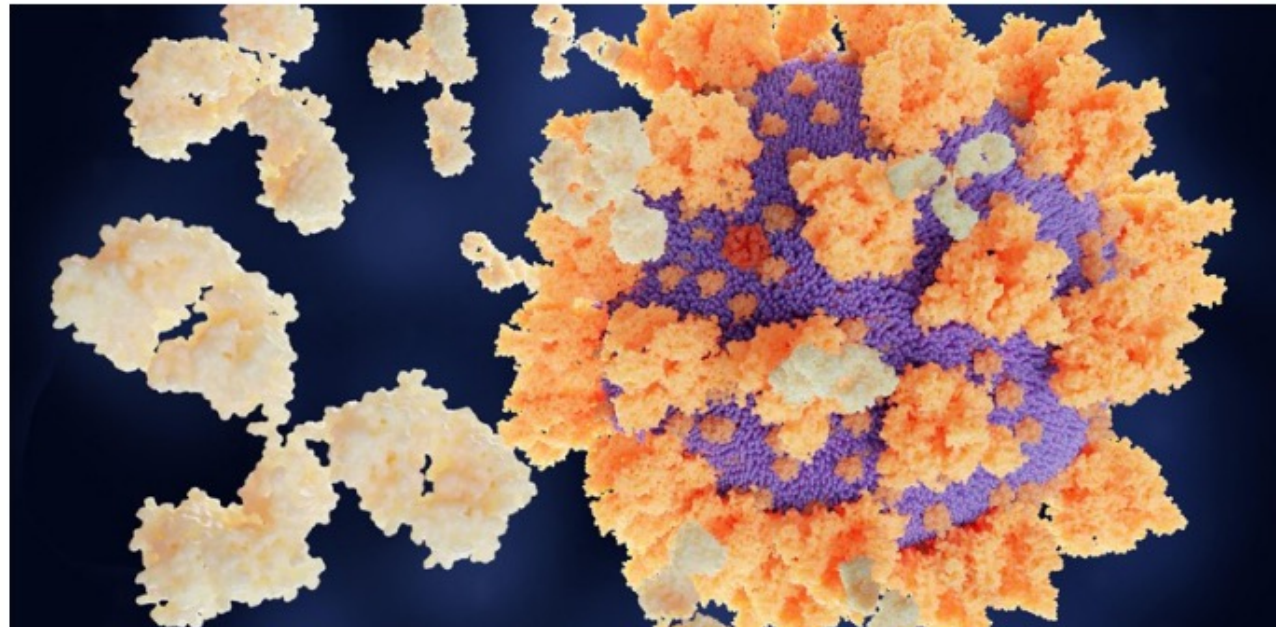


# Toward a Computational Precision Medicine

Mickaël Guedj



# Background

---



1999 - 2004

[INSA Lyon](#)

2004 - 2007

PhD at [Genopole](#) / [Merck-Serono](#)



2007 - 2009

*Computational Biologist*  
[Ligue Nationale contre le Cancer](#)

2009 - 2018

*Chief Data Officer*  
[Pharnext](#)

2018 - 2021

*Head of Computational Medicine*  
[Servier](#)

since 2021

*Head of Biometrics, Data & Decision Sciences*  
[Nanobiotix](#)



2005 - 2011

[ENSAI Rennes](#)



[mickael.guedj@gmail.com](mailto:mickael.guedj@gmail.com)



<https://www.linkedin.com/in/mickaelguedj>

# Principles

---

## Computational Precision Medicine

Integrate understanding of **disease mechanisms** & **patient heterogeneity**

By opposition to *one-size-fits-all*

Highly data & model-driven

### *Related*

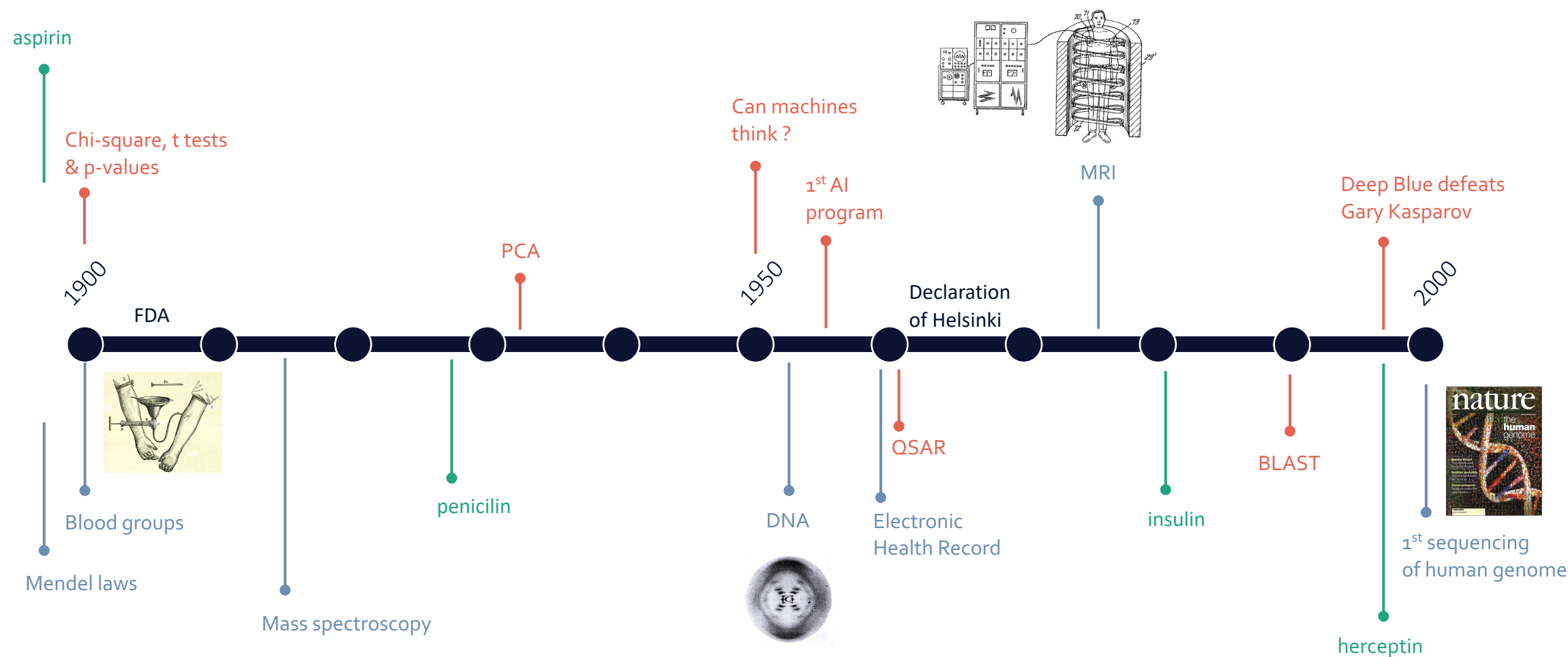
Stratified, Integrated, Systems, Network, In Silico, Digital, Data-driven, Translational

