Abordagens ômicas integrativas em uma perspectiva de biologia de sistemas
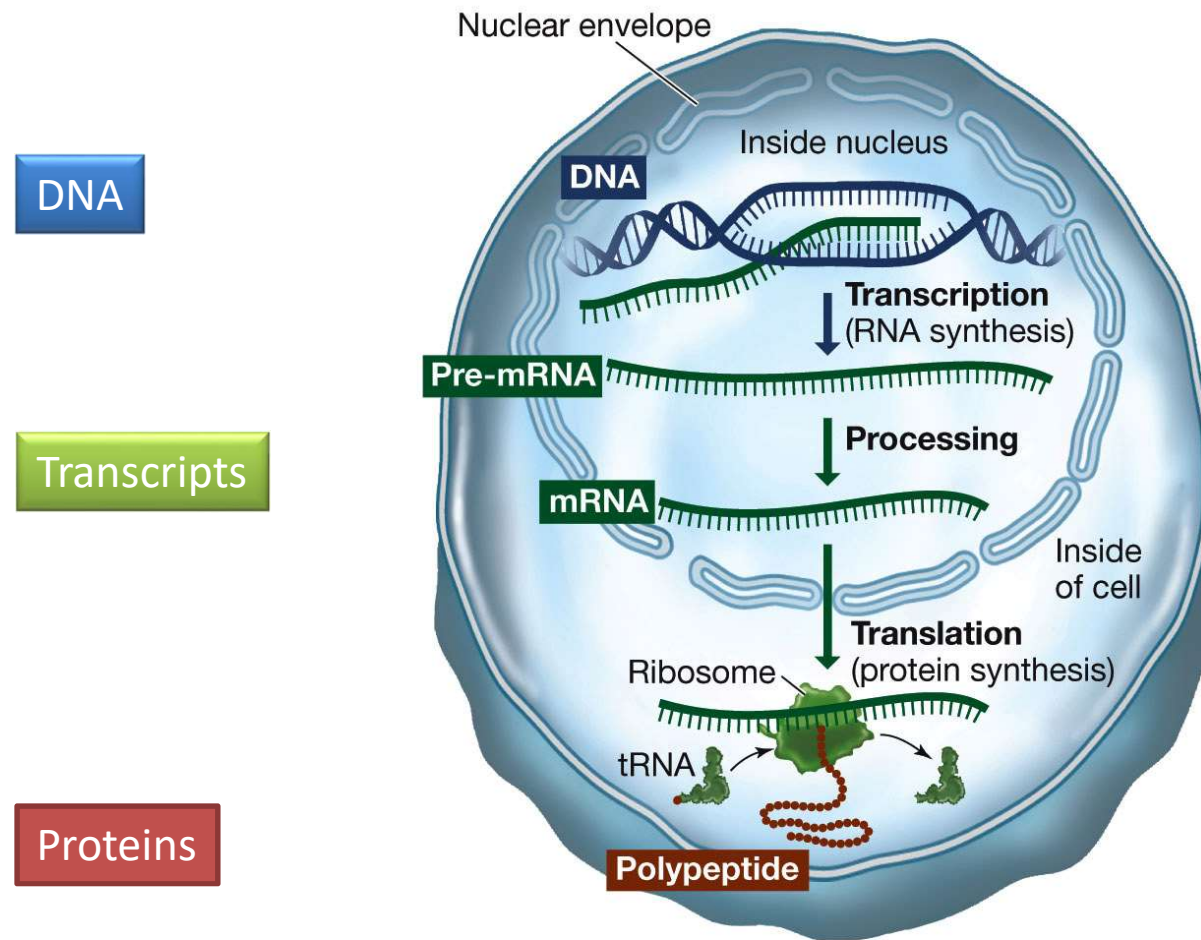
Flavia Vischi Winck

winck@iq.usp.br
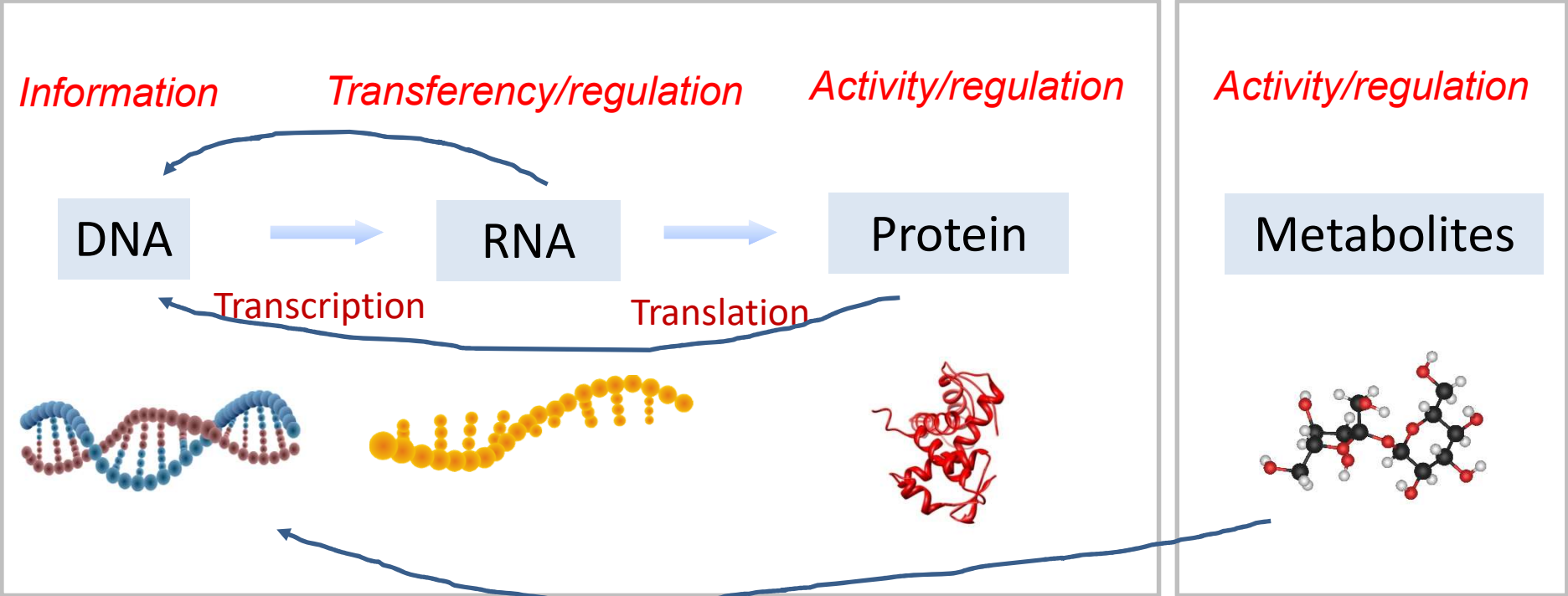
São Paulo, 31 de Maio de 2017

Art, system theory and biochemistry

Julian Trevelyan, 1936. *A Symposium.*

Biological functions are the translation of the inter relationships between molecules

DNA

Transcripts

Proteins

Nuclear envelope

Inside nucleus

DNA

Transcription (RNA synthesis)

Pre-mRNA

Processing

mRNA

Inside of cell

Translation (protein synthesis)

Ribosome

tRNA

Polypeptide

How cells control the composition of their biomass?

Information *Transferency/regulation* *Activity/regulation* *Activity/regulation*

DNA → RNA → Protein Metabolites

Transcription Translation

Cellular repertoire of macromolecules

Genome Transcriptome Proteome Metabolome

Integrative approach in a Systems Biology framework

„Systems biology is the science of discovering, modeling, understanding and ultimately engineering at the molecular level the **dynamic relationships** between the biological molecules that define living organisms. „

(Leroy Hood PhD, MD. President of Institute for Systems Biology, Seatle U.S.A.)

Systems biology (my view)

Holistic approach for studying biological systems with the aim of identifying and understanding systems level features through the integrative analysis of different layers of biological information.

System theory has long been discussed

> "The physical world is not simply a sum of spatial and temporal single worlds running one besides the others, and many phenomena escape [entziehen sich] the understanding when one does not consider a physical object [Gebilde] as a whole" (Max Planck 1929:17).

Bertalanffy (1928:69–70) characterized a systemic state ("Gestalt") as comprising properties that cannot be found by simple addition of the components' properties and that furthermore disappear when the "Gestalt" is destroyed [Systemzustand] (Bertalanffy 1929c:89).

"Systems biology" articles in a decade

Chuang et al., Ann. Rev. Cell. Dev. Biol. 2010. 26:721-44.
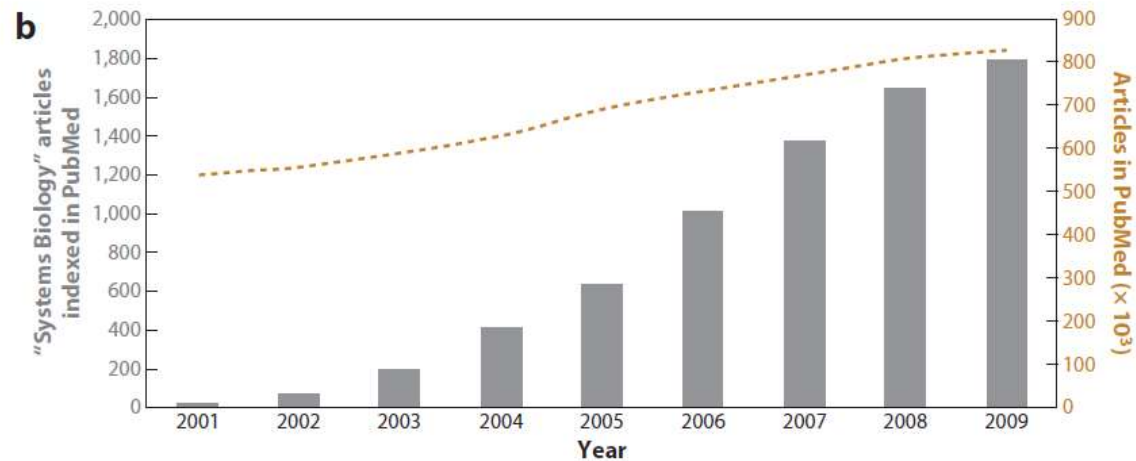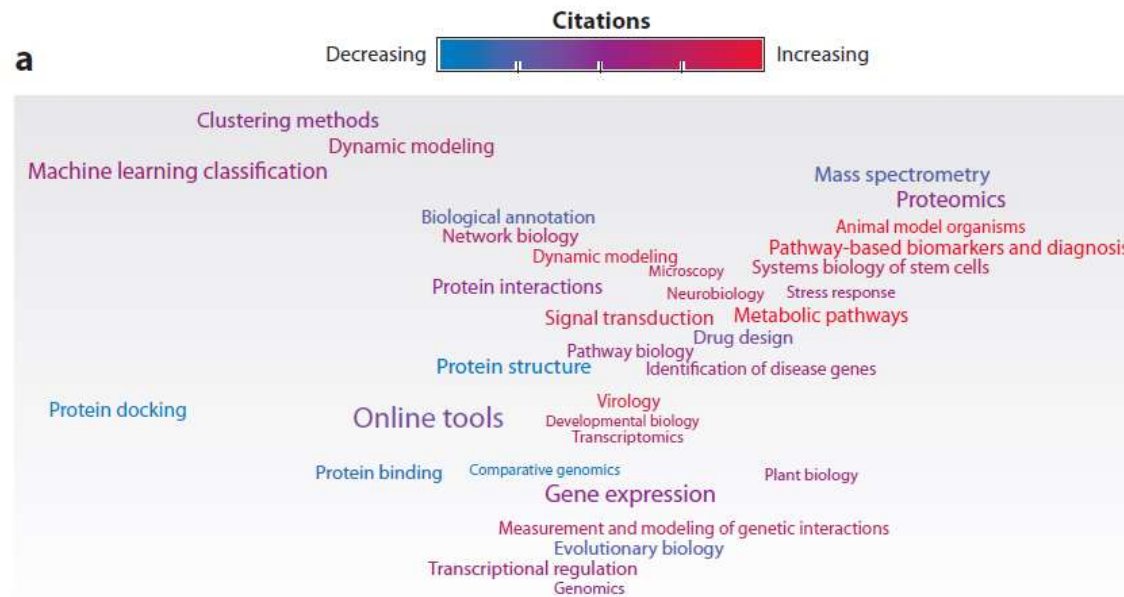
Wholeness behaviour

How to identify the laws concerning the relationships among ALL parts of the biological system?

substrates

product

enzyme
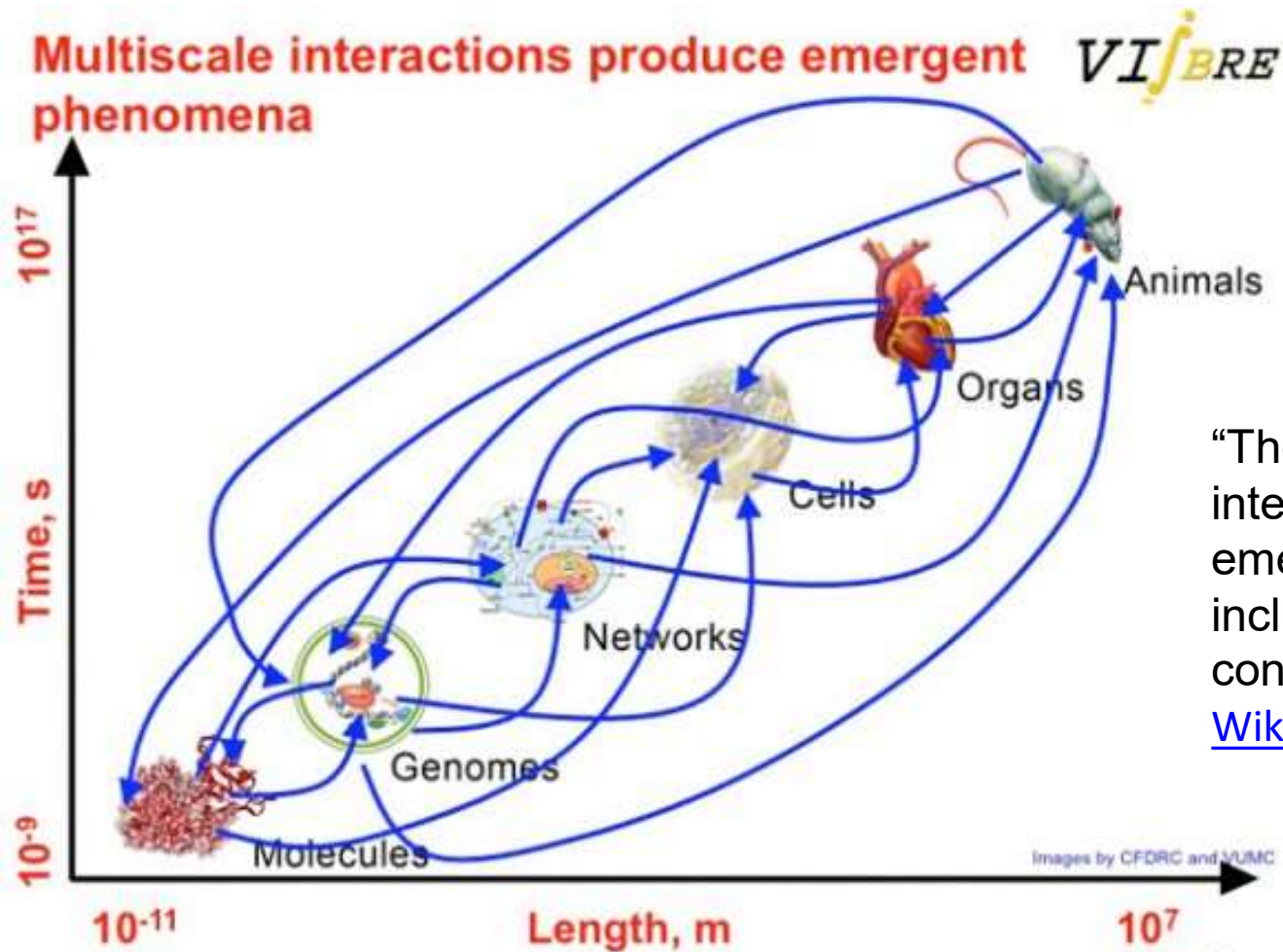
enzyme-substrate complex

enzyme
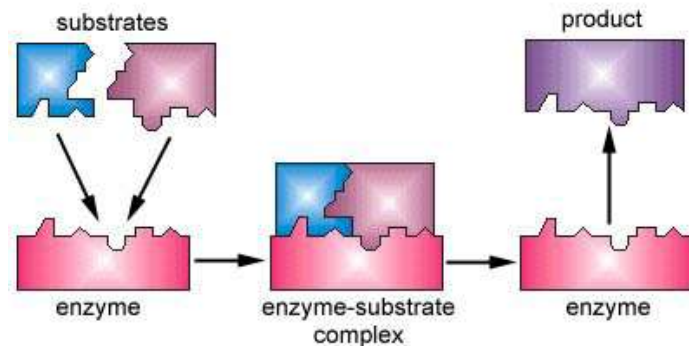
biowiki.ucdavis.edu

www.cellsignal.com

Relationship among parts

Relationship among ALL parts

Complex systems

Multiscale interactions produce emergent phenomena — VIBRE

"These multi-scale interactions produce emergent phenomena, including life and consciousness" (John P. Wikswo)

https://news.vanderbilt.edu/2011/10/13/robot-biologist/
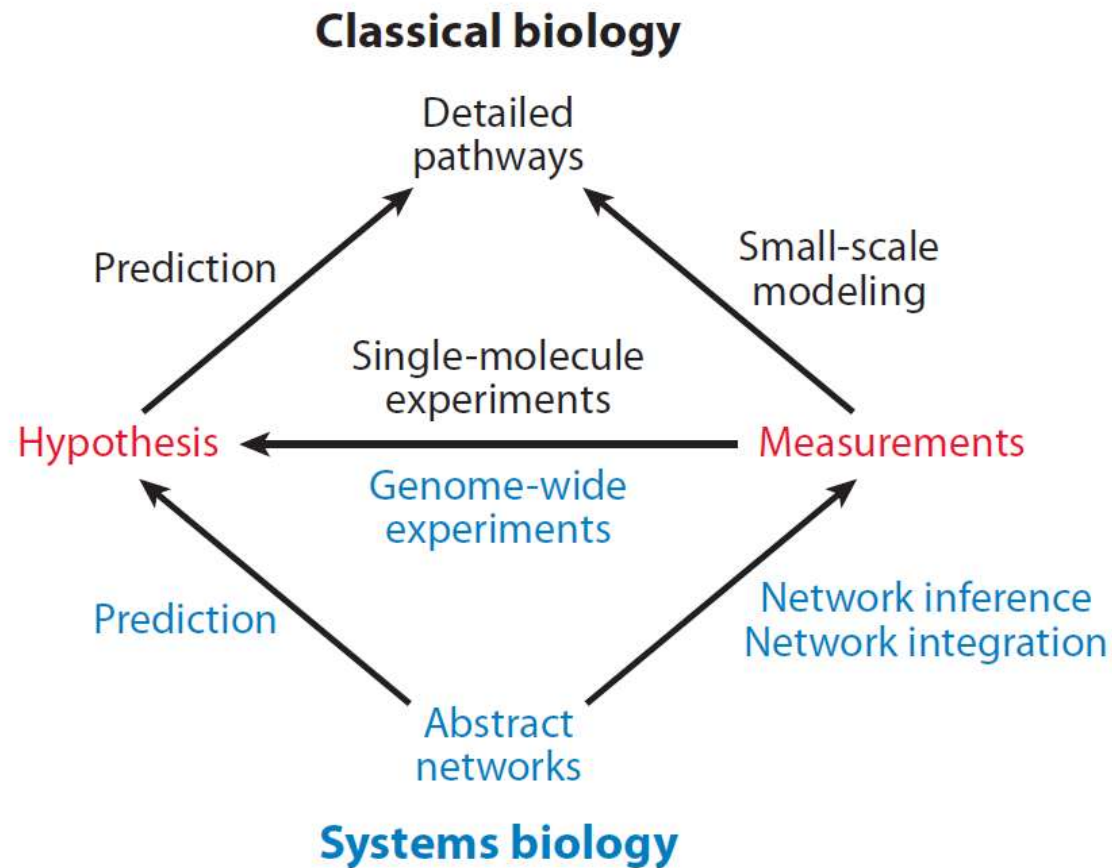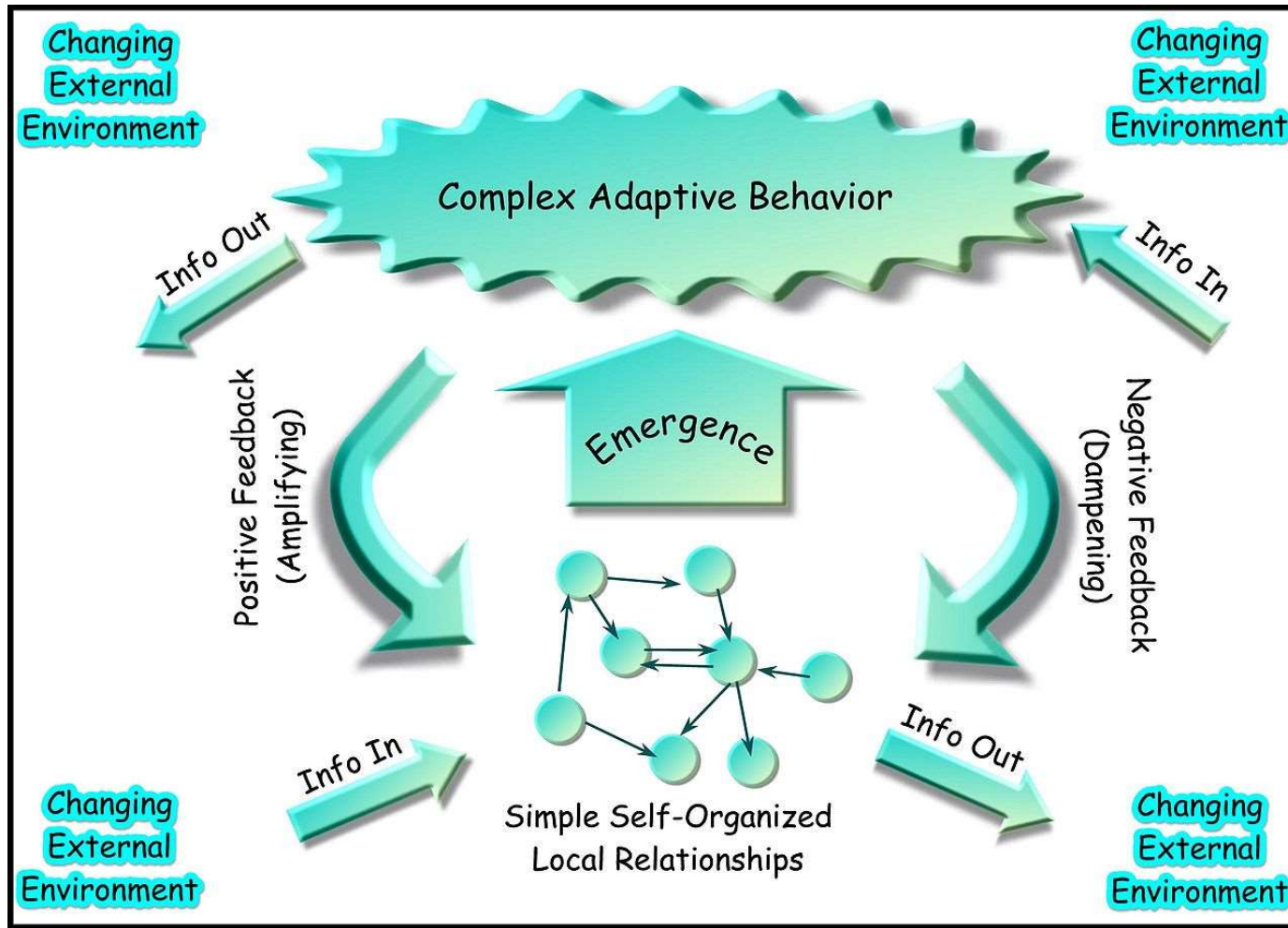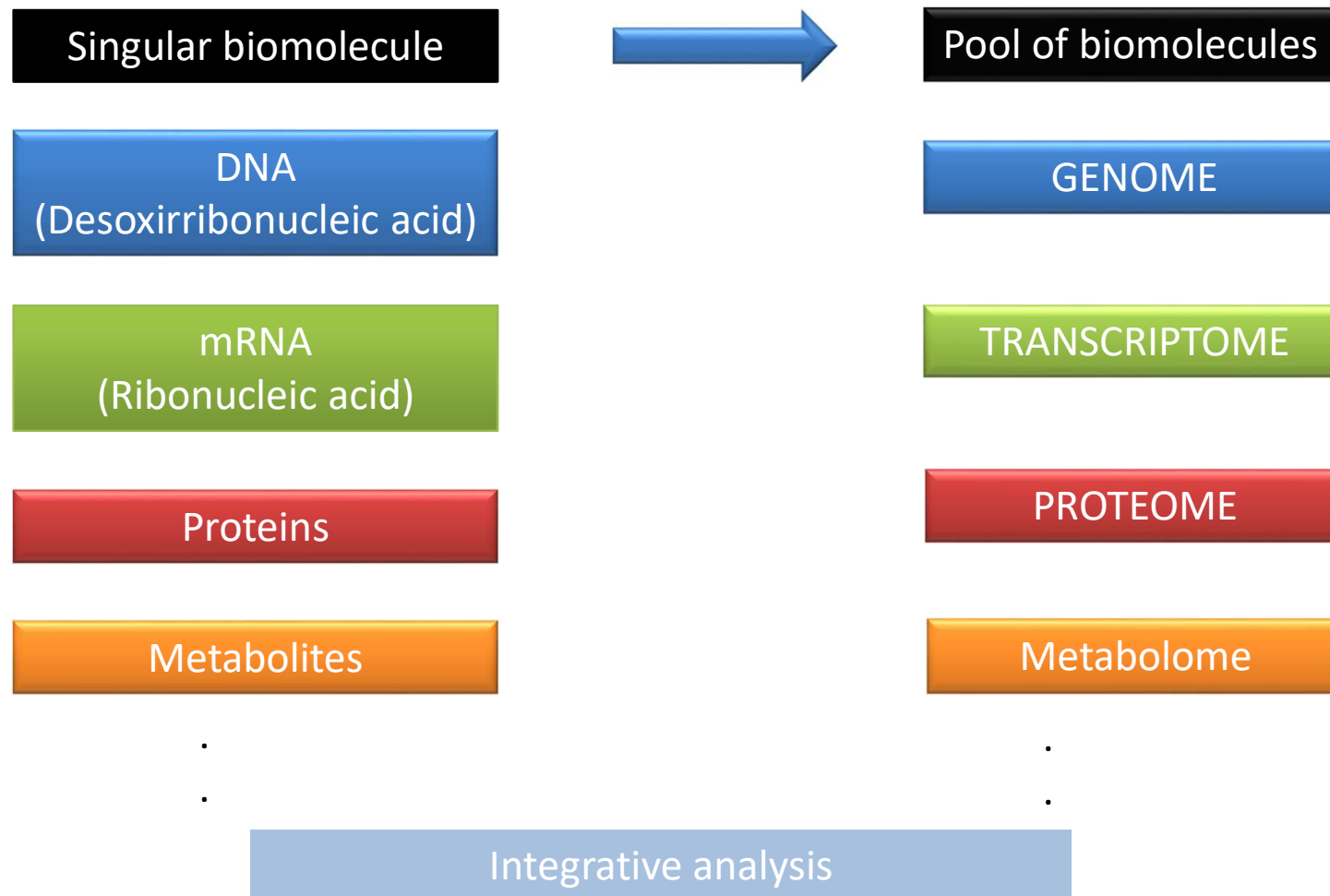
Complementary approaches

Figure 2

Overview of the experimental process in classical biology (*top*) versus systems biology (*bottom*).

Complex systems and emergent properties

Identification and quantification of pools of biomolecules – the "Omics" approaches

| Singular biomolecule | → | Pool of biomolecules |

| DNA (Desoxirribonucleic acid) | GENOME |

| mRNA (Ribonucleic acid) | TRANSCRIPTOME |

| Proteins | PROTEOME |

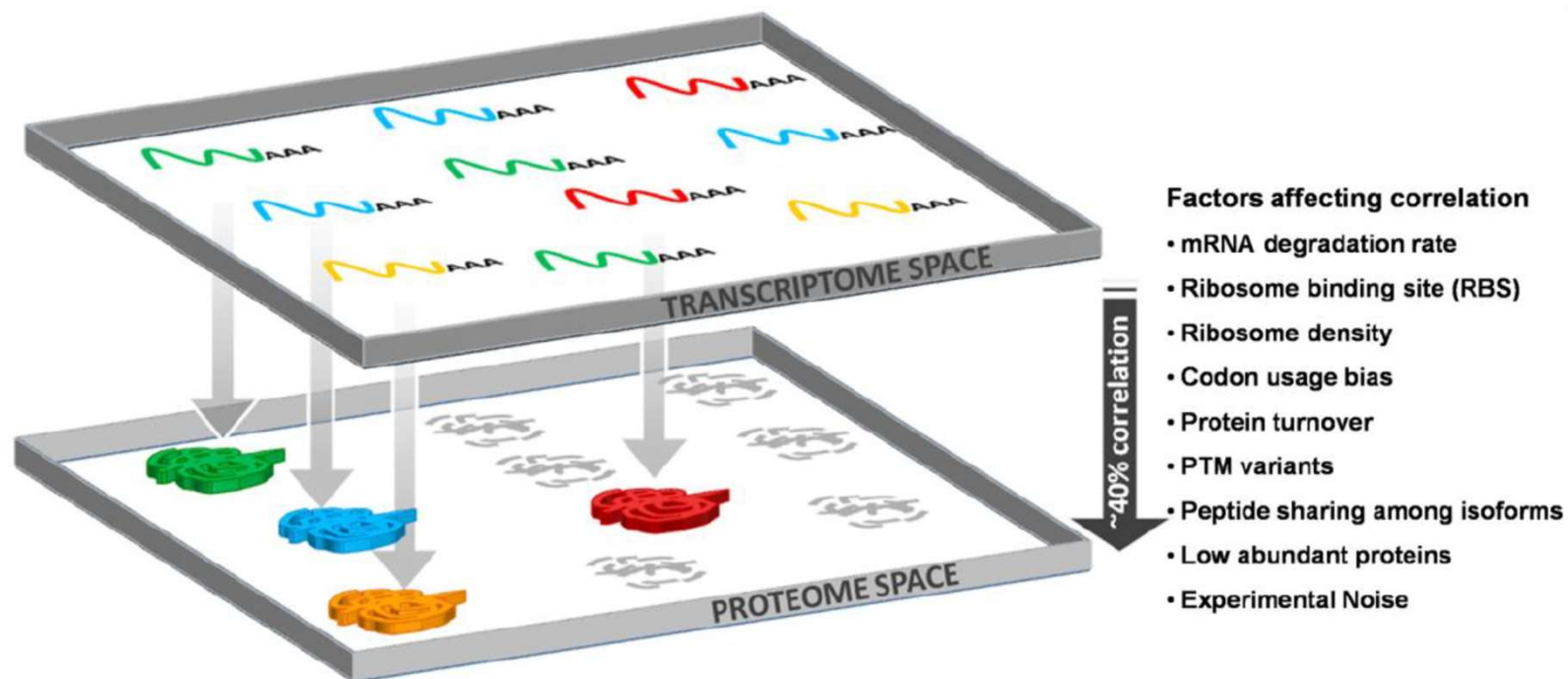| Metabolites | Metabolome |

.
.

.
.

