What efficacy and safety profiles can we expect

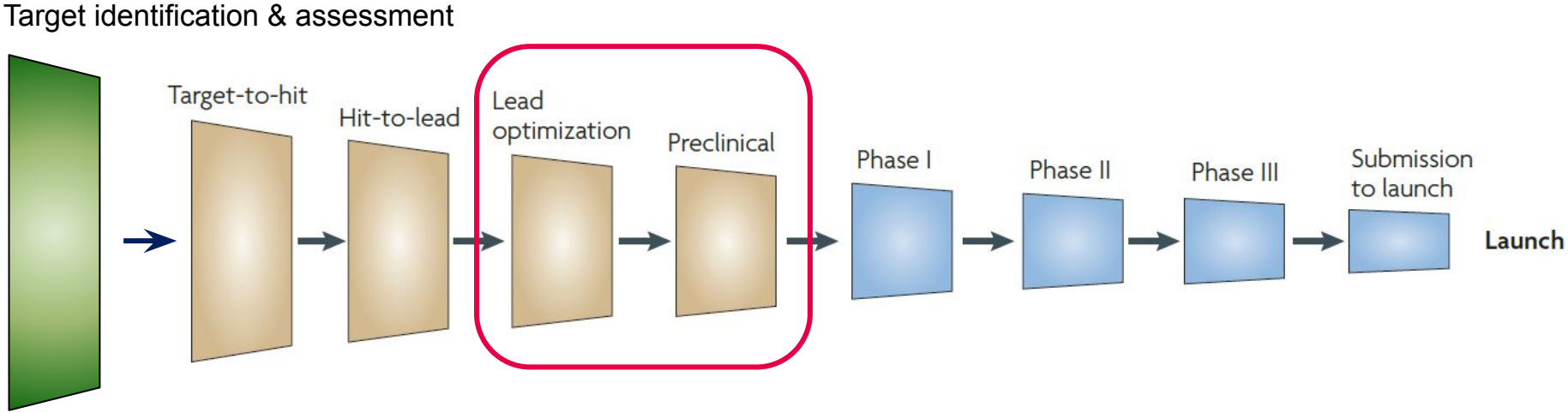
Mathematical and Computational Biology in Drug Discovery (MCBDD) Module IV

Dr. Jitao David Zhang
May 2022

Outline of Lecture 9

- Understanding pharmacology and toxicology with *in vitro*, *in vivo*, and *in silico* models
- Cell-type specific response to drugs
- Single-cell RNA sequencing for disease understanding and drug discovery

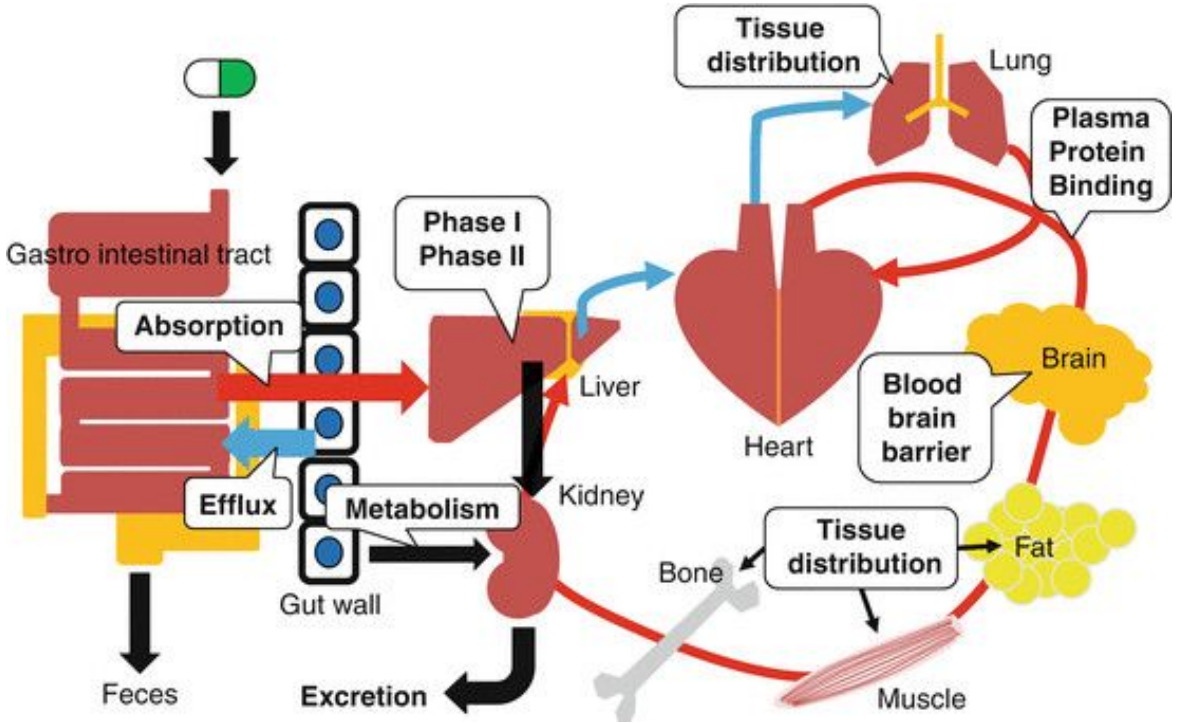
Where are we now



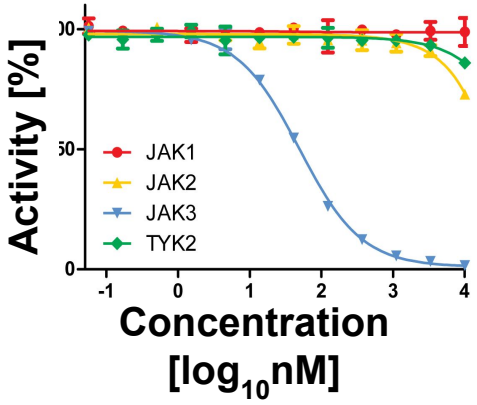
Goal: we want to select **one compound** from a few (~10^2-10^0) for entry in human.

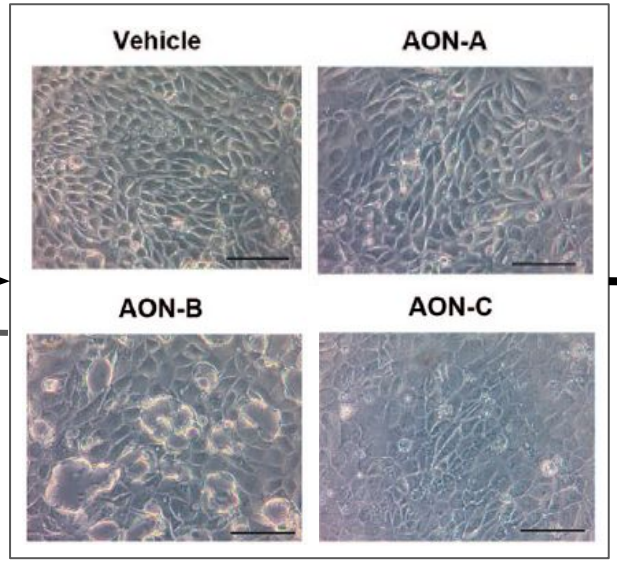
Factors that affect efficacy and safety profiles

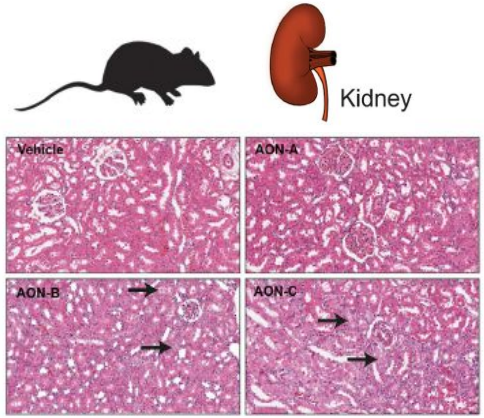
- Absorption
- Distribution
- **Pharmacology**
- **Toxicology**
- Metabolism
- Excretion



Classical workflow of efficacy and toxicity assessment







Biochemical & biophysical assays

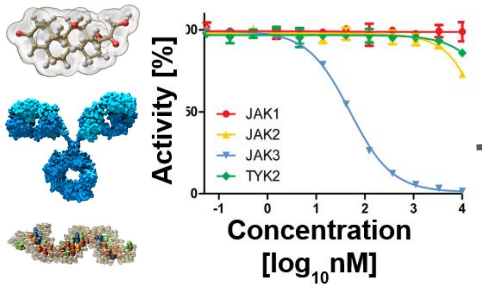
Cellular assays (*in vitro*)

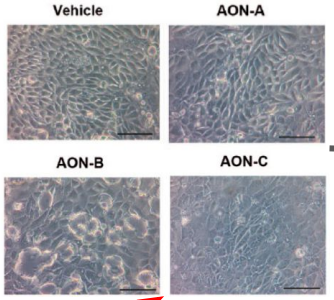
Animal experiments (*in vivo*)

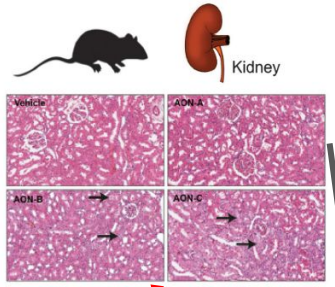
Usual workflow

Assay development

Computational methods empower efficacy and toxicity assessment

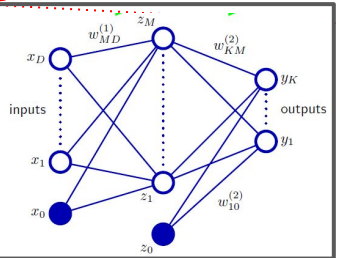






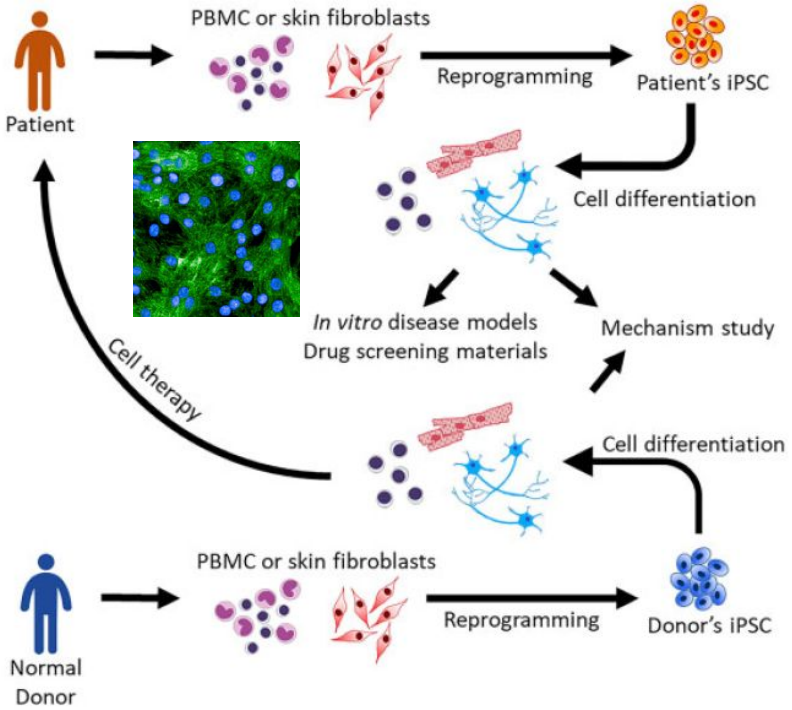
High-throughput technologies (omics, microscopy, etc.)

$$\frac{\partial u}{\partial t} = \underbrace{D\frac{\partial^2 u}{\partial x^2}}_{\text{Diffusion}} + \underbrace{ku}_{\text{Binding}}$$

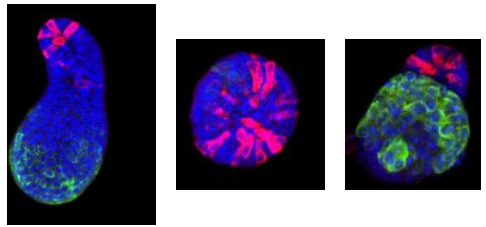


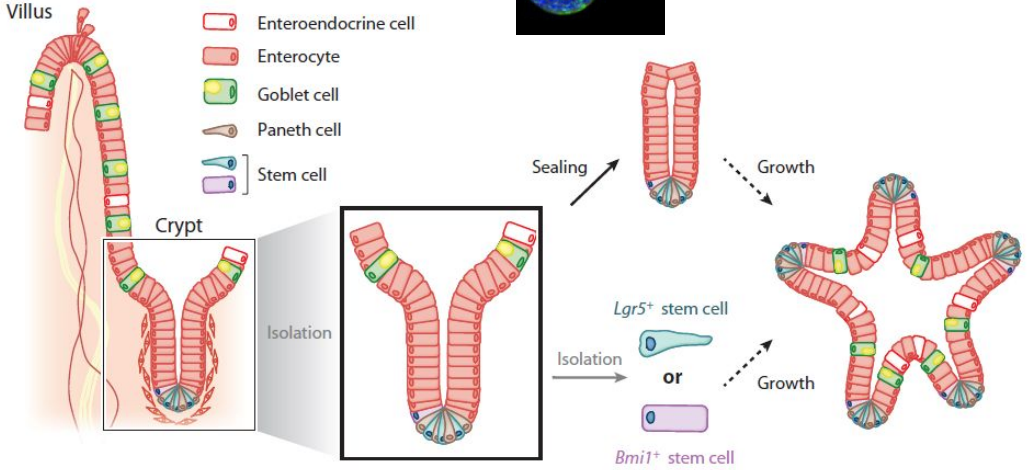
Mechanistic, causal, and statistical models

Stem cells and organoids empower efficacy and toxicity assessment



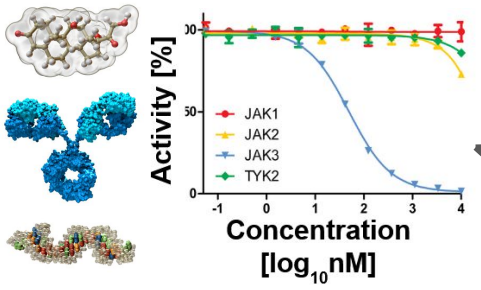
Induced pluripotent stem-cells

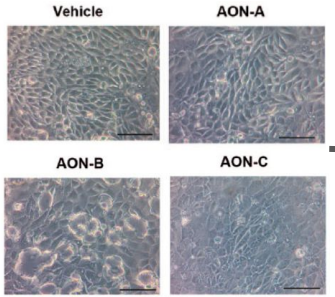


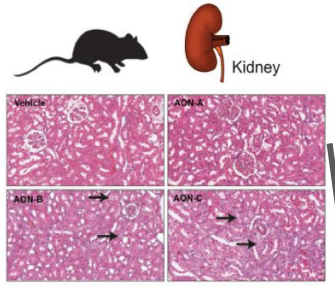


Small-intestinal organoids

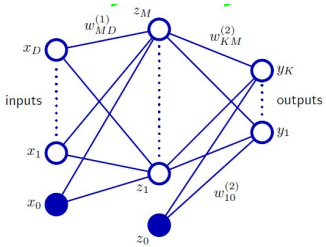
Computational methods and novel biological models empower efficacy and toxicity assessment

