Biologie des systèmes

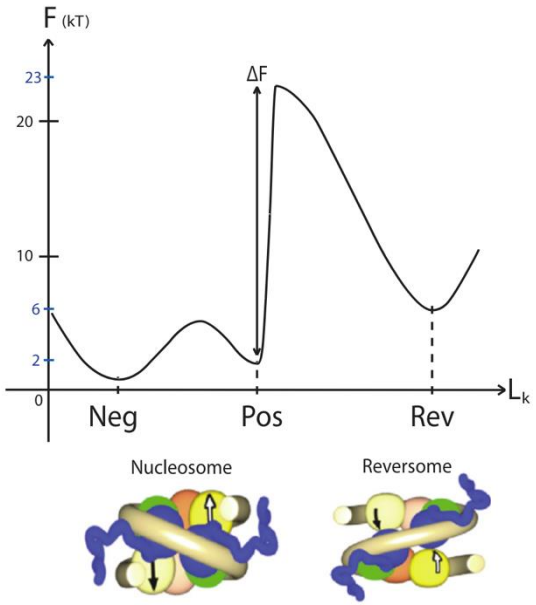
Biologie, Bio-informatique et IA

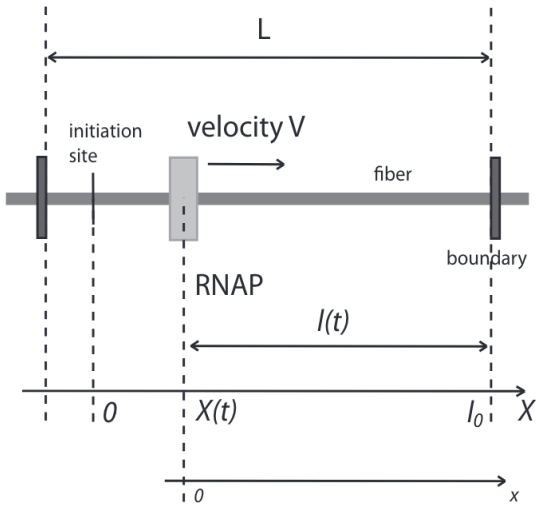
GB3 2024-2025

Christophe Bécavin

Maître de conférences UniCA

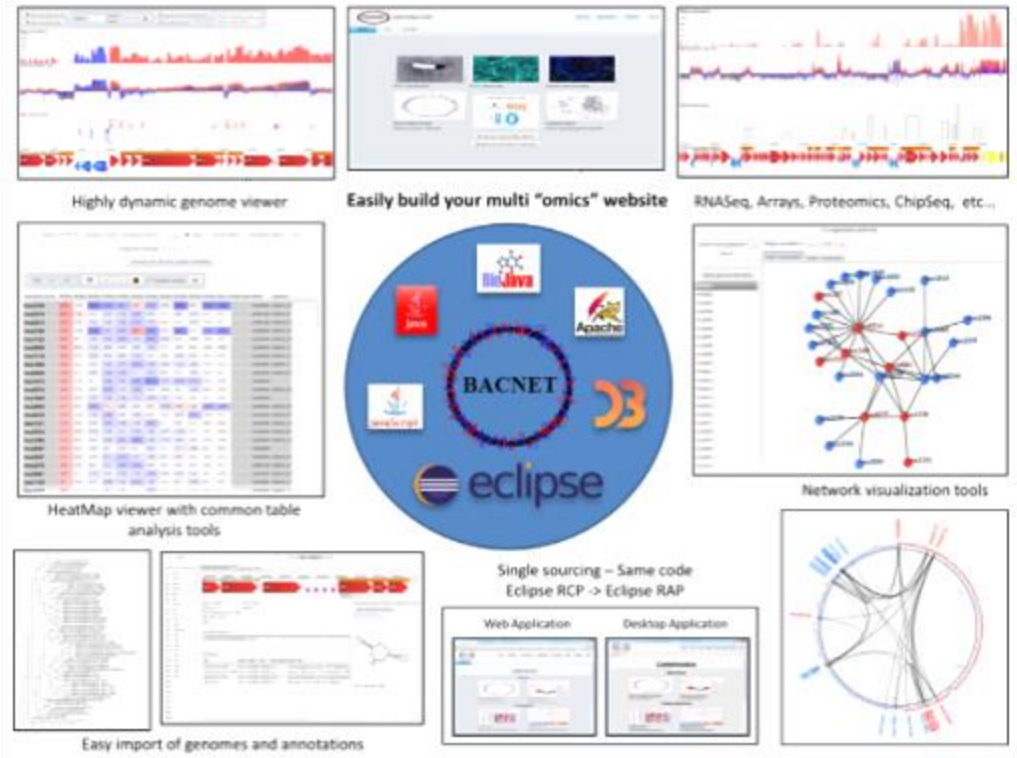
Développement de pipeline bioinformatique pour l'anayse de données multi-omiques





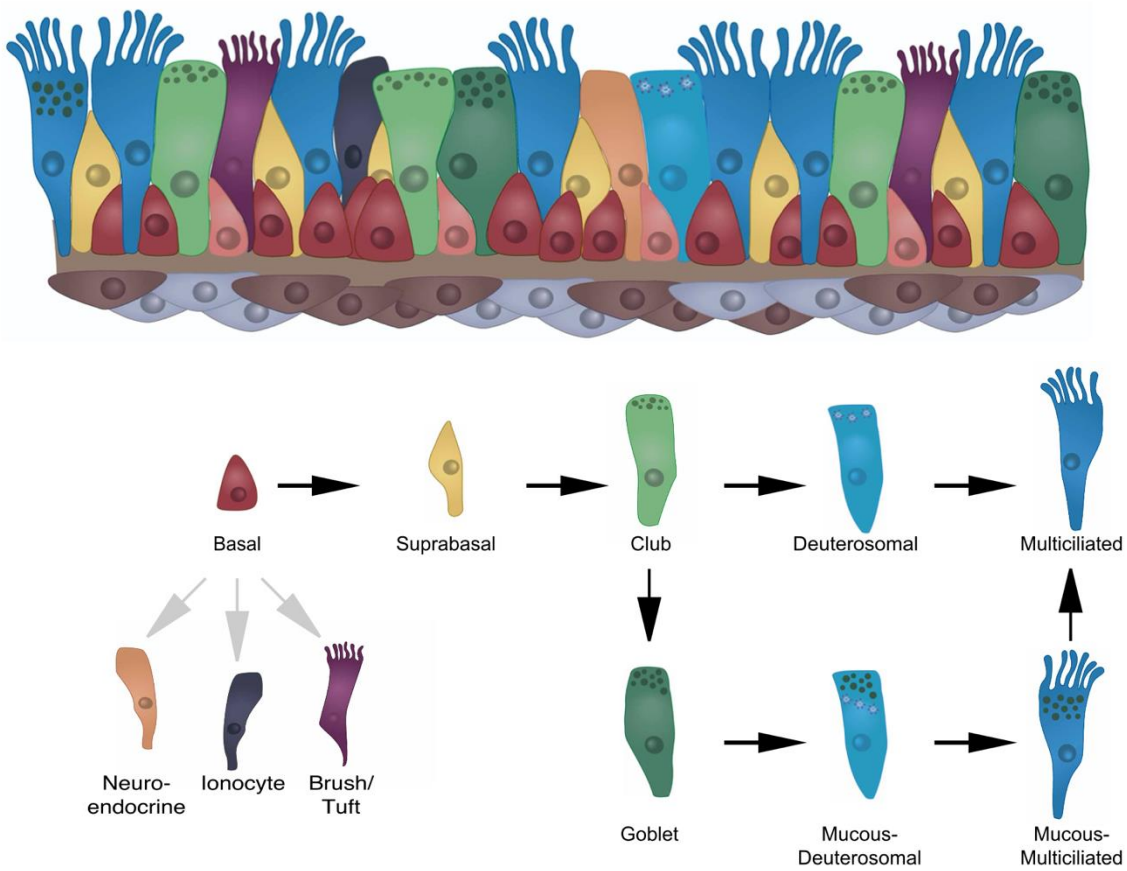
$$\frac{\partial \Theta}{\partial t} = D\frac{\partial^2 \Theta}{\partial x^2},$$

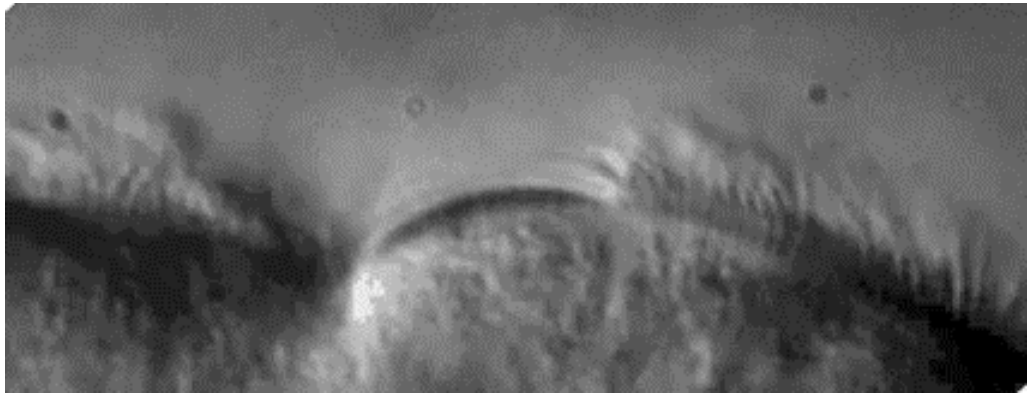
Becavin et al., BioPhysical Journal 2010



Bacnet: Une plateforme d'integration de données omiques (Danès et al. Bioinformatics 2021)

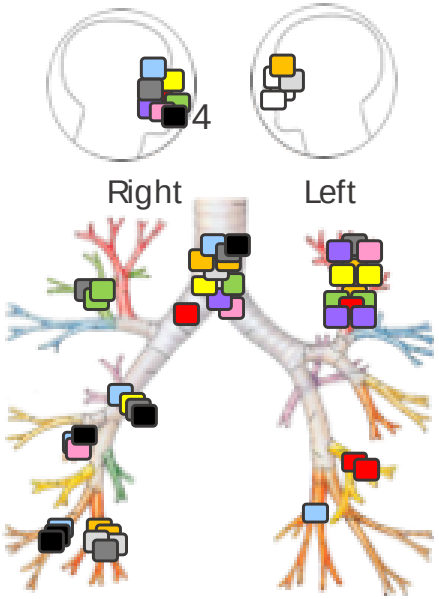
The human upper airway epithelium



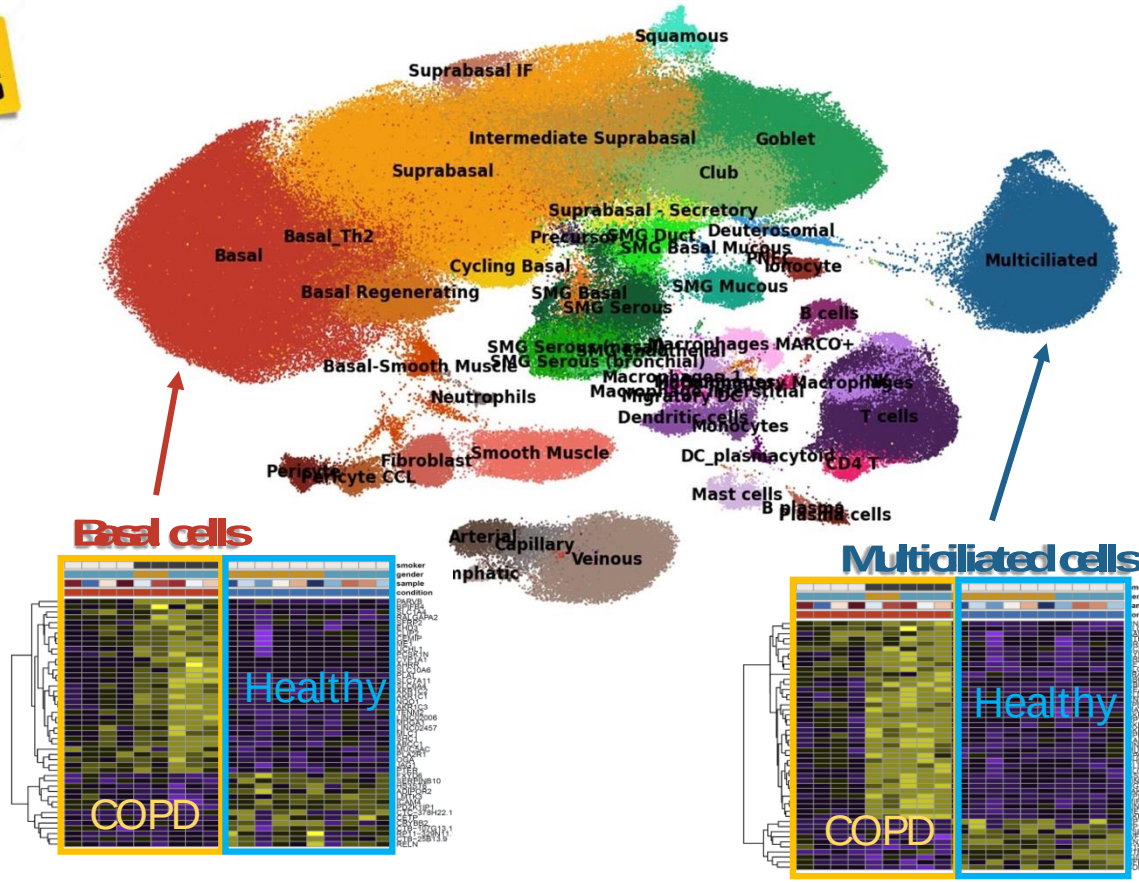


Human Lung Cell Atlas

HUMAN CELL ATLAS



(CHU of Nice)



Deprez et al, AJRCCM 2020
Wungnak et al., Nature Medecine 2020
Muus et al., Cell 2020
Sikkema, Nature Medecine (In Press)

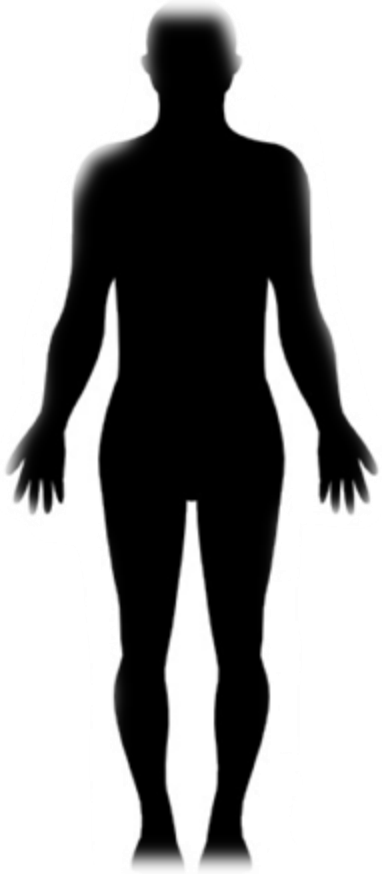
Reconstruction d'atlas cellulaire des voies respiratoires

Biologie des systèmes

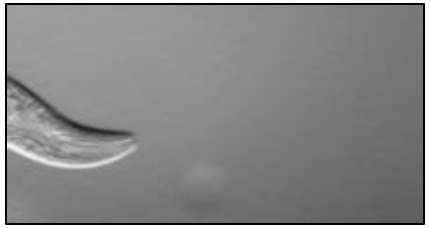
- **L’approche systémique en biologie**
- Bioinformatique et données omiques
- Reconstruire un réseau biologique
  - Les différents types de réseaux
  - Reconstruction de réseau biologique
  - Les obstacles à la reconstruction
  - Structure des réseaux biologiques
  - Etat de l’art des reseaux biologiques les plus étendus
- Virtual cell et digital twins

Du Phenotype au Genotype

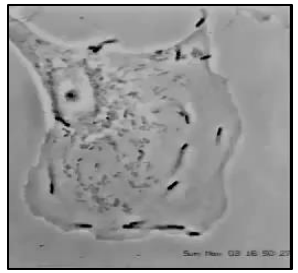
Comment caractériser un organisme vivant ?



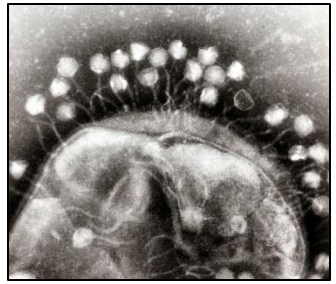
10^{-3} m



10^{-6} m

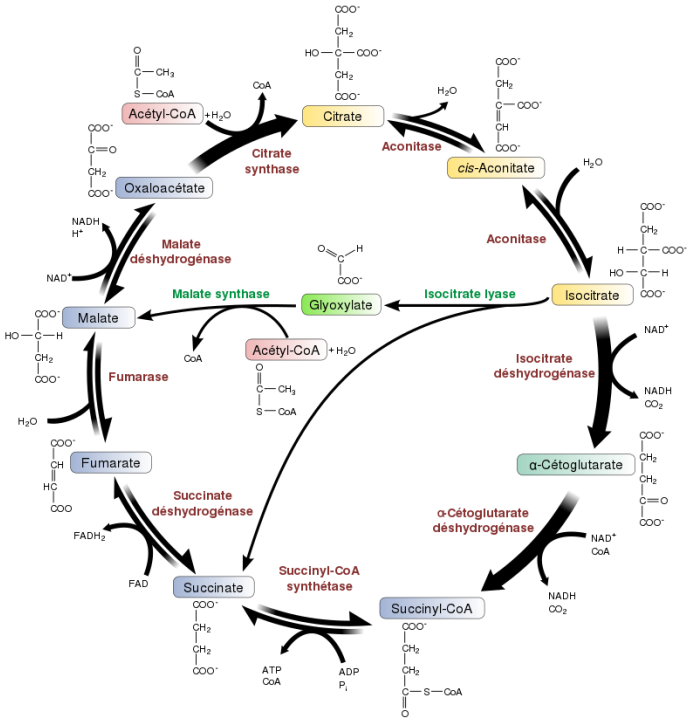


5.10^{-9} m



Le (les ?) phénotype

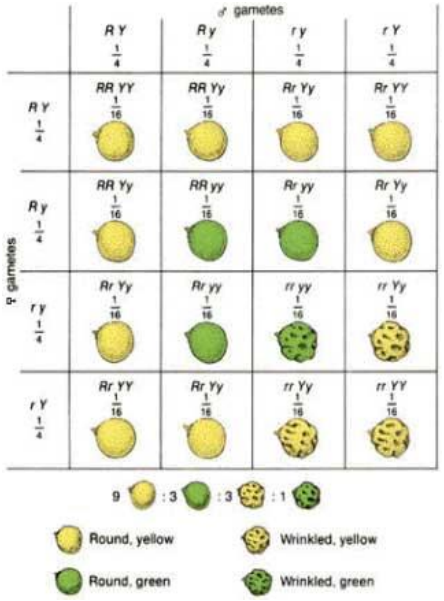
« Le phénotype c'est l'ensemble des traits observable » *Wikipedia*



Le gène comme élément héréditaire expliquant un phenotype

Gregor Mendel (1822-1884)

Le phénotype est héréditaire



Walter Sutton (1902)

Séparation des paires de chromosomes pendant la méiose

↓

Le génotype

Thomas Morgan (1910)

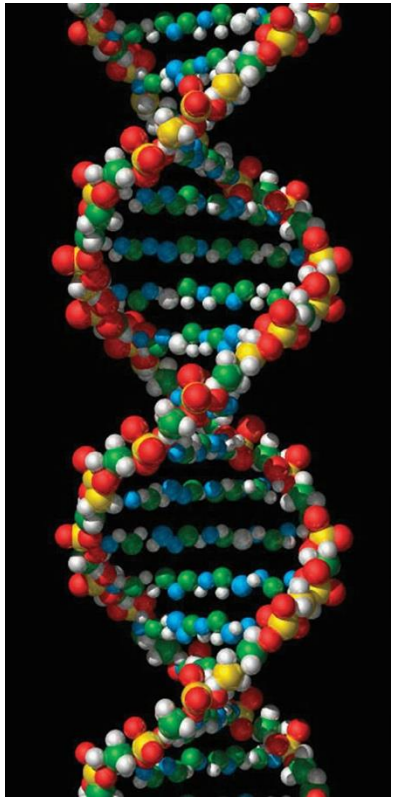
Théorie chromosomique de l'hérédité

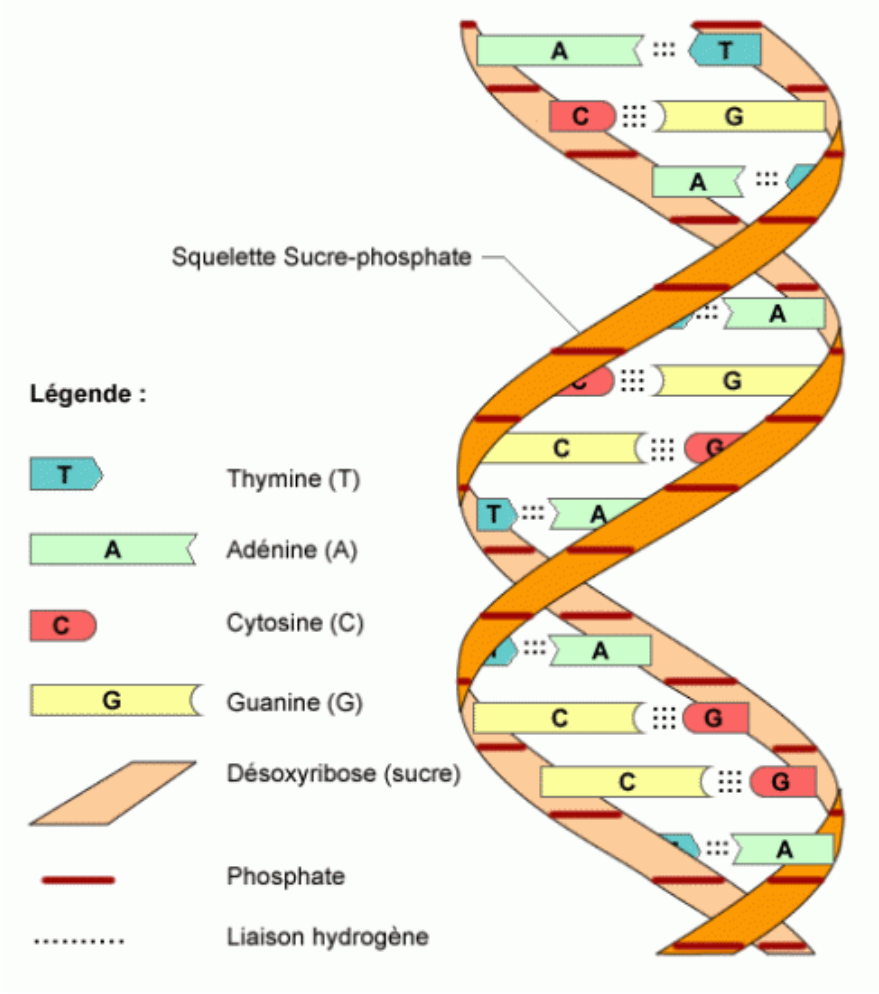
↓

Le génotype et le phénotype sont liés

L’ADN comme support DU GENOTYPE

Franklin, Watson, Crick (1953)





AGGGTCATTAAATATATATAAAGATCTATATAGAGATCTTTTTATTAGATCTACTATTAAGGAGCAGGATCTTTGTGGAT
AAGTGAAAAATGATCAACAAGATCATGCGATTCAGAAGGATCAGATCGTGTGATCAACCACTGATCTGTTCAAGGATTAG
CTGGGATCAAAAACCTATGTTATACACAGCCACCTTGGGATCTAAAACTTGTTATATGGATAACTATAGGAAGATCACCG
GATAATCGTATAGTTATCCACATGAGATTTGATTGAAAAAGCATCAATCAATTTTTTCACTACCGTTAAATTTATCCACA
ATCCAAAAAAAAAGAGCGGCATTAAGCCGCTCTGCATGGAATAGGTCATTATTTAGAAGCGATTGATGACGCGTTTGAGCC
AAGCTTCAGCGGCATCTTCAGGCACTGGGTGCTCTTGTACATCGATGGTAAAGCAGTTGGCCAGAGGTTTAGCACCAATA
TCCCCCAGCAGCTGATAGGCATGTTTACCTGCCGCGCAGAAAGTATCGTAGCTTGAATCACCAATCGCGACCACGGCATA
ACGTAGTGCAGAGGTATTCGGTGGTGTATTCTGCAGAGCCTGAATAAAGGGCTGGATATTATCCGGGTACTCACCAGCCC
CGTGGGTTGAGGTGATGATCAGCCAAGTCCCTTTAGCAGGGATCTCACTCATGTTGGGCTGGTTATGAATTTTGGTGTCA
AAGCCTTGTTCTTGCAGTAAATCACTCAGGTGGTCACCCACATATTCCGCACCGCCTAGGGTGCTGCCAGTAATGATATG
AATCATAGCGTTACTCTATTTCCCAATACAGAATGATGAAAAAATGCGGCCAAGCAGATCATCGGAGCTGAACTCGCCCG
TAATTTCGTTAAGGTGTTGCTGGGCTATACGCAGCTCTTCGGCGAGGATTTCTCCGGCCATATAGCCTTCAAGTTGTTGC
TGGCCAATCGCTAAGTGCTCTGCGGCTCGCTCTAGGGCATCGAGATGACGGCGGCGTGCCATAAAGCCACCTTCCTGATT
GCCTGAAAAACCCATGCACTCTTTGAGGTGCTGACGCAAGGCATCGACCCCTTGGCCTGTTTTGGCTGATAGGCGGATCA
AGGTGGGTTGATTAACATGGCAGATCCCAAGGGGCTCACCAGTTTGATCGGCTTTATTACGGATCACAGTGATCCCAATA
TTCTCTGGCAGTTTGTCAACAAAATCAGGCCAGATGTCCTGTGGATCGGTGGCCTCTGTGGTGGTGCCATCGACCATAAA
CAGTACGCGATCGGCTTGGCGGATCTCTTCCCATGCGCGCTCAATACCAATTTTTTCTACCGCATCAGAAGCGTCTCGTA
GTCCCGCAGTATCGATGATGTGCAGCGGCATCCCATCAATATGGATATGCTCACGCAGAACATCACGGGTGGTACCGGCA
ATGTCGGTAACGATGGCAGACTCTTTACCTGAAAGCGCATTGAGTAGGCTCGATTTACCCGCATTAGGACGCCCAGCAAT
CACCACCTTCATCCCTTCGCGCATAATGGCGCCTTGGTTGGCTTCACGGCGCACTGCGGCAAGATTATCTATGATGGTTT
GCAGATCAGCGGAAACCTTACCATCGGCCAGAAAATCGATCTCTTCTTCTGGGAAATCAATTGCGGCTTCAACATAGATG
CGCAGGTGAATCAGCGATTCCACCAAGGTATGGATGCGTTTAGAAAACTCGCCTTGCAGTGATTGCAGCGCGGATTTCGC
GGCTTGCTCAGAGCTGGCATCAATCAGGTCTGCGATGGCTTCCGCTTGGGTTAAATCCATCTTGTCATTGAGGAAAGCGC
GTTCTGAGAATTCACCGGGACGGGCTGGGCGCACTCCTTTAATCTGCAAAATACGGCGGATCAGCATATCCATGACGACC
GGGCCACCGTGACCTTGCAGCTCAAGCACATCTTCACCGGTAAATGAATGAGGATTGGGGAAAAACAGCGCAATGCCTTG
ATCGAGCTGTTGGCCATCTTCATCGGTGAAAGGCAAGTATTCGGCATAGCGGGGTCTGAGCGTGCGTCCAGTGACTGTCT
GCGCGACGTGGGCAGCCAGTGGGCCTGATACACGAATAATGCCGACACCACCACGGCCGGGTGCCGTAGCTTGCGCGACA
ATGGTATCTGTTGTCATAGTGTTACCTGAACAGGATTGAATTAGCGCCATTGTAATCAGCAGCCAACAAAAAGGCGACCT
TTTGGCCGCCTCTTTATTACTCAATCAAACTTACTTGGAGTGTAAGCCTTTTTTCTCTAGCGCTTTGTAGATCAGCGTTT
GCTGGATTAGCGTTACGATGTTCGACACCAACCAGTACAGAACCAGACCTGATGGGAACCACAGGAAGAAGAAAGTGAAC
ATGACTGGCATGAAGGTCATGATCTTCTGTTGCATTGGATCGGTGATCGTGGTCGGGCTCATCTTCTGGATAACGAACAT
GCTCGCACCCATCAGCAGTGGCAAGATGTAGTAAGGATCTTGCGCGGACAAGTCATGAATCCAACCAAAGAATGGAGAGT
GACGCAGCTCTACCGACTCCATCAGTGCCCAGTACAGCGCAATGAAAATCGGCATTTGCAGAAGGATAGGGAGACAGCCA
CCCAGTGGGTTTACTTTCTCTTTCTTATACAGCTCCATCATCTCTTGGCTCATGCGCTGGCGATCATCGCCAATCCGCTC
ACGCATCGCTTGCAGTTTAGGTTGCAGCATGCGCATTTTCGCCATTGAGGTGTACTGTGCTTTGGTCAGTGGGTACATCG
CACCACGAACAATAAAGGTTAAGCAGATGATGGCCACACCCCAGTTGCCAACAAAGGTCTGAATCACAGACAGCAACCAG
TGCAGTGGCTTAGCAATGAACCACAACCAACCATAATCAACCACCAGATCAAGATTAGGGGCTGTTGCGGCCATTTGGTC
TTGCAGTTTTGGGCCTACCCATAAGGTTGCTTCAAACTCTGCTTTATCACCGTTGGCAATGGTTTTGTTTGGTGTACGGA
TACCCATATCACCTAAATTATTGATAACGCGTGAGTAAAGCTGTGCTTGCGGCTCATCACGTGGGATCCATGCTGAAGCG
AAGTAGTGCTGGATCATCGCGACCCAGTTTACCGTCACATCGTTTGGTGCACTGAGGTTGCGATCTTTCATGTCGTCGAA
GCTGTACTTCTTGTAACGCACATCGCTGGTTGAGTAAGCACCACCACGGTAAGTTGGCATGGCGAGGTTGCCACCAGAAT

Du genotype au PHENOTYPE PAR la regulation genetique

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

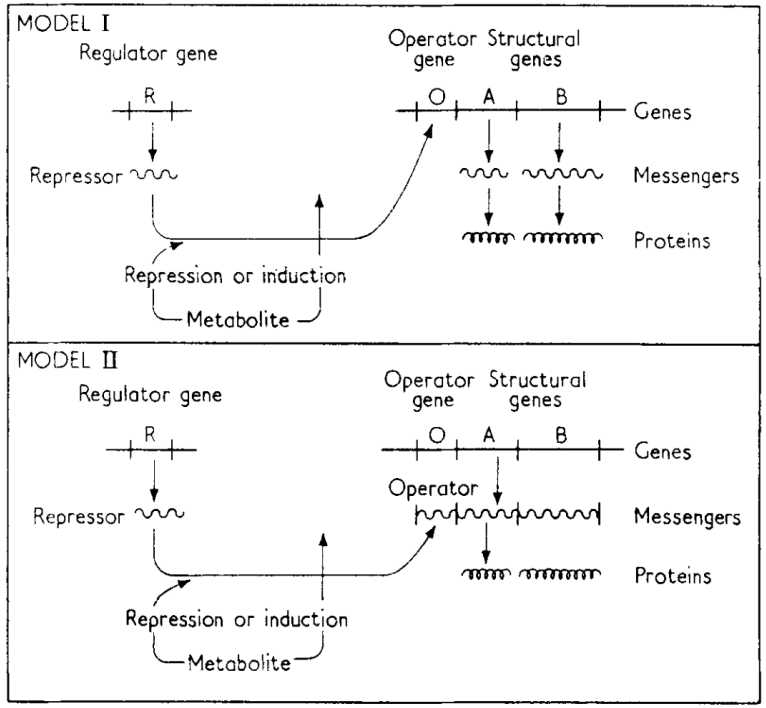
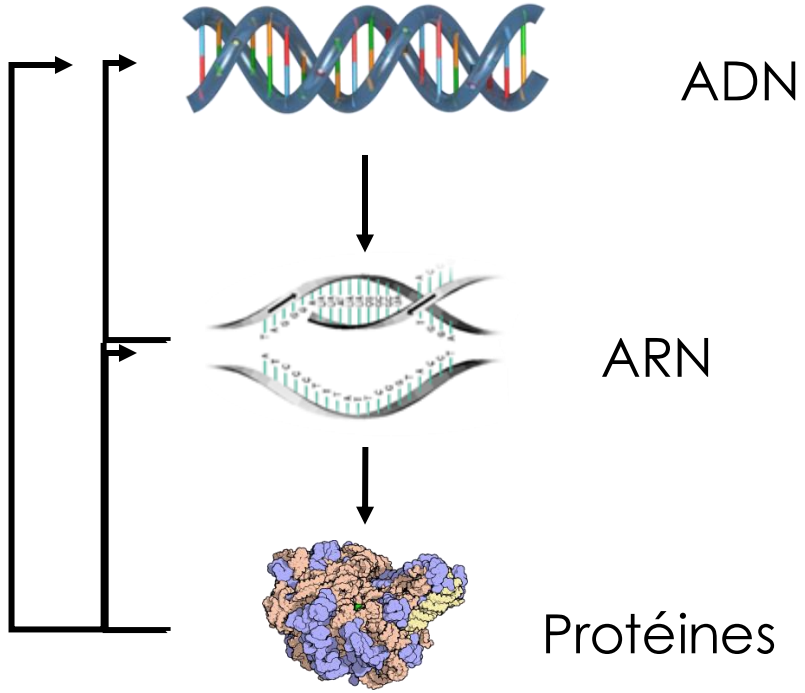


FIG. 6. Models of the regulation of protein synthesis.



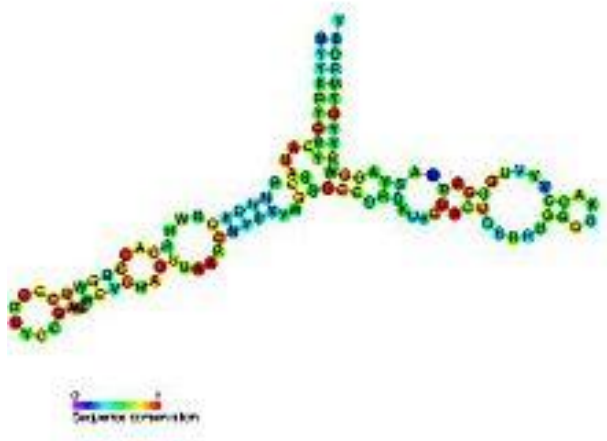
Décoder le Genotype

Frederick Sanger (1918 – 2013)

Protéine

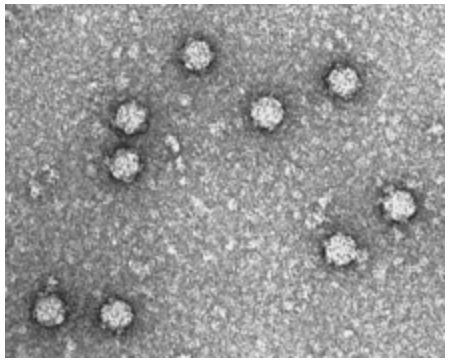
Insuline de bovin
1955

ARN



5S rRNA
120 nucleotides
1965

ADN



E. Virus ΦX174
5386 nucléotides
1977

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

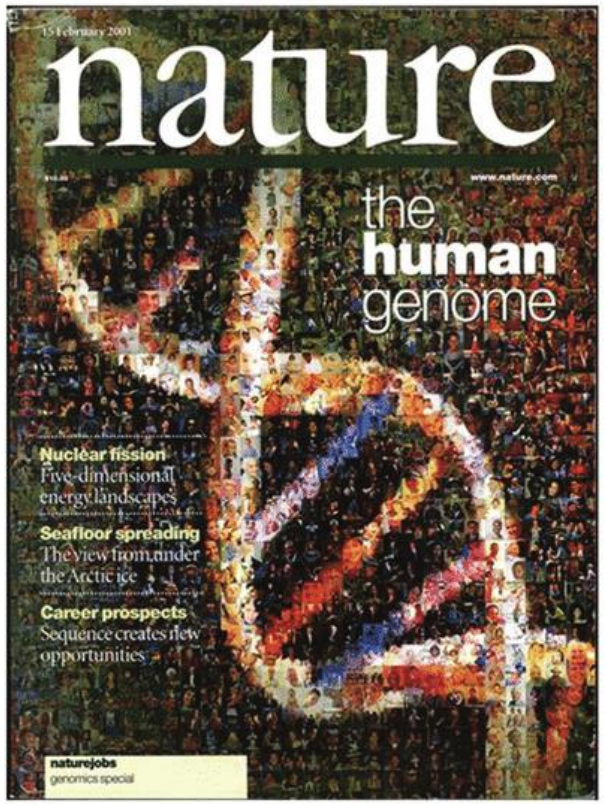
Le séquençage du génome humain

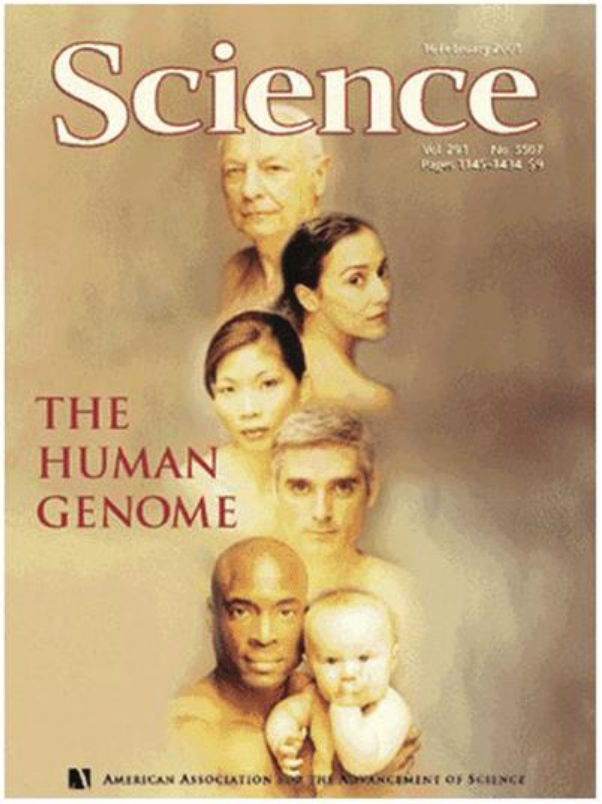
Human Genome Project

Lancé en 1985
par le gouvernement
américain

1989-2001

5 milliards de dollars
20 Instituts
6 pays





Février 2001

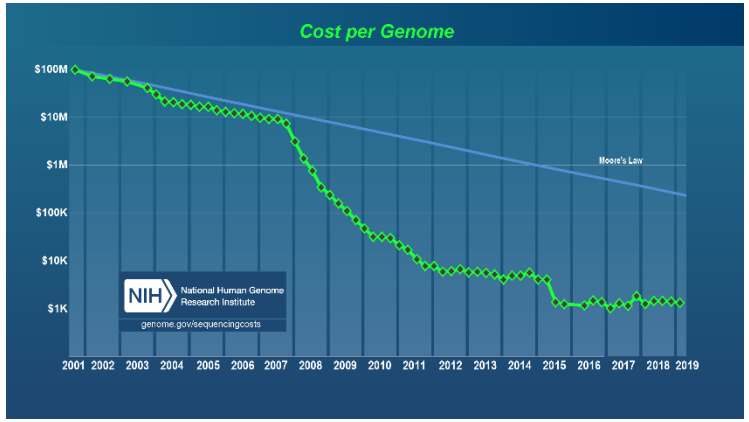
Celera Genomics
(J. Craig Venter)

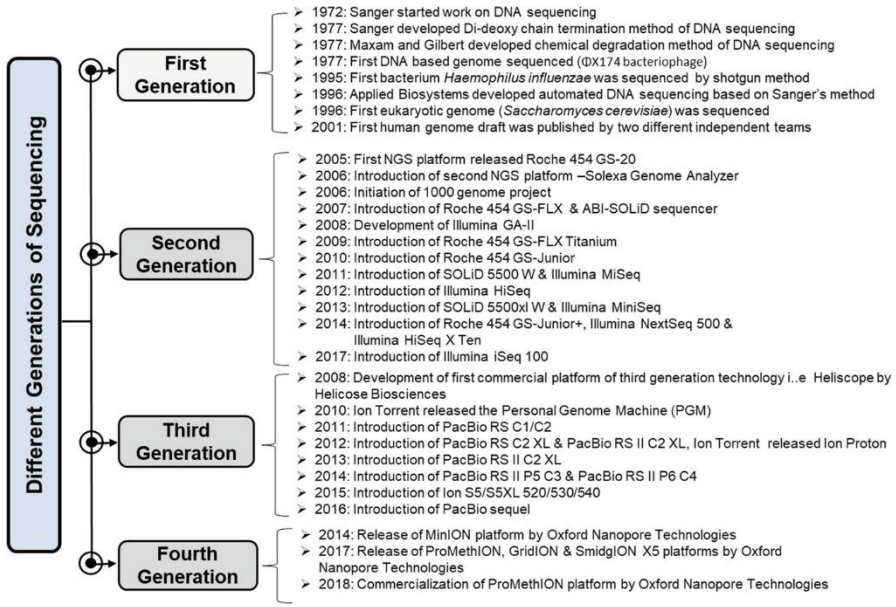
1998-2001

300 millions dollars
(Shotgun sequencing)

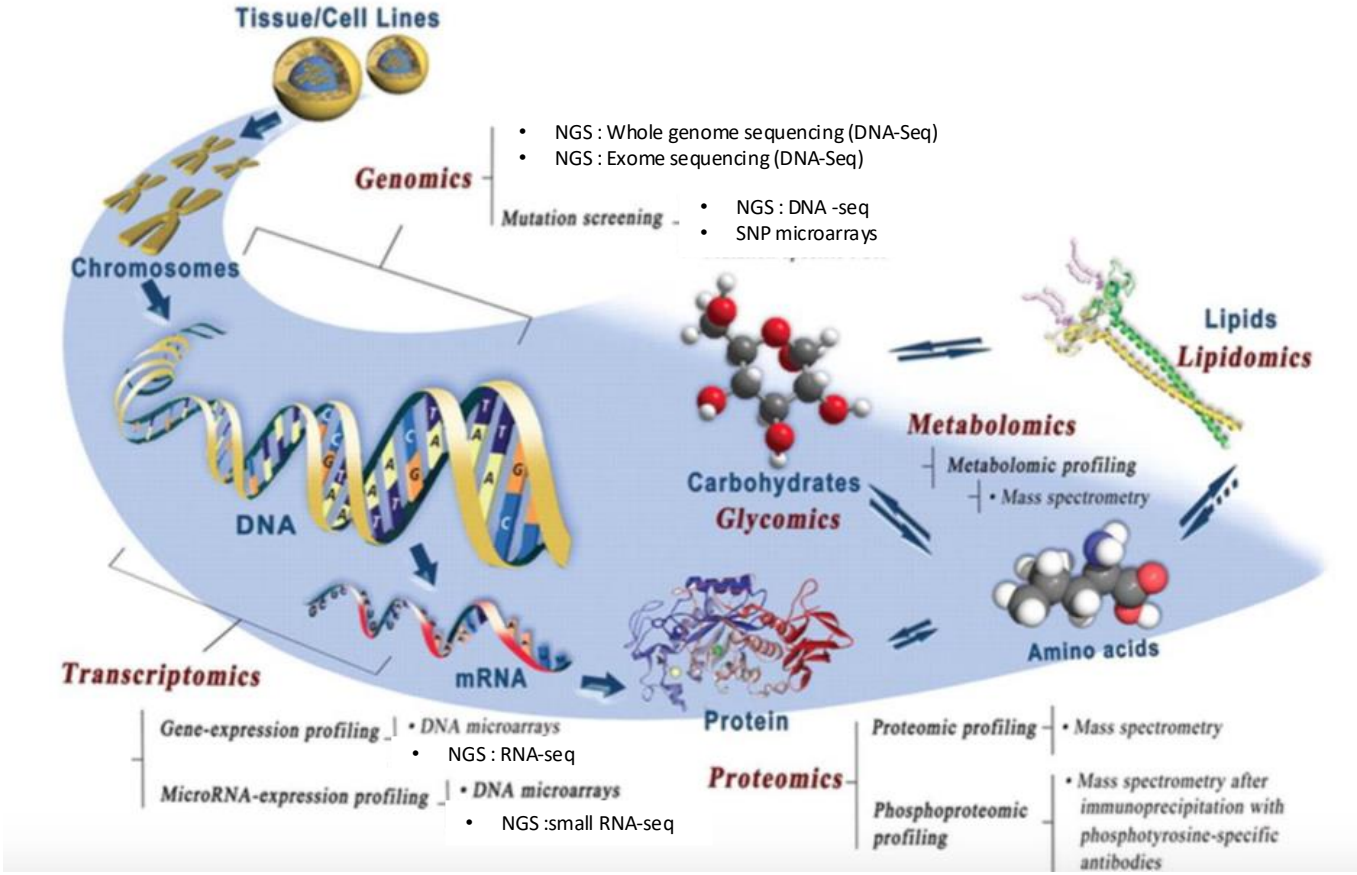
Du Phenotype au genotype

La révolution omique



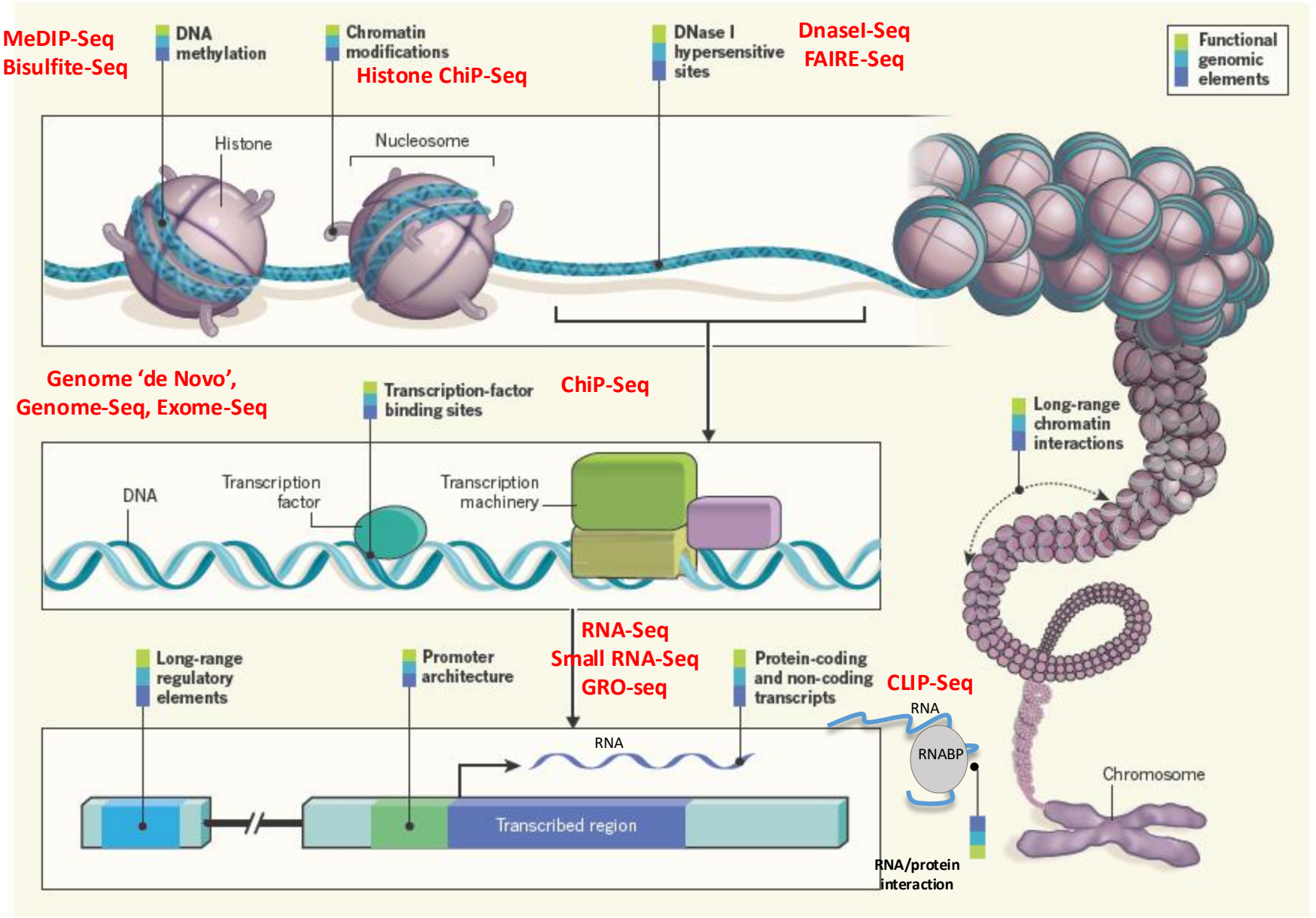


Microbial Technology for the Welfare of Society, 2019, Springer



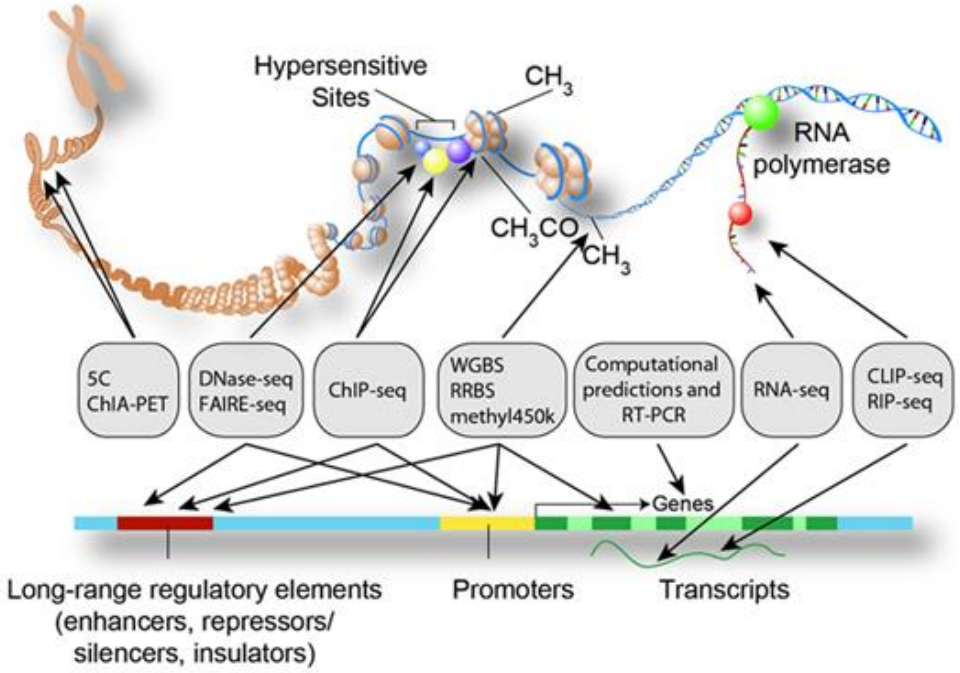
Wu R.Q., J. dent. Research, 2010

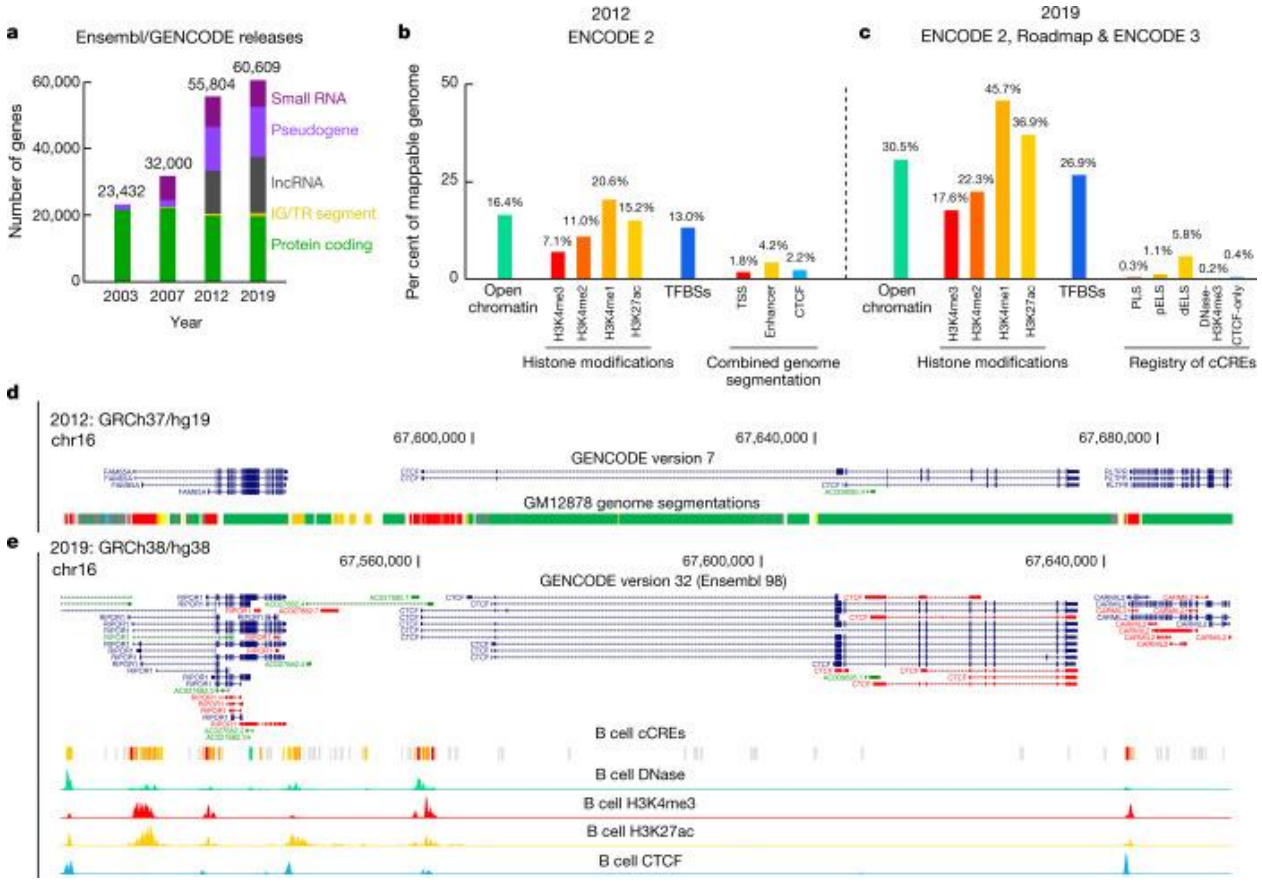
Les différentes échelles omiques



Le Projet ENCODE

Encyclopedia of DNA Element
2003 - Présent





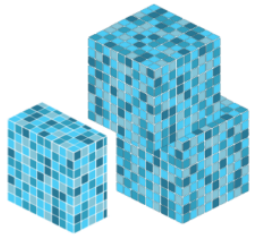
Big Science
TCGA – Atlas des tissus cancereux

