

# BIOLOGICAL NETWORKS AND HUMAN DISEASES

---

A BIOINFORMATICS VIEWPOINT

M2 TECH SANTE - GHISLAIN BIDAUT – 5 NOVEMBRE 2024

amu  
Aix Marseille Université

# GOALS

---

1. Molecular biology, disease and bioinformatics: Deluge of data
2. Networks: how they facilitate data integration and interpretation
3. Everybody can (and has to) do **bioinformatics**: publicly available databases and tools
4. Network interpretation and some mathematical tools
5. Conclusion and Perspectives

G BIDAUT, CRCM, Aix-Marseille Université

---

## WHAT IS BIOINFORMATICS ?

---



# BIOINFORMATICS

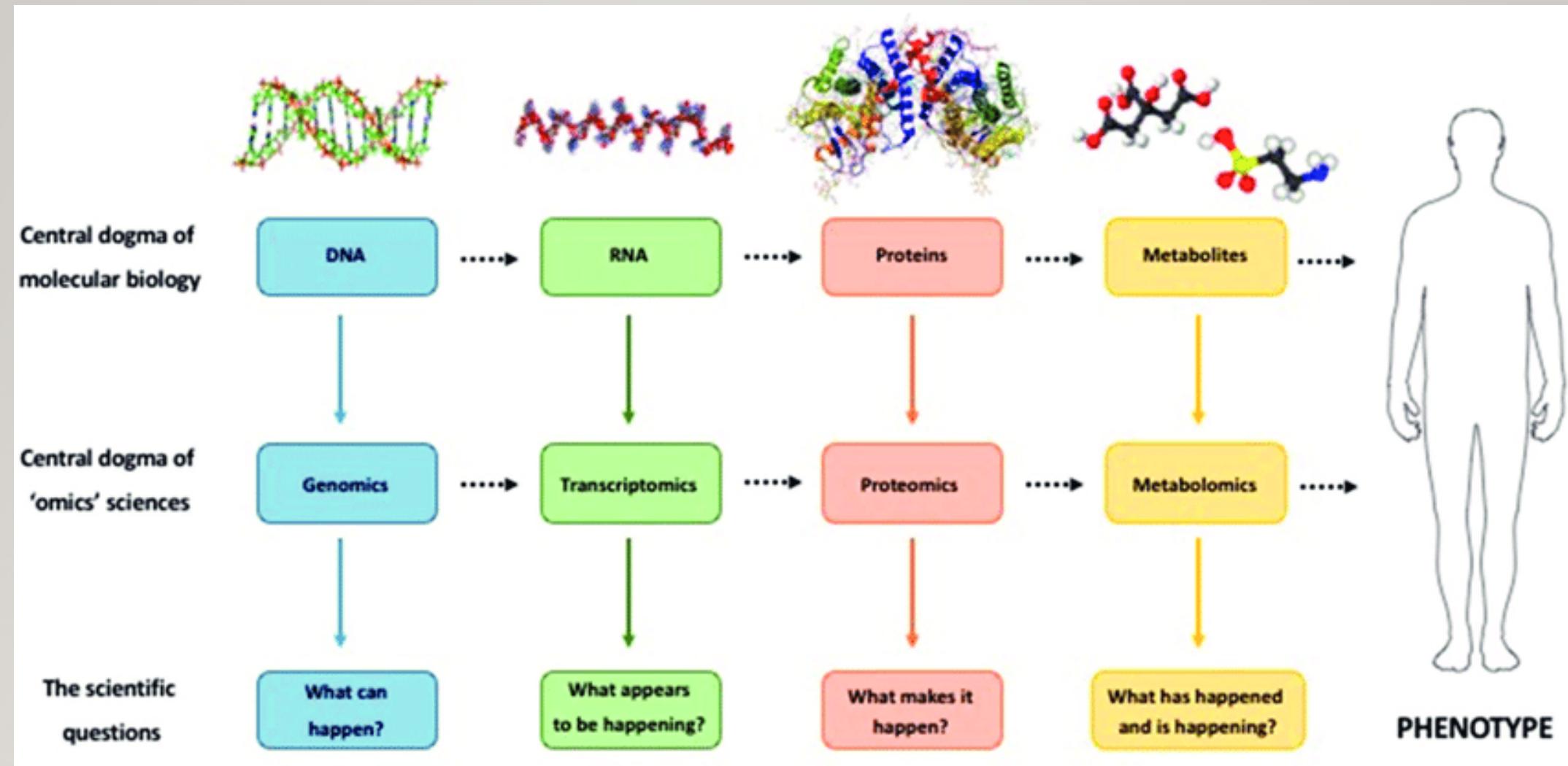
---

- Comparison of genomes among individuals : 99.4% identity
- Bioinformatics: discovery of differences
- Pattern recognition from data:

```
AGGCAGAGGTCATGTCTTATTCA  
TGAGTGGAGCAGAATTGCCAGGCACA  
TCTATCCTACTTCGTTTTCTGTTTCT  
TACAAGGCAGAACGACCTATGTGGGA  
GTGATTACAATTATCACTTAACTTAA  
ACCCTTTATCTGTTTGACAGTCTGGC  
GCAAACAAAATAAAATATCTGTGCAATA  
GAAGACAGAAACGACATGAGCACAGC  
ATACAGTATTGATGAATAATTAAAAAT  
ACTGAAC TGCA GTGGAAATAAGCTATT  
TTTACAGTAAAATGATAGTTATTCCA  
AGAAACCATAATTCAAATGTGTTATT  
CAGAAGACTGAAATATCCTTAAACCTT
```

# WHERE DOES THIS INFORMATION COMES FROM ?

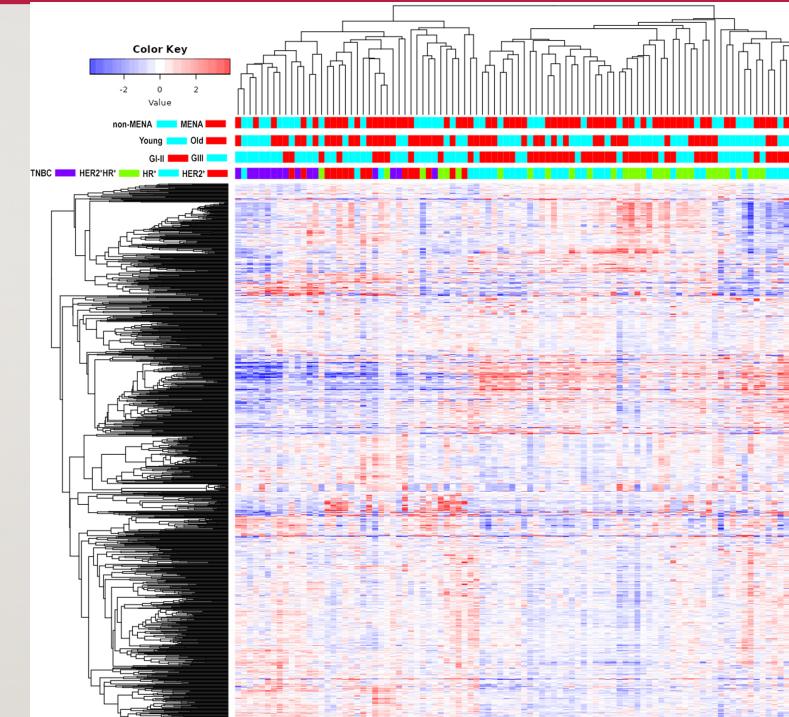




Dettmer et al. (2007)

# TRANSCRIPTOMICS

- Study of gene expression: comparison among different biological conditions:
- **Supervised comparison of experimental conditions** (example: ctrl vs drug).
- **Unsupervised discovery** of molecular profiles in cancers (cancer types, stratification)



Glioblatome subtypes: Elango et al Nature 2023

# SUPERVISED ANALYSIS OF GENE EXPRESSION IN CANCER

---

- Comparison of experimental or clinical conditions
- Supervised: statistical test to link gene expression to sample type
- **RNA-seq** to generate sequences. **Limma**, **EdgeR** or **DeSeq2** to perform statistics
- Public databases of data: The Cancer Genome Atlas. Data available at the **Genomics Data Commons**: <https://portal.gdc.cancer.gov/>
- Publicly available tool for analysis: **Phantasus**: <https://artyomovlab.wustl.edu/phantasus/>

<https://portal.gdc.cancer.gov/>

# Genomic Data Commons Data Portal

## Harmonized Cancer Datasets

A repository and computational platform for cancer researchers who need to understand cancer, its clinical progression, and response to therapy.

[Explore Our Cancer Datasets](#)

## Data Portal Summary

[Data Release 41.0 - August 28, 2024](#)



