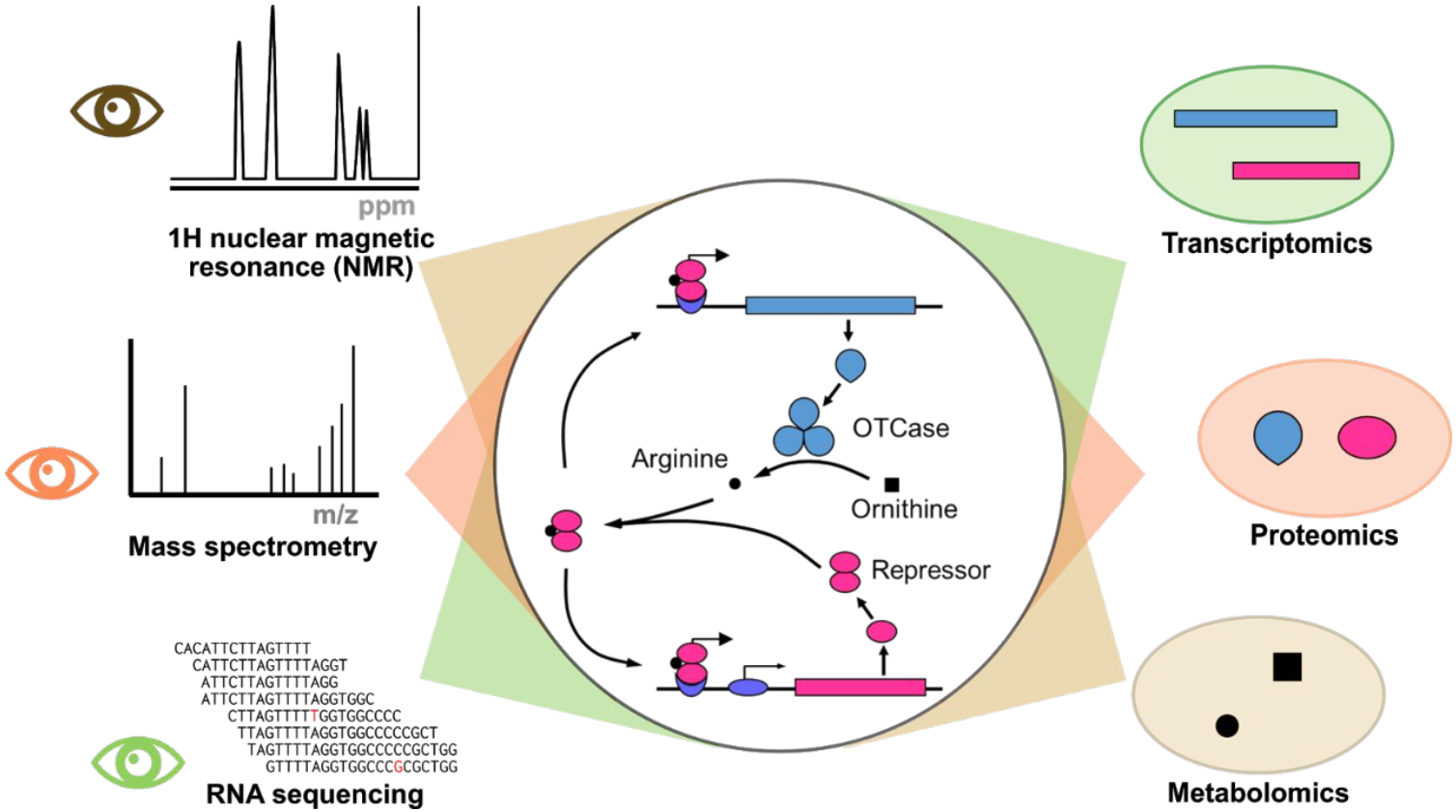
AMIDD Lecture 8: Omics- and cellular-level models



Omics data are projections of high-dimensional biological space. It is an *inverse problem* to infer a high-dimensional space from its projections.

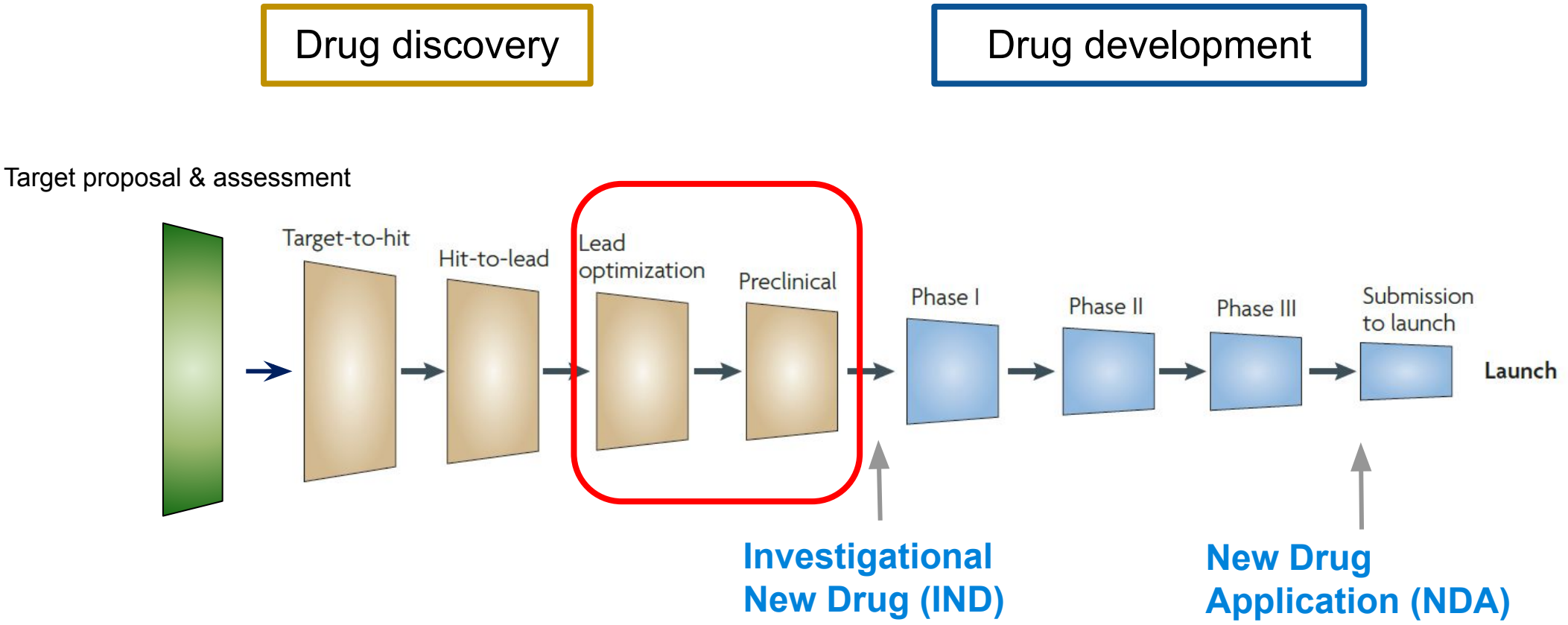
Multiscale Modelling of Drug Mechanism and Safety by Zhang, Sach-Peltason, Kramer, Wang and Ebeling, Drug Discovery Today, 2020

Dr. Jitao David Zhang, Computational Biologist

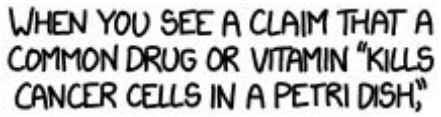
[1] Pharmaceutical Sciences, Pharma Research and Early Development, Roche Innovation Center Basel, F. Hoffmann-La Roche

[2] Department of Mathematics and Informatics, University of Basel

Omics and cellular models in drug discovery



Adapted from Paul *et al.* "How to Improve R&D Productivity: The Pharmaceutical Industry's Grand Challenge." Nature Reviews Drug Discovery, 2010



It is often easy to see what a compound does to cells or to animals.

It takes time and can be challenging to understand why it does so.

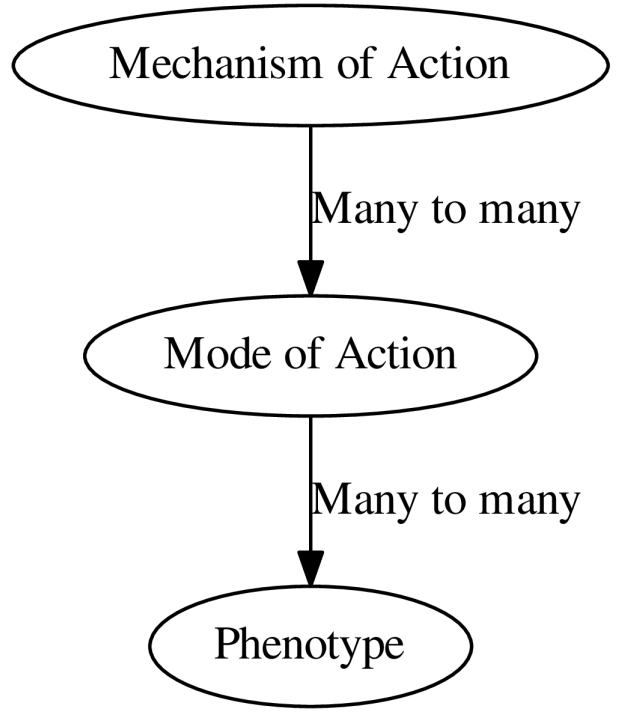
Outline

- **What is MoA, and how can we study it with cell- and omics models?**
- **What modality-specific approaches are available?**
  - Small molecules
  - Therapeutic antibodies
  - Antisense oligonucleotides
- **Quiz**

What is MoA and how can we study it?

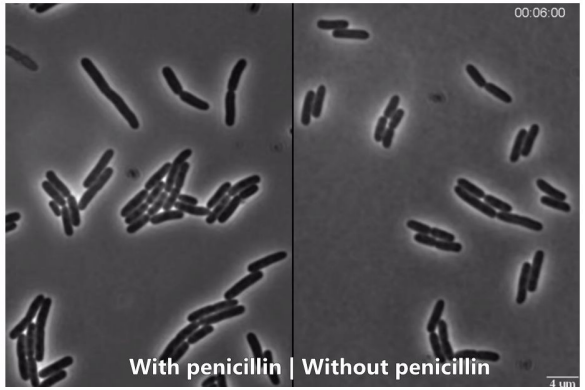
Mechanism of Action & Mode of Action

- **Mechanism of Action**: The specific biochemical interaction through which a drug substance produces its pharmacological effect, ***at the molecular level***.
- **Mode of Action**: Functional or anatomical changes, ***at the cellular level,*** resulting from the exposure of a living organism to a substance.
- For instance, a mechanism of action of a drug can be "*binding to epithelial growth factor receptor (EGFR)*" while its mode of action would be "*inhibition of proliferation*".
- In this talk we use the two terms interchangeably, since in many cases we want to understand ***both*** to make a good drug.

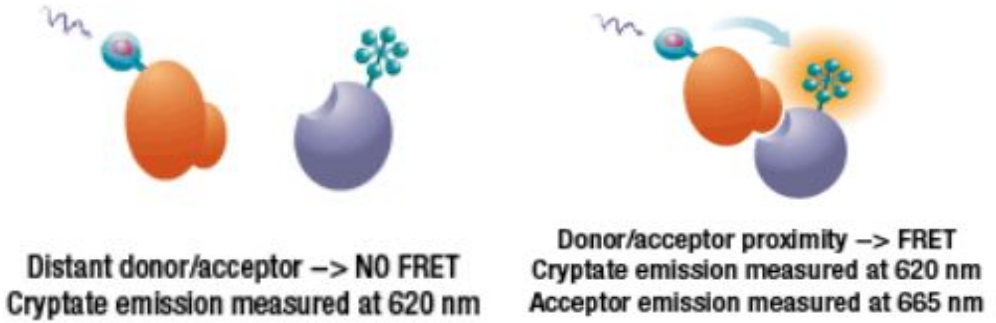


General approaches for MoA understanding

- **Microscopy-based methods**, *e.g.* bacteria phenotyping
- **Direct biochemical methods**, *e.g.* binding and TR-FRET (time-resolved fluorescence energy transfer) assays
- **Computer inference methods,** with chemoinformatics, computer-aided drug design, and bioinformatics tools
- **Omics based methods,** with genetics, transcriptomics, proteomics, etc.

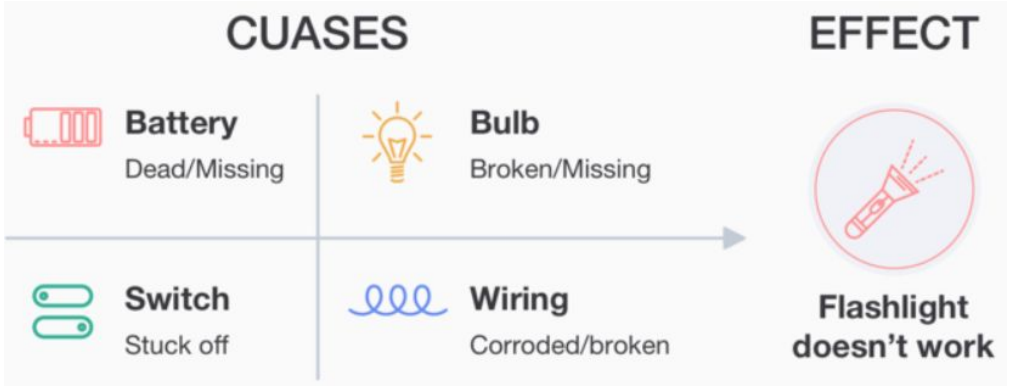


***Lysis of* E.coli, Etienne Maisonneuve & Kenn Gerdes, Center for Bacterial Cell Biology, University of Newcastle, UK**



Principles of TR-FRET, by cisbio.com

Why it can be challenging to understand the MoA of a compound? (I): Many Causes, One Effect



- Many different causes can lead to the same effect.
- The same principle applies to biological systems, where many different inputs can lead to highly similar outputs.

