

Biologie des systèmes

Biologie théorique et
bio-informatique

GB4 2023

Christophe Bécavin

Maître de conférences UCA



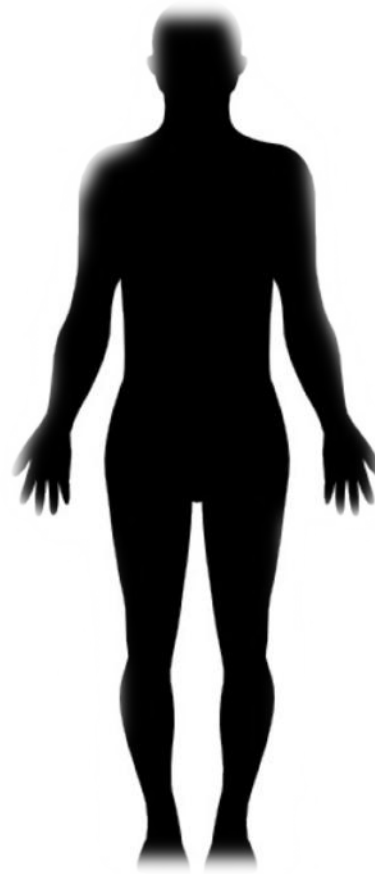
Biologie des systèmes

- **L'approche systémique en biologie**
- **Reconstruire un réseau biologique**
 - Les différents types de réseaux
 - Reconstruction directe
 - Reconstruction indirecte
 - Les obstacles à la reconstruction
- **Analyser un réseau biologique**
 - Théorie des graphes
 - Topologie des réseaux
 - Les réseaux sans échelles
- **Utiliser un réseau biologique**
 - Etat de l'art des réseaux biologiques les plus étendus
 - Bases de données en biologie des systèmes
- **Modéliser un réseau biologique**

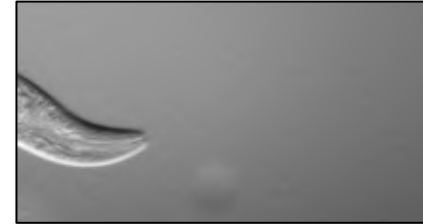


Du Phenotype au Genotype

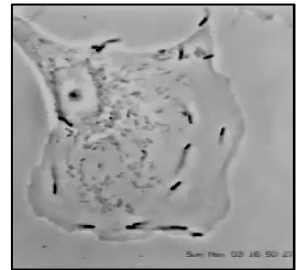
Comment caractériser un organisme vivant ?



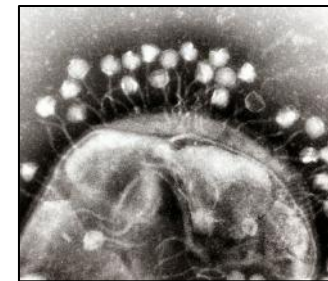
10^{-3} m



10^{-6} m

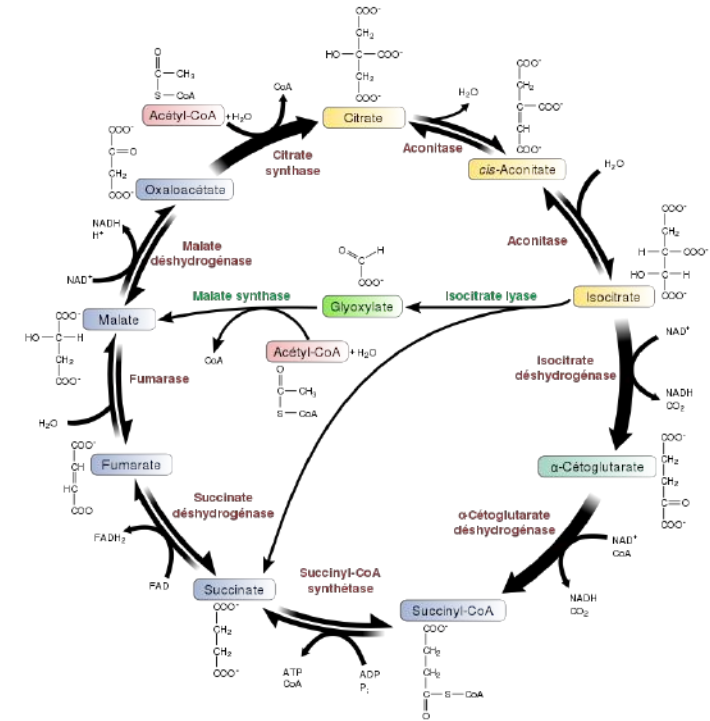
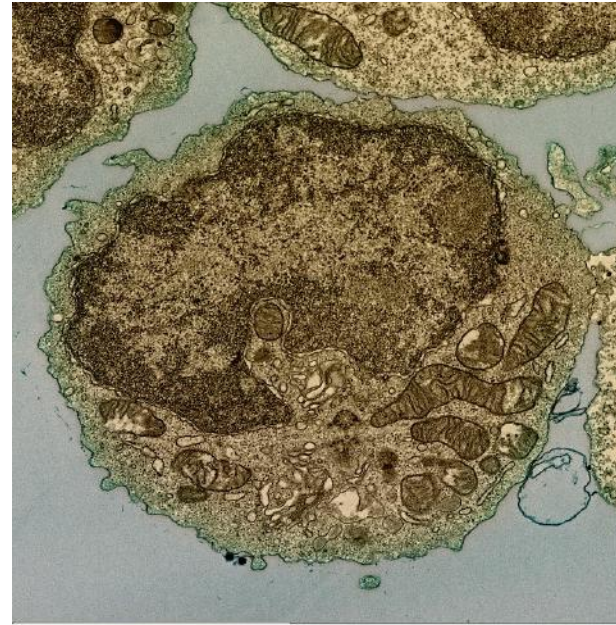


$5 \cdot 10^{-9}$ m



Le (les ?) phénotype

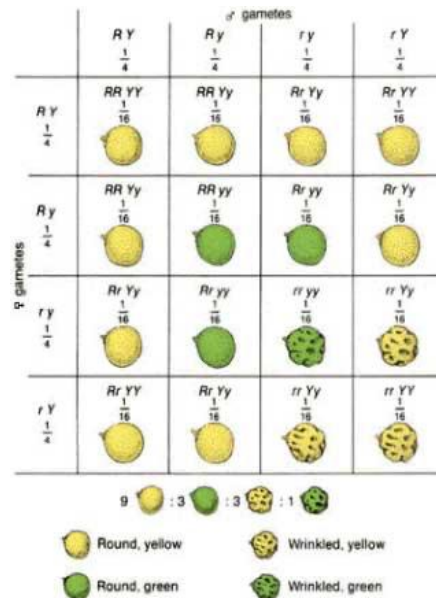
« Le phénotype c'est l'ensemble des traits observable » *Wikipedia*



Le gène comme élément héréditaire expliquant un phénotype

Gregor Mendel (1822-1884)

Le phénotype est héréditaire



Walter Sutton (1902)

Séparation des paires de chromosomes pendant la méiose

↓
Le génotype

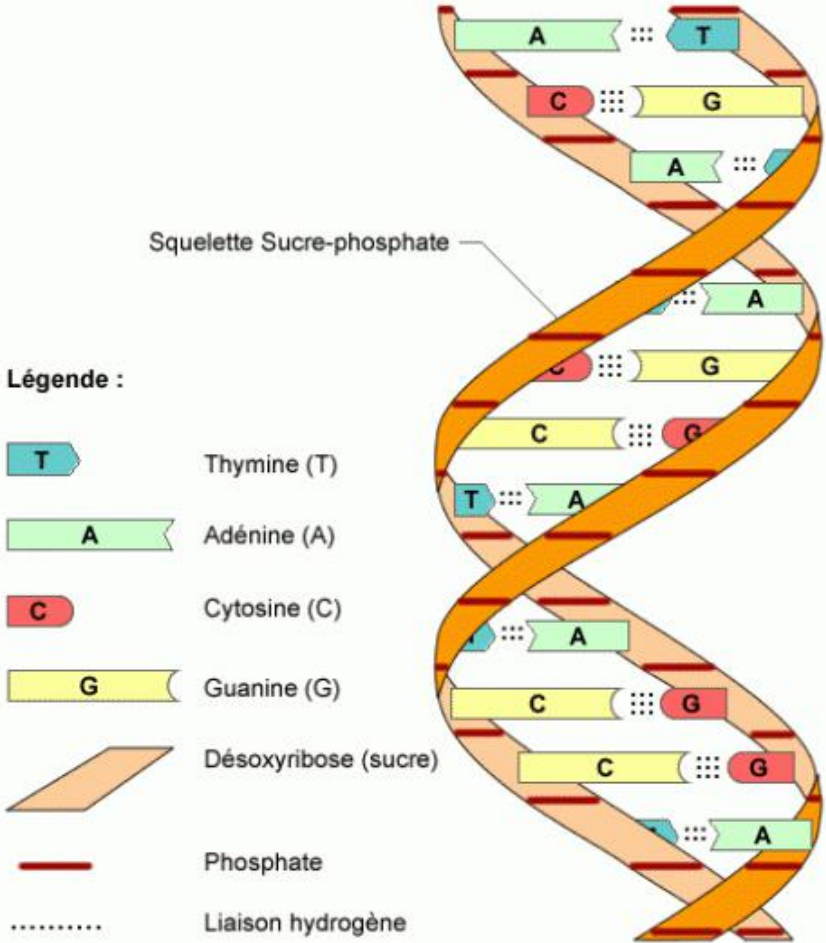
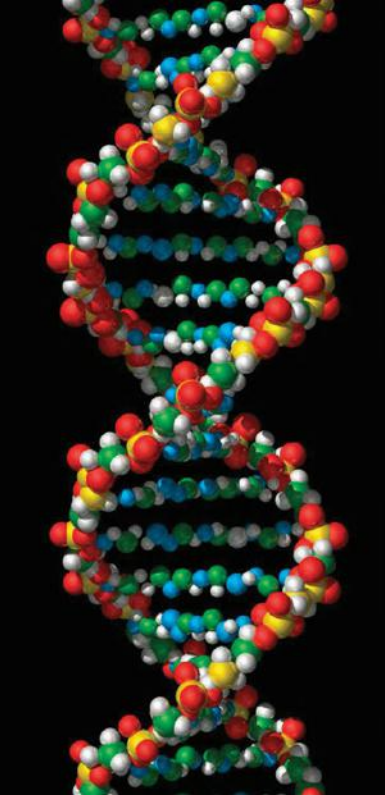
Thomas Morgan (1910)

Théorie chromosomique de l'hérédité

↓
Le génotype et le phénotype sont liés

L'ADN comme support DU GENOTYPE

Franklin, Watson, Crick (1953)



```

AGGGTCATTAATAATATATAAAGATCTATATAGAGATCTTTTTATTAGATCTACTATTAAGGAGCAGGATCTTTGGTGGAT
AAGTAAAAATGATCAACAAGATCATGCGATTGAGAAGGATCAGATCGTGTGATCAACCCTGATCTGTTCAAGGATTAG
CTGGGATCAAAAACCTATGTTTATACACAGCCACTTGGGATCTAAAACCTGTTATATGGATAACTATAGGAAGATCACCG
GATAATCGTATAGTTATCCACATGAGATTTGATTGAAAAAGCATCAATCAATTTTTCTACTACCGTTAAATTTATCCACA
ATCCAAAAAAGAGCGGCATTAAGCCGCTCTGATGGAATAGGTCAATTTTAGAAGCGATTGATGACGCGTTTGAGCC
AAGCTTCAGCGGCATCTTCAGGCCTGGGTGCTCTTGACATCGATTGTAAGCAGTTGCCAGAGGTTTACACCAATA
TCCCCCAGCAGCTGATAGGCATGTTTACCTGCGCGCAGAAAGTATCGTAGCTTGAATCACCATCCGACACCGGCATA
ACGTAGTGCAGAGGTTATCGGTGGTGTATTCTGCAGAGCCTGAATAAAGGGCTGGATATTATCCGGGTACTCACCAGCCC
CGTGGGTTGAGGTGATGATCAGCCAAGTCCCTTAGCAGGGATCTCACTCATGTTGGGCTGGTTATGAATTTGGTGTCA
AAGCCTTGTCTTCAGTAAATCACTCAGGTGGTCACCCACATATTCGACCCGCCTAGGGTGTCCAGTAATGATATG
AATCATAGCGTTACTCTATTTCCCAATACAGAATGATGAAAAATGCGGCCAAGCAGATCATCGAGCTGAACCTCGCCG
TAATTTCTGTAAGGTGTTGCTGGGCTACCGCAGCTCTTCGGCGAGGATTTCTCGGCCATATAGCTTCAAGTTGTTGC
TGCCCAATCGTAAGTGTCTCGCGCTCGCTAGGGCATCGAGATGACGGCGGCGTCCATAAAGCCACTTCTCGATT
GCCTGAAAAACCCATGCACTCTTTGAGGTGCTGACGCAAGGCATCGACCCTTGGCTGTTTTGGCTGATAGGCGGATCA
AGGTGGGTTGATTAACATGGCAGATCCCAAGGGCTCACAGTTTGTATCGGCTTTATTACGGATCACAGTATCCCAATA
TTCTCTGCGAGTTTGTCAACAAAATCAGGCCAGATGTCCTGTGGATCGGTGGCTCTGTGGTGGTCCATCGACCATAAA
CAGTAGCGATCGGCTTGGCGGATCTTCCCATGCGCGCTCAATACCAATTTTTCTACCGCATCAGAAGCGTCTCGTA
GTCCCGGATATCGATGATGTCAGCGGCATCCATCAATATGGATATGCTCACGGAGAACATCACGGGTGTACCGGCA
ATGTCGGTACGATGGCAGACTCTTACTGAAAGCGCATGAGTAGGCTCGATTACCGCATTAGGACGCCCAAGCAT
CACCCCTTCACTCCCTTCGCGCAATAAGCGCCTTGGTTGGCTCACGGCGACTGCGGCAAGTTATCTATGATGTTTT
CGAGTACGCGGAAACCTTACCATCGCCAGAAAATCGATCTCTTCTTGGGAAATCAATTCGGCTTCAACATAGATG
CGCAGGTGAATCAGCGATTCCACCAAGGTATGGATGCGTTTAGAAAACCTCGCCTTGCAGTATTGACGCGCGGATTTCCG
GGCTTGTCTCAGAGCTGGCATCAATCAGGCTCGCGATGGCTTCGCTTGGGTTAAATCCATCTTGTCAATTGAGGAAAGCC
GTTCTGAGAATTCACGGGACGGGCTGGCGCCTCTTAAATCTGCAAAAACGGCGGATCAGCATATCCATGACGACC
GGCCACCGTGACCTTGCAGCTCAAGCACATCTTACCCTGAAATGAATGAGGATTGGGAAAAACAGCGCAATGCCCTTG
ATCGAGCTGTTGGCCATCTTCACTCGGTGAAAGGCAAGTATTCGGCATAGCGGGGCTGAGCGTGCCTGACGTACTGCT
GCAGCAGTGGGCAAGCAGTGGGCTGATACACGAATAATGCGACACCACACGGCGGGTGGCTAGCTTGGCGGACA
ATGGTATCTGTTGATAGTGTACTGAAACAGGATTAATAGCGCCATGTAATCAGCAGCAACAAAAAGGCGACCT
TTTGGCCGCTCTTTATTACTCAATCAAACTACTTGGAGTGAAGCCTTTTTTCTAGCGCTTTGATAGTCAGCGTTT
GCTGGATTAGCGTTACGATGTCGACACCAACAGTACAGAACCAGACCTGATGGGAACACAGGAAGAAAGTGAAC
ATGACTGGCATGAAGGTCTATGATCTTCTGTCATGGATCGGTGATCGTGGTGGGCTCATCTTGGATACGCAACAT
GCTCGACCCATCAGCAGTGGCAAGATGATGAAAGTCTTGGCGGCAAGTATGAAATCAACAAAGAAATGGAGGAT
GAGCAGCTCTACCGACTCATCAGTCCCAAGTACAGCCCAATGAAATCGGCATTTGCAAGGATAGGAGAGACAGCCA
CCCAGTGGGTTACTTCTCTTTTATACAGCTCCATCATCTTGGCTCATGGCTGGCGATCATCGCCATCCGCTC
ACGCATCGCTGAGTTAGGTTGACGATGCGCATTTTCCGCTTGAAGTGTACTGTGCTTGGTCAAGTGGTACATCG
CACCCGAAACAATAAAGTTAAGCAGATGATGGCCACACCCAGTTGCCAACAAAGGTTGTAATCACAGACAGCAACAG
TGACGTGCTTAGCAATGAACACAAACCAACATAATCAACCCAGATCAAGATTAAGGGGCTGTTGGCGCATTTGGCT
TTGCAATTTGGGCTTCCATAAGGTTGCTTCAAACCTCTTACCGCTTGAATGCTTGGTAAAGTGGTTTGGTGGTACGGA
TACCATATCACCTAAATTTAGTAAACGCGTGAAGTAAAGTGTGCTTGGCGCTCATCAGTGGGATCATGCTGGAAGCG
AAGTAGTGTGGATCATCGGACCCAGTTTACCCTGATGTTGGTCACTGAGGTTGCGATCTTTCATCGTGGTCCAA
GCTGACTTCTTGAACGCACATCGCTGTTGATGAAGCACCACCGGTAAGTTGGCATGGCAGGTTGCCACCAAGAT
    
```

Du genotype au PHENOTYPE PAR la regulation genetique

Genetic Regulatory Mechanisms in the Synthesis of Proteins
F. Jacob, J. Monod, J. Mol. Biol. 1961

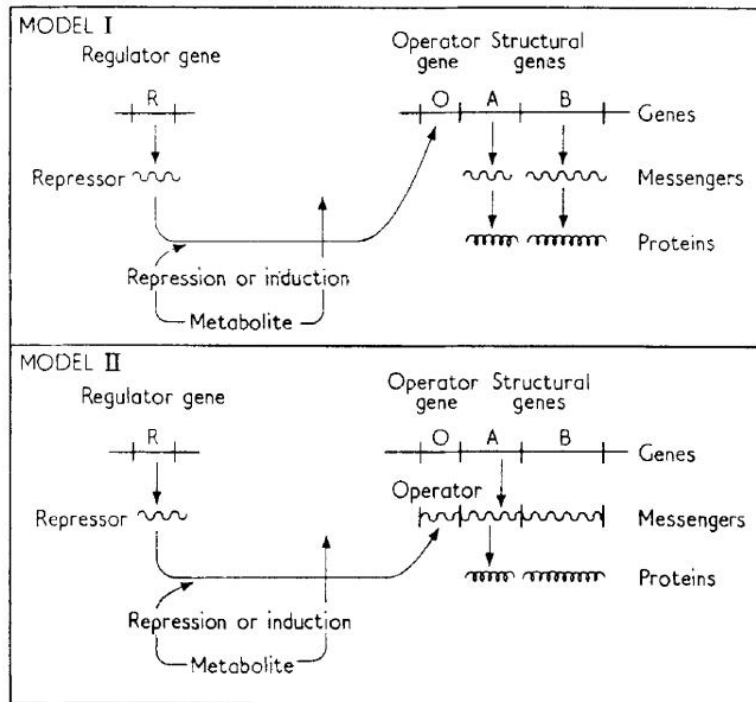
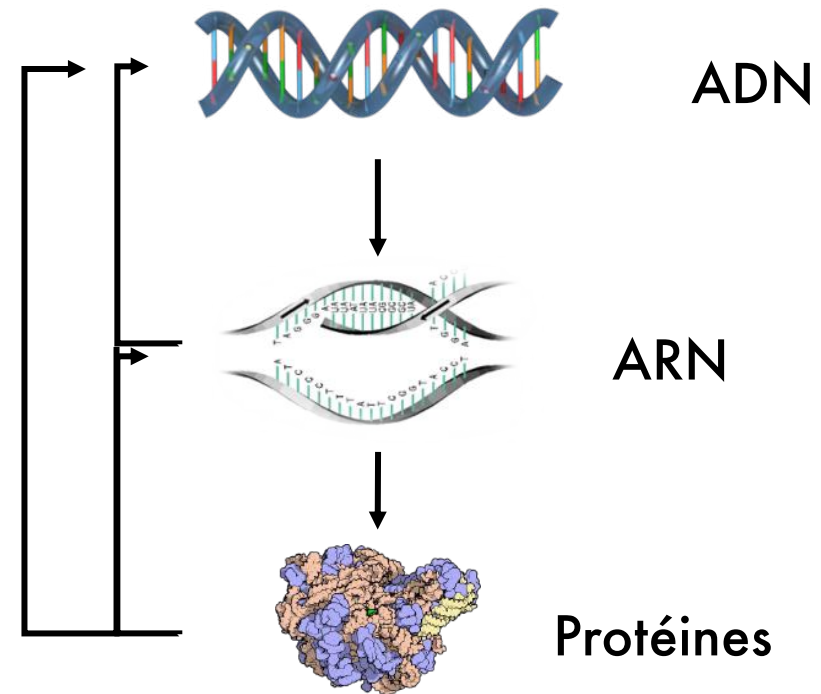


FIG. 6. Models of the regulation of protein synthesis.



Décoder le Genotype

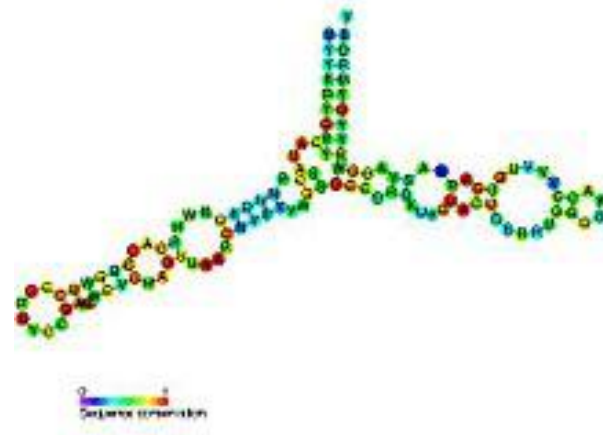
Frederick Sanger (1918 – 2013)

Protéine

Insuline de bovin
1955

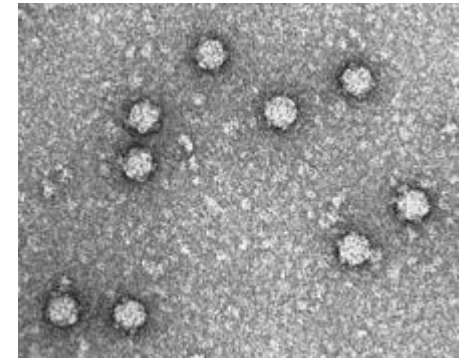
A chain	B chain	
Gly	Phe	1
Ile	Val	
Val	Asn	
Glu	Gln	
Gln	His	5
Cys	Leu	
Cys	Cys	
Ala	Gly	
Ser	Ser	
Val	His	10
Cys	Leu	
Ser	Val	
Leu	Glu	
Tyr	Ala	
Gln	Leu	15
Leu	Tyr	
Glu	Leu	
Asn	Val	
Tyr	Cys	20
Cys	Gly	
Asn	Glu	
	Arg	
	Gly	
	Phe	
	Phe	25
	Tyr	
	Thr	
	Pro	
	Lys	
	Ala	30

ARN



5S rRNA
120 nucleotides
1965

ADN



E. Virus Φ X174
5386 nucléotides
1977

“Sanger sequencing method”
Mitochondrie 16569 bp
Bacteriophage λ 48502 bp
1977

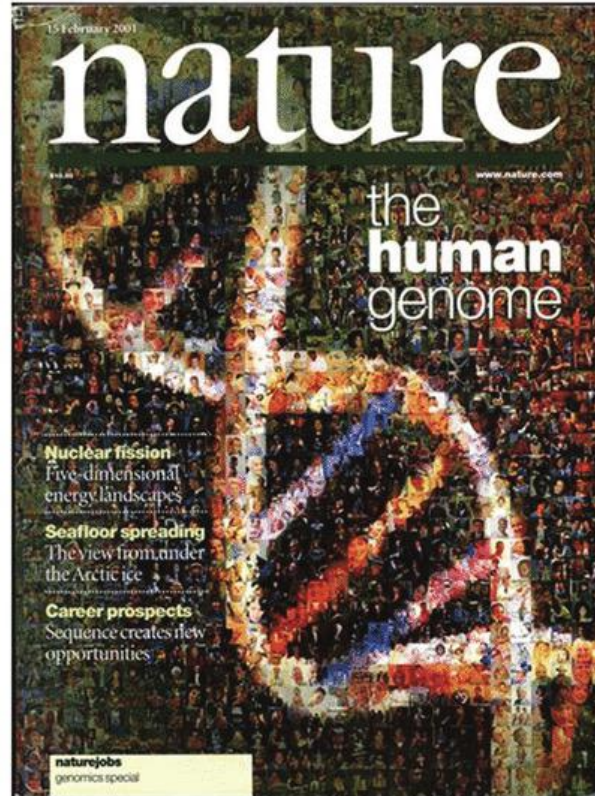
Le séquençage du génome humain

Human Genome Project

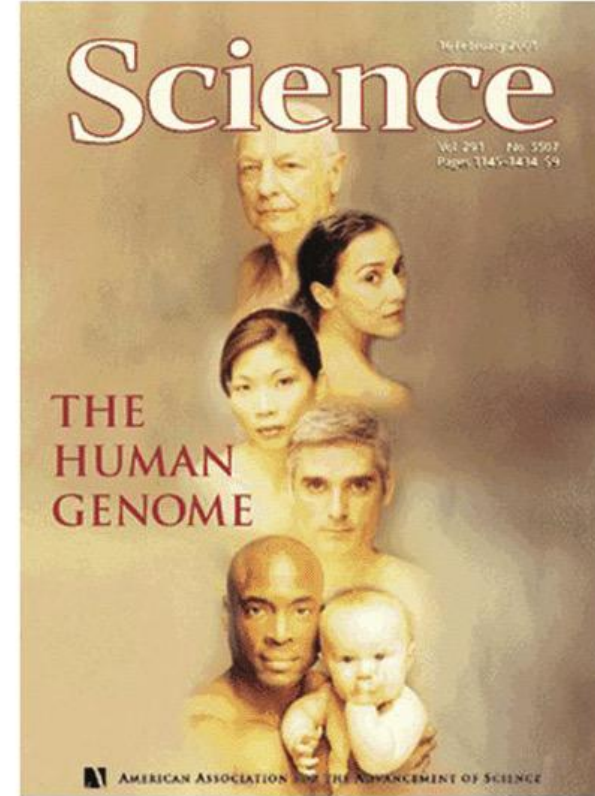
Lancé en 1985
par le gouvernement
américain

1989-2001

5 milliards de dollars
20 Instituts
6 pays



Février 2001



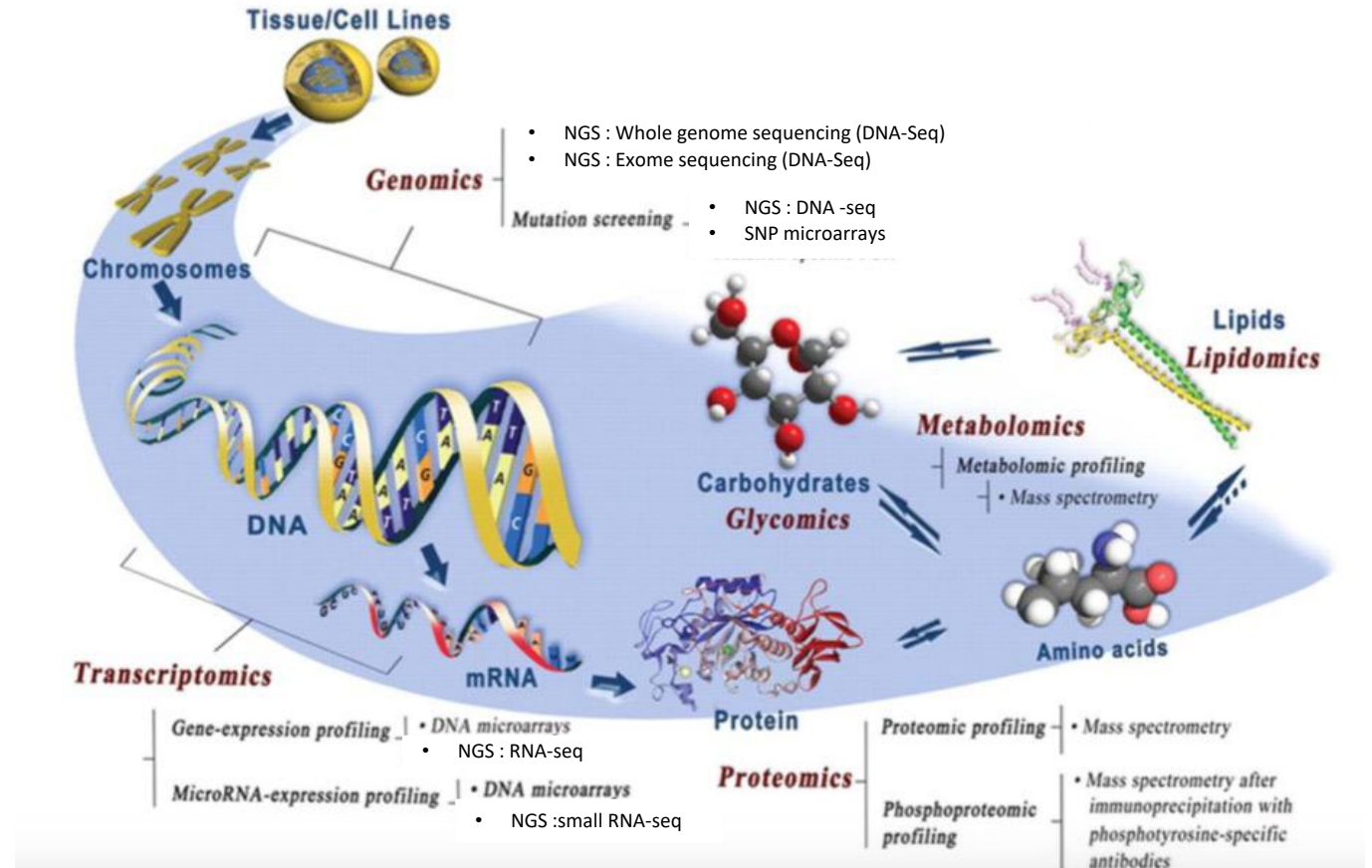
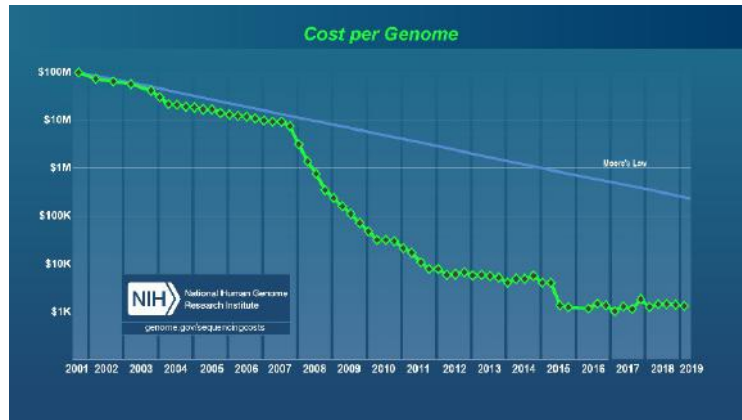
Celera Genomics
(J. Craig Venter)

1998-2001

300 millions dollars
(Shotgun sequencing)

Du Phenotype au genotype

La révolution omique



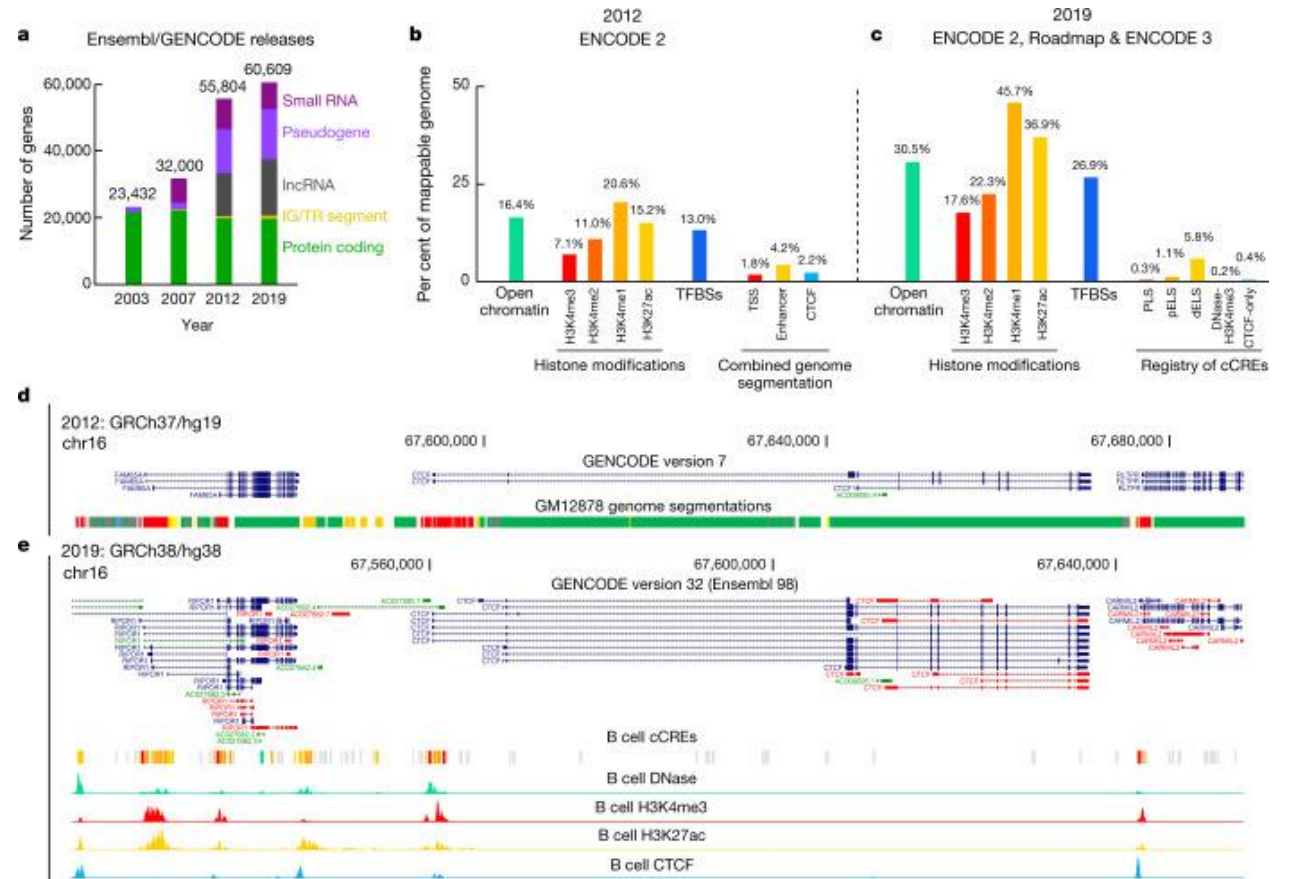
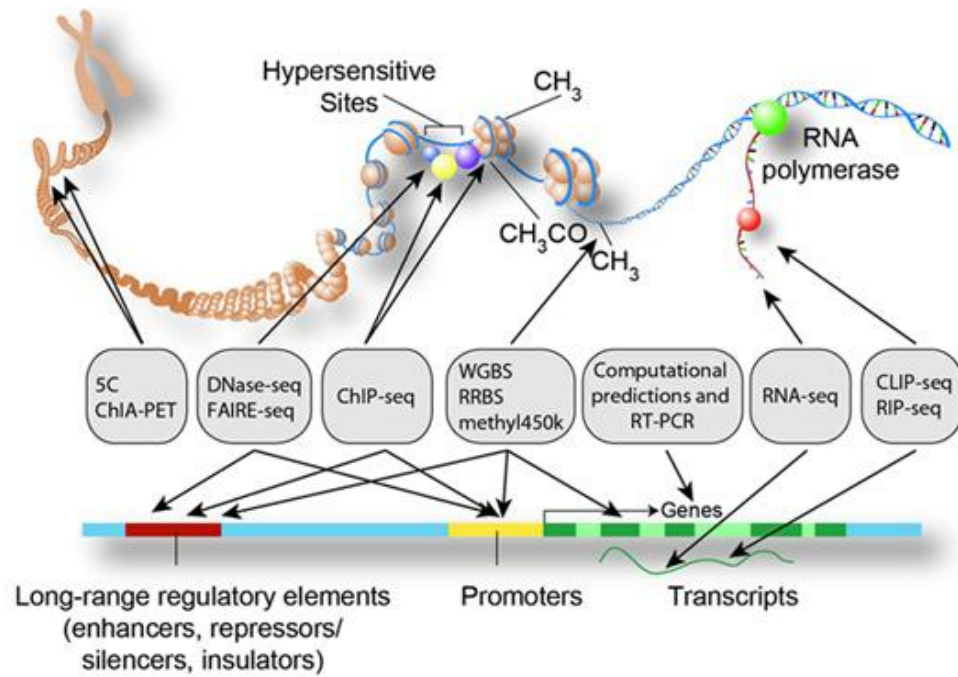
Different Generations of Sequencing

- First Generation**
 - 1972: Sanger started work on DNA sequencing
 - 1977: Sanger developed Di-deoxy chain termination method of DNA sequencing
 - 1977: Maxam and Gilbert developed chemical degradation method of DNA sequencing
 - 1977: First DNA based genome sequenced (φX174 bacteriophage)
 - 1995: First bacterium *Haemophilus influenzae* was sequenced by shotgun method
 - 1996: Applied Biosystems developed automated DNA sequencing based on Sanger's method
 - 1996: First eukaryotic genome (*Saccharomyces cerevisiae*) was sequenced
 - 2001: First human genome draft was published by two different independent teams
- Second Generation**
 - 2005: First NGS platform released Roche 454 GS-20
 - 2006: Introduction of second NGS platform –Solexa Genome Analyzer
 - 2006: Initiation of 1000 genome project
 - 2007: Introduction of Roche 454 GS-FLX & ABI-SOLID sequencer
 - 2008: Development of Illumina GA-II
 - 2009: Introduction of Roche 454 GS-FLX Titanium
 - 2010: Introduction of Roche 454 GS-Junior
 - 2011: Introduction of SOLID 5500 W & Illumina MiSeq
 - 2012: Introduction of Illumina HiSeq
 - 2013: Introduction of SOLID 5500xl W & Illumina MiniSeq
 - 2014: Introduction of Roche 454 GS-Junior+, Illumina NextSeq 500 & Illumina HiSeq X Ten
 - 2017: Introduction of Illumina iSeq 100
- Third Generation**
 - 2008: Development of first commercial platform of third generation technology i.e Helicosep by Helicose Biosciences
 - 2010: Ion Torrent released the Personal Genome Machine (PGM)
 - 2011: Introduction of PacBio RS C1/C2
 - 2012: Introduction of PacBio RS C2 XL & PacBio RS II C2 XL, Ion Torrent released Ion Proton
 - 2013: Introduction of PacBio RS II C2 XL
 - 2014: Introduction of PacBio RS II P5 C3 & PacBio RS II P6 C4
 - 2015: Introduction of Ion S5/S5XL 520/530/540
 - 2016: Introduction of PacBio sequel
- Fourth Generation**
 - 2014: Release of MinION platform by Oxford Nanopore Technologies
 - 2017: Release of ProMethION, GridION & SmidgION X5 platforms by Oxford Nanopore Technologies
 - 2018: Commercialization of ProMethION platform by Oxford Nanopore Technologies

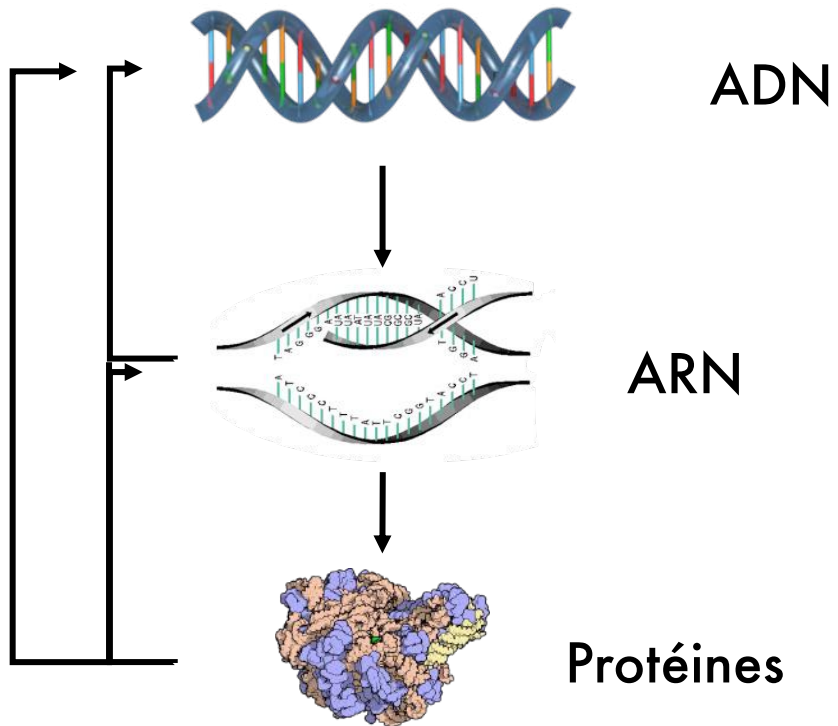
Wu R.Q., J. dent. Research, 2010

Le Projet ENCODE

Encyclopedia of DNA Element
2003 - Présent



Du Phenotype au Genotype Qu'avons-nous appris ?



- Il n'y a pas que 3 échelles omiques de régulation
- Les gènes ne sont pas les seuls éléments du génome
- Importance des parties non-codantes
- Les modes de régulation sont quasi-infinies

La biologie des systèmes

LacZ operon version 2021

https://www.genome.jp/kegg-bin/show_pathway?ko00052+K01190

Genetic Regulatory Mechanisms in the Synthesis of Proteins
 F. Jacob, J. Monod, J. Mol. Biol. 1961

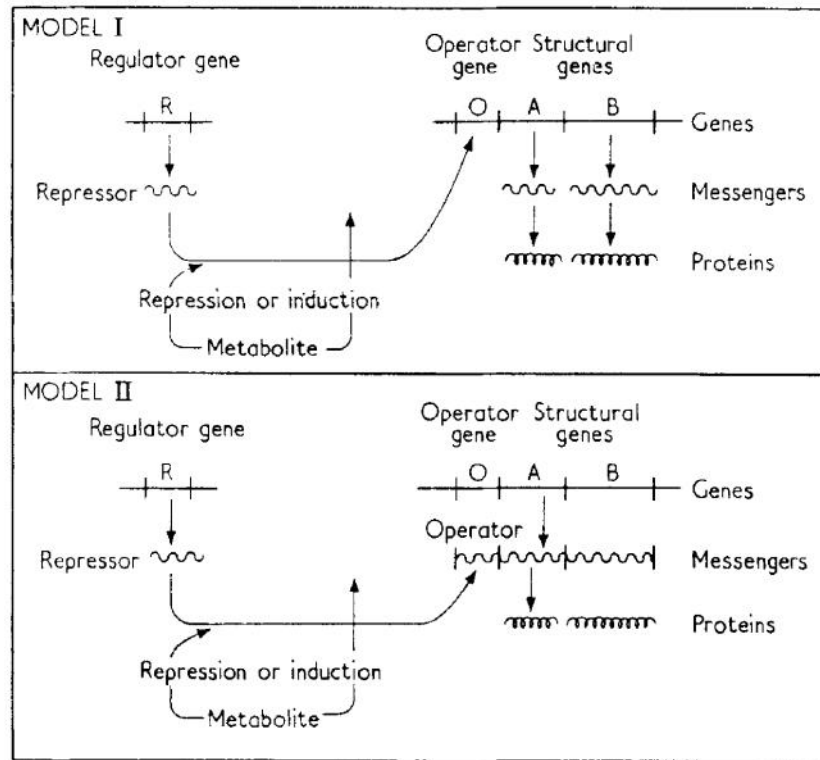
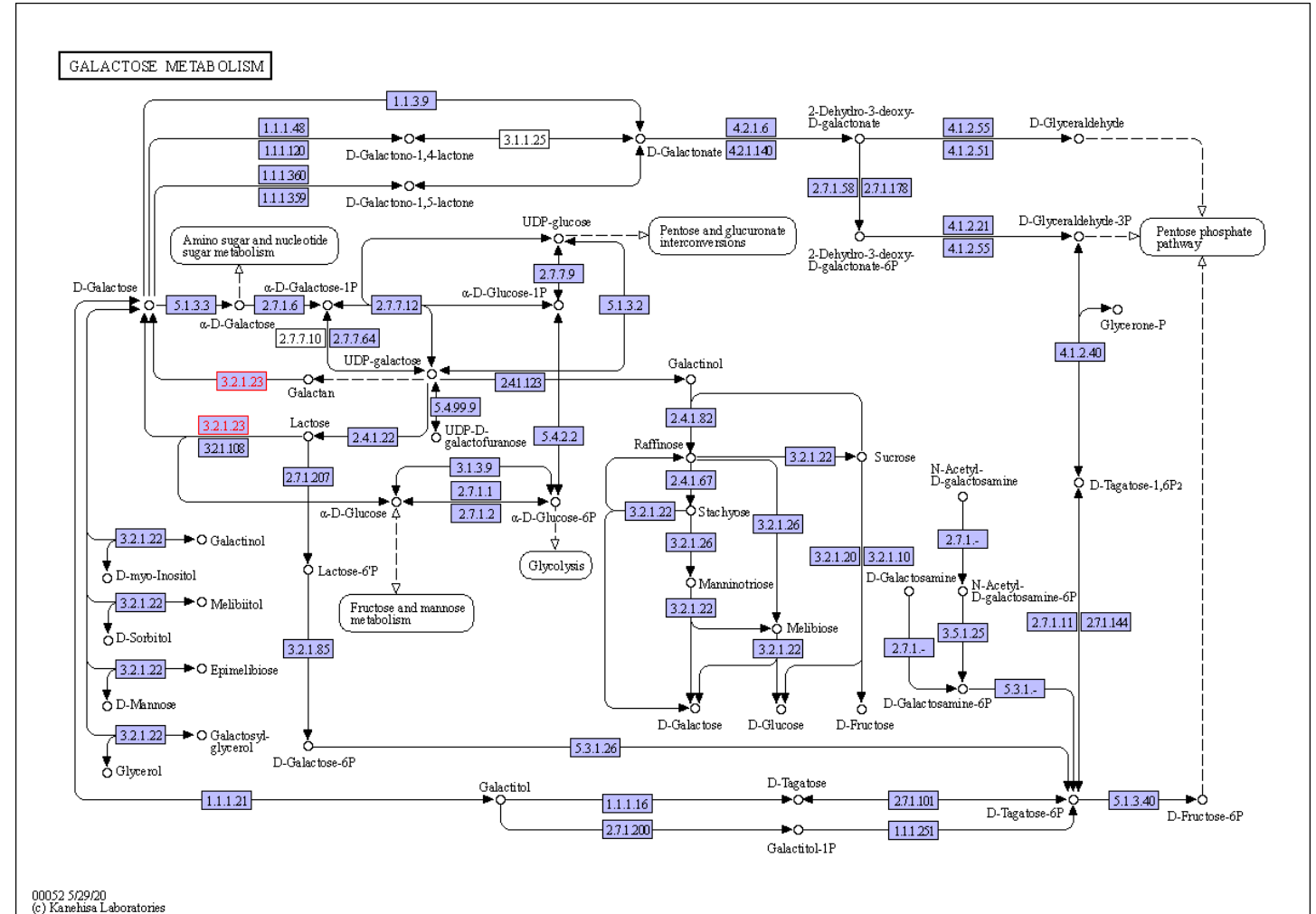
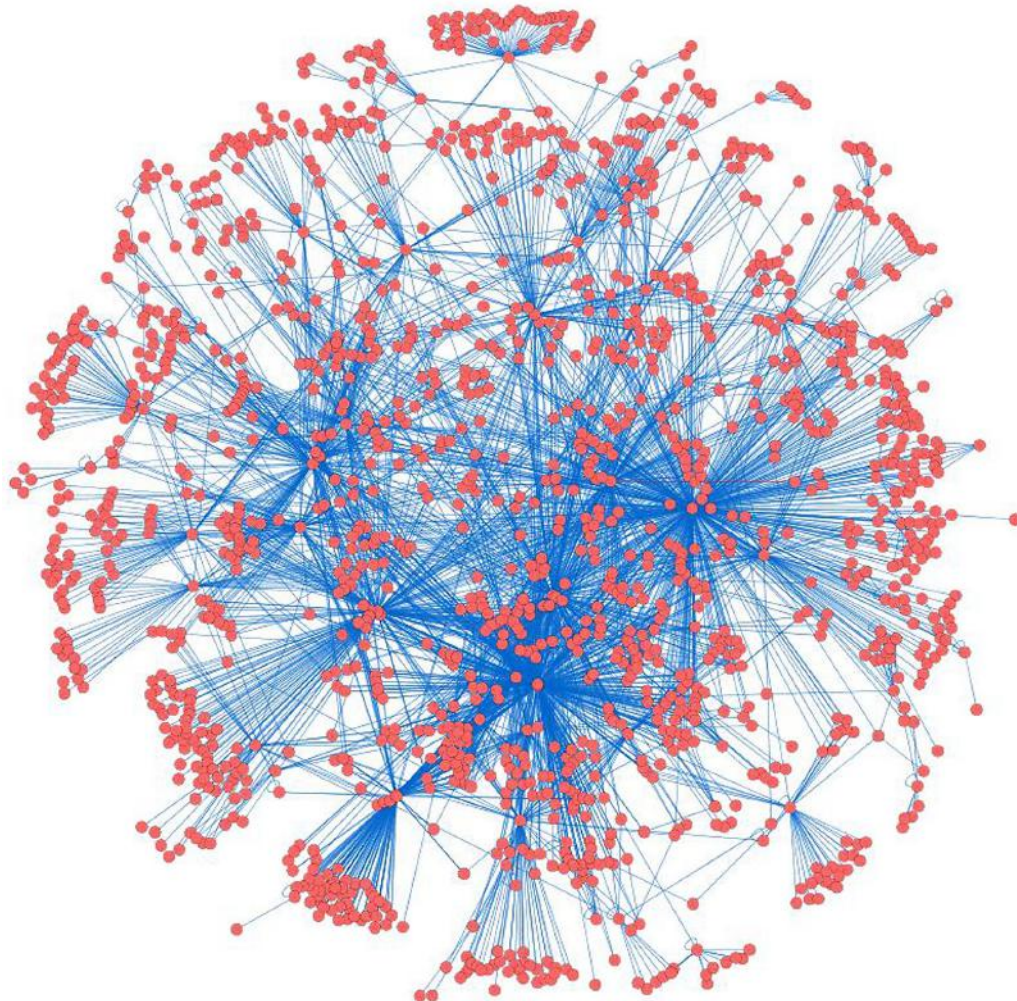


FIG. 6. Models of the regulation of protein synthesis.



