Chemoinformatics and network science
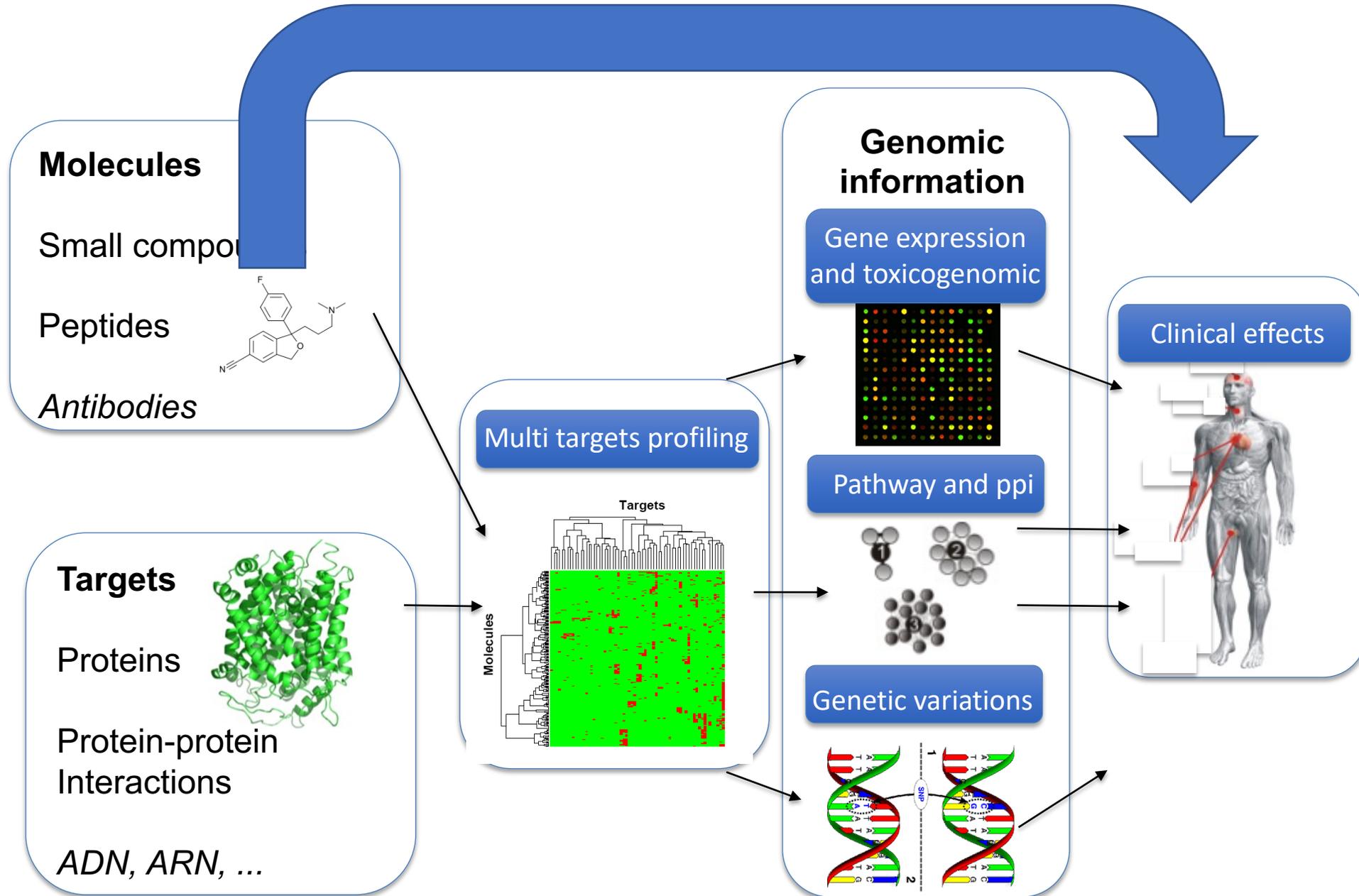
Pr. Olivier TABOUREAU
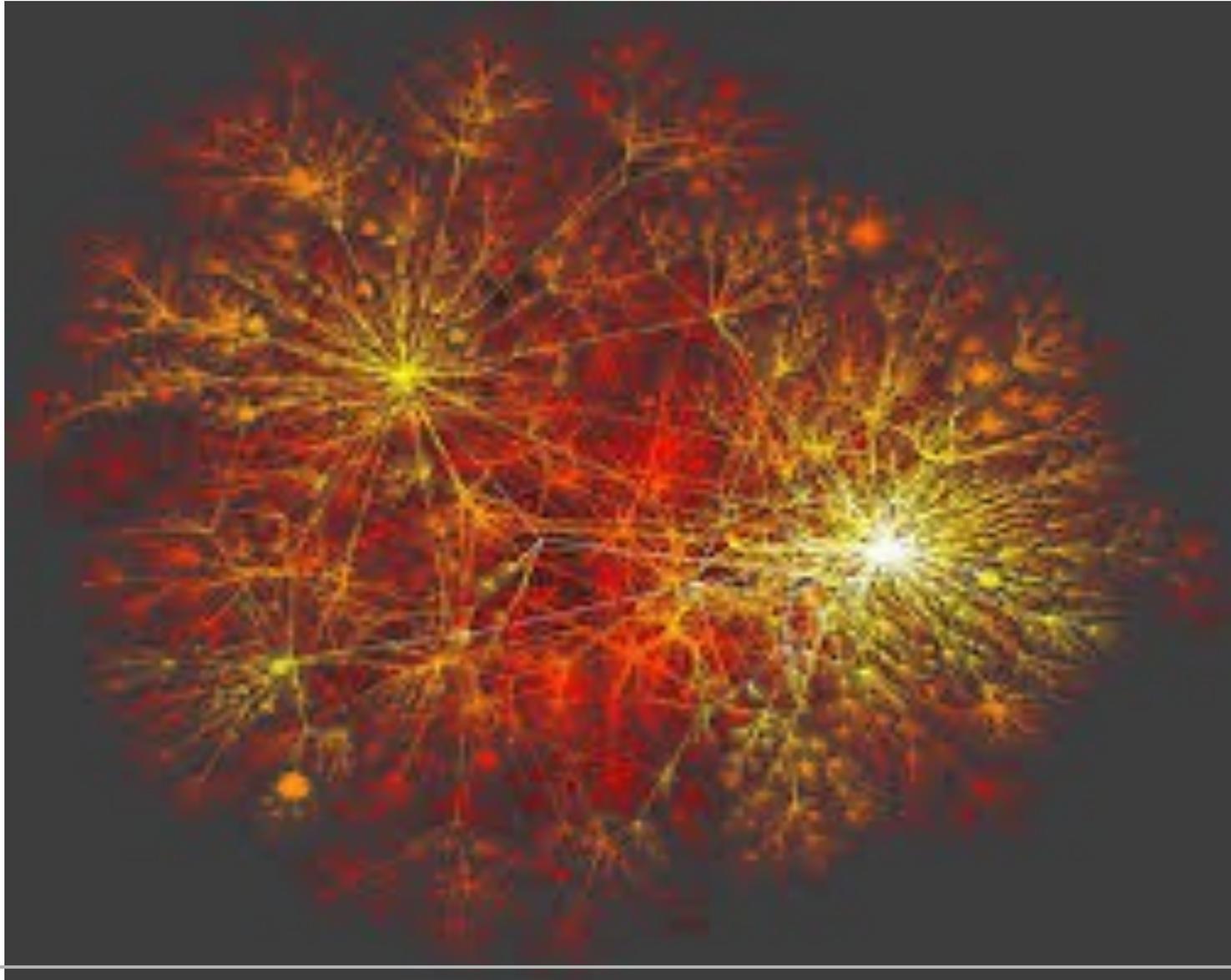
olivier.taboureau@univ-paris.fr

Objective: Understand the relationship between molecules (environmental chemicals, drugs, natural products, peptides) and proteins (or genes) from the molecular to more complex levels and in relation to diseases.


Molecules

Small compounds

Peptides

Antibodies

Targets

Proteins

Protein-protein Interactions

ADN, ARN, ...

Multi targets profiling

Genomic information

Gene expression and toxicogenomic

Pathway and ppi

Genetic variations

Clinical effects

Computational modeling of protein-ligand interactions (Inserm U1133)

Molecules

Small compounds

Peptides

Antibodies

Targets

Proteins

Protein-protein Interactions

ADN, ARN, ...

Multi targets profiling

Genomic information

Gene expression and toxicogenomic

Pathway and ppi

Genetic variations

Clinical effects

Network Science

Network Science

Network science is an academic field which studies complex networks such as telecommunication networks, computer networks, biological networks, cognitive and semantic networks, and social networks, **considering distinct elements or actors represented by *nodes* (or *vertices*) and the connections between the elements or actors as *links* (or *edges*).** The field draws on theories and methods including graph theory from mathematics, statistical mechanics from physics, data mining and information visualization from computer science, inferential modeling from statistics, and social structure from sociology.

Network = nodes + edges

- **Nodes** (vertices) are the objects in the network
- **Edges** are the links/interaction in the network

Network Science
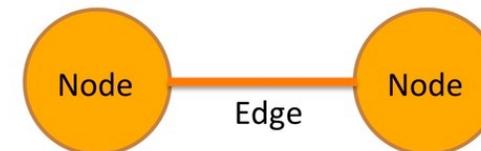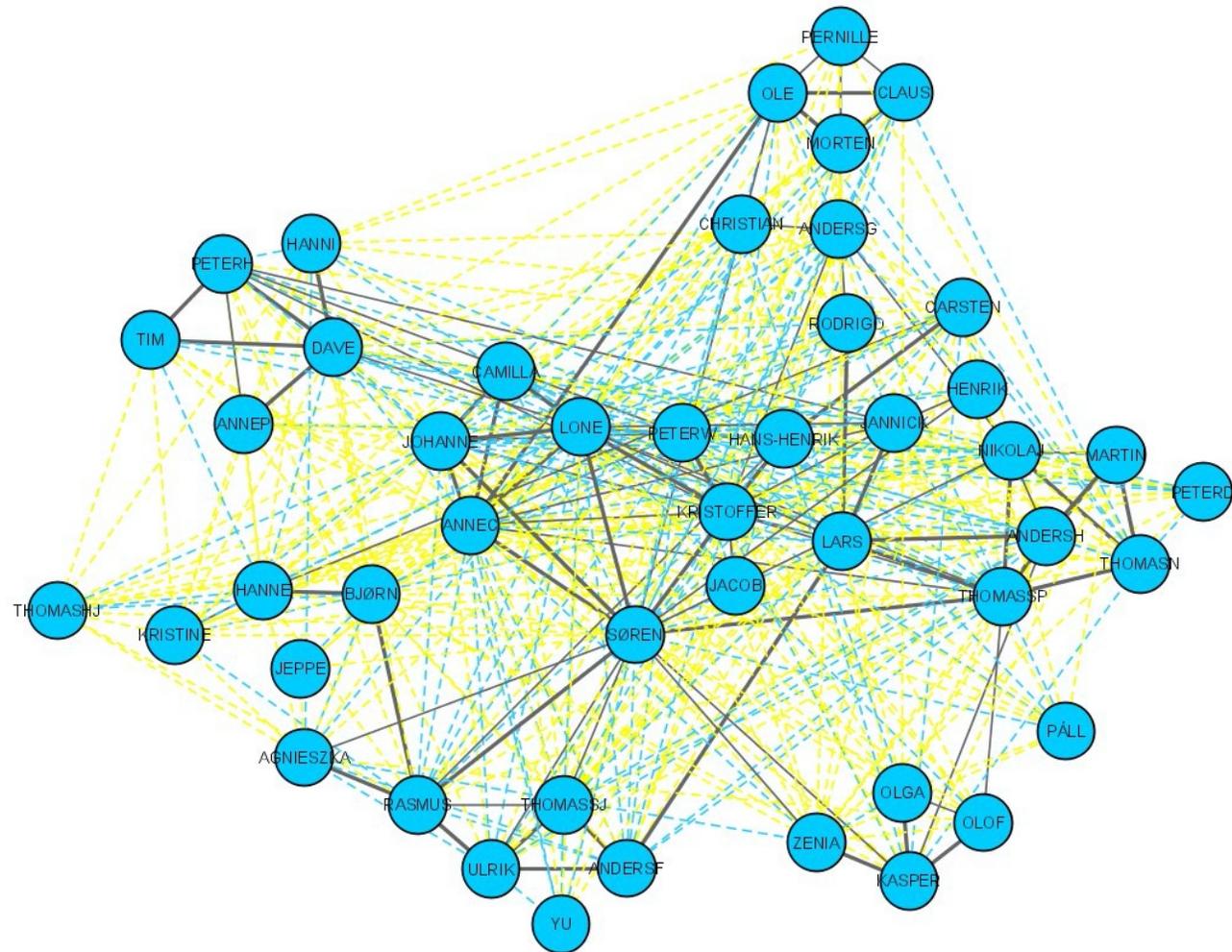
Social Network

