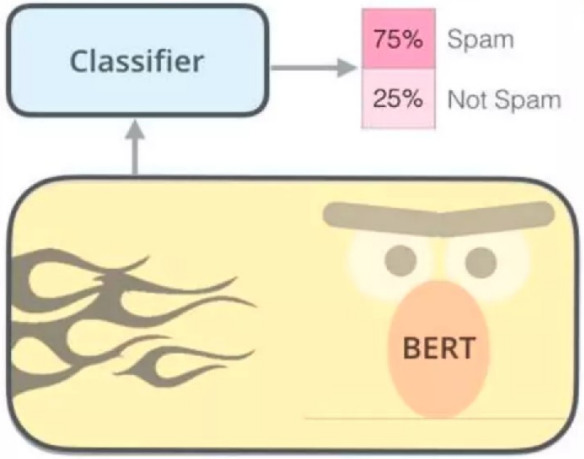
Objective:

Predict the masked word
(langauge modeling)

2 - Supervised training on a specific task with a labeled dataset.

Supervised Learning Step

Model:
(pre-trained in step #1)

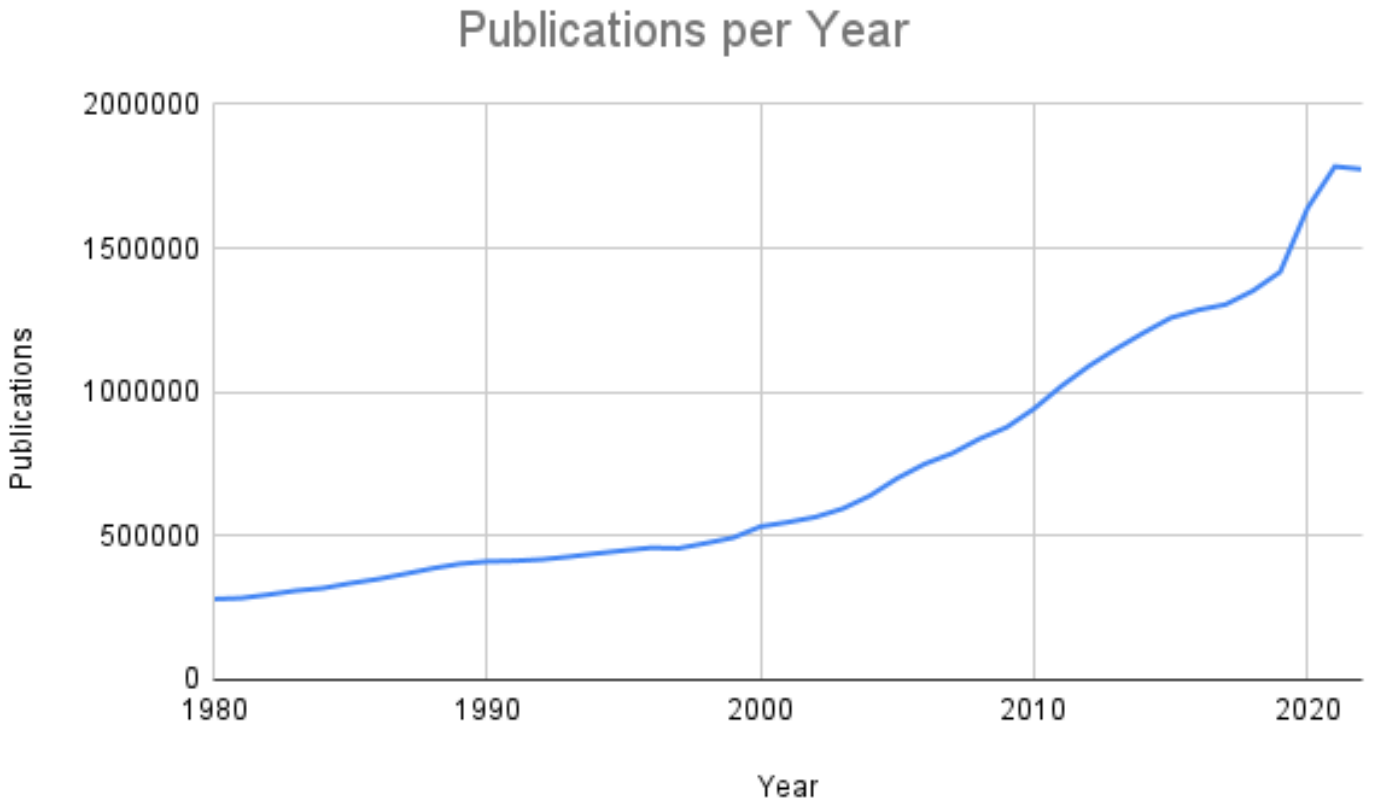


Dataset:

Email message	Class
Buy these pills	Spam
Win cash prizes	Spam
Dear Mr. Atreides, please find attached…	Not Spam

BERT was trained on Wikipedia (~2.5B words) and Google's BooksCorpus (~800M words).
These large informational datasets contributed to BERT's deep knowledge not only of the English language but also of our world

Motivation



More than 2 articles published in biomedical journals every minute!

- Make sure this knowledge is used to the benefit of patients
- Need to make it accessible to biomedical experts
- Automated answering of questions based on semantics
- Benchmarking these systems can achieve a multiplying effect

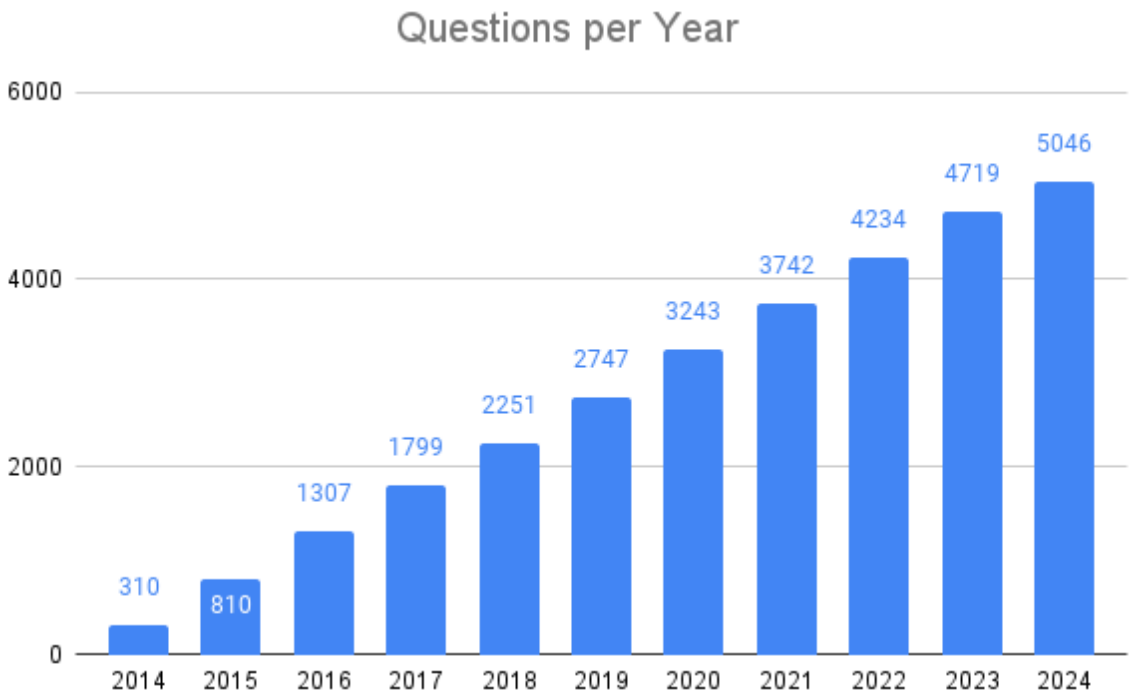
BioASQ: Biomedical Semantic Indexing and Question Answering

- BioASQ initiated 2012 an annual challenge on **biomedical semantic indexing** and **question answering** (QA).