2005 - Présent

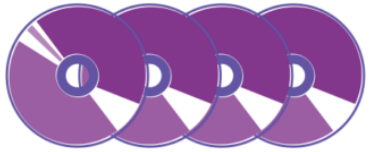
NATIONAL CANCER INSTITUTE
THE CANCER GENOME ATLAS

TCGA BY THE NUMBERS

TCGA produced over **2.5** PETABYTES of data



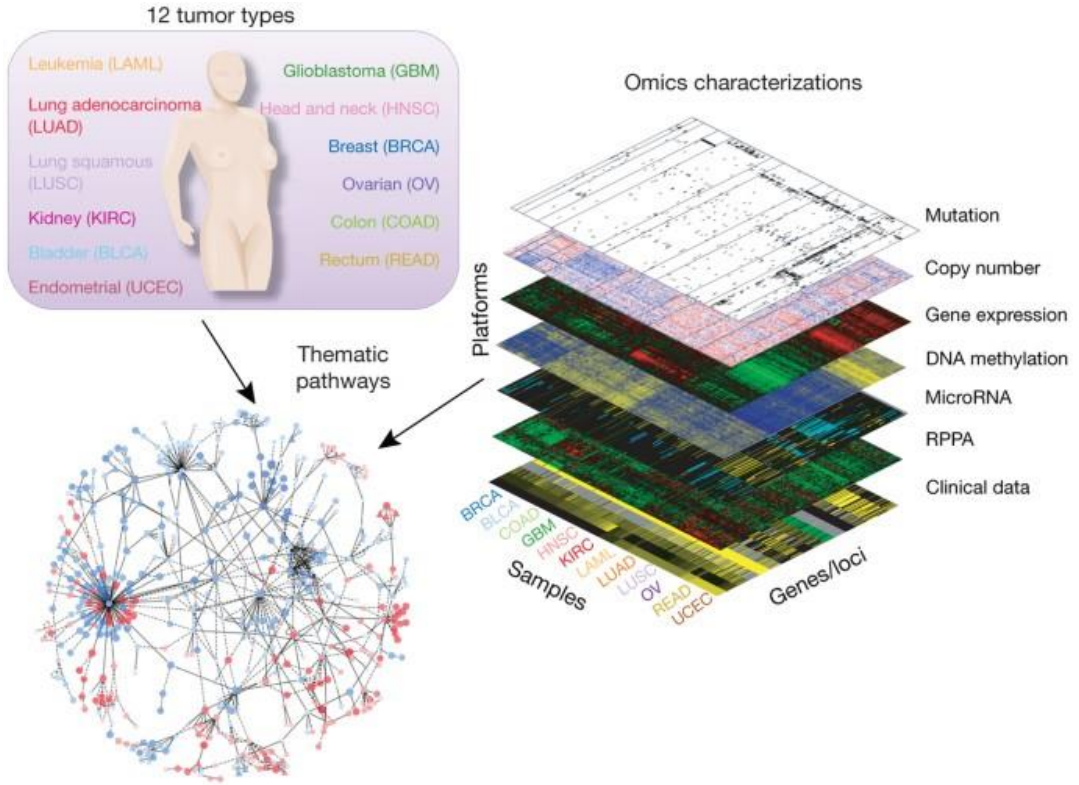
To put this into perspective, **1 petabyte** of data is equal to **212,000** DVDs



TCGA data describes **33** DIFFERENT TUMOR TYPES ...including **10** RARE CANCERS

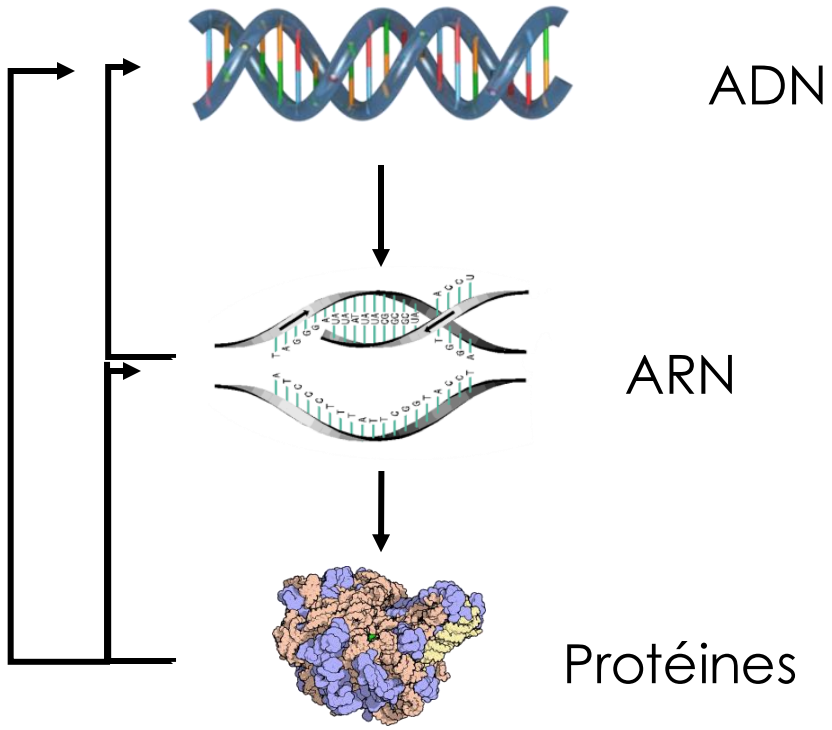
...based on paired tumor and normal tissue sets collected from **11,000** PATIENTS

...using **7** DIFFERENT DATA TYPES



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-infinis

La biologie des systèmes

LacZ operon version 2021

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

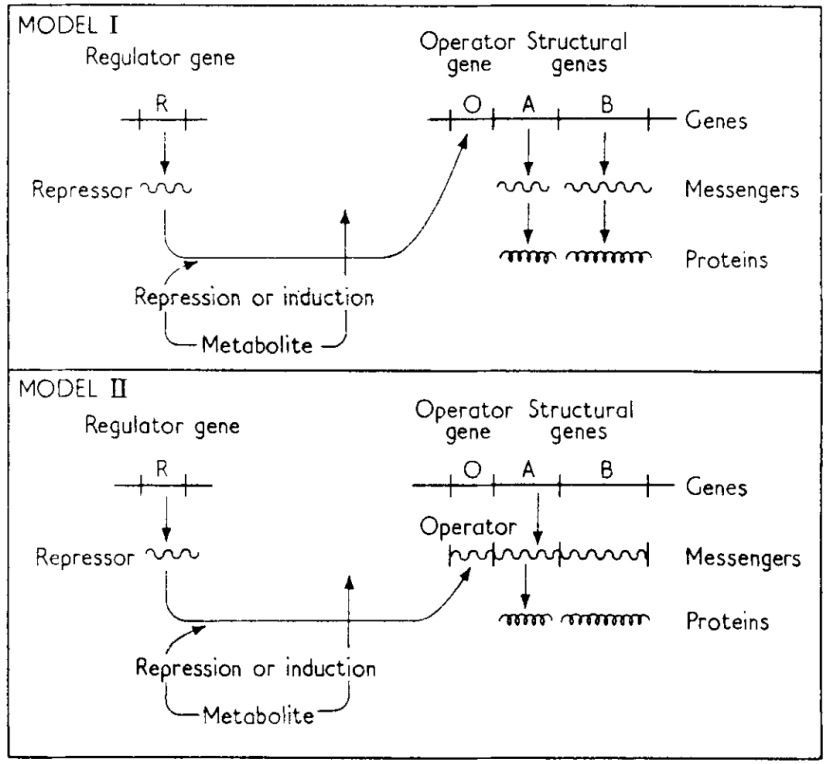
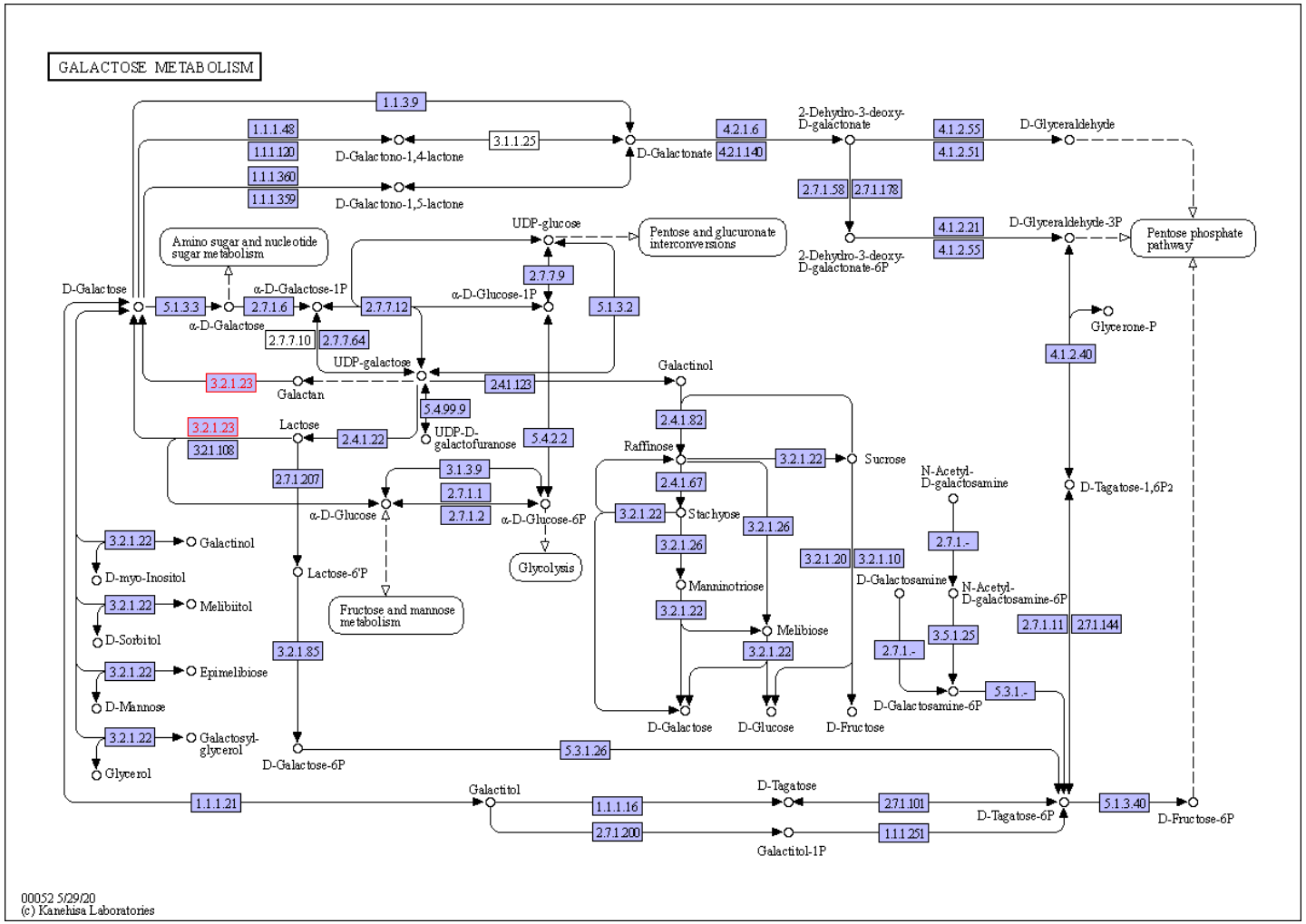
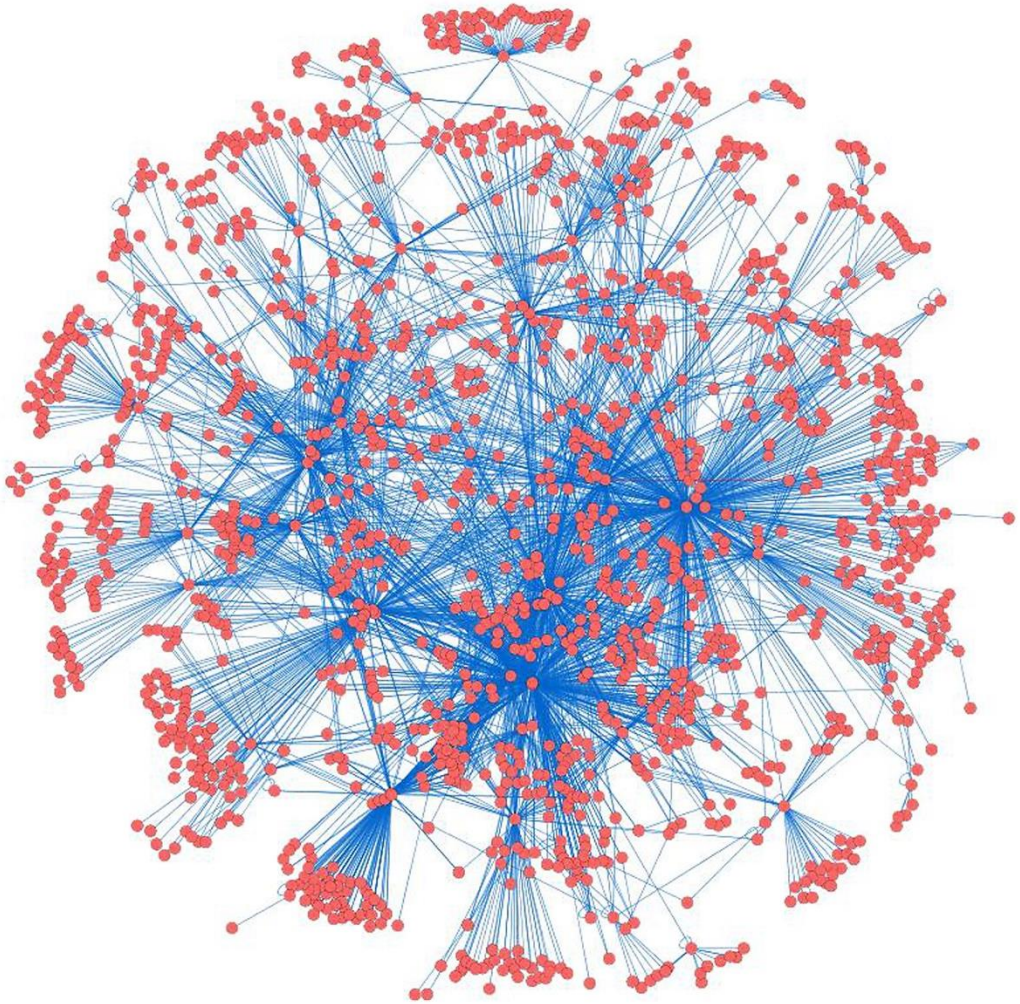


FIG. 6. Models of the regulation of protein synthesis.

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



Réseau de régulation génétique chez *Escherischia Coli*



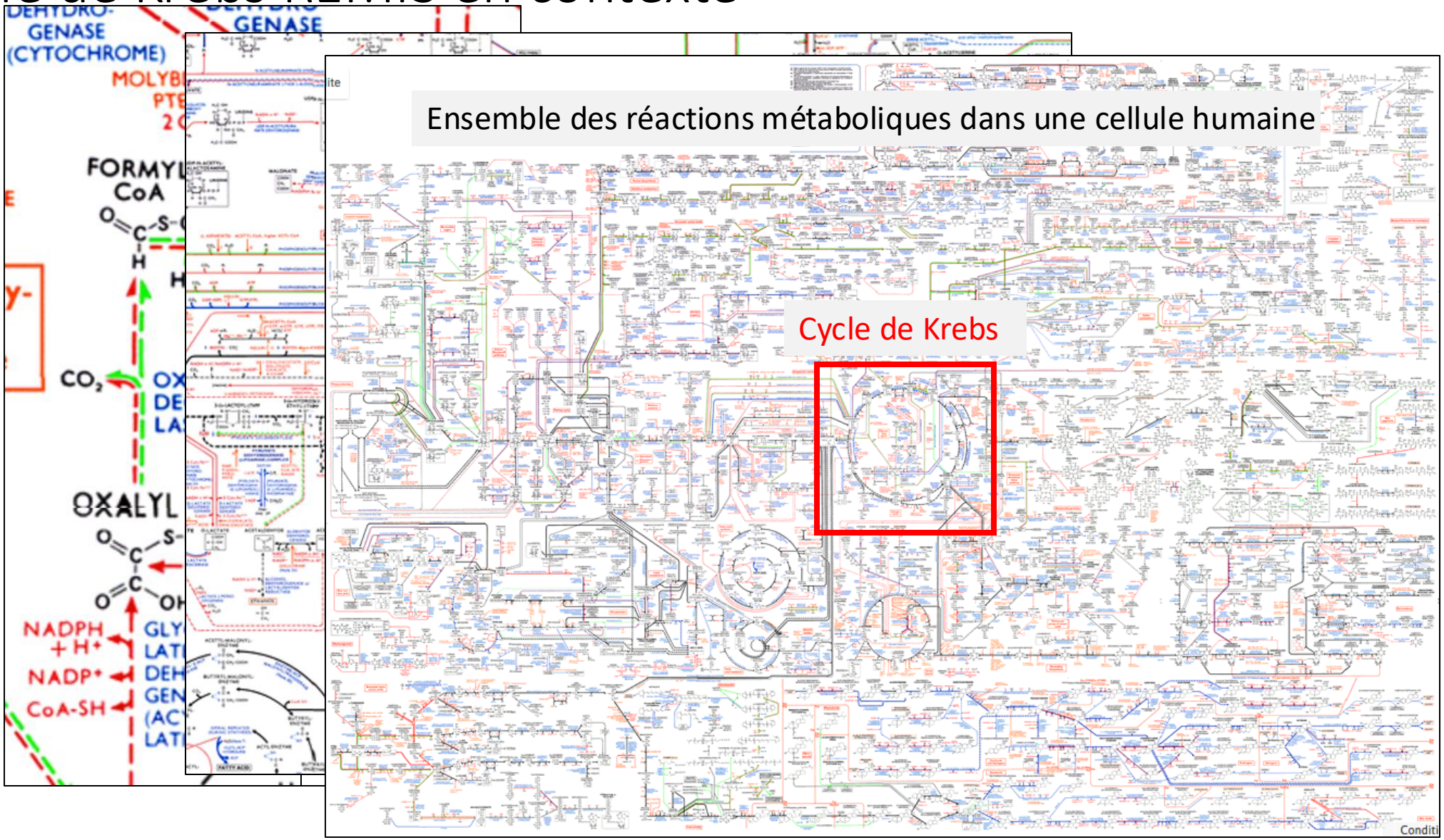
Chaque point est un gène
Chaque ligne est un lien de régulation entre gène

Il y a 4292 gènes

Le réseau a été **reconstruit** en utilisant :
524 puces ADN dans 264 conditions expérimentales

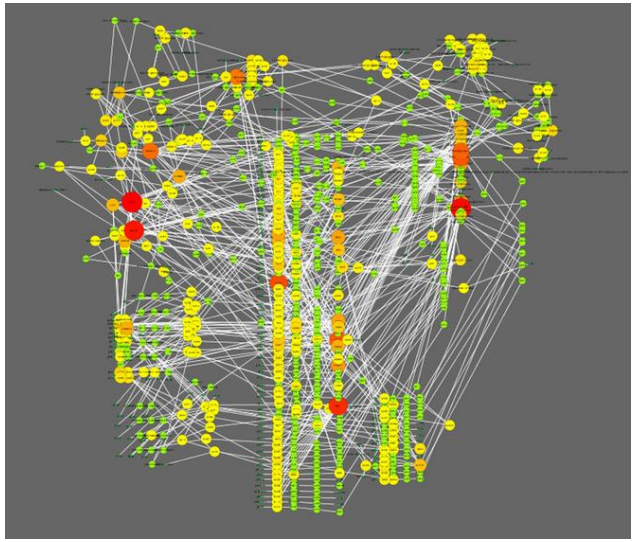
Allen et al., PlosOne, Jan 2012

Le cycle de Krebs REMIS en contexte

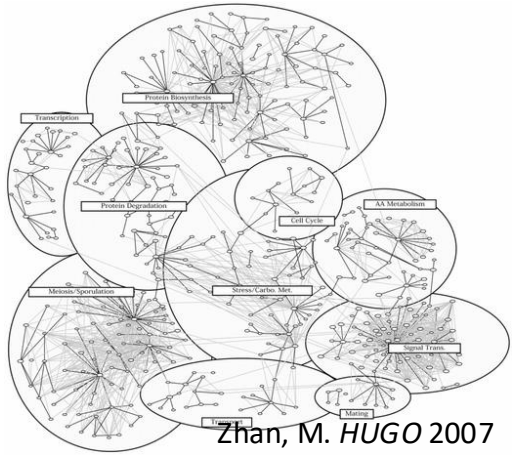


La biologie des systèmes

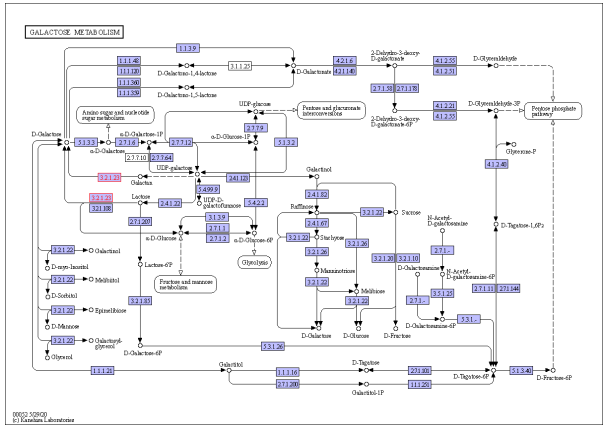
Cancer gene network – Institut Curie

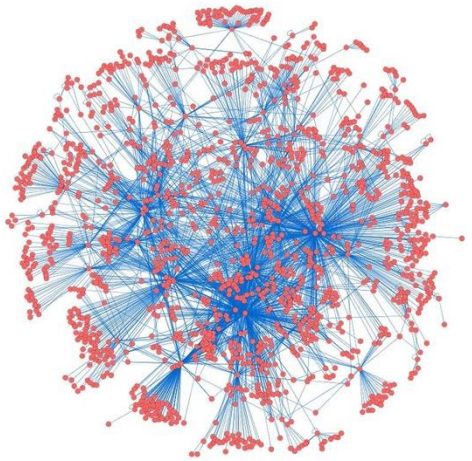


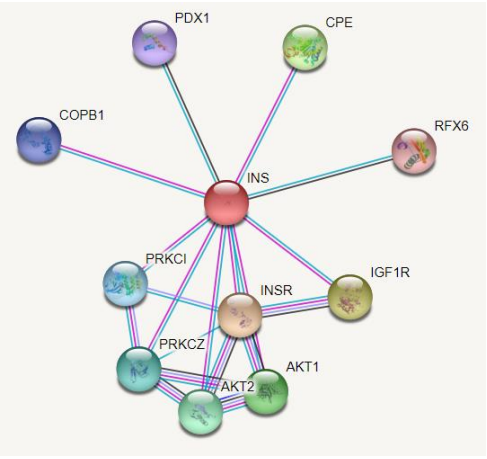
- Etudier les processus biologiques avec une approche de réseau
- Une approche « Top-down »
- Le tout est plus que la somme de ses parties



Zhan, M. *HUGO* 2007





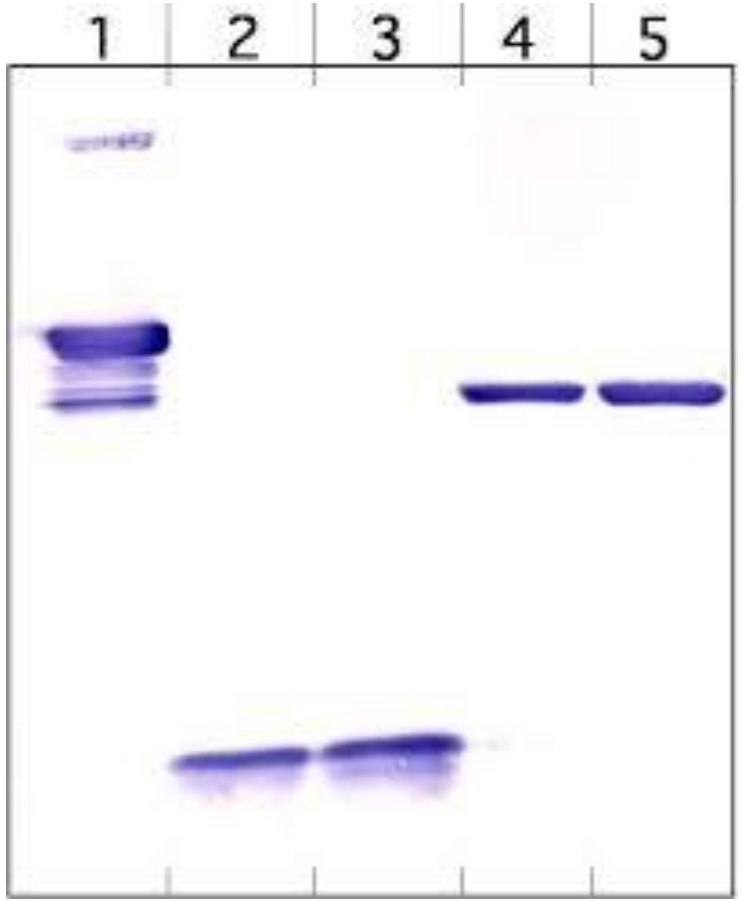


Biologie des systèmes

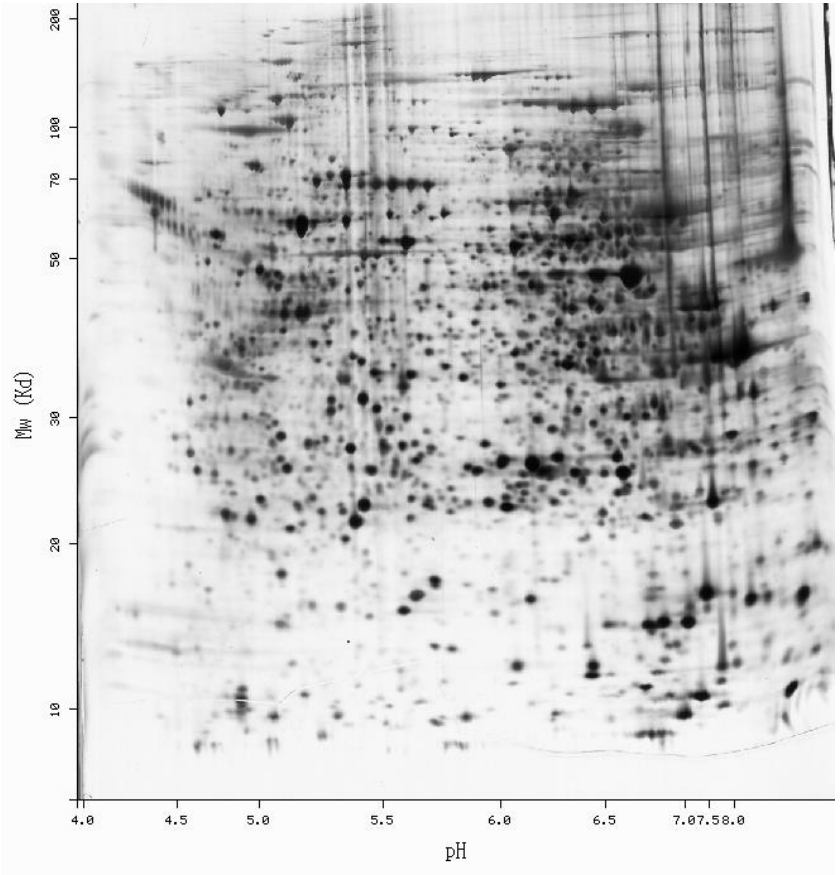
- L’approche systémique en biologie
- **Bioinformatique et données omiques**
- Reconstruire un réseau biologique
  - Les différents types de réseaux
  - Reconstruction de réseau biologique
  - Les obstacles à la reconstruction
  - Structure des réseaux biologiques
  - Etat de l’art des reseaux biologiques les plus étendus
- Virtual cell et digital twins

Evolution de la protéomique

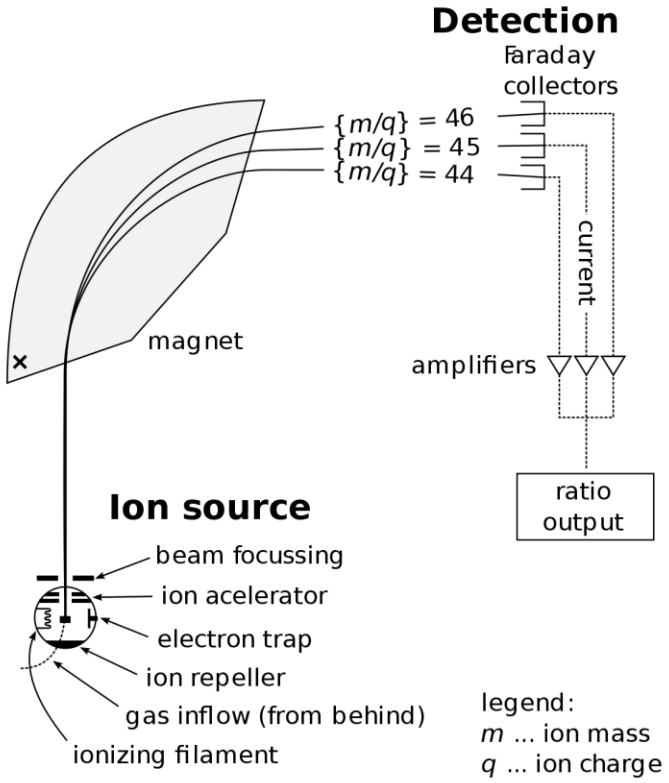
Western blot : Mesure de protéine unique



Gel 2D : Mesure de plusieurs protéines

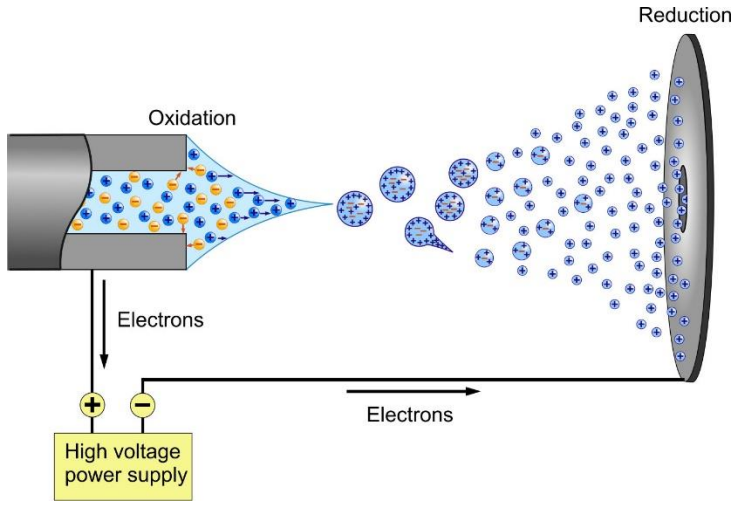


Evolution de la protéomique

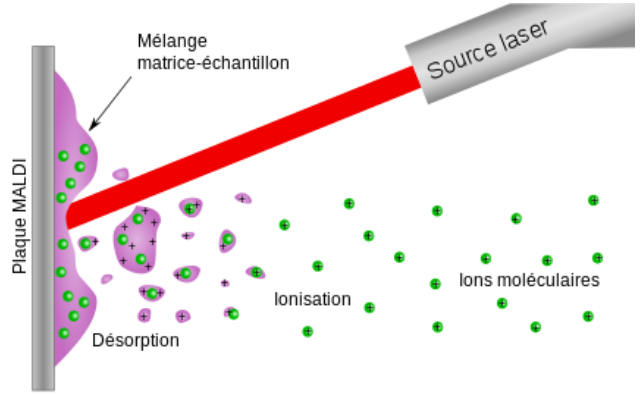


Spectrométrie de masse
1888 - 1918

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.

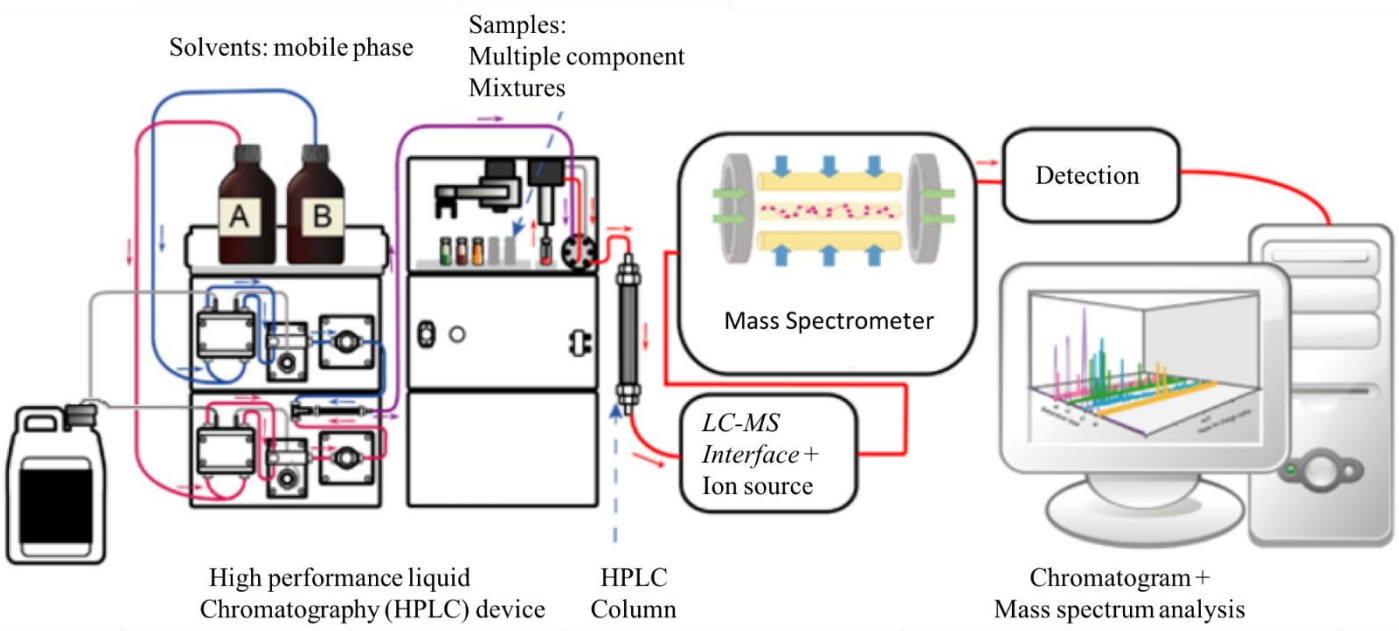


Schéma du LiquidChromatography/MassSpectrometry LC/MS - 2009

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

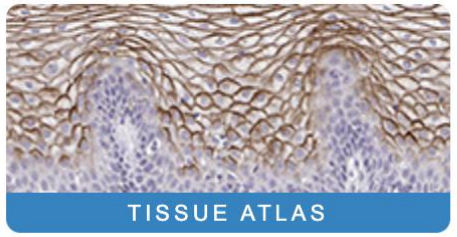
Big Science

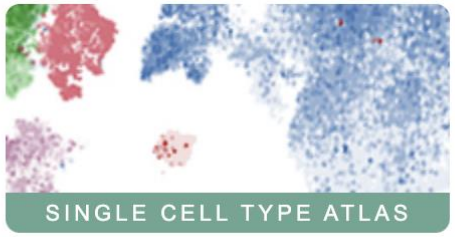
THE HUMAN PROTEIN ATLAS

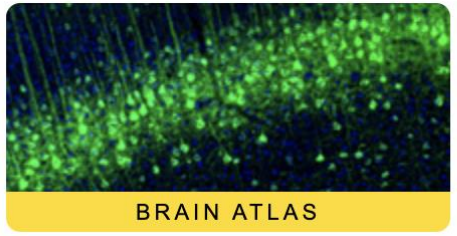
2000 - Présent

SEARCH[i]

e.g. ACE2, GFAP, EGFR



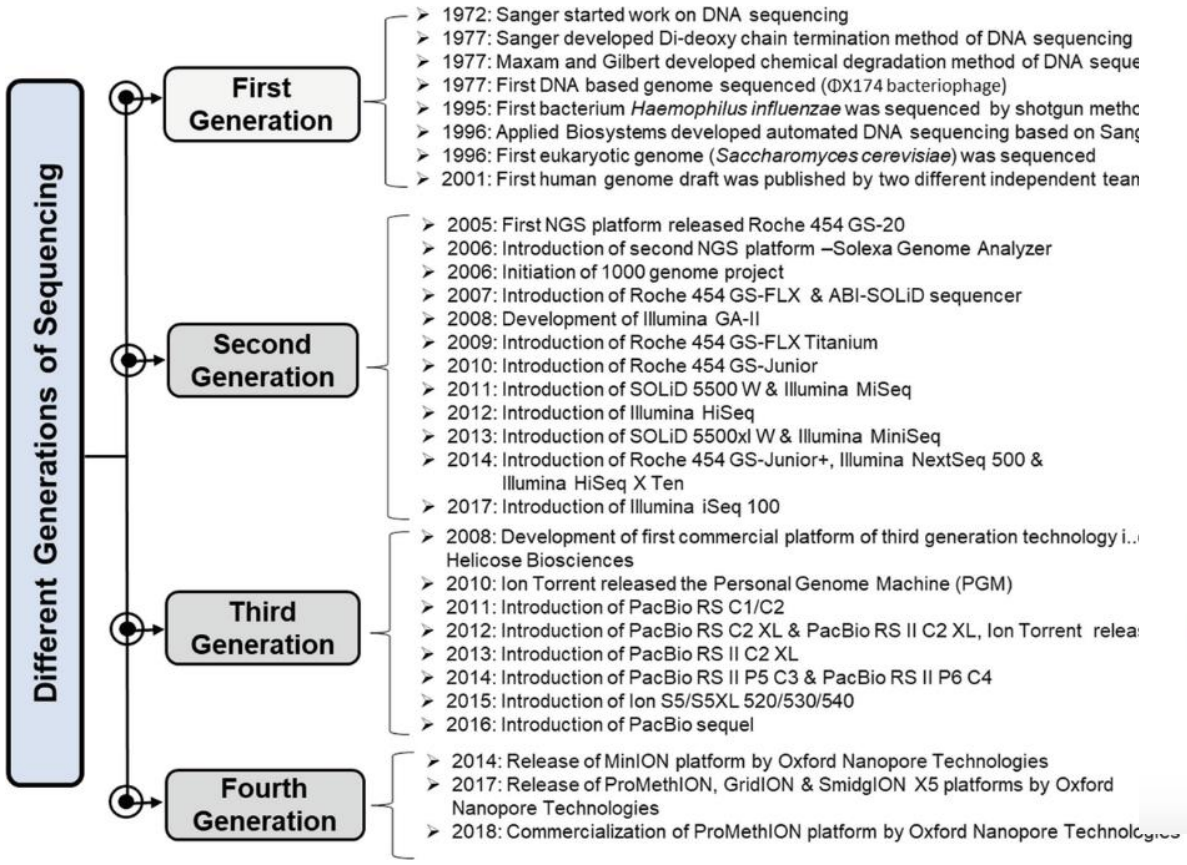




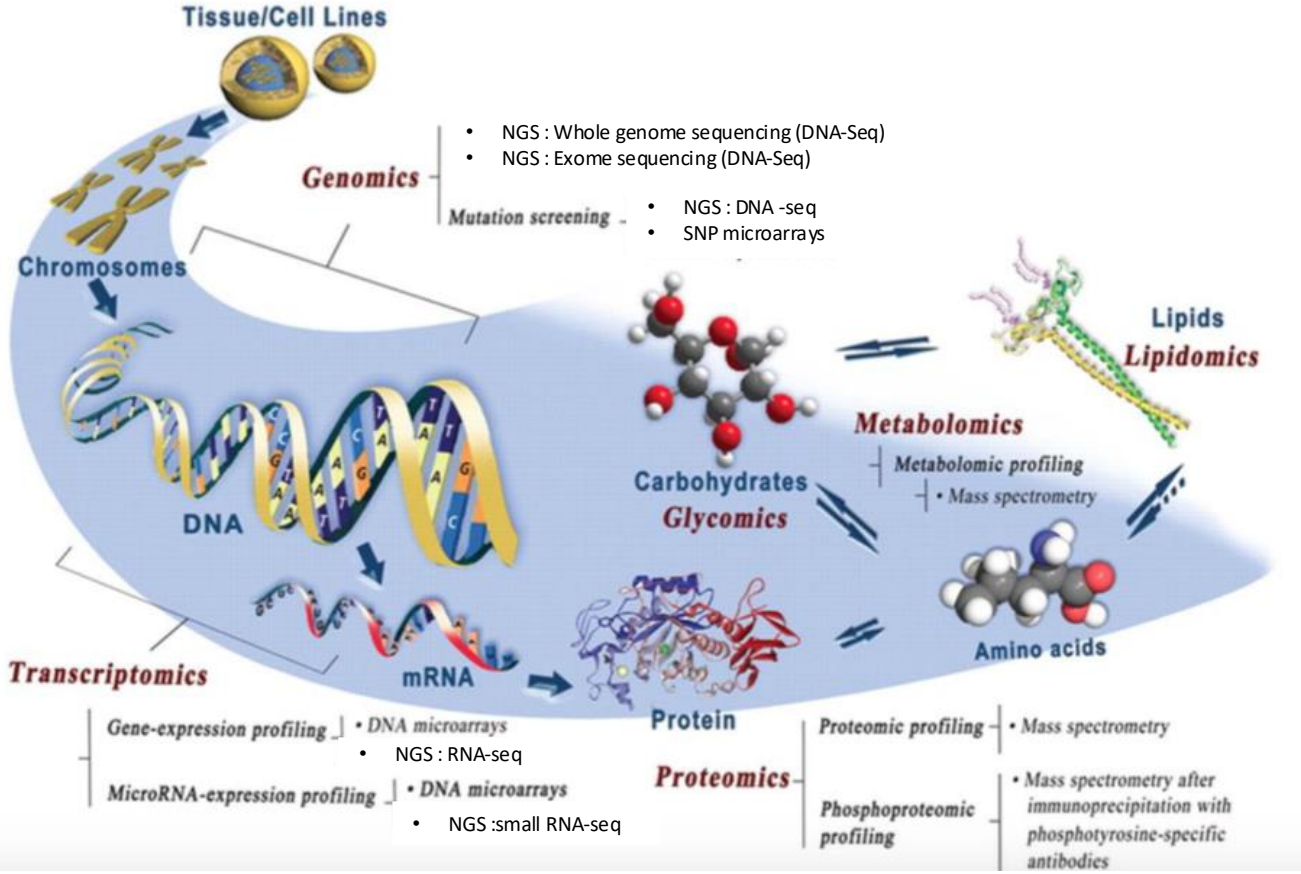
Scientific Milestones

Year	No	Title	Year	No	Title
1985	1	Affinity tags for protein purification	2012	19	Targeted proteomics
1989	2	Solid phase sequencing	2014	20	Integration of RNA and protein profiles
1993	3	Pyrosequencing	2015	21	The Tissue Atlas
1996	4	First concept of antibody-based proteomics	2016	22	Correlation of RNA and protein levels
2000	5	The Human Genome Project	2016	23	Human secretome resource
2000	6	Chromosome 21 pilot	2016	24	Antibody validation
2003	7	Start of the Human Protein Atlas program	2017	25	The Subcellular Atlas
2004	8	Tissue microarrays	2017	26	The Pathology Atlas
2004	9	The HPA data management system	2017	27	Systems medicine
2005	10	Launch of the Human Protein Atlas portal	2018	28	Wellness profiling and precision medicine
2006	11	Creation of an antibody resource	2018	29	Deep learning and citizen science
2007	12	Protein arrays	2019	30	Human secretome annotation
2008	13	Biomarkers for body fluids	2019	31	The Blood Atlas
2008	14	Epitope mapping of antibodies	2019	32	The HPA Kaggle Challenge
2008	15	Antibodypedia antibody portal	2020	33	The Brain Atlas
2009	16	Biomarker discovery in pathology	2020	34	The Metabolic Atlas
2010	17	Knowledge-based portal	2020	35	The fight against the novel coronavirus
2011	18	Therapeutic antibodies and Affibodies			

Comment détecter les régulations transcriptomiques ?