4P: Predictive, Preventive, Personalized & Participative



Credit: Lorenzo Gritti

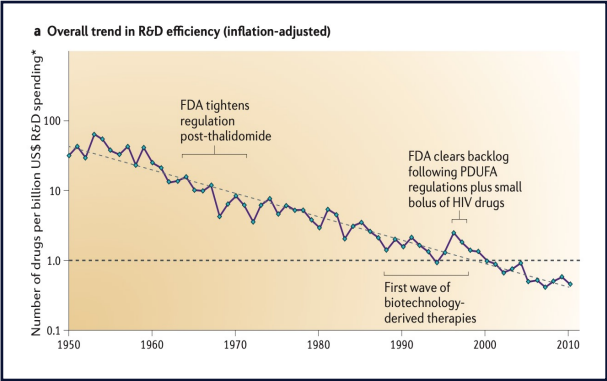
# Brief history





# Since 2000

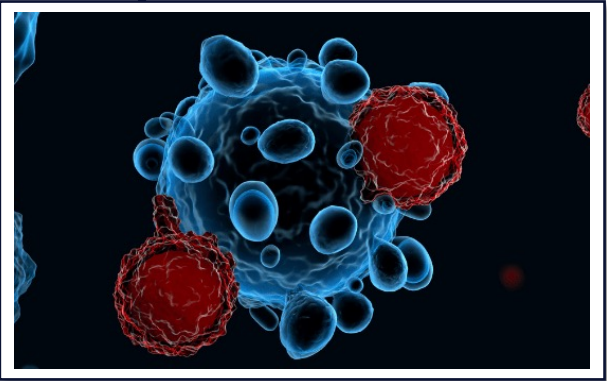
## Decline of R&D productivity



## Large-scale & multimodal patient profiling

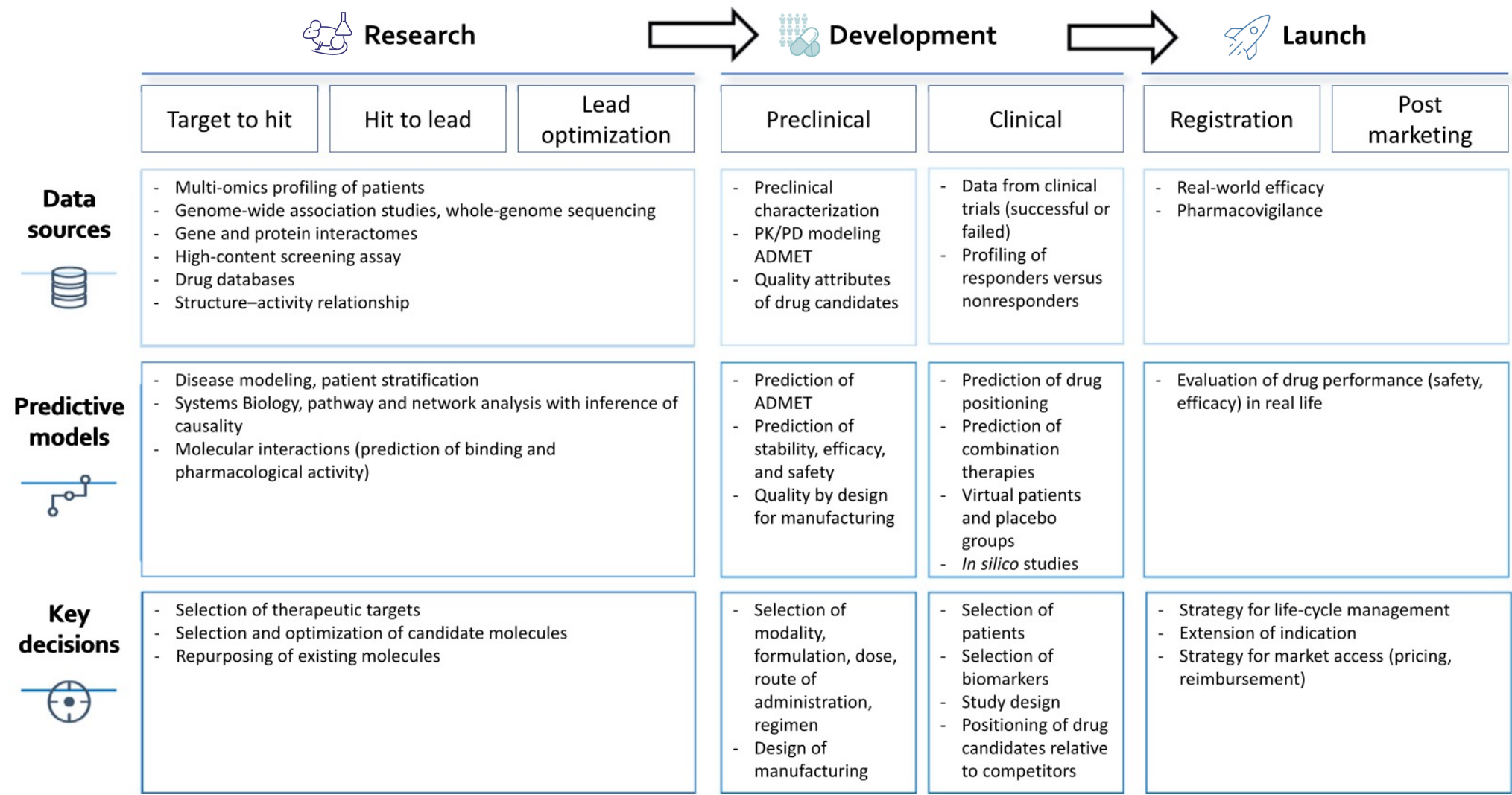


## Data storage, access & treatment capacities

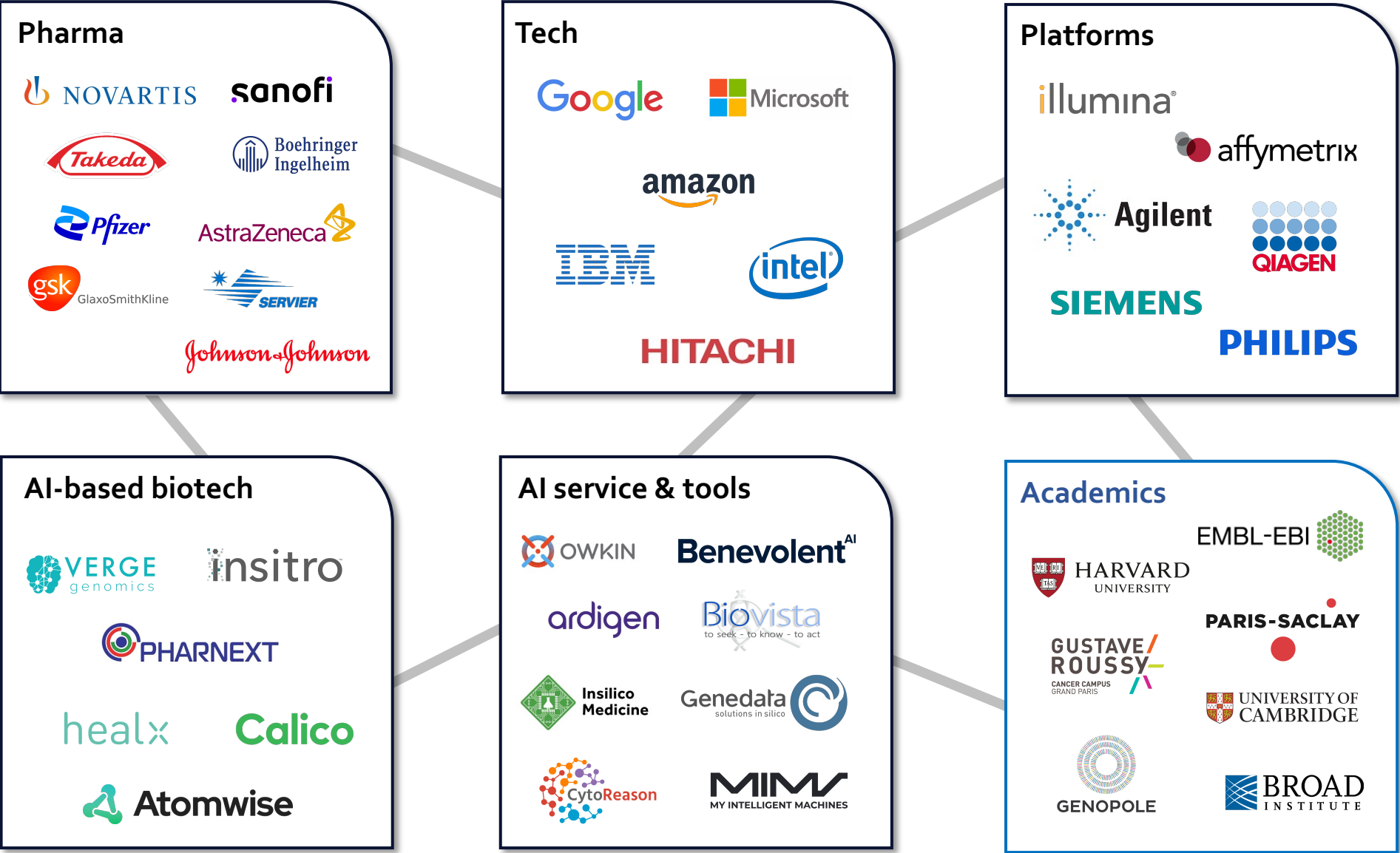


## Diversification of therapeutic strategies

# Data-driven decision making in drug discovery & development



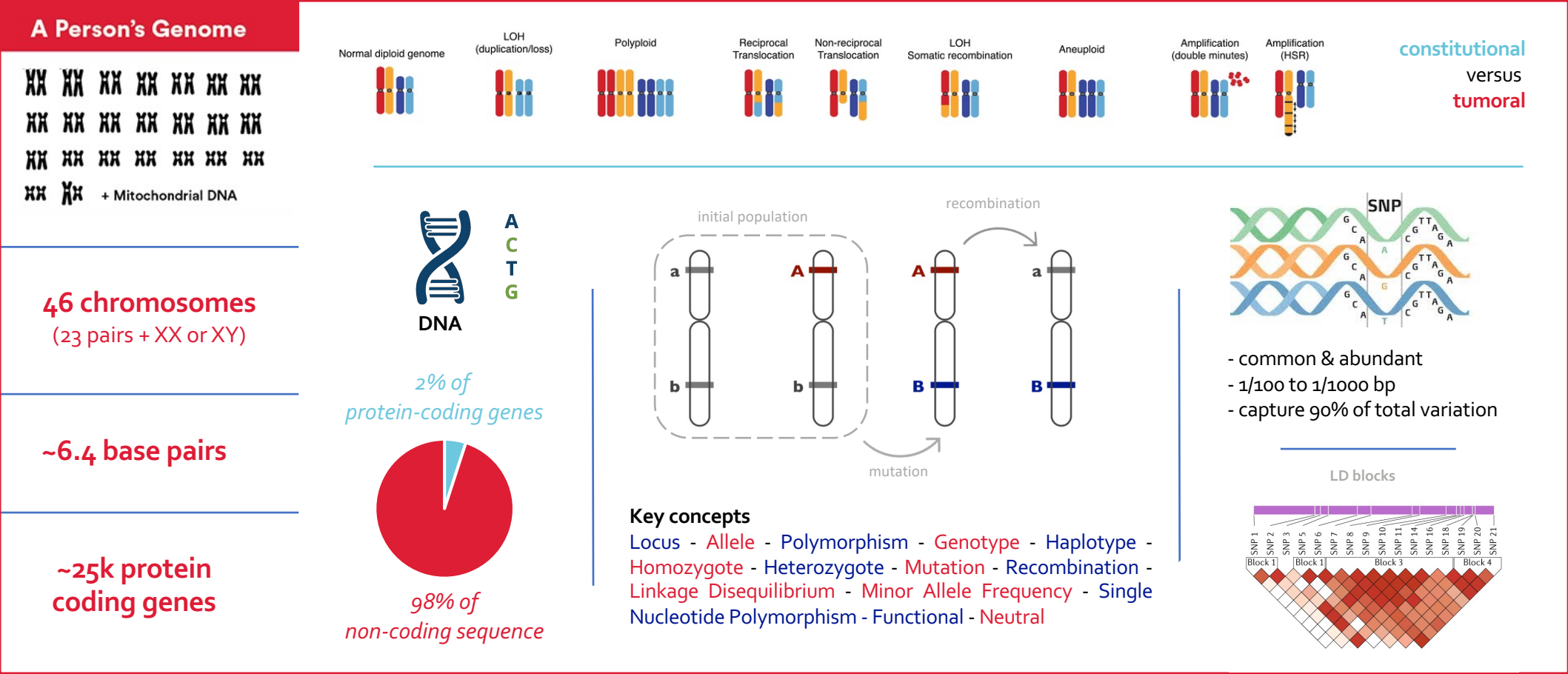
# Industrial ecosystem



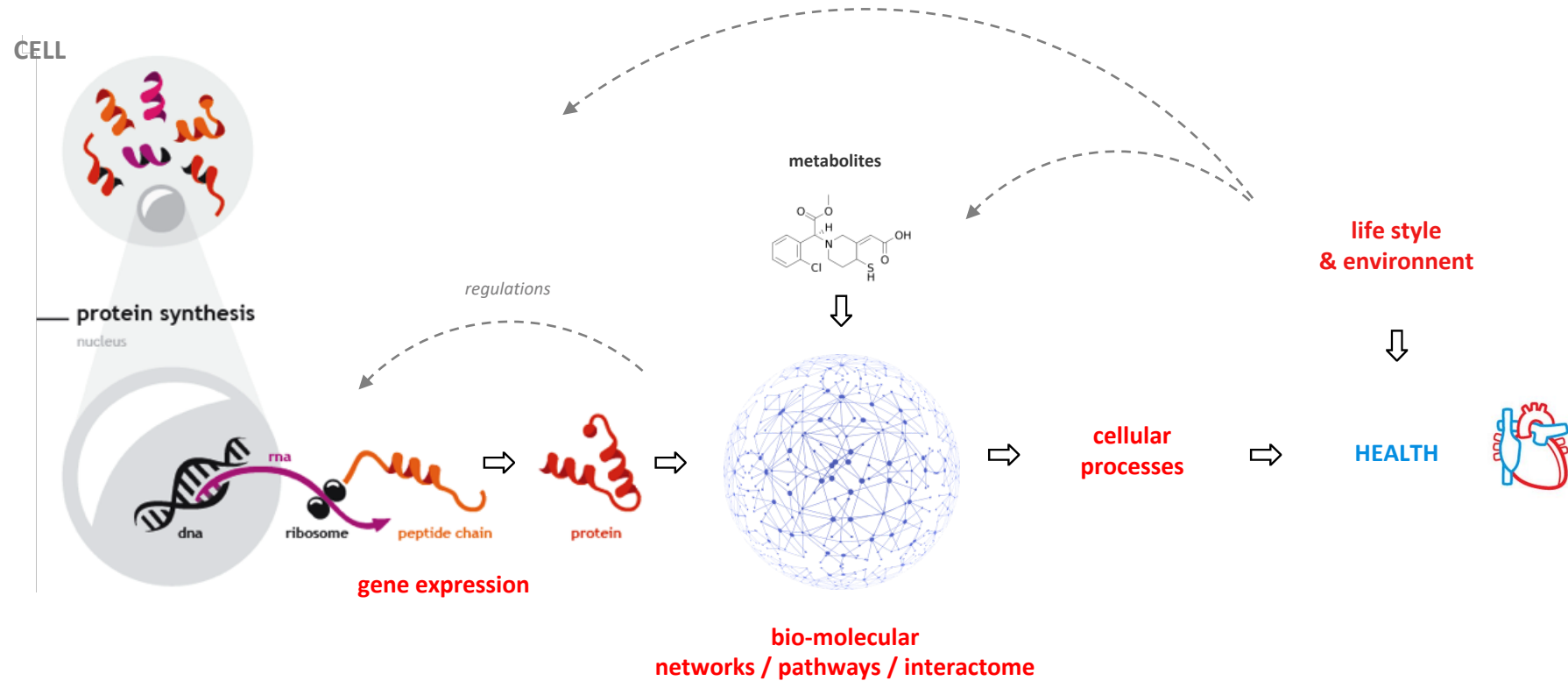
**Data**



# Genetics as a key entry



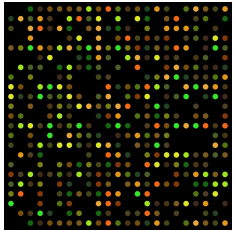
# Health as equilibrium of connected systems



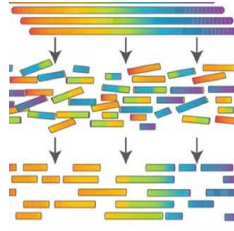
# Capturing the value of big biomedical data

## High-throughput technology platforms

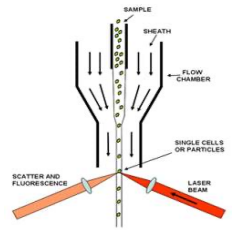
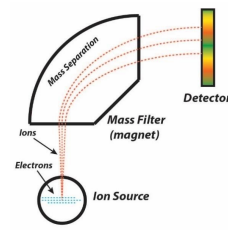
microarray



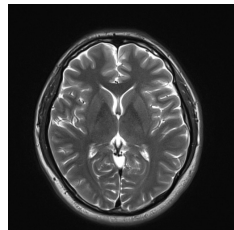
sequencing



spectroscopy



cytometry



imaging

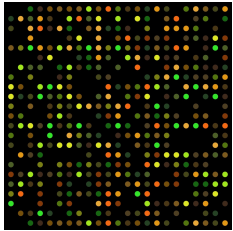


electronic  
capture

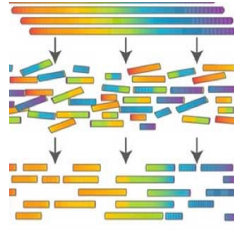
# Capturing the value of big biomedical data

## High-throughput technology platforms

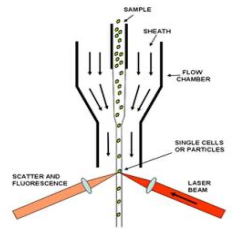
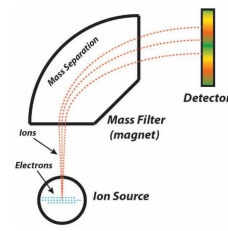
microarray



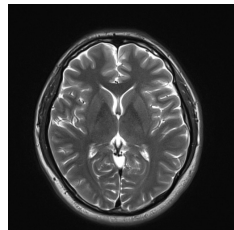
sequencing



spectroscopy



cytometry



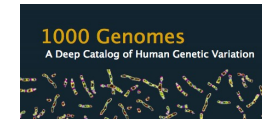
imaging



electronic  
capture

## Big data collecting initiatives

public, private, national, international, population-based, disease-centric

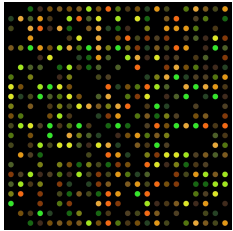




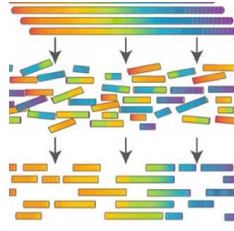
# Capturing the value of big biomedical data

## High-throughput technology platforms

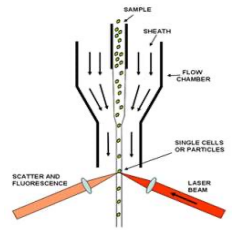
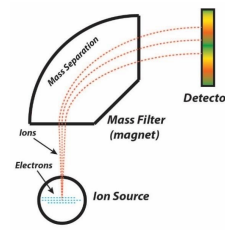
microarray



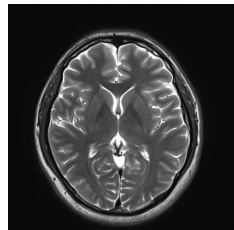
sequencing



spectroscopy



cytometry



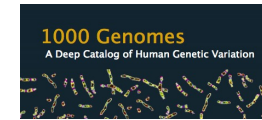
imaging



electronic  
capture

## Big data collecting initiatives

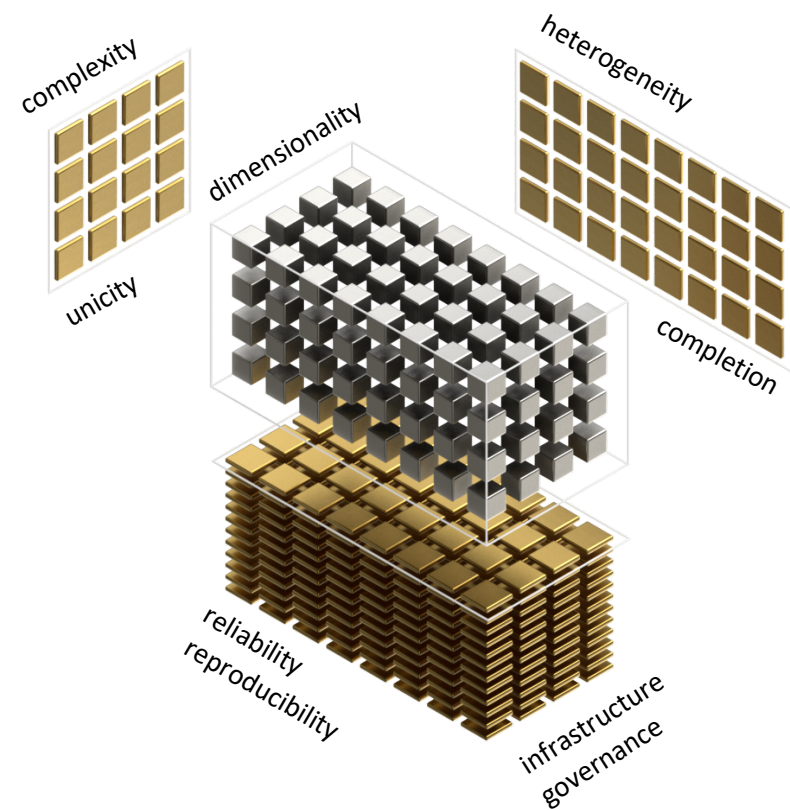
public, private, national, international, population-based, disease-centric



## Structured data / knowledge bases



# Data-associated challenges

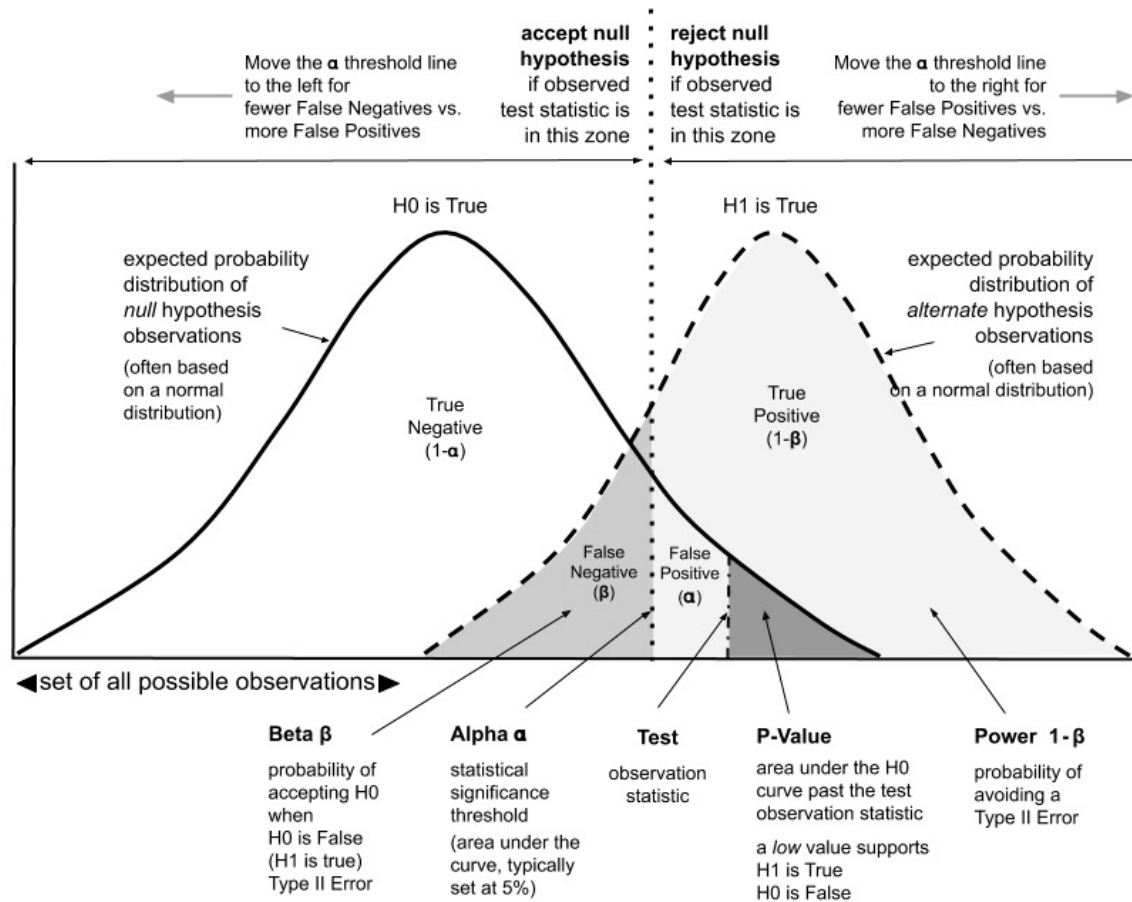


Emerging standards  
& guidance



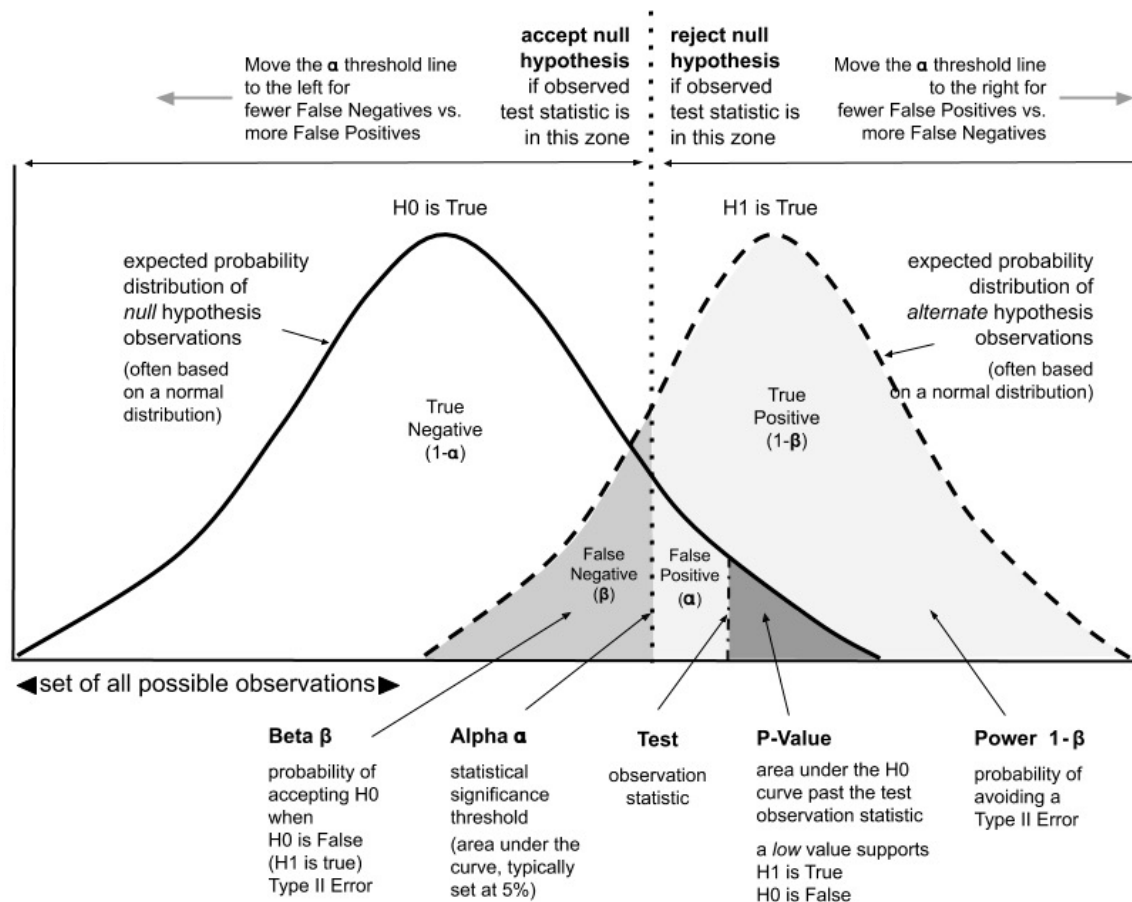
# Computational tool box

# Statistical testing & modelling





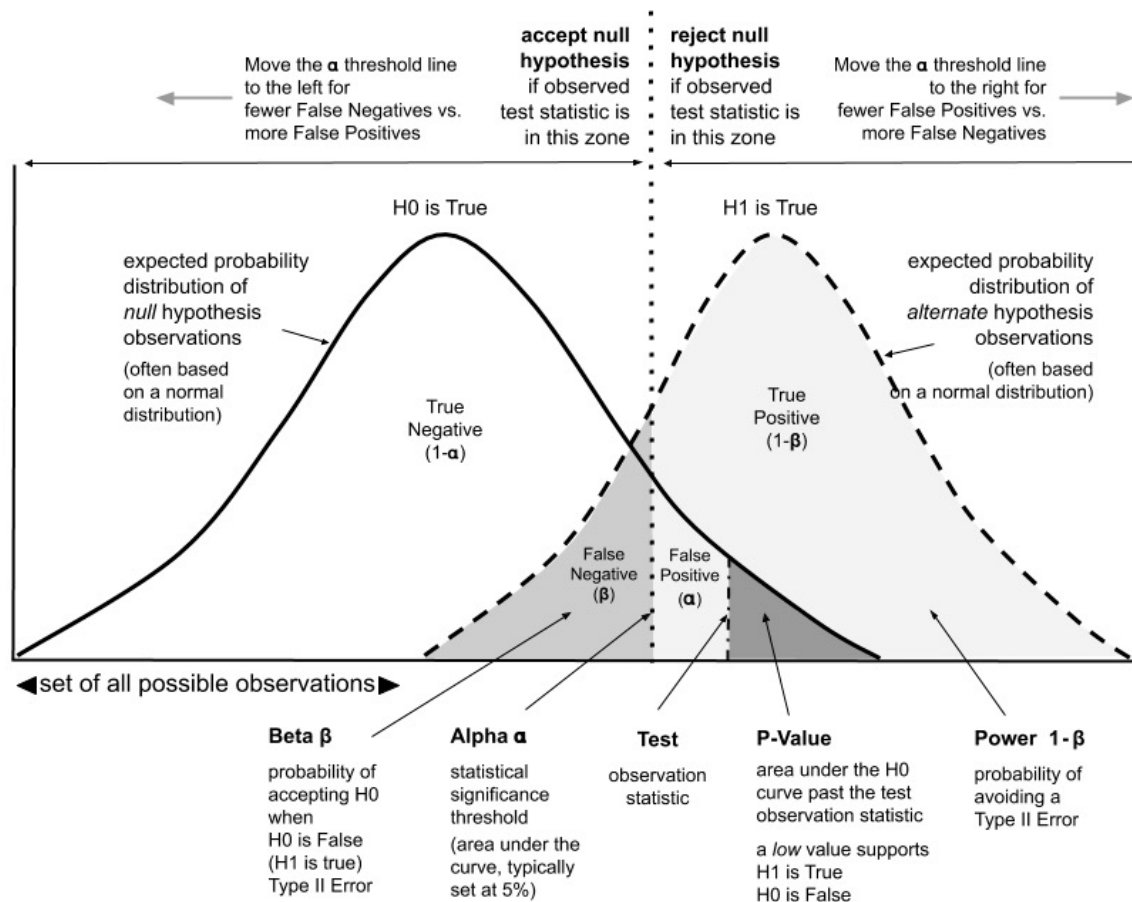
# Statistical testing & modelling



## Multiple-testing

- FWER, FDR,  $q$ -value, local FDR
- Bonferroni, Benjamini-Hochberg

# Statistical testing & modelling

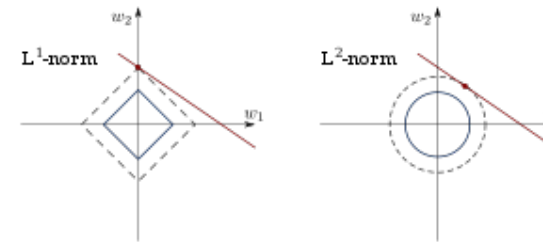


## Multiple-testing

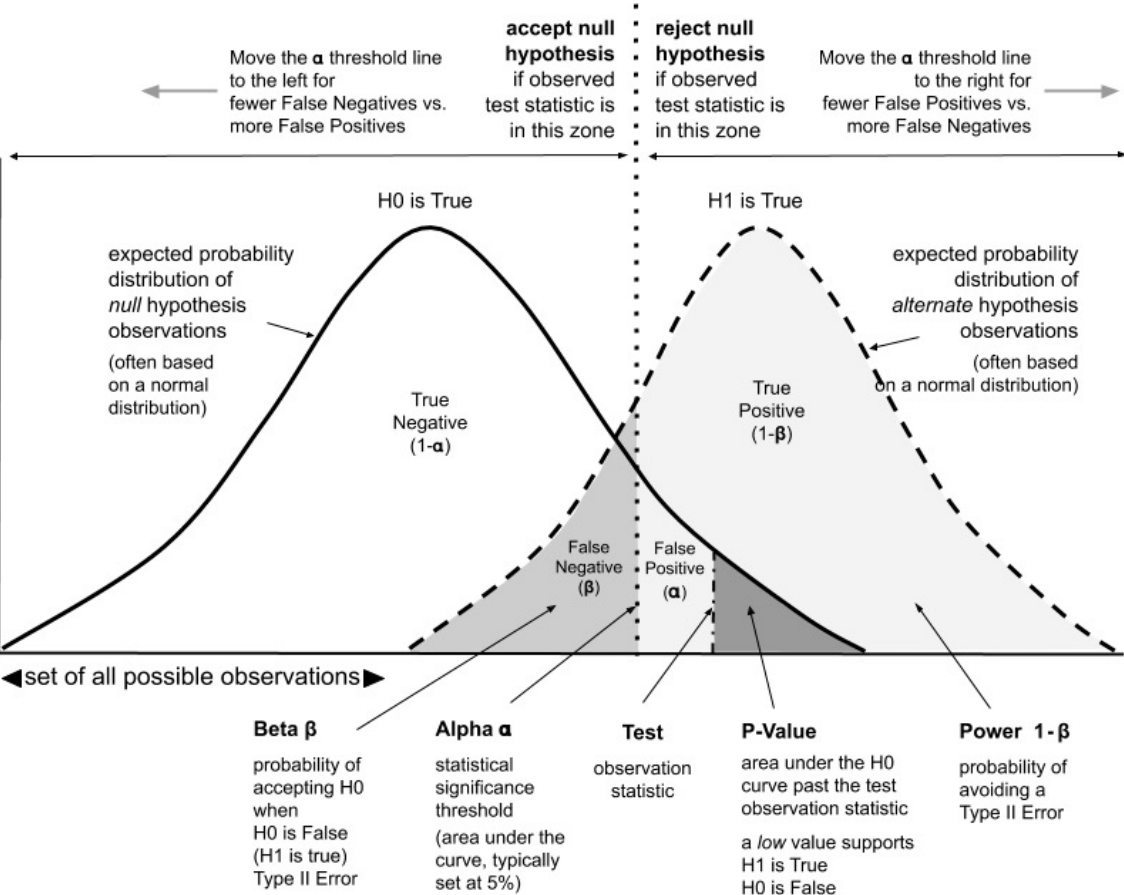
- FWER, FDR,  $q$ -value, local FDR
- Bonferroni, Benjamini-Hochberg