86  
Projects



69  
Primary Sites



44736  
Cases



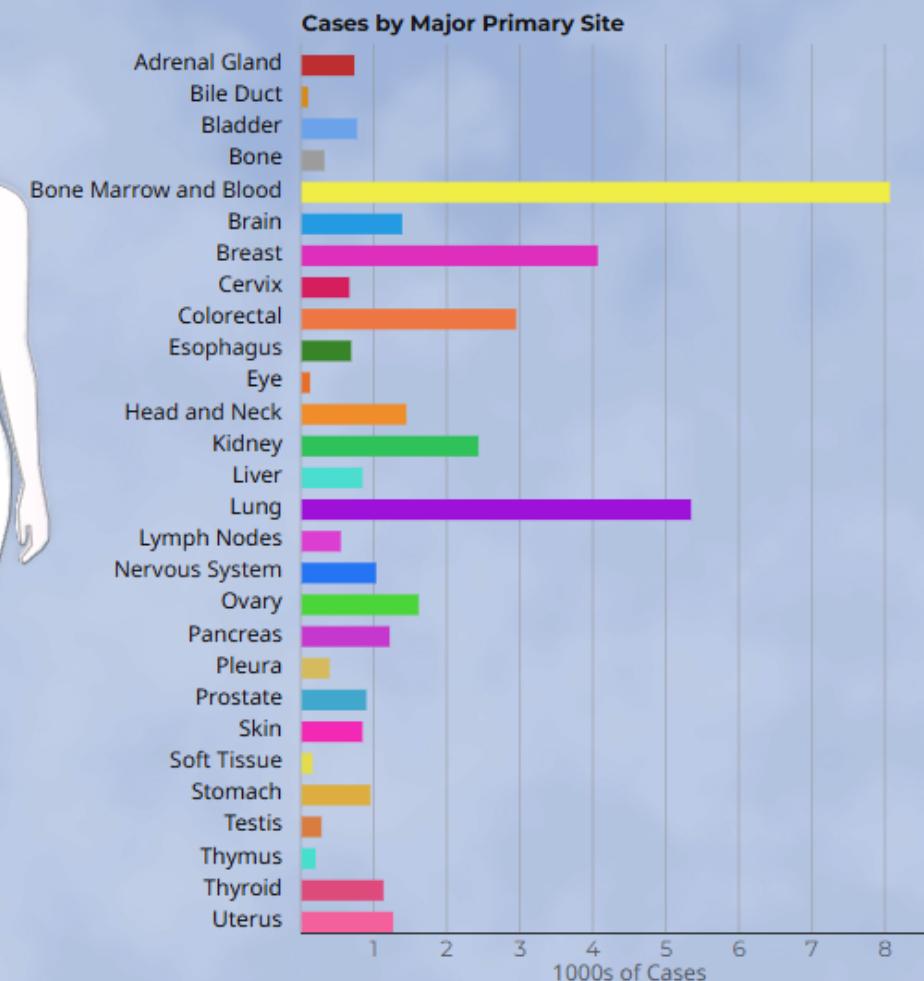
1027517  
Files



22534  
Genes



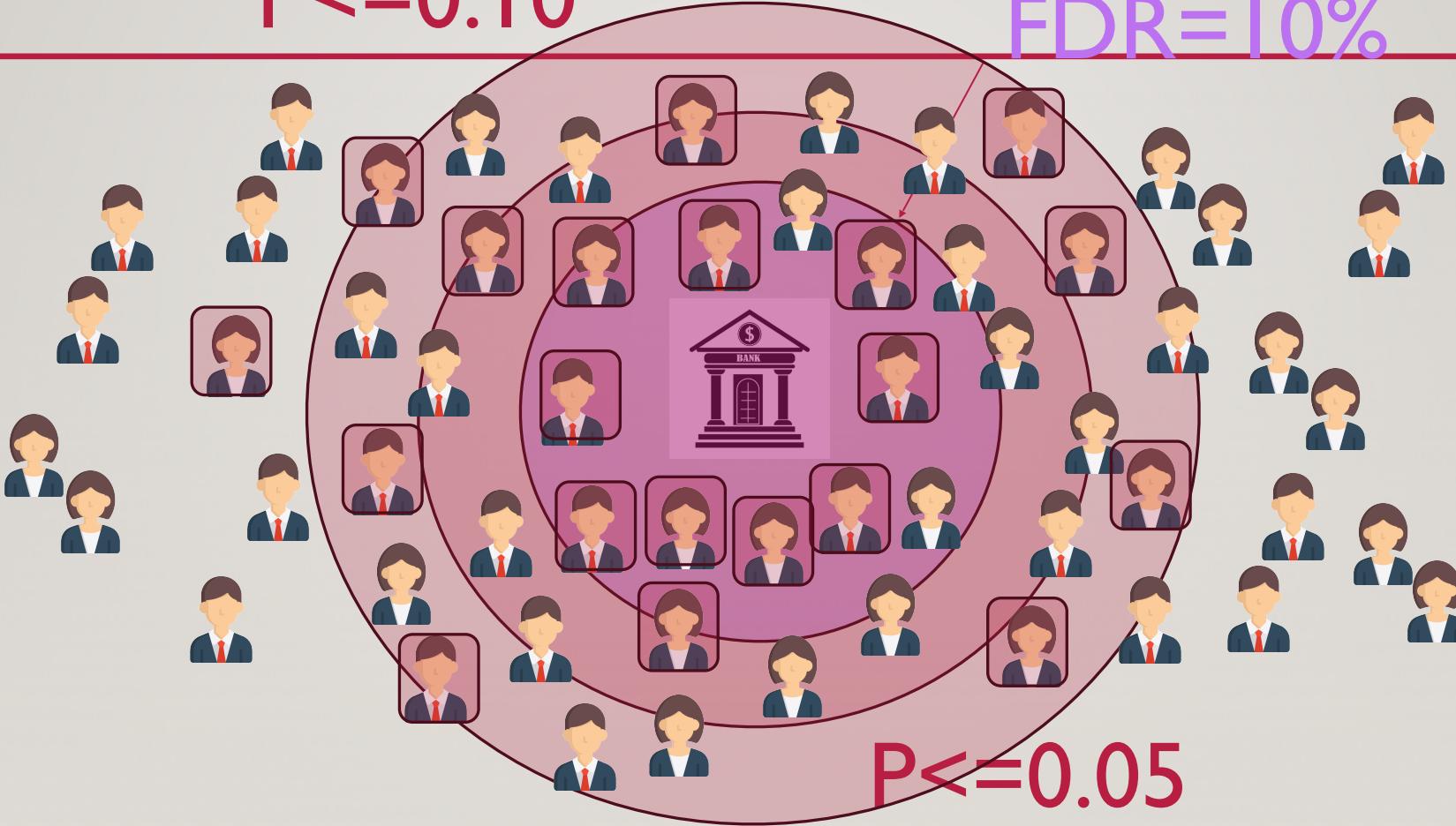
2940240  
Mutations



# BIOINFORMATICS: CRIME INVESTIGATION

$P \leq 0.10$

FDR = 10%



# USUAL SUSPECTS IN CANCEROLOGY

---

## ONCOGENES

- Growth factors
- Signaling
- Cell cycle genes
- Angiogenesis
- Telomeres
- KRAS, EGFR, MYC, VEGF, ...
- **Disease cannot be reduced to a single gene!**

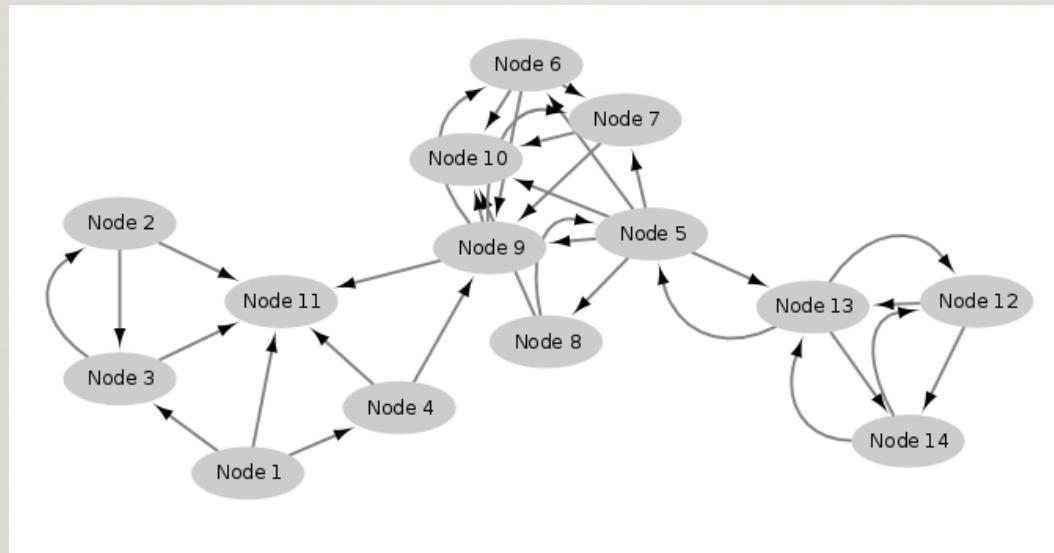
## TUMOR SUPPRESSORS

- Cell cycle inhibitors
- DNA repair enzymes
- Telomere maintenance
- Programmed cell death
- P53, BRCA1,2, RBI, PTEN,...
- Global interaction with oncogenes.

# WHAT IS A BIOLOGICAL NETWORK ?

---

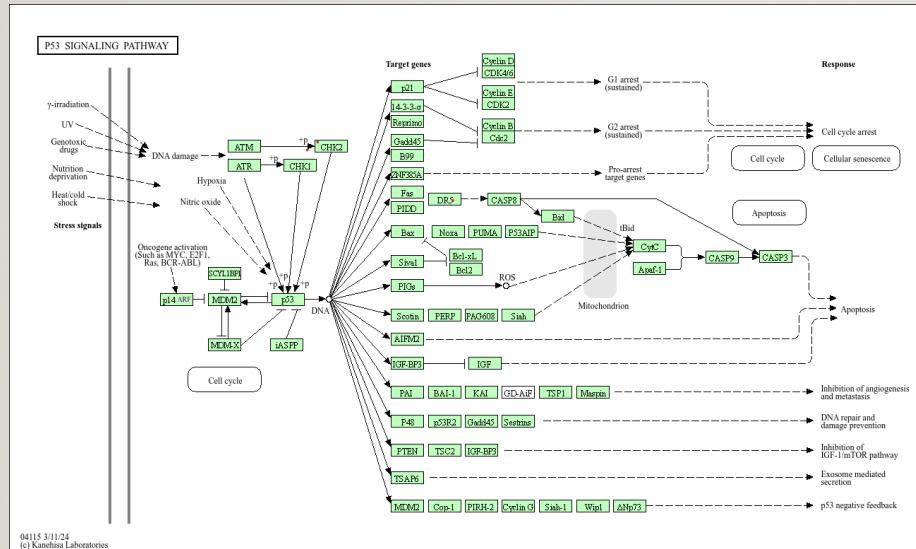
- **Définition :** A biological network represents interactions between components of a biological system
- **For example,** proteins, genes, or metabolites can be the **nodes**, and the interactions, the **links**.