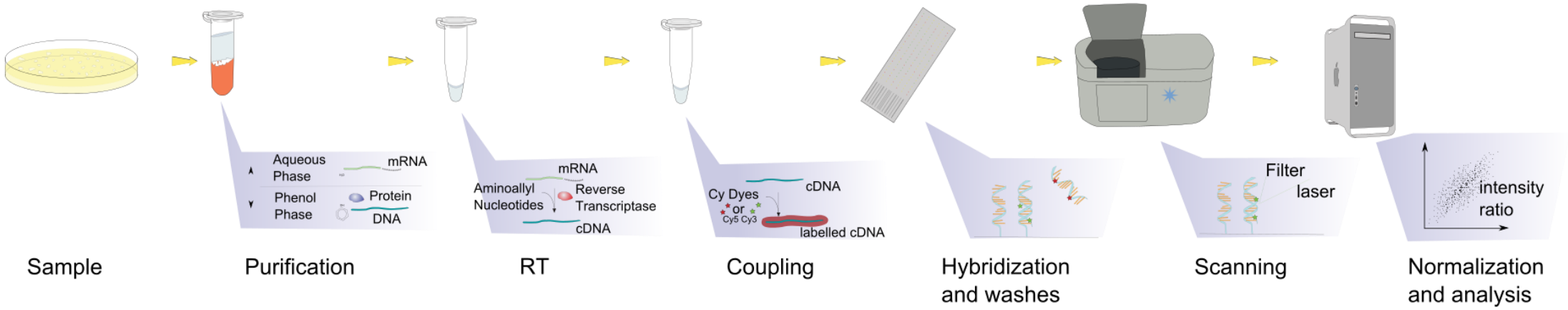
Microbial Technology for the Welfare of Society, 2019, Springer



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

Evolution de la transcriptomique

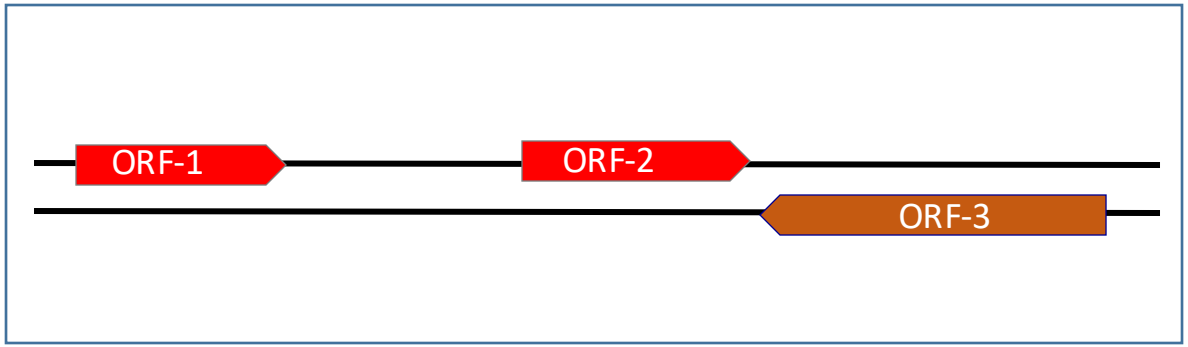
Les prémices avec les puces ADN



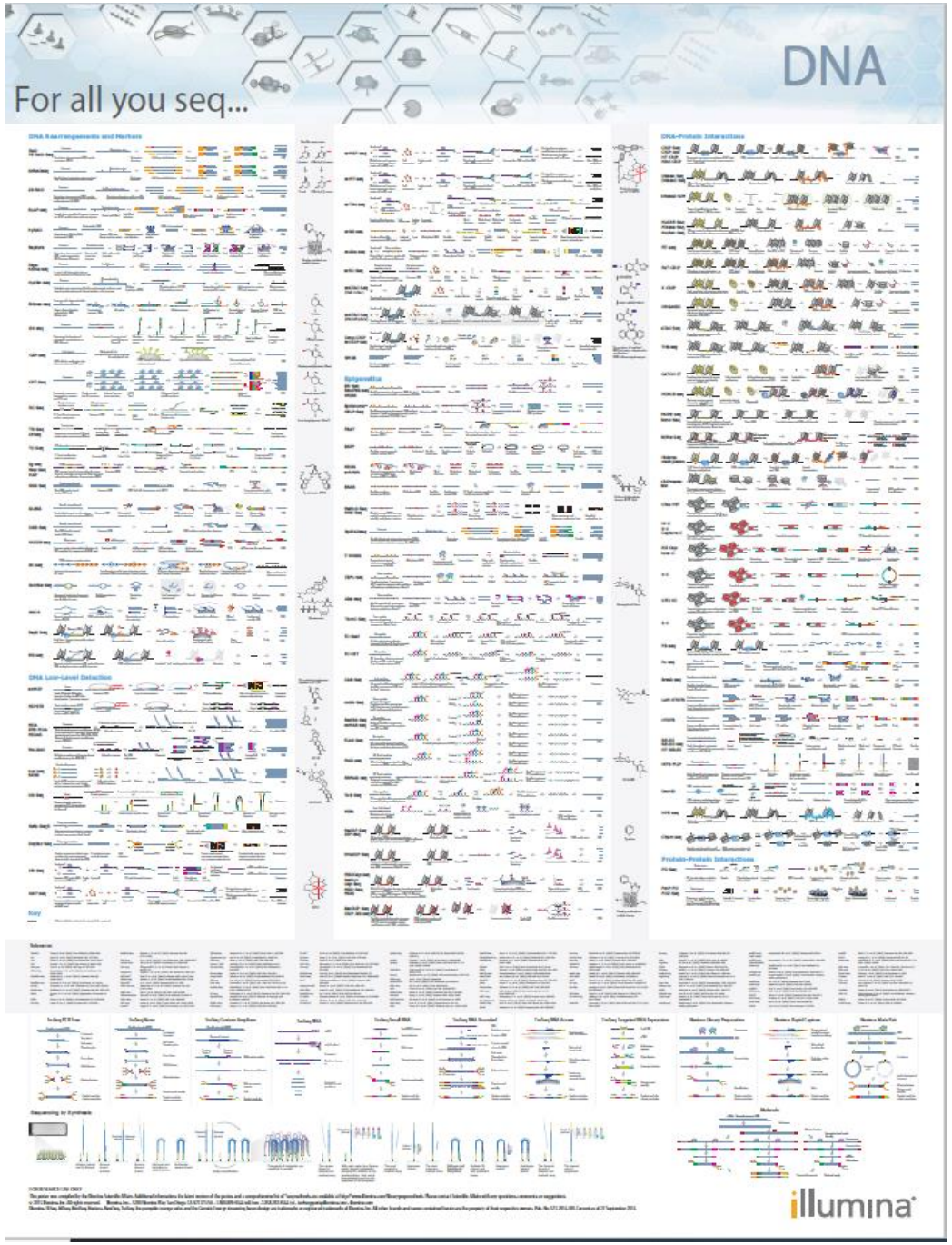
- Expression des ARNs
- Comparaison de génomes
- SNP
- ChIP

Séquençage haut débit - short-read

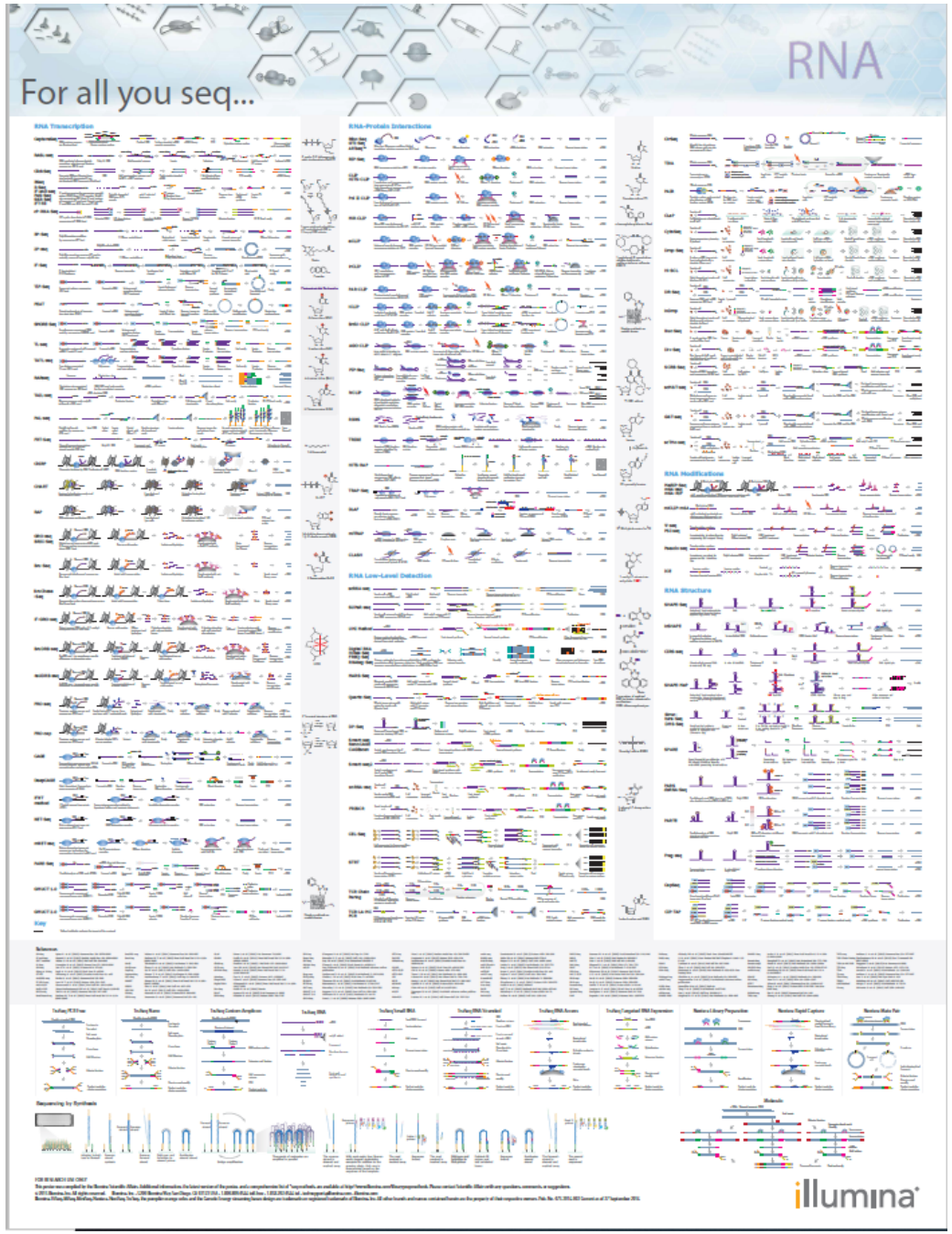
mRNA extraction



Reads mapping



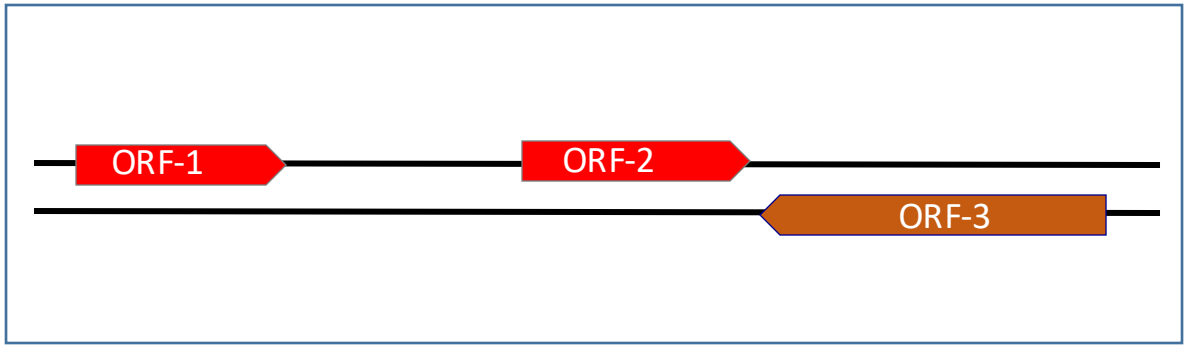
https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/ForAllYouSeqMethods.pdf



https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/ForAllYouSeqMethods.pdf

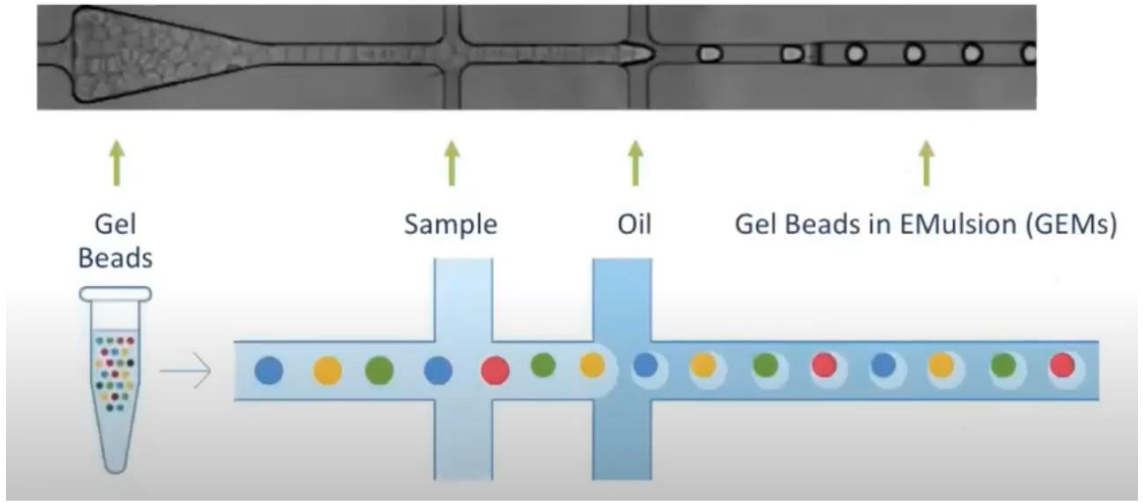
Séquençage haut débit - long-read

mRNA extraction

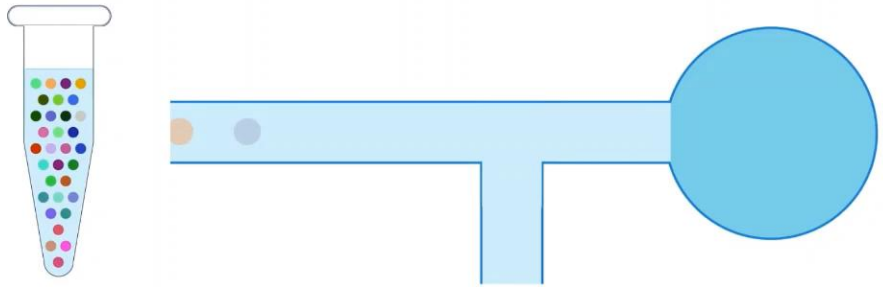


Reads mapping

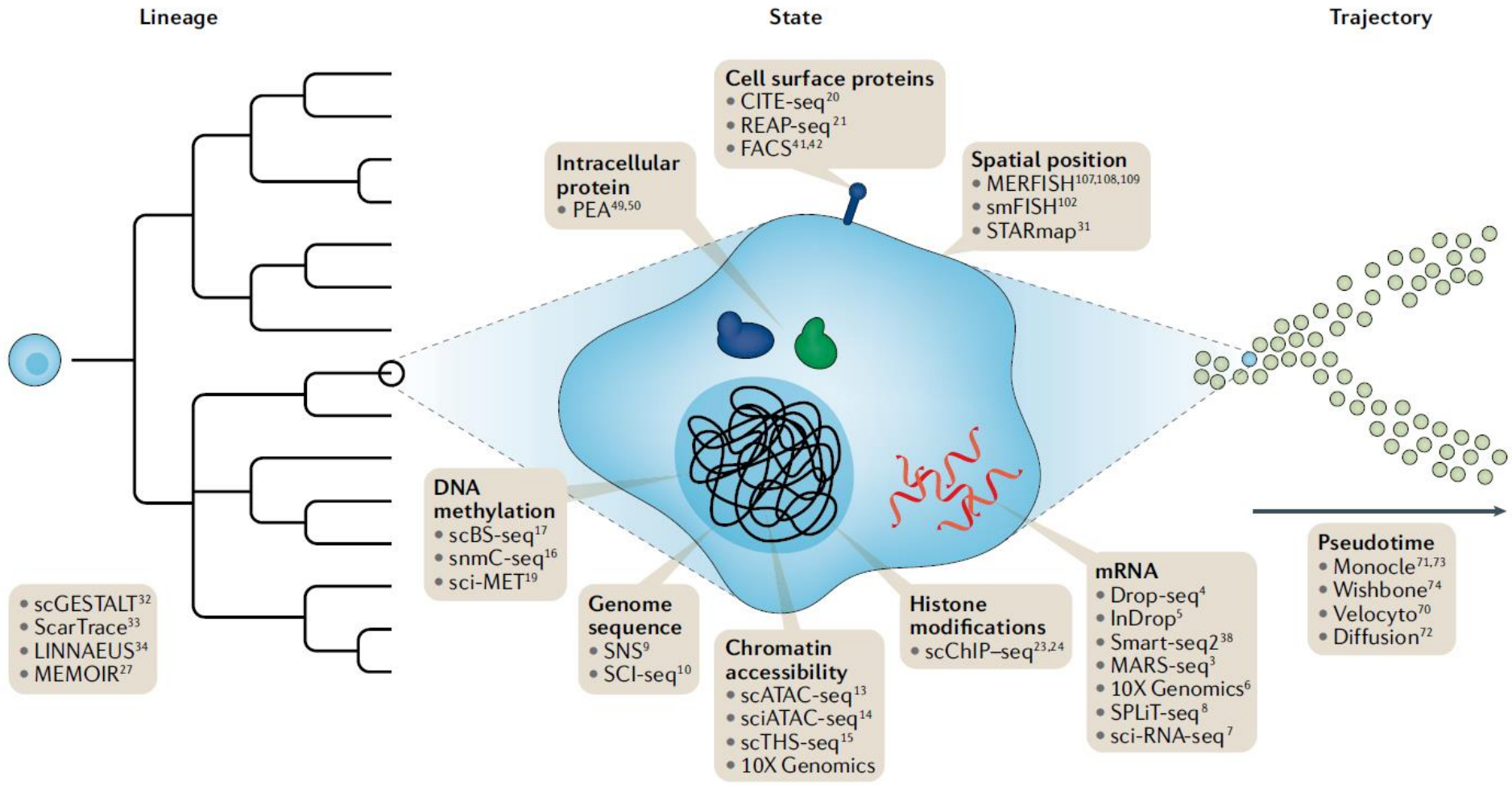
Séquençage à cellule unique



Adapted from 10x Genomics



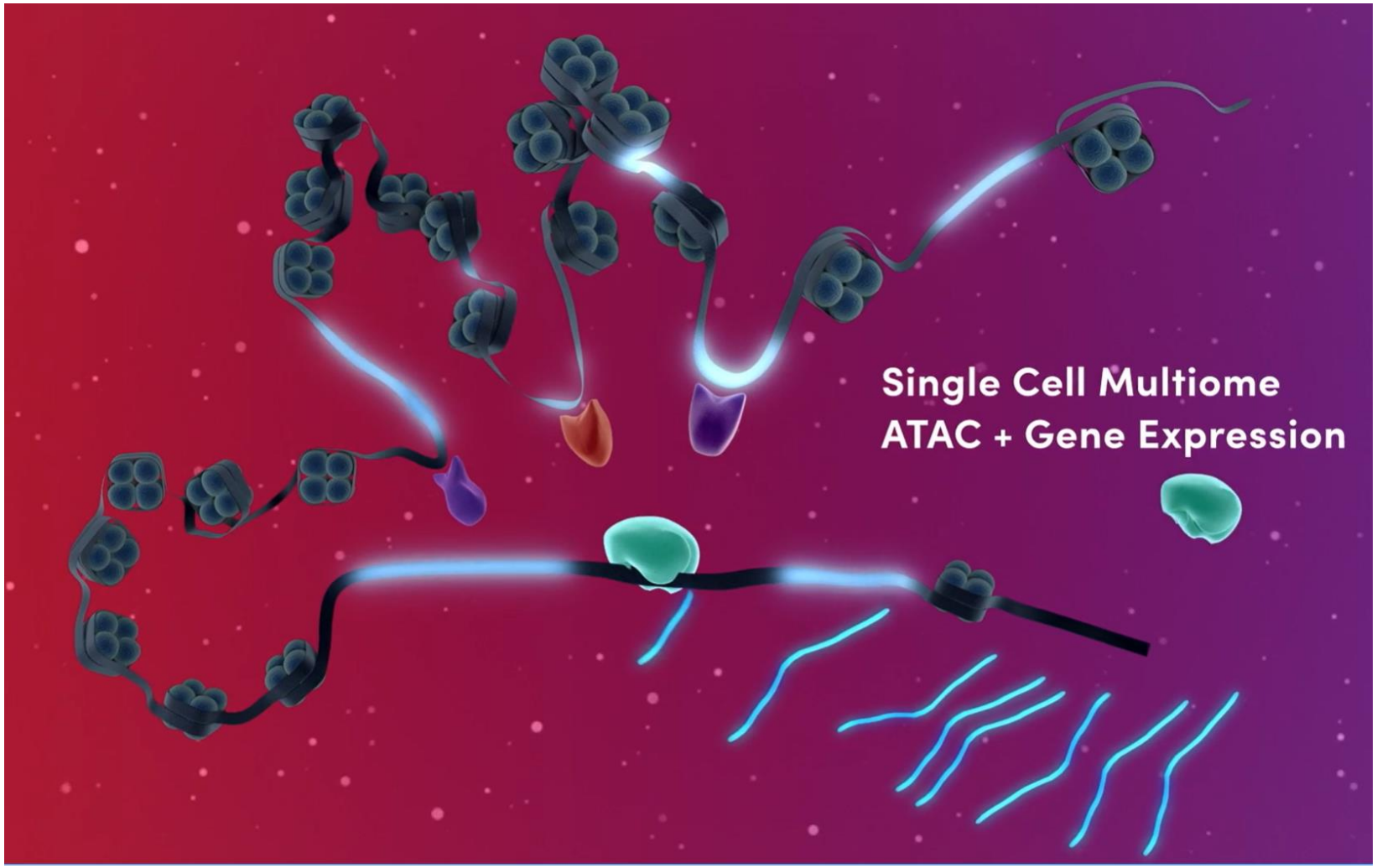
Une variété de techniques single-cell

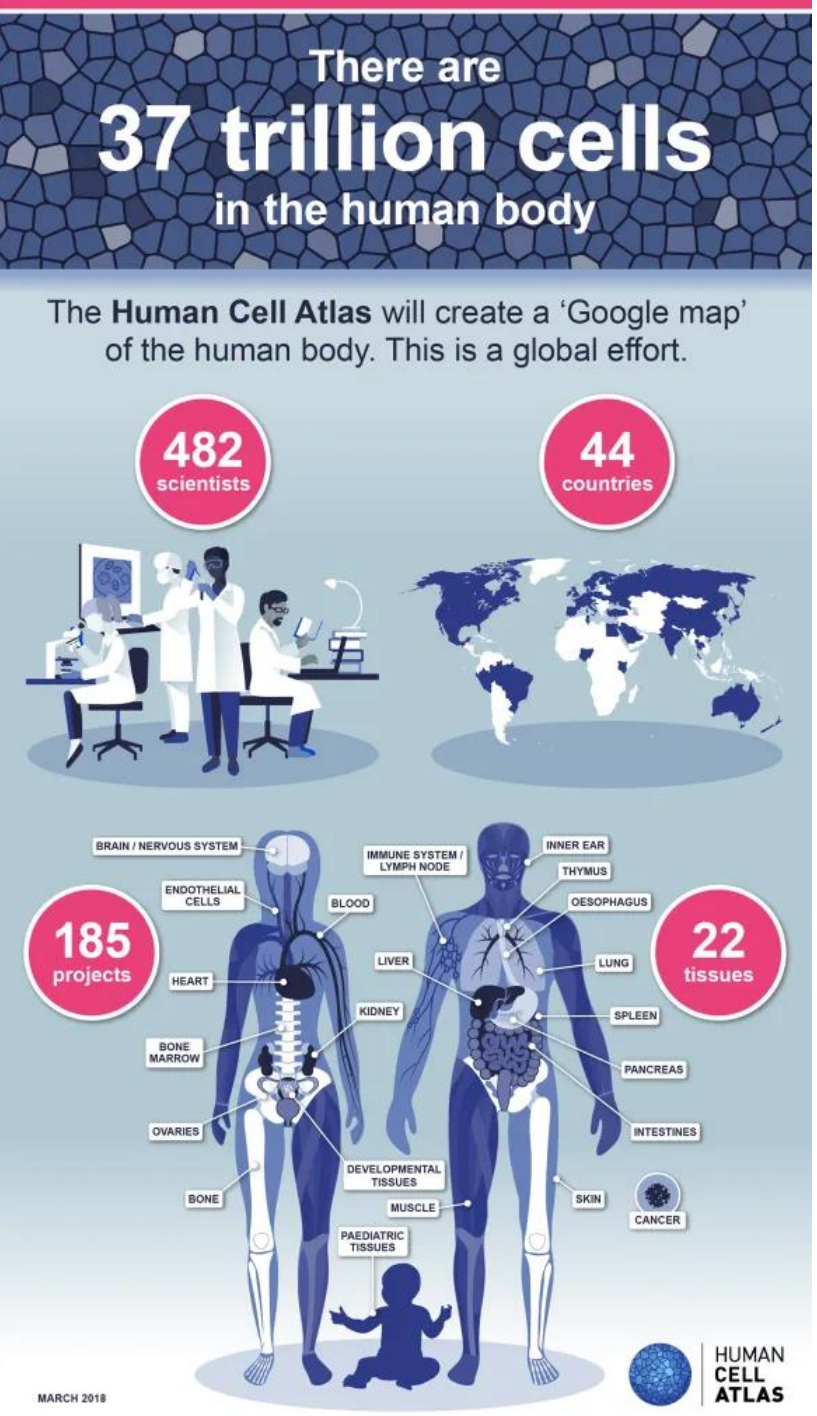


Integrative single cell analysis, Stuart & Satija, Nature Rev. Genetics. 2019

Mesure simultanée sur cellule unique

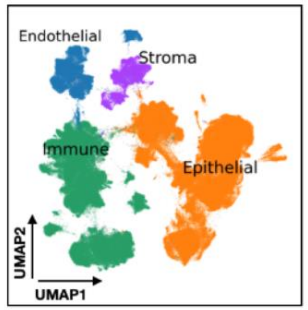
Exemple : 10X Multiome sequencing



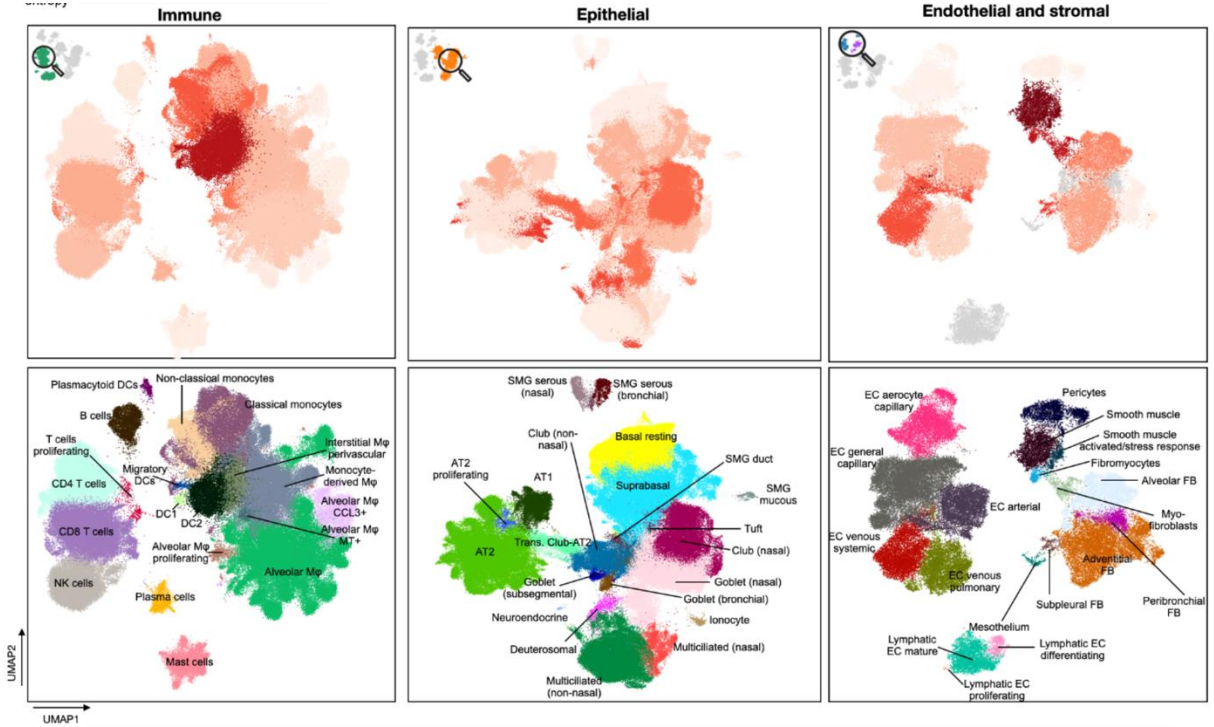


Big Science
Human Cell Atlas

2017-present

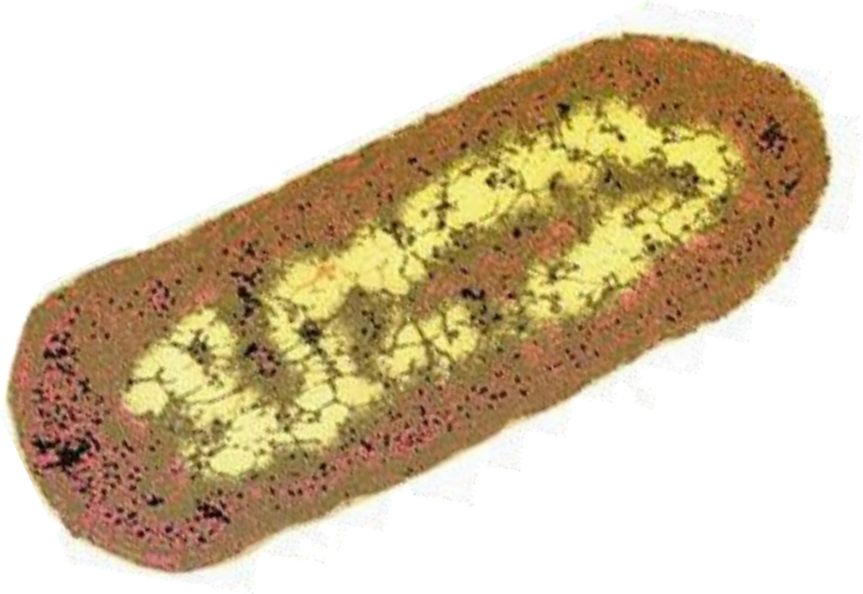


Atlas des voies respiratoires
46 jeux de données
+ 2,2 millions de cellules



Sikkema et al. Nature Medecine 2023

La bioinformatique pour analyser les données omiques

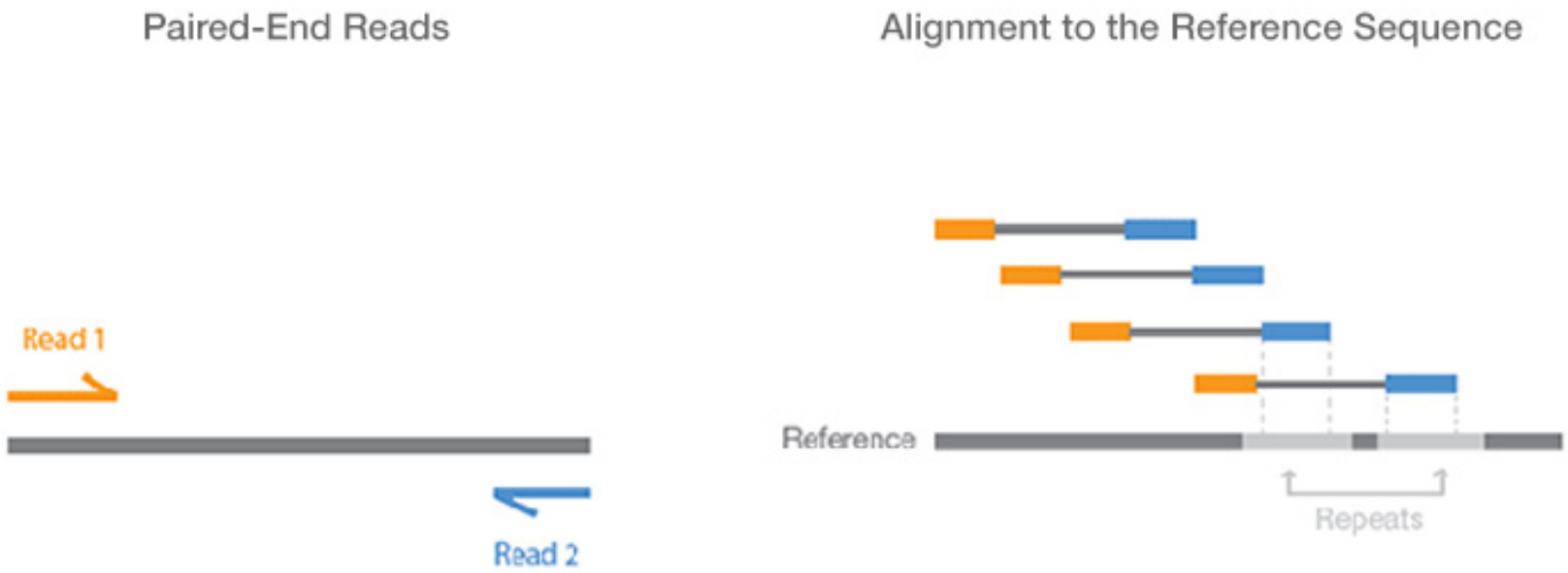


Escherichia coli K12 substr. MG1655
4641652 bp with 1 chroimosome

Exemple: Analyse RNA-Seq de *Escherichia coli GB4 BIMB*

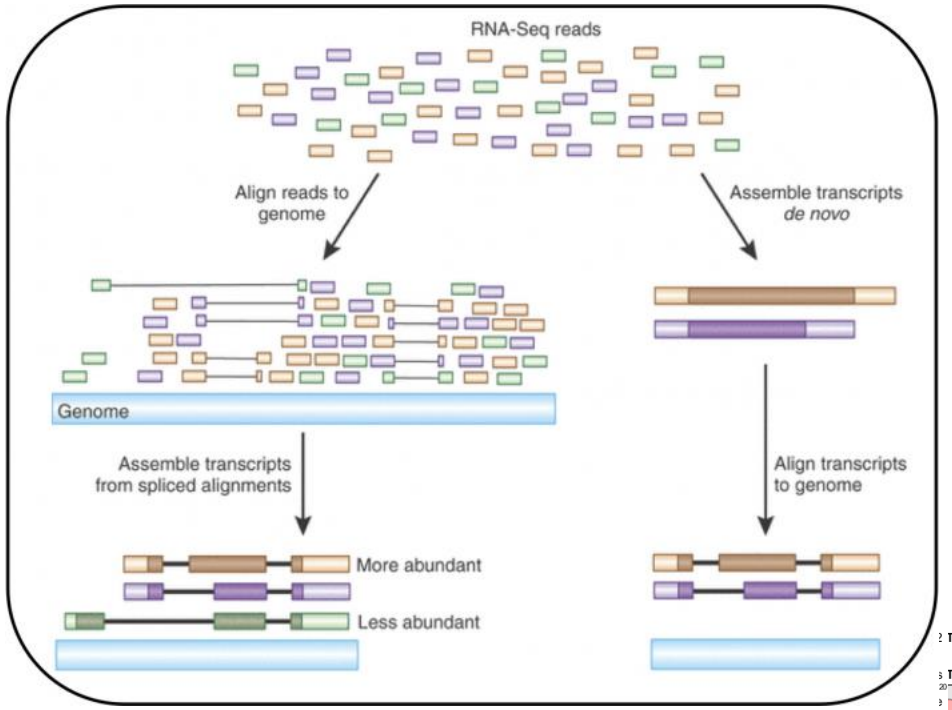
- 2 technical replicates of WT strain
- 2 technical replicates of MazF expressing strain

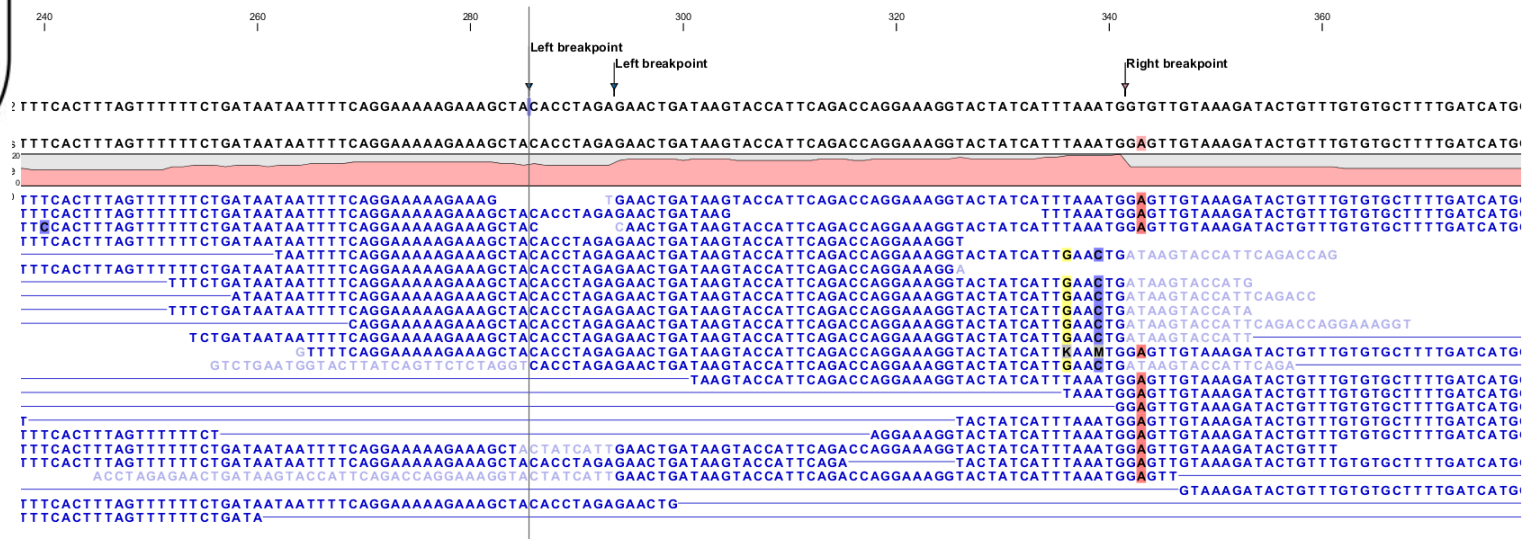
Paired-end sequencing



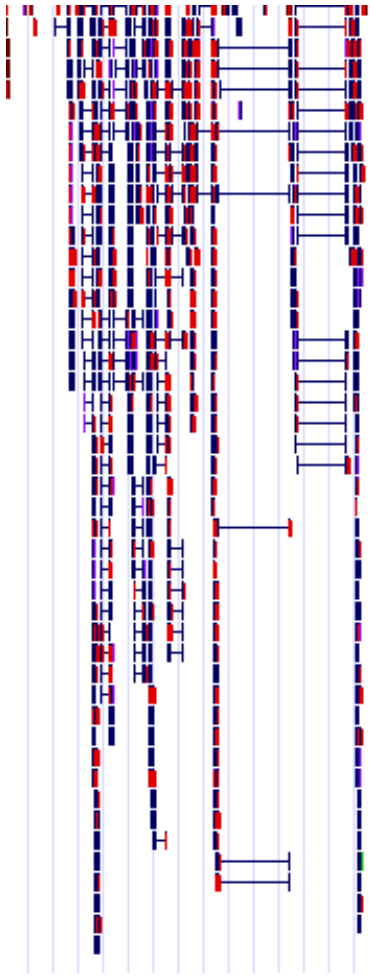
Paired-end sequencing enables both ends of the DNA fragment to be sequenced. Because the distance between each paired read is known, alignment algorithms can use this information to map the reads over repetitive regions more precisely. This results in much better alignment of the reads, especially across difficult-to-sequence, repetitive regions of the genome.

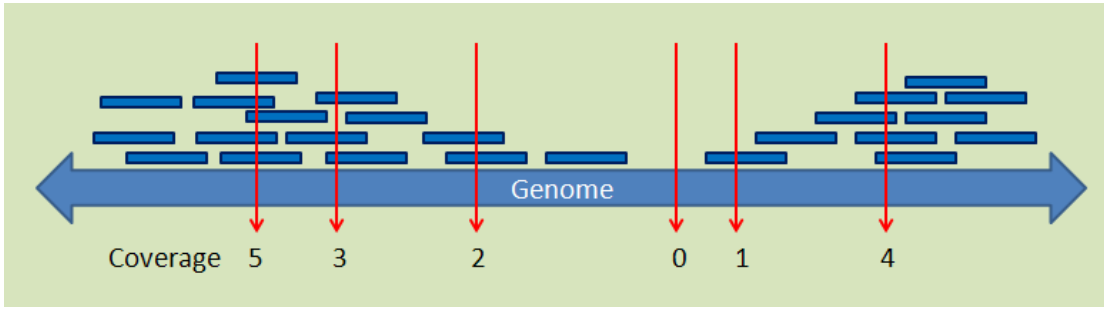
Mapping des reads





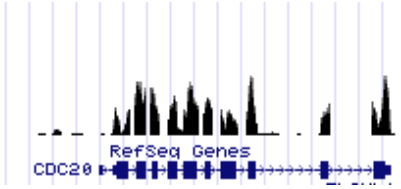
Création des fichiers de couvertures





Fichier .bam

Alignements des reads le long du génome par leurs coordonnées génomiques



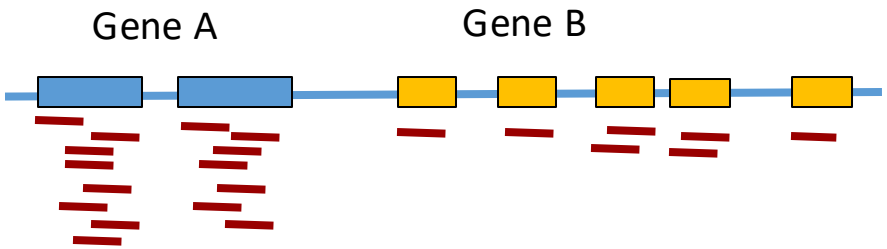
Fichier .bw

Coverage (couverture)
= Somme de reads alignés à chaque position génomique

Transcrit annoté

Comptage du nombre de reads par gène (ou exon)

Echantillon 1



Echantillon 2

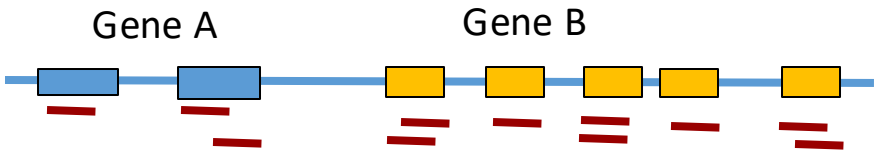


Tableau de counts bruts:

RNA-seq (NGS)

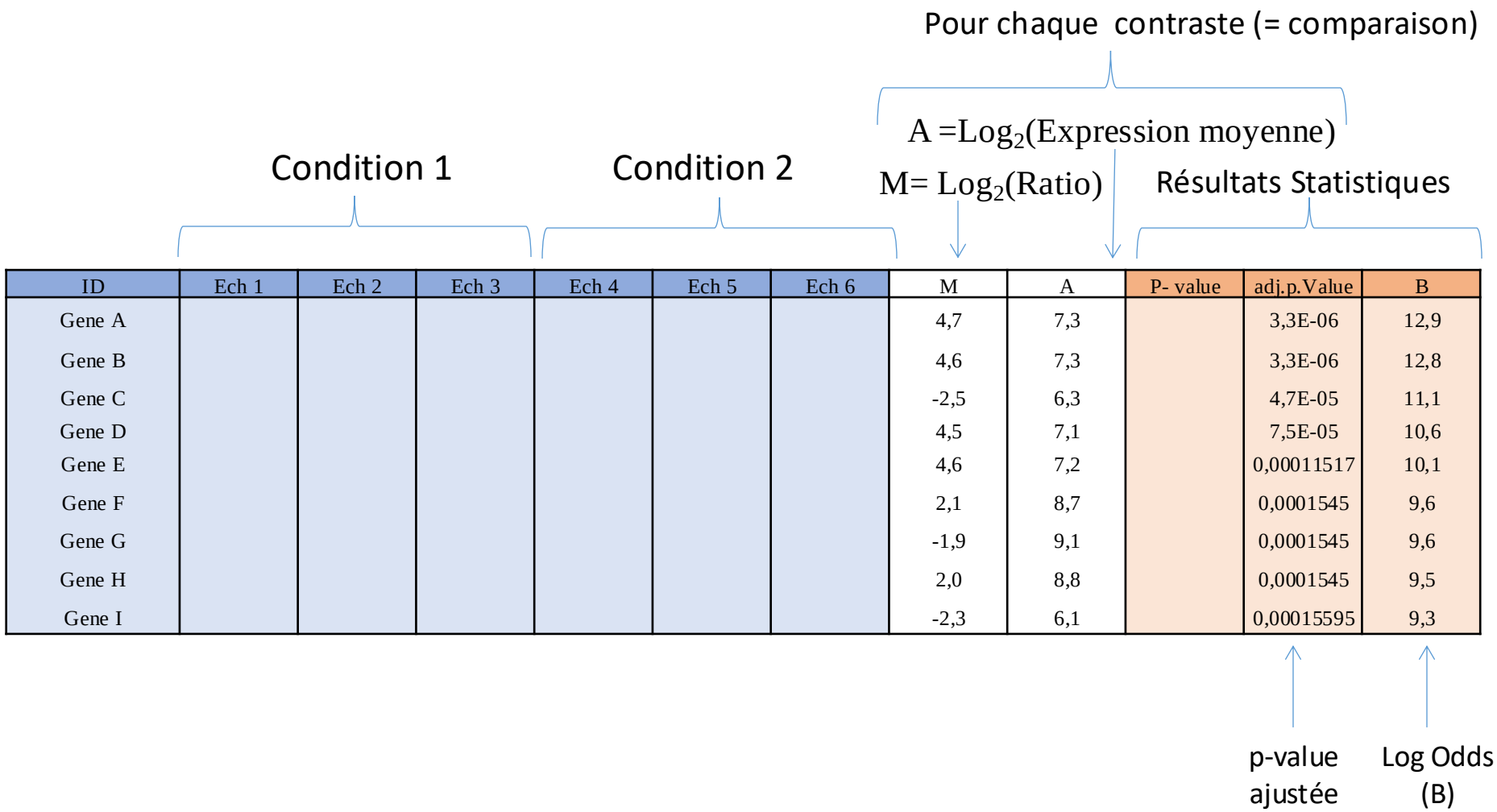
= Comptage des reads associés à chaque gène (ou exon) après alignement sur le génome

	Echantillon 1	Echantillon 2	...
Gene A	15	3	
Gene B	7	8	

...

Analyse d'expression différentielle

Exemple Design = 3 réplicats pour chaque condition

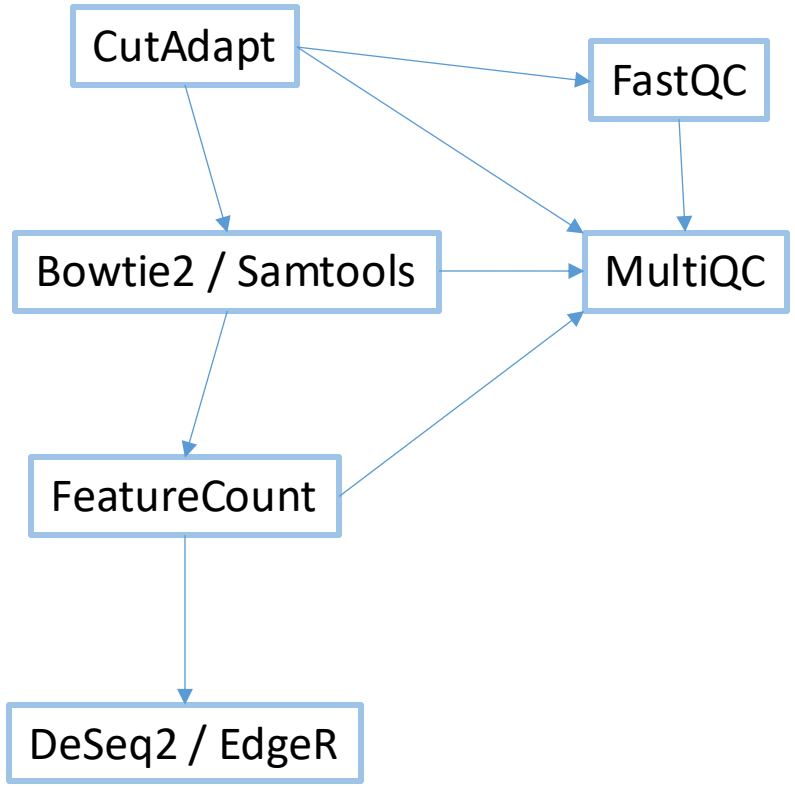


ID	Ech 1	Ech 2	Ech 3	Ech 4	Ech 5	Ech 6	M	A	P- value	adj.p.Value	B
Gene A							4,7	7,3		3,3E-06	12,9
Gene B							4,6	7,3		3,3E-06	12,8
Gene C							-2,5	6,3		4,7E-05	11,1
Gene D							4,5	7,1		7,5E-05	10,6
Gene E							4,6	7,2		0,00011517	10,1
Gene F							2,1	8,7		0,0001545	9,6
Gene G							-1,9	9,1		0,0001545	9,6
Gene H							2,0	8,8		0,0001545	9,5
Gene I							-2,3	6,1		0,00015595	9,3

RNASeq analysis workflow

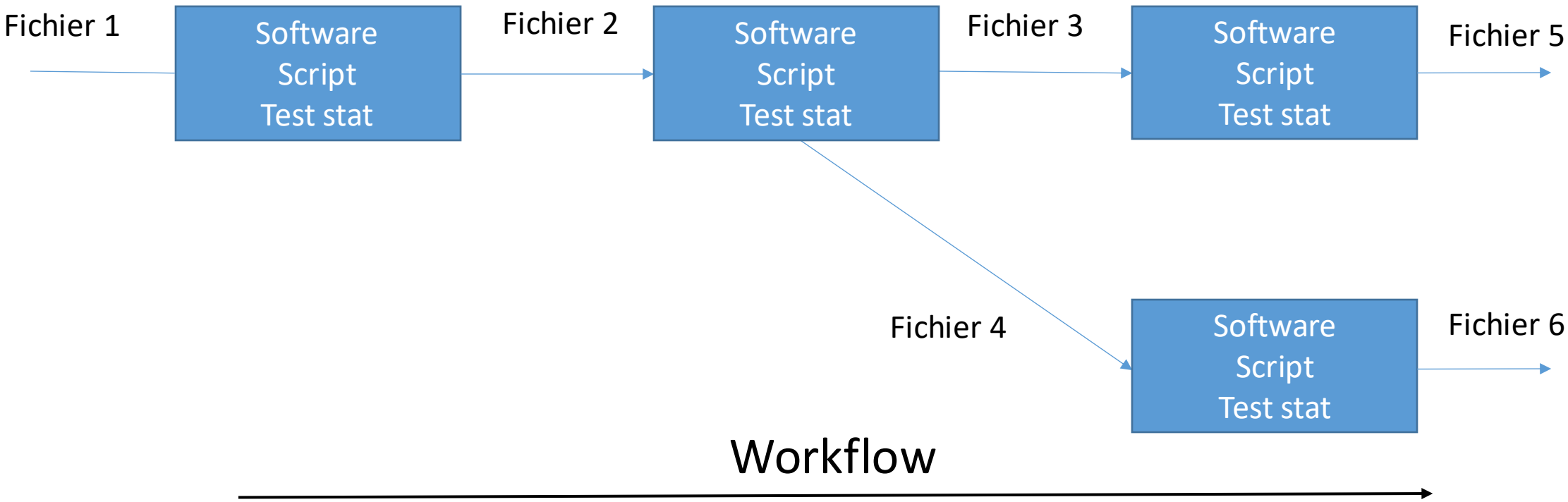
- Trimming of reads - CutAdapt
- Reads mapping – Bowtie2
- Counting read per genes - FeatureCount
- QC tools - MultiQC
- Differential expression analysis – DeSeq2

Workflow implementation

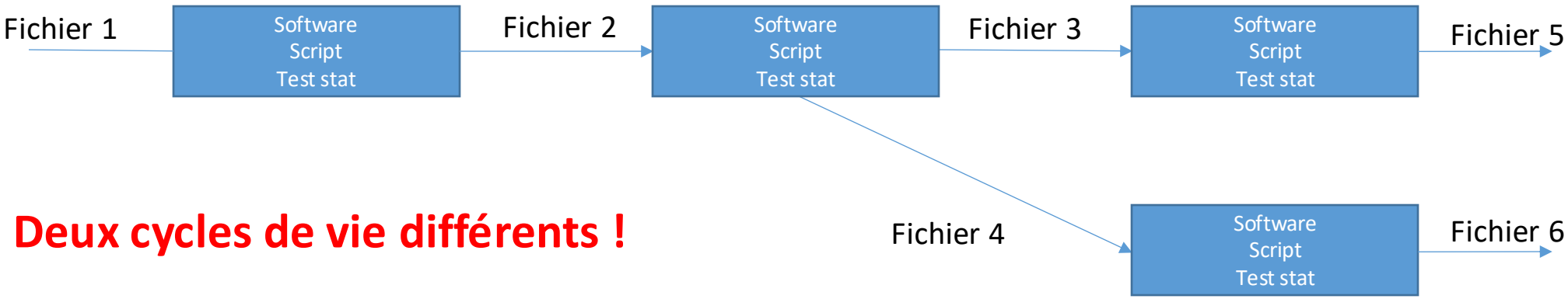


Workflow bioinformatique

Tout pipeline d'analyse bioinformatique est un workflow d'échange et modification de données

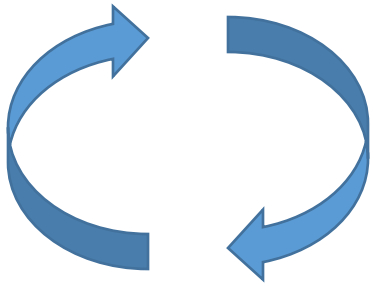


Le developpement de projet et la gestion des versions



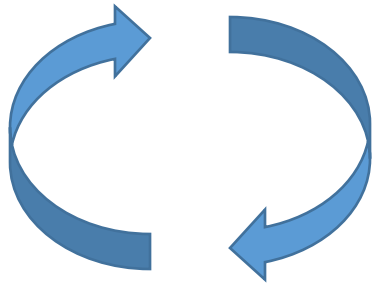
Deux cycles de vie différents !

Software/Script/Test stat



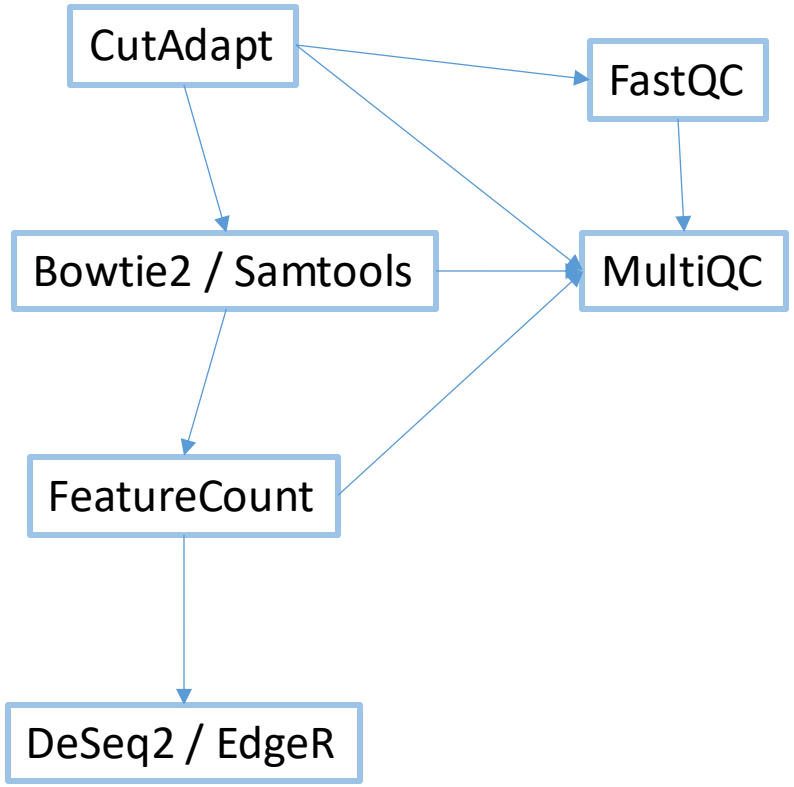
Software 1
Test 1
Script 1
Script 2

Datasets



Data 1
Data 2
Data 3
Data 4

Temps d’exécution



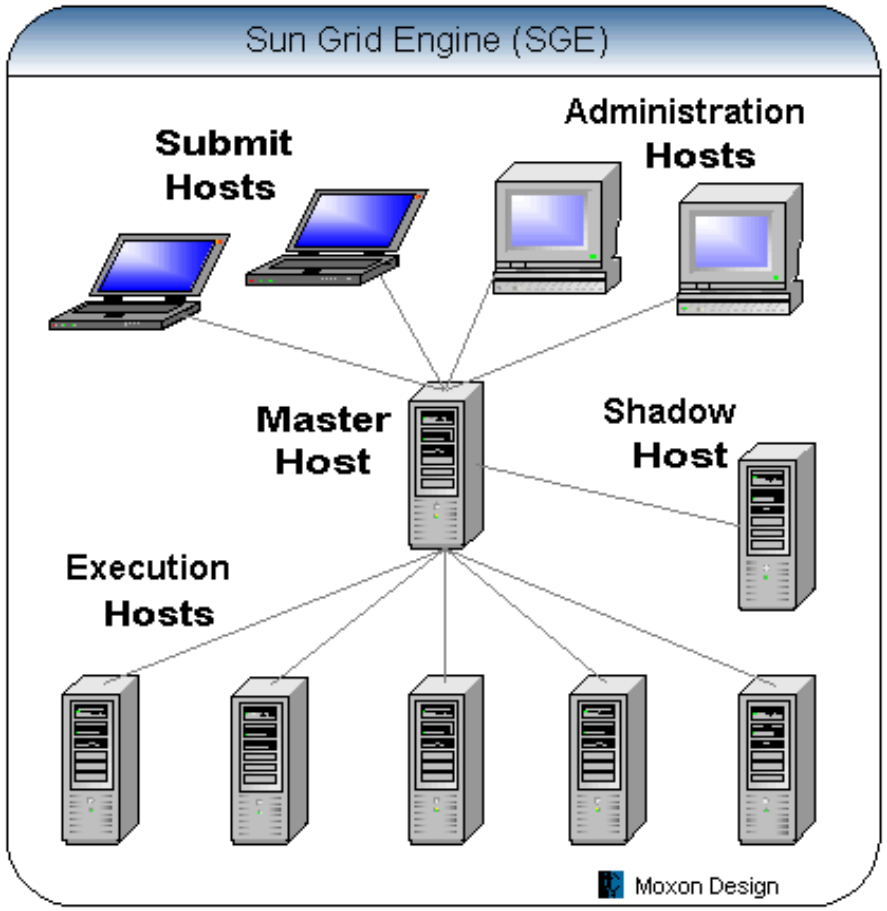
CutAdapt – 5 minutes
Bowtie2 – 30 minutes
FeatureCount – 5 minutes
FastQC – 2 minutes
MultiQC – 2 minutes
DeSeq2 – 1 minute

Total: 45 minutes for 1 dataset

4 datasets = 3h
8 datasets = 6h
16 datasets = 12h
…

Ecoli = 4.6 Mbase pair
Human = 3.4 Gbase pair !!!
x10 more data !
Need to have more POWER

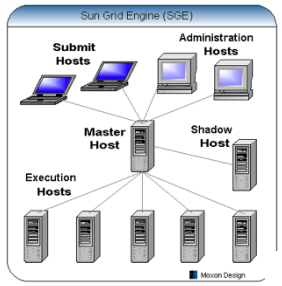
Le cluster de calcul



- Submit hosts (notre ordinateur)
- Master hosts (job scheduler running on SGE)
- Execution hosts (array of computers)

!!! Faire attention à la mémoire partagée entre nœuds et celle unique à chaque nœud !!!