## Model estimation & variable selection

- Ridge, LASSO, Elastic Net



# Statistical testing & modelling

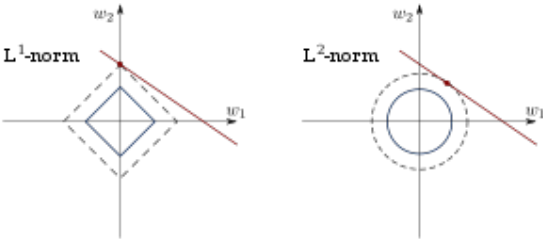


## Multiple-testing

- FWER, FDR,  $q$ -value, local FDR
- Bonferroni, Benjamini-Hochberg

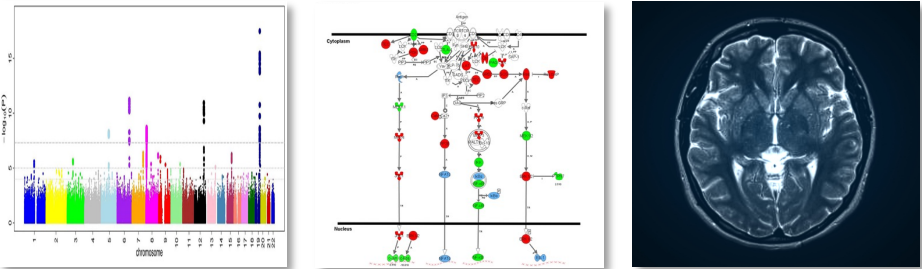
## Model estimation & variable selection

- Ridge, LASSO, Elastic Net



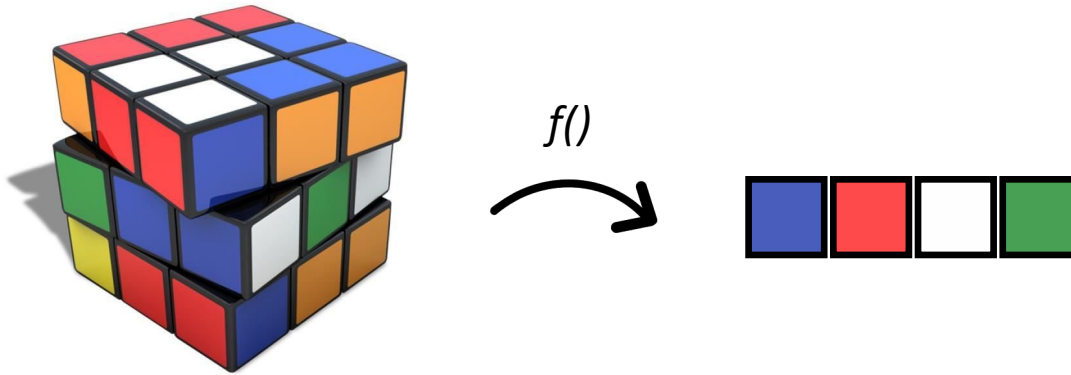
## Spatiality

- sliding-window, local score, enrichment

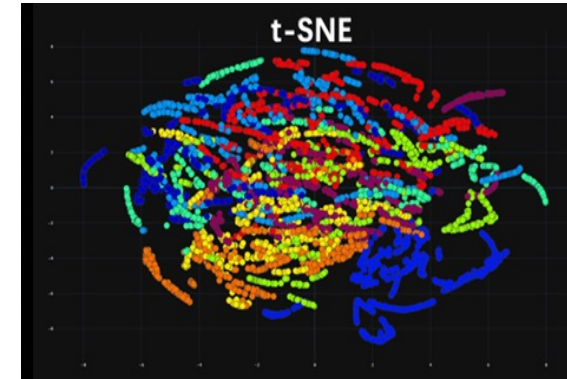
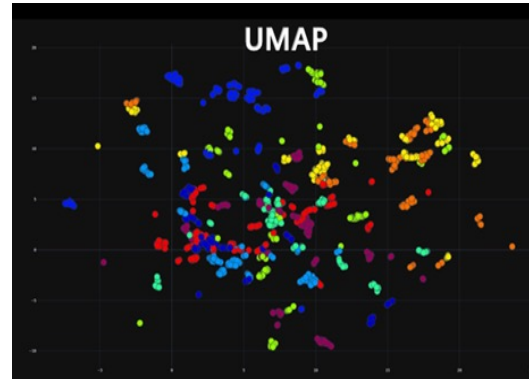
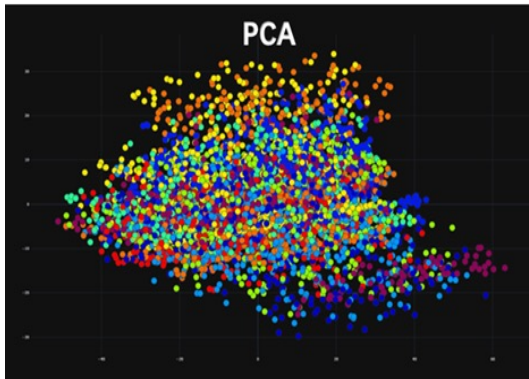


# Dimension reduction

Transformation of data from a high-dimensional space into a low-dimensional space so that it retains some meaningful properties

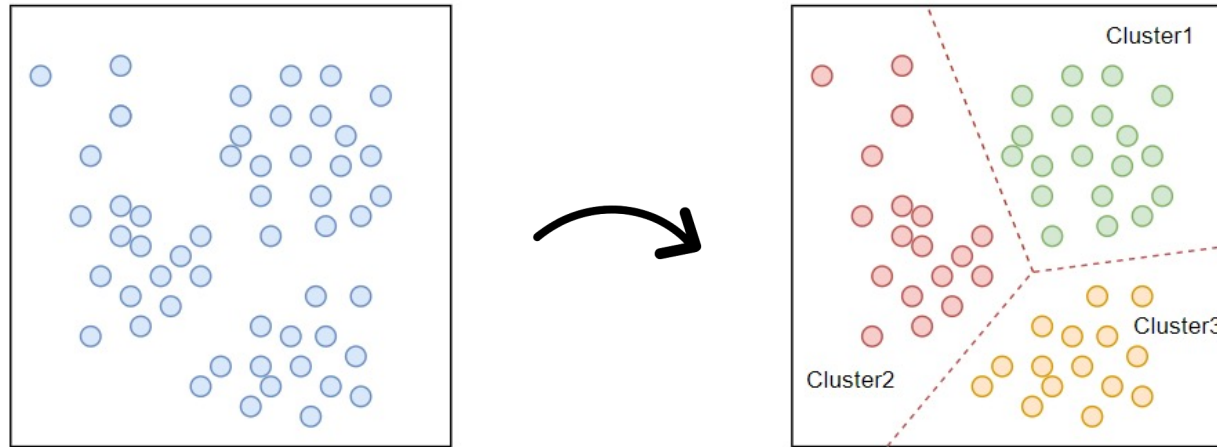


- Principal Component Analysis (**PCA**)
- Uniform Manifold Approximation and Projection (**UMAP**)
- t-distributed stochastic neighbour embedding (**t-SNE**)
- **Auto-encoders**





# Unsupervised classification / clustering



- **Connectivity approach**

Hierarchical clustering

- **Centroid approach**

K-means

- **Distribution approach**

Gaussian mixture models

- **Graph-based approach**

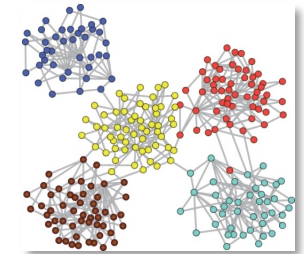
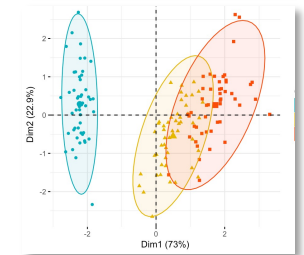
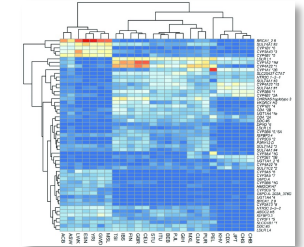
WGCNA

- **Decomposition approach**

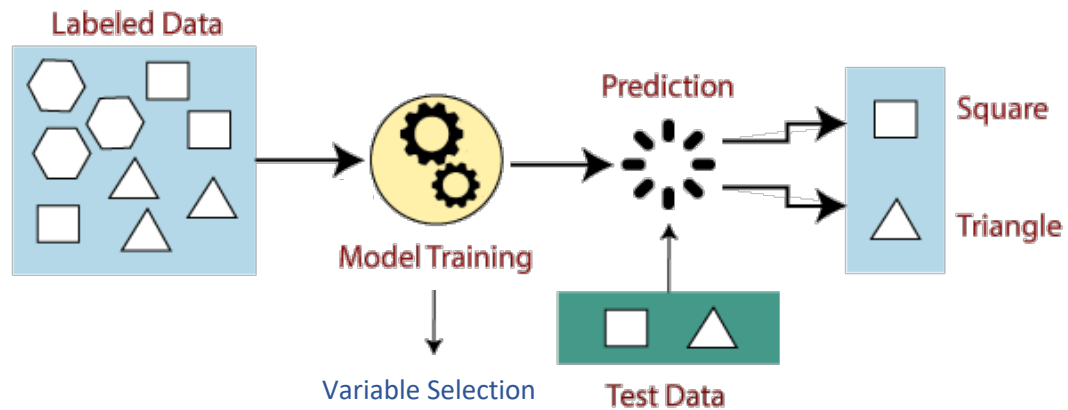
Spectral clustering

- **Neural networks**

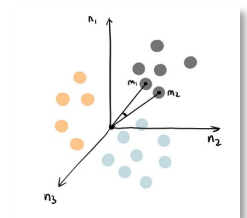
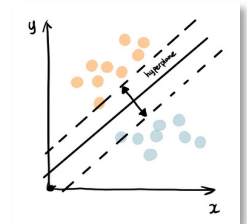
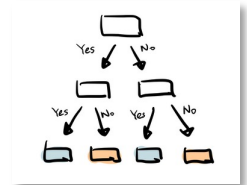
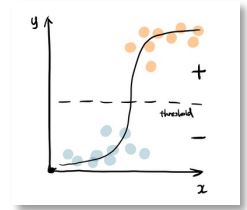
Self-organizing map



# Supervised classification



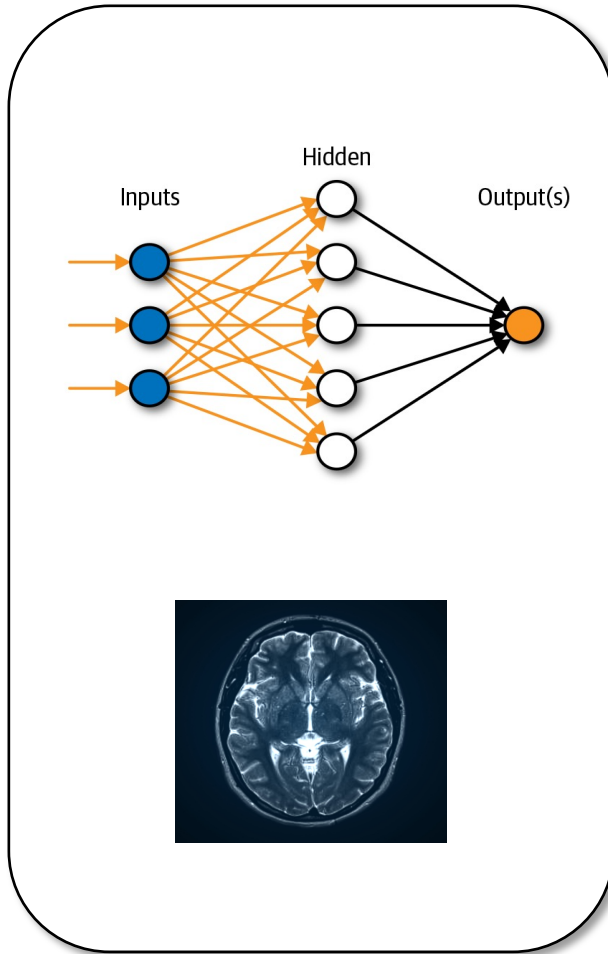
- Logistic regression
- Decision tree, random forests
- Naïve bayes classifier
- Support vector machine
- K-nearest neighbour
- Neural networks



# Bio-inspired algorithms

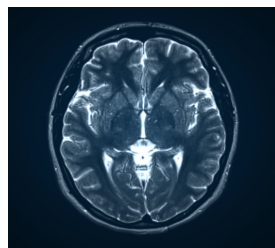
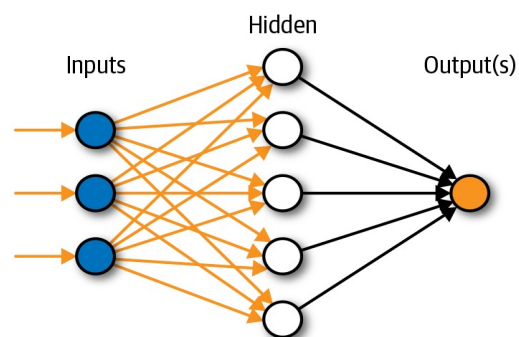
---

## Neural networks

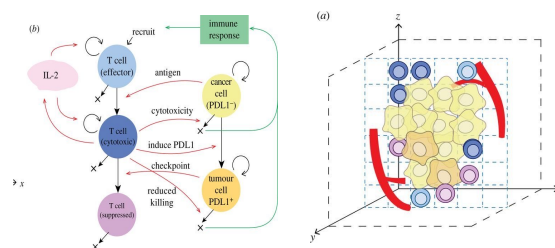
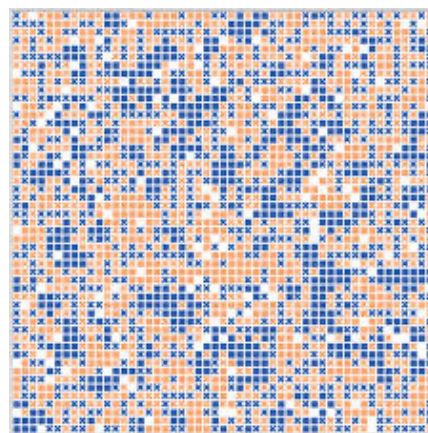


# Bio-inspired algorithms

## Neural networks

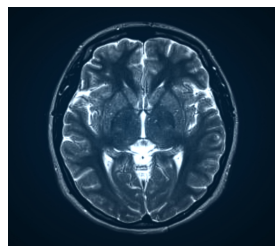
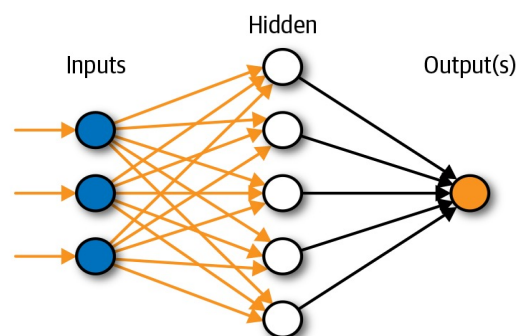


## Multi-agent systems

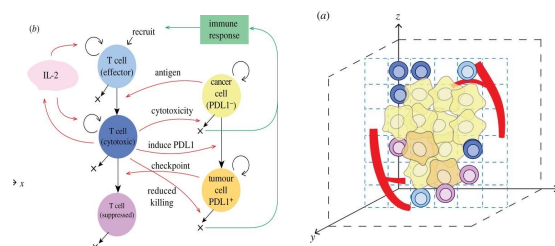
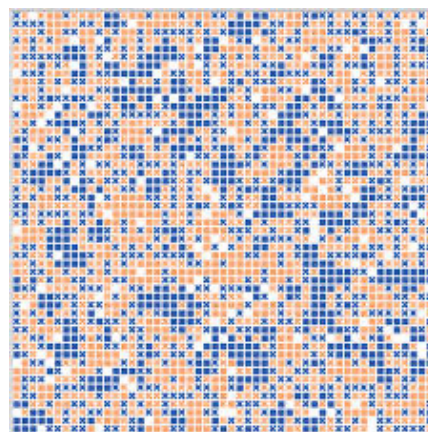


# Bio-inspired algorithms

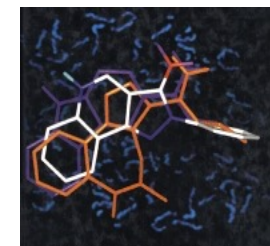
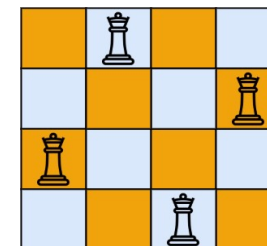
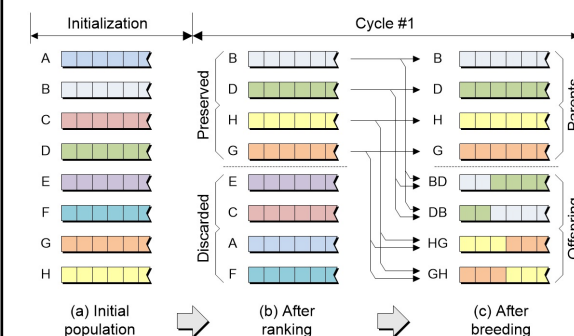
## Neural networks