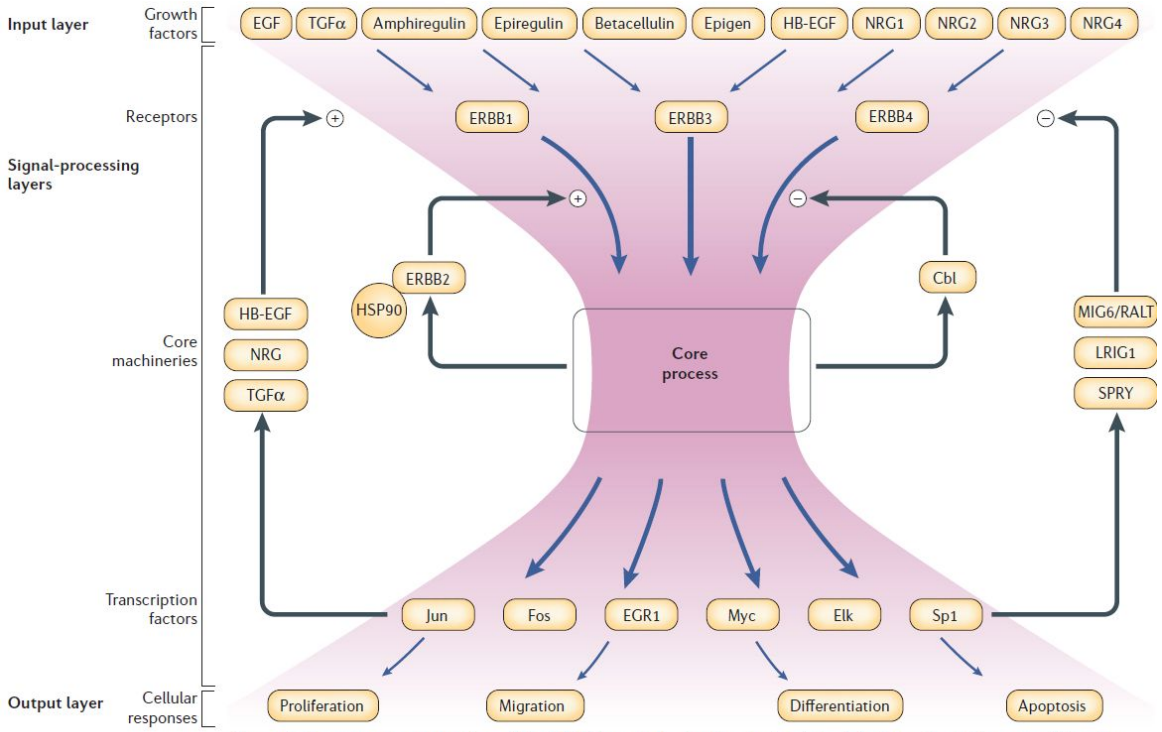
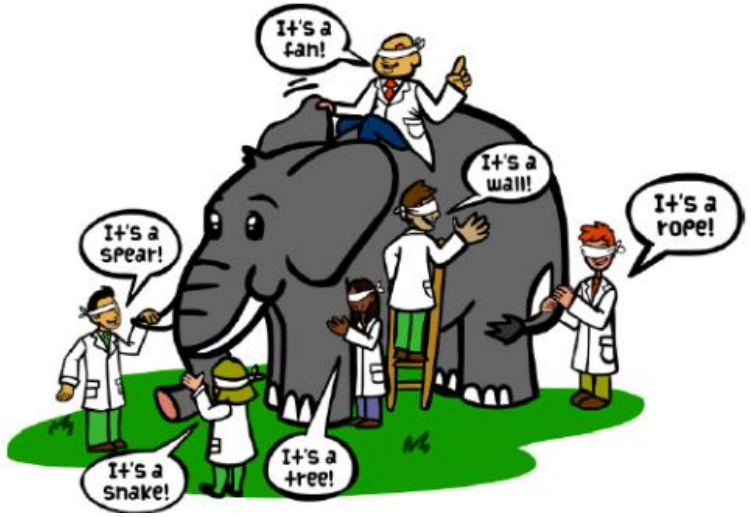


Figure 1 | **A systems perspective of the ERBB network.** A reductionist view of the bow-tie-architectured signalling

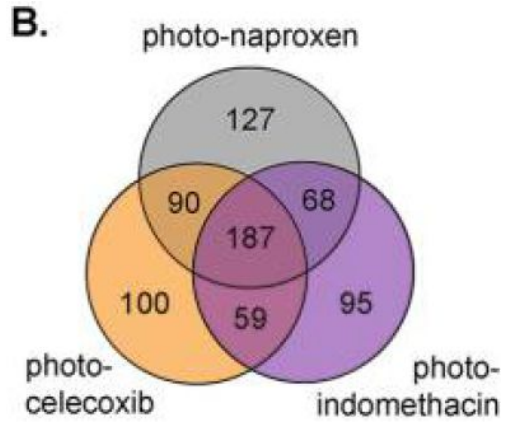
Citri, Ami, and Yosef Yarden. "EGF–ERBB Signalling: Towards the Systems Level." *Nature Reviews Molecular Cell Biology* 7, no. 7 (July 2006): 505–16

Many causes, One Effect **makes MoA understanding challenging**

Why it can be challenging to understand the MoA of a compound? (II): The *One MoA* assumption may be wrong

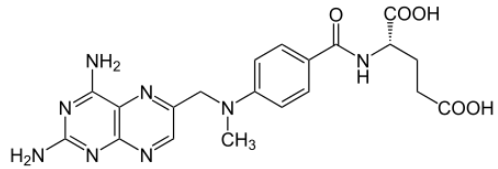


- A drug may have multiple MoAs. Most methods study only one type of effect.
- Recent findings in medicinal chemistry, pharmacology and bioinformatics proffer a 'multi-MoA' view.



Three commonly used NSAIDs are found bound to a surprisingly high number of proteins in cells. Gao *et al.*, *J. Am. Chem. Soc.* 2018, 140, 4259−4268

Methotrexate

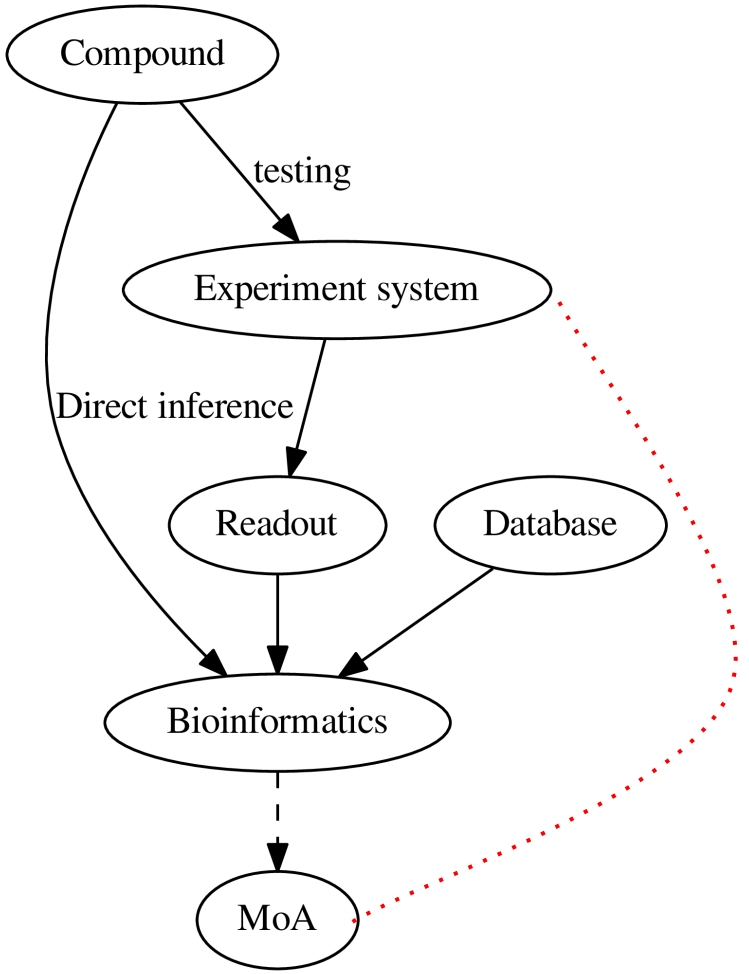


- **[As chemotherapy agent]** Inhibiting dihydrofolate reductase (DHFR) and consequently DNA synthesis.
- **[As immunosuppressant]** Multiple mechanisms, *e.g.* (1) inhibiting purine metabolism, (2) inhibiting methyltransferase, and (3) inhibiting IL-1b binding to its receptor.

Drugs may have more than one MoA, which makes MoA understanding challenging

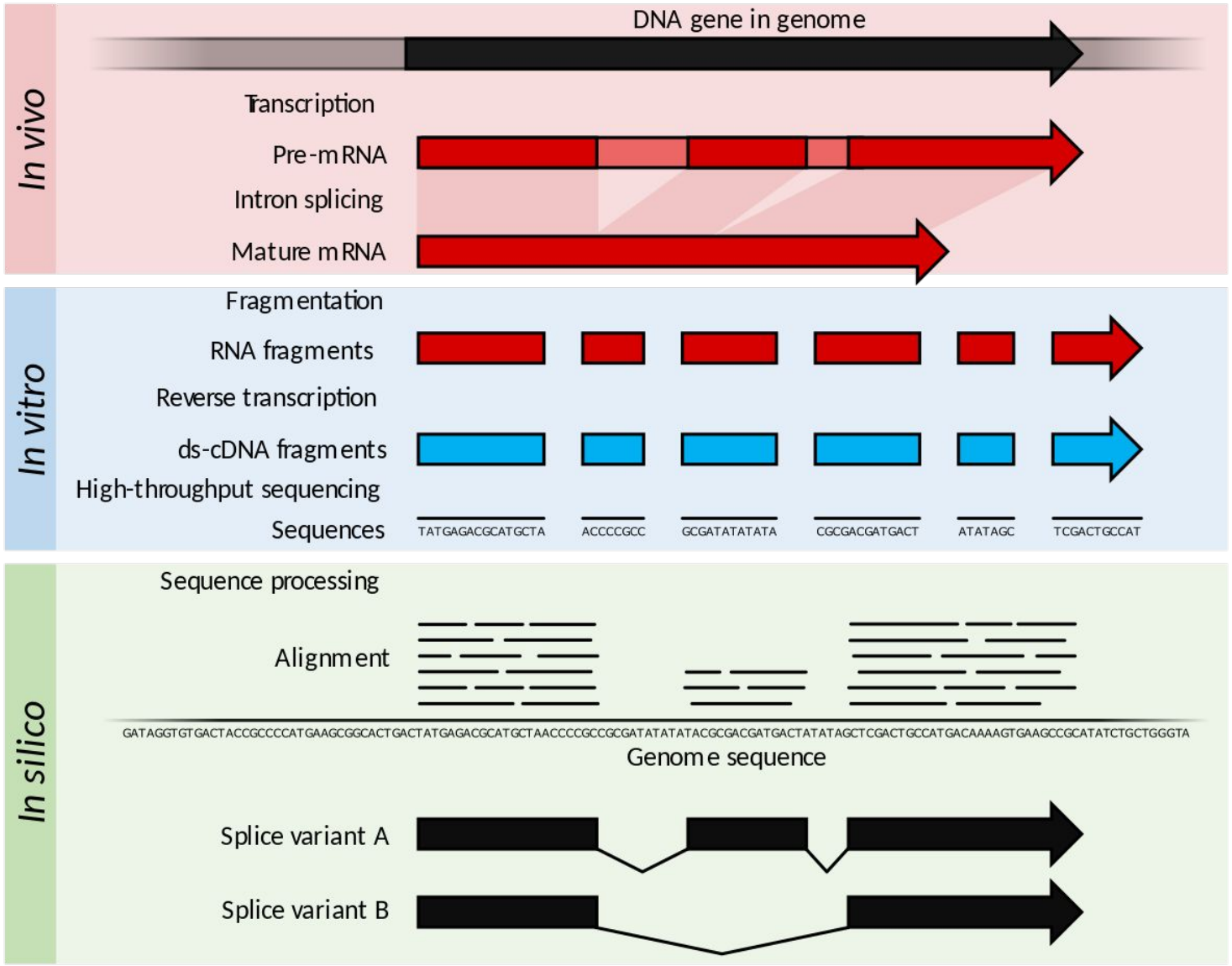
Bioinformatics contributes to MoA understanding of many modalities by integrating information

- MoA can be inferred either with the information of the compound alone, or with the data generated by testing the compound in *in vitro* or *in vivo* experiment systems. Prior knowledge encoded in databases is often of great help.
- The process is usually iterative with hypothesis-testing cycles.
- Many bioinformatics approaches are applicable to virtually all modalities, for instance:
  - Experiment design
  - Sequence analysis
  - Analysis of RNA-sequencing and other omics data
  - Statistical data modelling
  - Network analysis
- Modality-specific approaches are illustrated later.