6 degrees of separation
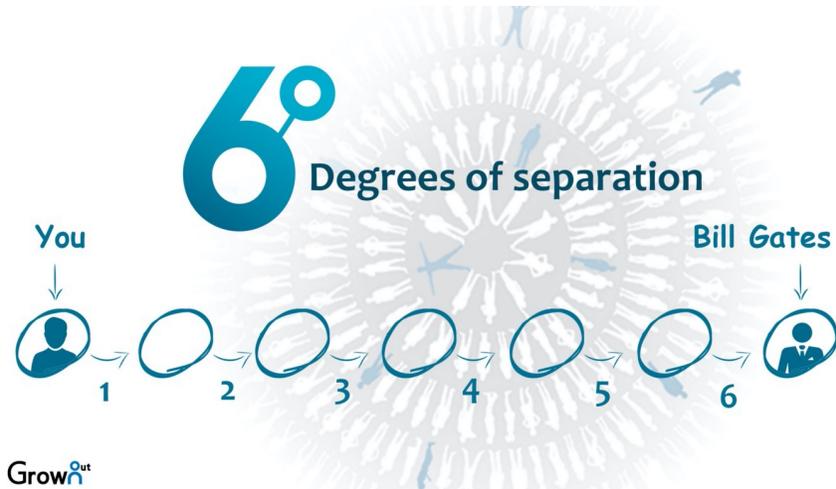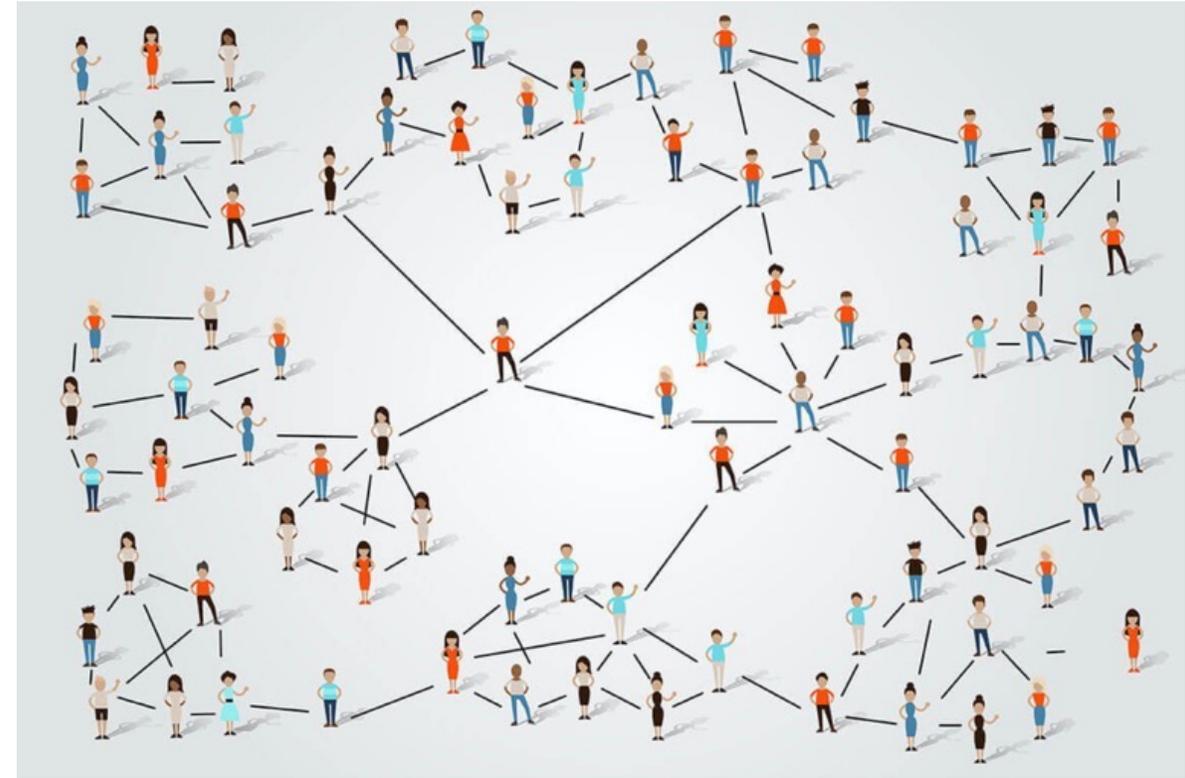
All people are six or fewer social connections away from each other

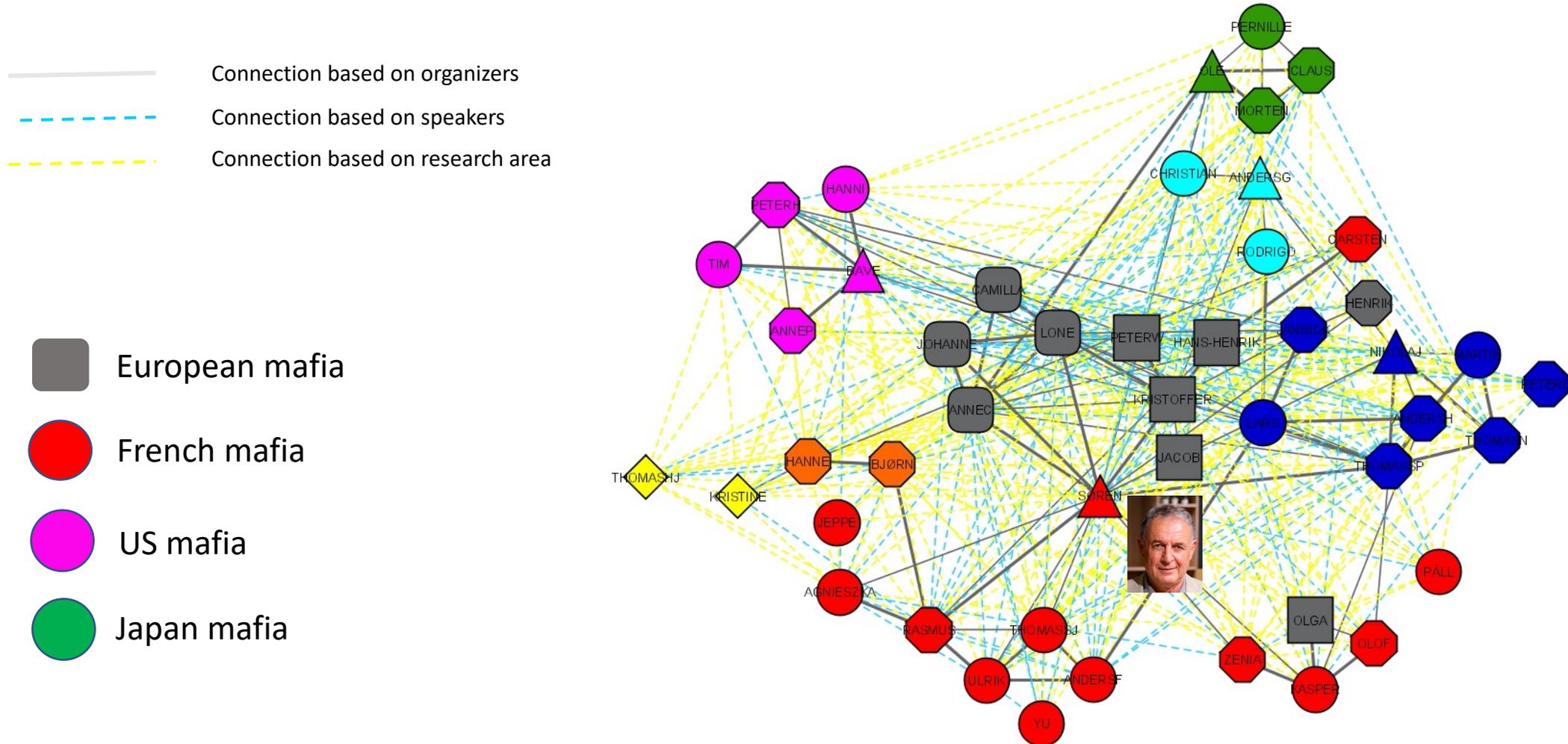## The chemoinformatics summer school network

Connection based on organizers

Connection based on speakers

Connection based on research area

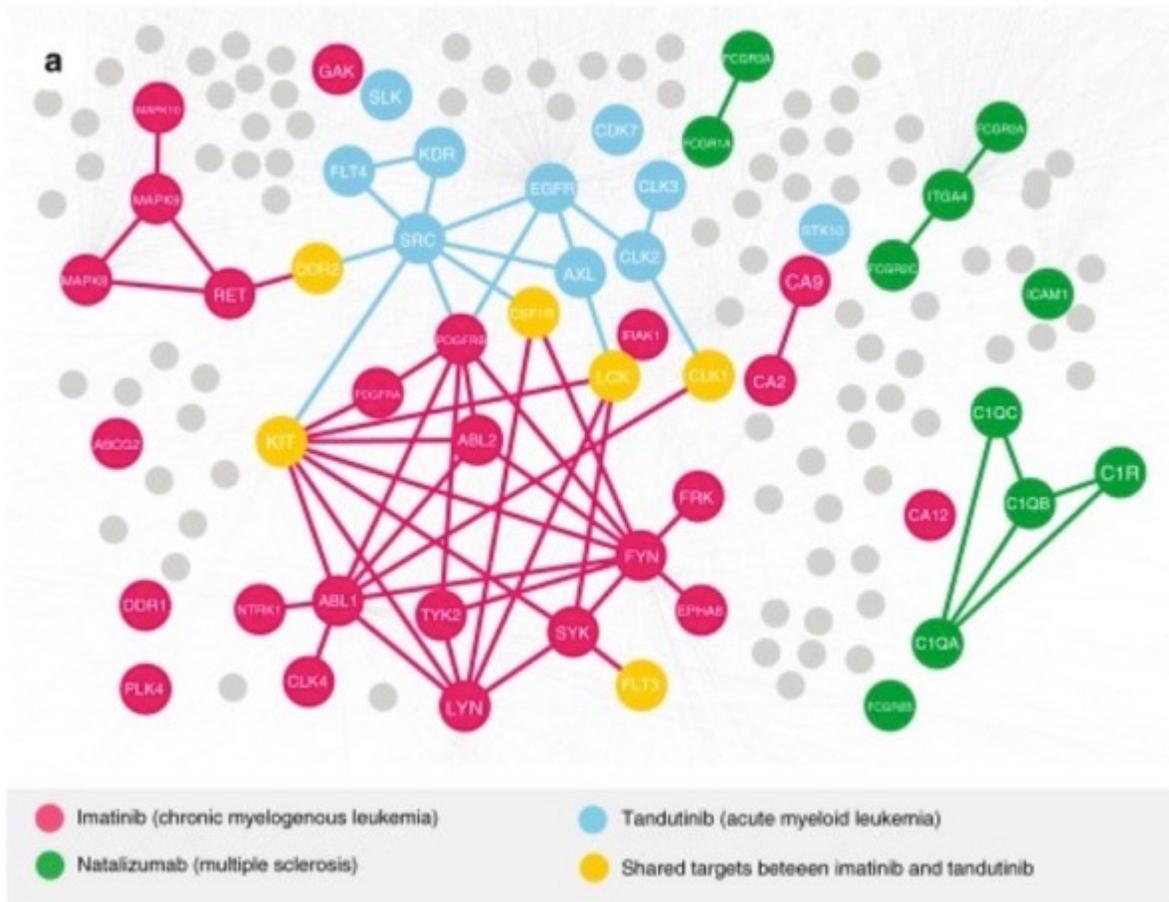- Can you cluster them?
- Who is the central connection?