Réseau de régulation génétique chez *Escherichia Coli*



Chaque point est un gène

Chaque ligne est un lien de régulation entre gène

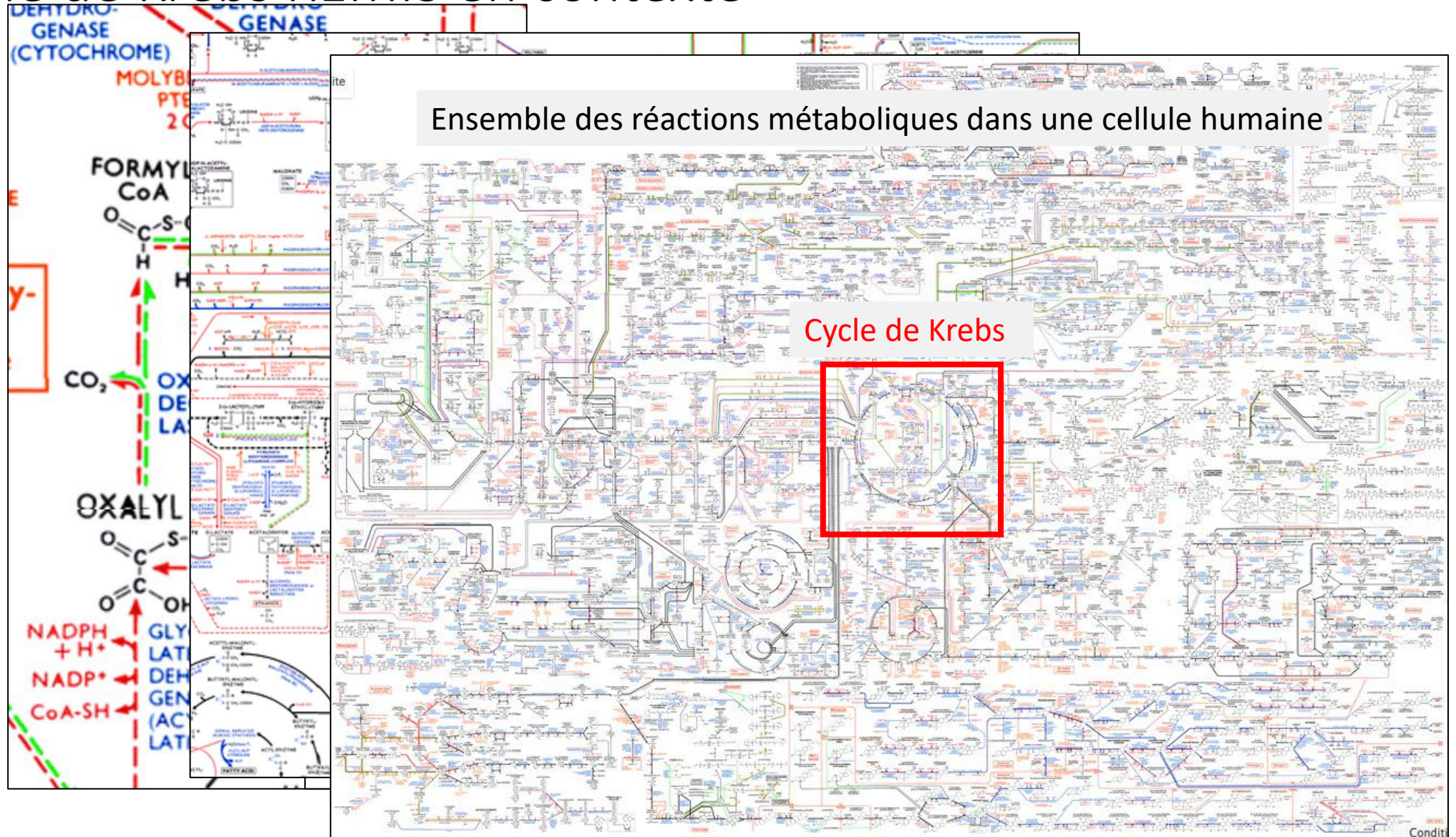
Il y a 4292 gènes

Le réseau a été **reconstruit** en utilisant :

524 puces ADN dans 264 conditions expérimentales

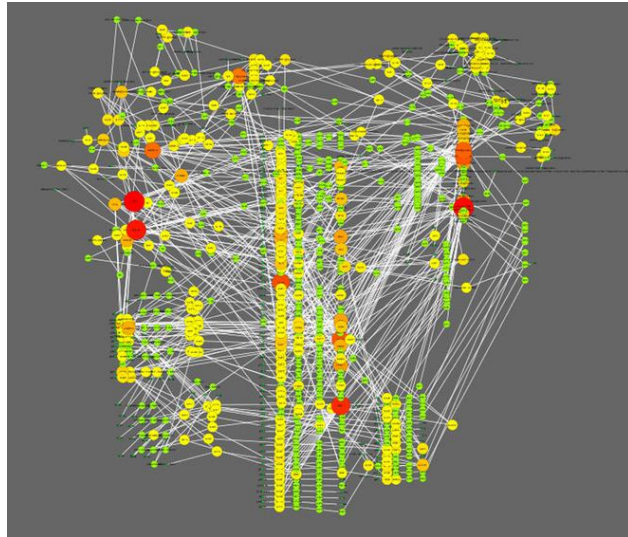
Allen et al., PlosOne, Jan 2012

Le cycle de Krebs REMIS en contexte

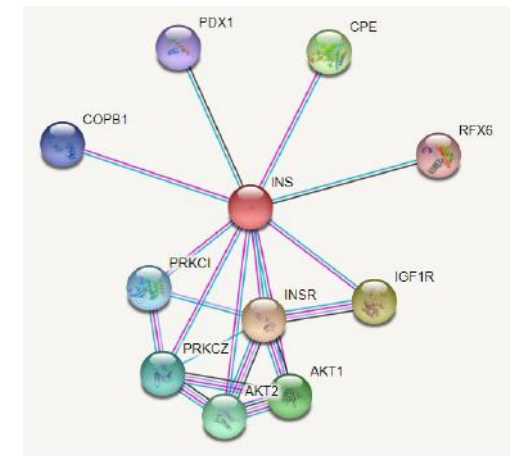
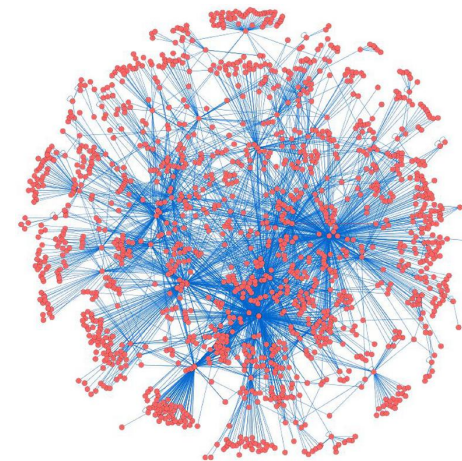
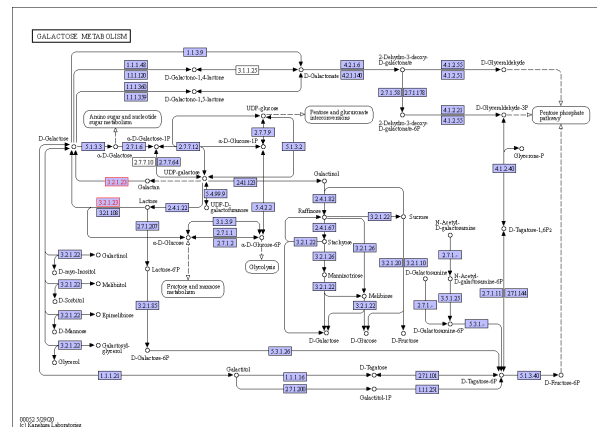
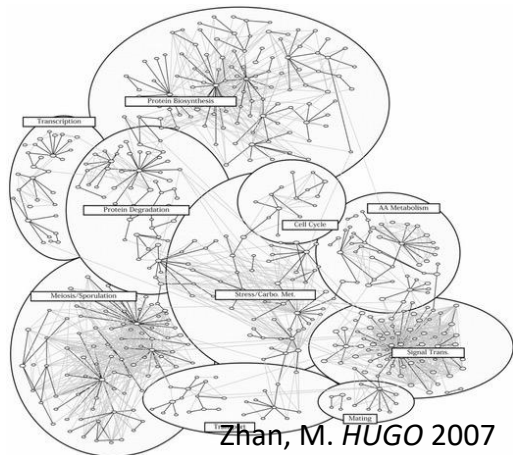


La biologie des systèmes

Cancer gene network – Institut Curie



- Etudier les processus biologiques avec une approche de réseau
- Une approche « Top-down »
- Le tout est plus que la somme de ses parties



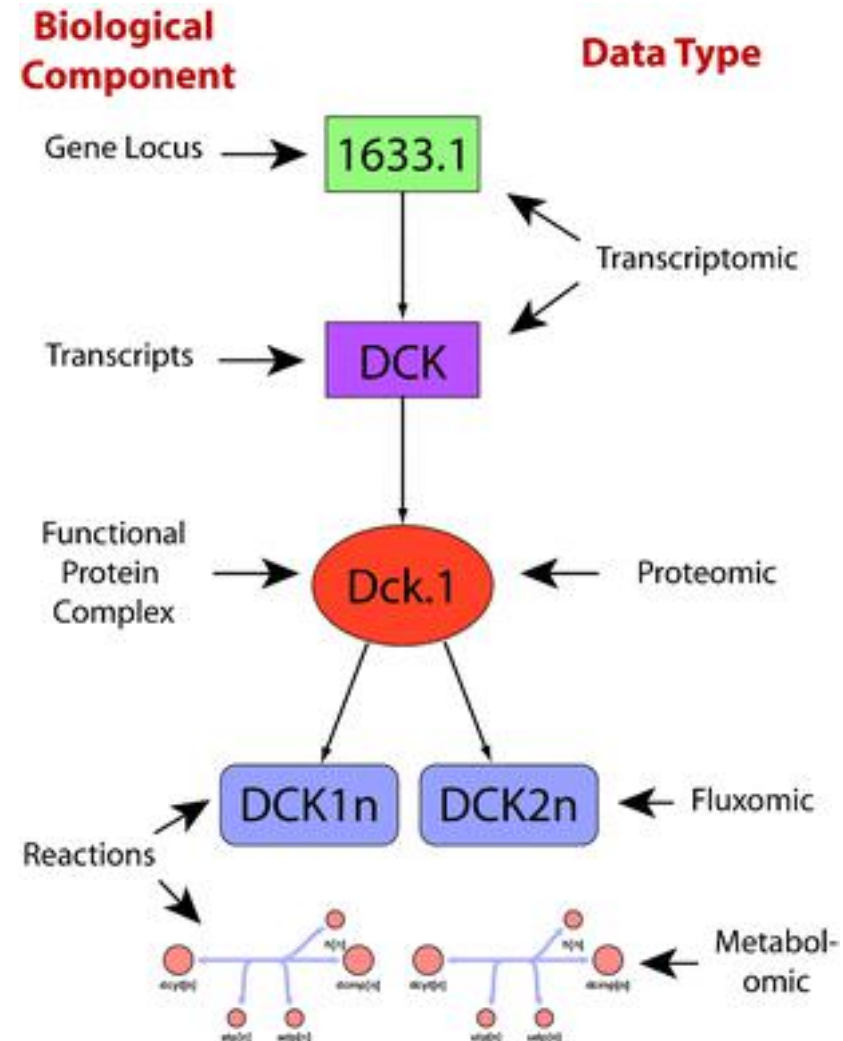
Biologie des systèmes

- L'approche systémique en biologie
- **Reconstruire un réseau biologique**
 - Les différents types de réseaux
 - Reconstruction directe
 - Reconstruction indirecte
 - Les obstacles à la reconstruction
- Analyser un réseau biologique
 - Théorie des graphes
 - Topologie des réseaux
 - Les réseaux sans échelles
- Utiliser un réseau biologique
 - Etat de l'art des réseaux biologiques les plus étendus
 - Bases de données en biologie des systèmes
- Modéliser un réseau biologique



Les éléments fondamentaux des réseaux biologiques

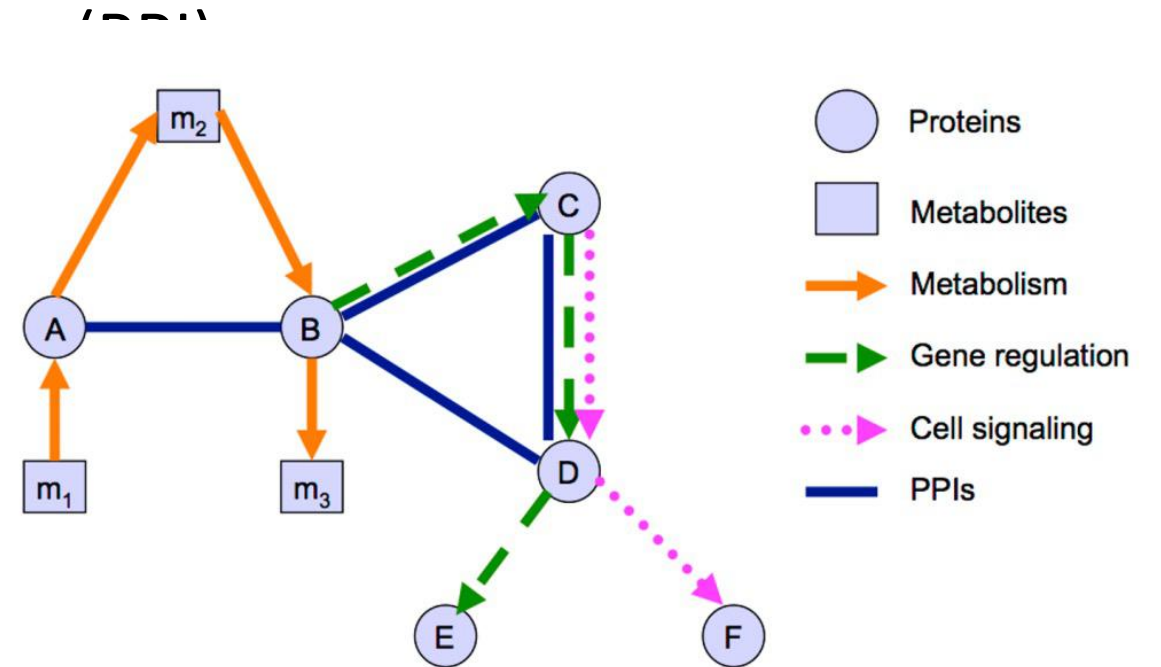
- Les gènes
- Les transcrits
- Les protéines
- Les complexes de protéines
- Les métabolites
- Les réactions enzymatiques



Mol. BioSyst., 2007, **3**, 598-603

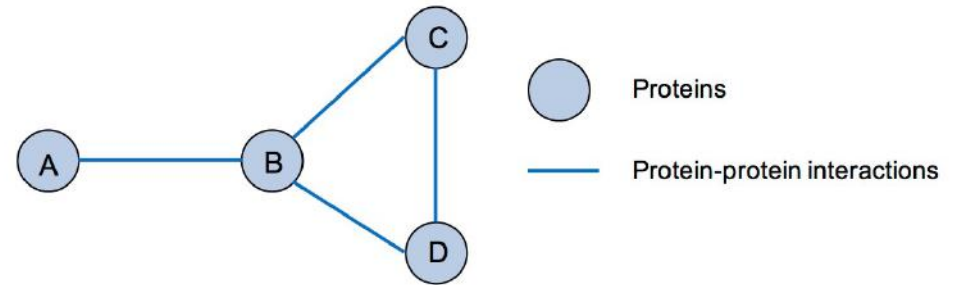
Types de réseaux biologiques

- Réseau d'interaction protéine-protéine
- Réseau métabolique
- Réseau de régulation génétique
- Réseau de signalisation cellulaire



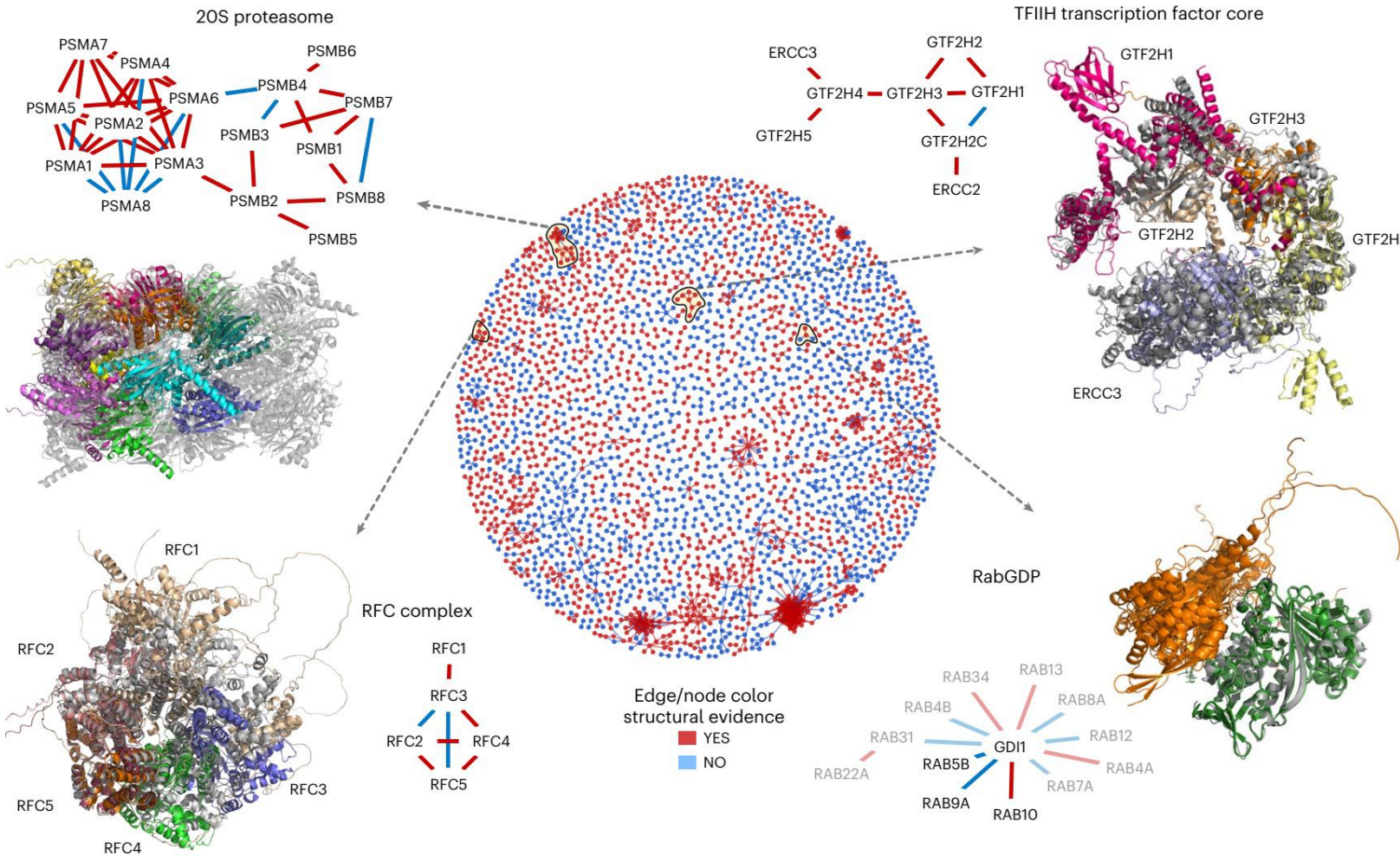
Réseau d'interaction protéine-protéine (PPI)

- Chaque nœud est une protéine
- Chaque arête représente un lien d'interaction physique entre protéines
- Le réseau est non orienté



Exemple de PPI

(Protein protein interaction)



3137 interactions de protéines

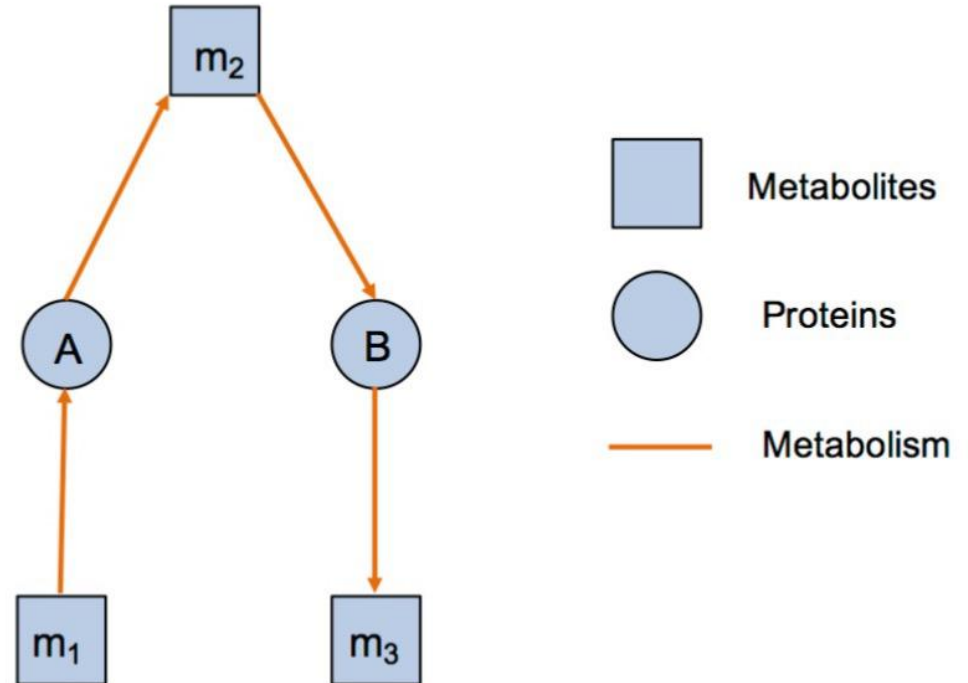
➤ 5000 protéines

Prédiction faites en utilisant
FoldDock et AlphaFold

Burke et al.
Nature Structural &
Molecular Biology 2023

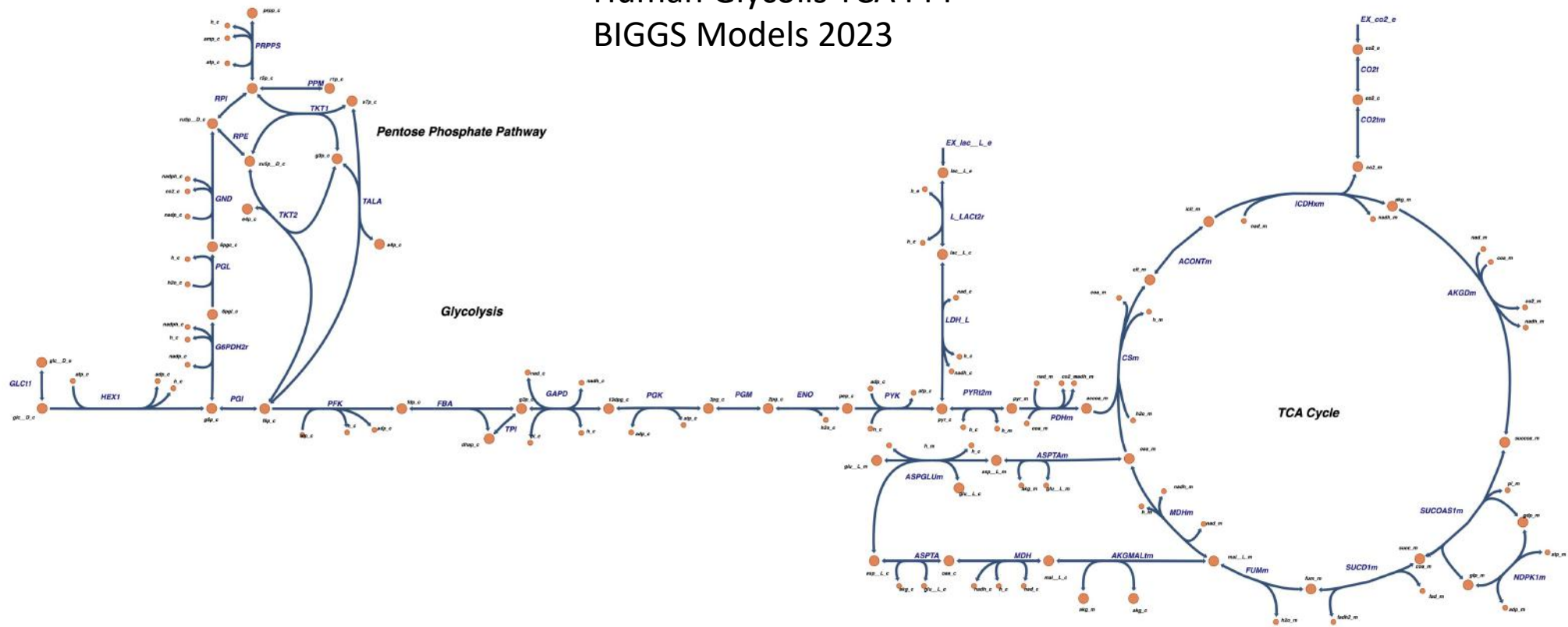
Réseau métabolique

- Les nœuds sont les métabolites et les enzymes
- Les arêtes représentent les réactions métaboliques et leurs flux
- Le réseau est orienté



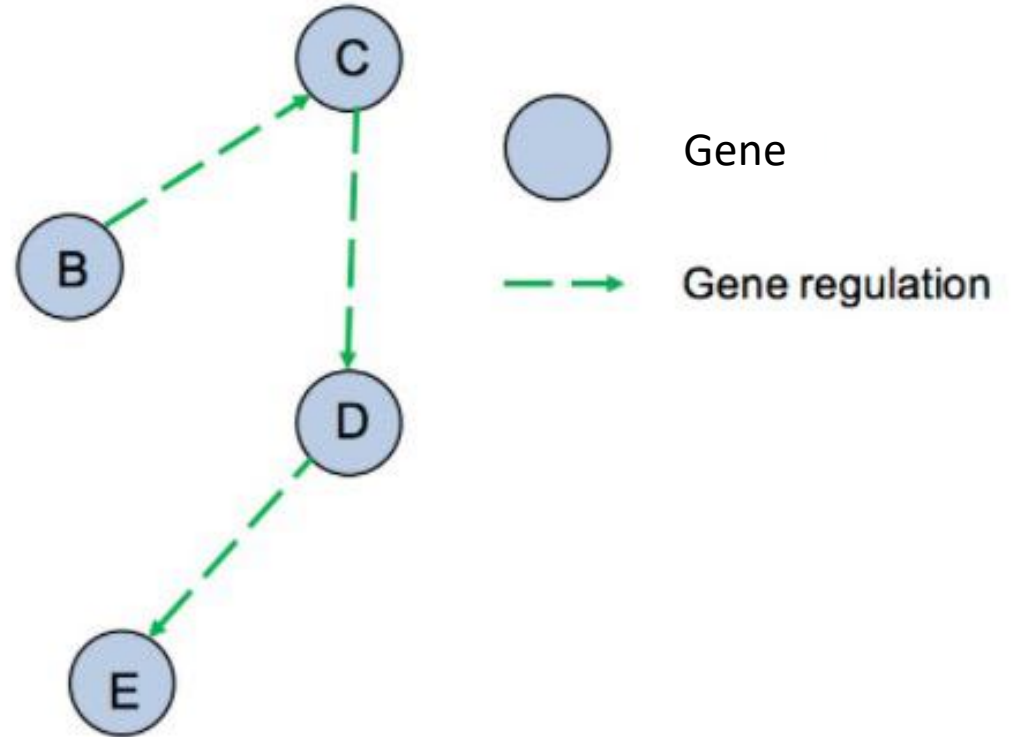
Exemple de réseau métabolique

Human Glycolis TCA PPP
BIGGS Models 2023

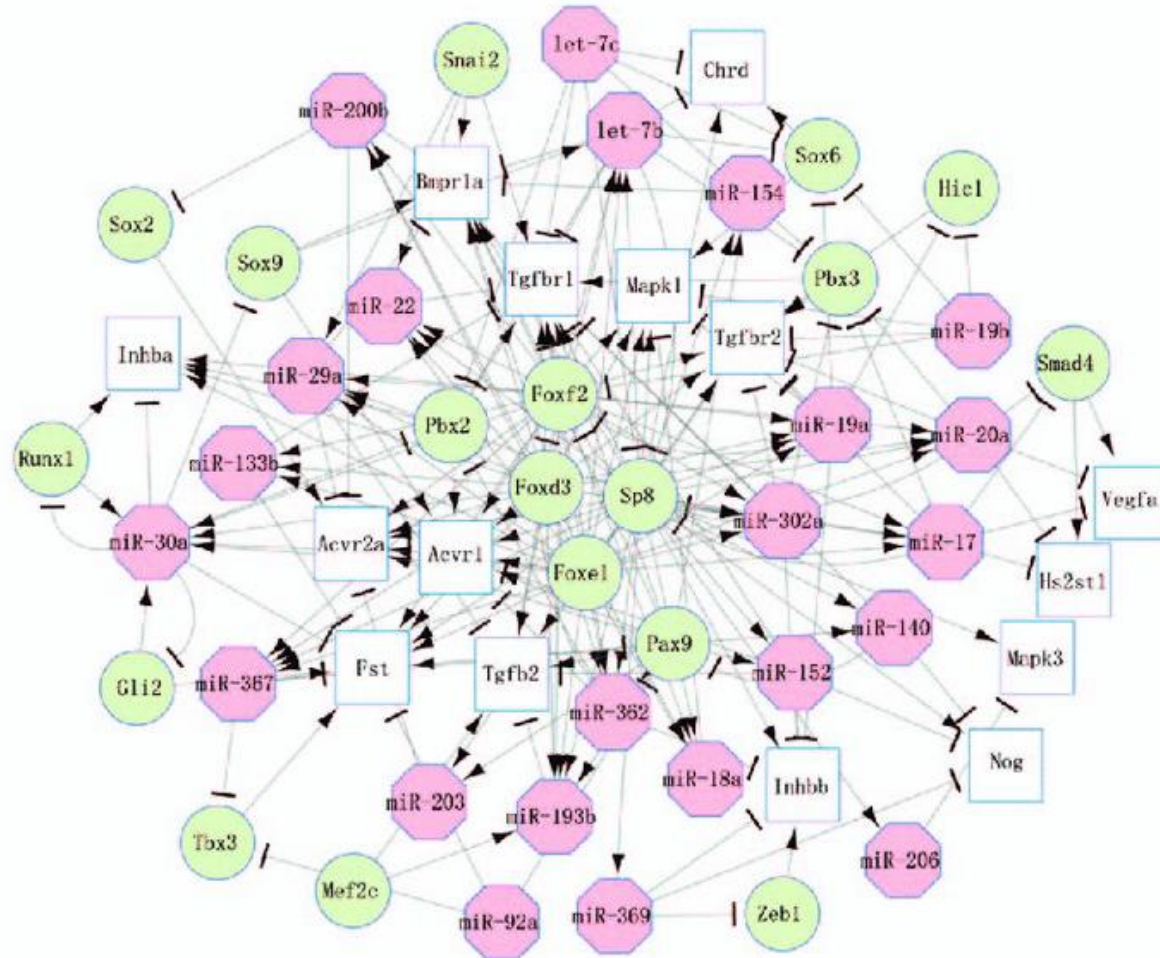


Réseau de régulation génétique (GRN)

- Les nœuds sont les gènes et les facteurs de transcription
- Les arêtes représentent les régulations transcriptionnelles
- Le réseau est orienté



Exemple de GRN (Gene regulation Network)



Un réseau de régulation génétique chez la souris

Les gènes codants sont en blanc

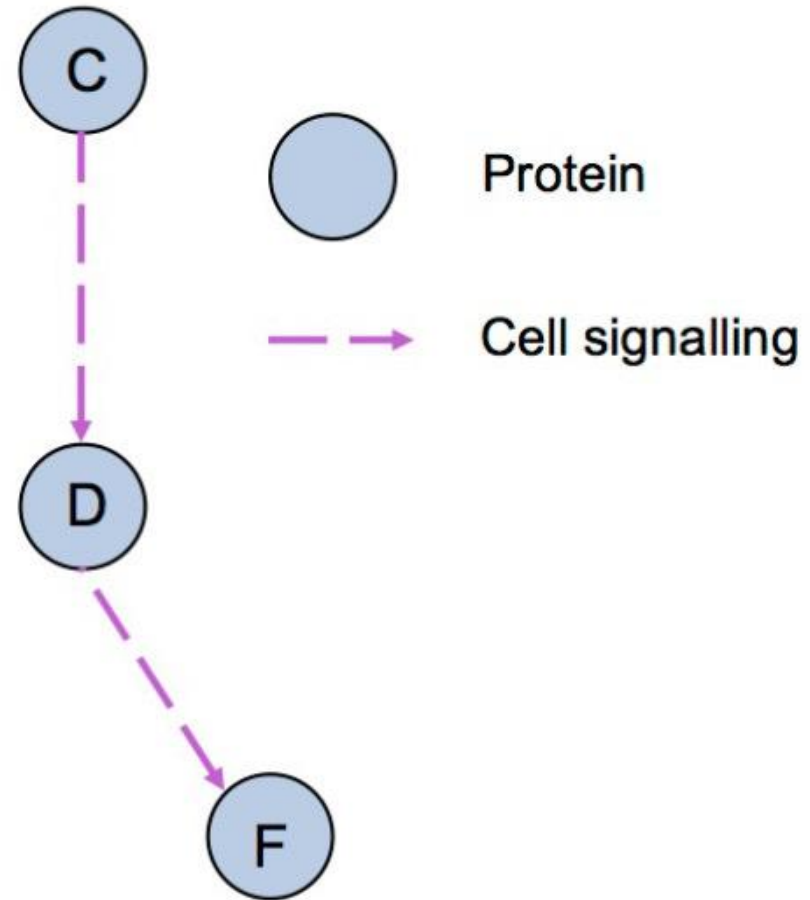
miRNA en rose

TF = facteur de transcription en vert

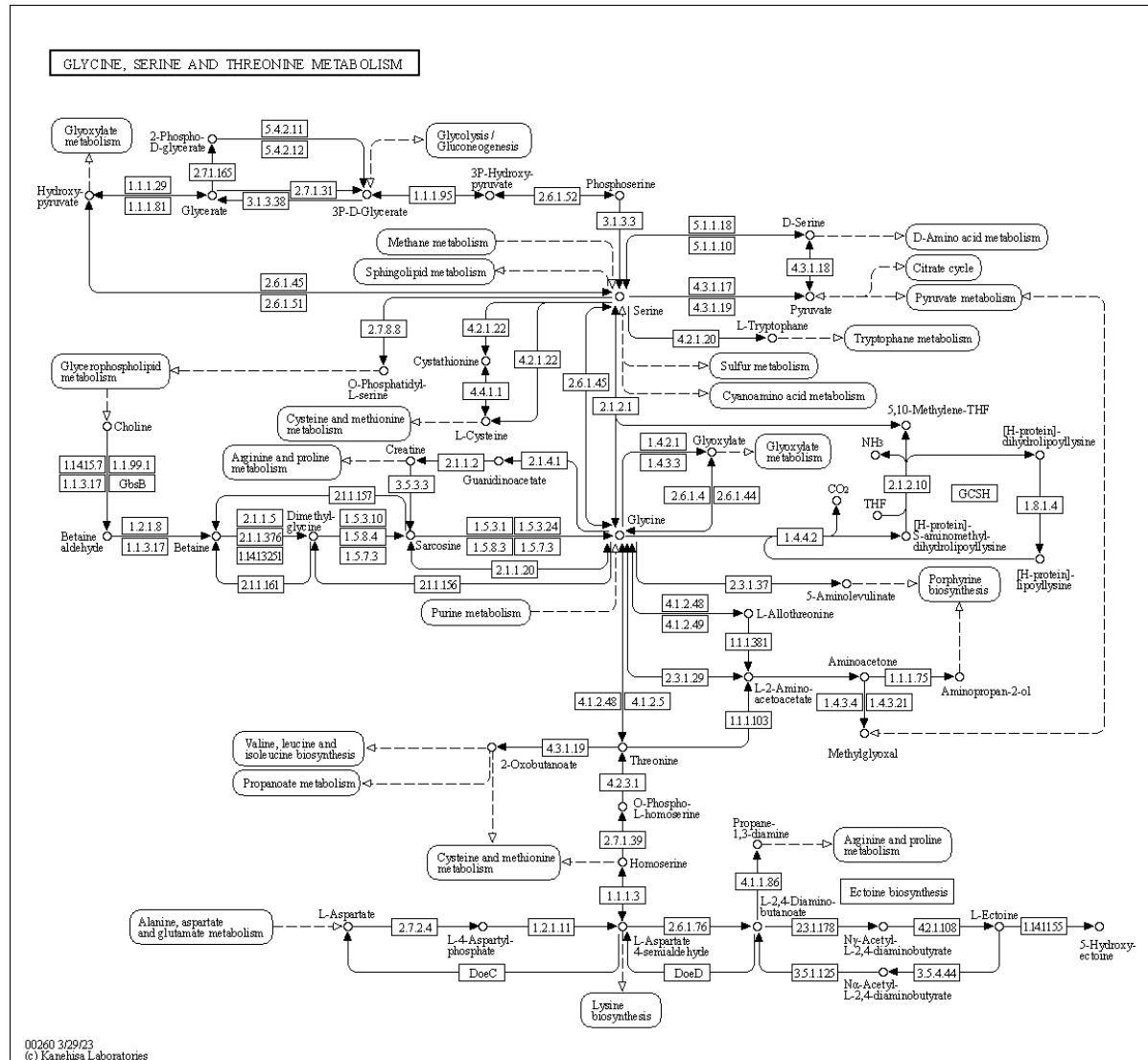
Li et al., Briefings in Bioinformatics, 2019

Réseau de signalisation cellulaire (pathway)

- Les nœuds sont les protéines, mais aussi les gènes et métabolites
- Les arêtes orientées représentent les voies de signalisation cellulaire
- Les autres réseaux peuvent être vu comme des sous-graphes de celui-ci



Example de pathway



KEGG Pathway database 2023

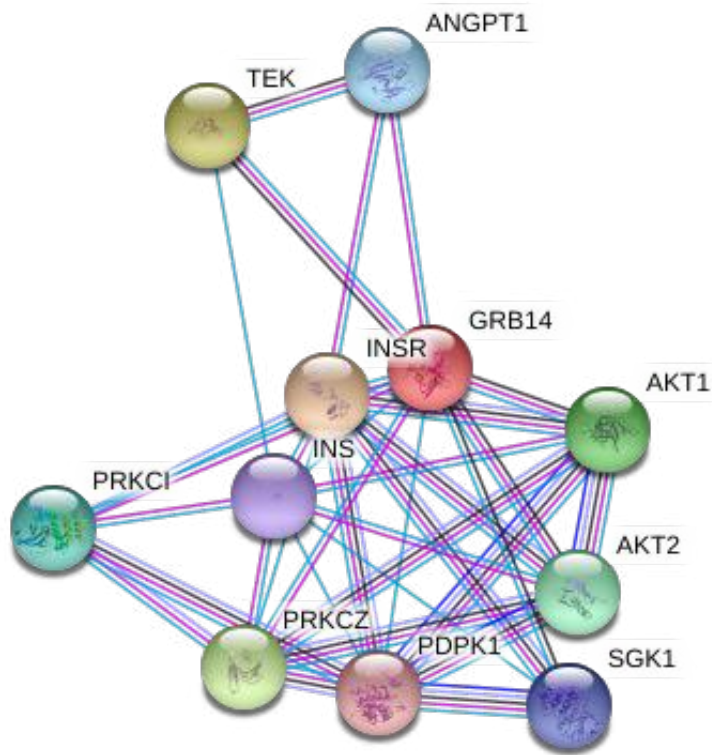
Glycine, Serine, and threonine metabolism

Genes = rectangles



Metabolites = les points

Autres pathway = ellipses




Reconstruction des réseaux biologiques



Known Interactions

-  *from curated databases*
-  *experimentally determined*




Predicted Interactions

-  *gene neighborhood*
-  *gene fusions*
-  *gene co-occurrence*

L'exemple de l'interaction protéines-protéines

Réseau d'interaction protéine-protéine
de l'insuline prédite par string-db.org

Others

-  *textmining*
-  *co-expression*
-  *protein homology*

Reconstruction des réseaux biologiques

L'exemple d'interaction protéines-protéines

Reconstruction directe

Known Interactions



from curated databases



experimentally determined

Reconstruction indirecte

Predicted Interactions



gene neighborhood



gene fusions



gene co-occurrence

Others



textmining



co-expression



protein homology

Reconstruction directe des réseaux métaboliques

Top-level EC numbers^[5]

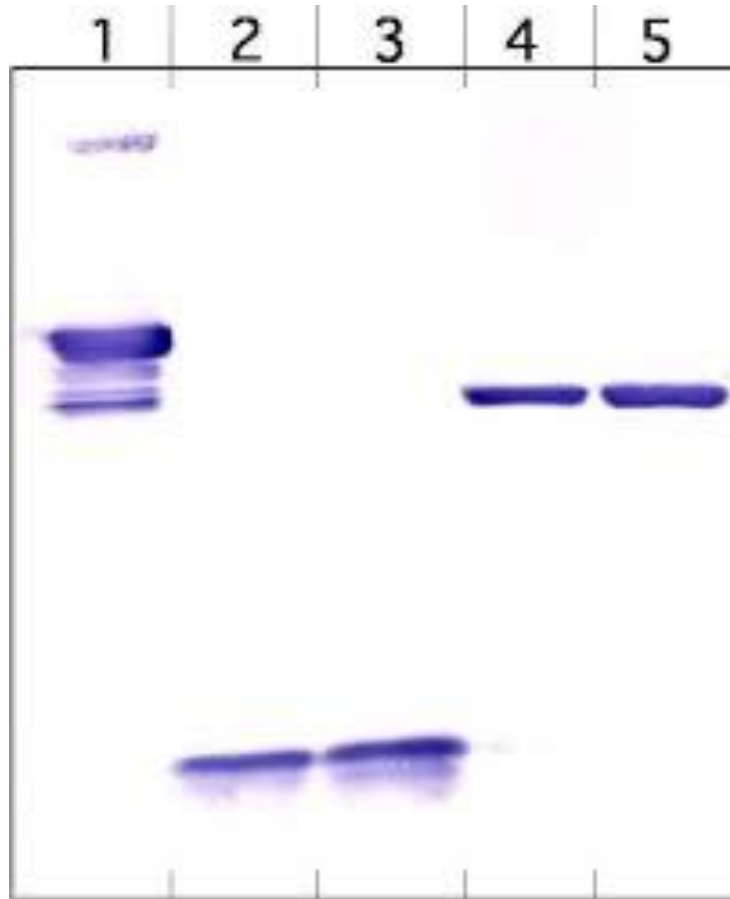
Class	Reaction catalyzed	Typical reaction	Enzyme example(s) with trivial name
EC 1 <i>Oxidoreductases</i>	To catalyze oxidation/reduction reactions; transfer of H and O atoms or electrons from one substance to another	$AH + B \rightarrow A + BH$ (reduced) $A + O \rightarrow AO$ (oxidized)	Dehydrogenase, oxidase
EC 2 <i>Transferases</i>	Transfer of a functional group from one substance to another. The group may be methyl-, acyl-, amino- or phosphate group	$AB + C \rightarrow A + BC$	Transaminase, kinase
EC 3 <i>Hydrolases</i>	Formation of two products from a substrate by hydrolysis	$AB + H_2O \rightarrow AOH + BH$	Lipase, amylase, peptidase, phosphatase
EC 4 <i>Lyases</i>	Non-hydrolytic addition or removal of groups from substrates. C-C, C-N, C-O or C-S bonds may be cleaved	$RCOCOOH \rightarrow RCOH + CO_2$ or $[X-A+B-Y] \rightarrow [A=B + X-Y]$	Decarboxylase
EC 5 <i>Isomerases</i>	Intramolecule rearrangement, i.e. isomerization changes within a single molecule	$ABC \rightarrow BCA$	Isomerase, mutase
EC 6 <i>Ligases</i>	Join together two molecules by synthesis of new C-O, C-S, C-N or C-C bonds with simultaneous breakdown of ATP	$X + Y + ATP \rightarrow XY + ADP + P_i$	Synthetase

Enzyme Commission number for enzymes

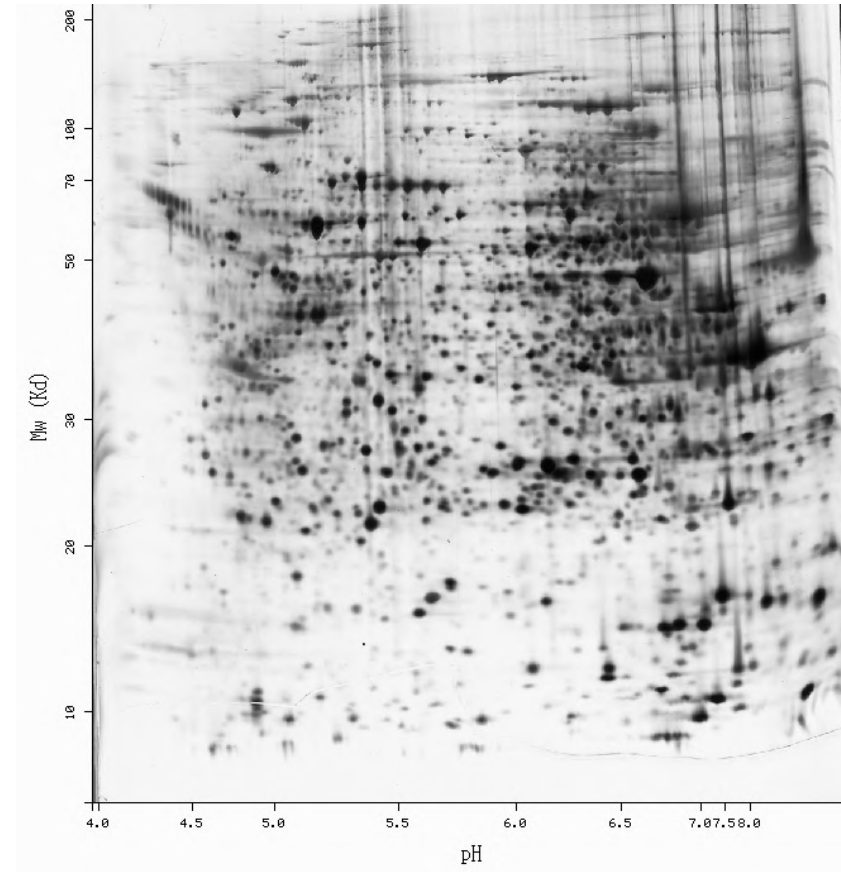
https://en.wikipedia.org/wiki/List_of_enzymes

Reconstruction directe : Mesure d'interaction protéique

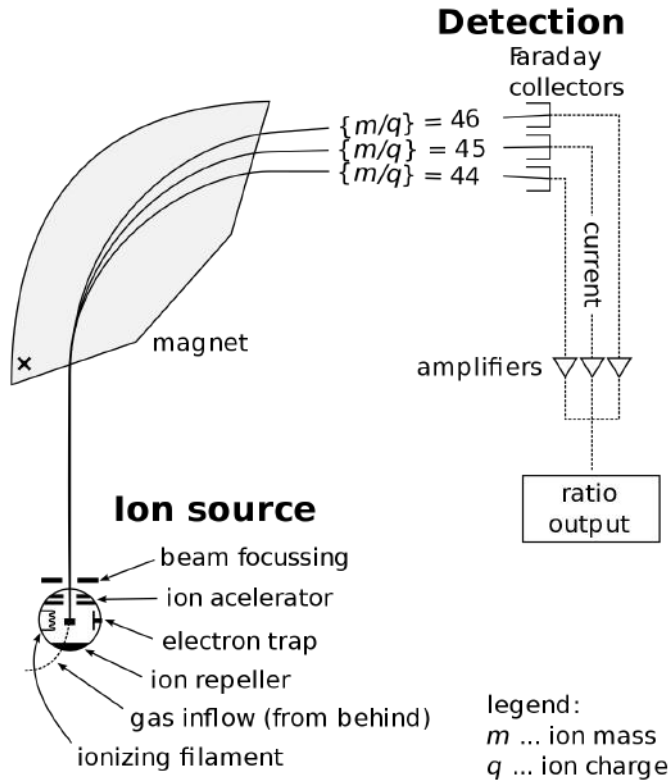
Western blot : Mesure de protéine unique



Gel 2D : Mesure de plusieurs protéines

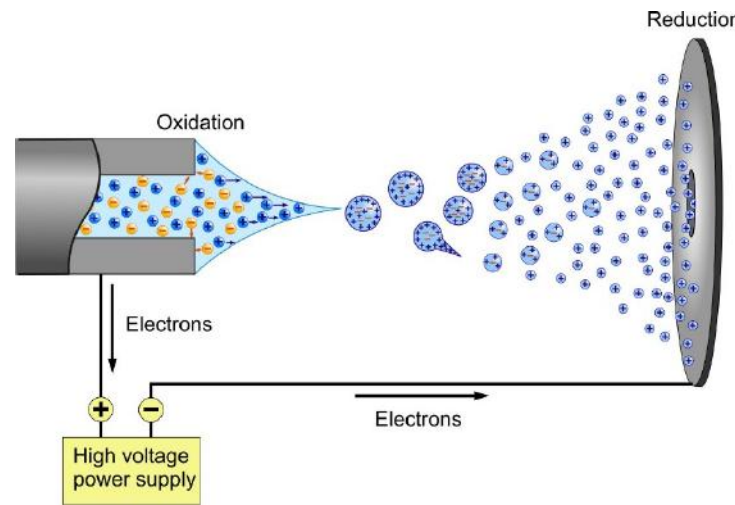


Evolution de la protéomique

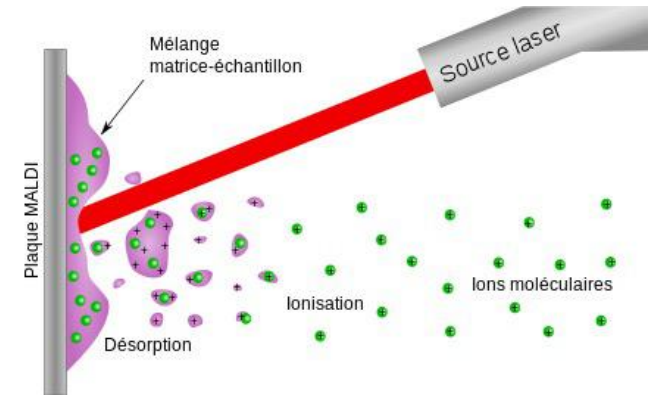


Spectrométrie de masse
1888 - 1918

Pour pouvoir passer les protéines dans un spectromètre il faut les ioniser



Ionisation des protéines par spray
ESI - 1968



Ionisation des protéines par laser
MALDI - 1985

Evolution de la protéomique

Les protéines ionisées sont injectées dans le spectromètre de masse.

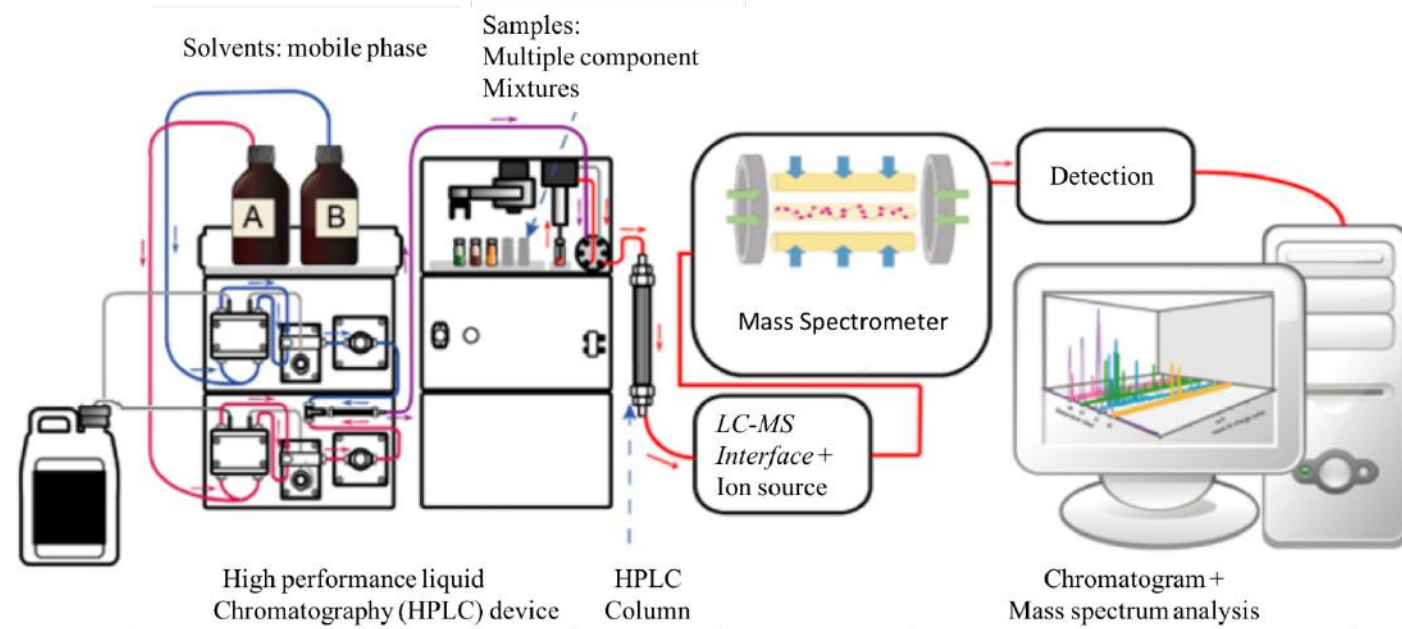


Schéma du LiquidChromatography/MassSpectrometry LC/MS - 2009

On rajoute un fractionnement des peptides pour avoir une meilleure précision
LC/MS/MS

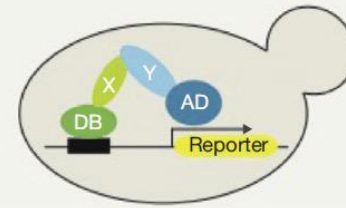
Reconstruction directe des interactions protéines-protéines

Base de données : Reactome, PDB, KEGG

(a)

Binary mapping

Yeast two-hybrid (Y2H)

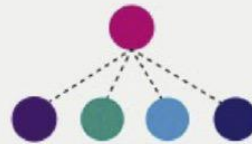


Interaction 2 à 2

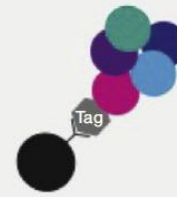
(b)

Co-complex mapping

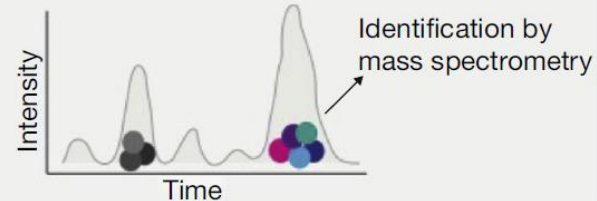
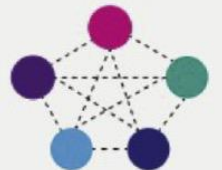
Affinity purification followed by mass spectrometry (AP-MS)



Identification by mass spectrometry



Co-fractionation followed by mass spectrometry

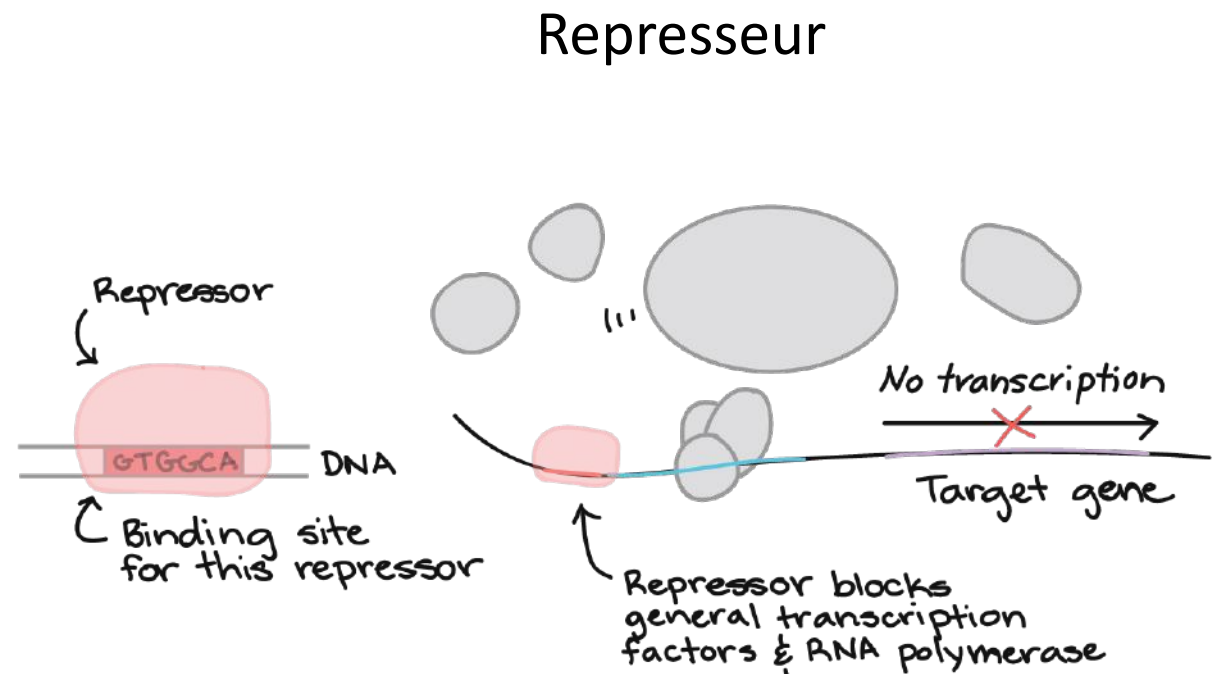
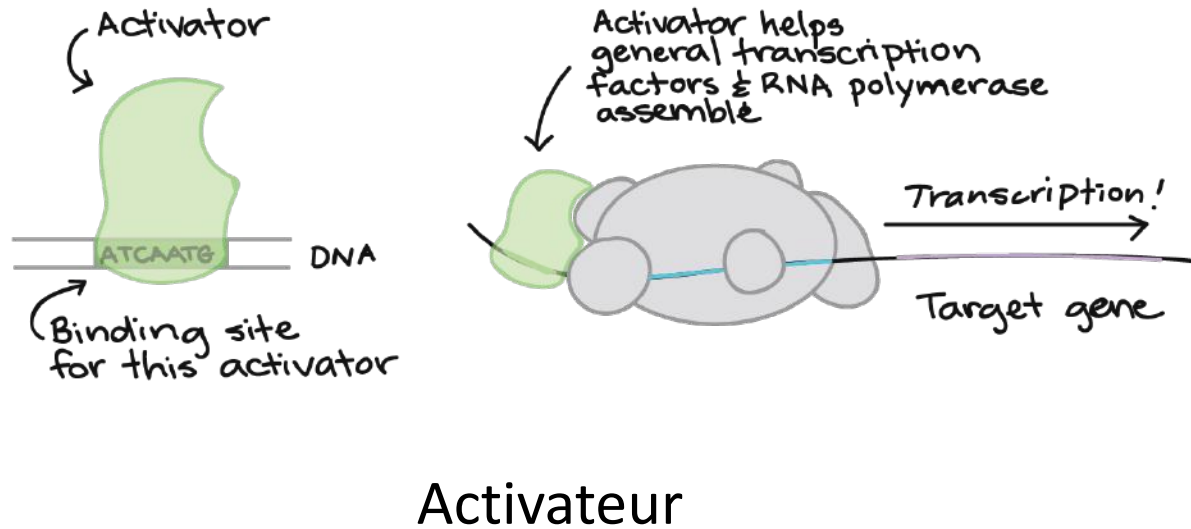


Complexe protéique

● Protein — Direct physical interaction - - - - Protein association

Reconstruction Directe des Gene Regulatory Network







Détection des facteurs de transcription



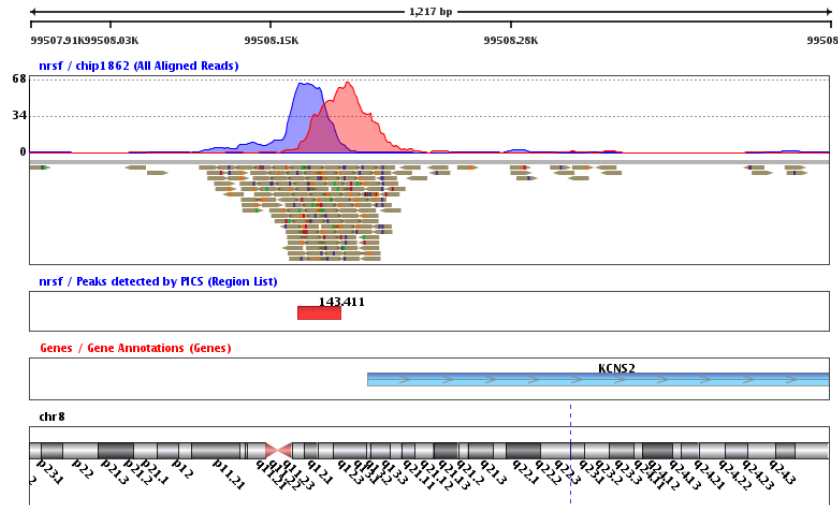
Reconstruction Directe des Gene Regulatory Network

Détection des facteurs de transcription

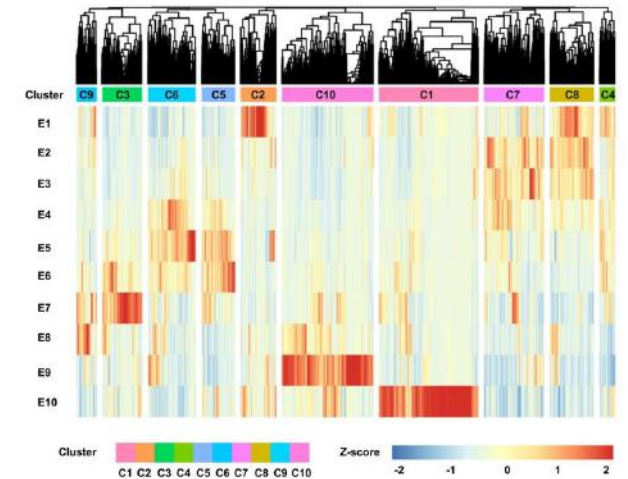
Recherche de motif

Transcription Factor	Motif	Motif Fold Enrichment
<i>Neurod</i> family		2.39
<i>Lhx / Lmx</i> family		2.42
<i>Nfi</i> family dimer		4.14
<i>Rfx</i> family dimer		3.33
Novel <i>Hox</i> dimer		2.32
Novel <i>Nfi</i> dimer		2.06

Méthode CHIPSeq



Validation des TF



Comment détecter les régulations transcriptomiques ?