# CANONICAL PATHWAYS VS EXPERIMENTALLY DEFINED NETWORKS

## CANONICAL: KEGG PATHWAYS (KYOTO ENCYCLOPEDIA OF GENES AND GENOMES)

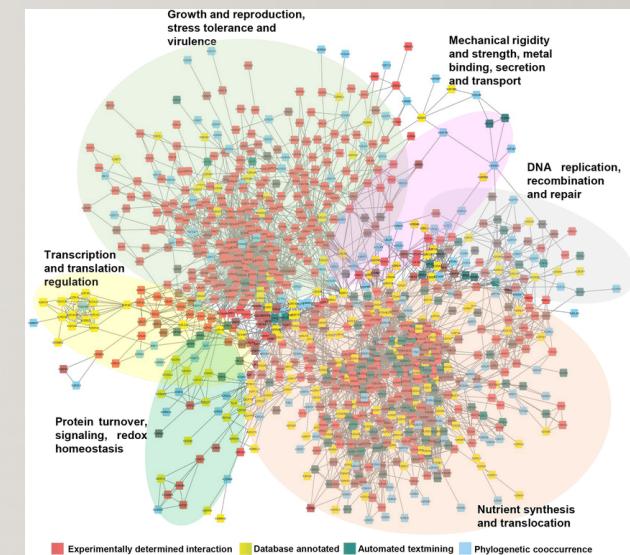
- KEGG: <https://www.genome.jp>
- Generic among organisms
- Theoretical (books, **Robust**)
- **Few** data (dozens of interactions)
- Easy to visualize



<https://www.genome.jp/pathway/hsa04115>

## EXPERIMENTAL: PROTEIN-PROTEIN INTERACTION NETWORKS

- Mathematical representation of association between proteins in the cell
- Organism-specific
- Directly measured in the lab (**Noisy**)
- **Massive** data (Thousands of interactions)
- Not easy to visualize



Arafat et al 2022 Journal of Proteins and Proteomics

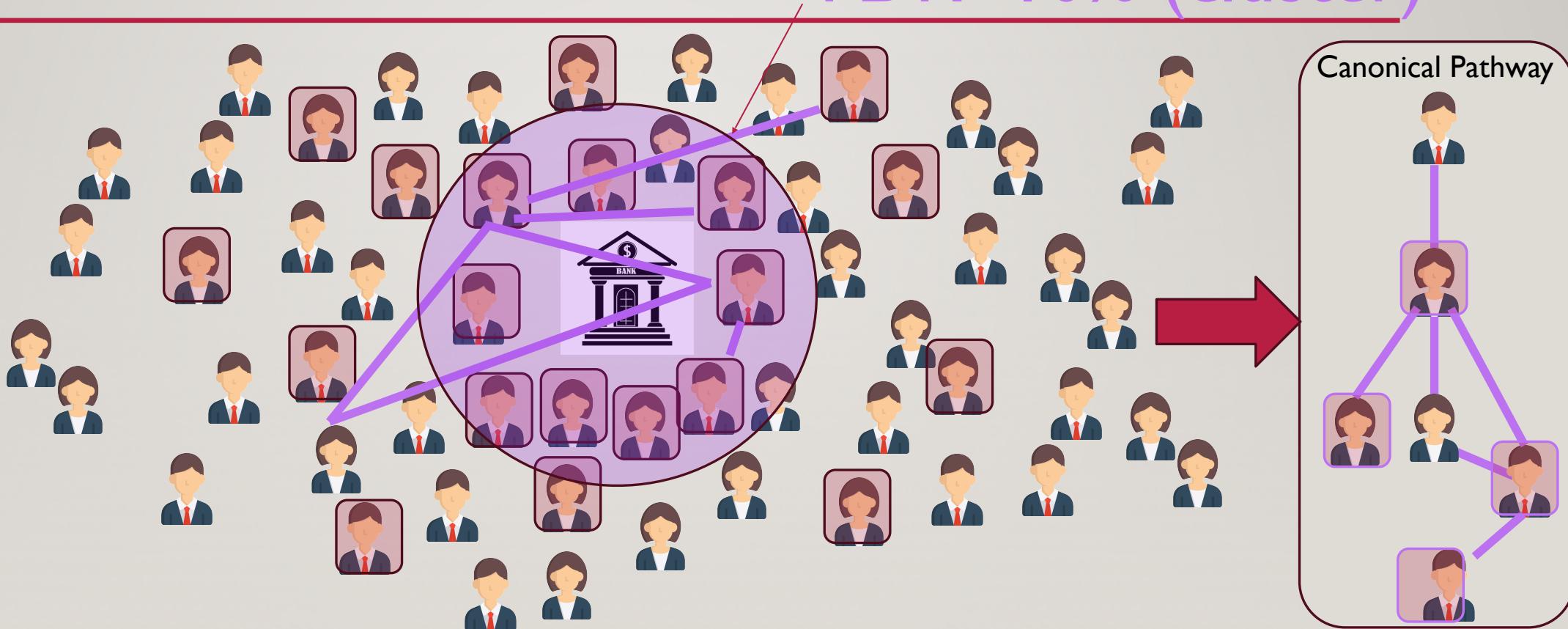
# WHY USING NETWORKS IN BIOLOGY ?

---

- ***They are everywhere !*** Molecular networks, cell-cell communication, social networks, etc...
- ***They are powerful!***
  - **Reduce** complexity compared to a table
  - **Intuitive**
  - Great for **data visualisation/ integration**
- ***Why are they useful ?*** They allow **discovering genes function or role in a disease** through « **guilt by association** »

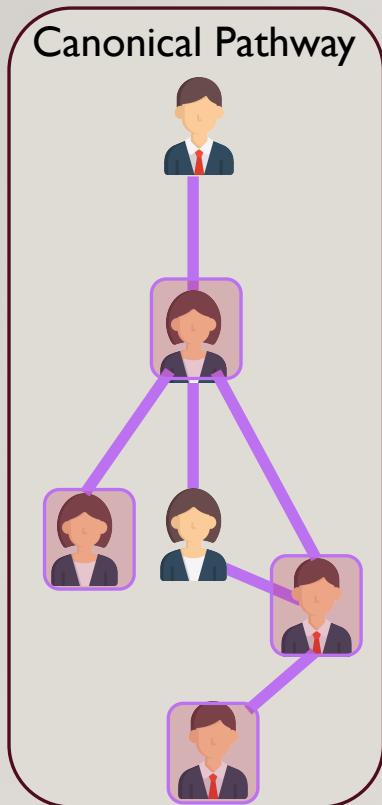
# BIOINFORMATICS: CRIME INVESTIGATION

FDR=10% (cluster)



# PATHWAY ENRICHMENT

---



## ENRICHMENT

- 10 genes detected in Cluster
- 4 members of the Cancer Pathway detected in cluster
- Size of the Canonical pathway = 6 genes
- 150 genes studied (25k genes for pan genome analysis)

# CONTINGENCY TABLE

---

	Cluster	Non Cluster
Cancer Pathway	4	2
Non Cancer Pathway	6	100

- **Easy Fisher Exact Test Calculator:** <https://www.socscistatistics.com/tests/fisher/default2.aspx>
- The Fisher exact test statistic value is **0.0005**. The result is significant at  $p < .05$ .
- What does that means ?

# REAL LIFE PATHWAY ENRICHMENT

Author Manuscript

Author Manuscript



## HHS Public Access

Author manuscript

*Cancer Res.* Author manuscript; available in PMC 2019 September 01.

Published in final edited form as:

*Cancer Res.* 2018 September 01; 78(17): 4971–4983. doi:10.1158/0008-5472.CAN-17-3822.

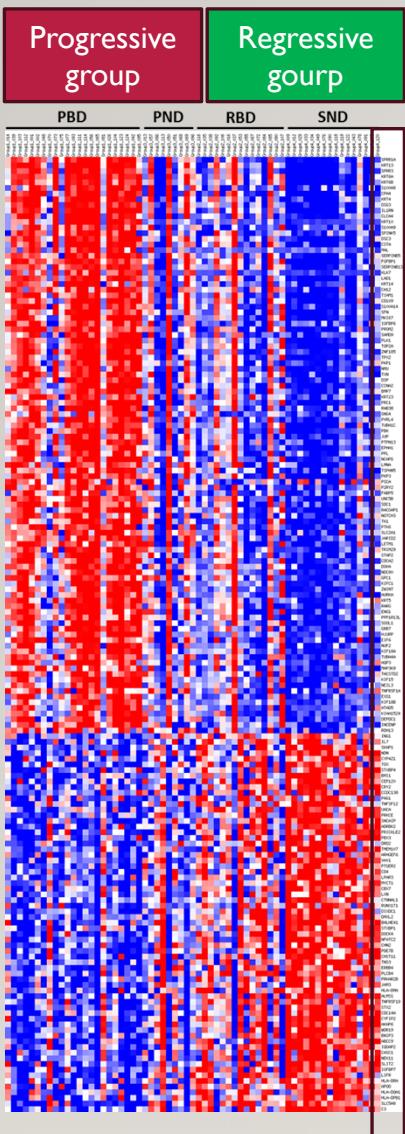
### Altered cell-cycle control, inflammation and adhesion in high-risk persistent bronchial dysplasia

Daniel T. Merrick<sup>1</sup>, Michael G. Edwards<sup>2</sup>, Wilbur A. Franklin<sup>1</sup>, Michio Sugita<sup>1</sup>, Robert L. Keith<sup>3,4</sup>, York E. Miller<sup>3,4</sup>, Micah B. Friedman<sup>1</sup>, Lori D. Dwyer-Nield<sup>3,5</sup>, Meredith A. Tennis<sup>4</sup>, Mary C. O'Keefe<sup>6</sup>, Elizabeth J. Donald<sup>1</sup>, Jessica M. Malloy<sup>1</sup>, Adrie van Bokhoven<sup>1</sup>, Storey Wilson<sup>1</sup>, Peter J. Koch<sup>7</sup>, Charlene O'Shea<sup>7</sup>, Christopher Coldren<sup>8</sup>, David J. Orlicky<sup>1</sup>, Xian Lu<sup>9</sup>, Anna E. Baron<sup>9</sup>, Greg Hickey<sup>4</sup>, Timothy C. Kennedy<sup>4</sup>, Roger Powell<sup>5</sup>, Lynn Heasley<sup>10</sup>, Paul A. Bunn<sup>11</sup>, Mark Geraci<sup>12</sup>, and Raphael A. Nemenoff<sup>4,13</sup>