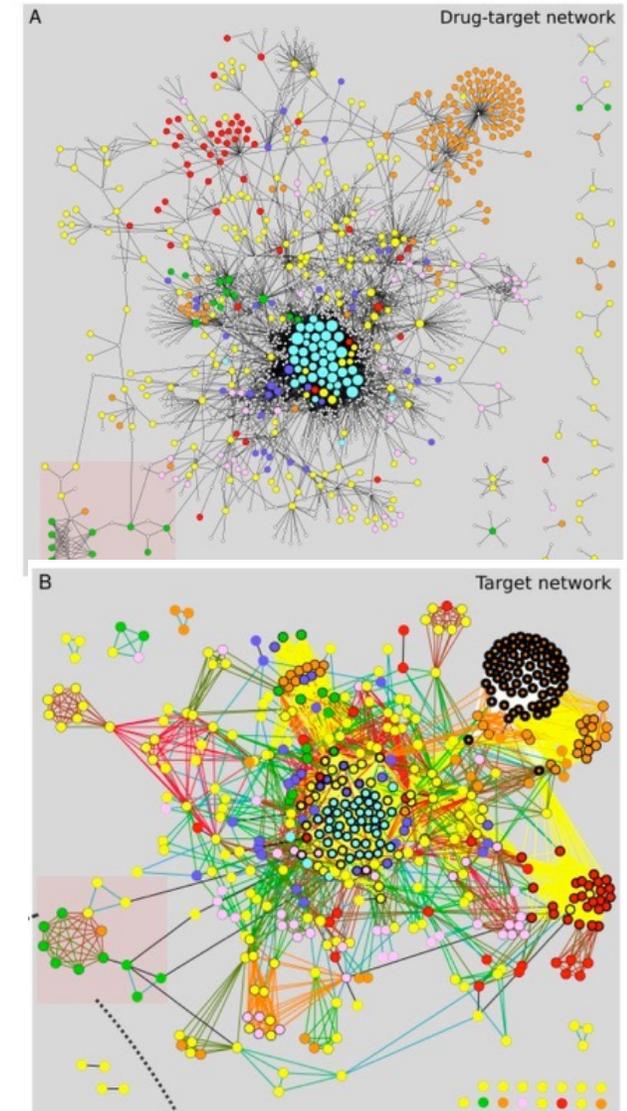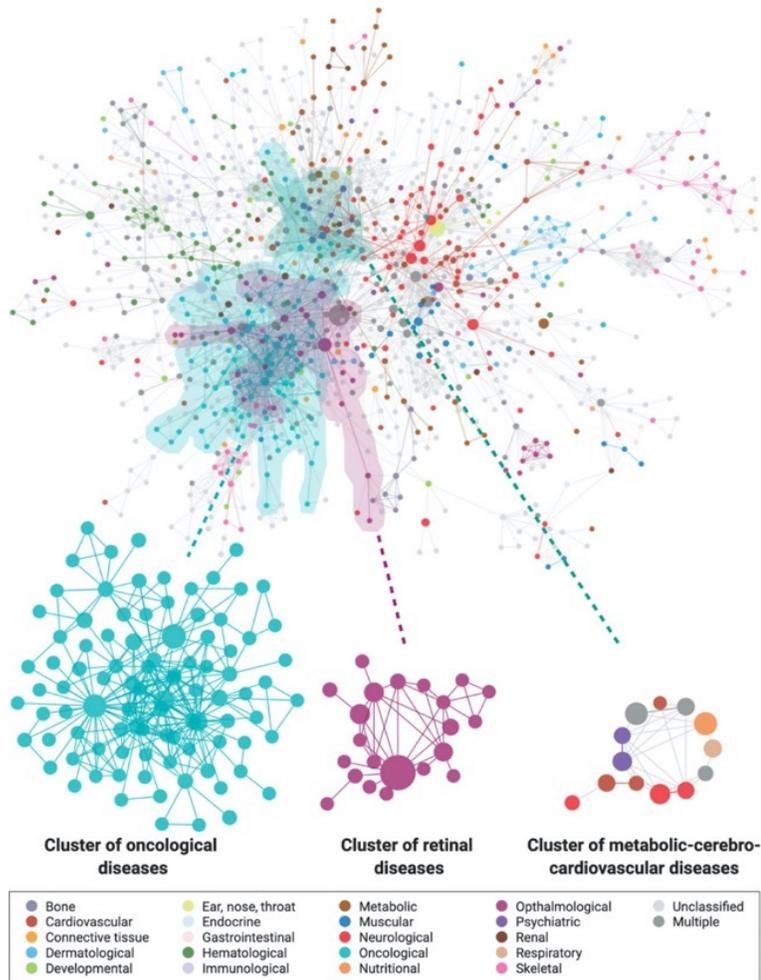## The chemoinformatics summer school Network

Connection based on organizers

Connection based on speakers

Connection based on research area

European mafia

French mafia

US mafia

Japan mafia

The chemoinformatics summer school Network

Connection based on organizers
Connection based on speakers
Connection based on research area

European mafia

French mafia

US mafia

Japan mafia

Network pharmacology

Cheng F. et al. Nat. Commun. 2019

Vogt I. et al. Mol.Inf 2019

Human disease network

Cluster of oncological diseases

Cluster of retinal diseases

Cluster of metabolic-cerebro-cardiovascular diseases

Legend:
- Bone
- Cardiovascular
- Connective tissue
- Dermatological
- Developmental
- Ear, nose, throat
- Endocrine
- Gastrointestinal
- Hematological
- Immunological
- Metabolic
- Muscular
- Neurological
- Oncological
- Nutritional
- Opthalmological
- Psychiatric
- Renal
- Respiratory
- Skeletal
- Unclassified
- Multiple

Nogales C. et al. Trends Pharmacol Sci. 2022

Covid-19 research

Viral-Human Protein-Protein Interaction

Human-Human Protein-Protein Interaction

Drug-Human Protein-Protein Interaction

Viral Interactome

Human Interactome

https://www.networkscienceinstitute.org/covid-19

Exposome research

Regulatory Network

Protein Interaction Network

Metabolic Network

B (-)-epigallocatechin 3- (EGCG) Found in green tea
- EGCG Protein Targets
- Type II Diabetes Disease Proteins

C Trichloroethylene (TCE)

Vermeulen R. et al. Sciences. 2020

Network Science : Protein-Protein association network -> inferred network

Protein-chemical associations
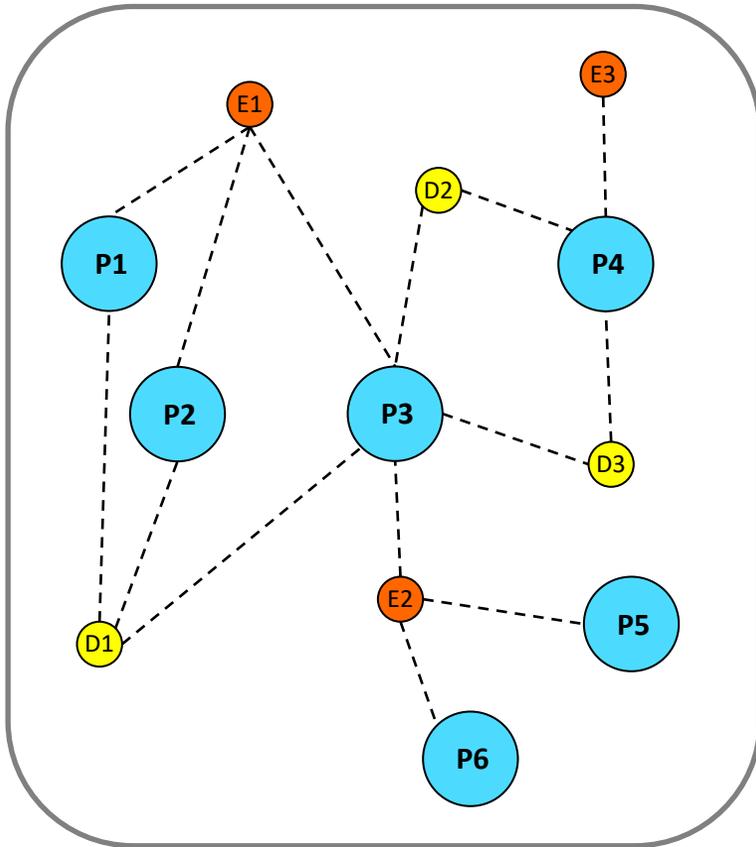
Protein-protein associations
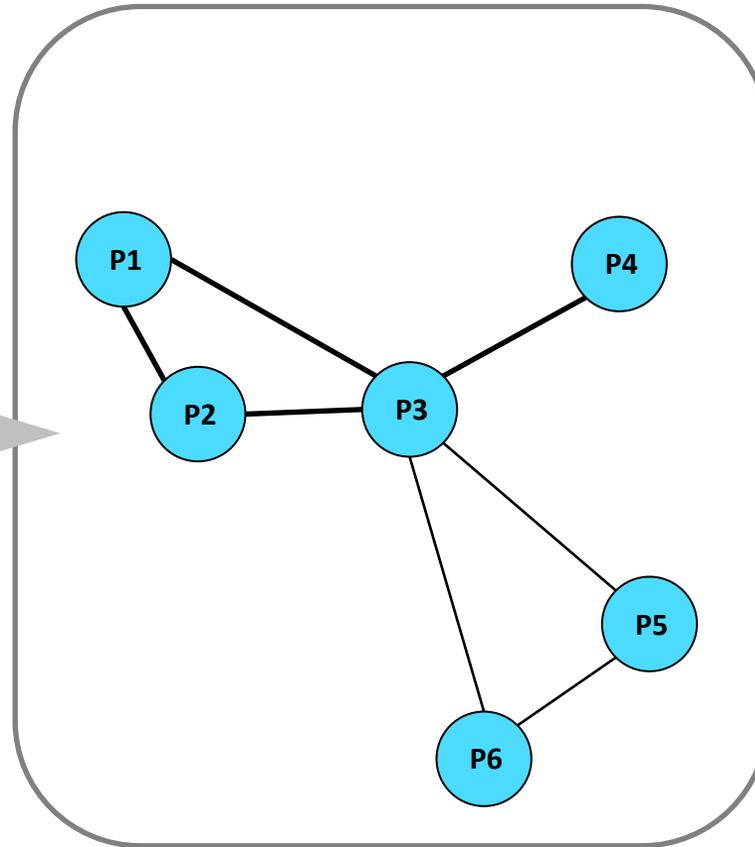
reliability

scores

Guilt by association theory

Guilty   Guilty

Guilty

Based on the guilt by association theory

Audouze et al. (2010) PLoS Comp. Biol.