Integrative analysis

Integrative OMICS analysis is on the horizon

We define *data integration* as the use of multiple sources of information (or data) to provide a better understanding of a system/situation/association/etc. (Gomez-Cabrero et al_2014, **DOI:** 10.1186/1752-0509-8-S2-I1)

Current Opinion in Chemical Biology
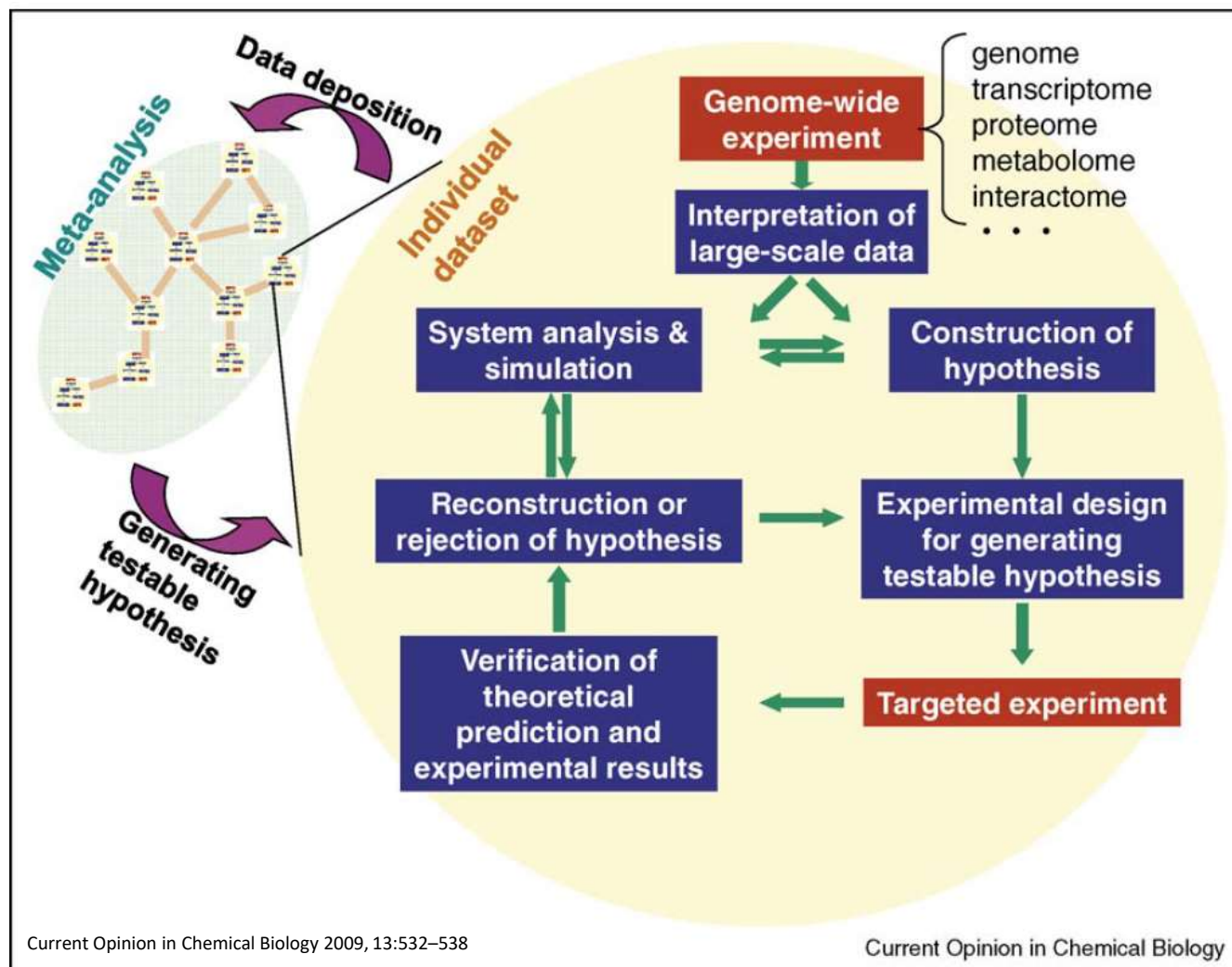
Different OMICS datasets are not always directly correlated

Figure 2: Factors influencing the correlation between mRNA-protein quantities.

Factors affecting correlation

- mRNA degradation rate
- Ribosome binding site (RBS)
- Ribosome density
- Codon usage bias
- Protein turnover
- PTM variants
- Peptide sharing among isoforms
- Low abundant proteins
- Experimental Noise

Systems biology approach (one of many)

Chlamydomonas reinhardtii

Perturbation of single parameter

Top-down data
- complete
- quantitative
- accurate
- reproducible
- high-throughput
- kinetics

Theoretical Model

Prediction
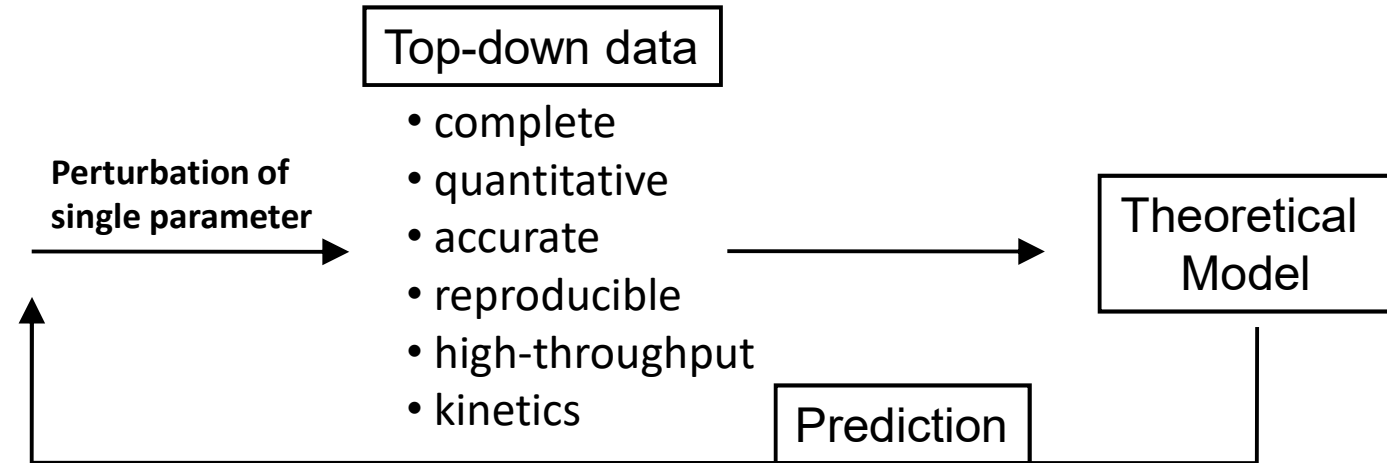
Experimental datasets:
transcriptomics,
translatomics,
proteomics,
metabolomics,
physiological and biochemical parameters
(gas exchange, chlorophyll fluorescence,
enzyme activities, ...)

Generation of models of biological systems from systematic measurements, not necessarily "omics" data.

Overview of our scientific and experimental approaches

Experimental design

Spatial-temporal profiling

"Transcriptome"   "Regulome"   Proteome   Metabolome

| qRT-PCR/ | FAIRE-seq | LC-MS | GC-MS |

High throughput methods for relative quantification

Analysis methods

- Statistical methods
- Functional annotation
- Relative quantification
- Network analysis

Functional signatures
- Genes/proteins
- Epigenetic modifiers
- Regulatory motifs
- Metabolites

Biological networks
- Functional modules
- Pathways
- Axis of regulation
- Bootlenecks

Applications

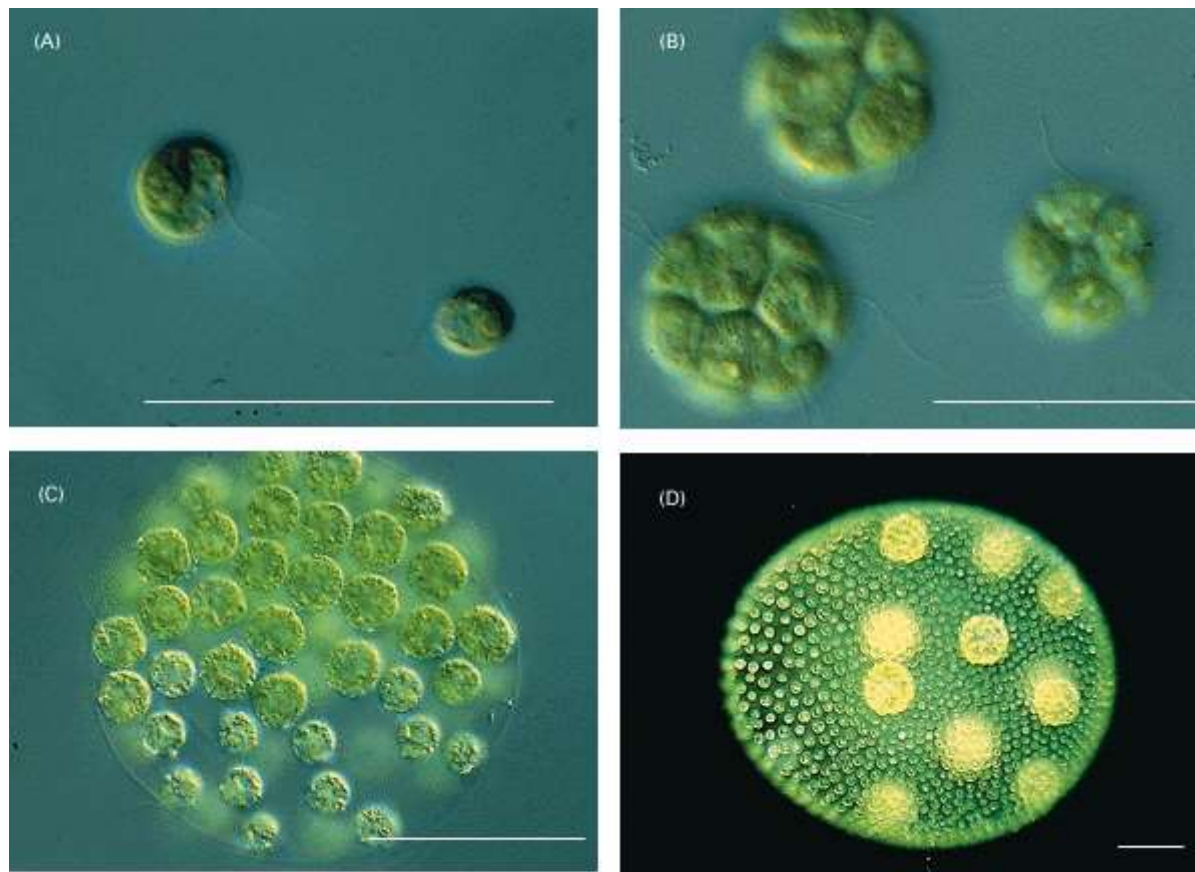Candidates Discovery   Response Classification   Interactions   Synthesizer Systems   Cellular mechanisms

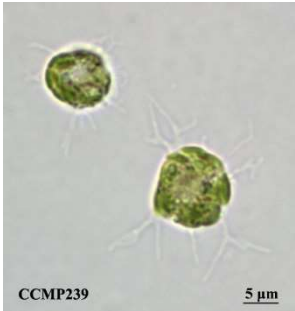# Microalgae can contribute to sustainable applications
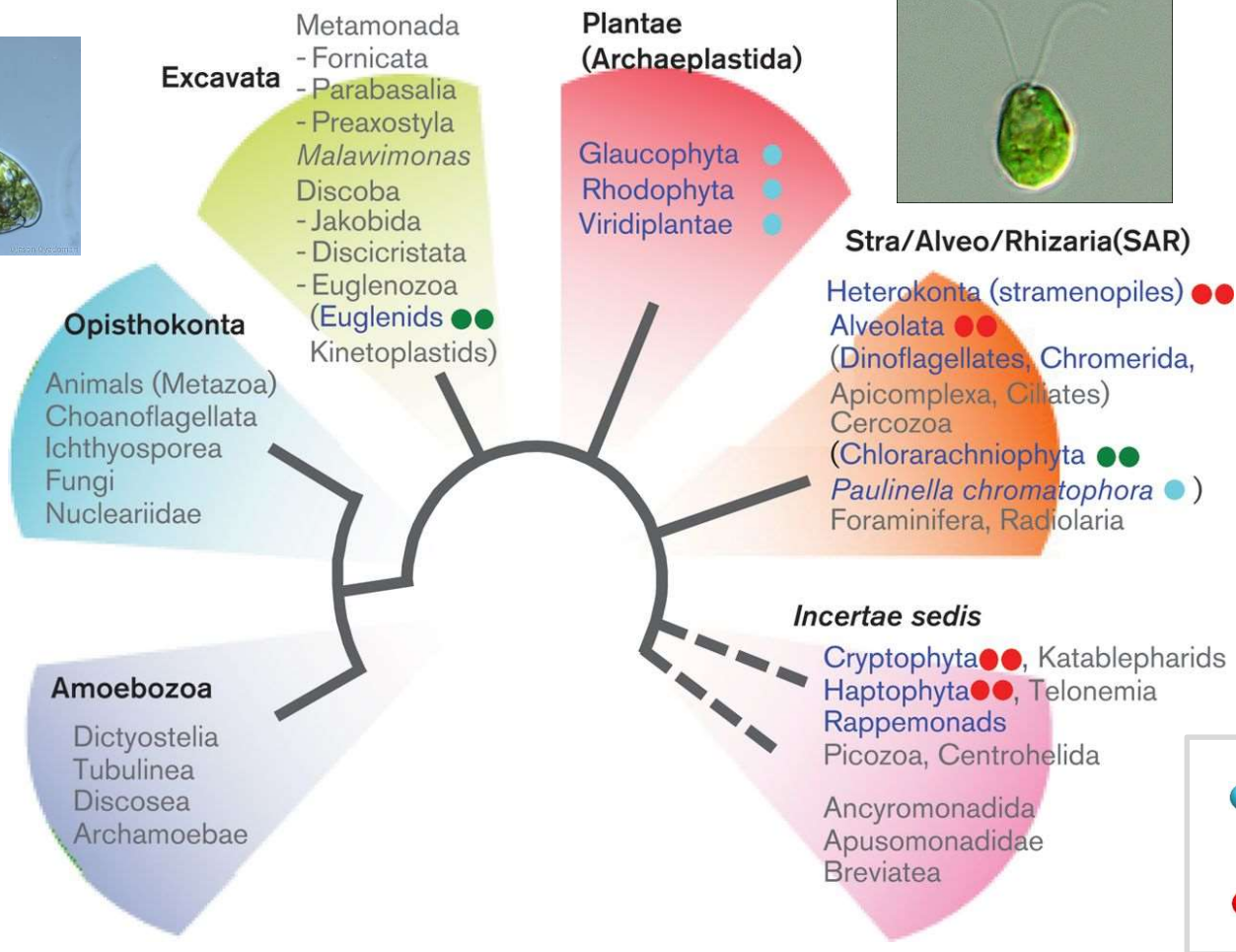

Winck et al, 2013. *J.Proteomics*
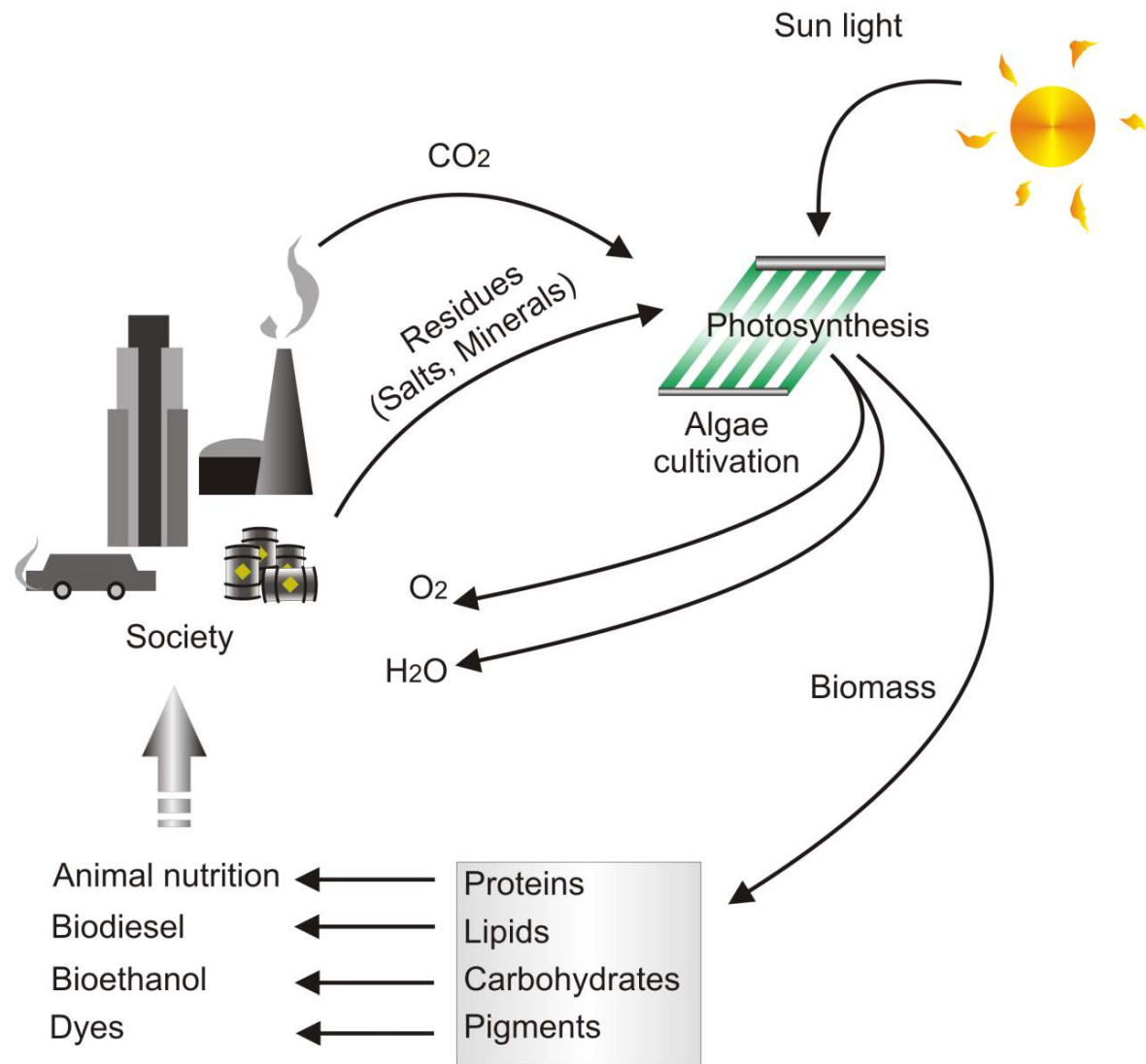
Microalgae diversity: uni and multicellularity

the scale bar shown represents 50 µm in each case

Distribution on supergroups

The five eukaryotic supergroups that have been identified using molecular data.

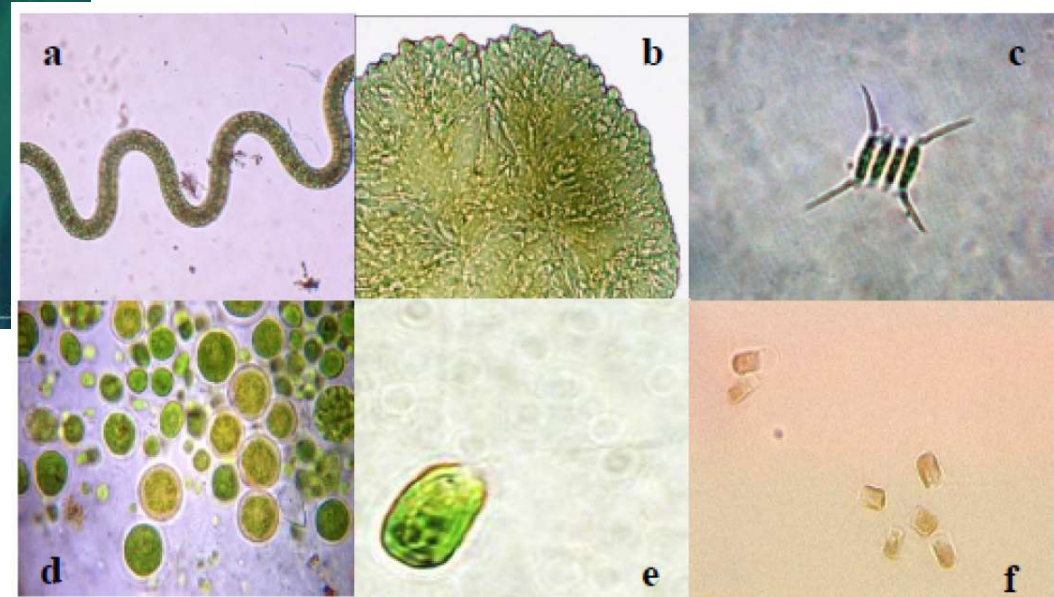# Microalgae in the oceans, rivers and soil

Figure 1: Few commercially important microalgal strains. (a) *Arthrospira maxima* (b) *Botryococcus brauanii* (c) *Scenedesmus quadricauda* (d) *Chlorella vulgaris* (e) *Dunaliella salina* (f) *Chaetoceros muelleri.*

Raja et al., Oceanography 2014, 2:1

http://earthobservatory.nasa.gov/IOTD/view.php?id=87083&src=ve

20

NASA Earth Observatory image by Joshua Stevens, using Landsat data from the U.S. Geological Survey. Caption by Adam Voiland.

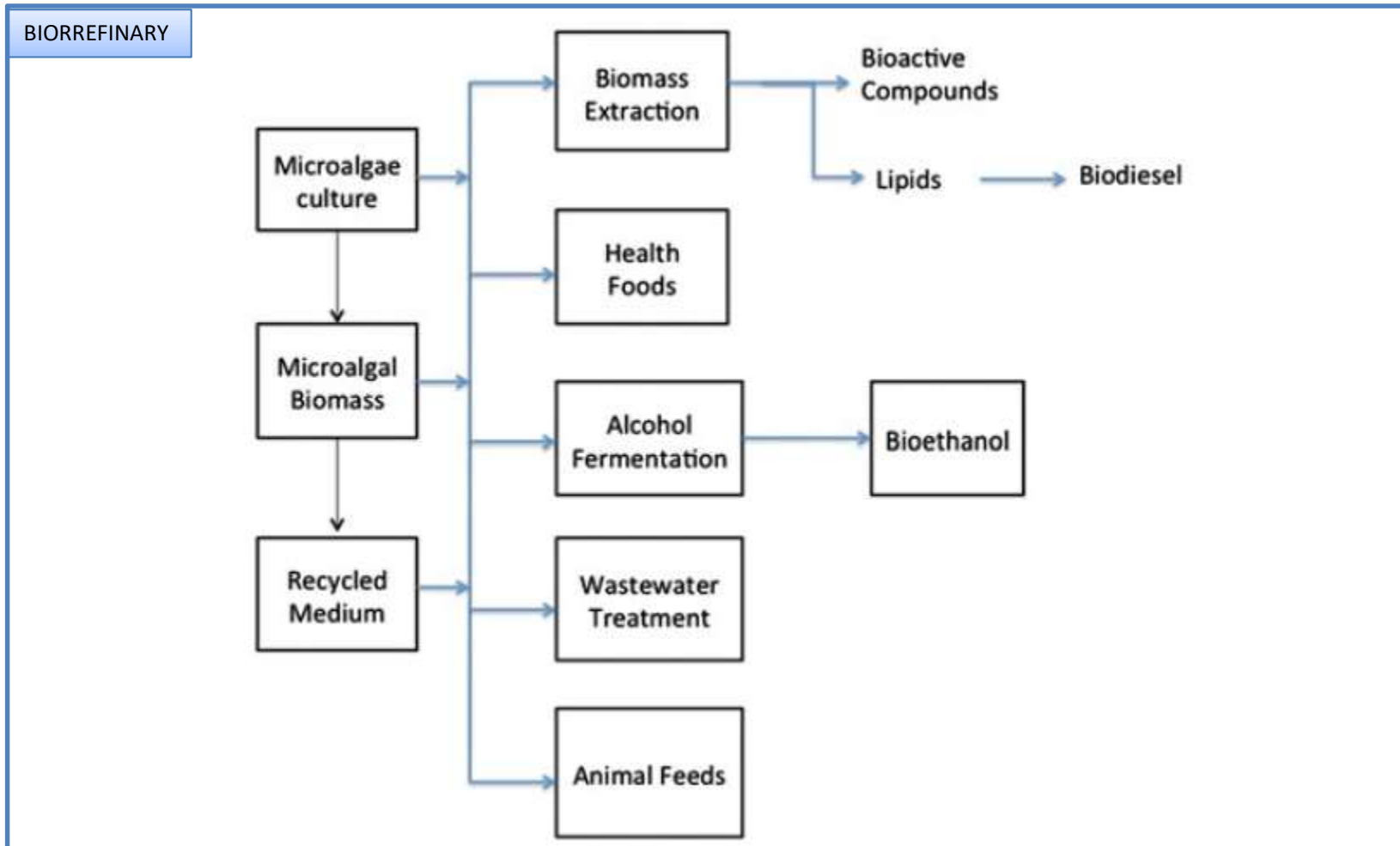# Microalgae biomass production in scale

Figure 2: (a) Culture scale-up for open raceway ponds (b) A typical lab-scale photobioreactor (c) French press (side view) (d) Open raceway pond (e) Flocculated culture shows algal clumping (f) CO_2 cylinder (g) Culture storage tank (h) French press (front view) (i) Wet algal biomass collected from French press (j) Hot air dryer (Courtesy: Aquatic Energy LLC, Louisiana).

Raja et al., Oceanography 2014, 2:1

Microalgae cultivation

Bioreactors
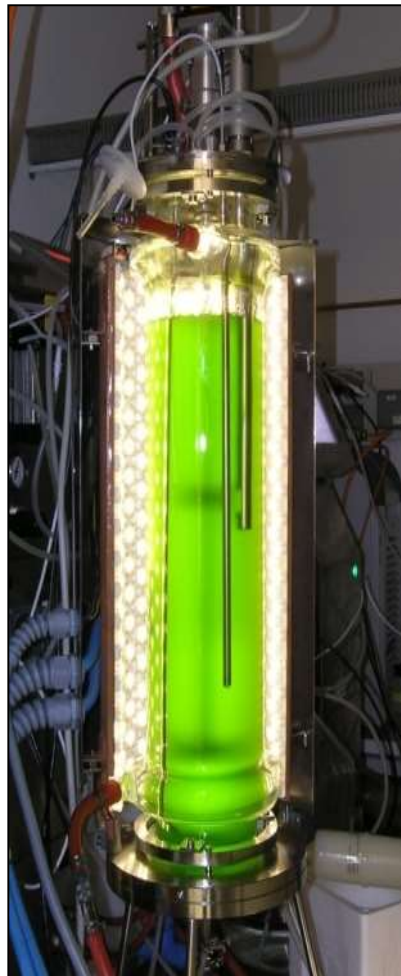
Microalgae biomass in a sustainable scenario

Strong development is needed in the field of microalgae biomass production in order to make it economically competitive. Further research is needed.

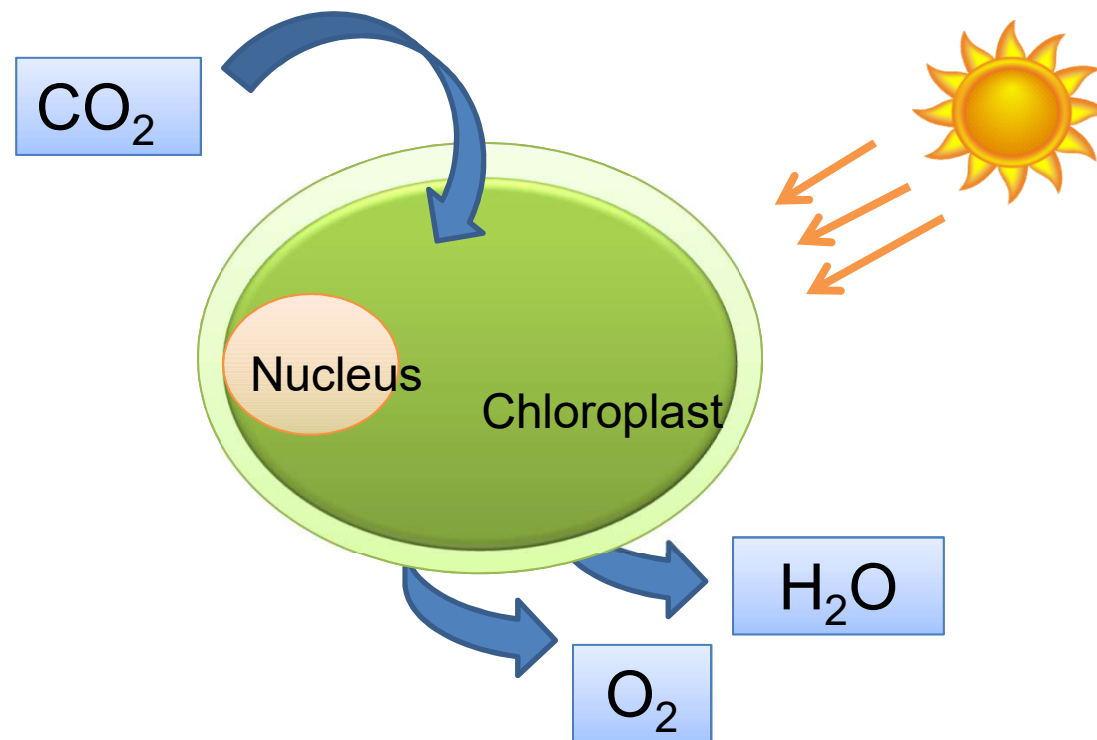# Microalgae adjust and optimize carbon accumulation for proper biomass production and growth
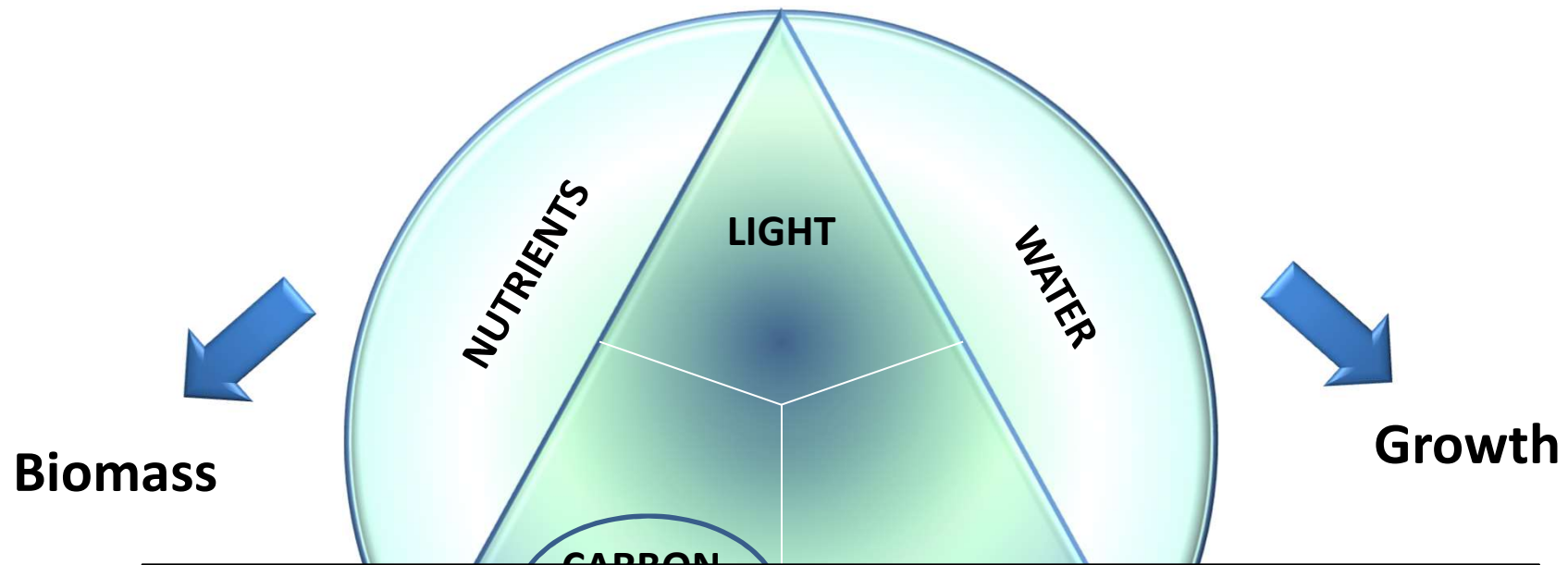
Photosynthesis ➡️ Biomass ➡️ Growth

CO_2

Nucleus

Chloroplast

H_2O

O_2

Environmental factors affect biomass production

LIGHT

NUTRIENTS

WATER

CARBON

Biomass

Growth

How cells control the biomass production and biomass composition ?