Different Generations of Sequencing

First Generation

- 1972: Sanger started work on DNA sequencing
- 1977: Sanger developed Di-deoxy chain termination method of DNA sequencing
- 1977: Maxam and Gilbert developed chemical degradation method of DNA sequencing
- 1977: First DNA based genome sequenced (ΦX174 bacteriophage)
- 1995: First bacterium *Haemophilus influenzae* was sequenced by shotgun method
- 1996: Applied Biosystems developed automated DNA sequencing based on Sanger method
- 1996: First eukaryotic genome (*Saccharomyces cerevisiae*) was sequenced
- 2001: First human genome draft was published by two different independent teams

Second Generation

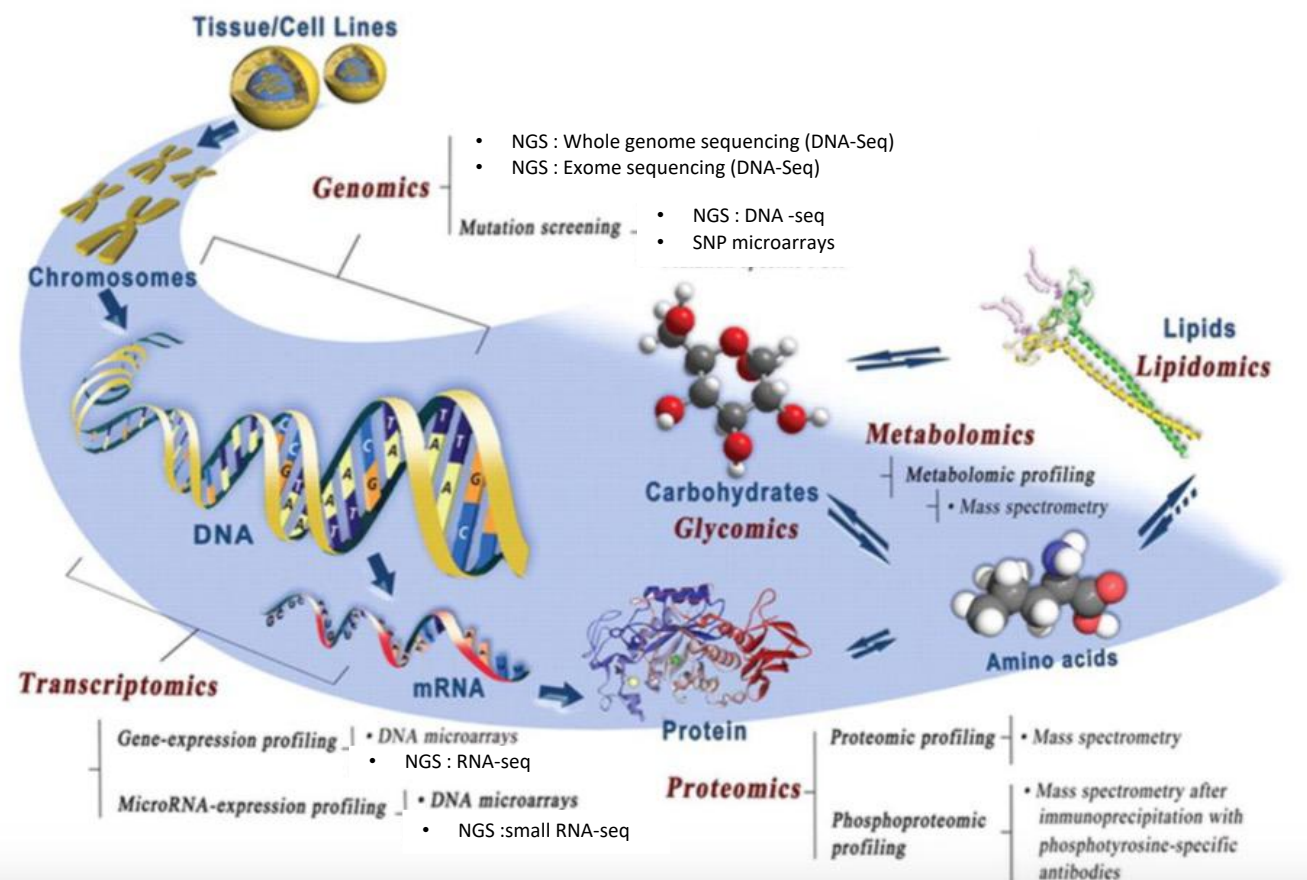
- 2005: First NGS platform released Roche 454 GS-20
- 2006: Introduction of second NGS platform –Solexa Genome Analyzer
- 2006: Initiation of 1000 genome project
- 2007: Introduction of Roche 454 GS-FLX & ABI-SOLID sequencer
- 2008: Development of Illumina GA-II
- 2009: Introduction of Roche 454 GS-FLX Titanium
- 2010: Introduction of Roche 454 GS-Junior
- 2011: Introduction of SOLiD 5500 W & Illumina MiSeq
- 2012: Introduction of Illumina HiSeq
- 2013: Introduction of SOLiD 5500xl W & Illumina MiniSeq
- 2014: Introduction of Roche 454 GS-Junior+, Illumina NextSeq 500 & Illumina HiSeq X Ten
- 2017: Introduction of Illumina iSeq 100

Third Generation

- 2008: Development of first commercial platform of third generation technology i.e. Helicose Biosciences
- 2010: Ion Torrent released the Personal Genome Machine (PGM)
- 2011: Introduction of PacBio RS C1/C2
- 2012: Introduction of PacBio RS C2 XL & PacBio RS II C2 XL, Ion Torrent released the Proton
- 2013: Introduction of PacBio RS II C2 XL
- 2014: Introduction of PacBio RS II P5 C3 & PacBio RS II P6 C4
- 2015: Introduction of Ion S5/S5XL 520/530/540
- 2016: Introduction of PacBio sequel

Fourth Generation

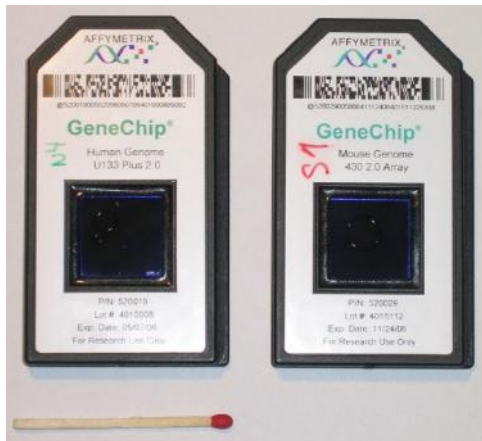
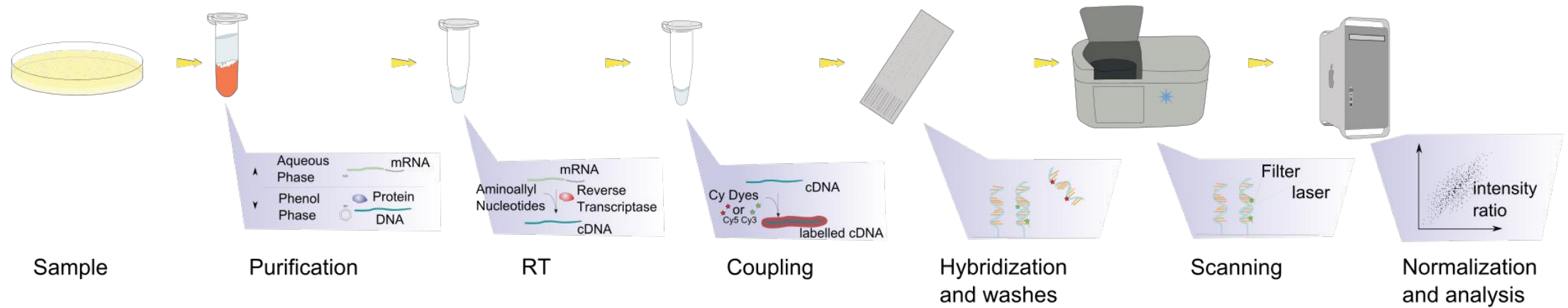
- 2014: Release of MinION platform by Oxford Nanopore Technologies
- 2017: Release of ProMethION, GridION & SmidgION X5 platforms by Oxford Nanopore Technologies
- 2018: Commercialization of ProMethION platform by Oxford Nanopore Technologies



Wu R.Q., J. dent. Research, 2010

Evolution de la transcriptomique

Les prémices avec les puces ADN

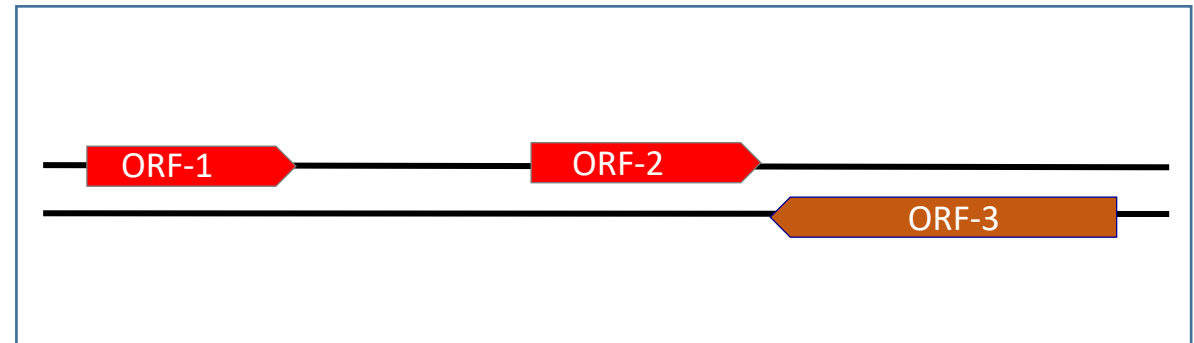


- Expression des ARNs
- Comparaison de génomes
- SNP
- ChIP

Séquençage haut débit - short-read



mRNA extraction



Reads mapping

For all you seq...

DNA

This infographic is divided into several sections detailing sequencing methods and protocols. The top left section, 'DNA Sequencing Methods', lists various techniques such as Sanger sequencing, Next-Generation Sequencing (NGS), and Single-Molecule Sequencing (SMS), each with a small schematic. The top right section, 'DNA-Protein Interactions', illustrates methods like ChIP-seq and ATAC-seq. The middle section, 'Sequencing', provides a detailed overview of the sequencing process, from library preparation to data analysis, with numerous sub-diagrams and text blocks. The bottom section, 'Sequencing by Synthesis', shows the underlying mechanism of the sequencing process on a flow cell, with diagrams of DNA synthesis and signal detection. The entire infographic is densely packed with text, diagrams, and icons, providing a comprehensive overview of DNA sequencing technologies.

This section of the infographic focuses on 'Sequencing by Synthesis' and 'Methods'. It includes several sub-diagrams illustrating the step-by-step process of DNA synthesis on a flow cell, from the initial binding of DNA fragments to the final detection of fluorescence signals. Below these diagrams, there are several small tables and lists of reagents and components used in the sequencing process. The bottom right corner of this section features the Illumina logo.

RNA

For all you seq...

RNA Transcription

RNA-Protein Interactions

RNA Low-Level Detection

RNA Modifications

RNA Structure

Reporting by ReadPair

Reporting by ReadStack

illumina

© 2014 Illumina, Inc. All rights reserved. This document is for informational purposes only and does not constitute an offer of any product or service. For more information, please visit www.illumina.com.
 The information in this document is subject to change without notice. The information in this document is provided as is and without warranty of any kind, including without limitation, the implied warranties of merchantability and fitness for a particular purpose. The information in this document is provided on an "as is" basis and without warranty of any kind, including without limitation, the implied warranties of merchantability and fitness for a particular purpose. The information in this document is provided on an "as is" basis and without warranty of any kind, including without limitation, the implied warranties of merchantability and fitness for a particular purpose. The information in this document is provided on an "as is" basis and without warranty of any kind, including without limitation, the implied warranties of merchantability and fitness for a particular purpose.

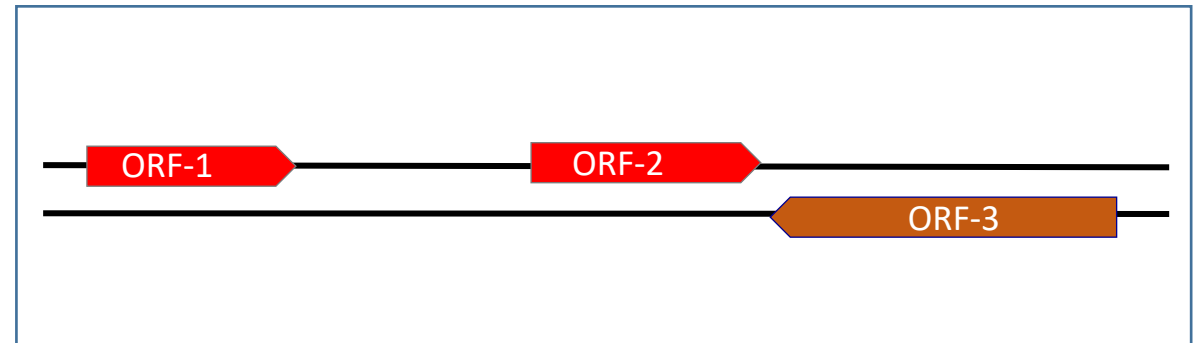
Séquençage haut débit - long-read



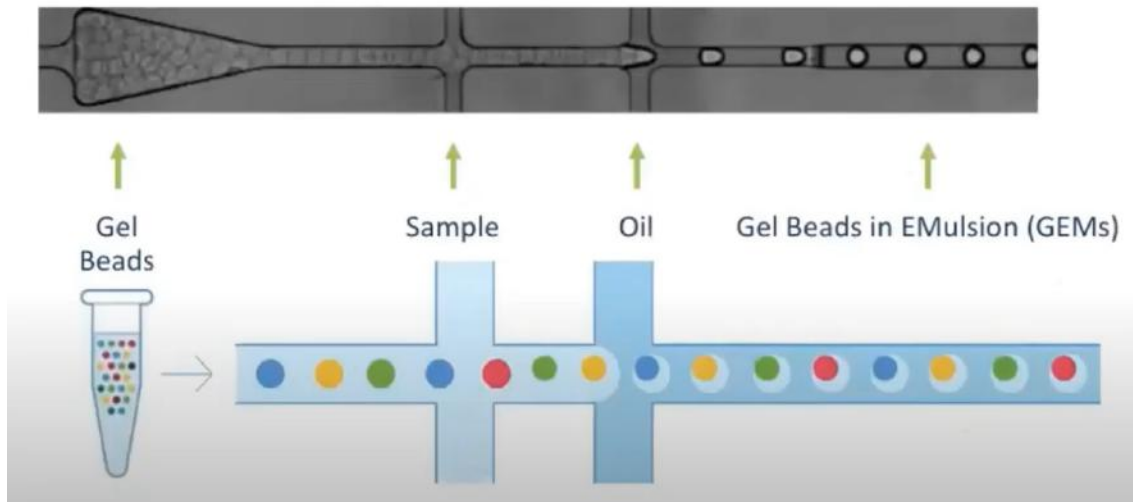
mRNA extraction



Reads mapping

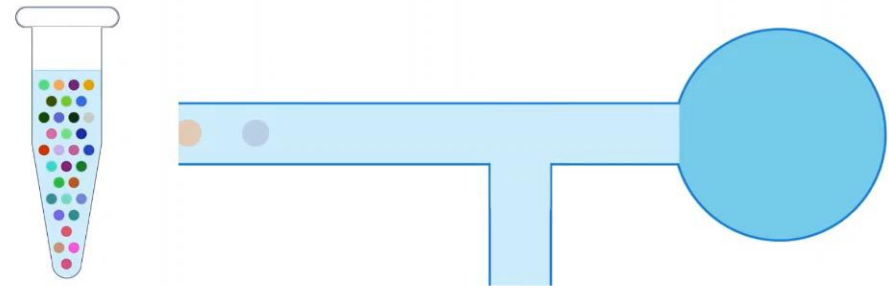


Séquençage à cellule unique

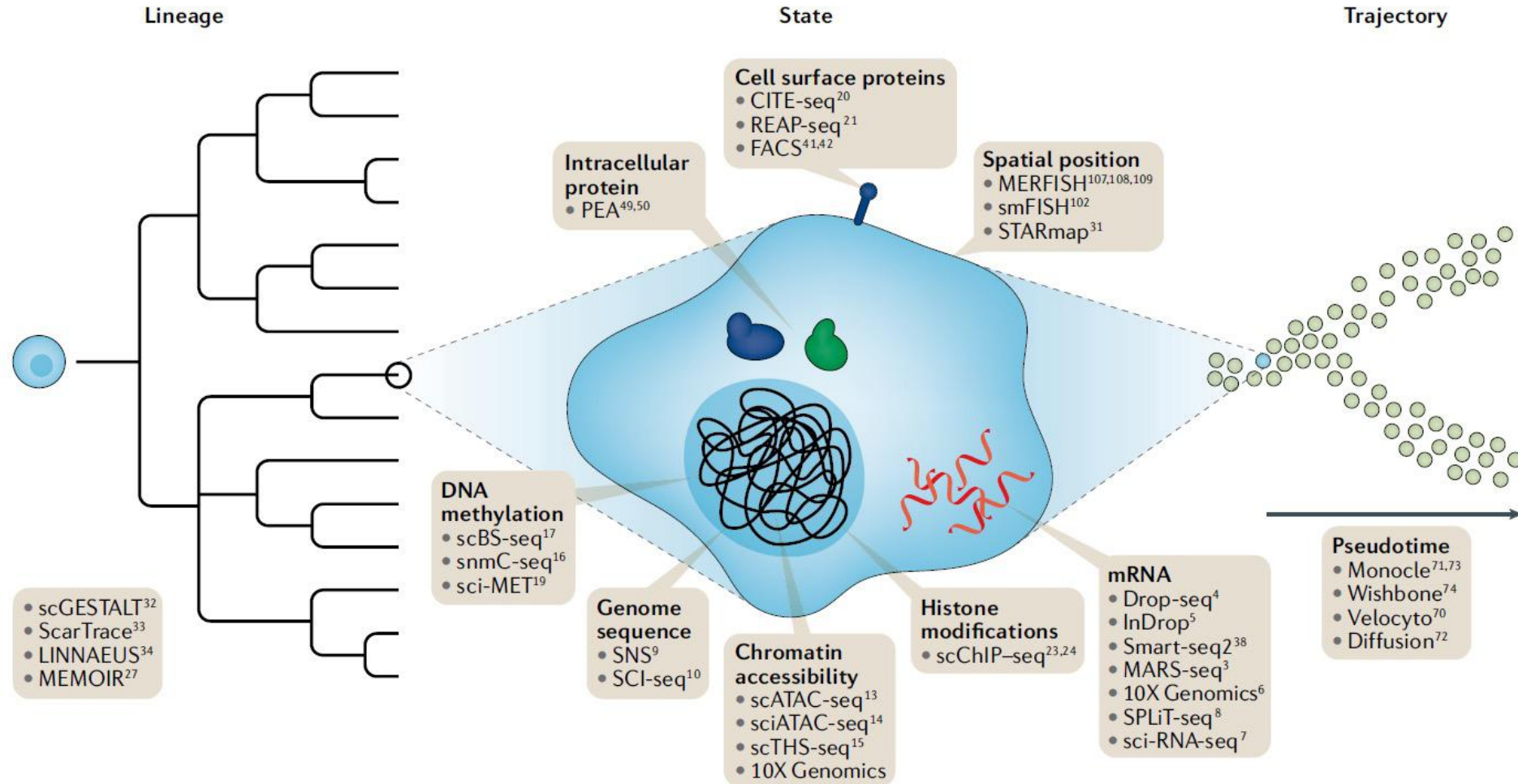


Adapted from 10x Genomics

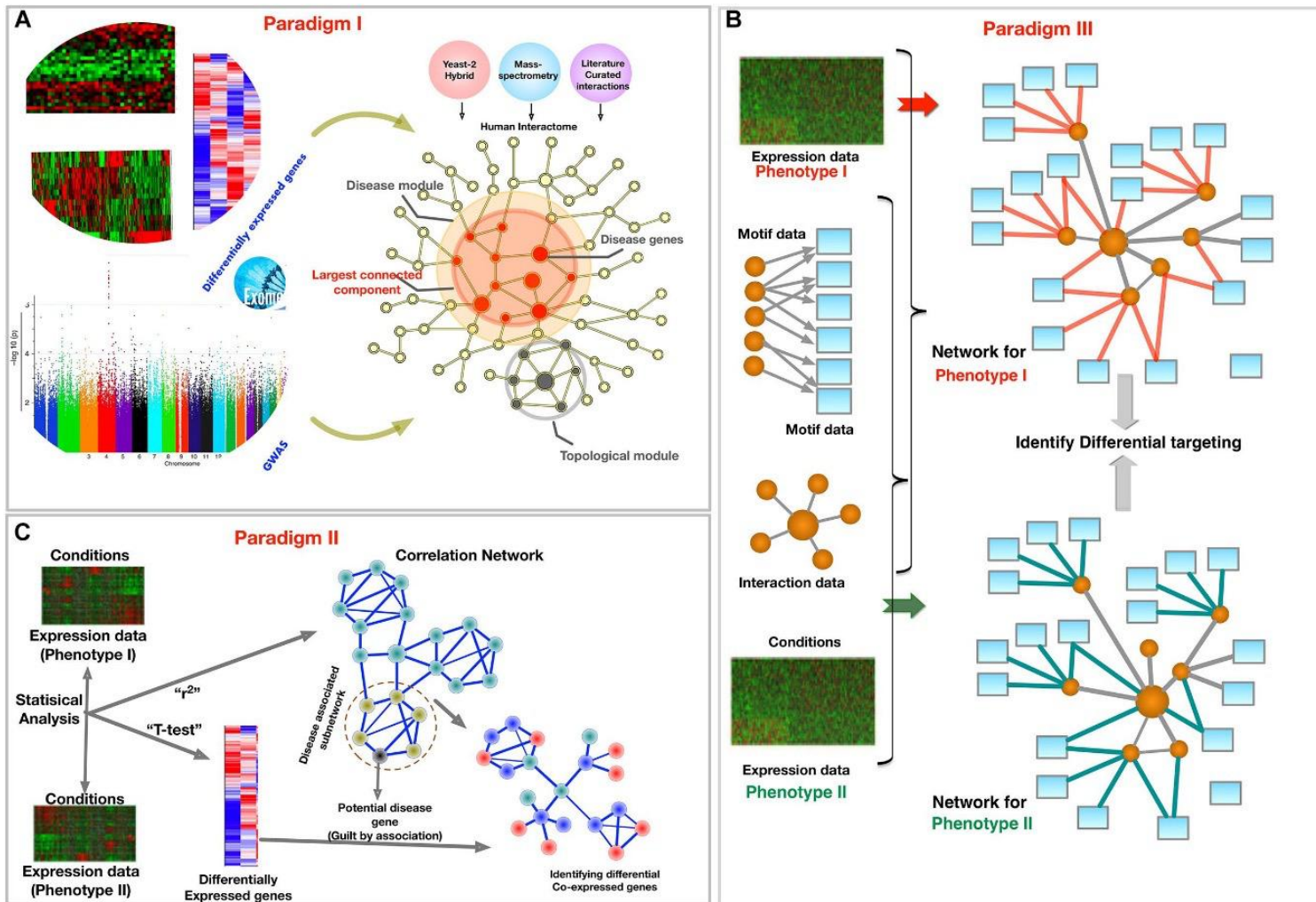
10x Next GEM Technology for Single Cell Partitioning



Une variété de techniques single-cell

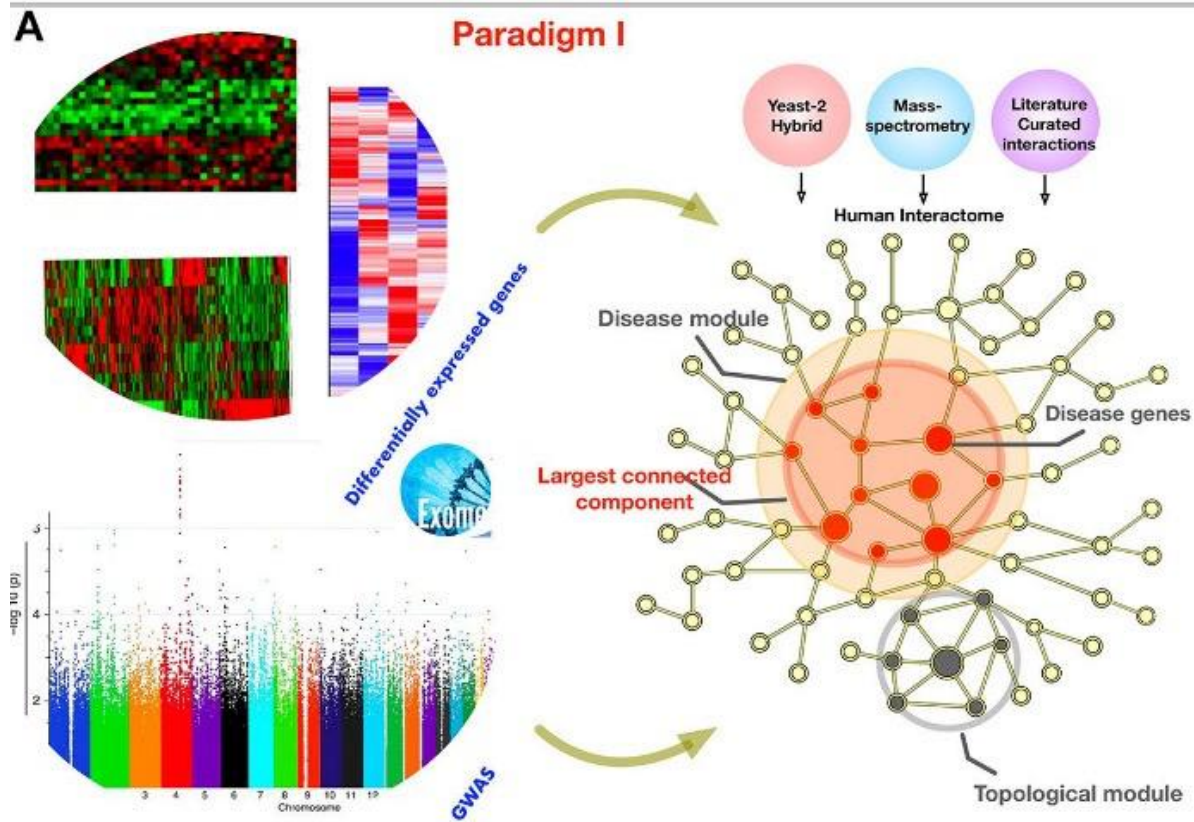


Reconstruction indirecte des réseaux biologiques



Network Medicine in the Age of Biomedical Big Data
 Front. Genet., 11 April 2019

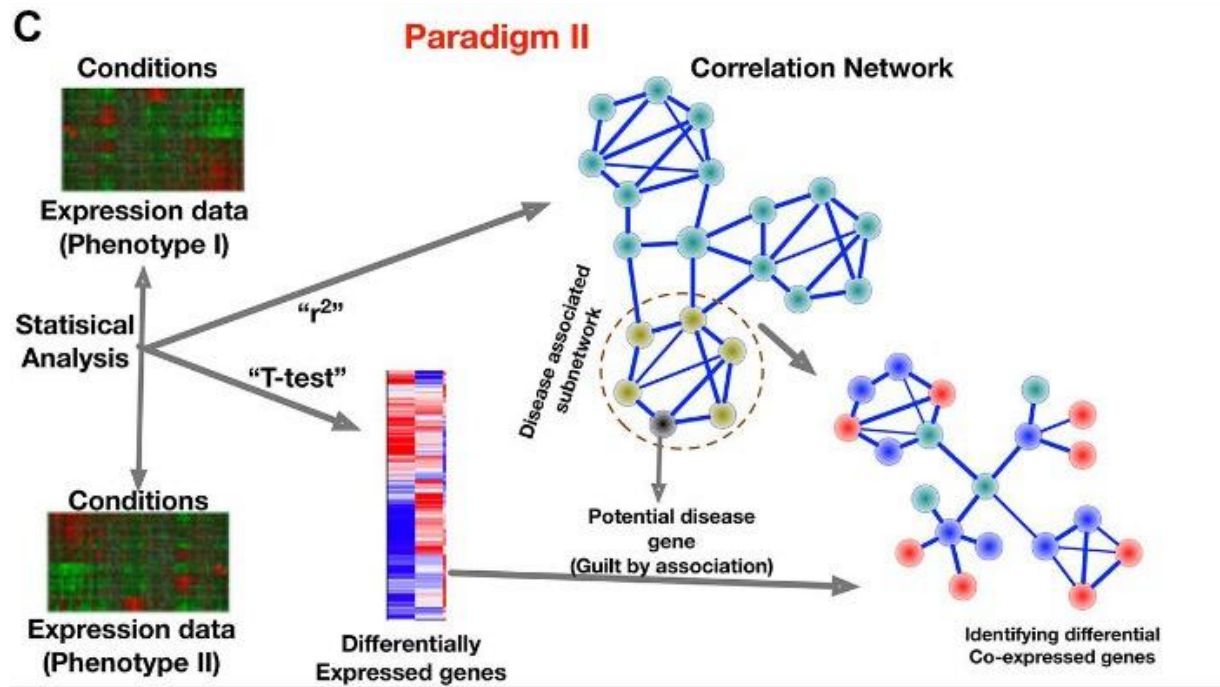
Reconstruction indirecte des réseaux biologiques



Corrélation de données multi-omiques

- Réseau d'interaction protéine-protéine
- Données de transcriptomiques
- Présence de SNP caractéristique

Reconstruction indirecte des réseaux biologiques

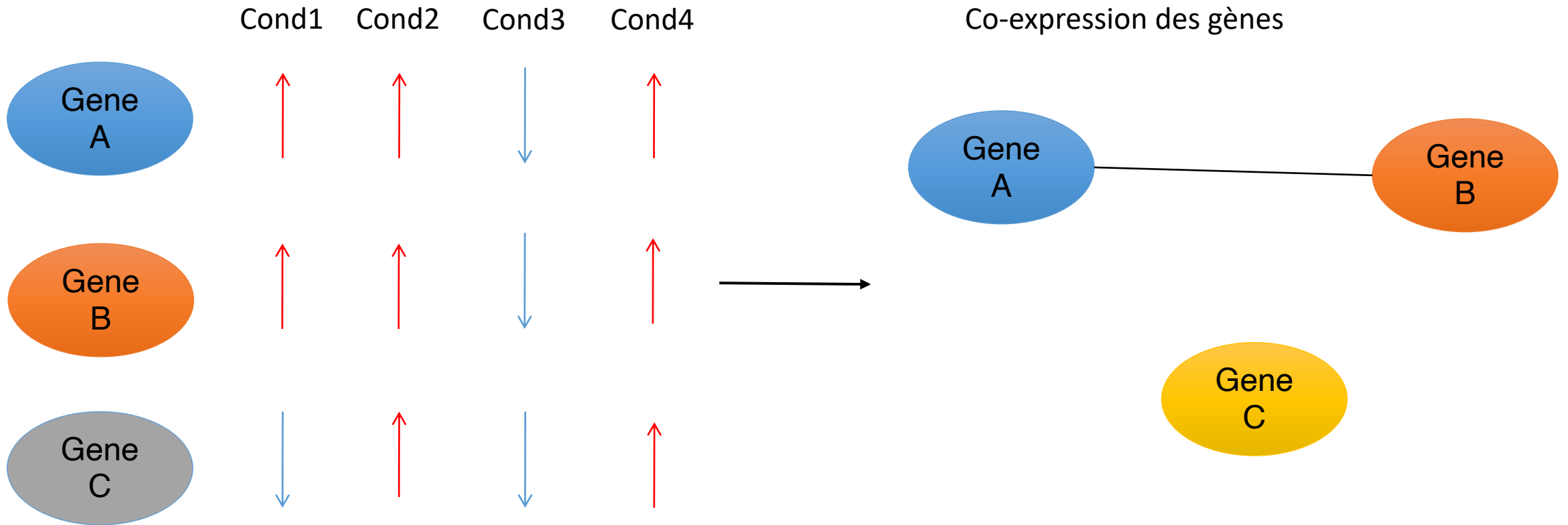


Réseau de co-expression

A partir de données de transcriptomiques on calcule
Une valeur de corrélation pour regrouper
ensemble les gènes ayant un « profile d'expression »
Commun

« Guilt by association »

Réseau de co-expression



Workflow de reconstruction d'un réseau de co-expression

Matrice de comptes genes x samples

	S ₁	S ₂	S ₃
G ₁	43.26	40.89	5.05
G ₂	166.6	41.87	136.65
G ₃	12.53	39.55	42.09
G ₄	28.77	191.92	236.56
G ₅	114.7	79.7	99.76
G ₆	119.1	80.57	114.59
G ₇	118.9	156.69	186.95
G ₈	3.76	2.48	136.78
G ₉	32.73	11.99	118.8
G ₁₀	17.46	56.11	21.41

Gene expression values

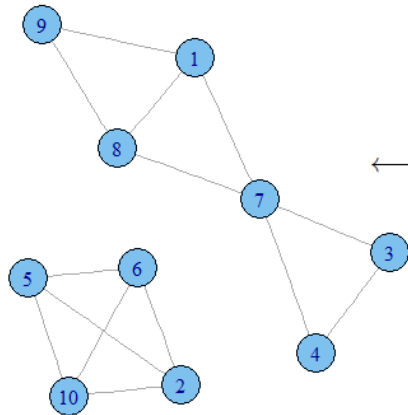
$|r(G_i, G_j)|$
Pearson correlation

	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆	G ₇	G ₈	G ₉	G ₁₀
G ₁	1.00	0.23	0.61	0.71	0.03	0.35	0.86	1.00	0.97	0.37
G ₂	0.23	1.00	0.63	0.52	0.98	0.99	0.29	0.30	0.46	0.99
G ₃	0.61	0.63	1.00	0.99	0.77	0.53	0.93	0.56	0.41	0.51
G ₄	0.71	0.52	0.99	1.00	0.69	0.41	0.97	0.66	0.52	0.40
G ₅	0.03	0.98	0.77	0.69	1.00	0.95	0.48	0.09	0.27	0.94
G ₆	0.35	0.99	0.53	0.41	0.95	1.00	0.17	0.41	0.57	1.00
G ₇	0.86	0.29	0.93	0.97	0.48	0.17	1.00	0.83	0.72	0.16
G ₈	1.00	0.30	0.56	0.66	0.09	0.41	0.83	1.00	0.98	0.42
G ₉	0.97	0.46	0.41	0.52	0.27	0.57	0.72	0.98	1.00	0.58
G ₁₀	0.37	0.99	0.51	0.40	0.94	1.00	0.16	0.42	0.58	1.00

Similarity (Co-expression) score

Matrice de corrélation genes x genes

Graphe du réseau



	G ₁	G ₂	G ₃	G ₄	G ₅	G ₆	G ₇	G ₈	G ₉	G ₁₀
G ₁	0	0	0	0	0	0	1	1	1	0
G ₂	0	0	0	0	1	1	0	0	0	1
G ₃	0	0	0	1	0	0	1	0	0	0
G ₄	0	0	1	0	0	0	1	0	0	0
G ₅	0	1	0	0	0	1	0	0	0	1
G ₆	0	1	0	0	1	0	0	0	0	1
G ₇	1	0	1	1	0	0	0	1	0	0
G ₈	1	0	0	0	0	0	1	0	1	0
G ₉	1	0	0	0	0	0	0	1	0	0
G ₁₀	0	1	0	0	1	1	0	0	0	0

Network adjacency matrix

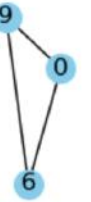
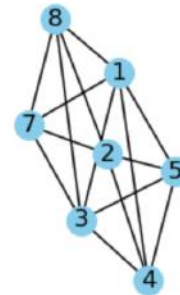
$|r(G_i, G_j)| \geq 0.8$
Significance threshold

Matrice d'adjacence genes x genes

Reconstruction de Reseau de co-expression (exemple en python)

```
# Read the table
data = pd.read_table("CoExpr-Cours3.txt", index_col=0, decimal=',')
# Calculate Pearson correlation coefficients
cor_matrix = data.T.iloc[:, :].corr()
# Define a cutoff value
cutoff = 0.5
# Create an adjacency matrix
adjacency_matrix = np.where(abs(cor_matrix.values) >= cutoff, 1, 0)
np.fill_diagonal(adjacency_matrix, 0)
# Create a network graph from the adjacency matrix
G = nx.Graph(adjacency_matrix)
# Plot the co-expression network
pos = nx.spring_layout(G) # You can choose different layout algorithms
nx.draw(G, pos, with_labels=True, node_size=200, node_color='skyblue')
plt.title("Co-expression Network with cutoff 0.5")
plt.show()
```

Co-expression Network with cutoff 0.5



Reconstruction de Reseau de co-expression (exemple en R)

```
# Read table
data <- read.table(file="CoExpr-Cours3.txt", sep="\t", dec = ",",
                  header = TRUE, row.names = 1)

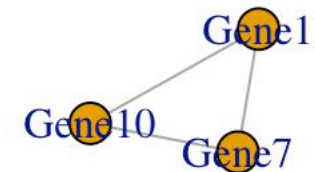
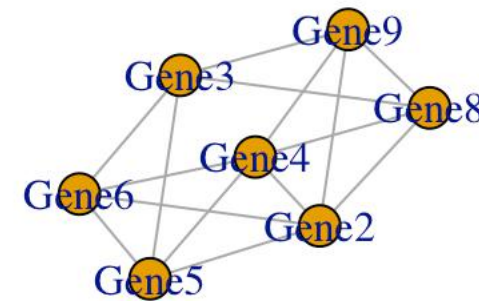
# Calculate Pearson correlation coefficients
cor_matrix <- cor(t(data), method = "pearson")

# Create an adjacency matrix
adjacency_matrix_0_5 <- ifelse(abs(cor_matrix) >= 0.5, 1, 0)
diag(adjacency_matrix_0_5) <- 0

# Create a graph object from the adjacency matrix
graph <- graph_from_adjacency_matrix(adjacency_matrix_0_5, mode = "undirected")

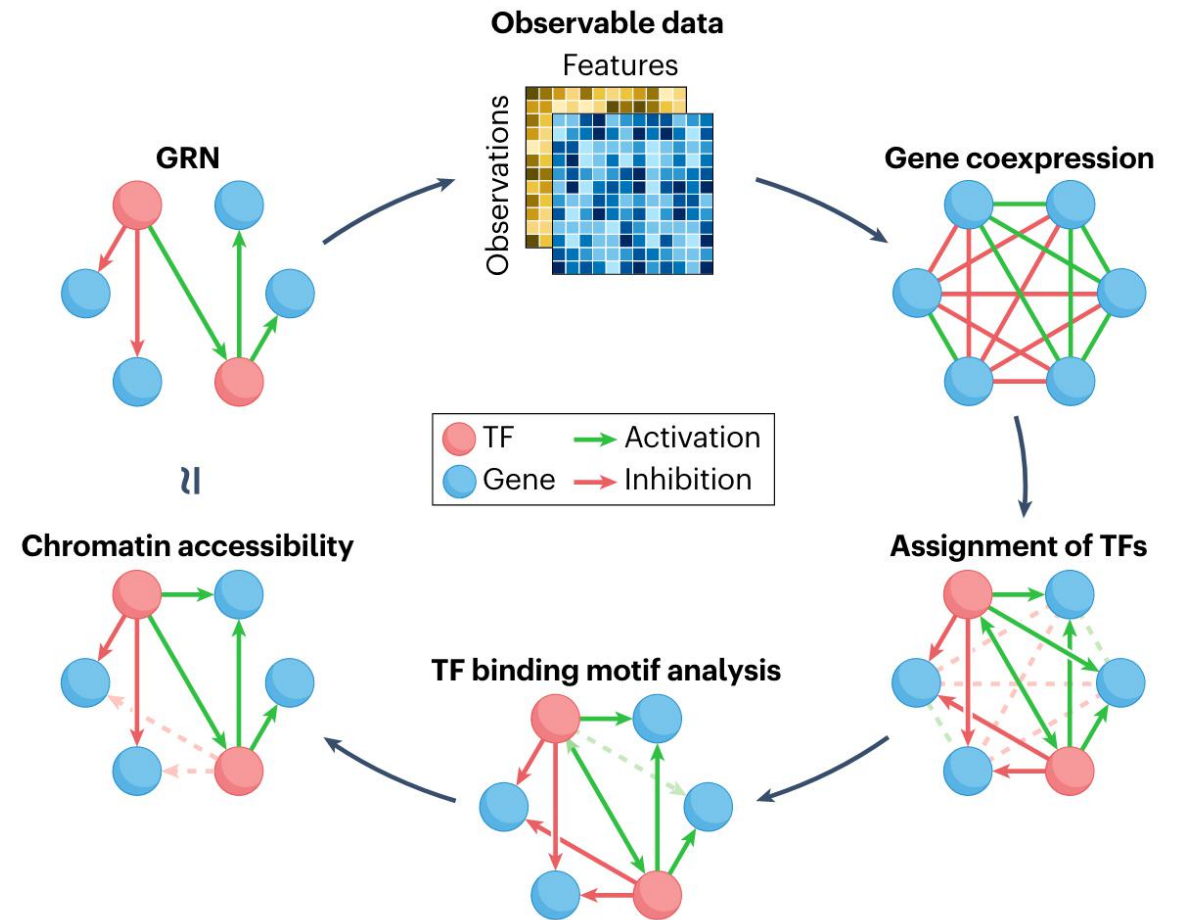
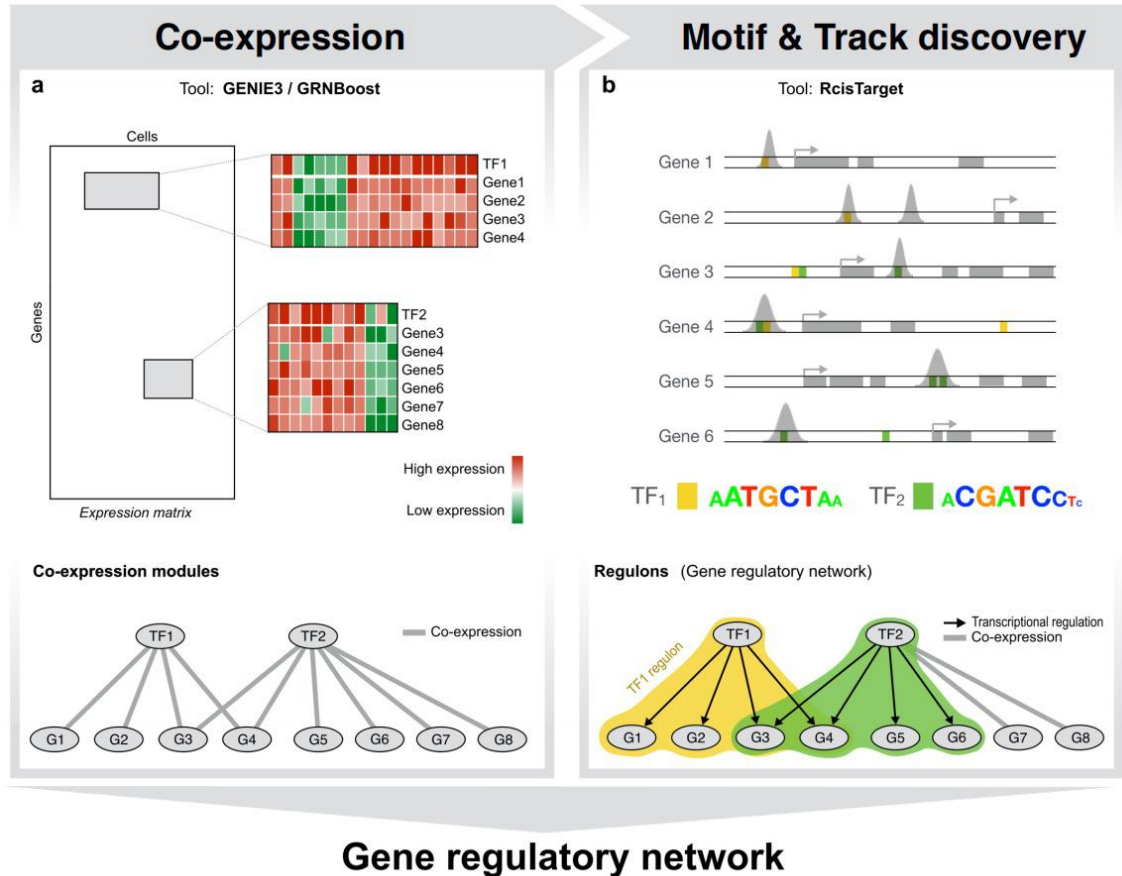
# Plot the network
plot(graph, layout = layout_with_fr, vertex.label = V(graph)$name,
     main="Pearson co-expression with cutoff 0.5")
```

Pearson co-expression with cutoff 0.5



Ajout des facteurs de transcriptions dans la reconstruction

Outil pyScenic



Gene regulatory network inference

nature reviews genetics

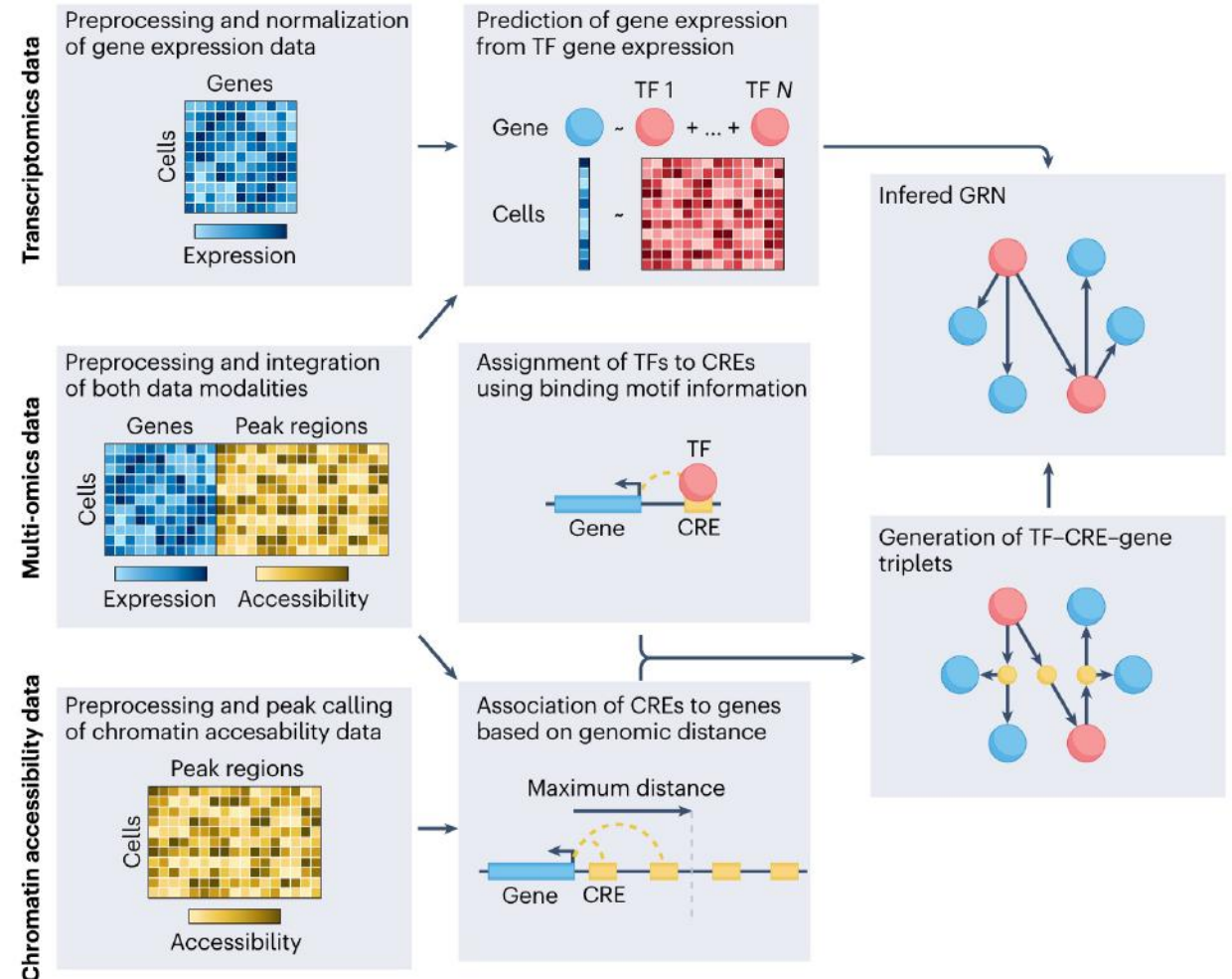
<https://doi.org/10.1038/s41576-023-00618-5>

Review article

Check for updates

Gene regulatory network inference in the era of single-cell multi-omics

Pau Badia-i-Mompel¹, Lorna Wessels^{1,2}, Sophia Müller-Dott¹, Rémi Trimbouret^{1,3}, Ricardo O. Ramirez Flores¹, Ricard Argelaguet⁴ & Julio Saez-Rodriguez¹✉