Many type of biological interactions

- **Protein-drug interactions**

- **Protein-protein interactions**

- **Protein-DNA interactions**

- **Protein-disease interactions**

- **Co-expressions interactions**

- **Metabolic interactions**

…

Association drugs, targets and clinical outcomes

Drug is more likely to cause side effects in the organ/tissue where it is more likely to accumulate
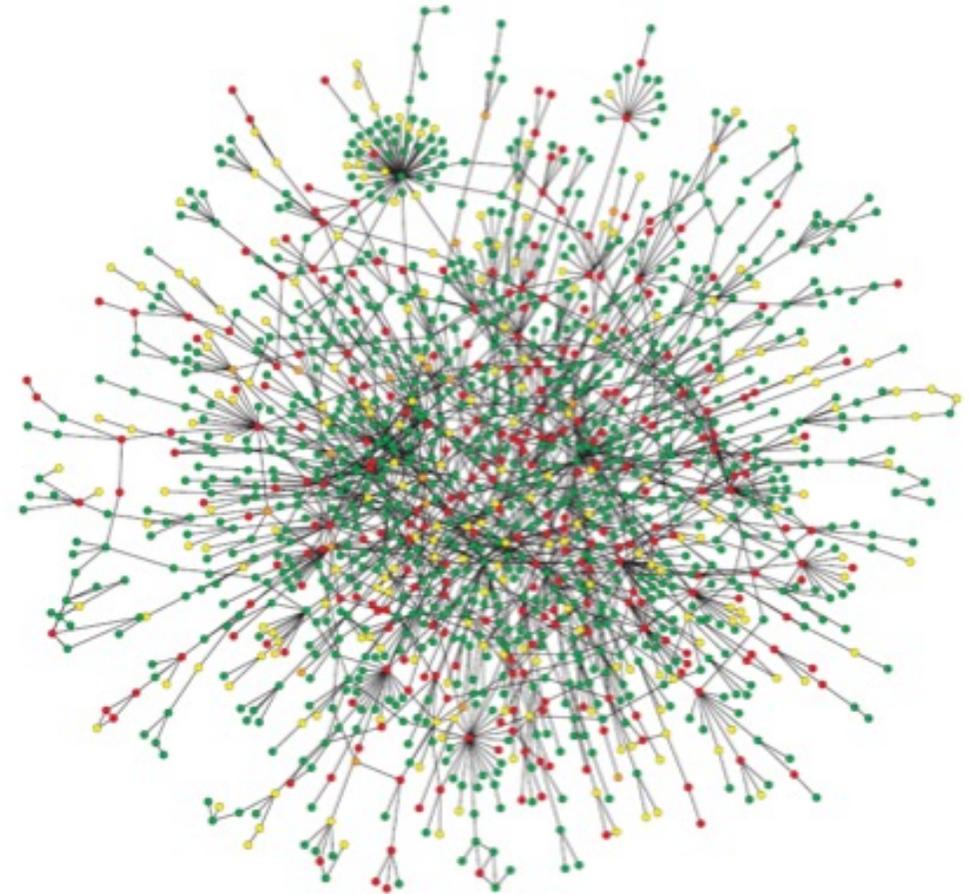
Network Science

More information can be found about network science theory

Albert-László Barabási
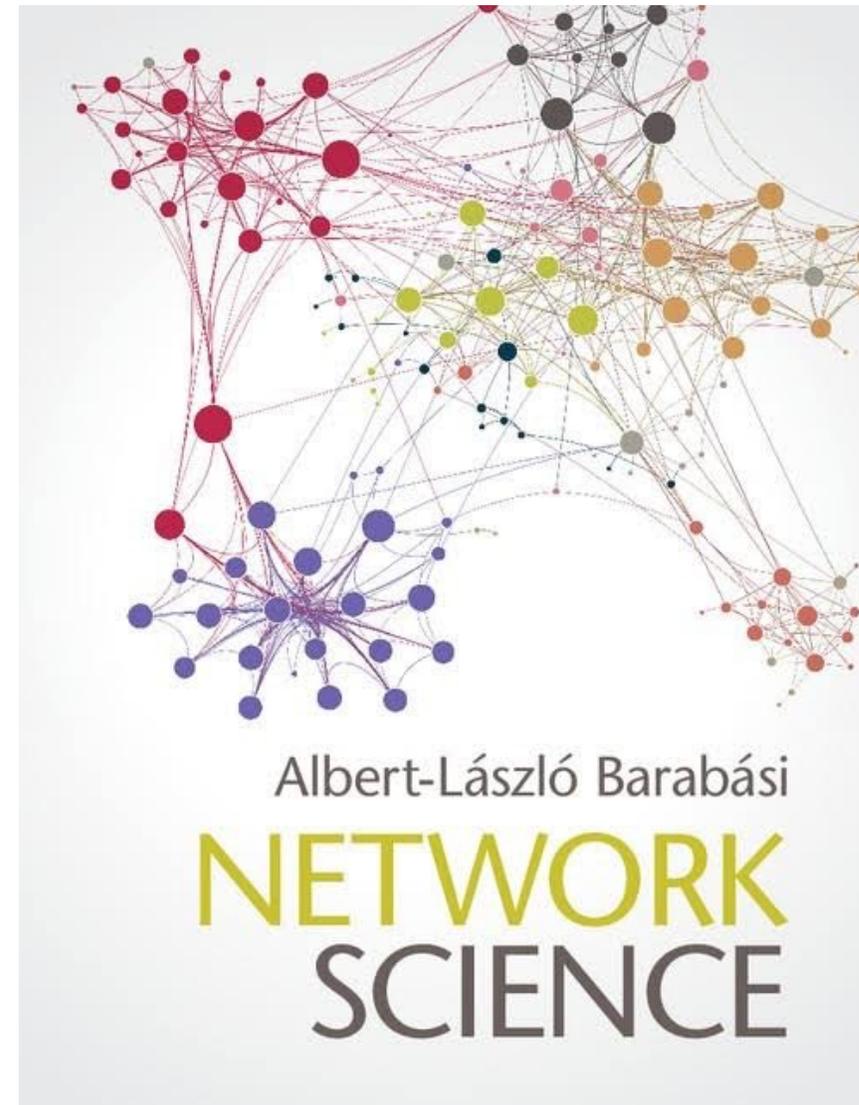
NETWORK
SCIENCE

Network Science in toxicology and adverse drug reactions

Environmental exposure in Europe

430,000 /year [1]

Adverse Drug Reactions in Europe

197,000 /year [2]

Toxicology profiling is failing

Disability-adjusted life years (environmental exposure) Worldwide

86m /year

[3]

Environmental exposure deaths Worldwide

4,9m /year

[3]

[1] https://www.eea.europa.eu/media/newsreleases/many-europeans-still-exposed-to-air-pollution-2015/premature-deaths-attributable-to-air-pollution
[2] Jacoline C. Bouvy, et al. 2015, Drug Saf.
[3] Annette Prüss-Ustün, et al. 2011, Environ. Health

Network Science : An example with adverse drug reactions

Drugs – Adverse Drug Reactions (ADR) - System Organ Classification (SOC) network

- 1000 Drugs and 6 164 ADRs classified in 27 SOC
  - Drugs have ADRs belonging to different SOC

- Most common ADRs are : 'Nausea', 'Vomiting'
  - resp. 41% and 40% of drugs respectively

- Specific ADRs : 1 626 ADRs are only related to one drug

- Drugs with the highest number of ADRs : methotrexate, alendronic acid and prednisone
  - resp 1 561, 1 422 and 1 377 ADRs

- 155 drugs are associated to only one ADR

- **Knowing drug-target and drug-ADR -> Target-ADR can be suggested**
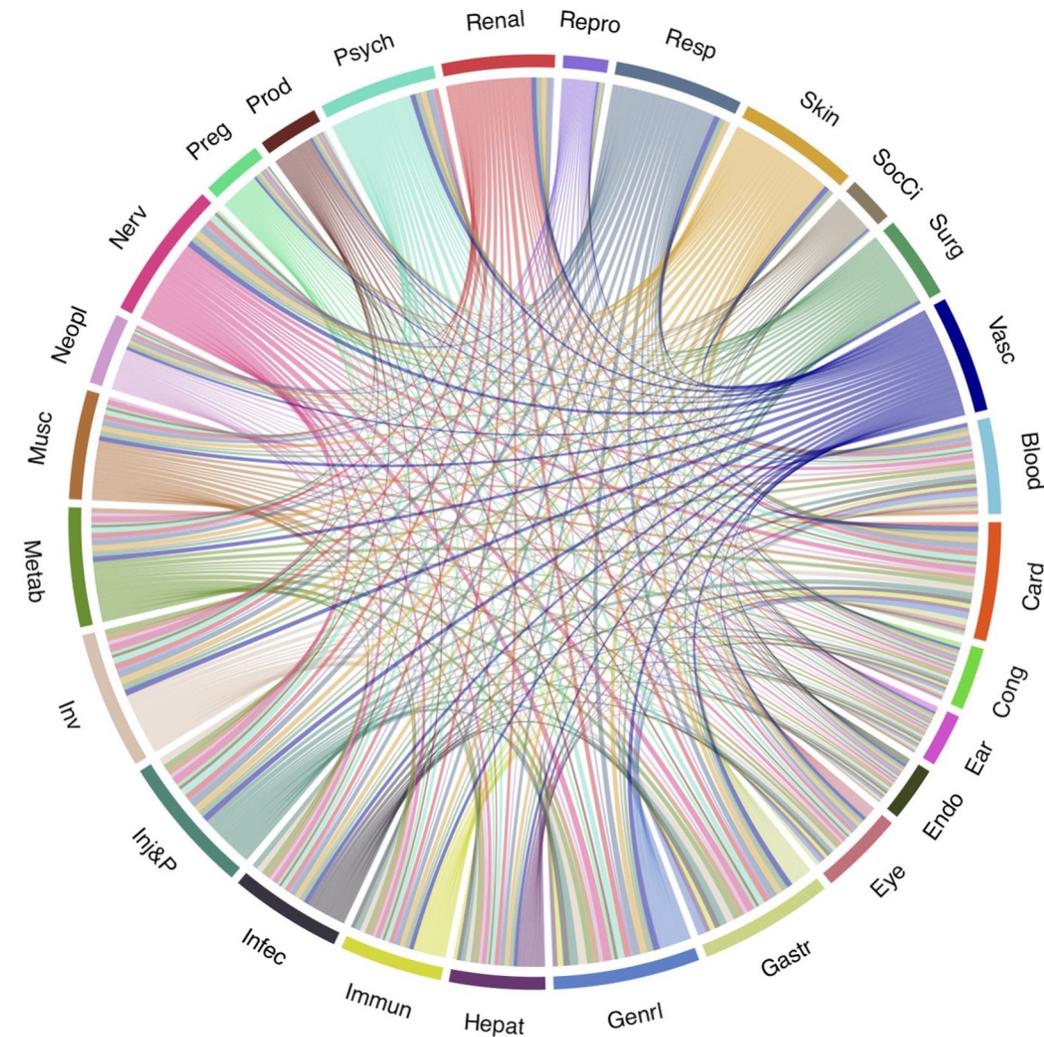
Network Science : An example with adverse drug reactions

Drugs – Adverse Drug Reactions (ADR) - System Organ Classification (SOC) network

Identification and prioritization of drug-target involved in specific ADRs, as well as in more general terms using SOC