http://www.icis.com

25

Complex mechanisms control biomass accumulation and composition

Activation of Carbon Concentrating Mechanism

Cah1

$HCO_3^- \rightleftharpoons CO_2$ — Cah8

Periplasmic space

$HCO_3^- \quad Cah9 \quad CO_2$

Cytoplasm

$HCO_3^- \quad Cah6 \quad CO_2$

Chloroplast stroma

HCO_3^-

H^+ Ca

Pyrenoid

Cell wall | Plasma membrane | Chloropl... | Thylakoid membrane
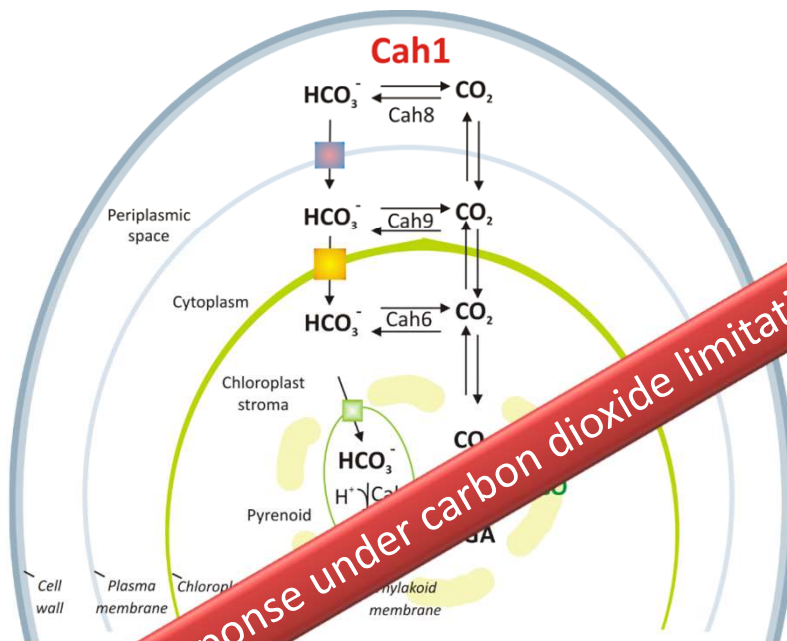
Cellular response under carbon dioxide limitation

Limited uptake on high carbon dioxide
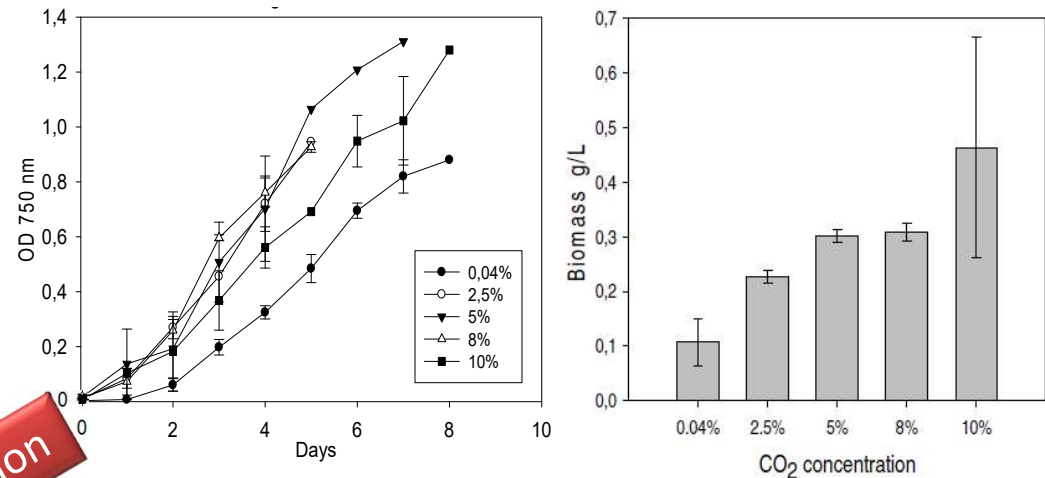
Páez Melo, et al. (2014) *Advances in Computational Biology*

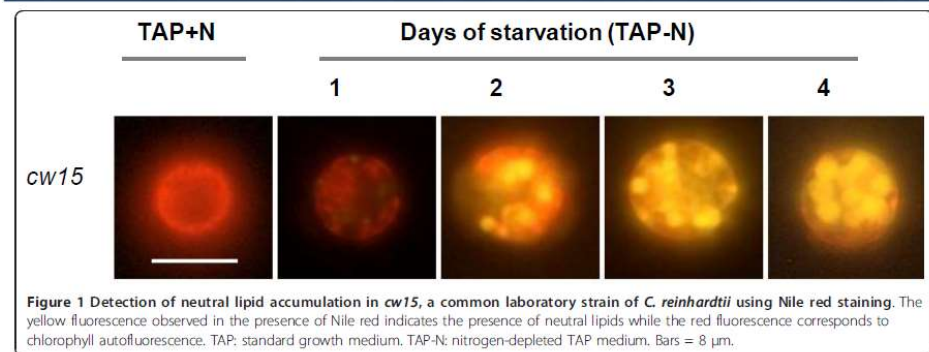**Nutrient stress-induced cellular responses**

Figure 1 Detection of neutral lipid accumulation in *cw15*, a common laboratory strain of *C. reinhardtii* using Nile red staining. The yellow fluorescence observed in the presence of Nile red indicates the presence of neutral lipids while the red fluorescence corresponds to chlorophyll autofluorescence. TAP: standard growth medium. TAP-N: nitrogen-depleted TAP medium. Bars = 8 µm.
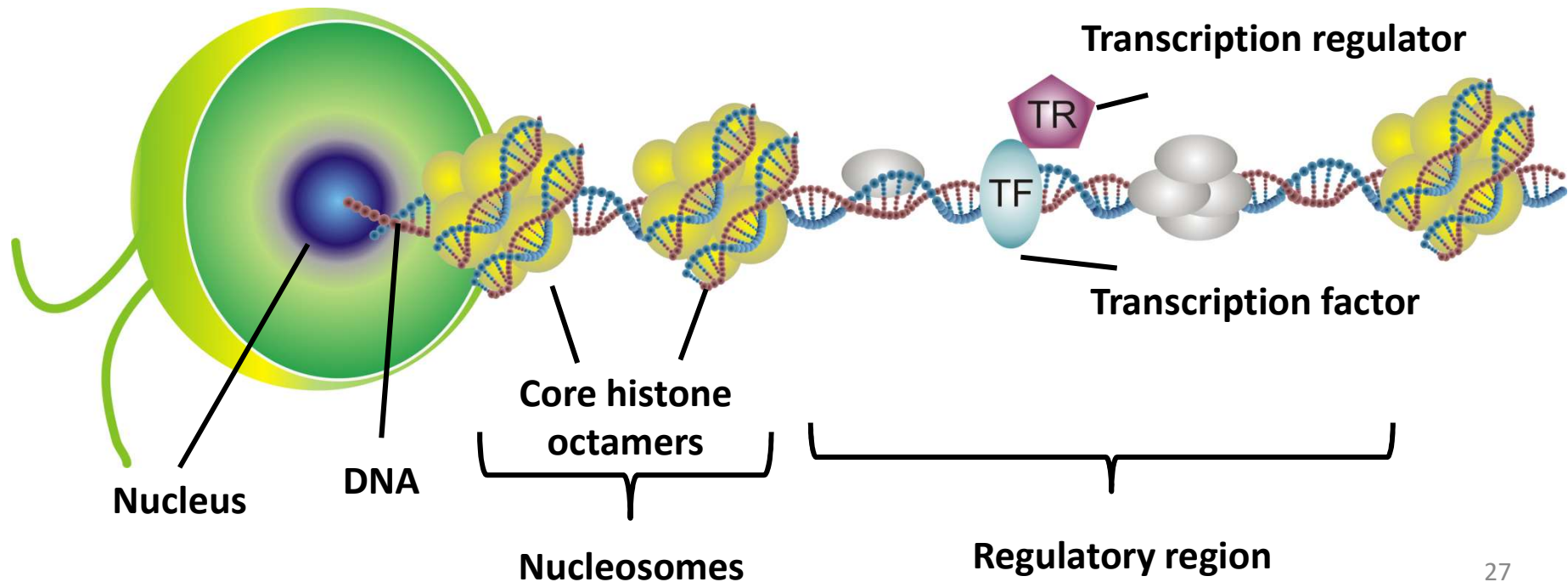
Winck, F.V., et al (2013). *J Proteomics*, **94C**, 207-218.

Siaut et al., 2011 *BMC Biotechnology.*

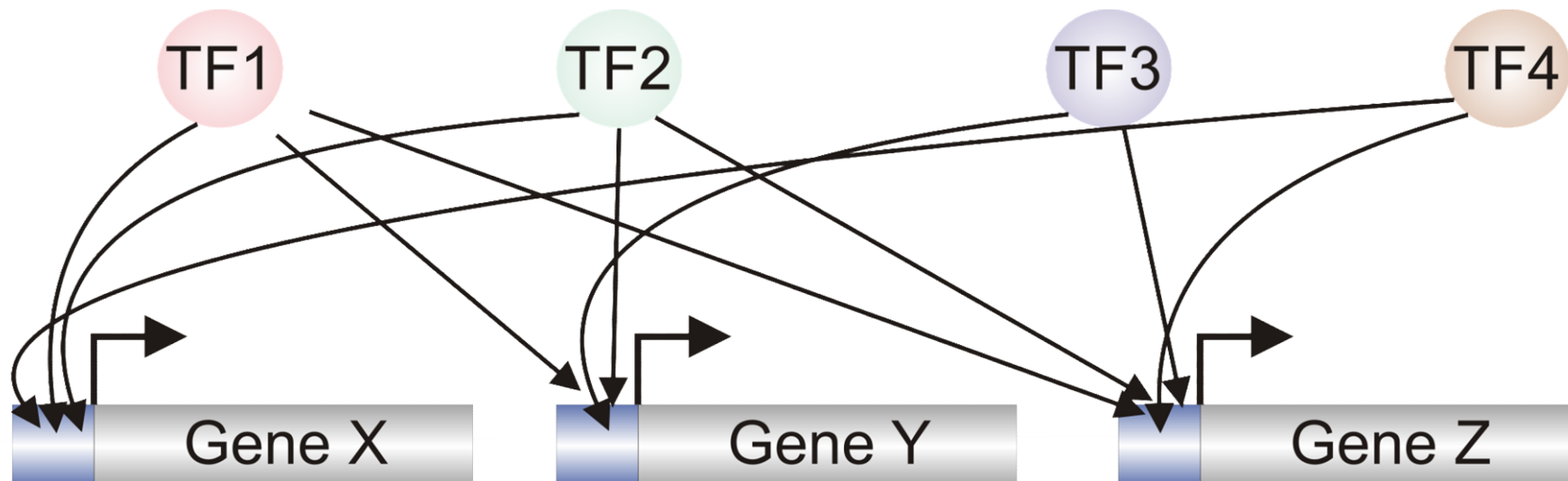# Transcriptional control influences cellular responses

Exogenous and endogenous stimuli (e.g. CO_2, light, temperature, hormones)

Regulatory mechanisms (epigenetic and genetic)

Transcription regulator

TR

TF

Transcription factor

Nucleus

DNA

Core histone octamers
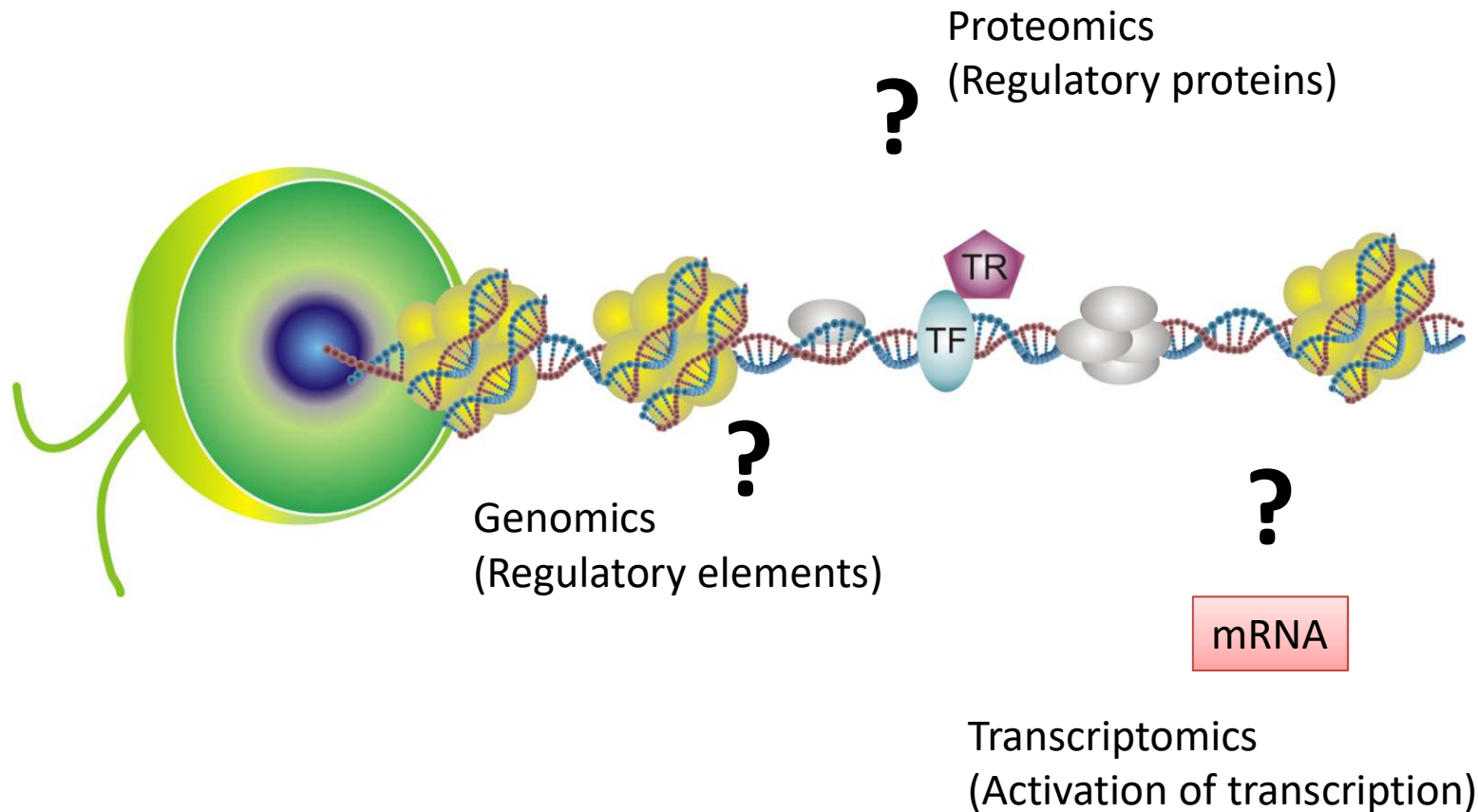
Nucleosomes

Regulatory region

# Regulatory proteins modulate gene transcription by multi-combinatorial mechanisms

TF = Transcription factor

The combination of different layers of information can help us to understand the transcriptional control of biomass accumulation
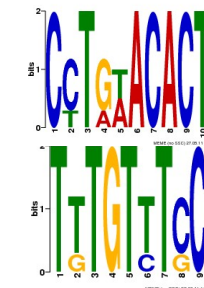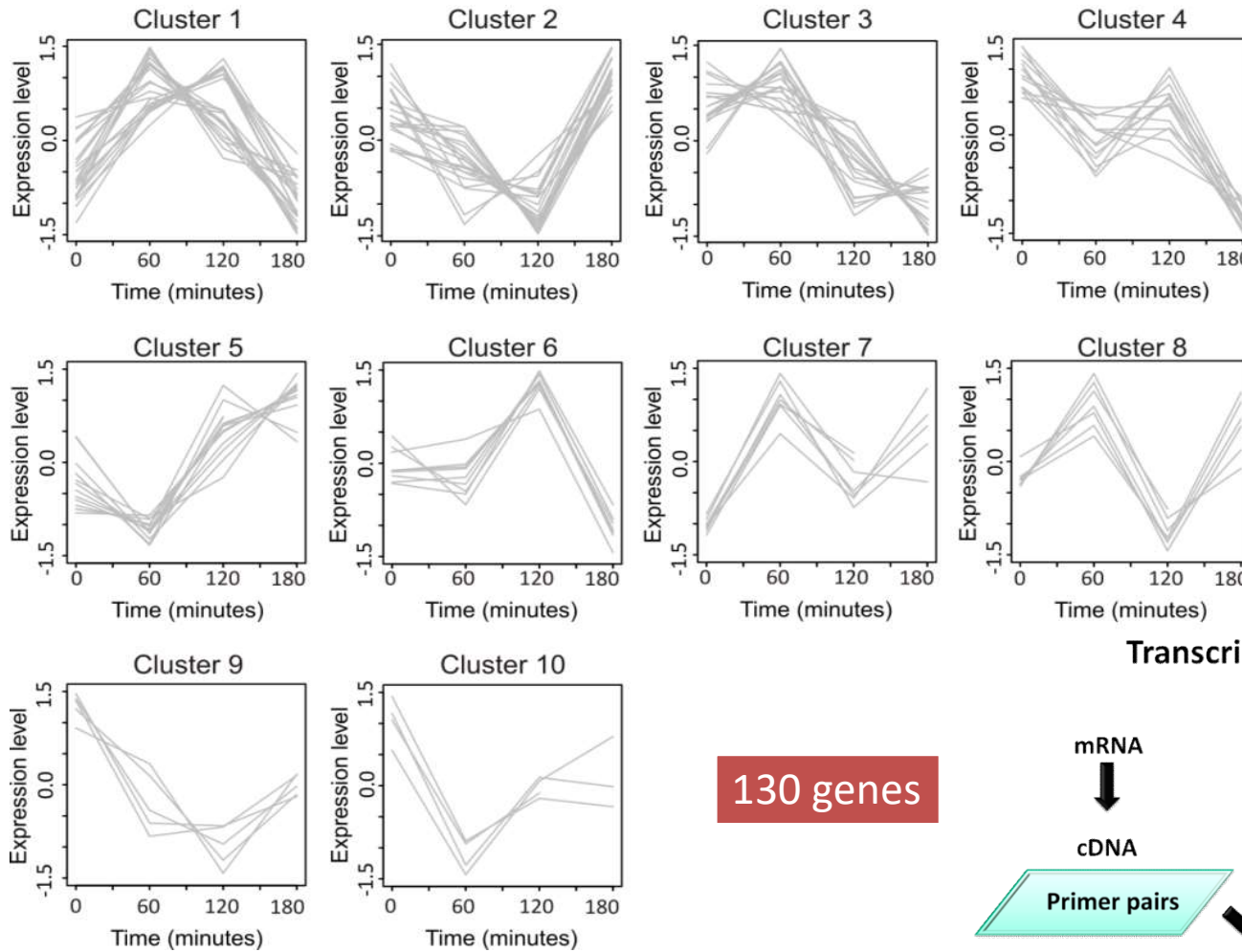
Proteomics
(Regulatory proteins)

?

?

Genomics
(Regulatory elements)

?

TR

TF

mRNA

Transcriptomics
(Activation of transcription)

Experimental analysis of microalgae cellular responses under carbon dioxide limitation
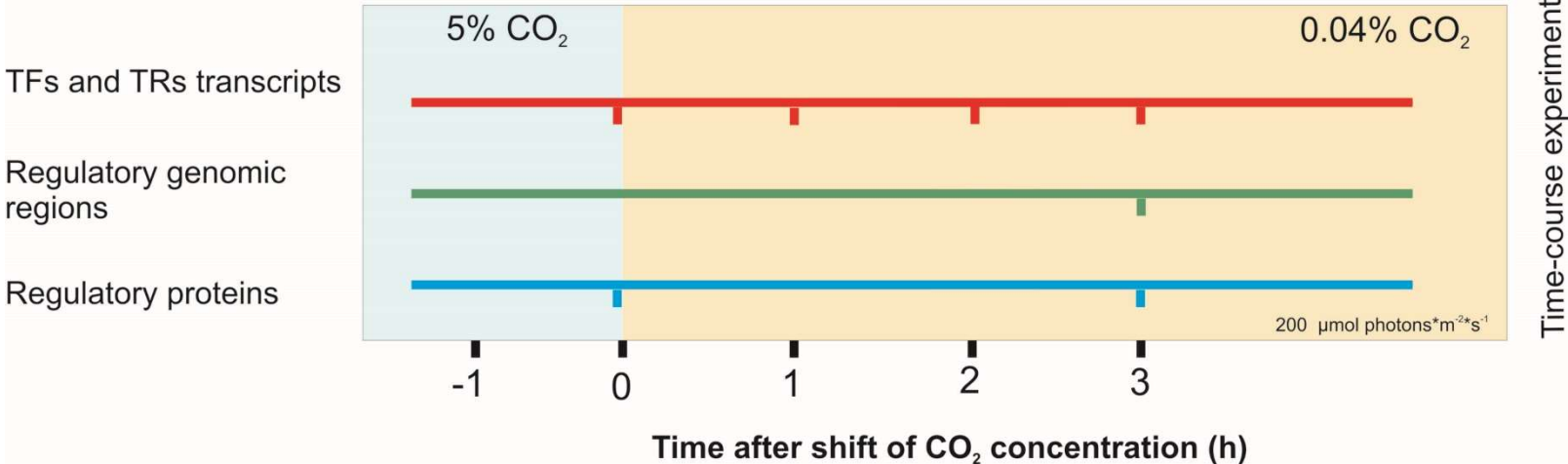
Chlamydomonas reinhardtii

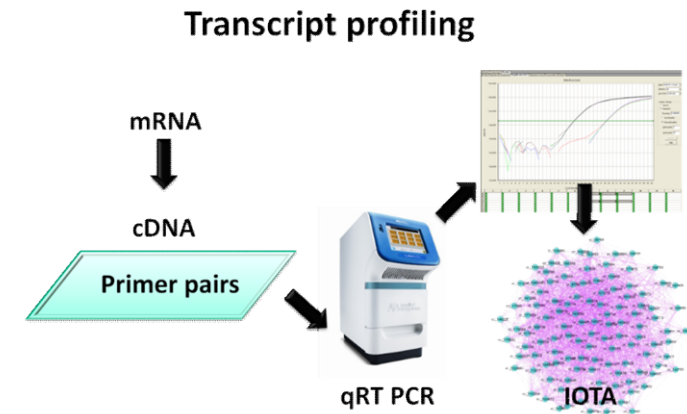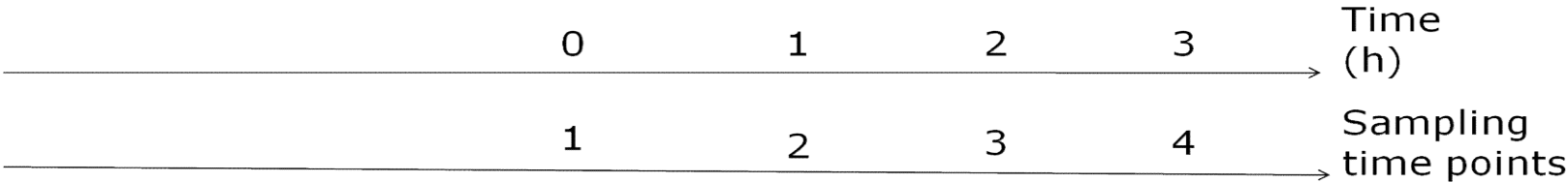# Expression profiles of transcription factors were identified

$p < 10^{-5}$

130 genes

Transcript profiling

Quality threshold (QT) clustering method

Vischi Winck F, Arvidsson S, Riaño-Pachón DM, Hempel S, et al. (2013) Genome-Wide Identification of Regulatory Elements and Reconstruction of Gene Regulatory Networks of the Green Alga Chlamydomonas reinhardtii under Carbon Deprivation. PLoS ONE 8(11): e79909.

Time-series analysis of cellular responses
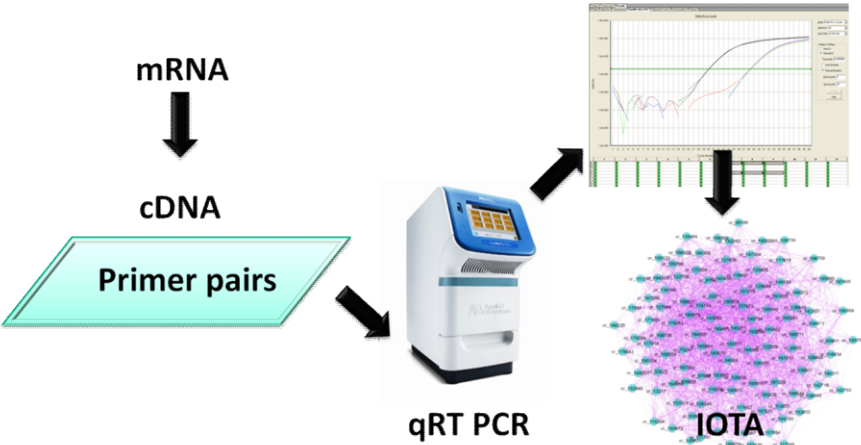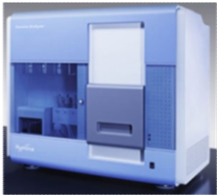
Transcript profiling: the abundance of gene transcripts in one specific moment

https://en.wikipedia.org/wiki/Reverse_transcription_polymerase_chain_reaction   •CC BY-SA 3.0
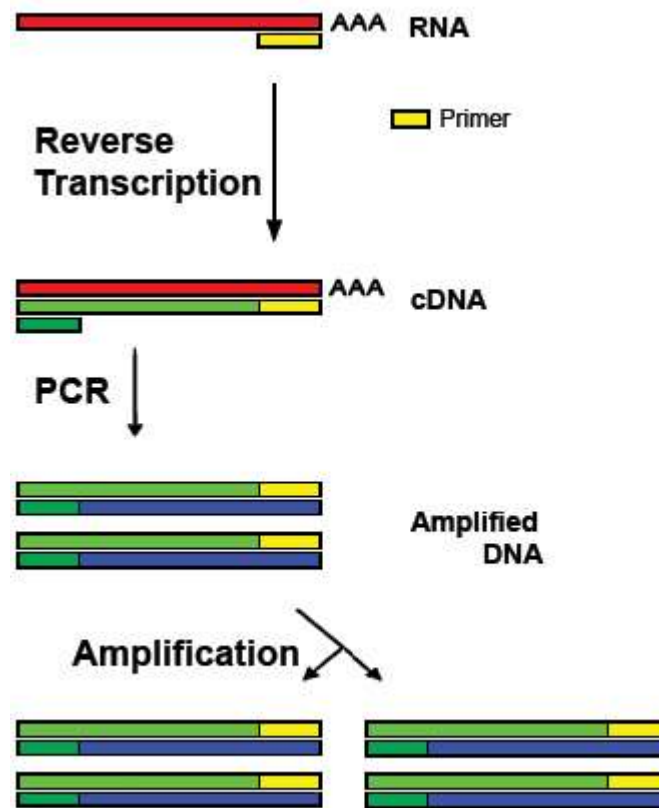
Transcript profiling revealed transcripts responsive to reduced CO_2 concentration

Gene A

Transcript of gene A

Gene B

Transcript of gene B
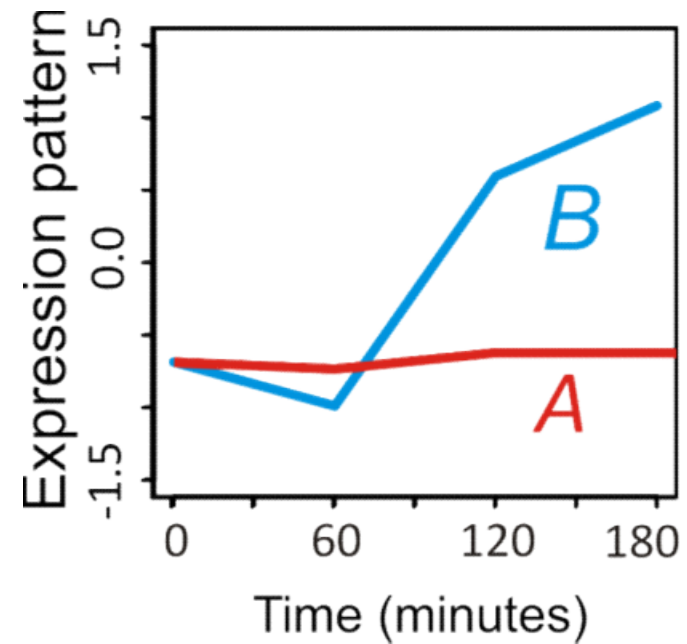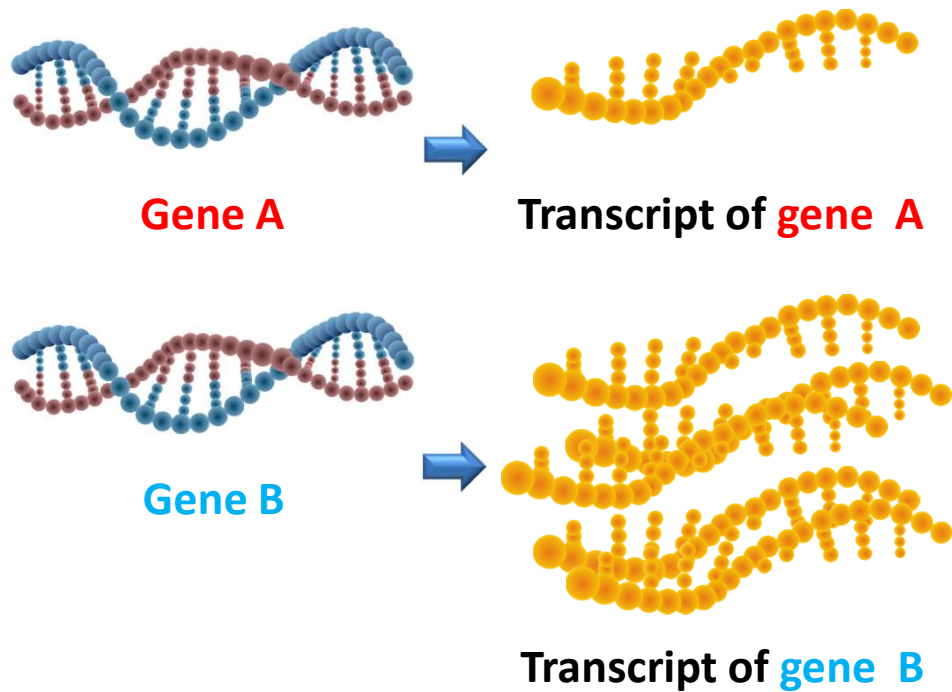
Directed gene regulatory network was inferred based on experimental data and mathematical modeling
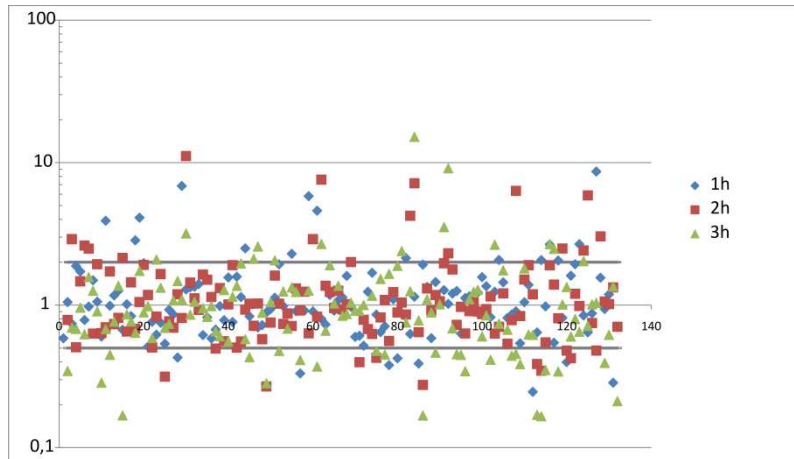
Expression data
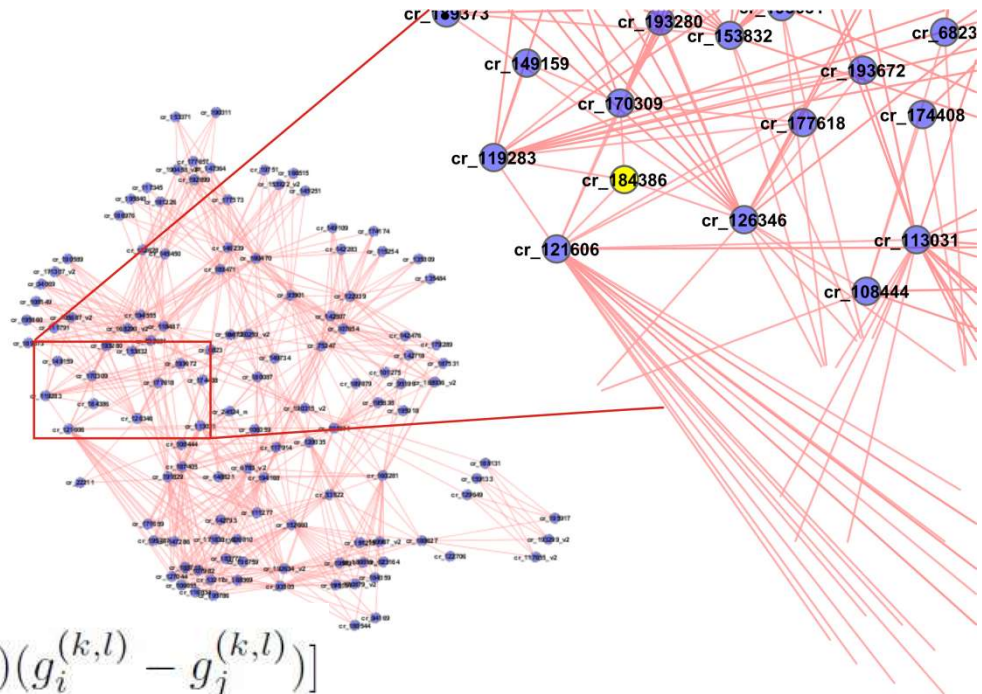
$$\iota^{(l\to k)} = 1 - \frac{\sum_{i=1}^{n-2}\sum_{j=i+1}^{n-1} w_{ij}\Theta[(g_{j+1}^{(k,l)} - g_i^{(k,l)})(g_i^{(k,l)} - g_j^{(k,l)})]}{\Delta}$$

Matemathical modeling
(Inner Composition Alignment- IOTA* method)

* Hempel S, Koseska A, Kurths J, Nikoloski Z. Inner composition alignment for inferring directed networks from short time series. Phys Rev Lett. 2011 Jul 29;107(5):054101.