Le calcul Haute-performance - HPC

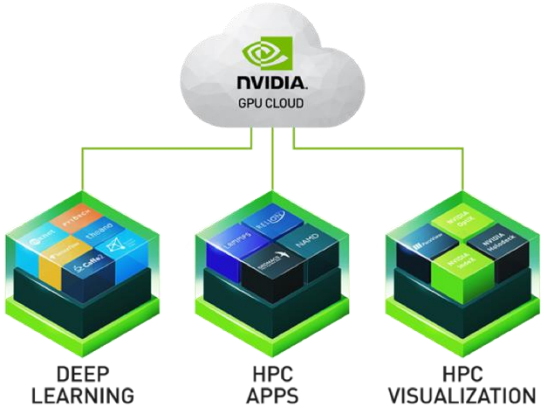


slurm workload manager

PBS Works™

Cluster de calcul CPU

Cluster de calcul GPU

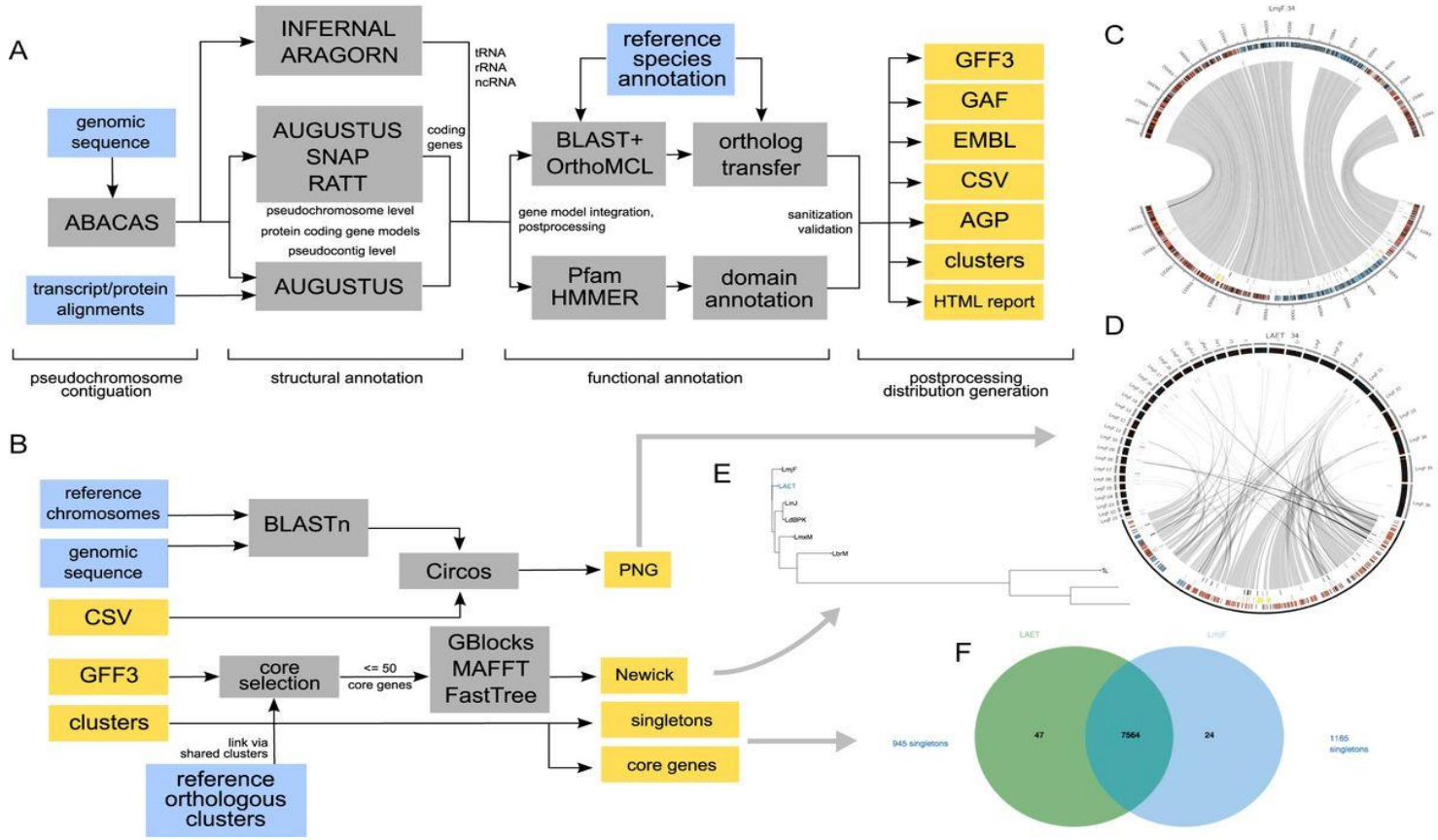


Cluster de calcul cloud

Exemple de pipeline bioinformatique

https://academic.oup.com/nar/article/44/W1/W29/2499316

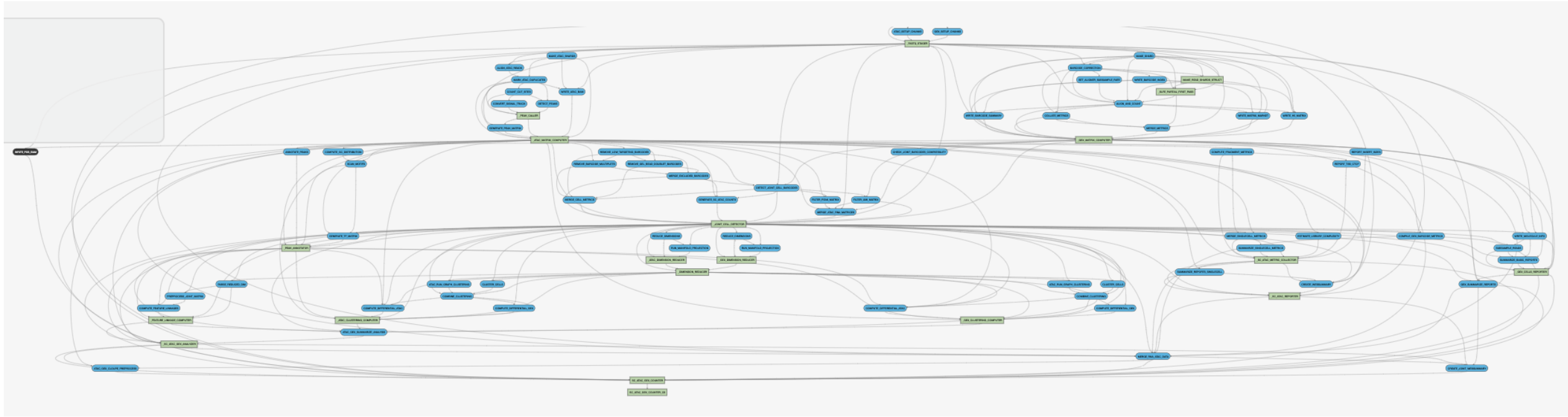
Pipeline d'annotation de génome bactérens



Exemple de pipeline bioinformatique

10X Single cell Multiome (RNA-Seq + ATAc-Seq)

https://support.10xgenomics.com/single-cell-multiome-atac-gex/software/pipelines/latest/map/cr-arc

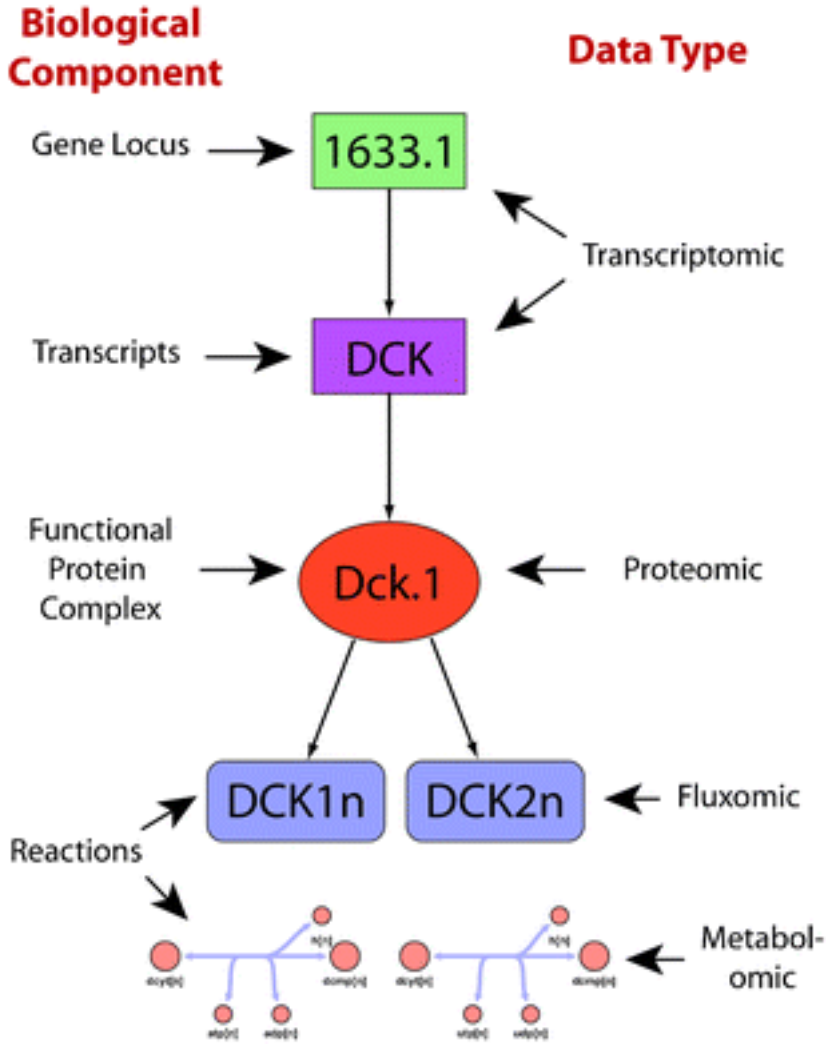


Biologie des systèmes

- L’approche systémique en biologie
- Bioinformatique et données omiques
- **Reconstruire un réseau biologique**
  - Les différents types de réseaux
  - Reconstruction de réseau biologique
  - Les obstacles à la reconstruction
  - Structure des réseaux biologiques
  - Etat de l’art des reseaux biologiques les plus étendus
- Virtual cell et digital twins

Les éléments fondamentaux des réseaux biologiques

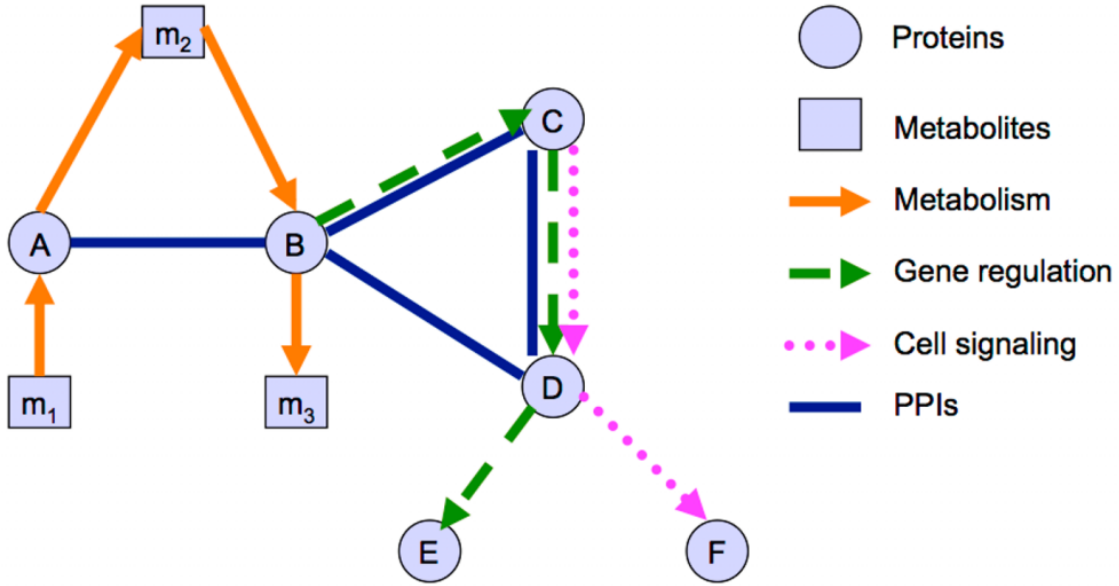
- Les gènes
- Les transcrits
- Les protéines
- Les complexes de protéines
- Les métabolites
- Les réactions enzymatiques



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

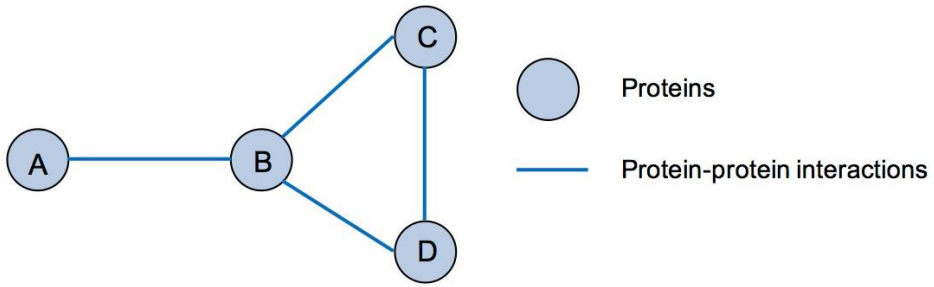
Types de réseaux biologiques

- Réseau d'interaction protéine-protéi (PPI)
- Réseau métabolique
- Réseau de régulation génétique
- Réseau de signalisation cellulaire



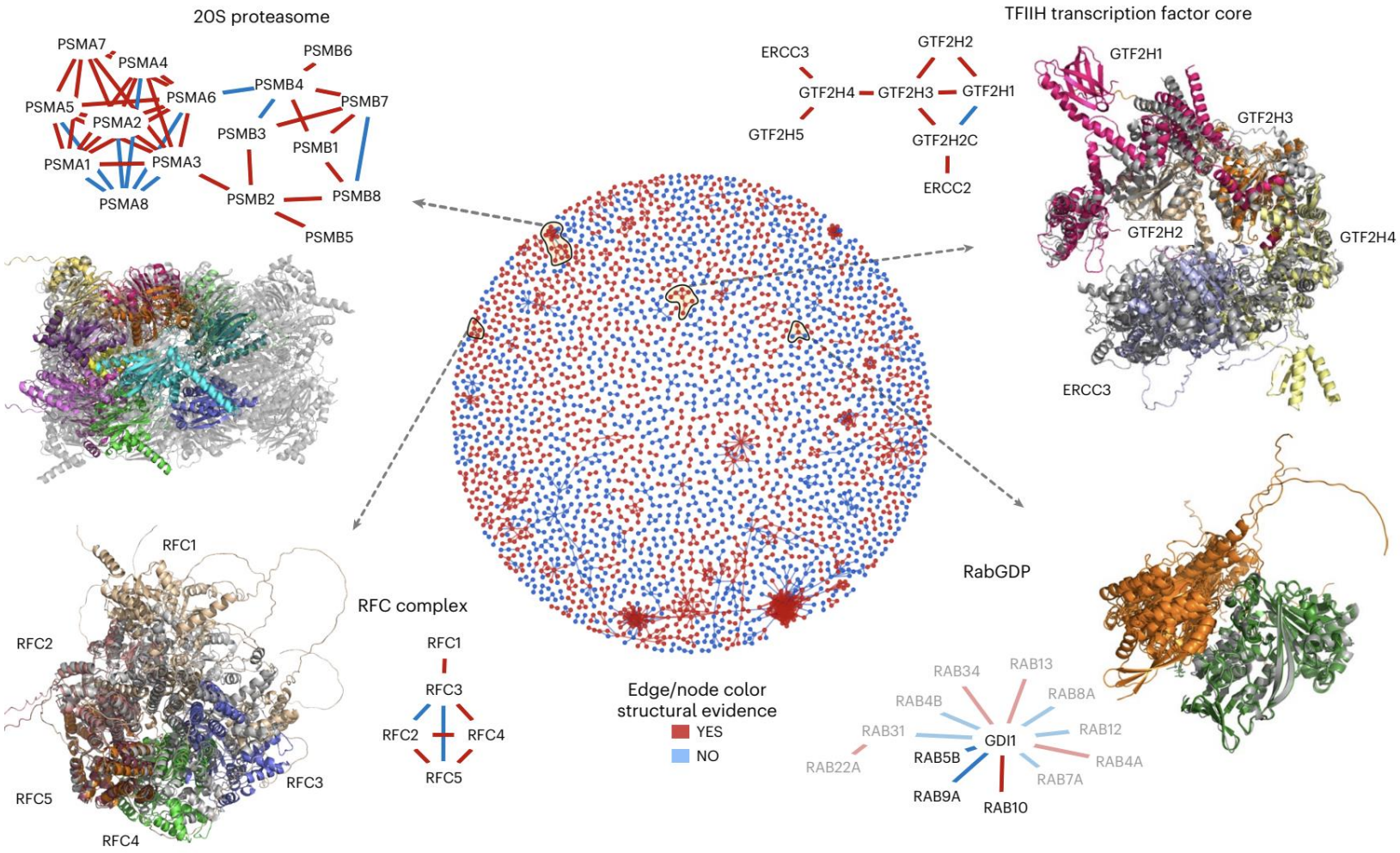
Réseau d’interaction protéine-protéine (PPI)

- 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

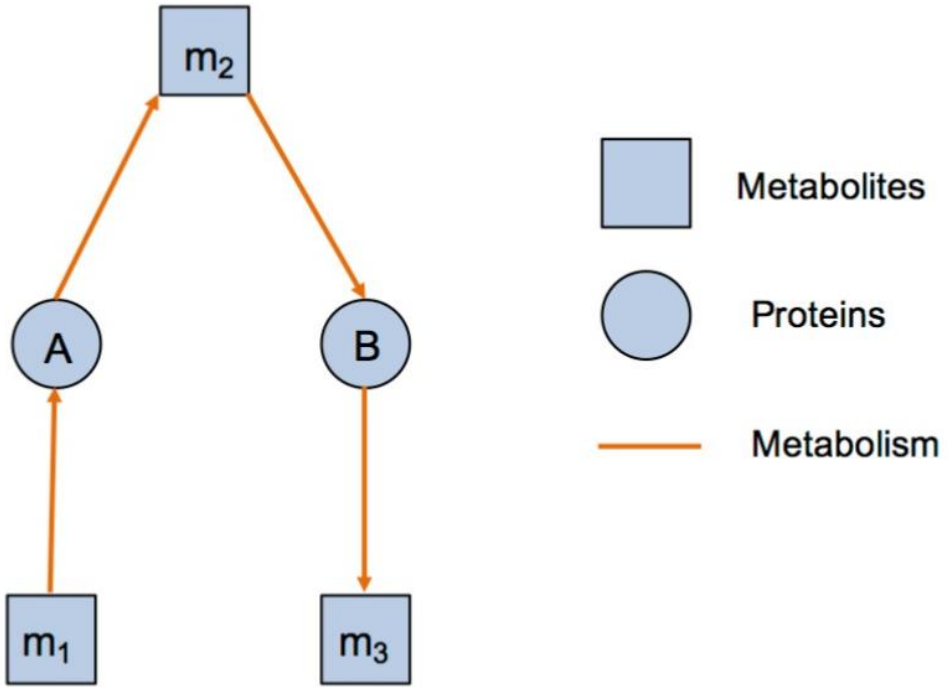
➢ 5000 protéines

Prédiction faites en utilisant FoldDock et AlphaFold

Burke et al.
Nature Structural & Molecular Biology 2023

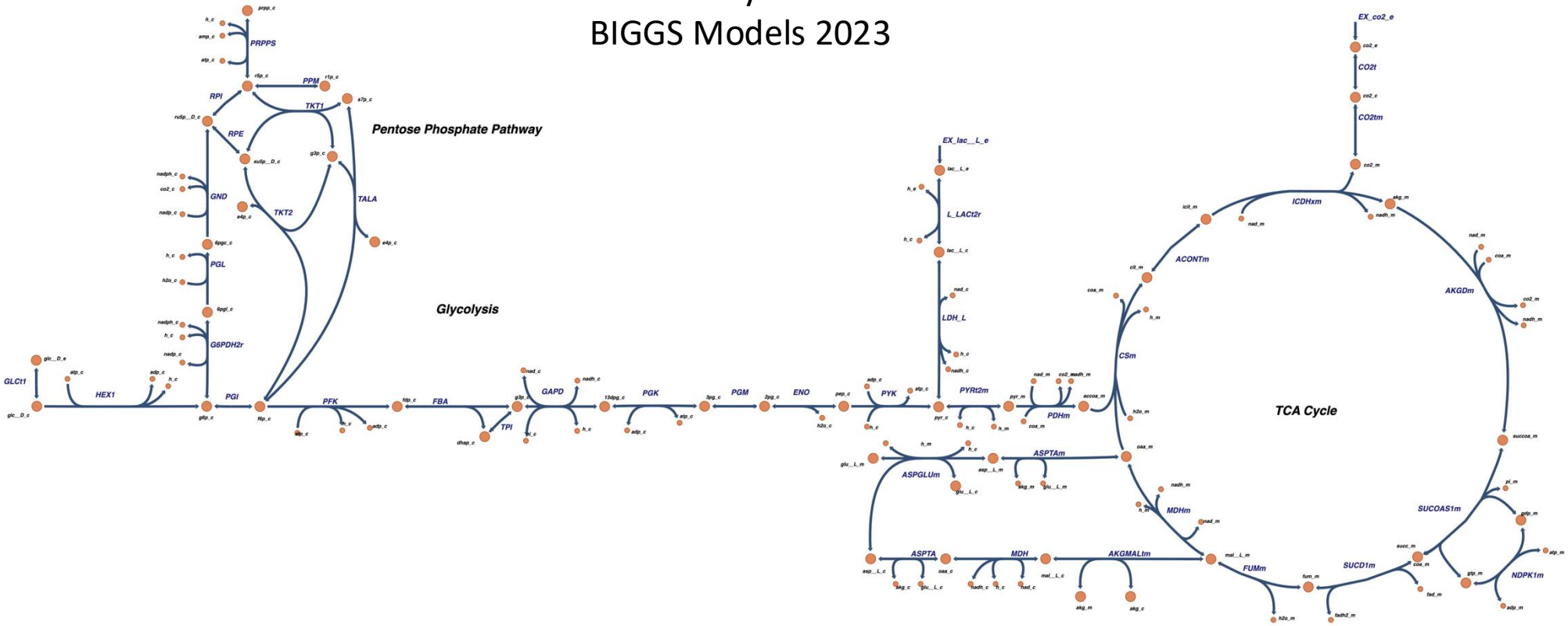
Réseau métabolique

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



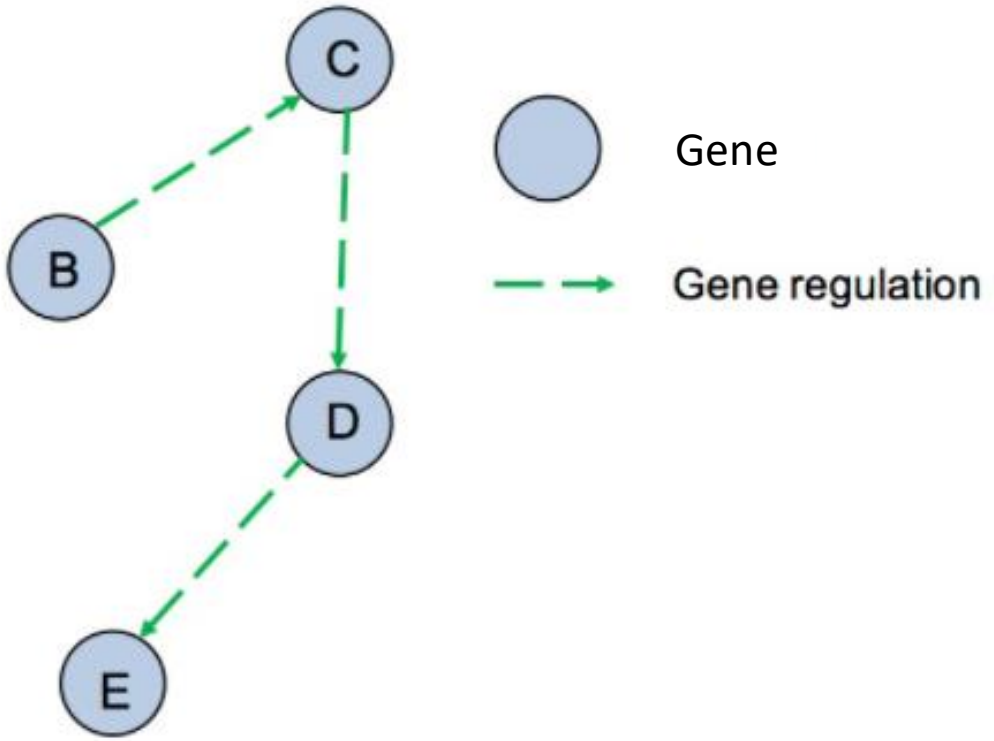
Exemple de réseau métabolique

Human Glycolis TCA PPP
BIGGS Models 2023

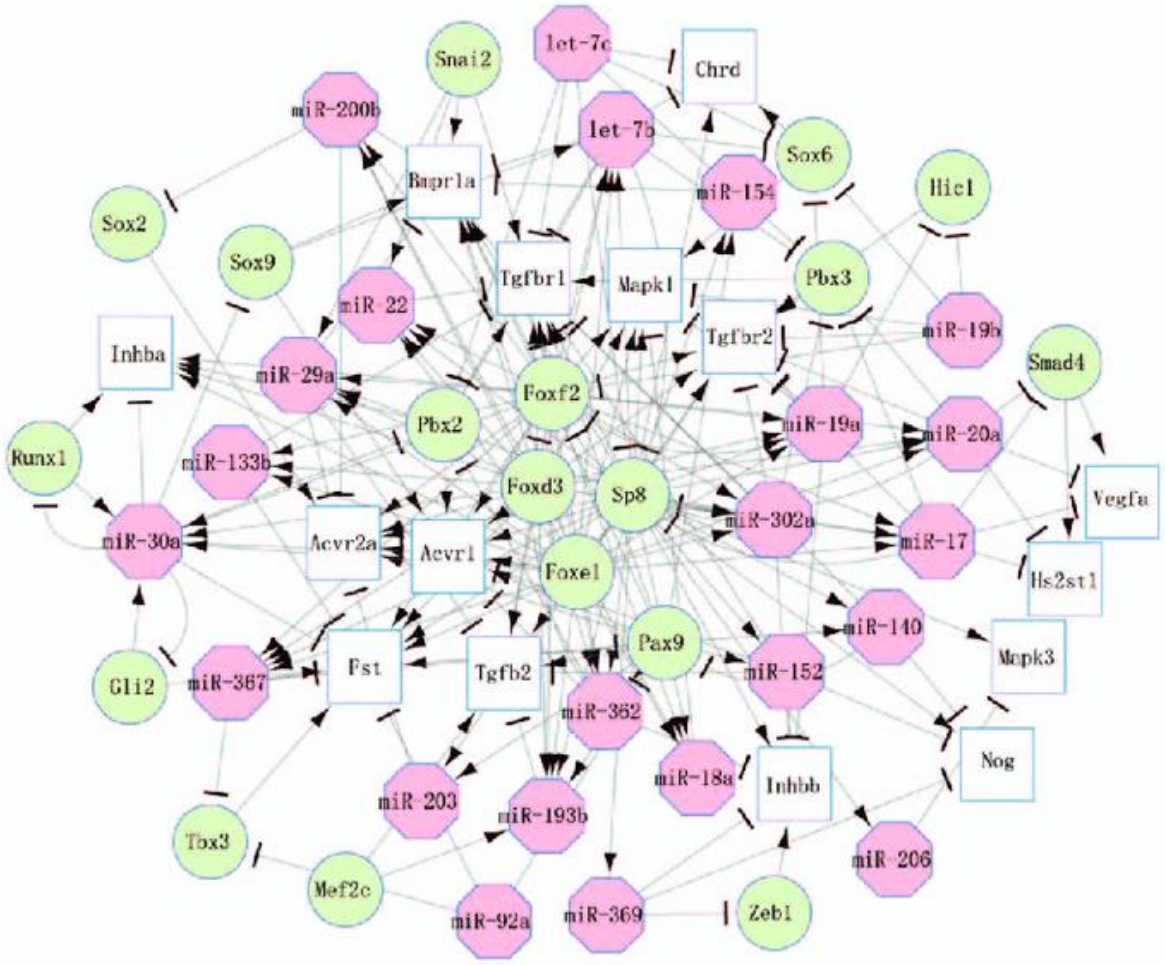


Réseau de régulation génétique (GRN)

- 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

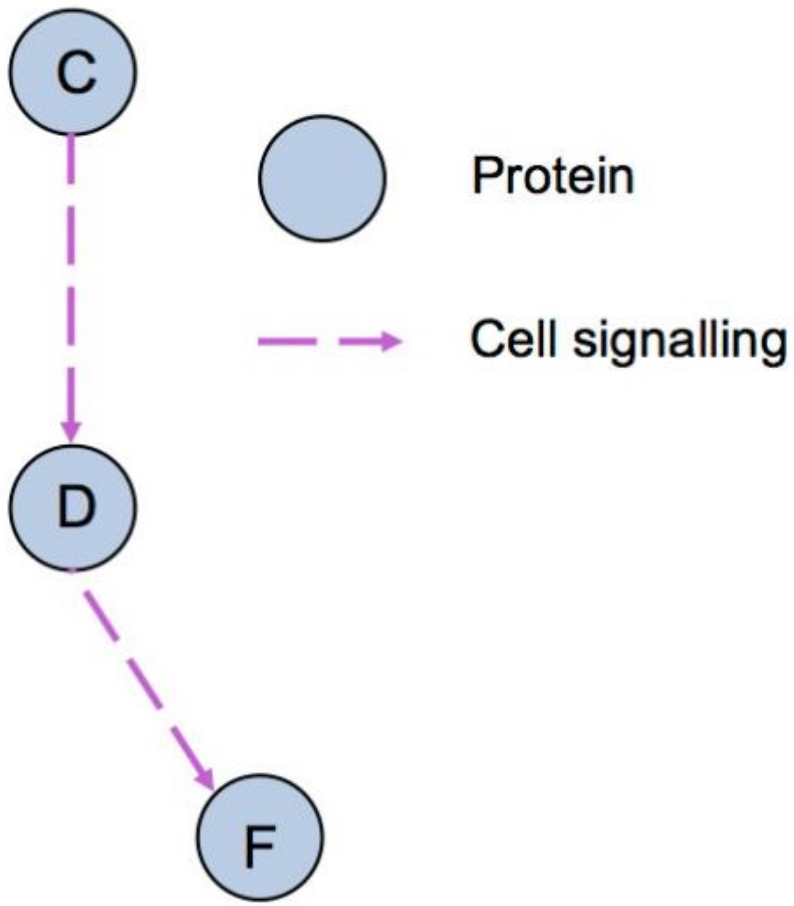
miRNA en rose

TF = facteur de transcription en vert

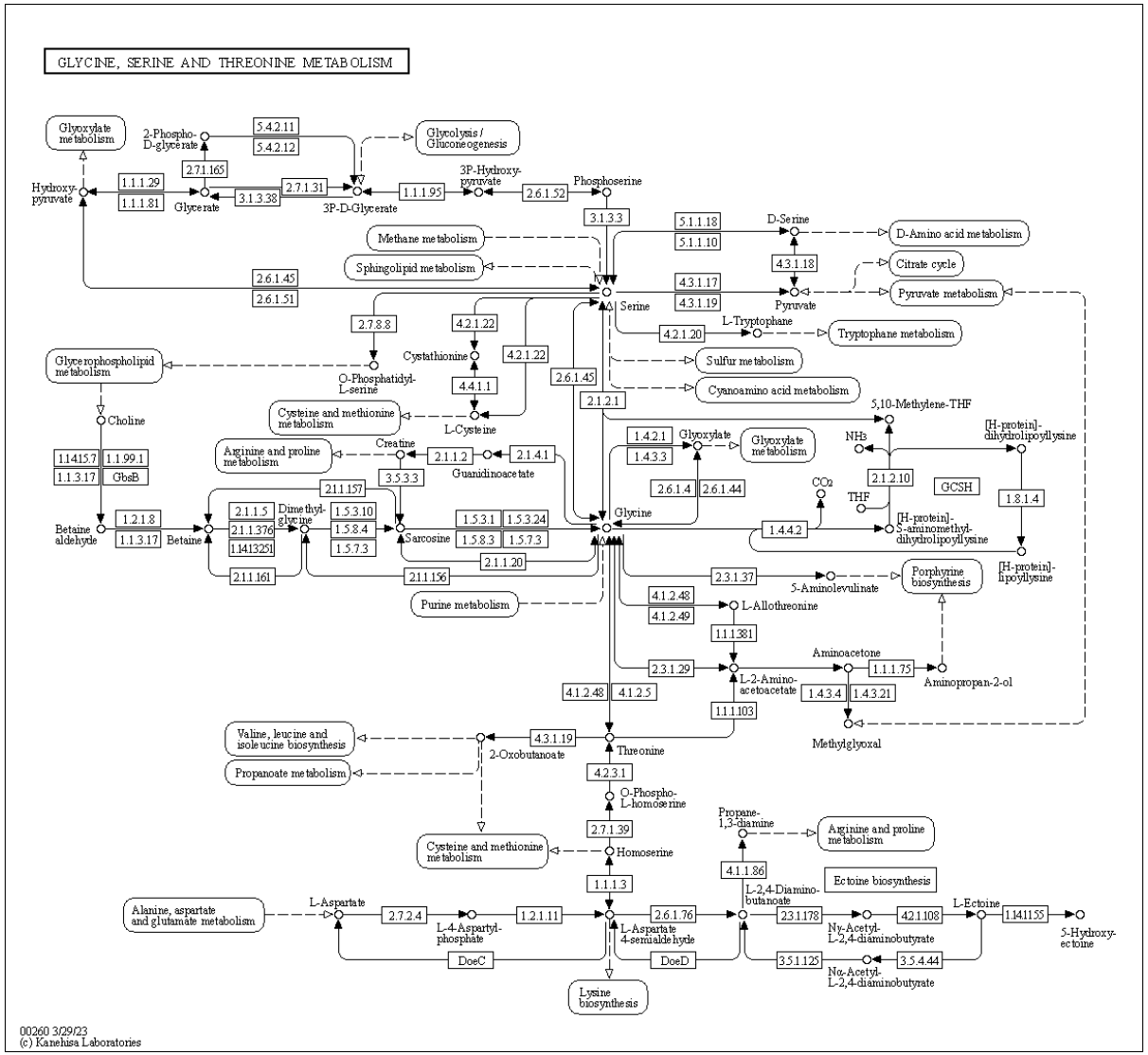
Li et al., Briefings in Bioinformatics, 2019

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

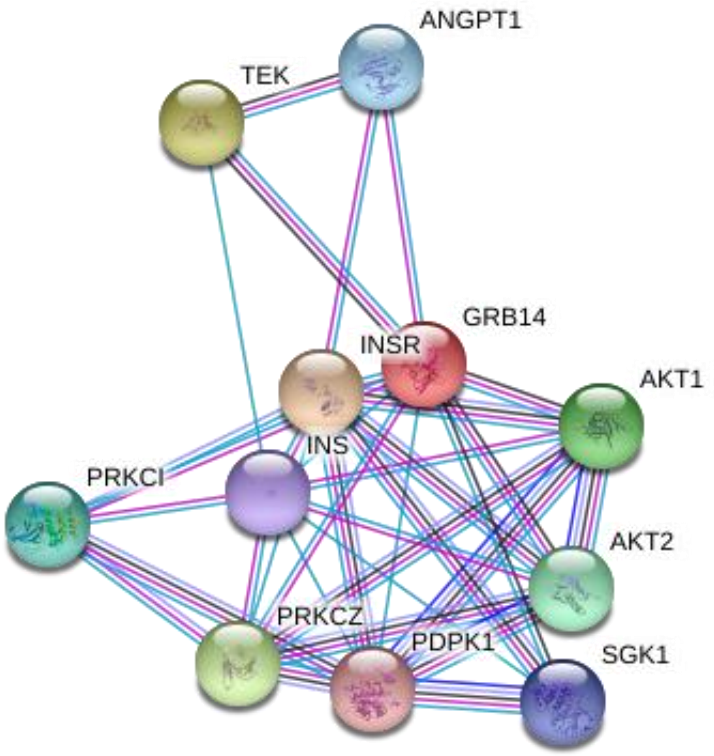
Metabolites = les points

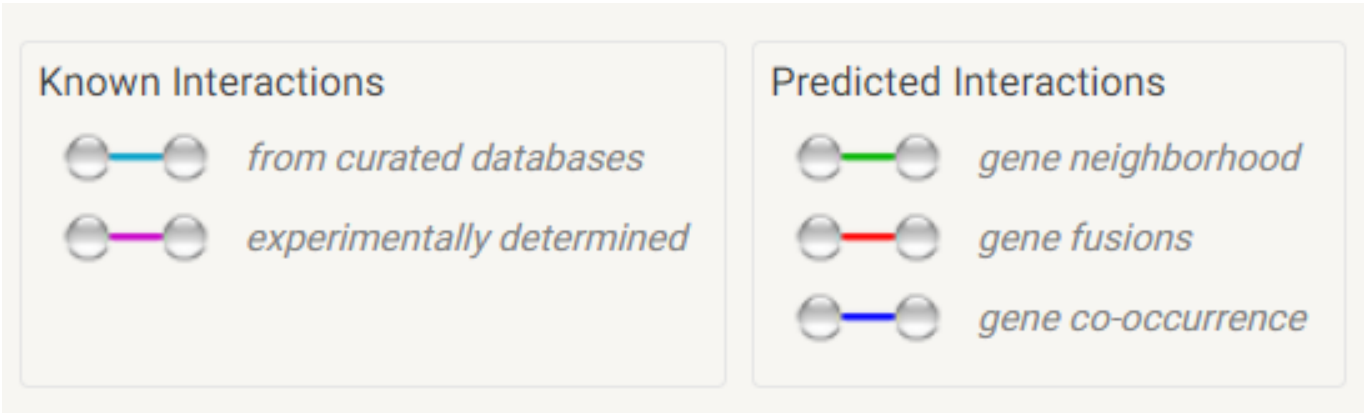
Autres pathway = ellipses

Biologie des systèmes

- L’approche systémique en biologie
- Bioinformatique et données omiques
- **Reconstruire un réseau biologique**
  - Les différents types de réseaux
  - Reconstruction de réseau biologique
  - Les obstacles à la reconstruction
  - Structure des réseaux biologiques
  - Etat de l’art des reseaux biologiques les plus étendus
- Virtual cell et digital twins

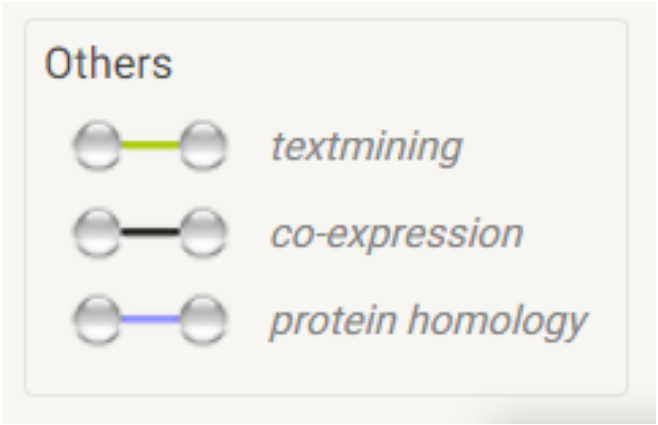
Reconstruction des réseaux biologiques





L'exemple de l'interaction protéines-protéines

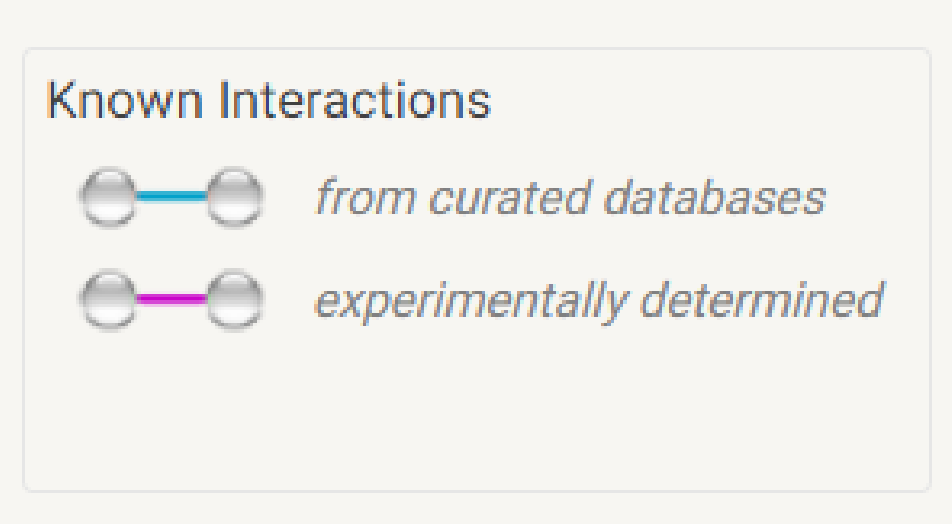
Réseau d'interaction protéine-protéine
de l'insuline prédite par string-db.org



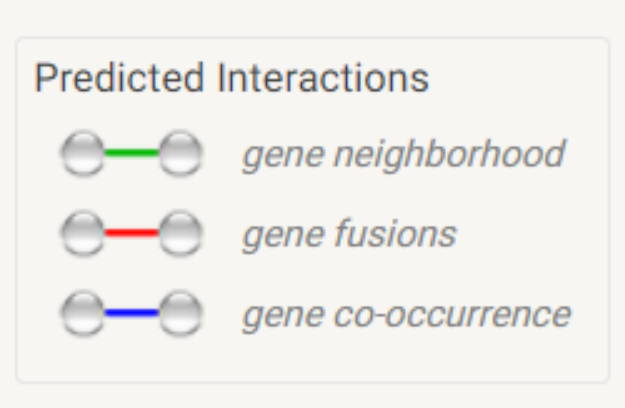
Reconstruction des réseaux biologiques

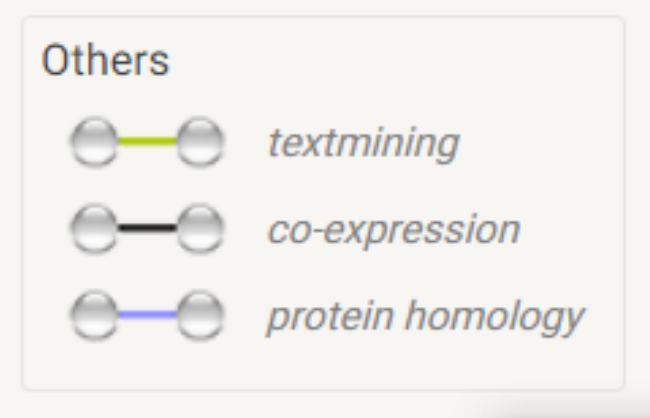
L'exemple d'interaction protéines-protéines

Reconstruction directe



Reconstruction indirecte





Reconstruction directe des réseaux métaboliques

Top-level EC numbers[5]

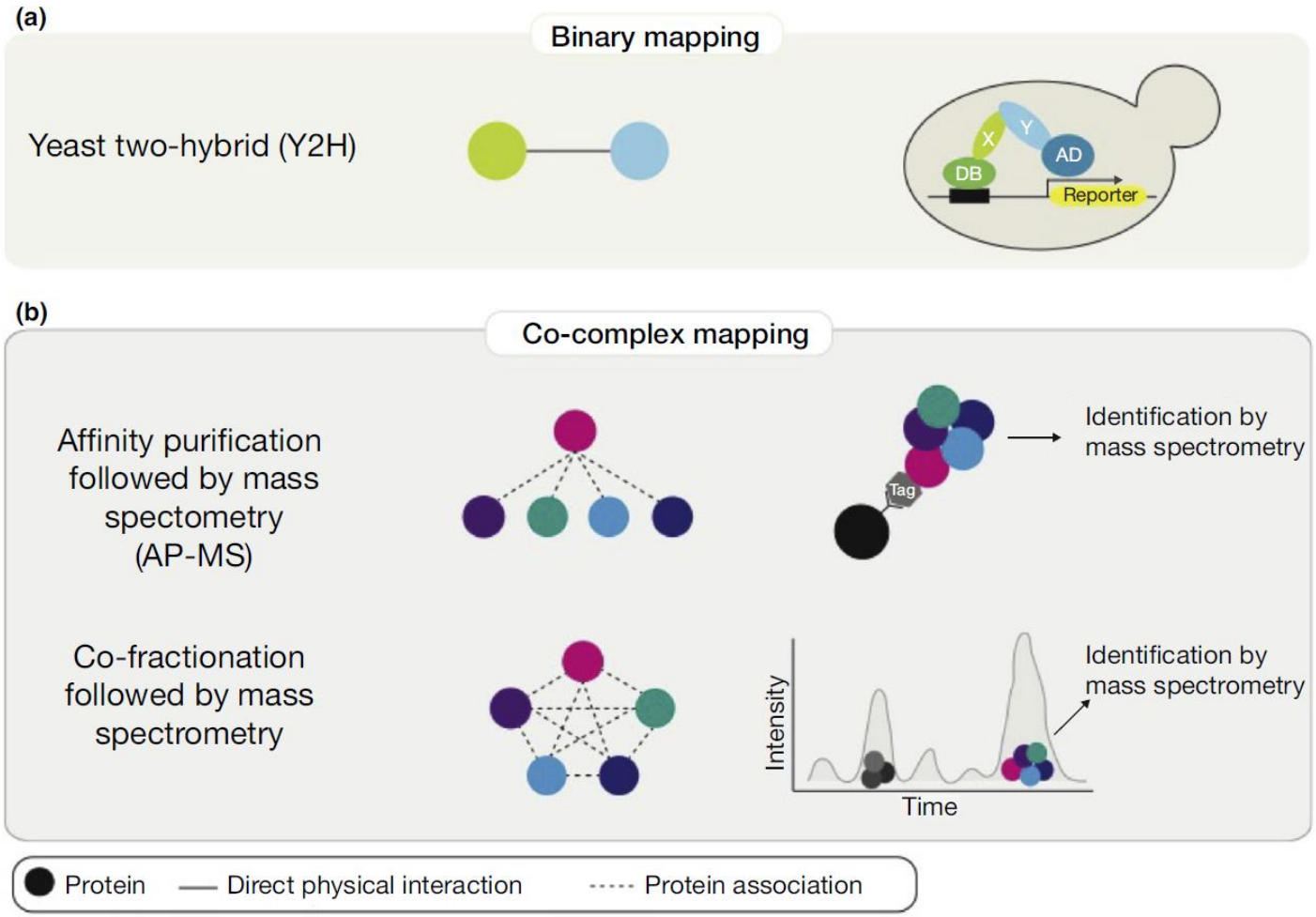
Class	Reaction catalyzed	Typical reaction	Enzyme example(s) with trivial name
EC 1 ***Oxidoreductases***	To catalyze oxidation/reduction reactions; transfer of H and O atoms or electrons from one substance to another	$AH + B \rightarrow A + BH$ (**reduced**) $A + O \rightarrow AO$ (**oxidized**)	Dehydrogenase, oxidase
EC 2 ***Transferases***	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 ***Hydrolases***	Formation of two products from a substrate by hydrolysis	$AB + H_2O \rightarrow AOH + BH$	Lipase, amylase, peptidase, phosphatase
EC 4 ***Lyases***	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 $[X\text{-}A{+}B\text{-}Y] \rightarrow [A{=}B + X\text{-}Y]$	Decarboxylase
EC 5 ***Isomerases***	Intramolecule rearrangement, i.e. isomerization changes within a single molecule	$ABC \rightarrow BCA$	Isomerase, mutase
EC 6 ***Ligases***	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 des interactions protéines-protéines

Base de données : Reactome, PDB, KEGG

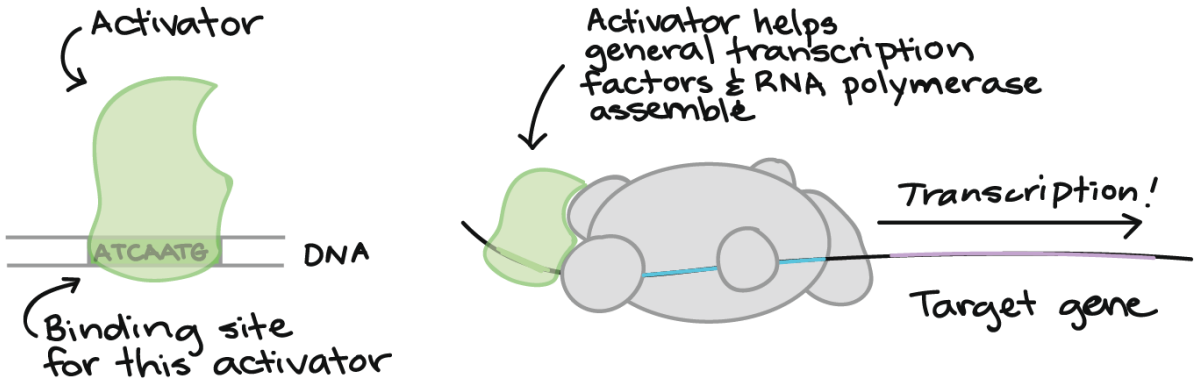


Interaction 2 à 2

Complexe protéique

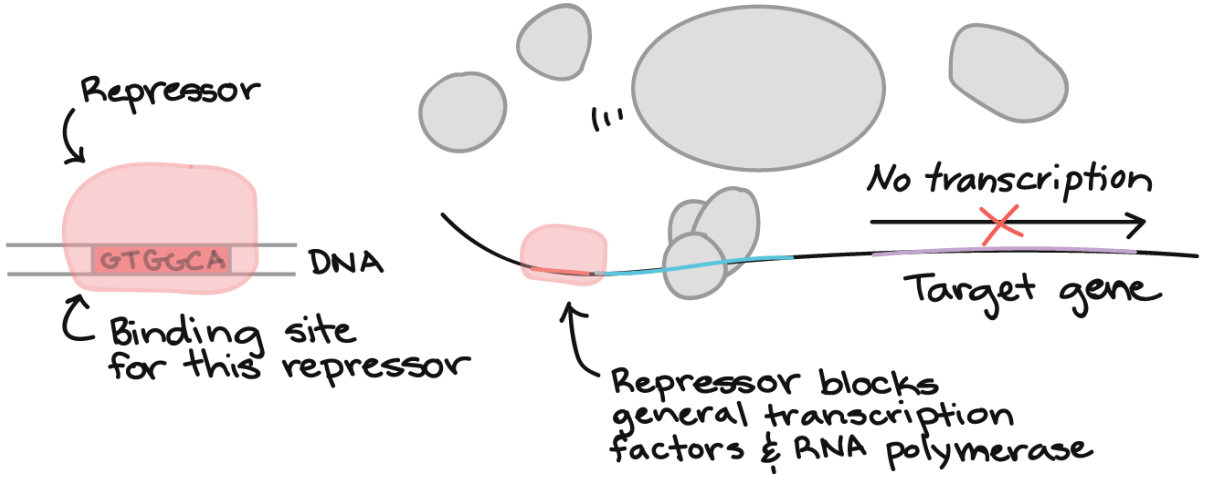
Reconstruction Directe des Gene Regulatory Network

Détection des facteurs de transcription



Activateur

Represseur



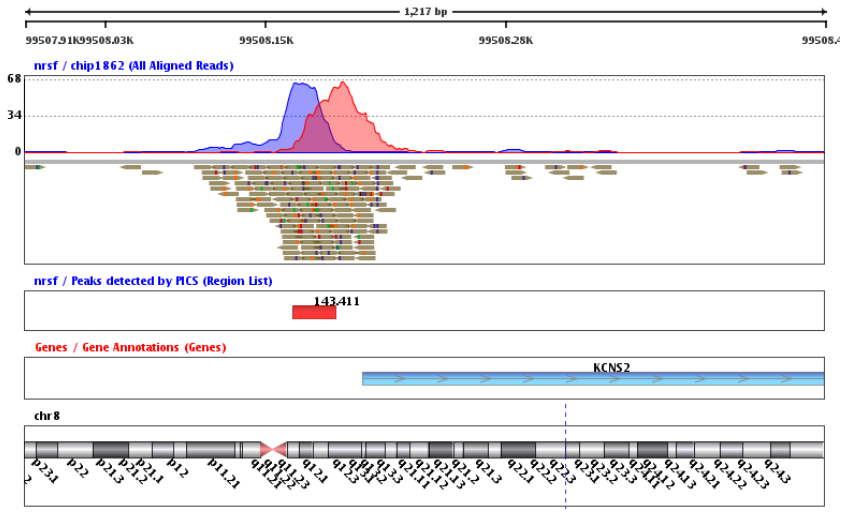
Reconstruction Directe des Gene Regulatory Network