## Multi-agent systems



## Genetic algorithms

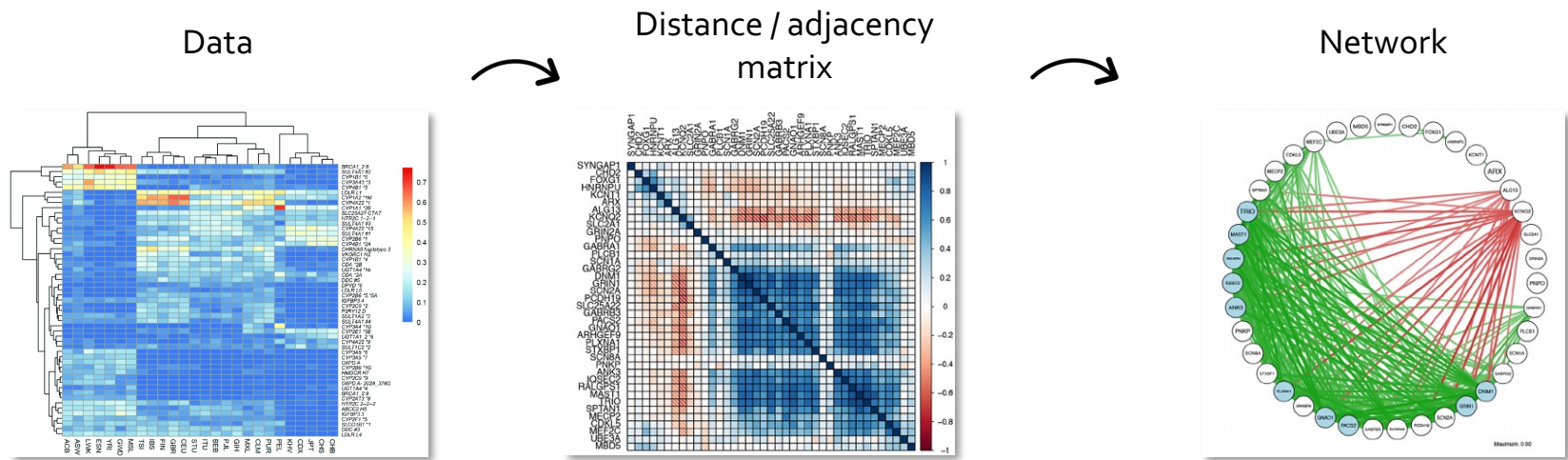


2 4 1 3

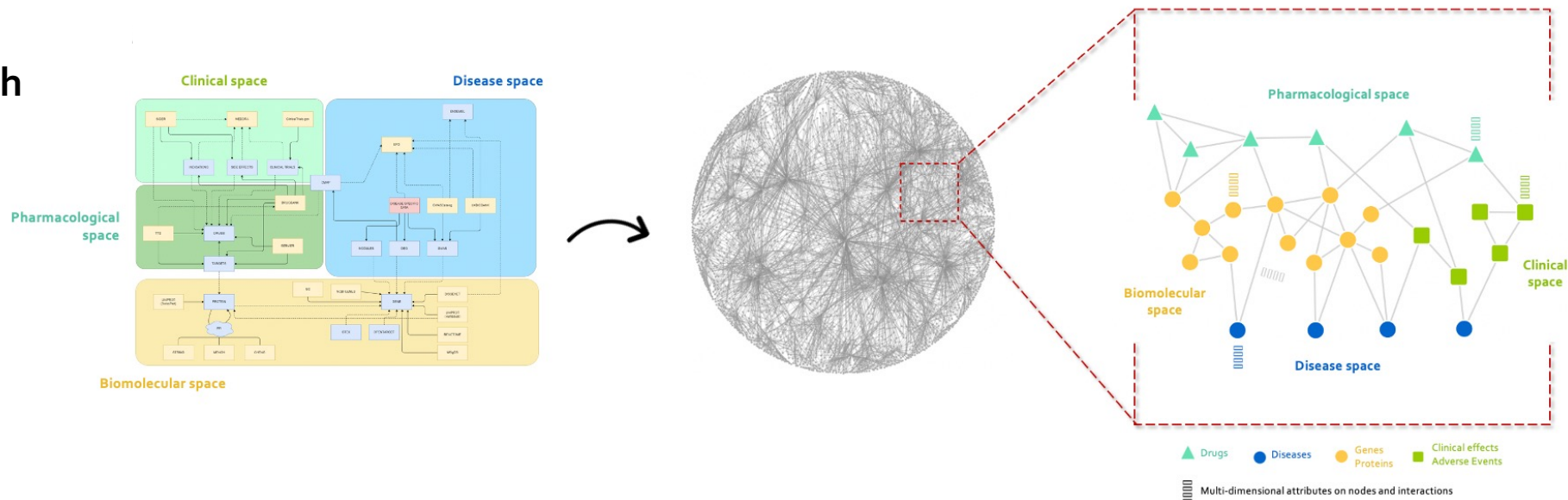


# Network reconstruction

## Network inference



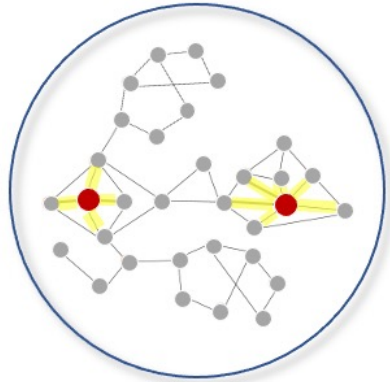
## Knowledge graph



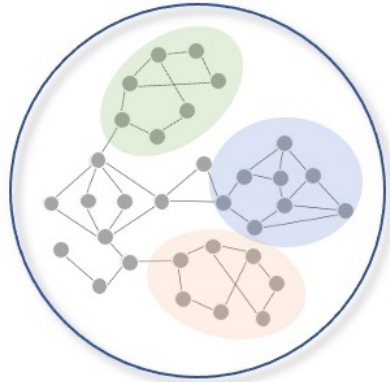


# Network analysis

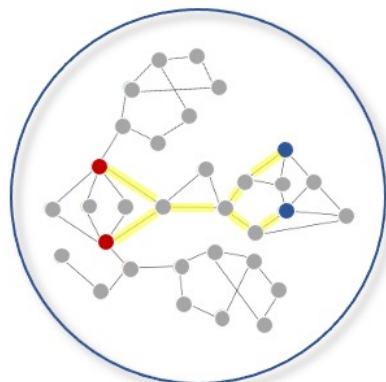
---



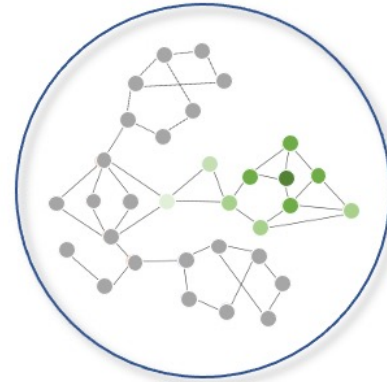
**Hubs**



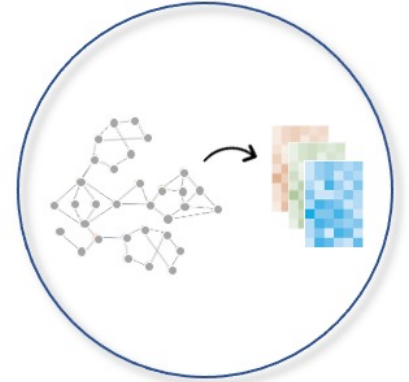
**Clusters**



**Distance**



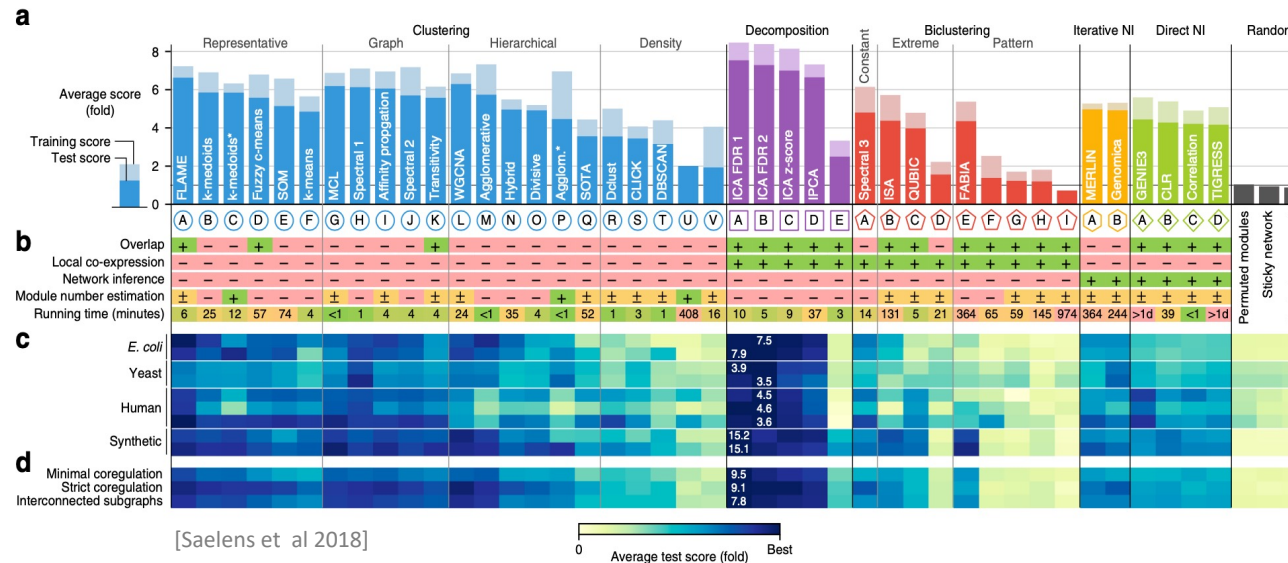
**Diffusion**



**Deep learning**

# Need for benchmarks

## Comparison studies



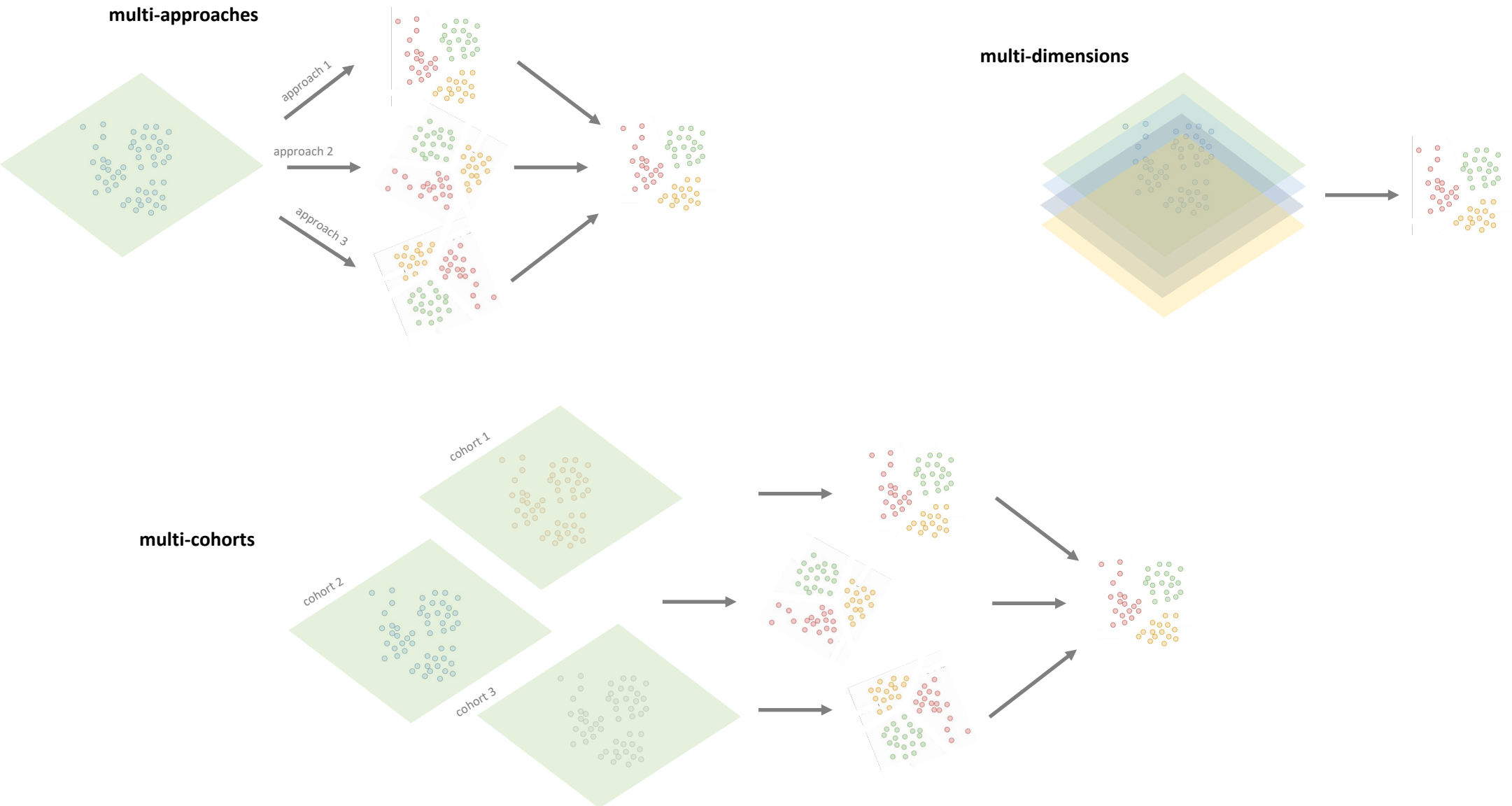
## Data challenges



precisionFDA 

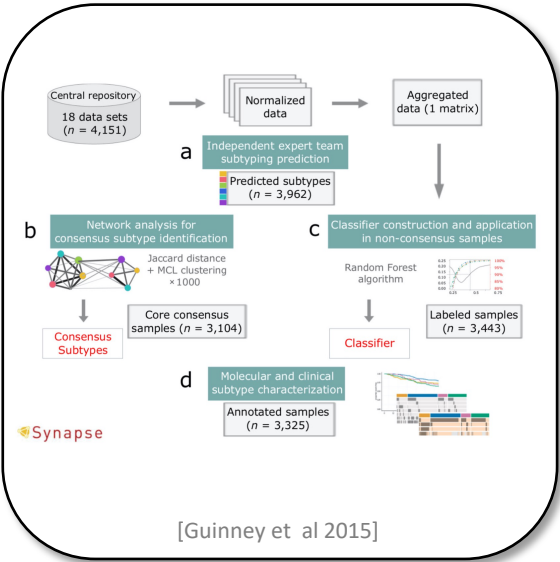
## genedisco challenge

# Integration of multi-sources

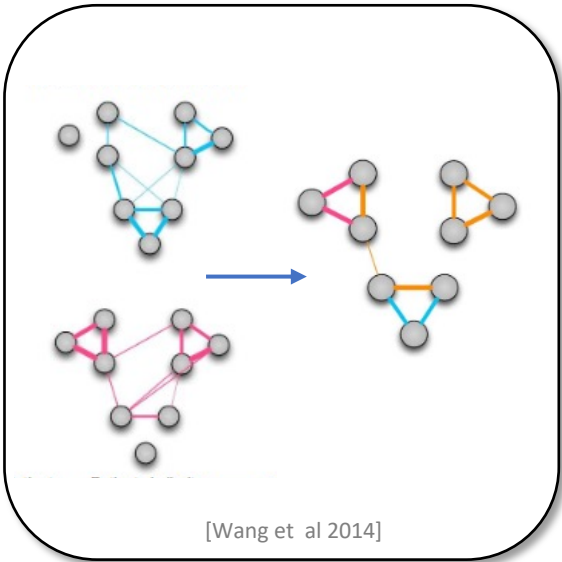


# Integration of multi-sources

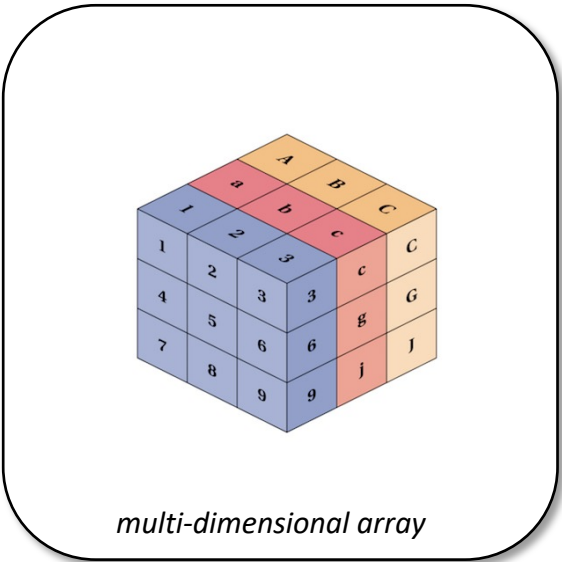
## Consensus



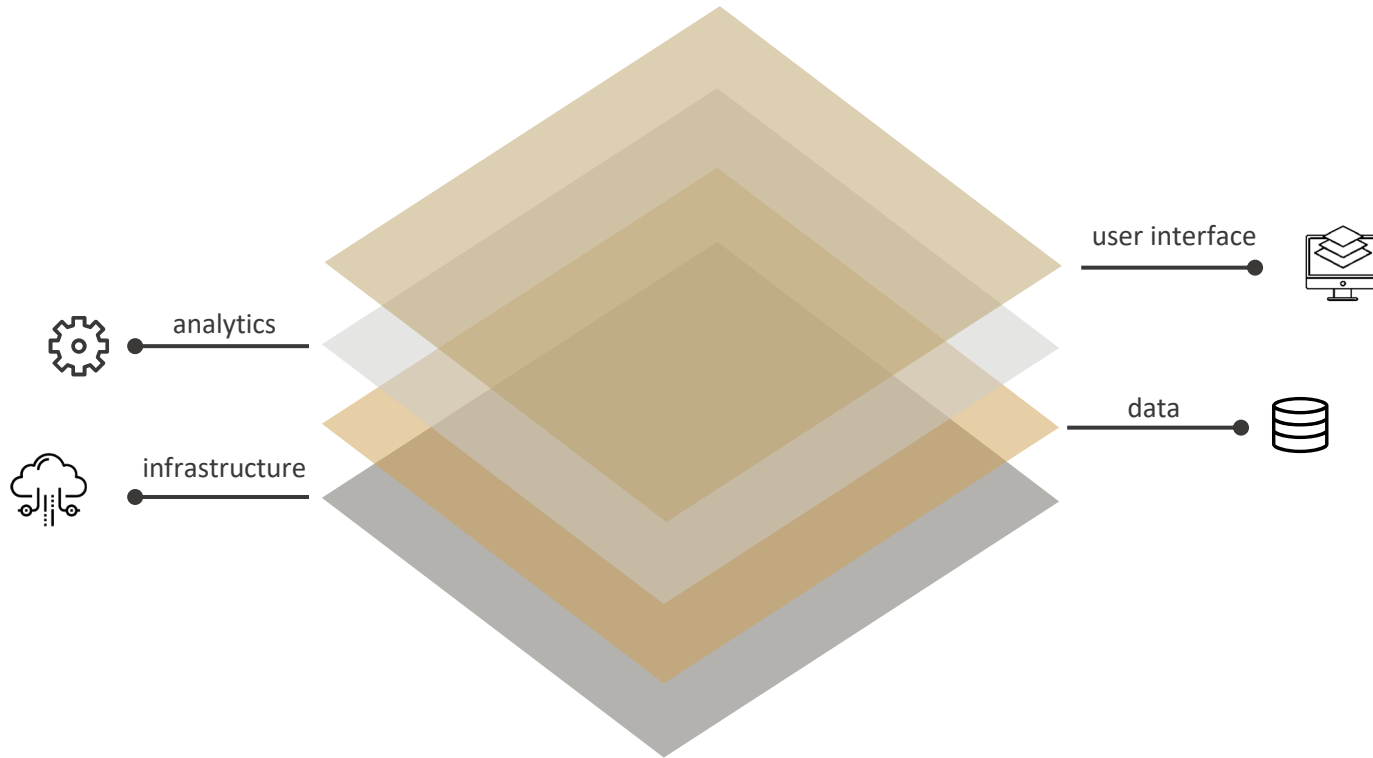
## Data fusion



## Tensors



# All-in-one computing platform



## Programming



## Dev Tools



## Versioning



## Database



## Interactive Data Viz



## Cloud



## Networks

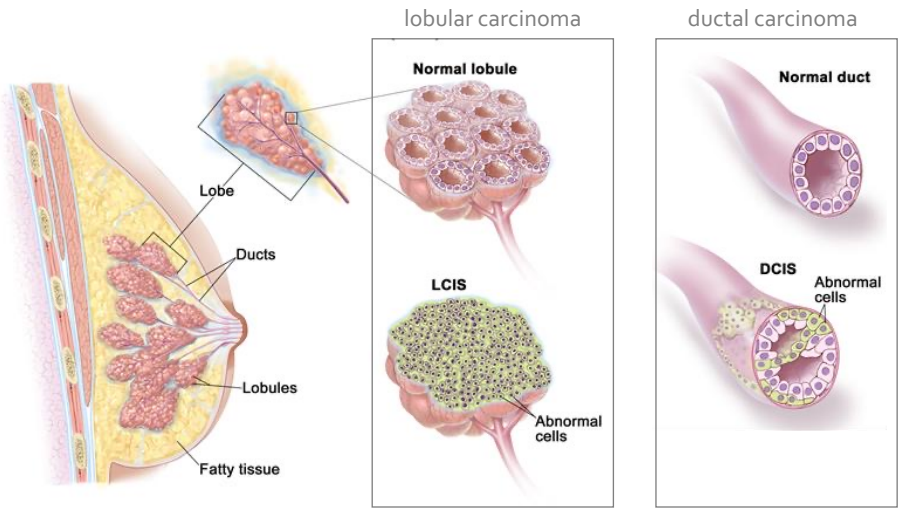


# Applications

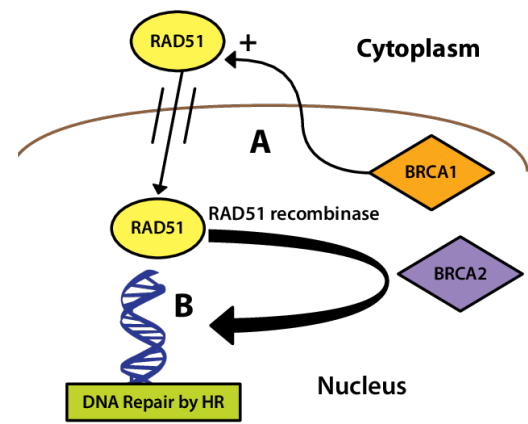


# Molecular classification of breast cancer

## Heterogeneous disease



## Genetic predisposition



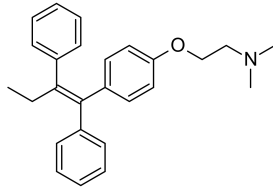
## Standard therapeutic strategies

- Surgery
- Chemo
- Radio

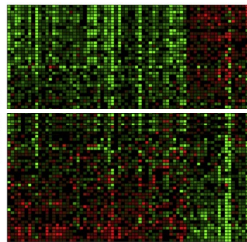
## Pioneer in Computational Precision Medicine