Permutation-based reconstruction of gene regulatory networks

$$\iota^{(l \to k)} = 1 - \frac{\displaystyle\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} w_{ij} \Theta\left[(g_{j+1}^{(k,l)} - g_i^{(k,l)})(g_i^{(k,l)} - g_j^{(k,l)})\right]}{\Delta}$$
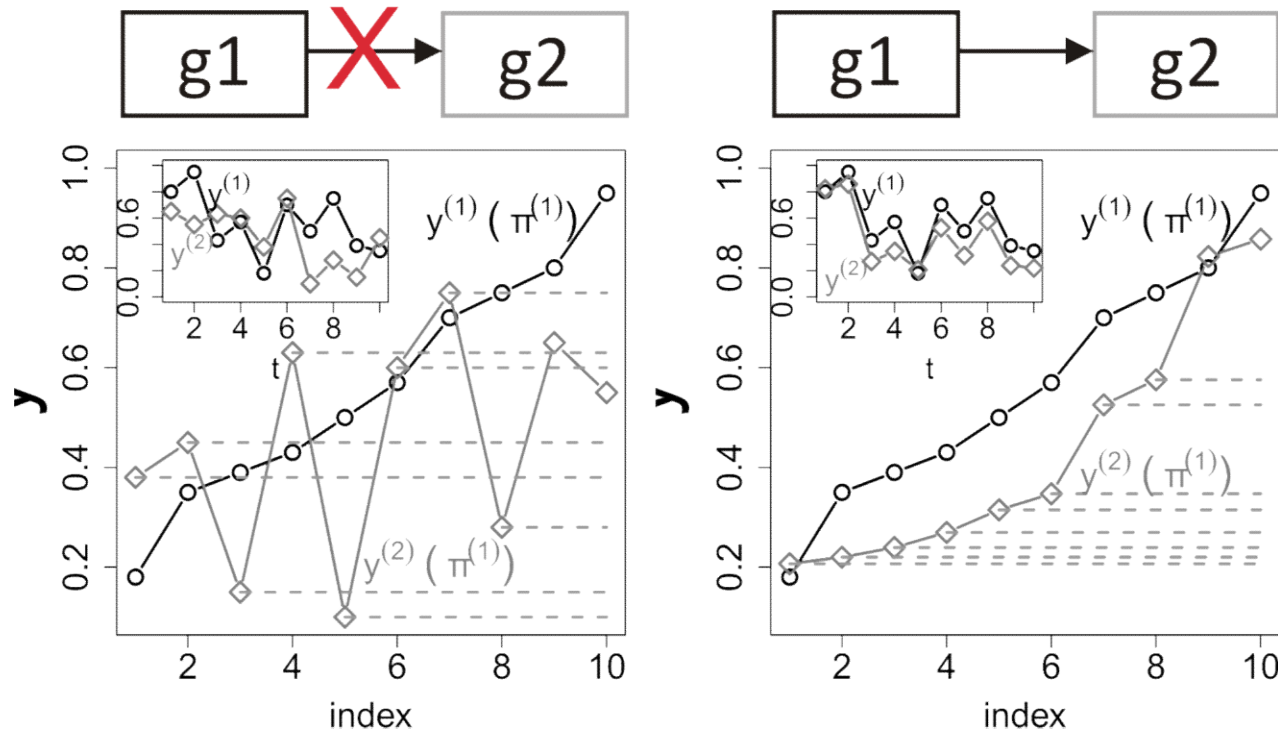
$$\Delta = \frac{(n-1)(n-2)}{2}$$

Networks in biology

Complex biological systems may be represented and analyzed as computable networks

Examples of networks

➢ Protein–protein interaction networks
➢ Gene regulatory networks (molecular regulators networks)
➢ Gene co-expression networks (transcript–transcript association networks)
➢ Metabolic networks (reactions and enzymes networks)
➢ Signaling networks (usually integrate protein–protein interaction networks, gene regulatory networks, and metabolic networks)
➢ Neuronal networks
➢ Between-species interaction networks
➢ Within-species interaction networks

Systems biology and biological networks

Biological networks are one of the many forms of modeling and representing relationships between biological components, which are dynamical units of the system (e.g., genes, proteins, metabolites, etc.).

Edges or lines
(links between units)
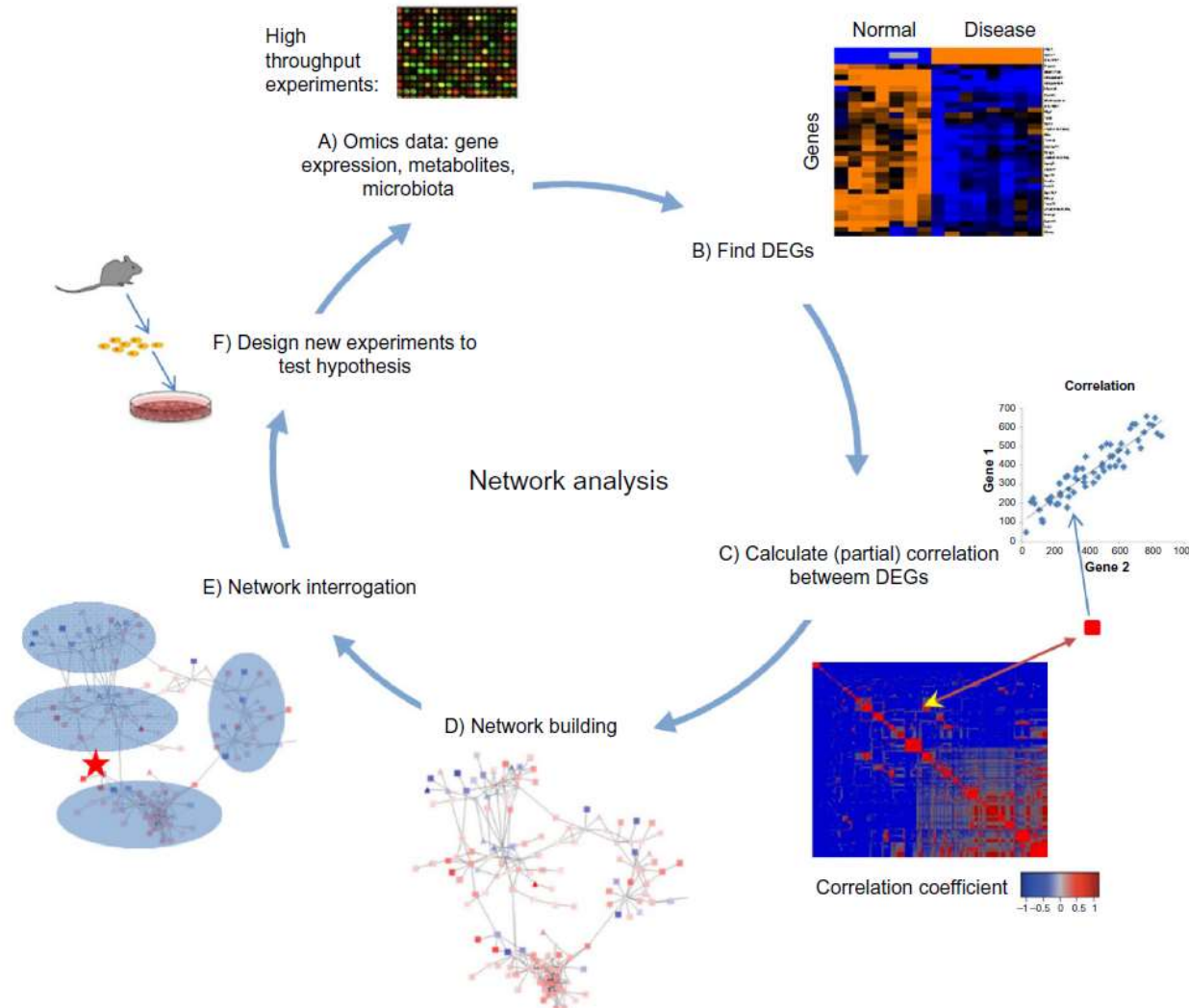
Nodes or vertices
(dynamical units)
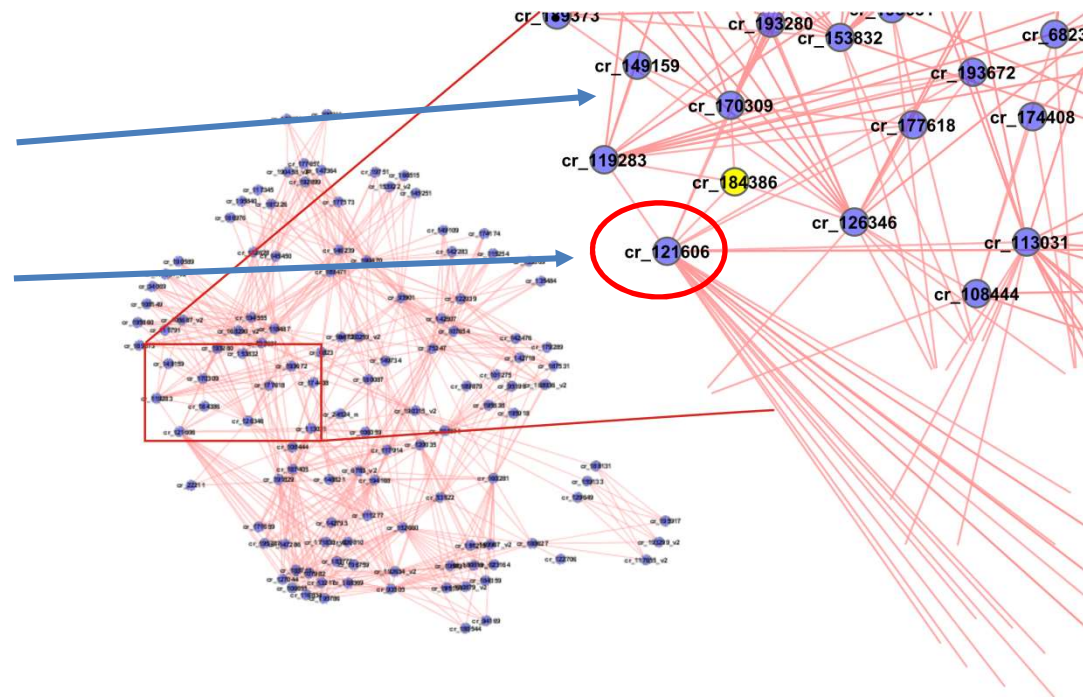
Networks used for visualization and modeling purposes

Systems biology and the prediction of biological networks

Dong et al. Reverse enGENEering of Regulatory Networks from Big Data: ARoadmap for Biologists. *Bioinformatics and Biology Insights* 2015:9 61–74 doi: 10.4137/BBI.S12467

Expected and unexpected correlations in biological states

Correlation of expression values
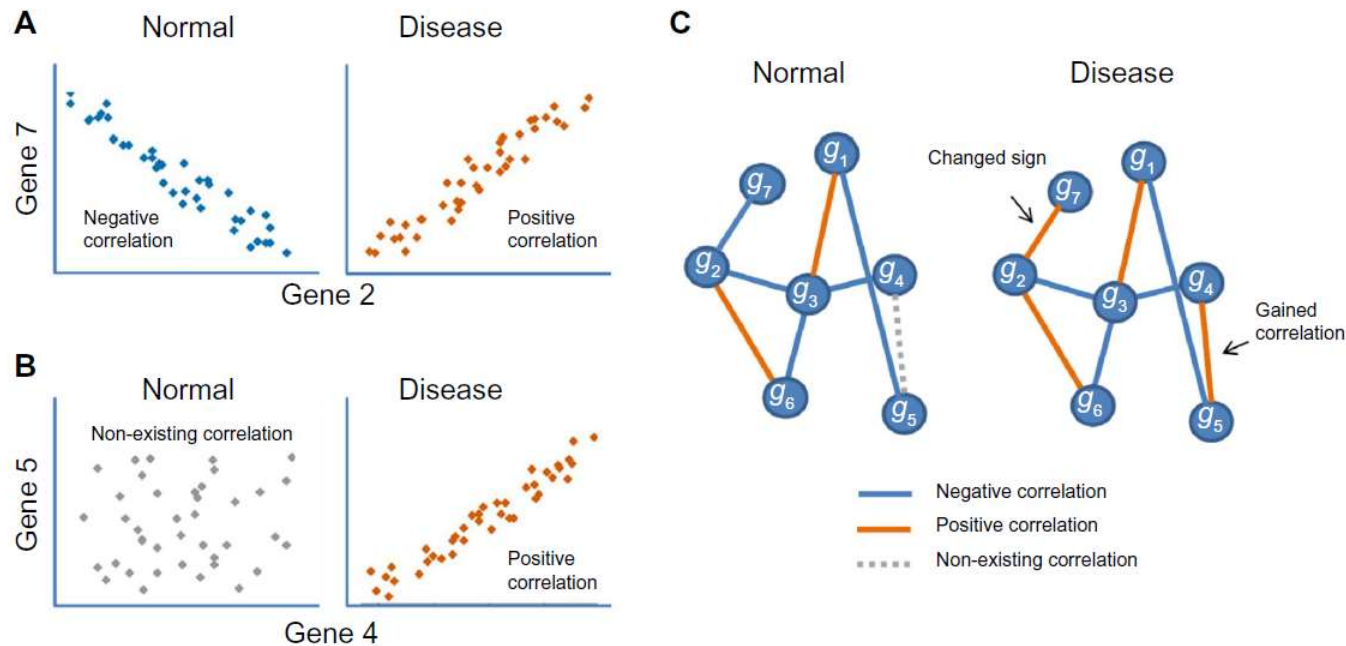
Figure 4. (**A**) Gene 2 and gene 7 correlate with each other in both normal and disease conditions, but the signs of the correlation coefficient are opposite. (**B**) In normal condition, there is no correlation between gene 4 and gene 5, but they gain positive correlation when the biological system transitioned to disease. (**C**) Example of visualization of a network transitioning between normal and disease conditions. Red lines represent positive correlation, blue line represent negative correlation, and dotted gray lines represent nonexisting correlations in one condition that strongly appear in the other condition (on this case, becomes positively correlated).

Correlation to causation in reconstructing biological networks

42
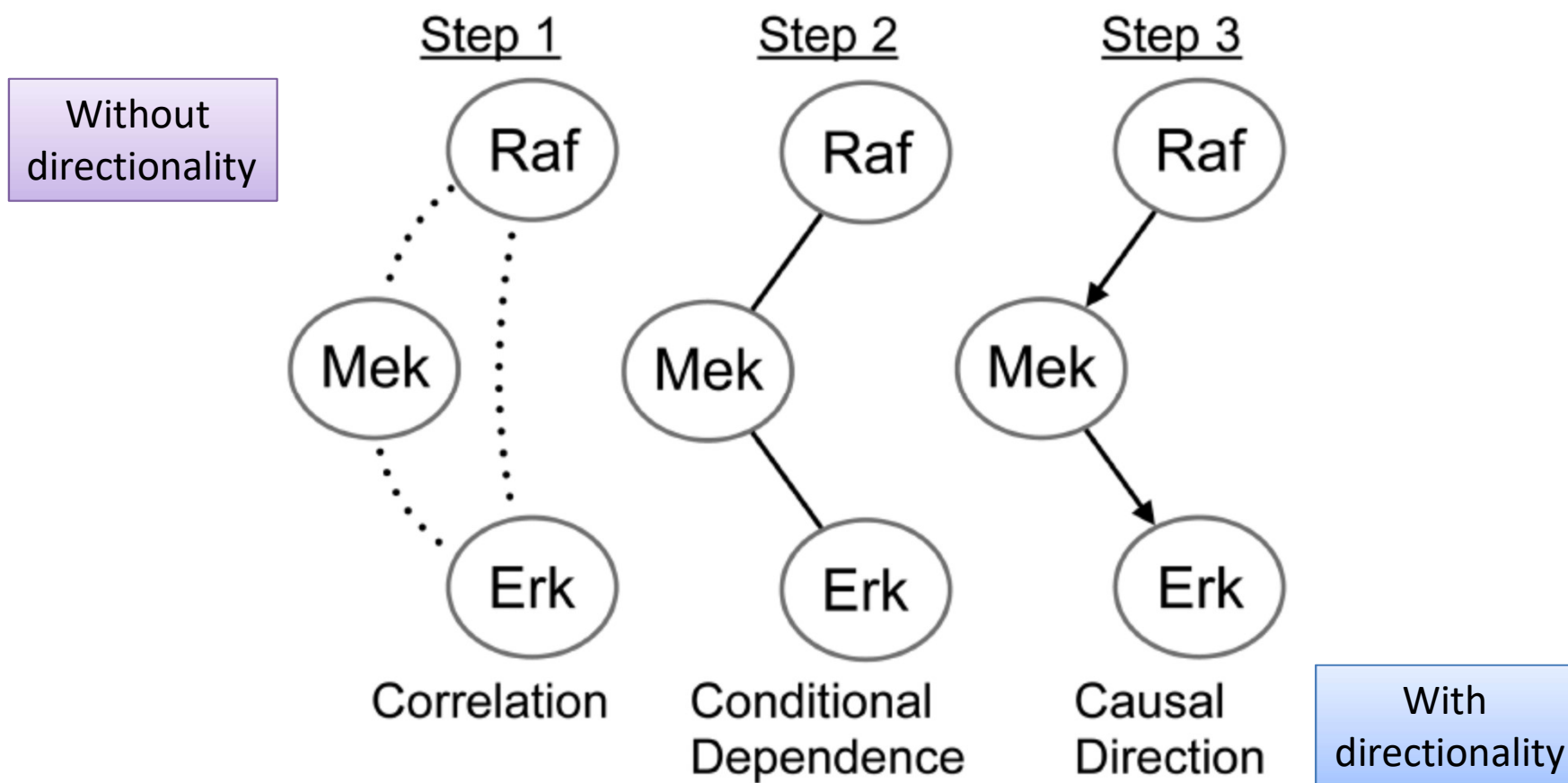
Directionality in biological networks

RANDOM DIRECTIONALITY

MAXIMAL DIRECTIONALITY
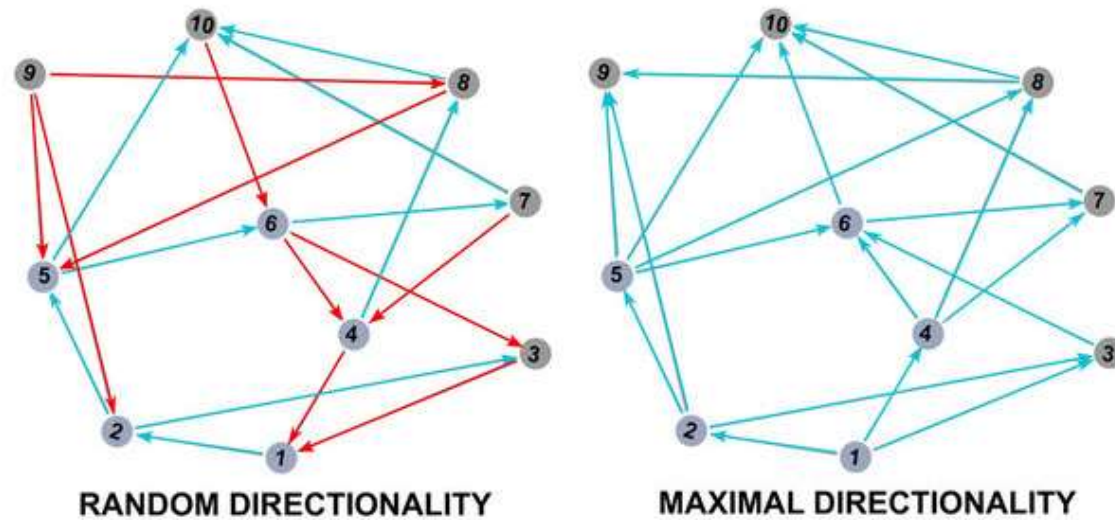
Fidelity and dynamics (Time-series)

Gene promoter motifs were predicted for expressed genes

DNA sequence of the promoter regions of TFs showing alteration on transcript level

⬇

MEME- Motif finder

(http://meme.sdsc.edu/meme4_3_0/intro.html)

⬇

MAST-Specificity analysis of overrepresented motifs against the whole set of promoter sequences (14,598)
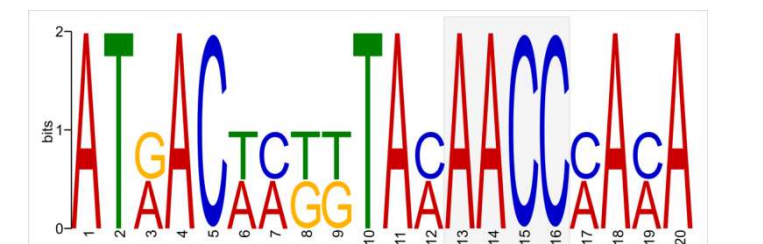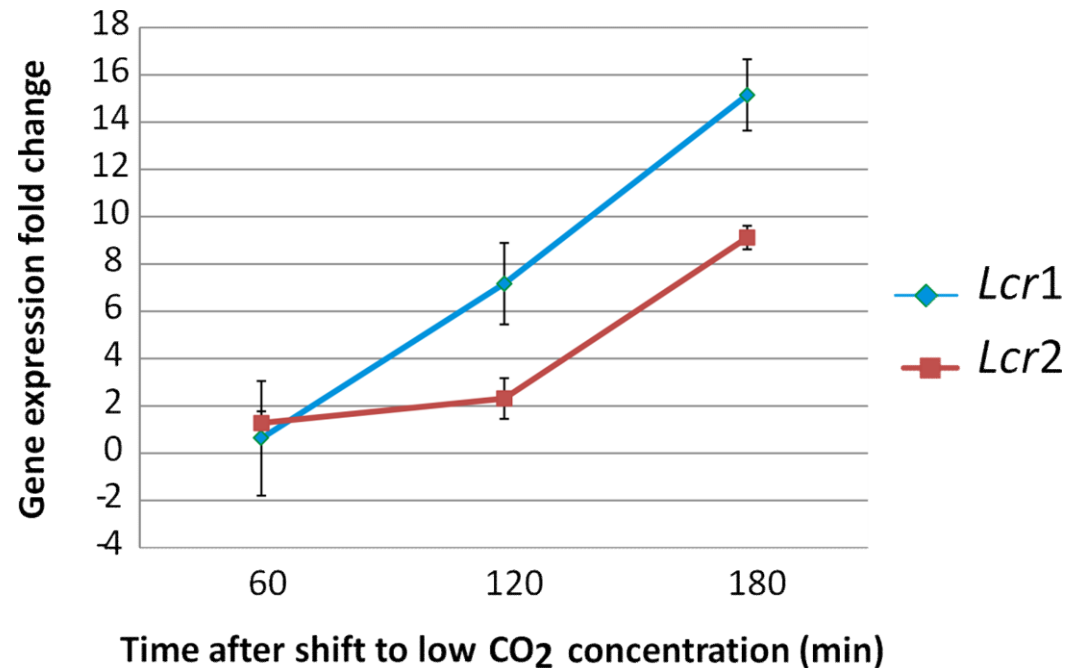
Protein ID	Motif p-value	MAST p-value
188627	9.05E-12	2.3E-10
184359	1.77E-11	3.3E-10

Genes regulated by a common set of TFs may have similar regulatory regions in their promoters

Down-regulated genes: SBP, C3H, FHA, Orphan
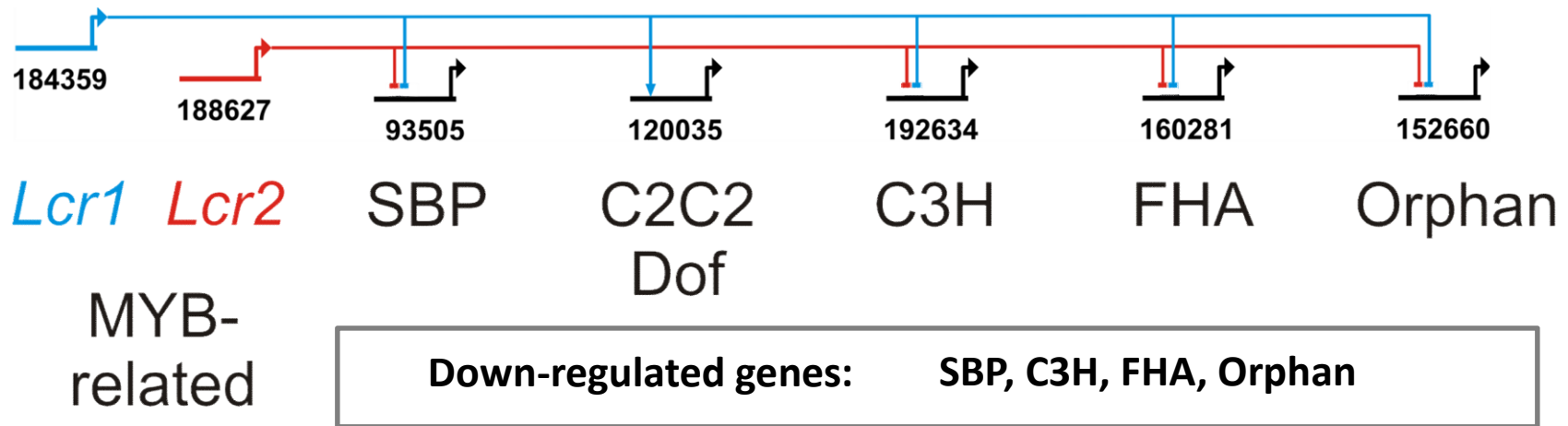
Co-occurred motifs in co-regulated genes
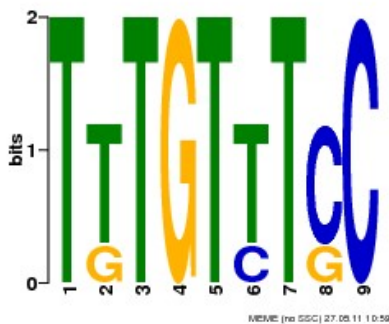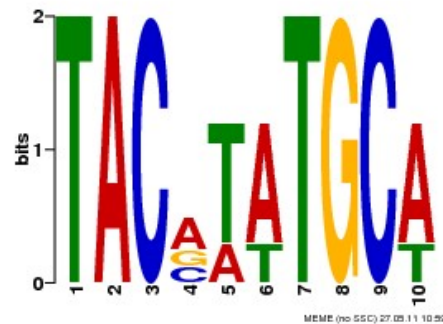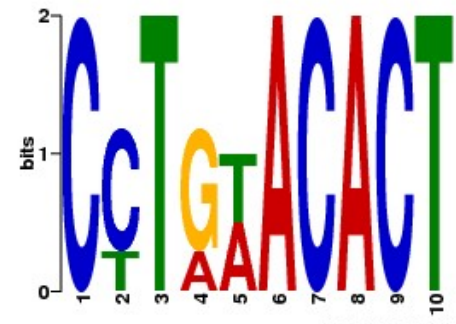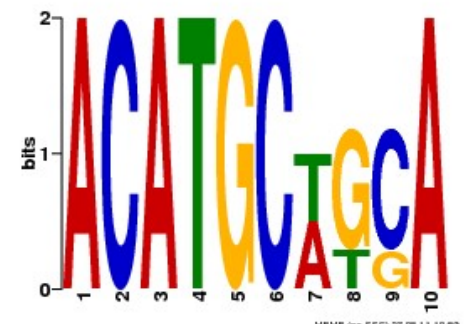
$p < 10^{-5}$ $p < 10^{-5}$ $p < 10^{-5}$ $p < 10^{-5}$

Early-responsive genes may have a role in chromatin remodeling
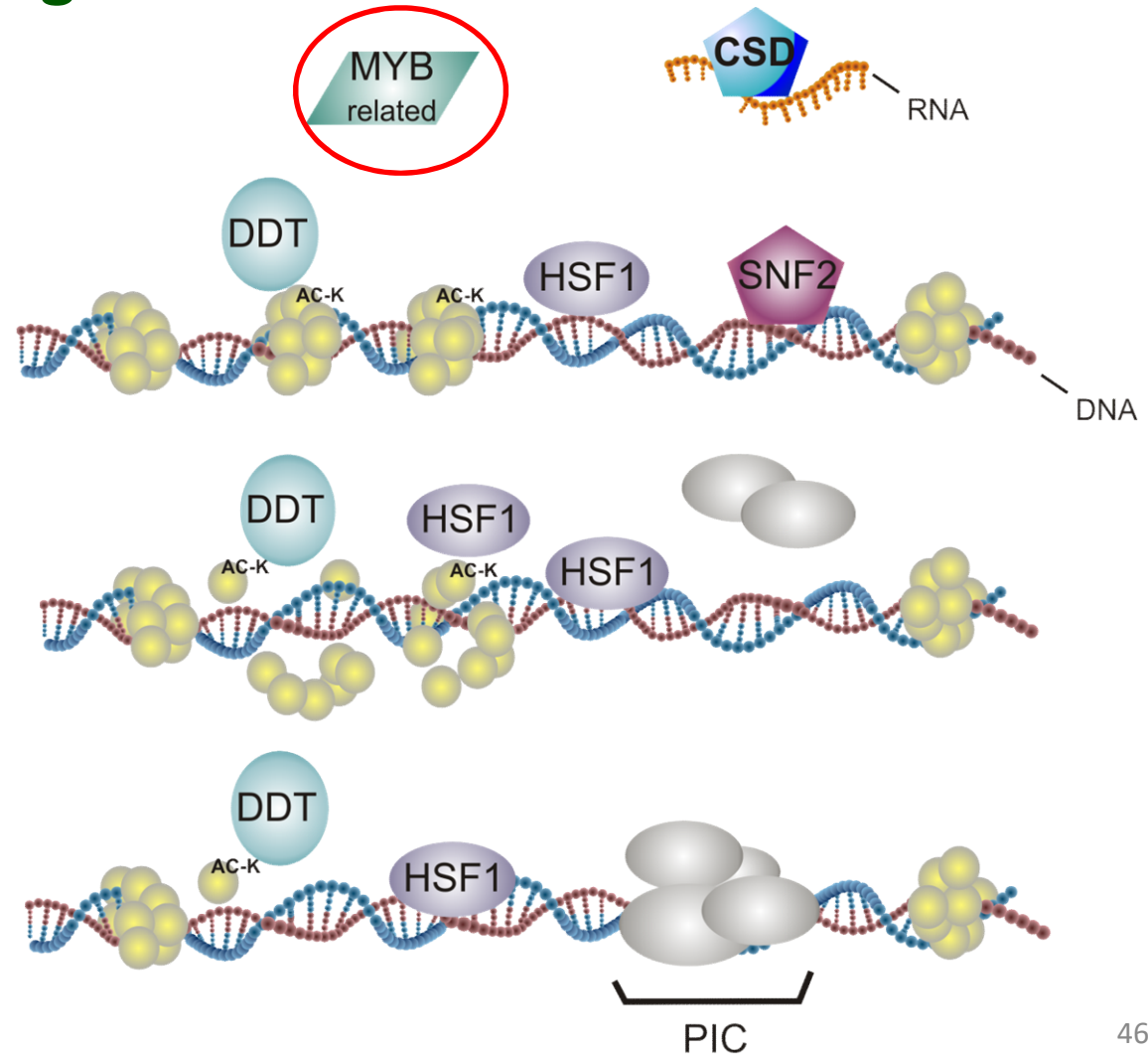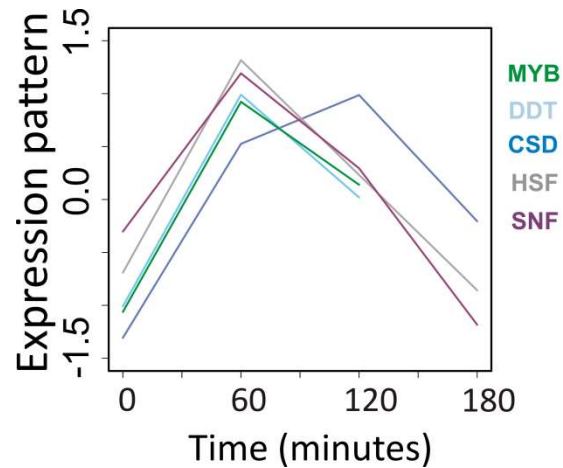
MYB = Myeloblastosis viral oncogene
DDT = DNA different transcription
CSD = Cold shock domain
HSF1 = Heat shock factor 1
SNF2 = Sucrose nonfermenting 2