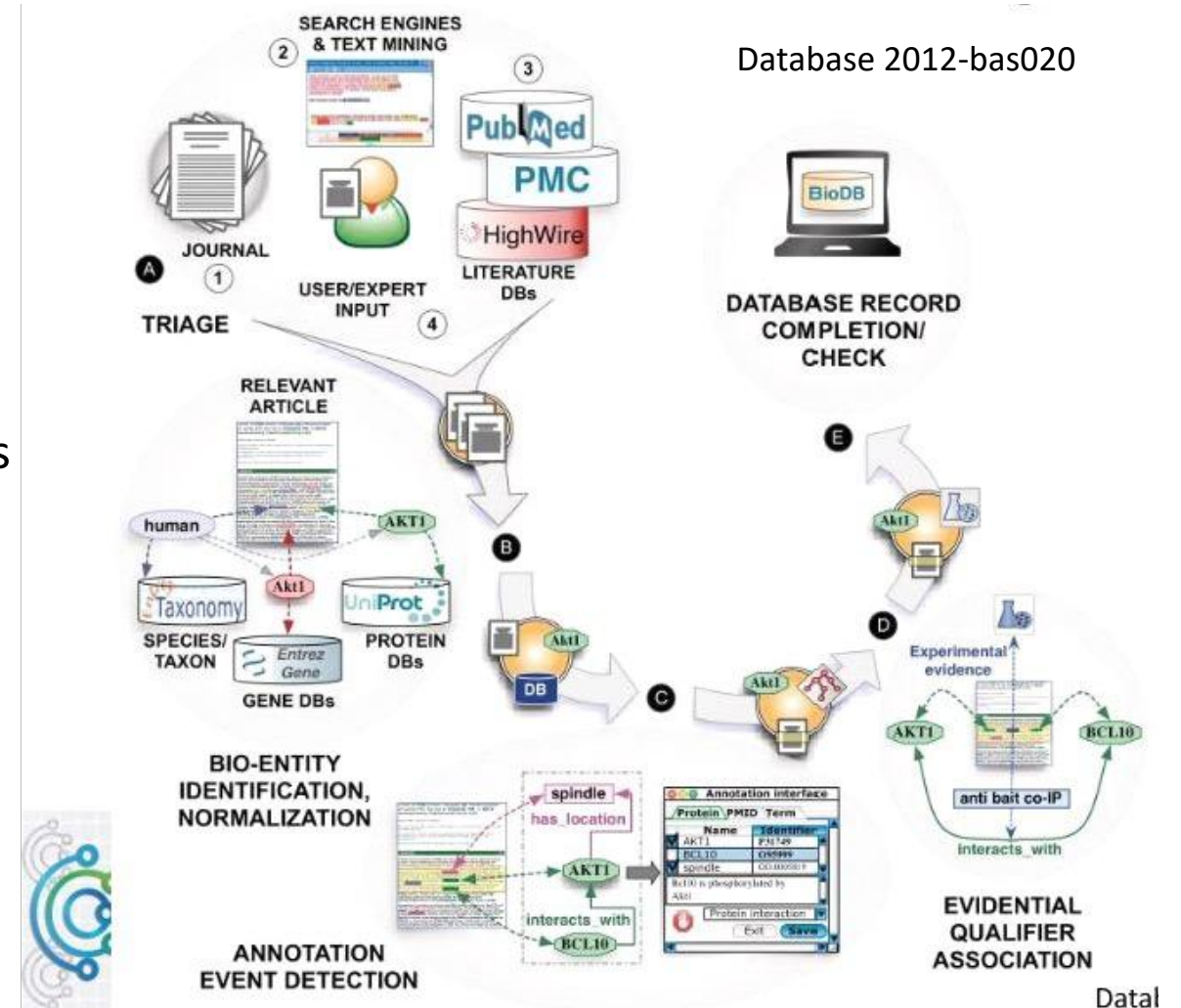
Reconstruction à l'aide d'information génétique

- Gene co-occurrence
- Gene fusion
- Protein Homology
- Synthetic Genetic Array

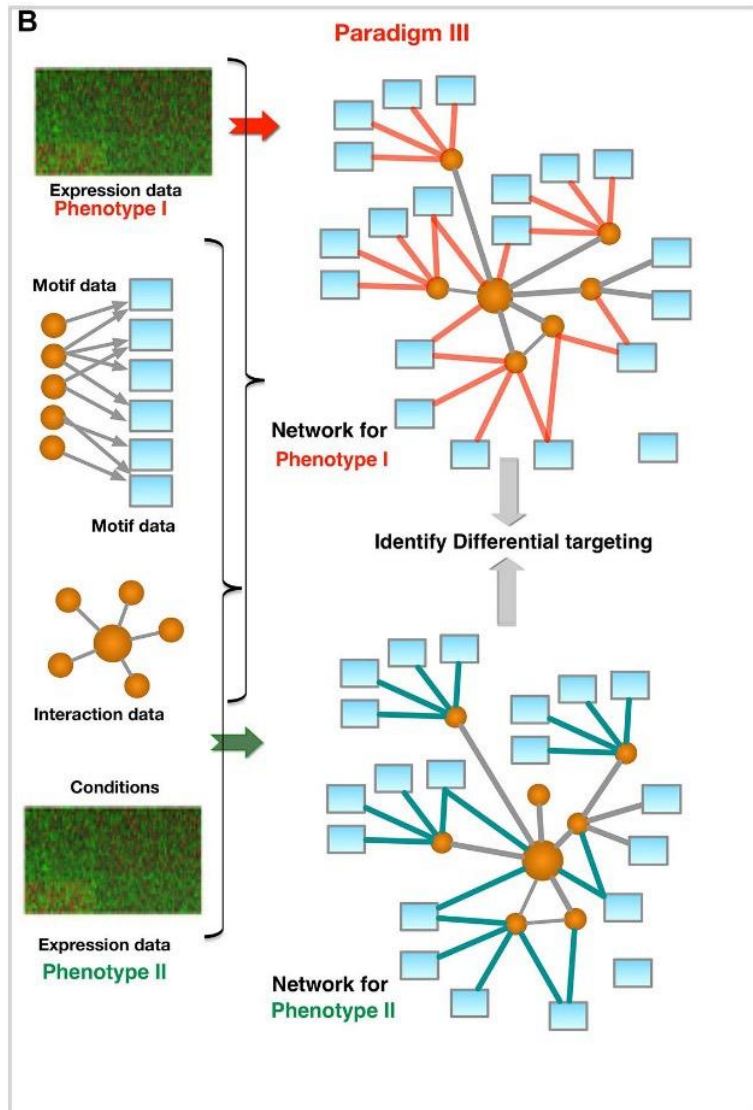
TheCellMap.org: A Web-Accessible Database for Visualizing and Mining the Global Yeast Genetic Interaction Network. Usaj et al., G3 (Bethesda) 2017

Reconstruction indirecte par Text-Mining

- Extraire les mots clés dans les publications et bases de données
- Nettoyer la liste de mots clés = Molécules, gènes, protéines
- Annoter ces éléments avec informations disponibles dans bases de données
- Chercher les co-occurrence dans les publications
- Mettre à jour les bases de données avec ces nouvelles informations



Reconstruction indirecte des réseaux biologiques

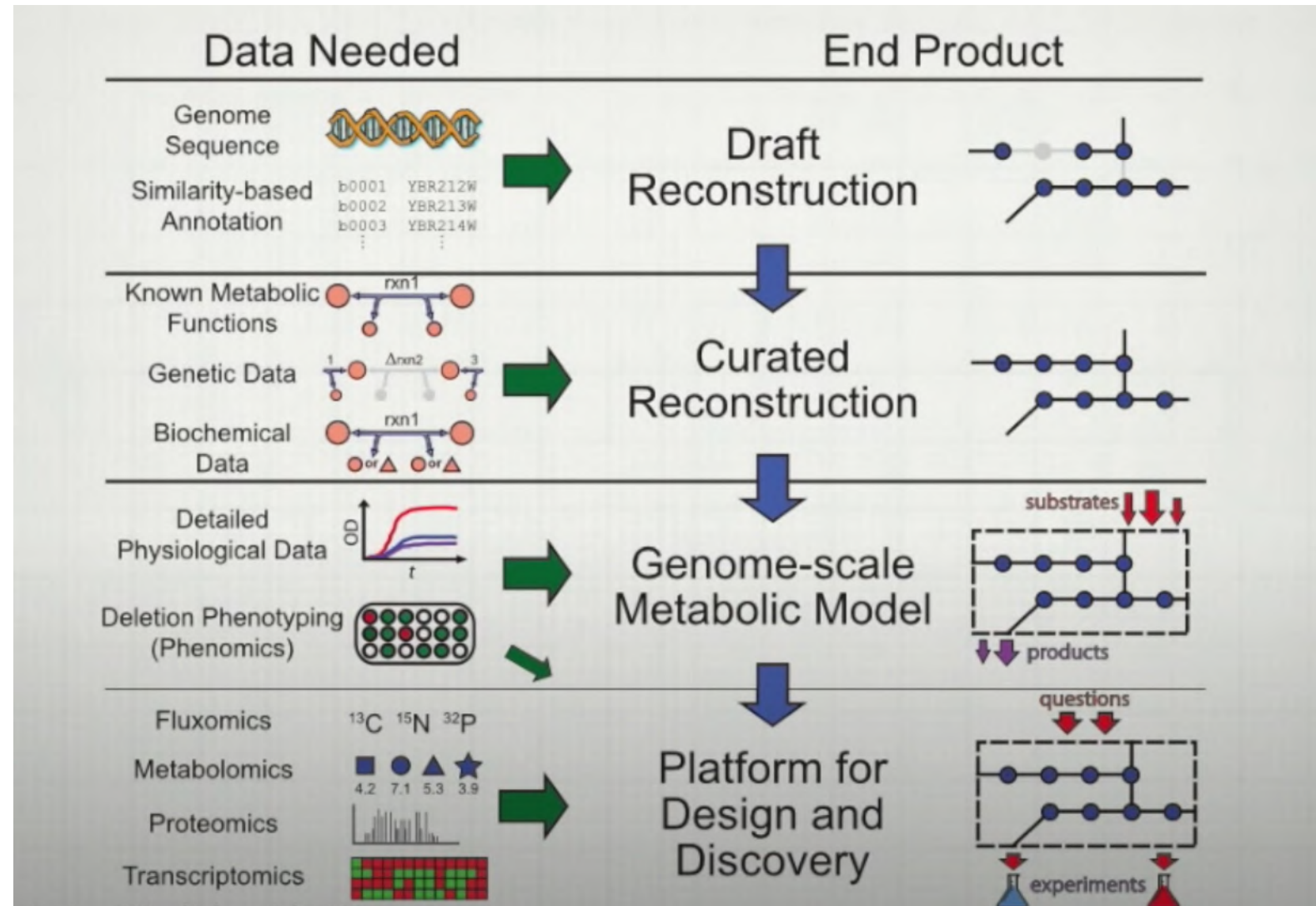


Corrélation de différents réseaux biologiques

On peut par exemple corrélérer un réseau d'interaction protéine-protéine avec un réseau de co-expression des gènes, en ajoutant une information de motif trouvées dans ces gènes

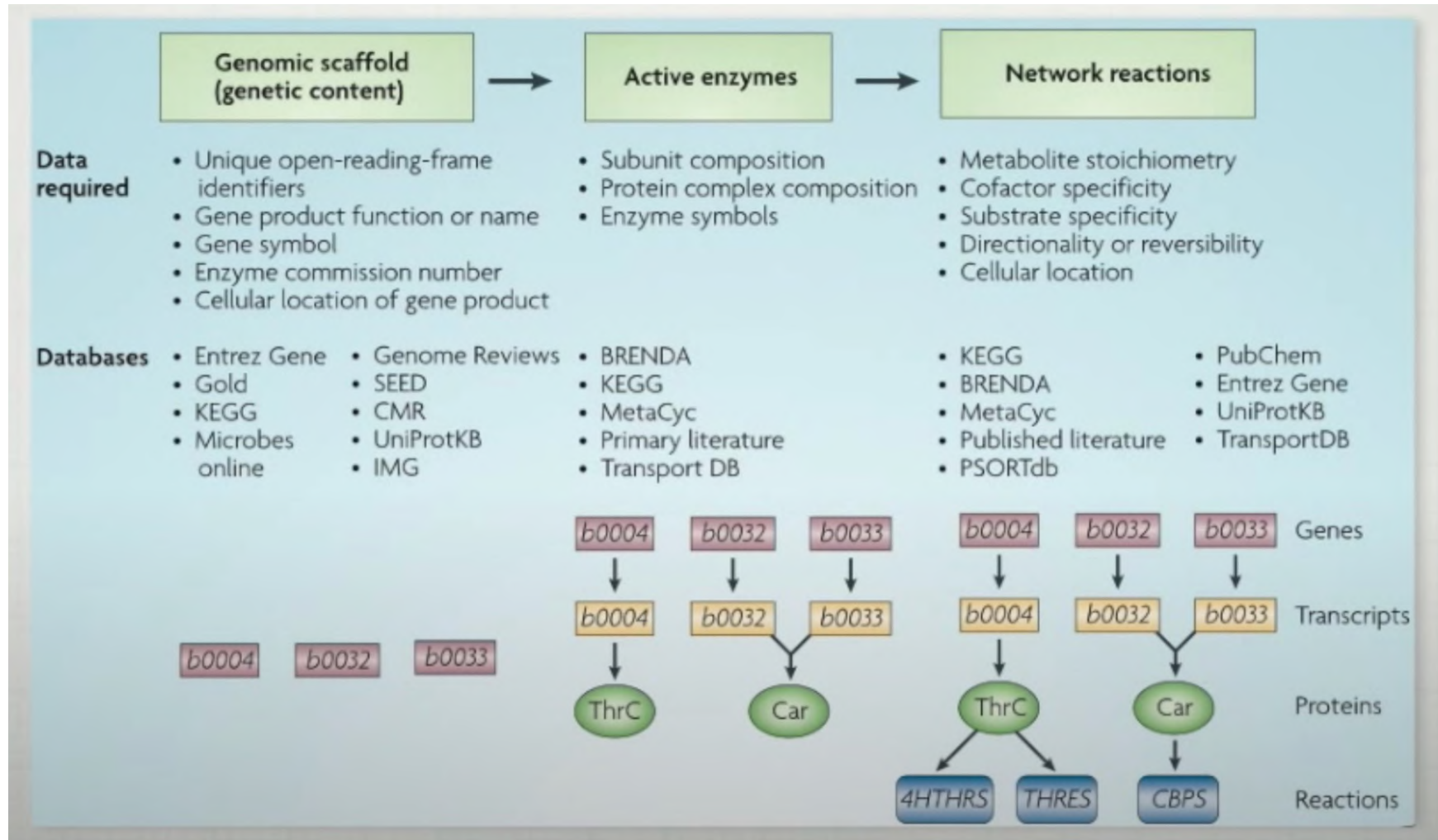
Reconstruction du réseau biochimique d'un organisme

Palsson et al., Nature Review Microbiology, 2009

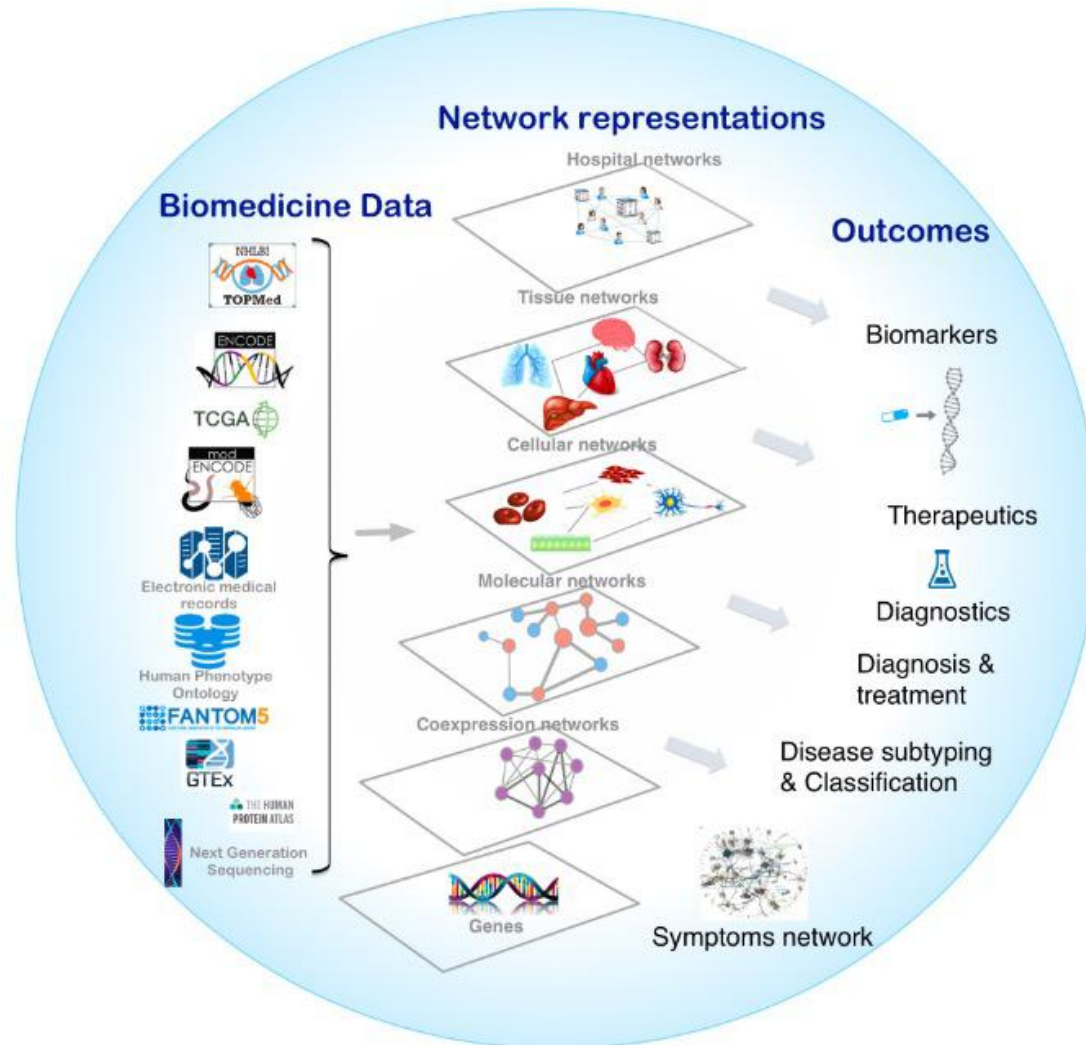


Reconstruction du réseau biochimique d'un organisme

Palsson et al., Nature Review Microbiology, 2009



Reconstruction des réseaux biologiques



Il faut multiplier les types de reconstruction

Il n'y a finalement qu'UN réseau global à reconstruire

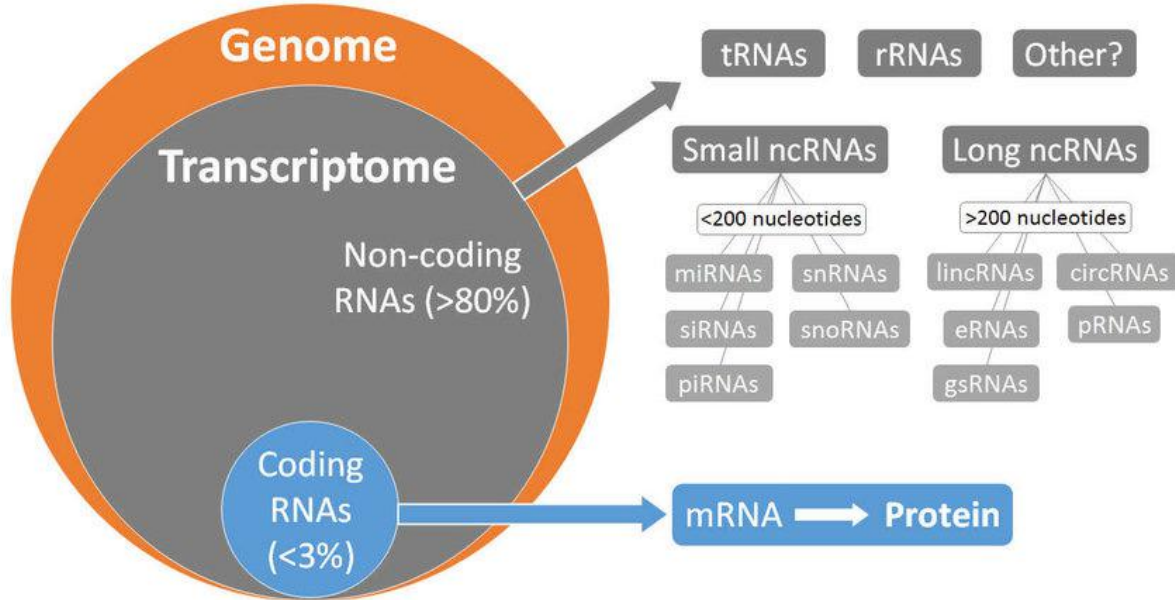
Network Medicine in the Age of Biomedical Big Data
Front. Genet., 11 April 2019

Les obstacles à la reconstruction

- Disponibilité de l'annotation (=les nœuds du réseau)
- Disponibilité des données omiques pour reconstruire les arêtes
- Biais de mesures
- Biais de connaissances
- Impossibilité d'avoir une vision en « instantanée » d'un organisme

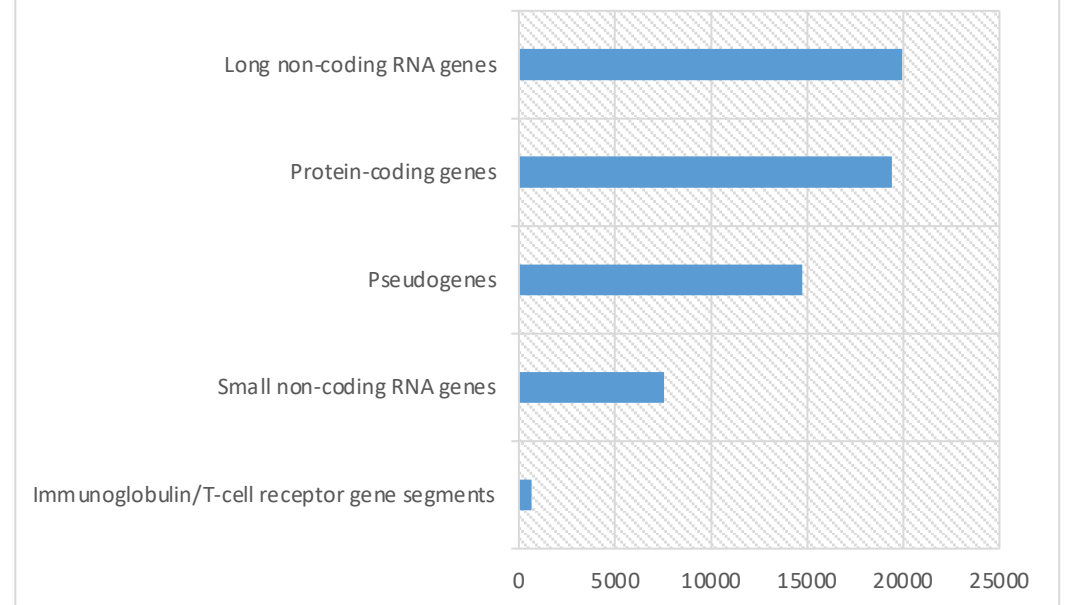
Annotation des génomes

Pedrosa et al., The CardioRNA COST Action 2019



GENCODE 21, Frankish et al., N.A.R. 2021

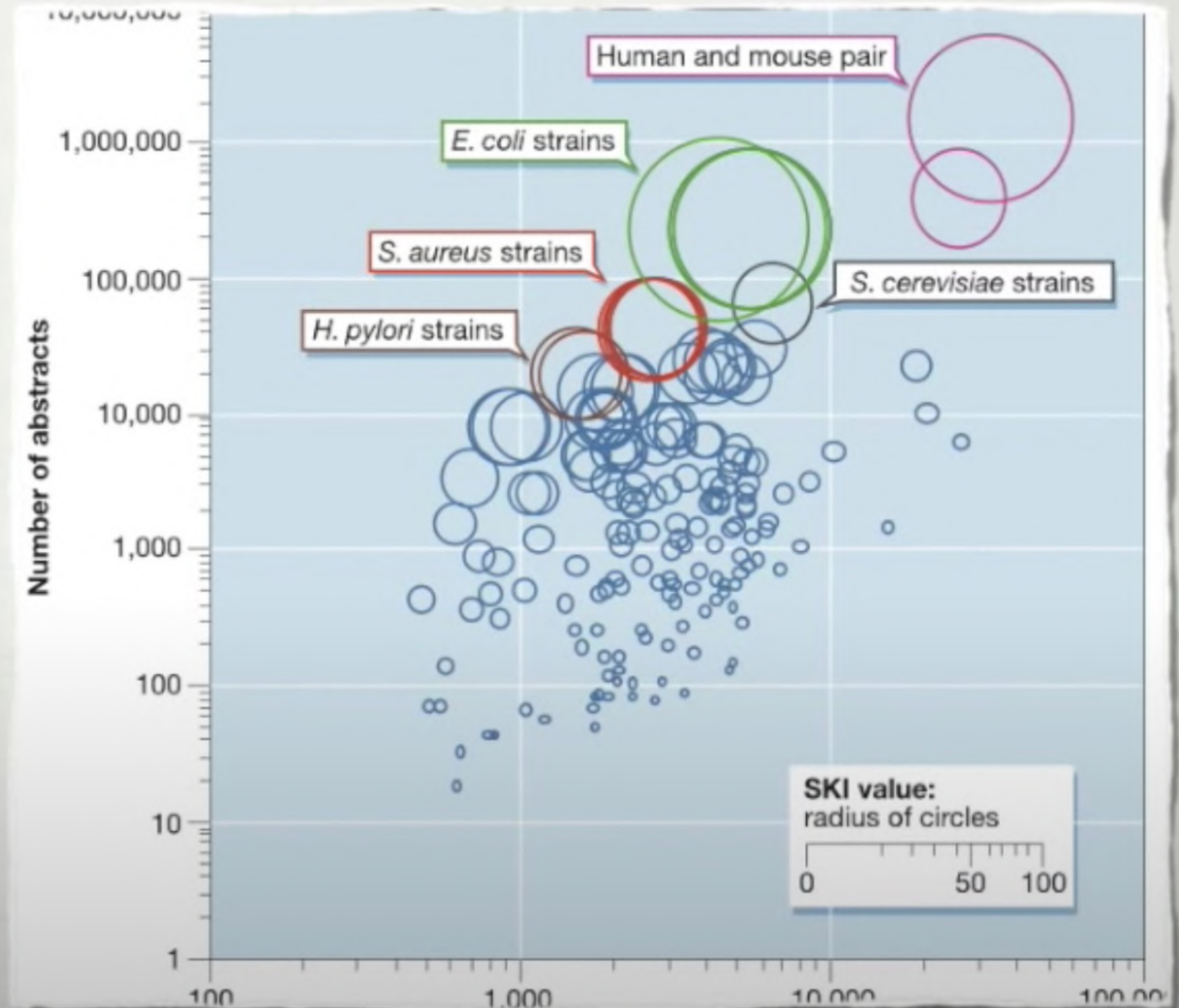
GENCODE Human release (version 43)
62703 genes



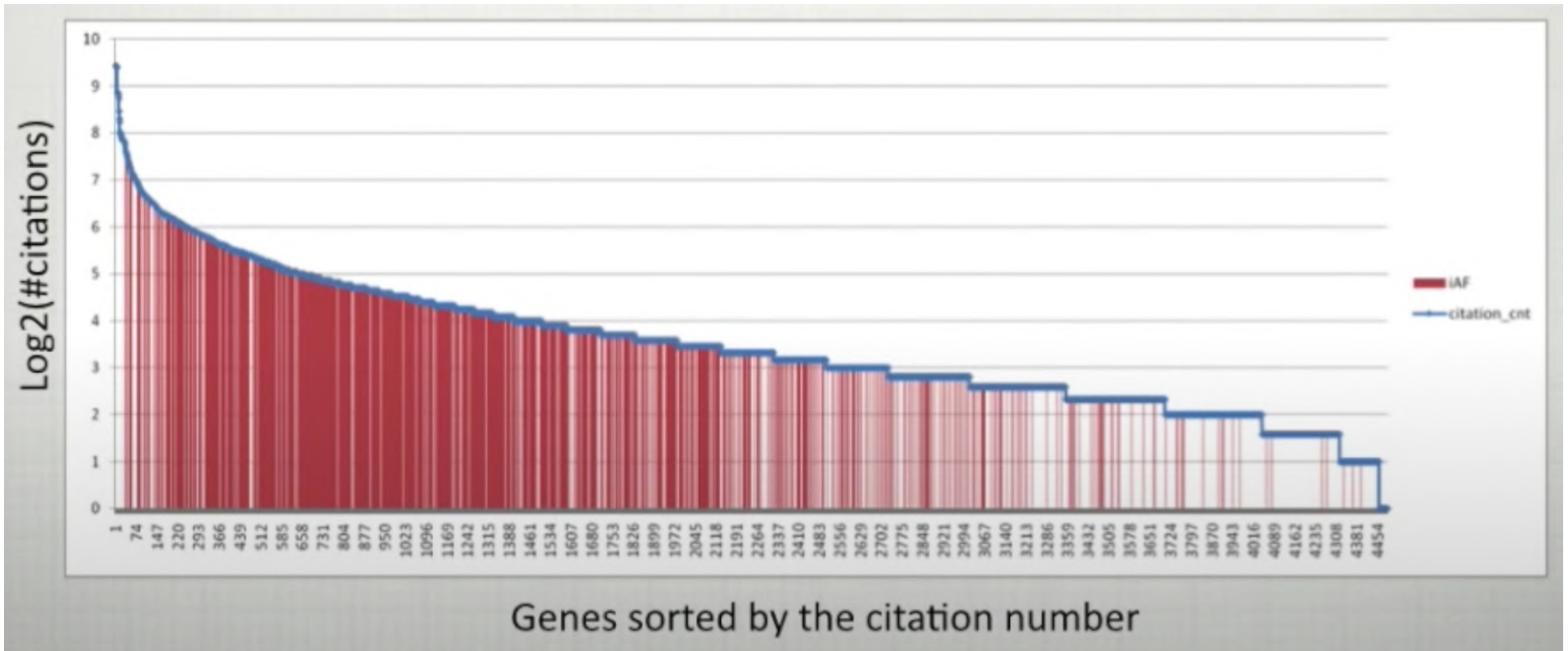
Biais de connaissance dans l'ensemble des gènes

$$\text{SKI} = \frac{\text{No. Abstracts}}{\text{No. Genes}}$$

<i>E. coli</i>	55.1
Human	48.5
<i>S. aureus</i>	16-17
Mouse	15.6
<i>H. pylori</i>	13
<i>S. cerevisiae</i>	10.6



Biais de connaissance dans l'ensemble des gènes

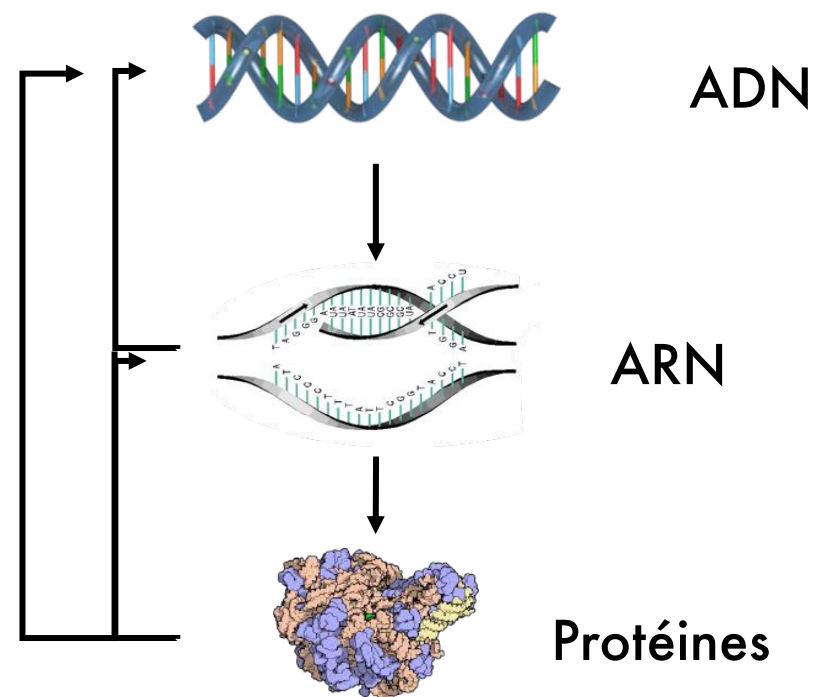


Les différences d'échelles temporelles

Division, Replication, Transcription, Translation & Degradation Rates

at 37°C with a temperature dependence Q10 of $\approx 2-3$

9. Cell cycle time (exponential growth in rich media): *E. coli* $\approx 20-40$ min; yeast 70-140 min; human cell line (Hela): 15-30 hours
10. Rate of replication by DNA polymerase *E. coli* $\approx 200-1000$ bases/s; human ≈ 40 bases/s. Transcription by RNA polymerase 10-100 bases/s
11. Translation rate by ribosome 10-20 aa/s
12. Degradation rates (proliferating cells): mRNA half life < cell cycle time; protein half life \approx cell cycle time



Problème du multi-omique

Transcriptomique vs Protéomique

Transcriptomique

- Seuil de détection bas
- Si la molécule est présente on doit la détecter
- On peut amplifier le signal facilement (PCR)

On mesure une grande majorité des ARNs présents dans l'échantillon

Protéomique

- Seuil de détection plus haut
- Certaines molécules ne vont pas être détectées
- Plus compliqué d'amplifier le signal

On mesure moins de la moitié des protéines présentes dans l'échantillon

Comment corréler les échelles omiques dans ce contexte ?

Comparing protein abundance and mRNA expression levels on a genomic scale

[Dov Greenbaum](#), [Christopher Colangelo](#), [Kenneth Williams](#) ✉ & [Mark Gerstein](#) ✉

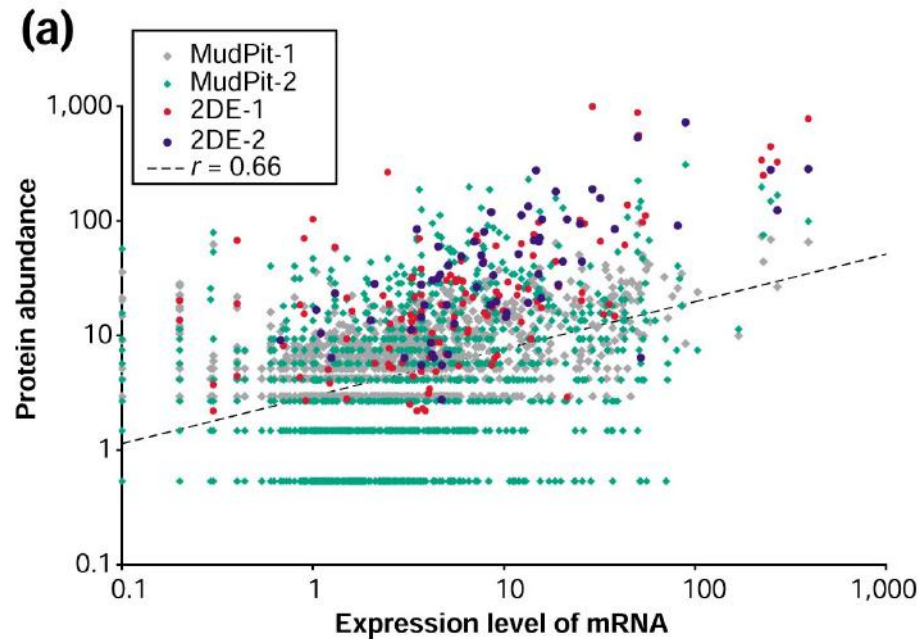
[Genome Biology](#) **4**, Article number: 117 (2003) | [Cite this article](#)

67k Accesses | 1144 Citations | 4 Altmetric | [Metrics](#)

Multi-omique

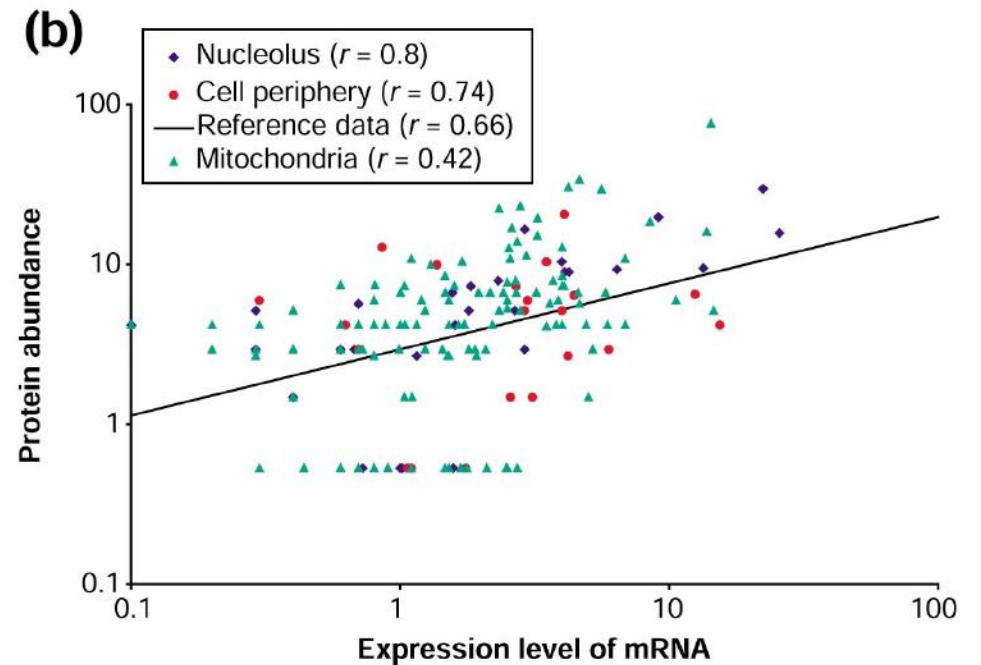
$R = 0.66$

Il s'agit de la corrélation globale calculées sur toutes les protéines



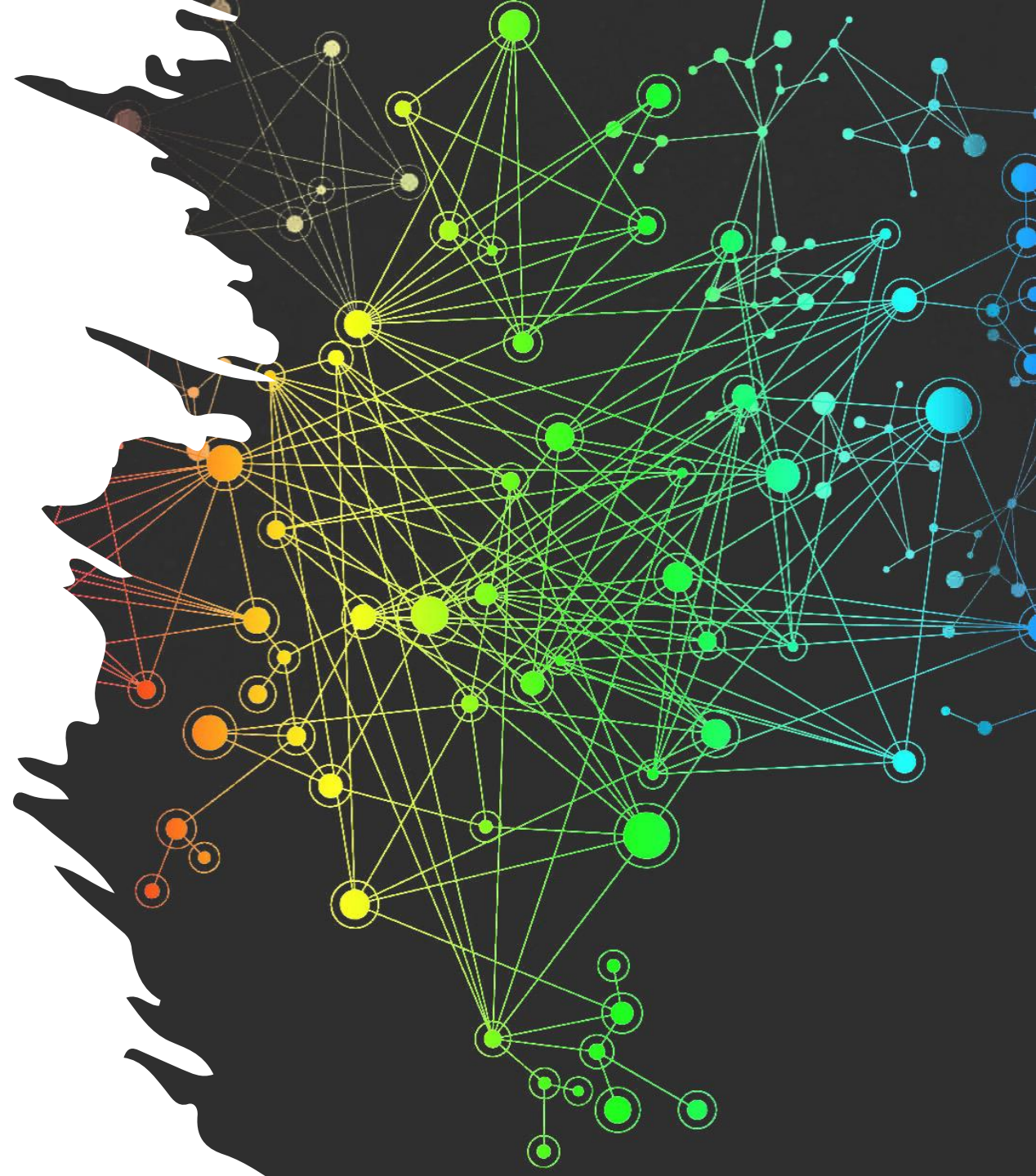
$0.42 < R < 0.8$

Pour certains groupes GO la corrélation augmente

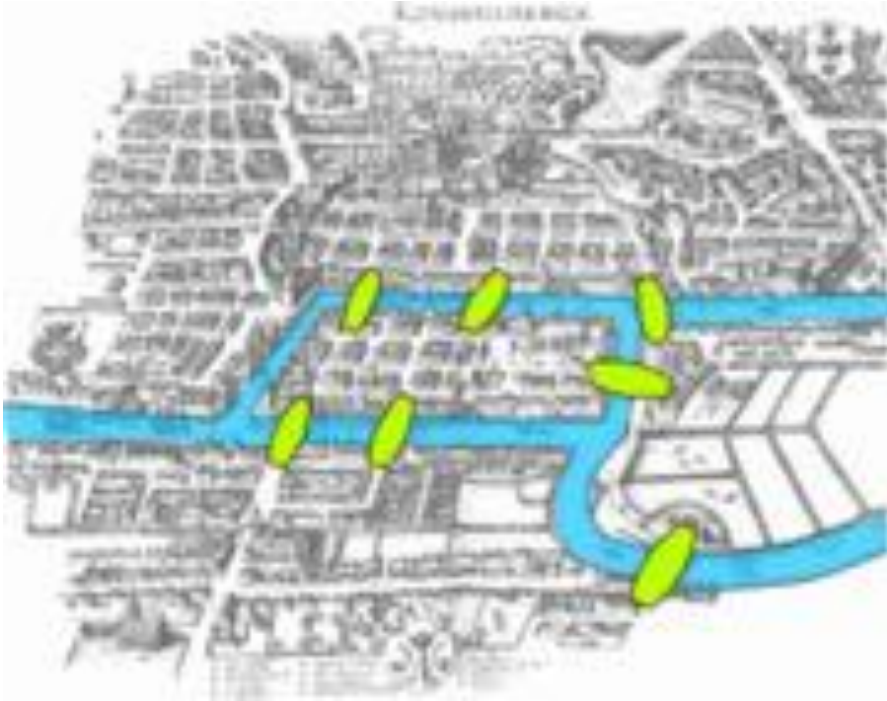


Biologie des systèmes

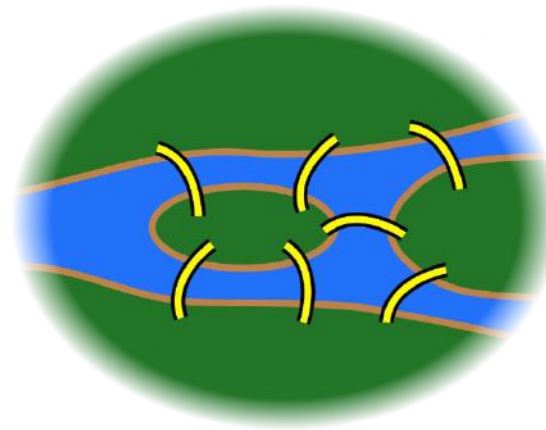
- L'approche systémique en biologie
- Reconstruire un réseau biologique
 - Les différents types de réseaux
 - Reconstruction directe
 - Reconstruction indirecte
 - Les obstacles à la reconstruction
- Analyser un réseau biologique
 - Théorie des graphes
 - Topologie des réseaux
 - Les réseaux sans échelles
- Utiliser un réseau biologique
 - Etat de l'art des réseaux biologiques les plus étendus
 - Bases de données en biologie des systèmes
- Modéliser un réseau biologique



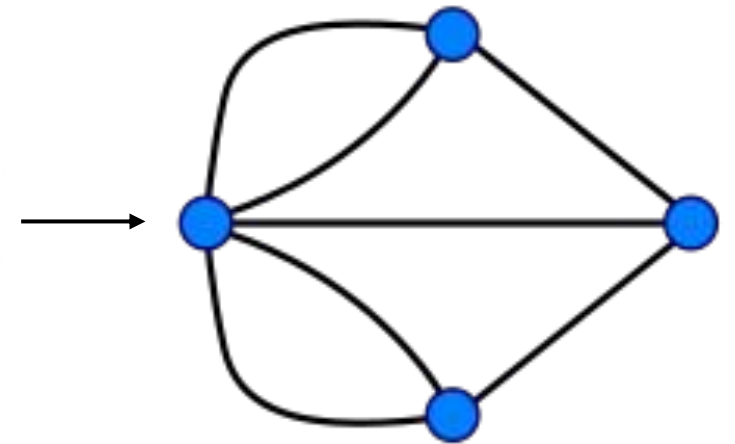
Théorie des graphes



Le Problème des sept ponts de Königsberg
Leonhard Euler - 1735



Réponse : Non !



Théorie des graphes

Existe-t-il une solution, à partir d'un point de départ au choix,
de passer une et une seule fois par chaque pont, et de revenir à son point de départ ?

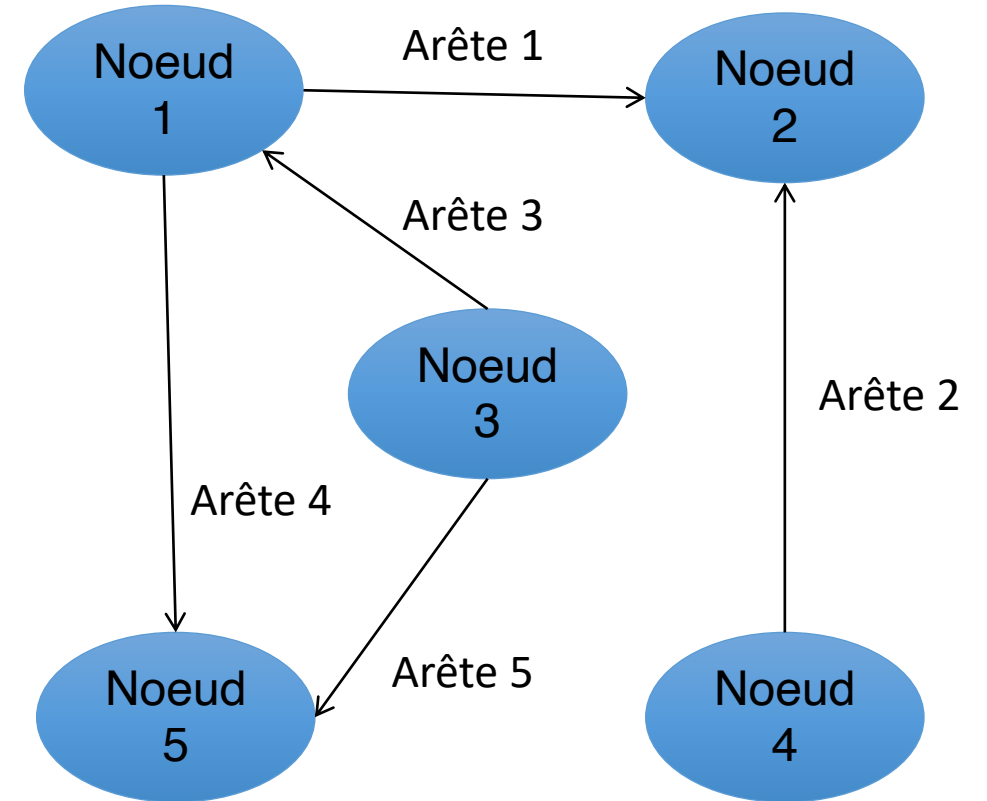
Définition d'un graphe orienté avec boucles

Un graphe est un triplet $G = (V, E, \Phi)$

- V est un ensemble de nœuds (sommets, points)
- E est un ensemble d'**arêtes orientées (flèches)**
- Φ est une fonction d'incidence décrivant chaque

arête orientée comme une paire de sommets

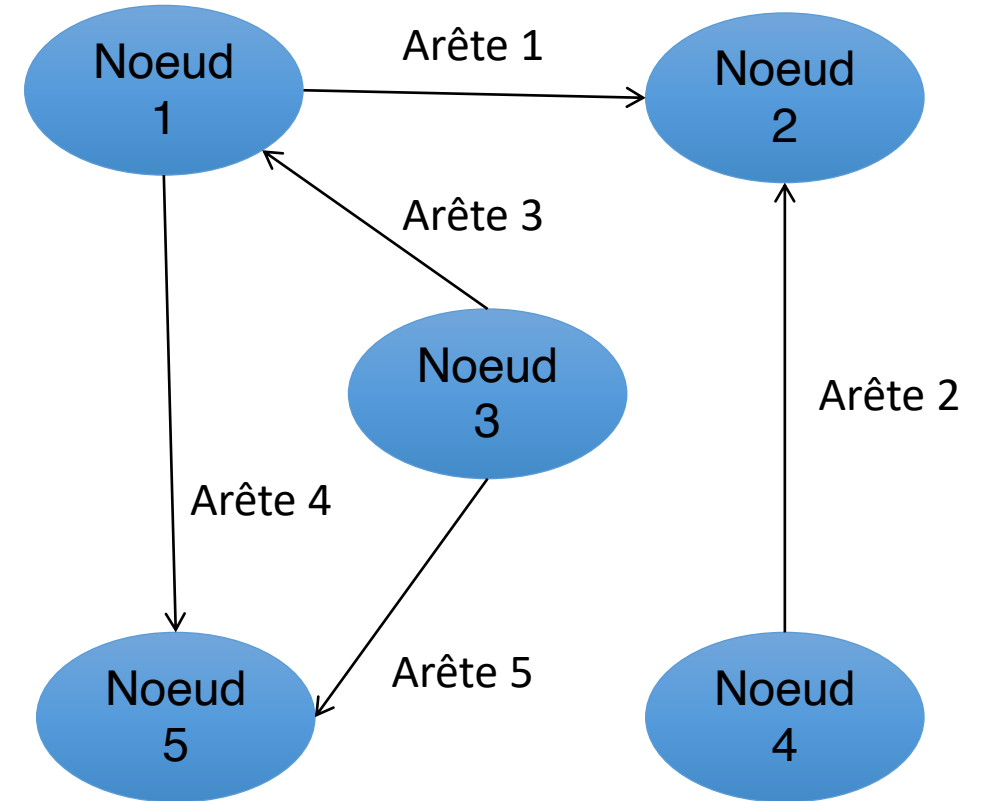
$$\Phi: E \rightarrow \{ \{x, y\} \mid (x, y) \in V^2 \}$$



Théorie des graphes

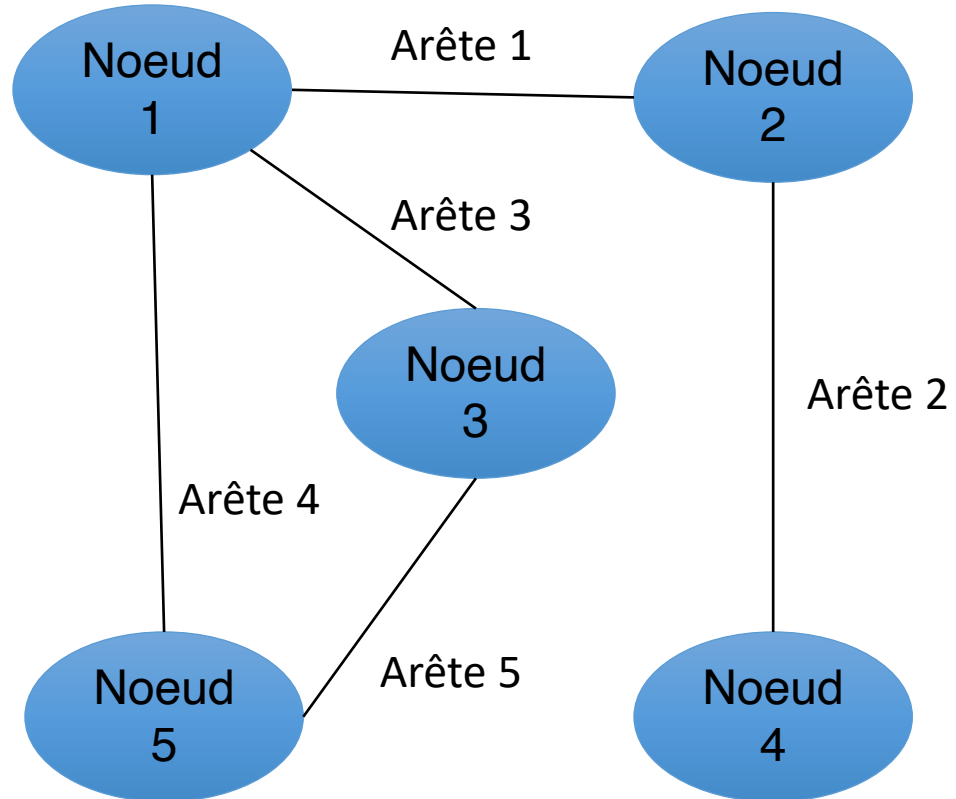
Définitions

- **L'ordre** d'un graphe est son nombre de noeuds
- Deux noeuds reliés par une arête sont **adjacents**
- Le **degré** d'un noeud est le nombre d'arêtes dont ce noeud est une extrémité
 - (exemple : Le noeud 1 a pour degré 3)
- Une **chaîne** est une suite d'arêtes reliant deux noeuds
 - (exemple : La chaîne {Arête 3, Arête 2, Arête 1} qui relie le noeud 3 au noeud 4)
- La **longueur d'une chaîne** est le nombre d'arêtes qui la composent
- Un **cycle** est une chaîne dont les extrémités coïncident.