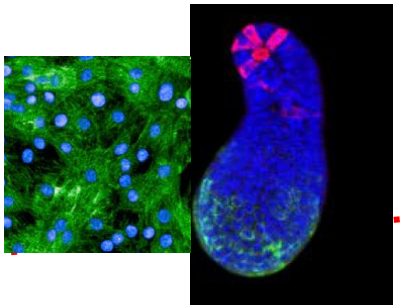




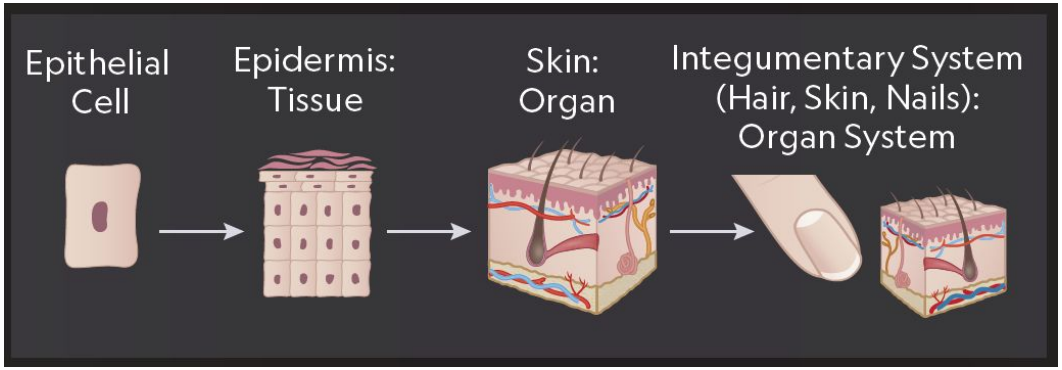


$$\frac{\partial u}{\partial t} = \underbrace{D\frac{\partial^2 u}{\partial x^2}}_{\text{Diffusion}} + \underbrace{ku}_{\text{Binding}}$$





Complexity Increases Through a System

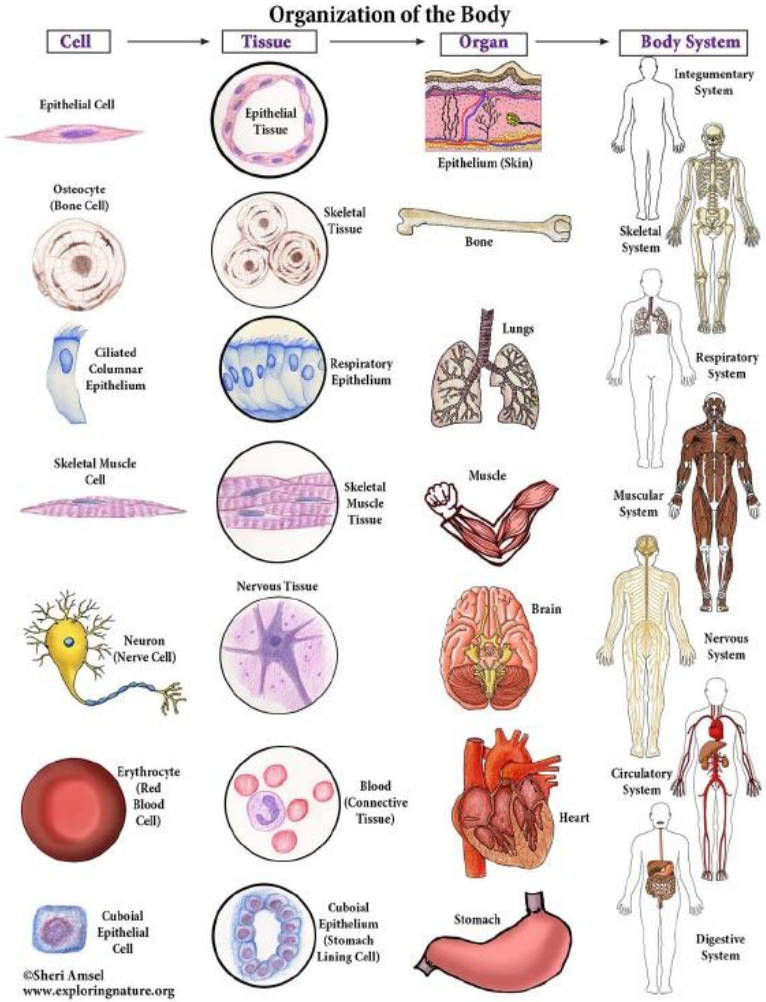


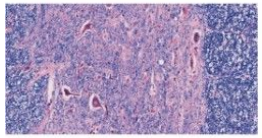
Cells: basic building blocks, variable morphologies and functions

Tissues: groups of specialized cells that communicate and collaborate

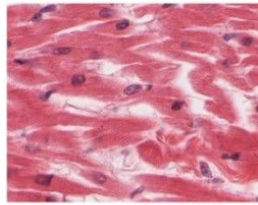
Organ: group of tissues to perform specific functions

Organ systems: group of organs and tissues



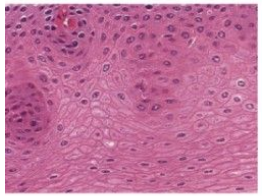


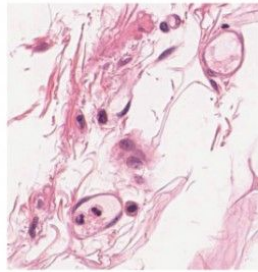
Nervous tissue
Brain
Spinal cord
Nerves



Muscle tissue
Cardiac muscle
Smooth muscle
Skeletal muscle

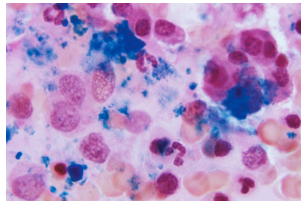
Epithelial tissue
Lining of GI tract organs and other hollow organs
Skin surface (epidermis)

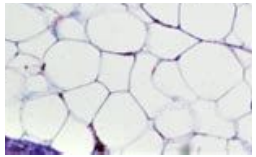


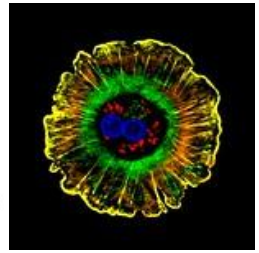


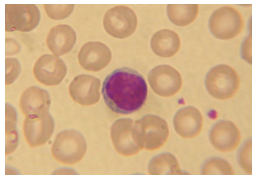
Connective tissue
Fat and other soft padding tissue
Bone
Tendon

Four major tissue types







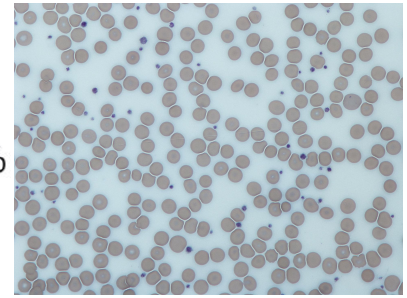


adipocytes 0.2%

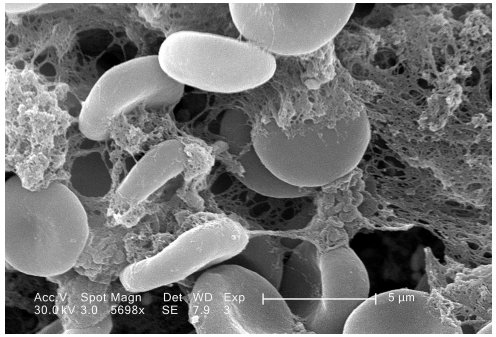
lymphocytes 1.5%

bone marrow cells 2.5%

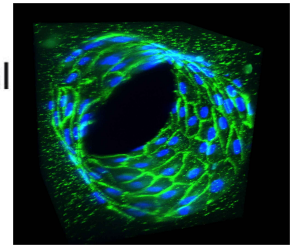
hepatocytes 0.8%



platelets 4.9%

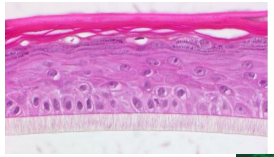


erythrocytes, 84%
25×10^{12} of 30×10^{12}
total human cell count



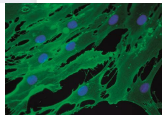
bronchial endothelial cells 0.5%

vascular endothelial cells 2.1%



epidermal cells 0.5%

respiratory interstitial cells 0.5%

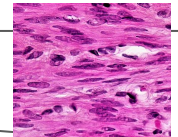


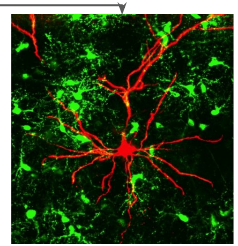
dermal fibroblasts 0.1%

muscle cells 0.001%

other 2.0%

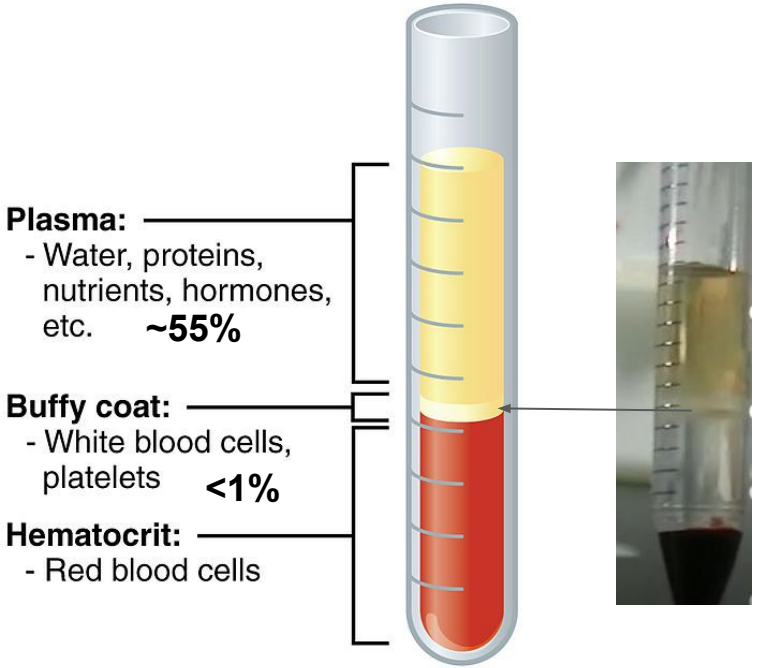
neurons and glia 0.6%

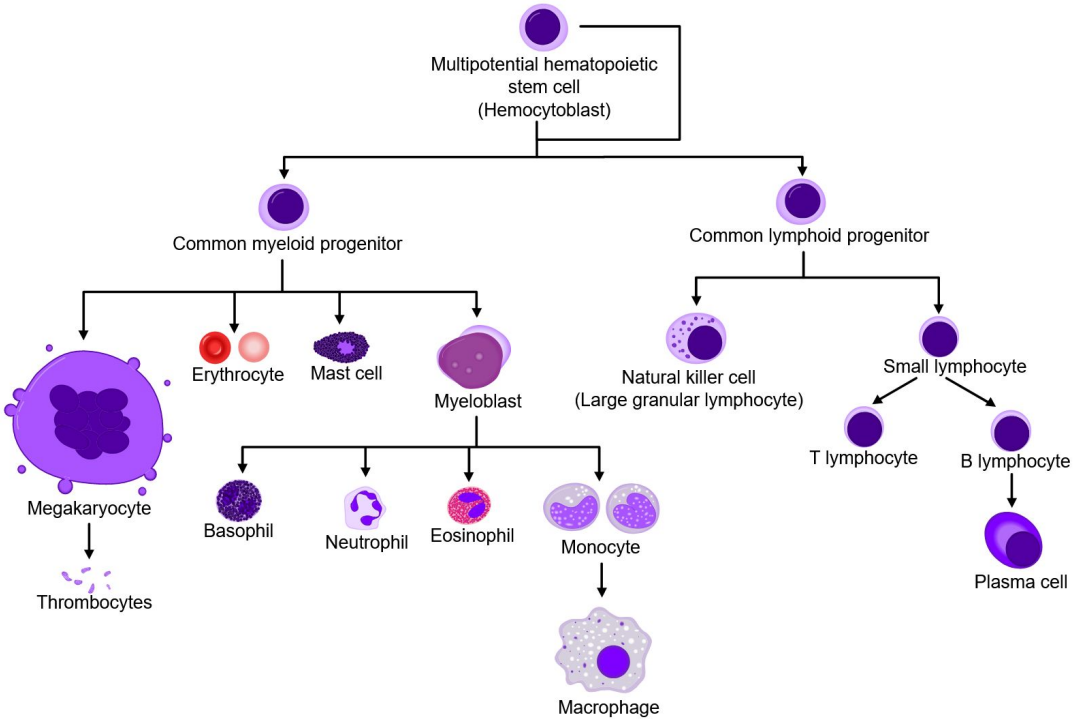




UNI BASEL

What's in a drop of blood? Ask a doctor or a biologist!

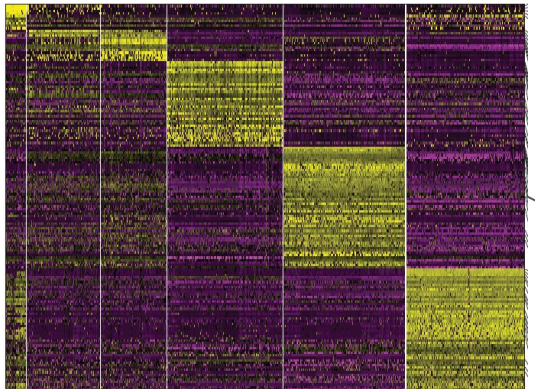




What's in a drop of blood? Count the genes!