PIC = Pre-initiation complex

Late responsive TF gene products may modulate important CCM-related genes
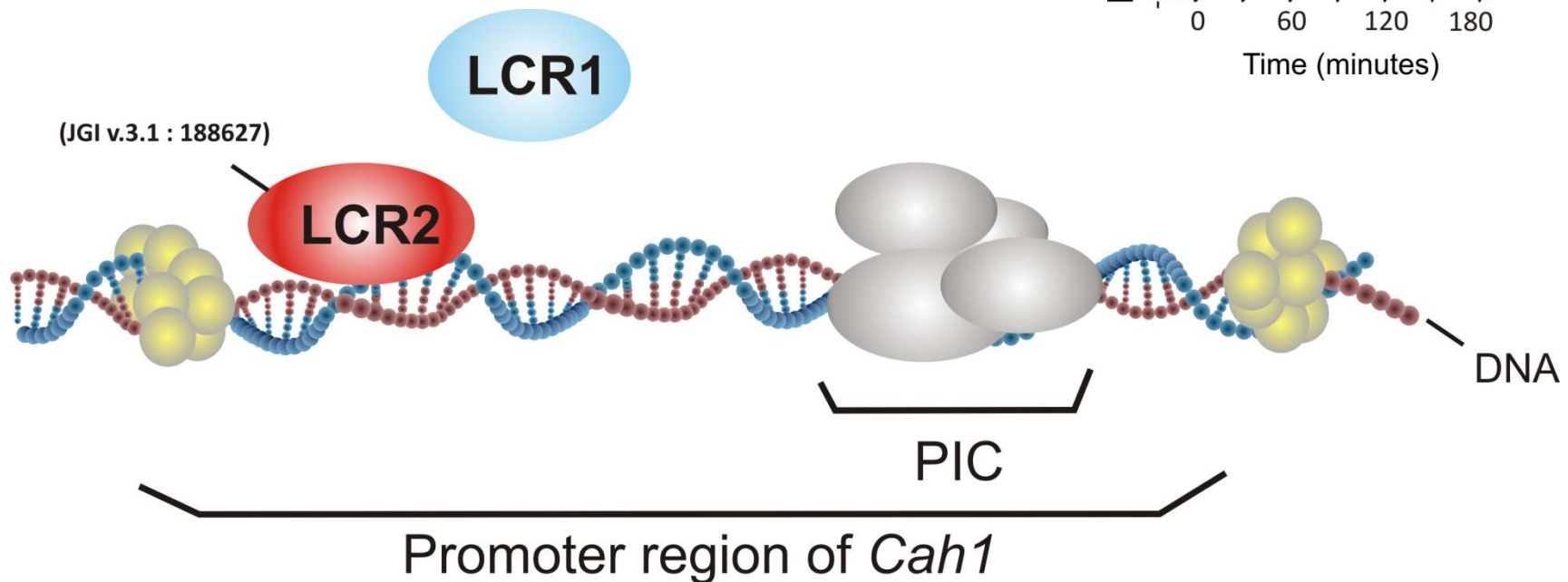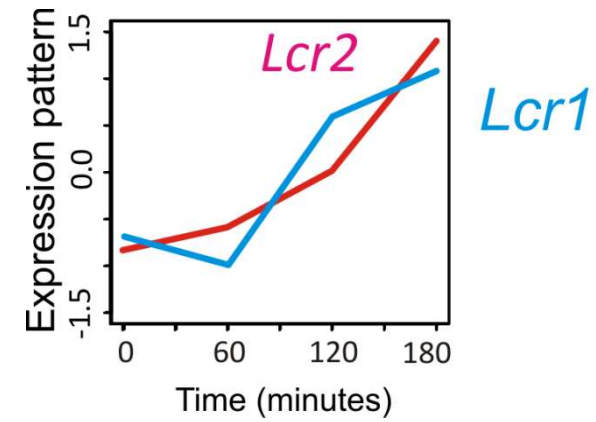
LCR = Low CO_2 stress response regulator

Identification of regulatory elements
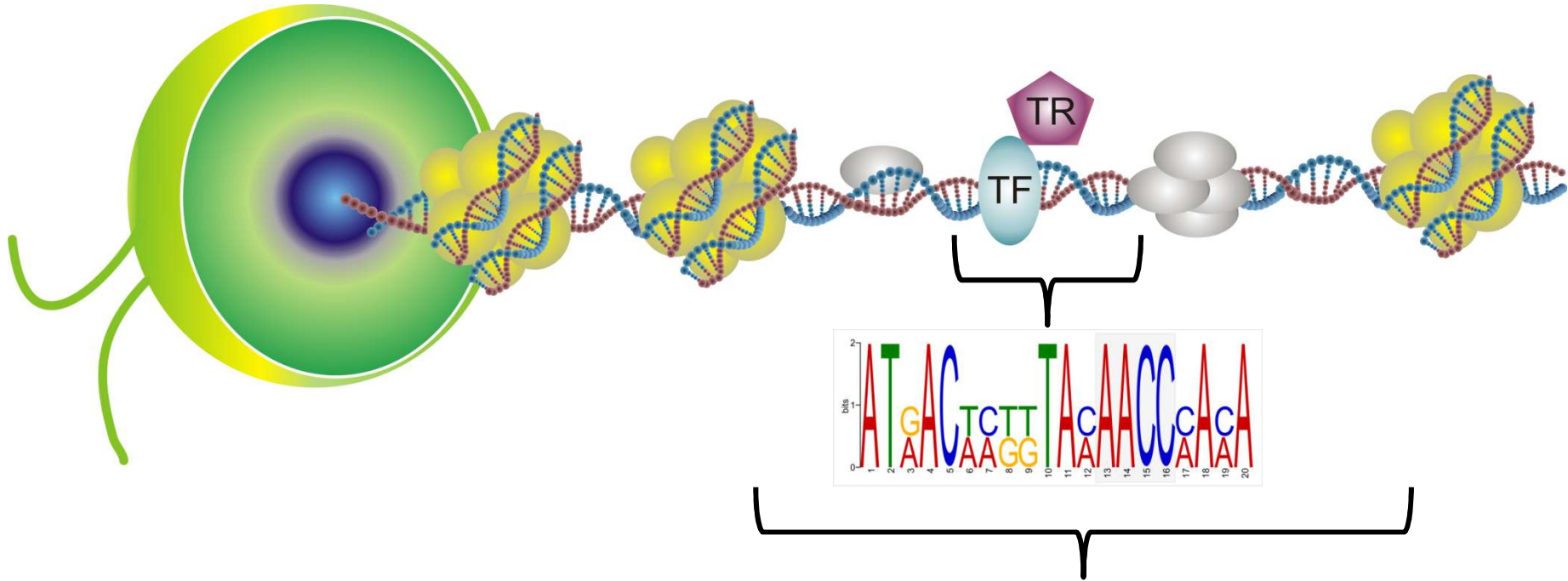
Regulatory element

Nucleosome-depletion is correlated with transcription

Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE)

Modified from Giresi *et al.,* 2007 *Genome Res.*
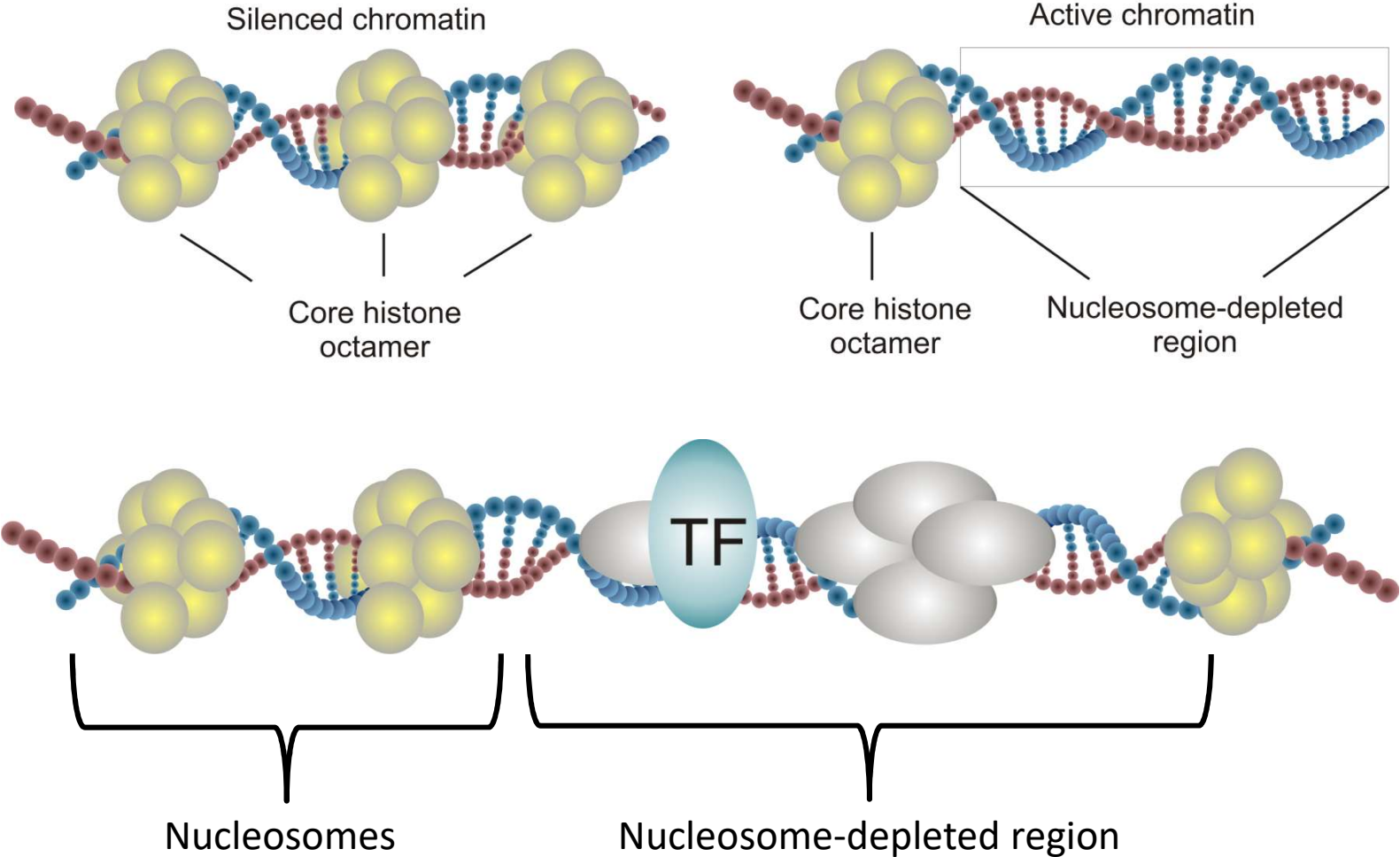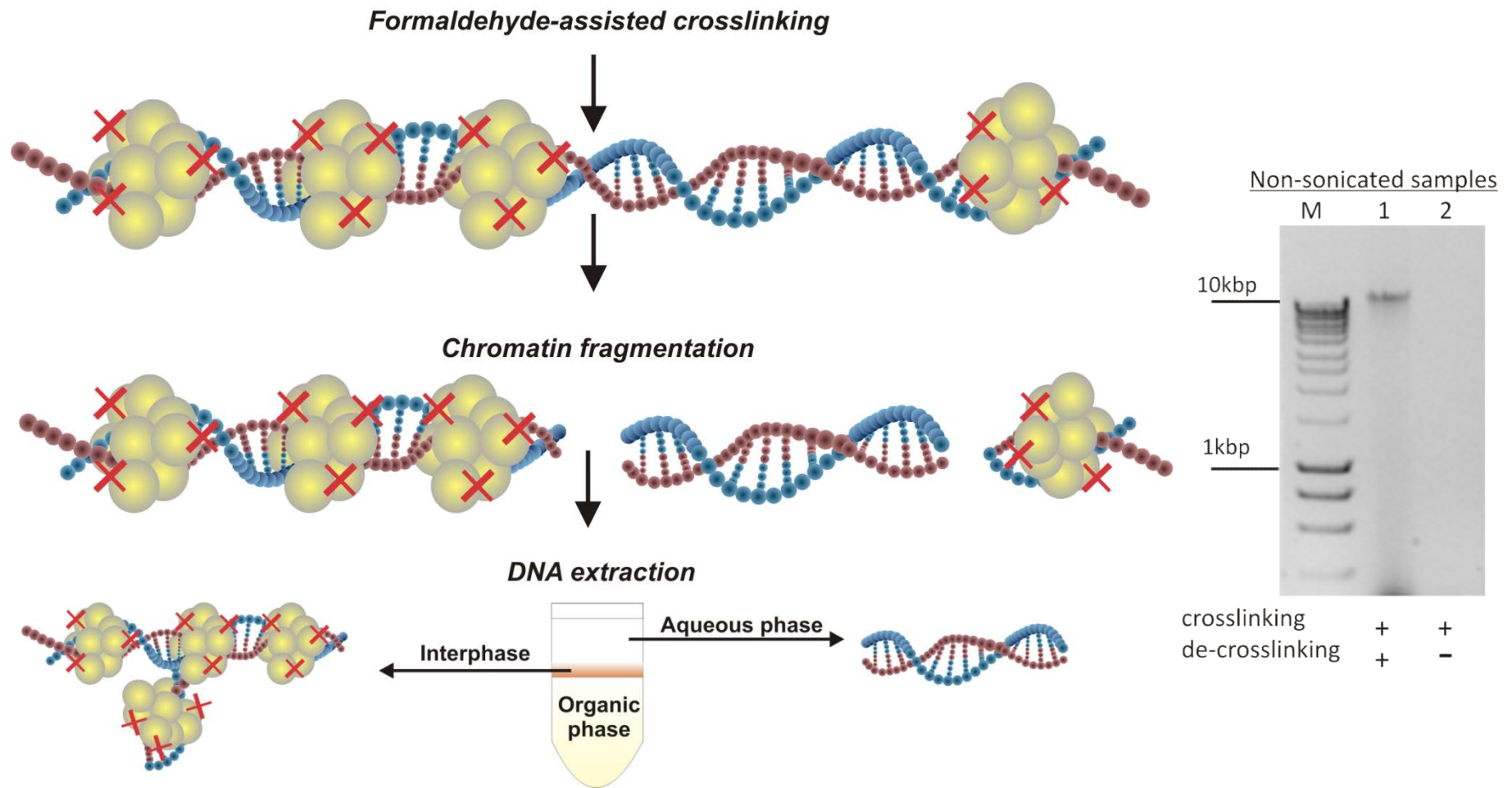
Regulatory regions of the genome were isolated from cells under carbon dioxide limitation

Formaldehyde-assisted crosslinking

Chromatin fragmentation

DNA extraction

Interphase

Aqueous phase

Organic phase

FAIRE-seq

Formaldehyde assisted isolation of regulatory elements (FAIRE)

Ilumina seq

Identification of enriched regions (MACS tool)
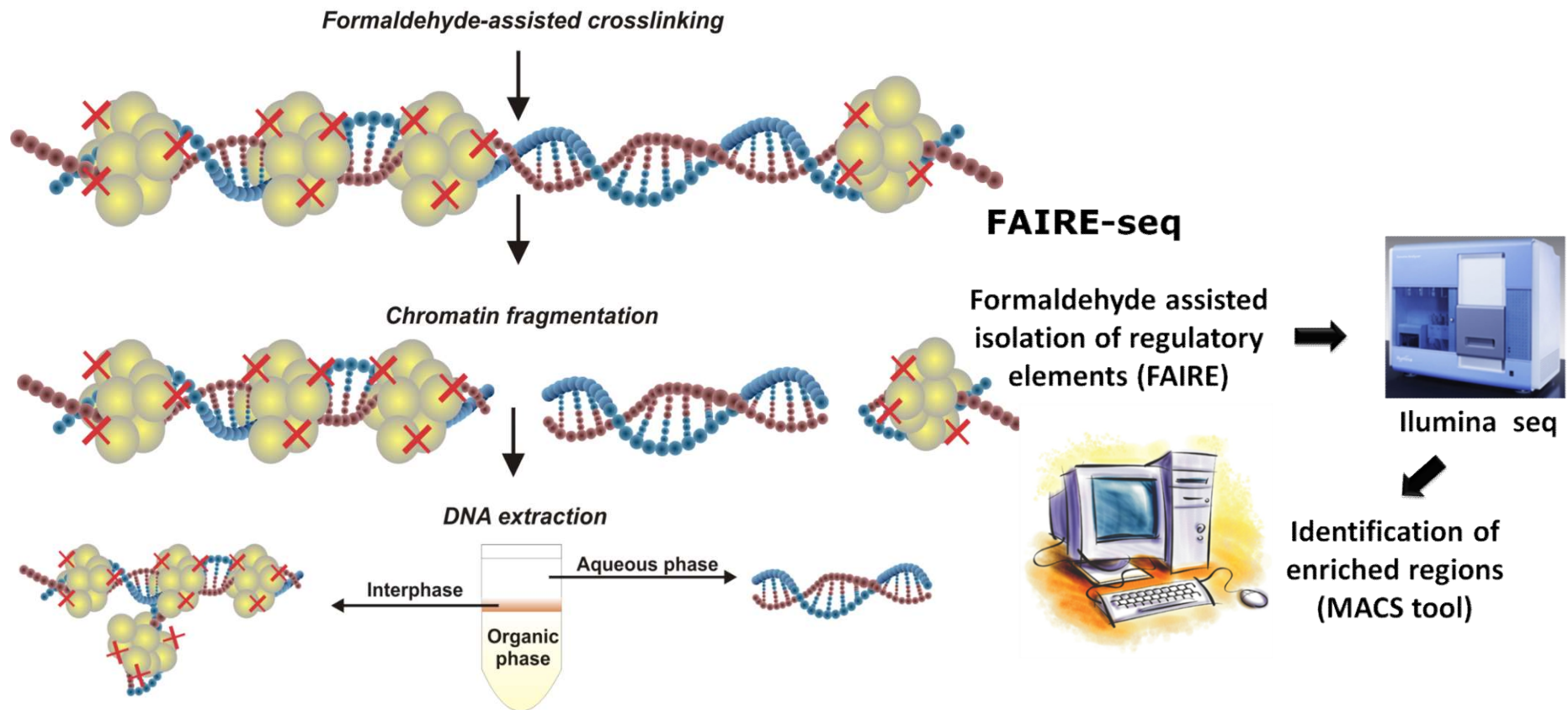
There are two candidate regulatory regions in the *Cah1* locus
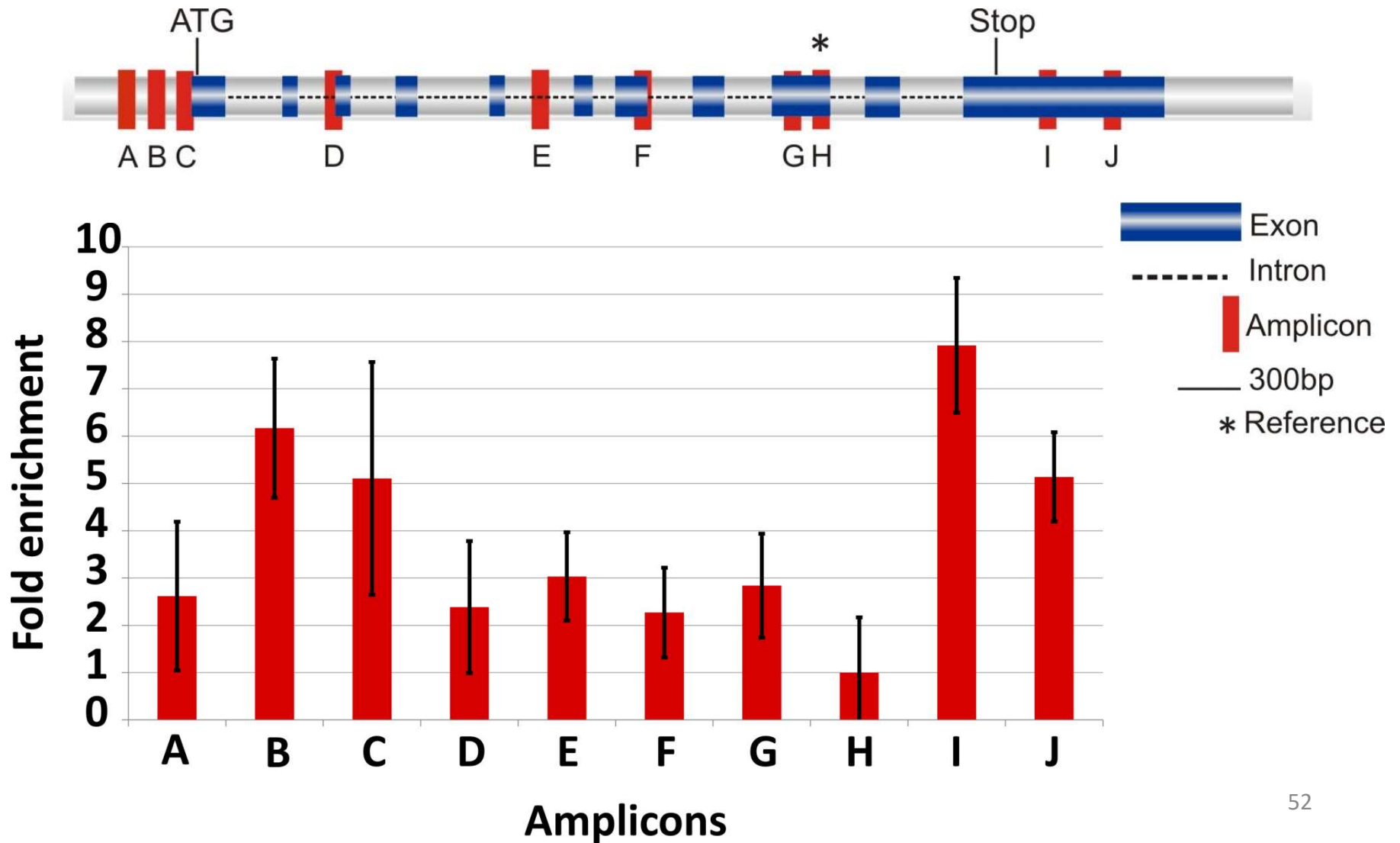
FAIRE-seq: Enriched FAIRE fragments were identified in a genome-wide manner

50 bp paired-end reads (Illumina Hiseq 2000)
Detection of enriched fragments using the MACS tool

Winck F.V., Arvidsson S, Riaño-Pachón DM, Hempel S, et al. (2013) Genome-Wide Identification of Regulatory Elements and Reconstruction of Gene Regulatory Networks of the Green Alga Chlamydomonas reinhardtii under Carbon Deprivation. PLoS ONE 8(11): e79909.

Regulatory regions were identified using FAIRE-seq

Cah1 gene transcript

FAIRE peaks

HSP70A gene transcript

DNAse I hypersensitive site
(-7bp to +67bp from the TSS)
(Schroda *et al.*, 2002 *Plant J.*)

FAIRE summit

http://tartarus.uniandes.edu.co/cgi-bin/gbrowse/Chlamydomonas_v4/
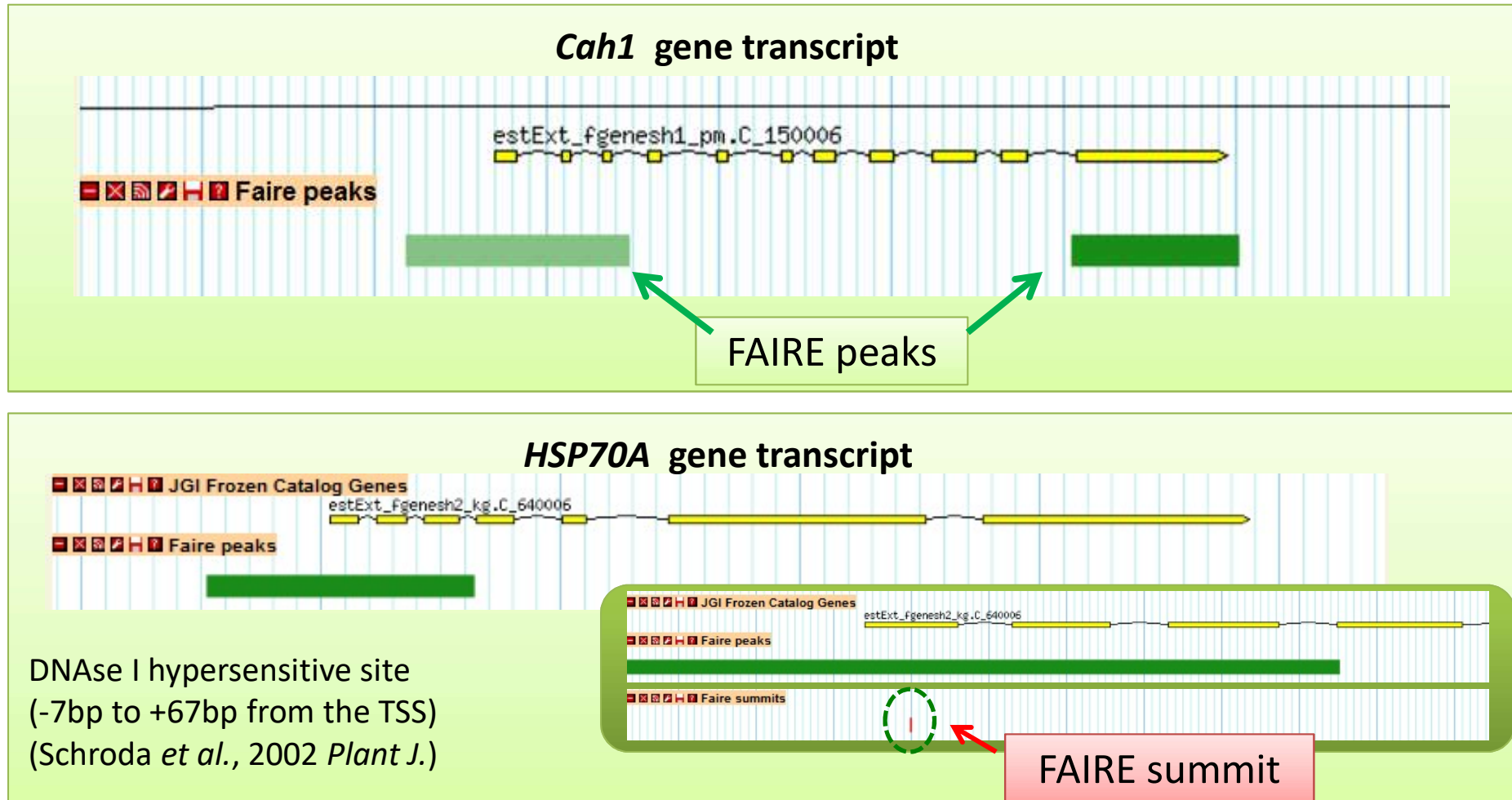
FAIRE peaks of regulatory genes responsive to carbon deprivation
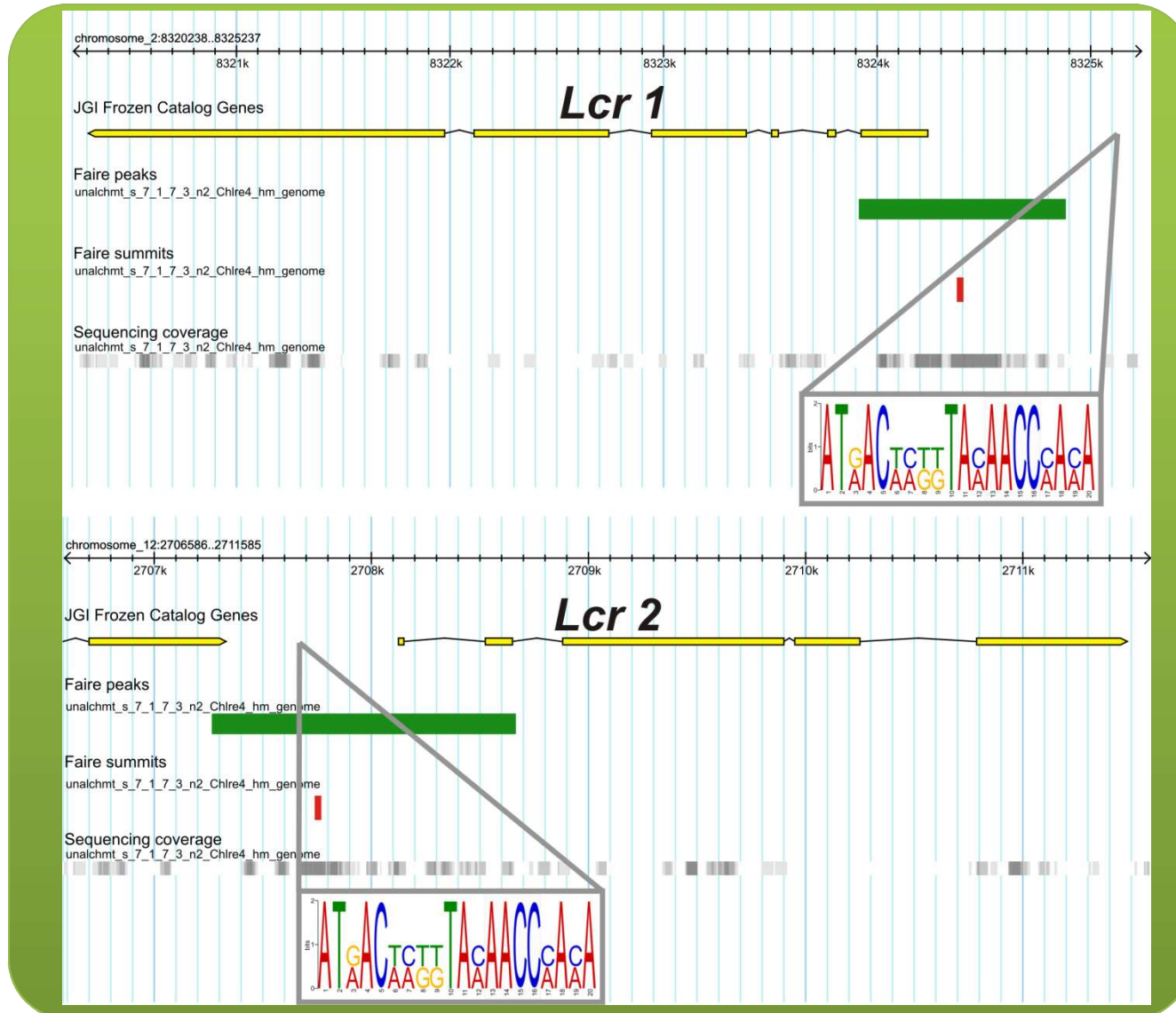
FAIRE summits tend to appear close to coding sequences
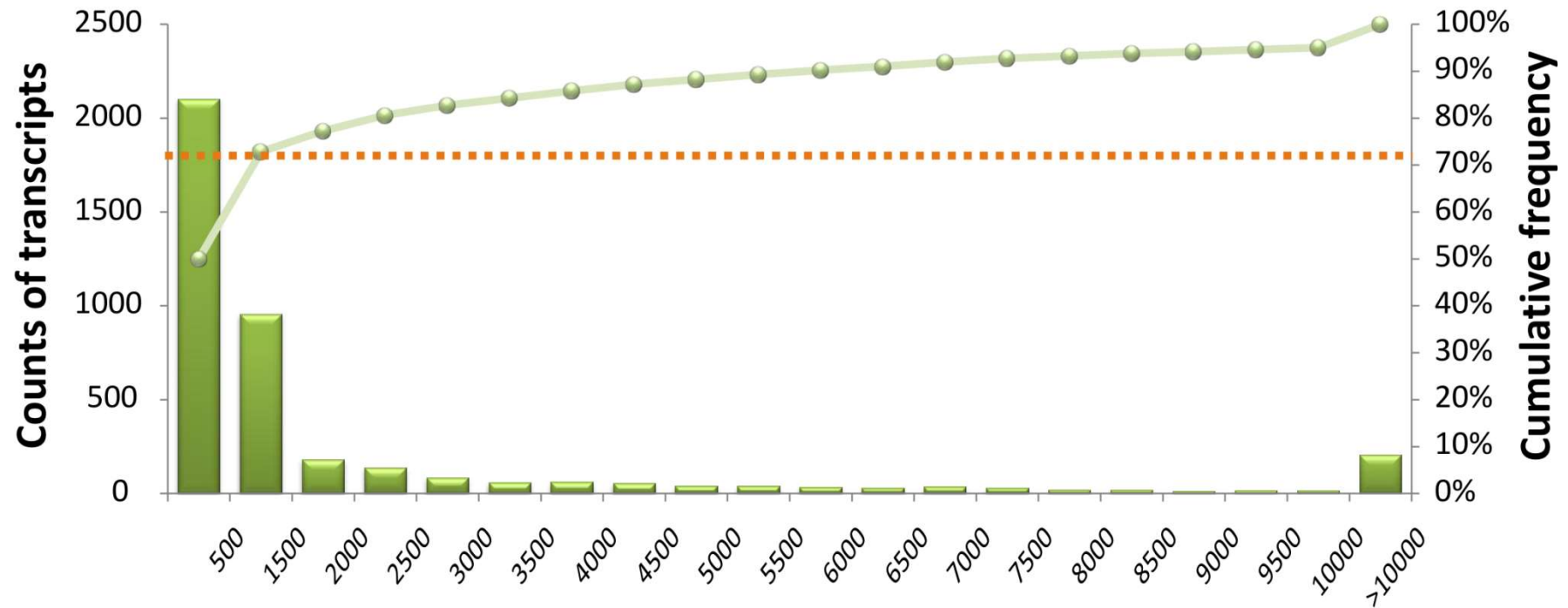
~ 70 % of the FAIRE summits are located within 1500 bp from 5′ or 3′ end

FAIRE peaks identified are genome-wide distributed

Correlation of FAIRE peaks

Positive for TF and TR genes with FC > 2

No correlation to the direction of the expression change (Up or Down)

Winck F.V., Arvidsson S, Riaño-Pachón DM, Hempel S, et al. (2013) Genome-Wide Identification of Regulatory Elements and Reconstruction of Gene Regulatory Networks of the Green Alga Chlamydomonas reinhardtii under Carbon Deprivation. PLoS ONE 8(11): e79909.