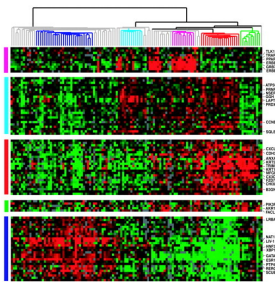
1998: herceptin



1998: tamoxifen



2002 : 1st prognostic signature  
(van 't Veer et al | MammaPrint approved by FDA)



2003: 1st molecular classification  
(Sorlie et al)

# Molecular classification of breast cancer

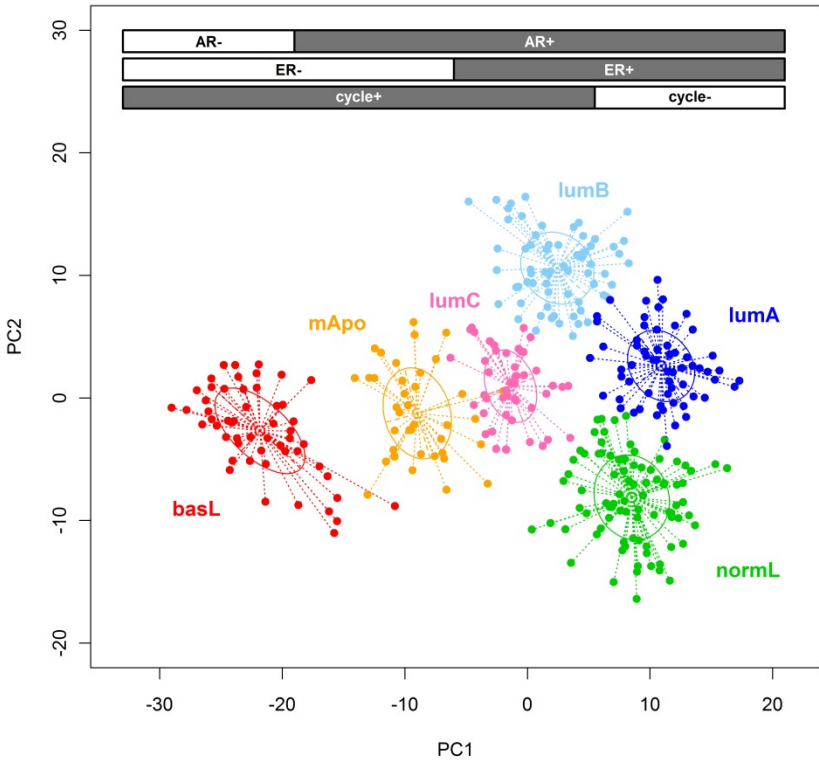
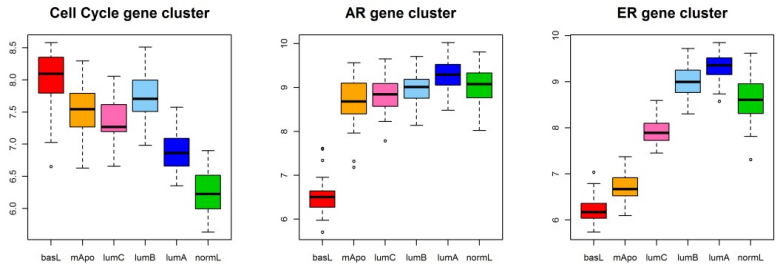
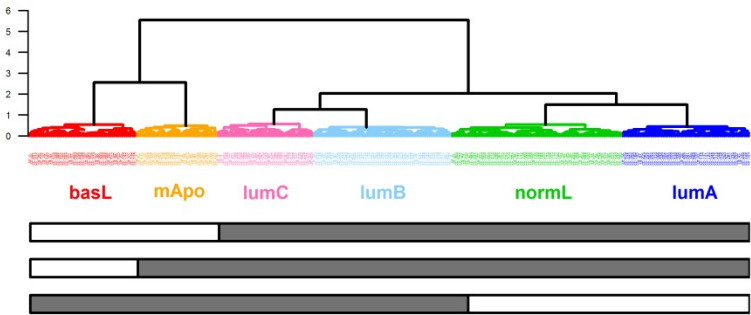
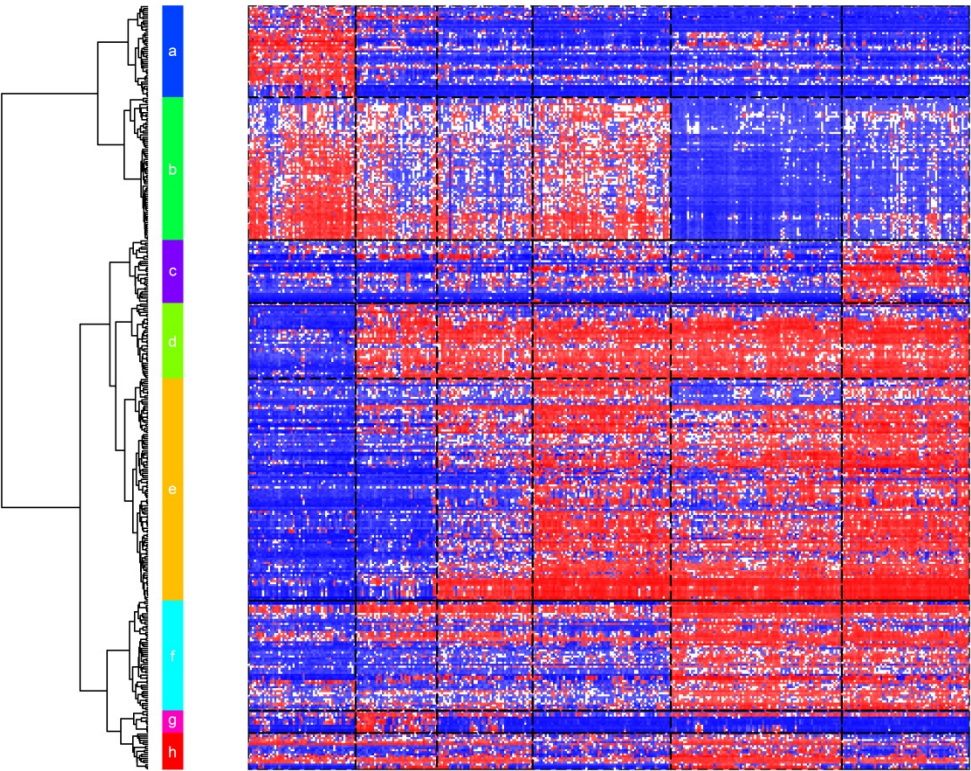
500 patients

Consensus of 3 clustering methods

6 patient clusters

8 gene modules

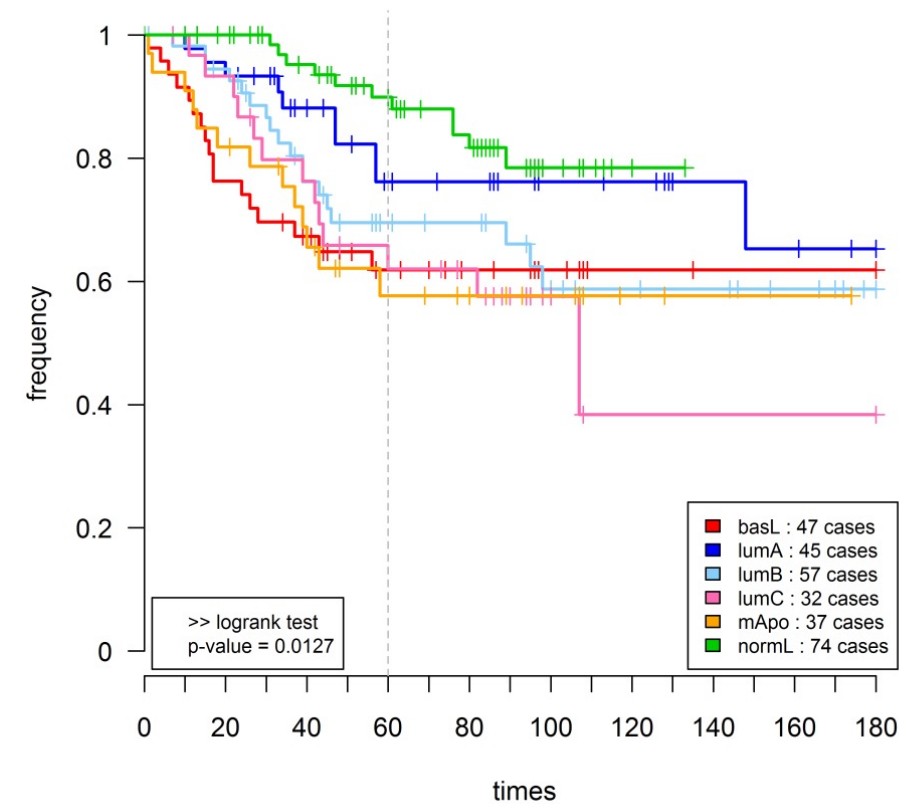
ESR1  
AR  
Cycle





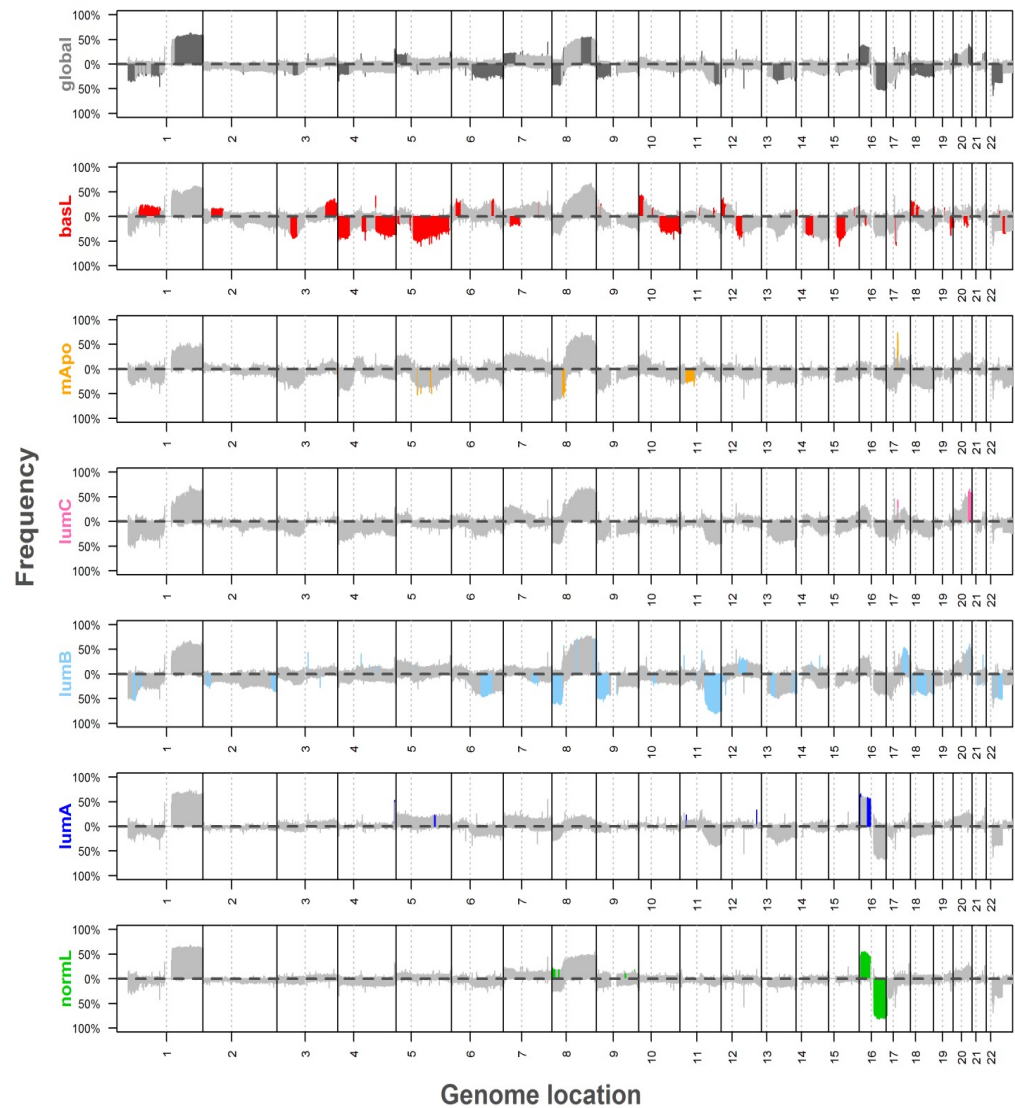
# Molecular classification of breast cancer

CIT classification							
Variable	pv	BasL	mApo	LumC	LumB	LumA	NormL
Total		53	39	48	66	61	88
ER + (IHC)	1.00E-50	5 (10%)	1 (3%)	37 (84%)	63 (98%)	58 (97%)	81 (93%)
ER - (IHC)		46 (90%)	35 (97%)	7 (16%)	1 (2%)	2 (3%)	6 (7%)
ER + (EXP)	6.00E-68	3 (6%)	2 (5%)	48 (100%)	66 (100%)	61 (100%)	87 (99%)
ER - (EXP)		50 (94%)	37 (95%)	0 (0%)	0 (0%)	0 (0%)	1 (1%)
PR + (IHC)	2.00E-25	4 (8%)	1 (3%)	25 (54%)	43 (67%)	53 (88%)	62 (71%)
PR - (IHC)		48 (92%)	34 (97%)	21 (46%)	21 (33%)	7 (12%)	25 (29%)
PR + (EXP)	1.00E-37	5 (9%)	5 (13%)	32 (67%)	47 (71%)	58 (95%)	85 (97%)
PR - (EXP)		48 (91%)	34 (87%)	16 (33%)	19 (29%)	3 (5%)	3 (3%)
ERBB2 + (IHC)	9.00E-19	3 (7%)	19 (68%)	10 (26%)	5 (11%)	0 (0%)	0 (0%)
ERBB2 - (IHC)		43 (93%)	9 (32%)	28 (74%)	41 (89%)	37 (100%)	74 (100%)
ERBB2 + (EXP)	4.00E-31	2 (4%)	29 (74%)	20 (42%)	2 (3%)	0 (0%)	5 (6%)
ERBB2 - (EXP)		51 (96%)	10 (26%)	28 (58%)	64 (97%)	61 (100%)	83 (94%)
AR + (EXP)	2.00E-57	2 (4%)	32 (82%)	47 (98%)	63 (95%)	61 (100%)	88 (100%)
AR - (EXP)		51 (96%)	7 (18%)	1 (2%)	3 (5%)	0 (0%)	0 (0%)
P53mut	1.00E-15	29 (83%)	13 (72%)	24 (69%)	5 (16%)	1 (4%)	1 (5%)
P53wt		6 (17%)	5 (28%)	11 (31%)	27 (84%)	27 (96%)	21 (95%)
Ductal	0.05	51 (98%)	32 (84%)	39 (87%)	54 (84%)	50 (83%)	61 (77%)
Lobular	0.004	1 (2%)	1 (3%)	3 (7%)	3 (5%)	5 (8%)	15 (19%)
Other	0.1	0 (0%)	5 (13%)	3 (7%)	7 (11%)	5 (8%)	3 (4%)
SBR Grade 1	8.00E-11	0 (0%)	0 (0%)	0 (0%)	0 (0%)	7 (12%)	23 (27%)
SBR Grade 2	2.00E-13	6 (11%)	8 (21%)	21 (47%)	38 (58%)	44 (77%)	53 (62%)
SBR Grade 3	4.00E-26	47 (89%)	30 (79%)	24 (53%)	28 (42%)	6 (11%)	9 (11%)
Age (median)	4.00E-07	50	56	54	57	62	52
MR 5year	0.001	17 (36%)	14 (38%)	11 (34%)	15 (26%)	9 (20%)	6 (8%)
MR 15year	0.01	17 (36%)	14 (38%)	13 (41%)	18 (32%)	10 (22%)	11 (15%)
Bones	0.01	4 (24%)	8 (57%)	7 (54%)	14 (78%)	7 (70%)	9 (82%)
Brain	0.06	5 (29%)	3 (21%)	1 (8%)	0 (0%)	0 (0%)	2 (18%)
Liver	0.7	5 (29%)	6 (43%)	7 (54%)	8 (44%)	3 (30%)	3 (27%)
Lung	0.9	6 (35%)	4 (29%)	6 (46%)	8 (44%)	3 (30%)	4 (36%)
Other	0.1	4 (24%)	1 (7%)	7 (54%)	8 (44%)	3 (30%)	3 (27%)