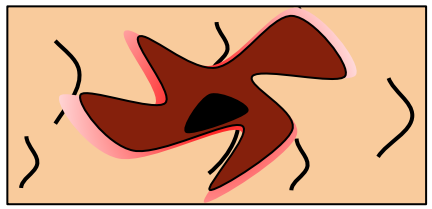
MoA understanding with bioinformatics works by information integration and iterations

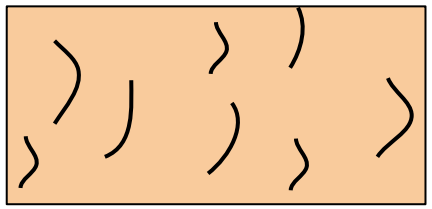
Principle of RNA sequencing



RNA Sequencing

Bulk





Condition 1 (compound treatment)

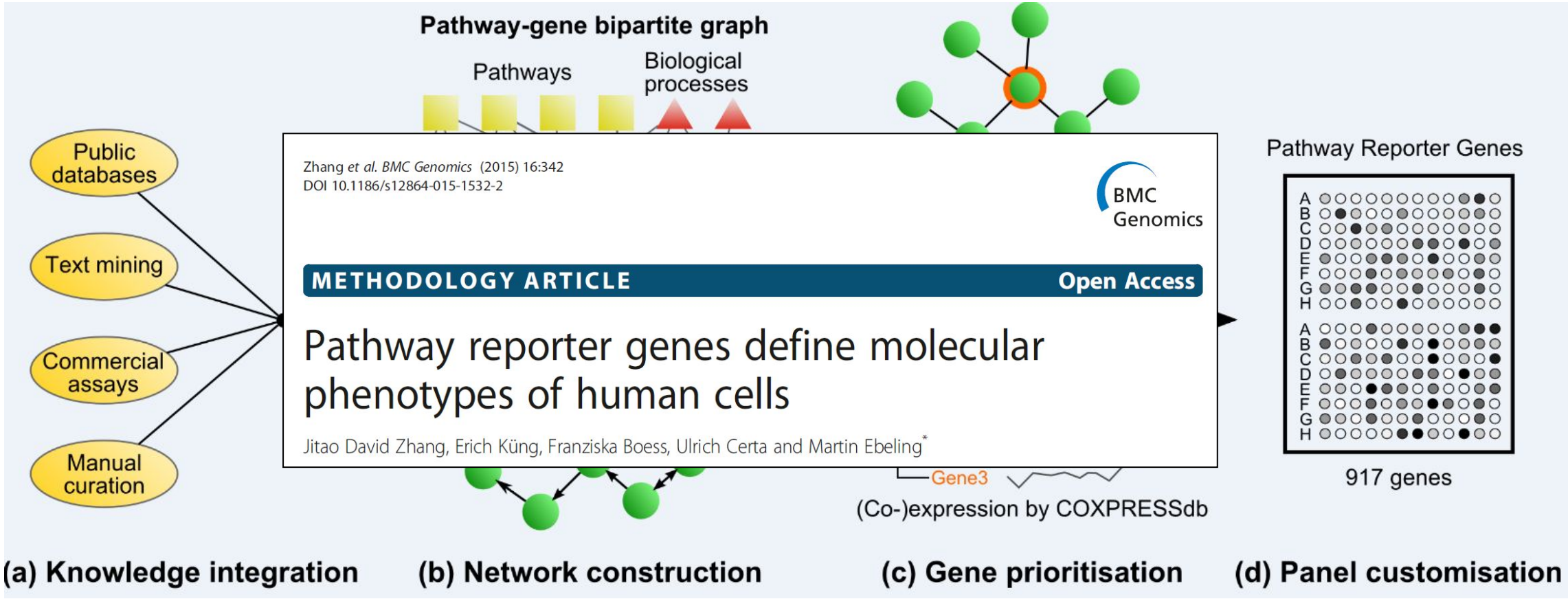
Condition 2 (control)

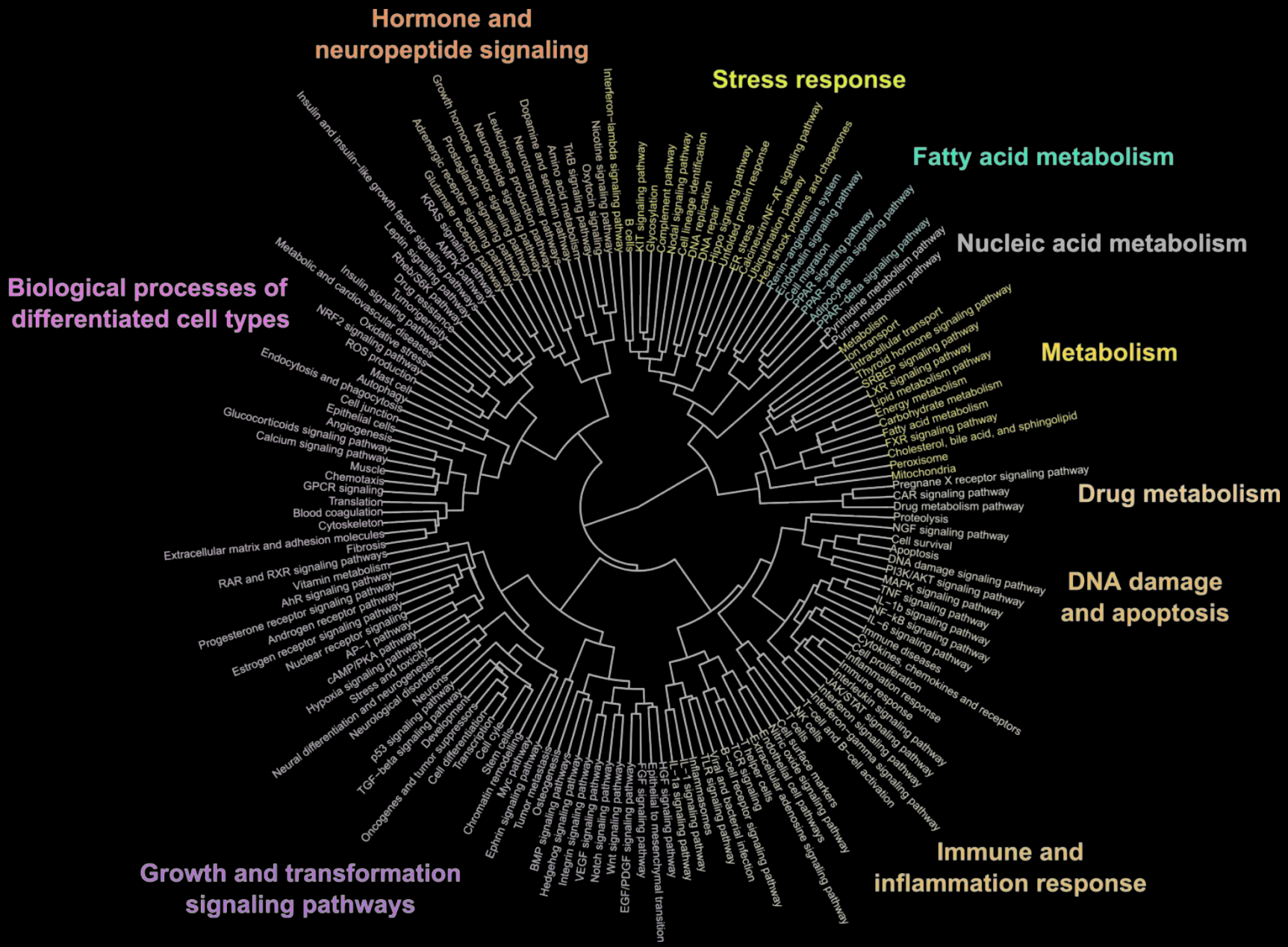
RNA → Sequencing → Reads → Mapping → Quantification → Samples

Genes

Expression matrix

Pathway Reporter Genes

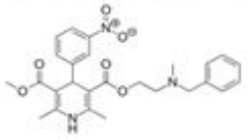



Hormone and neuropeptide signaling
Stress response
Fatty acid metabolism
Nucleic acid metabolism
Metabolism
Drug metabolism
DNA damage and apoptosis
Immune and inflammation response
Growth and transformation signaling pathways
Biological processes of differentiated cell types
Interferon−lambda signaling pathway
B cells
KIT signaling pathway
Glycosylation
Complement pathway
Nodal signaling pathway
Cell lineage identification
DNA replication
DNA repair
Hippo signaling pathway
ER stress
Unfolded protein response
Calcineurin/NF−AT signaling pathway
Ubiquitination pathway
Heat shock proteins and chaperones
Renin−angiotensin system
Endothelin signaling pathway
Cell migration
PPAR signaling pathway
PPAR−gamma signaling pathway
Adipocytes
PPAR−delta signaling pathway
Pyrimidine metabolism pathway
Purine metabolism
Metabolism
Ion transport
Intracellular transport
Thyroid hormone signaling pathway
SREBP signaling pathway
LXR signaling pathway
Lipid metabolism pathway
Energy metabolism
Carbohydrate metabolism
Fatty acid metabolism
FXR signaling pathway
Cholesterol, bile acid, and sphingolipid
Peroxisome
Mitochondria
Pregnane X receptor signaling pathway
CAR signaling pathway
Drug metabolism pathway
Proteolysis
NGF signaling pathway
Cell survival
Apoptosis
DNA damage signaling pathway
PI3K/AKT signaling pathway
MAPK signaling pathway
TNF signaling pathway
IL−1b signaling pathway
NF−kB signaling pathway
IL−6 signaling pathway
Immune diseases
Cytokines, chemokines and receptors
Cell proliferation
Inflammation response
Immune response
Interleukin signaling pathway
JAK/STAT signaling pathway
Interferon−gamma signaling pathway
T−cell and B−cell activation
NK cells
T cells
Cell surface markers
Nitric oxide signaling pathway
Endothelial cell pathways
Extracellular adenosine signaling pathway
Thymic cells
TCR signaling
B−cell receptor signaling pathway
Viral and bacterial infection
TLR signaling pathway
Inflammasomes
IL−1a signaling pathway
HGF signaling pathway
Epithelial to mesenchymal transition
FGF signaling pathway
EGF/PDGF signaling pathway
Wnt signaling pathway
Notch signaling pathway
VEGF signaling pathway
Integrin signaling pathway
Hedgehog signaling pathway
BMP signaling pathways
Osteogenesis
Tumor metastasis
Myc pathway
Ephrin signaling pathway
Chromatin remodeling
Stem cells
Cell Cycle
Transcription
Cell differentiation
Oncogenes and tumor suppressors
TGF−beta signaling pathway
p53 signaling pathway
Development
Neurons
Neural differentiation and neurogenesis
Neurological disorders
Stress and toxicity
Hypoxia signaling pathway
cAMP/PKA pathway
AP−1 pathway
Nuclear receptor signaling
Estrogen receptor signaling pathway
Androgen receptor pathway
Progesterone receptor signaling pathway
AhR signaling pathway
Vitamin metabolism
RAR and RXR signaling pathways
Fibrosis
Extracellular matrix and adhesion molecules
Cytoskeleton
Blood coagulation
Translation
GPCR signaling
Chemotaxis
Muscle
Calcium signaling pathway
Glucocorticoids signaling pathway
Angiogenesis
Epithelial cells
Cell junction
Autophagy
Mast cell
Endocytosis and phagocytosis
ROS production
NRF2 signaling pathway
Oxidative stress
Tumorigenicity
Drug resistance
Diabetic and cardiovascular diseases
Insulin signaling pathway
RhebSGK pathway
Leptin signaling pathway
AMPK pathway
KRAS signaling pathway
Glutamate receptor signaling pathway
Adrenergic receptor signaling pathway
Prostaglandin signaling pathway
Growth hormone receptor signaling pathway
Neuropeptide signaling pathway
Leukotrienes production pathway
Neurotransmitter pathways
Dopamine and serotonin pathways
Amino acid metabolism
TrkB signaling pathway
Oxytocin signaling
Nicotine signaling pathway
Insulin and insulin−like growth factor signaling pathway

Molecular phenotyping reveals modulation of human pathway activities by compounds

A workflow to quantify expression of pre-defined pathway reporter genes at early time points after perturbation to infer pathway activities, which may predict late-onset cellular phenotypes

Small molecule



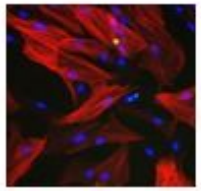
Antibodies

Antisense oligos

Therapeutic candidates

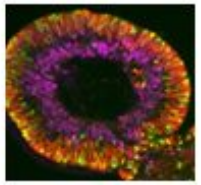
Early time point (3-12h)

Human *in vitro* disease models



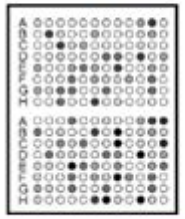
Cell lines/ primary cells

iPS-derived cells (opt. genome editing)