Integration of gene expression data with regulatory genomics data

TFs with a FAIRE peak in the TSS region and Fold change >2

TFs with Fold change >2 and no FAIRE peak in the TSS.

Transcriptional or post transcriptional control of gene expression levels

CO2 (input -ON/OFF)

TF

Gene promoter

58
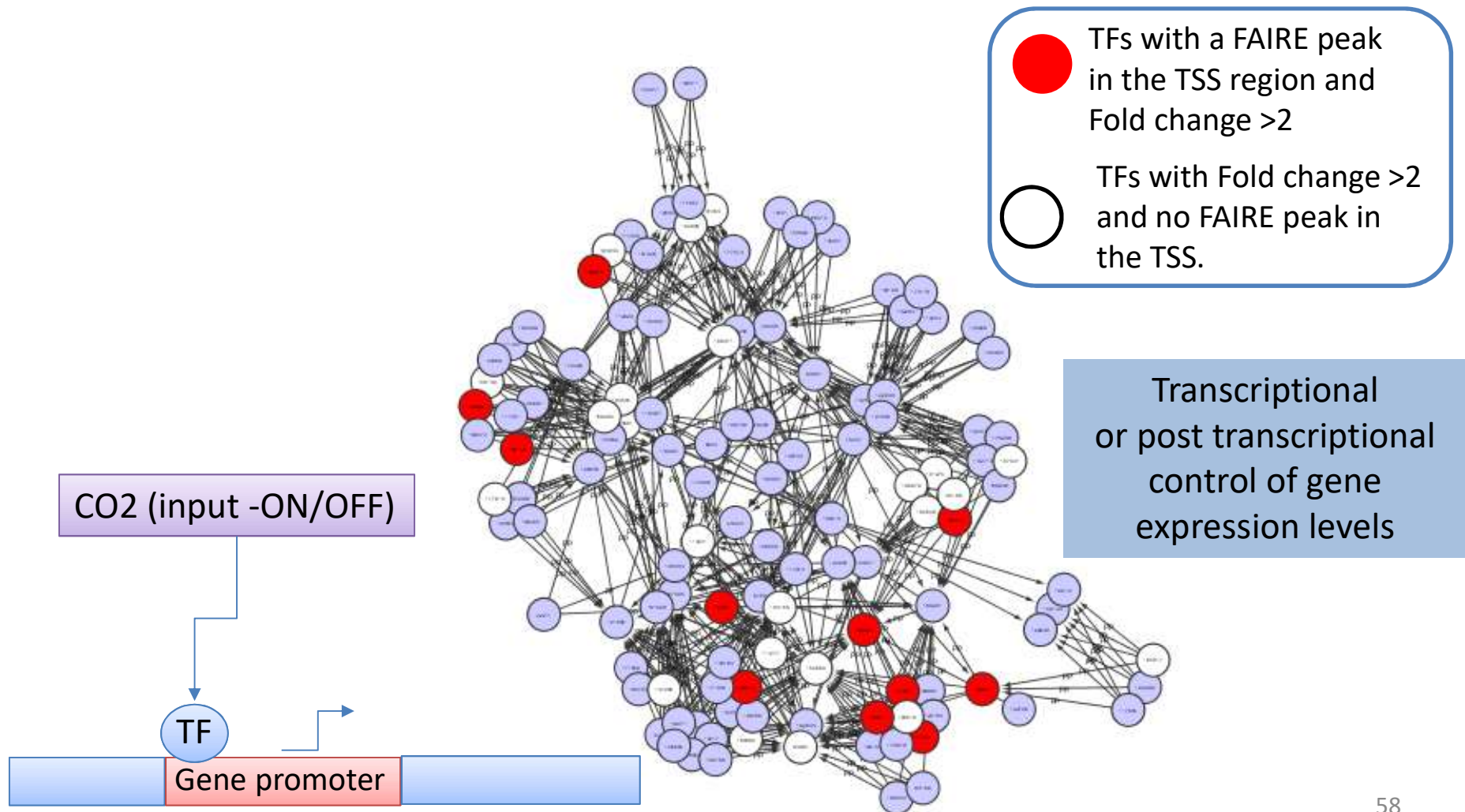
Different OMICS data can give us better insights

Chromatin de-crosslinking and protein extraction

LC-MS/MS

Deep-sequencing (Ilumina Hi-seq2000)

p-FAIRE

FAIRE-seq

Transcriptome

Correlation analysis
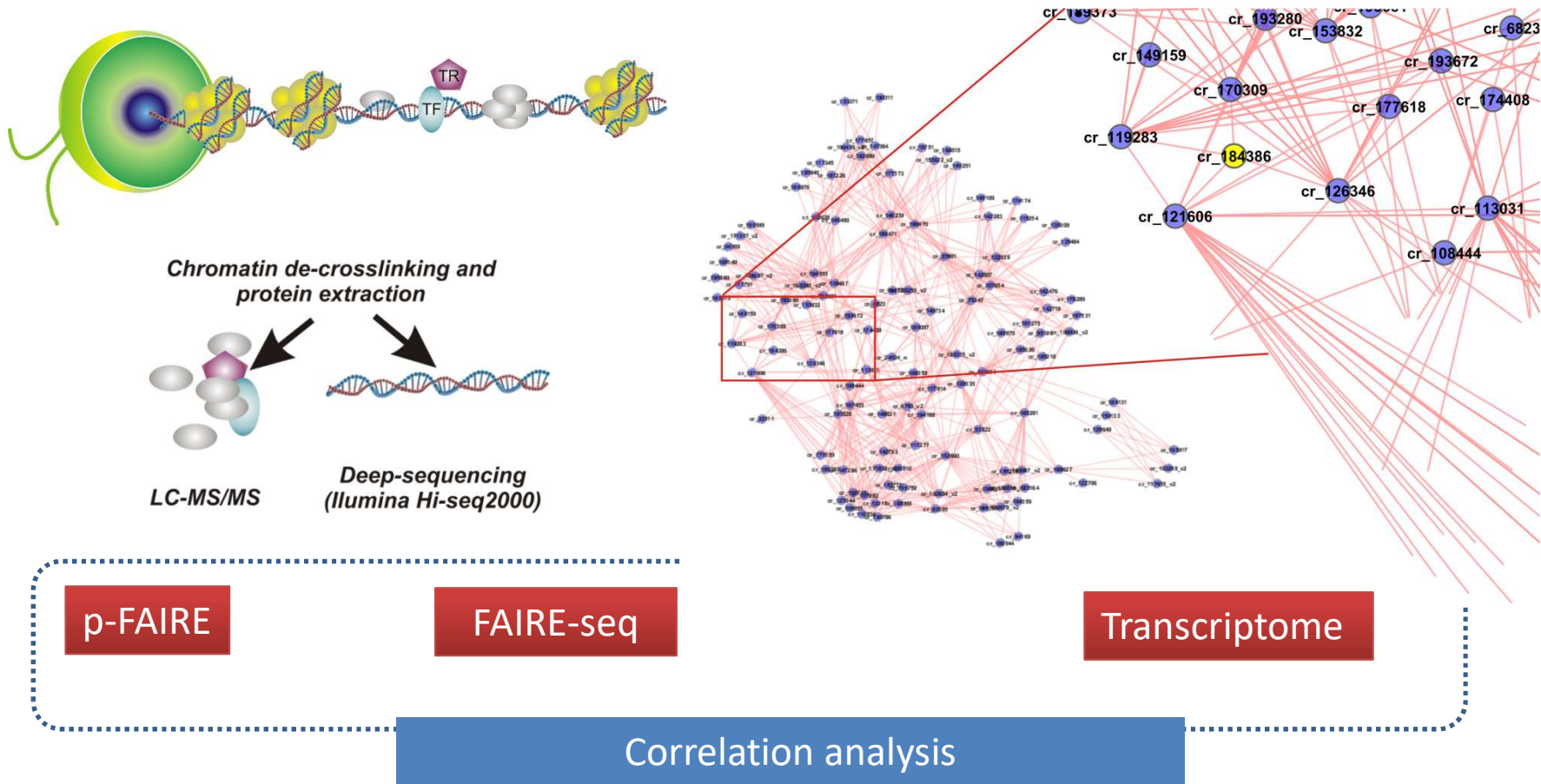
Inference of timeline for cellular responses
(Early-responsive genes and late-responsive genes)

DDT = DNA different transcription
HSF1 = Heat shock factor 1
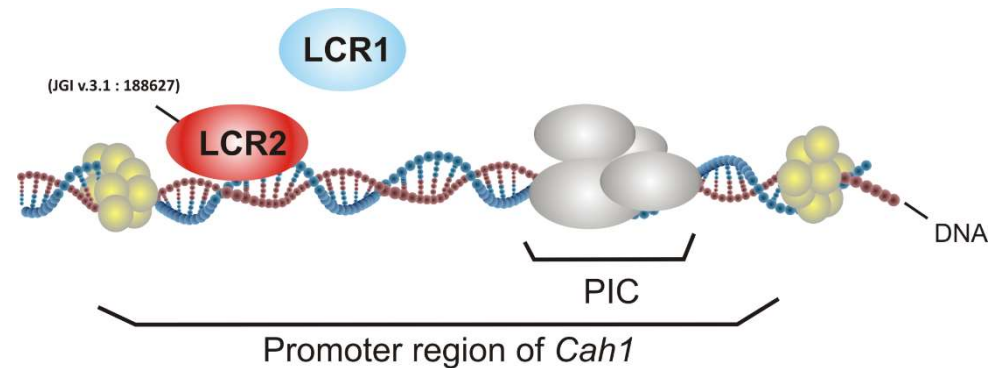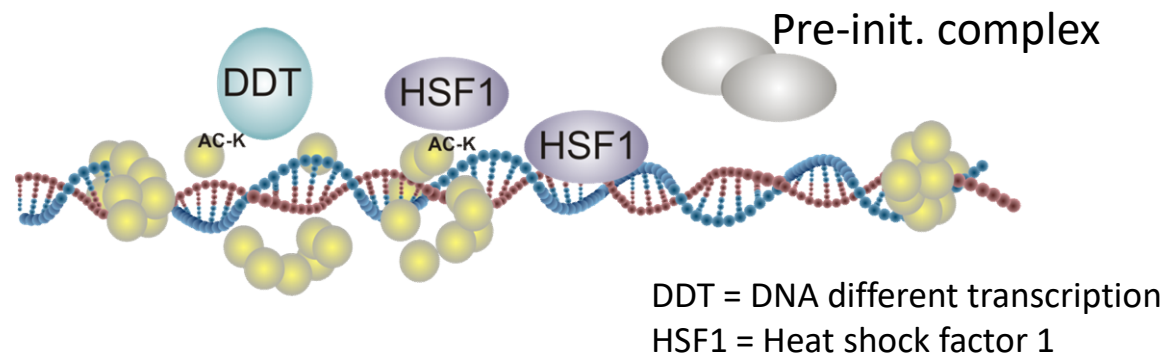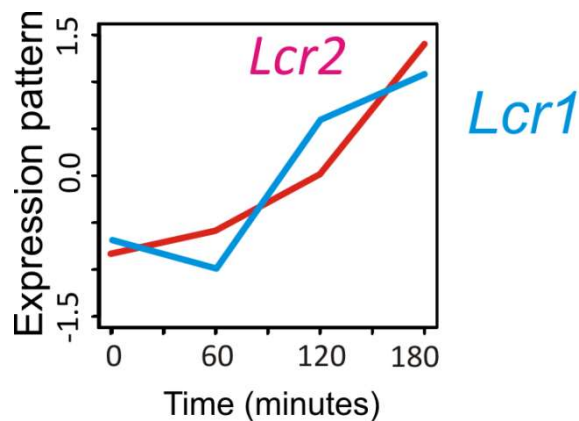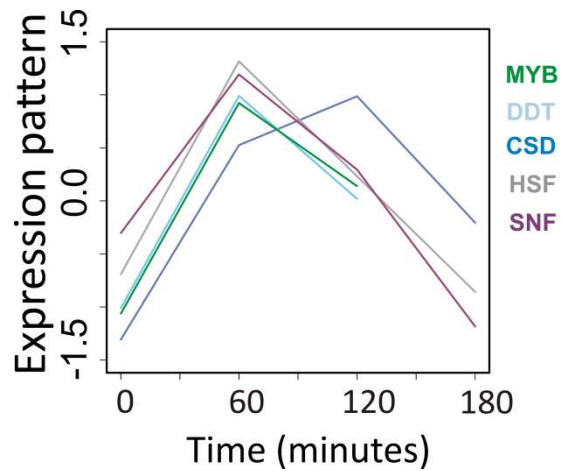
LCR = Low CO_2 stress response regulator

PIC = Pre-initiation complex

Proteome analysis of regulatory regions revealed the identity of annotated regulatory proteins

Formaldehyde-assisted crosslinking

Chromatin fragmentation and extraction

Chromatin de-crosslinking and protein extraction

p-FAIRE

Unpublished data

LC-MS/MS

Deep-sequencing
(Ilumina Hi-seq2000)
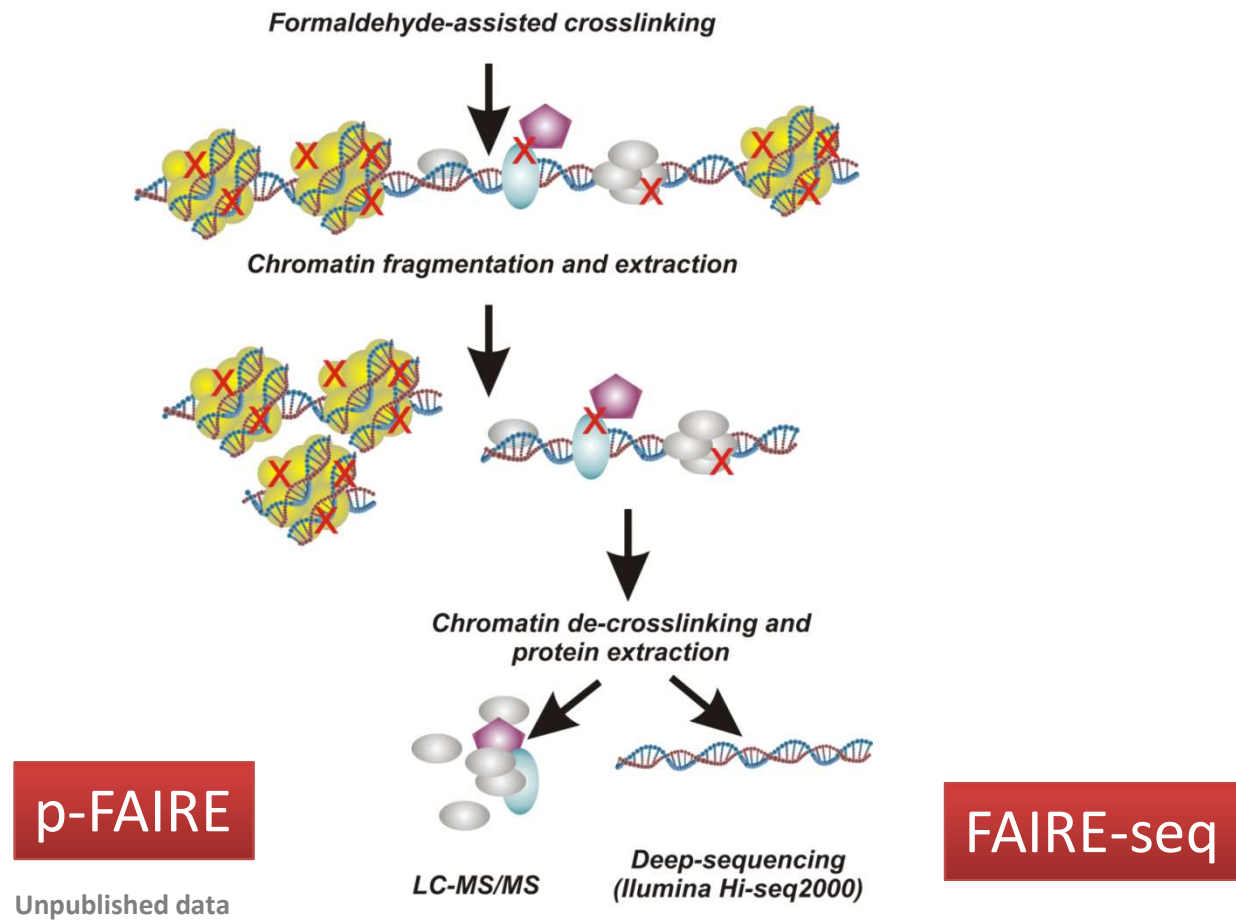
FAIRE-seq

FAIRE shotgun proteomics revealed identity of regulatory proteins responsive to variations in the CO_2 concentration

Complex protein sample

Mass spectrometry (Orbitrap XL)

Protein Identification (MaxQuant)

Control

Proteins found in chromatin fraction without formaldehyde crosslinking

141 "unspecific" DNA-binding proteins

- HSP70
- Ribosomal L19,L7Ae,S9,L17
- Nucleosome assembly protein (NAP)
- Glyceraldehyde 3-phosphate dehydrogenase
- Histone core
- Calcium-binding EF-hand
- Prefoldin
- Helicase superfamily 1 and 2
- Translation elongation factor P

(Unpublished data)

Specific regulatory proteins were identified under low CO2 concentration condition

High CO_2 (5%) vs Low CO_2 (0.04%)

⬇

189 stress responsive DNA-binding proteins

- 39 Transcription factors/regulators (SNFs, bZIP, Jumonji)
- Initiation factor eIF-4 gamma;
- Argonaute and Dicer proteins;
- Zinc finger;
- GATA-type;
- Ankyrin;
- Tetrapyrrole biosynthesis,
- hydroxymethylbilane synthase

(Unpublished data)

Integration of gene expression information, regulatory genomics and proteomics data of carbon limitation
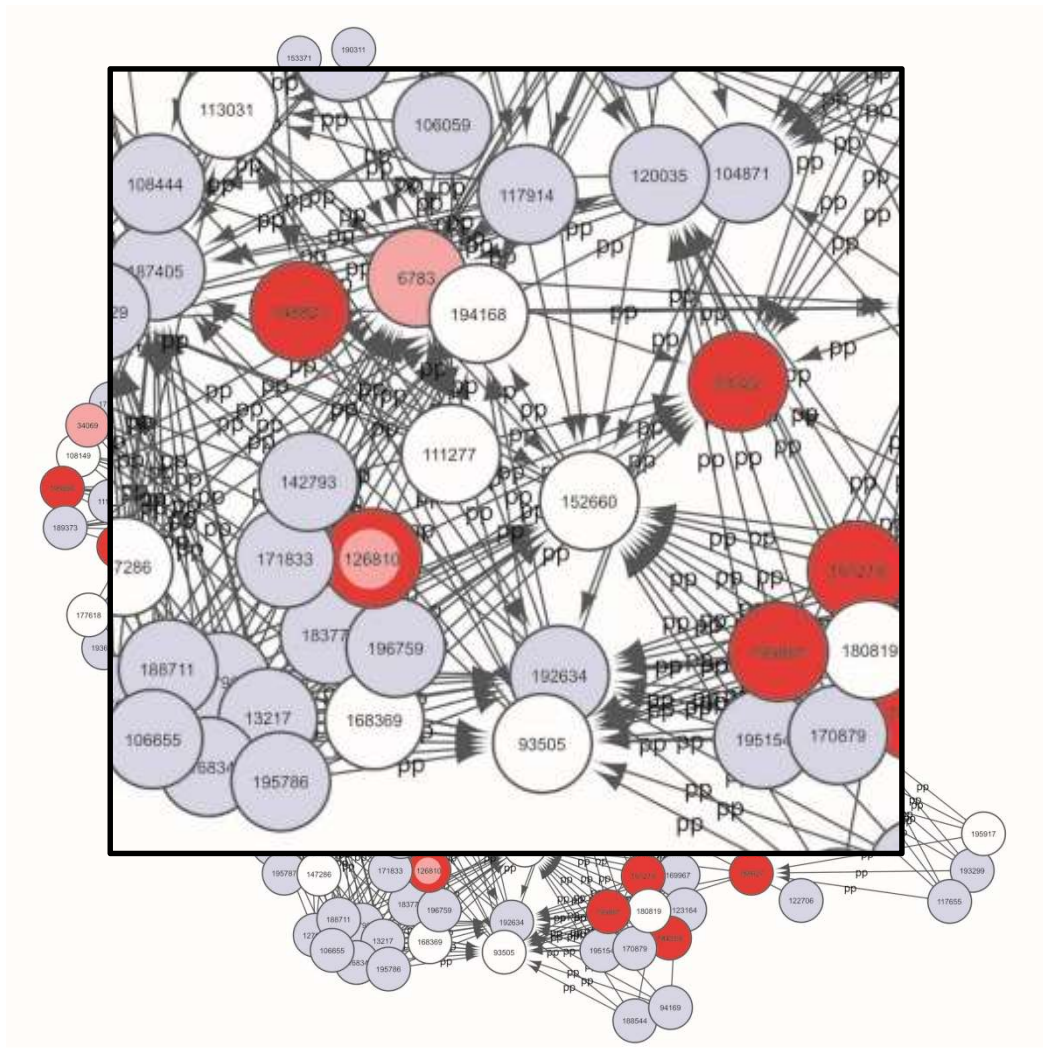
Legend:
- 🔴 TFs with a FAIRE peak in the TSS region and Fold change >2
- ⚪ TFs with Fold change >2 and no FAIRE peak in the TSS.
- 🟠 TFs identified in the Proteome of chromatin
- 🟢 TFs identified in the Proteome of chromatin, No FAIRE peak

Integrative analysis gave us testable hypothesis

1h

FHA

C3H

TR

TF

?

MYB

Regulatory genes
+
Regulatory proteins

3h

TR

TF

Regulatory regions
+
Regulatory proteins
+
Regulatory genes

bHLH

RWP-RK
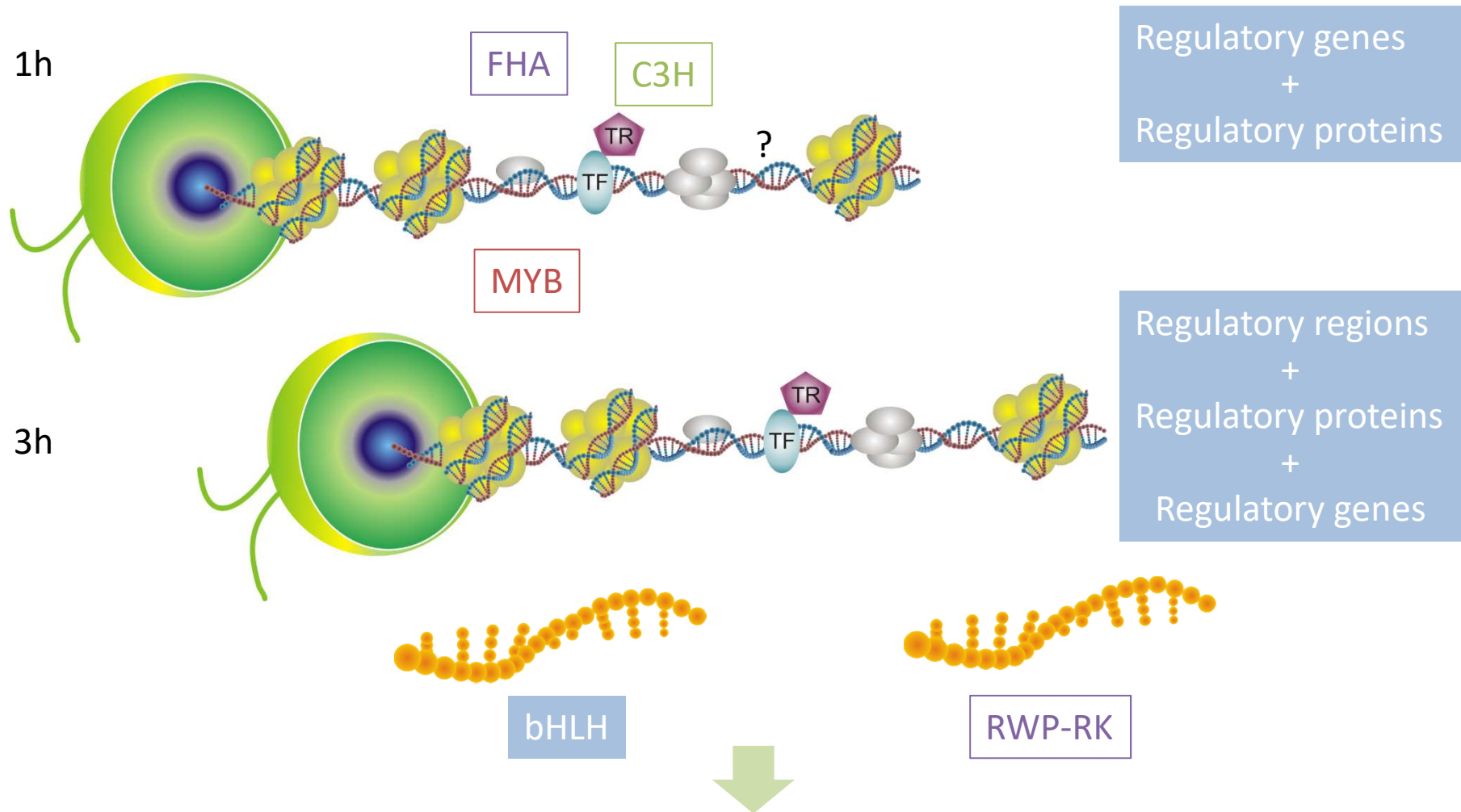
Testable hypothesis of transcriptional control of the CCM mechanism

65

Gene expression network and proteome analysis revealed

Transcripts and proteins association					
Identifier	gene family	Fold change 1h	Fold change 2h	Fold change 3h	FAIRE peak
6783	SNF2	0.845556467199114	5.89144285360181	0.71443470567	
34069	FHA	1.93629295615854	0.985677031777896	0.6502876073	
101275	CCAAT	0.585306980608999	0.786567180219919	0.34460629804	
126810	CSD	6.85803552567784	11.127986144712	3.1816773588	
129649	MYB	1.33362368604303	1.11777881115331	1.0558646708	
135484	SNF2	0.616329401178817	1.50848377331645	0.82714320871	
145251	HMG	0.76076099161463	0.50358781730018	1.3640979323	
146239	C3H	1.13833479961758	0.930144334461022	0.57714902489	
149734	C2H2	0.887126470785533	0.754275300871498	1.0465798934	
187531	bZIP	0.586500467357807	1.17644290555762	0.46354668714	
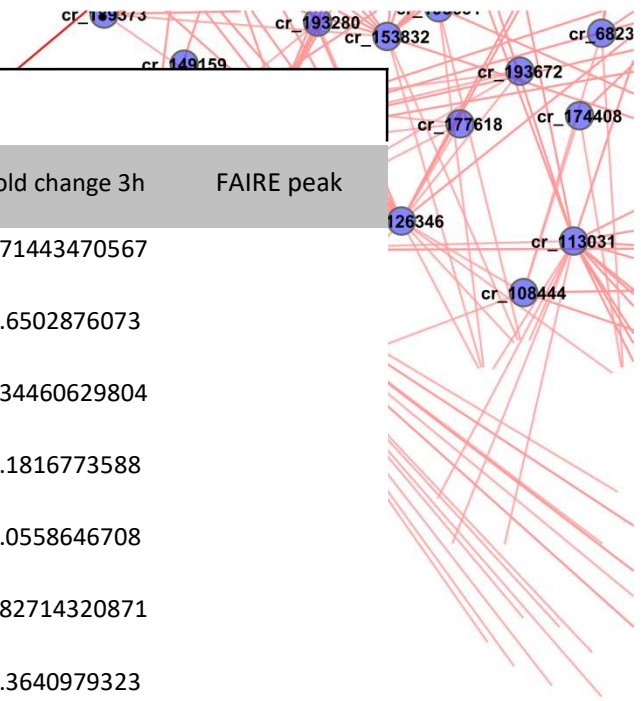
Integrative analysis can give us better insights on gene regulatory networks
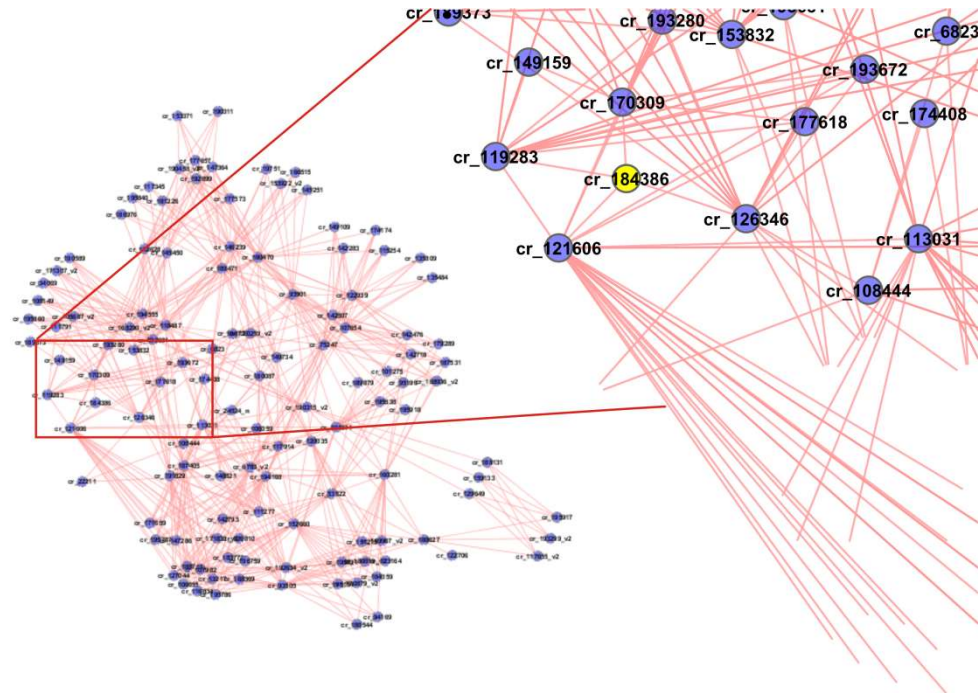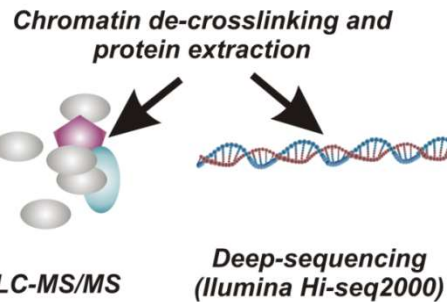
Proteomics (p-FAIRE)

Regulatory genomics (FAIRE-seq)

Transcriptomics (TFs qRT-PCR)

Correlation analysis

Carnielli, C.M., Winck, F.V., Paes Leme, A.F. (2015) Functional annotation and biological interpretation of proteomics data. *Biochimica et Biophysica Acta* 1854, 46–54.