# Molecular classification of breast cancer

## Genomic alterations

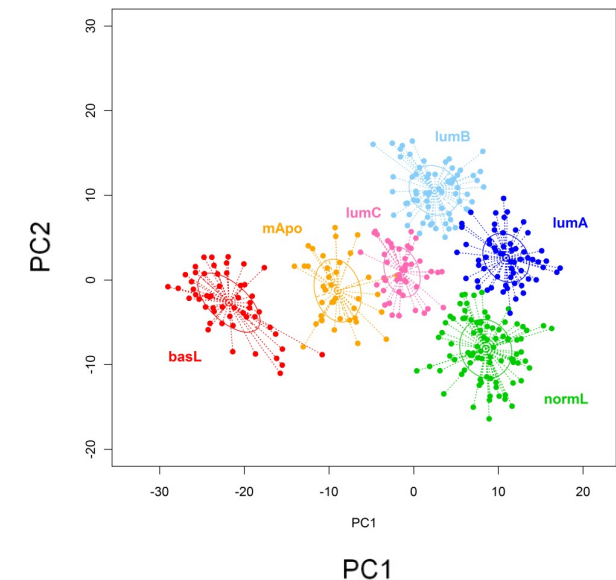


Category	Pathways	Bas-L	m-Apo	Lum-C	Lum-B	Lum-A	Norm-L
Cell communication	Adherens junction						
	Focal adhesion						
Motility	cell motility						
	Regulation of actin cytoskeleton						
Cell growth and death	Apoptosis						
	Cell cycle						
Replication and repair	p53 signaling pathway						
	Base excision repair						
	DNA replication						
	Mismatch repair						
Lipid metabolism	Nucleotide excision repair						
	Androgen and estrogen metabolism						
Endocrine system	Fatty acid metabolism						
	GnRH signaling pathway						
Signal transduction	Insulin signaling pathway						
	Renin-angiotensin system						
	androgen receptor signaling						
	Calcium signaling pathway						
	ErbB signaling pathway						
	estrogen receptor signaling						
	mTOR signaling pathway						
	Phosphatidylinositol signaling						
	PTEN cell cycle arrest and apoptosis						
	TGF-beta signaling						
Immune system	Wnt signaling pathway						
	Antigen processing and presentation						
	B cell receptor signaling						
	Hematopoietic cell lineage						
	Natural killer cell mediated cytotoxicity						
	T cell receptor signaling						
	Toll-like receptor signaling						

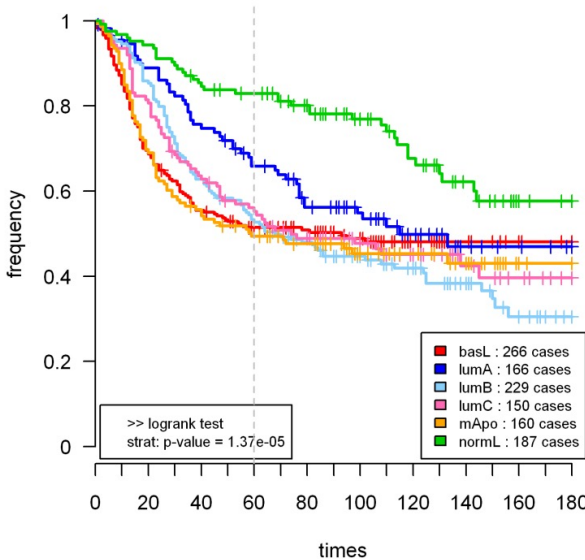
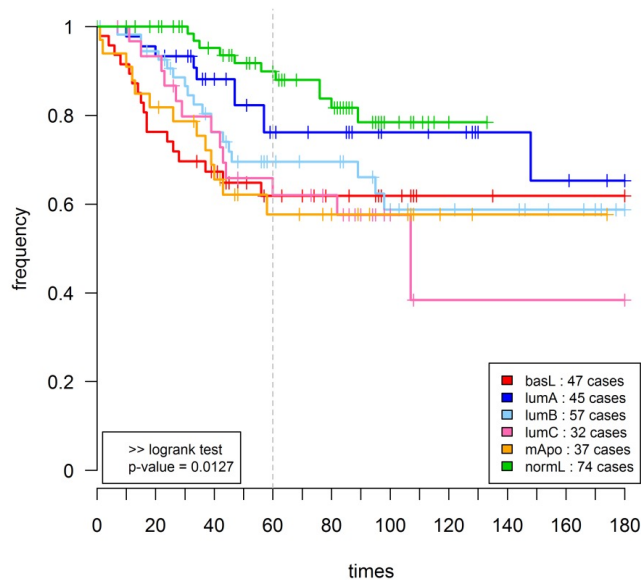
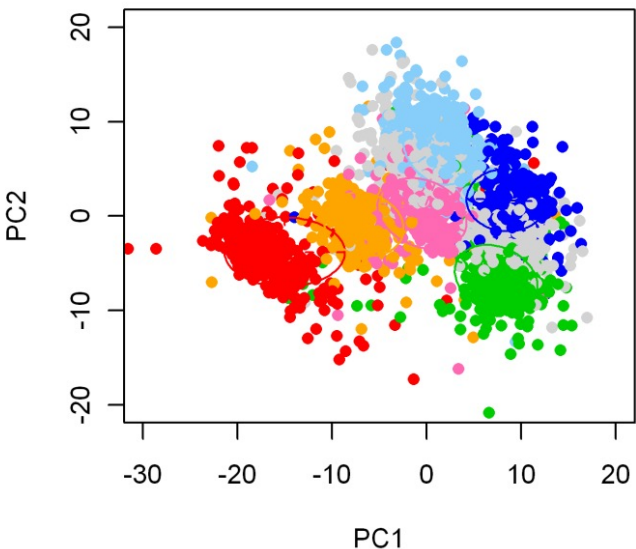
Pathways associations

# Molecular classification of breast cancer

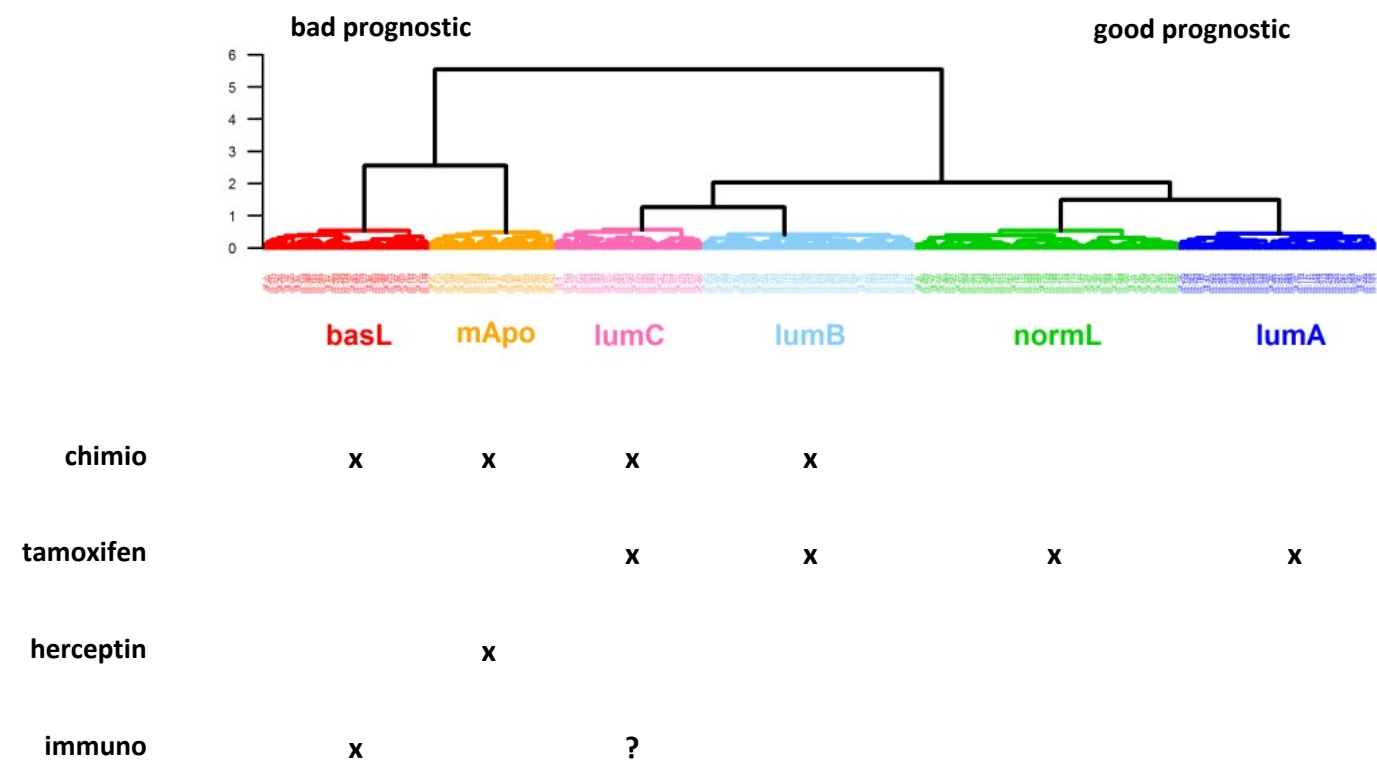
Discovery set  
500 samples



Validation set  
3000 public samples



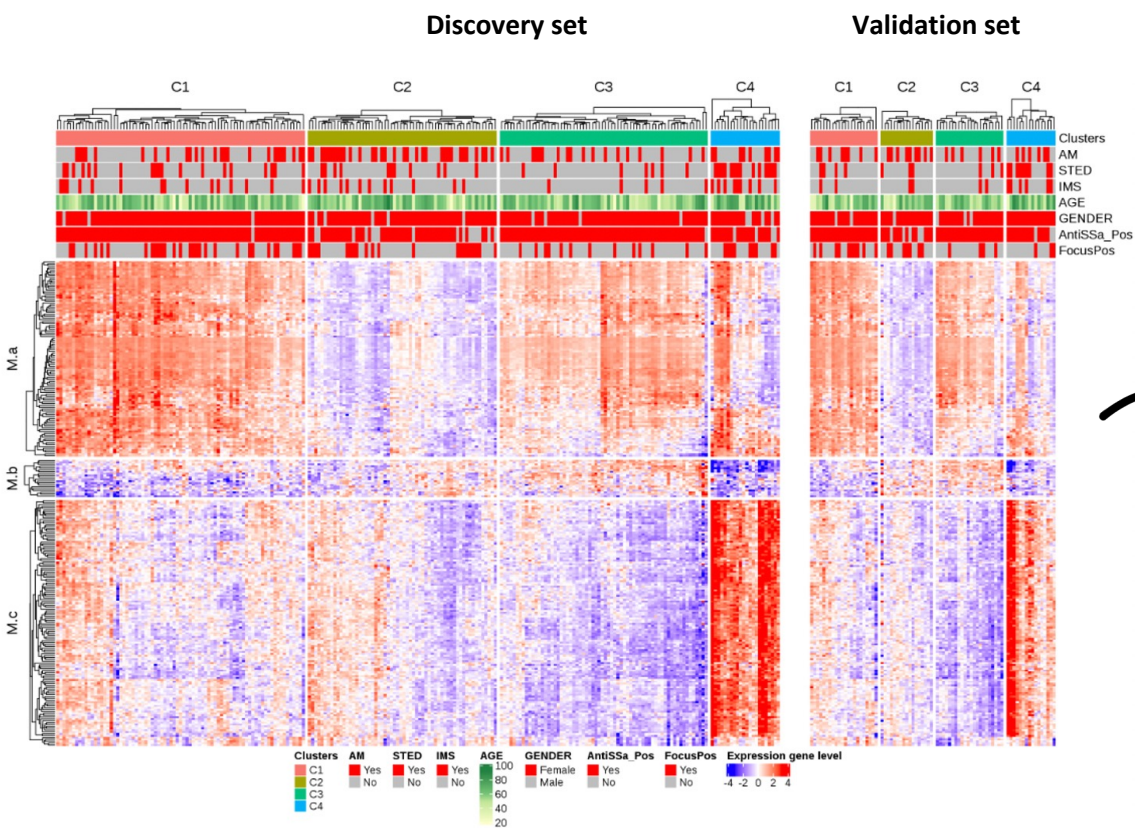
# Molecular classification of breast cancer





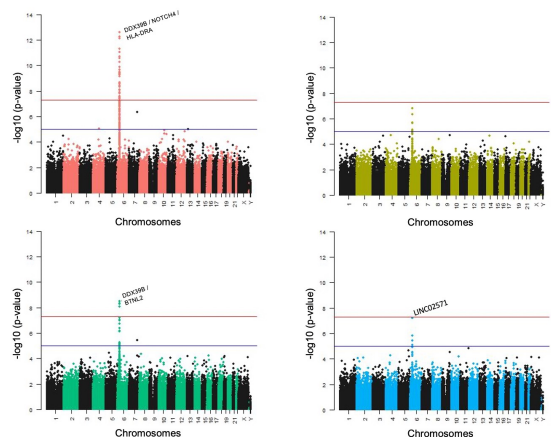
# Molecular classification of Sjögren

341 RNA-seq on whole blood samples



## Multi-omics characterization

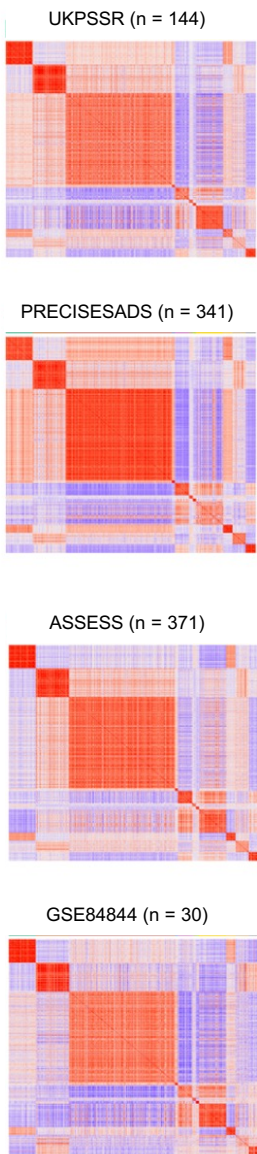
- Canonical pathways & immunologic modules analysis
- Methyloomics
- Flow cytometry / deconvolution
- GWAS



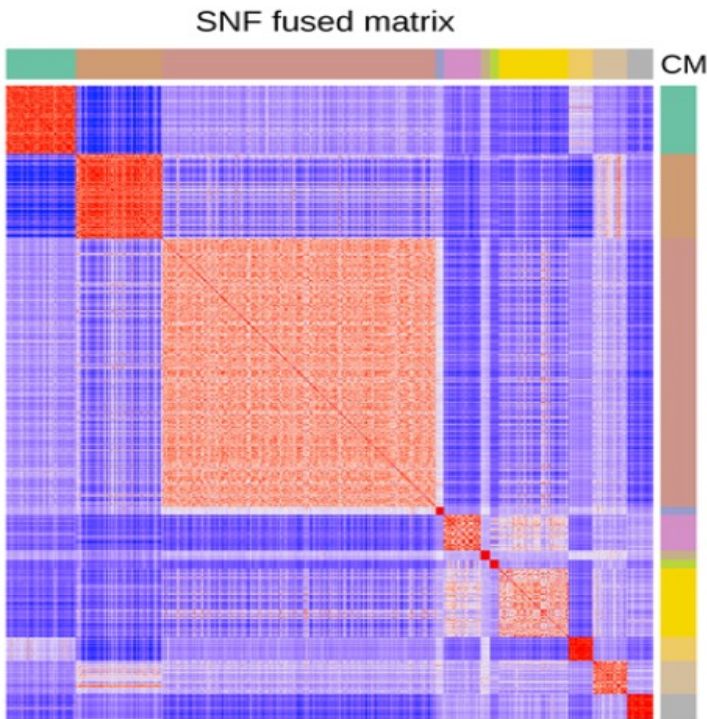
Target identification,  
drug repurposing,  
drug combinations