Détection des facteurs de transcription

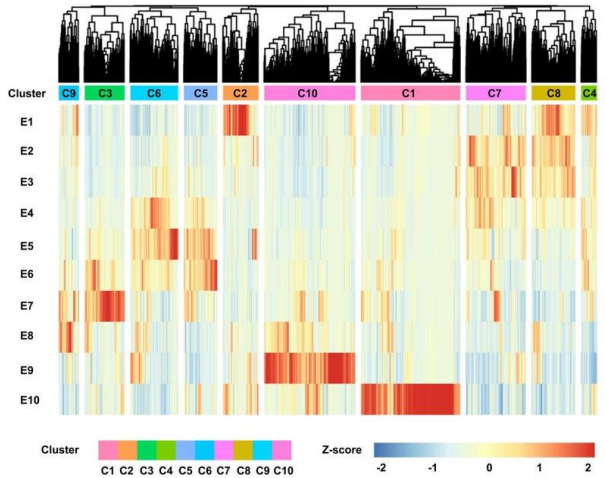
Recherche de motif

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

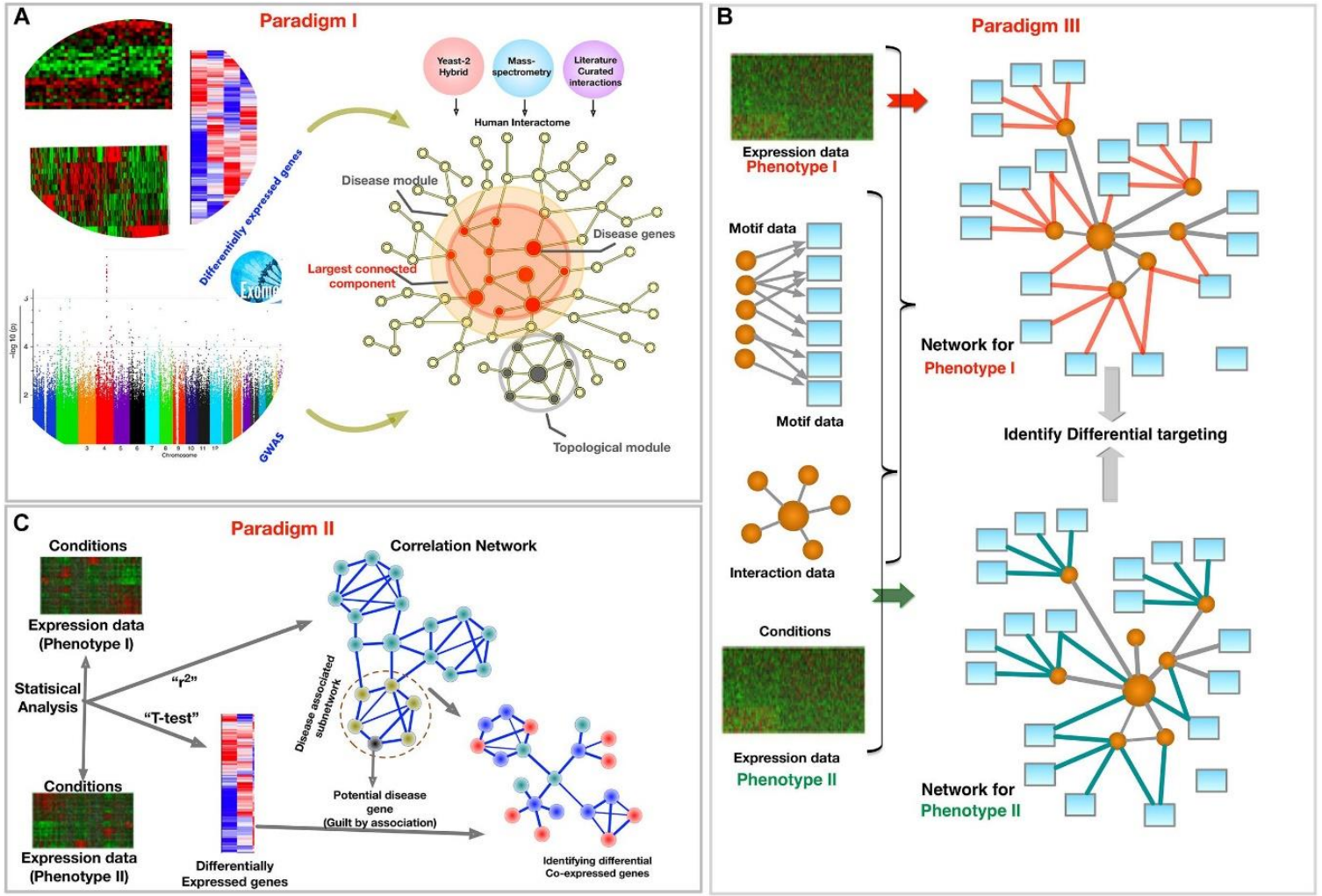
Méthode CHIPSeq



Validation des TF

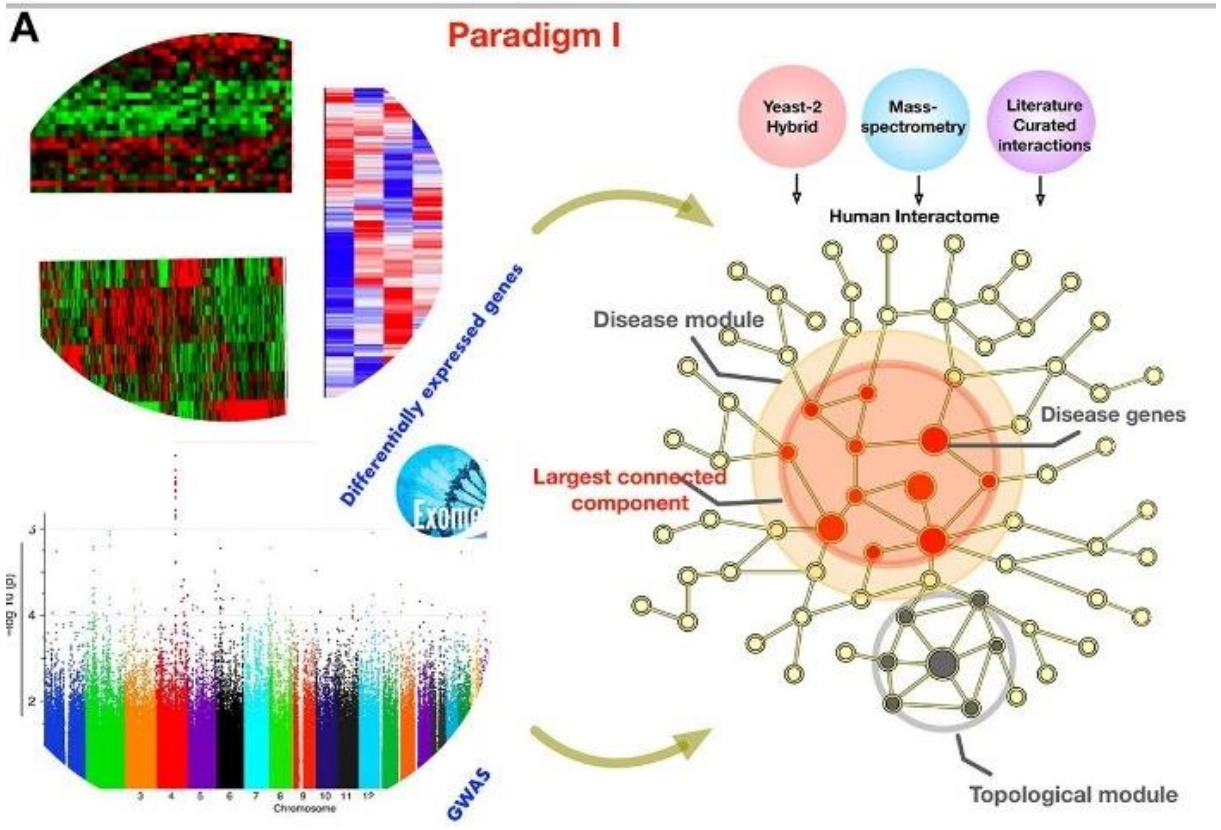


Reconstruction indirecte des réseaux biologiques



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

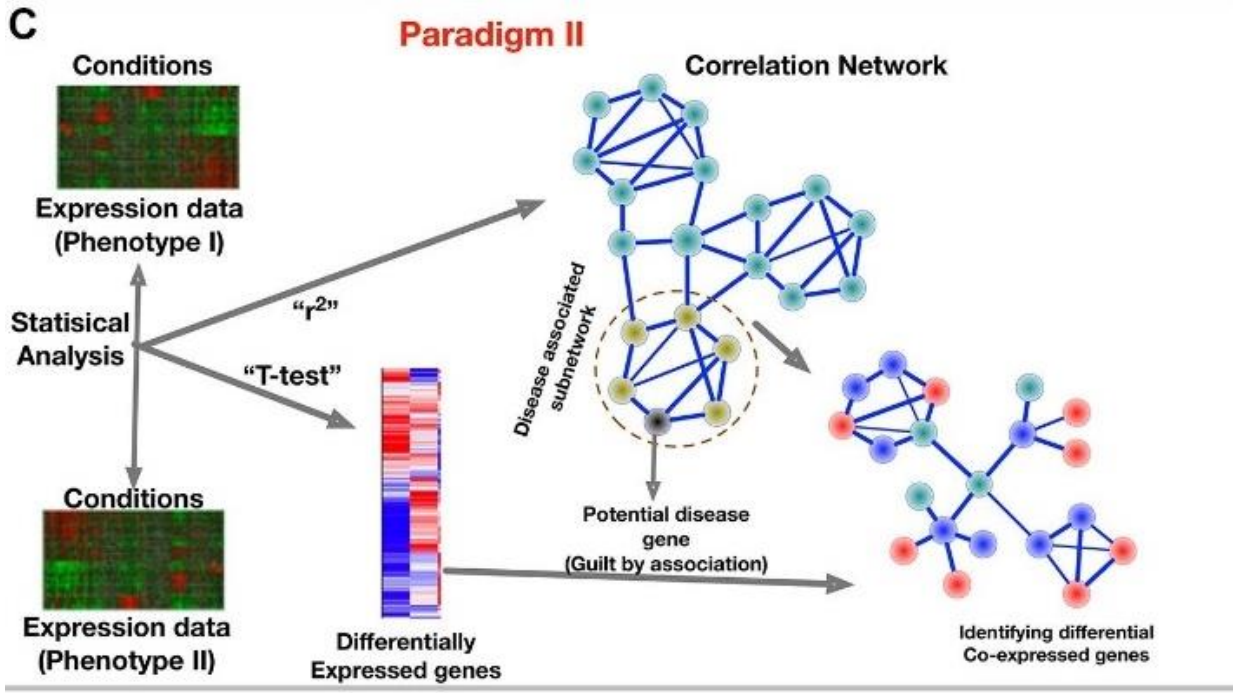
Reconstruction indirecte des réseaux biologiques



Corrélation de données multi-omiques

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

Reconstruction indirecte des réseaux biologiques

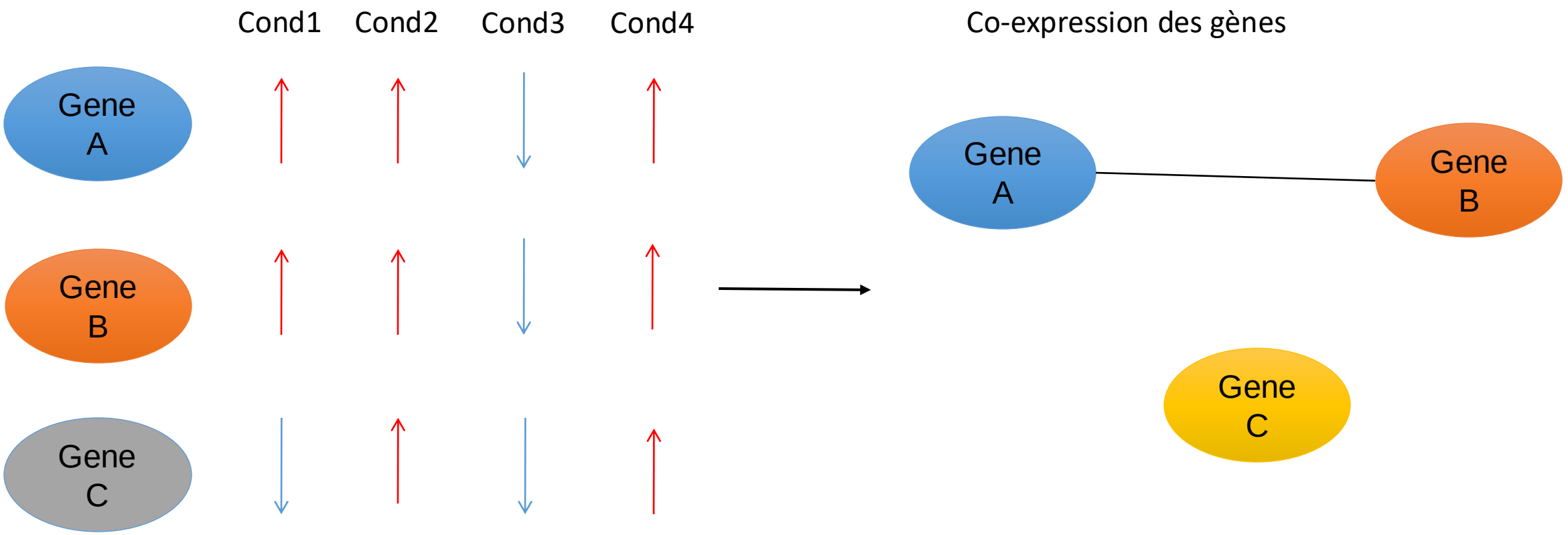


Réseau de co-expression

A partir de données de transcriptomiques on calcule
Une valeure 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

$|r(G_i, G_j)|$

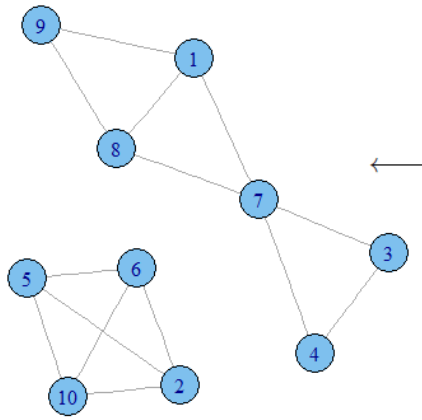
Pearson correlation

	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

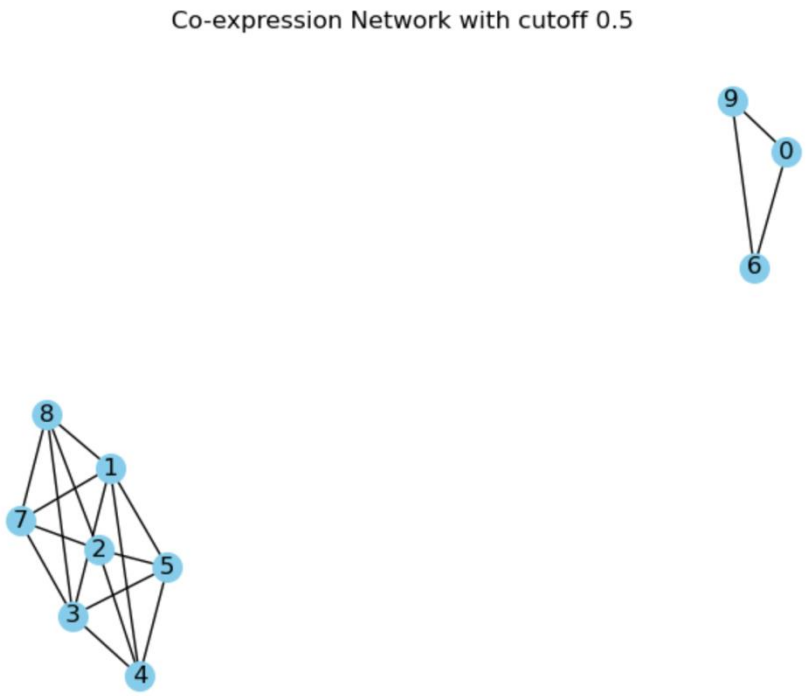
$|r(G_i, G_j)| >= 0.8$

Significance threshold

Matrice d'adjacence genes x genes

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

```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()
```



Gene regulatory network inference

nature reviews genetics

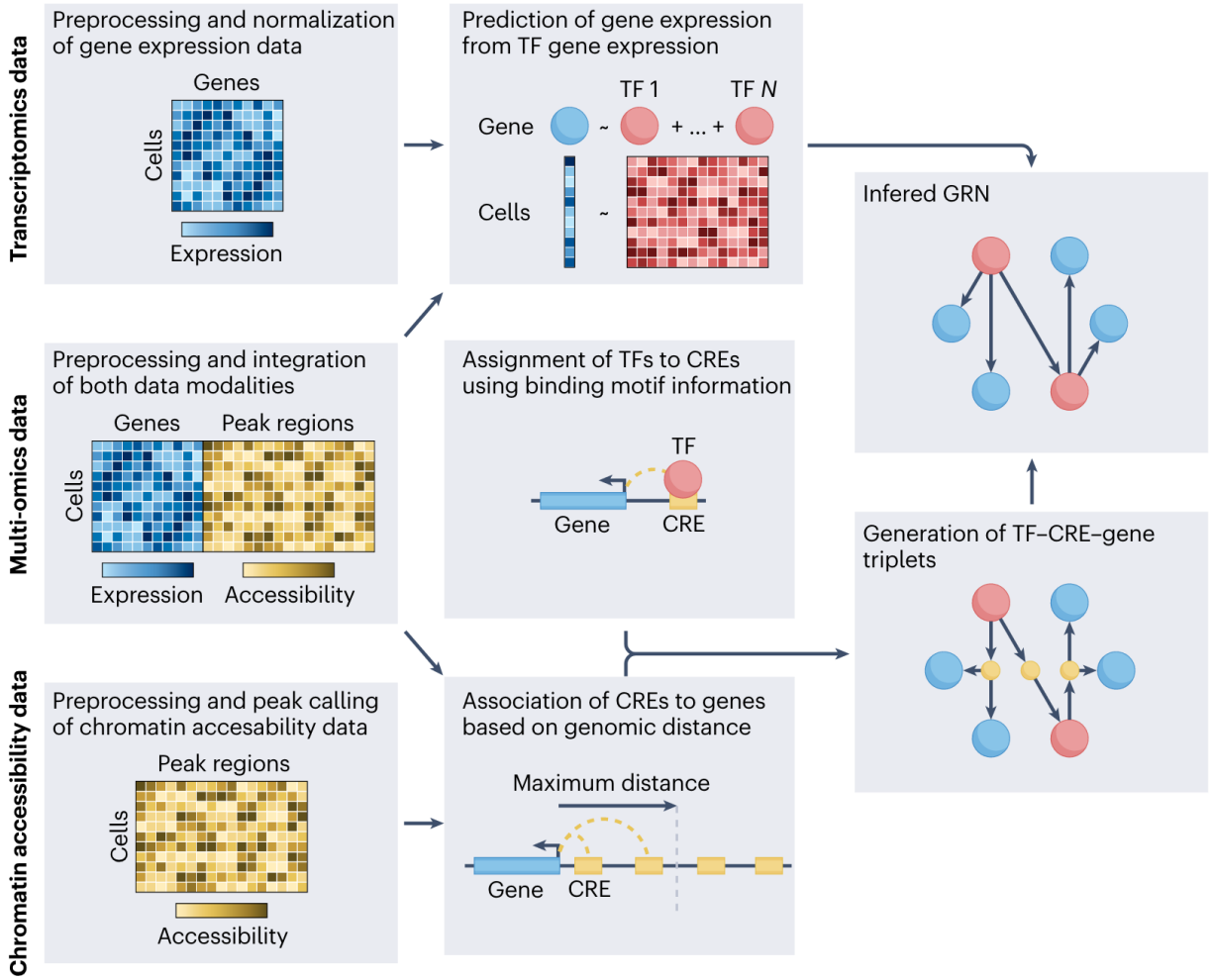
https://doi.org/10.1038/s41576-023-00618-5

Review article

Check for updates

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

Pau Badia-i-Mompel[1], Lorna Wessels[1,2], Sophia Müller-Dott[1], Rémi Trimbour[1,3], Ricardo O. Ramirez Flores[1], Ricard Argelaguet[4] & Julio Saez-Rodriguez[1]



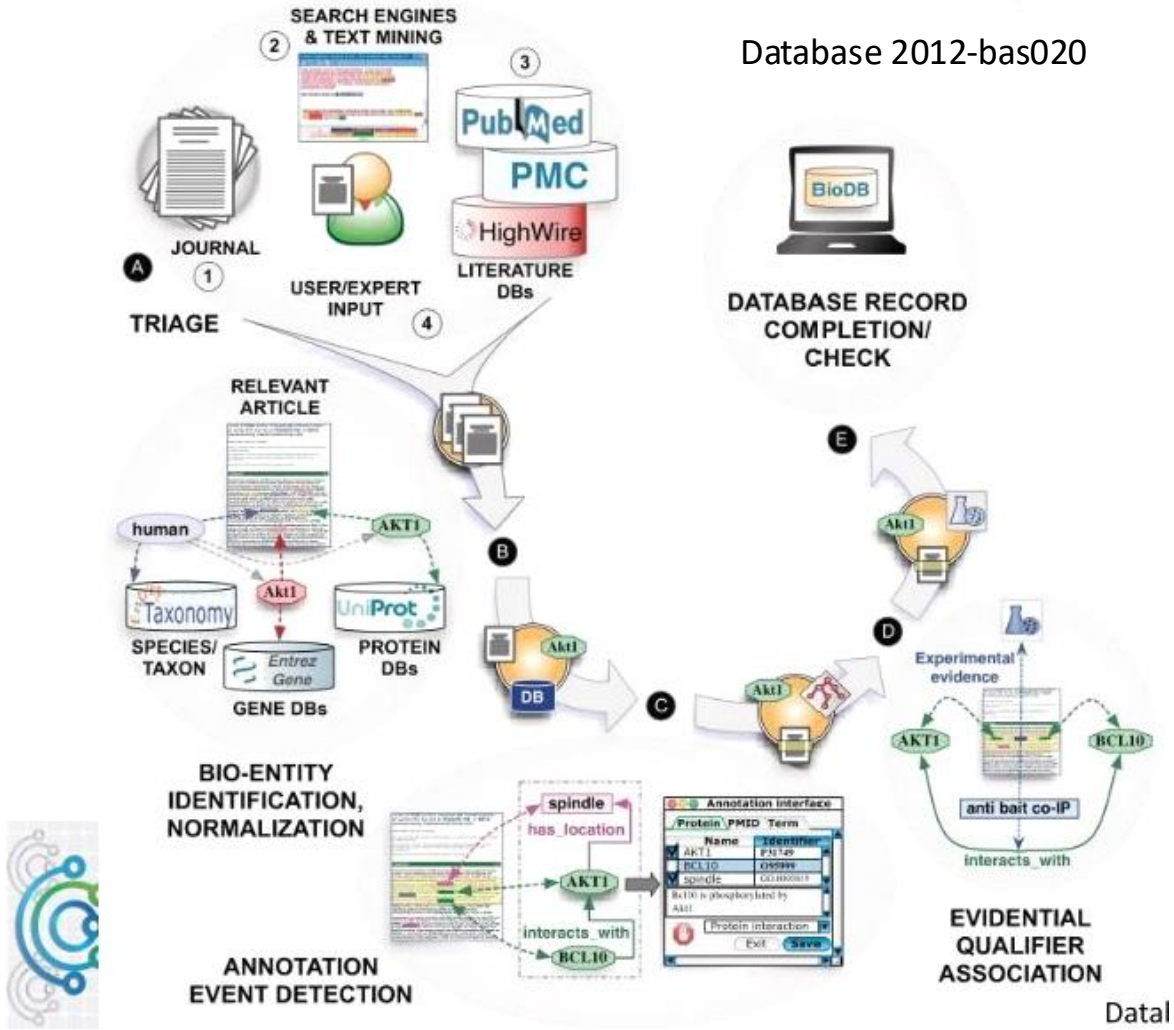
Reconstruction à l'aide d'information génétique

- 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

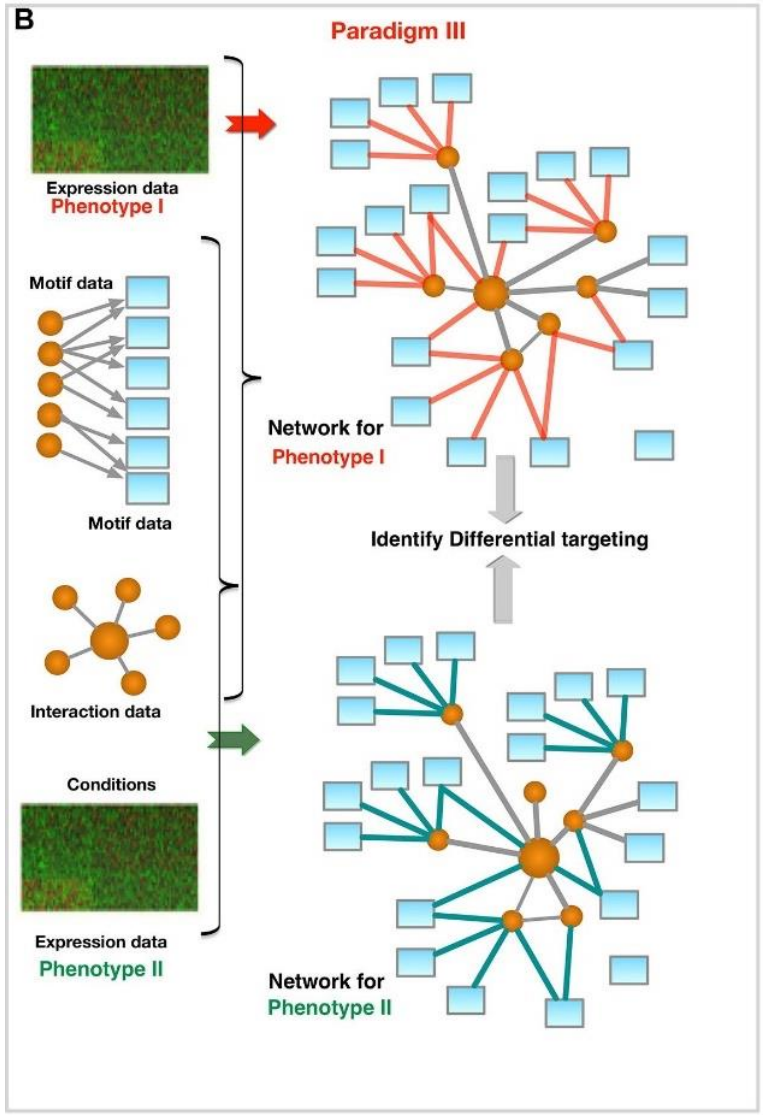
- Extraire les mots clés dans les publications et bases de données
- Nettoyer la liste de mots clés = Molécules, gènes, protéines
- Annoter ces élèments avec informations disponibles dans bases de données
- Chercher les co-occurrence dans les publications
- Mettre à jour les bases de données avec ces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



Database 2012-bas020

Datal

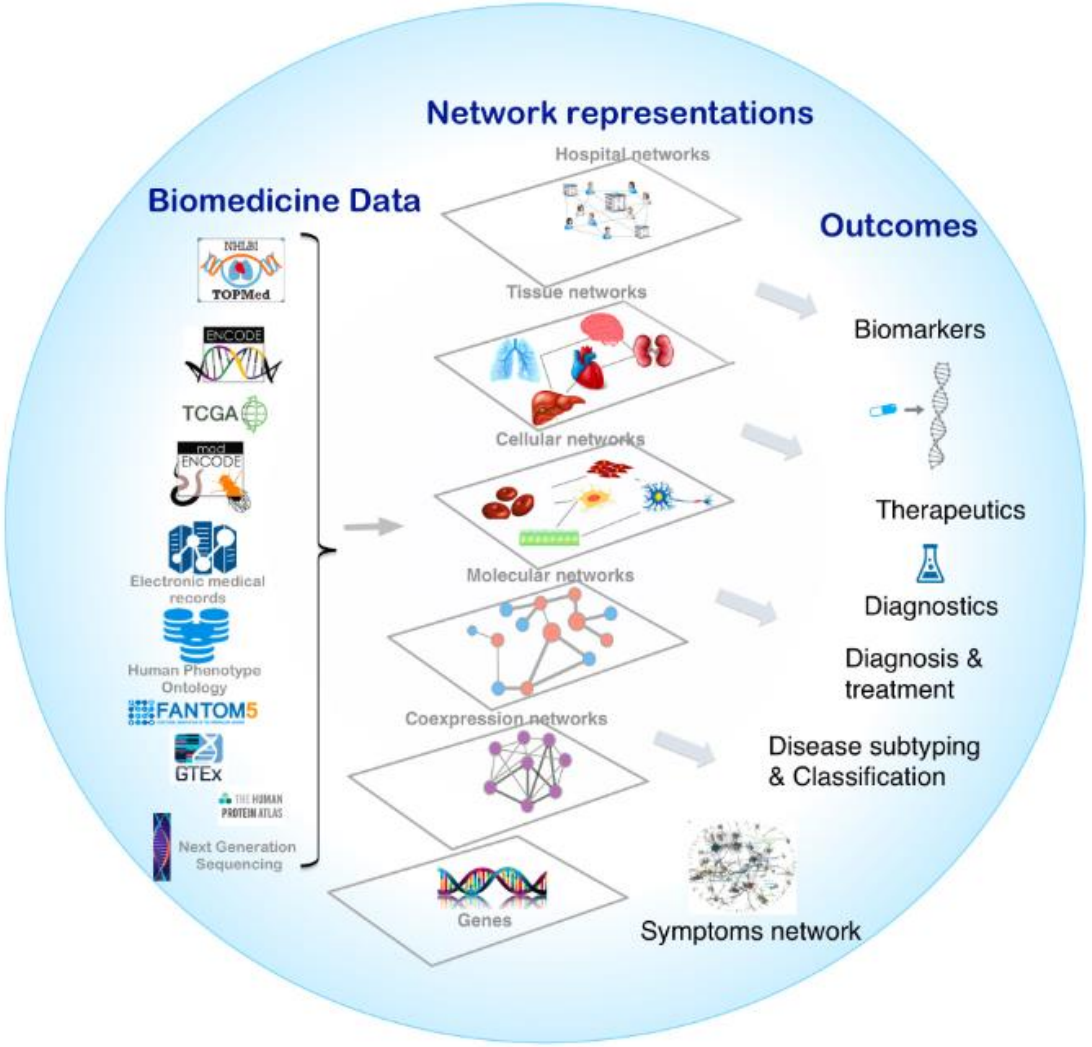
Reconstruction indirecte des réseaux biologiques



Corrélation de différents réseaux biologiques

On peut par exemple corréler 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 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

Biologie des systèmes

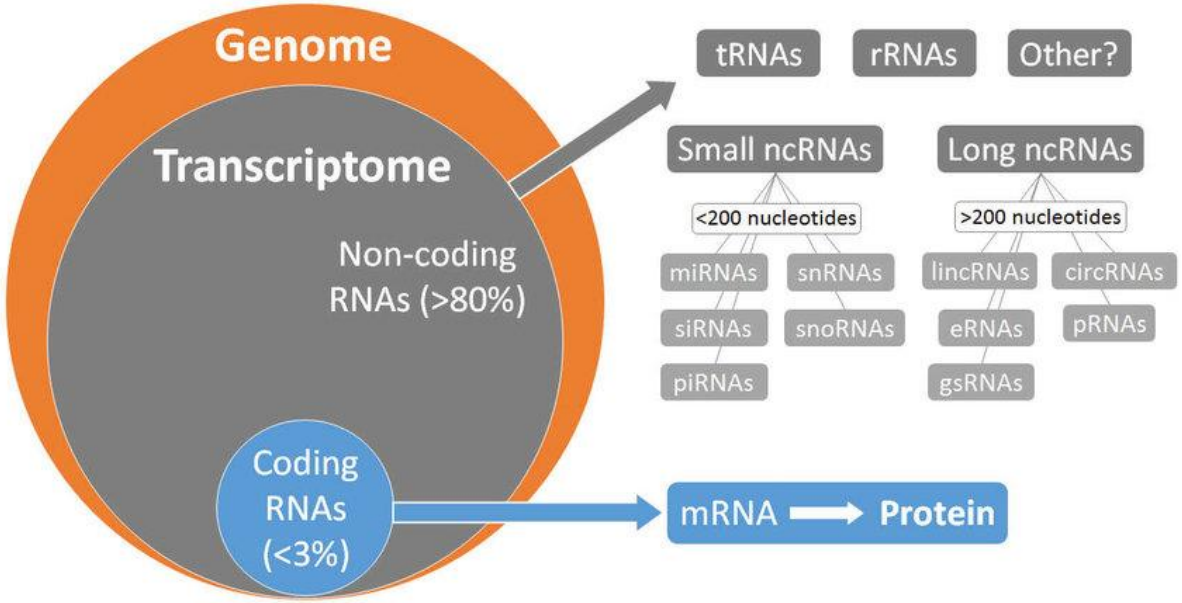
- L’approche systémique en biologie
- Bioinformatique et données omiques
- **Reconstruire un réseau biologique**
  - Les différents types de réseaux
  - Reconstruction de réseau biologique
  - Les obstacles à la reconstruction
  - Structure des réseaux biologiques
  - Etat de l’art des reseaux biologiques les plus étendus
- Virtual cell et digital twins

Les obstacles à la reconstruction

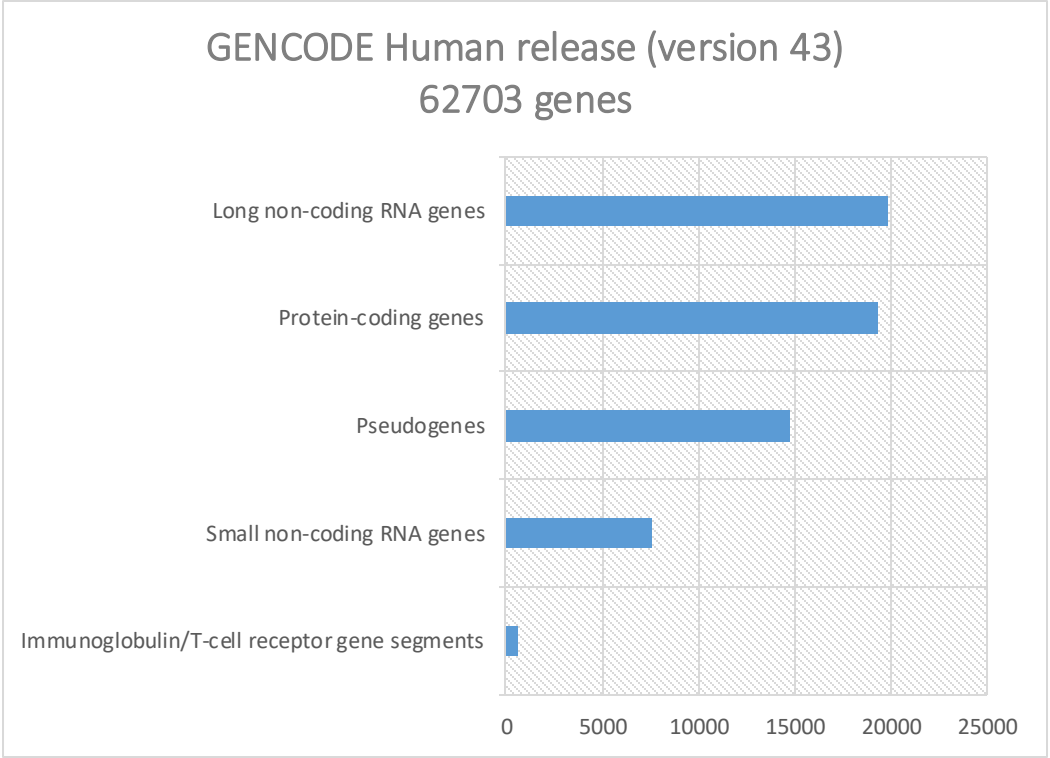
- Disponibilité de l'annotation (=les nœuds du réseau)
- Disponibilité des données omiques pour reconstruire les arêtes
- Biais de mesures
- Biais de connaissances
- Impossibilité d'avoir une vision en « instantanée » d'un organisme

Annotation des génomes

Pedrosa et al., The CardioRNA COST Action 2019



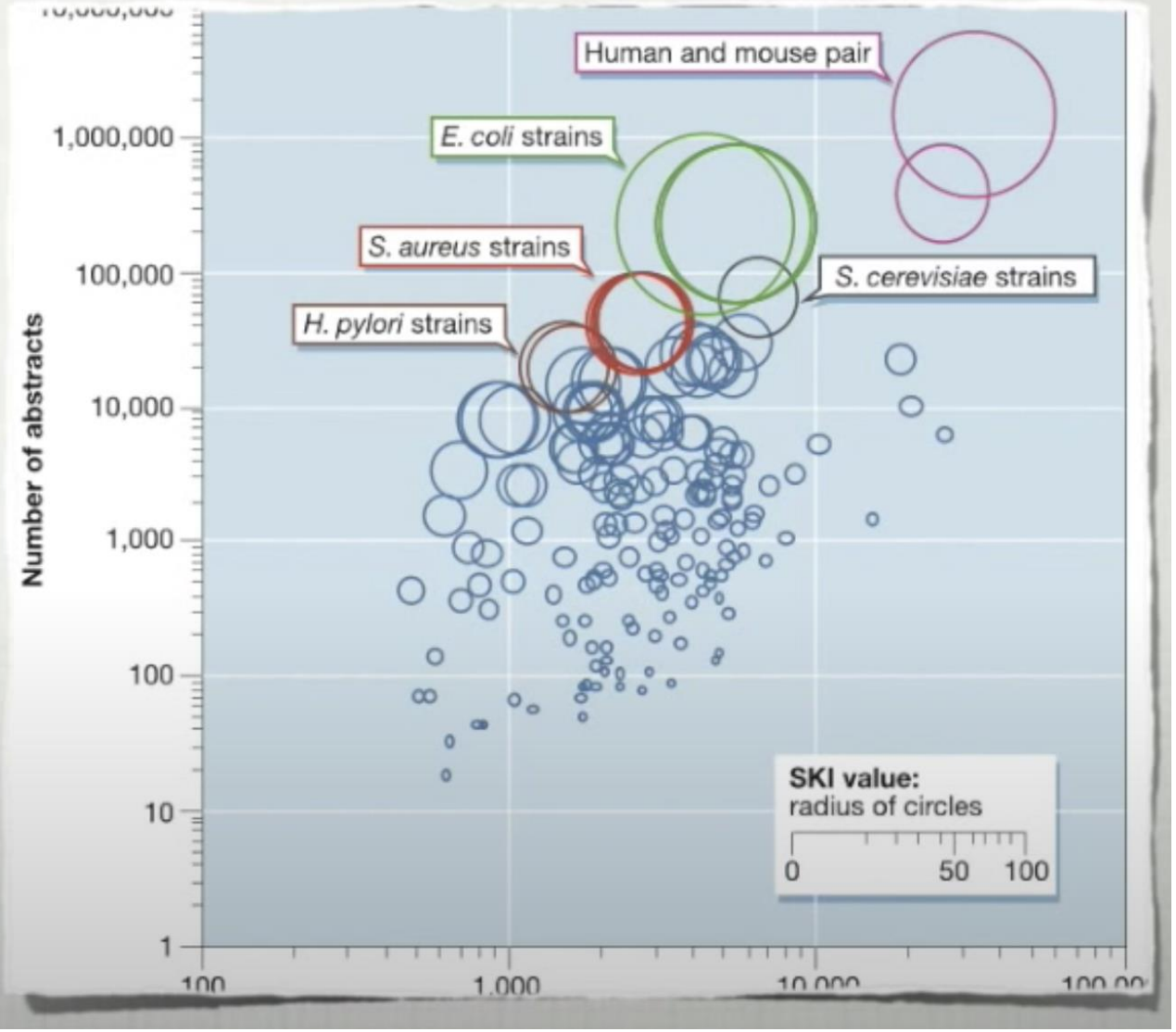
GENCODE 21, Frankish et al., N.A.R. 2021



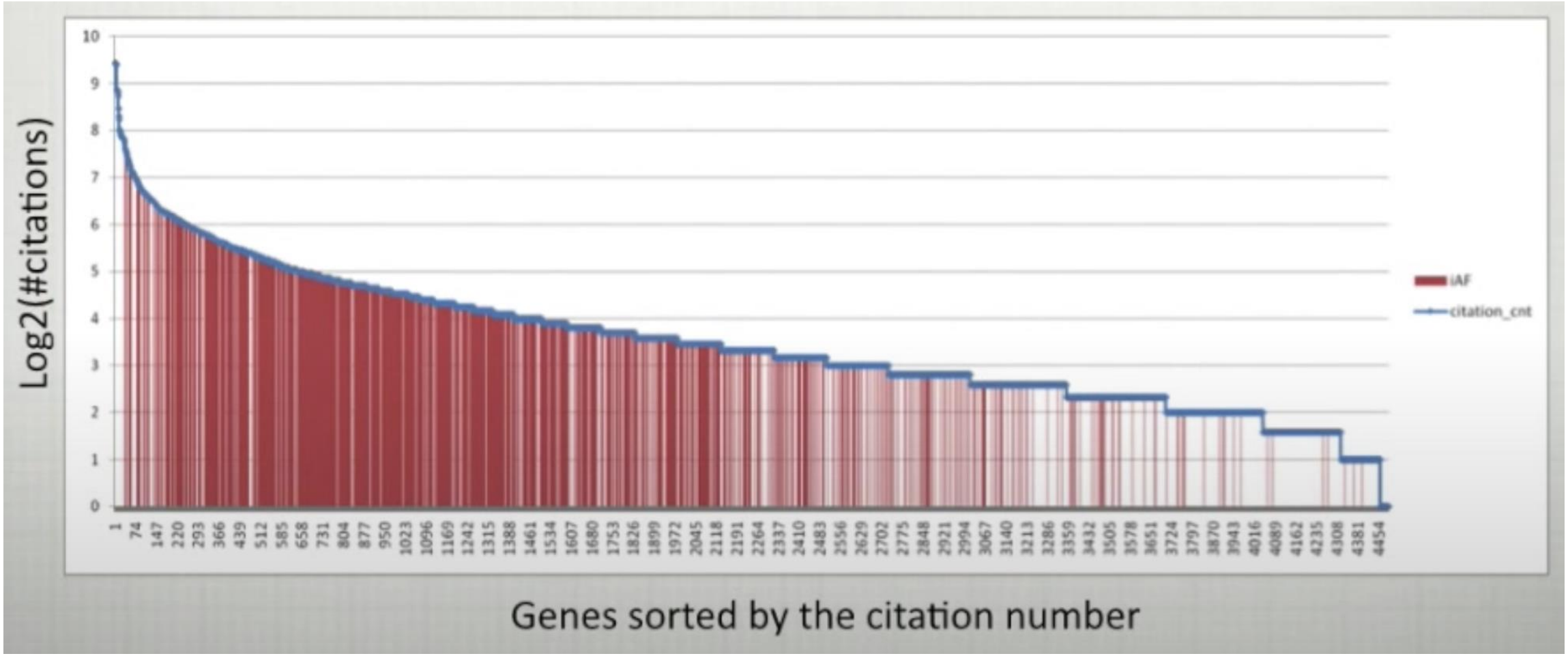
Biais de connaissance dans l'ensemble des gènes

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

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



Biais de connaissance dans l’ensemble des gènes

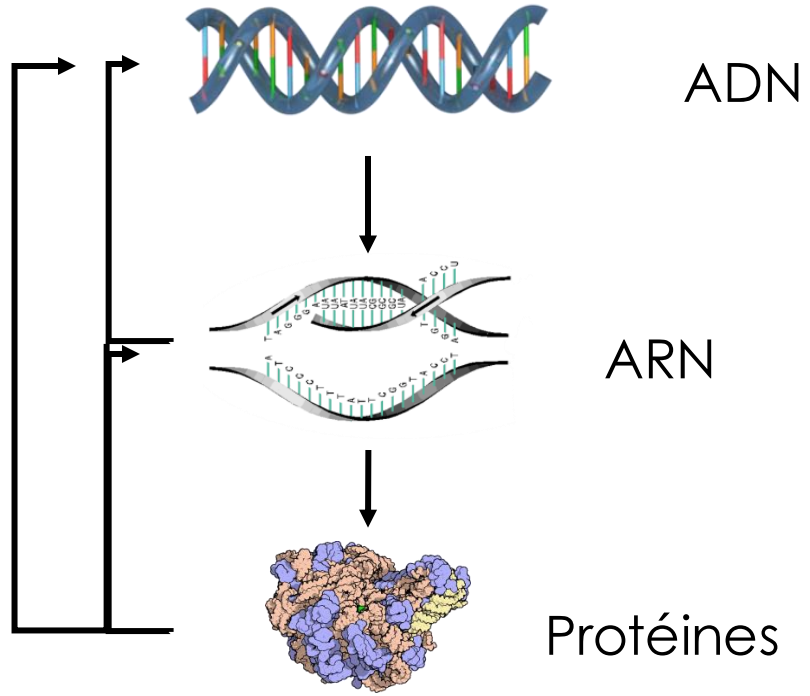


Les différences d'échelles temporelles

Division, Replication, Transcription, Translation & Degradation Rates
at 37°C with a temperature dependence Q10 of ≈2-3

9. Cell cycle time (exponential growth in rich media): *E. coli* ≈20-40 min; yeast 70-140 min; human cell line (Hela): 15-30 hours
10. Rate of replication by DNA polymerase *E. coli* ≈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 ≥ cell cycle time

B10NUMB3R5
THE DATABASE OF USEFUL BIOLOGICAL NUMBERS



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 detectées
- 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 ?

Multi-omique

Opinion | Published: 29 August 2003

Comparing protein abundance and mRNA expression levels on a genomic scale

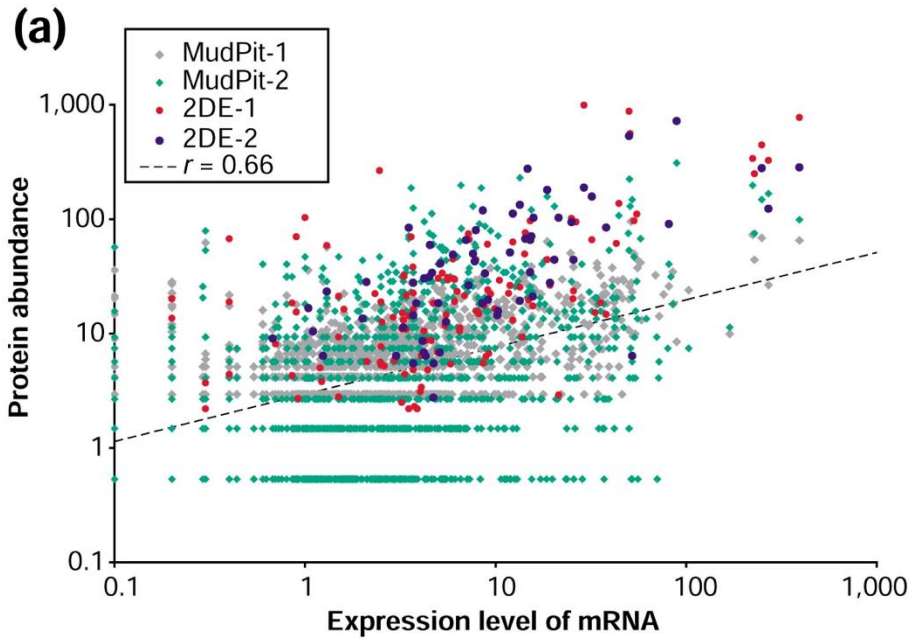
Dov Greenbaum, Christopher Colangelo, Kenneth Williams & Mark Gerstein

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

67k Accesses | **1144** Citations | **4** Altmetric | Metrics

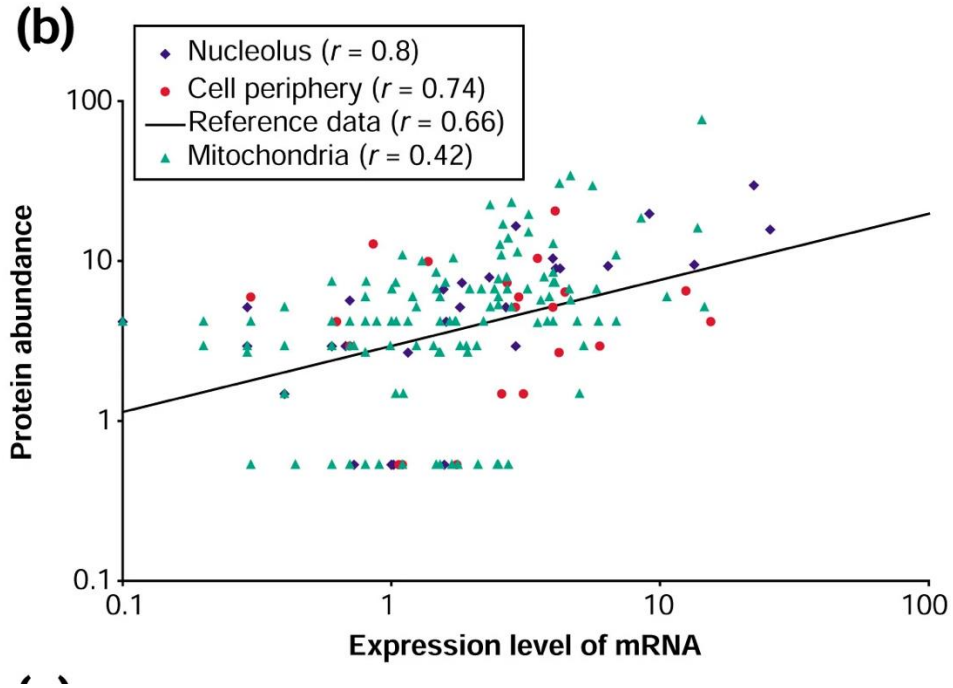
R = 0.66

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



0.42 < R <0.8

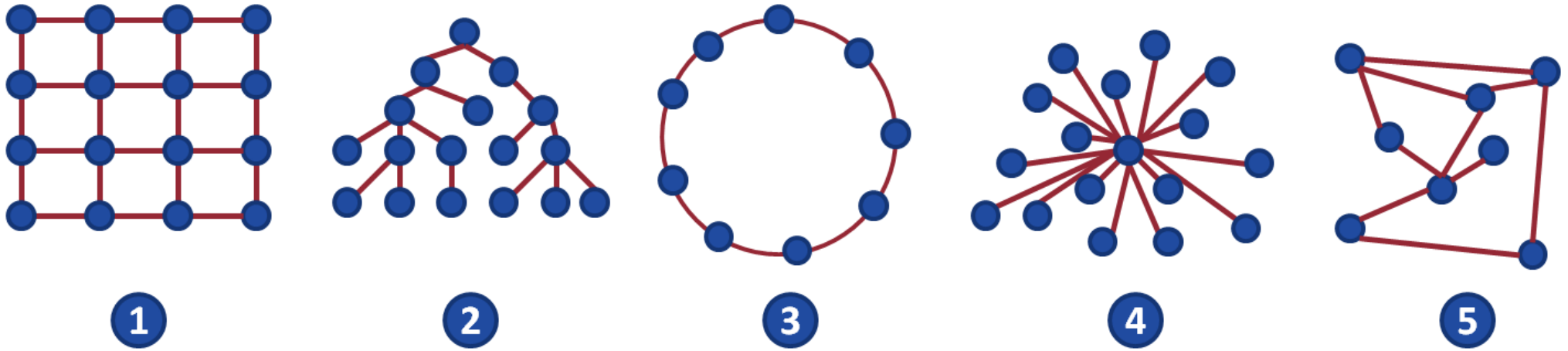
Pour certains groupes GO la corrélation augmente



Biologie des systèmes

- L’approche systémique en biologie
- Bioinformatique et données omiques
- **Reconstruire un réseau biologique**
  - Les différents types de réseaux
  - Reconstruction de réseau biologique
  - Les obstacles à la reconstruction
  - Structure des réseaux biologiques
  - Etat de l’art des reseaux biologiques les plus étendus
- Virtual cell et digital twins