Identification of critical networks and elements related to regulation of biomass accumulation
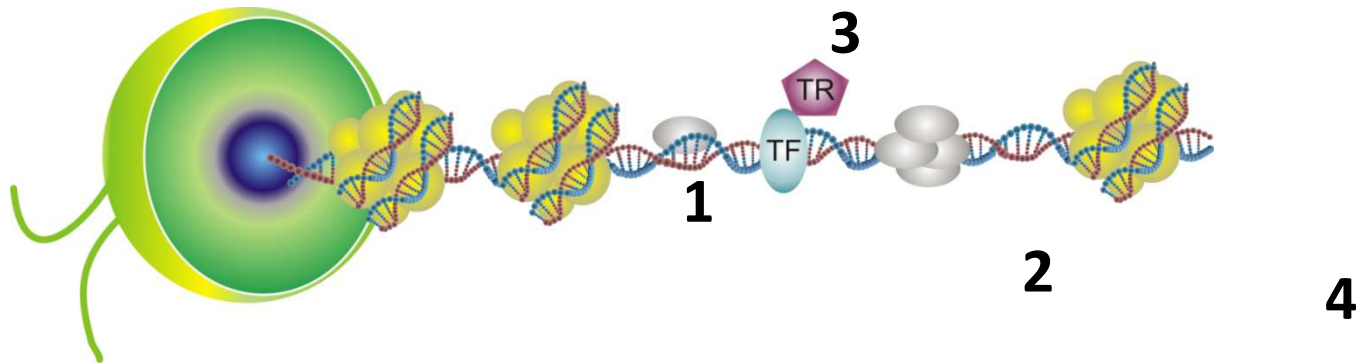
| **1** Genomics | **2** Transcriptomics | **3** Proteomics | **4** Metabolomics |

Quantitative models, modeling and simulation

Selection of subnetworks for transcriptional engineering

Melo D., Moncada R.-P., **Winck F.** and Gonzalez Barrios, A.F. (2014) In Silico Analysis for Biomass Synthesis under Different CO2 Levels for *Chlamydomonas reinhardtii* Utilizing a Flux Balance Analysis Approach. In Castillo, L.F.*, et al.* (eds), *Advances in Computational Biology*. Springer International Publishing, pp. 279-285.

Information that the integrative analysis provided

Which TFs may play a role in biomass accumulation

When TFs exert regulatory role

Where TFs interact with chromatin

Further analysis performed

What effects TFs generate on the metabolism

How TFs affect metabolic sensitive genes

Metabolism

Gene expression

TFs
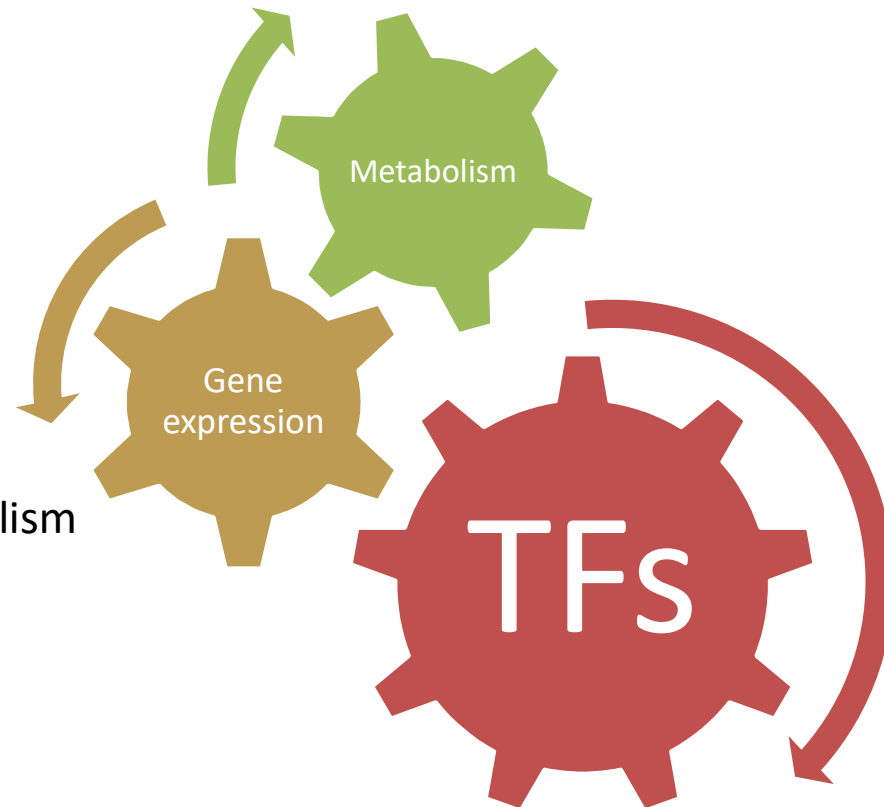
Integrative analysis helped us to reduce the number of possible candidate subnetworks that may control biomass accumulation

CO2 (input -ON/OFF)

TF
promoter
gene
transcript
protein
Biomass (Output)
Metabolite
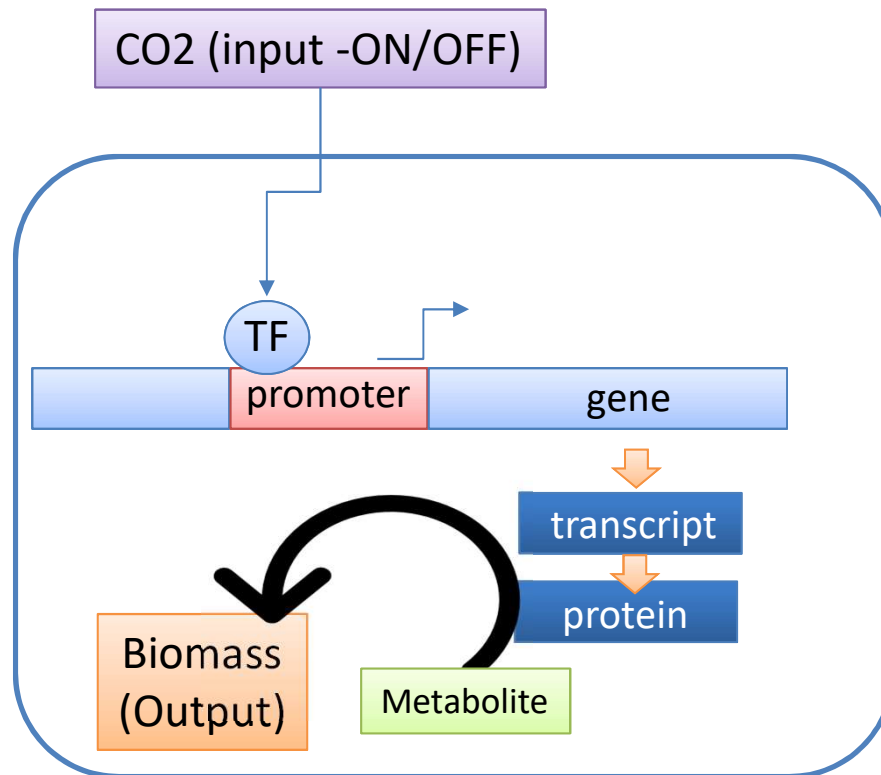
Input: levels of CO2 (Low/High)

Regulatory regions: TF-DNA interactions

Regulatory genes: Modulated regulators

Target genes: Metabolic targets (sensitive genes)

Output: High biomass/Lipids/Pigments

Relevance of integration schemes

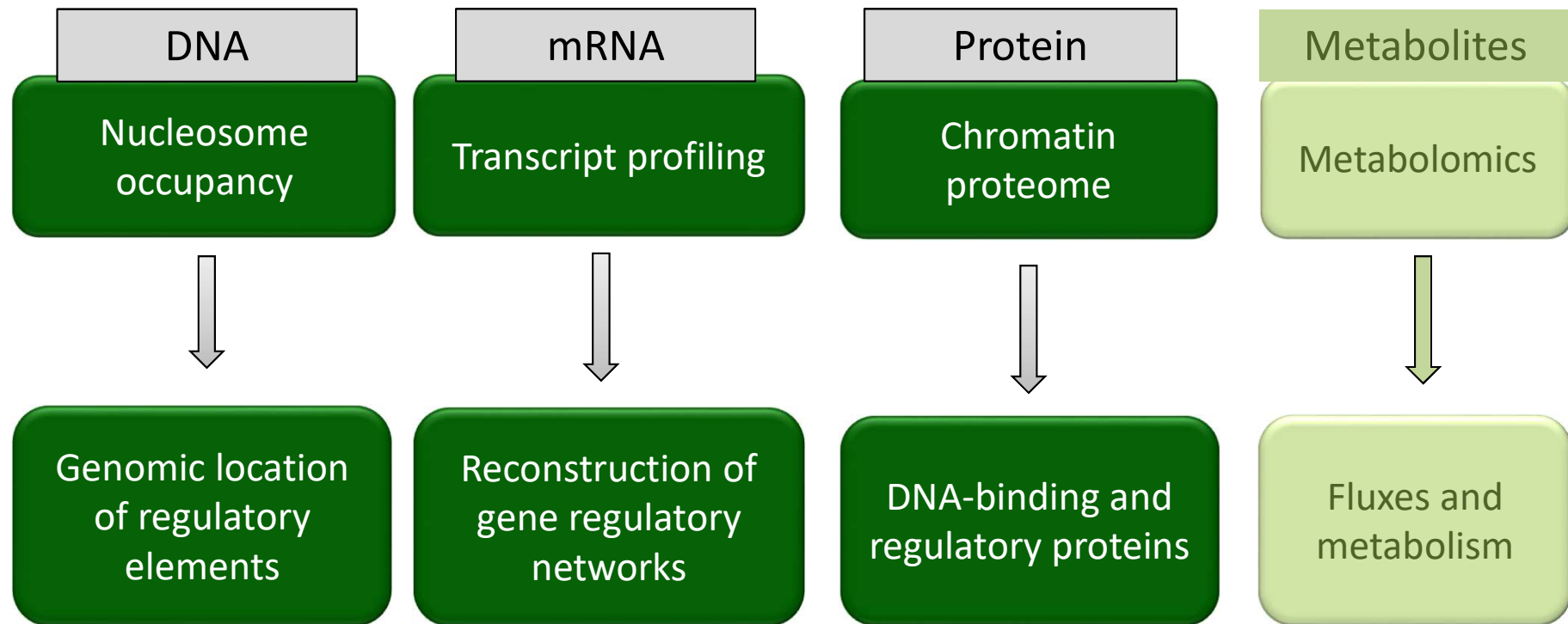|	RNA-Seq	ncRNA	ChIP-Seq Histone	ChIP-Seq TF	CpG DNA Methylation	DNase-Seq	Complete DNA sequencing	Exome sequencing	Proteomics	Metabolomics	Chromatin Conformation	Clinical Data	Co-morbidities	Other
RNA-Seq		29.6%	24.8%	29.6%	32.8%	16.0%	21.6%	22.4%	36.8%	21.6%	14.4%	28.0%	10.4%	0.0%
ncRNA	6.4%		8.0%	7.2%	10.4%	4.0%	6.4%	8.0%	5.6%	4.0%	1.6%	10.4%	4.0%	0.0%
ChIP-Seq Histone	6.4%	0.8%		16.0%	16.0%	11.2%	3.2%	4.8%	7.2%	4.0%	8.8%	5.6%	2.4%	0.0%
ChIP-Seq TF	6.4%	0.8%	0.8%		12.0%	16.0%	5.6%	7.2%	9.6%	4.0%	10.4%	7.2%	2.4%	0.0%
CpG DNA Methylation	11.2%	2.4%	3.2%	2.4%		8.8%	9.6%	7.2%	6.4%	4.0%	9.6%	12.0%	4.8%	0.0%
DNase-Seq	4.0%	0.8%	1.6%	2.4%	4.8%		4.0%	5.6%	4.8%	4.0%	10.4%	9.6%	2.4%	0.0%
Complete DNA sequencing	8.8%	1.6%	1.6%	1.6%	2.4%	4.0%		10.4%	13.6%	10.4%	2.4%	20.0%	5.6%	0.0%
Exome sequencing	17.6%	0.8%	1.6%	0.8%	2.4%	0.8%	6.4%		12.0%	8.8%	0.0%	20.0%	7.2%	0.0%
Proteomics	15.2%	1.6%	0.8%	0.8%	1.6%	2.4%	4.8%	8.0%		27.2%	5.6%	16.8%	5.6%	1.6%
Metabolomics	16.8%	2.4%	2.4%	1.6%	3.2%	2.4%	6.4%	4.8%	10.4%		2.4%	17.6%	6.4%	0.8%
Chromatin Conformation	0.8%	0.0%	2.4%	2.4%	0.8%	0.0%	0.8%	0.0%	0.0%	0.8%		4.0%	2.4%	0.0%
Clinical Data	31.2%	8.0%	7.2%	9.6%	15.2%	9.6%	20.0%	21.6%	16.8%	20.0%	4.0%		14.4%	3.2%
Co-morbidities	8.8%	4.0%	3.2%	5.6%	6.4%	4.8%	7.2%	5.6%	2.4%	5.6%	0.8%	16.0%		1.6%
Other	0.8%	0.0%	0.0%	0.0%	0.8%	0.0%	0.8%	0.0%	0.0%	0.0%	0.0%	2.4%	0.8%	

	RNA-Seq	ncRNA	ChIP-Seq Histone	ChIP-Seq TF	CpG DNA Methylation	DNase-Seq	Complete DNA sequencing	Exome sequencing	Proteomics	Metabolomics	Chromatin Conformation	Clinical Data	Co-morbidities	Other
Same data Type in Basic Science	14.4%	6.4%	5.6%	6.4%	4.8%	3.2%	5.6%	4.0%	7.2%	4.8%	2.4%	4.0%	3.2%	1.6%
Same data type in Clinical Environment	5.6%	0.0%	0.0%	0.8%	0.8%	0.0%	2.4%	0.0%	1.6%	4.0%	0.0%	5.6%	0.8%	0.0%

(N=125 participants)

Integrative approaches may give us broader insights

DNA	mRNA	Protein	Metabolites
Nucleosome occupancy	Transcript profiling	Chromatin proteome	Metabolomics
↓	↓	↓	↓
Genomic location of regulatory elements	Reconstruction of gene regulatory networks	DNA-binding and regulatory proteins	Fluxes and metabolism

Models of biological systems from systematic measurements

Winck, F.V., Paez Melo, D.O. and Gonzalez Barrios, A.F. (2013) Carbon acquisition and accumulation in microalgae Chlamydomonas: Insights from "omics" approaches, *J Proteomics*, **94C**, 207-218
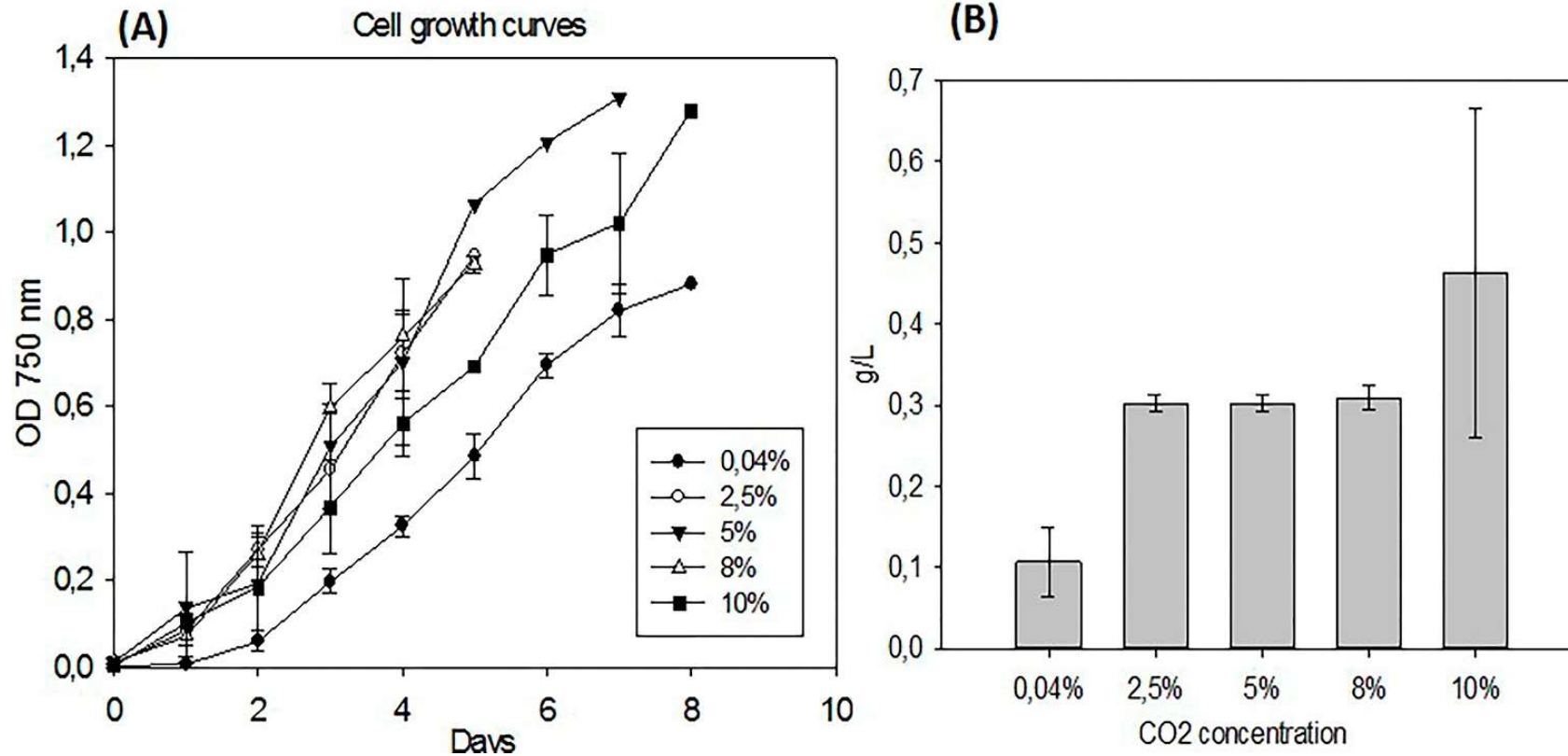
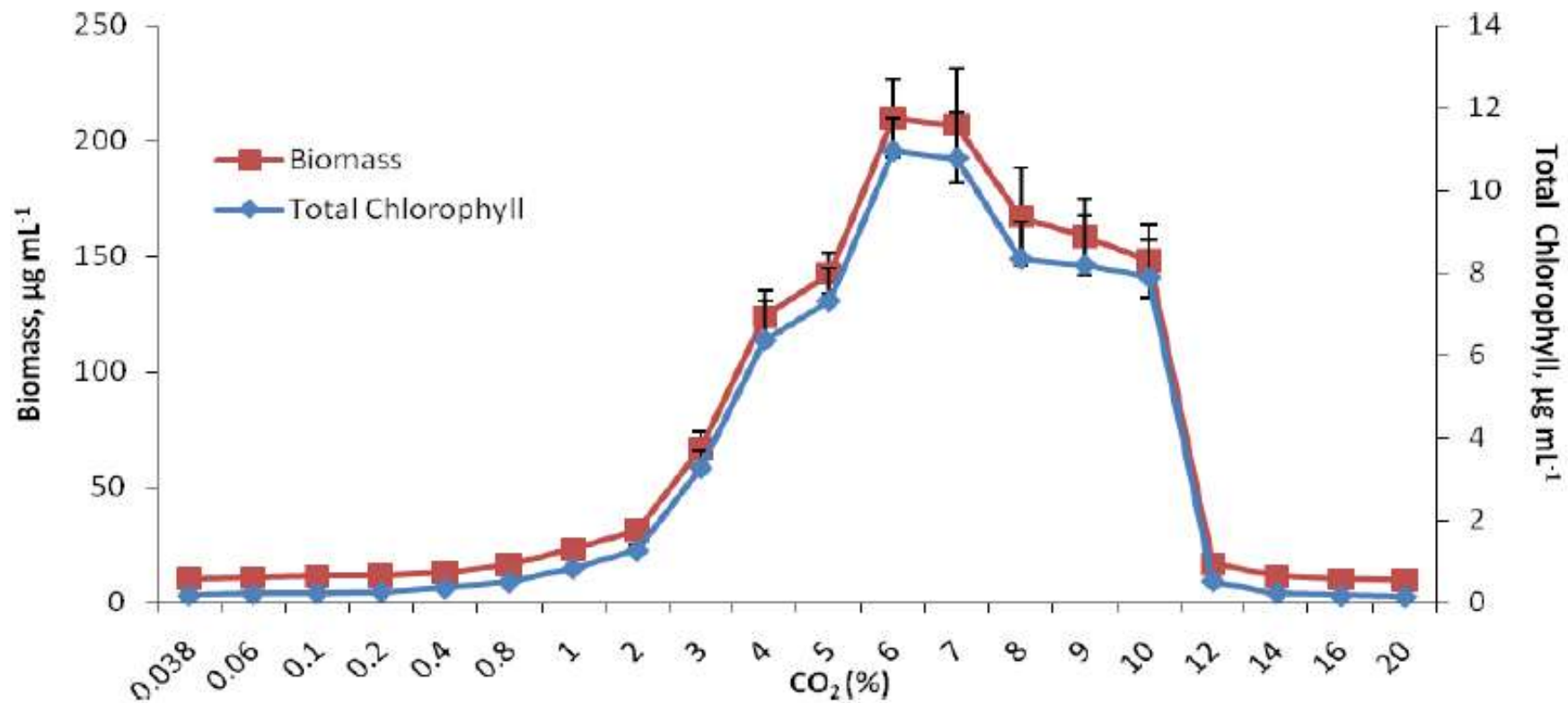# Cellular responses to varying CO_2 may reveal pathways for improving biomass accumulation
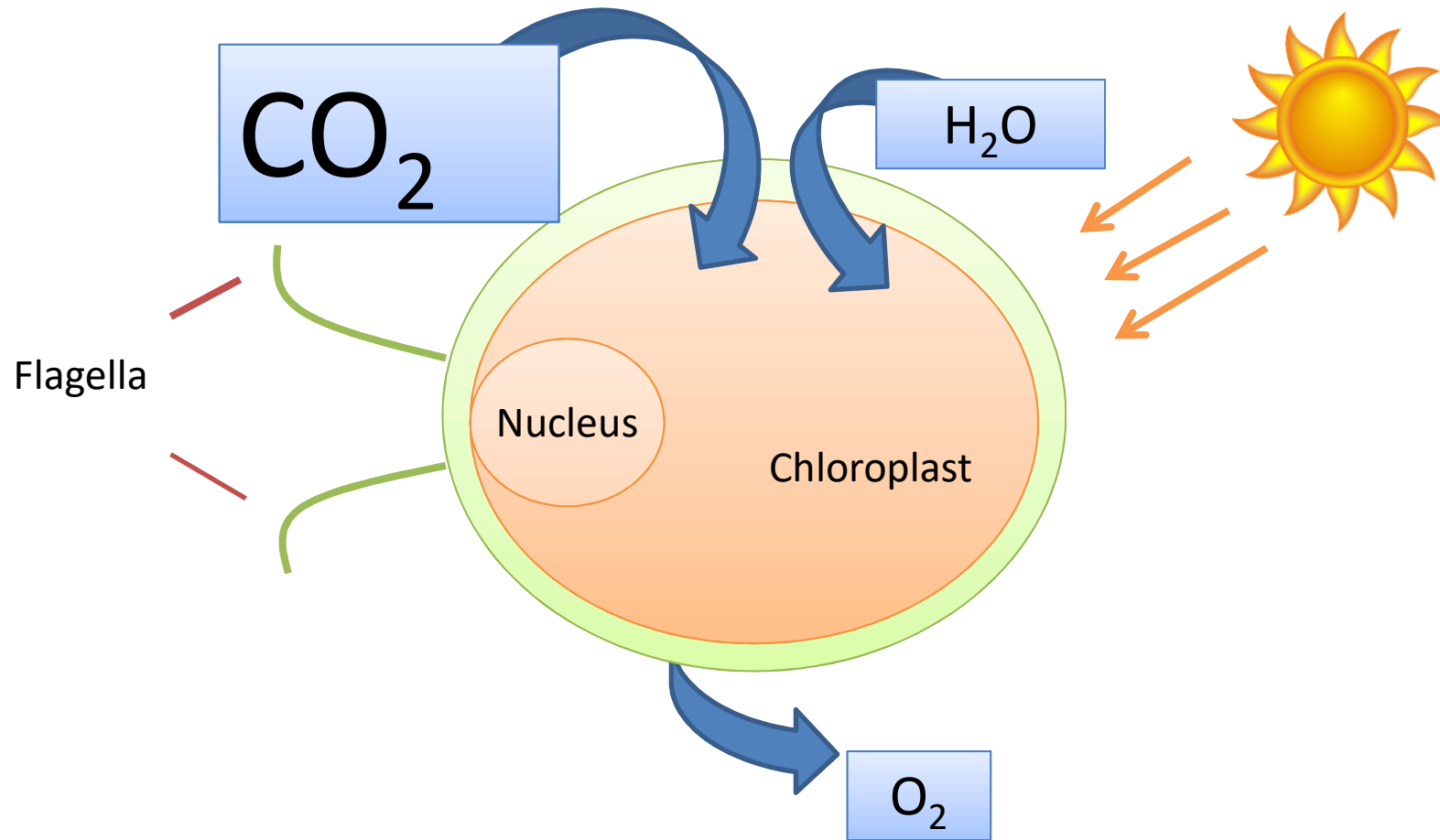
# Higher CO_2 concentration lead to enhanced biomass accumulation in microalgae
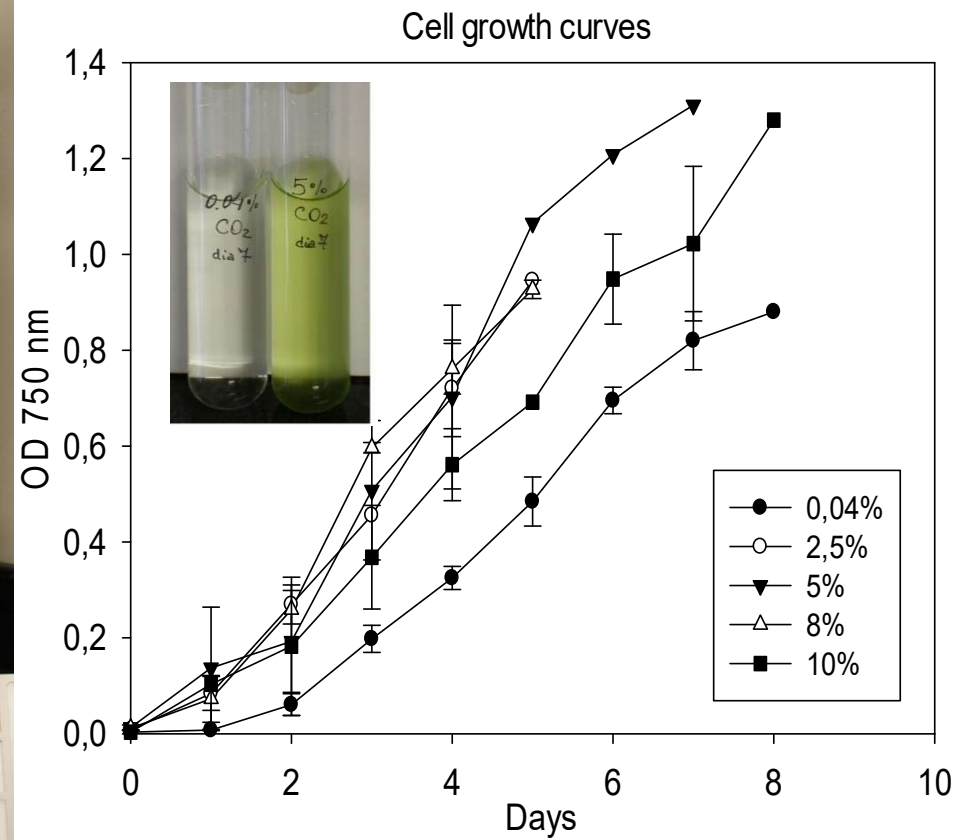
More carbon dioxide is not always better

(A) Cell growth curves — OD 750 nm vs Days. Legend: 0,04%, 2,5%, 5%, 8%, 10%

(B) g/L vs CO2 concentration: 0,04%, 2,5%, 5%, 8%, 10%

Winck, F.V., Páez Melo, D.O., Riaño-Pachón, D.M., Caldana, C., Martins, M., González Barrios, A.F. (2015) Analysis of sensitive CO_2 pathways and genes related to carbon uptake in *Chlamydomonas reinhardtii* through genomic scale modeling and experimental validation (Unpublished data)

Chlamydomonas under high availability of carbon dioxide suppress biomass accumulation

Chinnasamy et al.2009 *Int. J. Mol. Sci.*

Metabolic network models
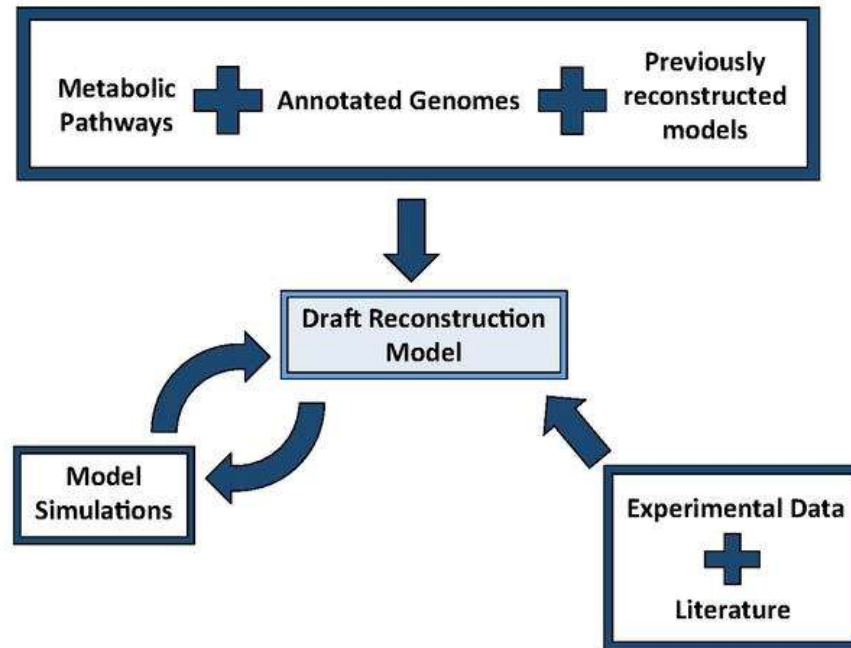
Allison Yaguchi, 21:54, 3 November 2012

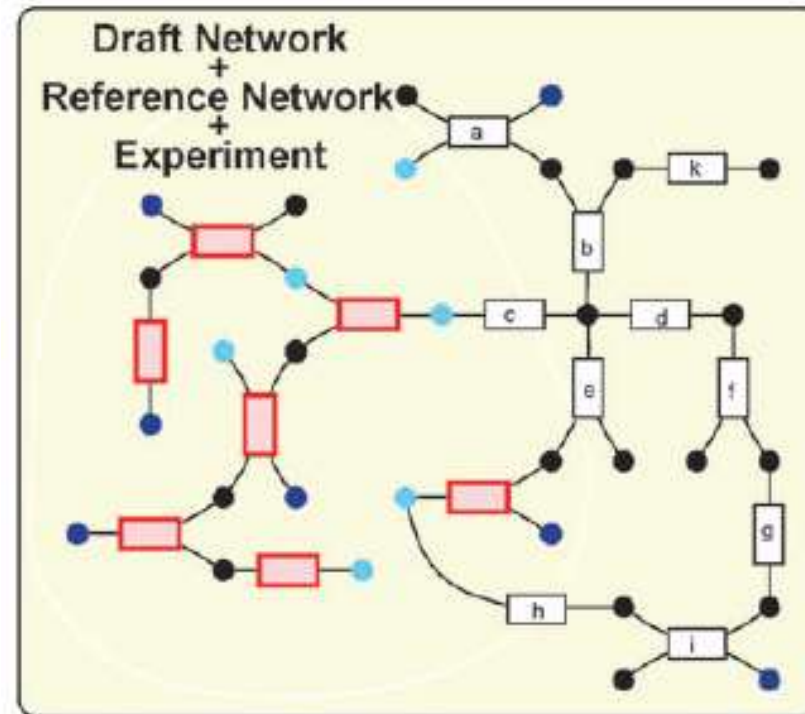| No. Reaction	Equation	LB	UB	KEGG RxnNum	Observation
2894	[c] : gdpmann + h2o + (2) nad --> gdpdm + (2) nadh + (2) h	-1000	1000	R00880	Found by homology
2895	[c] : gdpdm + (2) nadh + (2) h --> gdpmann + h2o + (2) nad	-1000	1000	R00880	Found by homology
2896	[c] : h + nadh + ru5p-D --> nad + dr5p	-1000	1000	R01524	Found by homology
2897	[h] : h + nadh + ru5p-D --> nad + dr5p	-1000	1000	R01524	Found by homology
2898	[c] : nad + dr5p --> h + nadh + ru5p-D	-1000	1000	R01524	Found by homology
2899	[h] : nad + dr5p --> h + nadh + ru5p-D	-1000	1000	R01524	Found by homology
2900	[c] : h + nadph + ru5p-D --> nadp + dr5p	-1000	1000	R01525	Found by homology