- Participants are required to compose **informative answers** to biomedical natural language questions.
- Includes tasks on question answering, summarization, entity detection, and more
- Funded by the EU and co-organized by several research institutions and companies
- Provides benchmark datasets, evaluation metrics, and prizes
- Attracts participants from various domains, such as NLP, ML, IR, and bioinformatics
- The best indexing system is now used by NIH to assign MeSHterms

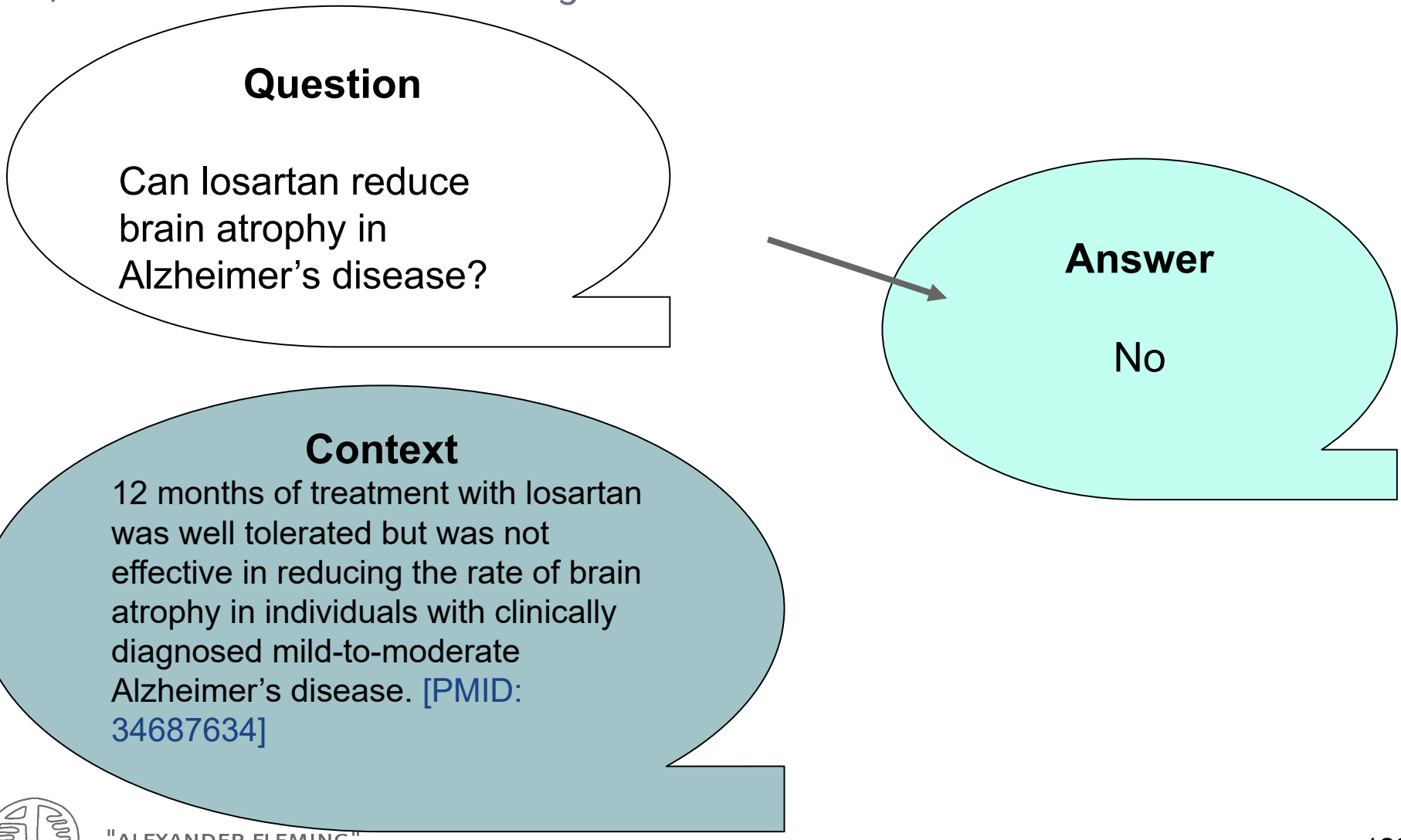
BioASQ Dataset

The development dataset consists of biomedical questions in English, along with their gold concepts, articles, snippets, RDF triples, "exact" answers, and "ideal" answers in JSON format

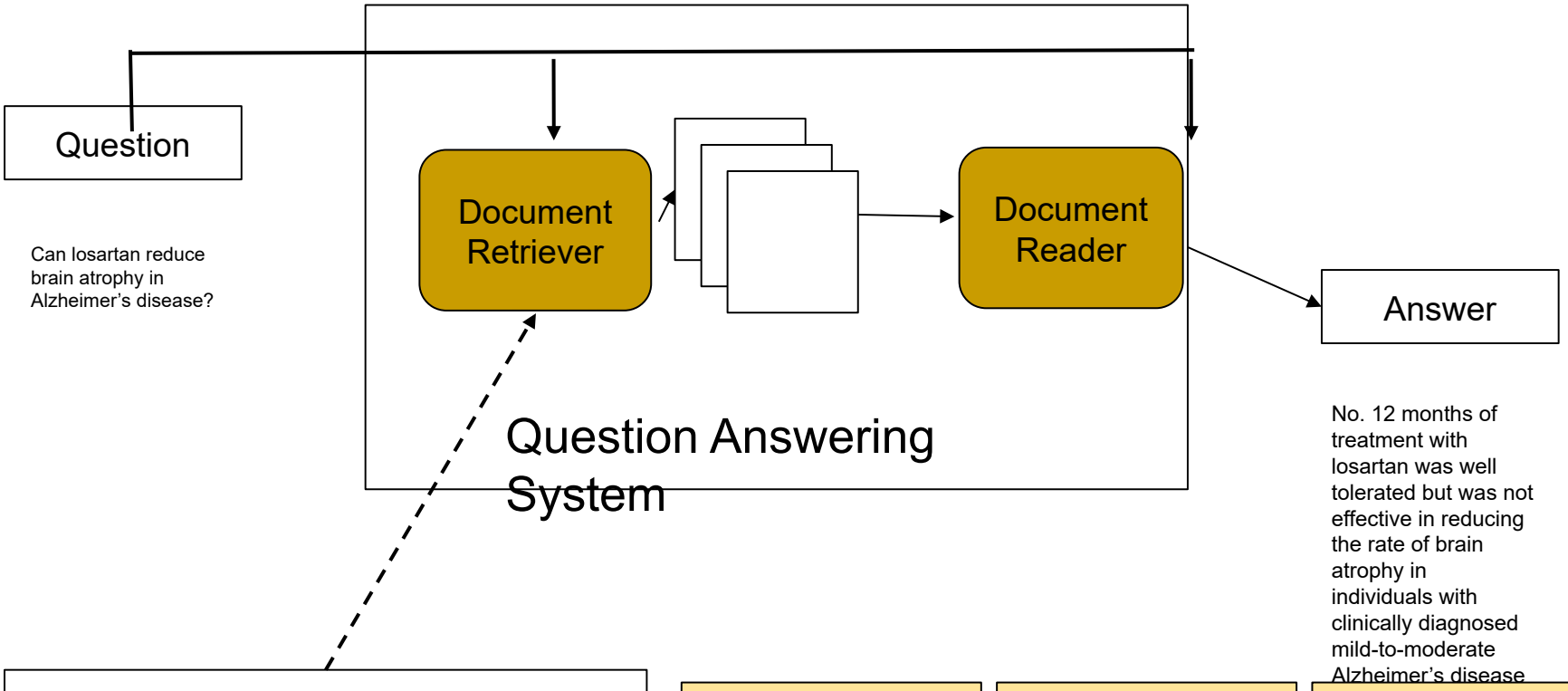


Question-Answer pairs example

Question Answering models can retrieve the answer to a question from a given text, which is useful for searching for an answer in a document.