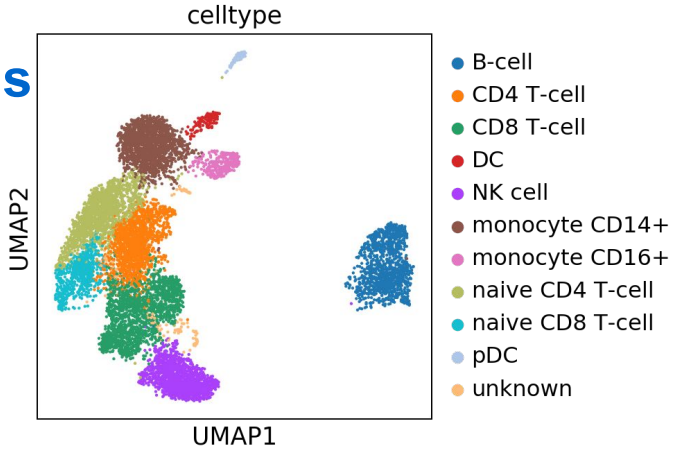
Sequencing

Genes



Cells

Data analysis



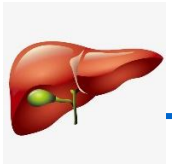
Low Expression

High Expression

Single-cell sequencing (scSeq) workflow

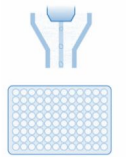
Tissue dissociation

Single cell capture and transcriptome sequencing



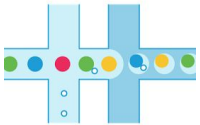
Solid tissue

Dissociated cells/nuclei



Flow cytometry 96-384 cells

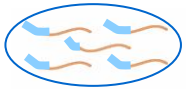
Cell/nuclei suspension



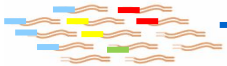
Droplet microfluidics 1000s of cells

Throughput

Single cells or nuclei



cDNA



Pooled cDNA libraries

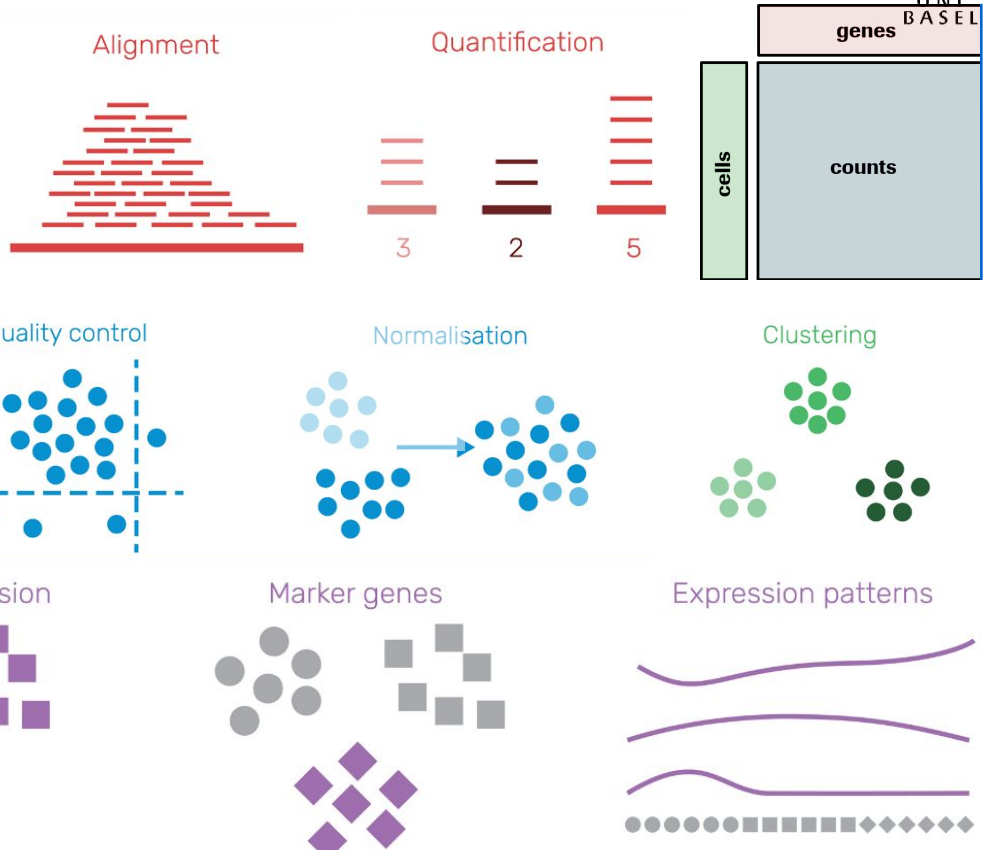
Next-Generation Sequencing

A linearized workflow of scSeq data analysis

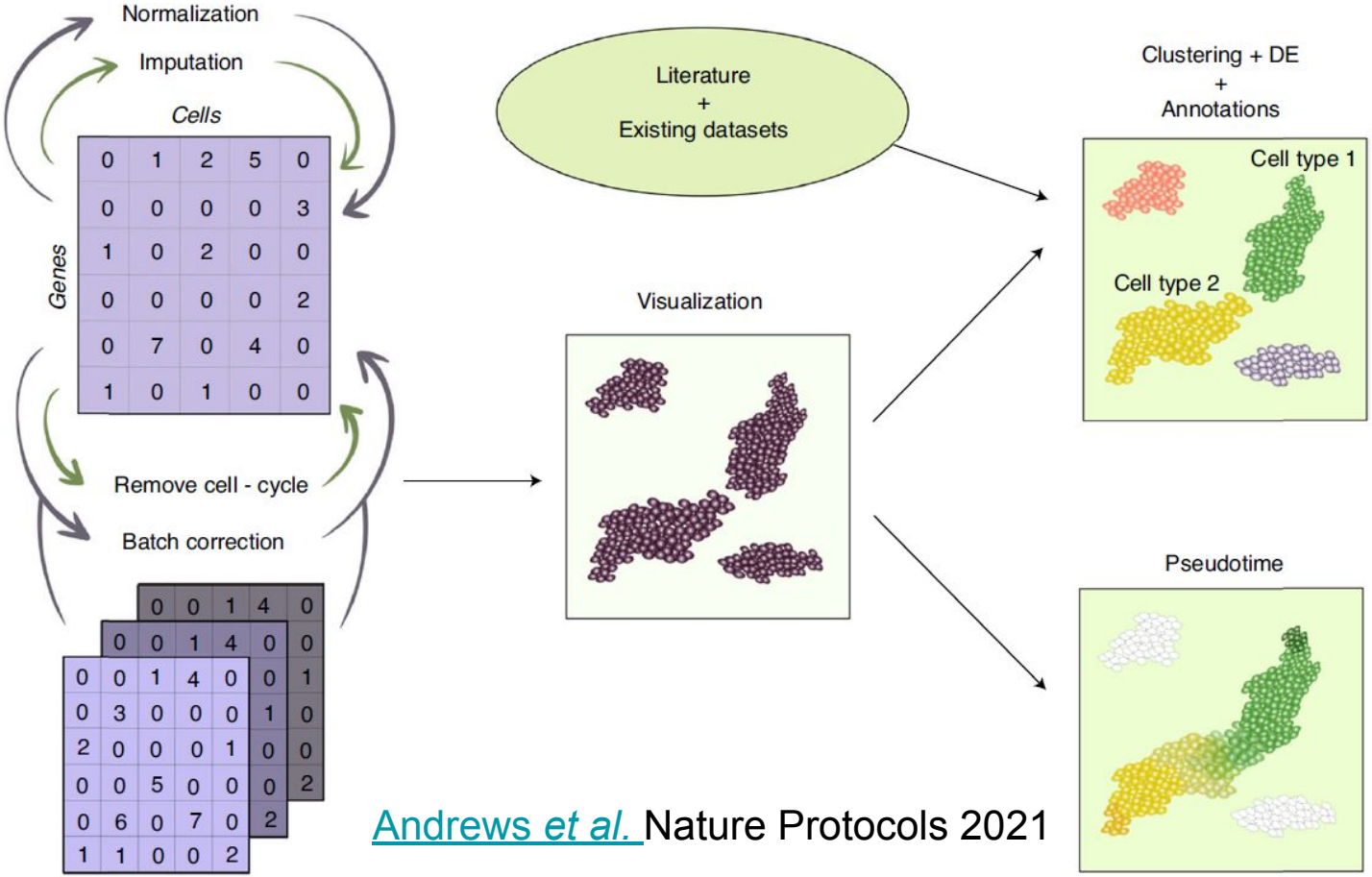
From short reads to gene-cell matrix

QC, filtering & normalization, dimensionality reduction, and clustering

Downstream analysis

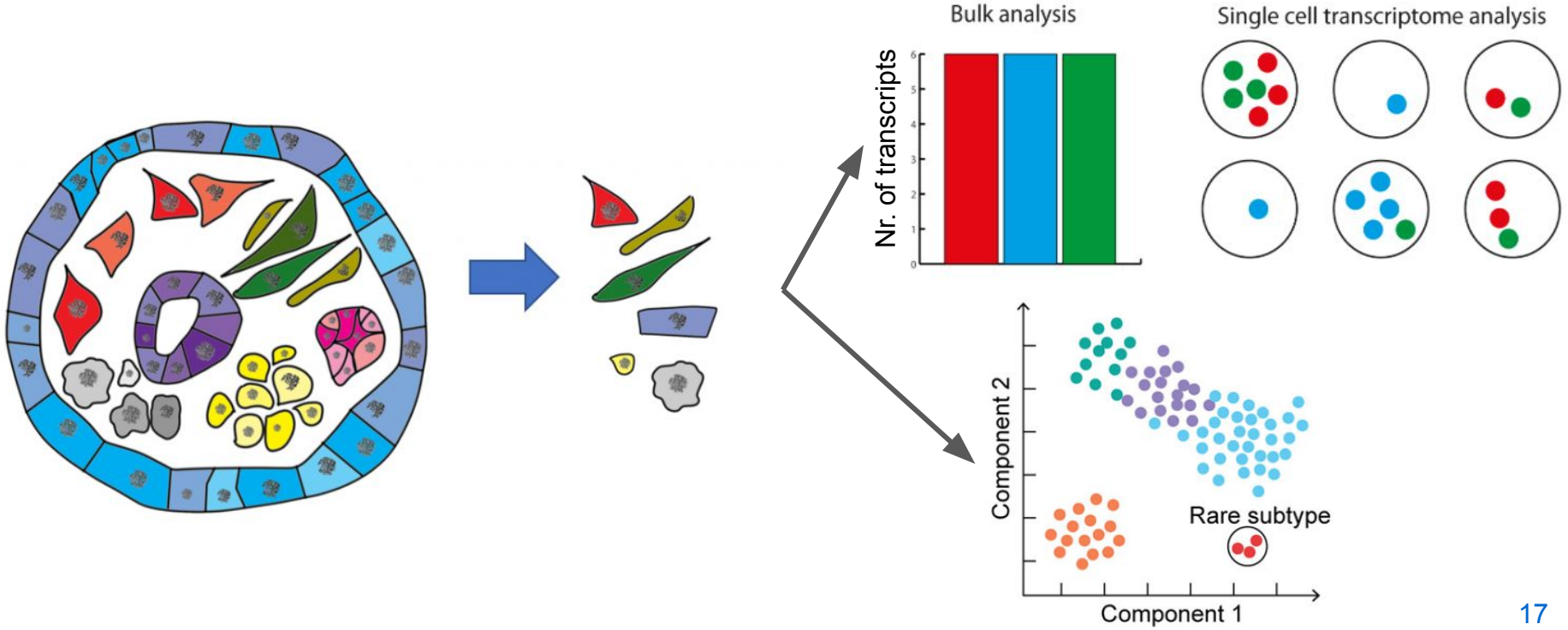


Overview of the computational workflow



Andrews *et al.* Nature Protocols 2021

Single-cell biology benefits both disease understanding and drug discovery

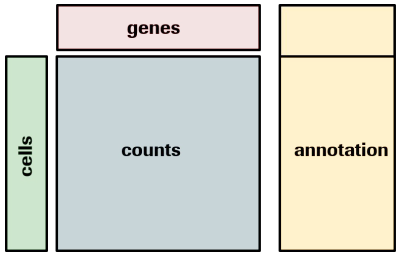


BESCA: An open-source Python package for single-cell gene expression analysis

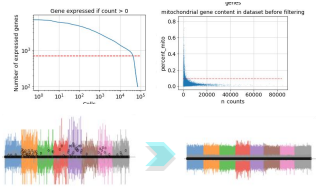
An automatized standard workflow

Data loading → Preprocessing → Clustering → Visualization → Analysis

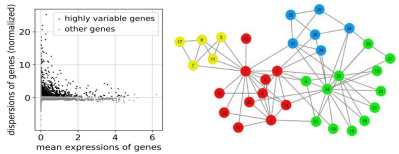
From FAIR format



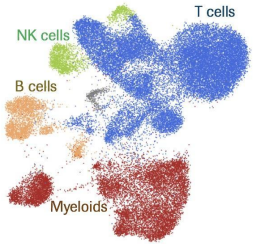
Quality Control
Filtering,
Normalization



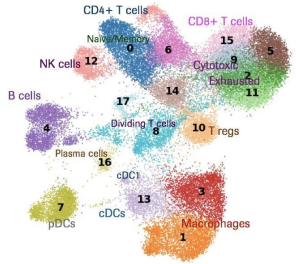
Dimensionality
reduction
Clustering



UMAP



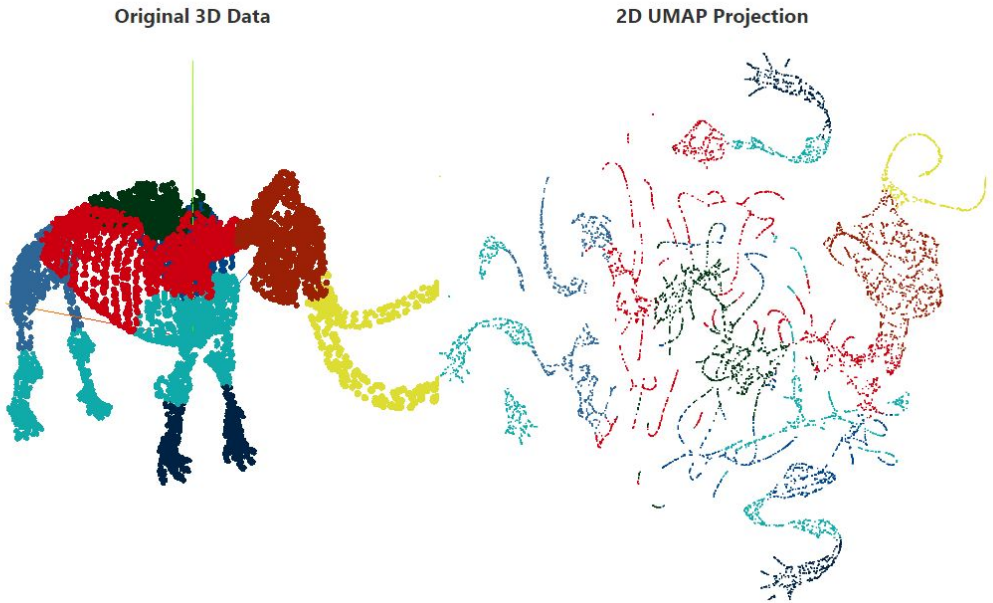
Cell type
characterisation

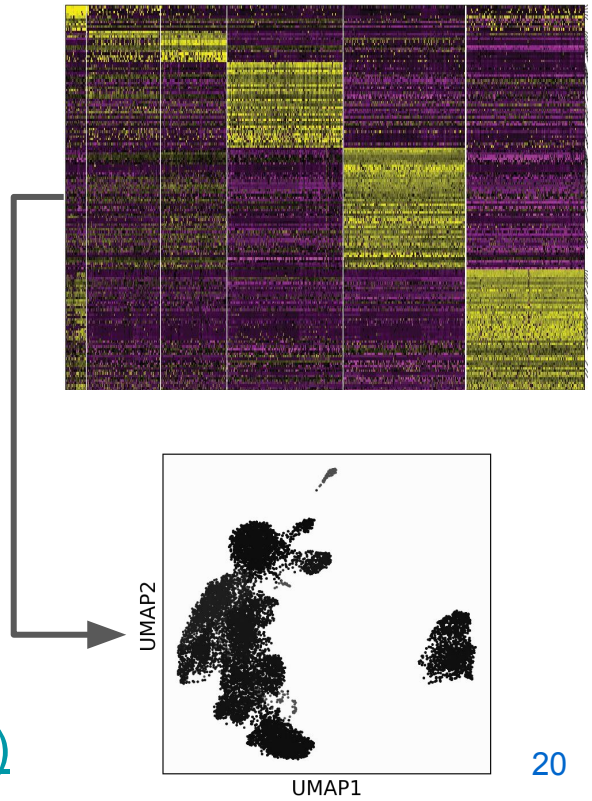


How to represent voxels with pixels?