Advanced models

~1000 pathway reporter genes



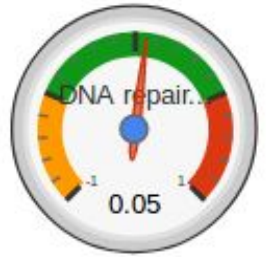
Next-generation sequencing

Molecular phenotyping

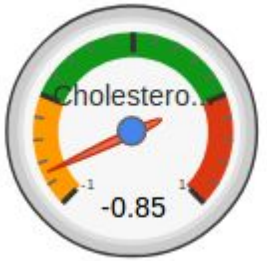
Molecular Phenotyping was established within Roche and is used to study MoA of compounds

An illustrative example of molecular phenotyping results

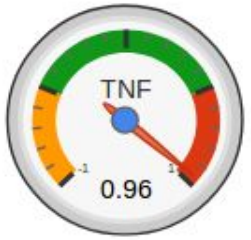
Inferred activity of 154 human pathways



Little change of genes induced by DNA repair



Inhibition of reporter genes of cholesterol synthesis



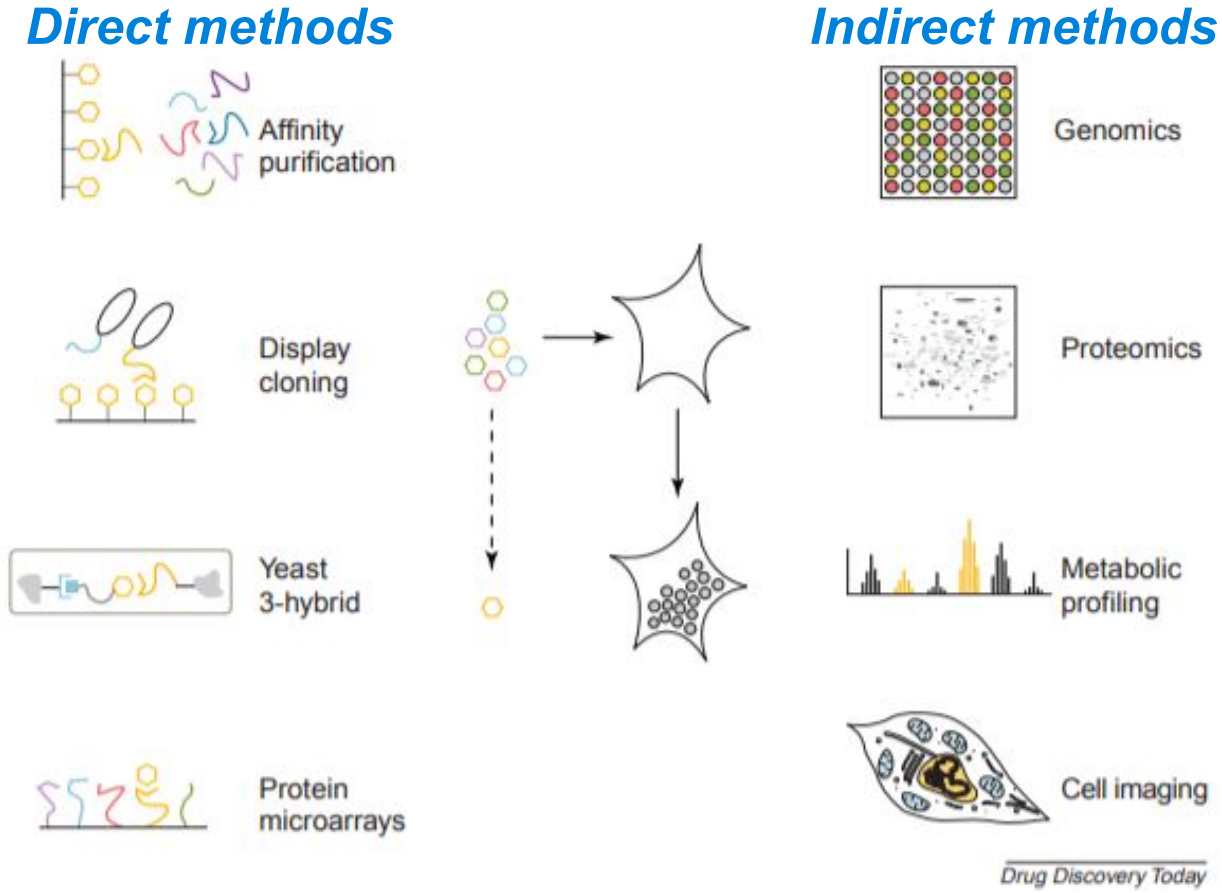
Induction of genes downstream of TNF-alpha signalling

Molecular phenotyping reveals what pathways are modulated by each compound

Understanding MoA of small molecules

MoA inference for small molecules

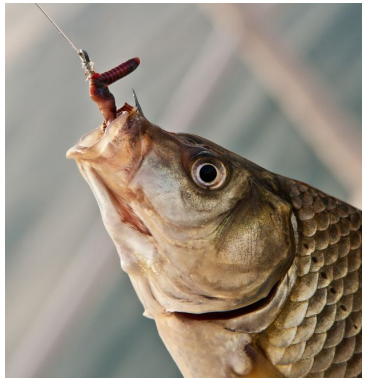
- Small-molecule drug candidates can be discovered via target-based or phenotypic approach. **Target information and MoA are wished for both types of molecules.**
- Comess *et al.* (AbbVie), Journal of Medicinal Chemistry (2018), gave a solid review.
- Prunotto *et al.* (Roche+Genentech, Pfizer, Eli Lilly, Novartis), Nature Reviews Drug Discovery (2017), discussed target ID for phenotypic drug discovery in more details.



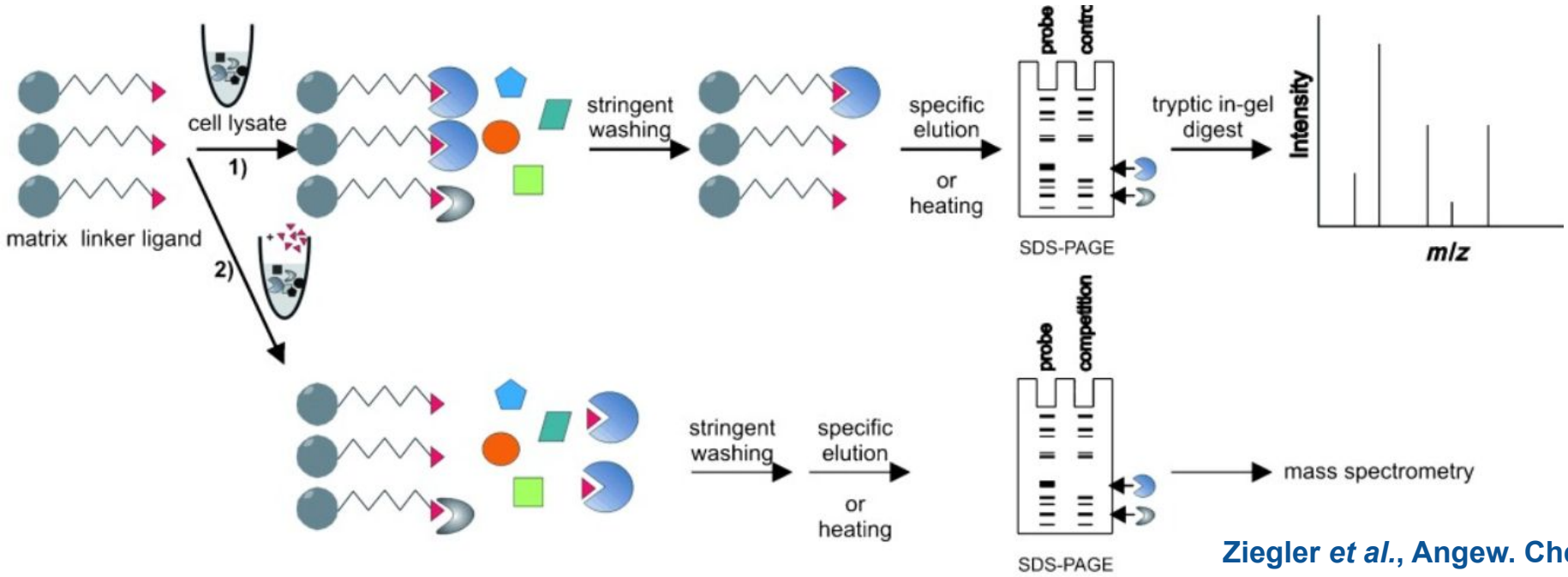
Simple figure, complex issue?
This figure (Hart, Drug Discovery Today, 2005, slightly adapted) provides nevertheless a good overview.

Bioinformatics helps translating data complexity into meaningful decision making

Chemoproteomics methods



- Chemoproteomics methods are based on two principles: (1) **bait/prey** and (2) **competition**.
- Commonly used methods to identify binding partners of small molecules include affinity-based profiling (shown below), activity-based profiling, SILAC, *etc.*



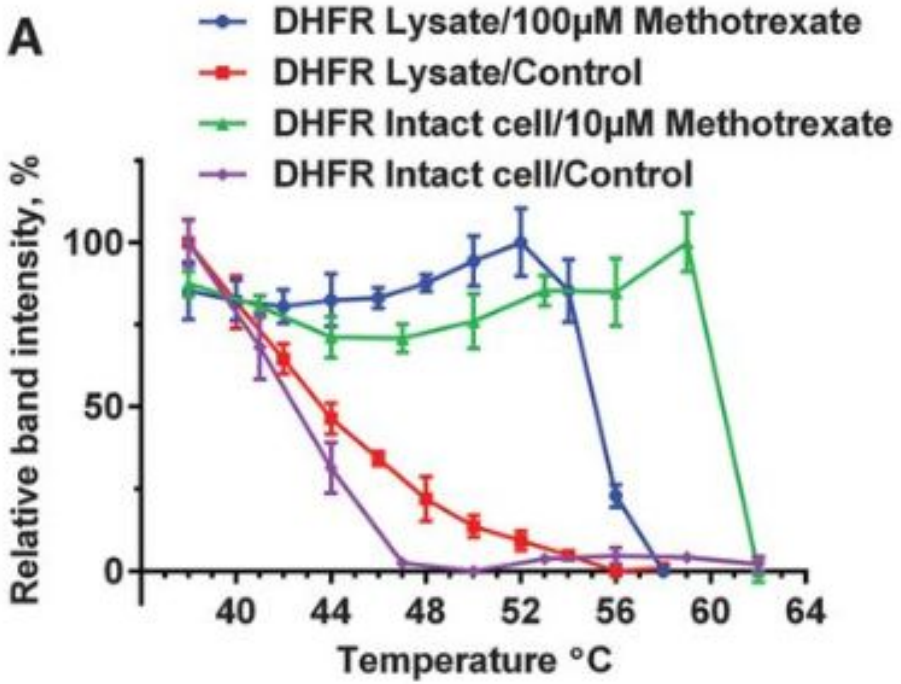
Ziegler *et al.*, Angew. Chemie Int. Ed., 2013

Bioinformatics empowers chemoproteomics by statistical analysis and data interpretation

Protein stability-based methods

DON'T EAT, NOT EVEN COOKED!
The *death cap* contains *amatoxin*, a thermal stable toxin.

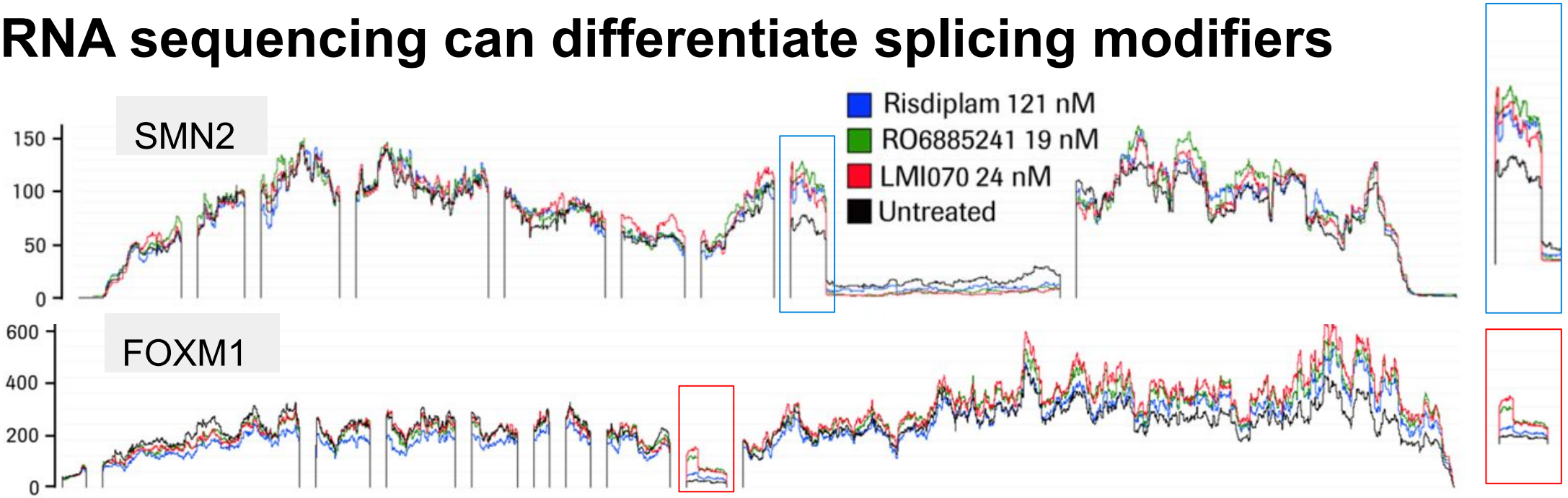
- Proteins are usually stabilized by ligands binding to them.
- This principle can be used to identify protein targets of a ligand without modification of the ligand (label-free)
- Currently prohibitively expensive due to patents.