Matrice d'adjacence d'un graphe

Visualisation graphique



Matrice d'adjacence

Matrice A de taille $n \times n$
où n est l'ordre du graphe (nb de nœuds)

	No1	No2	No3	No4	No5
No1	0	1	1	0	1
No2	1	0	0	1	0
No3	1	0	0	0	1
No4	0	1	0	0	0
No5	1	0	1	0	0

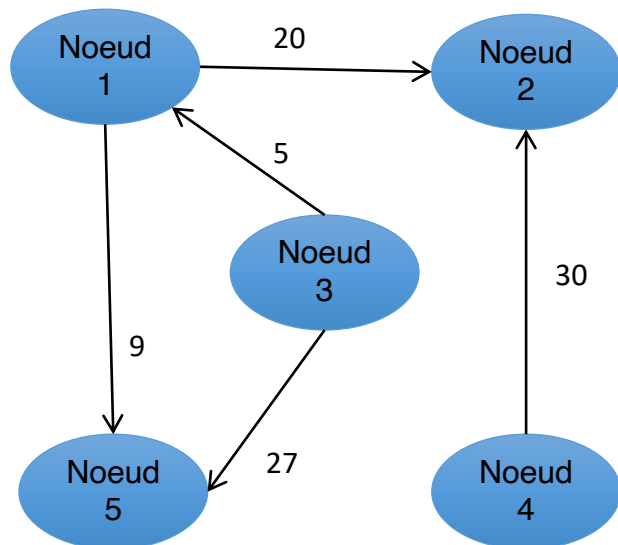
$$A = [a_{i,j}] = \text{nb. arêtes } \{i, j\}$$

Les graphes pondérés

Un graphe est dit pondéré si ses arêtes sont affectées d'un nombre positif

- La pondération est nommée le **poids d'une arête**
- **Le poids d'une chaîne** est la somme des poids des arêtes

Une plus petite chaîne entre deux sommets est la chaîne avec le poids minimal



Matrice d'adjacence

Matrice A de taille n x n

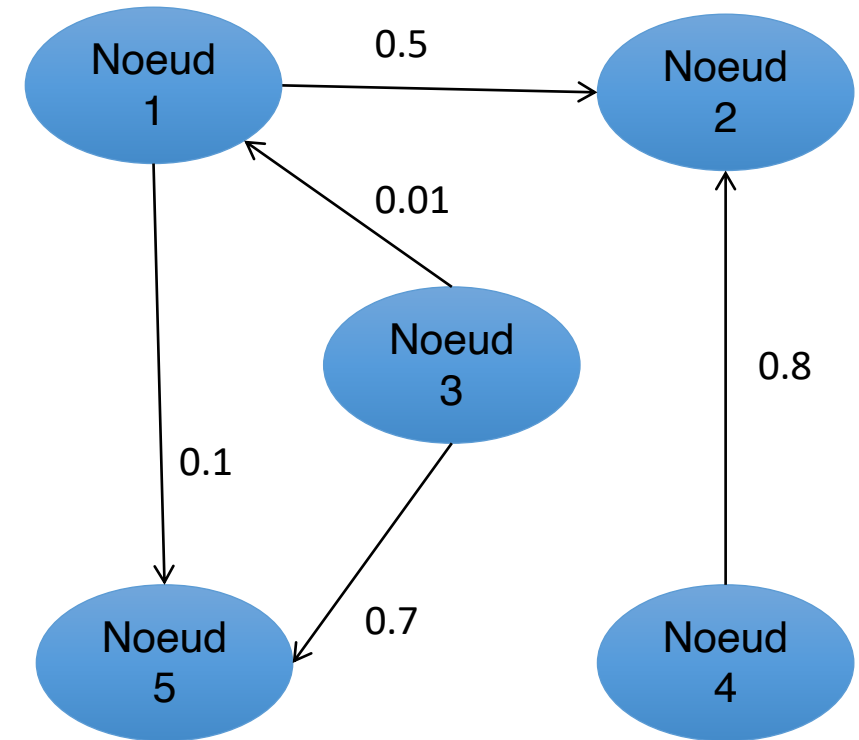
où n est l'ordre du graphe (nb de nœuds)

	No1	No2	No3	No4	No5
No1	0	20	0	0	9
No2	0	0	0	0	0
No3	5	0	0	0	27
No4	0	30	0	0	0
No5	0	0	0	0	0

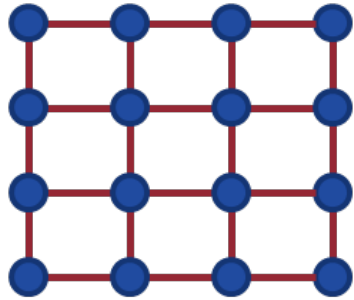
$$A=[a_{i,j}] = nb. arêtes \{i,j\}$$

Quelques Problèmes classiques en théorie des graphes

- la chaîne (ou chemin) la plus courte entre deux nœuds ?
- la chaîne (ou chemin) la plus longue entre deux nœuds ?
- Problème de coloration de graphe ?
- La distribution des degrés des nœuds ?

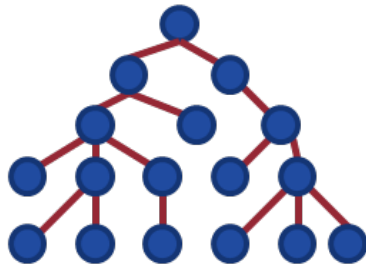


Topologie des réseaux



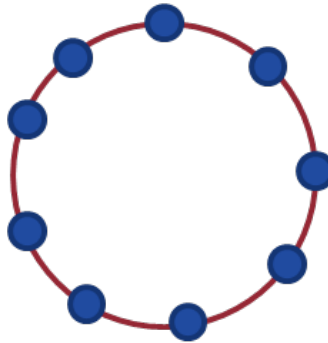
1

Graphe homogènes



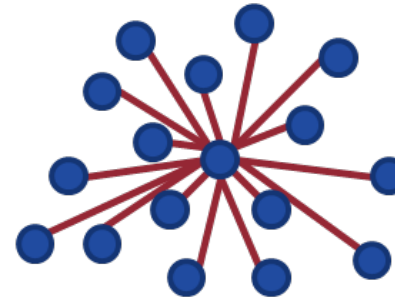
2

Graphe hiérarchique



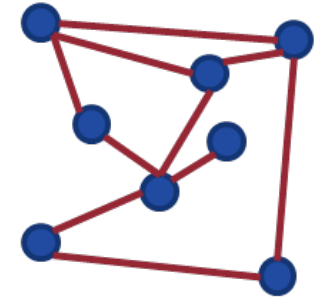
3

Graphe cyclique



4

Graphe centralisé

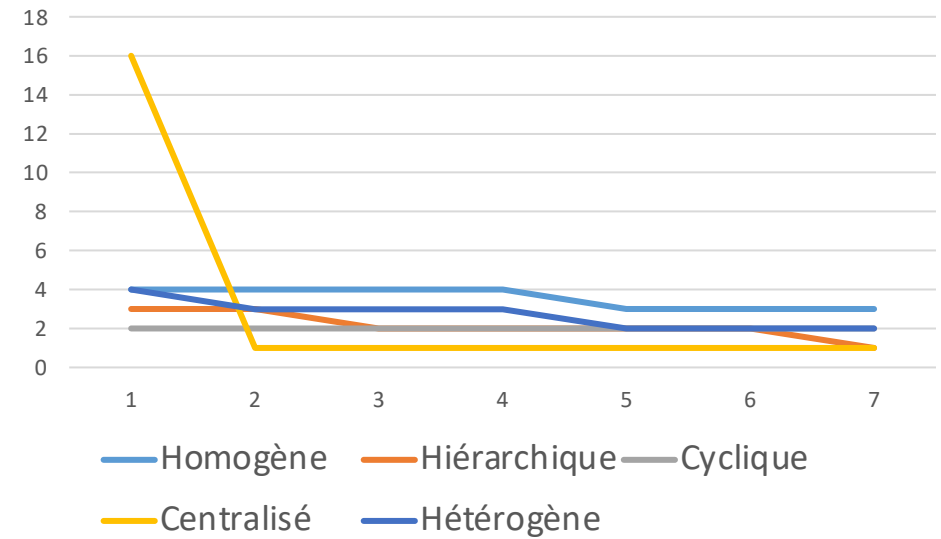
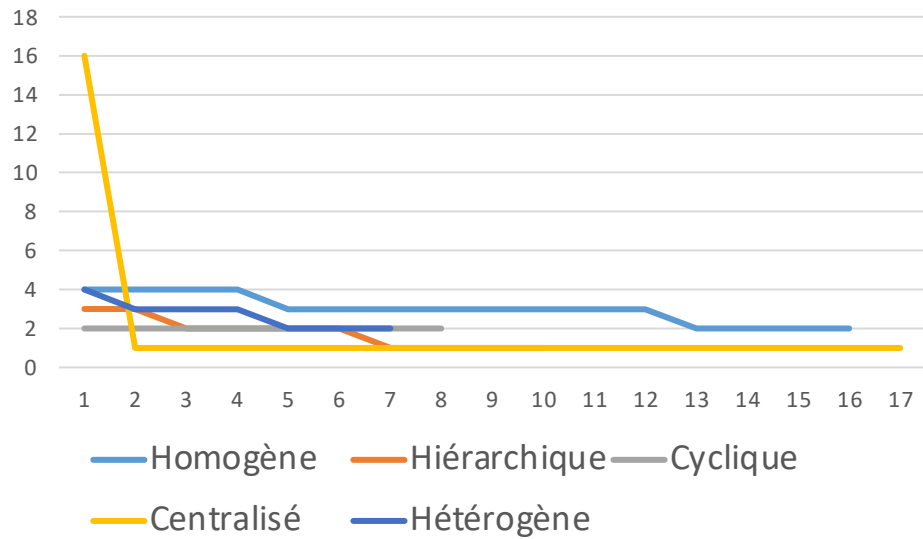
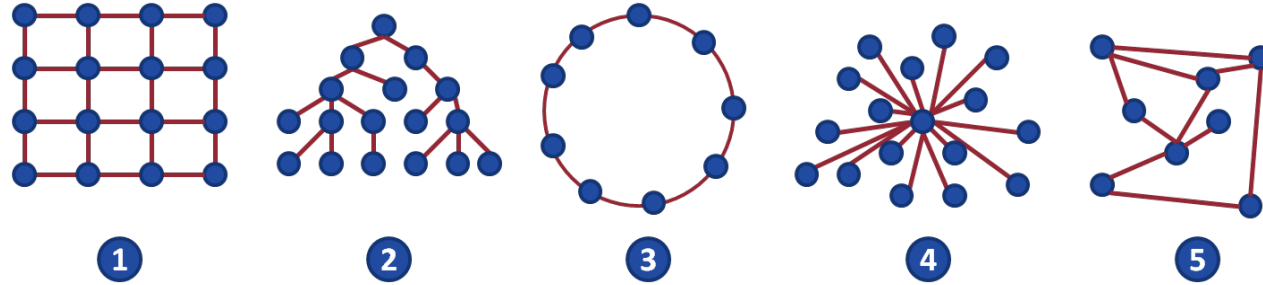


5

Graphe hétérogène

Mesure de la topologie

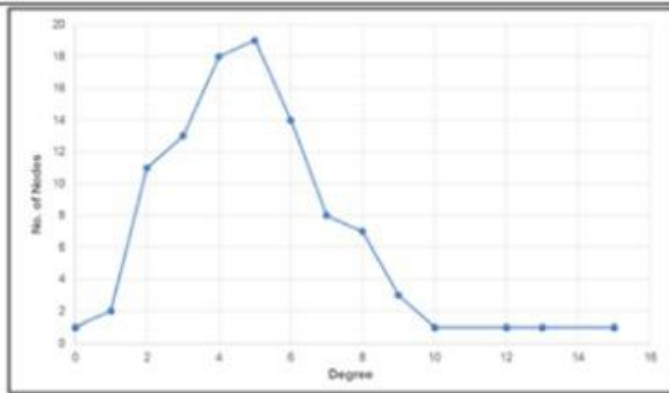
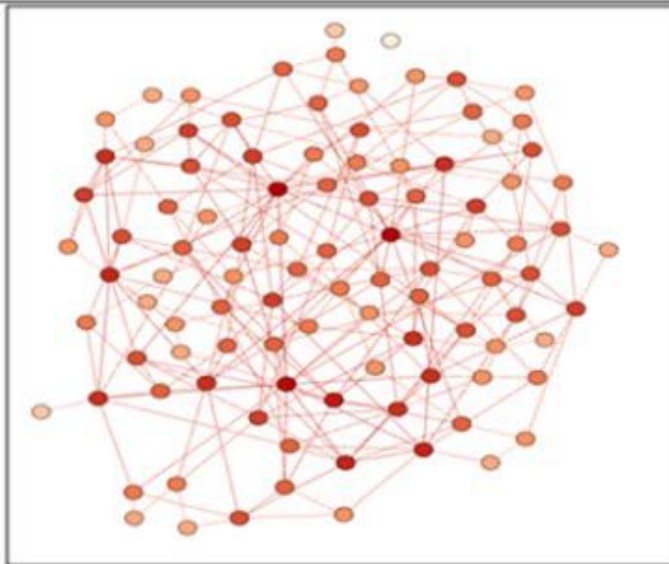
distribution des degrés



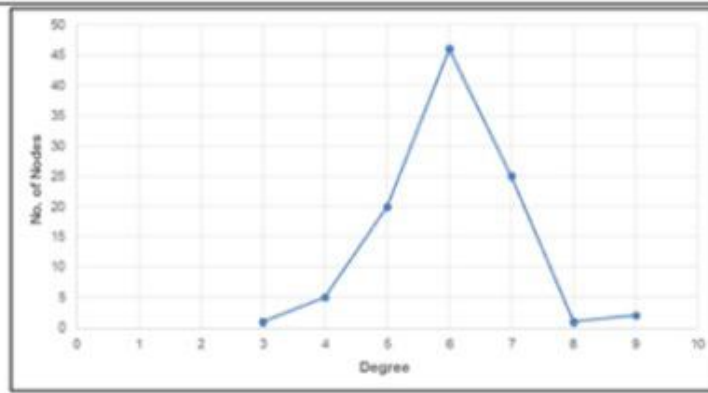
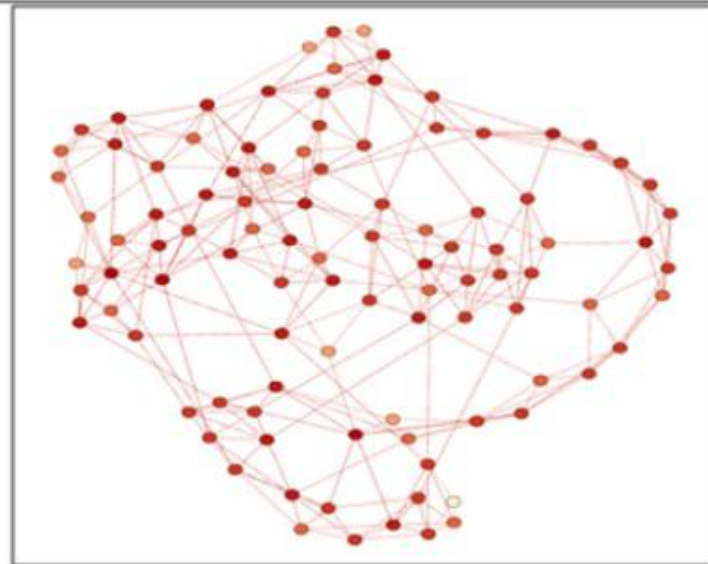
Mesure de la topologie

distribution des degrés

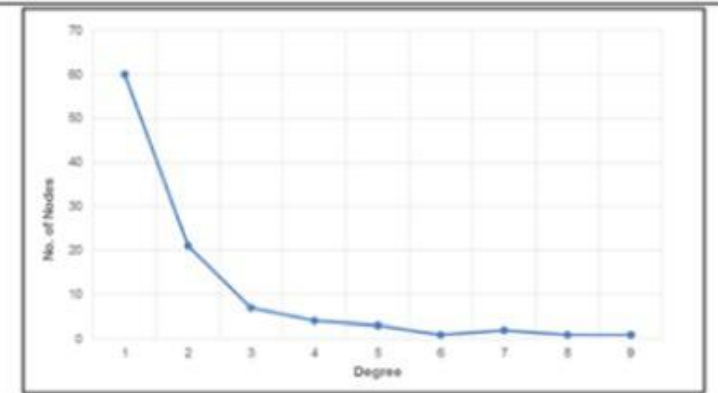
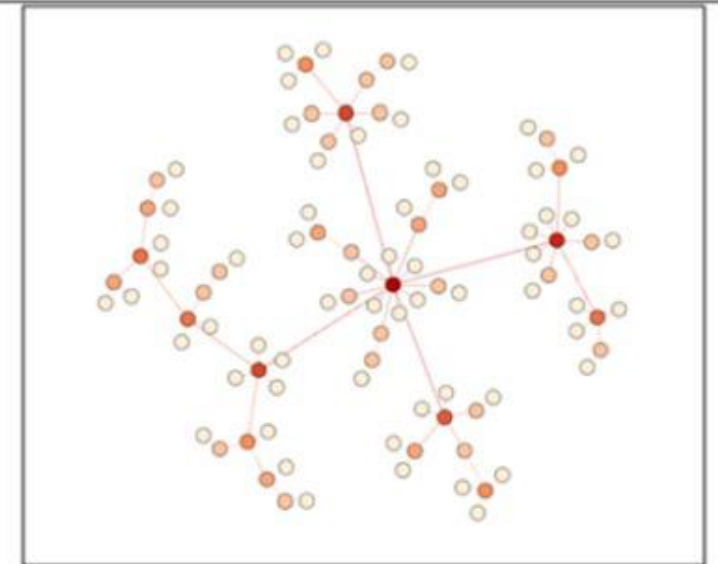
Réseau aléatoire



Réseau small world

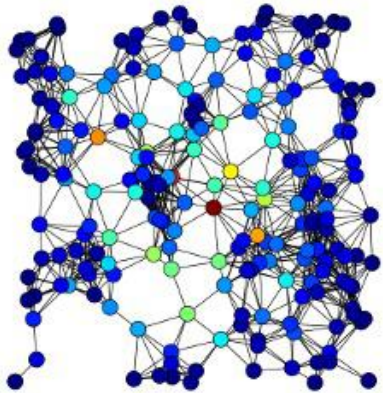


Réseau centralisé
(sans échelles)

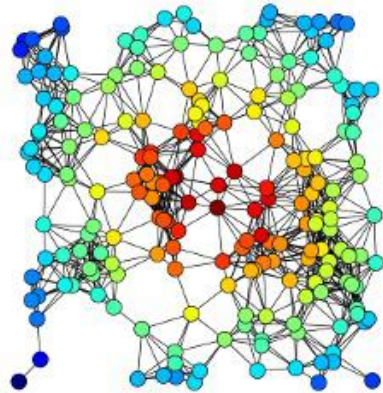


Mesure de la topologie

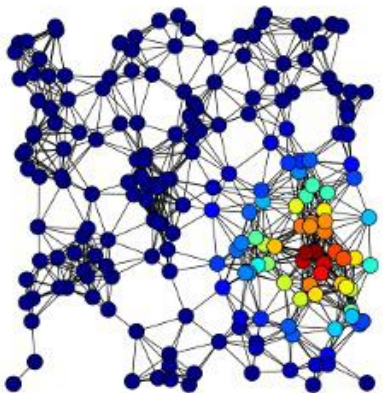
Notion de centralité



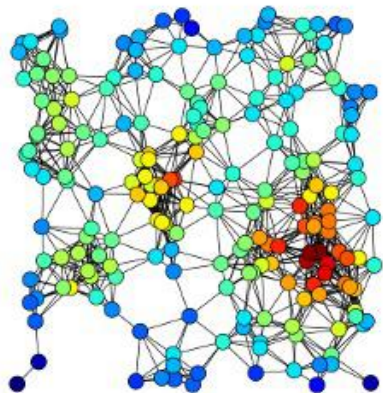
A



B



C



D

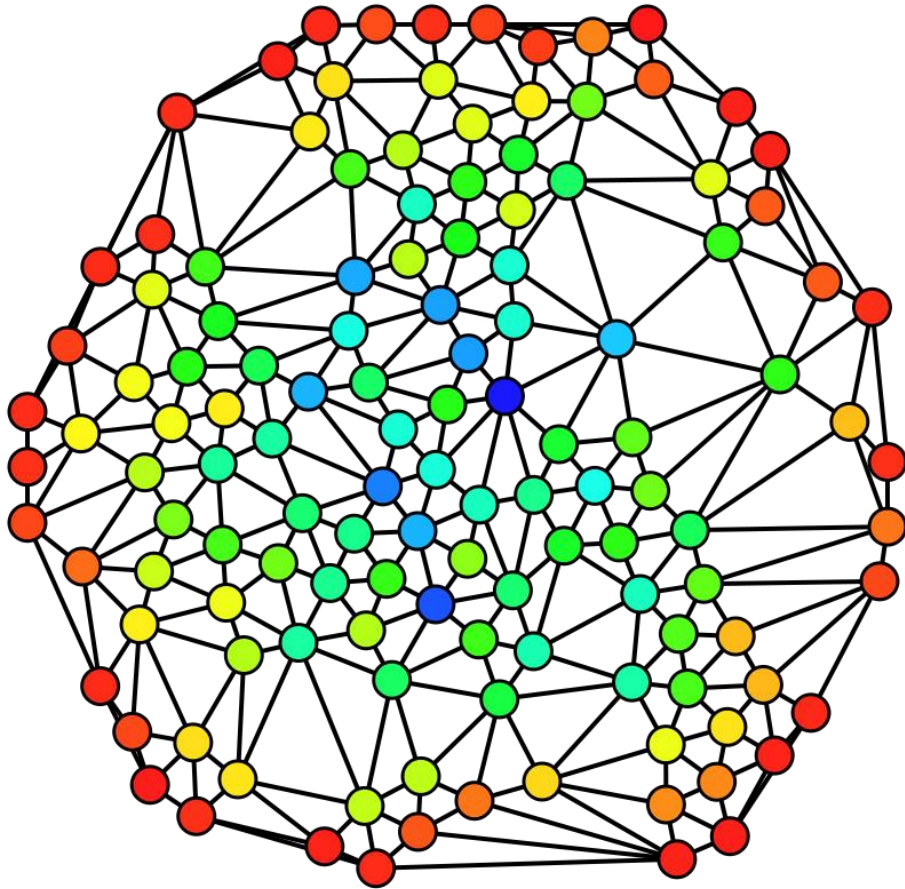
A - Centralité d'intermédiation

B - Centralité de proximité

C - Centralité de vecteur propre

D - Centralité de degré

Centralité d'intermédiation



L'intermédiation d'un nœud v est calculé comme suit :

- On calcule les chemins reliant deux nœuds s et t
- On détermine la proportion de chemins entre tous les nœuds qui passent par v
- On somme cette fraction sur tous les couples (s,t) de noeuds.

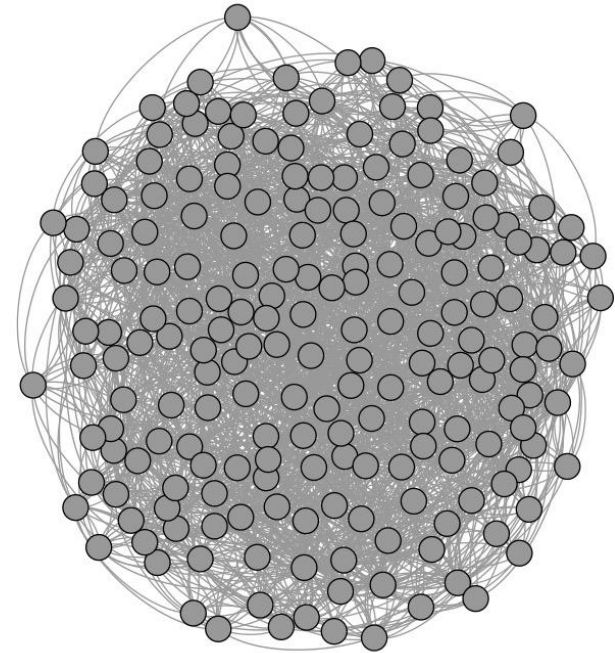
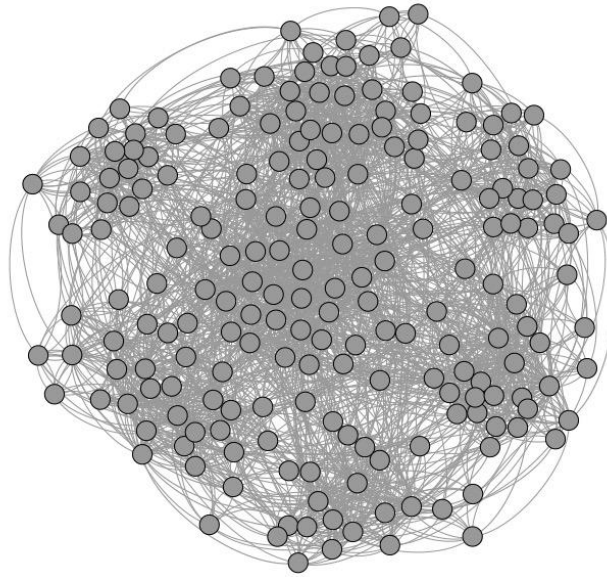
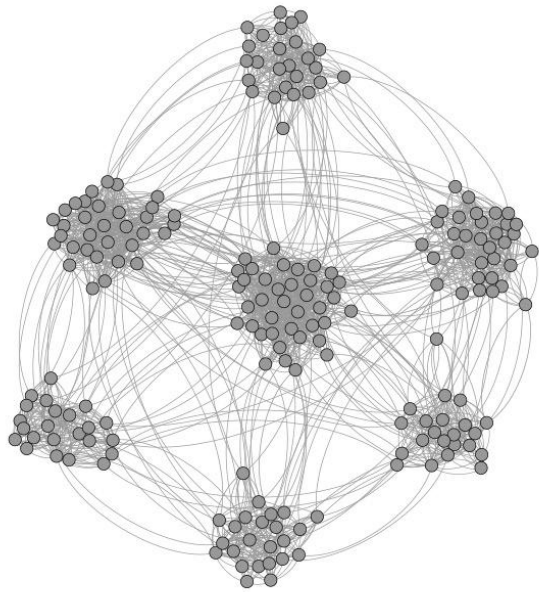
$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

σ_{st} est le nombre de chemins entre le noeud s et le noeud t

$\sigma_{st}(v)$ est le nombre de chemins, passant par v , entre le noeud s et le noeud t

Module, Cluster, Motifs

Clustering de réseau

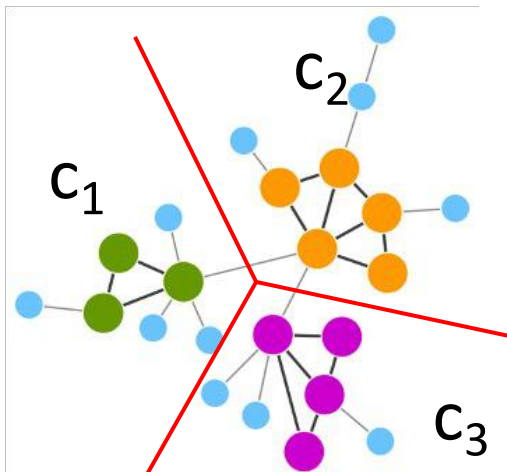


Module, Cluster, Motifs

Exemple de clustering : Algorithme de Louvain 2008

Optimisation de la modularité

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{i,j} - \frac{k_i * k_j}{2m} \right] \delta(c_i, c_j)$$



Trouver le « meilleur » partitionnement c = trouver l'ensemble des c_i

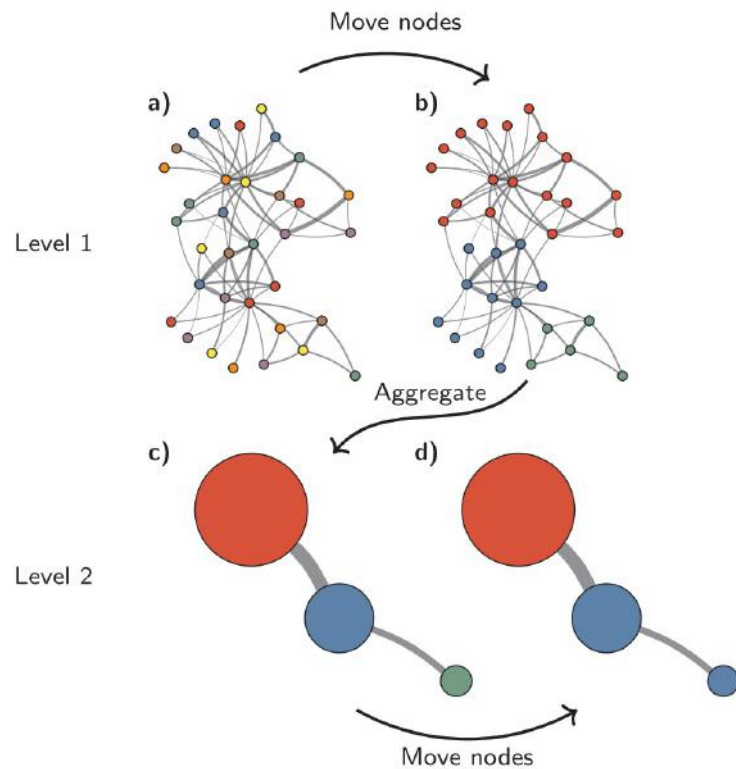
Algorithme heuristique

Deux phases :

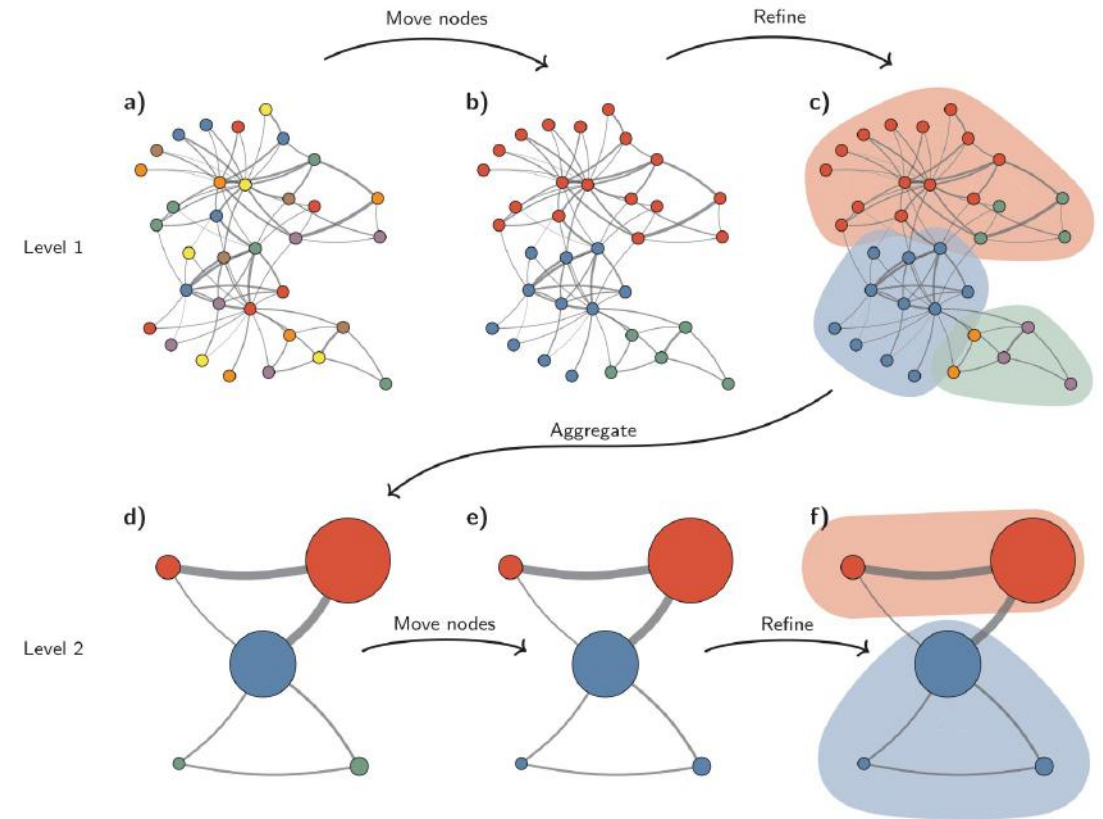
- On change l'appartenance des nœuds à leur classe et on calcule le changement de modularité que ça induit. On sélectionne les Nouvelles assignations augmentant la modularité.
- On optimise le partitionnement en créant un graphe des classes et Optimisant ce graphe avec la méthode de la première phase

Module, Cluster, Motifs

Exemple de clustering : Algorithme de Louvain 2008



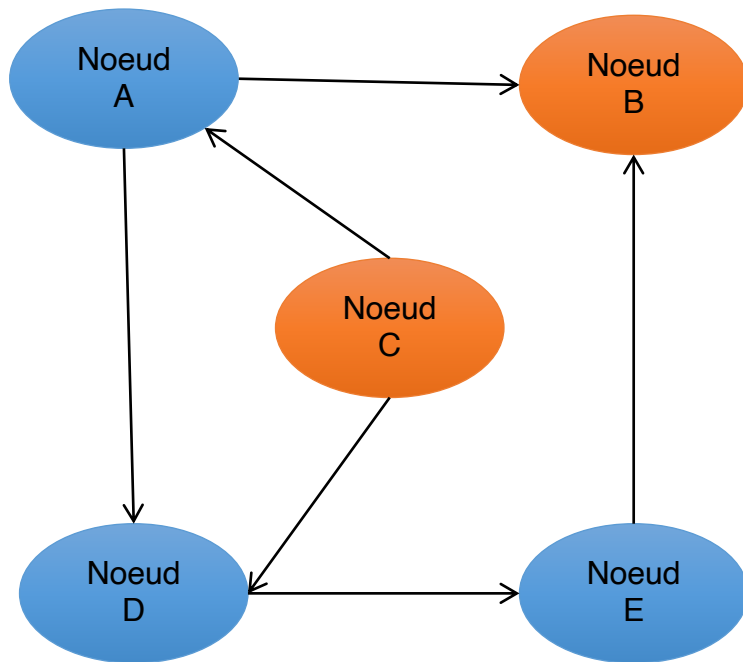
Louvain (2008)



Leiden (2018)

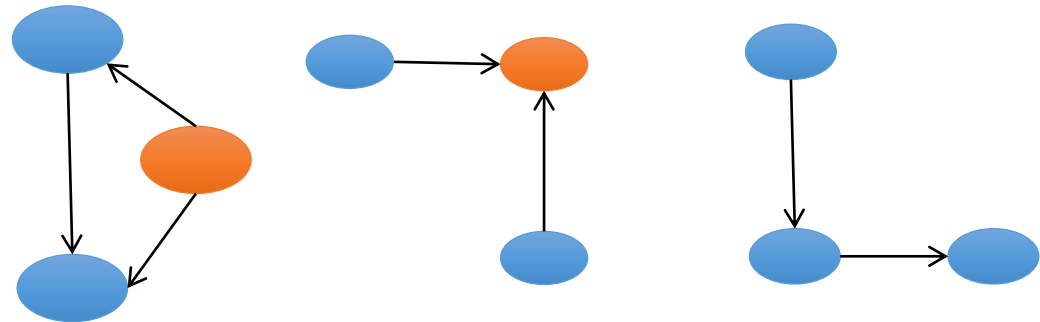
Module, Cluster, Motifs

Les motifs



Les motifs sont l'ensemble des sous-graphes de k nœuds

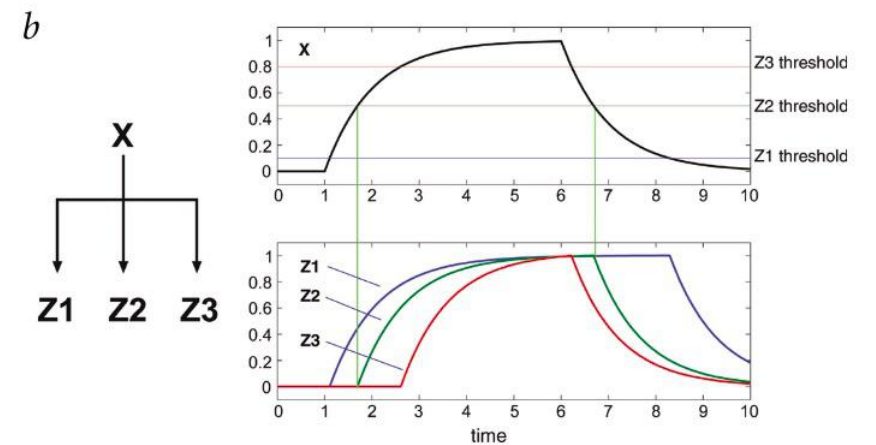
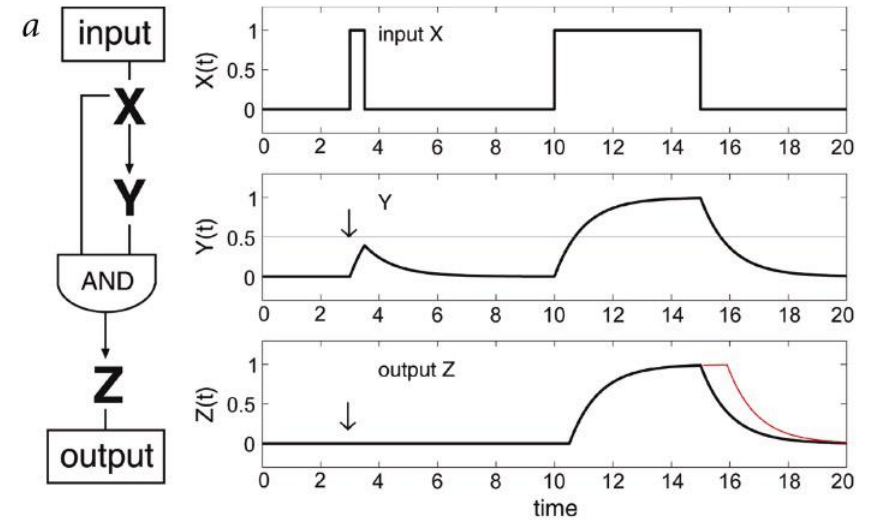
Exemple pour $k = 3$:



Réseau de régulation chez Escherichia Coli

Shen-Orr et al., Nature 2002

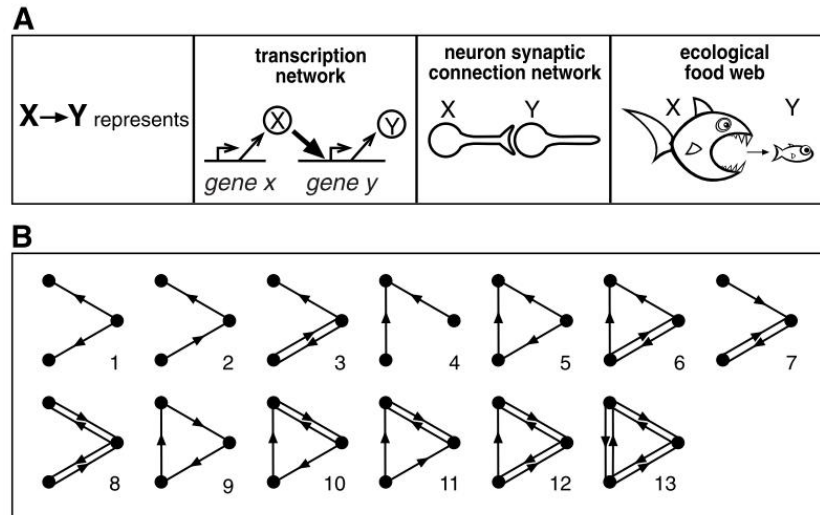
- 577 interactions
- 424 operons
- 116 facteurs de transcriptions
- Extrait de RegulonDB
- Que des facteurs de transcription validés directement

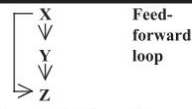
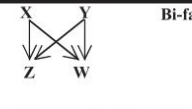
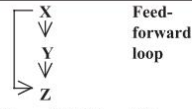
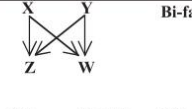
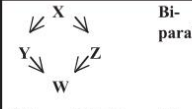
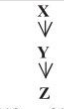
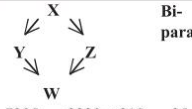
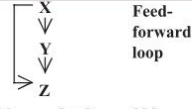
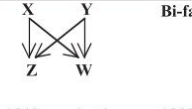
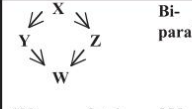
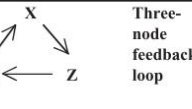
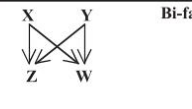
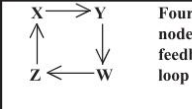
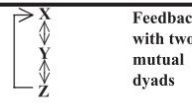
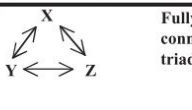
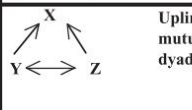


Les motifs privilégiés dans les réseaux

Milo et al., Science 2002

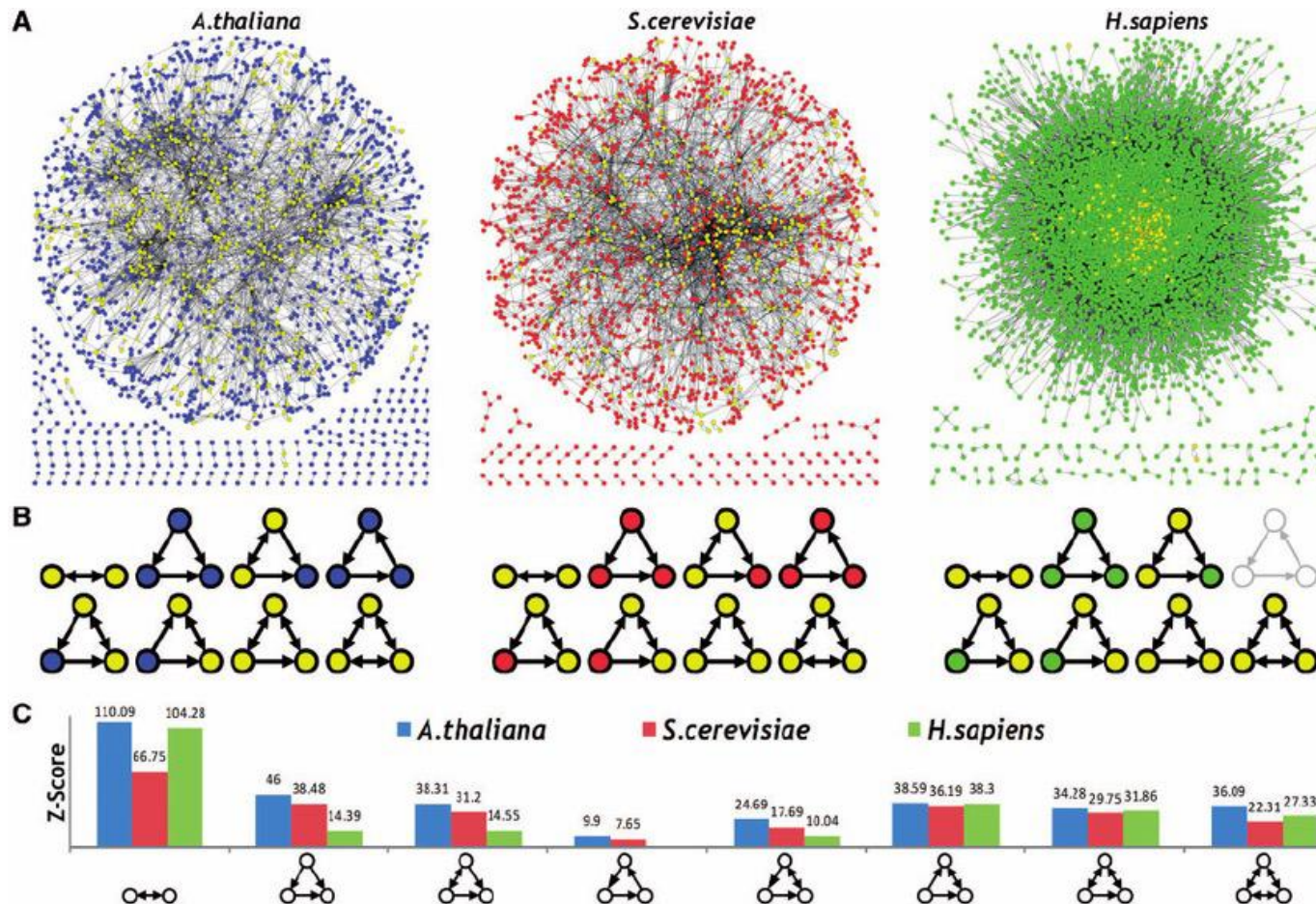
Les motifs sont-ils les briques élémentaires de construction des réseaux ?



Network	Nodes	Edges	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score	N_{real}	$N_{rand} \pm SD$	Z score
Gene regulation (transcription)				Feed-forward loop		Bi-fan					
<i>E. coli</i>	424	519	40	7 ± 3	10	203	47 ± 12	13			
<i>S. cerevisiae</i> *	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons				Feed-forward loop		Bi-fan		Bi-parallel			
<i>C. elegans</i> †	252	509	125	90 ± 10	3.7	127	55 ± 13	5.3	227	35 ± 10	20
Food webs				Three chain		Bi-parallel					
Little Rock	92	984	3219	3120 ± 50	2.1	7295	2220 ± 210	25			
Ythan	83	391	1182	1020 ± 20	7.2	1357	230 ± 50	23			
St. Martin	42	205	469	450 ± 10	NS	382	130 ± 20	12			
Chesapeake	31	67	80	82 ± 4	NS	26	5 ± 2	8			
Coachella	29	243	279	235 ± 12	3.6	181	80 ± 20	5			
Skipwith	25	189	184	150 ± 7	5.5	397	80 ± 25	13			
B. Brook	25	104	181	130 ± 7	7.4	267	30 ± 7	32			
Electronic circuits (forward logic chips)				Feed-forward loop		Bi-fan		Bi-parallel			
s15850	10,383	14,240	424	2 ± 2	285	1040	1 ± 1	1200	480	2 ± 1	335
s38584	20,717	34,204	413	10 ± 3	120	1739	6 ± 2	800	711	9 ± 2	320
s38417	23,843	33,661	612	3 ± 2	400	2404	1 ± 1	2550	531	2 ± 2	340
s9234	5,844	8,197	211	2 ± 1	140	754	1 ± 1	1050	209	1 ± 1	200
s13207	8,651	11,831	403	2 ± 1	225	4445	1 ± 1	4950	264	2 ± 1	200
Electronic circuits (digital fractional multipliers)				Three-node feedback loop		Bi-fan		Four-node feedback loop			
s208	122	189	10	1 ± 1	9	4	1 ± 1	3.8	5	1 ± 1	5
s420	252	399	20	1 ± 1	18	10	1 ± 1	10	11	1 ± 1	11
s838‡	512	819	40	1 ± 1	38	22	1 ± 1	20	23	1 ± 1	25
World Wide Web				Feedback with two mutual dyads		Fully connected triad		Uplinked mutual dyad			
nd.edu§	325,729	1.46e6	1.1e5	2e3 ± 1e2	800	6.8e6	5e4 ± 4e2	15,000	1.2e6	1e4 ± 2e2	5000

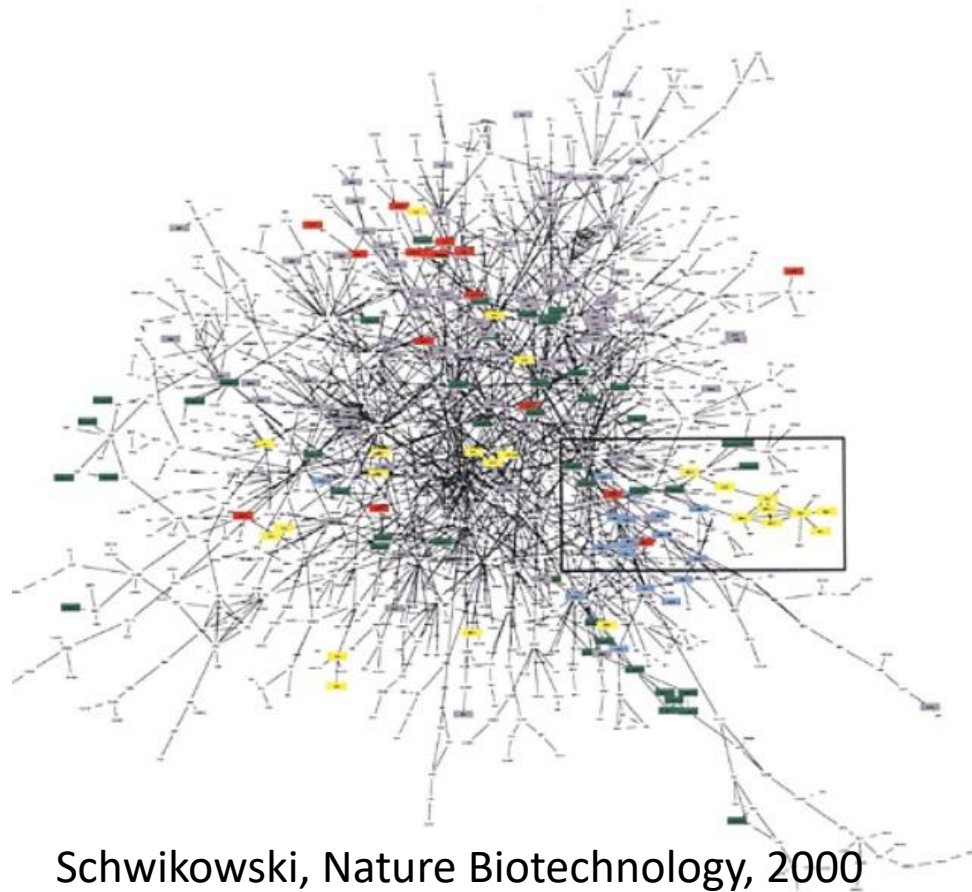
L'évolution = la sélection de motifs de régulation ?