The elephant bull *Tusker* at Zolli Basel plays with a tree trunk on a post (2022)

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction

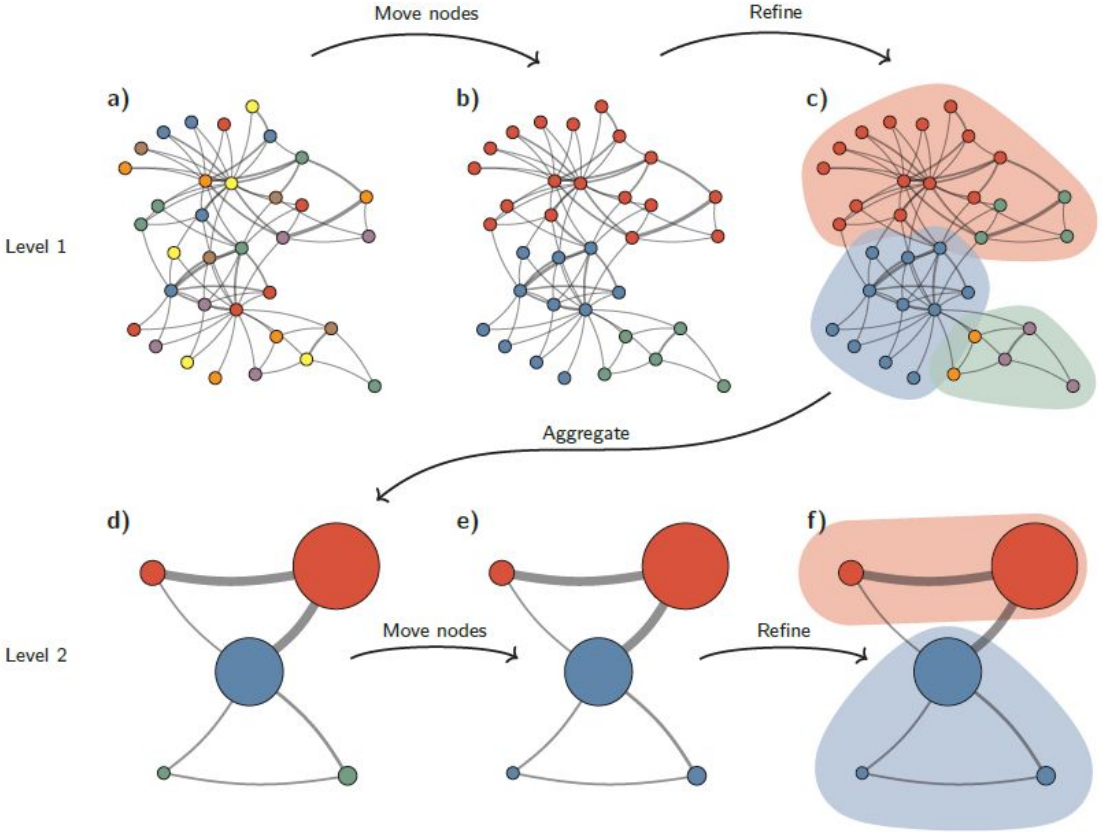


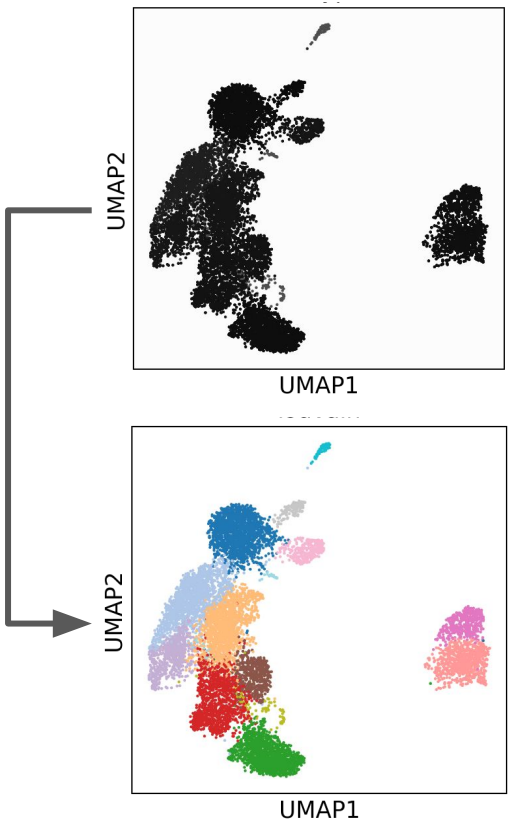


Understanding UMAP by A. Coenen and A. Pearce

UMAP by Leland McInnes on SciPy 2018 (YouTube)

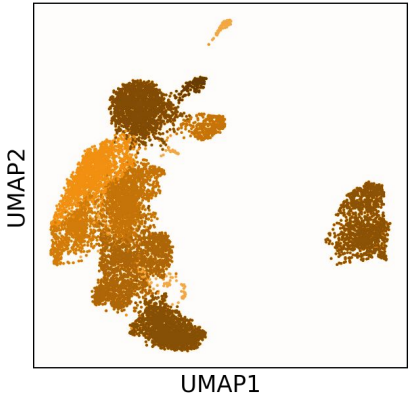
The Leiden Algorithm for Community Detection



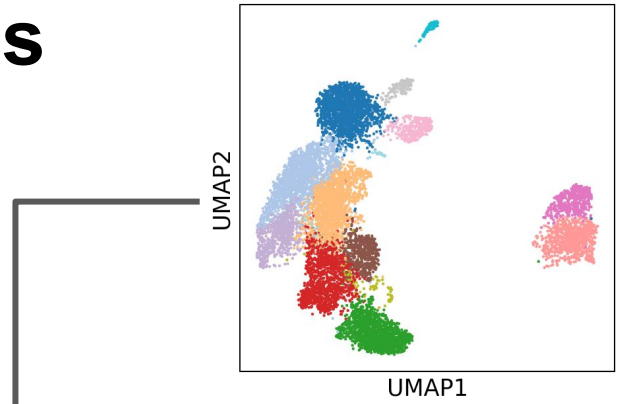


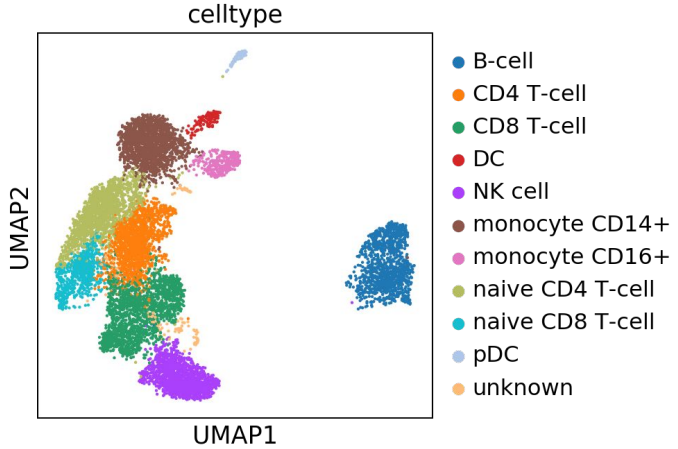
Biological knowledge and visual inspection is used to annotate cell types

Heatmap of gene X

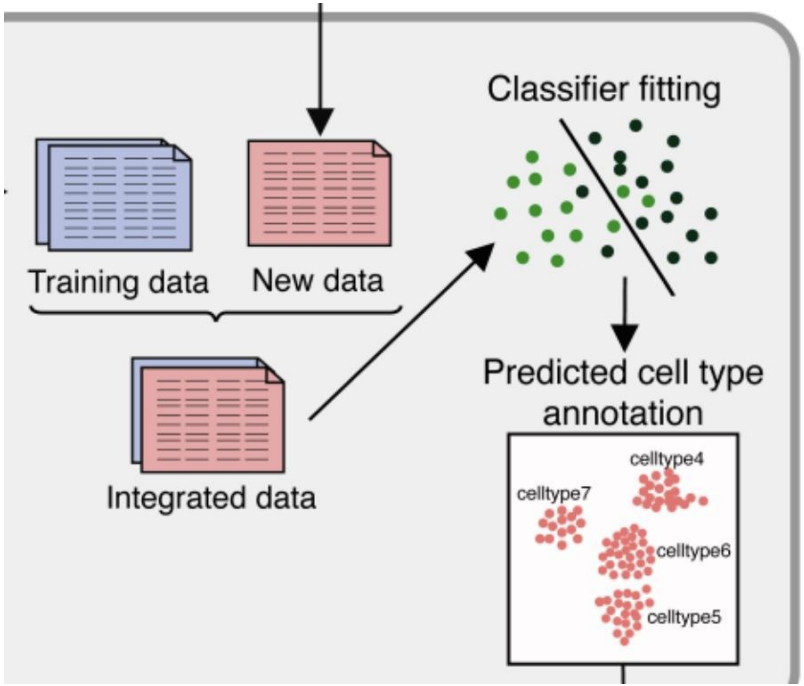


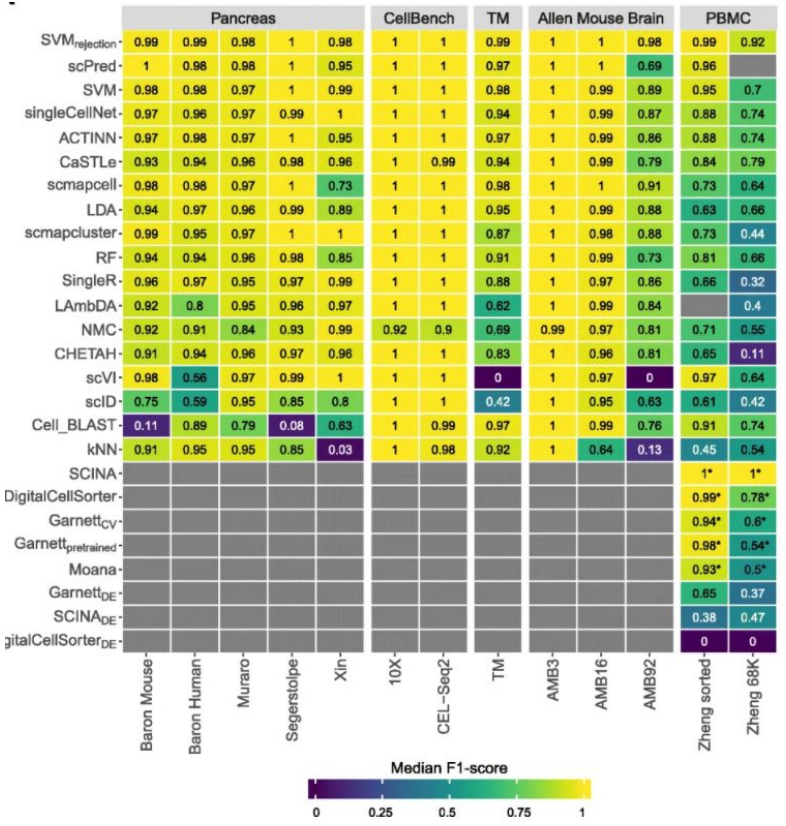
lymphocyte	PTPRC							
myeloid	S100A8	S100A9	CST3					
Bcell	CD19	CD79A	MS4A1					
Tcells	CD3E	CD3G	CD3D					
CD4	CD4							
CD8	CD8A	CD8B						
NKcell	NKG7	GNLY	NCAM1					
monocyte	CST3	CSF1R	ITGAM	CD14	FCGR3A	FCGR3B		
macrophage	CD14	IL1B	LYZ	CD163	ITGAX	CD68	CSF1R	FCGR3A





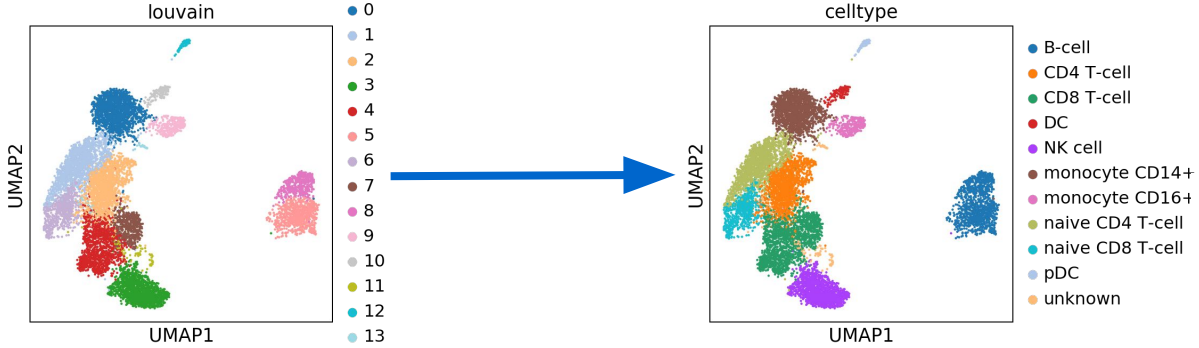
Cell type annotation with machine learning





An intern project: Cell type annotation

From unsupervised clustering and cluster based annotation



Luis Wyss
RAAN intern 2019

	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Label
Training Cell 1	10	50	0	12	4	Celltype A
Training Cell 2	8	45	78	3	23	Celltype B
Training Cell 3	14	55	78	65	55	Celltype B
Training Cell 4	78	12	13	9	58	Celltype A
Training Cell 5	45	23	65	98	11	Celltype C