# Molecular classification of Sjögren

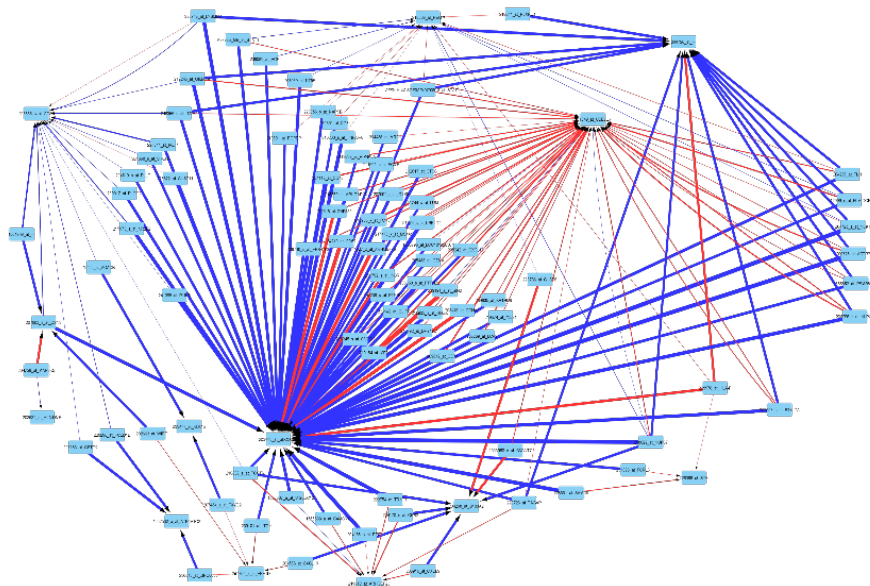


Matrix fusion

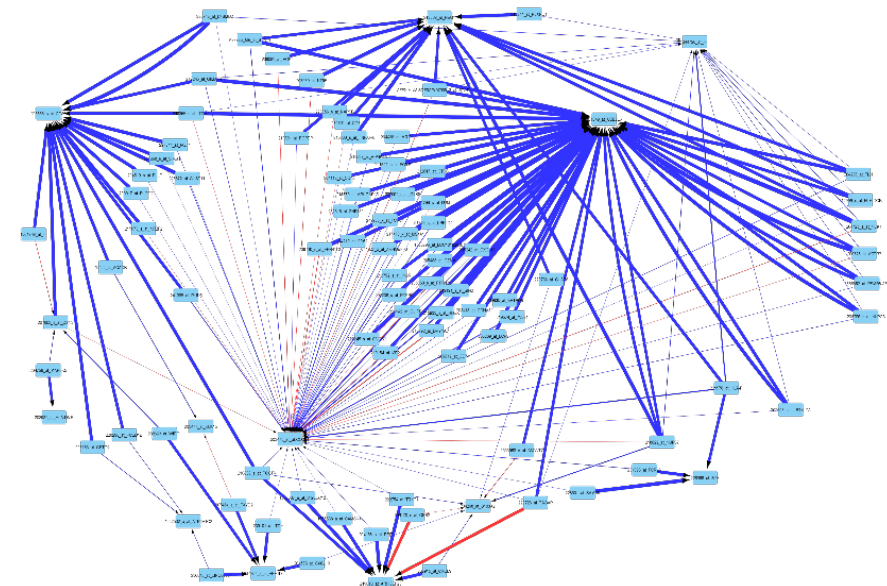


13 gene modules

# Molecular classification of Sjögren



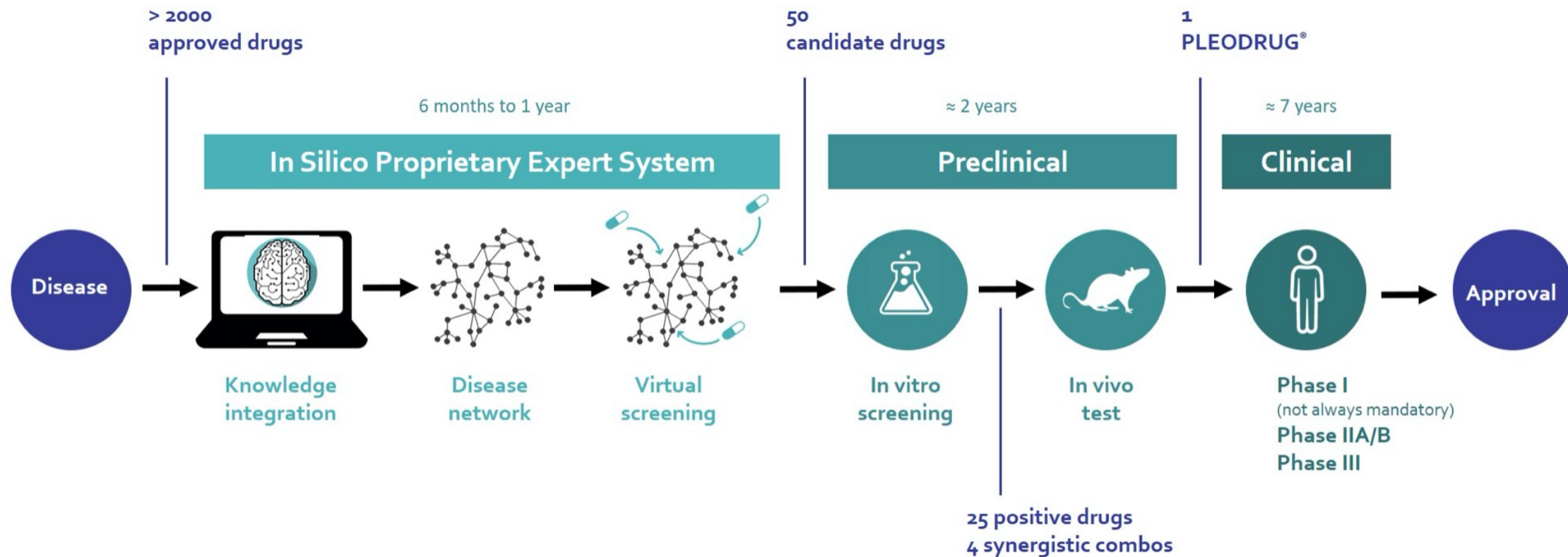
Sjögren

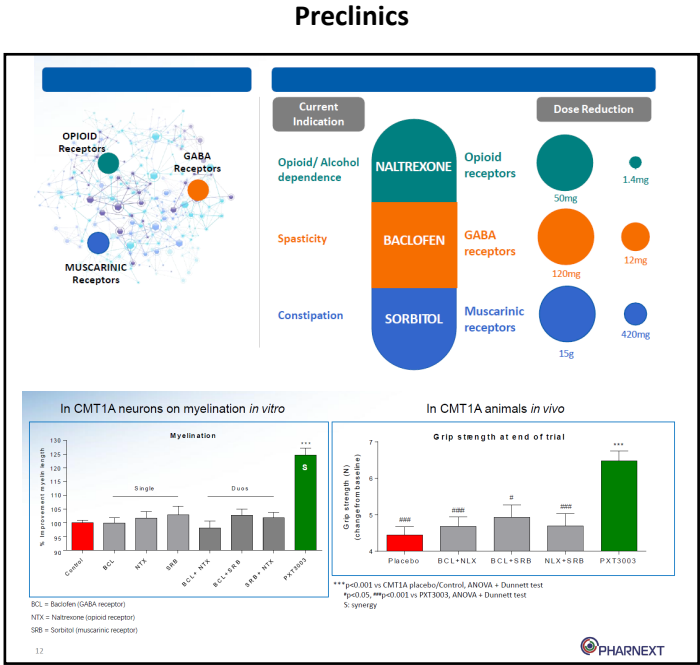
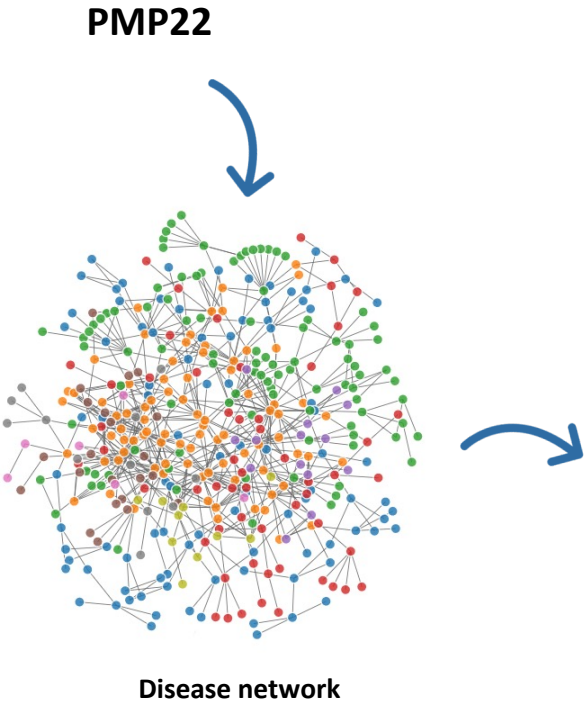


Control



# Combination of repurposed drugs

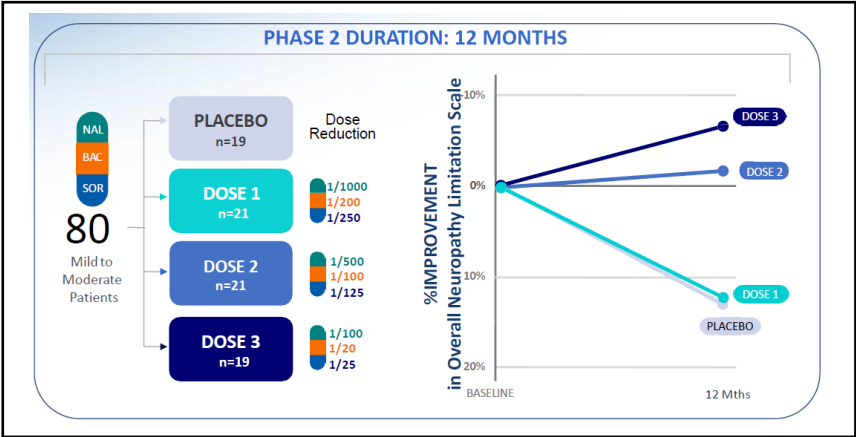




[Chumakov et al 2014]

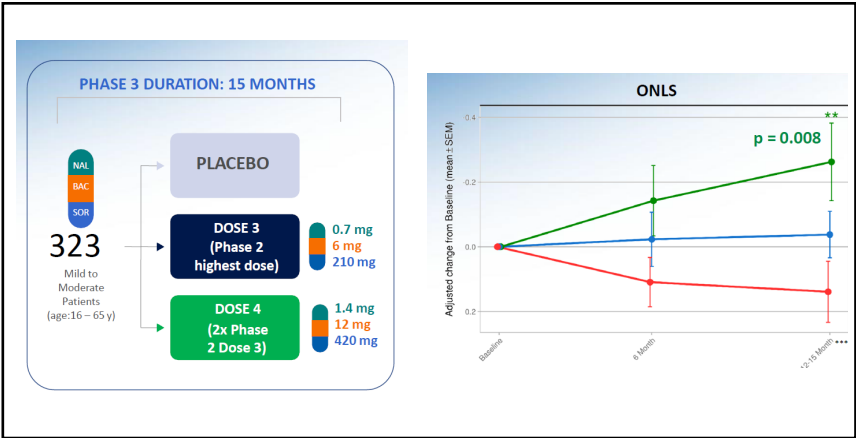
No phase 1

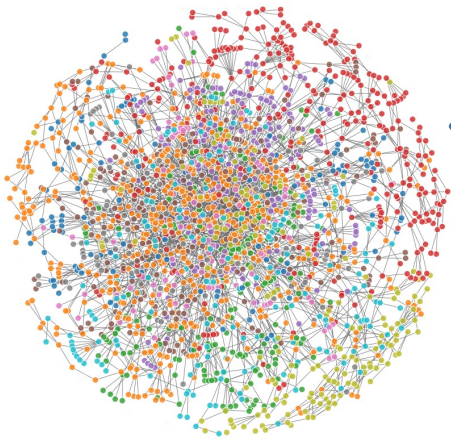
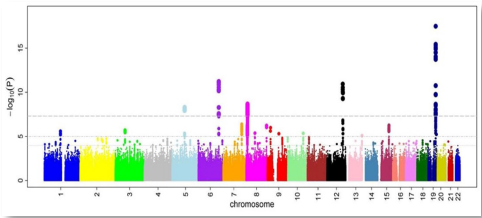
## Direct to Phase 2



[Attarian et al 2014]

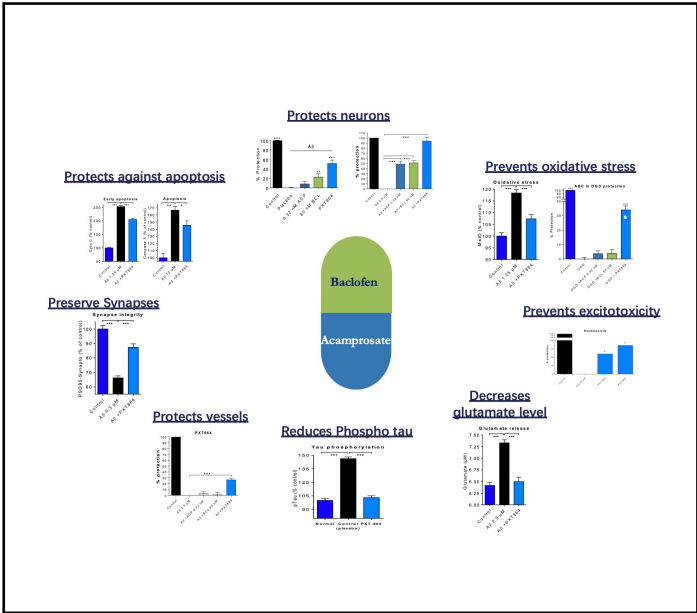
## Phase 3





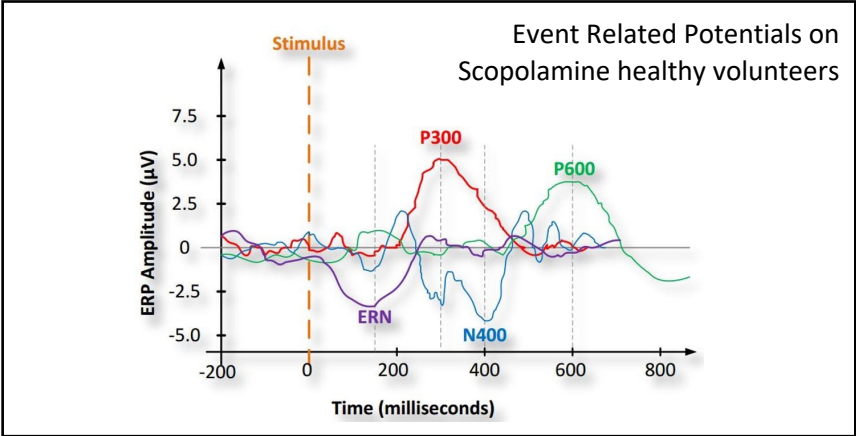
Disease network

### Preclinics

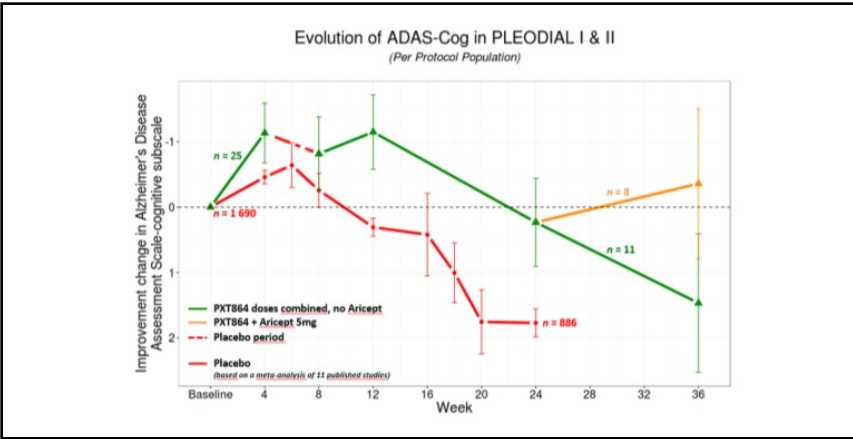


[Chumakov et al 2015]  
[Nabirovichin et al, in revision]

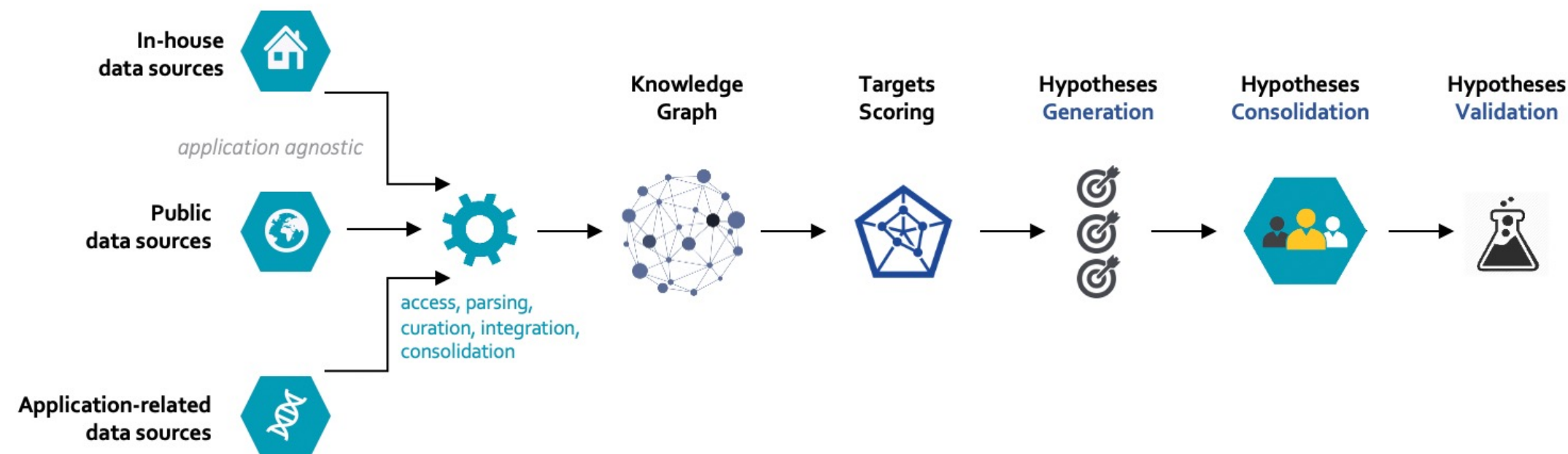
### Phase 1



### Phase 2



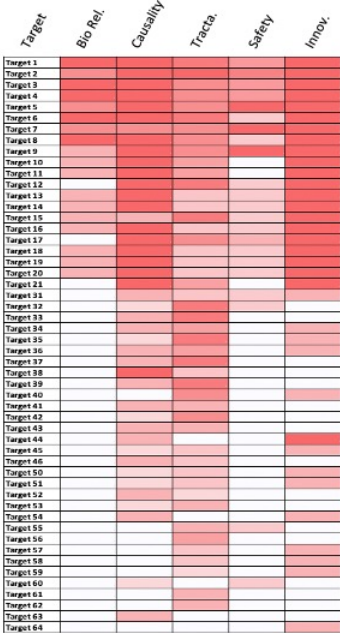




A - Target assessment

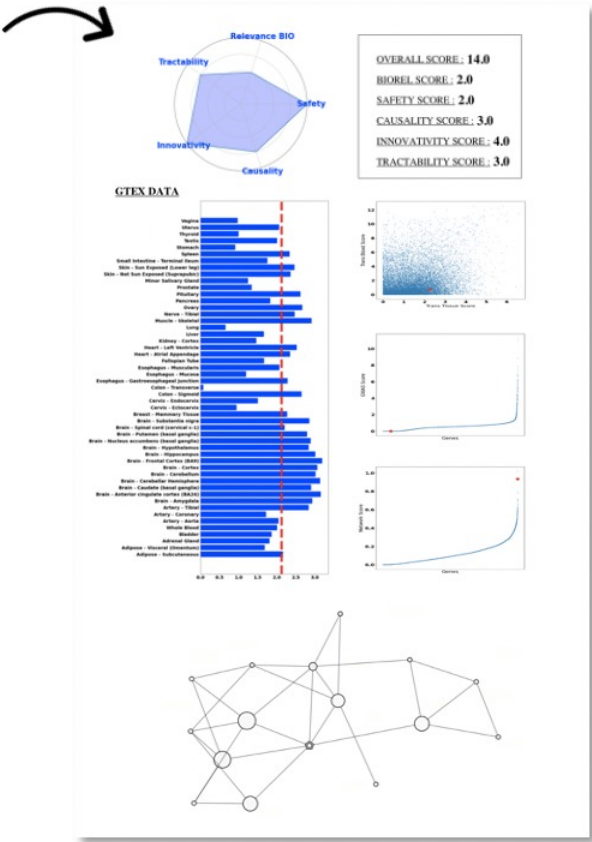


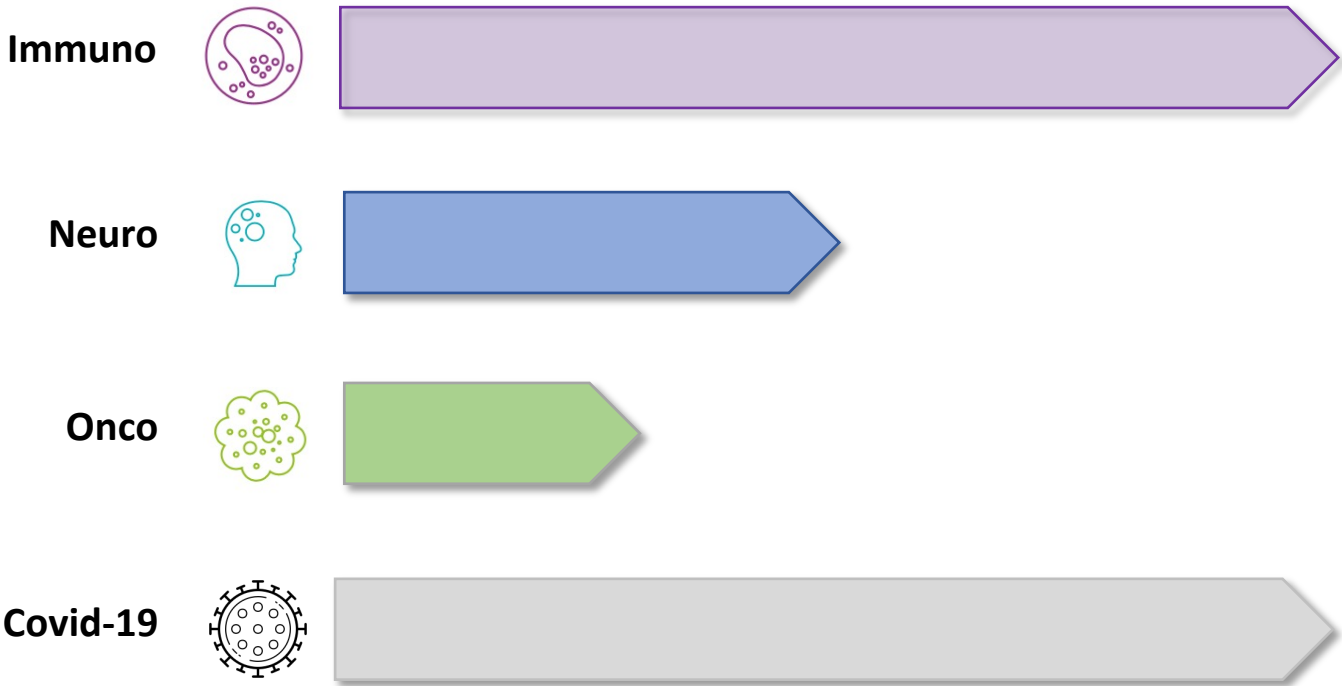
B - Target prioritization



low interest

C - Target ID cards





FINALISTE

## SERVIER Patrimony

R&D : l'IA  
au service de  
la recherche  
médicamentuse



La recherche médicamenteuse et le développement de nouvelles molécules est un processus long et coûteux. Pour y remédier, l'idée de Patrimony est de **capitaliser sur le patrimoine de données « dormantes »** de Servier et de faire parler ces données pour **générer grâce au machine learning de nouvelles hypothèses thérapeutiques**. Un projet jusqu'à lors irréalisable en partie en raison de contraintes informatiques et juridiques fortes, résolues grâce à la mise en place d'un **environnement cloud sur mesure**.

Aujourd'hui, Patrimony intègre une trentaine de sources de données internes et externes et permet d'étudier plus de 200 000 interactions.

À la clef, un gain de temps évident pour les biologistes et une pertinence des hypothèses générées garantie. L'intérêt de cette approche computationnelle est aussi de pouvoir **couvrir un plus grand nombre de pathologies**, en se penchant sur des pathologies rares à moindre coût, et de **répondre plus rapidement à des événements soudains** comme des épidémies. Car Patrimony cherche aussi à systématiser une démarche qui autrefois relevait du hasard ou de l'intuition heureuse d'un biologiste : le **repositionnement de médicaments**. Cette **démarche de « re-use »** présente l'énorme avantage de permettre une mise à disposition plus rapide auprès des patients puisque la molécule est déjà connue et une partie des essais cliniques déjà réalisée.