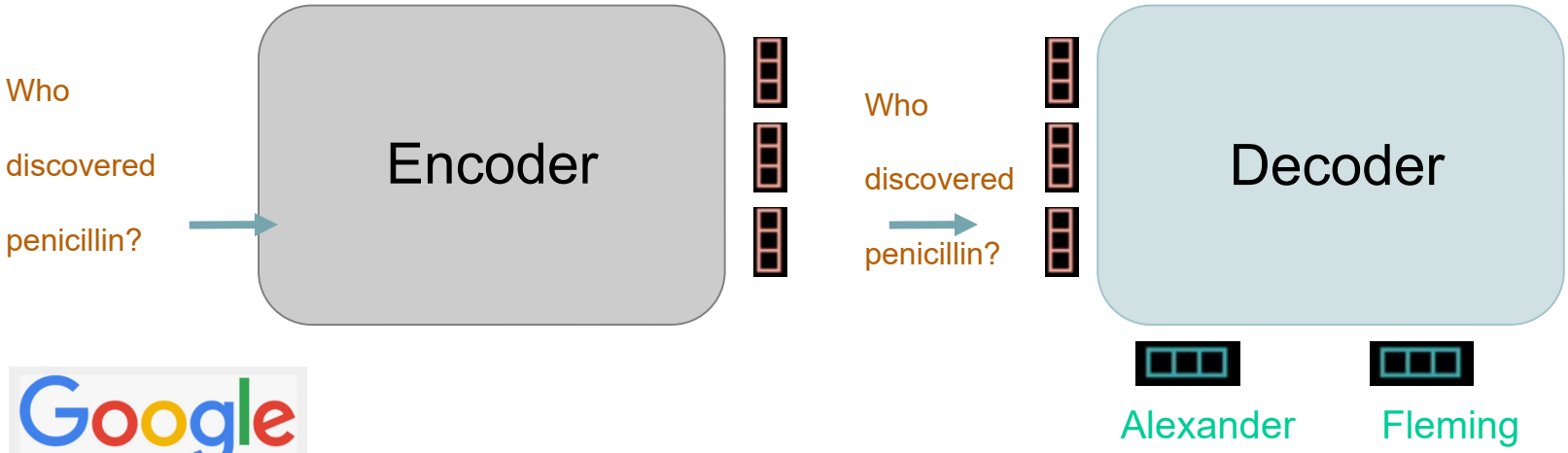


General Question Answering System



Using Transformers

- groundbreaking algorithm pushing the state of the art in many areas
- used in many applications like machine language translation, conversational chatbots, better search engines, image annotation and **is the key algorithm for Alphafold2**



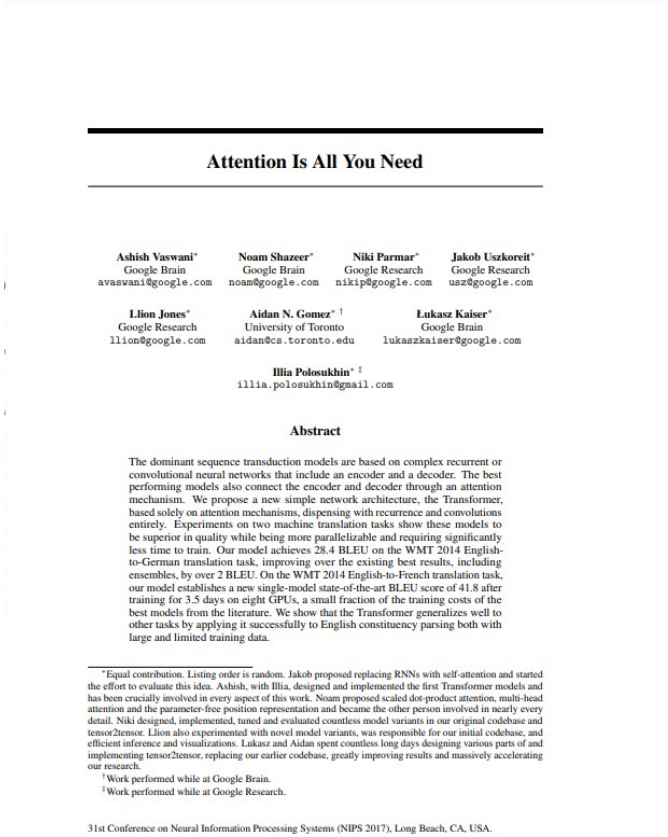
Google

OpenAI GPT-4

Bidirectional **E**ncoder **R**epresentations from **T**ransformers (BERT)

Generative **P**re-trained **T**ransformer (GPT)

Attention in Transformers



Attention Is All You Need

Ashish Vaswani* Google Brain avaswani1@google.com

Noam Shazeer* Google Brain noam@google.com

Niki Parmar* Google Research nikip@google.com

Jakob Uszkoreit* Google Research usz@google.com

Llion Jones* Google Research llion@google.com

Aidan N. Gomez* † University of Toronto aidan@cs.toronto.edu

Łukasz Kaiser* Google Brain lukaszkaiser@google.com

Illia Polosukhin* ‡ illia.polosukhin@gmail.com

Abstract

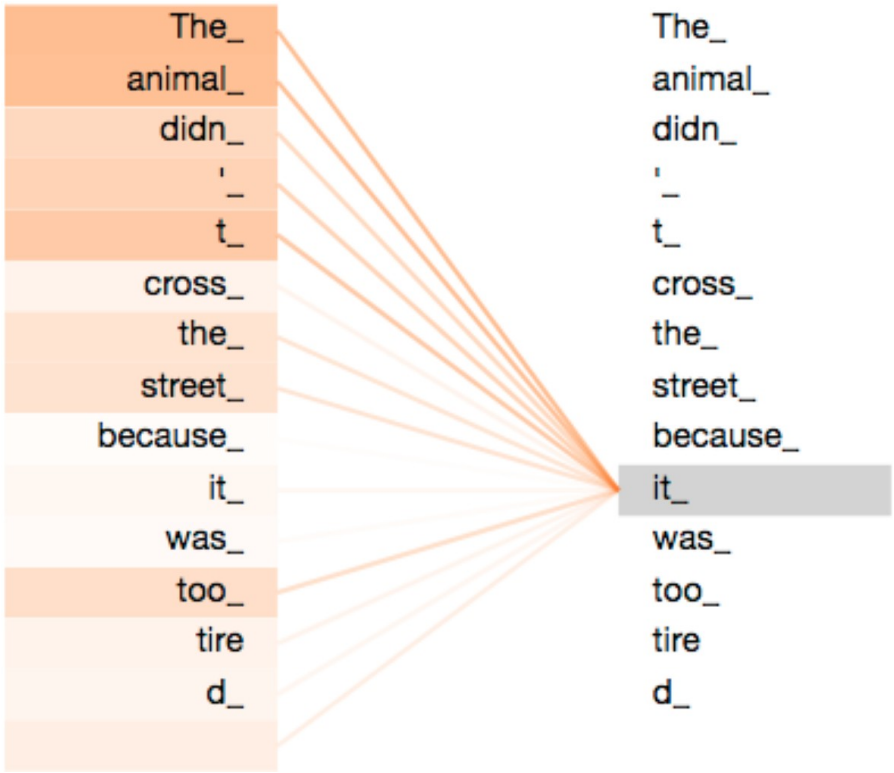
The dominant sequence transduction models are based on complex recurrent or convolutional neural networks that include an encoder and a decoder. The best performing models also connect the encoder and decoder through an attention mechanism. We propose a new simple network architecture, the Transformer, based solely on attention mechanisms, dispensing with recurrence and convolutions entirely. Experiments on two machine translation tasks show these models to be superior in quality while being more parallelizable and requiring significantly less time to train. Our model achieves 28.4 BLEU on the WMT 2014 English-to-German translation task, improving over the existing best results, including ensembles, by over 2 BLEU. On the WMT 2014 English-to-French translation task, our model establishes a new single-model state-of-the-art BLEU score of 41.8 after training for 3.5 days on eight GPUs, a small fraction of the training costs of the best models from the literature. We show that the Transformer generalizes well to other tasks by applying it successfully to English constituency parsing both with large and limited training data.

*Equal contribution. Listing order is random. Jakob proposed replacing RNNs with self-attention and started the effort to evaluate this idea. Ashish, with Illia, designed and implemented the first Transformer models and has been crucially involved in every aspect of this work. Noam proposed scaled dot-product attention, multi-head attention and the parameter-free position representation and became the other person involved in nearly every detail. Niki designed, implemented, tuned and evaluated countless model variants in our original codebase and tensor2tensor. Llion also experimented with novel model variants, was responsible for our initial codebase, and efficient inference and visualizations. Lukasz and Aidan spent countless long days designing various parts of and implementing tensor2tensor, replacing our earlier codebase, greatly improving results and massively accelerating our research.

†Work performed while at Google Brain.

‡Work performed while at Google Research.

31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA.