<sup>1</sup>Department of Pathology, University of Colorado Anschutz Medical Campus

<sup>2</sup>Department of Medicine/Division of Pulmonary Medicine, University of Colorado Anschutz Medical Campus

- Comparison of Persistent and Regressive Bronchial Dysplasia (BD) in Lung Cancer (SCC).
- Bioinformatics Analysis:
  - Differentially expressed genes
  - Pathway Enrichment



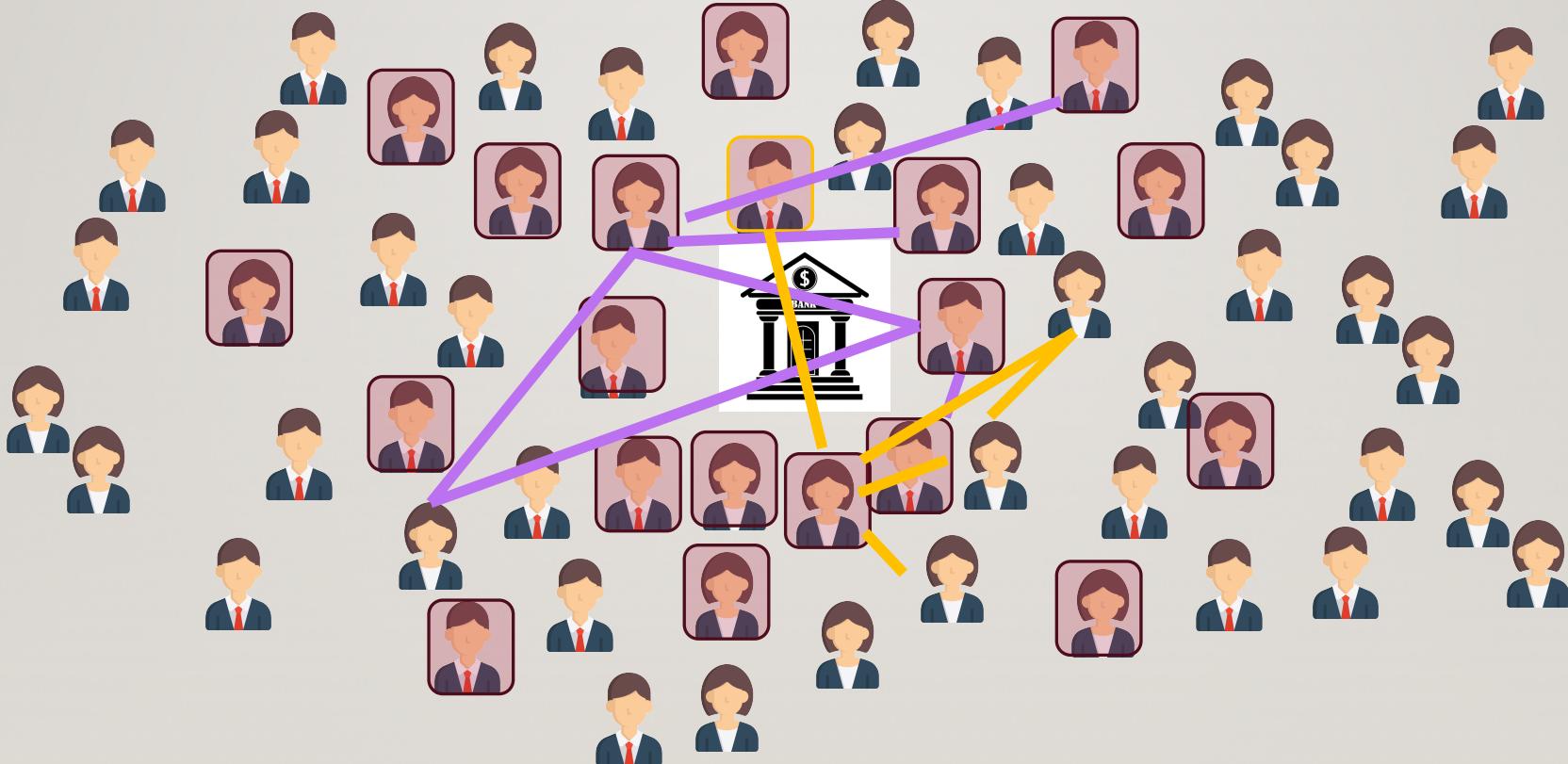
```
graph LR; A[Gene List] --> B[Enrichment]
```

Pathway	p-value	Genes
Antigen Presentation Pathway	3.16E-04	HLA-DPB1,HLA-DMA,HLA-DQA1,HLA-DPA1,HLA-DRA
Calcium-induced T Lymphocyte Apoptosis	6.03E-04	PRKCE,NFATC2,HLA-DMA,HLA-DQA1,CD4,HLA-DRA
Cdc42 Signaling	1.70E-03	HLA-DPB1,IQGAP3,ARHGEF6,HLA-DMA,HLA-DQA1,HLA-DPA1,IQGAP2,VAV1,HLA-DRA
B Cell Development	2.19E-03	HLA-DMA,HLA-DQA1,HLA-DRA,IL7
PKCθ Signaling in T Lymphocytes	3.16E-03	NFATC2,HLA-DMA,HLA-DQA1,CD4,VAV1,MAP3K9,HLA-DRA
OX40 Signaling Pathway	3.31E-03	HLA-DPB1,HLA-DMA,HLA-DQA1,HLA-DPA1,CD4,HLA-DRA
Gluconeogenesis I	7.59E-03	ENO4,MDH1B,ENO1
Graft-versus-Host Disease Signaling	7.59E-03	IL1RN,HLA-DMA,HLA-DQA1,HLA-DRA
Cell Cycle: G2/M DNA Damage Checkpoint Regulation	8.13E-03	TOP2A,SFN,AURKA,PLK1
iCOS-iCOSL Signaling in T Helper Cells	8.51E-03	NFATC2,HLA-DMA,HLA-DQA1,CD4,VAV1,HLA-DRA
14-3-3-mediated Signaling	1.23E-02	PRKCE,SFN,TNFRSF1A,PLCB4,TUBA4A,TUBA1C
CD28 Signaling in T Helper Cells	1.29E-02	NFATC2,HLA-DMA,HLA-DQA1,CD4,VAV1,HLA-DRA
Allograft Rejection Signaling	1.32E-02	HLA-DPB1,HLA-DMA,HLA-DQA1,HLA-DPA1,HLA-DRA
Role of IL-17A in Psoriasis	1.82E-02	S100A8,S100A9

Merrick et al. *Cancer Research* 2019

# SPECIALIZED NETWORKS: DISCOVERY OF NEW PATHWAYS THROUGH CONNECTION

---



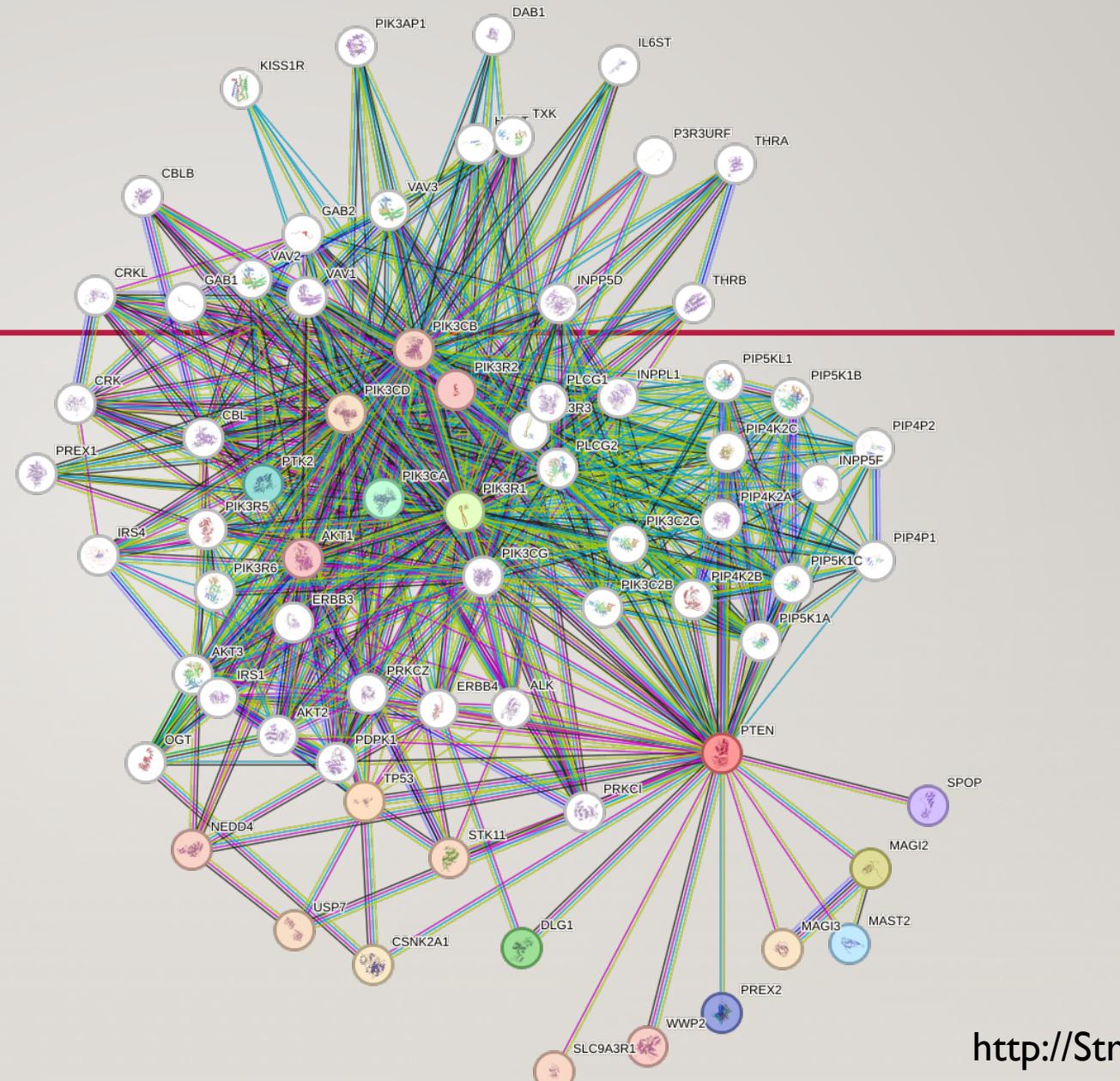
# SPECIALIZED NETWORKS

---

- These networks are activated on **special circumstance** as the cell are in '**cancer**' or '**genetic disease**' mode
- Also, sometimes cell need to **recruit genes** very quickly – they don't have time to **produce it**, so there is **no change in expression** of these quickly recruited genes, since they will use mRNAs **that are already present**.
- These networks contains important molecules that **are not necessarily differentially expressed**. These can be detected by networks tools.