Network of 515 959 interactions
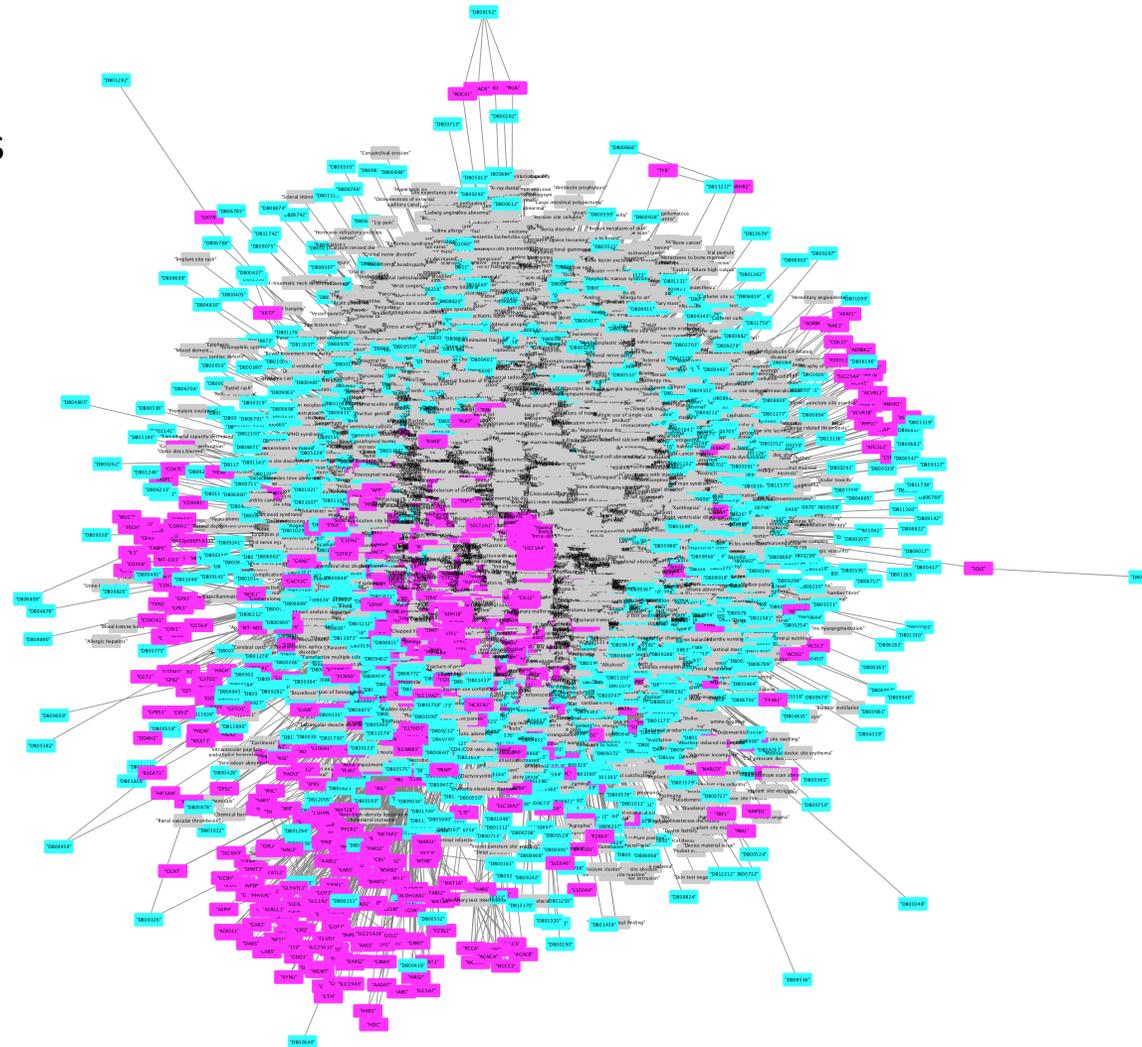
- Drug
- Protein
- ADR

Drugs – Adverse Drug Reactions (ADR) - System Organ Classification (SOC) network

- Targets associated with the most ADRs are cytochromes and transporters.

- Possible to assess the role of proteins in specific ADRs

  - HPD (hydroxyphenyl pyruvate dioxygenase) might be highly related to 'Hepatocellular carcinoma'

  - SQLE (Squalene exposidase) is connected to the ADR 'Blister'
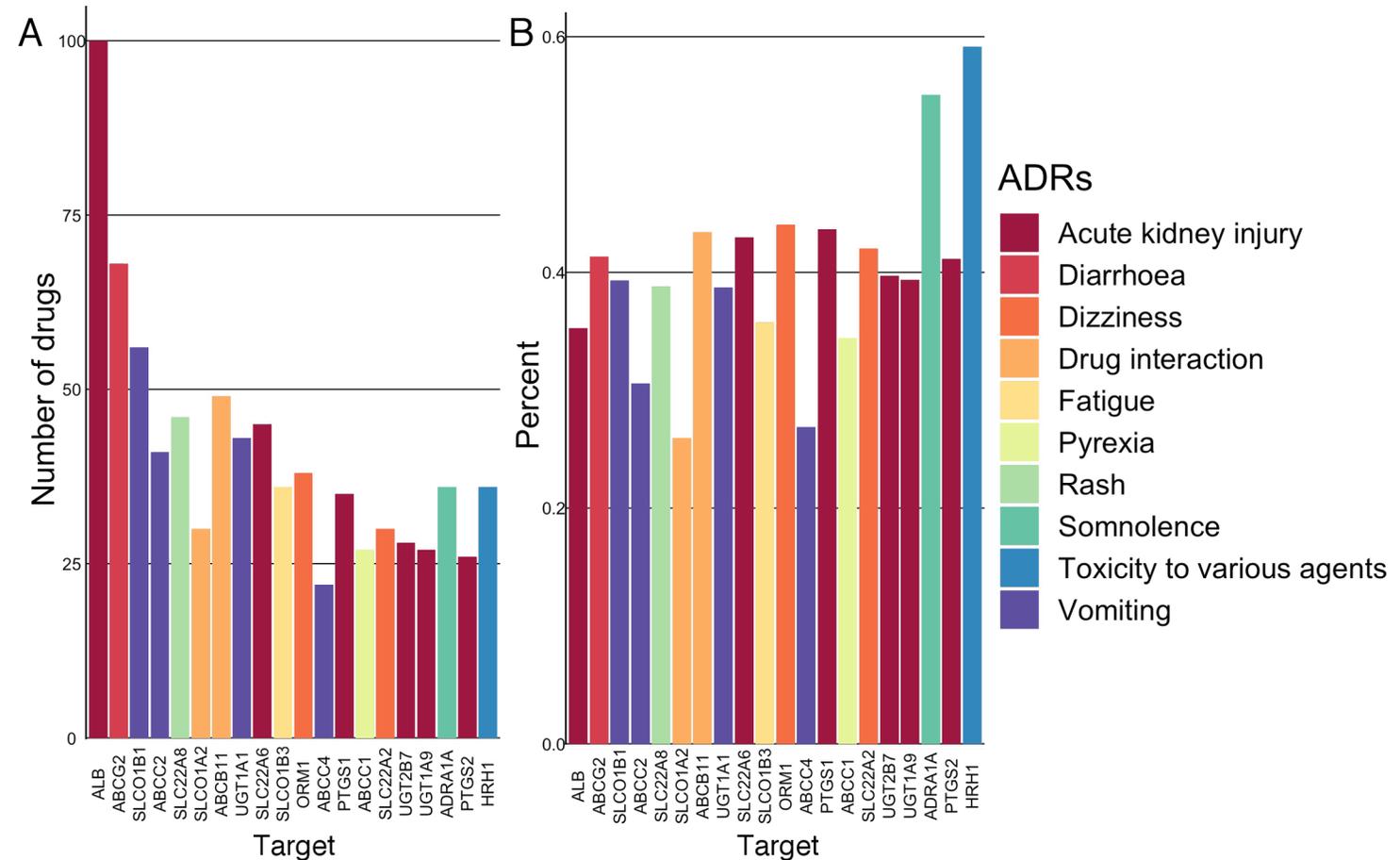
Figure : A) The 20 first proteins associated the most to drugs and ADRs. B) The proportion of these 20 proteins associated with an ADR among the others proteins linked to this ADR.

Network Science : An example with adverse drug reactions

Drugs – Adverse Drug Reactions (ADR) - System Organ Classification (SOC) network

- The majority of proteins do not seem to be specific to one SOC

- Development of a scoring function to estimate the contribution of each protein to an ADR

Scoring function

$$ADRx= (\sum_{g=1}^{h} \frac{1}{T_{D_g}} D_g)Protein_1 + (\sum_{i=1}^{j} \frac{1}{T_{D_i}} D_i)Protein_2 + \ldots + (\sum_{k=1}^{l} \frac{1}{T_{D_k}} D_k)Protein_n$$

$D_{(g,i,k)}$: Drugs known to cause ADR_x, that also interact with the protein associated with ADR_x
$T_{D(g,i,k)}$: Total number of proteins interacting with D_i
$Protein_{(1\ldots n)}$: Protein associated with ADR_x
g,i,k can represent the same drug if it targets multiple proteins for ADR_x

- 'Prostate cancer stage IV' = 0.5xGNRHR + 0.5xCYP3A4 (1 drug = leuprolide)

- 'Bone fragmentation' = 2.033xFDPS + 0.5xPTGS2 + 0.333xABCC1 + 0.333xGGPS1 + 0.2xATP6V1A + 0.2xPTPN4 + 0.2xPTPRE + 0.2xPTPRS (4 drugs = alendronic, ibandronate, risedronic, zoledronic)

→ **Such network can suggest the role and the importance of targets involved in an ADR**

ADR – ADR network

- 1000 Drugs
- 6164 ADRs

Network Science : Environmental disease network (EDN)

A semi-quantitative network-based analysis to predict disease-disease associations based on chemical-disease data

Example with Type 2 Diabetes and hormonal change

Disease-disease associations:

Chemical-Target-Disease associations:

Chemical-protein-disease network

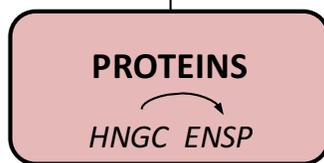# Network Science : Integrative systems chemical toxicology
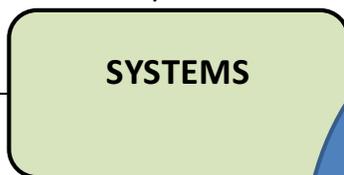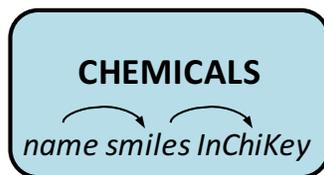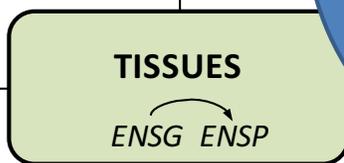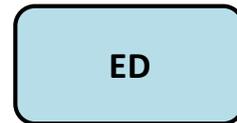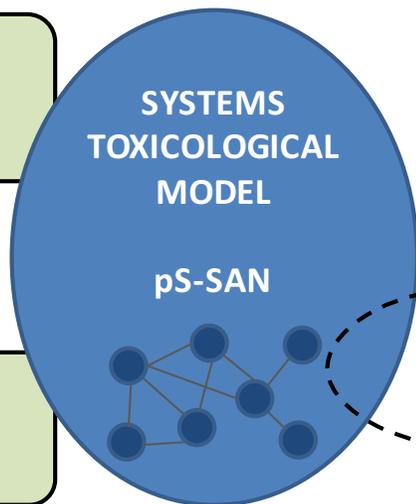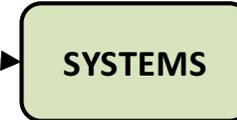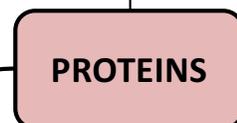
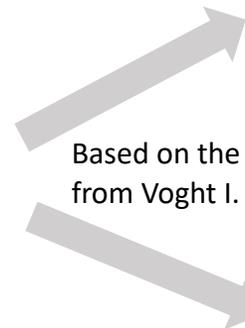**Chemical-protein-system toxicology network**
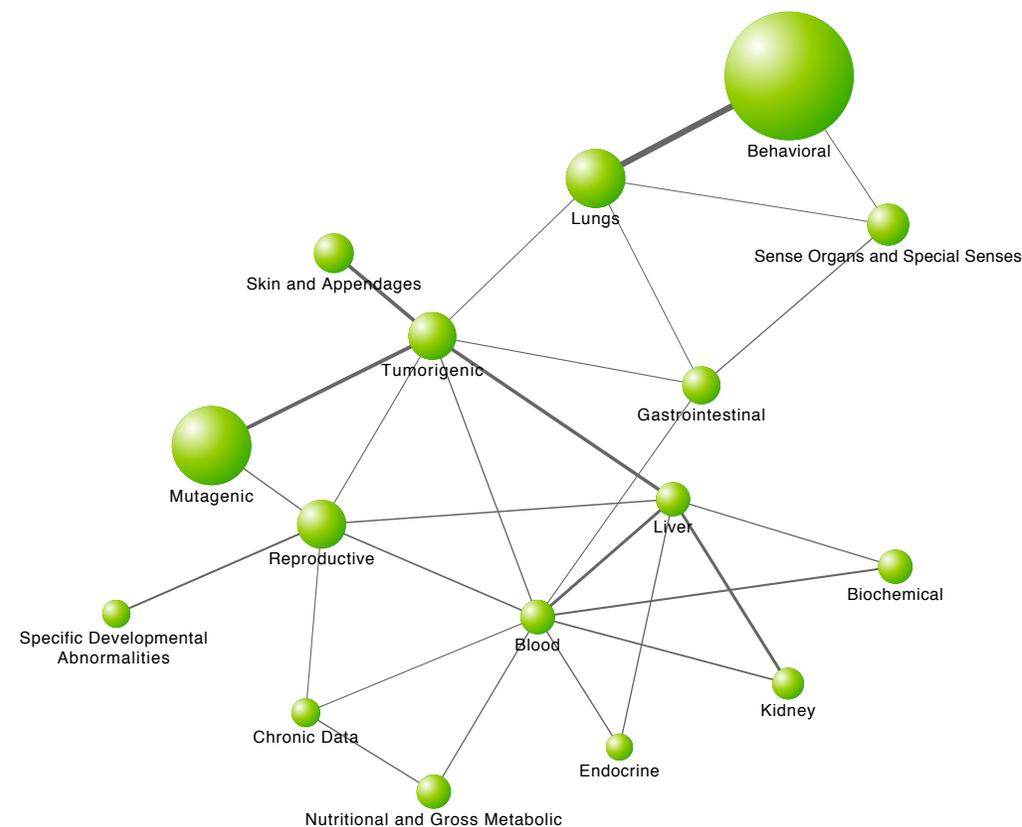
Network Science : Measuring the loss of coverage in network biology
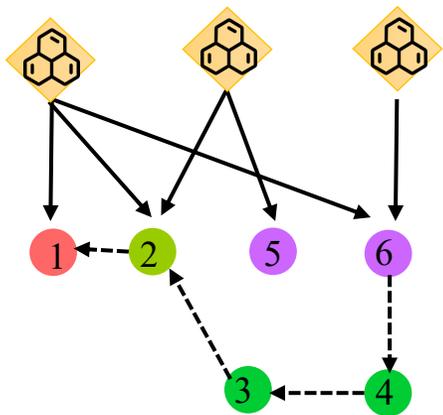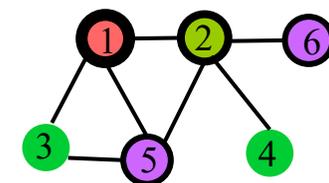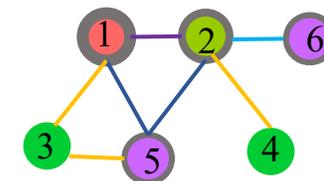