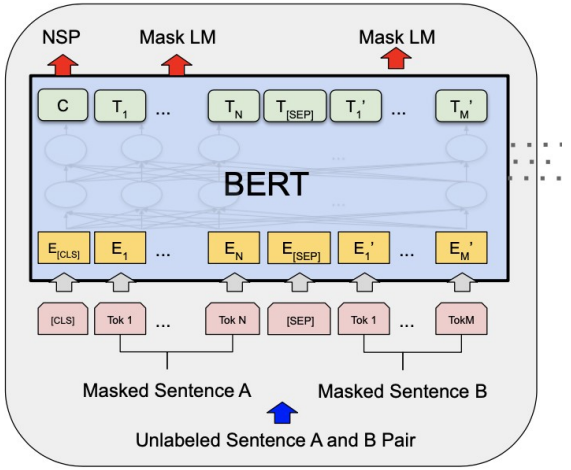
Attention mechanism can detect all relations in input sequences

Transformer models

1. BERT
Bidirectional **E**ncoder **R**epresentations from **T**ransformers, 2018

BERT was trained on Wikipedia (~2.5B words) and Google's BooksCorpus (~800M words).
These large informational datasets contributed to BERT's deep knowledge not only of the English language but also of our world

- Masked Language Model
- Next Sentence Prediction
- Transfer learning

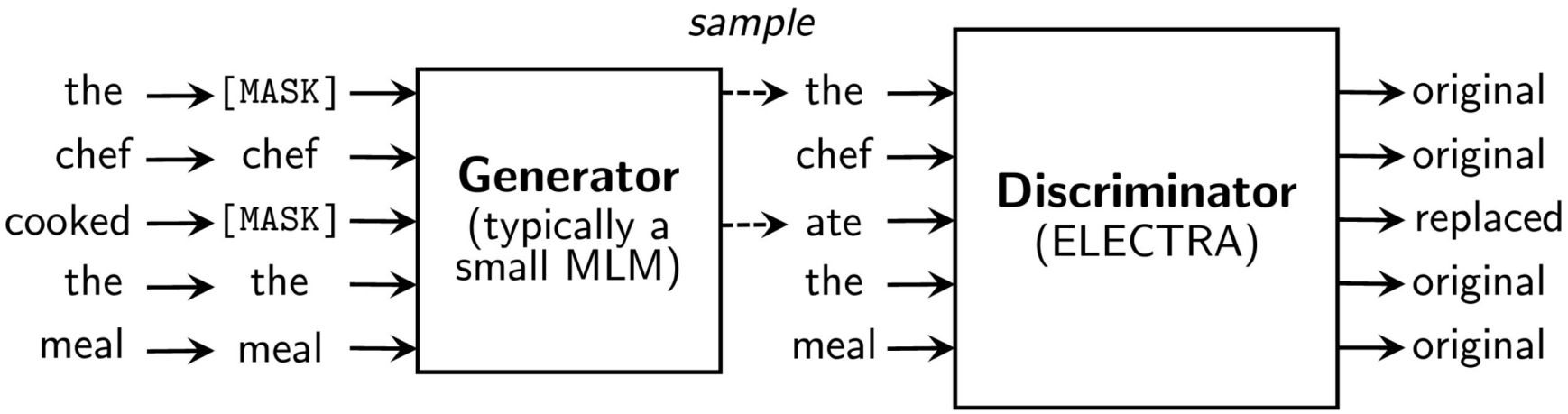


Pre-training

=> Finetune to final task

1. ELECTRA

Efficiently Learning an Encoder that Classifies Token Replacements Accurately
(K. Clark et al. 2020)



2. ALBERT

(Z. Lan et al. 2020)

- Factorized embedding parameterization
- Cross-layer parameter sharing
- Inter-sentence coherence loss / Sentence order prediction (SOP)

speed

performance

1. + 2. = ELECTR-A-LBERT = ELECTROLBERT

"ALEXANDER FLEMING"
Biomedical Sciences Research Center

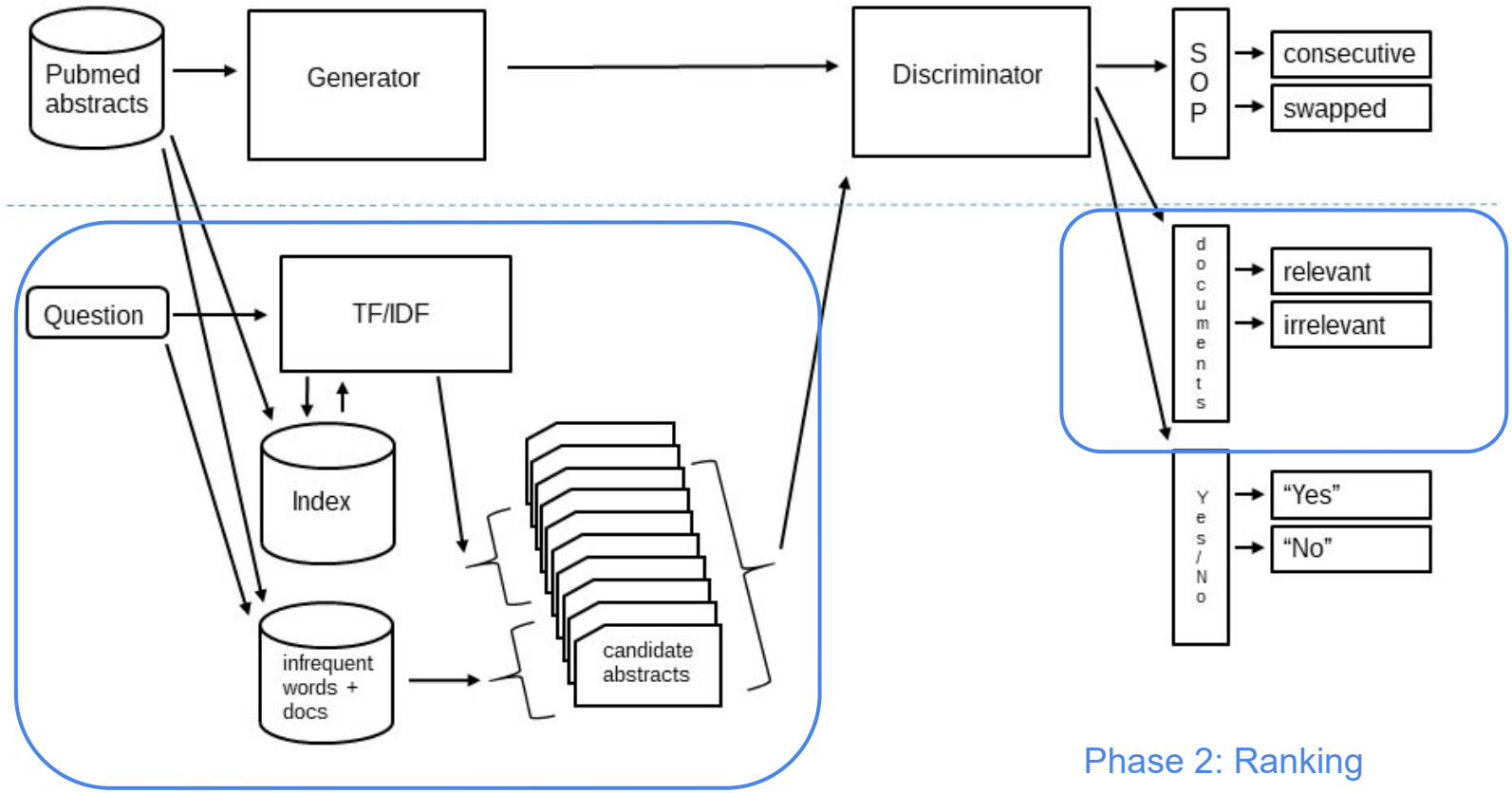
Question-Answer pairs in real abstracts

Coherence detected by sentence order prediction (SOP) facilitates answer identification. The topic-shift detection obtained with next sentence prediction (NSP) is less specific.

PMID	Question Answer
34884907	Neurogenic Inflammation in the Context of Endometriosis-What Do We Know? Endometriosis (EM) is an estrogen-dependent disease characterized by the presence of epithelial, stromal, and smooth muscle cells outside the uterine cavity.
34894155	The question to ask is, is this prescribed load regimen congruent with Wolff's law, and does it provide an adequate mechanical stimulus to maintain the functional health of periodontal complex? This question was answered by studying the effects of mice chewing on soft food (SF) and hard food (HF) while undergoing experimental tooth movement (ETM).
34893939	Will Artificial Intelligence (AI) re-humanize or de-humanize medicine? As AI becomes pervasive in clinical medicine, we argue that the ethical framework that sustains a responsible implementation of such technologies should be reconsidered.