# STRING DB

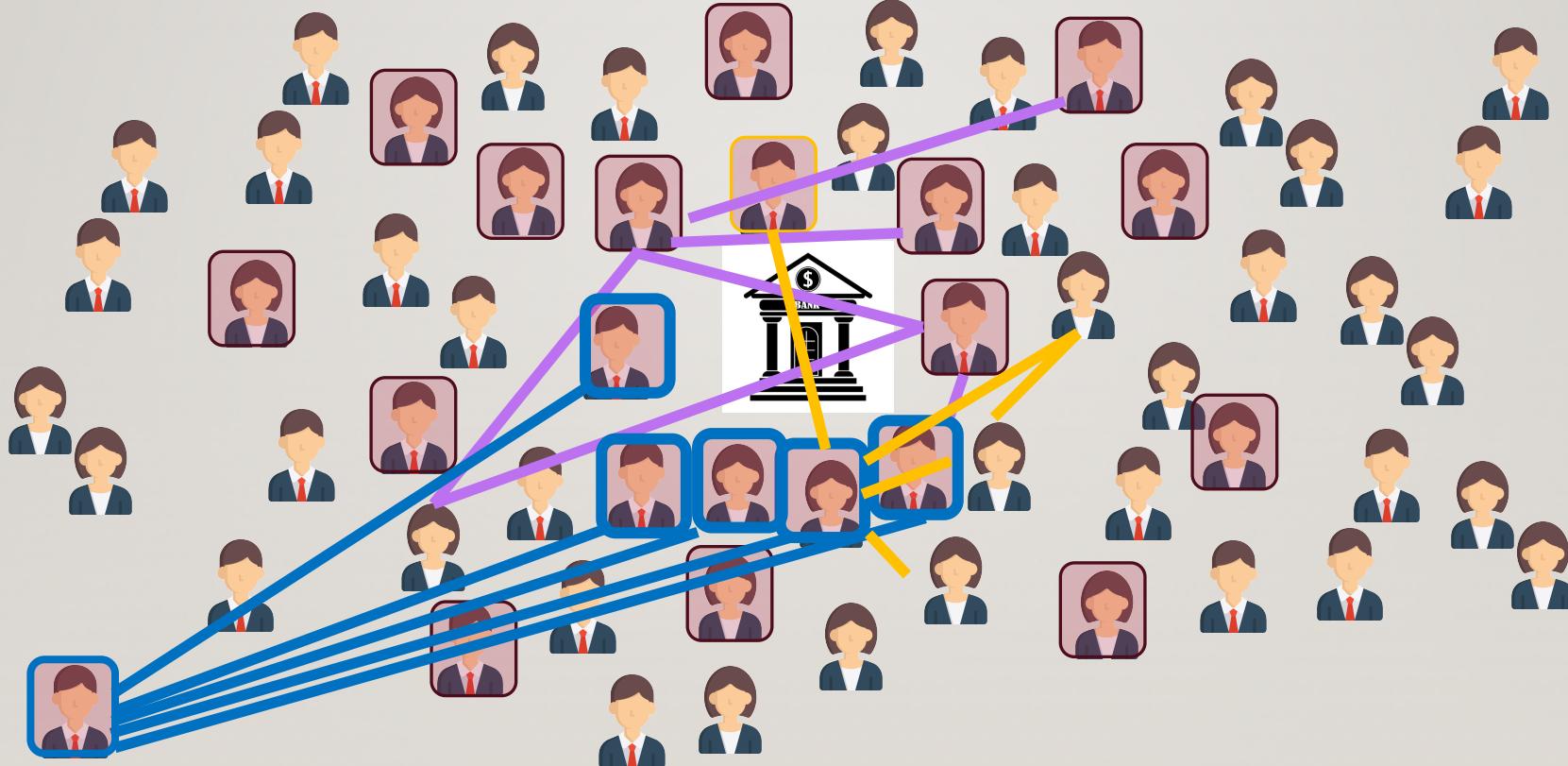
- Network build around ‘PTEN’
- Co –expression
- Co occurrence
- Database
- Text-mining
- Neighborhood
- Gene fusion
- Experiments



<http://String-db.org>

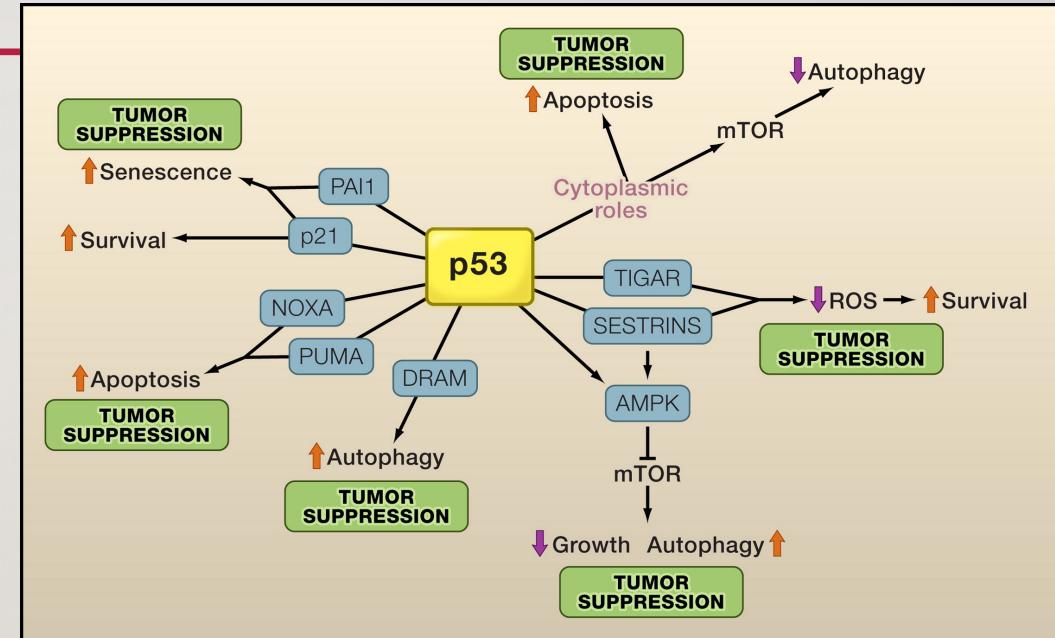
# MASTER REGULATORS

---



# MASTER REGULATORS

- These are genes that are **not necessary differentially expressed** but **regulate a large number** of differentially expressed genes.
- **Ex1: Cytokines:** works in networks, and in cascade (producing other cytokines). Also work in an additive manner. IL1, IL2, etc...
- **Ex2: Transcription factors, such as P53 (Tumor Suppressors)**: The control of cell survival, proliferation, and death by p53 is mediated by the regulation of expression of p53 target genes (some examples shown in blue). **Most of these p53 responses have the potential to contribute to tumor suppression.**



Youssden & Prives Cell 2009

# PROTEIN-PROTEIN INTERACTION NETWORKS (PPI)

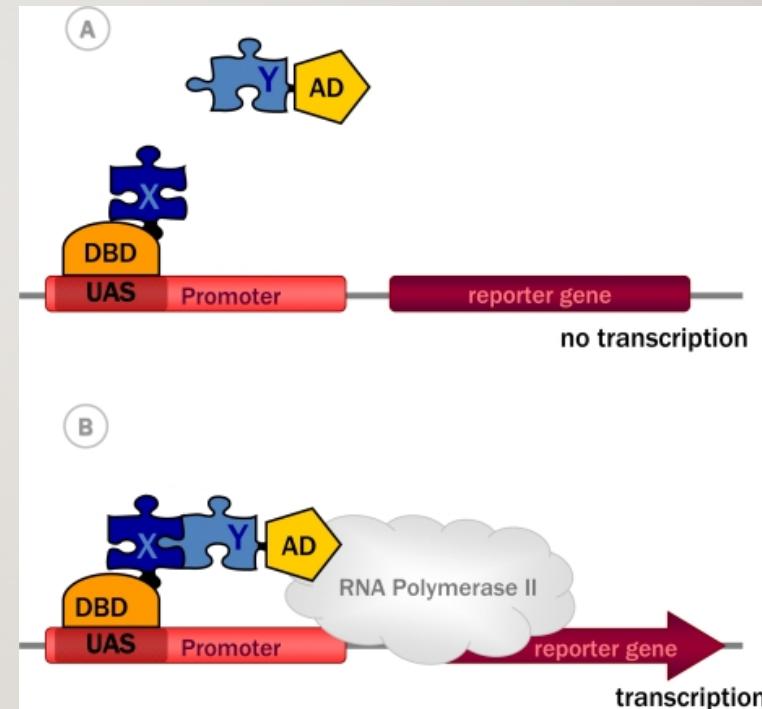
---

- **Description:** A protein-protein interaction (PPI) network represents the **physical interactions** between proteins in a cell. These interactions are **crucial** for most cellular functions, such as signaling, transport, and enzymatic reactions.
- **They allow**
  - Understanding which protein are present in a specific biological pathways.
  - Understanding how a mutation in a protein affect other proteins
  - Detecting key regulatory proteins in a networks

# PROTEIN-PROTEIN INTERACTION NETWORKS (PPI)

---

- **The Y2H (Yeast 2-Hybrid) technique allows detection of interacting proteins in living yeast cells.** Interaction between two proteins, called bait and prey, activates reporter genes that enable growth on specific media or a color reaction. **Y2H can be easily automated** for high-throughput studies of protein interactions on a genome-wide scale.
- **Limitation:** False positive and false negatives



Brückner et al Int J Mol Sci 2009: Yeast Two-Hybrid, a Powerful Tool for Systems Biology

# GENE REGULATORY NETWORKS

- **Description:** This type of network shows how genes regulate each other's expression through interactions between transcription factors and regulatory DNA sequences.
- Example:
  - **Cell cycle regulatory network:** Proteins like Rb (retinoblastoma) and E2F regulate the expression of genes involved in cell cycle progression; disruption of this network can lead to cancer.
  - **Hox gene network:** This network controls embryonic development and body segmentation, and mutations can result in congenital malformations.