Results of Cellular Thermal Shift Assay (CETSA) to verify DHFR as a target of methotrexate. Molina *et al.*, Science, 2013.

Label-free methods despite of limitations have great potentials

RNA sequencing can differentiate splicing modifiers



Ratni, Ebeling, Baird, *et al.*, J Med Chem (2018)

- Example above: number of reads mapped to human genomic loci *SMN2* (on target) and *FOXM1* (off target) in patient fibroblasts. Each peak corresponds to an exon present in mature mRNA.
- While risdiplam (blue) and the competitor compound (red) show similar on-target effects, the off-target effects of risdiplam are much less pronounced than those of the competitor compound.

RNA-sequencing and bioinformatics analysis are essential to study splicing modifiers

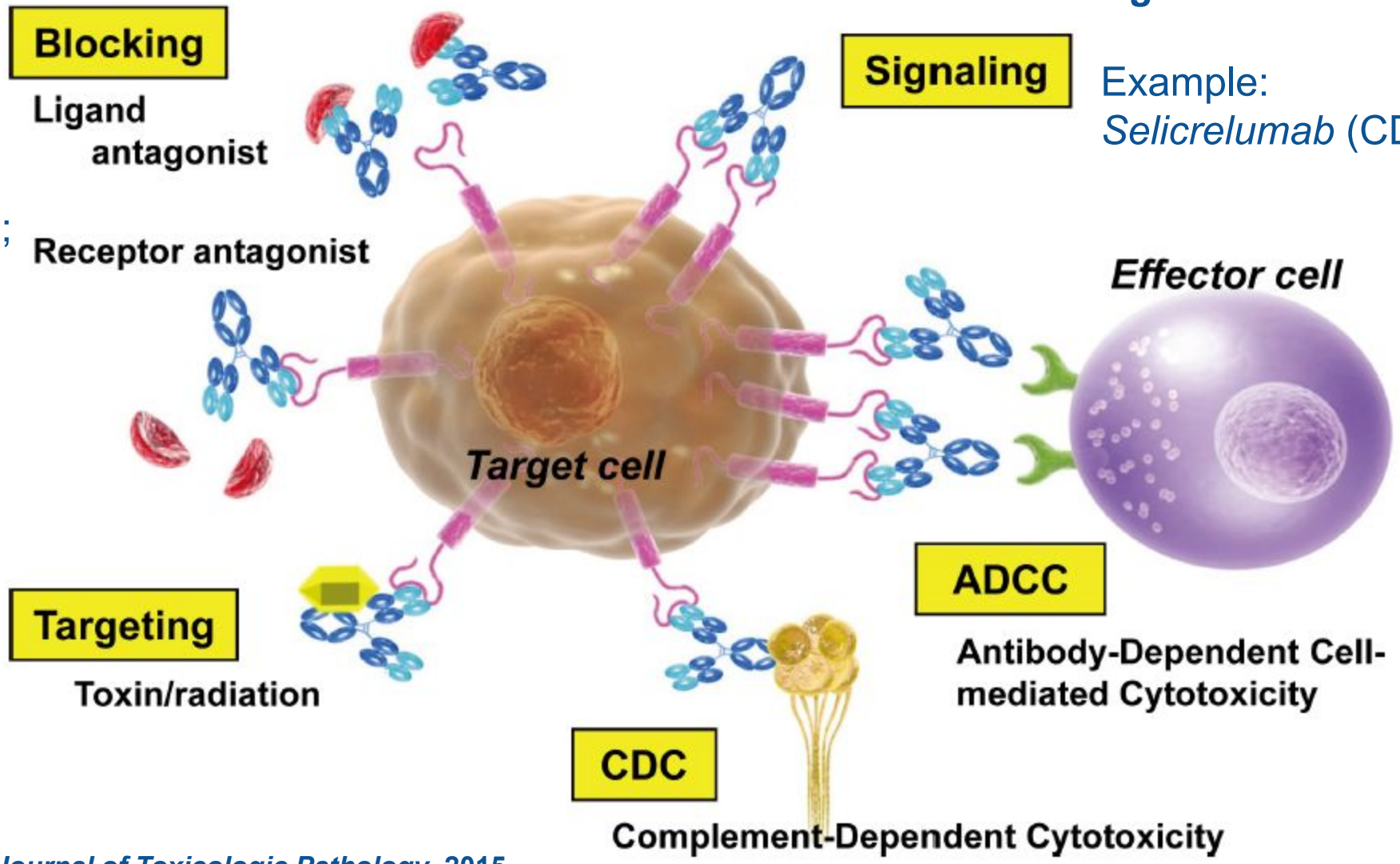
Understanding MoA of antibodies

Therapeutic antibodies

Antagonistic

Examples:
Trastuzumab (blocking Her2 ligand binding);
Pertuzumab (blocking Her dimerization)

Example: T-cell bispecific antibody, which brings T cells and tumor cells together and activates T cells



Agonistic

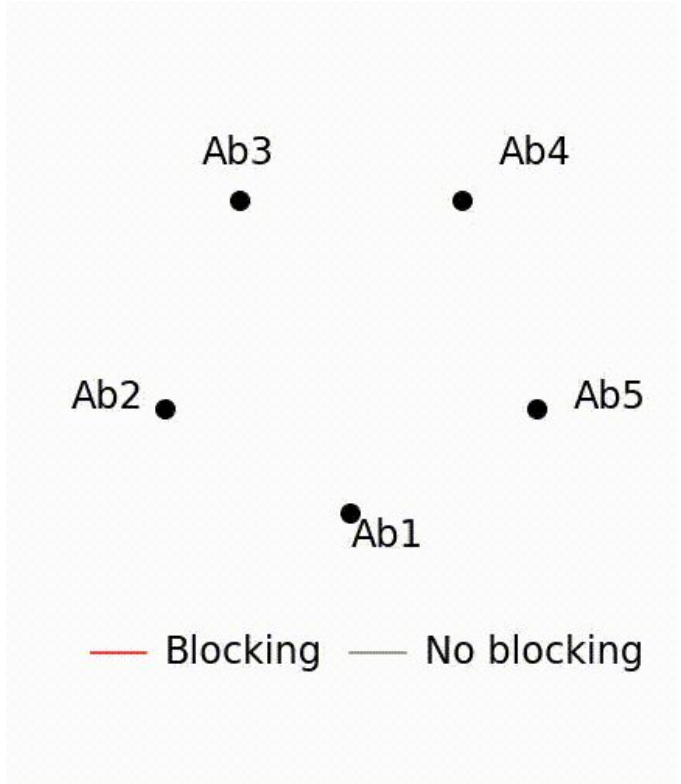
Example:
Selicrelumab (CD40 agonist)

CDC and ADCC contribute to the efficacy and safety profiles of many antibody drugs

Suzuki, Masami *et al. Journal of Toxicologic Pathology*, 2015

Bioinformatics approaches are integral to MoA understanding of therapeutic antibodies

- **Epitope binning and epitope mapping**: antibodies of a target are tested pairwise against each other to see whether antibodies block one another's binding to the epitope of an antigen. Antibodies are *binned* by the competitive blocking profiles. The information of epitope binding is important to understand MoA of the compound as well as for differentiation.
- Other bioinformatics topics in antibody drug discovery:
  - **Sequence analysis** and comparative genomics;
  - **Study of avid effects**, for antibodies with two paratopes, using surface plasmon resonance (SPR);
  - Study and prediction of **immunogenicity** and **anti-drug antibodies (ADA)**.

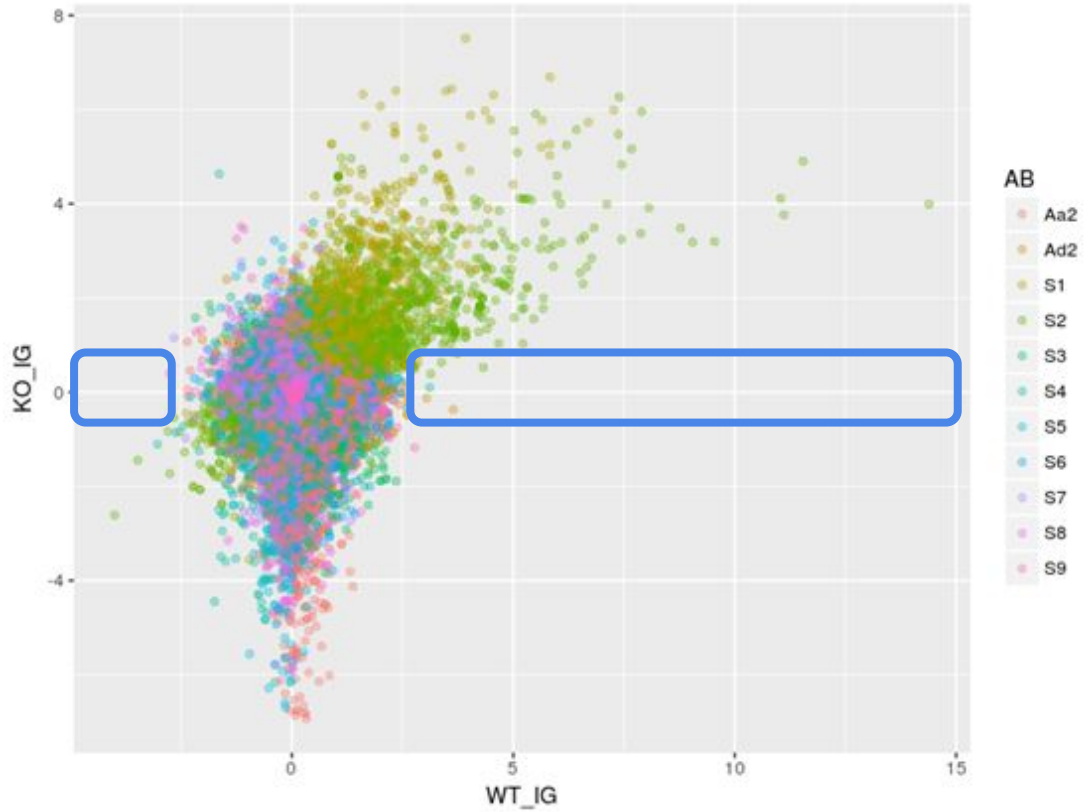


The principle of epitope binning. In this toy example, {A1, A3, A5} are binned together, and {A2, A4} are binned together.

A variety of bioinformatics approaches contribute to MoA understanding of antibodies

Molecular phenotyping revealed unexpected effects of antibodies

- In this experiment, 11 antibodies were characterized by their effects on human pathways in THP-1 cells (either wild-type [WT], or target knockout cells [KO]).
- The right panel summarizes effects on pathways. Each point corresponds to one pathway. The colors encode the antibodies used.
- We are looking for antibody effects that are caused by effects on the target, which means they should be (a) absent in KO cells, (b) visible when comparing versus the unspecific IgG1 (the blue boxes).
- Surprisingly, we found that some antibodies showed similar effects in WT and KO cells, while others either show little effects in both, or show stronger effects in KO than in WT cells. The observations raise question about MoA and quality of the antibodies.



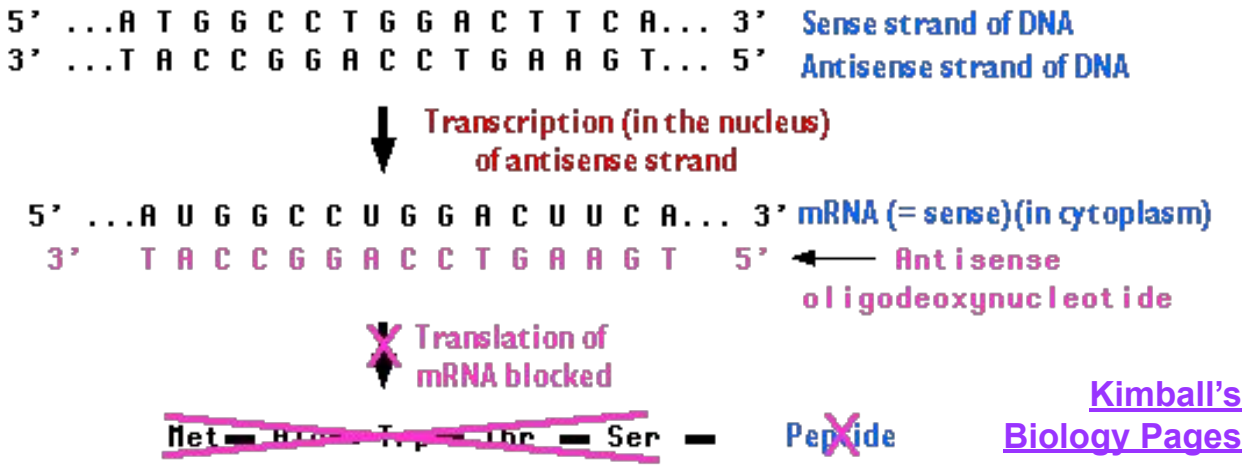
On the X axis, we plot a score indicating effects in WT compared to IgG1-treated controls (positive/negative scores indicate up-/down-regulation compared with IgG1). On the Y axis, the same is shown for effects compared to IgG1 controls in KO cells. Scores are overall lower but far from insignificant. *Courtesy of Martin Ebeling.*

These unexpected effects are currently being investigated to gain new insights in biology

Understanding MoA of antisense oligonucleotides

Sequence-dependent binding of oligonucleotides induces both on- and off-target effects

- Antisense Oligonucleotides (ASOs) work by binding to mRNA transcripts in a **sequence-dependent** way.
- ASO-mRNA binding is a chemical reaction with a spectrum of affinities. For simplification (!), we often use the following classification:
  - **On-target,** usually of one mRNA species.
  - **Off-targets** potentially of many undesired mRNA species.
  - **Non-targets,** hardly bound by the ASO, though they can be potentially regulated by secondary effects.