drug-event bipartite network (from AOP)
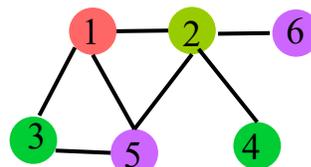
event- event network

Based on the work from Voght I. 2019

Increase in uncertainty

loss of coverage

Data compilation & Bipartite network development
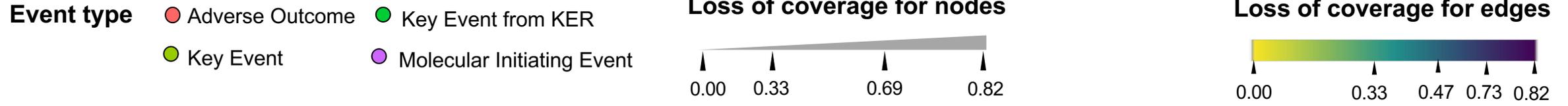
Monopartite network development

Network analysis

- Molecular initiating event
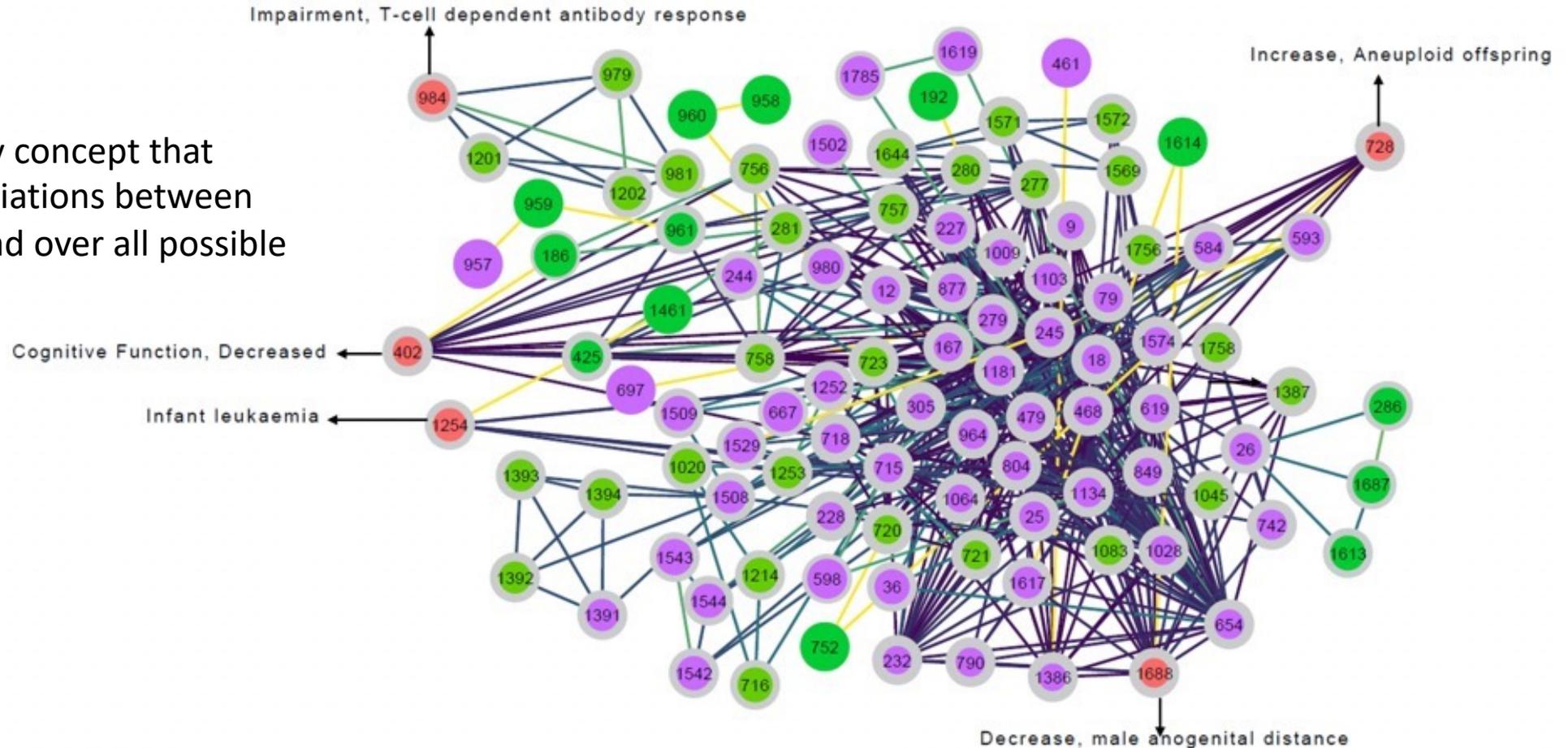- Key event
- Key event from KER
- Adverse outcome
- Drug

→ Drug-event associations

⇢ Key event relationships (KER)

— Event-event relationships

Network Science : Measuring the loss of coverage in network biology

Based on Shannon entropy concept that measures how many associations between drugs and events are spread over all possible connections in a network