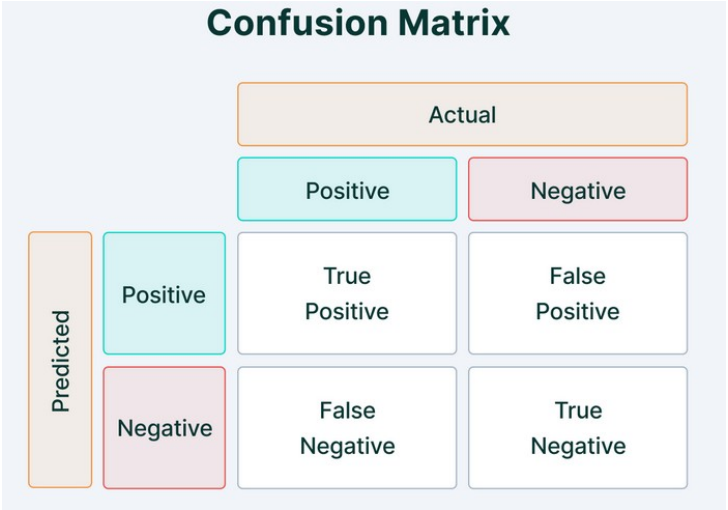
Coherent scientific text has a semantic flow, each sentence builds upon the previous

ELECTROLBERT Architecture



Evaluation: Mean Average Precision (MAP)

Mean Average Precision(mAP) is the current benchmark metric used by the computer vision research community to evaluate the robustness of



- Calculate the confusion matrix—TP, FP, TN, FN.
- Calculate the precision and recall metrics.

Precision=TP/(TP+FP) Recall=TP/(TP+FN)

$$AP = \frac{\sum_{r=1}^{|L|} P(r) \cdot rel(r)}{|L_R|} \qquad MAP = \frac{1}{n} \cdot \sum_{i=1}^{n} AP_i$$

- Measure the average pre

ELECTROLBERT performance in BioASQ10, 2022

"ALEXANDER FLEMING"
Biomedical Sciences Research Center

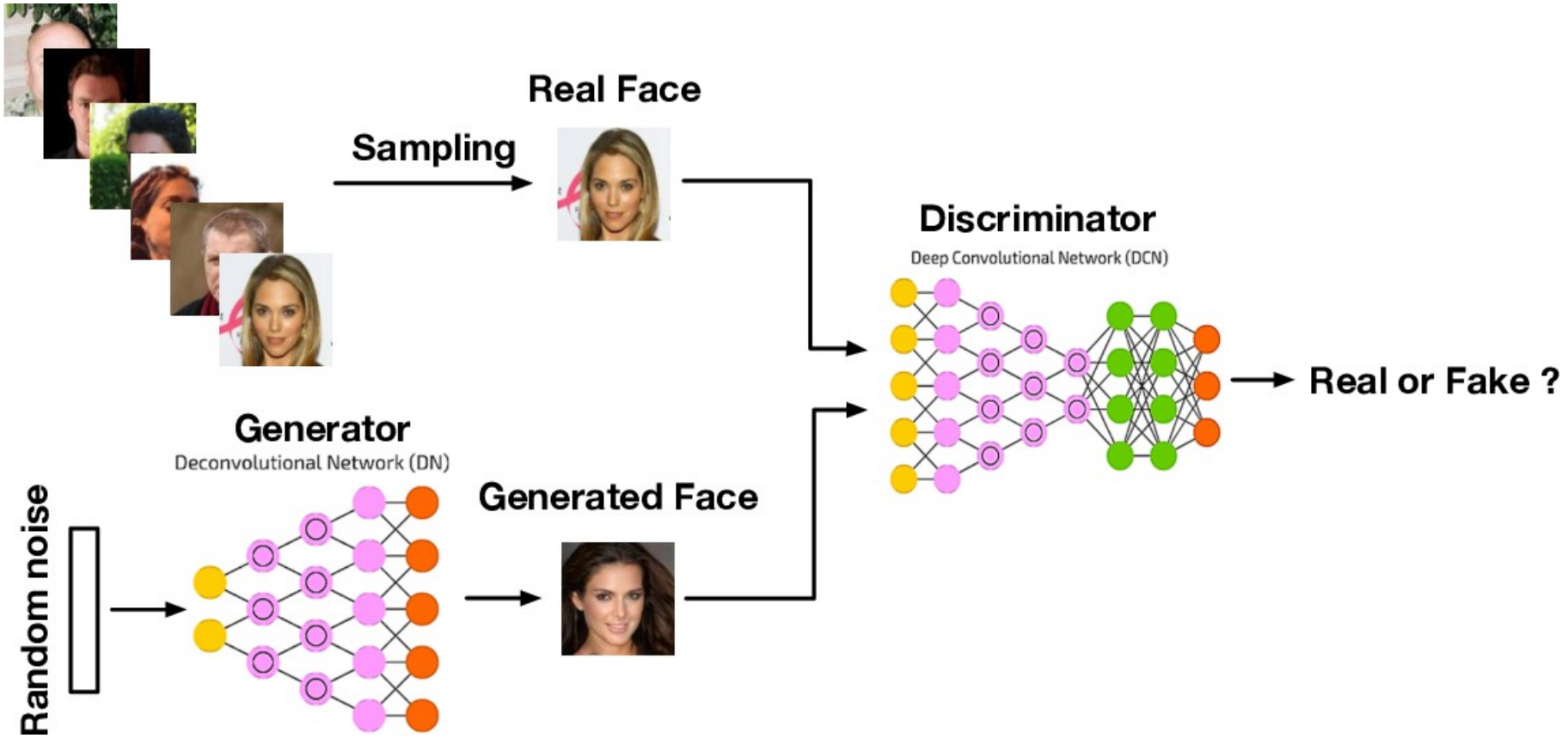
Document relevance and "yes/no" task. For documents, the mean average precision (MAP) is used for evaluation.

	documents				"yes/no" type questions	
	BioASQ submission		final system			
batch	MAP	per team rank	MAP	per team rank	accuracy	per team rank
1	0.1121	7	0.3649	5	-	-
2	0.1632	9	0.3090	5	-	-
3	0.3209	8	0.3666	6	0.76	10
4	0.3101	6	0.3140	6	0.75	11
5	0.3242	4	0.3242	4	0.6429	10
6[2]	0.0977	3	0.0977	3	**1.0**	**1**

[2]Batch 6 consisted of questions posed by new biomedical experts interested in material and answers that can be automatically provided by state-of-the-art IR and QA systems. It is not part of the official evaluation.

More details at http://ceur-ws.org/Vol-3180/paper-24.pdf

Generative Adversarial Networks (GANs)



GANBERT

Why to use GANBERT?

- In the era of information overload, answering questions accurately is crucial.
- Obtaining high-quality annotated data is expensive
- Fine-tuning of pretrained Large Language Models (LLM) helps, but risks overfitting
- Semi-supervised technique enhances generalization capabilities using extensive unlabeled text data, expanding LLM scope to handle alternative semantic formulations

GANs

How does it work?

- GANBERT extends BERT by incorporating unlabeled data within a Generative Adversial Network (GAN) framework
- Works by pitting two neural networks against each other
  - generator (G) is trained to produce internal BERT representations resembling the distribution of unlabeled data
  - discriminator (D) learns to distinguish between generator samples and real instar