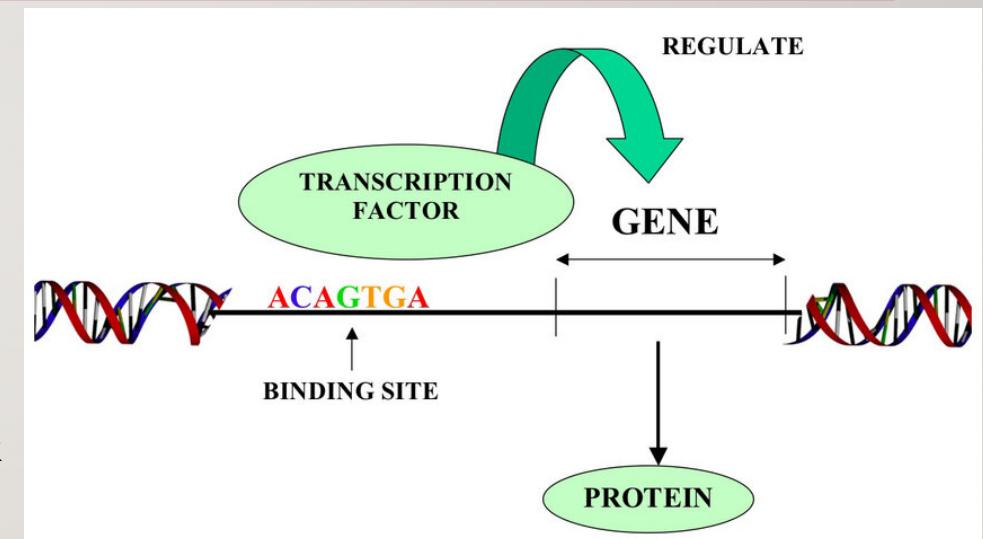


Figure by Saurabh Sinha.

# METABOLIC NETWORKS

---

- Description: This network represents all the metabolic pathways in an organism. The nodes represent metabolites (small molecules like glucose or pyruvate), and the edges are the reactions catalyzed by enzymes.
- Example: Glycolysis pathway: This key metabolic pathway converts glucose into pyruvate, producing energy in the form of ATP. Dysregulation of enzymes involved in glycolysis is often observed in cancer cells.
- Krebs cycle: This cycle is essential for cellular energy production. Dysregulation can lead to metabolic disorders such as lactic acidosis.

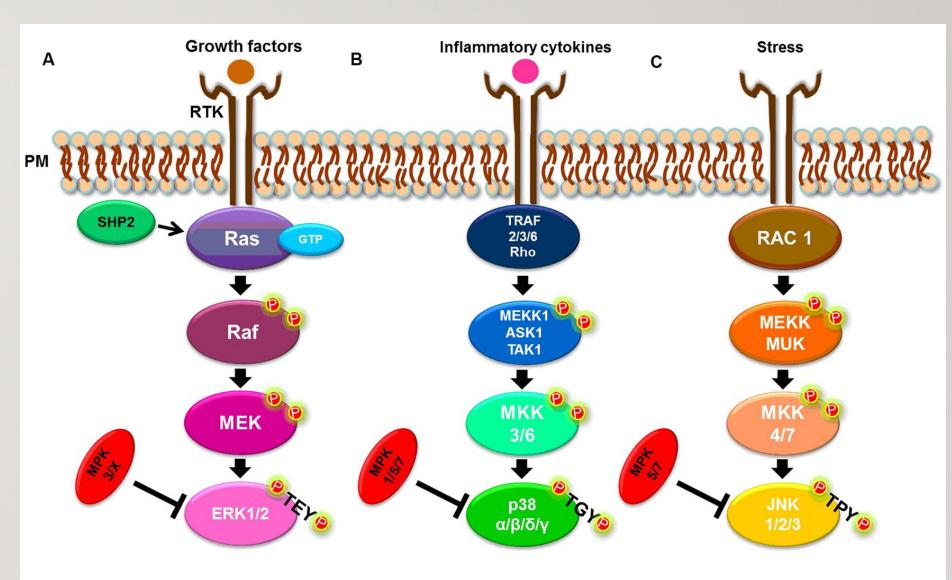
# CELL SIGNALING NETWORKS

---

- Description: These networks describe how cells communicate through receptors and ligands to regulate cellular responses to external stimuli (growth, stress, etc.).
- Example:
  - MAP kinase pathway: This signaling network is involved in regulating proliferation, differentiation, and apoptosis. In cancer, this pathway is often uncontrollably activated.
  - Insulin signaling: Insulin signaling regulates glucose metabolism. Disruptions in this network lead to type 2 diabetes.

# MAP KINASE PATHWAY

- There are three well-known MAPK pathways in mammalian cells:
  - The ERK1/2,
  - The c-JUN N-terminal kinase 1, 2 and 3 (JNK1/2/3)
  - The p38 MAPK  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  pathways.
- ERK, JNK, and p38 isoforms are grouped according to their activation motif, structure and function (Owens and Keyse, 2007; Raman et al., 2007; Zhang and Dong, 2007). ERK1/2 is activated in response to growth factors, hormones and proinflammatory stimuli, while JNK1/2/3 and p38 MAPK  $\alpha$ ,  $\beta$ ,  $\delta$ , and  $\gamma$  are activated by cellular and environmental stresses, in addition to proinflammatory stimuli (Owens and Keyse, 2007; Kyriakis and Avruch, 2012; Figure 1).



Soares Silva et al. 2016: The Mitogen-Activated Protein Kinase (MAPK) Pathway: Role in Immune Evasion by Trypanosomatids

# GENE CO-EXPRESSION NETWORKS

---

- Description: These networks show genes that are co-expressed under similar conditions, which can reveal new functional modules that are co-regulated.
- Example:
  - **Co-expression network in cancer:** Analysis of gene co-expression in cancer cells can reveal molecular signatures specific to different cancer types and help identify new therapeutic markers.
  - **Co-expression network in neurodegenerative diseases:** Studying co-expressed genes in diseases like Parkinson's helps to understand the underlying mechanisms of neurodegeneration.

# PRACTICAL CASE: DISCOVER KEY CANCER PLAYERS WITH NETWORKS

---

- Case I: **Discover networks** that discriminate between **ER-negative** and **ER-positive** patients in **breast cancer** from transcriptomics data.
- Complete pipeline of network inference:
  - Extract **differentially expressed** genes from TCGA breast cancer (Breast Invasive Carcinoma) data with **Phantasus**: <https://artyomovlab.wustl.edu/phantasus/>
  - Extract **network** from differentially-expressed genes with **String-db**: <https://string-db.org/>
  - Extract **enriched pathways** with **String-db**

# PRACTICAL CASE: R2+STRING

---

- Case 2: We want to analyse a network of genes correlated to a differentially-expressed gene in breast cancer (ER+/ER): the A2ML1 gene: ***alpha-2-macroglobulin like 1***
- Pipeline of discovery of correlated gene modules with A2ML1:
  - Search for genes **correlated with A2ML1** using **R2 Genomics and Visualisation** platform: [https://hgserver1.amc.nl/cgi-bin/r2/main.cgi?open\\_page=login](https://hgserver1.amc.nl/cgi-bin/r2/main.cgi?open_page=login)
  - **Extract network** from differentially-expressed genes with **String-db**: <https://string-db.org/>
  - Search for **enrichment of number of edges** compared to random chance

---

---

# SOME MATHEMATICAL BACKGROUND



**Biology: Scale-free networks (power-law distributed)**



**Network centralisation measures: detection of highly connected nodes**

- Betweenness centrality
- Closeness centrality
- Degree
- Eigennetworks

# PAUSE FOR A MOMENT !

---

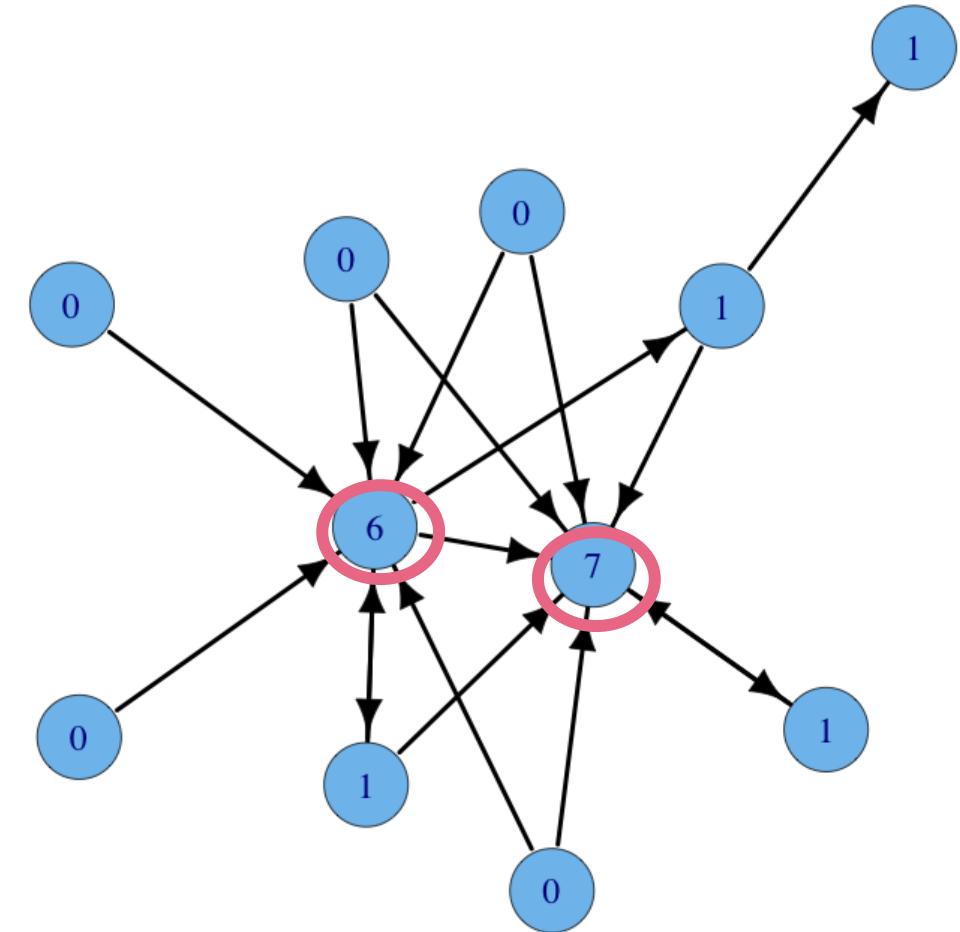
FOR A LIST OF NUMBERS, WHAT IS ?