Kim, Tae-Hwan et al. Evolutionary design principles and functional characteristics based on kingdom-specific network motifs. Bioinformatics 2010



- Robustesse
- Multistabilité
- Homeostaticité

Le réseau d'interaction protéine-protéine chez la levure



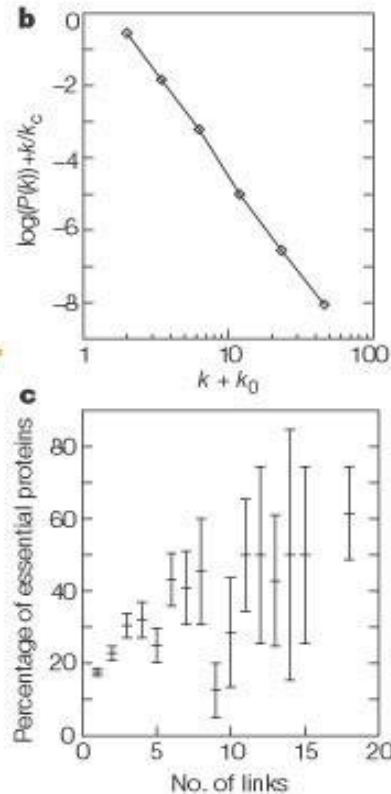
1548 / 6000 protéines – 2358 interactions

Le réseau a été reconstruit en combinant :

- Analyses d'un ensemble de mutants
- Réseau de coexpression
- Co-evolution
- Yeast two hybrid

Le réseau d'interaction protéine-protéine chez la levure

Jeong, Nature 2001



$P(k)$ = le nombre de protéines avec un degré k

La distribution des degrés suit une loi de puissance de la forme :

$$P(k) \sim k^{-3}$$

C'est ce que l'on appelle un réseau sans échelles ! (scale-free network)

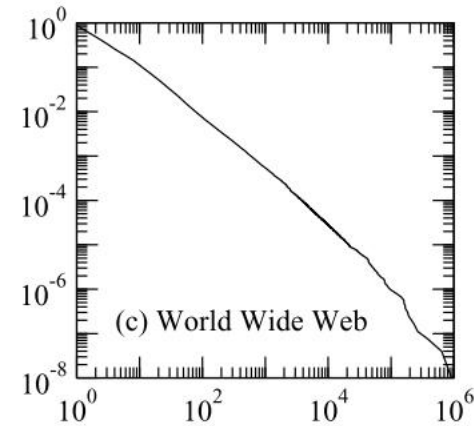
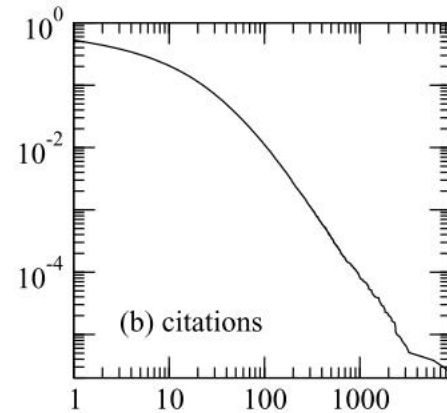
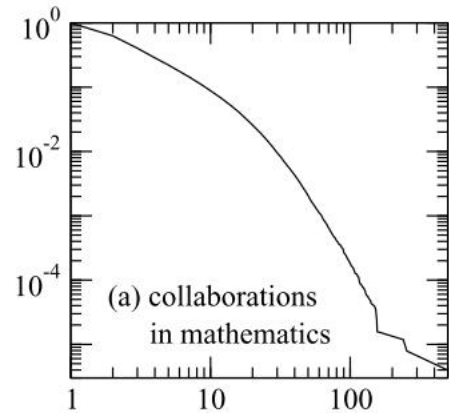
- Un petit nombre de nœud a un degré très grand
- Un grand nombre de nœud a un degré très petit

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Les réseaux sans échelles sont partout

Distribution des degrés de différents réseaux

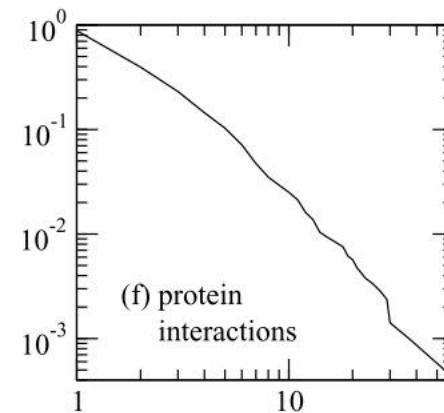
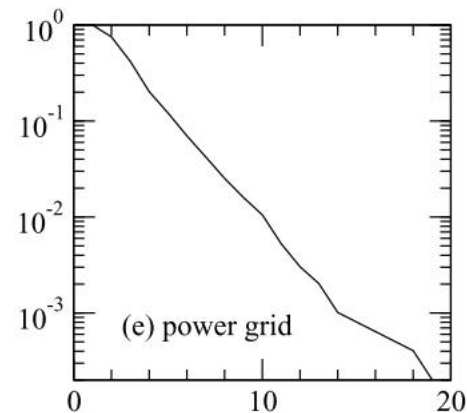
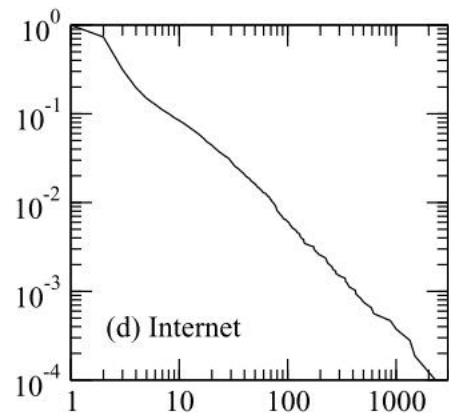


Loi de probabilité d'un réseau sans échelles

$$P(k) \sim k^{-\gamma}$$

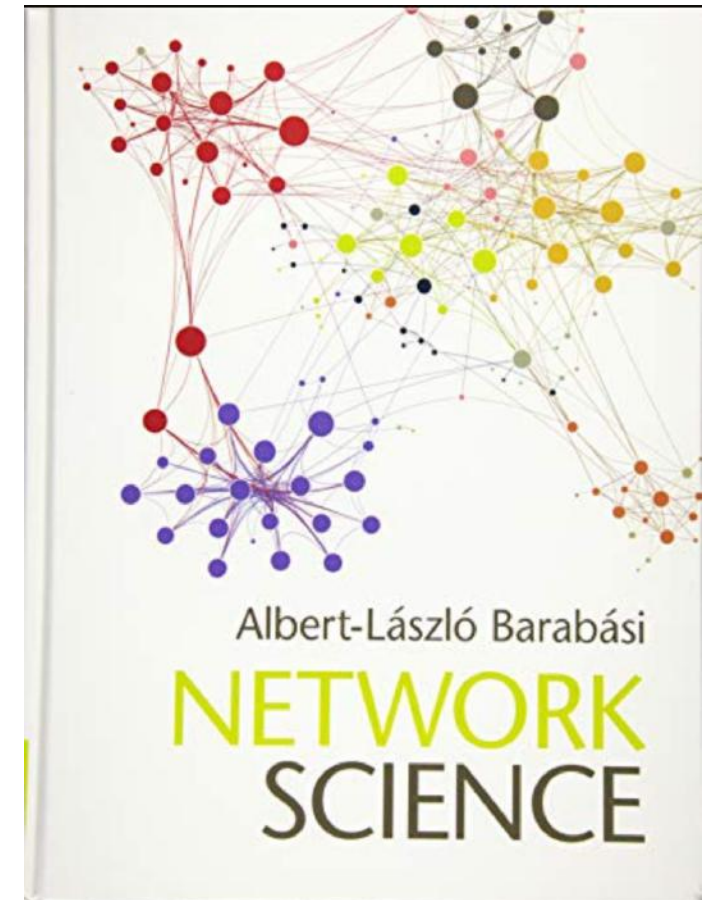
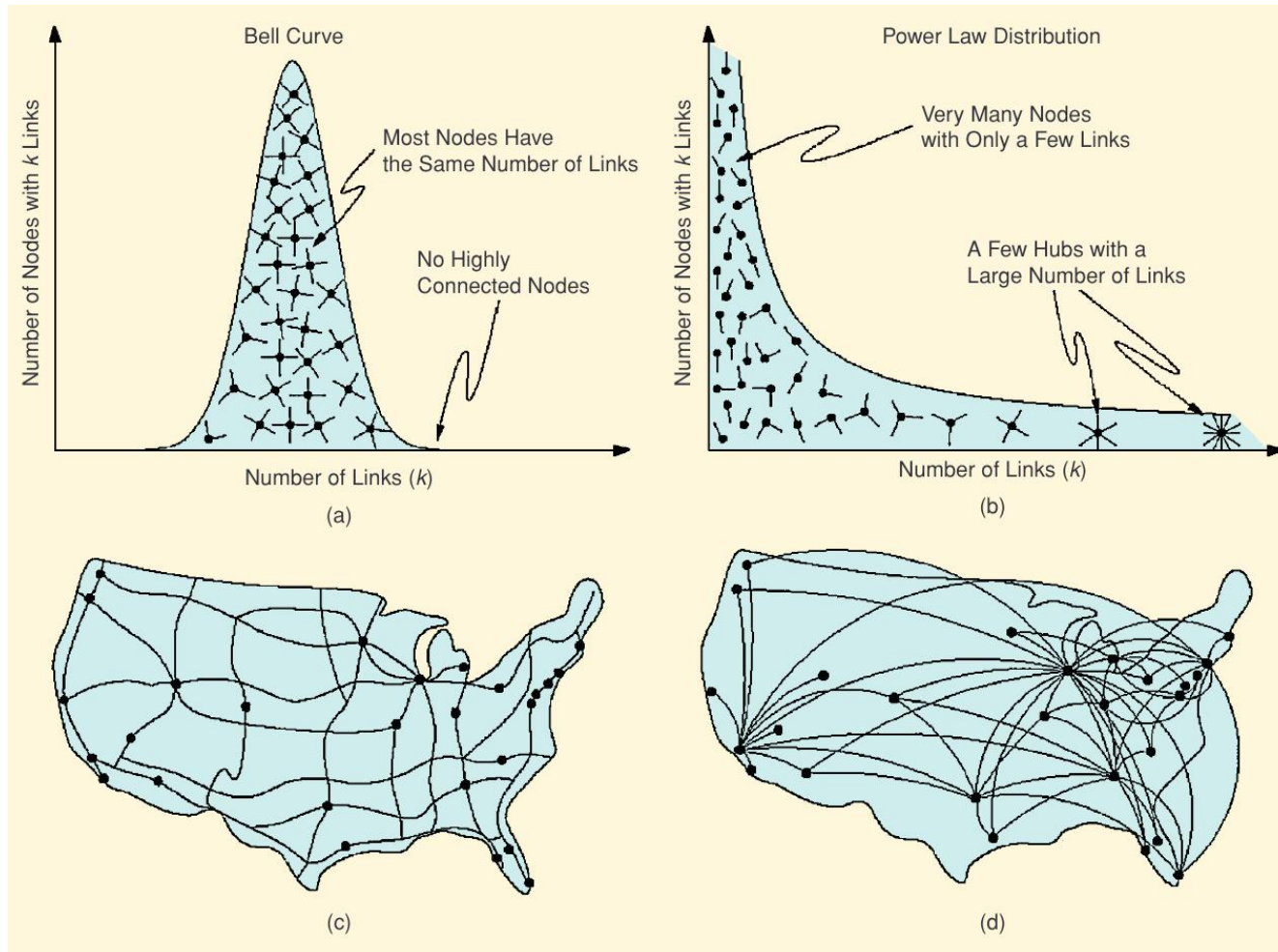
M. Newman, Siam Review, 2003

Avec $2 < \gamma < 3$



La structure de réseau sans échelles

Albert-Laszlo Barabasi, The architecture of complexity, IEEE 2007



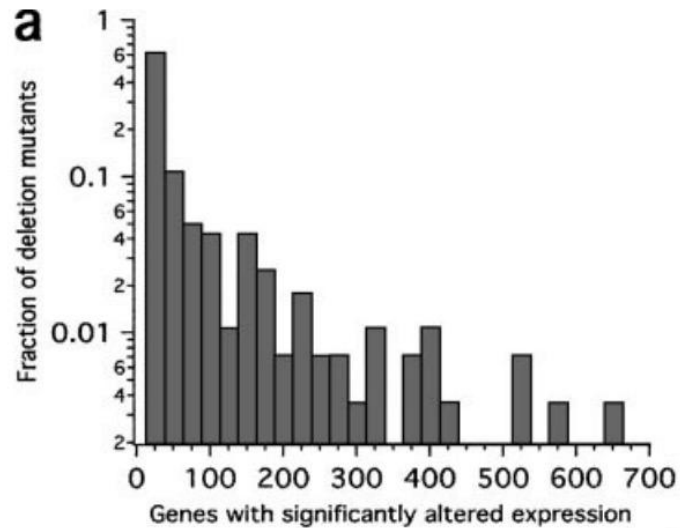
Exemple de réseau sans échelles

Gene expression network

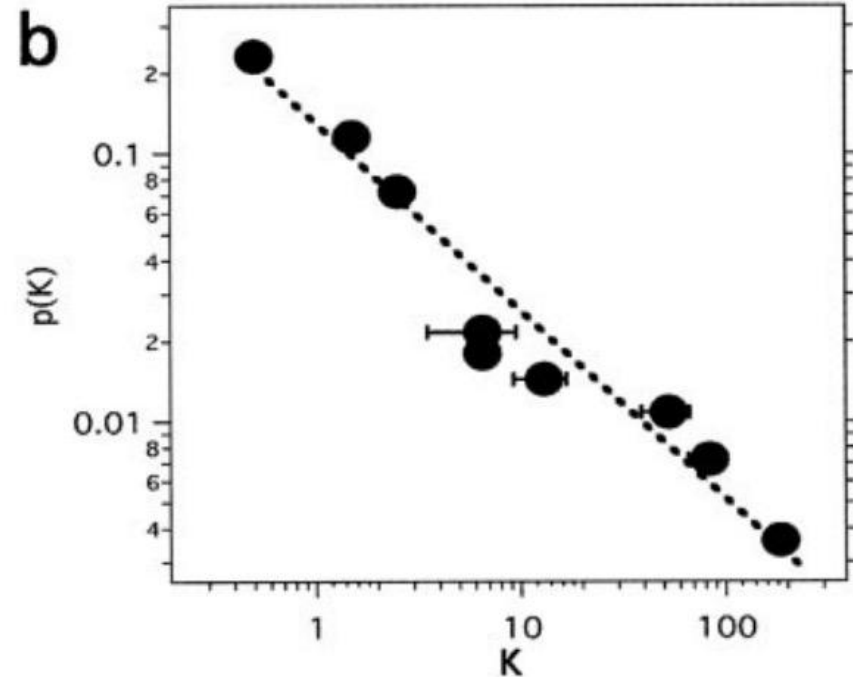
Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network

Bioessays - 2002

David E. Featherstone* and Kendal Broadie

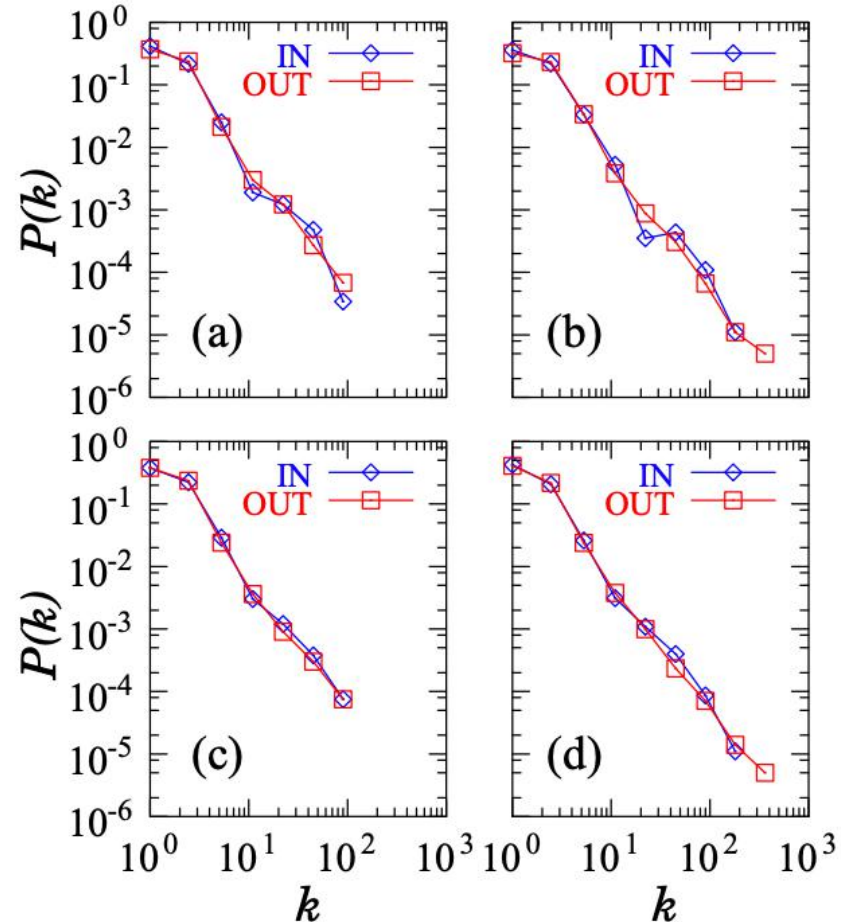


$$P(k) \sim k^{-0.7}$$



Exemple de réseau sans échelles

Les réseaux métaboliques



The large-scale organization of metabolic networks

H. Jeong¹, B. Tombor², R. Albert¹, Z. N. Oltvai² and A.-L. Barabási¹

Nature 2000

(a) *A. fulgidus* (Archae)

(b) *E. coli* (Bacterium)

(c) *C. Elegans* (Eukaryote)

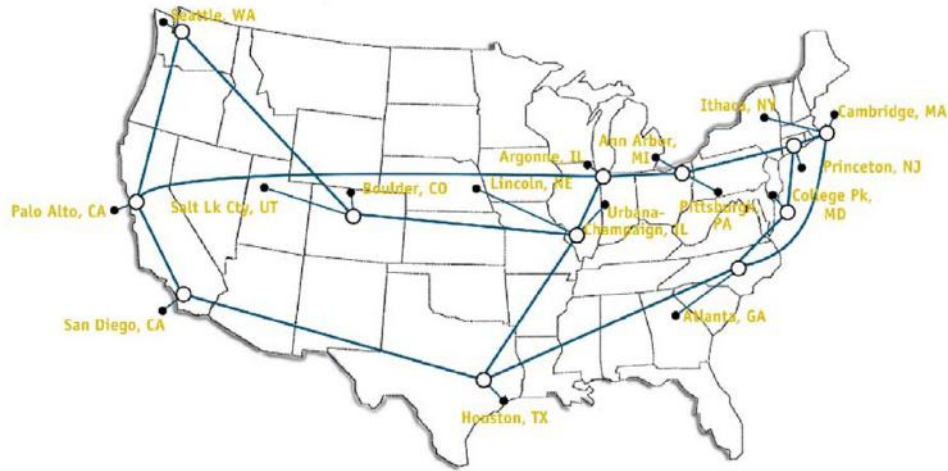
(d) Moyenne sur 43 organismes

$$P(k) \sim k^{-\gamma}$$

Avec $2 < \gamma < 3$

Des réseaux partout !

NSFNET T3 Network 1992



Réseau SNCF 2020

Interaction protéine-protéine
chez la Levure
Jeong, Nature 2001



La science des réseaux

Collective dynamics of 'small-world' networks

Duncan J. Watts* & Steven H. Strogatz

Department of Theoretical and Applied Mechanics, Kimball Hall,
Cornell University, Ithaca, New York 14853, USA

1998

Emergence of Scaling in Random Networks

Albert-László Barabási* and Réka Albert

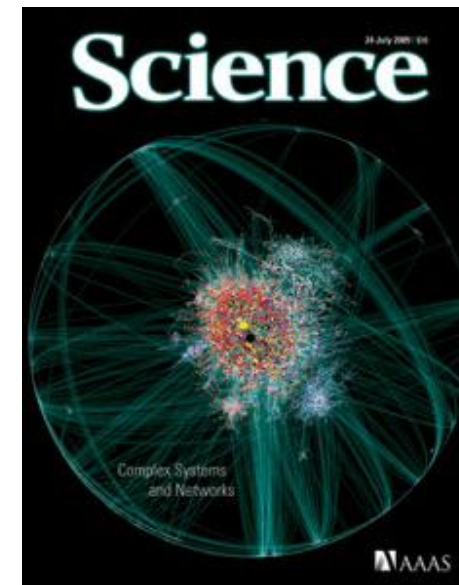
Department of Physics, University of Notre-Dame, Notre-Dame, IN 46556

1999

Network motifs in the transcriptional regulation network of *Escherichia coli*

Shai S. Shen-Orr¹, Ron Milo², Shmoolik Mangan¹ & Uri Alon^{1,2}

2002



Edition spéciale de
Science en 2009

Modèle de génération des réseaux

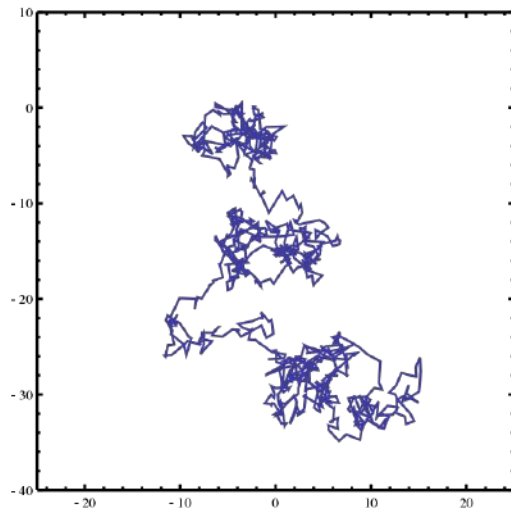
- Existe-t-il une loi générale permettant de construire les réseaux ?
- Existe-t-il une propriété générale aux réseaux « naturelles » ?
- Peut-on prédire le comportement d'un système à partir de la structure de son réseau ?

L'exemple de la physique statistique

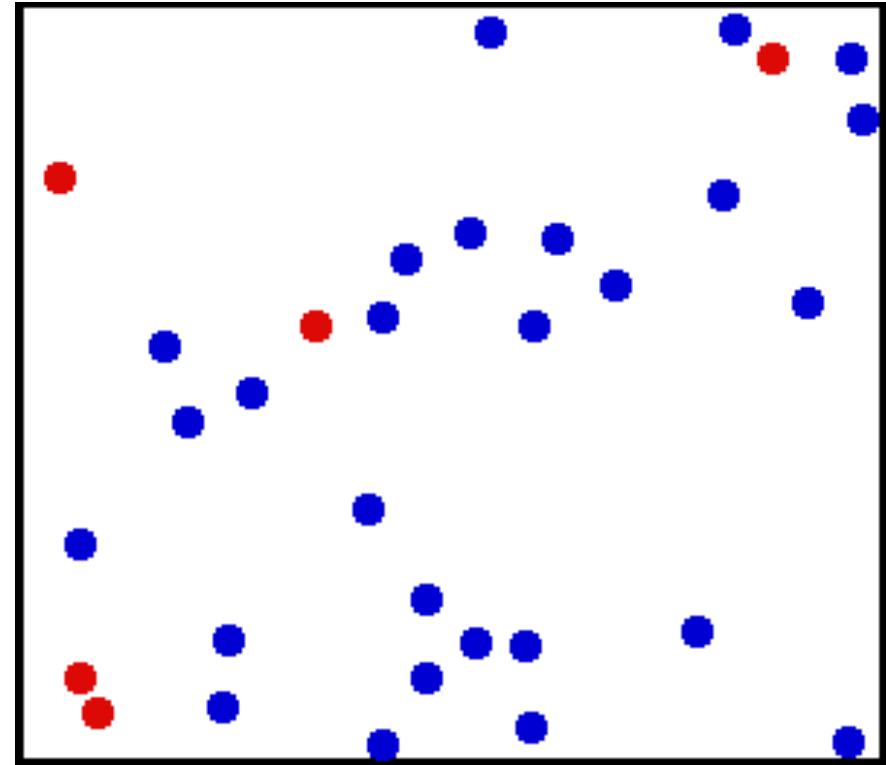
Quel est le lien entre :

Théorie cinétique des gaz

Mouvement brownien des atomes



ET la température d'un gaz



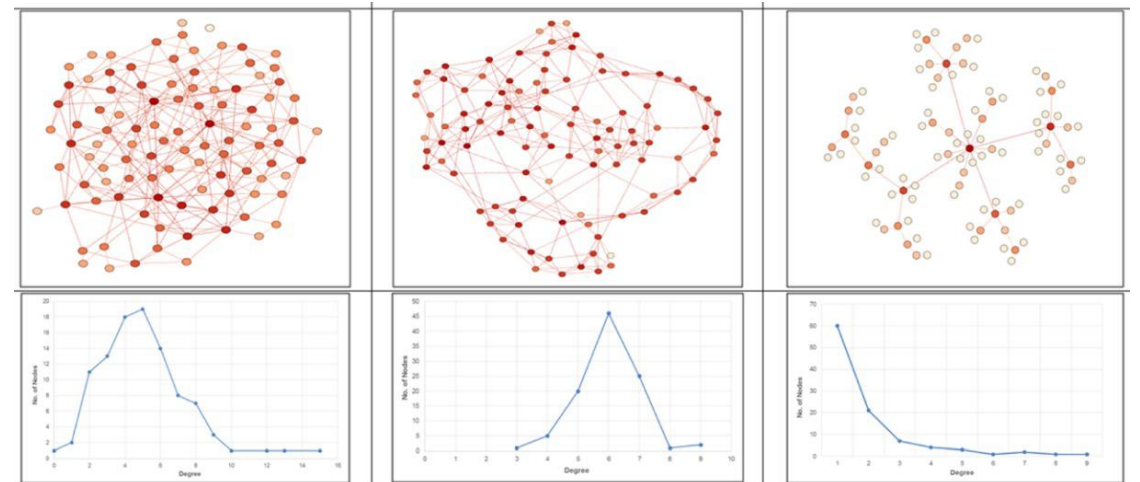
Physique statistique et théorie des réseaux

Description microscopique

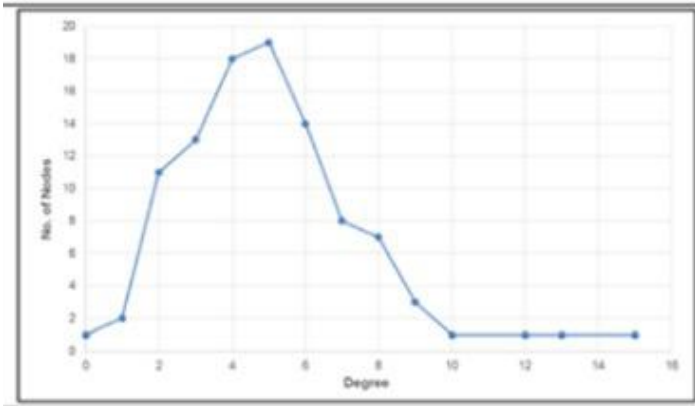
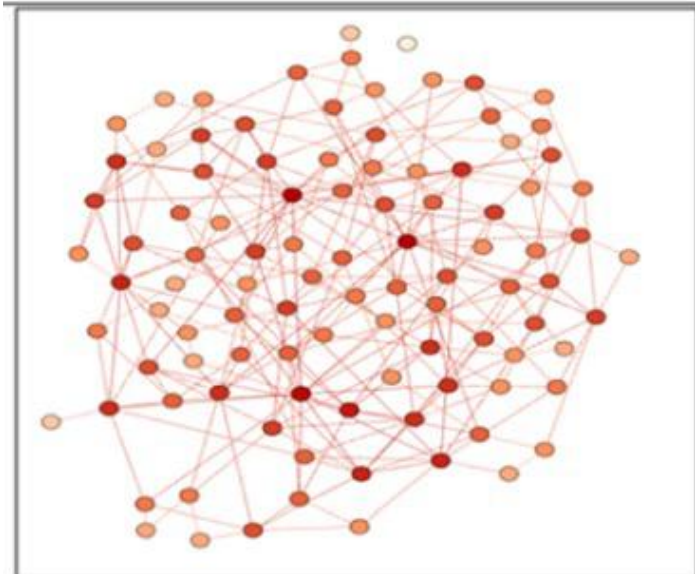
- Ensemble des états d'un nœud ?
- Ensemble des états d'une arête ?
- Distribution statistique des connexions ?
- Distribution statistique des chemins ?

Description macroscopique

La distribution des degrés



Modèle Erdos-Renyi - 1959



On établit deux paramètres :

n le nombre total de sommets

P la probabilité que deux nœuds soient connectés

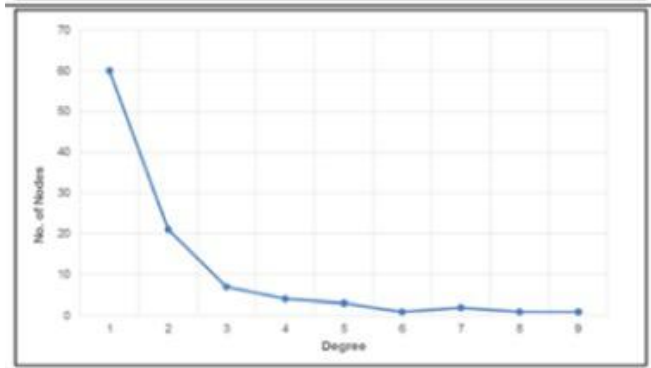
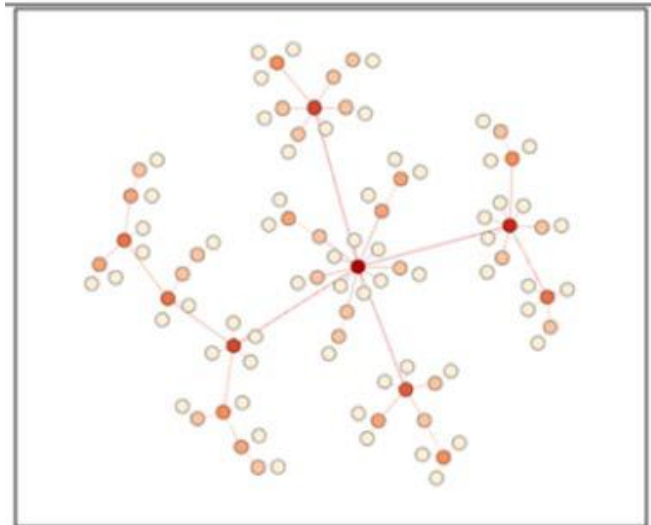
Le modèle d'Erdos-Renyi nous permet de générer un réseau avec une distribution des degrés représentant une loi binomiale

$$P(\text{deg}(v) = k) = \binom{n-1}{k} p^k (1-p)^{n-1-k}$$

Modèle sans échelle de Barabasi-Albert

Génération de réseau comportant des « hub »

- Un petit nombre de nœud a un degrés très grand
- Un grand nombre de nœud a un degré très petit



- On initialise le réseau avec un ensemble m de nœuds
- On ajoute des nouveaux nœuds un par un avec une probabilité de connexion aux autres nœuds i : $p_i = \frac{k_i}{\sum_j k_j}$ où k_i est le degré du nœud i

Ce modèle nous permet de générer des réseaux avec une distribution de degrés en loi puissance $P(k) \sim k^{-3}$

Ces réseaux sont dit « sans échelle »

Une loi naturelle ?

- « Les riches deviennent plus riches ? »
- Spécialisation et centralité
- Mécanisme de l'évolution
- Robustesse ?



Edition spéciale de
Science en 2009

Lethality and centrality in protein networks

The most highly connected proteins in the cell are the most important for its survival.

Jeong, Nature 2001

Biologie des systèmes

- **L'approche systémique en biologie**
- **Reconstruire un réseau biologique**
 - Les différents types de réseaux
 - Reconstruction directe
 - Reconstruction indirecte
 - Les obstacles à la reconstruction
- **Analyser un réseau biologique**
 - Théorie des graphes
 - Topologie des réseaux
 - Les réseaux sans échelles
- **Utiliser un réseau biologique**
 - **Etat de l'art des réseaux biologiques les plus étendus**
 - **Bases de données en biologie des systèmes**
- **Modéliser un réseau biologique**



LacZ operon version 2023

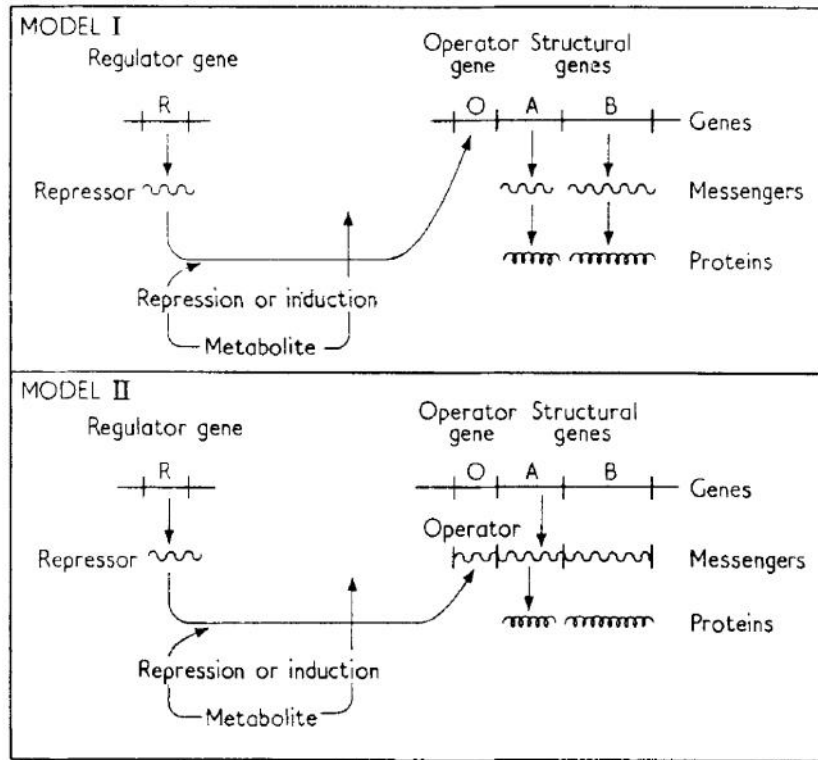
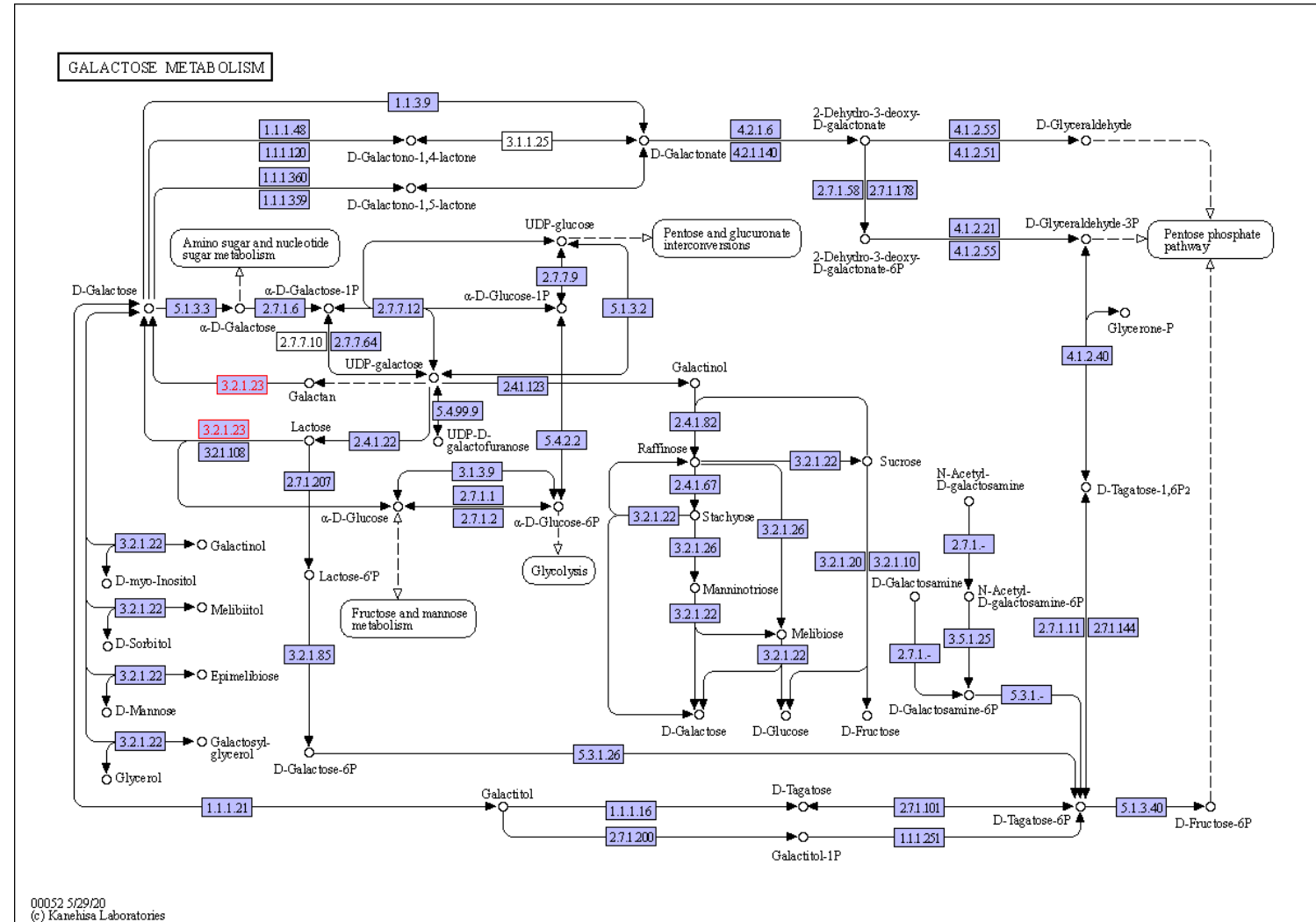


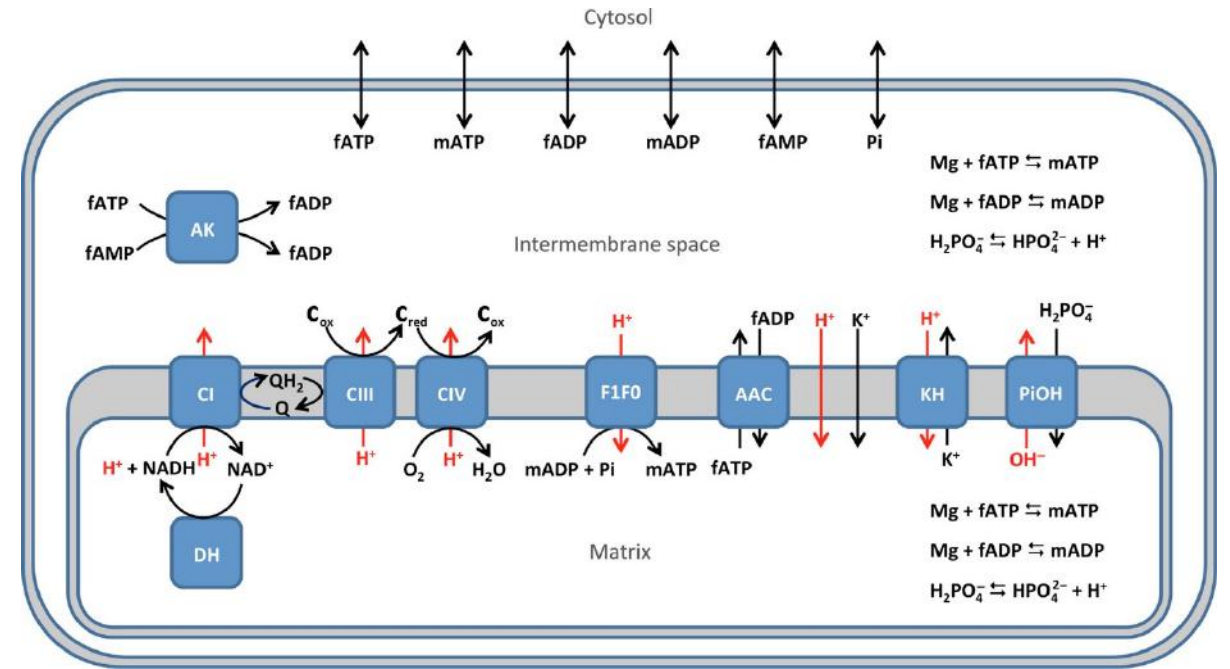
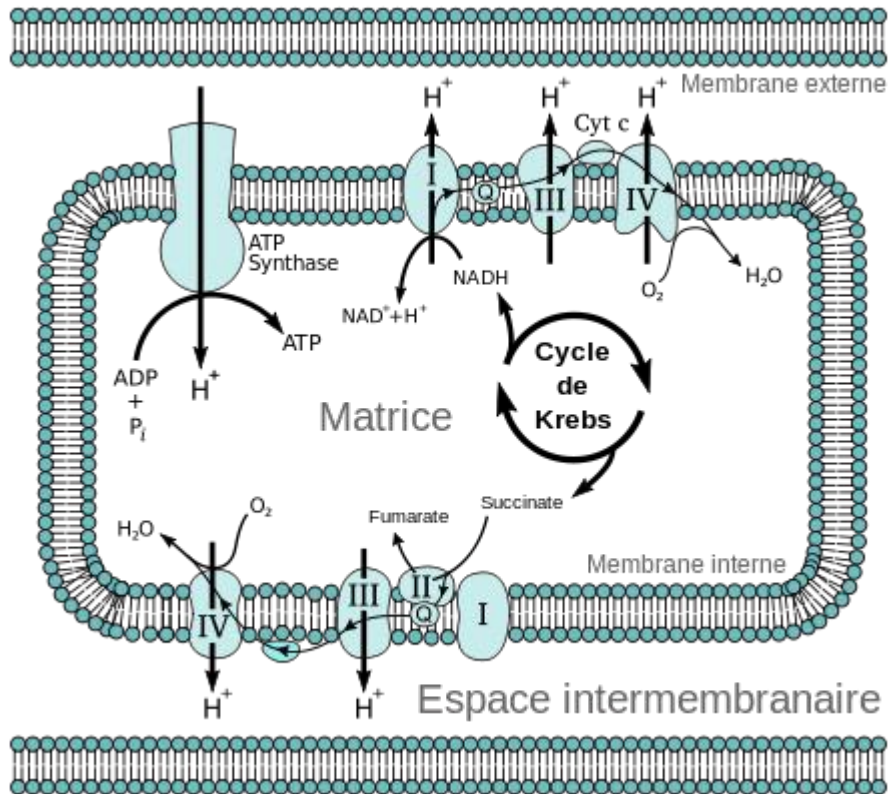
FIG. 6. Models of the regulation of protein synthesis.

Genetic Regulatory Mechanisms in the Synthesis of Proteins
 F. Jacob, J. Monod, J. Mol. Biol. 1961



https://www.genome.jp/kegg-bin/show_pathway?ko00052+K01190

Le cycle de Krebs en 2023 : Phosphorylation oxydative



Heiske et al., FEBS Journal 2017

Phosphorylation de l'ADP en ATP
dans les mitochondries
(wikipedia)

Le cycle de Krebs en 2023 : Phosphorylation oxydative

Matrix		
DH	Dehydrogenase reaction	$NAD_x + H_x \rightleftharpoons NADH_x$
MgATP _x	Mg ²⁺ binding on fATP	$fATP_x + Mg_x \rightleftharpoons mATP_x$
MgADP _x	Mg ²⁺ binding on fADP	$fADP_x + Mg_x \rightleftharpoons mADP_x$
Pdiss _x	Pi dissociation	$H_2PO_{4x} \rightleftharpoons HPO_{4x} + H_x$
Inner membrane		
C1	Complex I	$NADH_x + Q_m + 5H \rightleftharpoons NAD_x + QH_{2im} + 4H_{is}$
C3	Complex III	$QH_{2im} + 2Co_x + 2H_x \rightleftharpoons Q_m + 2Cred_{is} + 4H_{is}$
C4	Complex IV	$Cred_{is} + 0.25 O_2 + 2H_x \rightleftharpoons Cox_{is} + 0.5H_2O + H_{is}$
F1F0	ATP synthase	$mADP_x + P_i + n_A H_{is} + H_x \rightleftharpoons mATP_x + n_A H_x$
AAC	ADP/ATP carrier	$fATP_x + fADP_{is} \rightleftharpoons fATP_{is} + fADP_x$
PiOH	Pi/OH antiporter	$H_2PO_{4is} + OH_x \rightleftharpoons H_2PO_{4x} + OH_{is}$
KH	K ⁺ /H ⁺ antiporter	$K_{is} + H_x \rightleftharpoons K_x + H_{is}$
Hleak	H ⁺ leak	$H_{is} \rightleftharpoons H_x$
Kleak	K ⁺ leak	$K_{is} \rightleftharpoons K_x$
Intermembrane space		
AK	Adenylate kinase	$mATP_{is} + fAMP_{is} \rightleftharpoons mADP_{is} + fADP_{is}$
MgATP _{is}	Mg ²⁺ binding on fATP	$fATP_{is} + Mg_{is} \rightleftharpoons mATP_{is}$
MgADP _{is}	Mg ²⁺ binding on fADP	$fADP_{is} + Mg_{is} \rightleftharpoons mADP_{is}$
Pdiss _{is}	Pi dissociation	$H_2PO_{4is} \rightleftharpoons HPO_{4is} + H_{is}$
Outer membrane		
fATP _{om}	fATP diffusion	$fATP_e \rightleftharpoons fATP_{is}$
mATP _{om}	mATP diffusion	$mATP_e \rightleftharpoons mATP_{is}$
fADP _{om}	fADP diffusion	$fADP_e \rightleftharpoons fADP_{is}$
mADP _{om}	mADP diffusion	$mADP_e \rightleftharpoons mADP_{is}$
fAMP _{om}	fAMP diffusion	$fAMP_e \rightleftharpoons fAMP_{is}$
Pi _{om}	Pi diffusion	$P_{ie} \rightleftharpoons P_{is}$
Mg _{om} *	Mg ²⁺ diffusion	$Mg_e \rightleftharpoons Mg_{is}$
External space/cytosol		
MgATP _e *	Mg ²⁺ binding on fATP	$fATP_e + Mg_e \rightleftharpoons mATP_e$
MgADP _e *	Mg ²⁺ binding on fADP	$fADP_e + Mg_e \rightleftharpoons mADP_e$

$$\frac{d[H]_x}{dt} = x_{\text{diff}} \cdot \left(\frac{+v_{DH} - 5v_{C1} - 2v_{C3} - 4v_{C4}}{+(n_A - 1)v_{F1F0} + 2v_{PiOH} + v_{leak} - v_{KH}} \right) / W_x \quad (35)$$

$$\frac{d[K]_x}{dt} = (+v_{KH} + v_K) / W_x \quad (36)$$

$$\frac{d[Mg]_x}{dt} = -v_{MgATP_x} - v_{MgADP_x} \quad (37)$$

$$\frac{d[NADH]_x}{dt} = (+v_{DH} - v_{C1}) / W_x \quad (38)$$

$$\frac{d[fATP]_x}{dt} = -v_{AAC} / W_x - v_{MgATP_x} \quad (39)$$

$$\frac{d[mATP]_x}{dt} = +v_{F1F0} / W_x + v_{MgATP_x} \quad (40)$$

$$\frac{d[fADP]_x}{dt} = +v_{AAC} / W_x - v_{MgADP_x} \quad (41)$$

$$\frac{d[mADP]_x}{dt} = -v_{F1F0} / W_x + v_{MgADP_x} \quad (42)$$

$$\frac{d[P_i]_x}{dt} = (-v_{F1F0} + v_{PiOH}) / W_x \quad (43)$$

$$\frac{d[QH_2]_{im}}{dt} = (+v_{C1} - v_{C3}) / W_{im} \quad (44)$$

$$\frac{d[Cred]_{im}}{dt} = (+2v_{C3} - 2v_{C4}) / W_{is} \quad (45)$$

$$\frac{d[fATP]_{is}}{dt} = (+v_{ATP_{om}} + v_{AAC_{om}} - v_{AK_{is}}) / W_{is} - v_{MgATP_{is}} \quad (46)$$

$$\frac{d[mATP]_{is}}{dt} = +v_{mATP_{om}} / W_{is} + v_{MgATP_{is}} \quad (47)$$

$$\frac{d[fADP]_{is}}{dt} = (+v_{ADP_{om}} - v_{AAC_{om}} + 2v_{AK_{is}}) / W_{is} - v_{MgADP_{is}} \quad (48)$$

$$\frac{d[mADP]_{is}}{dt} = +v_{mADP_{om}} / W_{is} + v_{MgADP_{is}} \quad (49)$$

$$\frac{d[fAMP]_{is}}{dt} = (+v_{AMP_{om}} - v_{AK_{is}}) / W_{is} \quad (50)$$

$$\frac{d[P_i]_{is}}{dt} = (-v_{PiOH} + v_{Pi_{om}}) / W_{is} \quad (51)$$

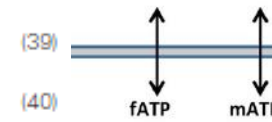
$$\frac{d\Delta\Psi}{dt} = (+4v_{C1} + 2v_{C3} + 4v_{C4} - n_A \cdot v_{F1F0} - v_{AAC} - v_{leak} - v_K) / C \quad (52)$$

$$[NAD]_x = N_{tot_x} - [NADH]_x \quad (53)$$

$$[Q]_{im} = Q_{tot_{im}} - [QH_2]_{im} \quad (54)$$

$$[Cox]_{is} = C_{tot_{is}} - [Cred]_{is} \quad (55)$$

$$[mADP]_e = \frac{1}{2} \left((K_{d_{mADP}} + ADP_{tot_e} + Mg_{tot_e}) - \sqrt{(K_{d_{mADP}} + ADP_{tot_e} + Mg_{tot_e})^2 - 4(Mg_{tot_e} \cdot ADP_{tot_e})} \right) \quad (56)$$



$$[fADP]_e = ADP_{tot_e} - [mADP]_e \quad (57)$$

$$[Mg]_e = Mg_{tot_e} - [mADP]_e \quad (58)$$

$$[Mg]_{is} = [Mg]_e \quad (59)$$

$$[P_i]_e = [P_i]_e \quad (60)$$

$$[K]_e = [K]_e \quad (61)$$

$$[H]_e = [H]_e \quad (62)$$

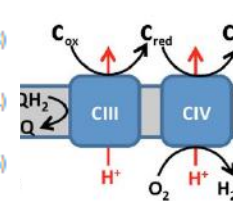
$$[H_2PO_4]_{is} = \frac{[H_{is}] \cdot [P_i]_{is}}{[H_{is}] + K_{d_{H_2PO_4}}}$$

$$[H_2PO_4]_x = \frac{[H_x] \cdot [P_i]_x}{[H_x] + K_{d_{H_2PO_4}}} \quad (63)$$

$$[OH]_{is} = \frac{10^{-14}}{[H_{is}]} \quad (64)$$

$$[OH]_x = \frac{10^{-14}}{[H_x]} \quad (65)$$

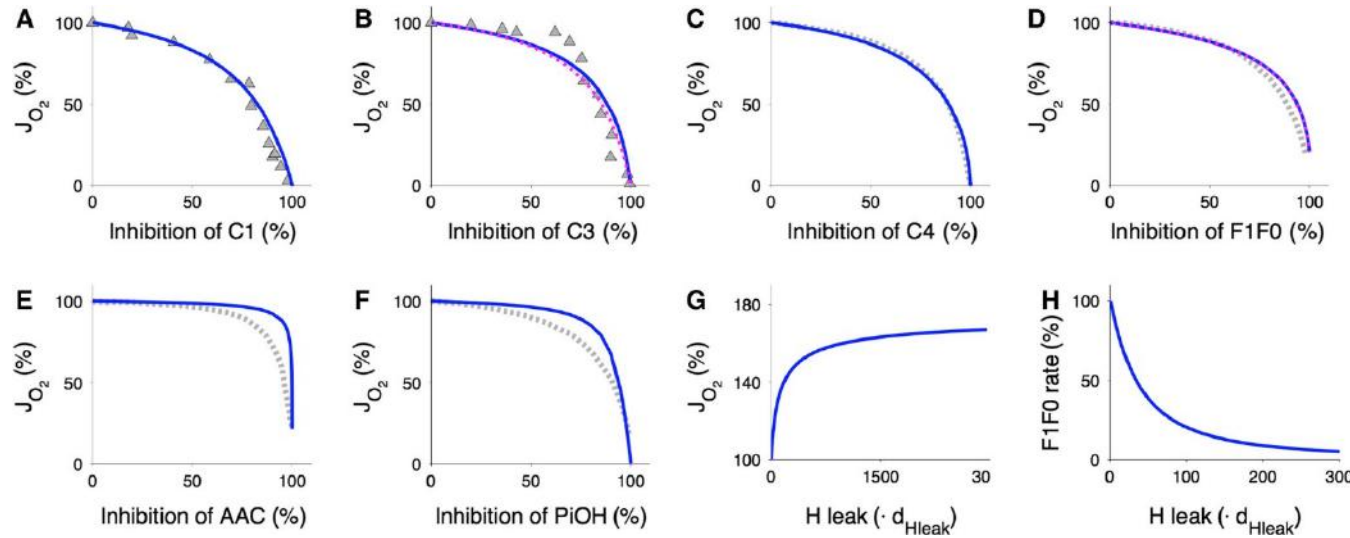
$$[OH]_x = \frac{10^{-14}}{[H_x]} \quad (66)$$



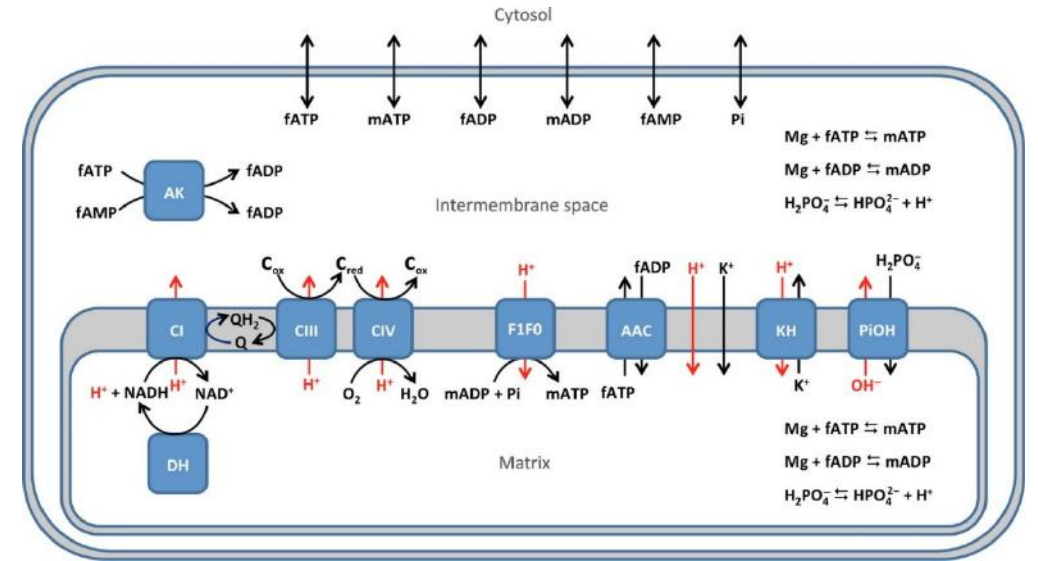
Heiske et al., FEBS Journal 2017

Le cycle de Krebs en 2023 : Phosphorylation oxydative

Optimisation non-linéaire sous contrainte



On obtient une modélisation de l'évolution de toutes les concentrations



Heiske et al., FEBS Journal 2017

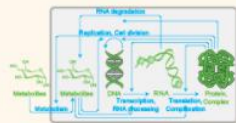
- Réglage fin de tous les paramètres du modèle
- **20 pages de description du modèle !**
- Le modèle reste « simple » ce n'est pas un organisme entier

Modèle d'organisme entier



Models: Comprehensive computational models of individual cells

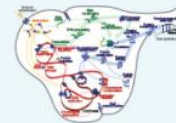
Archetypal bacterium



The archetypal bacterium model generator is a tool for generating WC models that represent user-specified numbers of genes, RNA, proteins, and reactions. The models generated by the model generator represent the metabolism, replication, transcription, translation, RNA and protein degradation, and cell division of a typical bacterium. The archetypal bacterium model generator is particularly useful for driving the development of WC modeling tools, as well as teaching WC modeling.

Availability: In development
Author: Karr Lab, Sinai

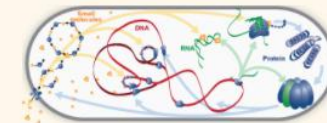
Mycoplasma genitalium



The *M. genitalium* WC model was the first model that represented each characterized gene function of an organism. The model is composed of 28 submodels of 28 cellular processes. In total, the model represents the functions of 401 genes, 722 compounds, 1,857 reactions, and 1,836 parameters. In addition to demonstrating the feasibility of WC models, the model has been used to gain new insights into cell cycle regulation, learn unknown parameters, and suggest new uses of existing antibiotics for *Mycoplasmas*.

Availability: Download
Author: Covert Lab, Stanford
More info: Docs | Source | License | Tests
Reference: Karr JR et al. *Cell* 2012

Mycoplasma pneumoniae



The *M. pneumoniae* WC model will be the most comprehensive, most systematically constructed, and most extensible WC model to date. The model will represent all of the major cellular functions of *M. pneumoniae*, including the function each characterized gene. The model will be based primarily on *M. pneumoniae* genomic data. The model will be used to drive the development of WC modeling methods, as well as to help design a reliable, energy efficient, fast-growing chassis for future bioengineering.

Availability: In development
Author: Karr Lab, Sinai

Escherichia coli



The *E. coli* WC model represents the core cellular functions of *E. coli*. The model is the most detailed and most thoroughly tested WC model to date.

Availability: In development
Author: Covert Lab, Stanford

Homo sapiens (H1-hESC)



The H1 human embryonic stem cell (hESC) model is the first step toward WC models of human cells. The model will represent the core cellular functions of all human cells including their metabolism, DNA replication, transcription, translation, protein complexation, RNA and protein degradation, and division. The model focuses on H1-hESCs because ESCs behave as individual cells, because ESC lines are karyotypically normal, because ESCs grow quickly, and because H1-hESC has been extensively characterized. In addition to demonstrating the feasibility of human WC models and driving the development of WC modeling tools, the model will be used to gain insights into how stem cells maintain pluripotency.