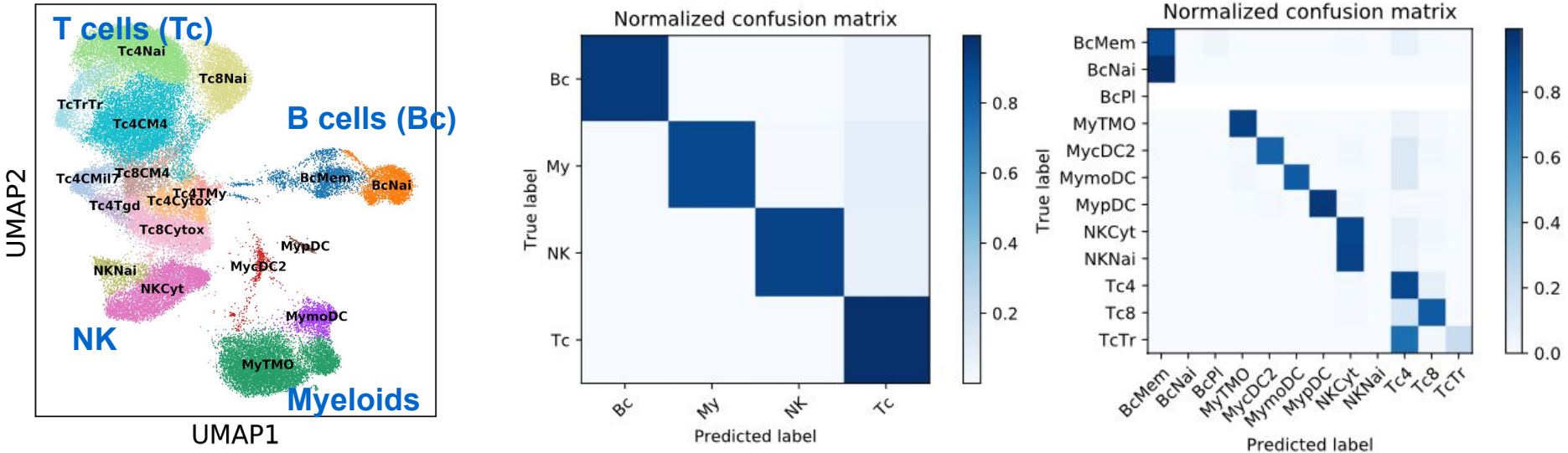
To supervised annotation at single-cell level:

	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5
Cell 1	45	45	8	56	3
Cell 2	65	120	78	45	12
Cell 3	79	12	34	65	88
Cell 4	7	59	32	47	62

	Gene 1	Gene 2	Gene 3	Gene 4	Gene 5	Prediction
Cell 1	45	45	8	56	3	Celltype A
Cell 2	65	120	78	45	12	Celltype B
Cell 3	79	12	34	65	88	Celltype C
Cell 4	7	59	32	47	62	Celltype B

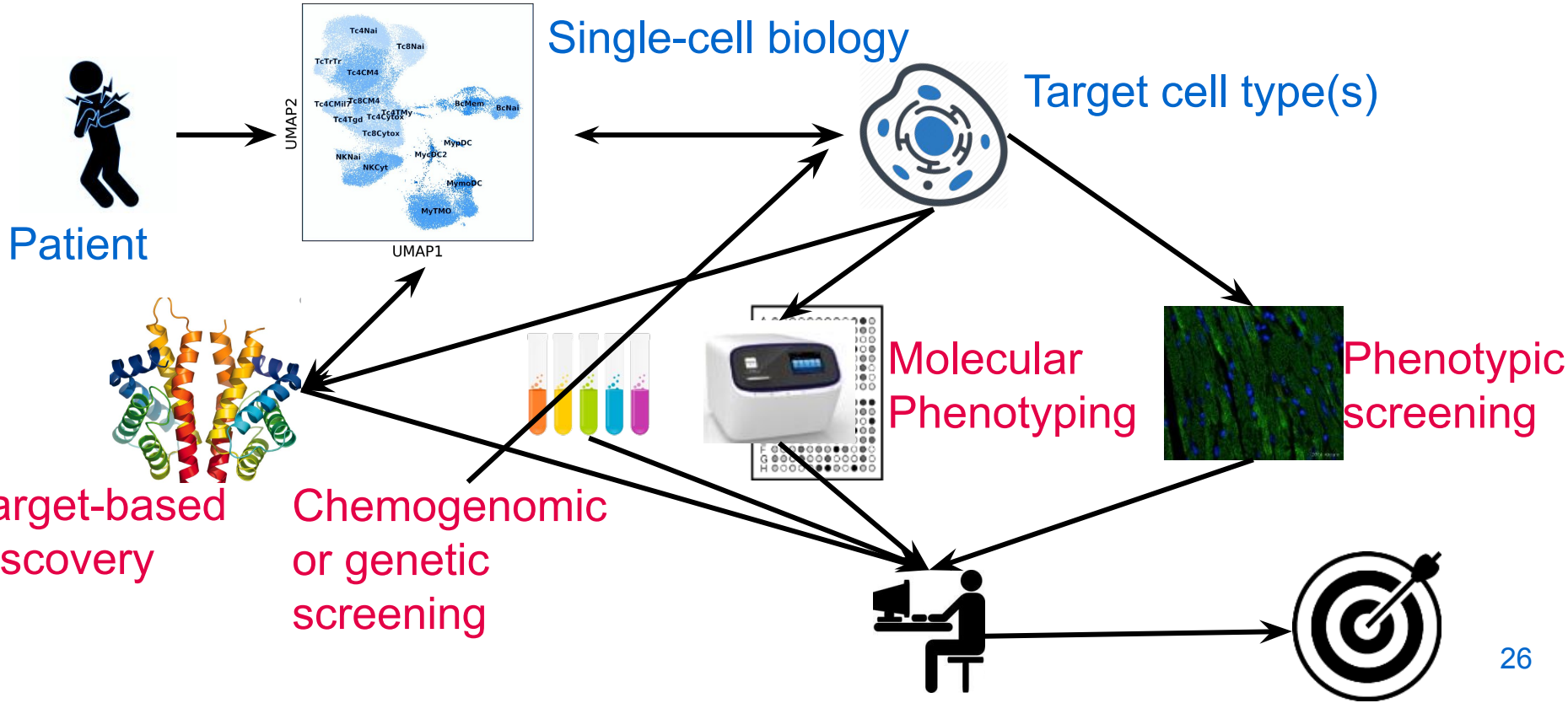
Advantages: (1) automation, (2) annotation independent from clustering, and (3) we can estimate the confidence of prediction

A PBMC example of cell type annotation



- Broad level cell types, including B cells (Bc), Myeloid (My), NK cells (NK) and T cells (Tc), are successfully predicted.
- Missing and highly similar cell types cause challenges with increased granularity. Essential: reference data quality and knowledge of cell types.

Computational biologists work with experimentalists to empower drug discovery



We are living ecosystems

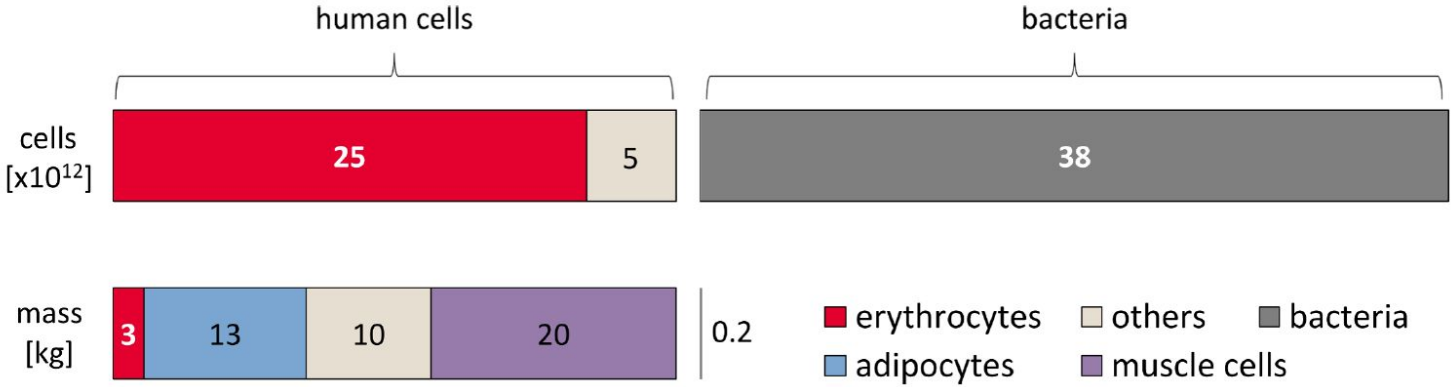
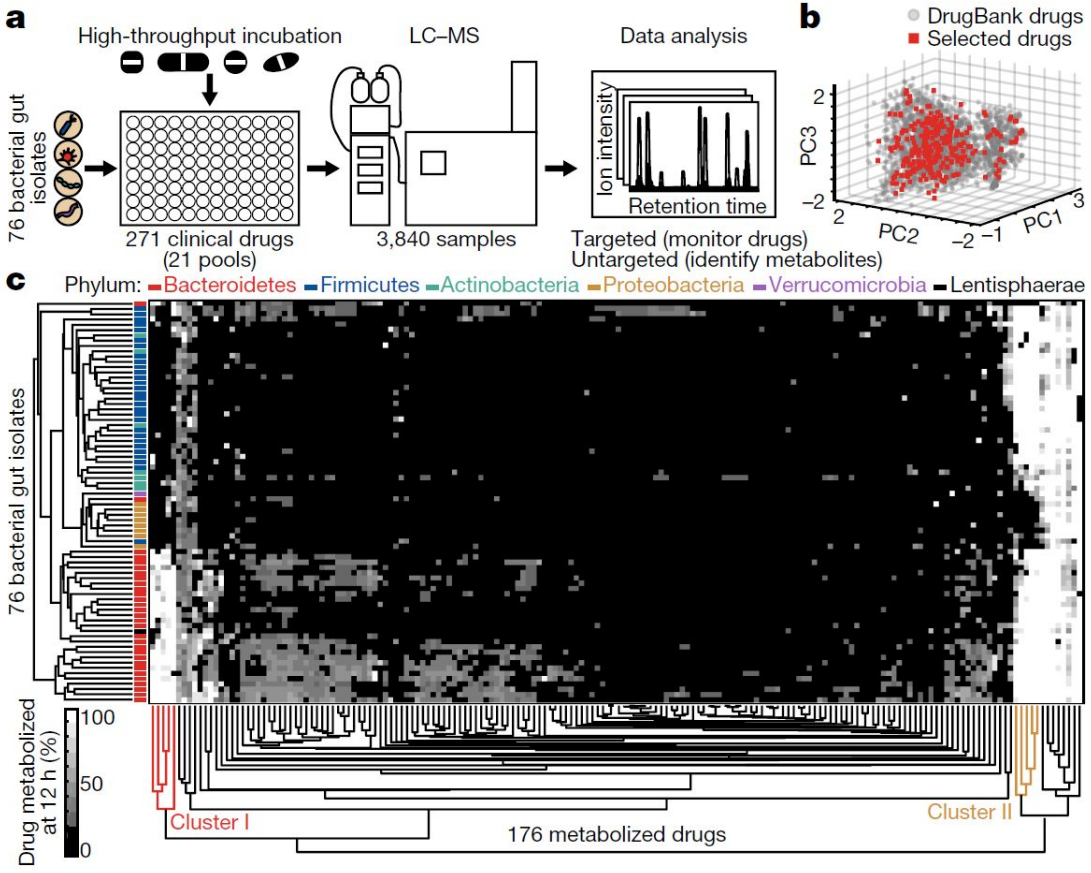
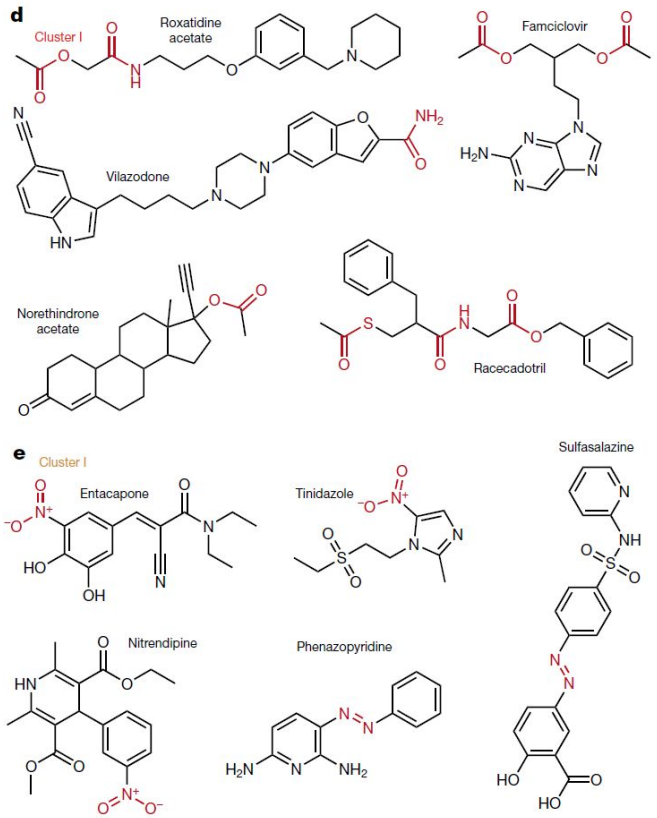


Table 3. B/H ratio for different population. See Table B in S1 Appendix for full references.

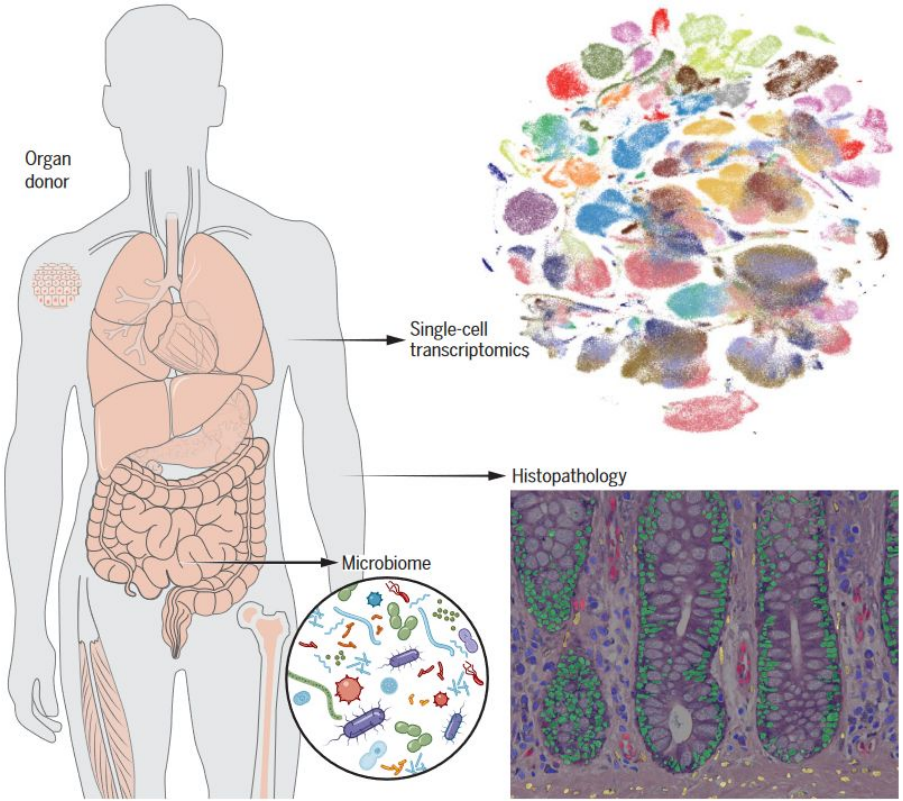
population segment	body weight [kg]	age [y]	blood volume [L]	RBC count [10^{12}/L]	colon content [g]	bac. conc. [10^{11}/ g wet] (1)	total human cells [10^{12}] (2)	total bacteria [10^{12}]	B:H
ref. man	70	20–30	4.9	5.0	420	0.92	30	38	**1.3**
ref. woman	63		3.9	4.5	480	0.92	21	44	**2.2**
young infant	4.4	4 weeks	0.4	3.8	48	0.92	1.9	4.4	**2.3**
infant	9.6	1	0.8	4.5	80	0.92	4	7	**1.7**
elder	70	66	3.8 (3)	4.8	420	0.92	22	38	**1.8**
obese	140		6.7	5.0 (4)	610 (5)	0.92	40	56	**1.4**