Topologie des réseaux



Graphe homogènes

Graphe hiérarchique

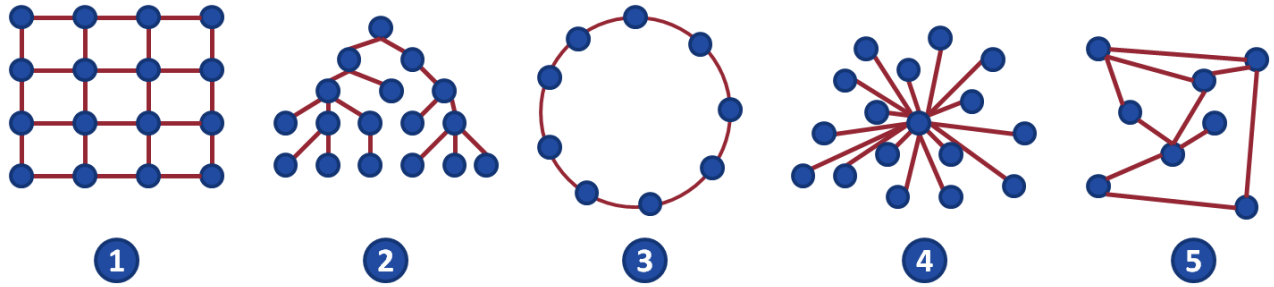
Graphe cyclique

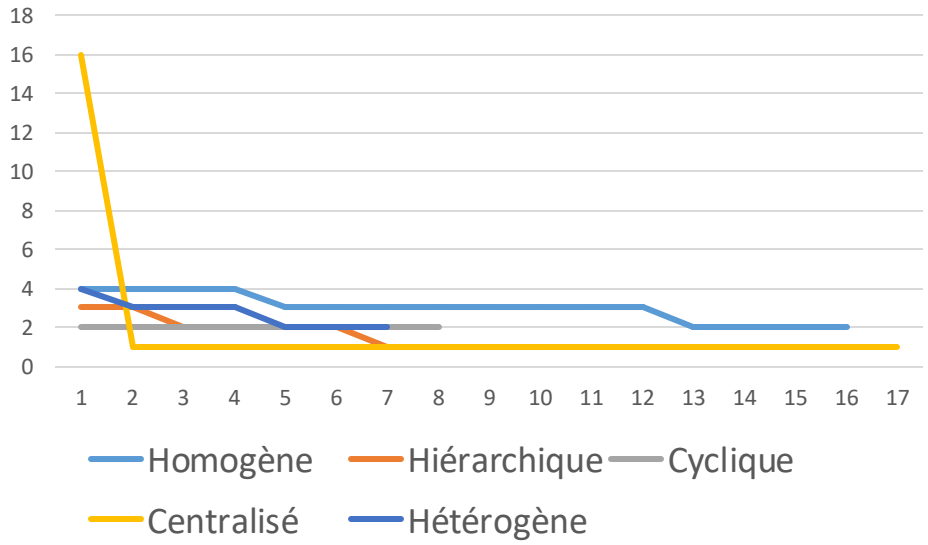
Graphe centralisé

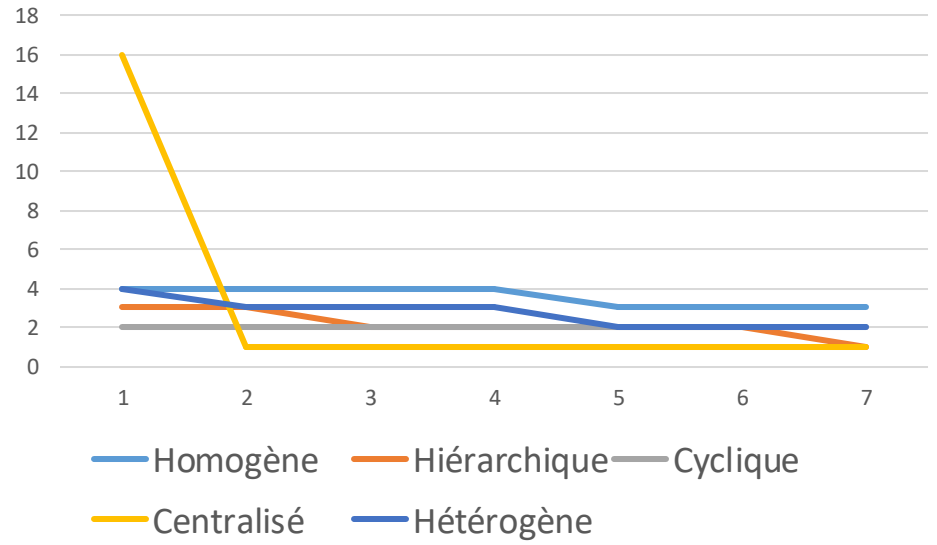
Graphe hétérogène

Mesure de la topologie

distribution des degrés







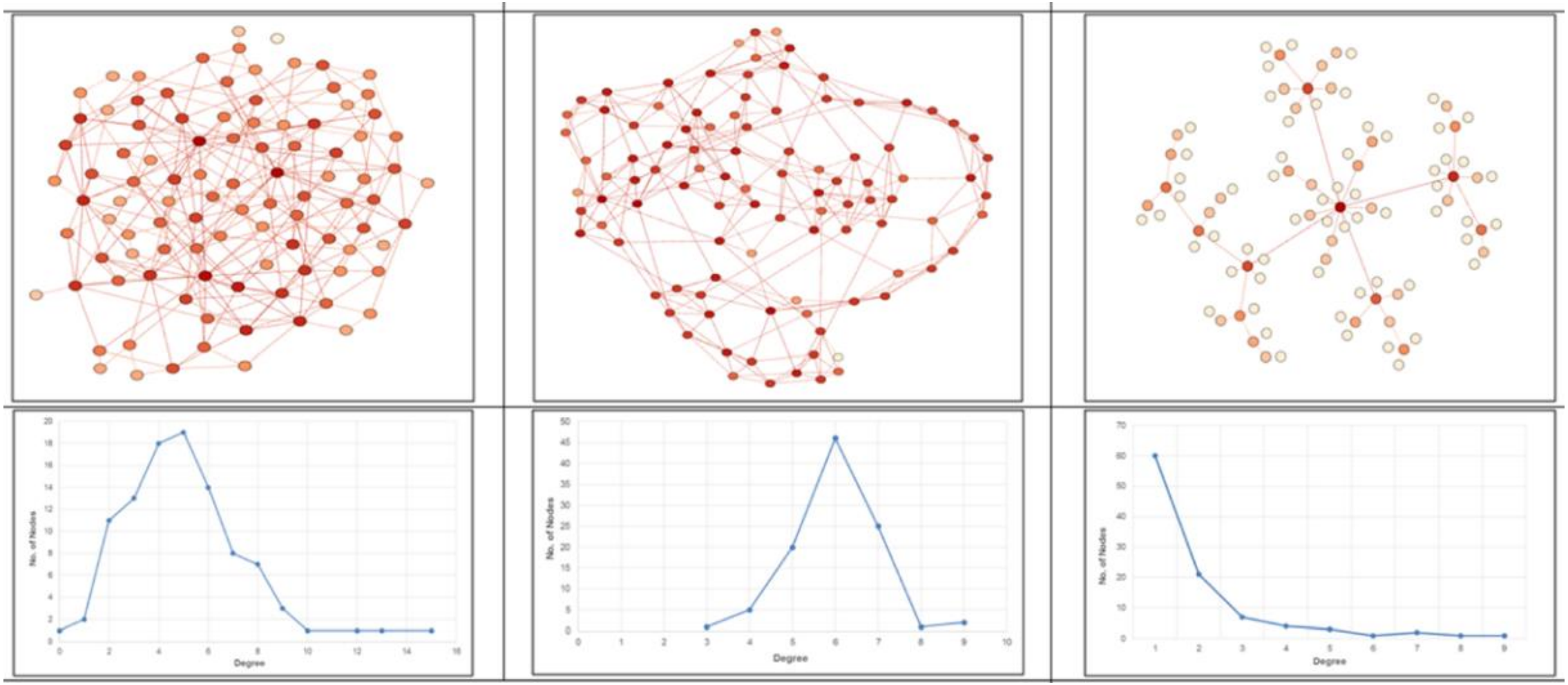
Mesure de la topologie

distribution des degrés

Réseau aléatoire

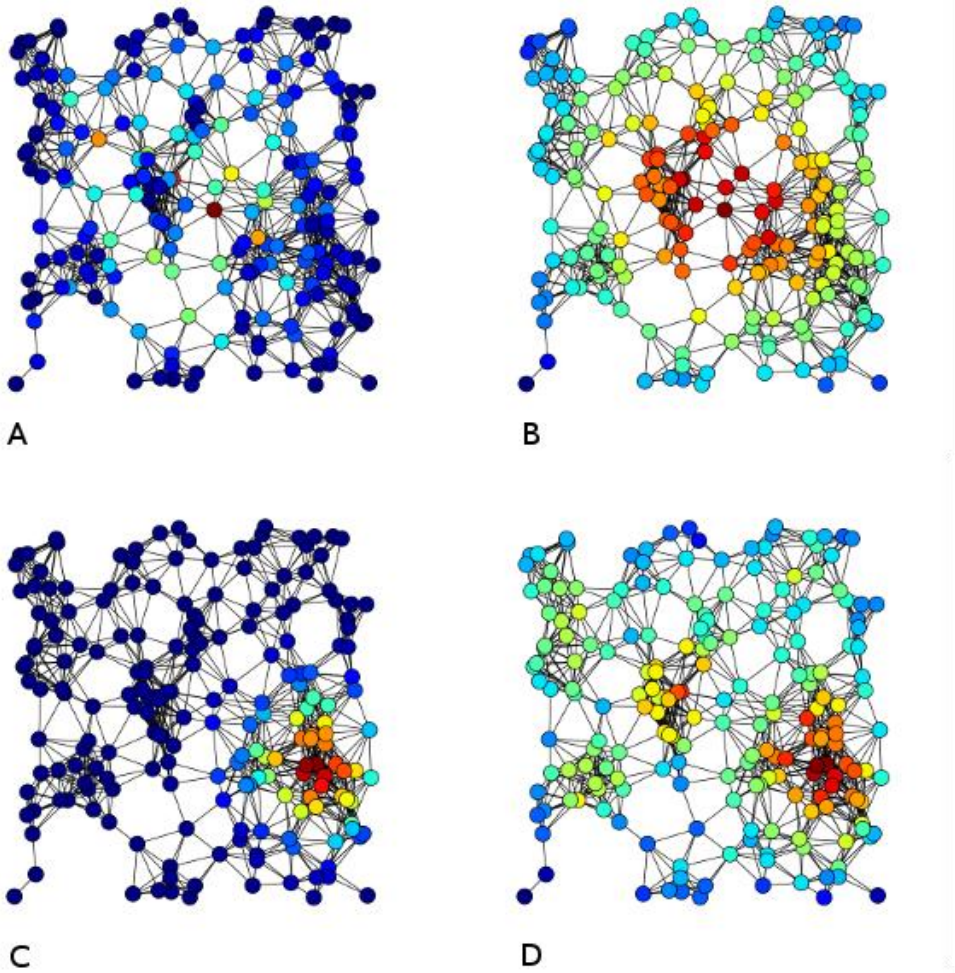
Réseau small world

Réseau centralisé
(sans échelles)



Mesure de la topologie

Notion de centralité



A - Centralité d'intermédiarité

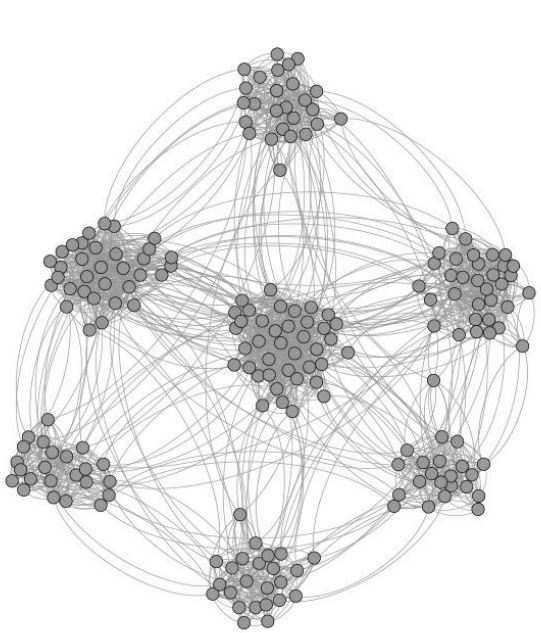
B - Centralité de proximité

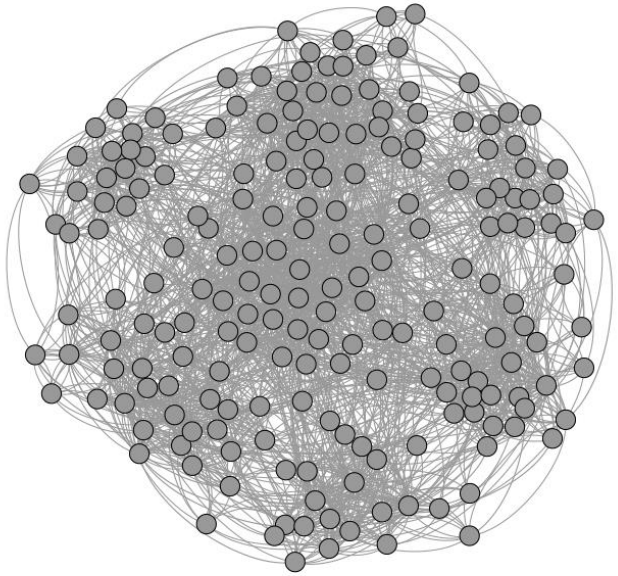
C - Centralité de vecteur propre

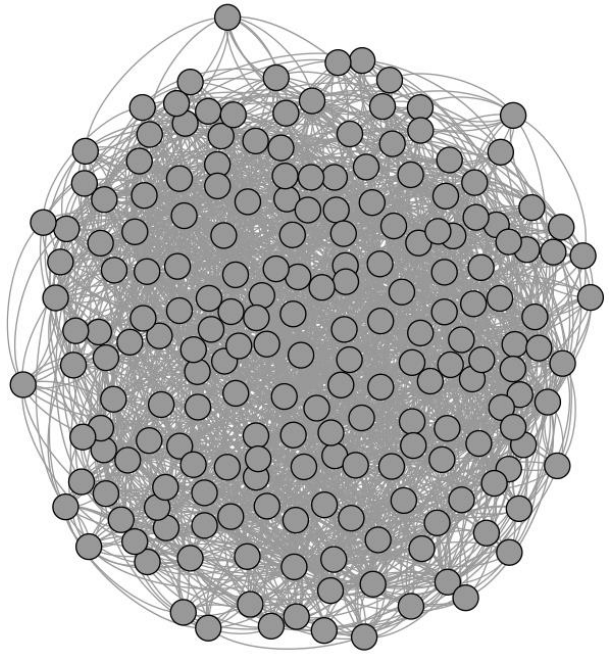
D - Centralité de degré

Module, Cluster, Motifs

Clustering de réseau





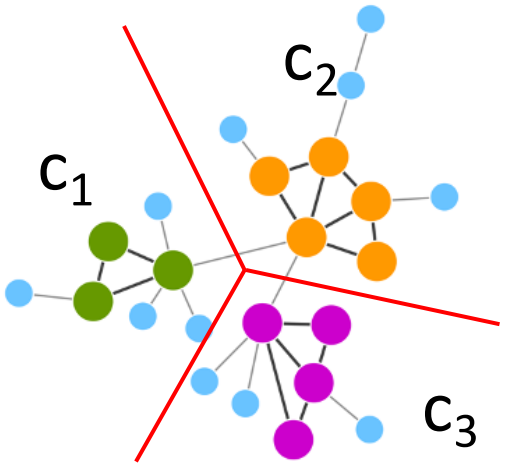


Module, Cluster, Motifs

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

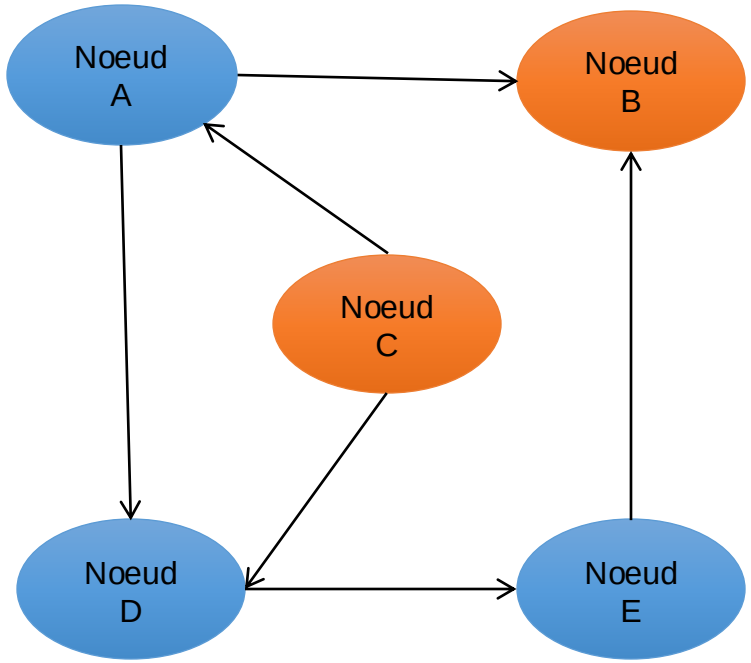
Deux phases :

- On change l'appartenance des nœuds à leur classe et on calcule le changement de modularité que ça induit. On selectionne 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

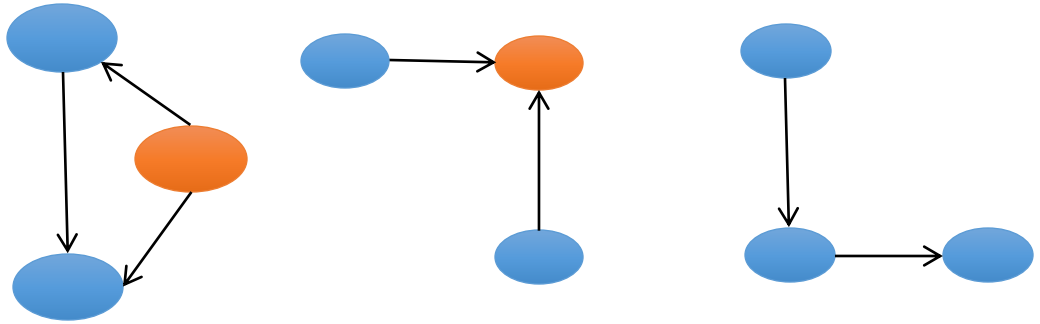
Module, Cluster, Motifs

Les motifs



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

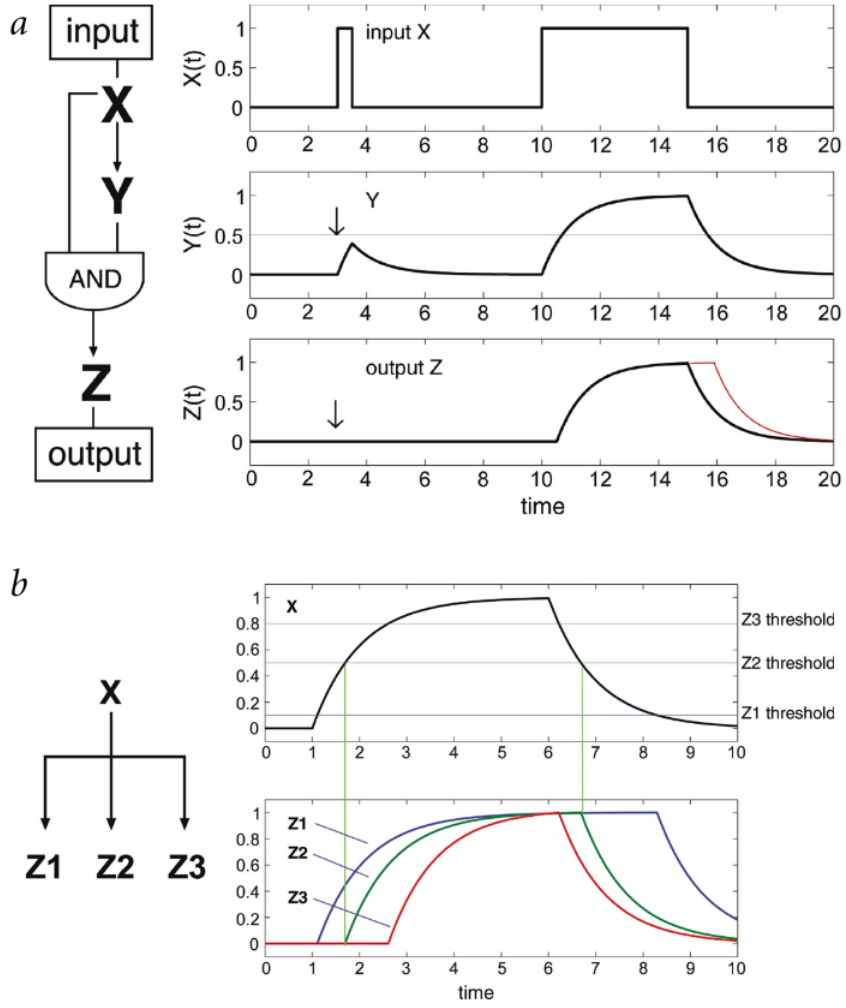
Exemple pour k = 3 :



Réseau de régulation chez Escherichia Coli

Shen-Orr et al., Nature 2002

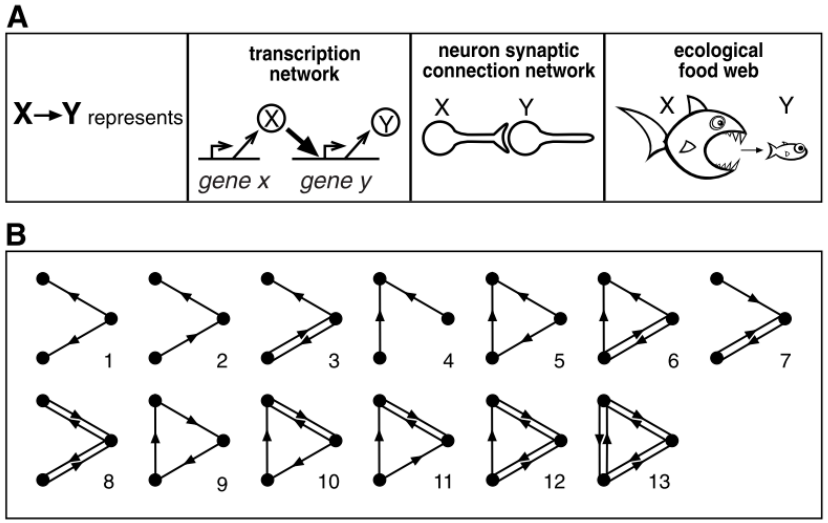
- 577 interactions
- 424 operons
- 116 facteurs de transcriptions
- Extrait de RegulonDB
- Que des facteurs de transcription validés directement



Les motifs privilegiés dans les réseaux

Milo et al., Science 2002

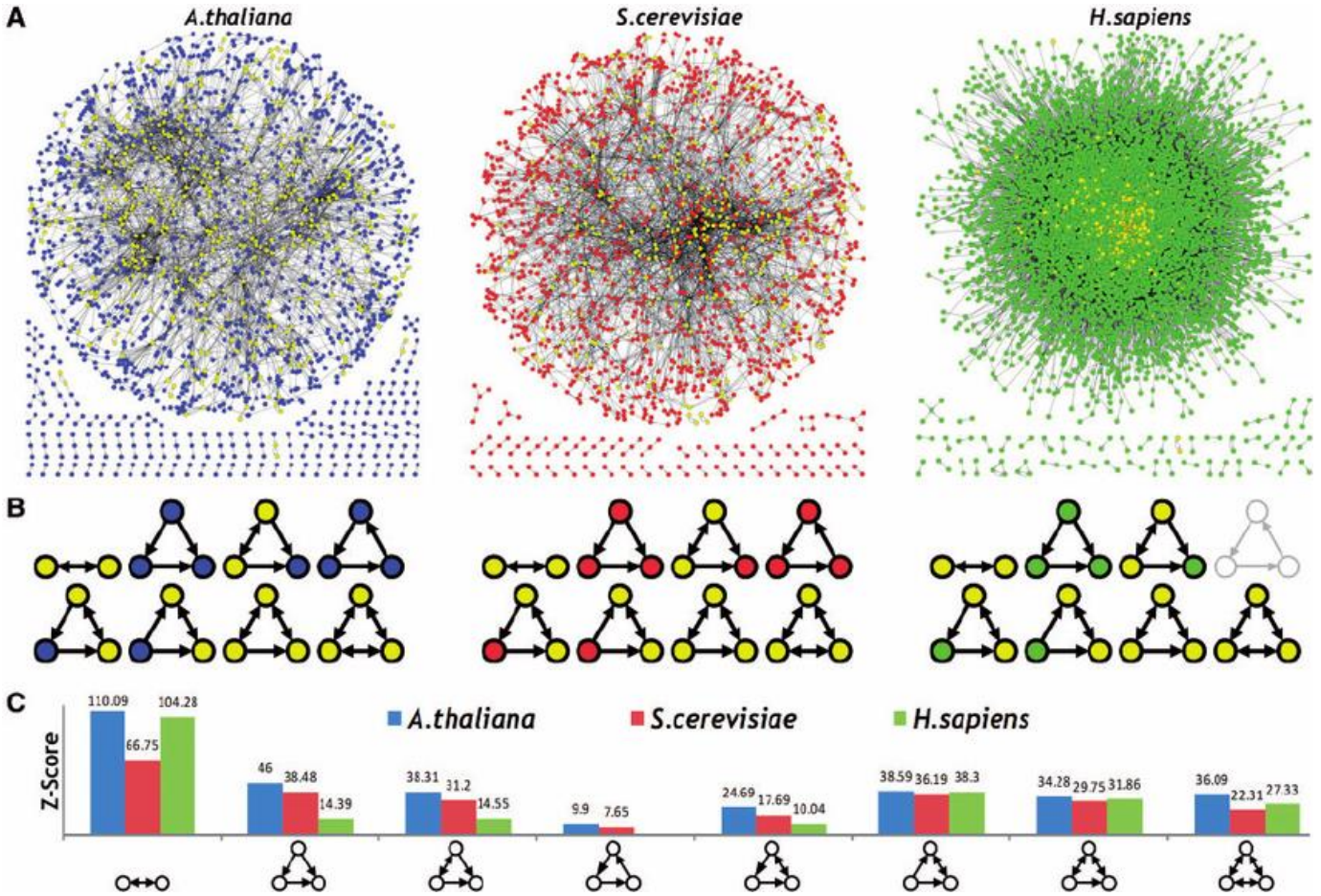
Les motifs sont-ils les briques élémentaires de construction des réseaux ?



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**			
E. coli	424	519	40	7 ± 3	10	203	47 ± 12	13			
*S. cerevisiae**	685	1,052	70	11 ± 4	14	1812	300 ± 40	41			
Neurons					**Feed-forward loop**			**Bi-fan**			**Bi-parallel**
C. elegans†	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

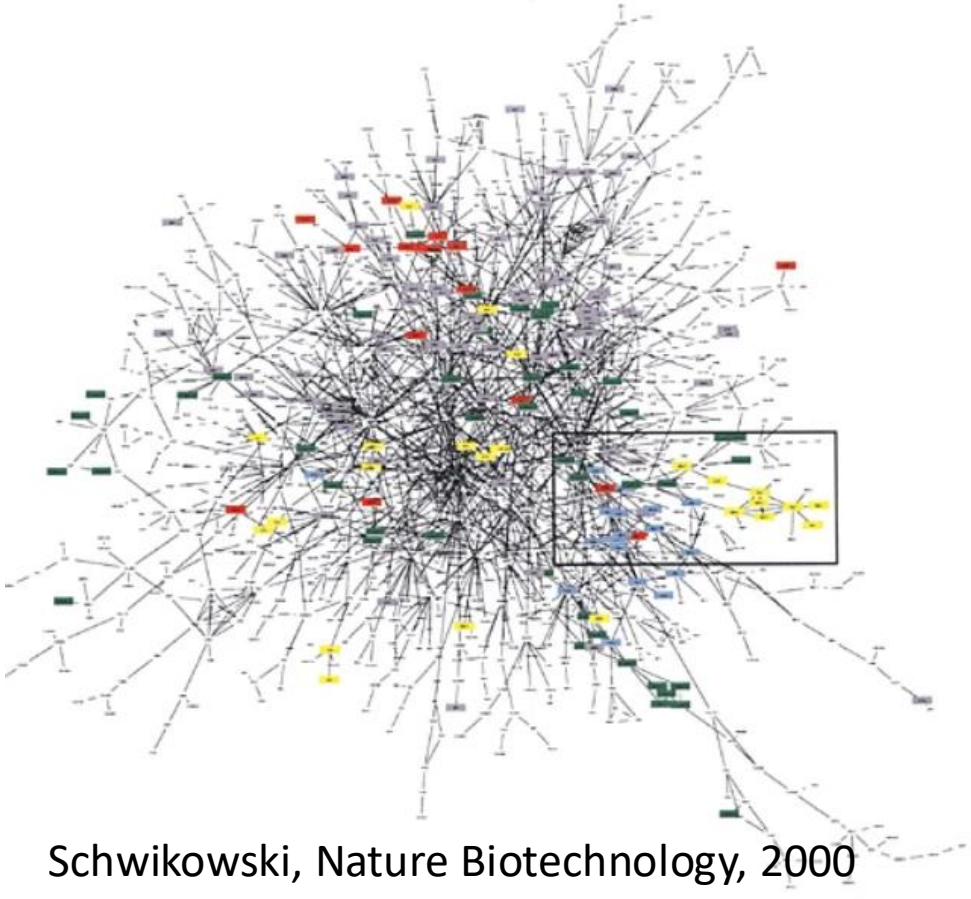
L’évolution = la sélection de motifs de régulation ?

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



Schwikowski, Nature Biotechnology, 2000

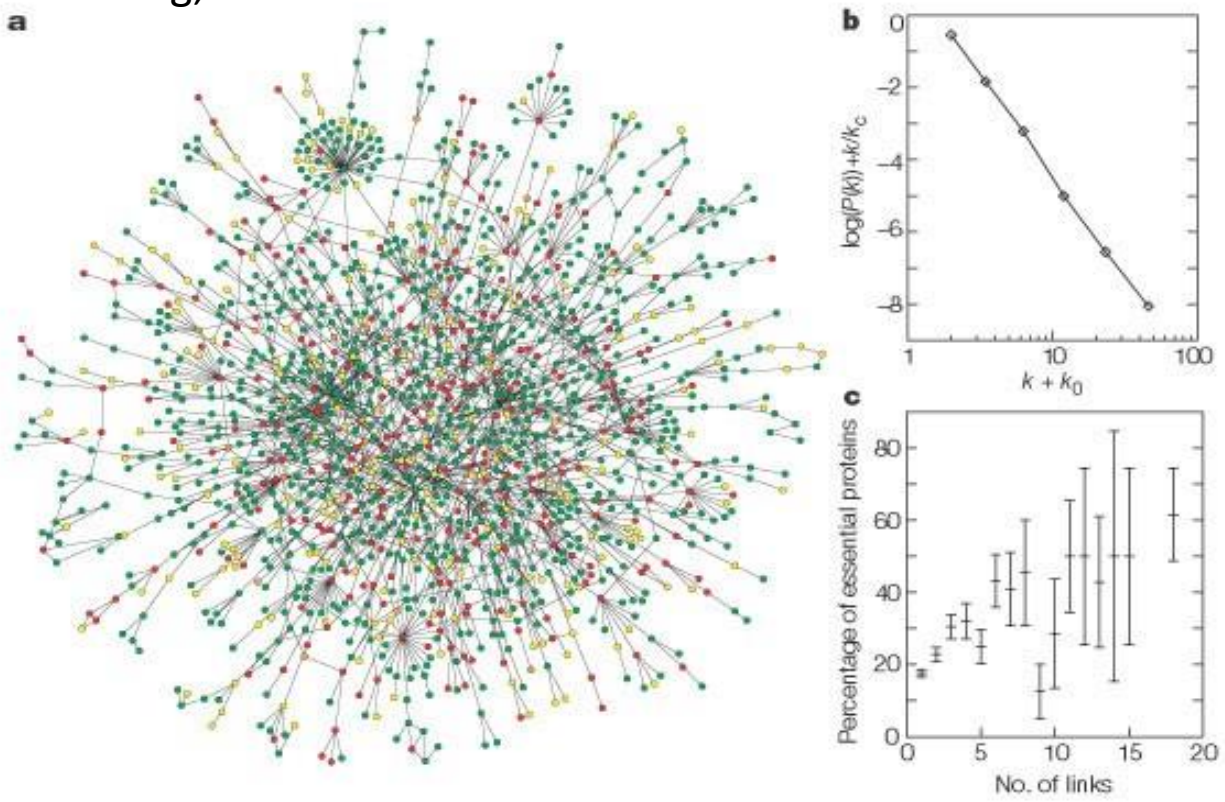
1548 / 6000 protéines – 2358 interactions

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 :

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

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

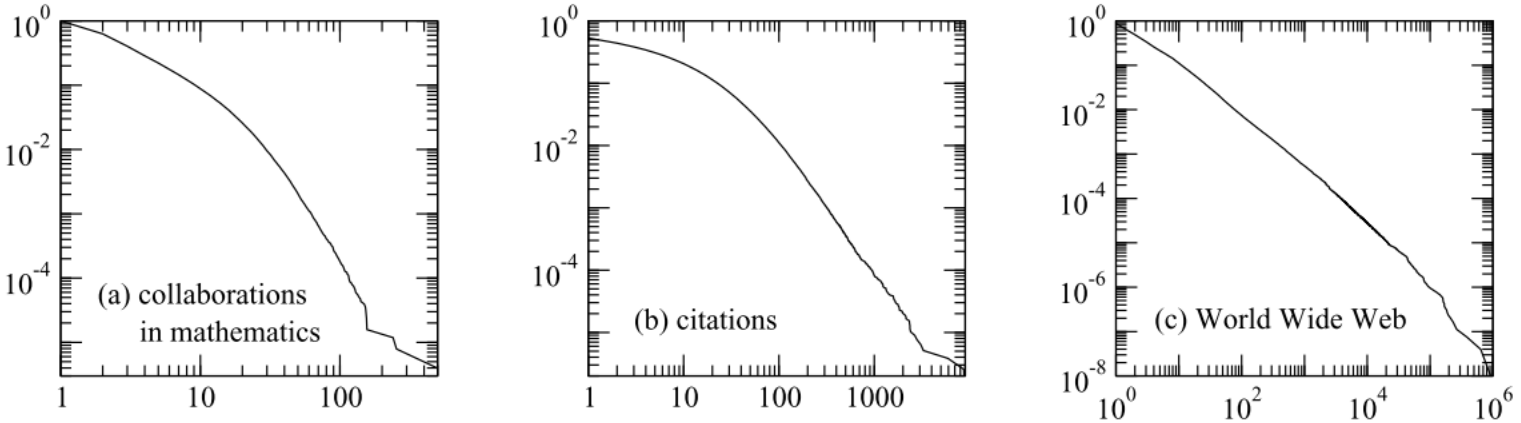
- Un petit nombre de nœud a un degré très grand
- Un grand nombre de nœud a un degré très petit

Lethality and centrality in protein networks

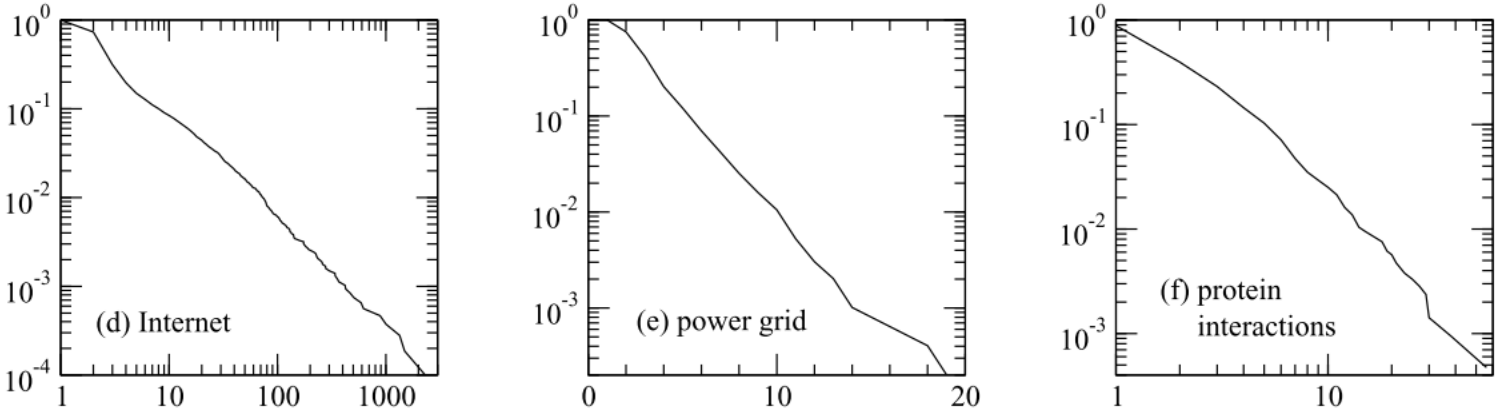
The most highly connected proteins in the cell are the most important for its survival.

Les réseaux sans échelles sont partout

Distribution des degrés de différents réseaux



M. Newman, Siam Review, 2003



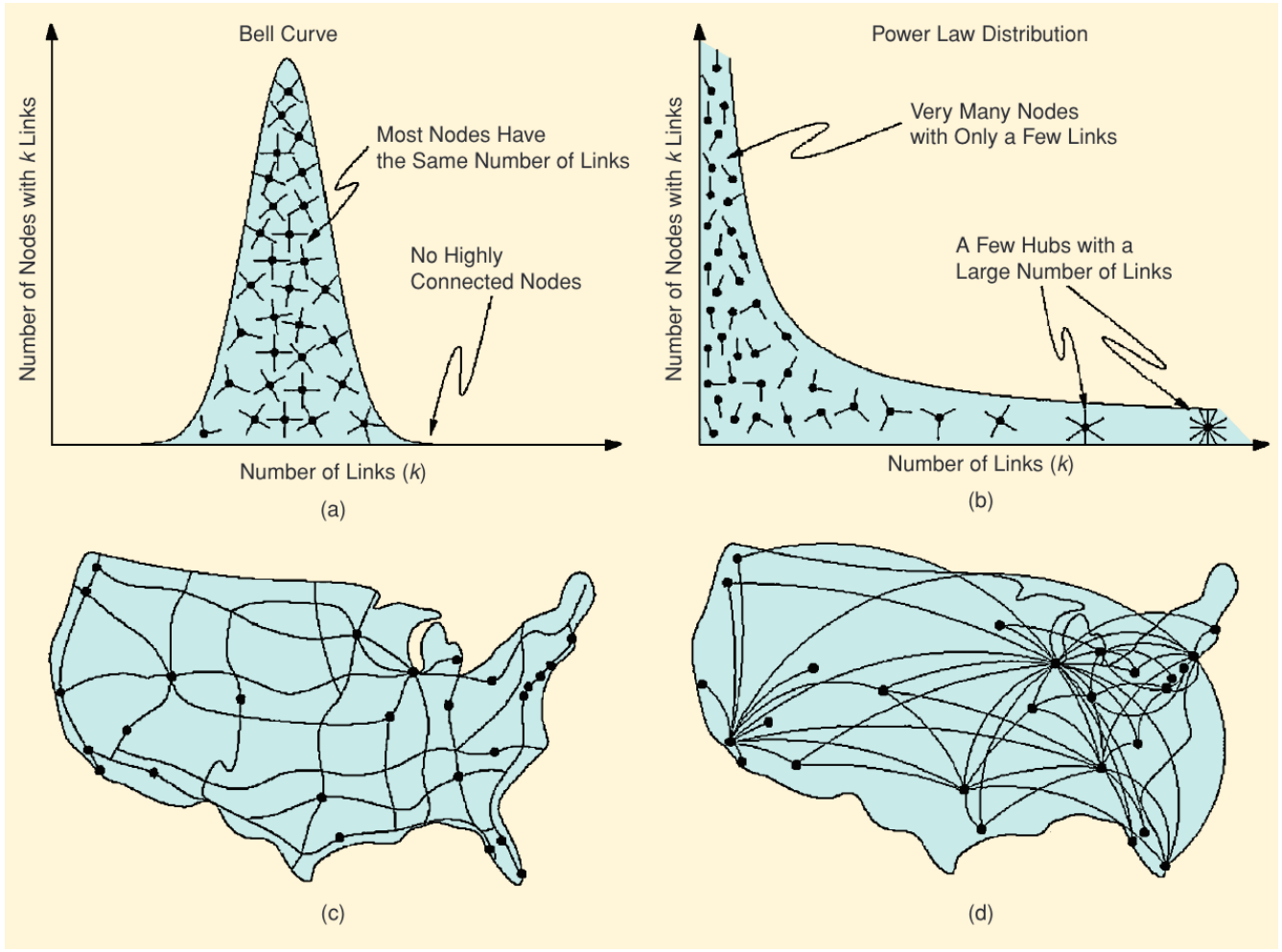
Loi de probabilité d'un réseau sans échelles

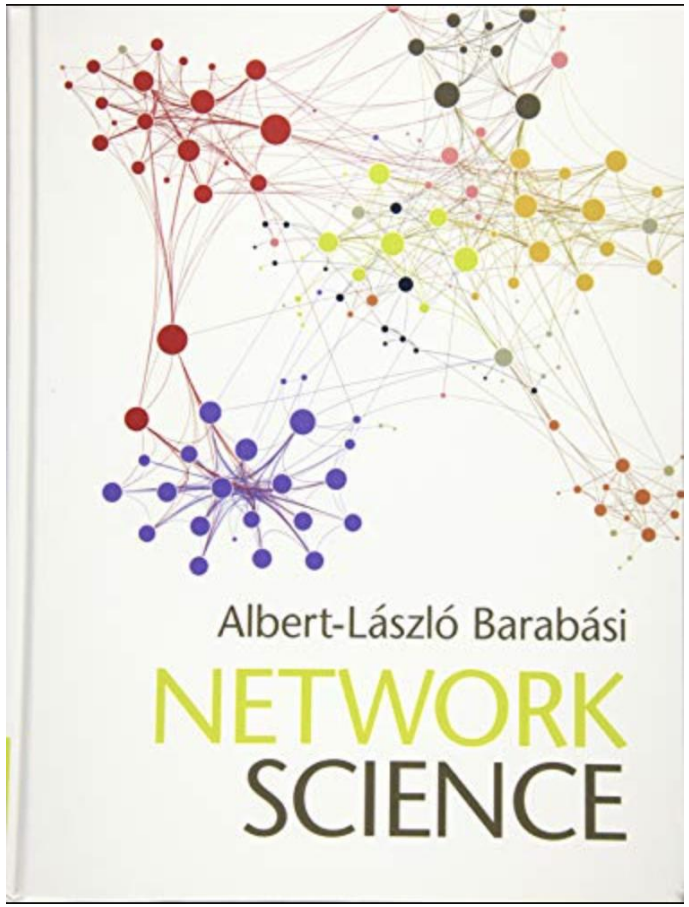
$$\boldsymbol{P(k) \sim k^{-\gamma}}$$

Avec $2 < \gamma < 3$

La structure de réseau sans échelles

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





Exemple de réseau sans échelles

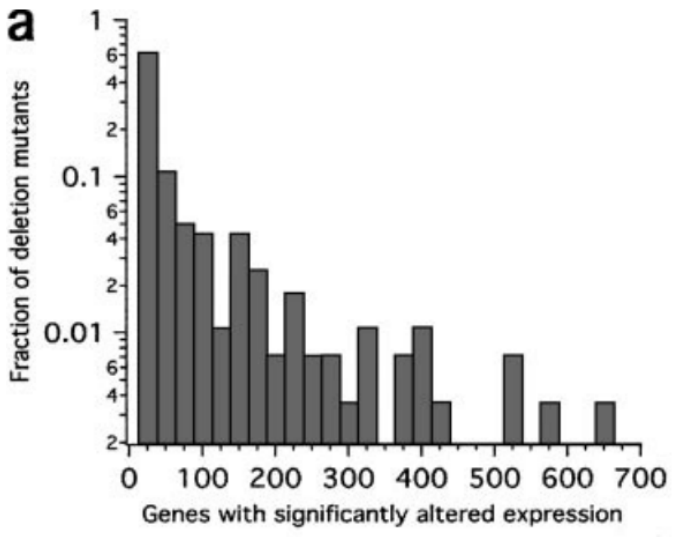
Gene expression network

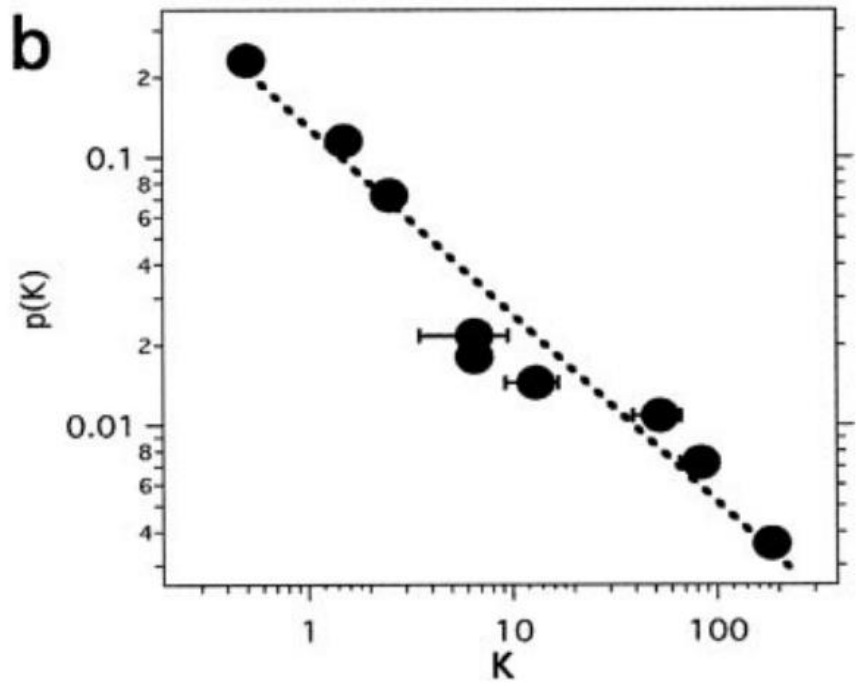
Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network

Bioessays - 2002

David E. Featherstone* and Kendal Broadie

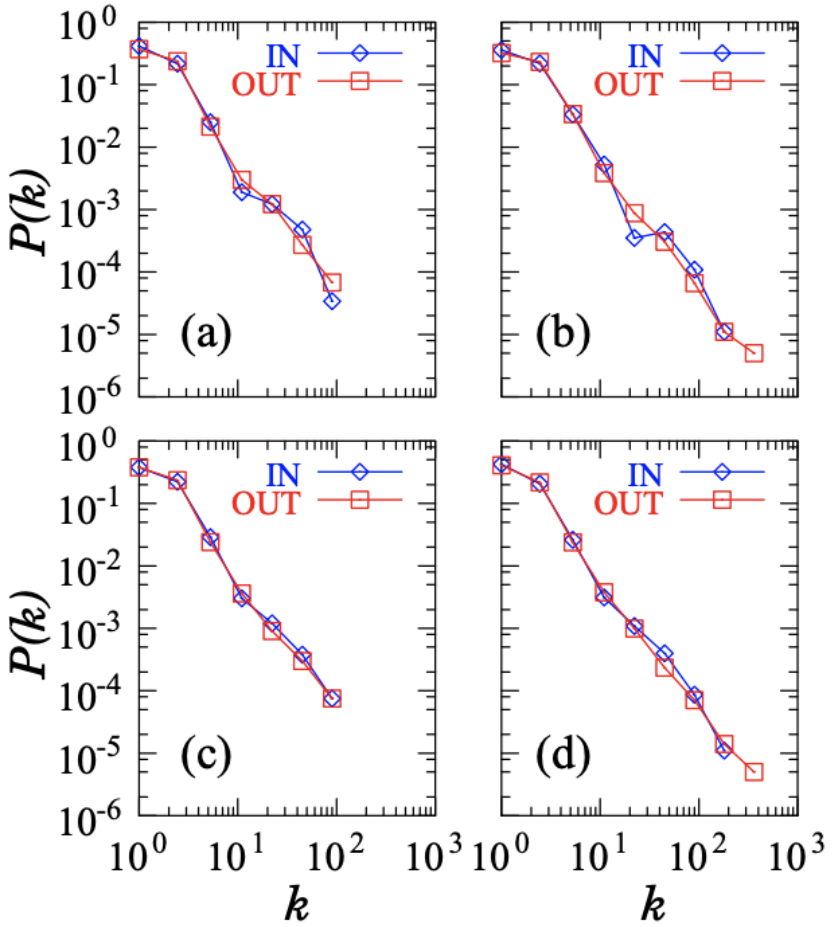
$$P(k) \sim k^{-0.7}$$





Exemple de réseau sans échelles

Les réseaux métaboliques



The large-scale organization of metabolic networks

H. Jeong[1], B. Tombor[2], R. Albert[1], Z. N. Oltvai[2] and A.-L. Barabási[1]

Nature 2000

(a) A. fulgidus (Archae)

(b) E. coli (B acterium)

(c) C. Elegans (Eukaryote)

(d) Moyenne sur 43 organismes

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

Avec $2 < \gamma < 3$

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

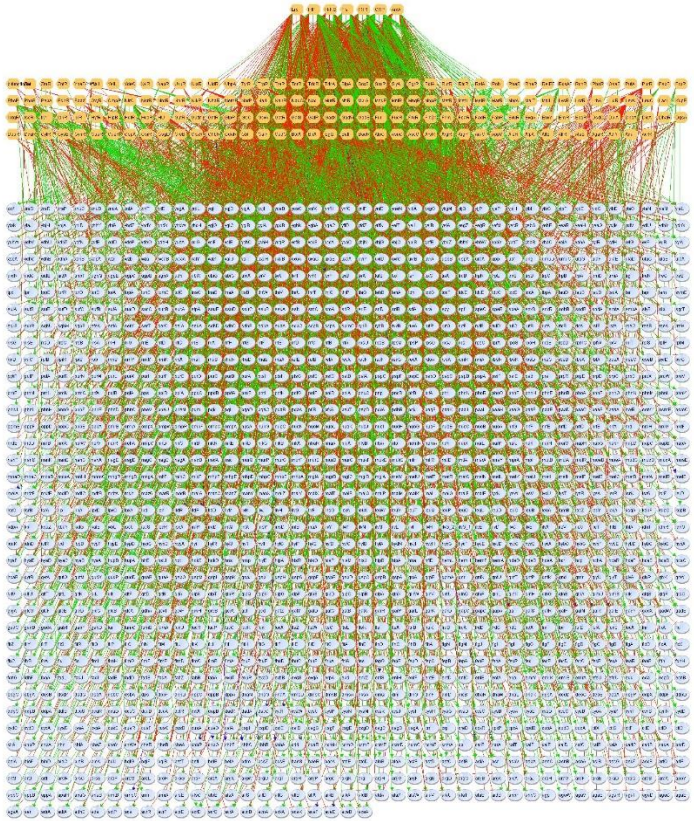
- L’approche systémique en biologie
- Bioinformatique et données omiques
- **Reconstruire un réseau biologique**
  - Les différents types de réseaux
  - Reconstruction de réseau biologique
  - Les obstacles à la reconstruction
  - Structure des réseaux biologiques
  - Etat de l’art des reseaux biologiques les plus étendus
- Virtual cell et digital twins

« Base de données la plus complète sur terre »

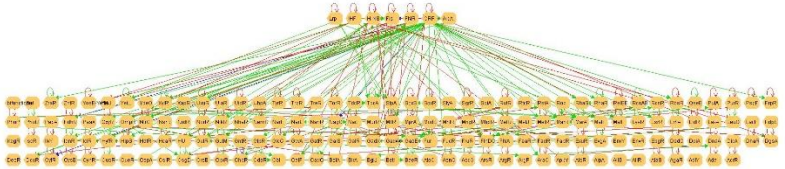
RegulonDB

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

RegulonDB

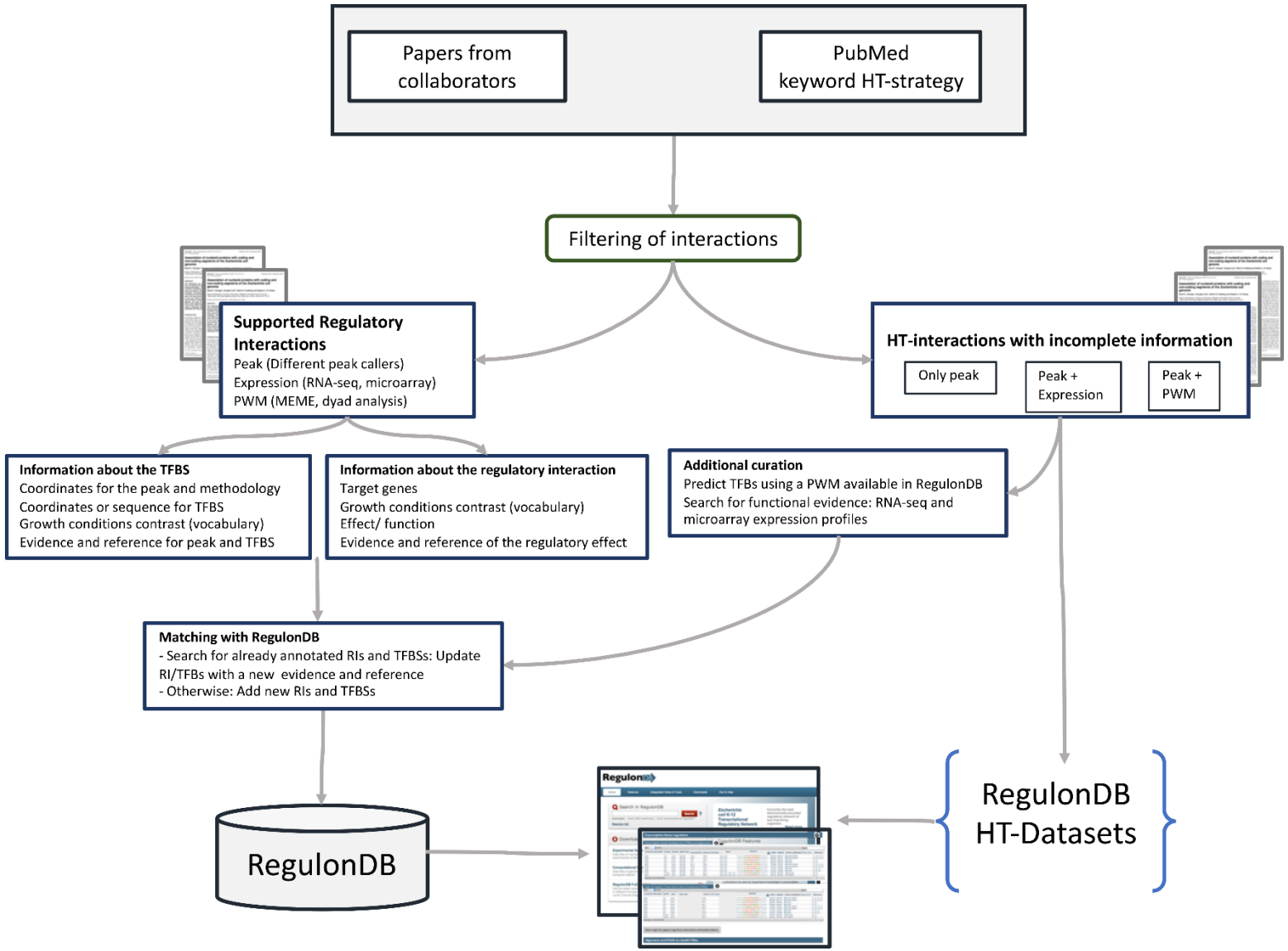
***Escherichia coli* K-12 Transcriptional Regulatory Network**