- Mean or Average ?
- Median ?
- Mode ?
- **Mean or average** = sum/number of element in the list
- **Median** = sort list, take the middle element
- **Mode** = element with the maximum frequency

# WHAT IS CENTRALITY ?

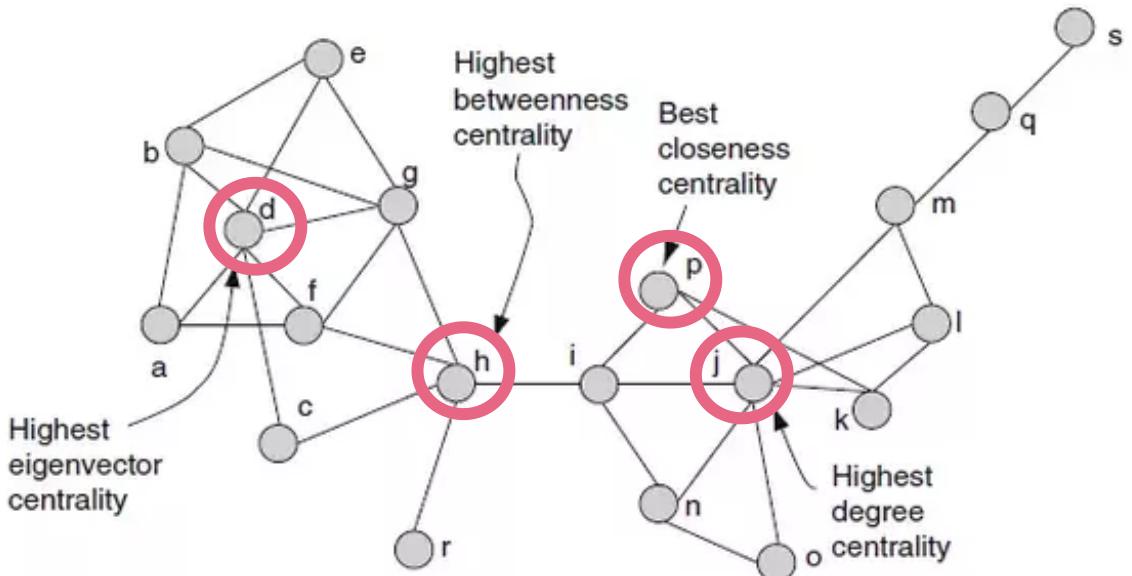
---

- Which nodes are important ?



# AGAIN, WHICH NODES ARE IMPORTANT ?

---



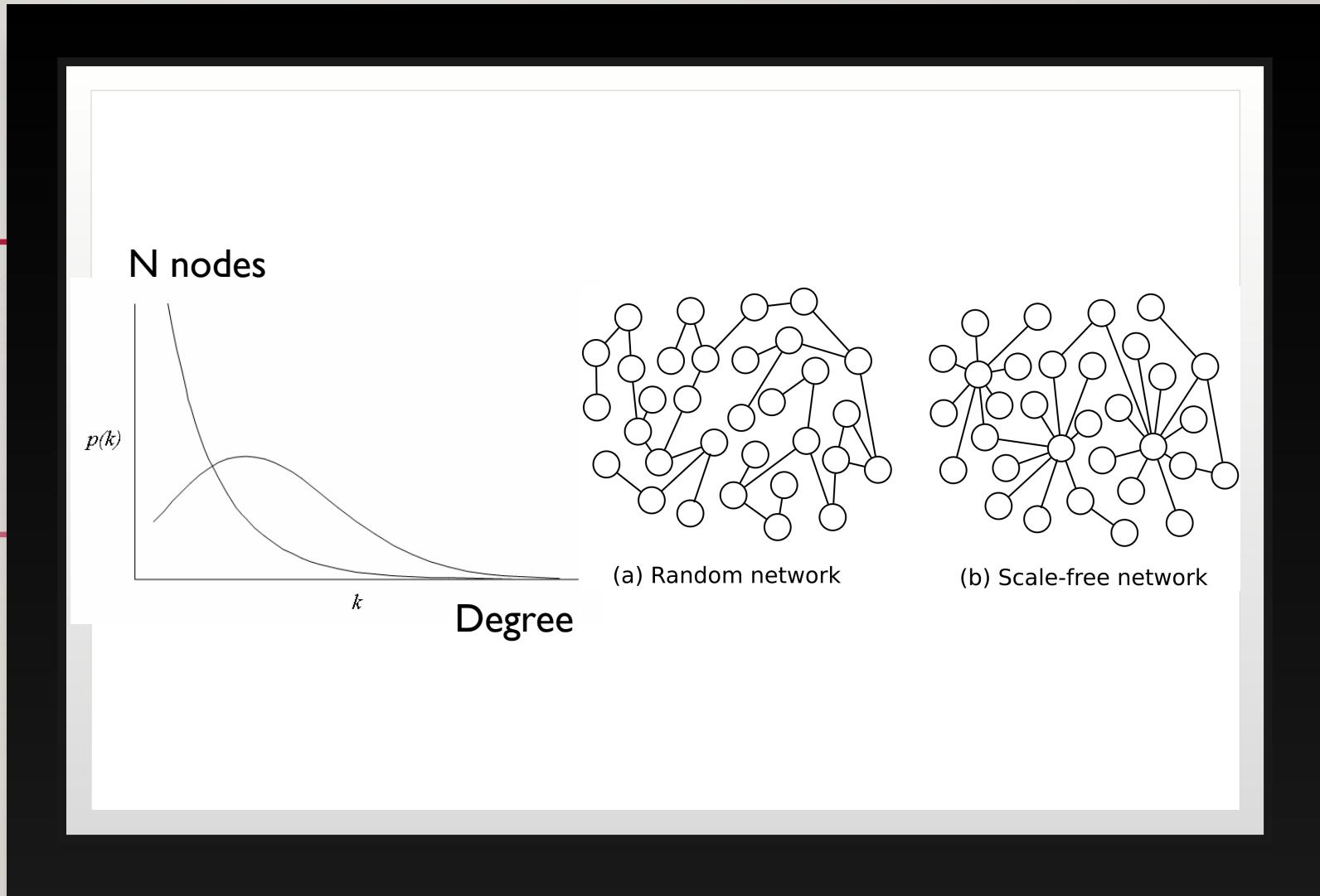
- Highest degree: **highest frequency** of connection (mode)
- Best closeness: **lower average distance** to any other node (average)
- Highest betweenness frequency: **highest frequency** of falling on the **shortest path** between nodes (median)
- Eigenvector centrality: how likely you are to be **connected to highly connected** nodes