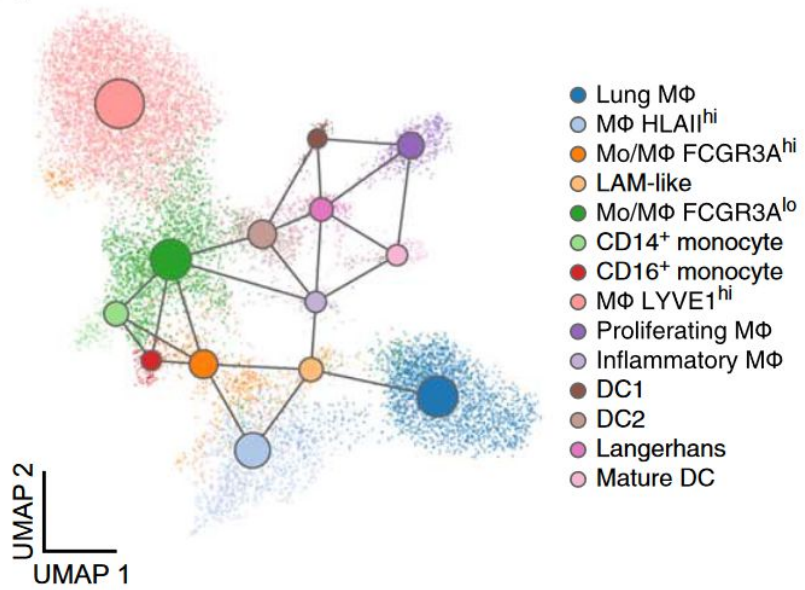
Gut microbiome can metabolize drugs differently





The *Tabula Sapiens* and other community projects offer reference expression data in healthy donors





Left: the *Tabula Sapiens*. Right: Myeloid (Mϕ=macrophages, Mo/monocytes, LAM=lipid-associated macrophages, DC=dendritic cells) gene expression

Conclusions

- Single-cell biology can identify rare cell populations associated with diseases, and investigate cell-type-specific perturbations caused by drug candidates.
- Algorithms for dimensionality reduction, clustering, and semi-automated cell type annotation allow us interpret and integrate single-cell datasets.

Offline activities of Module IV (optional)

Perform your own single-cell data analysis to get first-hand experience working with high-dimensional biological data.

- If you are new to the topic, please use the PBMC tutorial of *Scanpy* (python) or the PBMC tutorial of Seurat (R).
- If you have experience with such data already, checkout the NBIS workshop on single-cell sequencing data analysis to cover advanced topics such as spatial transcriptomics and trajectory inference.

Single-cell biology is important in drug discovery

Disease understanding: disease-specific cell types and states

Biomarker and patient stratification: which genes should we measure in which cell type(s)?

Target identification: expression pattern in health and disease across cell types

MoA and safety modelling: perturbation effect at single-cell level

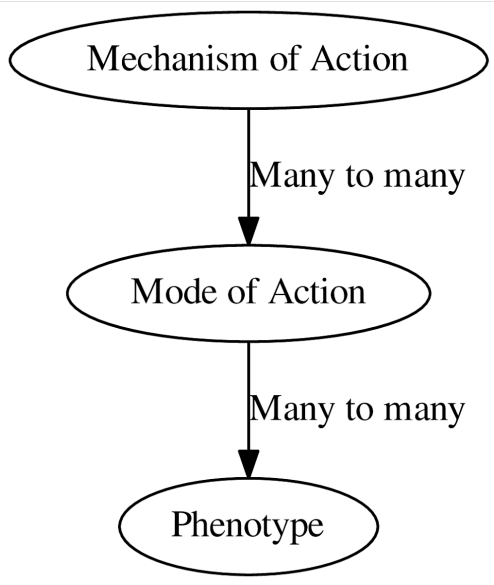
Outline of Lecture 10

- We predict efficacy and safety profiles of drugs by studying the mechanism and mode of action (MoA).
- Molecular modelling and (single-cell) RNA sequencing analysis are essential tools for understanding MoA of nucleotide-based modalities.
- Molecular modelling and proteomics based on mass spectrometry (MS) are essential tools for understanding MoA of small molecules and antibodies.

Mechanism of Action at the molecular level and *Mode of Action* at cellular and system levels

Mechanism of Action: The biochemical interactions through which a drug exerts its pharmacology and toxicity.

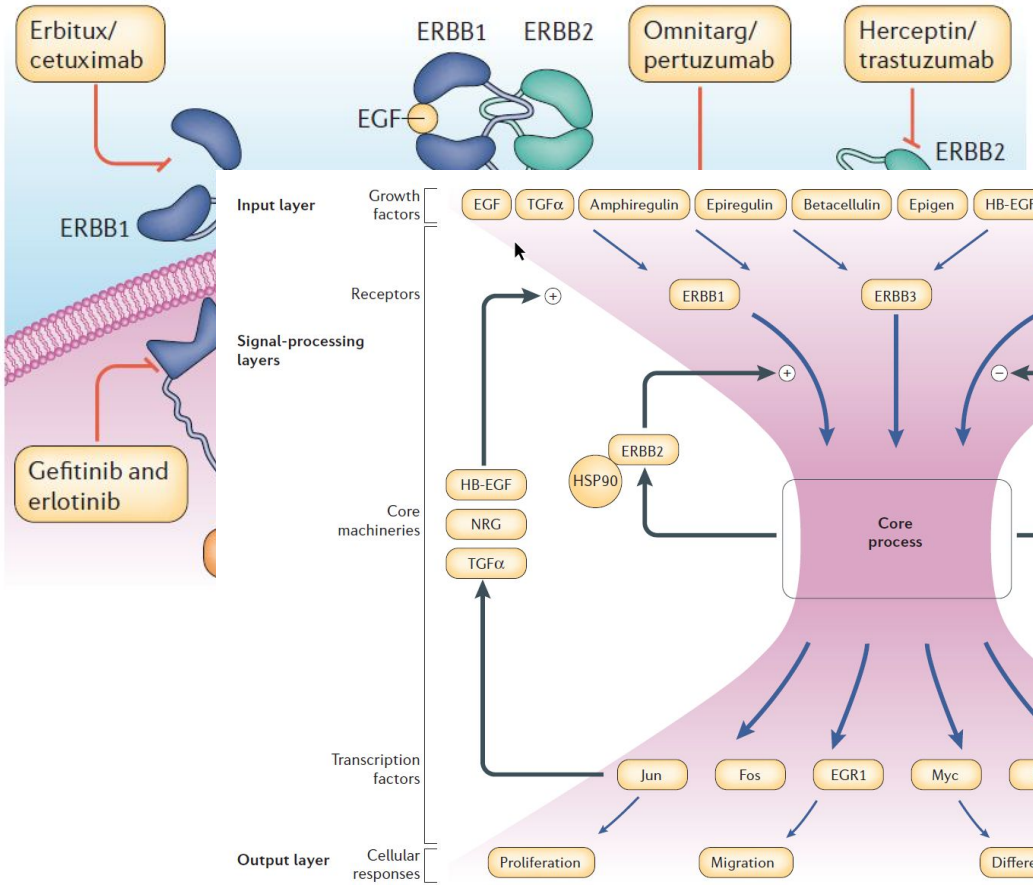
Mode of Action: Functional or anatomical changes of cells, or organ and tissue systems resulting from the exposure to a drug.



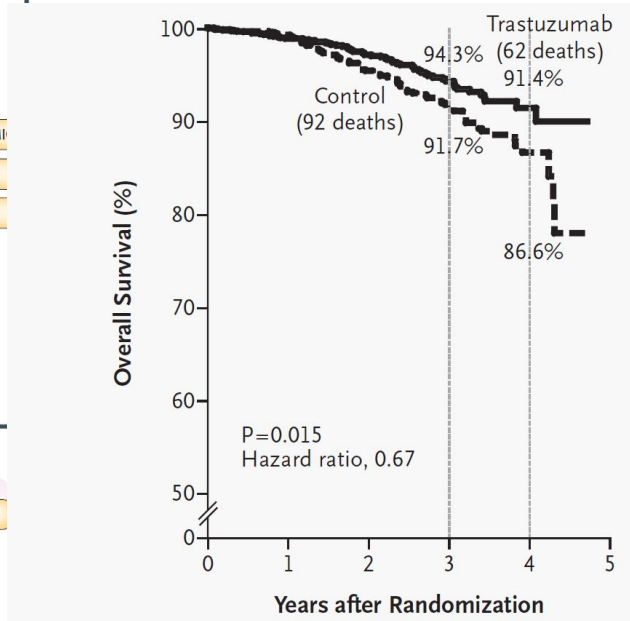
Mechanism of Action

UNI BASEL

Mode of Action



Phenotype

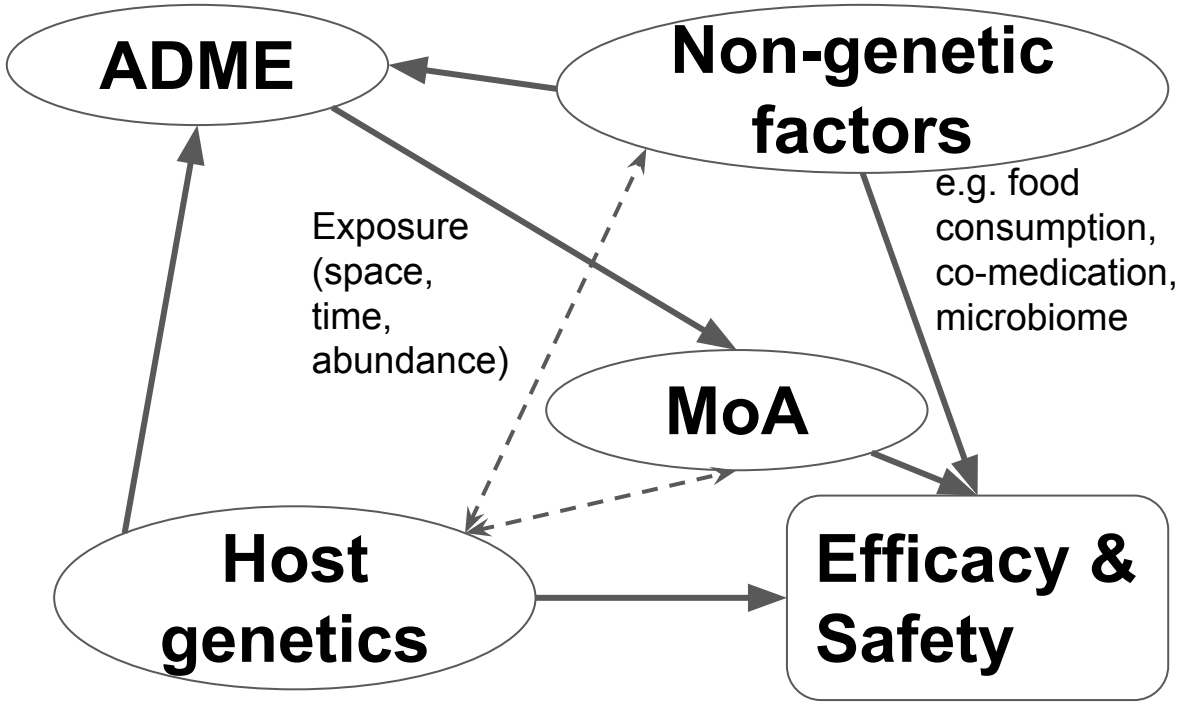


No. at Risk	3351	2441	1571	908	165	0
Control	1679	1200	766	448	83	0
Trastuzumab	1672	1241	805	460	82	0

Mechanism and Mode of Action determine phenotypes

Host genetics, non-genetic factors, MoA, and ADME together influence efficacy and safety

In this lecture, *MoA* refers to both Mechanism and Mode of action, because we need to understand **both** to make a good drug.



Four approaches for MoA understanding

- Imaging-based methods
- Direct biochemical methods

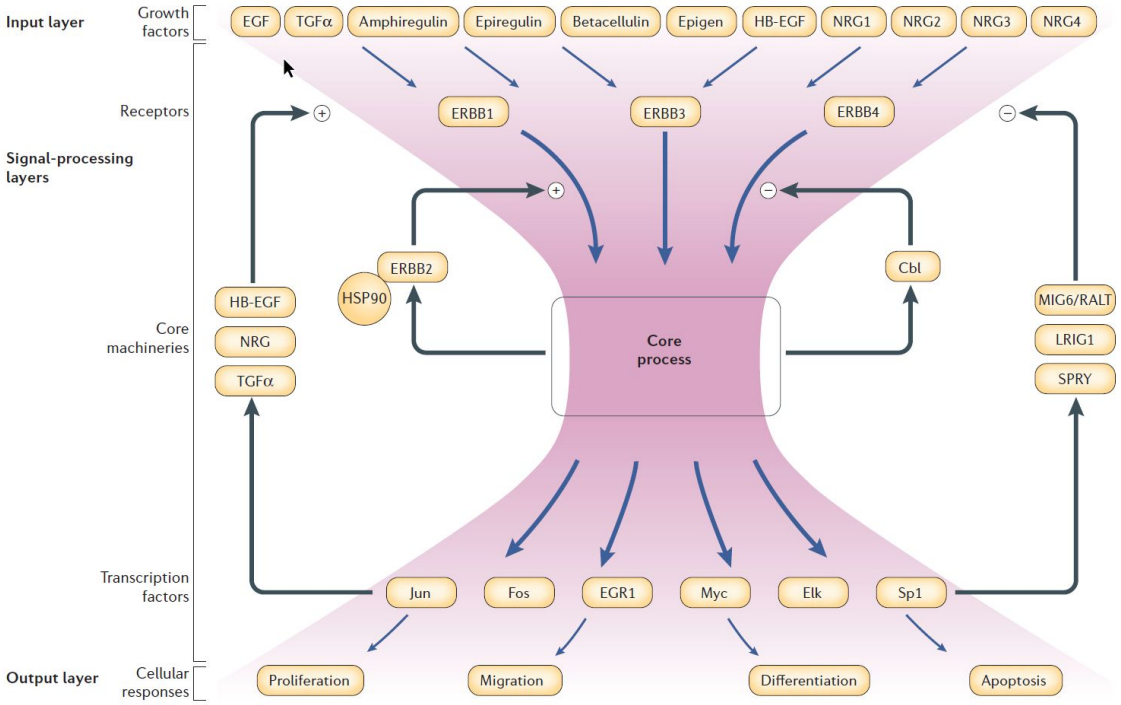
} Covered before

- **Computer-assisted inference methods**, *e.g.* sequence analysis and molecular modelling
- **Omics based methods**, *e.g.* transcriptomics (RNA-seq) and **proteomics (mass spectrometry)**

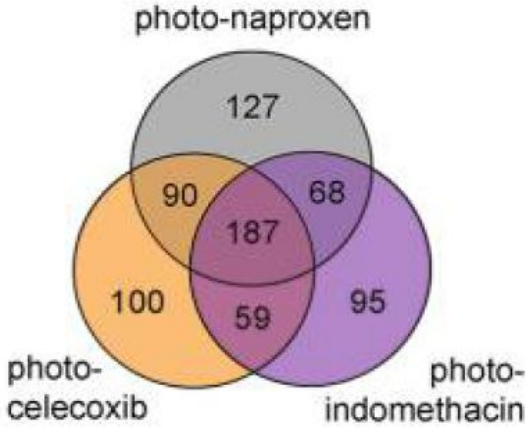
} **Focus today**

Challenge #1: Many Causes, Same Effect

Different Mechanisms of Action *may* or *may not* lead to the same Mode of Action.