Wu Q. et al. Front Public Health. 2021

Network Science : Measuring the loss of coverage in network biology

Network Science : Chemical Disease Network
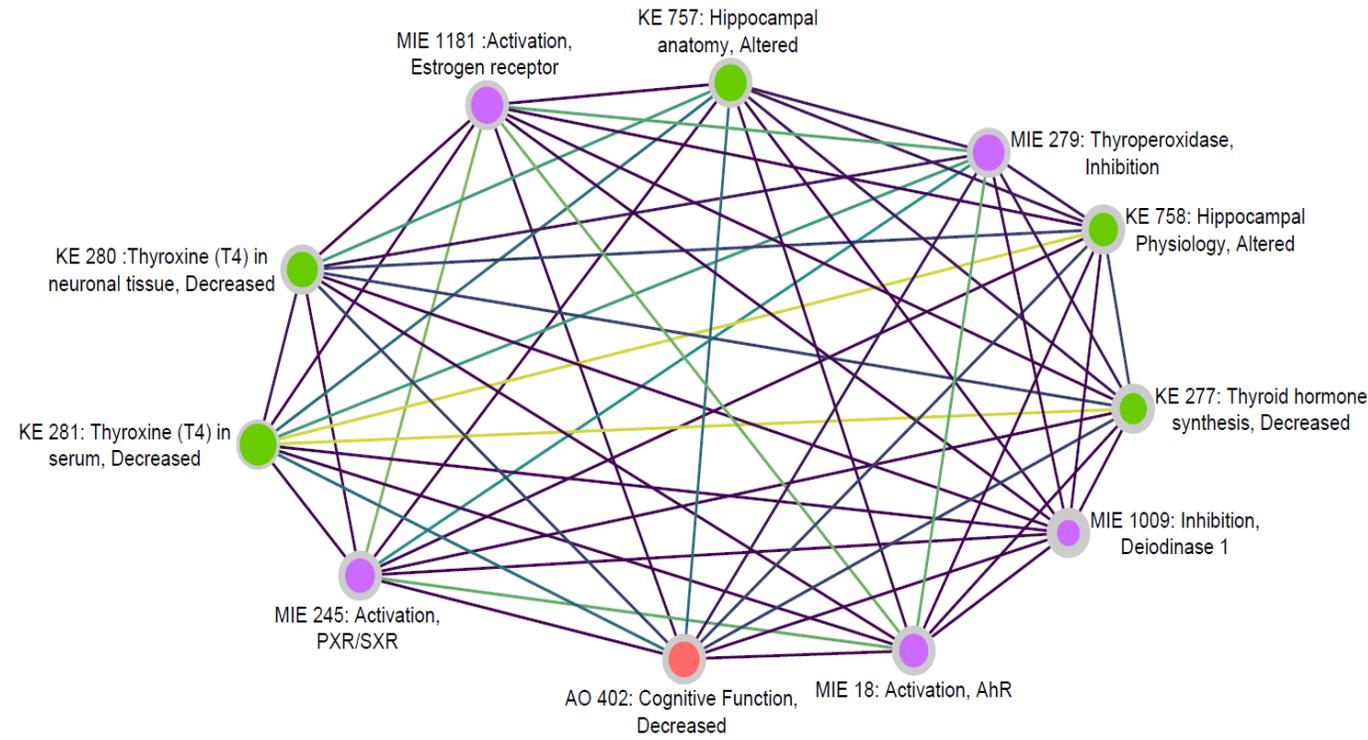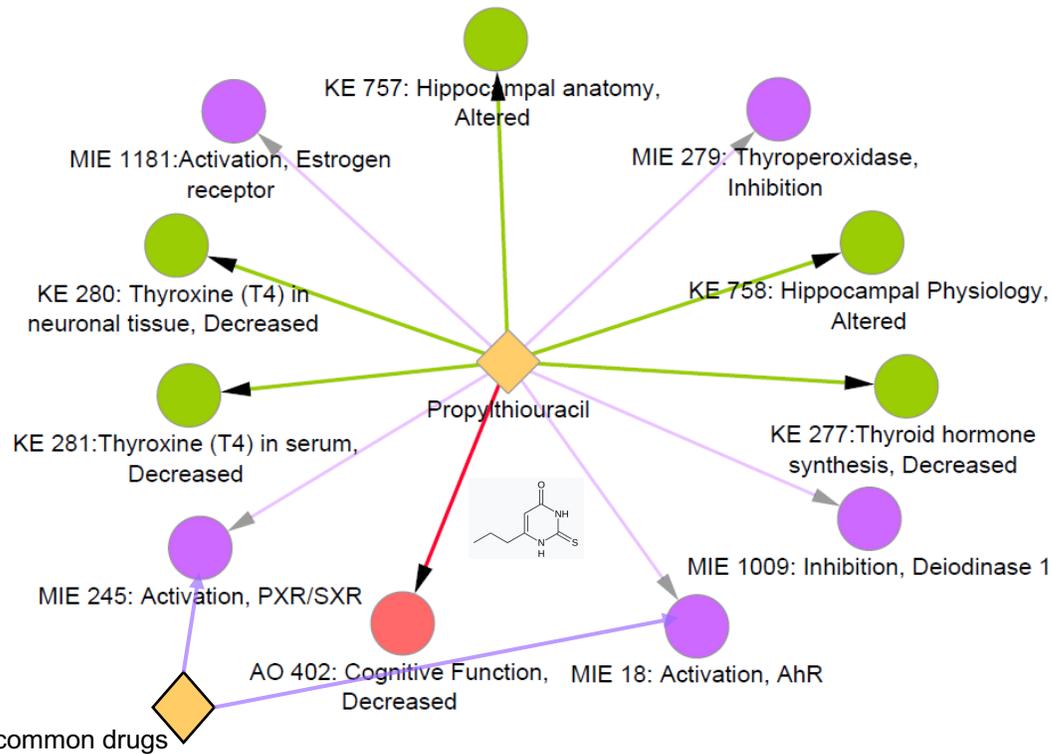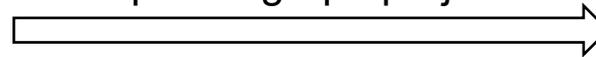
> **Link the proteome perturbation by a compound leading to morphological cell changes and diseases.**
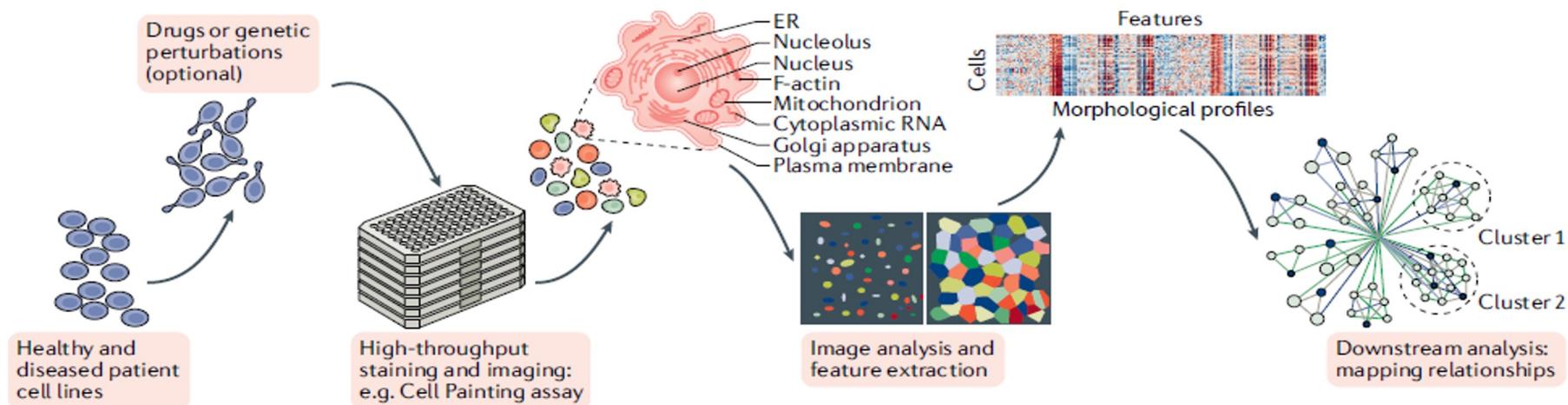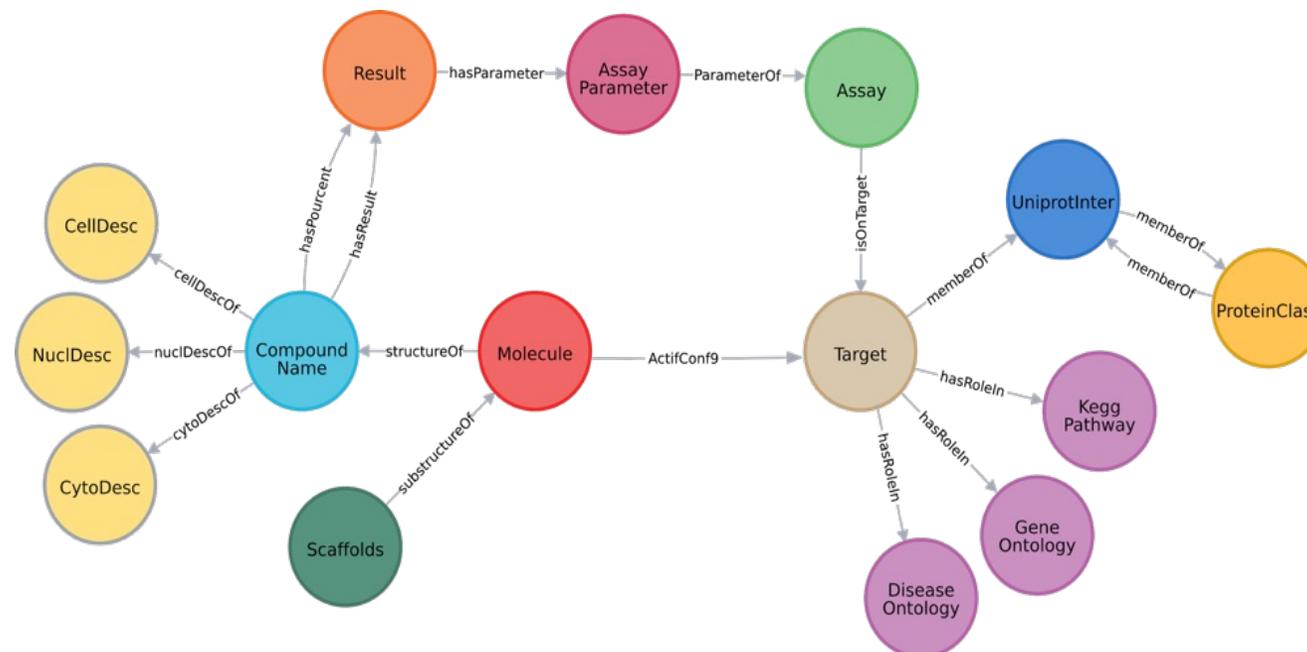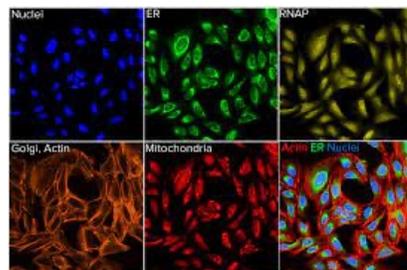

High Content imaging => Morphological cell perturbation (Cell painting)

- 29935 compounds tested on osteosarcoma cells (U2OS)

- Aim to colour the 7 major cell compartments with fluorescent dyes

- Fluorescence capture and automatic image analysis

- Extraction of the cell parameters: shape, texture, intensity, *etc*. using CellProfiler (http://cellprofiler.org/)

- Cell Painting morphological profiling assay [Bray MA, *et al*. Nat Protoc 2016]
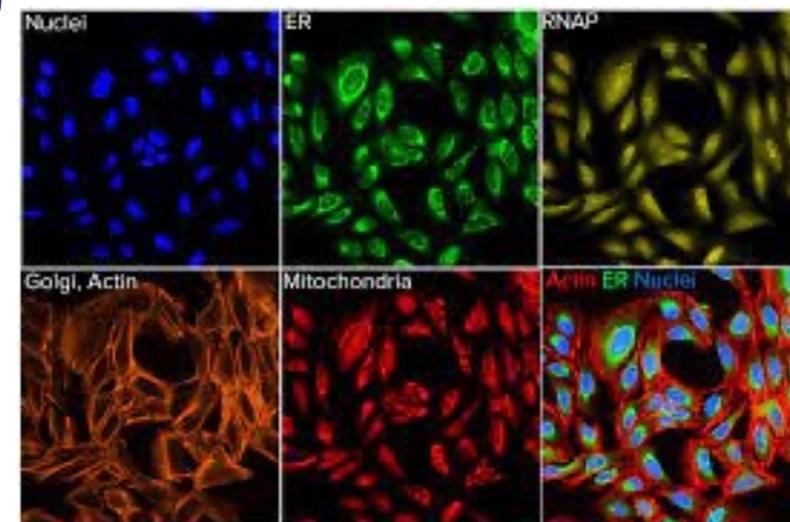
Link the proteome perturbation by a compound leading to morphological cell changes and diseases.
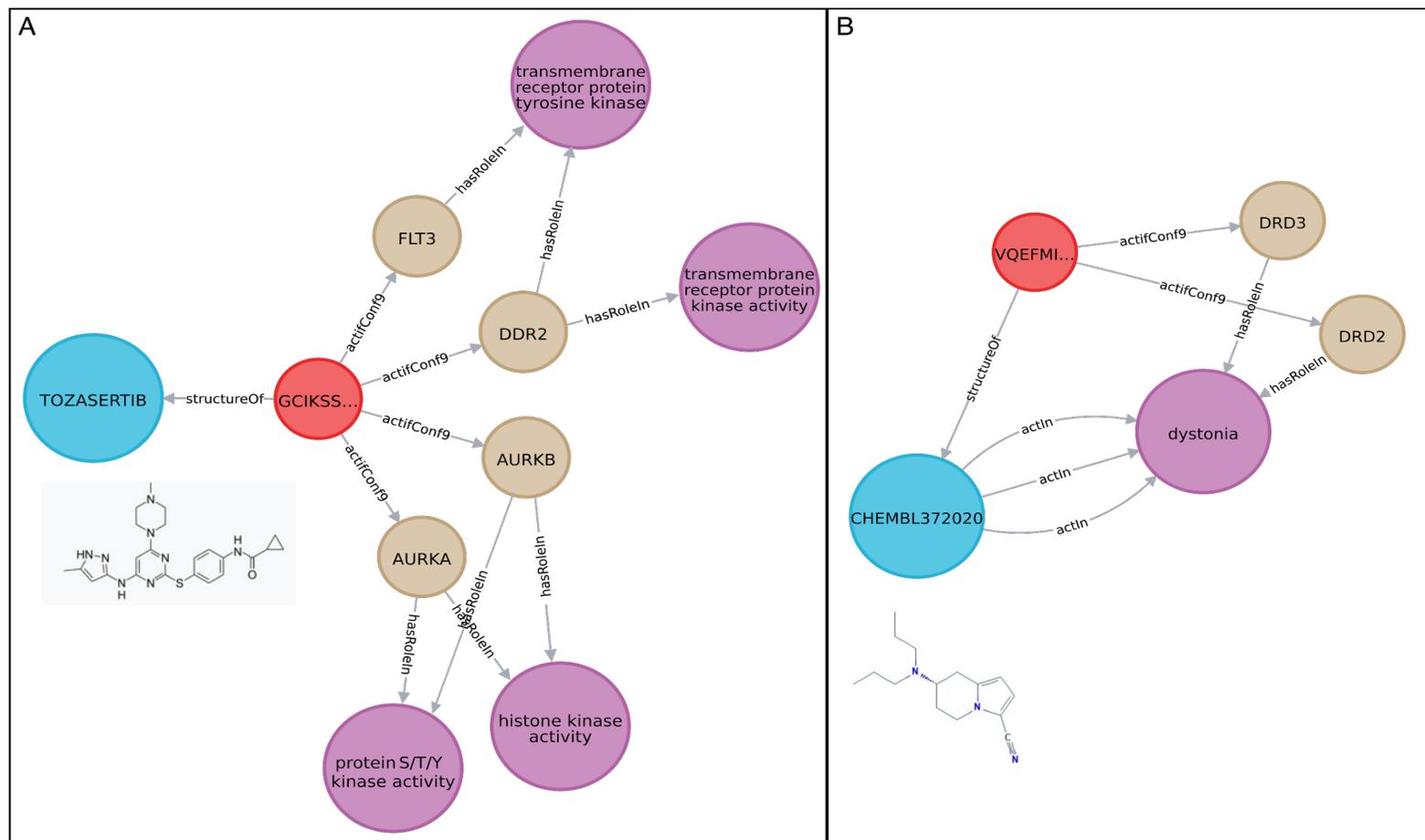
Dafniet et al., J. Cheminf. 2021