Patrimony est un projet dont l'industrialisation – du POC jusqu'à l'intégration dans les opérations « normales » de l'entreprise – a été pensée sur quatre ans. Une réelle ambition donc de **placer la médecine computationnelle au cœur des métiers de Servier** !

Prix Change 2020

Review

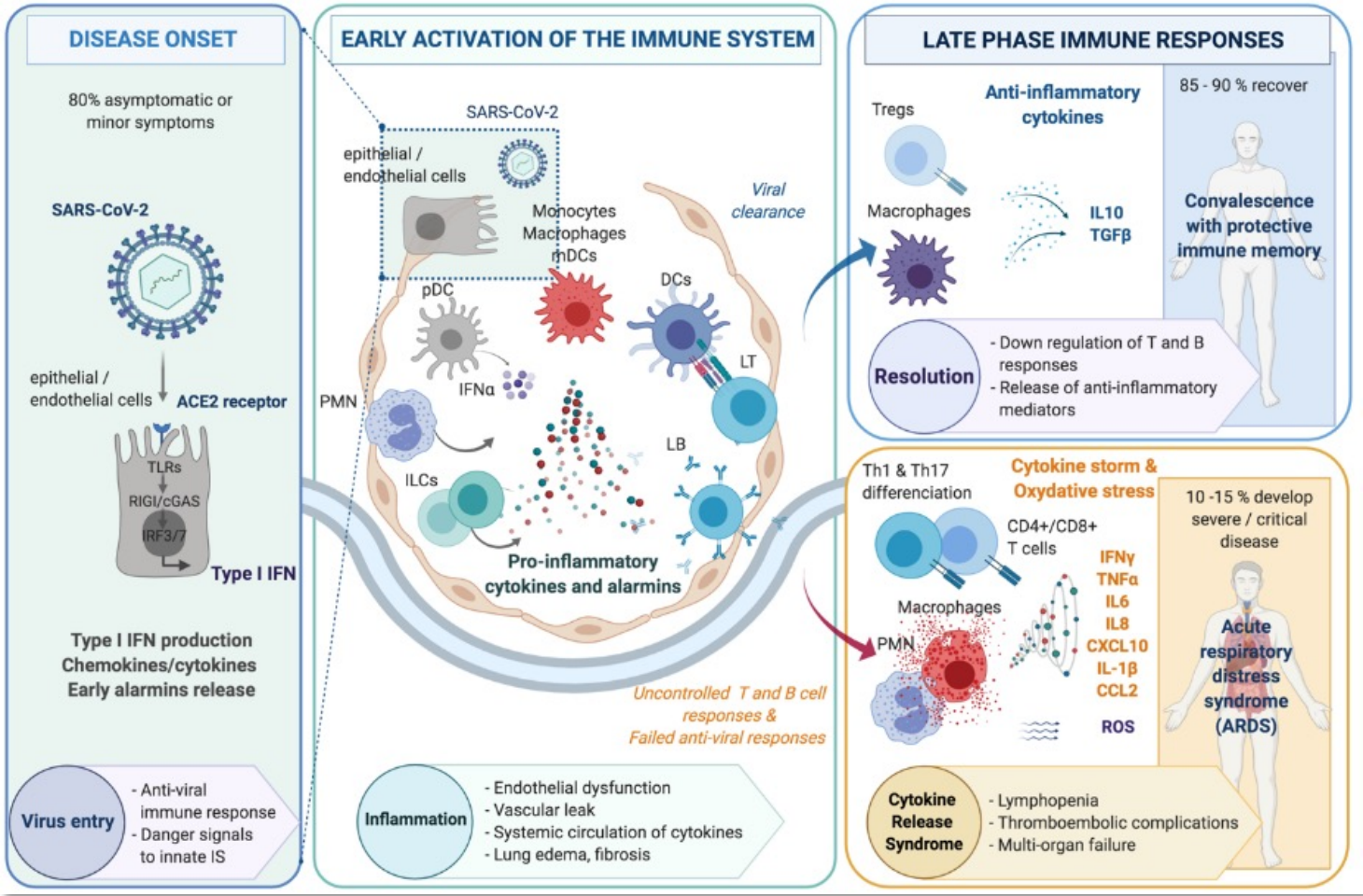
Artificial intelligence in COVID-19 drug repurposing

Yudi Zhou\*, Fei Wang\*, Jian Tang\*, Ruth Nussinov, Feibao Cheng

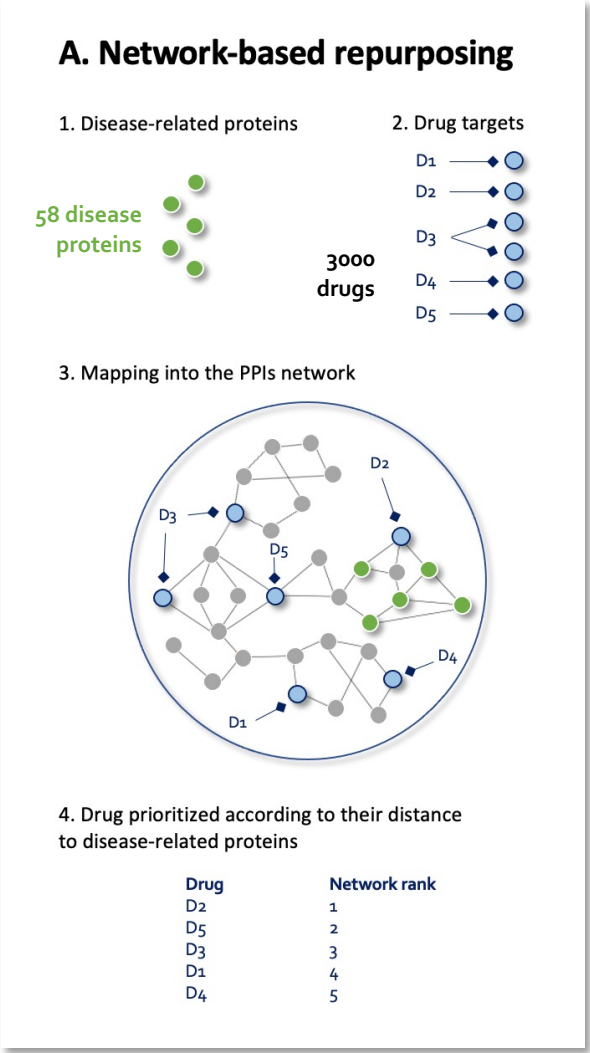
Drug repurposing or repositioning is a technique whereby existing drugs are used to treat emerging and challenging diseases, including COVID-19. Drug repurposing has become a promising approach because of the opportunity for reduced development timelines and overall costs. In the big data era, artificial intelligence (AI) and network medicine offer cutting-edge application of information science to defining disease, medicine, therapeutics, and identifying targets with the least error. In this Review, we introduce guidelines on how to use AI for accelerating drug repurposing or repositioning, for which AI approaches are not just formidable but are also necessary. We discuss how to use AI models in precision medicine, and as an example, how AI models can accelerate COVID-19 drug repurposing. Rapidly developing, powerful, and innovative AI and network medicine technologies can expedite therapeutic development. This Review provides a strong rationale for using AI-based assistive tools for drug repurposing medications for human disease, including during the COVID-19 pandemic.

Least Digital Health 2020; 4:460-76  
Published Online September 18, 2020  
http://dx.doi.org/10.1016/j.lsdh.2020.09.002  
© 2020 The Author(s)  
General Medicine Institute, Cleveland Clinic, Cleveland, OH, USA  
(Y. Zhou PhD, Prof F. Cheng PhD);  
F. Cheng, MD, PhD

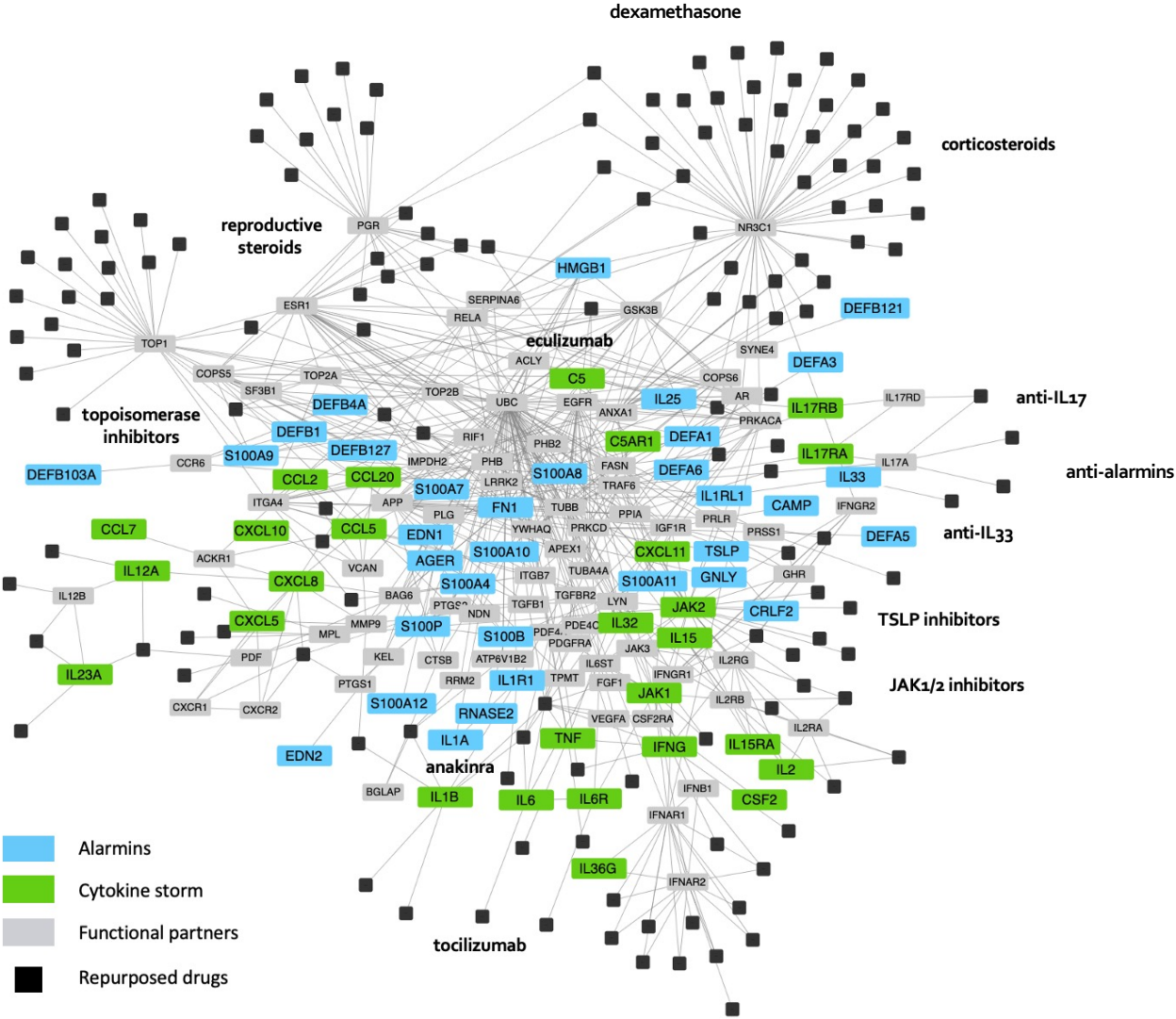
Benevolent<sup>AI</sup> >> baricitinib







1000 known drug targets



## **Concluding remarks**



# Expected impact

## FROM THE ANALYST'S COUCH

### AI in small-molecule drug discovery: a coming wave?

Madura K. P. Jayatunga, Wen Xie, Ludwig Ruder, Ulrik Schulze and Christoph Meier

Artificial intelligence (AI) offers the potential to transform drug discovery. Over the last few years, AI-enabled drug discovery has grown substantially through technological progress, such as the use of neural networks to design molecules and the application of knowledge graphs to understand target biology.

Several AI-native drug discovery companies have progressed molecules into clinical trials, in some cases reporting greatly accelerated timelines and reduced costs, raising high expectations in the R&D community. In addition, many established pharmaceutical companies have formed discovery partnerships with AI companies to explore the technology. Despite this progress, it is still early days for AI in drug discovery, with many open questions about its impact and future potential.

We see several dimensions for AI to create value in drug discovery, including greater productivity (faster speed and/or lower cost), broader molecular diversity and improved chances of clinical success. Here, we present an analysis of the impact of AI along these dimensions using publicly available data. We focused mainly on small-molecule drug discovery, for which AI approaches are relatively more established.

#### Impact in small-molecule drug discovery

**Pipeline growth.** We focused our analysis on 24 'AI-native' drug discovery companies, for which AI is central to their discovery strategy (see Supplementary information for a list and analysis strategy). For a subset of 20 of these companies, we were able to reconstruct their pipelines between 2010 and 2021 using public databases. During this time, AI drug discovery companies had rapid pipeline growth, with an average annual growth rate of around 36%. This is driven mainly by assets and programmes at the discovery and preclinical stage (FIG. 1a), reflecting the early-stage nature of AI-native companies. Today the combined pipeline of these 20 AI companies contains ~160 disclosed discovery programmes and preclinical assets and about 15 assets in clinical development.

For comparison, the combined in-house-originated pipeline of the top 20 pharma

companies contains ~330 disclosed discovery programmes and preclinical assets, and ~430 assets in phase I clinical development (using the same public data sources and excluding partnered assets or programmes; FIG. 1b). So, AI companies appear to have a combined pipeline equivalent to 50% of the in-house discovery and preclinical output of 'big pharma'. Even if we assume under-reporting of discovery programmes and preclinical assets by pharma companies and over-reporting by AI companies, this seems an impressive picture. Nevertheless, it remains to be seen how many of the AI-enabled preclinical programmes reach the clinical trial stage, and how successful AI-derived assets will be in clinical trials.

**Pipeline composition of AI drug discovery companies.** We further analysed the current pipelines of the full list of 24 AI-native drug discovery companies with regards to therapeutic areas and target classes. Detailed target information was available for only about a quarter of AI-enabled R&D programmes and assets, but analysis of this partial dataset suggests that AI-native drug discovery companies often focus on well-established target classes (FIG. 2a). For example, more than 60% of all disclosed targets of AI companies are enzymes such as kinases, and other well-known drug target classes such as

## NEWS & ANALYSIS



Credit: Ayman-Alakhras/Stock/Getty Images Plus

G-protein-coupled receptors also make up a high proportion.

This strong emphasis on well-established targets as appropriate testing grounds could be driven by multiple factors, including a desire to de-risk internal pipelines by focusing on targets with validated biology, to prove the viability of their technology platforms and to address well-known challenges such as selectivity issues for well-characterized targets with rich data (often including structural information). In contrast, top-20 pharma companies tend to have pipelines that balance both emerging and established target classes (FIG. 2a).

Despite these trends, there are some reported examples of potential first-in-class AI-derived compounds for novel targets, including protein tyrosine phosphatase SHP2, DNA helicase WRN and paracaspase MALT1, for which AI-derived compounds are among the first for which first-in-human studies or studies to enable an investigational new drug (IND) application have been initiated (see Supplementary information for details).

In terms of therapy area, most of the disclosed AI discovery programmes and assets are in the oncology and central nervous system areas, probably due to the high unmet medical need and many well-characterized targets (FIG. 2b).

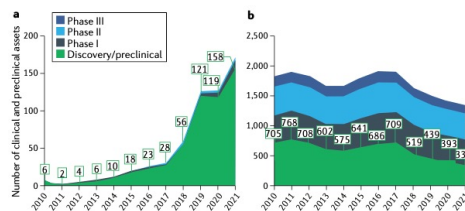
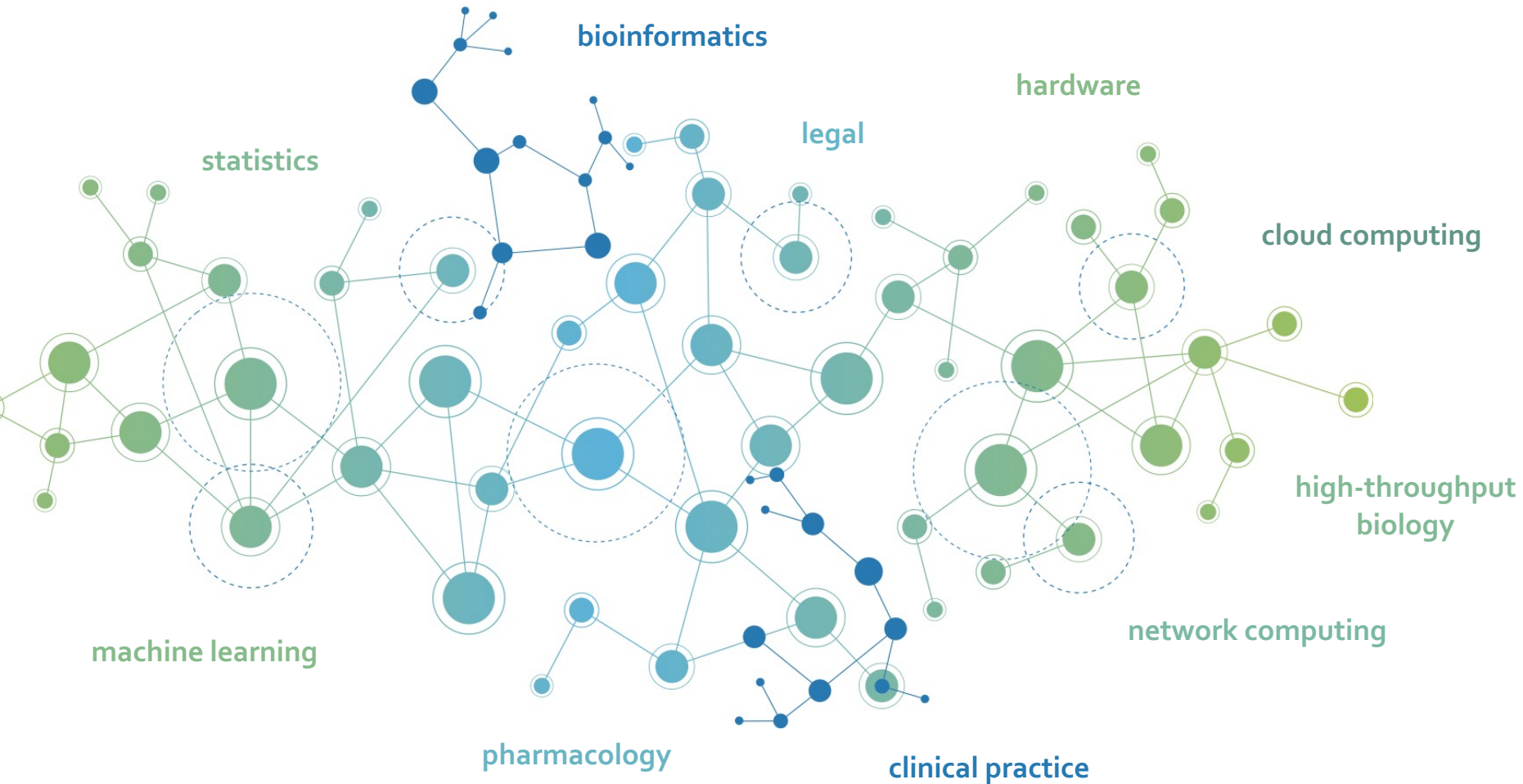


Fig. 1 | Number of annual R&D programmes and assets over time, showing the growth of AI-enabled drug discovery. a | AI-native drug discovery companies. b | For comparison, top-20 pharma companies. See Supplementary information for details.

- faster speed
- lower cost
- broader molecular diversity
- less animals in preclinical studies
- less patients in clinical studies
- improved chances of clinical success

# Need for transversality & interface

---



# What is a good model

---

*« All models are wrong but some are useful. »*

G. Box

simplicity

interpretability

robustness

reproducibility

# Acknowledgement

---

ALL COLLABORATORS FROM

*GENOPOLE*

*MERCK-SERONO*

*LIGUE CONTRE LE CANCER*

*PHARNEXT*

*SERVIER*

*NANOBIOTIX*

*INSA*

*ENSAI*

*STATOMIQUE*

**Thank you !**