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 »

RegulonDB

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	Dow
	TF - operon interactions	Dow
	TF - TU interactions	Dow
	TF - TF interactions	Dow
	Sigma - gene interactions	Dow
	Sigma - TU interactions	Dow
	Alon and MA interactions	Dow
	sRNA - gene interactions	Dow

Promoters	All Promoters	Download
	Sigma 70	Download
	Sigma 54	Download
	Sigma 38	Download
	Sigma 32	Download
	Sigma 28	Download
	Sigma 24	Download
	Sigma 19	Download
	Unknown	Download
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

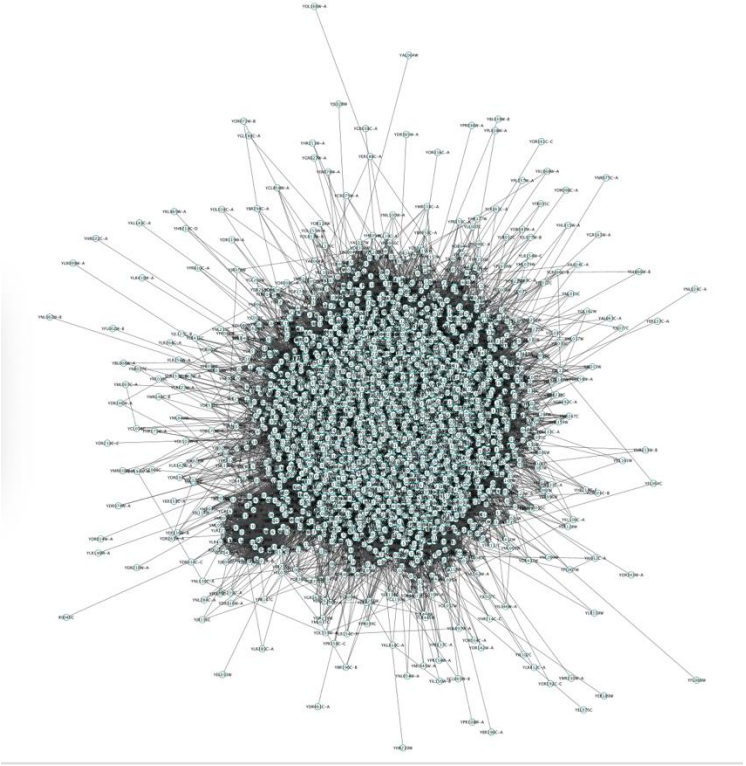
Description	Method
Promoter predictions	"We observed that real promoters occur mostly within regions with high densities of overlapping putative promoters. We evaluated several strategies to identify promoters. The best one uses an intrinsic score of the -10 and -35 hexamers that form the promoter as 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 *Escherichia coli* 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 *Escherichia coli* 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 *E. coli* 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 Escherichia coli 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.
; 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

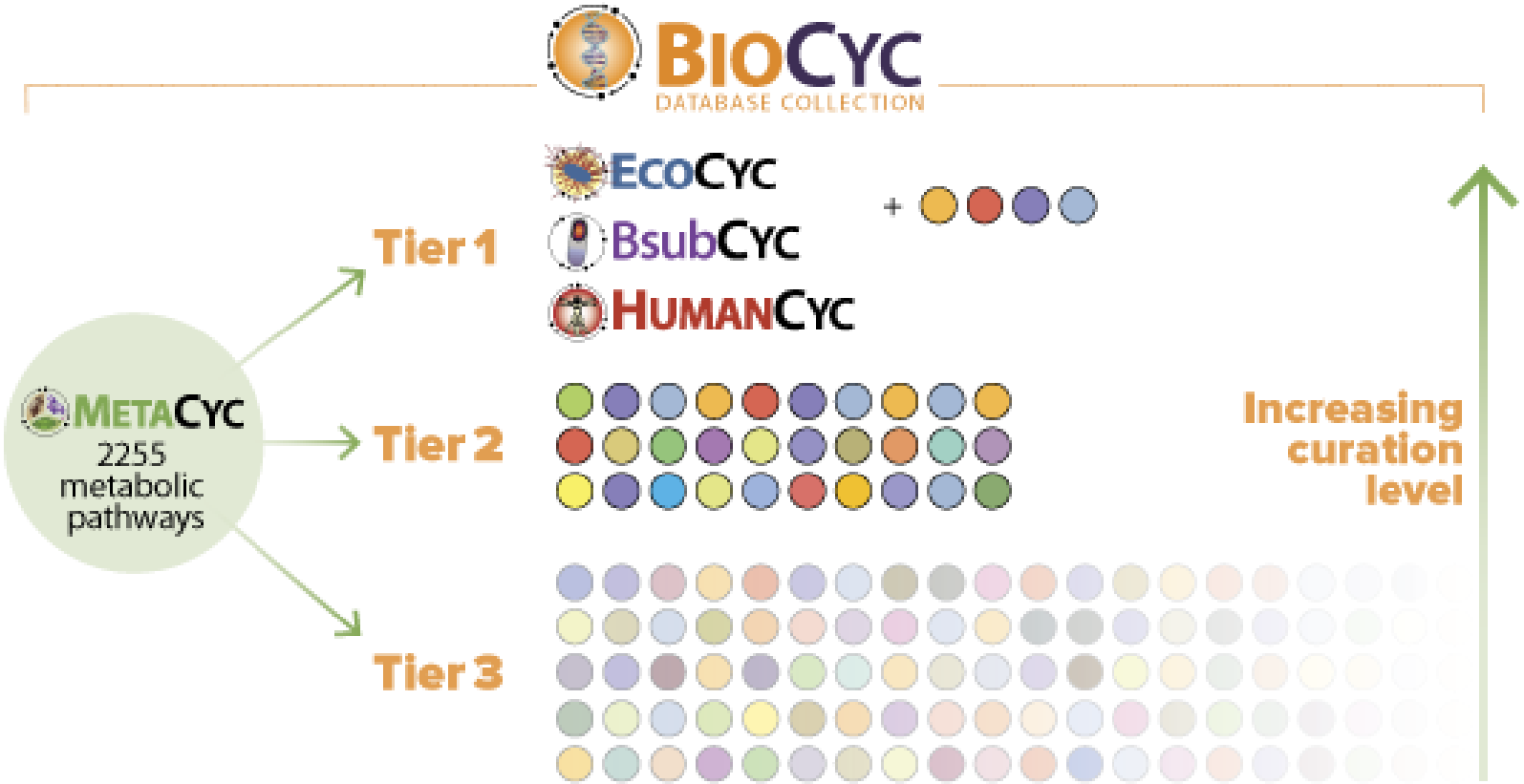


Kim et al.,
Nucleic Acid Research 2014

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



Nettoyé activement
par la communauté
5 espèces

Nettoyé pendant 1-3 mois
par la communauté
30 espèces

Génération automatisée
16764 espèces

Etat de l’art : BioCyc

EcoCyc

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

BsubCyc

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

HumanCyc

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

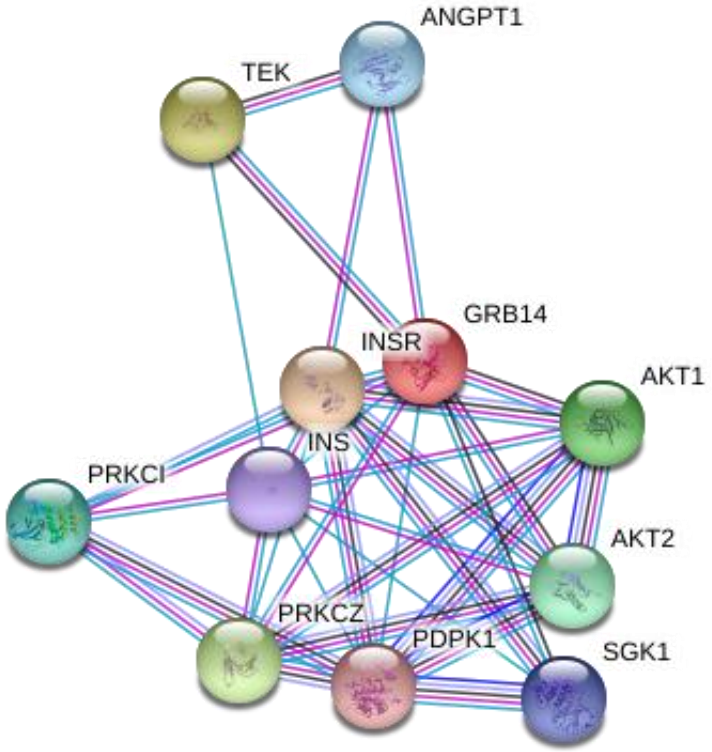
Etat de l’art : MetaCyc

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

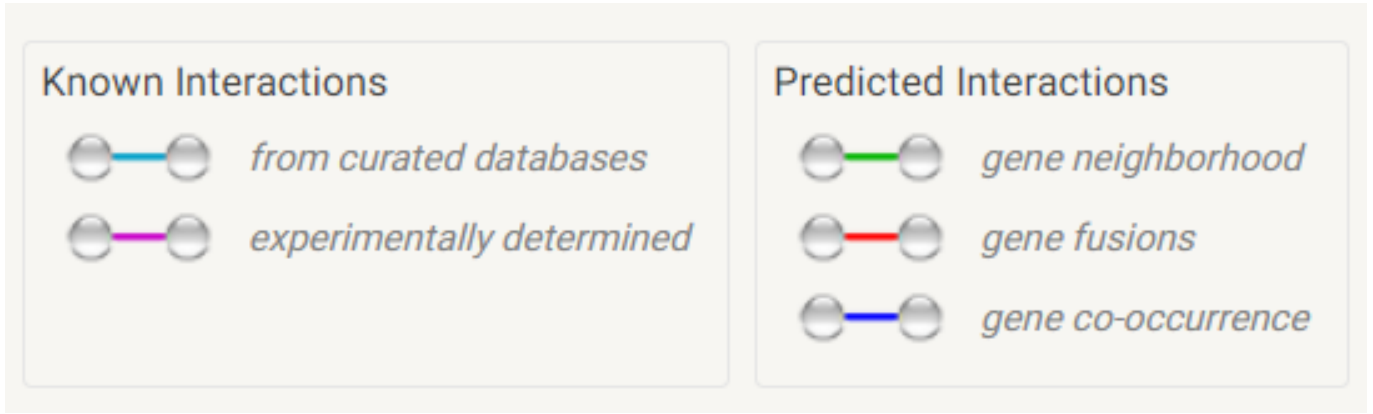
MetaCyc contains pathways involved in both primary and secondary metabolism, as well as associated metabolites, reactions, enzymes, and genes. The goal of MetaCyc is to catalog the universe of metabolism by storing a representative sample of each experimentally elucidated pathway.

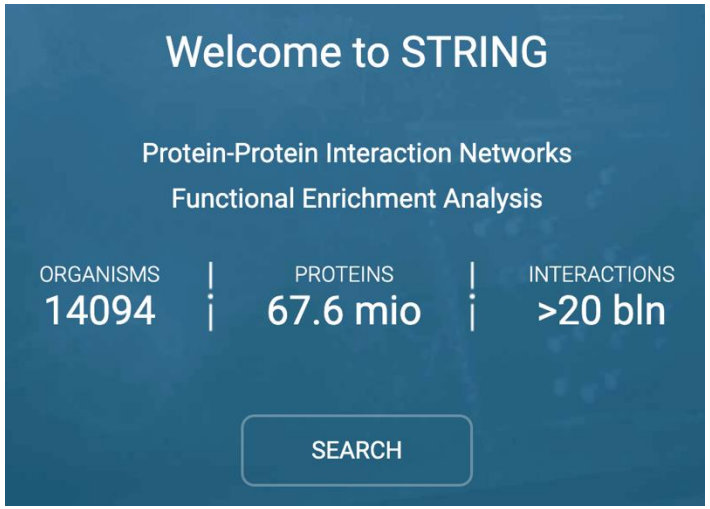
Genes:	14,976	SmartTable	Ontology
Pathways:	3,063	SmartTable	Ontology
Enzymatic Reactions:	18,285		
Transport Reactions:	910		Ontology
Polypeptides:	16,731	SmartTable	Ontology
Protein Complexes:	4,756	SmartTable	Ontology
Enzymes:	14,028	SmartTable	
Transporters:	638	SmartTable	
Compounds:	18,452	SmartTable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

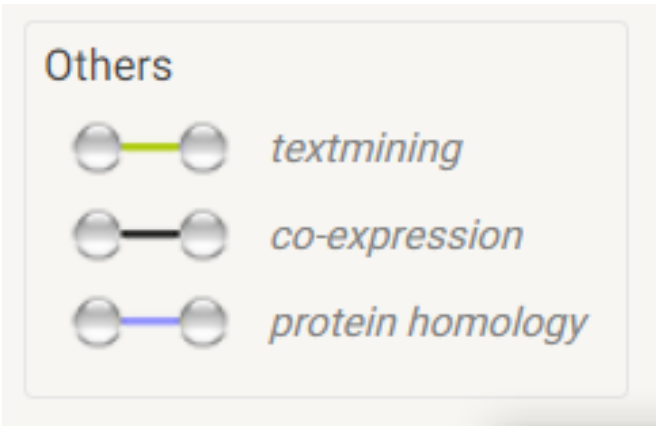
Le serveur String-DB.org



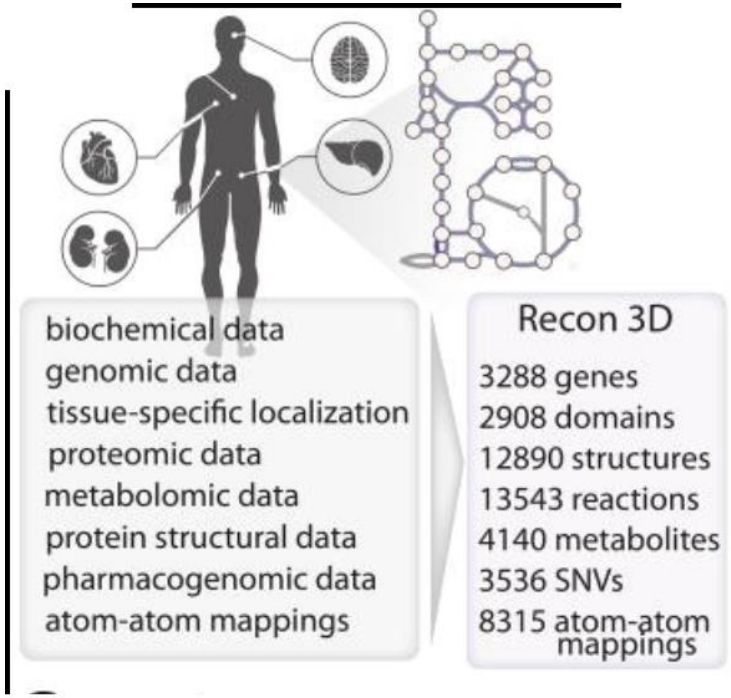
Réseau d'interaction protéine-protéine de l'insuline prédite par string-db.org

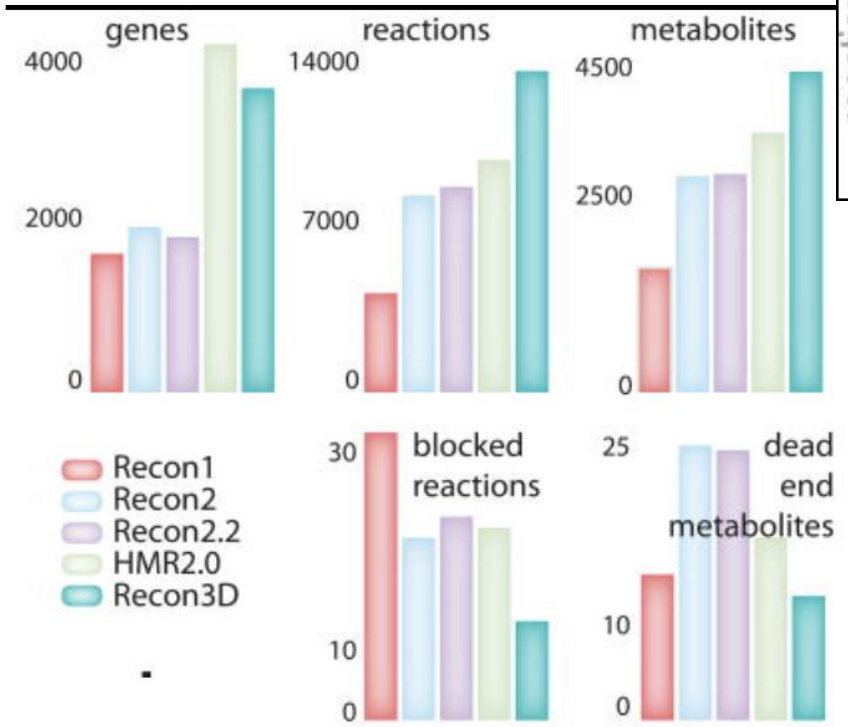


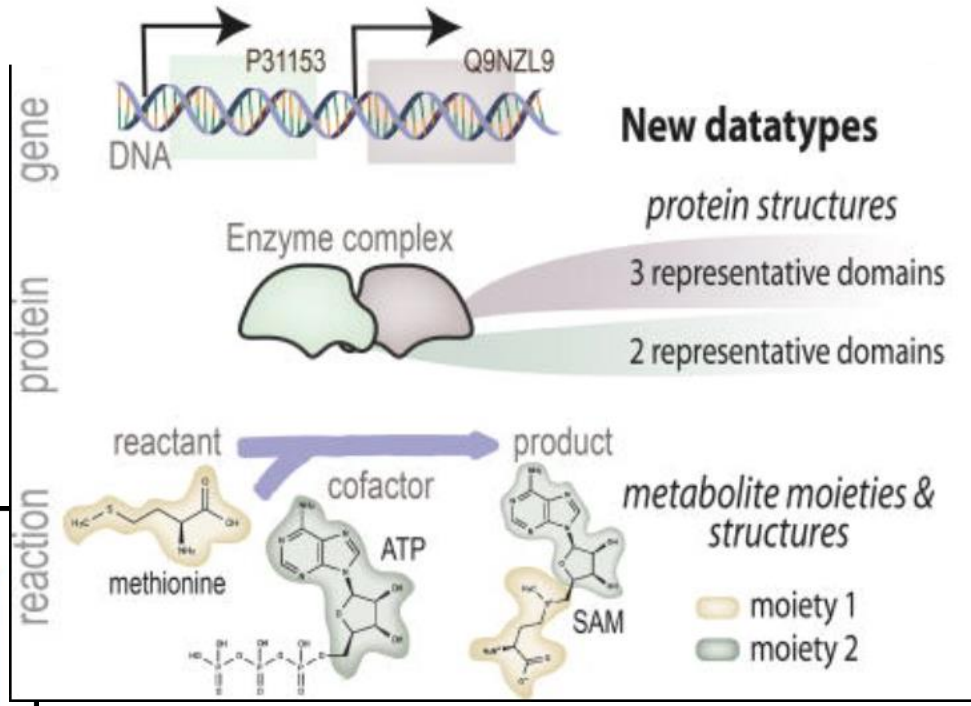




Recon3D : Human metabolism







Brunk et al., Nature Biotechnology 2018

Biologie des systèmes

- L’approche systémique en biologie
- Bioinformatique et données omiques
- Reconstruire un réseau biologique
  - Les différents types de réseaux
  - Reconstruction de réseau biologique
  - Les obstacles à la reconstruction
  - Structure des réseaux biologiques
  - Etat de l’art des reseaux biologiques les plus étendus
- **Virtual cell et digital twins**

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

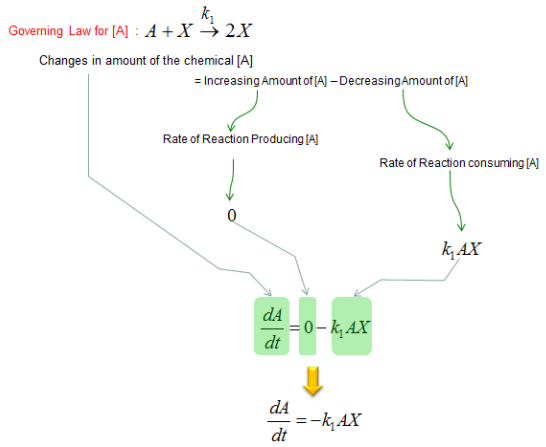
Application de la modélisation

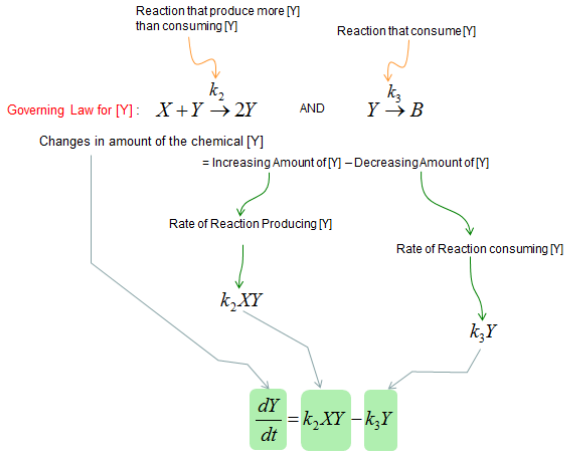
Modéliser = Partir des interactions définies dans le réseau et décrire/prédire le comportement du système

- 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
- …

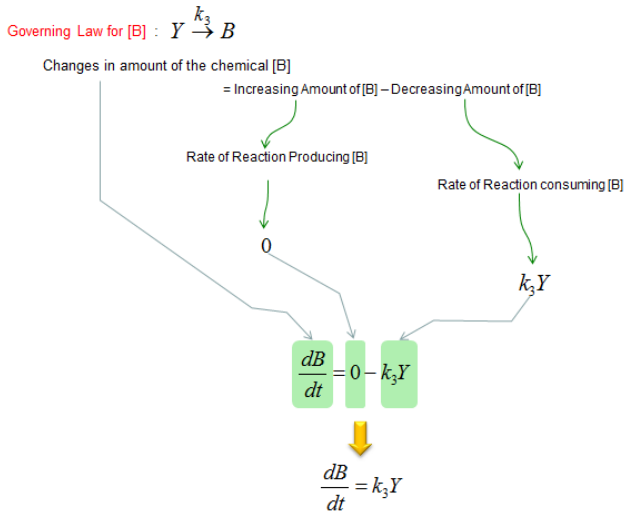
Modélisation d'équation chimique

$$\begin{cases} A + X \xrightarrow{k_1} 2X & \text{---------- } 1 \\ X + Y \xrightarrow{k_2} 2Y & \text{---------- } 2 \\ Y \xrightarrow{k_3} B & \text{---------- } 3 \end{cases}$$

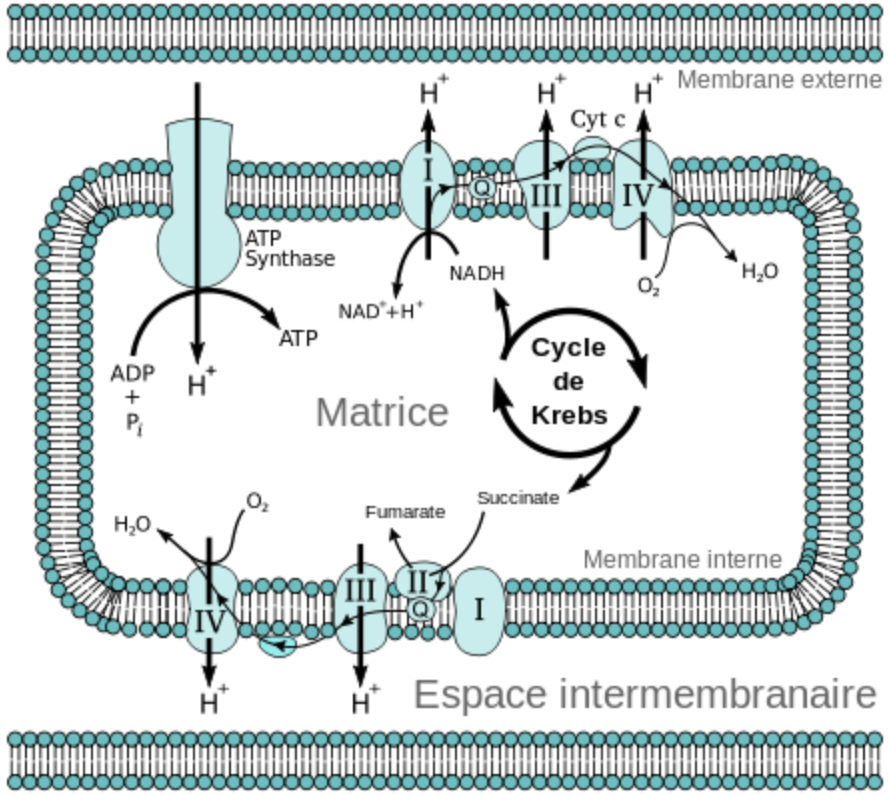




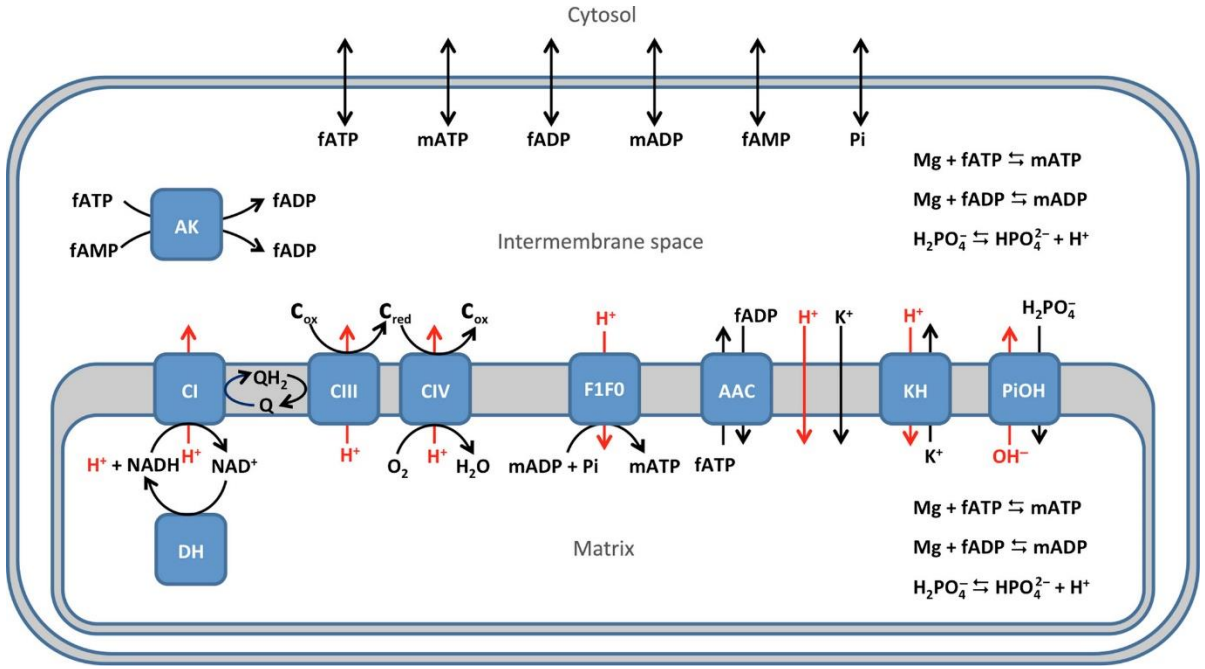
$$\begin{cases} \frac{dA}{dt} = -k_1AX \\ \frac{dX}{dt} = k_1AX - k_2XY \\ \frac{dY}{dt} = k_2XY - k_3Y \\ \frac{dB}{dt} = k_3Y \end{cases}$$



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	$QH_{2im} + 2Cox_{is} + 2H_x \rightleftharpoons Q_m + 2Cred_{is} + 4H_{is}$
C4	Complex IV	$Cred_{is} + 0.25\ O_2 + 2H_x \rightleftharpoons Cox_{is} + 0.5H_2O + H_{is}$
F1F0	ATP synthase	$mADP_x + P_{i_x} + n_A\ H_{is} + H_x \rightleftharpoons mATP_x + n_A\ H_x$
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	$P_{i_e} \rightleftharpoons P_{i_{is}}$
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	$fADP_e + Mg_e \rightleftharpoons mADP_e$

$$\frac{d[H]_x}{dt} = x_{buff} \cdot \begin{pmatrix} +v_{DH} - 5v_{C1} - 2v_{C3} - 4v_{C4} \\ +(n_A - 1)v_{F1F0} + 2v_{PiOH} + v_{Hleak} - v_{KH} \end{pmatrix} / W_x \quad (35)$$

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

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

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

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

$$\frac{d[mATP]_x}{dt} = +v_{F_1F_0}/W_x + v_{MgATP_x} \quad (40)$$

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

$$\frac{d[mADP]_x}{dt} = -v_{F_1F_0}/W_x + v_{MgADP_x} \quad (42)$$

$$\frac{d[Pi]_x}{dt} = (-v_{F_1F_0} + v_{PiHt})/W_x \quad (43)$$

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

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

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

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

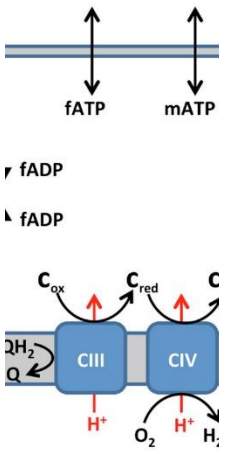
$$\frac{d[fADP]_{is}}{dt} = (+v_{ADP_{om}} - v_{AAC_{im}} + 2v_{AK_{is}})/W_{is} - v_{MgADP_{is}} \quad (48)$$

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

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

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

$$\frac{d\Delta\Psi}{dt} = (+4v_{C1} + 2v_{C3} + 4v_{C4} - n_A \cdot v_{F_1F_0} - v_{AAC} - v_{Hleak} - v_K)/C \quad (52)$$



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

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

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

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

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

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

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

$$[Pi]_s = [Pi]_e \quad (60)$$

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

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

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

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

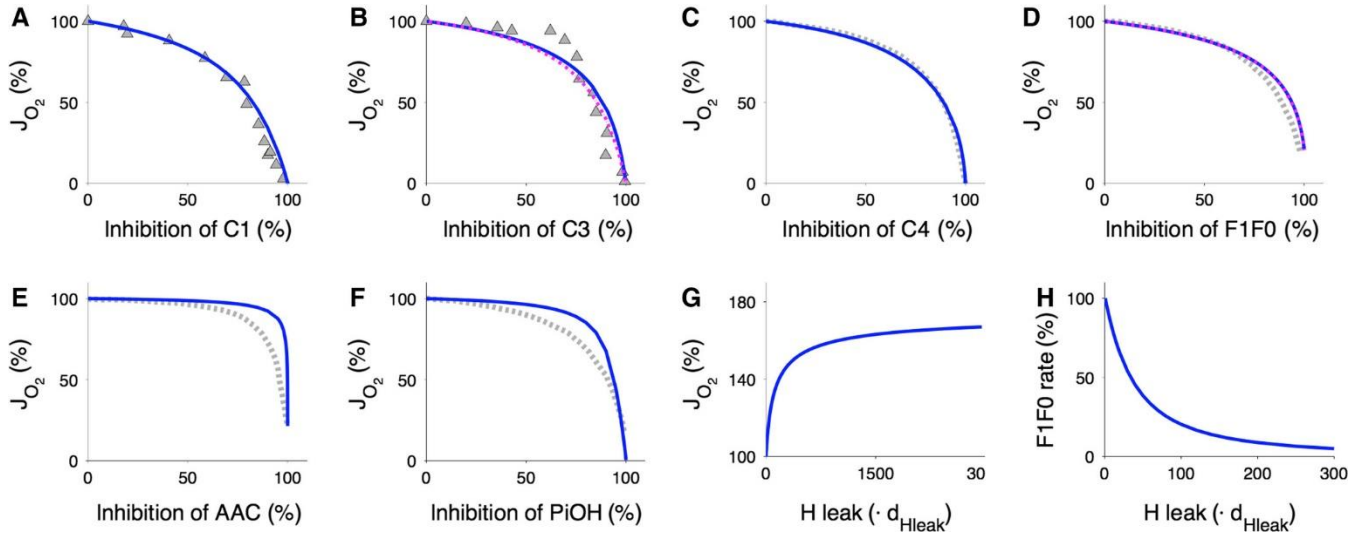
$$[OH]_{is} = \frac{10^{-14}}{[H]_{is}} \quad (65)$$

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

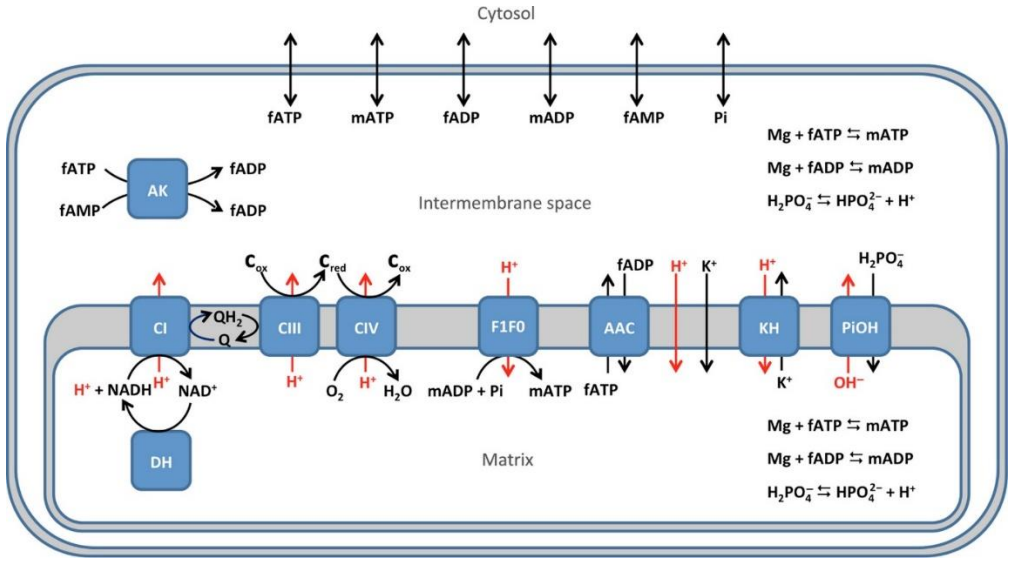
Heiske et al., FEBS Journal 2017

Exemple : Phosphorylation oxydative

Optimisation non-linéaire sous contrainte



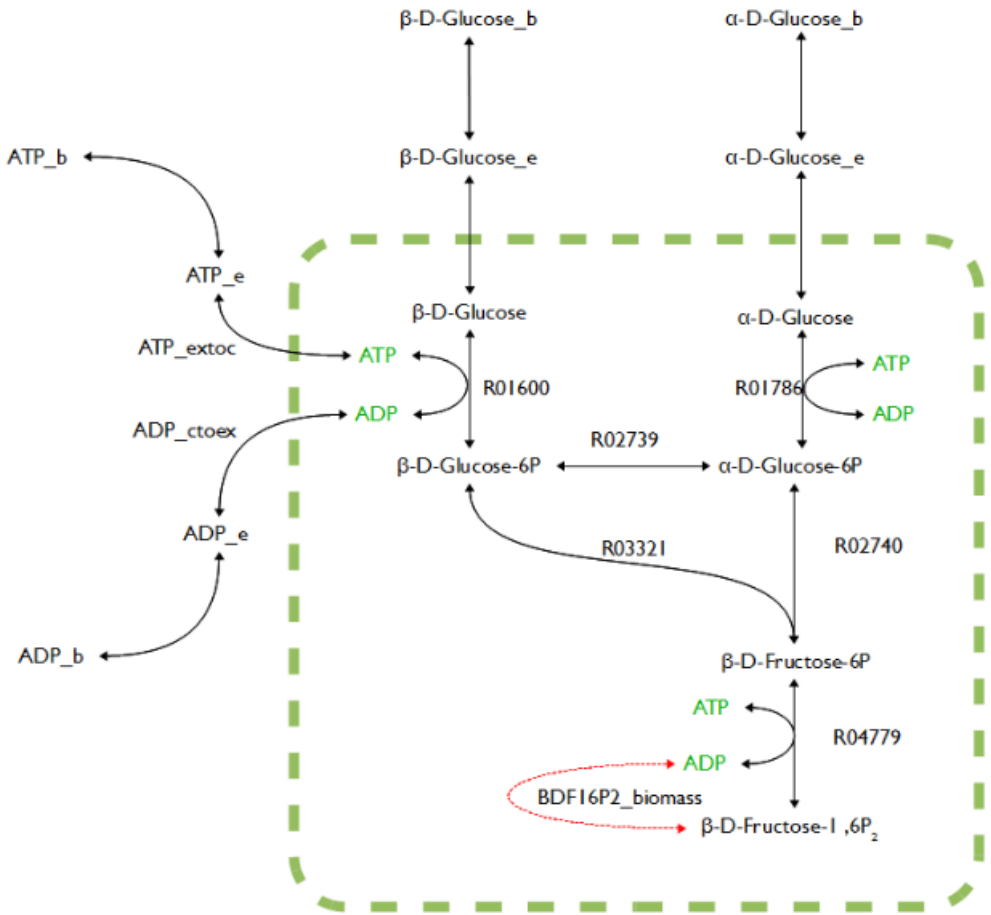
On obtient une modélisation de l'évolution de toutes les concentrations



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

openCOBRA

The COnstraint-Based Reconstruction and Analysis Toolbox

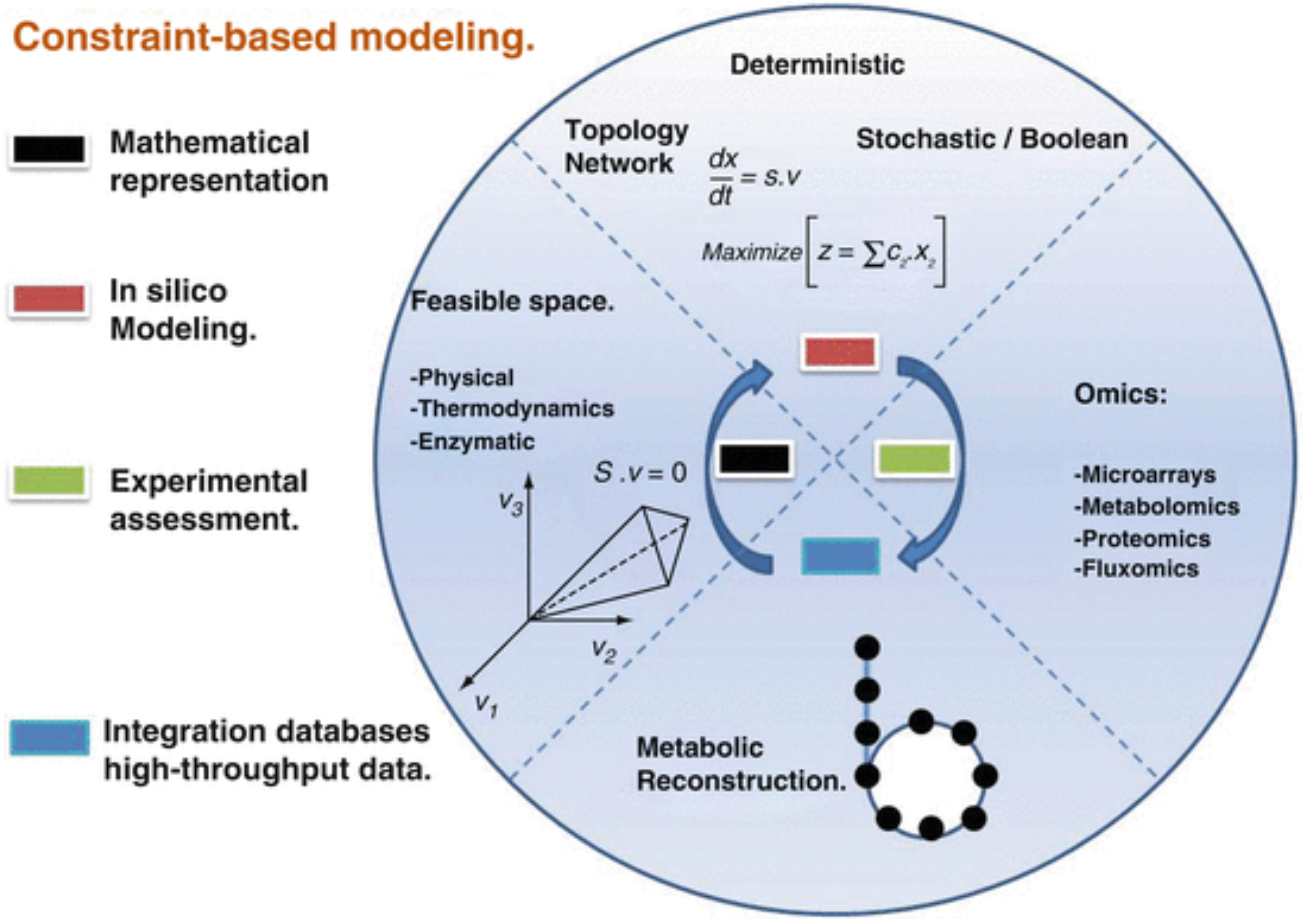
Il est trop couteux de modéliser l'évolution de tous les constituants du système

↓

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

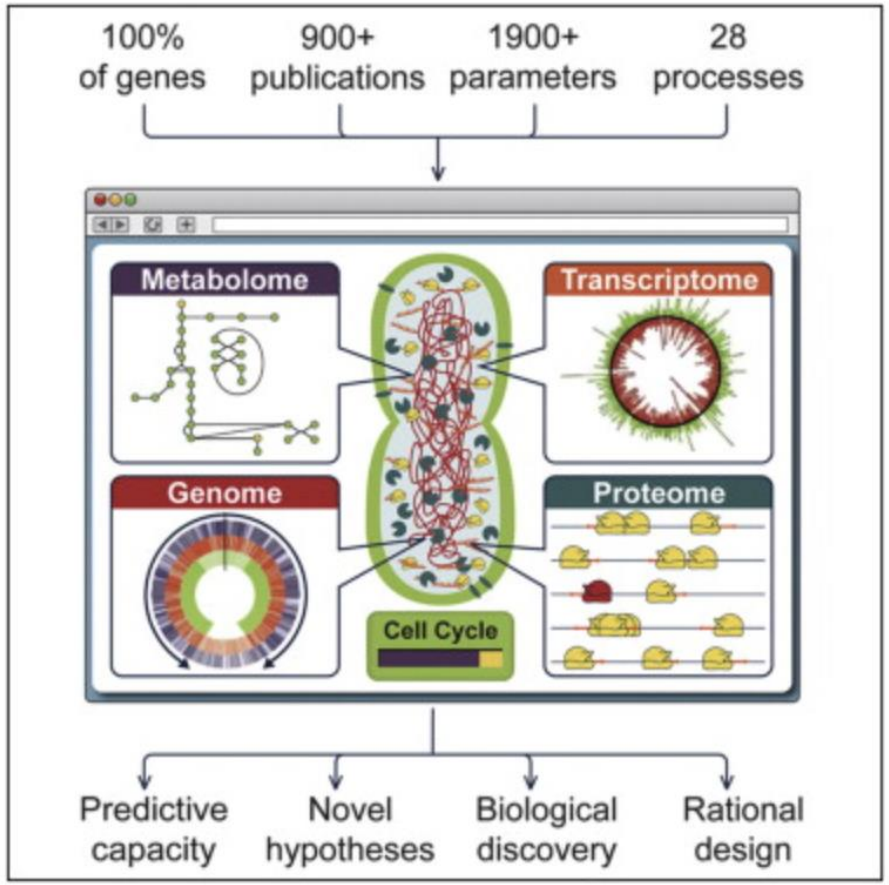
↓

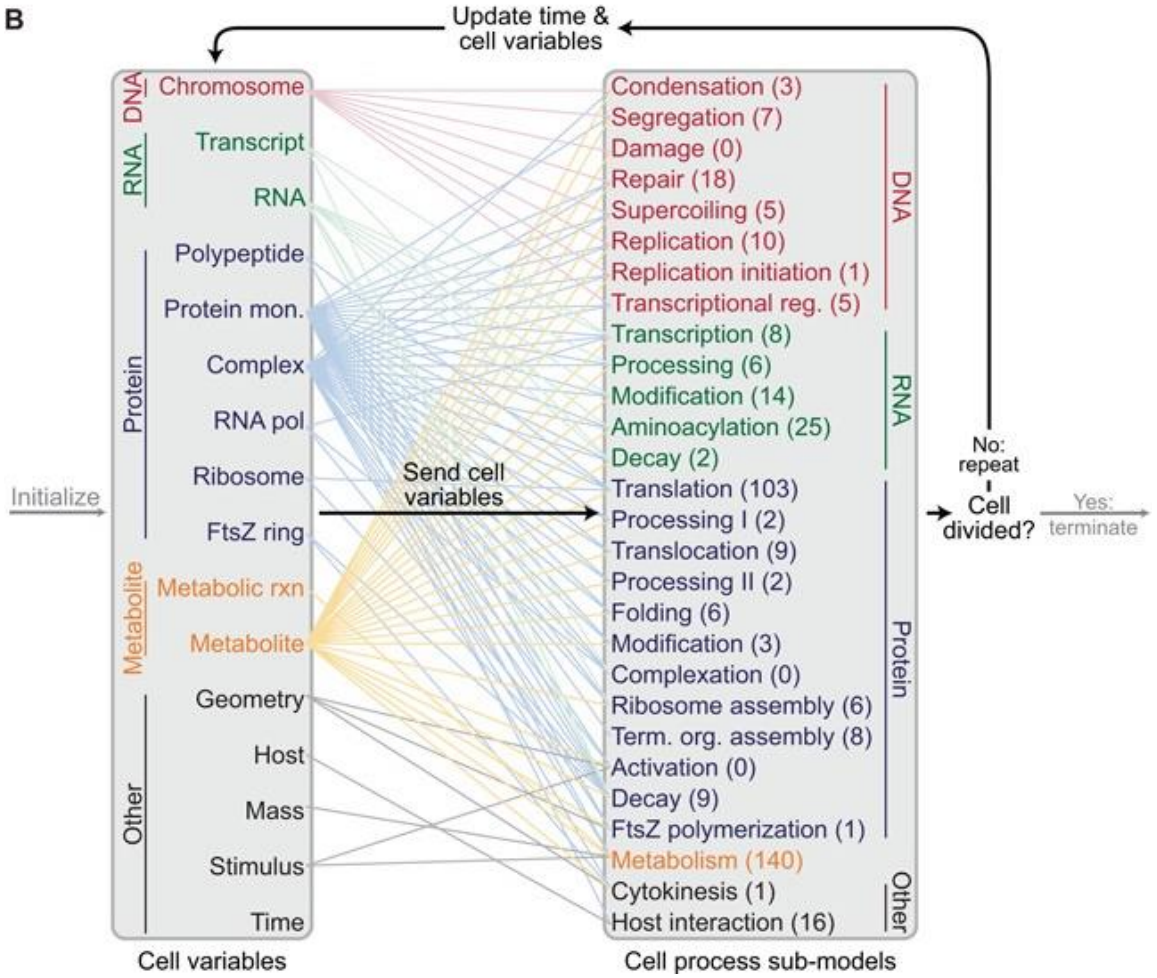
La modélisation deviendra alors possible en un temps raisonnable



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

Mycoplasma genitalium Reconstruction ?





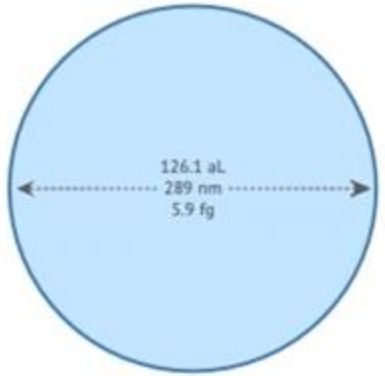
J. R. Karr, et al., M. W. Covert,
Cell, 2012

Mycoplasma genitalium
WholeCellViz
Home | Edit | Download | Help

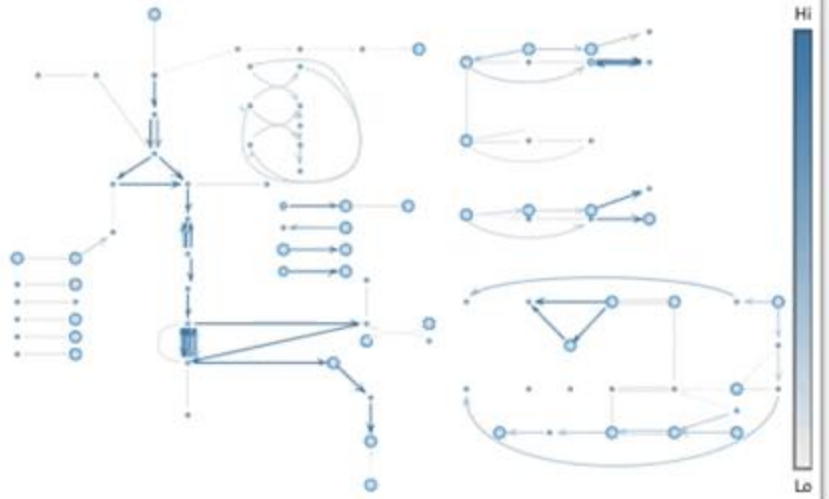
28 sous-modèles de 28 processus cellulaires
401 gènes, 722 molécules, 1,857 réactions, and 1,836 paramètres

Example views
Cell growth | Cell cycle | Replication
Omics | Synthesis | Population | Knockouts

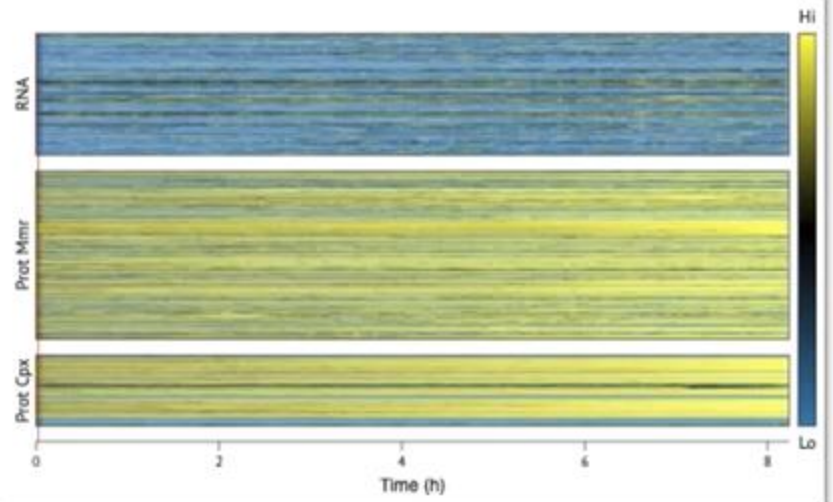
Cell shape
Simulation: Wild-type #1



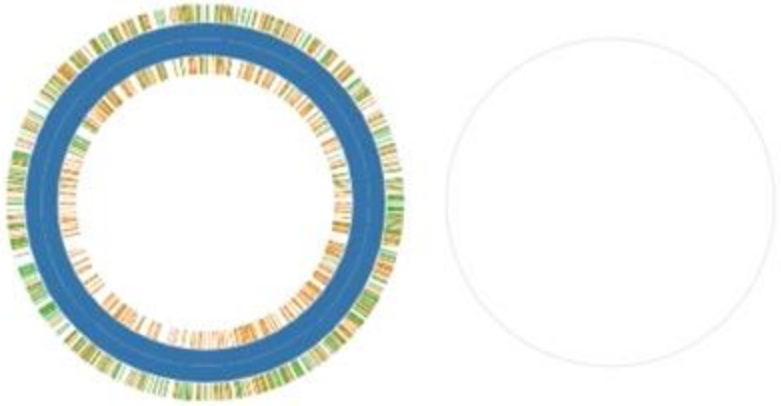
Metabolism
Simulation: Wild-type #1



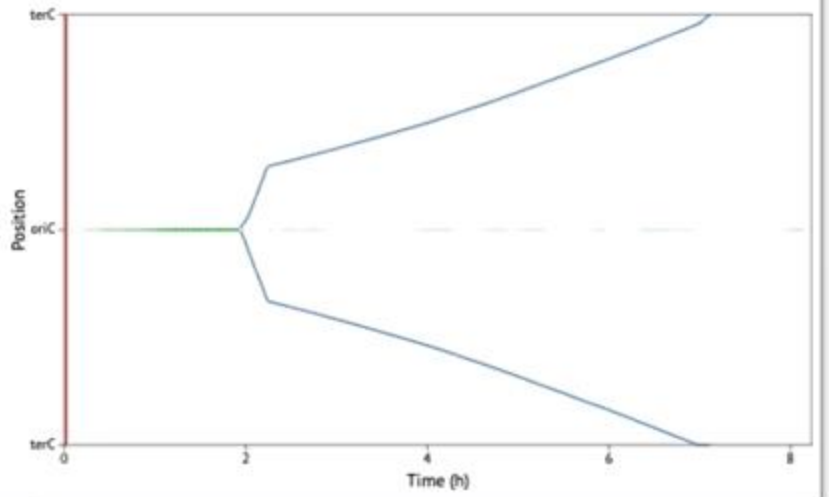
Gene expression
Simulation: Wild-type #1



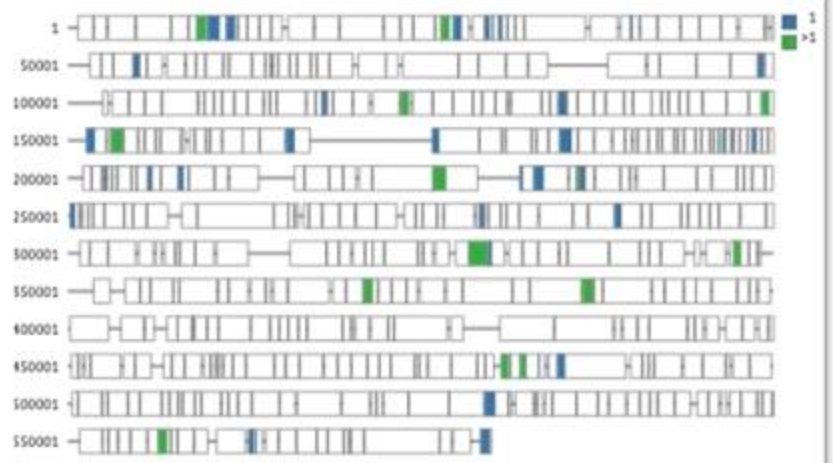
DNA replication, protein occupancy, methylation, & damage
Simulation: Wild-type #1



DNA replication
Simulation: Wild-type #1



Translation
Simulation: Wild-type #1



00:01:

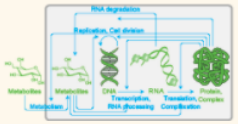
Modèle d'organisme entier

Whole-Cell Modeling
PREDICTING PHENOTYPE FROM GENOTYPE FOR SCIENCE, MEDICINE & ENGINEERING

Models Tools Team Learn more Get involved News About

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.