Availability: In development
Author: Karr Lab, Sinai

Homo sapiens cancer signaling (MCF10A)



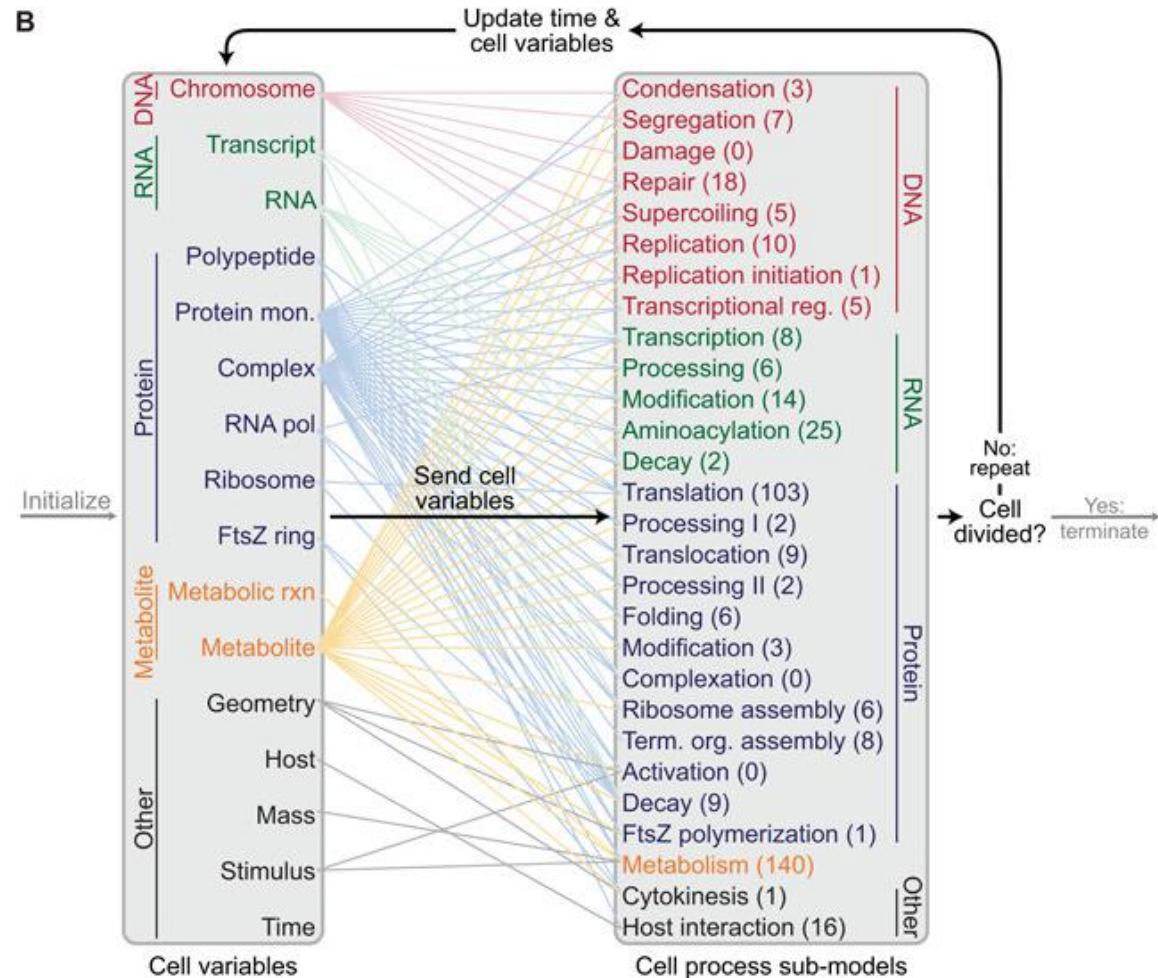
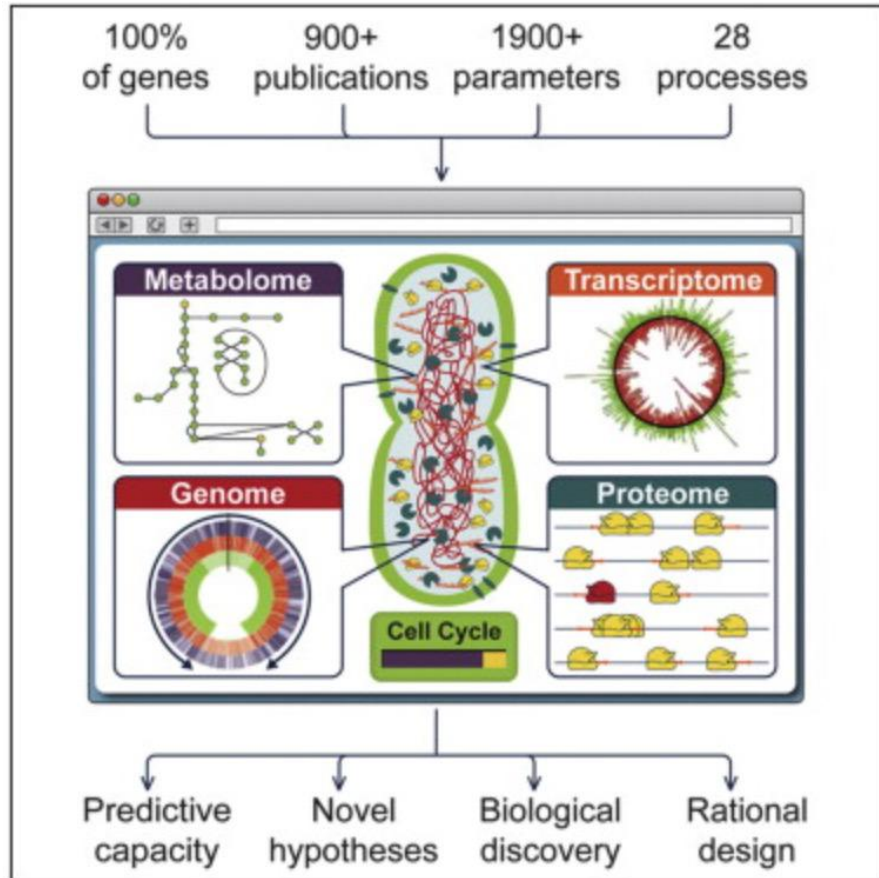
A mechanistic ordinary differential equation model describing the interactions between commonly mutated pan-cancer signaling pathways—receptor tyrosine kinases, Ras/RAF/ERK, PI3K/AKT, mTOR, cell cycle, DNA damage, and apoptosis.

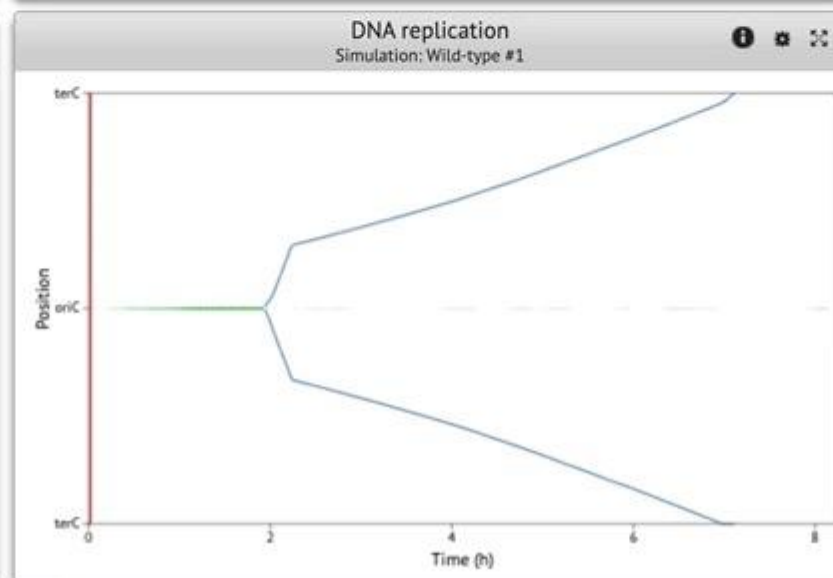
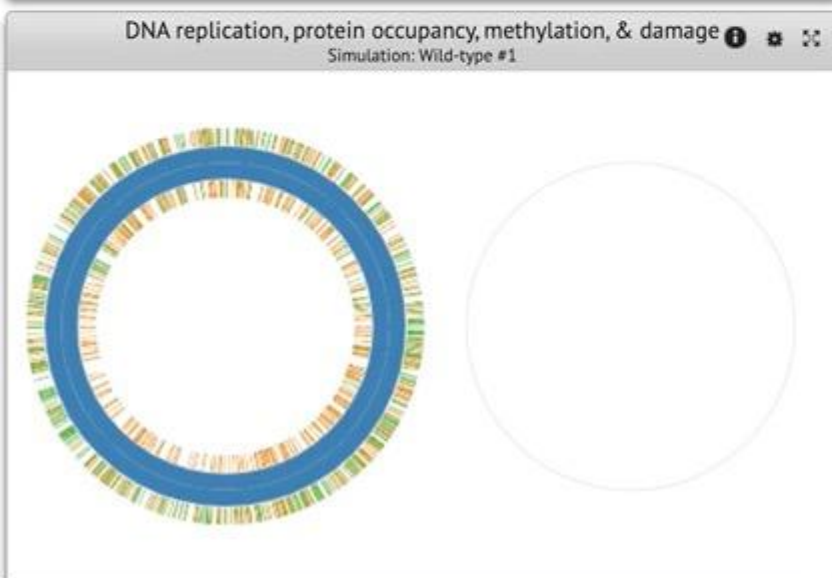
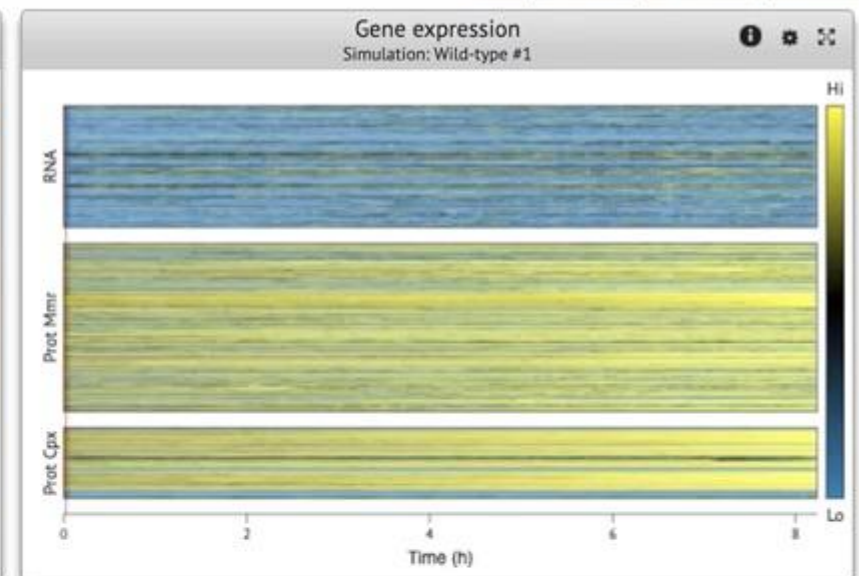
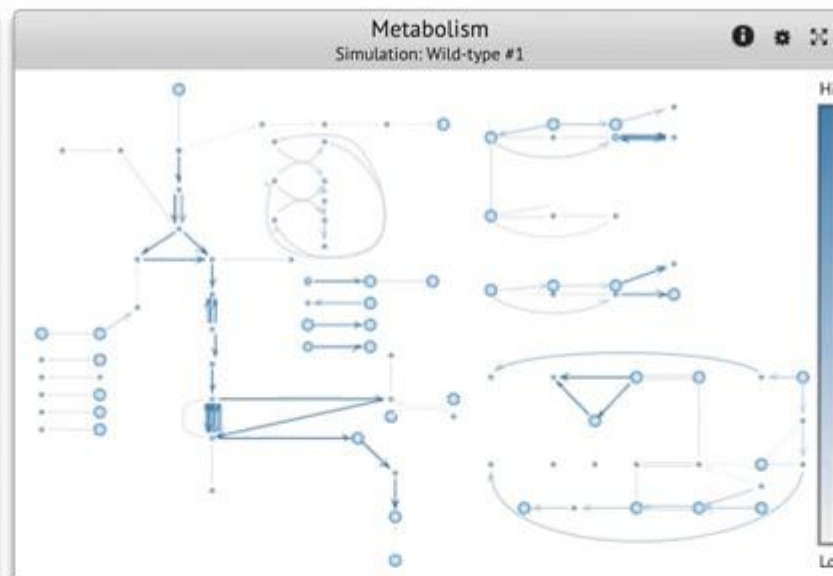
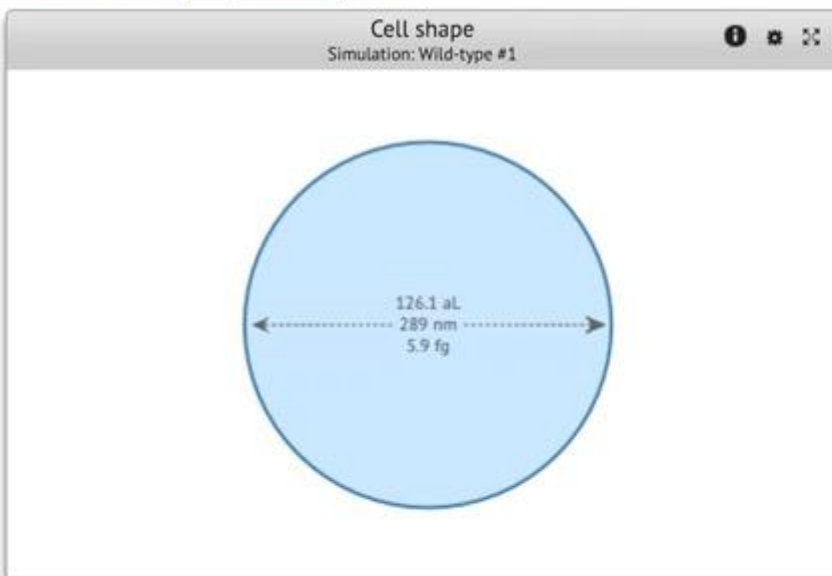
Availability: Bouhaddou M et al. A mechanistic pan-cancer pathway model informed by multi-omics data interprets stochastic cell fate responses to drugs and mitogens. *PLoS Comput Biol* 26: e1005985 (2018).

Author: Birtwistle Lab, Clemson

28 sous-modèles de 28 processus cellulaires
 401 gènes, 722 molécules, 1,857 réactions, and 1,836 paramètres

Mycoplasma genitalium Reconstruction ?



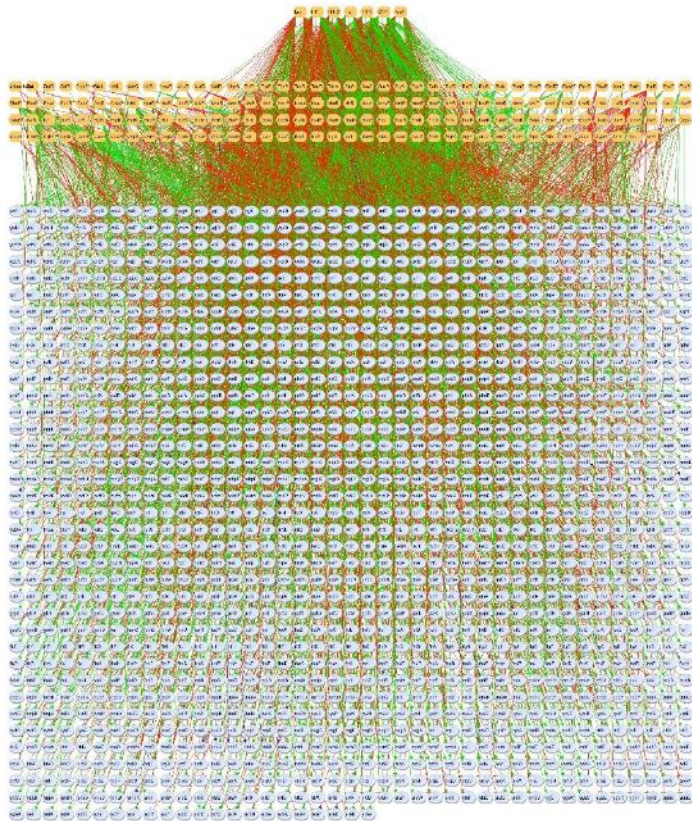


« Base de données la plus complète sur terre »

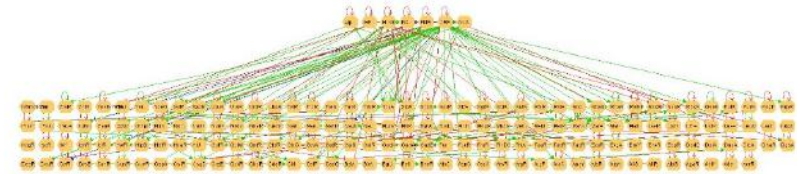
RegulonDB

Escherichia coli

Réseau TF vs gène



Réseau TF vs TF



- Réseau TF vs opéron
- Réseau Facteur sigma vs gène
- Réseau facteur sigma vs opéron

Regulon DB en 2023



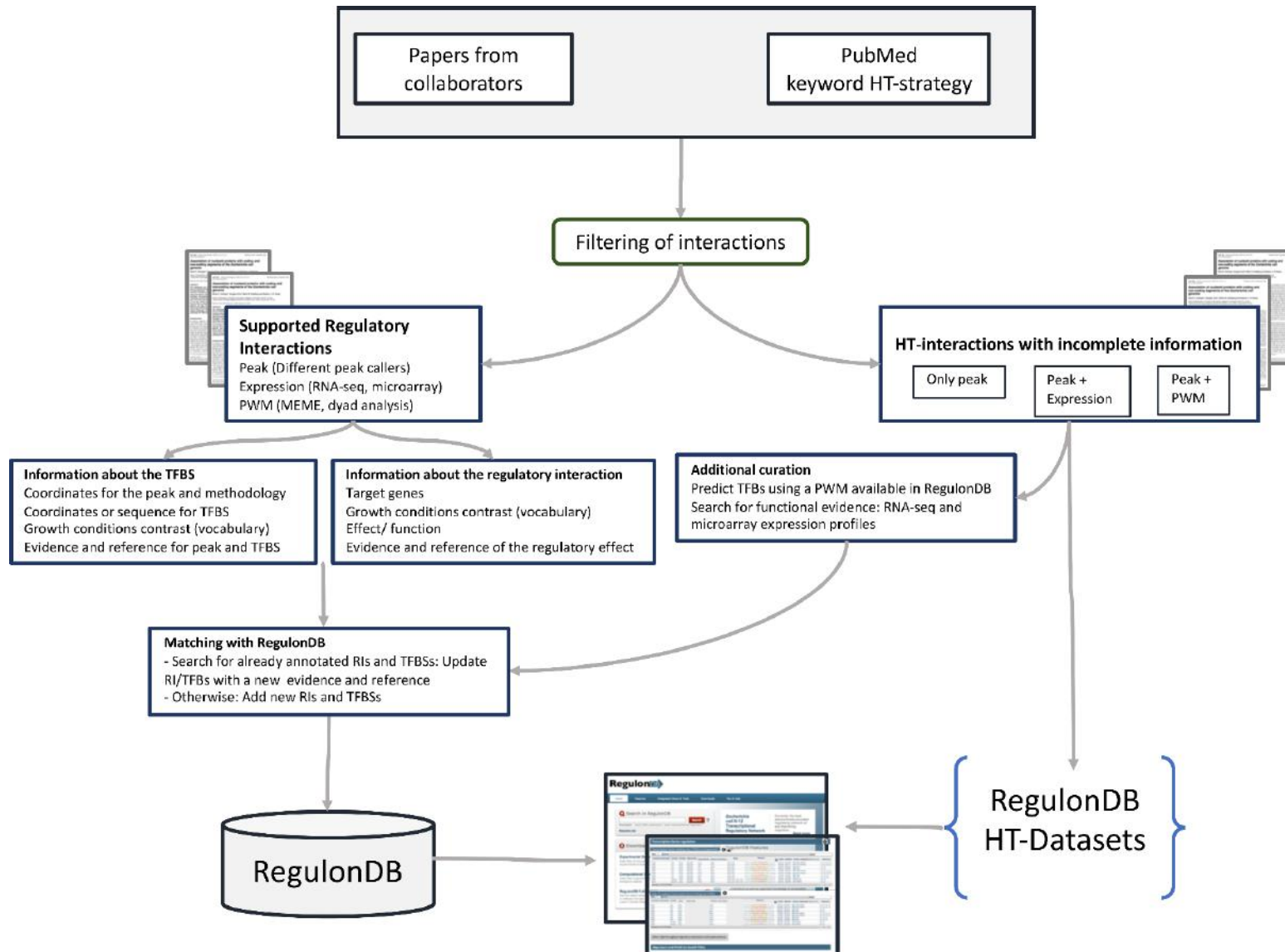
***Escherichia coli* K-12
Transcriptional
Regulatory Network**

Currently the best electronically-encoded regulatory network of any free-living organism.

[Read more](#)

Object	Total	Weak Evidence	Strong Evidence	Confirmed Evidence	Without Evidence
Transcription Units:	3696	2768	525		403
Genes:	4736				4736
Promoter:	8795	3099	5680	16	
Operon:	2592				2592
TF binding Sites:	6958	5030	1558		370
Regulatory Interactions:	3951	3169	86		696
small RNA Interactions:	247	171	76		
Terminators:	366				366
RBSs - Shine-Dalgarno:	179				179
Transcriptional Factors:	229	96	130		3
Simple Regulons:	124				
Complex Regulons:	432				
Effectors:	138				138
Attenuators:	751				751
Riboswitches:	51				51
Synonyms:	30794				
Growth Global Conditions:	16				
Experiment Conditions:	82				
Affected Genes in Different Experimental Conditions:	316				
Gensor Unit:	53				

RegulonDB en 2023













« Base de données la plus complète sur terre »





















Escherichia coli

RegulonDB en 2023

Experimental datasets

Description	File
coli K-12 genome sequence used into RegulonDB	E. coli K-12 genome sequence raw format
	E. coli K-12 genebank
	E. coli K-12 genebank refseq
Sequences	Gene Sequence
	5' and 3' UTR sequence of TUs
Gene - Product	All gene products
	Gene Product Identifiers
	sRNA genes
Transcriptional Factors - Functional conformation	 Download
Regulatory Interactions	 Download
Regulatory Network Interactions	TF - gene interactions 
	TF - operon interactions 
	TF - TU interactions 
	TF - TF interactions 
	Sigma - gene interactions 
	Sigma - TU interactions 
	Alon and MA interactions 
	sRNA - gene interactions 

Promoters	All Promoters 
	Sigma 70 
	Sigma 54 
	Sigma 38 
	Sigma 32 
	Sigma 28 
	Sigma 24 
	Sigma 19 
	Unknown 
Transcription start sites experimentally determined in the laboratory of Dr. Morett	High-throughput transcription initiation mapping. Illumina directional RNA-seq experiments were total RNA received different treatments to enrich for 5' monophosphate or 5' triphosphate ends. Version 3.0. See the file description. 
	High-throughput transcription initiation mapping. See the file description. 
	5'-RACE transcription initiation mapping with specific primers. See the file description. 
Transcription Factor Weight Matrix	TF-Matrix browser 
	 Download
Active and Inactive Transcription Factor Conformations	 Download
Transcription Units	 Download
Operons	 Download
Growth Conditions	 Download

RegulonDB en 2023

Computational Predictions datasets

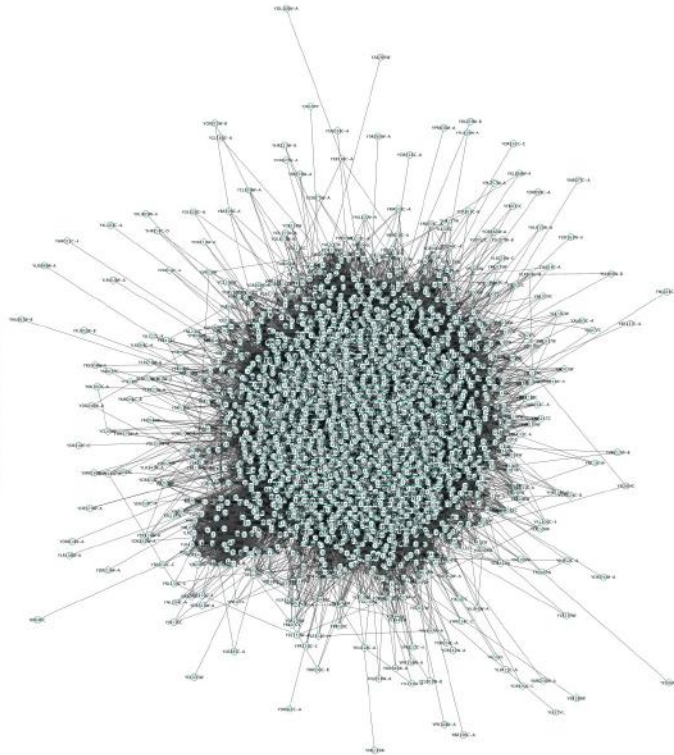
Description	Method
Promoter predictions	"We observed that real promoters occur mostly within regions with high densities of overlapping putative promoters. We evaluated several strategies to identify promoters. The best one uses an intrinsic score of the -10 and -35 hexamers that form the promoter as well as an extrinsic score that uses the distribution of promoters from the start of the gene. This high signal density is found mainly within regions upstream of genes, contrasting with coding regions and regions located between convergently transcribed genes." A.M. Huerta, J. Collado-Vides, J Mol Biol. 333:261-78 (2003).
Operon predictions	Operon prediction on (intergenic) distances Operon predictions based on (intergenic) distances and Riley's functional classification.
	We have previously demonstrated that genes within experimentally characterized operons of <i>Escherichia coli</i> are conserved together in other genomes more frequently than genes located at the borders of transcription units. We also show the relationship between our analyses of conservation and the inference of functional relationships from a genomic context
TF binding sites predictions	We have taken advantage of the phylogenetic proximity of <i>Escherichia coli</i> and other 16 organisms of this subdivision and the intensive search of the space sequence provided by a pattern-matching strategy. Using this approach, we complement predictions of regulatory sites made by using statistical models currently stored in Tractor_DB, and increase the number of transcriptional regulators with predicted binding sites up to 86. The original prediction approach, based on the representation of binding sites through statistical models was complemented by a new approach that uses known <i>E. coli</i> regulatory sites as the basis for a pattern matching search of regulatory sites. The use of both approaches together resulted in a more intensive exploration of the sequence space of each regulator's binding site.
	Computationally predicted transcription factor binding sites (TFBSs) using the evaluated weight matrix (see http://regulondb.ccg.unam.mx/menu/download/datasets/index.jsp). We scanned all upstream regions of every single gene, from +50 to -400 or from +50 to the closest upstream ORF, whatever happens first. (see the methodology)

tors Predictions	"Regulatory proteins in <i>Escherichia coli</i> with a helix-turn-helix (HTH) DNA binding motif show a position-function correlation such that repressors have this motif predominantly at the N terminus, whereas activators have the motif at the C-terminus extreme. Evidence is presented supporting a common history at the origin of this correlation. These results suggest that if shuffling of motifs occurred in Bacteria, it occurred only early in the history of these proteins, as opposed to what is observed in eukaryotic regulators." Pérez-Rueda E, Collado-Vides J. J Mol Evol. 2001 Sep;53(3):172-9.
s Prediction	For each group of orthologous proteins, the upstream regions of the first gene of each operon are taken and searched for motifs using MEME (Figure 1a). Each motif is then refined by several cycles of locating it among all upstream regions from all bacteria using MAST, and redefining a more specific motif with MEME (Figure 1b). Sequences with motifs can then be analyzed to see if they present evidence of conserved secondary structure (Figure 1c). Predicted motifs are also compared against the Rfam database to locate known structured elements and against RegulonDB to find known transcription factor binding sites. Click here to see image.
Prediction	For each predicted operon, the upstream region of the first gene is taken (Figure 1a). For every run of Us present in this region (Figure 1b), a stable structure in the adjacent region is searched for (Figure 1c). If a terminator is found, an anti-terminator is searched for, since it must be overlapping with the terminator (Figure 1d). An anti-antiterminator can be analogously located by finding a structure that overlaps with the anti-terminator (Figure 1e). For the particular case of translational attenuators, a terminator is searched for, since it overlaps with the Shine-Dalgarno site. Click here to see image.

Saccharomyces cerevisiae – YeastNet v3

[YeastNet: a network by integration of all data-type-specific networks (CC, CX, DC, GN, GT, HT, LC PG, TS)]

YeastNet v.3		
Edge list download	5808 genes	362421 links



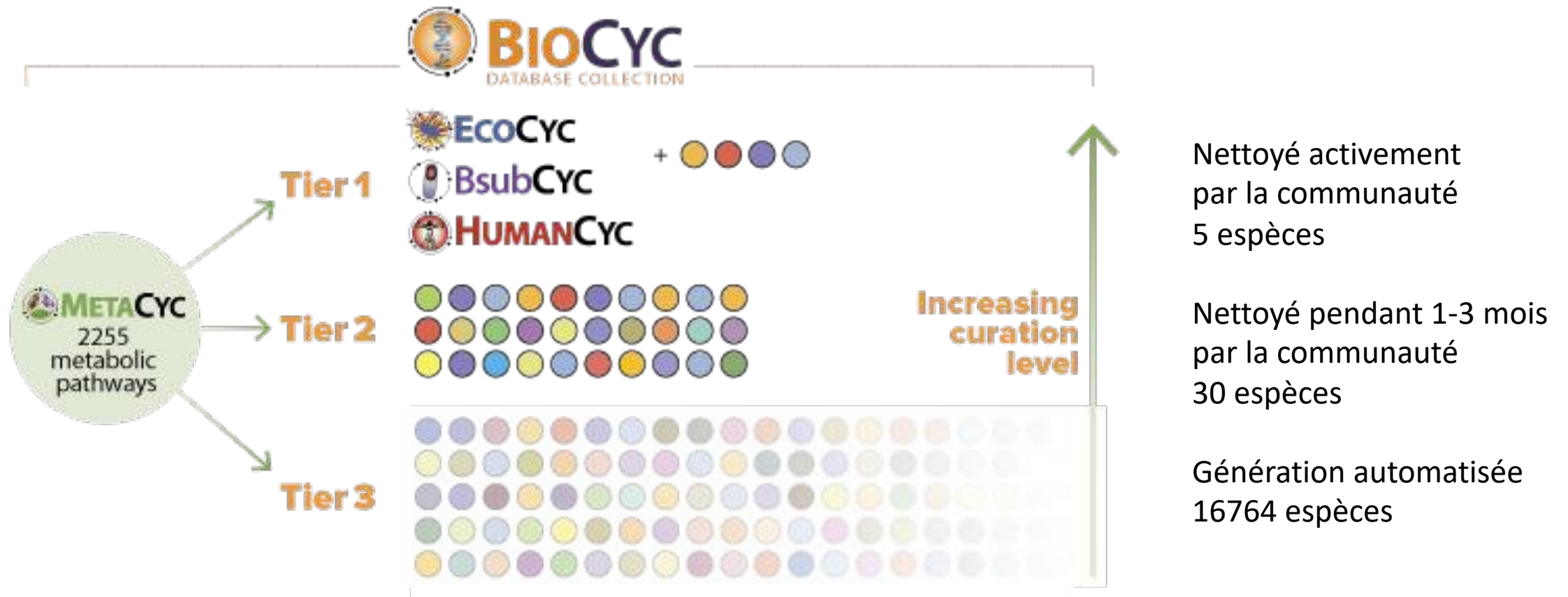
Kim et al.,
Nucleic Acid Research 2014

Saccharomyces cerevisiae – YeastNet v3

Evidence code	Data set description
CC	Inferred links by co-citation of two genes across 46,111 pubmed Medline article abstracts for yeast biology
CX	Inferred links by co-expression pattern of two genes (based on high-dimensional gene expression data)
DC	Inferred links by co-occurrence of protein domains between two coding genes
GN	Inferred links by similar genomic context of bacterial orthologs of two yeast genes

GT	Inferred links by similar profiles of genetic interaction partners
HT	Links by high-throughput protein-protein interactions
LC	Links by small/medium-scale protein-protein interactions (collected from protein-protein interaction data bases)
PG	Inferred links by similar phylogenetic profiles between two yeast genes
TS	Inferred links by 3-D protein structure of interacting orthologous proteins between two yeast proteins

Bases de données « nettoyées » par la communauté



Etat de l'art : BioCyc



Genes:	4,737	SmartTable	Ontology
Pathways:	365	SmartTable	Ontology
Enzymatic Reactions:	2,202		
Transport Reactions:	530		Ontology
Polypeptides:	4,466	SmartTable	Ontology
Protein Complexes:	1,166	SmartTable	Ontology
Enzymes:	1,714	SmartTable	
Transporters:	479	SmartTable	
Compounds:	2,967	SmartTable	Ontology
Transcription Units:	3,694	SmartTable	Ontology
tRNAs:	89		Ontology
Growth Media:	438		List
Transcriptional Regulation:	5,661		Ontology
Protein Features:	41,346		
Phenotype Microarray Datasets:	5		List
GO Terms:	71,124		Ontology
Gene Essentiality Datasets:	6		List



Genes:	4,541	SmartTable	Ontology
Pathways:	274	SmartTable	Ontology
Enzymatic Reactions:	1,529		
Transport Reactions:	92		Ontology
Polypeptides:	4,293	SmartTable	Ontology
Protein Complexes:	255	SmartTable	Ontology
Enzymes:	1,072	SmartTable	
Transporters:	632	SmartTable	
Compounds:	1,007	SmartTable	Ontology
Transcription Units:	1,648	SmartTable	Ontology
tRNAs:	86		Ontology
Growth Media:	1		List
Transcriptional Regulation:	864		Ontology
Protein Features:	22,636		
GO Terms:	34,078		Ontology
Gene Essentiality Datasets:	1		List



Genes:	20,997	SmartTable	Ontology
Pathways:	362	SmartTable	Ontology
Enzymatic Reactions:	2,895		
Transport Reactions:	145		Ontology
Polypeptides:	20,732	SmartTable	Ontology
Protein Complexes:	541	SmartTable	Ontology
Enzymes:	3,541	SmartTable	
Transporters:	773	SmartTable	
Compounds:	2,119	SmartTable	Ontology
tRNAs:	53		Ontology
Growth Media:	2		List
Transcriptional Regulation:	2		Ontology
Protein Features:	14		
GO Terms:	890,187		Ontology
Gene Essentiality Datasets:	1		List

Etat de l'art : MetaCyc

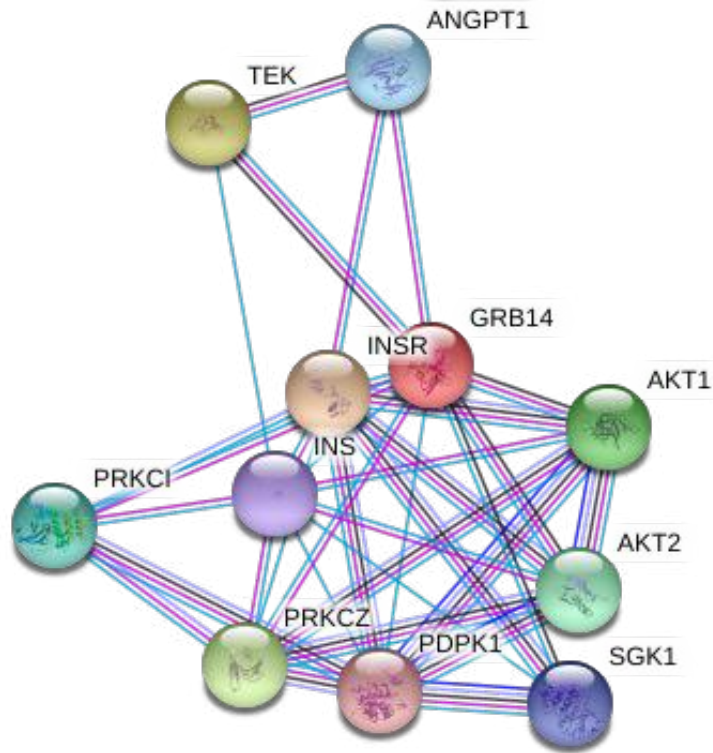


MetaCyc is a curated database of experimentally elucidated metabolic pathways from all domains of life.

MetaCyc contains pathways involved in both primary and secondary metabolism, as well as associated metabolites, reactions, enzymes, and genes. The goal of MetaCyc is to catalog the universe of metabolism by storing a representative sample of each experimentally elucidated pathway.

Genes:	14,976	SmartTable	Ontology
Pathways:	3,063	SmartTable	Ontology
Enzymatic Reactions:	18,285		
Transport Reactions:	910		Ontology
Polypeptides:	16,731	SmartTable	Ontology
Protein Complexes:	4,756	SmartTable	Ontology
Enzymes:	14,028	SmartTable	
Transporters:	638	SmartTable	
Compounds:	18,452	SmartTable	Ontology
tRNAs:	8		Ontology
Growth Media:	19		List
Protein Features:	30,982		
GO Terms:	72,124		Ontology

Le serveur String-DB.org



Réseau d'interaction protéine-protéine de l'insuline prédite par string-db.org

Known Interactions

- from curated databases*
- experimentally determined*

Predicted Interactions

- gene neighborhood*
- gene fusions*
- gene co-occurrence*

Welcome to STRING

Protein-Protein Interaction Networks
Functional Enrichment Analysis

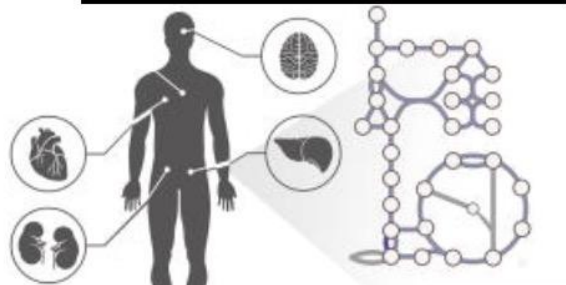
ORGANISMS	PROTEINS	INTERACTIONS
14094	67.6 mio	>20 bln

SEARCH

Others

- textmining*
- co-expression*
- protein homology*

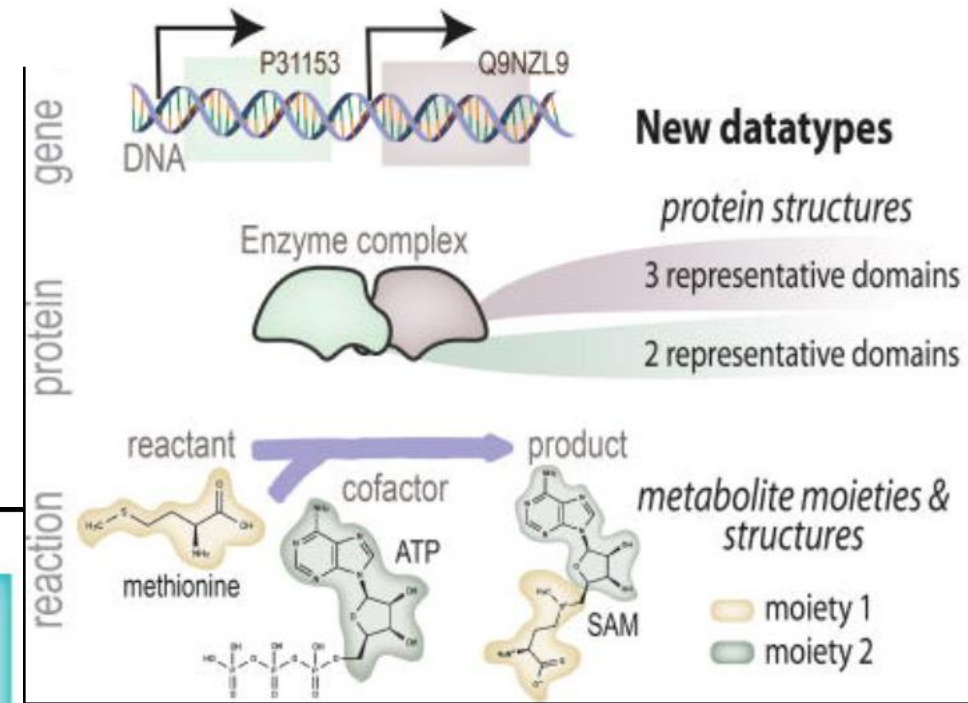
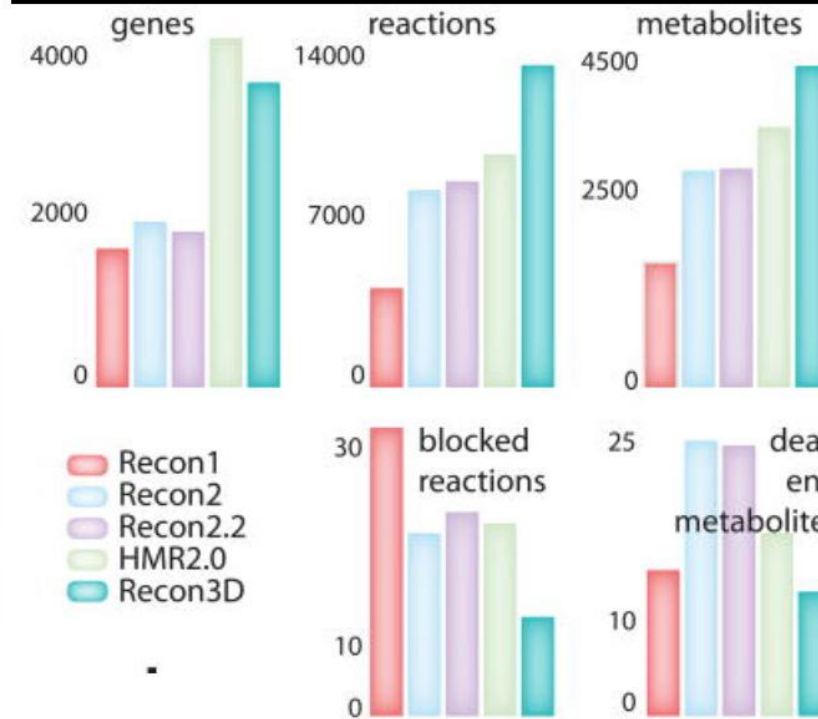
Recon3D : Human metabolism

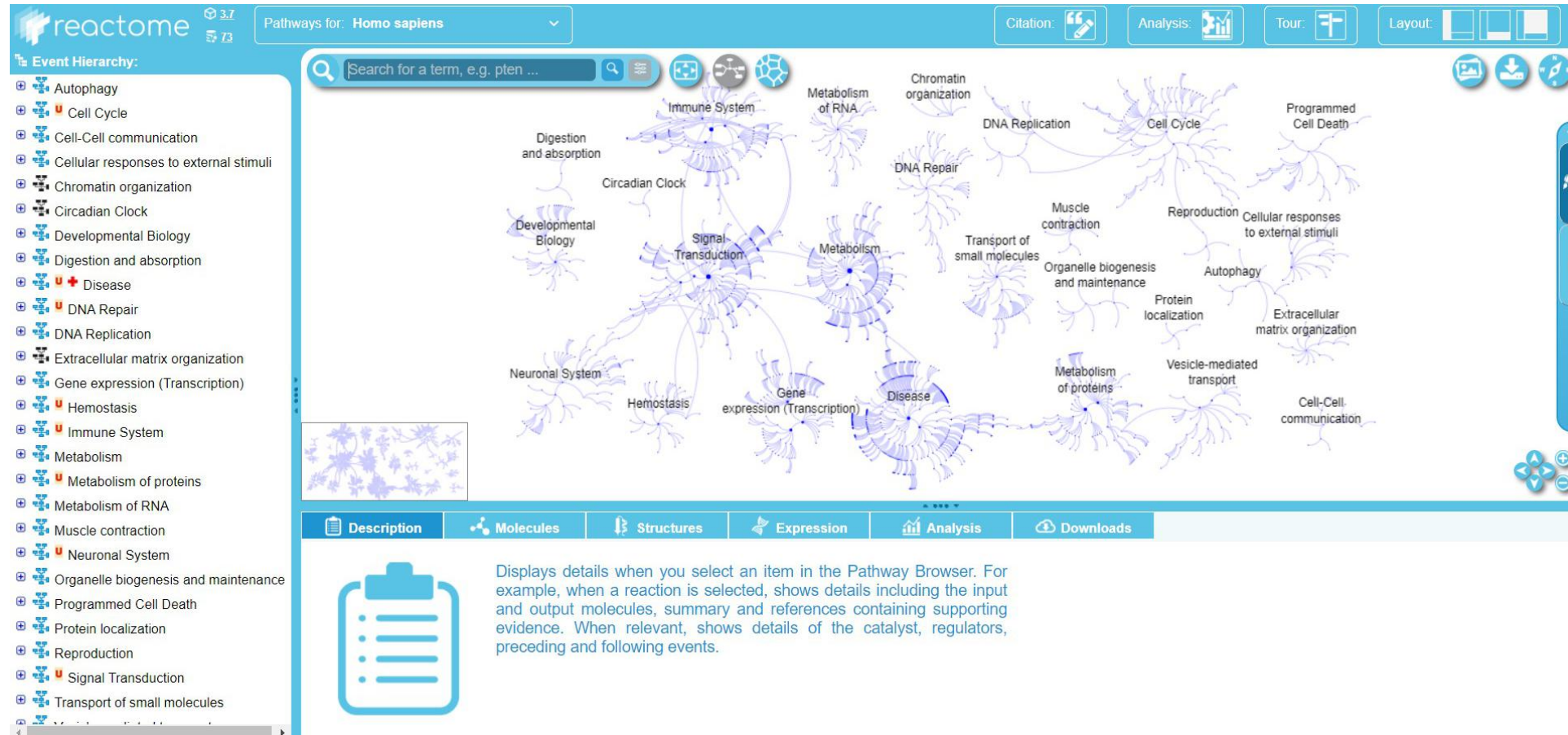


biochemical data
genomic data
tissue-specific localization
proteomic data
metabolomic data
protein structural data
pharmacogenomic data
atom-atom mappings

Recon 3D

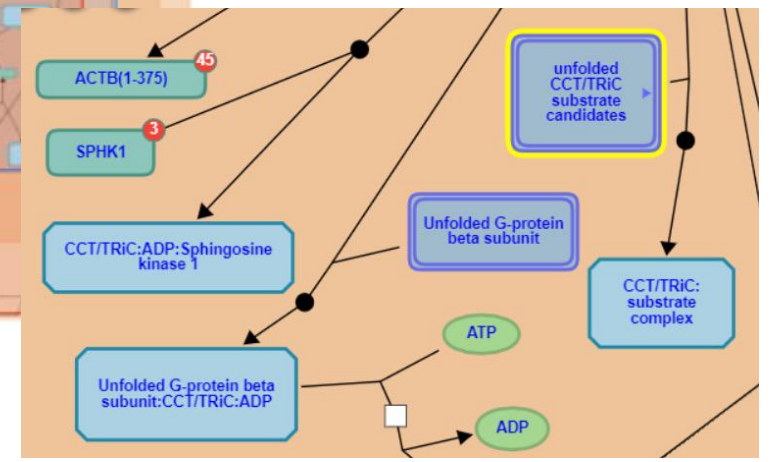
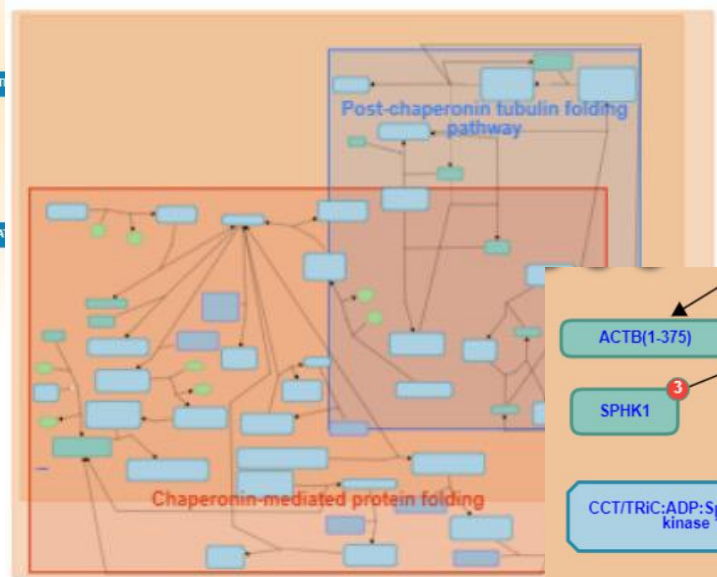
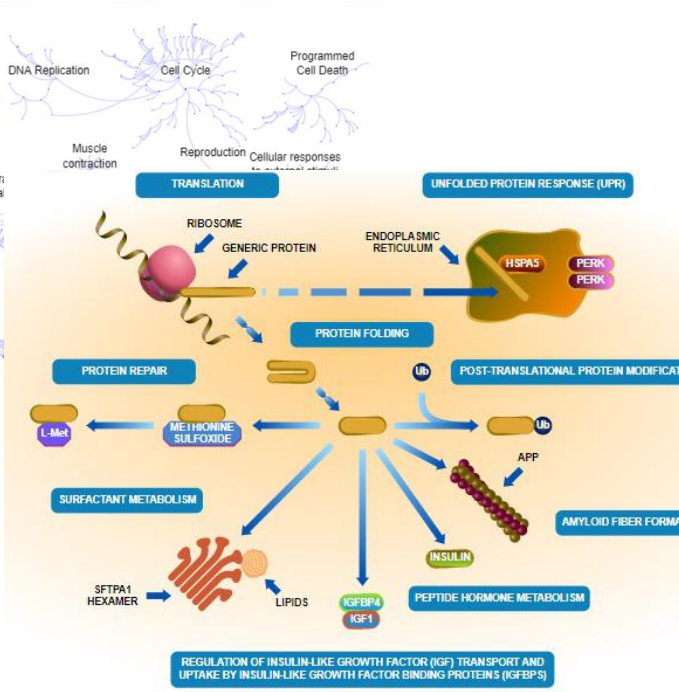
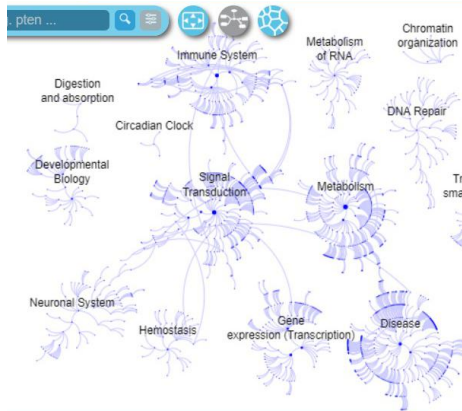
3288 genes
2908 domains
12890 structures
13543 reactions
4140 metabolites
3536 SNVs
8315 atom-atom mappings





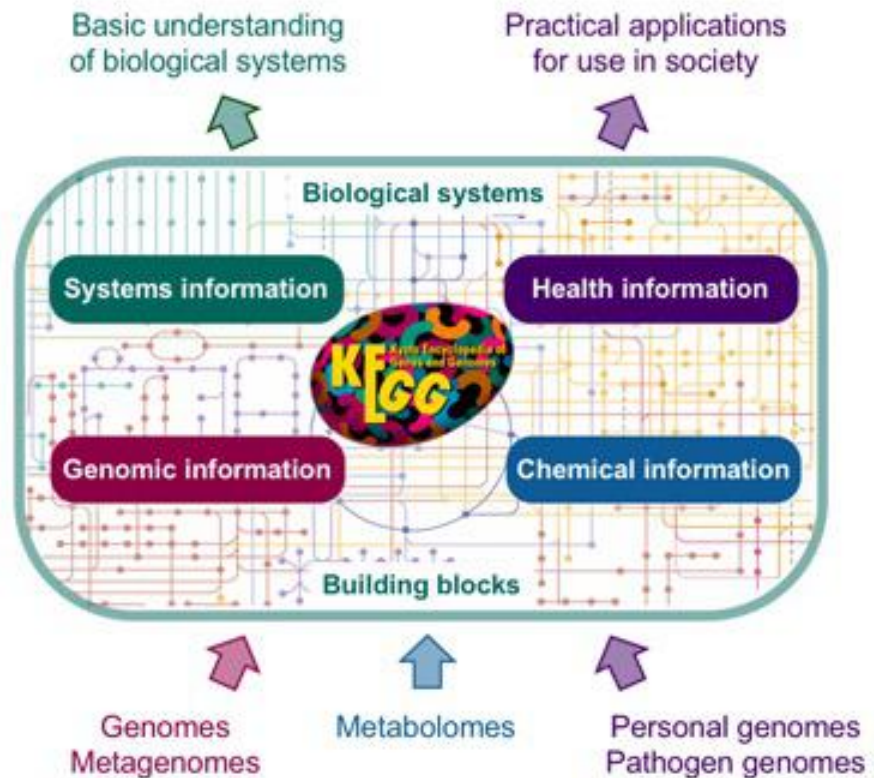
The screenshot displays the Reactome web interface for Homo sapiens. The top navigation bar includes the Reactome logo, version 3.7, and 73 pathways. A search bar is present with the text "Search for a term, e.g. pten ...". The main area features a hierarchical tree of biological processes, with "Signal Transduction" and "Metabolism" being prominent. A detailed description panel is visible at the bottom, showing a clipboard icon and text explaining the panel's function: "Displays details when you select an item in the Pathway Browser. For example, when a reaction is selected, shows details including the input and output molecules, summary and references containing supporting evidence. When relevant, shows details of the catalyst, regulators, preceding and following events."

Base de données de réaction chimique impliquées dans le métabolisme d'organisme modèles



On peut parcourir plusieurs niveaux de « réactions »

Kyoto Encyclopedia of Genes and Genomes



Category	Database	Content	Color
Systems information	KEGG PATHWAY	KEGG pathway maps	KEGG
	KEGG BRITE	BRITE hierarchies and tables	
	KEGG MODULE	KEGG modules and reaction modules	
Genomic information	KEGG ORTHOLOGY (KO)	Functional orthologs	KEGG
	KEGG GENES	Genes and proteins	
	KEGG GENOME	KEGG organisms and viruses	
Chemical information	KEGG COMPOUND	Metabolites and other chemical substances	KEGG
	KEGG GLYCAN	Glycans	
	KEGG REACTION	Biochemical reactions	
	KEGG RCLASS	Reaction class	
	KEGG ENZYME	Enzyme nomenclature	
Health information	KEGG NETWORK	Disease-related network variations	KEGG
	KEGG VARIANT	Human gene variants	
	KEGG DISEASE	Human diseases	
	KEGG DRUG	Drugs	
	KEGG DGROUP	Drug groups	

Chemical information category is collectively called **KEGG LIGAND**

Health information category integrated with drug labels is called **KEGG MEDICUS**

Gene Ontology

3 composants dans GO :

- Cellular component

Endosome, Mitochondrion, Apoptosome, Centriole...

Où ?

- Biological process

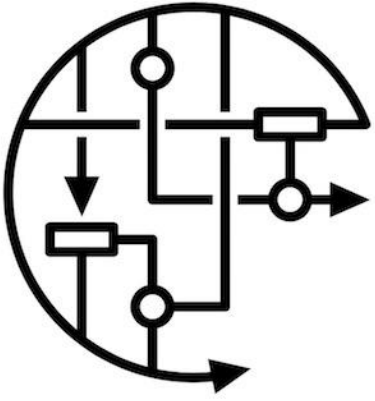
Notch signaling pathway, G1 phase, Serotonin biosynthesis...

Pour quoi faire ?

- Molecular function

Inositol 3-kinase activity, clathrin binding...

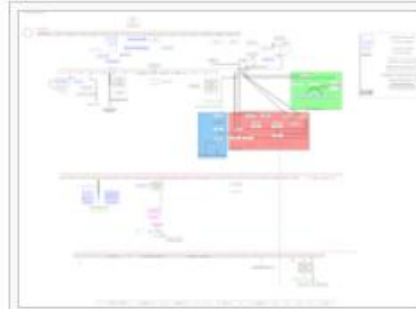
Comment ?