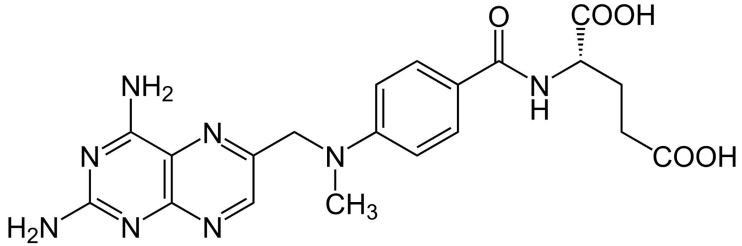
Challenge #2: Multiple MoAs are possible



Non-steroid anti-inflammatory drugs (NSAIDs) are thought to work by inhibiting proteins Cox-1 and Cox-2.

A recent study (Gao *et al.* 2018) reports that they bind to a surprisingly high number of proteins in cells.

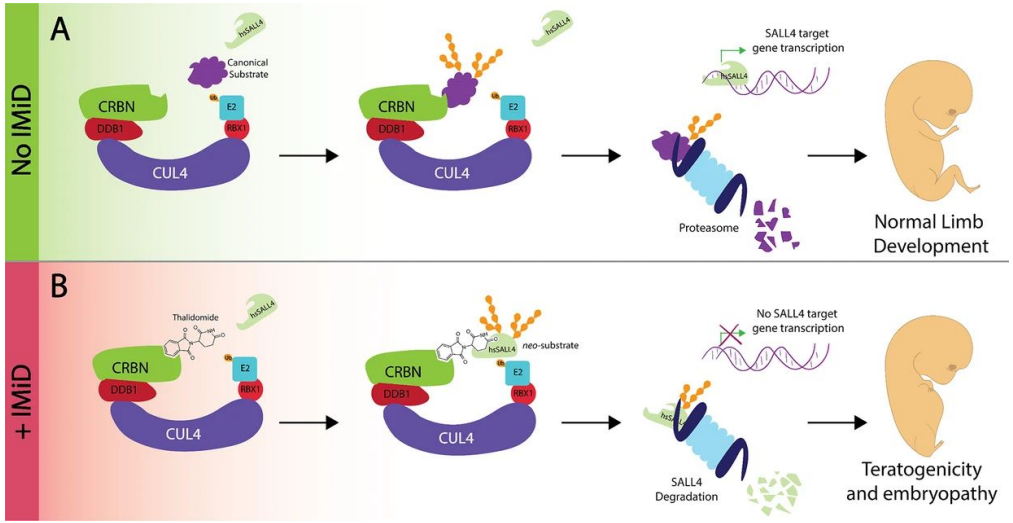
Methotrexate



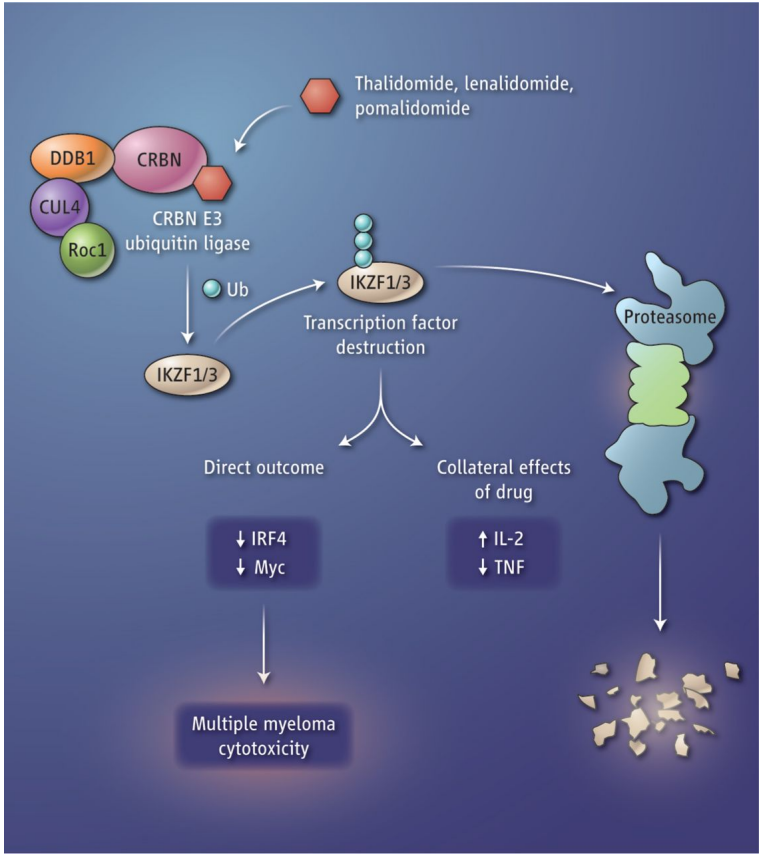
[As chemotherapy agent] Inhibiting dihydrofolate reductase (DHFR) and consequently DNA synthesis.

[As immunosuppressant] Multiple mechanisms, *e.g.* (1) inhibiting purine metabolism, (2) inhibiting methyltransferase, and (3) inhibiting IL-1b binding to its receptor.

Challenge #2: Multiple MoAs are possible

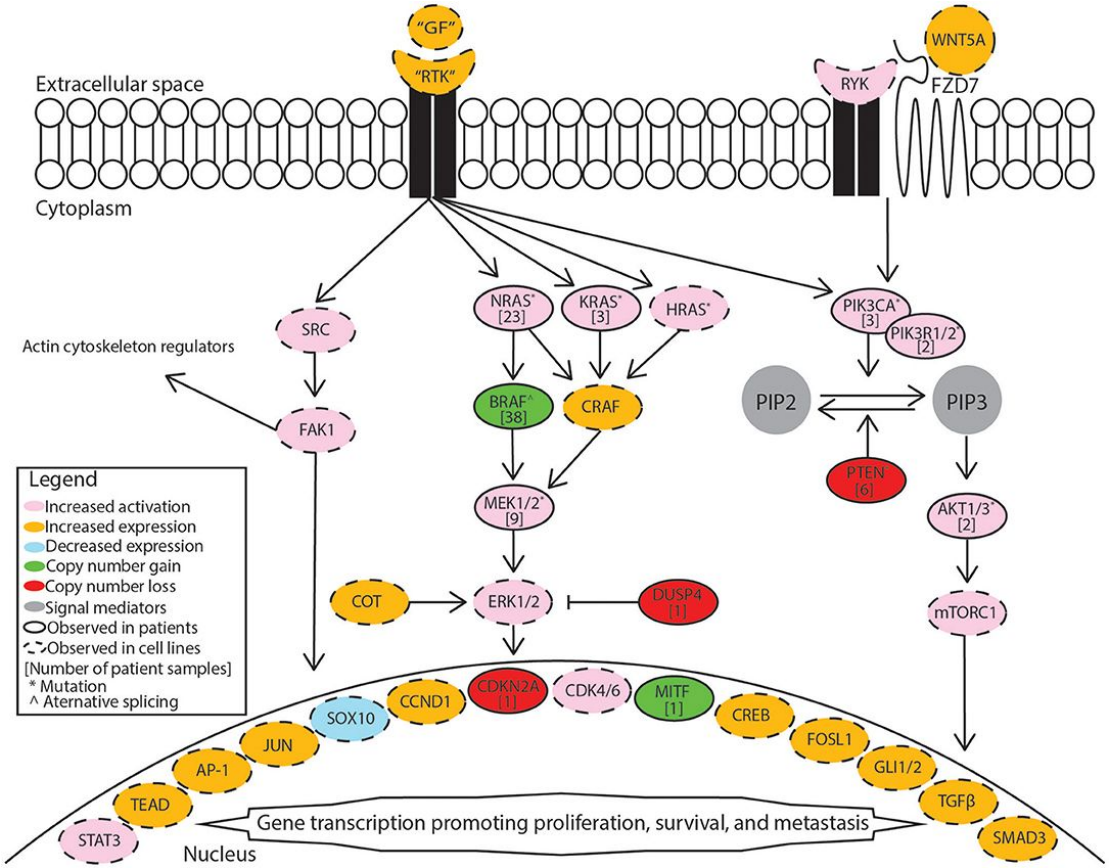


Thalidomide employs the same ubiquitination system to degrade different targets in teratogenicity (left) and in leukemia (right).



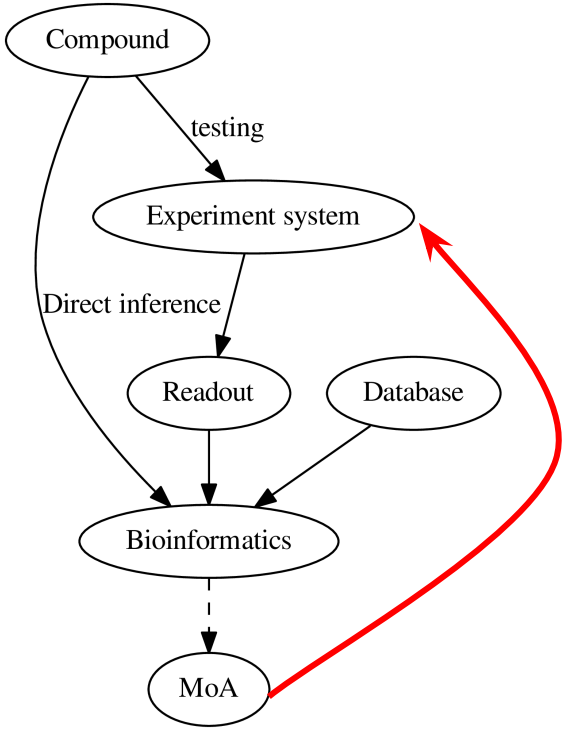
Challenge #3: Genetics may affect MoA

- Genetics may predispose individuals to different responses;
- Feedback loop and mutations may lead to drug resistance.



Computational biology contributes to MoA understanding by data analysis and integration

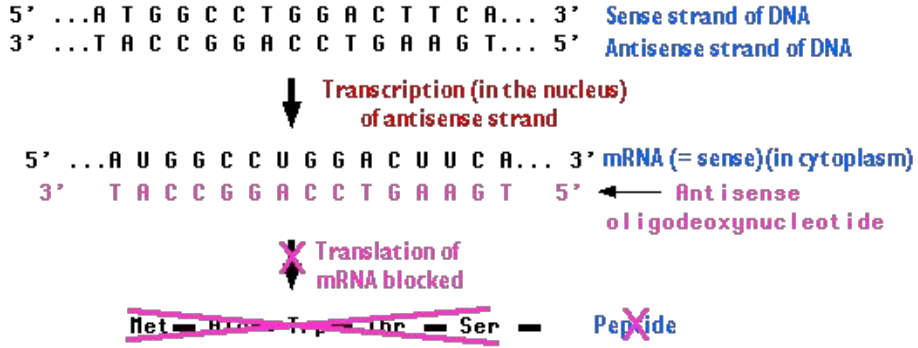
- MoA can be inferred either with the information of the compound alone *in silico*, or with the data generated *in vitro* or *in* vivo.
- Prior knowledge encoded in databases is often of great help.
- The process is always iterative with hypothesis-testing cycles.
- Below we illustrate modality-specific approaches



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.

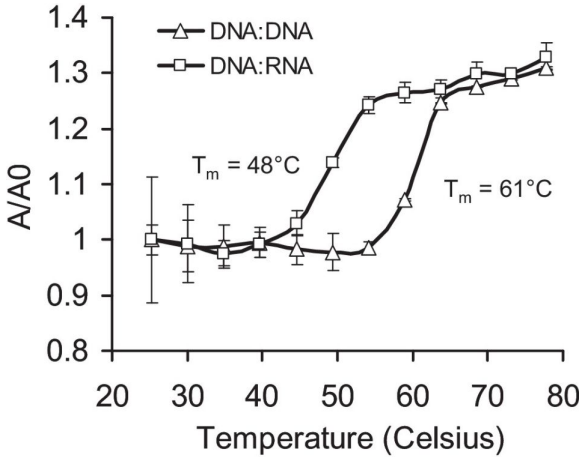


Human mRNAs

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

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.



- 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

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

Predicting melting temperature (i.e. binding affinity) of ASO-mRNA pairs with *free energy*

- It is possible to predict $T_{m,}$ using the nucleotide sequences and the principles of **nucleic acid thermodynamics.**
- The melting temperature is correlated with the free energy of the duplex (ΔG°), which can be predicted by dynamic-programming algorithms.
- **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')

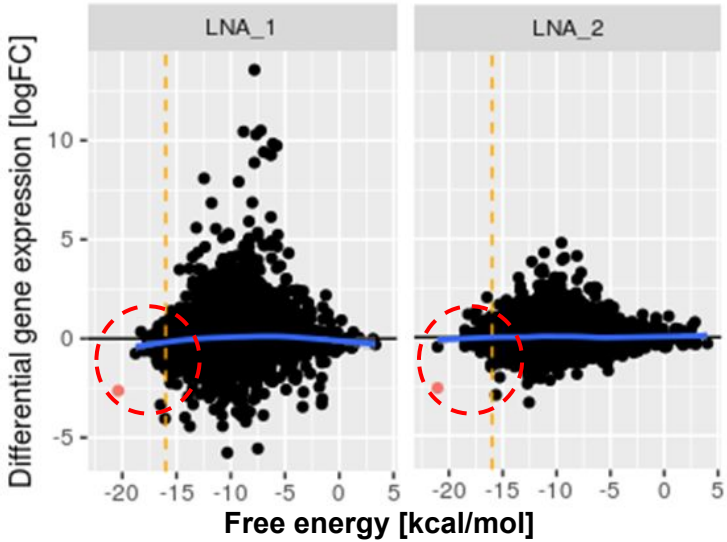
TGAAGTCCAGGCCAT

Human mRNAs	*Free Energy (kcal/mol)*
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?

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 for efficacy and safety studies.