John McCulloch. <https://www.youtube.com/watch?v=iiVeQkIELyc>

# SCALE FREE NETWORKS

**EXPONENTIAL RELATION**  
BETWEEN CONNECTIVITY AND  
FREQUENCY OF OCCURRENCE

**HUBS!**

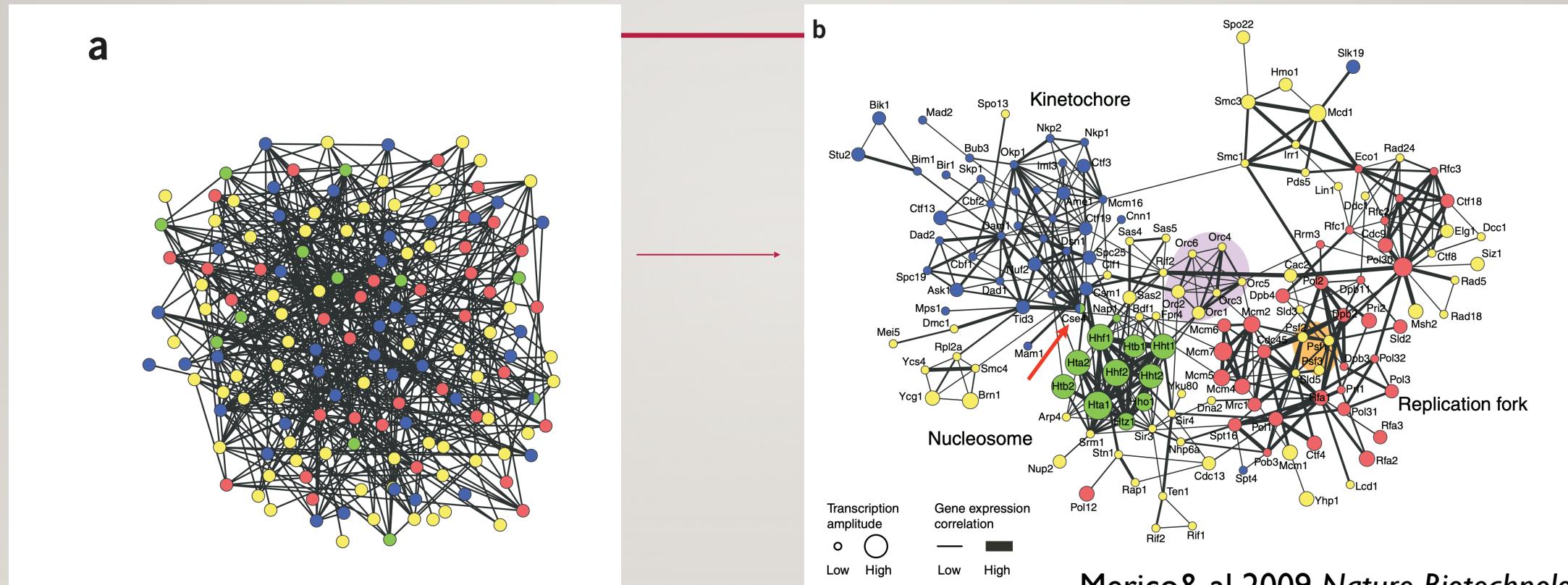


# NETWORK TOPOLOGY AND PROPERTIES

---

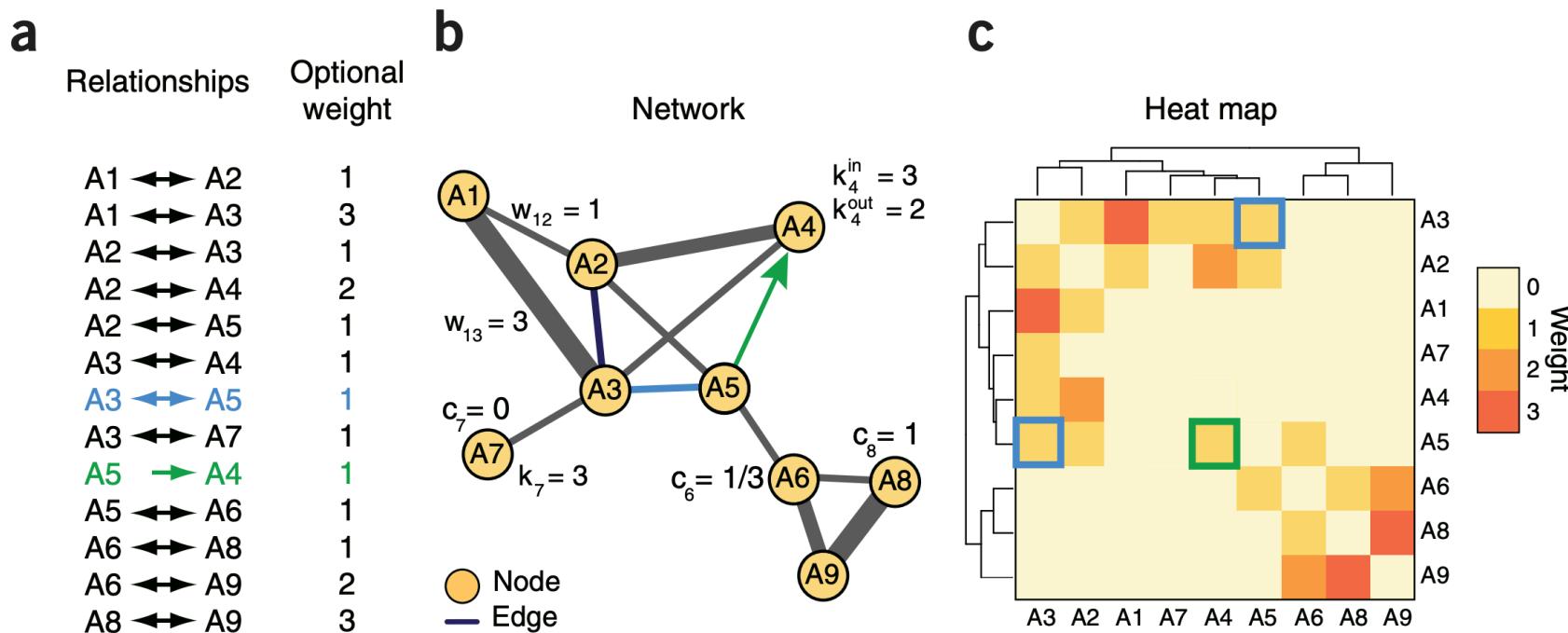
- **Key topological features:**
  - **Degree distribution:** Describes how many connections each node has.
  - **Clustering coefficient:** Measures how interconnected a node's neighbors are.
  - **Path length:** Average distance between nodes in the network.
- **Centrality measures:**
  - **Degree centrality:** Number of connections to a node (hubs).
  - **Betweenness centrality:** Importance of a node in connecting other nodes.
  - **Closeness centrality:** Nodes with shorter average distances to other nodes
  - **EigenNodes:** Nodes connected to highly connected nodes
- **Applications:**
  - **Hubs in PPI networks:** Identifying central proteins that are often targets in diseases like cancer.
  - **Network motifs:** Recurring patterns of interconnections that help to reveal the network's function.
- **Tools:**
  - **Cytoscape:** A key tool to visualize and analyze network topology.

# NETWORK INTERPRETATION: LAYOUT



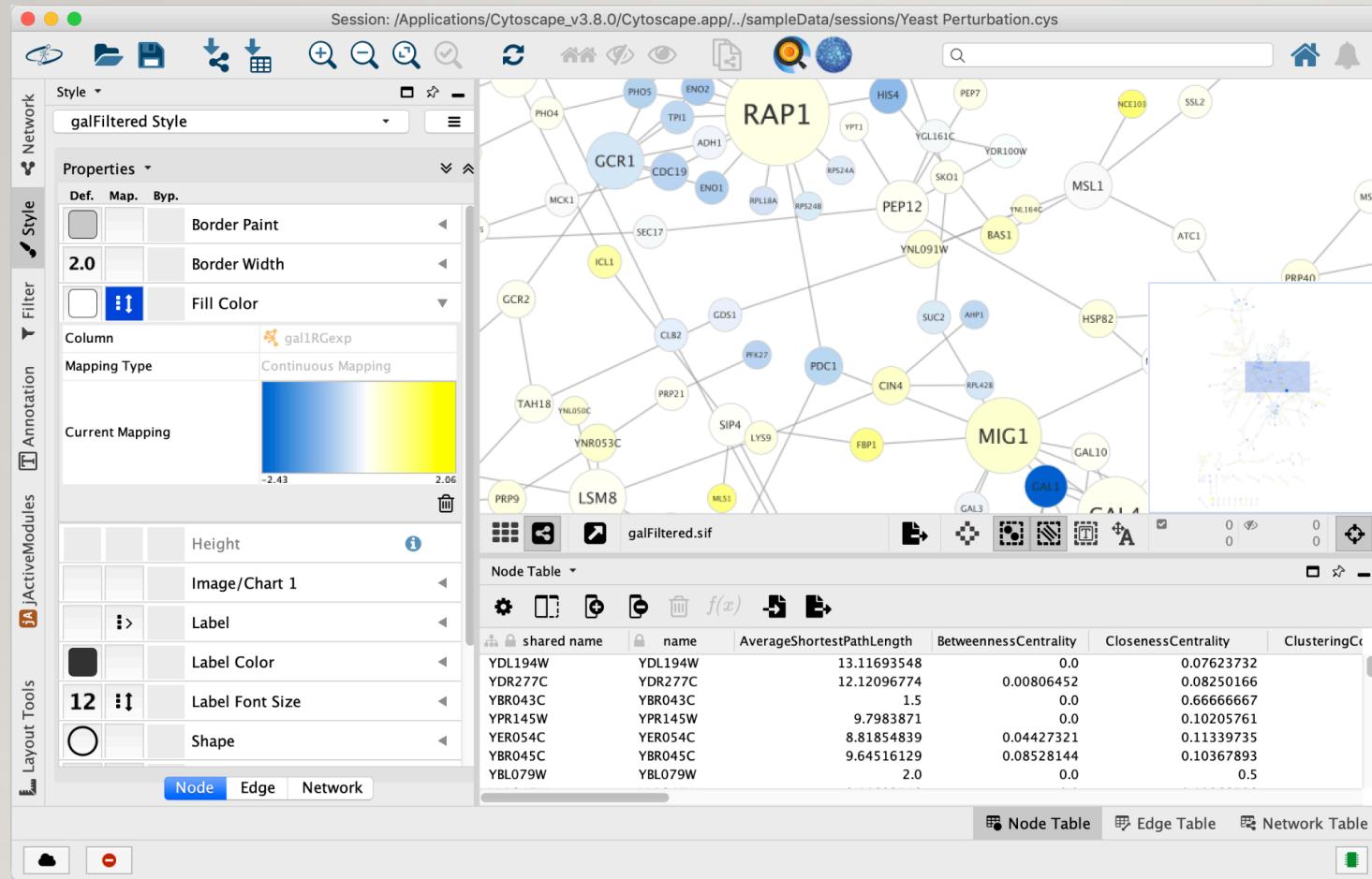
Network visualization of chromosome maintenance and duplication machinery in baker's yeast, *Saccharomyces cerevisiae*.

# DUALITY NETWORK-TABLE



Merico & al 2009 *Nature Biotechnology*

# OTHER TOOLS: CYTOSCAPE



<https://cytoscape.org>

# SOME NETWORK ANALYSIS TOOLS

---

- 1. STRING-DB:** For retrieving known and predicted protein-protein interactions.
- 2. Cytoscape:** Visualization and analysis of molecular interaction networks.
  1. Plugins: MCODE (module detection), ClueGO (functional analysis).
- 3. BioNetGen:** Rule-based modeling for simulating biochemical networks.
- 4. Ingenuity Pathways (Commercial):** Gene function prediction and gene interaction networks.
- 5. VisNetwork:** network visualisation tool

Bioinformatics: Bioinformatics talks by Michael Edwards (Youtube)

# CONCLUSION AND PERSPECTIVES

- Biological networks are essential to understanding complex diseases.
- Rich and more intuitive. Great for visualisation and pattern recognition.
- We explored the use of some tools/platforms that are useful to do bioinformatics **without code** and **without a bioinformatician**
  - R2 genomics
  - Phantasia
  - String-db
- Thanks for your attention!
- © Ghislain BIDAUT 2024 – CC-BY-NC CC-by-nc-sa (Attribution / Pas d'Utilisation Commerciale / Partage dans les mêmes conditions)