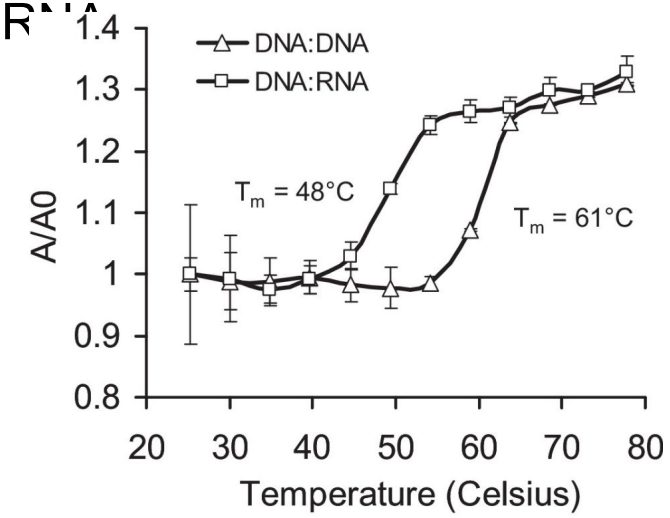
Human mRNAs

My silver-bullet oligo (3'-5'!)		
TACCGGACCTGAAGT	AUGGCCUGGACUUCA	On target
	AUGGCCUGGUCUUCA	Off target
	AUGGCCUGCUCUUCA	Off target
	AUGGCCACCACUUCA	Non target
	. . .	
	UACGUCGUAGUCUUC	Non target

Understanding the sequence of ASOs is critical to understand their MoAs

The binding affinity between RNA and ASO can be measured by the melting temperature T_m

- Binding affinity between RNA and ASOs can be measured by the duplex melting temperature (T_m), the temperature at which half of the ASOs are duplexed with RNA



A typical plot of measuring T_m. Zhou *et al.*, Colloids and Surfaces A, 2015

- The higher is the T_m, the stronger is the binding, when other conditions are constant.

Name	Target	Sequence (5′ to 3′)[a]	Length (nt)	T_m (°C)
T1	*Tradd*	GctcatactcgtaggcCA	18	66.8
T2	*Tradd*	GCtcatactcgtaggcCA	18	69.7
T3	*Tradd*	GCtcatactcgtaggCCA	18	72.1
T4	*Tradd*	GCTcatactcgtaggcCA	18	73.3
T5	*Tradd*	GCTcatactcgtaggCCA	18	76.3

Part of Table 1 of Hagedorn *et al*, NAR, 2018.

Question: when other conditions are constant, which ASO binds strongest to the target gene *Tradd*?

T_m can be used to characterize binding affinity between ASO and target/off-target RNAs

It is possible to predict melting temperature (i.e. binding affinity) of ASO-mRNA pairs with *free energy*

- It is a mature application of bioinformatics to predict $T_{m,}$ using the nucleotide sequences and the principles of **nucleic acid thermodynamics** and **dynamic programming.**
- The melting temperature is correlated with the free energy (ΔG°), which can be predicted by a fast and effective algorithm (Rehmsmeier *et al.*, RNA, 2004).
- **The more negative the free energy is** (i.e. the larger the absolute value is), the higher is T_m,namely **the ASO-mRNA pair is more likely to be stable** .

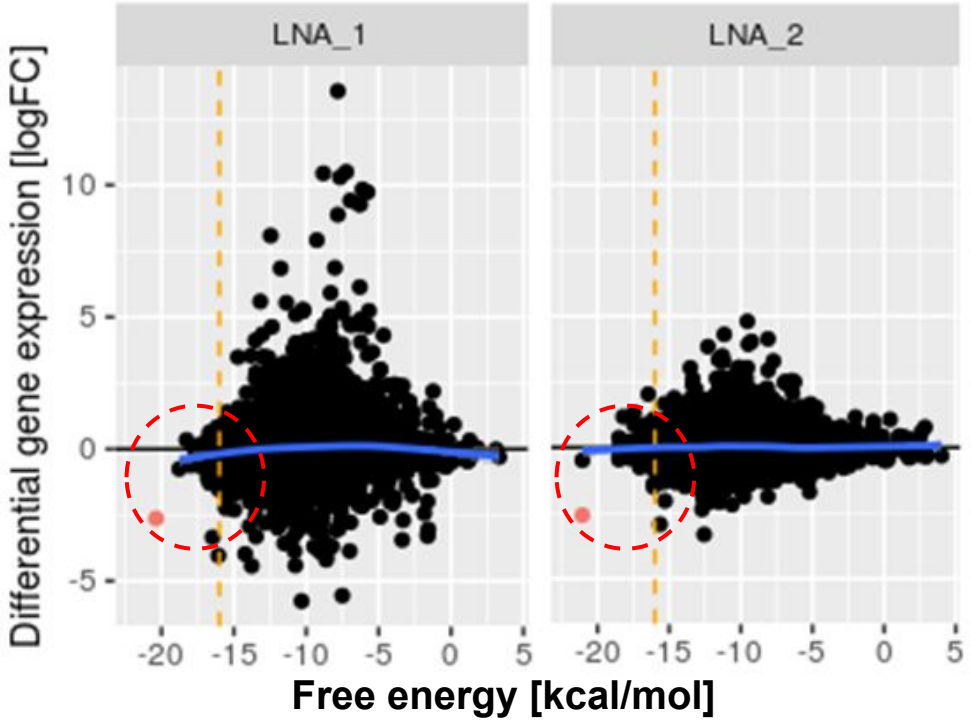
My silver-bullet oligo (5'-3')	*Human mRNAs*	*Free Energy (kcal/mol)*
TGAAGTCCAGGCCAT	AUGGCCUGGACUUCA	-32.8
	AUGGCCUGGUCUUCA	-28.5
	AUGGCCUGCUCUUCA	-23.7
	AUGGCCACCACUUCA	-20.2
	. . .	
	UACGUCGUAGUCUUC	-9.8

Question: Other conditions held constant, which mRNA has the highest predicted T_m given the data?

It is possible to predict the free energy of binding and T_m of any ASO-mRNA pair

Transcriptomics profiling allows simultaneous investigation of on- and off-target effects

- RNA-sequencing is able to quantify both on- and off-target effects of ASOs by measuring gene expression changes.
- Differential gene expression analysis can be used together with ASO-mRNA binding-affinity prediction to reveal off-target potentials of the tested ASOs.
- At the same time, RNA-sequencing can review pathway- and network-level changes induced by ASOs, to inform both efficacy and safety studies.



A declining trend at the left end (red dashed circle) is a warning sign: mRNAs that are predicted bound to the ASO are down-regulated, revealing potential off-target effects.

RNA-sequencing and bioinformatics analysis help prioritising LNA candidates

Quiz

Q: What is MoA?
A: MoA is the effect of a drug at the molecular or the cellular level.

Q:What is molecular phenotyping?
A: Infer pathway activities by quantifying expression of ~1000 pathway reporter genes (the gauges!).

Q: What specific methods are there for small-molecule MoA studies?
A: Chemoproteomics, protein-stability based methods, RNA-sequencing, …

Q: What specific methods are there for antibody MoA studies?
A: Epitope binning and epitope mapping, sequence analysis, SPR analysis, molecular phenotyping...

Q: What is the measure of binding affinity between ASOs and mRNAs?
A: Melting temperature and/or the free energy of binding.

Further readings

Molecular Phenotyping: Drawnel, Faye Marie, Jitao David Zhang, Erich Küng, Natsuyo Aoyama, Fethallah Benmansour, Andrea Araujo Del Rosario, Sannah Jensen Zoffmann, et al. “Molecular Phenotyping Combines Molecular Information, Biological Relevance, and Patient Data to Improve Productivity of Early Drug Discovery.” *Cell Chemical Biology* 18, no. 24(5), 2017: 624–34. https://doi.org/10.1016/j.chembiol.2017.03.016.

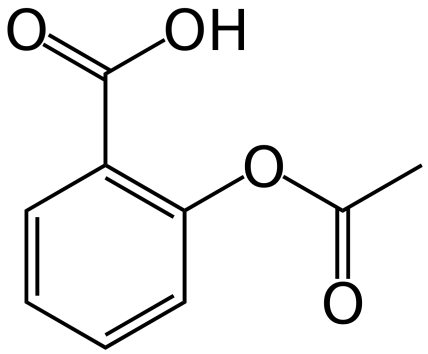
Small molecules: (a) Rix, Uwe, and Giulio Superti-Furga. “Target Profiling of Small Molecules by Chemical Proteomics.” *Nature Chemical Biology* 5, no. 9 : 616–24. https://doi.org/10.1038/nchembio.216 **(b)** Comess, Kenneth M., Shaun M. McLoughlin, Jon A. Oyer, Paul L. Richardson, Henning Stöckmann, Anil Vasudevan, and Scott E. Warder. “Emerging Approaches for the Identification of Protein Targets of Small Molecules - A Practitioners’ Perspective.” *Journal of Medicinal Chemistry*, May 2, 2018. https://doi.org/10.1021/acs.jmedchem.7b01921.

Therapeutic antibodies: Suzuki, Masami, Chie Kato, and Atsuhiko Kato. “Therapeutic Antibodies: Their Mechanisms of Action and the Pathological Findings They Induce in Toxicity Studies.” *Journal of Toxicologic Pathology* 28, no. 3, 2015: 133–39. https://doi.org/10.1293/tox.2015-0031; Brooks, Benjamin D. “The Importance of Epitope Binning for Biological Drug Discovery.” *Current Drug Discovery Technologies* 11, no. 2 (June 2014): 109–12. https://doi.org/10.2174/15701638106661311242338 27.

Antisense oligonucleotides: (a) Hagedorn, Peter H., Bo R. Hansen, Troels Koch, and Morten Lindow. “Managing the Sequence-Specificity of Antisense Oligonucleotides in Drug Discovery.” *Nucleic Acids Research* 45, no. 5 (March 17, 2017): 2262–82. https://doi.org/10.1093/nar/gkx056 **(b)** Hagedorn, Peter H., Malene Pontoppidan, Tina S. Bisgaard, Marco Berrera, Andreas Dieckmann, Martin Ebeling, Marianne R. Møller, et al. “Identifying and Avoiding Off-Target Effects of RNase H-Dependent Antisense Oligonucleotides in Mice.” *Nucleic Acids Research* 46, no. 11 (June 20, 2018): 5366–80. https://doi.org/10.1093/nar/gky397. Contact Lykke Pedersen at RICC for software source code to calculate affinity parameters of LNAs.

Bonus news outlet: Blog In the Pipeline by Derek Lowe

The road of MoA understanding can be 120 year long



Aspirin trademarked in 1899

Dai *et al*, Cell, 2019

Acetylation blocks cGAS activity and inhibits self-DNA-induced autoimmunity

- Acetylation suppresses cGAS activity
- Aspirin directly acetylates cGAS
- Aspirin inhibits cGAS-mediated interferon production
- Aspirin alleviates DNA-induced autoimmunity in AGS mouse models and patient cells