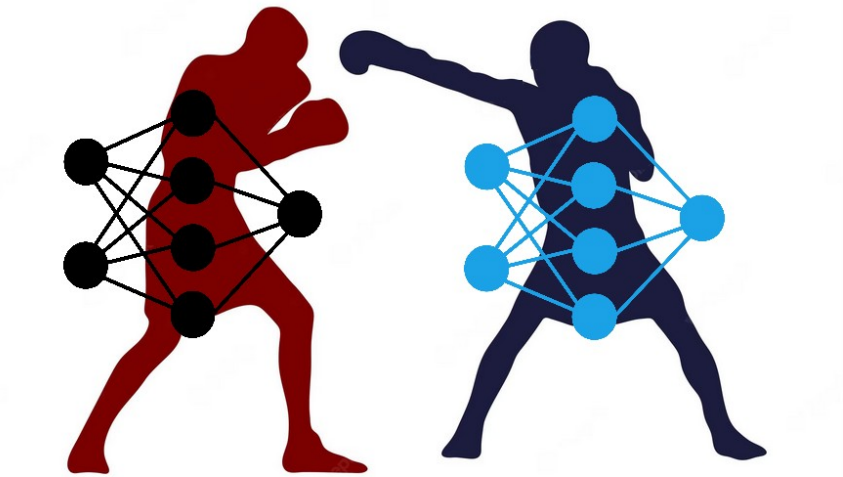
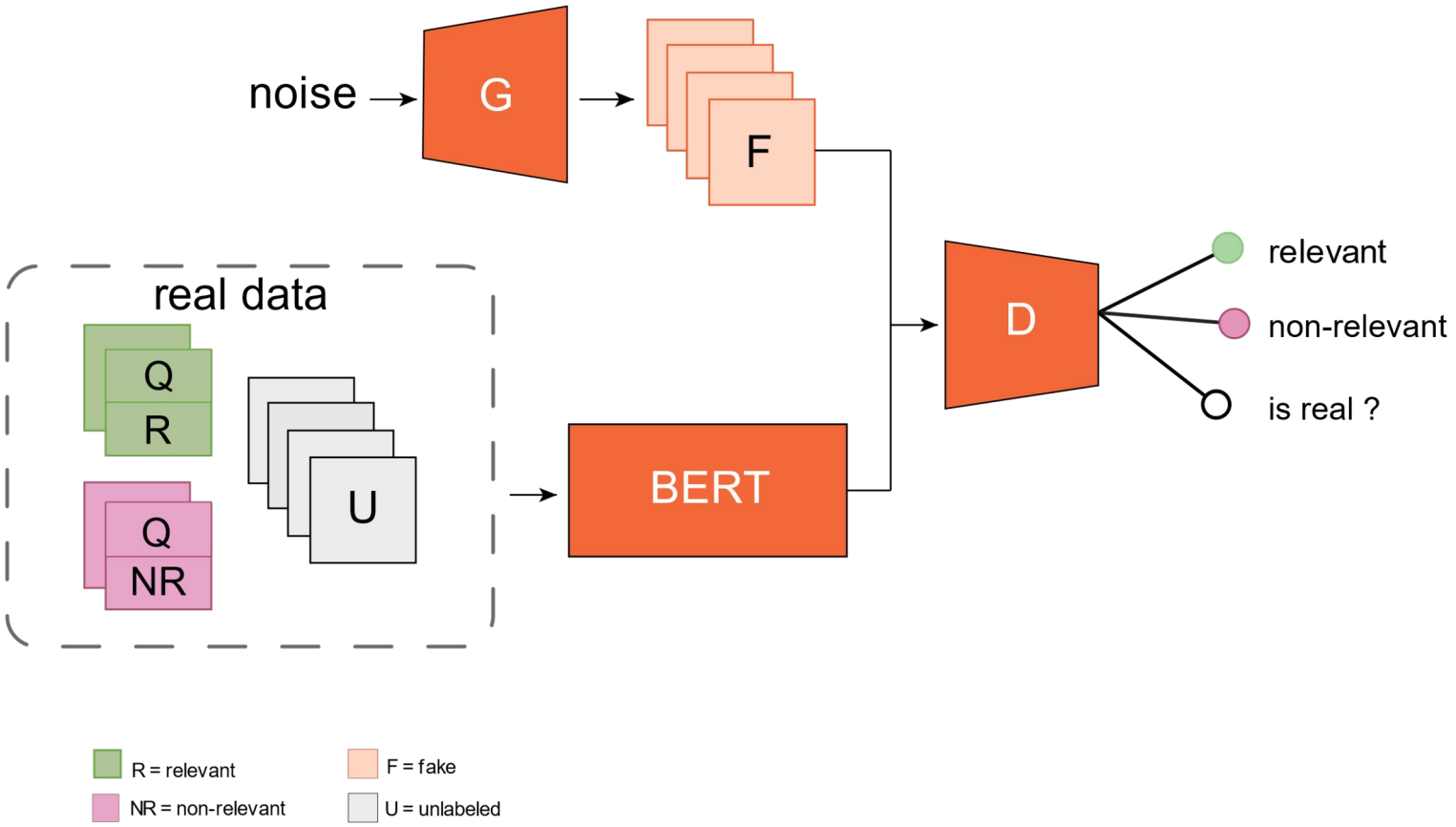


Image based on https://www.freepik.com/premium-vector/ & https://www.clipartmax.com/

GANBERT Architecture



GANBERT hyperparameter optimization

Model	LR	SLEN	LABEL_MASK	Unlabeled set	Noise generation	MAP_{b123}	MAP_{b4}
BERT	$2e-6$	200	-	-	-	0.3166	0.2231
GANBERT	**$2e-6$**	**200**	**0.02**	**BioASQ10**	**uniform[0,1]**	**0.3468**	0.2279
" , $SLEN_{predict} = 200$	"	"	"	"	"	0.3453	0.2223
" , $SLEN_{predict} = 150$	"	"	"	"	"	0.3456	0.2262
"	"	175	"	"	"	0.3459	0.2166
"	"	225	"	"	"	0.3363	0.2152
"	"	200	0.01	"	"	0.3395	**0.2301**
"	"	"	0.05	"	"	0.3380	0.2181
"	$1e-6$	"	0.02	"	"	0.3266	0.2066
"	$5e-6$	"	"	"	"	0.3334	0.2200
"	$2e-6$	"	"	BioASQ10+BoolQ	uniform[0,1]	0.3315	0.2166
"	"	"	"	BioASQ10	uniform[-1,1]	0.3333	0.2214
"	"	"	"	"	normal[0,1]	0.3289	0.2220

ELECTROLBERT & GANBERT performance in BioASQ11, 2023

batch	*MAP*	system	per team rank	model details
1	**0.4590**	bioinfo-0	1	
	0.3875	ELECTROLBERT-2,3	4	base model, $SLEN_{predict} = 200$, $ndocs = 11500$
	0.3732	ELECTROLBERT-0,1	4	base model, $SLEN_{predict} = 250$, $ndocs = 11500$
2	**0.3852**	bioinfo-4	1	
	0.3252	ELECTROLBERT-2	4	base model, $SLEN_{predict} = 175$, $ndocs = 6750$
	0.2942	ELECTROLBERT-0	4	base model, $SLEN_{predict} = 250$, $ndocs = 6750$
	0.2781	ELECTROLBERT-3	4	base model, $SLEN_{predict} = 275$, $ndocs = 6750$
	0.2513	ELECTROLBERT-1	4	Query expansion using Roccio's methond
				base model, $SLEN_{predict} = 400$, $ndocs = 300$
3	**0.3185**	dmiip2	1	
	0.2502	ELECTROLBERT-0	4	base model, $SLEN_{predict} = 175$, $ndocs = 60$
	0.2336	ELECTROLBERT-4	4	base model, $SLEN_{predict} = 175$, $ndocs = 16$
	0.2326	ELECTROLBERT-2	4	large model, $SLEN_{predict} = 200$, $ndocs = 13$
	0.2296	ELECTROLBERT-1	4	base model, $SLEN_{predict} = 150$, $ndocs = 300$
	0.2261	ELECTROLBERT-3	4	large model, $SLEN_{predict} = 150$, $ndocs = 11$
4	**0.3224**	dmiip3	1	
	0.2279	*ELECTROLBERT-1*	3	GANBERT4, $SLEN_{predict} = 175$, $ndocs = 10$
	0.2271	*ELECTROLBERT-4*	3	GANBERT3, $SLEN_{predict} = 175$, $ndocs = 10$
	0.2242	*ELECTROLBERT-3*	3	GANBERT2, $SLEN_{predict} = 175$, $ndocs = 11$
	0.2147	*ELECTROLBERT-2*	3	GANBERT1, $SLEN_{predict} = 175$, $ndocs = 16$
	0.1849	ELECTROLBERT-0	3	base model, $SLEN_{predict} = 175$, $ndocs = 60$