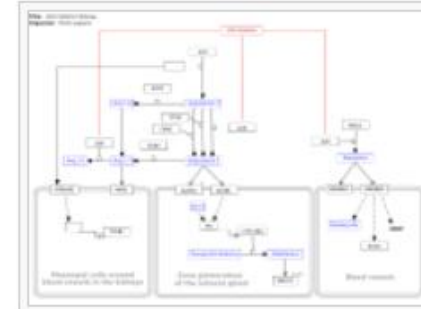
WIKIPATHWAYS
Pathways for the People

Browse pathways

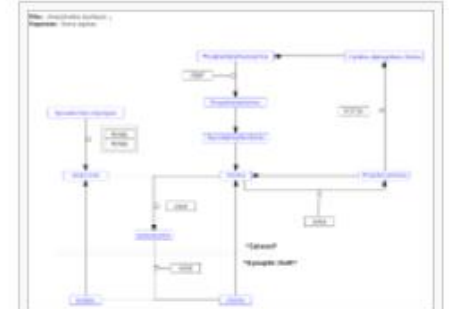
Species: Collection: View:



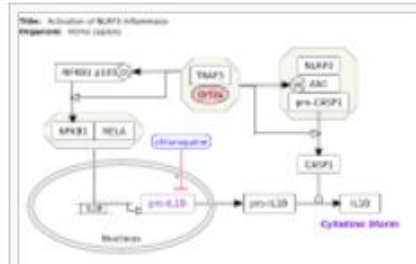
22q11.2 copy number variation syndrome



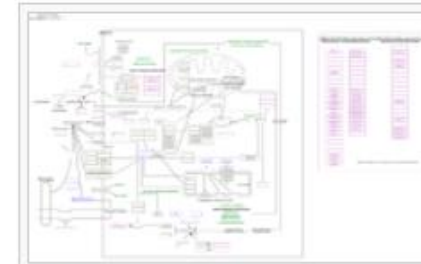
ACE Inhibitor Pathway



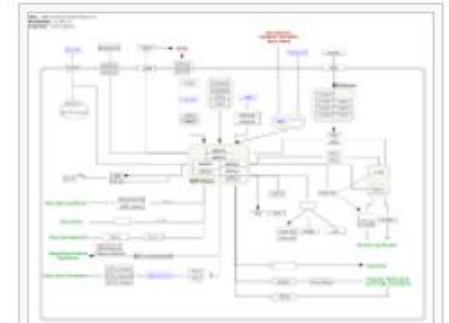
Acetylcholine Synthesis



Activation of NLRP3 Inflammasome by SARS-CoV-2



Alzheimers Disease



AMP-activated Protein Kinase (AMPK) Signaling



Base de données de voies de signalisation
« nettoyées » par la communauté

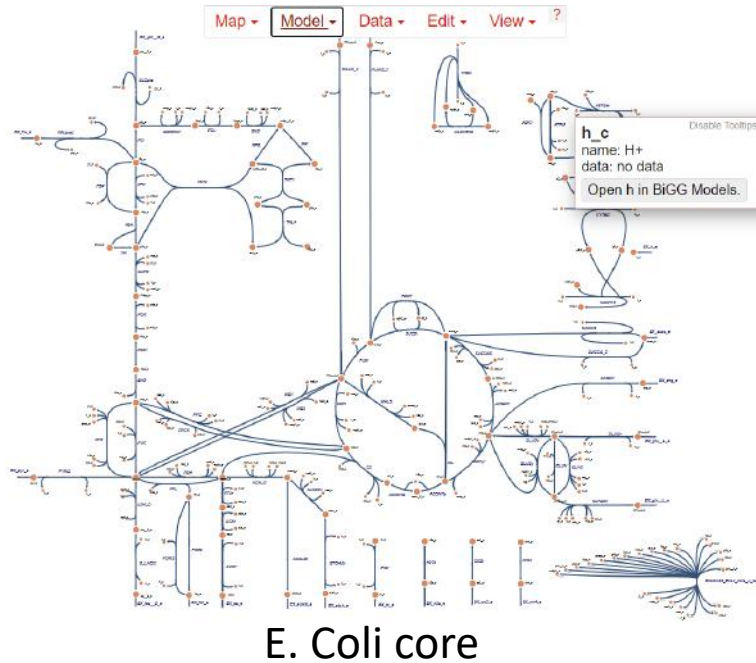
BiGG Models

Palsson et al.

Search the database by model, reaction, metabolite, or gene ?



The image shows three buttons arranged horizontally. The first button, 'View Models', features a diagram of a cell with labels for DNA, Ribosome, Enzyme, Membrane, Biomass precursors, and Actual biomass. The second button, 'View Metabolites', shows a 3D illustration of a bacterium with various chemical structures floating around it. The third button, 'View Reactions', displays a metabolic pathway diagram with enzymes like GLCptasep, PDW, PFL, PTAR, ACALD, and ALCOX labeled.



E. Coli core



openCOBRA

The COntstraint-Based Reconstruction and Analysis Toolbox

NDEx-The Network Data Exchange

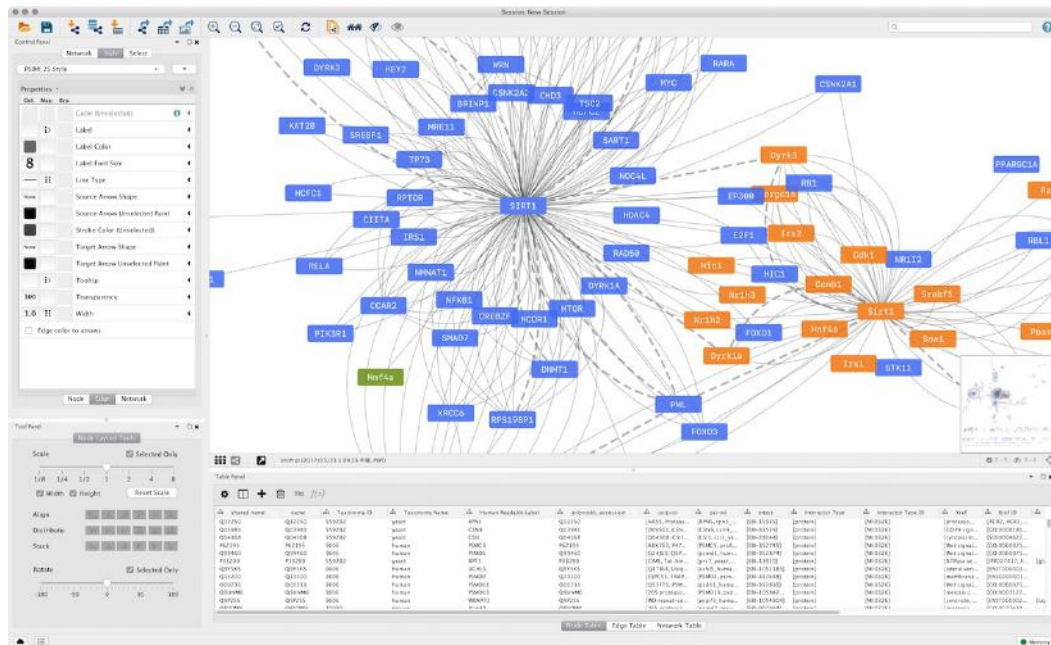
- Accéder à différentes bases de données de réseau
- Télécharger un réseau/modèle particulier
- Uploader un réseau/modèle particulier
- Recherche à travers les réseaux d'éléments (gène, opéron, etc)
- Opensource

Cytoscape

<https://cytoscape.org/download.html>

(Vous devez avoir installé Java 11)

- Importer des réseaux
- Parcourir des réseaux
- Visualiser des réseaux
- Modifier l'aspect des réseaux
- Exporter des réseaux



Biologie des systèmes

- **L'approche systémique en biologie**
- **Reconstruire un réseau biologique**
 - Les différents types de réseaux
 - Reconstruction directe
 - Reconstruction indirecte
 - Les obstacles à la reconstruction
- **Analyser un réseau biologique**
 - Théorie des graphes
 - Topologie des réseaux
 - Les réseaux sans échelles
- **Utiliser un réseau biologique**
 - Etat de l'art des réseaux biologiques les plus étendus
 - Bases de données en biologie des systèmes
- **Modéliser un réseau biologique**



Modéliser

Modéliser = Partir des interactions définies dans le réseau et décrire/prédire le comportement du système

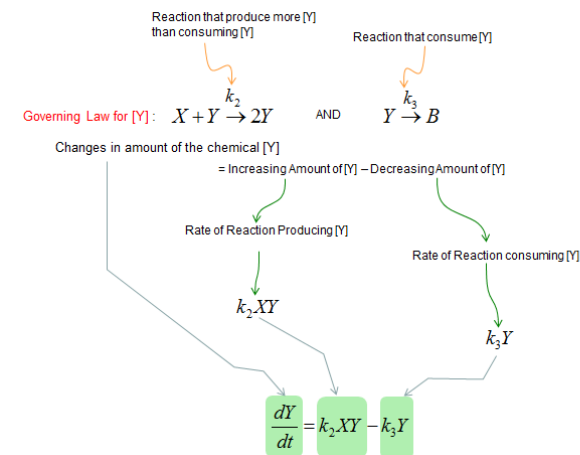
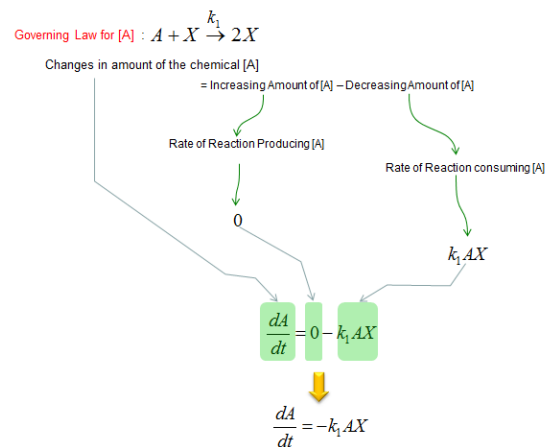
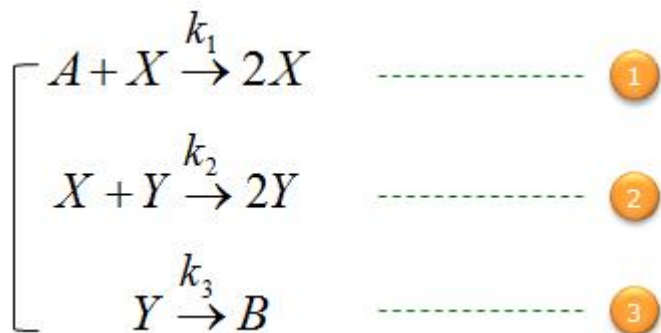
- Prédire un état d'équilibre
- Décrire comment le système va évoluer si on change une de ses propriétés
- Mieux comprendre la fonction de composants du système

Application de la modélisation

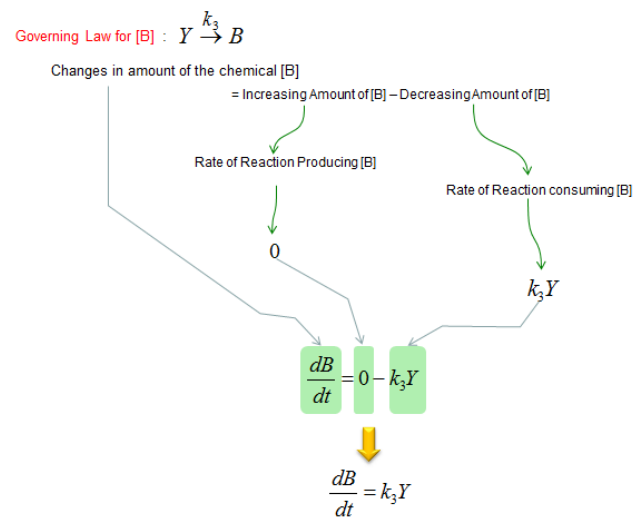
Modéliser = Partir des interactions définies dans le réseau et décrire/prédire le comportement du système

- En écologie : Comprendre le lien entre espèces
- En épidémiologie : Propagation de maladie au sein d'une population
- En biochimie : Trouver les concentrations idéales pour un rendement optimal
- En biologie synthétique : Comprendre quel élément du réseau de régulation modifier
- En pharmacologie : Prédire de nouvelles interactions, donc de nouvelles thérapeutiques
- ...

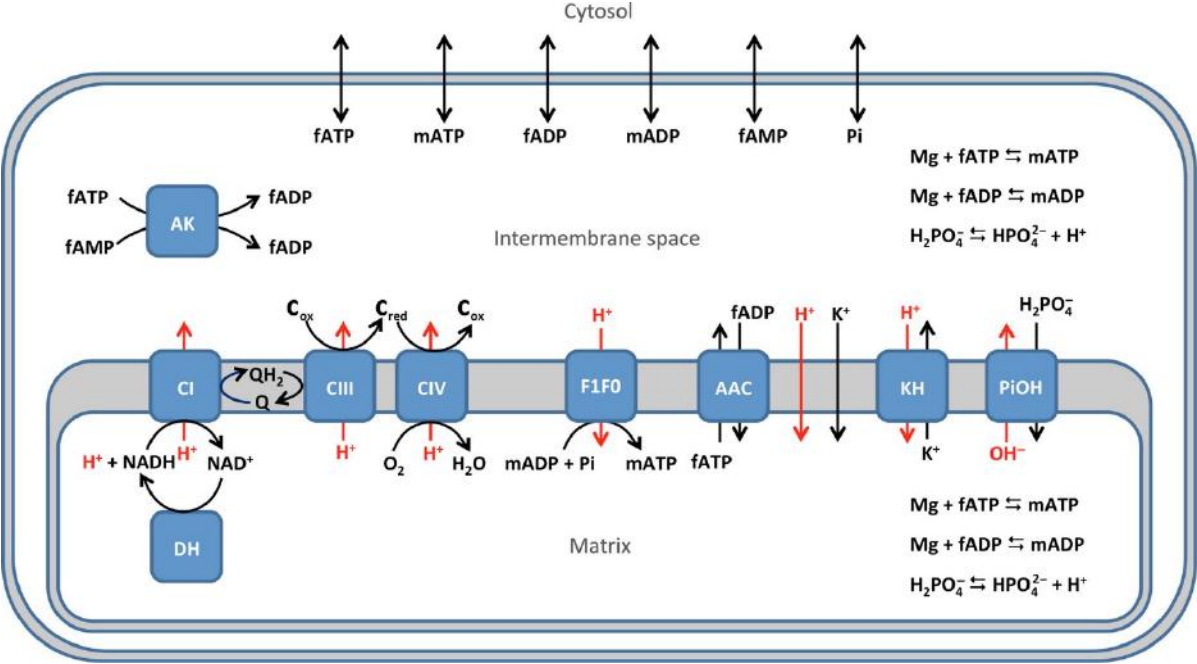
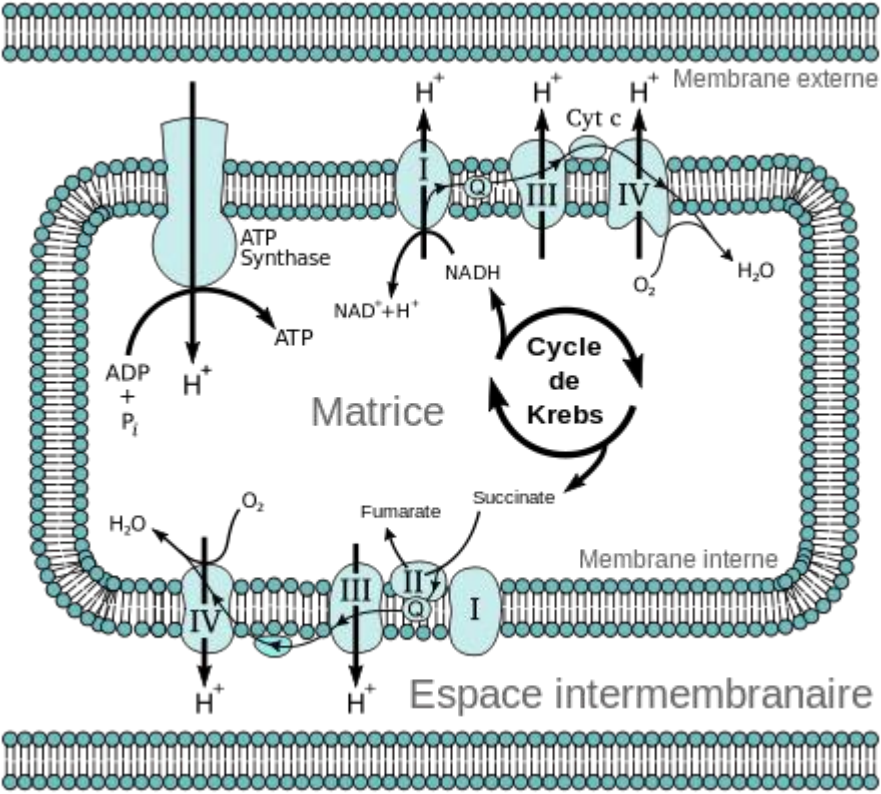
Modélisation d'équation chimique



$$\begin{cases}
 \frac{dA}{dt} = -k_1 AX \\
 \frac{dX}{dt} = k_1 AX - k_2 XY \\
 \frac{dY}{dt} = k_2 XY - k_3 Y \\
 \frac{dB}{dt} = k_3 Y
 \end{cases}$$



Exemple : Phosphorylation oxydative



Phosphorylation de l'ADP en ATP dans les mitochondries (wikipedia)

Heiske et al., FEBS Journal 2017

Exemple : Phosphorylation oxydative

Matrix		
DH	Dehydrogenase reaction	$NAD_x + H_x \rightleftharpoons NADH_x$
MgATP _x	Mg ²⁺ binding on fATP	$fATP_x + Mg_x \rightleftharpoons mATP_x$
MgADP _x	Mg ²⁺ binding on fADP	$fADP_x + Mg_x \rightleftharpoons mADP_x$
Pdiss _x	Pi dissociation	$H_2PO_{4x} \rightleftharpoons HPO_{4x} + H_x$
Inner membrane		
C1	Complex I	$NADH_x + Q_m + 5H \rightleftharpoons NAD_x + QH_{2im} + 4H_{is}$
C3	Complex III	$QH_{2im} + 2Co_{is} + 2H_x \rightleftharpoons Q_m + 2Cred_{is} + 4H_{is}$
C4	Complex IV	$Cred_{is} + 0.25 O_2 + 2H_x \rightleftharpoons Cox_{is} + 0.5H_2O + H_{is}$
F1F0	ATP synthase	$mADP_x + P_i + n_A H_x + H_x \rightleftharpoons mATP_x + n_A H_x$
AAC	ADP/ATP carrier	$fATP_x + fADP_{is} \rightleftharpoons fATP_{is} + fADP_x$
PiOH	Pi/OH antiporter	$H_2PO_{4is} + OH_x \rightleftharpoons H_2PO_{4x} + OH_{is}$
KH	K ⁺ /H ⁺ antiporter	$K_{is} + H_x \rightleftharpoons K_x + H_{is}$
Hleak	H ⁺ leak	$H_{is} \rightleftharpoons H_x$
Kleak	K ⁺ leak	$K_{is} \rightleftharpoons K_x$
Intermembrane space		
AK	Adenylate kinase	$mATP_{is} + fAMP_{is} \rightleftharpoons mADP_{is} + fADP_{is}$
MgATP _{is}	Mg ²⁺ binding on fATP	$fATP_{is} + Mg_{is} \rightleftharpoons mATP_{is}$
MgADP _{is}	Mg ²⁺ binding on fADP	$fADP_{is} + Mg_{is} \rightleftharpoons mADP_{is}$
Pdiss _{is}	Pi dissociation	$H_2PO_{4is} \rightleftharpoons HPO_{4is} + H_{is}$
Outer membrane		
fATP _{om}	fATP diffusion	$fATP_e \rightleftharpoons fATP_{is}$
mATP _{om}	mATP diffusion	$mATP_e \rightleftharpoons mATP_{is}$
fADP _{om}	fADP diffusion	$fADP_e \rightleftharpoons fADP_{is}$
mADP _{om}	mADP diffusion	$mADP_e \rightleftharpoons mADP_{is}$
fAMP _{om}	fAMP diffusion	$fAMP_e \rightleftharpoons fAMP_{is}$
Pi _{om}	Pi diffusion	$P_{ie} \rightleftharpoons P_{is}$
Mg _{om} *	Mg ²⁺ diffusion	$Mg_e \rightleftharpoons Mg_{is}$
External space/cytosol		
MgATP _e *	Mg ²⁺ binding on fATP	$fATP_e + Mg_e \rightleftharpoons mATP_e$
MgADP _e *	Mg ²⁺ binding on fADP	$fADP_e + Mg_e \rightleftharpoons mADP_e$

$$\frac{d[H]_x}{dt} = x_{cuff} \cdot \left(\frac{+v_{DH} - 5v_{C1} - 2v_{C3} - 4v_{C4}}{+(n_A - 1)v_{F1F0} + 2v_{PiOH} + v_{leak} - v_{KH}} \right) / W_x \quad (35)$$

$$\frac{d[K]_x}{dt} = (+v_{KH} + v_K) / W_x \quad (36)$$

$$\frac{d[Mg]_x}{dt} = -v_{MgATP_x} - v_{MgADP_x} \quad (37)$$

$$\frac{d[NADH]_x}{dt} = (+v_{DH} - v_{C1}) / W_x \quad (38)$$

$$\frac{d[fATP]_x}{dt} = -v_{AAC} / W_x - v_{MgATP_x} \quad (39)$$

$$\frac{d[mATP]_x}{dt} = +v_{F1F0} / W_x + v_{MgATP_x} \quad (40)$$

$$\frac{d[fADP]_x}{dt} = +v_{AAC} / W_x - v_{MgADP_x} \quad (41)$$

$$\frac{d[mADP]_x}{dt} = -v_{F1F0} / W_x + v_{MgADP_x} \quad (42)$$

$$\frac{d[P_i]_x}{dt} = (-v_{F1F0} + v_{PiOH}) / W_x \quad (43)$$

$$\frac{d[QH_2]_{im}}{dt} = (+v_{C1} - v_{C3}) / W_{im} \quad (44)$$

$$\frac{d[Cred]_{im}}{dt} = (+2v_{C3} - 2v_{C4}) / W_{is} \quad (45)$$

$$\frac{d[fATP]_{is}}{dt} = (+v_{ATP_{om}} + v_{AAC_{om}} - v_{AK_{is}}) / W_{is} - v_{MgATP_{is}} \quad (46)$$

$$\frac{d[mATP]_{is}}{dt} = +v_{mATP_{om}} / W_{is} + v_{MgATP_{is}} \quad (47)$$

$$\frac{d[fADP]_{is}}{dt} = (+v_{ADP_{om}} - v_{AAC_{om}} + 2v_{AK_{is}}) / W_{is} - v_{MgADP_{is}} \quad (48)$$

$$\frac{d[mADP]_{is}}{dt} = +v_{mADP_{om}} / W_{is} + v_{MgADP_{is}} \quad (49)$$

$$\frac{d[fAMP]_{is}}{dt} = (+v_{AMP_{om}} - v_{AK_{is}}) / W_{is} \quad (50)$$

$$\frac{d[P_{ie}]_e}{dt} = (-v_{PiOH} + v_{Pi_{om}}) / W_e \quad (51)$$

$$\frac{d\Delta\Psi}{dt} = (+4v_{C1} + 2v_{C3} + 4v_{C4} - n_A \cdot v_{F1F0} - v_{AAC} - v_{leak} - v_K) / C \quad (52)$$

$$[NAD]_x = N_{tot_x} - [NADH]_x \quad (53)$$

$$[Q]_{im} = Q_{tot_{im}} - [QH_2]_{im} \quad (54)$$

$$[Cox]_{is} = C_{tot_{is}} - [Cred]_{is} \quad (55)$$

$$[mADP]_e = \frac{1}{2} \left((K_{d_{mADP}} + ADP_{tot_e} + Mg_{tot_e}) - \sqrt{(K_{d_{mADP}} + ADP_{tot_e} + Mg_{tot_e})^2 - 4(Mg_{tot_e} \cdot ADP_{tot_e})} \right) \quad (56)$$

$$[fADP]_e = ADP_{tot_e} - [mADP]_e \quad (57)$$

$$[Mg]_e = Mg_{tot_e} - [mADP]_e \quad (58)$$

$$[Mg]_{is} = [Mg]_e \quad (59)$$

$$[P_i]_e = [P_i]_e \quad (60)$$

$$[K]_{is} = [K]_e \quad (61)$$

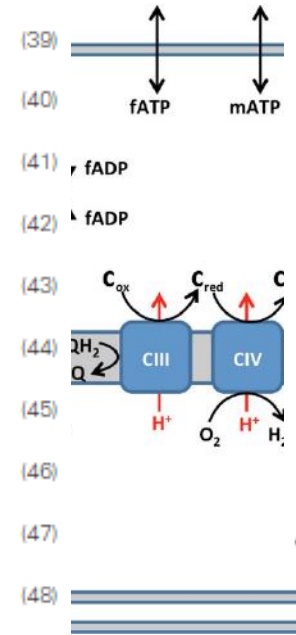
$$[H]_{is} = [H]_e \quad (62)$$

$$[H_2PO_4]_{is} = \frac{[H]_{is} \cdot [P_i]_{is}}{[H]_{is} + K_{d_{H_2PO_4}}} \quad (63)$$

$$[H_2PO_4]_x = \frac{[H]_x \cdot [P_i]_x}{[H]_x + K_{d_{H_2PO_4}}} \quad (64)$$

$$[OH]_{is} = \frac{10^{-14}}{[H]_{is}} \quad (65)$$

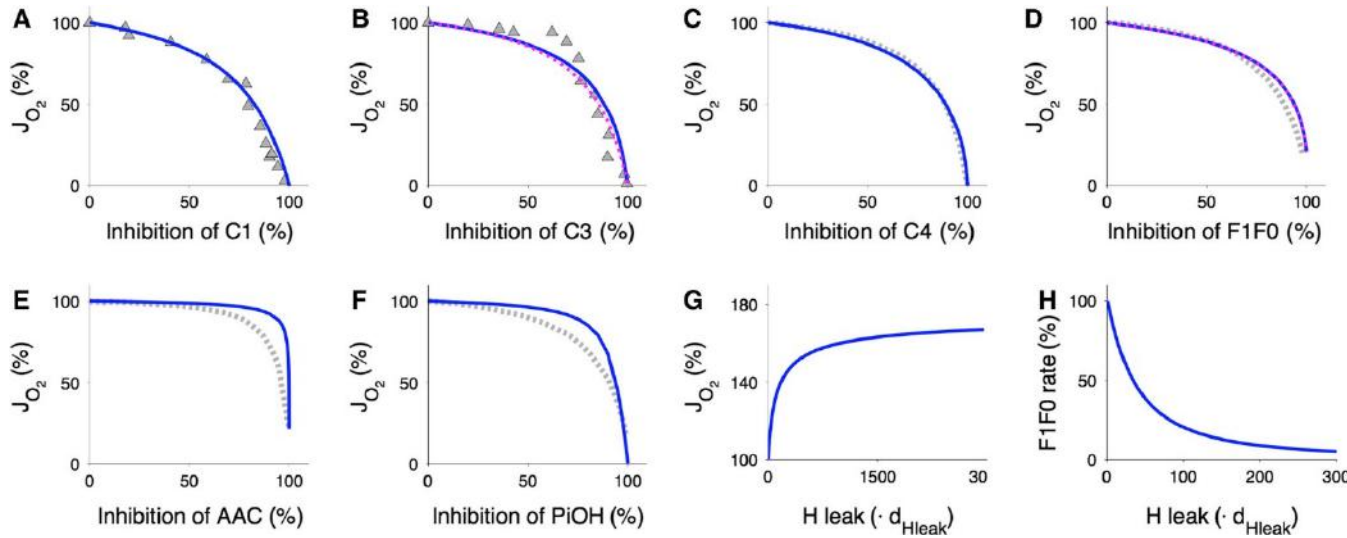
$$[OH]_x = \frac{10^{-14}}{[H]_x} \quad (66)$$



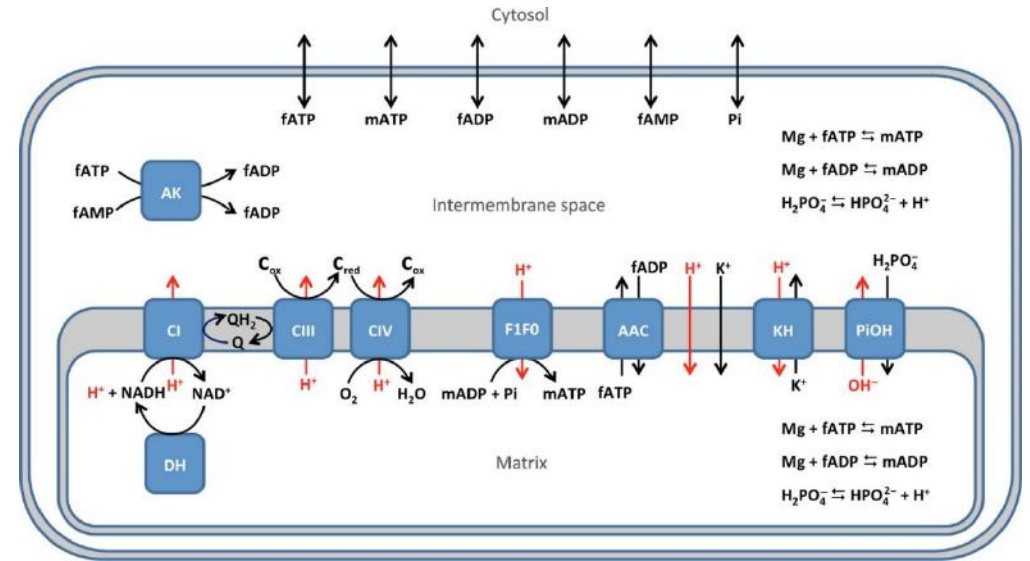
Heiske et al., FEBS Journal 2017

Exemple : Phosphorylation oxydative

Optimisation non-linéaire sous contrainte

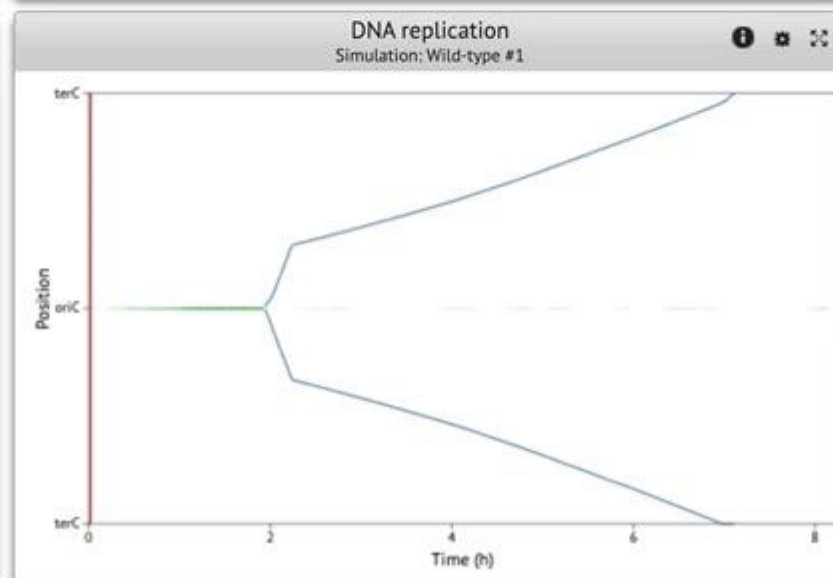
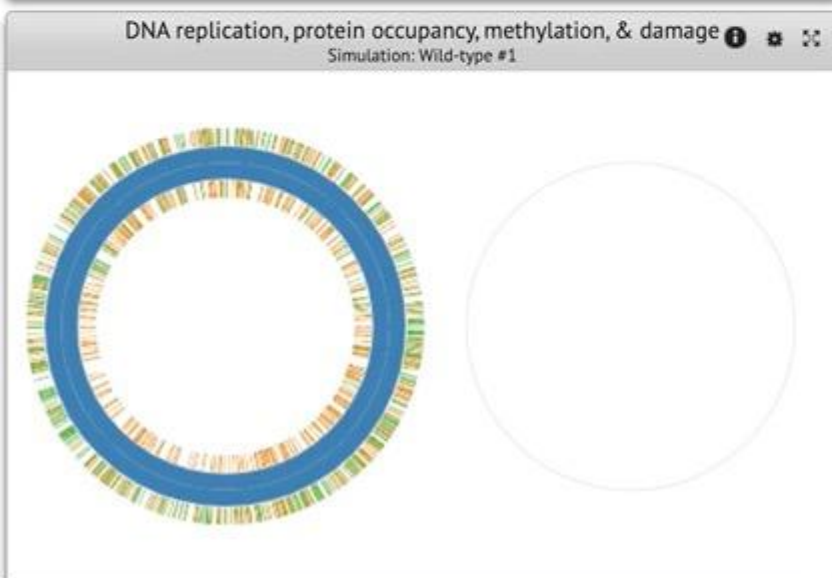
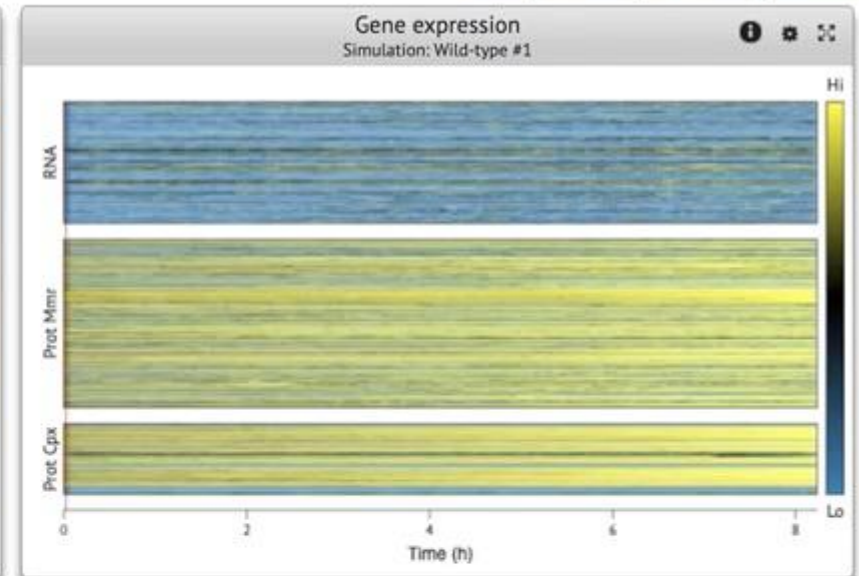
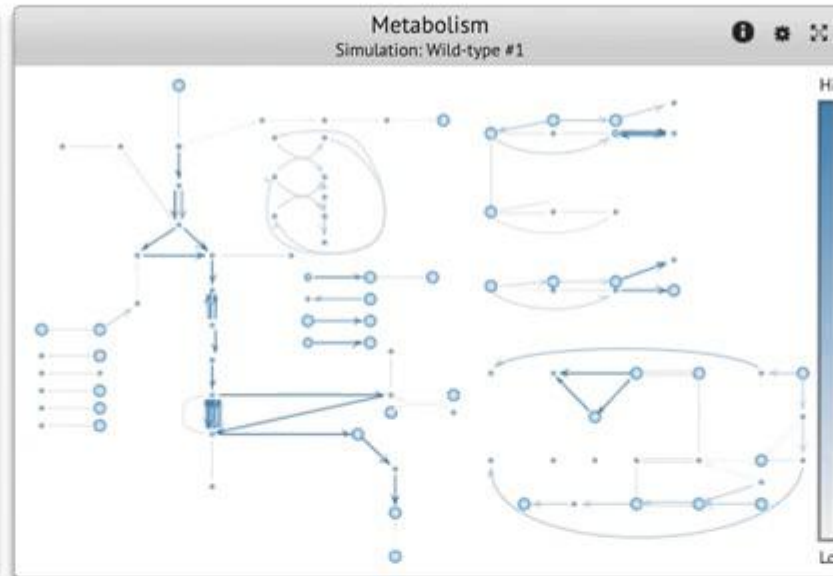
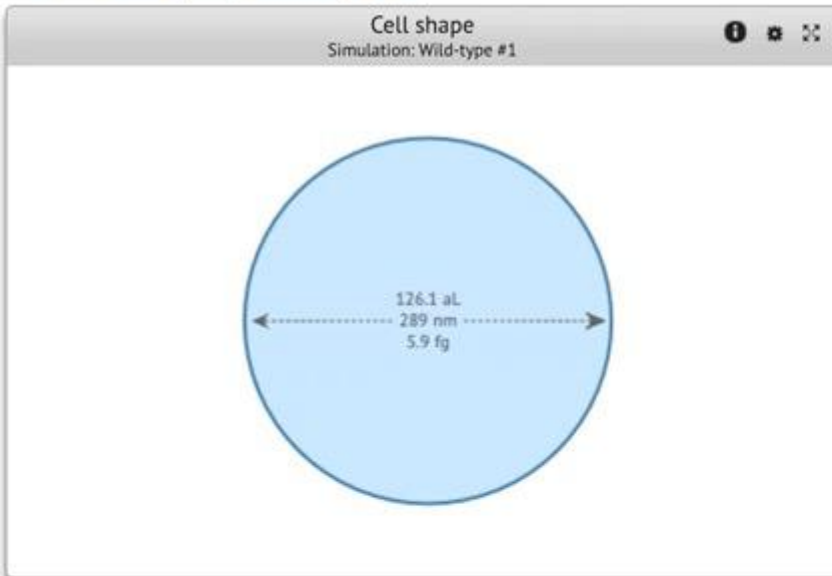


On obtient une modélisation de l'évolution de toutes les concentrations

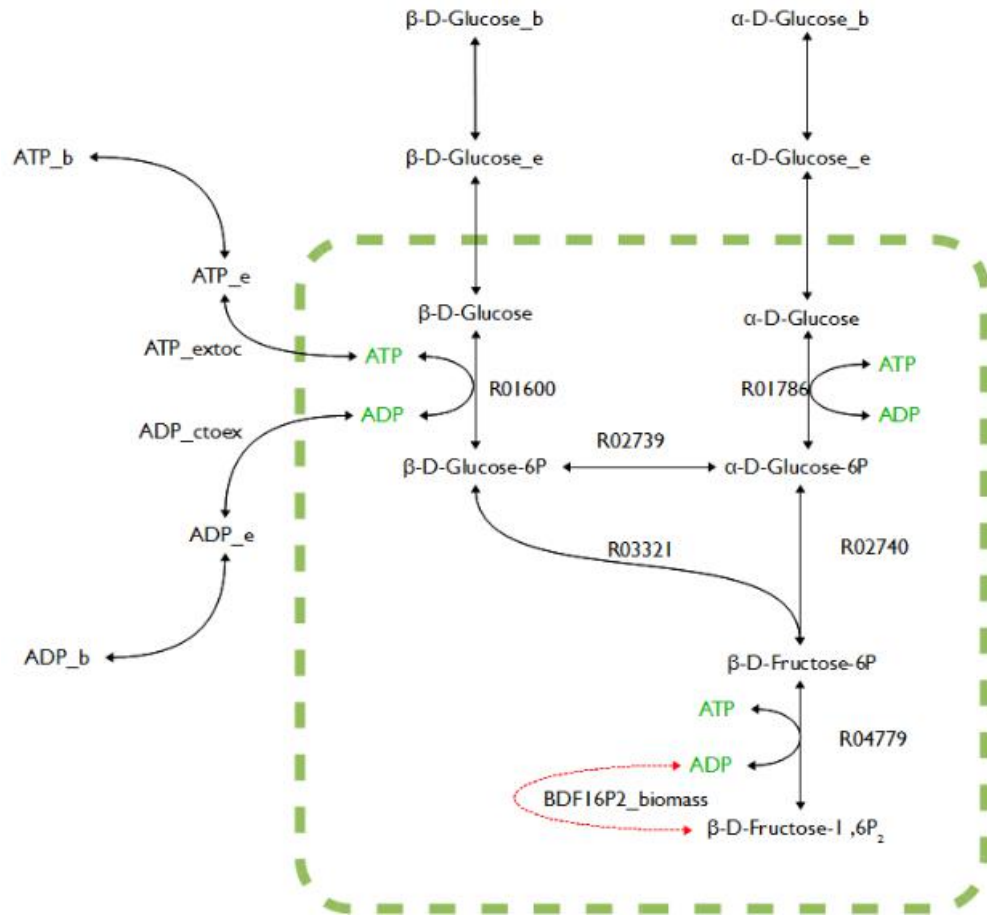


Heiske et al., FEBS Journal 2017

- Réglage fin de tous les paramètres du modèle
- **20 pages de description du modèle !**
- Le modèle reste « simple » ce n'est pas un organisme entier



Modéliser à une plus grande échelle



Un modèle =
ensemble
d'interaction
double ou multiple

Element A



Element B

On / Off

B si A

Flux A = Flux B

Variation de B
Dépend de A

Différentes façons de modéliser

- Modélisation de réseau booléen
- Modélisation par règles
- Modélisation par contraintes
- Modélisation par équations différentielles

Modélisation par contrainte

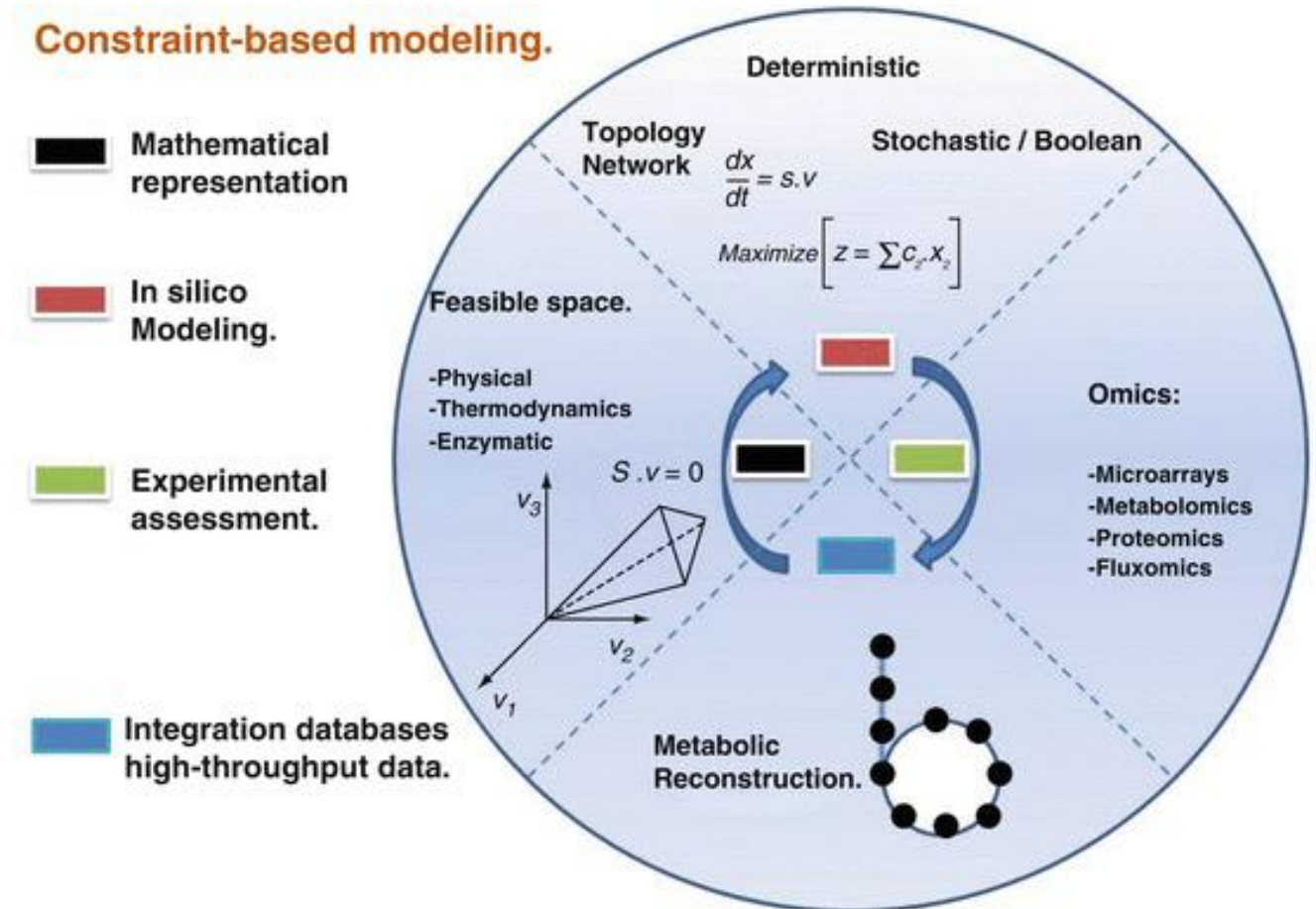
Il est trop couteux de modéliser l'évolution de tous les constituants du système



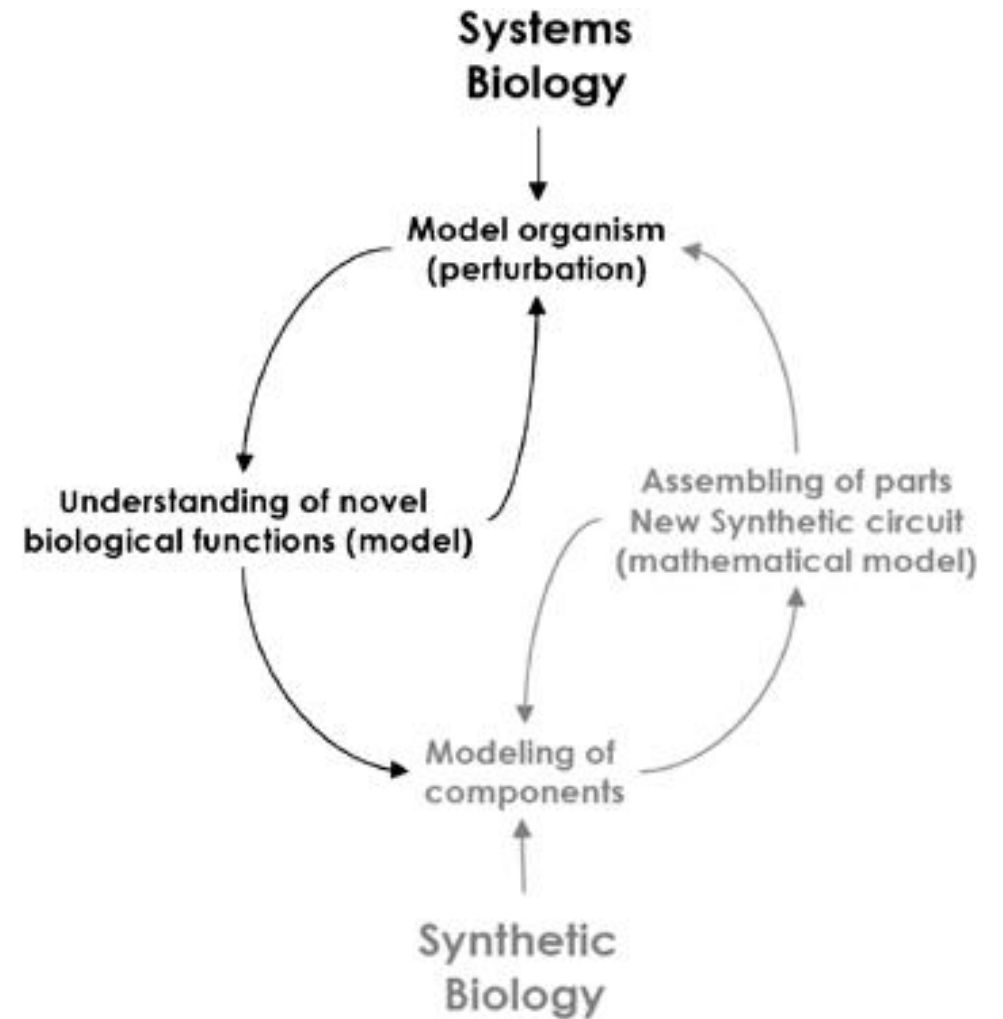
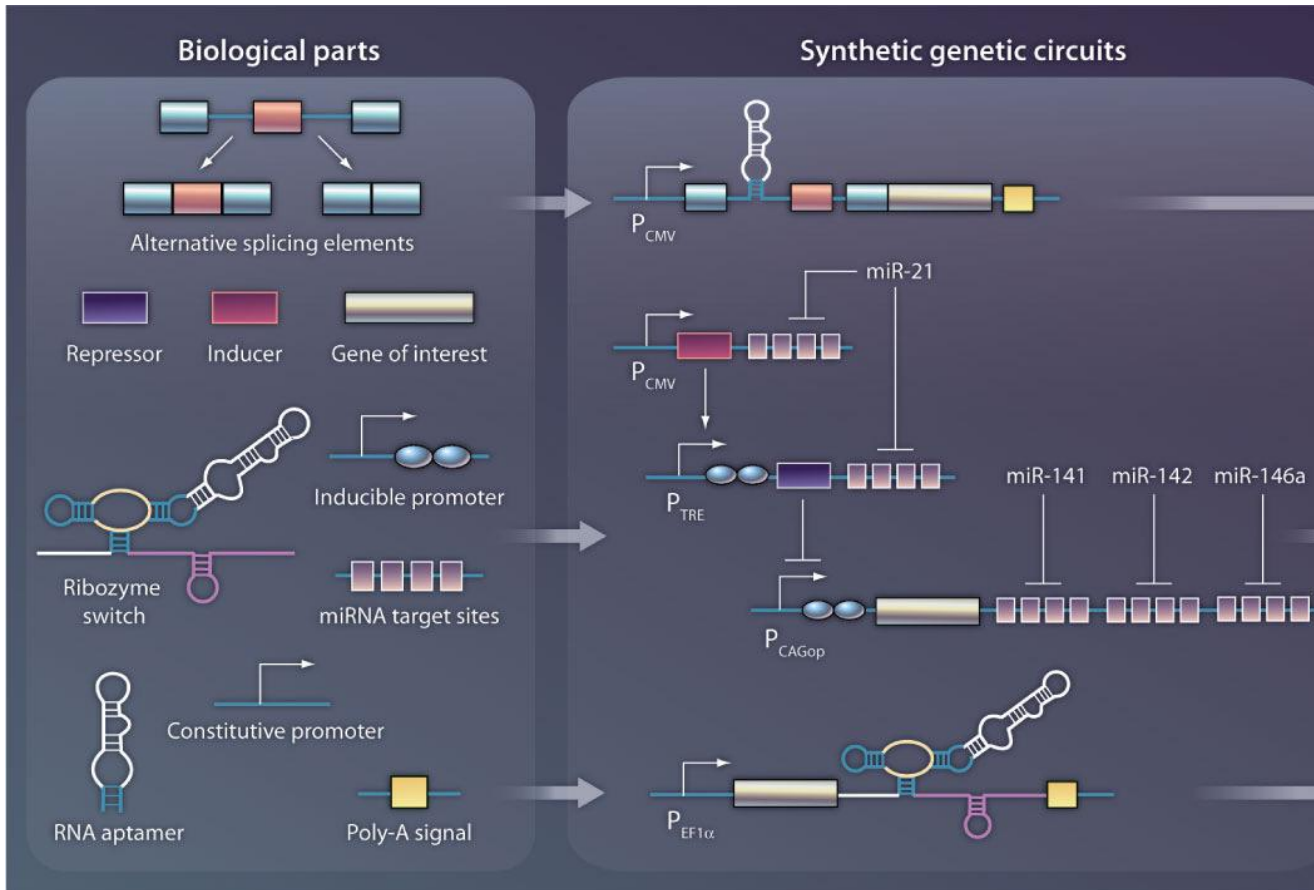
On impose une contrainte sur le système. Celle-ci va réduire la quantité d'état accessibles



La modélisation deviendra alors possible en un temps raisonnable

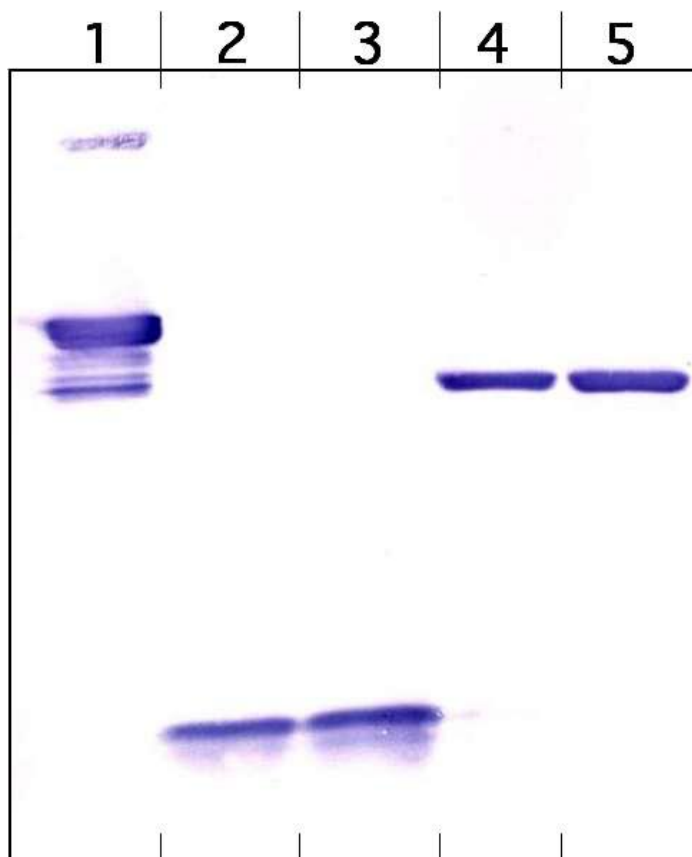


Modélisation et biologie synthétique

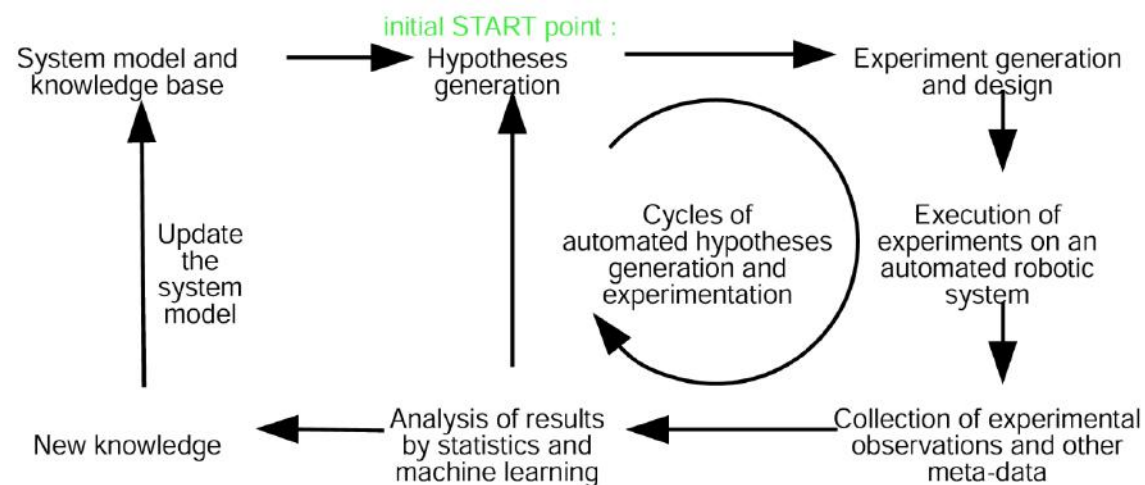


Validation des réseaux biologiques

Validation expérimentale



Validation automatisée



Ross D King, Science 2009

Un peu de lecture

E. Voit (2018) "A first course in systems biology" 2nd edition

U. Alon (2006)

"An introduction to systems biology : design principles of biological circuits"

B.O.Palsson (2015)

"Systems Biology: Constraint-based Reconstruction and Analysis" 2nd edition

M. W. Covert (2015)

"Fundamentals of Systems Biology: From Synthetic Circuits to Whole-cell Models"

A.-L. Barabási (2016)

"Network Science"