Understanding biomass accumulation by integrating experimental data into metabolic network models

- ● Metabolite
- ● Nutrient (experiment)
- ● Target (experiment)
- ☐ Reaction
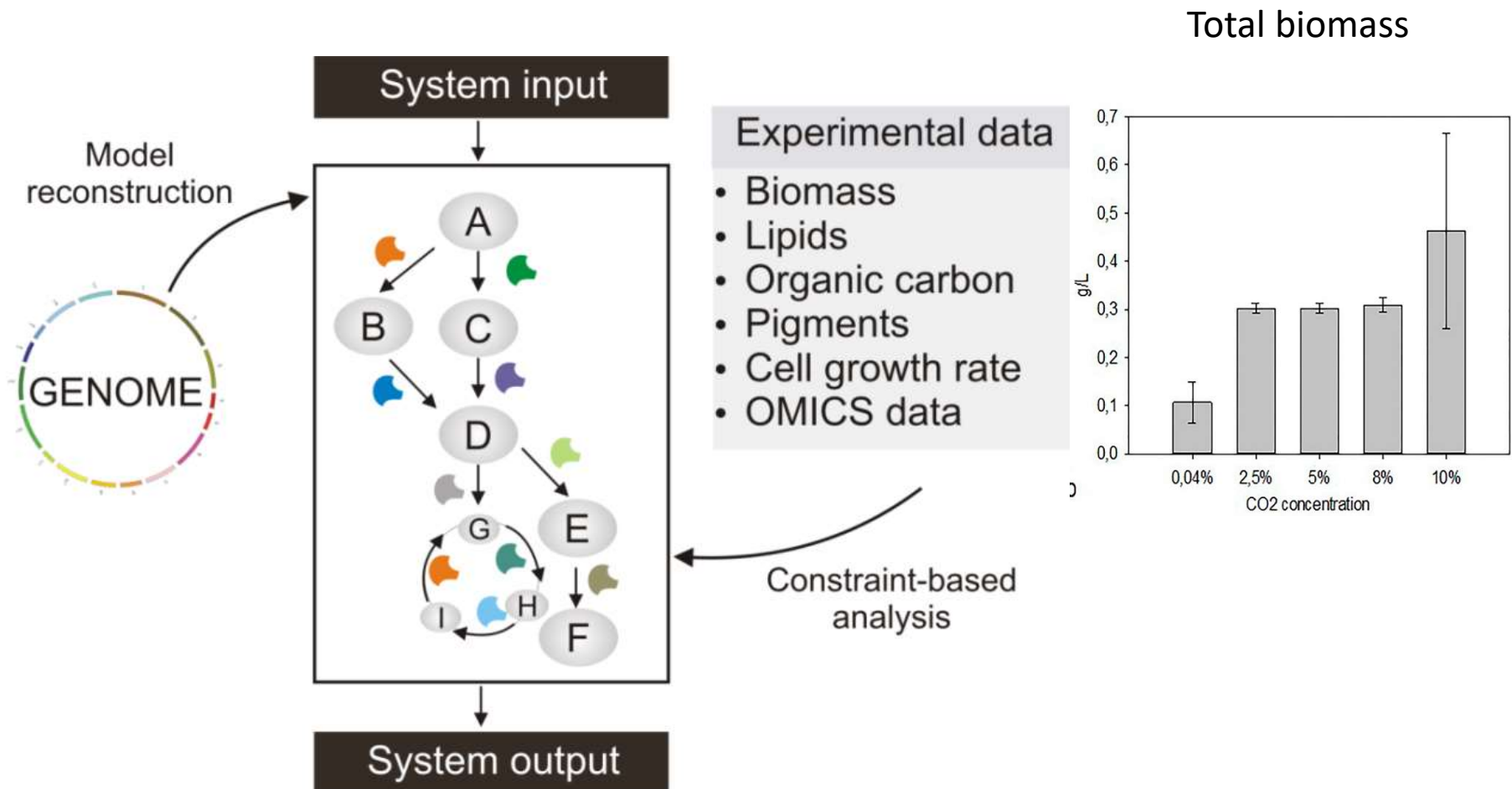
Experimental data
(total biomass)
Genome-scale network model

+

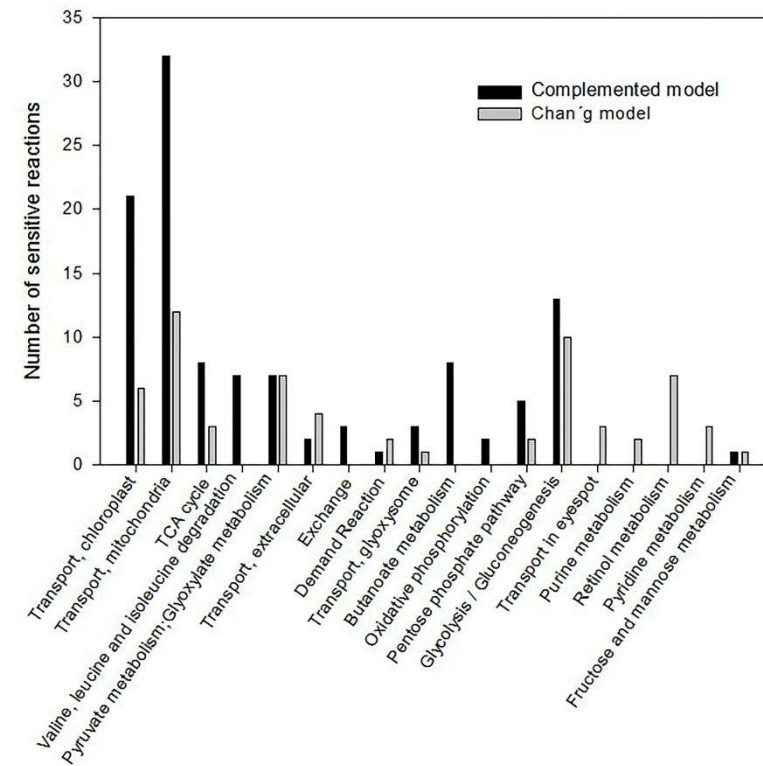Modeling and simulations
on reconstructed
metabolic network

Modified from *Mol.BioSyst., 2009, 5, 1889-1903.*

Melo D., Moncada R.-P., **Winck F.**, Gonzalez-Barrios A. (2014) In Silico Analysis for Biomass Synthesis under Different CO2 Levels for *Chlamydomonas reinhardtii* Utilizing a Flux Balance Analysis Approach. In Castillo, L.F.*, et al.* (eds), *Advances in Computational Biology*. Springer International Publishing, pp. 279-285.

Sensitive genes and reactions were identified based on In silico simulations which integrated experimental data

Total biomass

Winck, F.V., Paez Melo, D.O. and Gonzalez Barrios, A.F. (2013) Carbon acquisition and accumulation in microalgae Chlamydomonas: Insights from "omics" approaches, *J Proteomics*, **94C**, 207-218
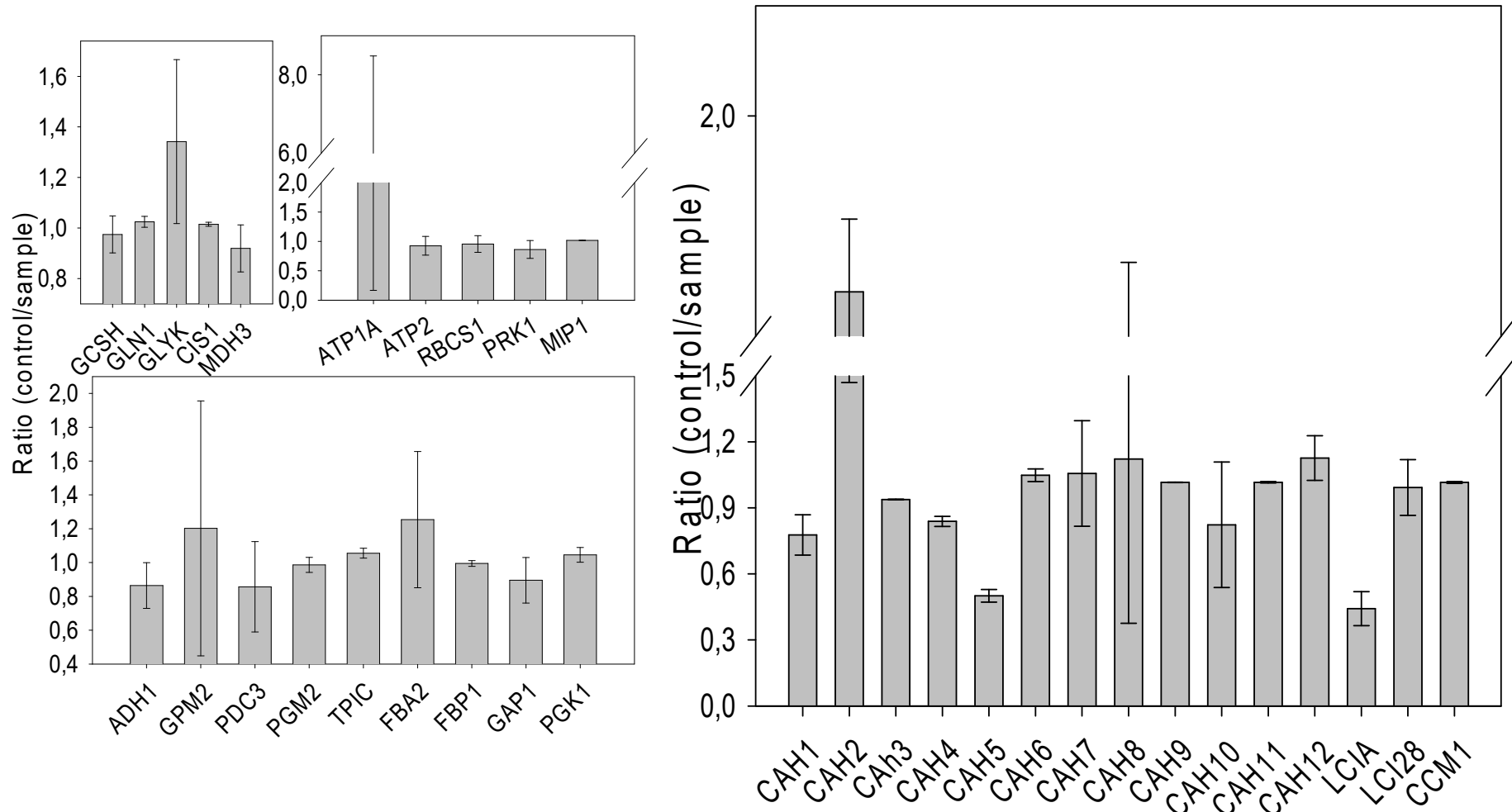
Detection of sensitive genes and pathways may indicate essential routes for biomass accumulation

Biological processes	Sensitive genes *
Transport, mitochondria	MITC14/ MITC28 / PTB8 / PTB7 / PRB1/ PRB12 / PTB4 / PTB2 / CRv4_Au5.213.g4507.t1
Phenylalanine tyrosine and tryptophan	AST4 / HIS5
TCA cycle/ CO2 fixation	ACH1/ IDH3 / SDH1 / SDH2 / OGD1
Valine, Leucine and isoleucine degradation	CRv4_Au5.s4.g11844.t1/ Crv4_Au5.s12.g3863.t1 / CRv4_Au5.s6.g13618.t1 / CRv4_Au5.s12.g3863.t1 / g1910.t1
Pyruvate metabolism; Glyoxylate metabolism	HYDA1 / MFDX / HYDA2 / PFL1 / ACK2 / AACK1 / ACK1 / PAT1 / PAT2 / CRv4_Au5.s6.g13230.t1/ CRv4_Au5.s2.g9723.t1
Alanine and aspartate metabolism; Glycerine, Serine and Threonine	AST3 / AST1
Carbon Fixation	AAT1 / AAT2 / MME3 / MME6 / MDH5 / MME2
Glycolisis, Gluconeogenesis, Valine, Leucine and isoleucine degradation	DLDH1 / PDC2 / PDH2 / ALSS1 / ALSL1 / PYK1 / PYK5 / PHG1 / GAP3/ GAP1 / PGM2 / PGM5 / PGM1B / PGK1 / TPIC / FBA1 / FBA2 / PGI1 / GPM2
Transport, Extracellular	NAR1.6 / NAR1.3 / NAR1.4
Pentose Phosphate Pathway	TAL1 / TRK1/ RPE1/ RPI1
Glycine, Serine and Threonine metabolism	Crv4_Au5.s10.g124.t2 / THD1 / SHMT3
Transport, Chloroplast	AOC6 / AOC5 / AOT7 / DAT1 / OMT1 / AOT5 / FBB13 / NAR1.5 / NAR1.2 / NAR1.1 / AAA3 / AAA1 / CRv4_Au5.s14.g5515.t1 / CRv4_Au5.s15.g5921.t1 / CRv4_Au5.g14736.t1 / MOT20 / MIP1 / MIP2
Butanoate Metabolism	CRv4_Au5.s7.g14479.t1 / CRv4_Au5.s16.g6952.t1
Oxidative Phosphorylation	NDA3 / NUO11 / NUO10 / NUO13 / NUO21 / NUO3 / NUO5 / NUO6 / NUO8 / NUO9 / IPY1 / IPY3
Propanoate Metabolism	PFL1
Nitrogen Metabolism	CGL77 / IBA57 / GCST

Winck, F.V., Páez Melo, D.O., Riaño-Pachón, D.M., Caldana, C., Martins, M., González Barrios, A.F. (2015) Analysis of sensitive CO_2 pathways and genes related to carbon uptake and accumulation in *Chlamydomonas reinhardtii* through genomic scale modeling and experimental validation (submitted).
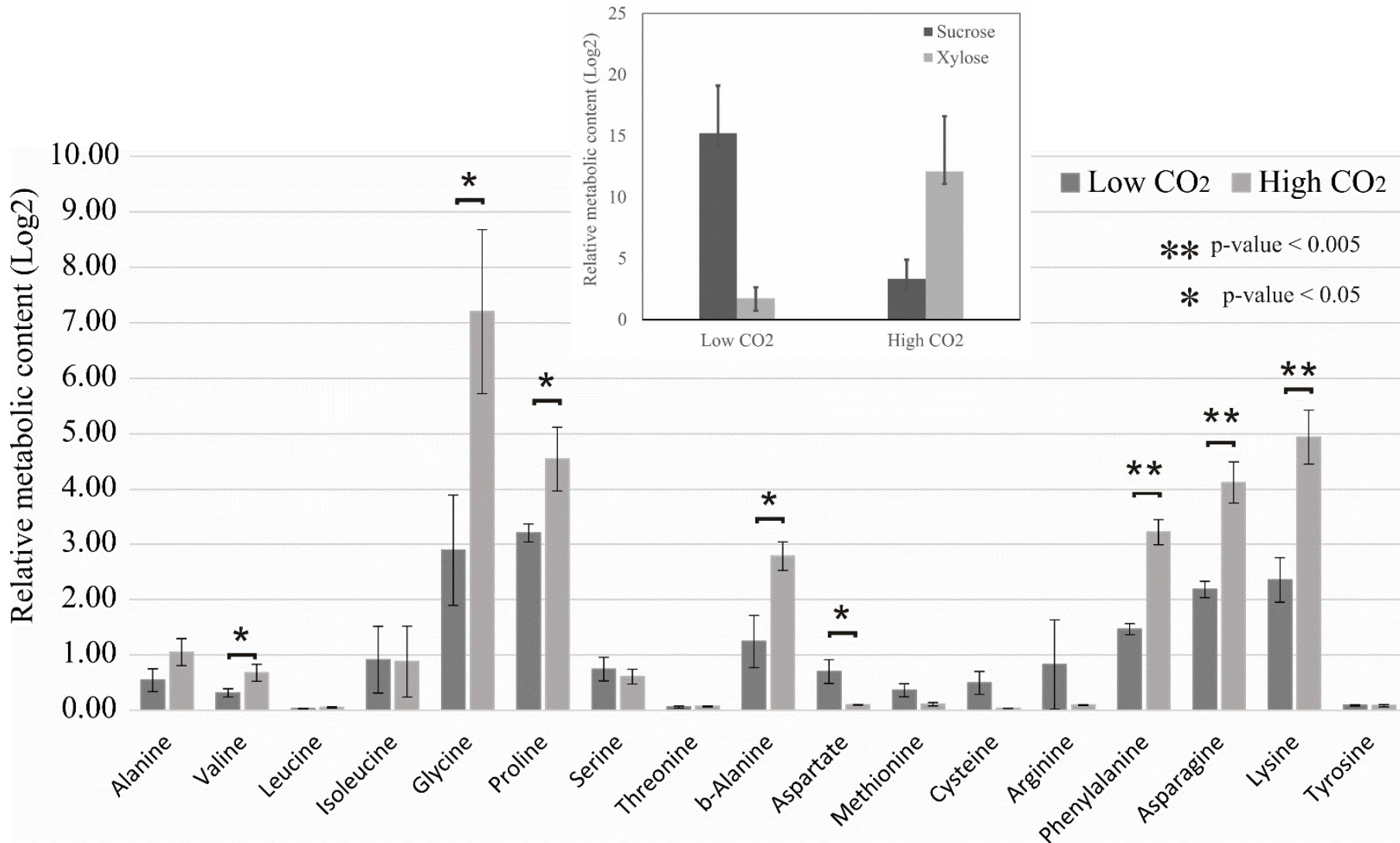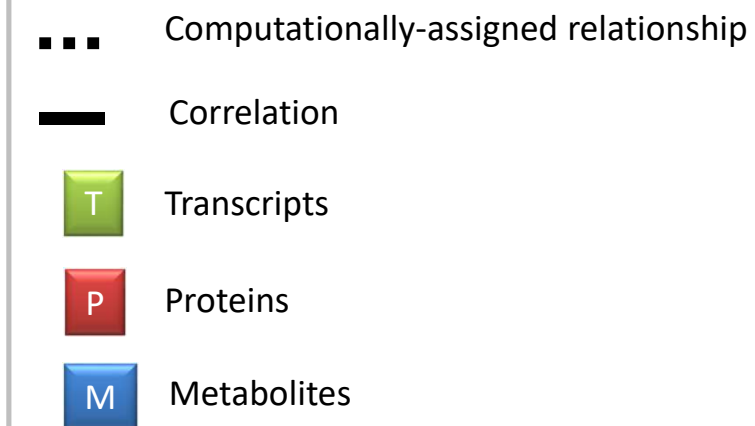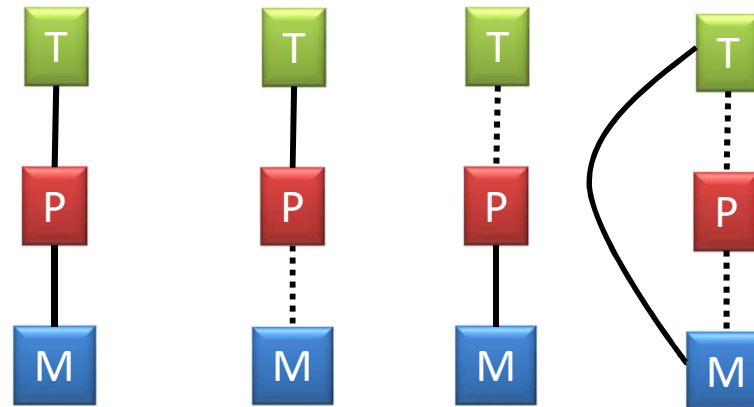
Gene expression analysis of sensitive genes validated candidates

Winck, F.V., Páez Melo, D.O., Riaño-Pachón, D.M., Caldana, C., Martins, M., González Barrios, A.F. (2015) Analysis of sensitive CO_2 pathways and genes related to carbon uptake in *Chlamydomonas reinhardtii* through genomic scale modeling and experimental validation (Unpublished data)

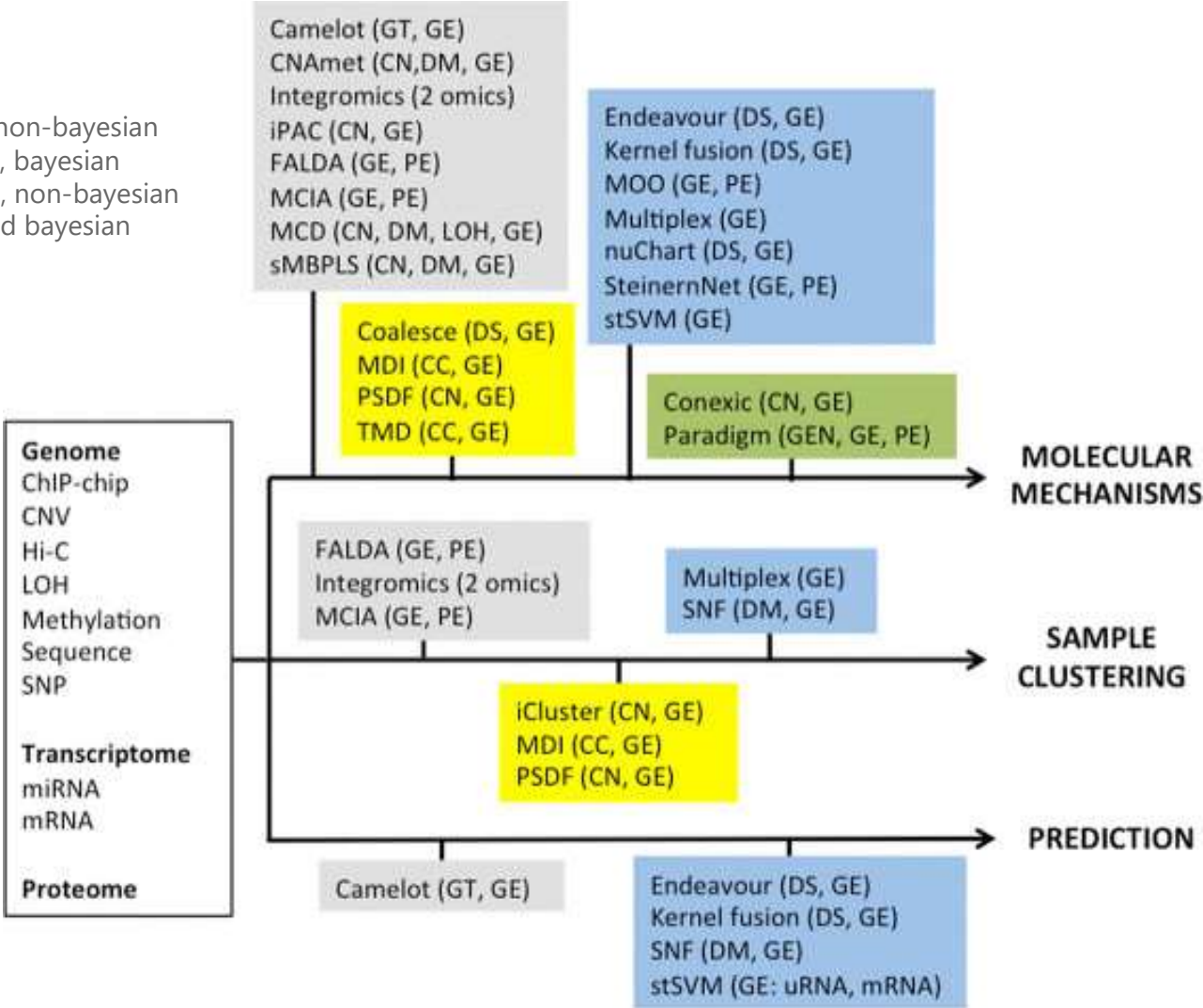# High carbon dioxide availability affected the metabolome

Inferring correlation between OMICS data

Legend:

... Computationally-assigned relationship

— Correlation

T Transcripts

P Proteins

M Metabolites

Methods and tools for integration of multi-omics data

Grey: network-free, non-bayesian
yellow: network-free, bayesian
*blue*: network-based, non-bayesian
green: network-based bayesian

Camelot (GT, GE)
CNAmet (CN, DM, GE)
Integromics (2 omics)
iPAC (CN, GE)
FALDA (GE, PE)
MCIA (GE, PE)
MCD (CN, DM, LOH, GE)
sMBPLS (CN, DM, GE)

Endeavour (DS, GE)
Kernel fusion (DS, GE)
MOO (GE, PE)
Multiplex (GE)
nuChart (DS, GE)
SteinernNet (GE, PE)
stSVM (GE)

Coalesce (DS, GE)
MDI (CC, GE)
PSDF (CN, GE)
TMD (CC, GE)

Conexic (CN, GE)
Paradigm (GEN, GE, PE)

Genome
ChIP-chip
CNV
Hi-C
LOH
Methylation
Sequence
SNP

Transcriptome
miRNA
mRNA

Proteome

FALDA (GE, PE)
Integromics (2 omics)
MCIA (GE, PE)

Multiplex (GE)
SNF (DM, GE)

iCluster (CN, GE)
MDI (CC, GE)
PSDF (CN, GE)

Camelot (GT, GE)

Endeavour (DS, GE)
Kernel fusion (DS, GE)
SNF (DM, GE)
stSVM (GE: uRNA, mRNA)

MOLECULAR MECHANISMS

SAMPLE CLUSTERING

PREDICTION

Integrative OMICS helps to improve genome annotation

Figure 1: Integrating the mRNA sequencing and peptide sequencing for proteogenomic discovery and genome annotation.
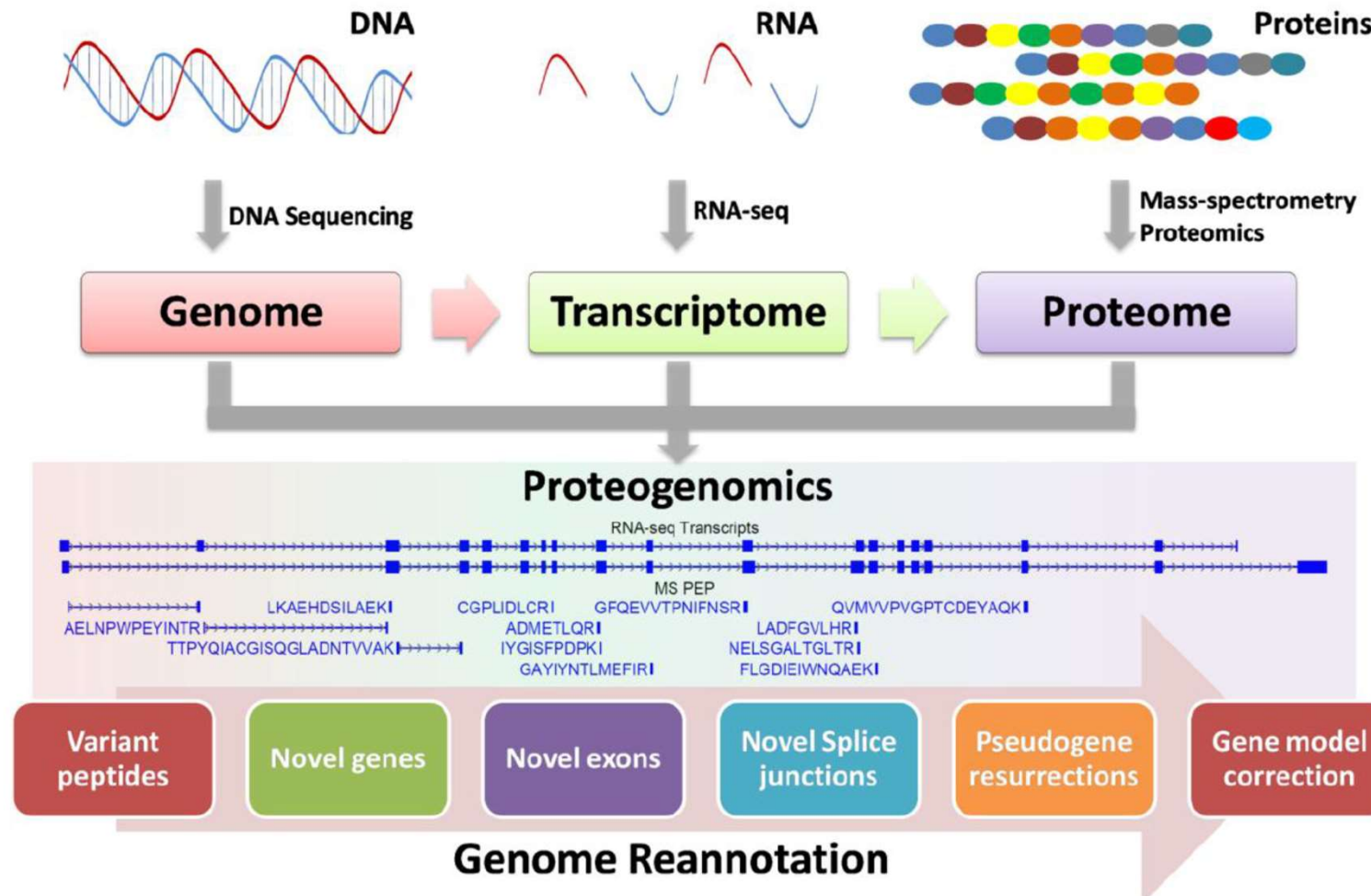
Design of proper devices for accelerating microalgae synthetic biology

From

Flasks

To
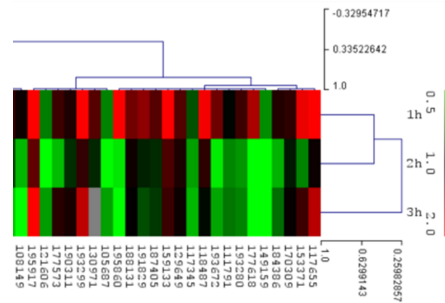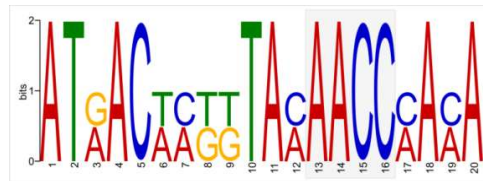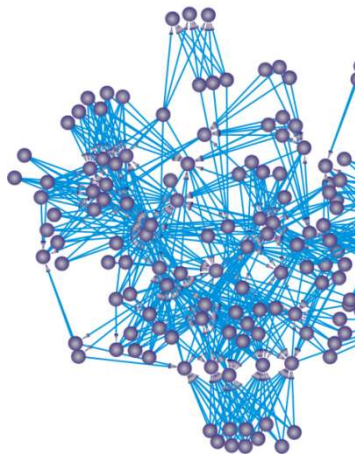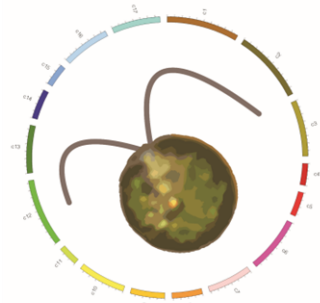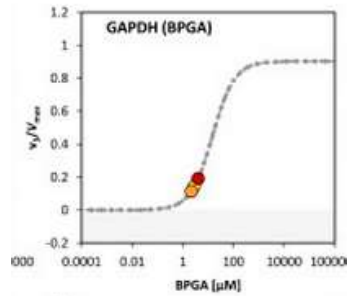
Farms

$$\iota^{(l \rightarrow k)} = 1 - \frac{\sum_{i=1}^{n-2} \sum_{j=i+1}^{n-1} w_{ij} \Theta[(g_{j+1}^{(k,l)} - g_i^{(k,l)})(g_i^{(k,l)} - g_j^{(k,l)})]}{\Delta}$$

International PhD in Bioenergy (USP, UNESP, UNICAMP)

http://genfis40.esalq.usp.br/pg_bio/

More info about microalgae biology and sustainable applications is freely available

Acknowledgments

Open positions for graduate students and postdocs
Contact: winck@iq.usp.br

Laboratory of
Regulatory
Systems
Biology
Institute of Chemistry
USP - BRAZIL

Follow us at Facebook:
https://www.facebook.com/regulatory.systems.biology/