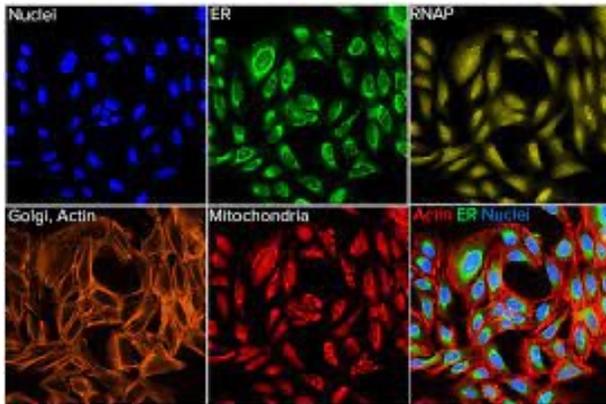Example 1: Ensemble of protein's target with IC50 < 1 μM

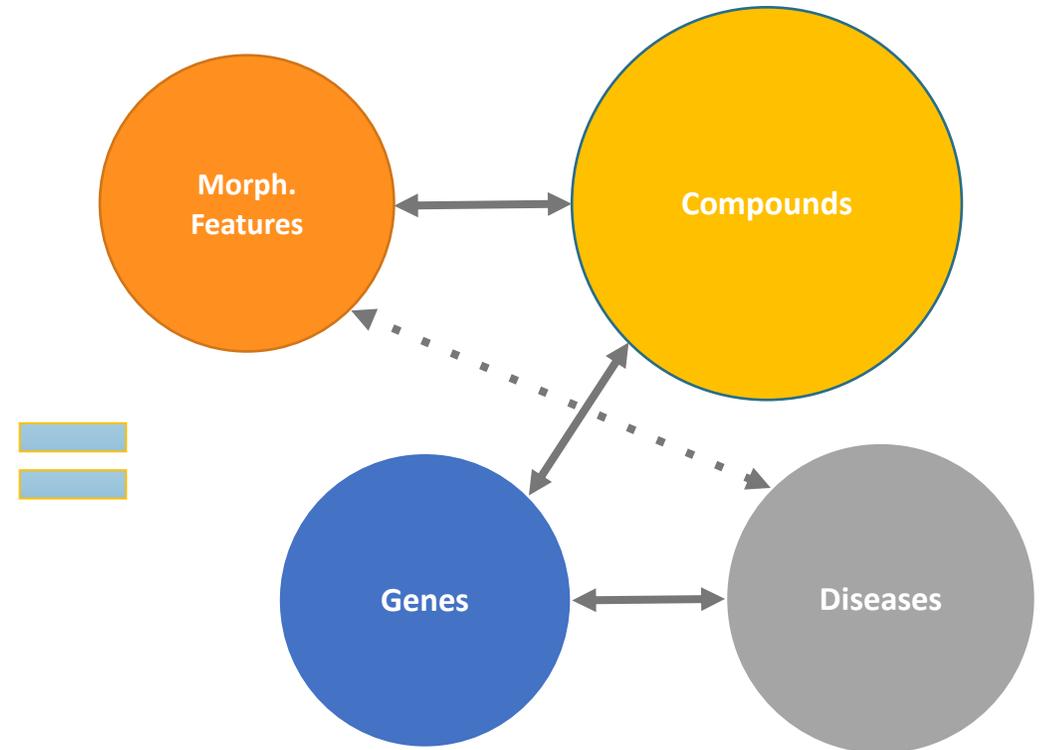## Example 2: Pathways and diseases enrichment associated to a chemical
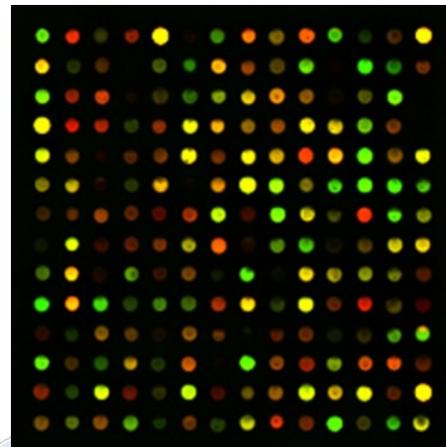
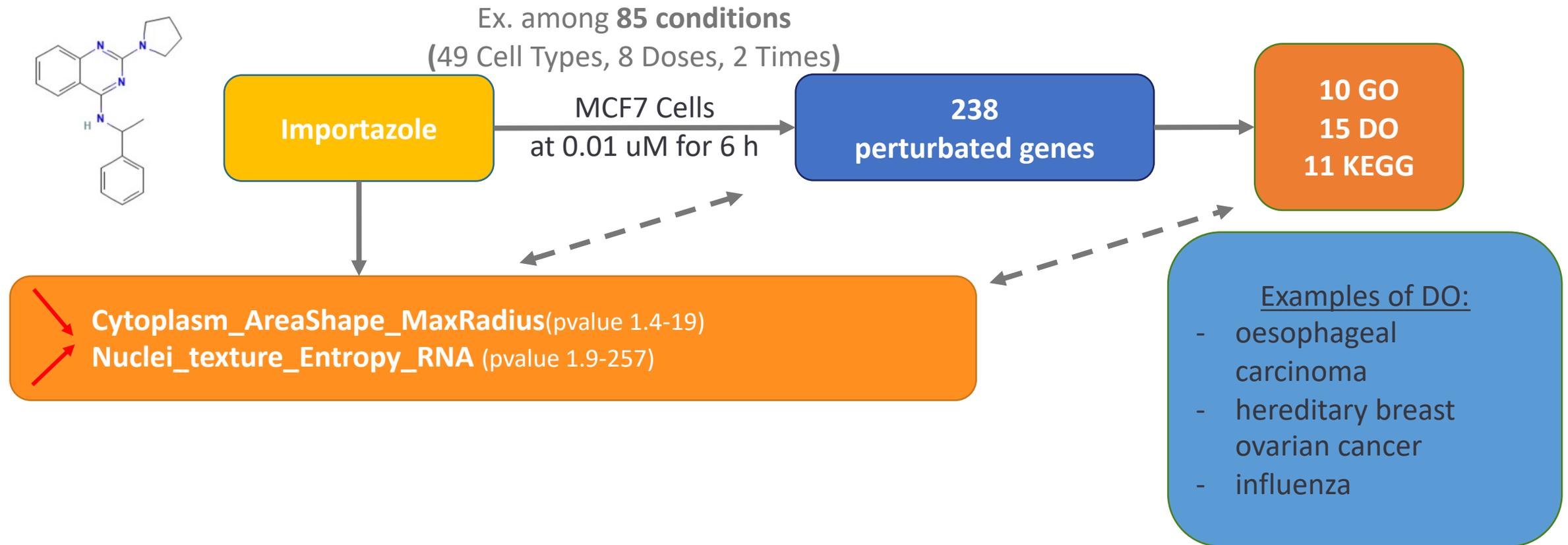**Link the gene deregulation by a compound to morphological cell changes and diseases.**

Morphological features

Transcriptomics

Link the gene deregulation by a compound to morphological cell changes and diseases.

Ex. among **85 conditions**
(**49 Cell Types, 8 Doses, 2 Times**)

Importazole

MCF7 Cells
at 0.01 uM for 6 h

**238
perturbated genes**

**10 GO
15 DO
11 KEGG**

Cytoplasm_AreaShape_MaxRadius(pvalue 1.4-19)
Nuclei_texture_Entropy_RNA (pvalue 1.9-257)

Examples of DO:
- oesophageal carcinoma
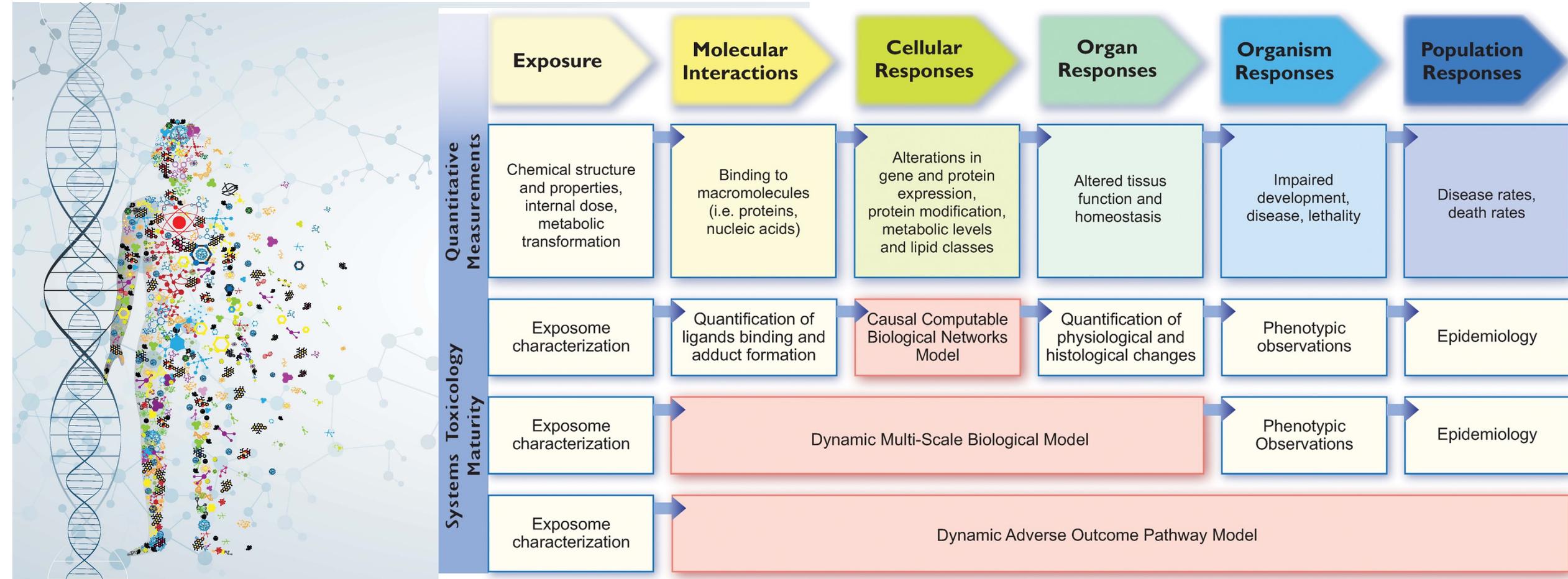- hereditary breast ovarian cancer
- influenza

▪ Suggestion of a relation between the morphological perturbation of some compartments in cells and diseases
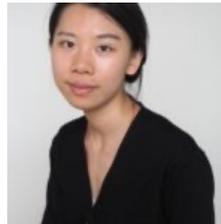
Work in progress

Conclusion

Data integration of biological information related to chemical is possible at different level of complexity (from molecular to population).

Acknowledgement

Questions?