MoA understanding can be a long process full of surprises

https://xkcd.com/1217/

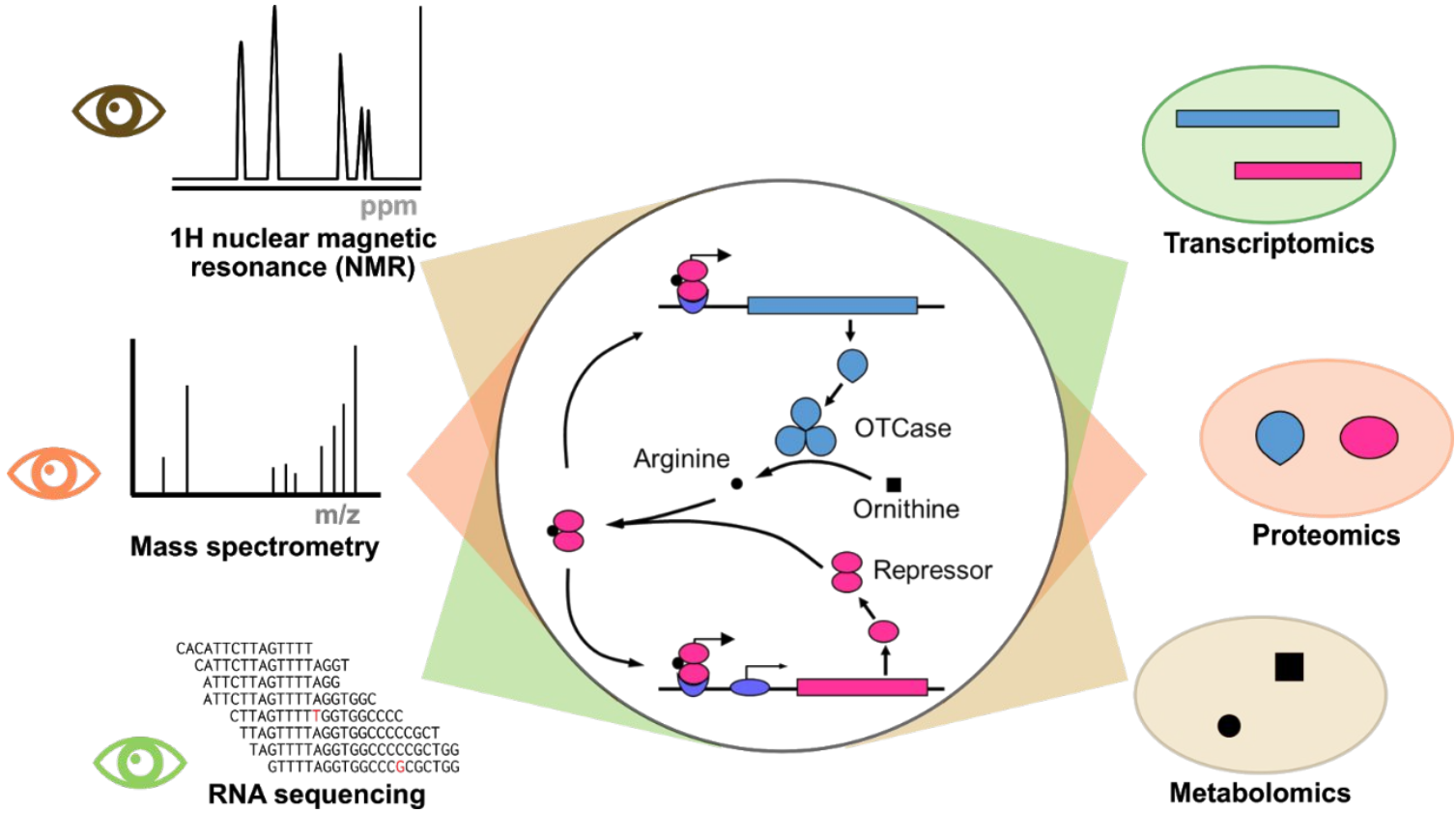
It is often easy to see what a compound does to cells or to animals.

It takes time and can be challenging to understand why it does so.

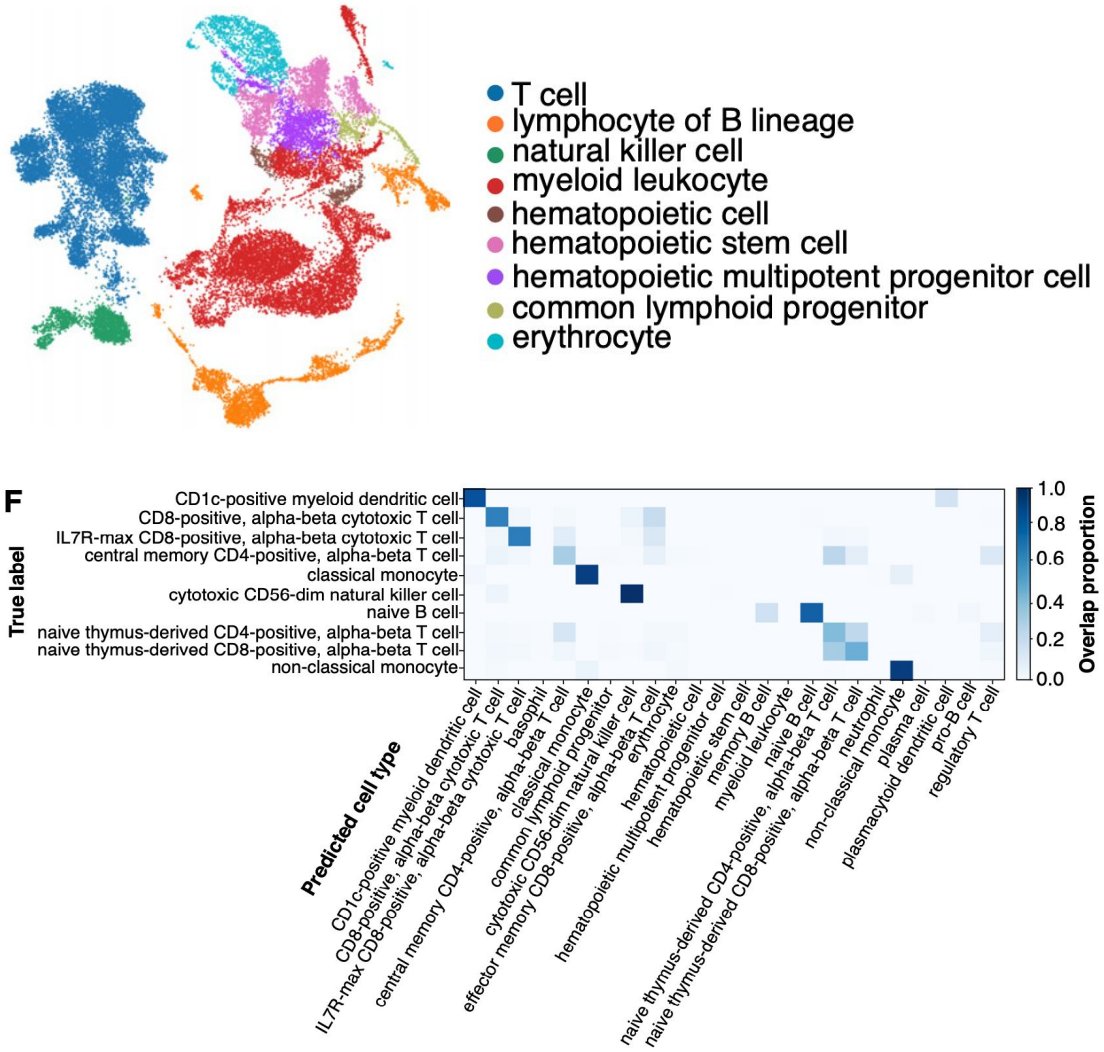
Take a deep breath, let's give it a try…

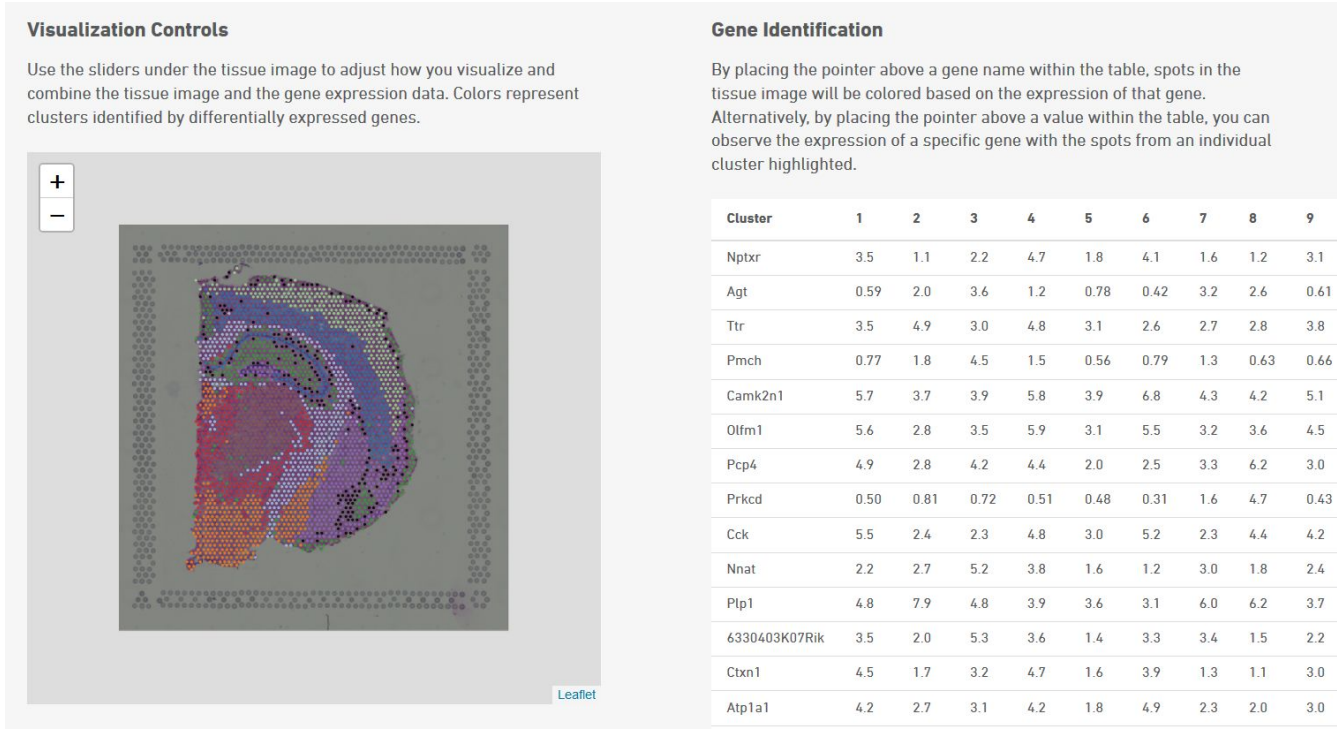
Summary

- **MoA profiling: a case study of omics and cellular modelling**
  - An example of modality-agonistic methods: RNA-sequencing
  - Examples of modality-specific methods: antibody binning and off-target effect prediction for oligonucleotides
- **Current research topics**
  - Single-cell sequencing
  - Spatio-omics profiling
  - Genome editing
  - Microbiome
  - High-content cellular imaging
  - Integrative modelling



From single-cell analysis to spatial-transcriptomics





Cluster	1	2	3	4	5	6	7	8	9
Nptxr	3.5	1.1	2.2	4.7	1.8	4.1	1.6	1.2	3.1
Agt	0.59	2.0	3.6	1.2	0.78	0.42	3.2	2.6	0.61
Ttr	3.5	4.9	3.0	4.8	3.1	2.6	2.7	2.8	3.8
Pmch	0.77	1.8	4.5	1.5	0.56	0.79	1.3	0.63	0.66
Camk2n1	5.7	3.7	3.9	5.8	3.9	6.8	4.3	4.2	5.1
Olfm1	5.6	2.8	3.5	5.9	3.1	5.5	3.2	3.6	4.5
Pcp4	4.9	2.8	4.2	4.4	2.0	2.5	3.3	6.2	3.0
Prkcd	0.50	0.81	0.72	0.51	0.48	0.31	1.6	4.7	0.43
Cck	5.5	2.4	2.3	4.8	3.0	5.2	2.3	4.4	4.2
Nnat	2.2	2.7	5.2	3.8	1.6	1.2	3.0	1.8	2.4
Plp1	4.8	7.9	4.8	3.9	3.6	3.1	6.0	6.2	3.7
6330403K07Rik	3.5	2.0	5.3	3.6	1.4	3.3	3.4	1.5	2.2
Ctxn1	4.5	1.7	3.2	4.7	1.6	3.9	1.3	1.1	3.0
Atp1a1	4.2	2.7	3.1	4.2	1.8	4.9	2.3	2.0	3.0

Left: Mädler, et al. 2020. "Besca, a Single-Cell Transcriptomics Analysis Toolkit to Accelerate Translational Research." NAR Genomics and Bioinformatics, 2021

Top: Spatial resolution of gene expression, which is becoming important for digital pathology and personalized healthcare, source: 10x Genomics