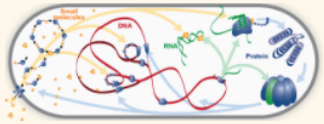
Availability: In development
Author: Karr Lab, Sinai

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.

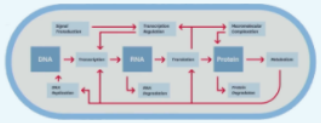
Mycoplasma pneumoniae



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

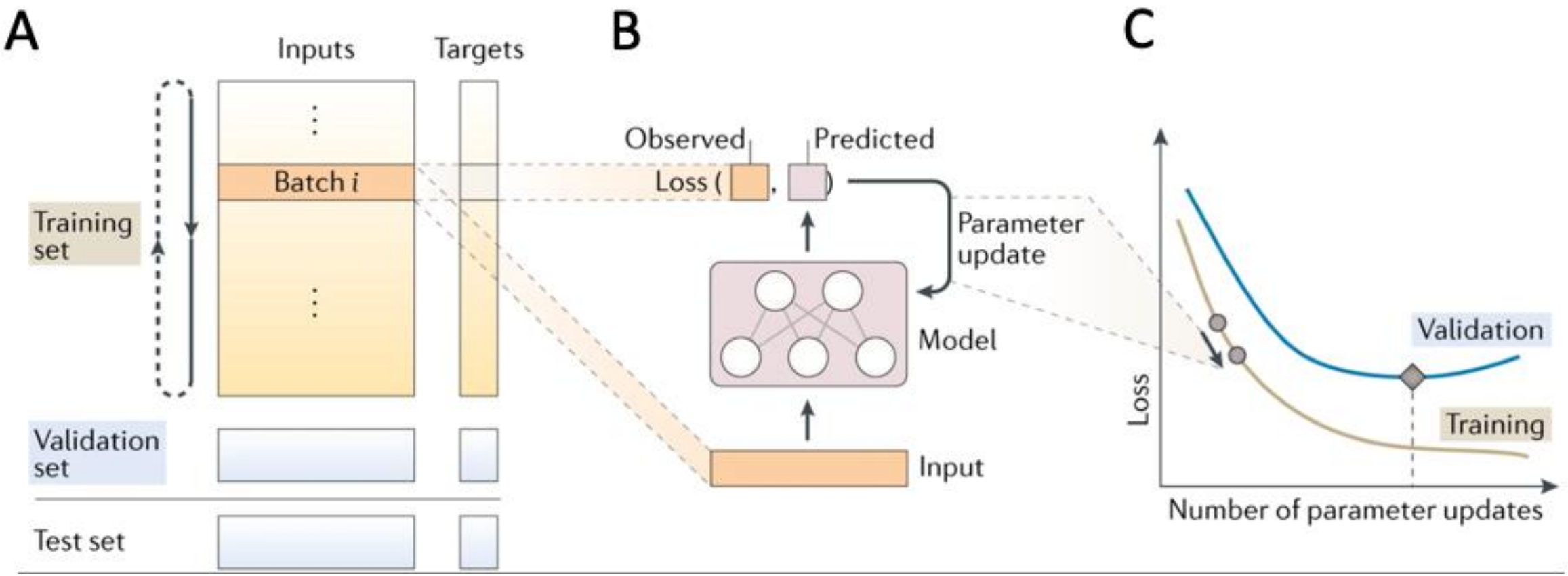
Homo sapiens cancer signaling (MCF10A)

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).

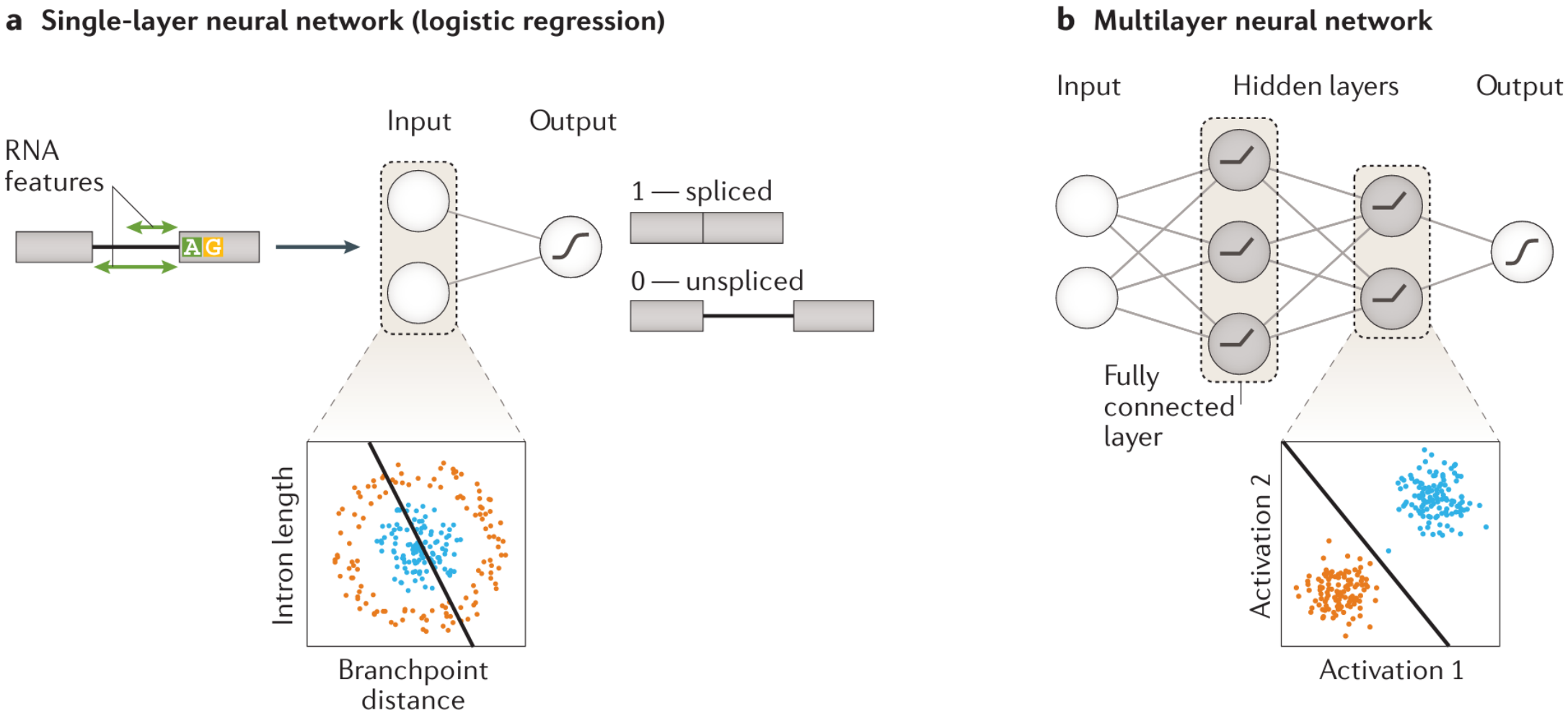
Author: Birtwistle Lab, Clemson

Le Machine Learning : Principe



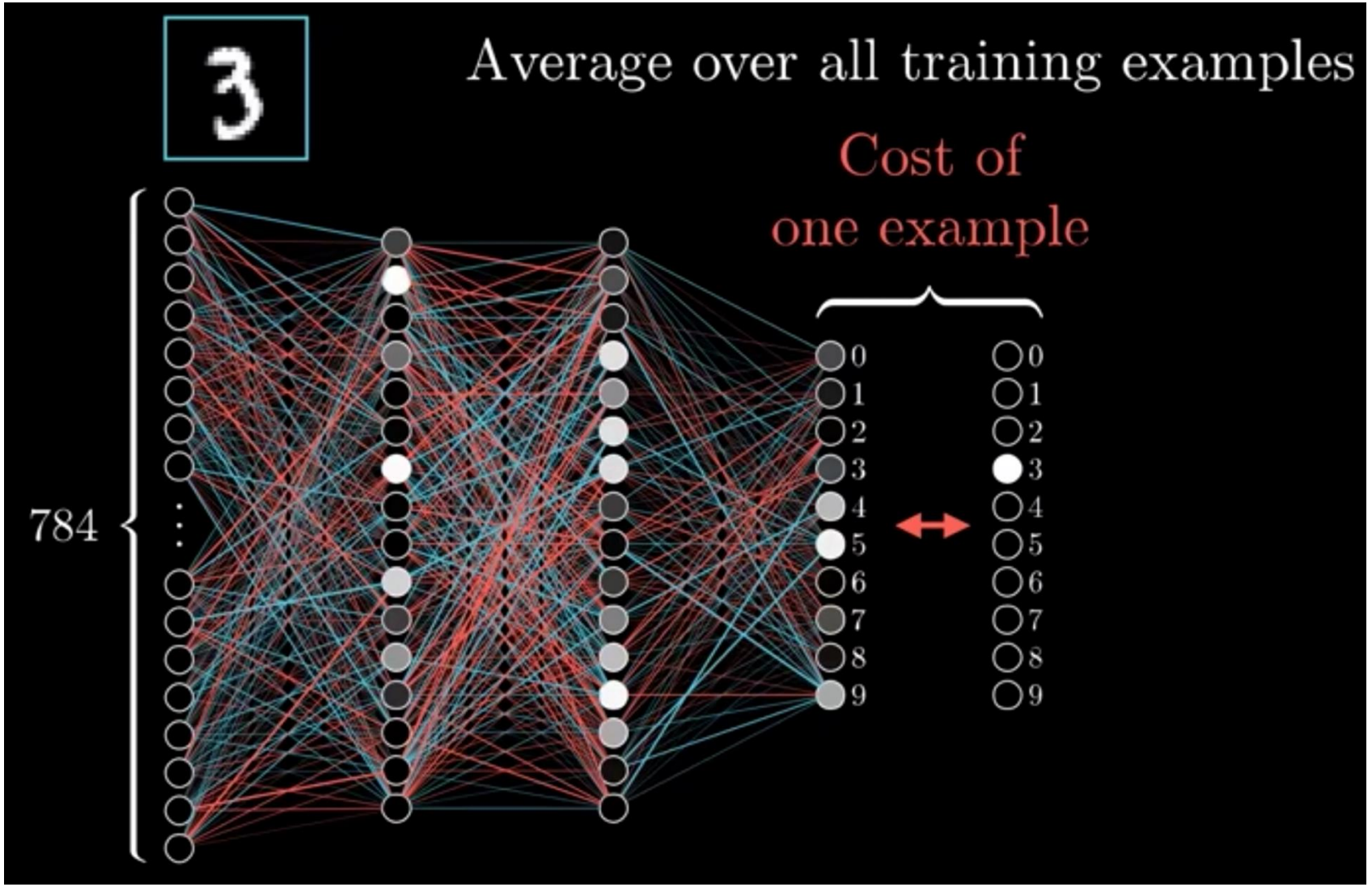
Eraslan, … , Theis, Nature Review Genetics, 2019

Le Deep Learning : Principe

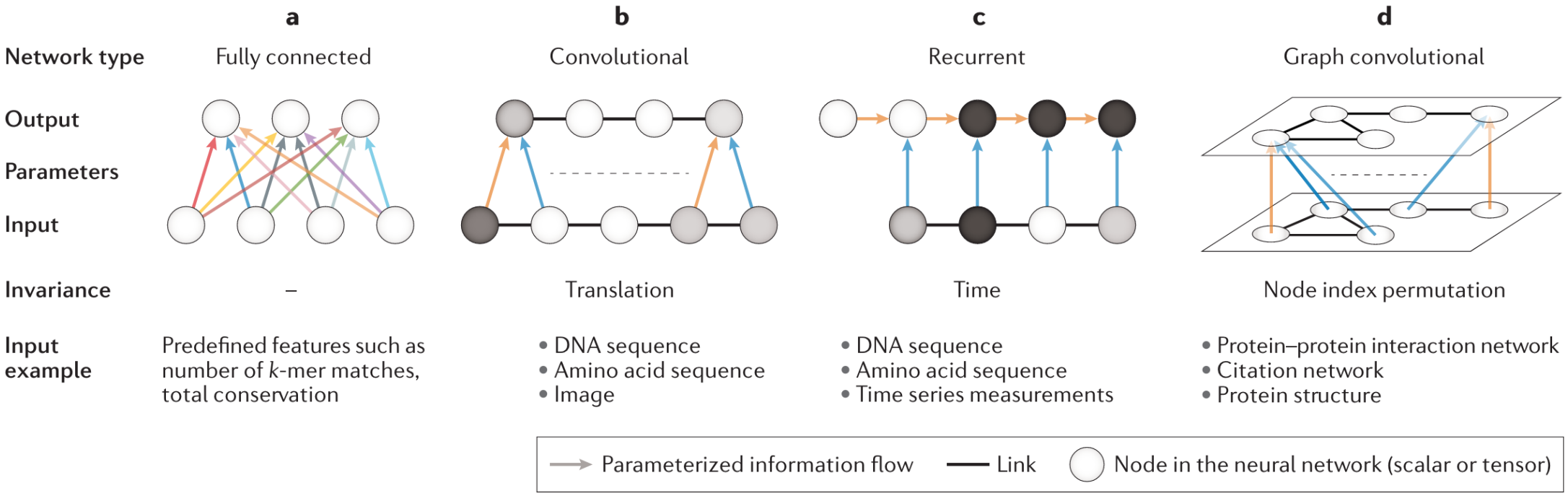


Eraslan, … , Theis, Nature Review Genetics, 2019

Entrainement et loss fonction

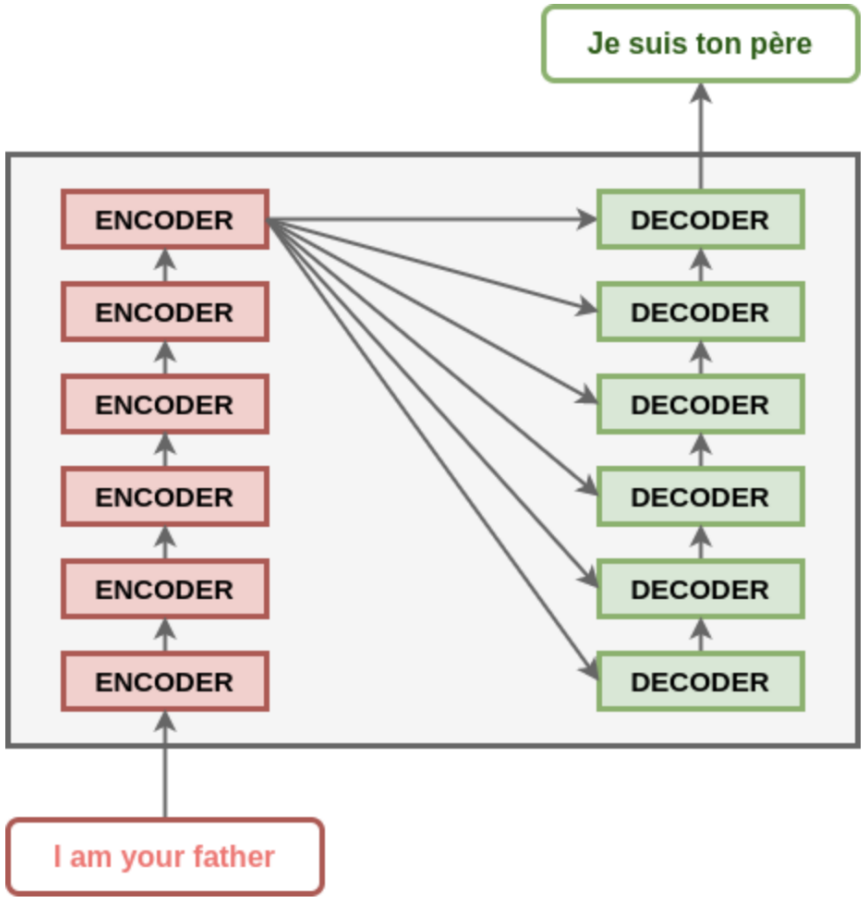


Les 4 architectures de références en deep learning (en 2019!)



Eraslan, … , Theis, Nature Review Genetics, 2019

Game changer : Le Transformer



Architecture du Transformer

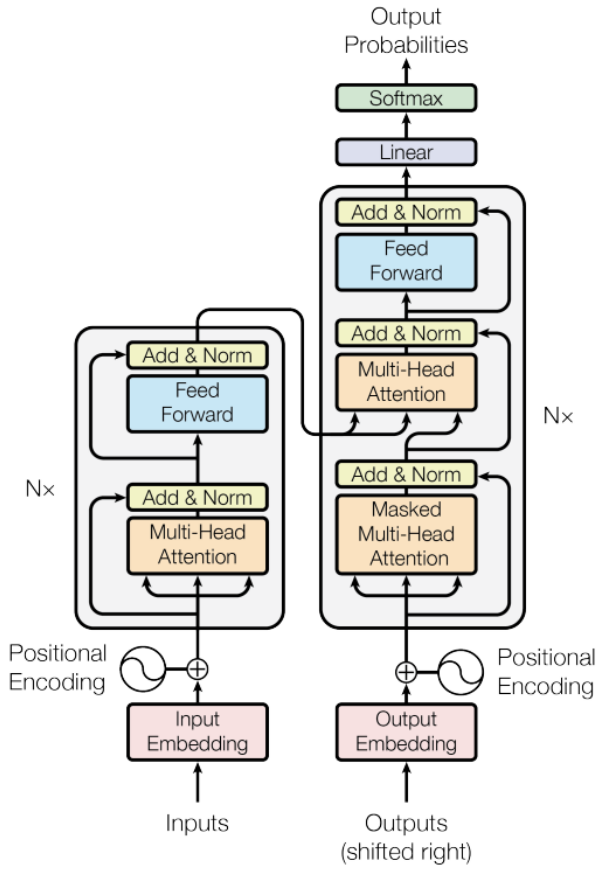


Figure 1: The Transformer - model architecture.

Attention is all you need
Vaswani et al., ArXiv, 2017

Large Language models

GPT-4

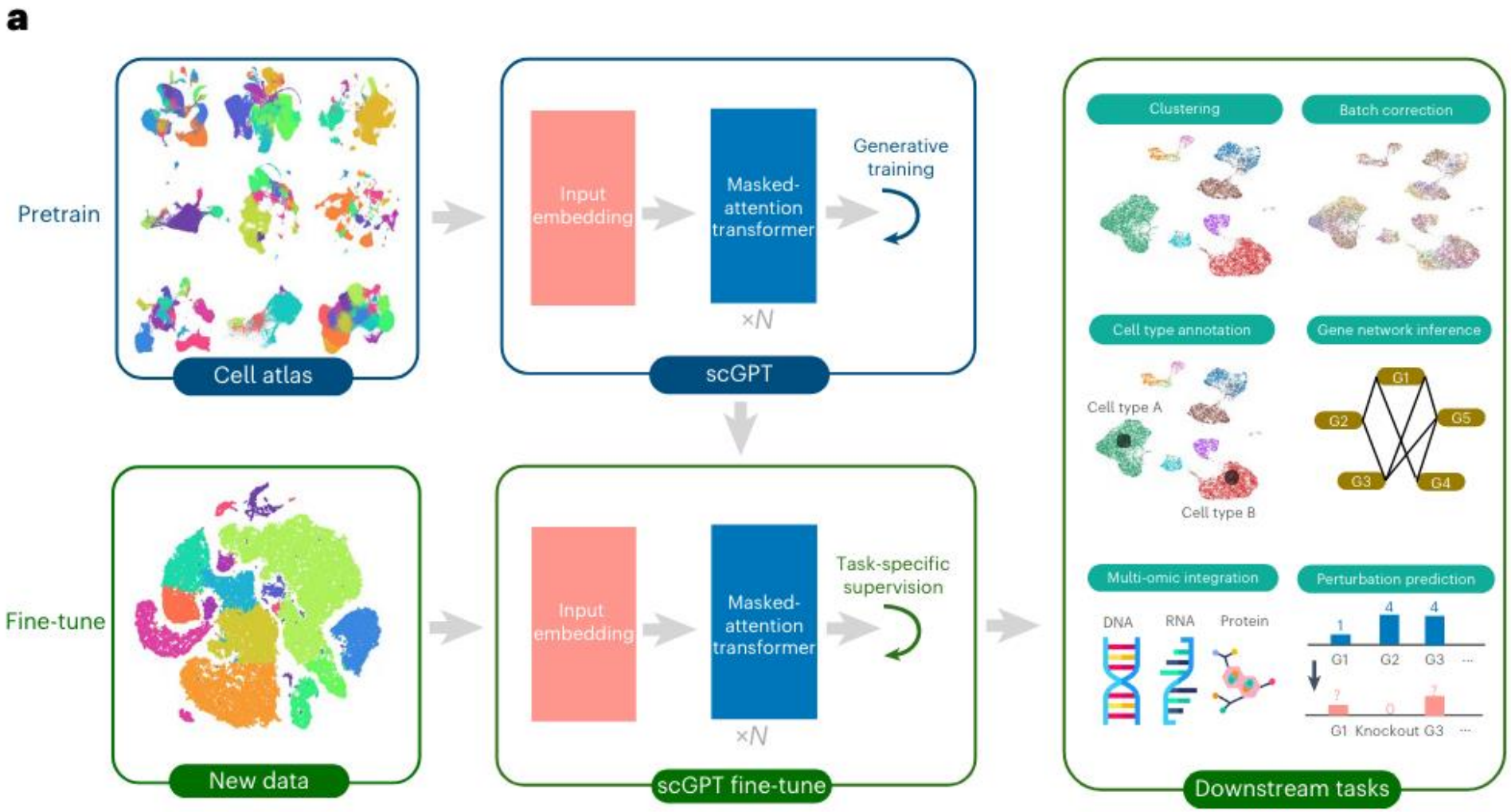
Gemini

deepseek

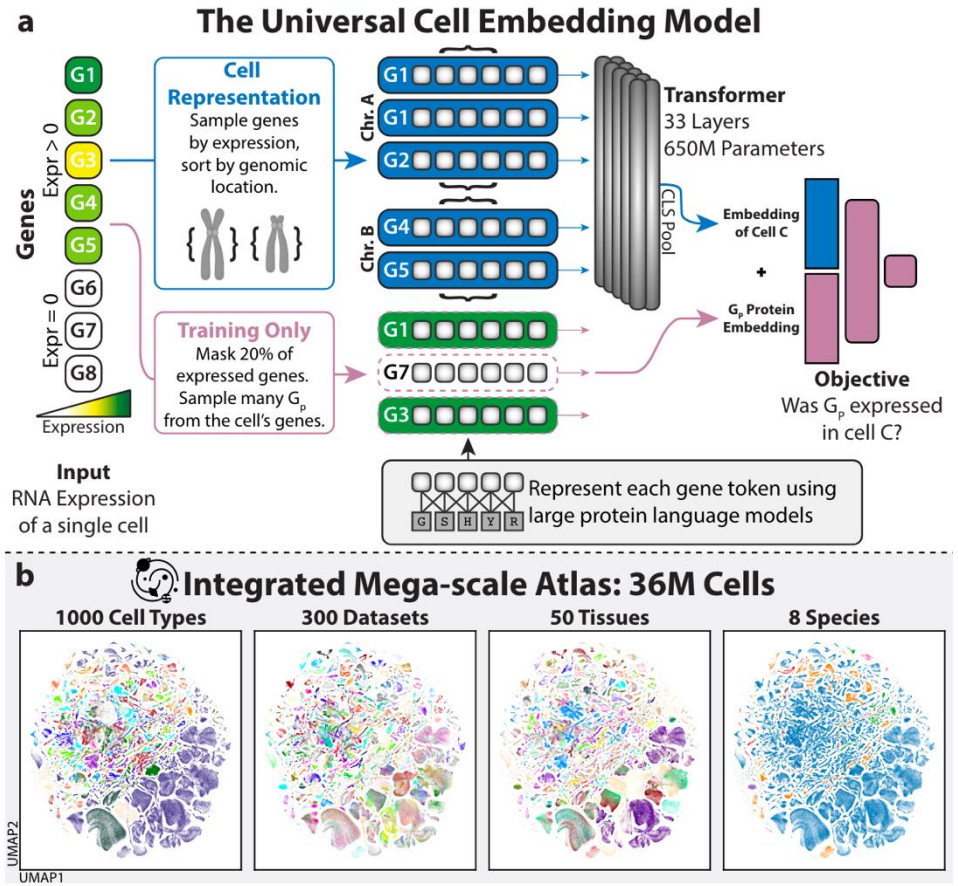
MISTRAL AI_

Foundation models for single-cell

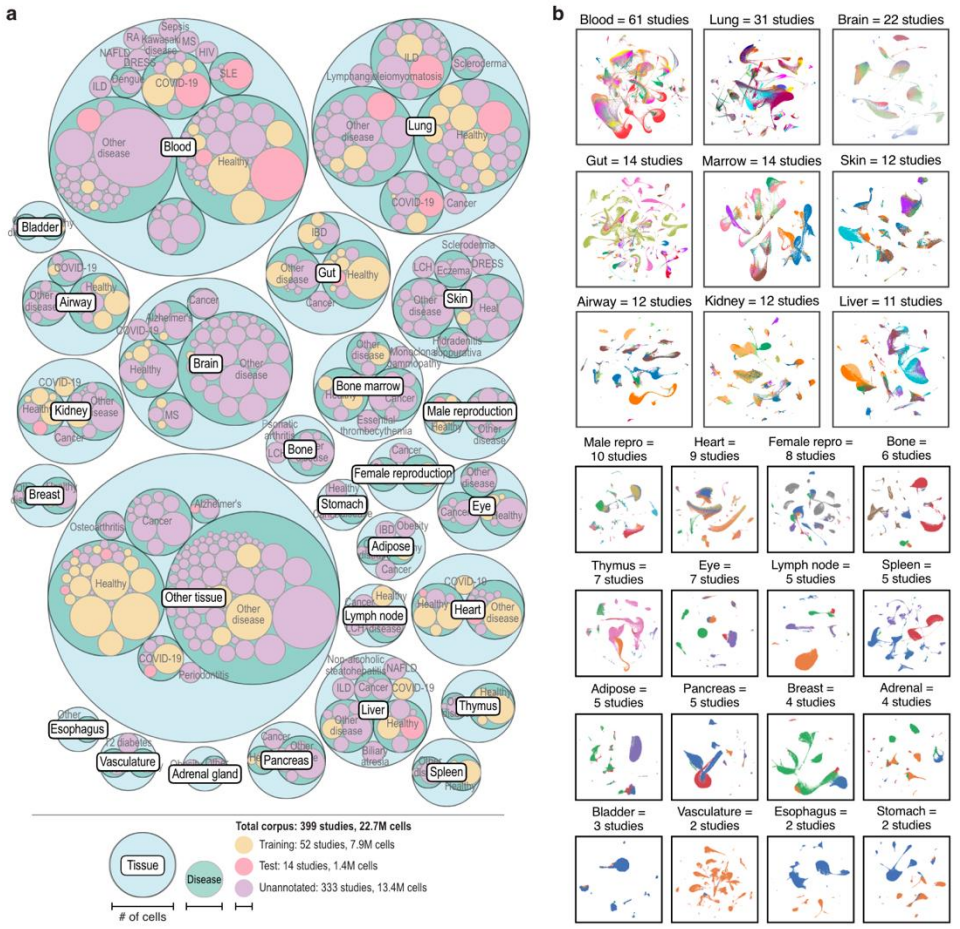
scGPT



Modèles de fondation



scImilarity



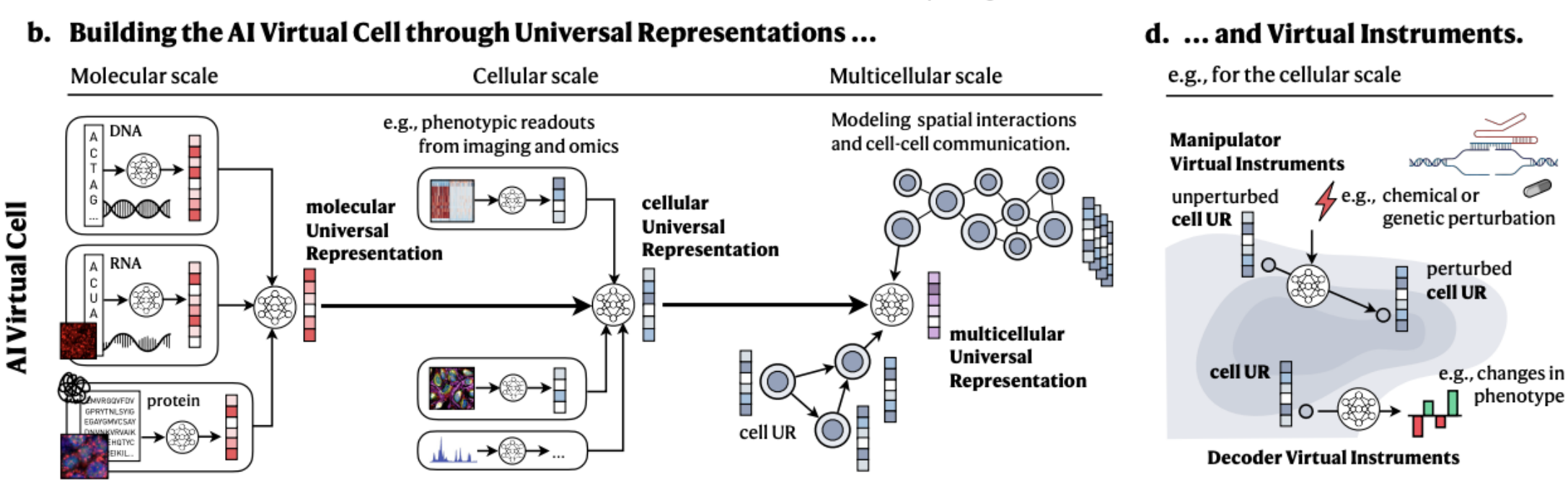
Digital twins / AI Virtual Cells

How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities

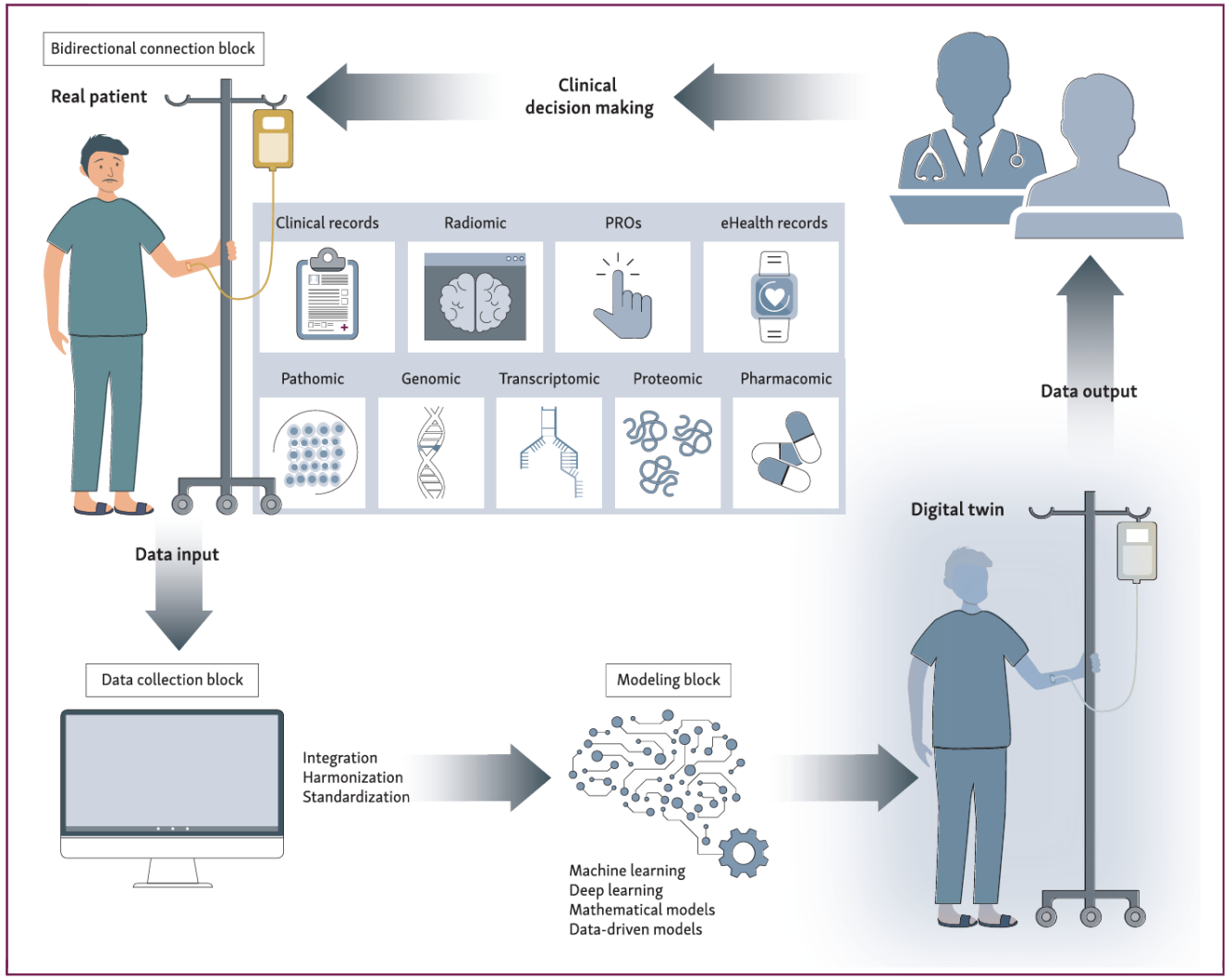
Sept 2024

Charlotte Bunne[1,2,3,4,*], Yusuf Roohani[1,3,5*], Yanay Rosen[1,3,*], Ankit Gupta[3,6], Xikun Zhang[1,3,7], Marcel Roed[1,3], Theo Alexandrov[8,9], Mohammed AlQuraishi[10], Patricia Brennan[3], Daniel B. Burkhardt[10], Andrea Califano[10,12,13], Jonah Cool[3], Abby F. Dernburg[14], Kirsty Ewing[3], Emily B. Fox[1,15,16], Matthias Haury[17], Amy E. Herr[16,18], Eric Horvitz[19], Patrick D. Hsu[5,18,20], Viren Jain[21], Gregory R. Johnson[22], Thomas Kalil[23], David R. Kelley[24], Shana O. Kelley[25,26], Anna Kreshuk[27], Tim Mitchison[28], Stephani Otte[17], Jay Shendure[29,30,31,32], Nicholas J. Sofroniew[33], Fabian Theis[34,35,36], Christina V. Theodoris[37,38], Srigokul Upadhyayula[14,16,39], Marc Valer[3], Bo Wang[40,41], Eric Xing[42,43], Serena Yeung-Levy[1,44], Marinka Zitnik[45,46,47], Theofanis Karaletsos[3,‡], Aviv Regev[2,‡], Emma Lundberg[3,6,7,48,‡], Jure Leskovec[1,3,‡], Stephen R. Quake[3,7,49,‡]

https://arxiv.org/pdf/2409.11654

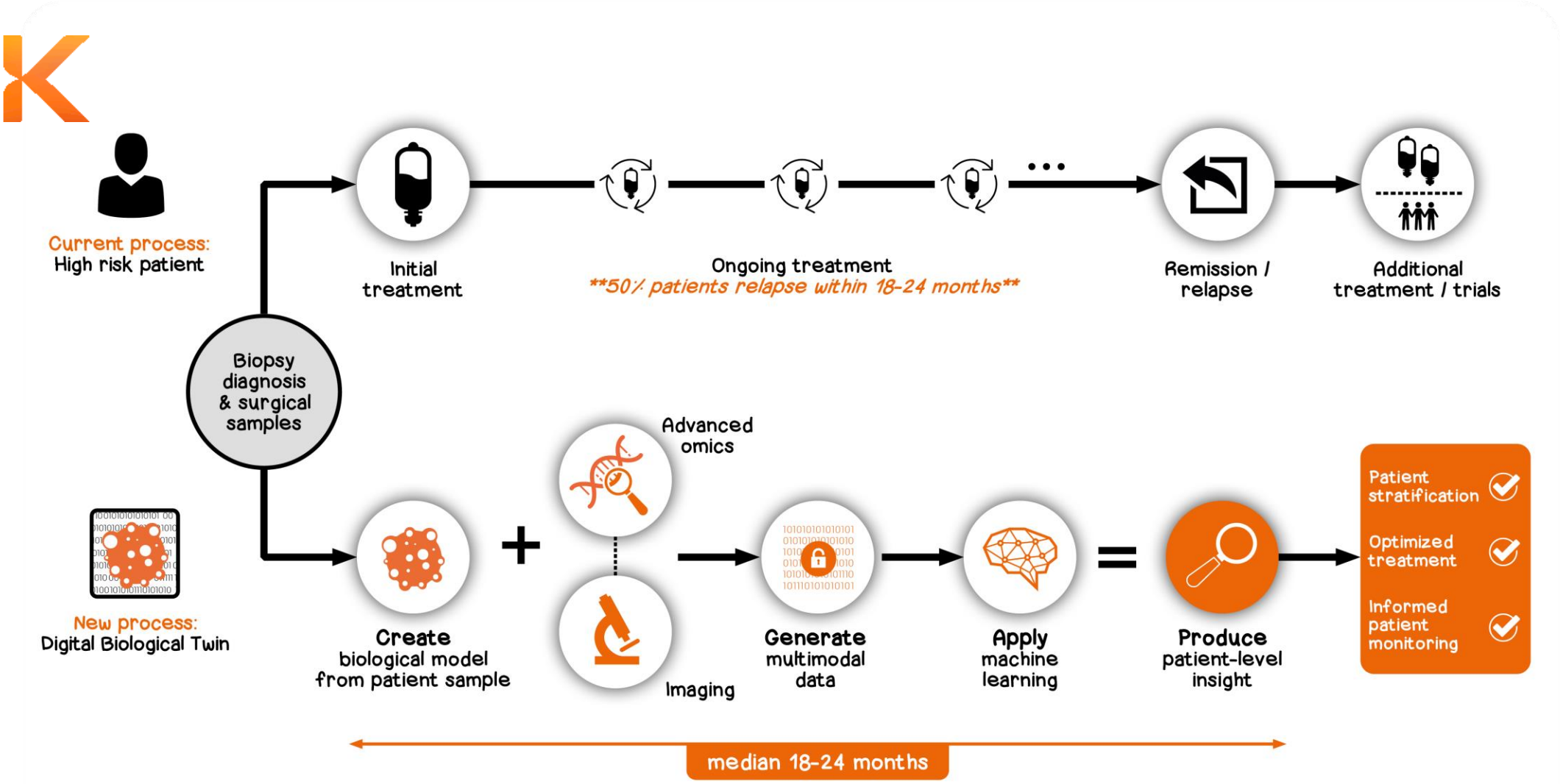


Digital twins en oncologie

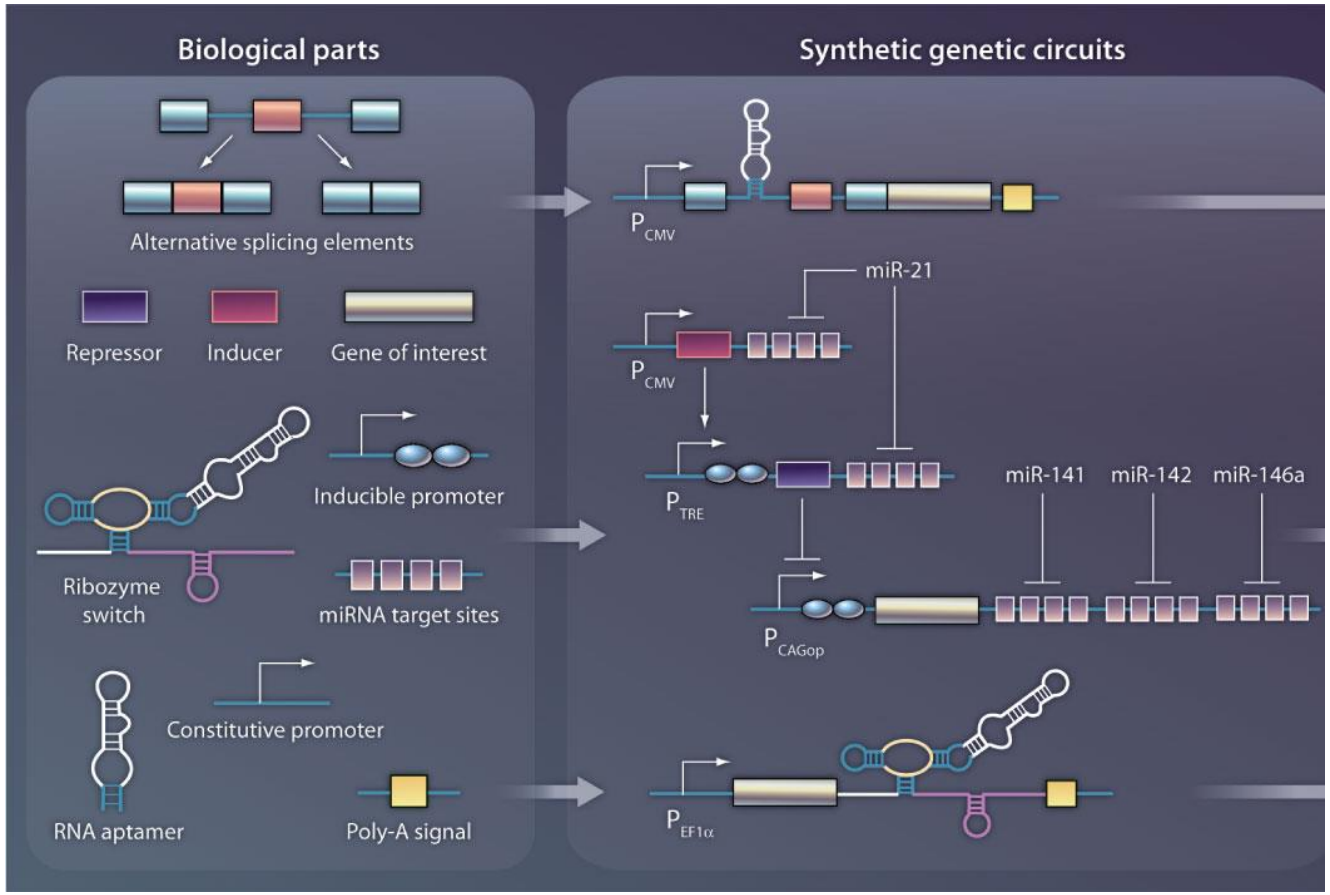


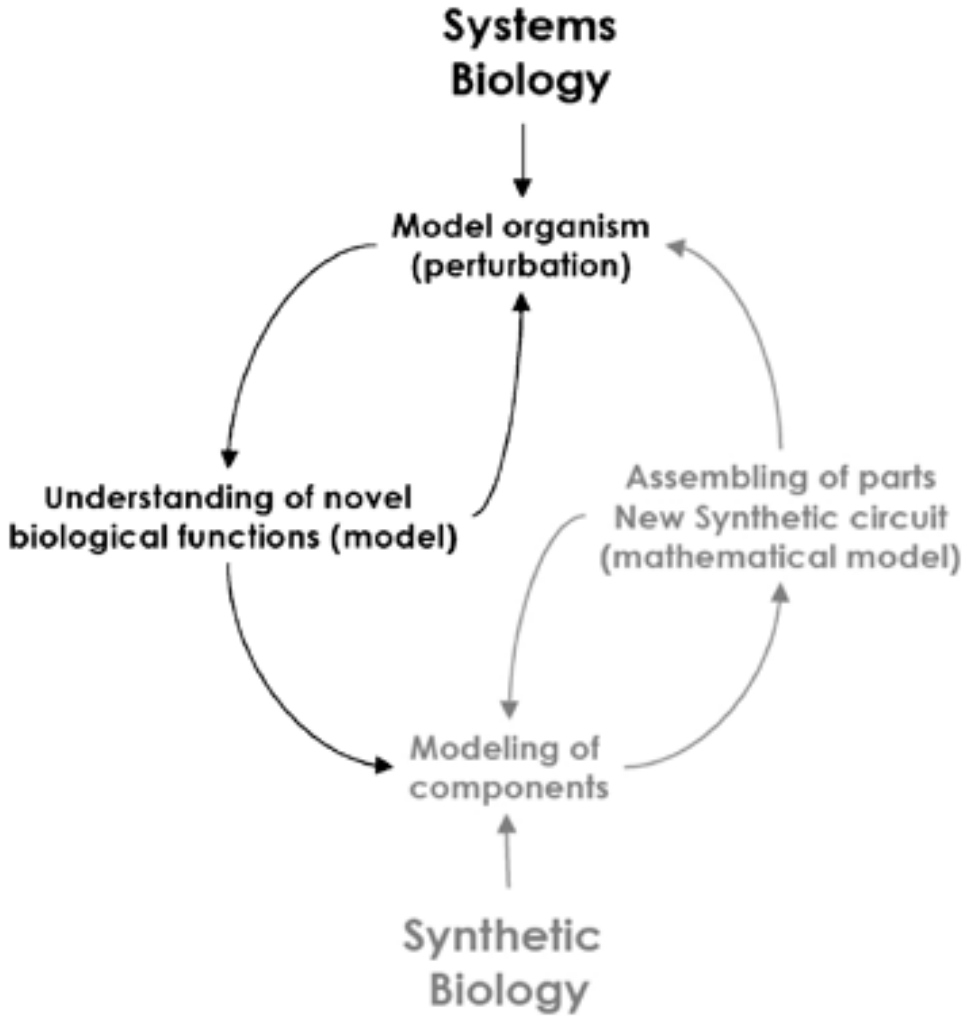
Digital twins: a new paradigm in oncology in the era of big data. ESMO 2024

Digital twins en oncologie



Digital twins pour la biologie synthétique

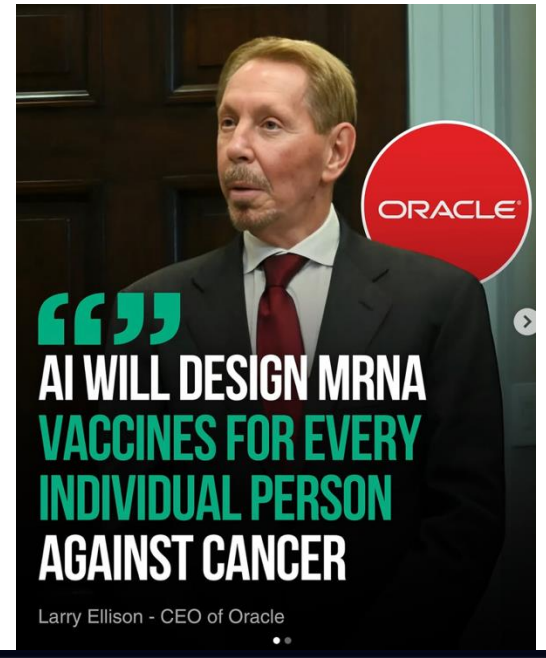




Digital twins / AI Virtual cells

Ex-Google DeepMind and Owkin scientists team up to create Bioptimus to build the first universal AI foundation model for biology

- The French startup has launched with a seed funding of $35M, led by Sofinnova Partners, with Bpifrance Large Venture, and additional funding from top global technology VCs
- Bioptimus will connect the different scales of biology with generative AI; from molecules to cells, tissues and whole organisms, to fuel scientific breakthroughs and accelerate innovation in biomedicine and beyond
- The world-class team, led by Professor Jean-Philippe Vert, brings together Google DeepMind alumni and Owkin experts



✨ OpenAI, Oracle, and SoftBank Usher in "Stargate" Era with $500 Billion AI Infrastructure Initiative

DeepLife is thrilled to announce a $10 million Series A round to propel our AI-powered Digital Twins of Cells platform!

Publié le : 11 décembre 2024

DeepLife.

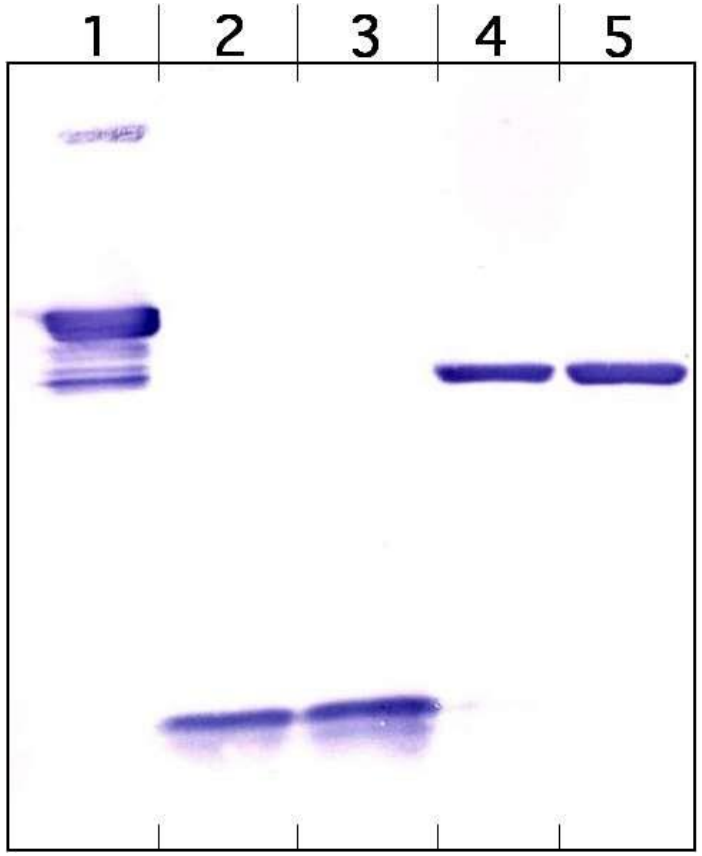
28 Jan 2025

UNIVERSITY ACTIVITIES

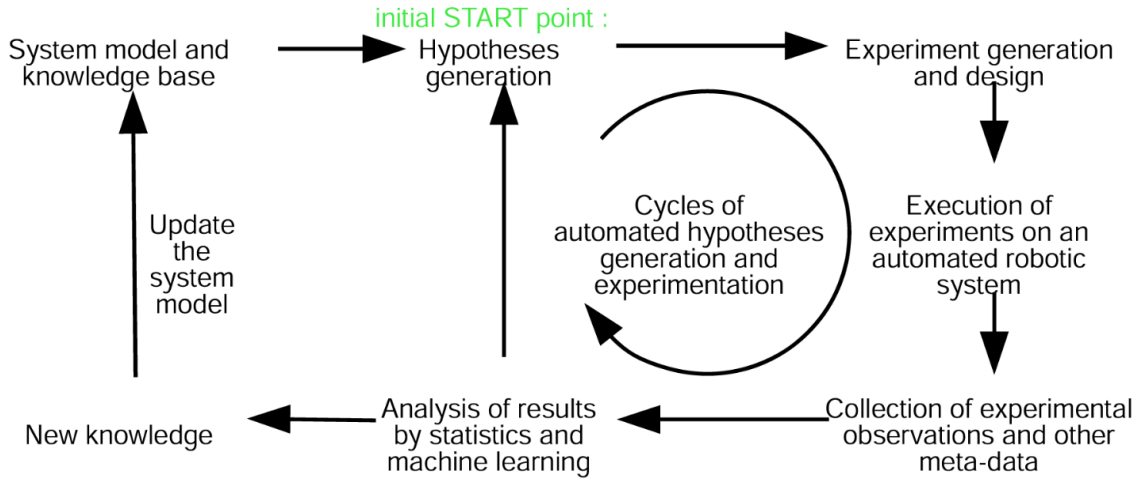
Cell Bauhaus receives $3 million grant for digital twin research into cell behaviour

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”