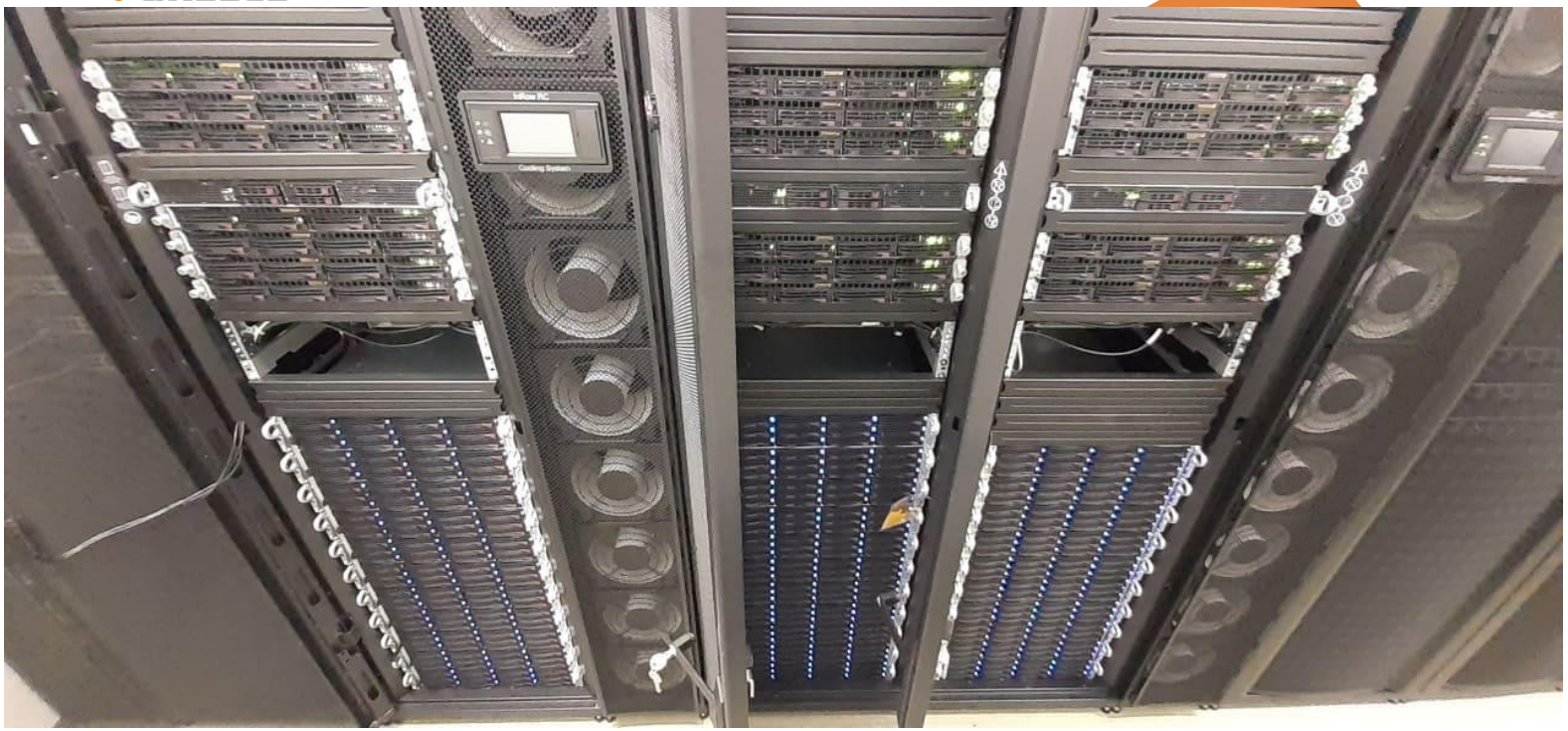
Base model: embedding size 768

Large model: embedding size 1024
pretrained for 30 million steps

GANBERT: finetuning for ~50K steps

Computational resources

GPU computations were offered by HYPATIA, the Cloud infrastructure that supports the computational needs of the Greek ELIXIR community



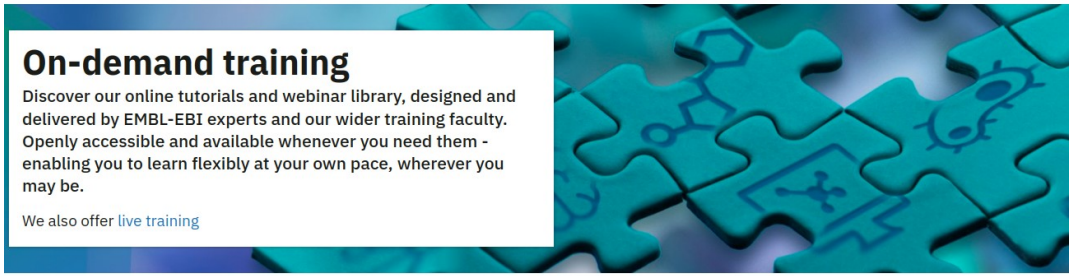
https://www.ebi.ac.uk/training/on-demand?query=single%20cell&domain=ebiw…

lomics bio dev med etc cs phys tools

EMBL-EBI Services Research Training About us EMBL-EBI

EMBL-EBI Training > On-demand training

Login Register



On-demand training

Discover our online tutorials and webinar library, designed and delivered by EMBL-EBI experts and our wider training faculty. Openly accessible and available whenever you need them - enabling you to learn flexibly at your own pace, wherever you may be.

We also offer live training

Refine

Type

- [ ] Course materials (18)
- [ ] Recorded webinar (7)
- [ ] Online tutorial (33)

Duration

- [ ] Less than 1 hour (10)
- [ ] 1 to 3 hours (27)
- [ ] More than 3 hours (3)

single cell

Search

Showing **1 - 10** out of **58** results for single cell in on-demand training

COURSE MATERIALS

Single-cell RNA-seq and network analysis using Galaxy and Cytoscape

These materials include those from the course Single-cell RNA-seq & network analysis using Galaxy and Cytoscape, which ran in April 2021. The materials utilise Galaxy pipelines, an...

RECORDED WEBINAR

Life cell by cell: Introduction to Single Cell Expression Atlas

Explore gene expression profiles across various species and conditions at single cell resolution. Single Cell Expression Atlas (SCEA) is an open resource committed to making the information from...

COURSE MATERIALS

Single-cell RNA-seq analysis using Galaxy

The 'Single-cell RNA-seq analysis using Galaxy' course, which focused on the droplet-based scRNA-seq analysis pipelines from raw reads to trajectories, for mammalian and plant genomes, ran in March...

RECORDED WEBINAR

Single cell RNA-seq analysis using a Galaxy interface

In this webinar, we will look at a Galaxy interface for single cell analysis. Specifically, we will run Scanpy (which would otherwise require Python programming skills) to analyse a Drop-seq dataset...