Backup slides

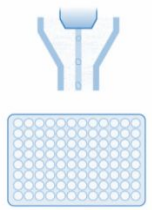
Single-cell sequencing (scSeq) workflow

Tissue dissociation

Single cell capture and transcriptome sequencing

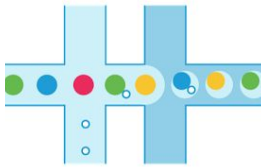
Solid tissue

Dissociated cells/nuclei



Flow cytometry 96-384 cells

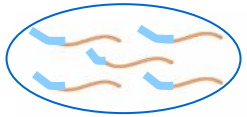
Cell/nuclei suspension



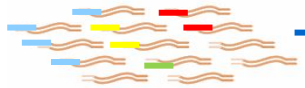
Droplet microfluidics 1000s of cells

Throughput

Single cells or nuclei



cDNA

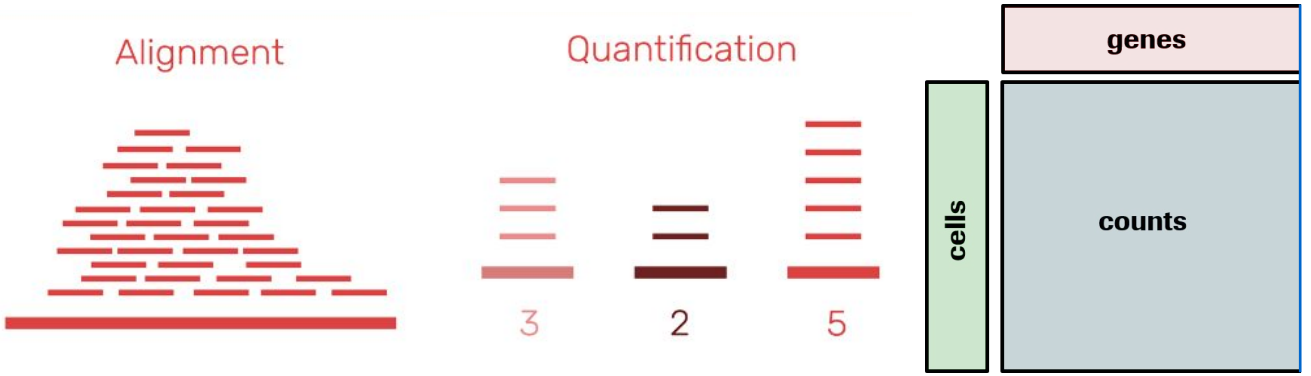


Pooled cDNA libraries

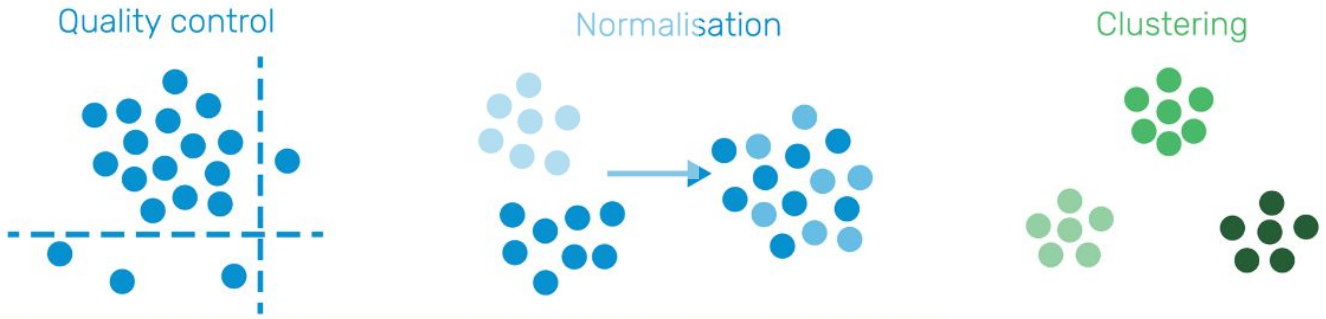
Next-Generation Sequencing

A linearized workflow of scSeq data analysis

From short reads to gene-cell matrix



QC, filtering & normalization, dimensionality reduction, and clustering

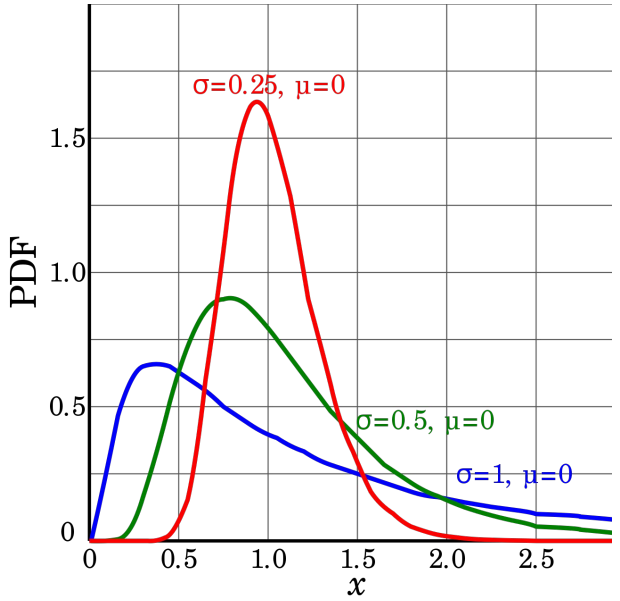


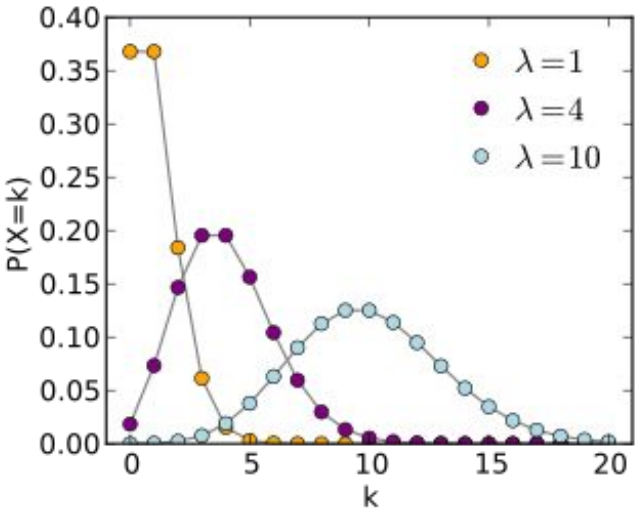
Downstream analysis

Zapia et al., PLoS Computational Biology 2018, adapted

Difference in statistical modelling of microarray data, next-generation sequencing, and single-cell RNA-seq data

- Microarray data: log-normal distributed, for instance implemented in the *limma* package of R/Bioconductor.
- Bulk RNA-sequencing data: Negative-Binomial distributed (or Poisson with overdispersion), for instance implemented in both *edgeR* and *DESeq2* package of R/Bioconductor.
- Single-cell data: some authors recently suggest that negative-binomial or Poisson distribution suffices if the cell population is homogenous (Kim, Tae Hyun, Xiang Zhou, and Mengjie Chen. 2020. "Demystifying 'Drop-Outs' in Single-Cell UMI Data." Genome Biology 21 (1): 196), though many tools assume zero-inflated negative-binomial model.





Top: Log-normal distribution with three rate parameters

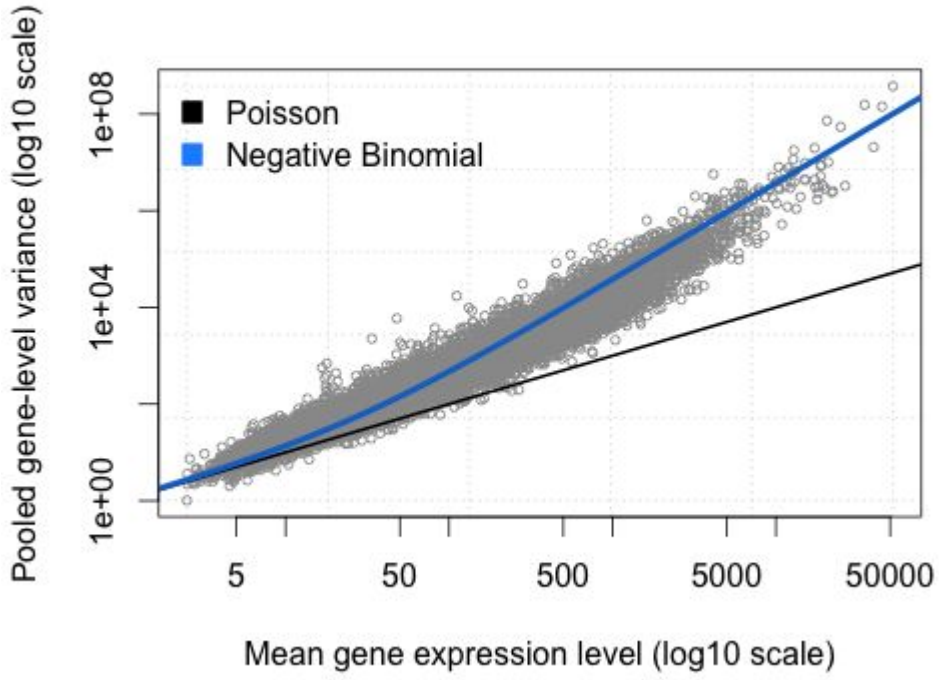
Bottom: Poisson distribution with three rate parameters

From Wikimedia, reused with the CC Attribution 3.0 license

From Poisson distribution to Negative Binomial Distribution

Two definitions of Negative-Binomial distribution

1. The number of failures seen before getting *n* successes (the inverse of *Binomial Distribution*, which the number of successes in *n* independent trials)
2. Poisson-Gamma mixture distribution, weighted mixture of *Poisson* distributions, where the rate parameter has an uncertainty modelled by a *Gamma* distribution.

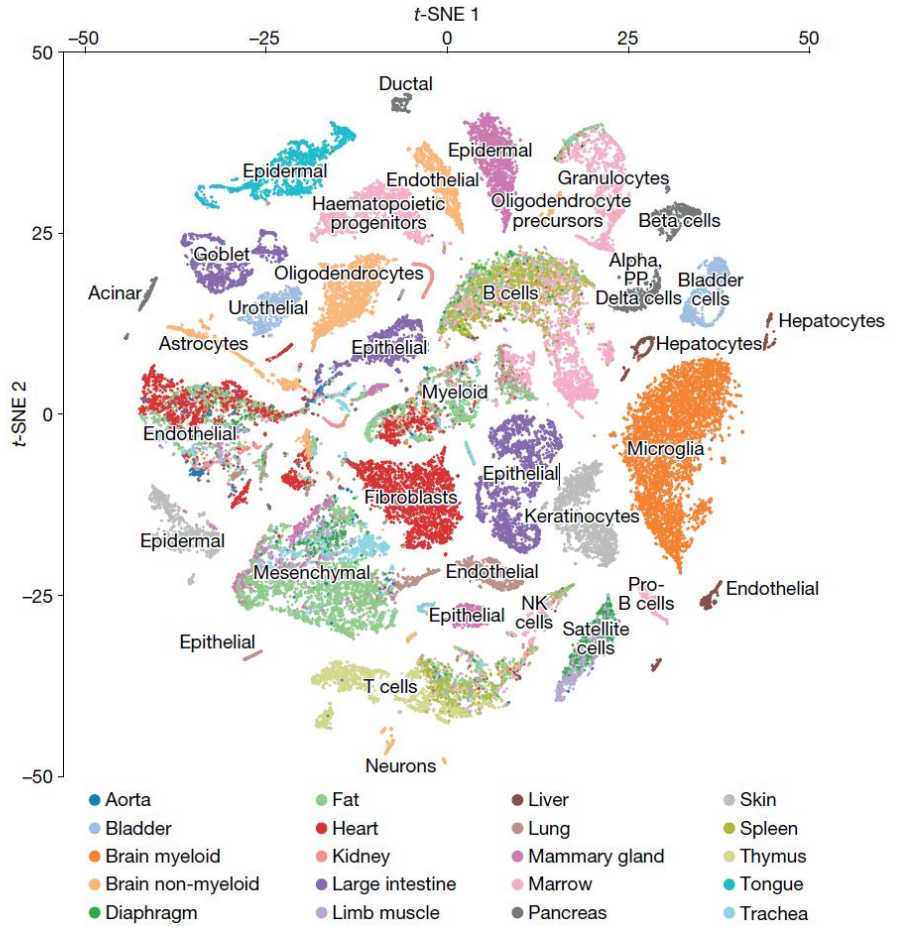


Credit of Jesse Lipp, bioramble.wordpress.com

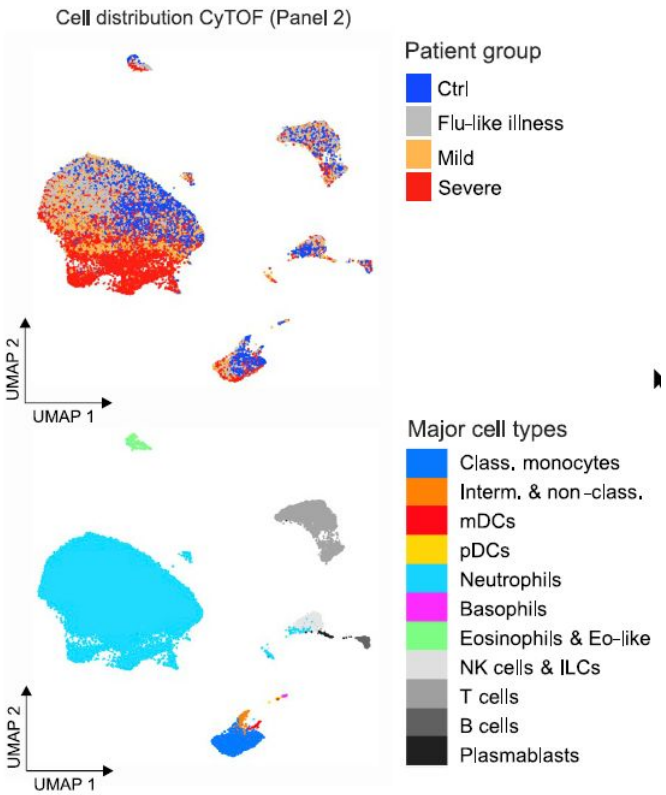
Commonly used dimensionality reduction techniques

- Principal component analysis (PCA)
- t-SNE (t-distributed Stochastic Neighbor Embedding)
- UMAP (Uniform Manifold Approximation and Projection) [A great talk by Leland McInnes, the developer of UMAP, a mathematician, Ph.D. In Profinite Lie Rings]

- For a recent overview of dimensionality reduction techniques and their applications in biology, see Nguyen, Lan Huong, und Susan Holmes. "Ten Quick Tips for Effective Dimensionality Reduction". *PLOS Computational Biology* 15, Nr. 6 (20. Juni 2019): e1006907.



The Tabula Muris Consortium. 2018. "Single-Cell Transcriptomics of 20 Mouse Organs Creates a Tabula Muris." *Nature* 562 (7727): 367.



Schulte-Schrepping, et al. 2020. "Severe COVID-19 Is Marked by a Dysregulated Myeloid Cell Compartment." *Cell* 182 (6): 1419-1440.e23.