COURSE MATERIALS

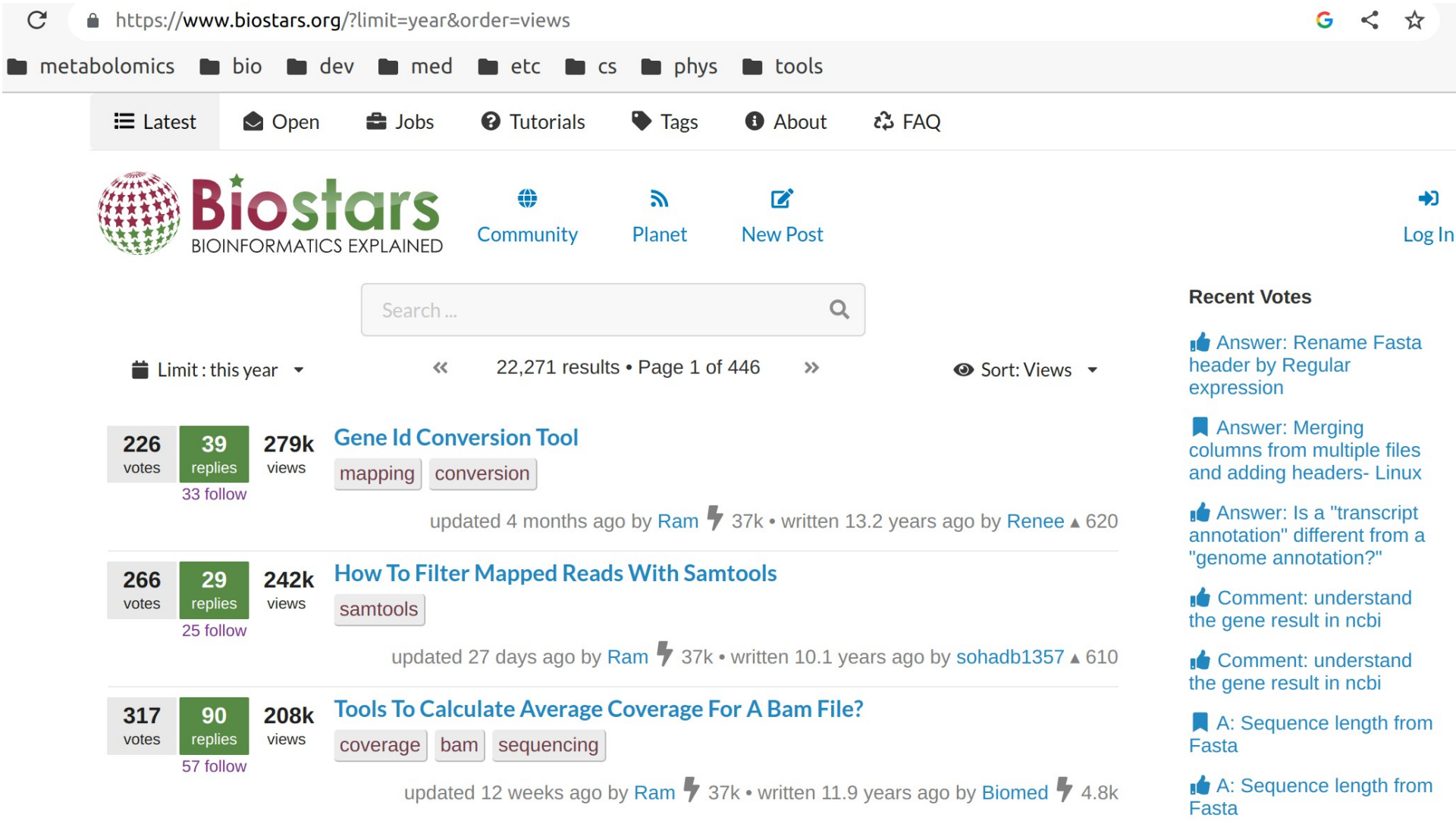
Single-cell RNA-seq analysis using R

This set of materials includes those from the 2022 course Single-Cell RNA-seq Analysis using R; it covers the analysis of scRNA-seq data using R and command line tools. Lectures and practical...

RECORDED WEBINAR

Identifying tumor cells at the single cell level through machine learning

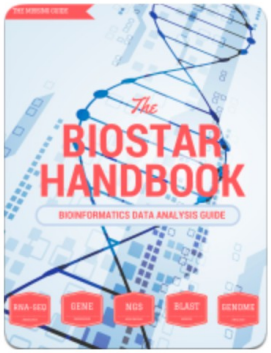
Tumours are highly complex tissues composed of cancerous cells, surrounded by a heterogeneous cellular microenvironment. A tumor's response to treatments is governed by an interaction of the cancer...


https://www.biostars.org/?limit=year&order=views
metabolomics bio dev med etc cs phys tools
Latest Open Jobs Tutorials Tags About FAQ
Biostars
BIOINFORMATICS EXPLAINED
Community Planet New Post Log In
Search ...
Limit : this year
22,271 results • Page 1 of 446
Sort: Views
226 votes 39 replies 279k views
33 follow
Gene Id Conversion Tool
mapping conversion
updated 4 months ago by Ram 37k • written 13.2 years ago by Renee ▲ 620
266 votes 29 replies 242k views
25 follow
How To Filter Mapped Reads With Samtools
samtools
updated 27 days ago by Ram 37k • written 10.1 years ago by sohadb1357 ▲ 610
317 votes 90 replies 208k views
57 follow
Tools To Calculate Average Coverage For A Bam File?
coverage bam sequencing
updated 12 weeks ago by Ram 37k • written 11.9 years ago by Biomed 4.8k
Recent Votes
Answer: Rename Fasta header by Regular expression
Answer: Merging columns from multiple files and adding headers- Linux
Answer: Is a "transcript annotation" different from a "genome annotation?"
Comment: understand the gene result in ncbi
Comment: understand the gene result in ncbi
A: Sequence length from Fasta
A: Sequence length from Fasta

The Biostar Handbook: 2nd Edition

Welcome to the Biostar Handbook

Last updated on **December 13, 2022**, ISBN: **978-0-578-80435-4**



The Biostar Handbook introduces readers to bioinformatics, the scientific discipline at the intersection of biology, computer science, and statistical data analytics dedicated to the digital processing of genomic information.

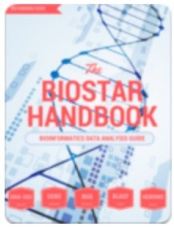
The Biostar Handbook has been developed, improved and refined over more than a half decade in a research university setting while used in an accredited Ph.D. level training program. The contents of this book have provided the analytical foundation to thousands of students, many of whom have become full-time bioinformaticians and work at the most innovative companies in the world.

The Biostar Handbook Collection

The Biostar Handbook is being reworked into separate, more manageable volumes of study.

- The Biostar Handbook - An introduction to Bioinformatics as a scientific field.
- The Art of Bioinformatics Scripting - Learn advanced Unix and Bash scripting skills.
- RNA-Seq by Example - Master RNA-Seq data analysis.
- Corona Virus Genome Analysis - Advanced topics devoted to the study of the Corona Virus.
- Biostar Workflows - Create automated bioinformatics workflows

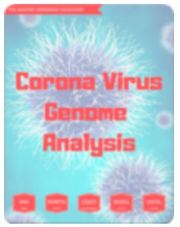
Access to new books and new content is included with your subscription!



The Biostar Handbook

Bioinformatics Scripting

RNA-Seq by Example



Corona Virus Genome Analysis

Biostar Workflows

Pick a paper to present on 13+20 January

	Paper	source/year	Presented by (pls enter your name in a free field)
1	The nature of the last universal common ancestor and its impact on the early Earth system (with overview of method paper IQ-TREE 2)	Nature Ecology & Evolution 2024	
2	RiNALMo: general-purpose RNA language models can generalize well on structure prediction tasks	Nature Communications 2025	
3	Learning the natural history of human disease with generative transformers	Nature 2025	
4	Metabolic modeling elucidates phenformin and atpenin A5 as broad-spectrum antiviral drugs against RNA viruses	Communications Biology 2025	
5	Systematic benchmark of state-of-the-art variant calling pipelines identifies major factors affecting accuracy of coding sequence variant discovery	BMC Genomics. 2022	
6	A universal snp and small-indel variant caller using deep neural networks	Nat Biotechnol. 2018	
7	Protocol for unbiased, consolidated variant calling from whole exome sequencing data	STAR Protocols 2022	
8	Quality assessment of RNA 3D structure models using deep learning and intermediate 2D maps	bioRxiv 2025	
9	Scaling Large Language Models for Next-Generation Single-Cell Analysis	bioRxiv 2025	
10	Continuous Thought Machines	arXiv 2025	
11	Lyra: An Efficient and Expressive Subquadratic Architecture for Modeling Biological Sequences	arXiv 2025	
12	Titans: Learning to Memorize at Test Time	arXiv 2025	
13	Biomolecular dynamics with machine-learned quantum-mechanical force fields trained on diverse chemical fragments	Science Advances 2024	
14	Discriminating physiological from non-physiological interfaces in structures of protein complexes: a community-wide study	Proteomics 2023	
15	A Survey of Large Language Models for Healthcare: from Data, Technology, and Applications to Accountability and Ethics	Proc of IEEE 2023	
16	scBERT as a large-scale pretrained deep language model for cell type annotation of single-cell RNA-seq data	Nature Machine Intelligence 2022	
17	Accelerating science with human-aware artificial intelligence	Nature Human Behaviour 2023	
18	Integrating end-to-end learning with deep geometrical potentials for ab initio RNA structure prediction	Nature Communications 2023	
19	Pairing a high-resolution statistical potential with a nucleobase-centric sampling algorithm for improving RNA model refinement	Nature Comm. 2021	
20	Accurate structure prediction of biomolecular interactions with AlphaFold 3	Nature 2024	
21	A Survey on Transformers in Reinforcement Learning	Machine Learning Research 2023	
22	Antibody-Antigen Docking and Design via Hierarchical Equivariant Refinement	ICML 2022	
23	A self-supervised deep learning method for data-efficient training in genomics	Communications Biology 2023	
24	Data-driven discovery of innate immunomodulators via machine learning-guided high throughput screening	Chemical Science 2023	
25	Simulating 500 million years of evolution with a language model	bioRxiv 2024	
26	Broadly applicable and accurate protein design by integrating structure prediction networks and diffusion generative models	bioRxiv 2023	
27	Predicting structures of large protein assemblies using combinatorial assembly algorithm and AlphaFold2	bioRxiv 2023	
28	Unifying Large Language Models and Knowledge Graphs: A Roadmap	arXiv 2023	

https://tinyurl.com/35ka9bmz

Seasons greetings with some deep bioinformatics...

https://kingteeshops.com/product/artificial-intelligence-machine-learning-christmas-tree-t-shirt/