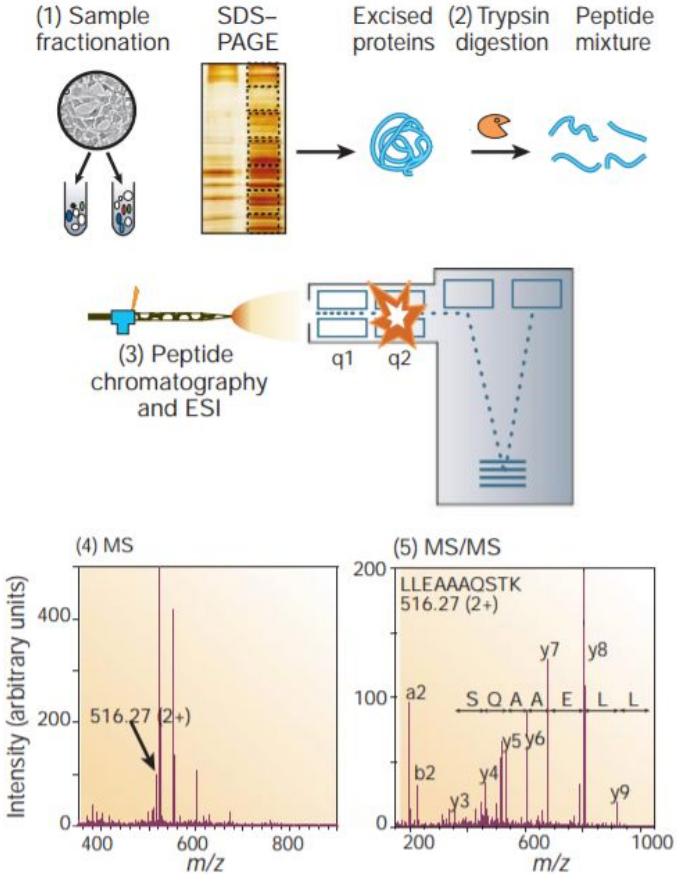


A declining trend at the left end (red dashed circle) is a warning sign: mRNAs that are predicted bound to the ASO are down-regulated, revealing potential off-target effects.

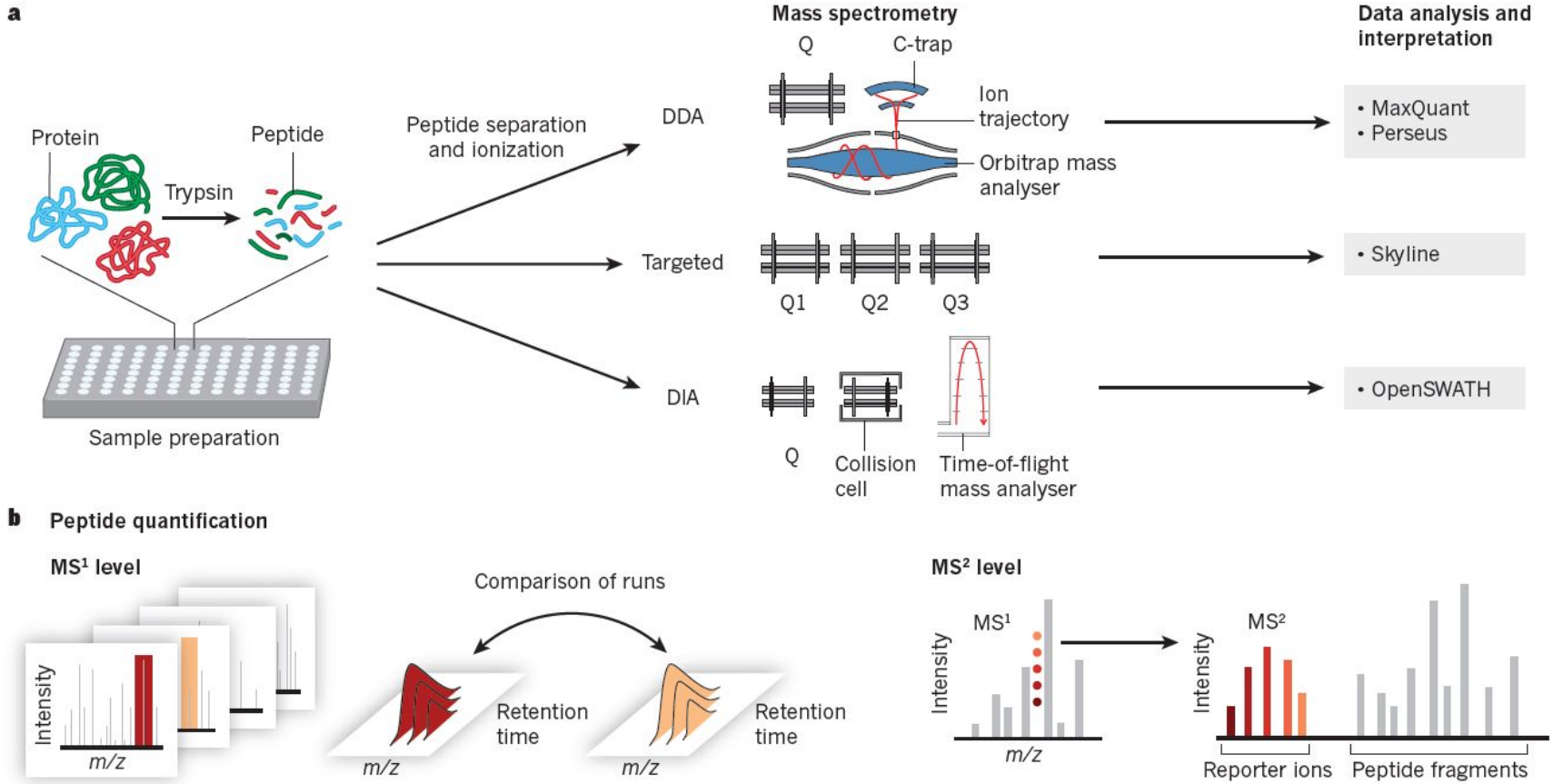
Understanding MoA of small molecules and antibodies with proteomics

Mass-spectrometry based Proteomics

- **SDS-PAGE**: Sodium Dodecyl Sulphate-Polyacrylamide Gel Electrophoresis
- **ESI**: Electrospray ionization
- **q1/q2**: selection/collision/separation cells
- **MS**: Mass spectrometry
- **MS/MS**: tandem mass spectrometry



Mass-spectrometry based proteomics

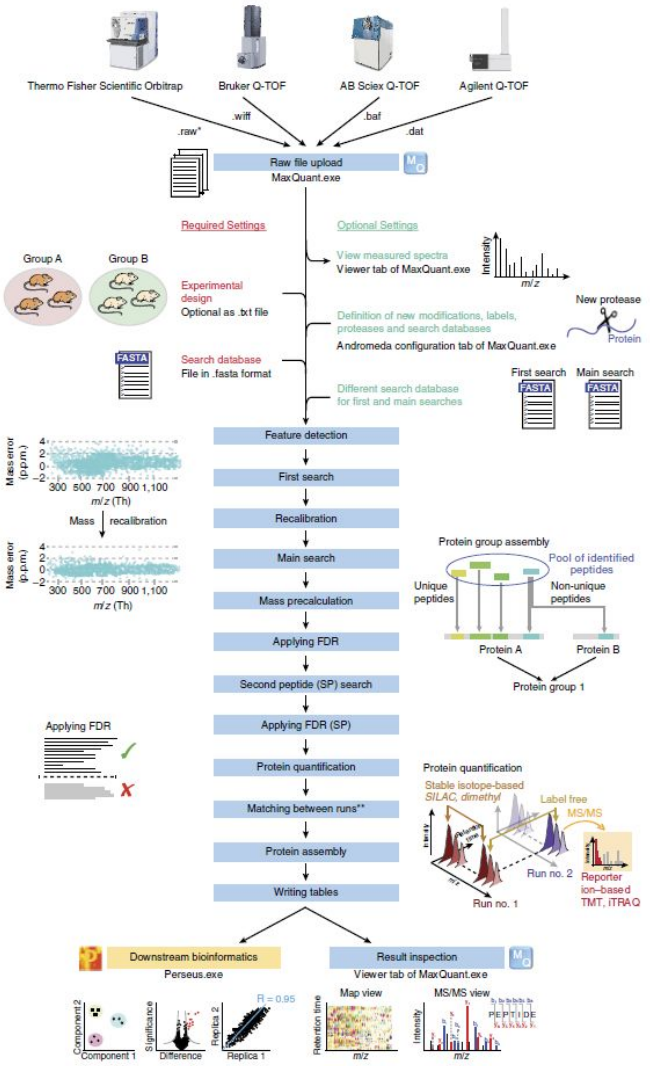


We use mature software to handle MS data

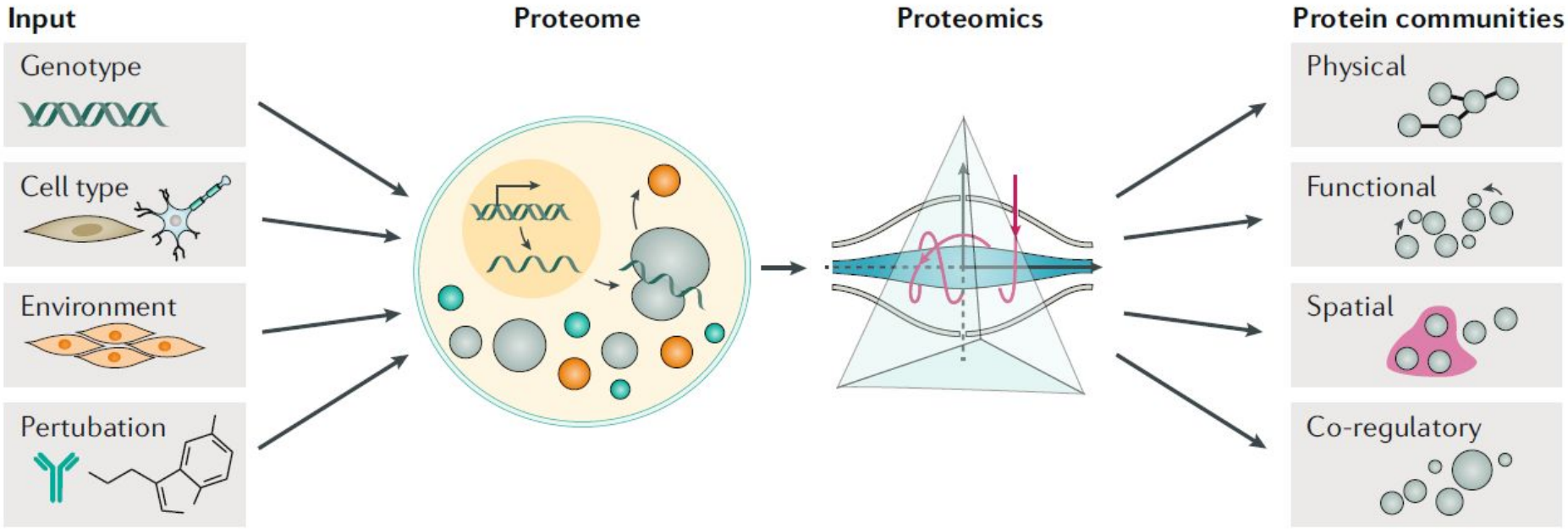
Here is an example of *MaxQuant*.

Additional work needs to be done:

- Experiment design
- Statistical modelling
- Pathway and network analysis
- Integration with other data



Proteomics enables the elucidation of protein relations in the protein communities



Proteomics approaches for drug discovery

Affinity purification

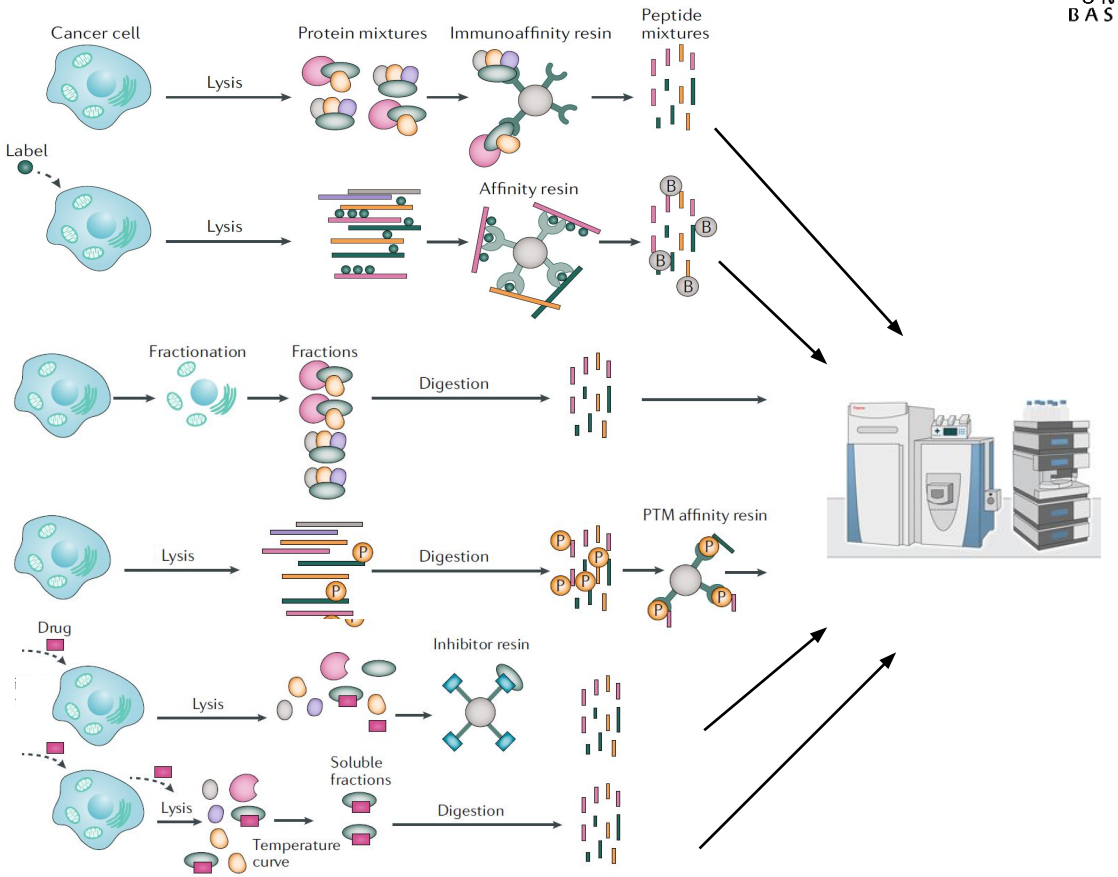
Proximity labelling

Organelle proteome profiling

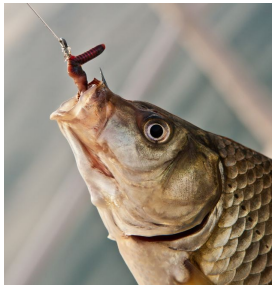
Post-translational modification (PTM) profiling

Chemoaffinity enrichment

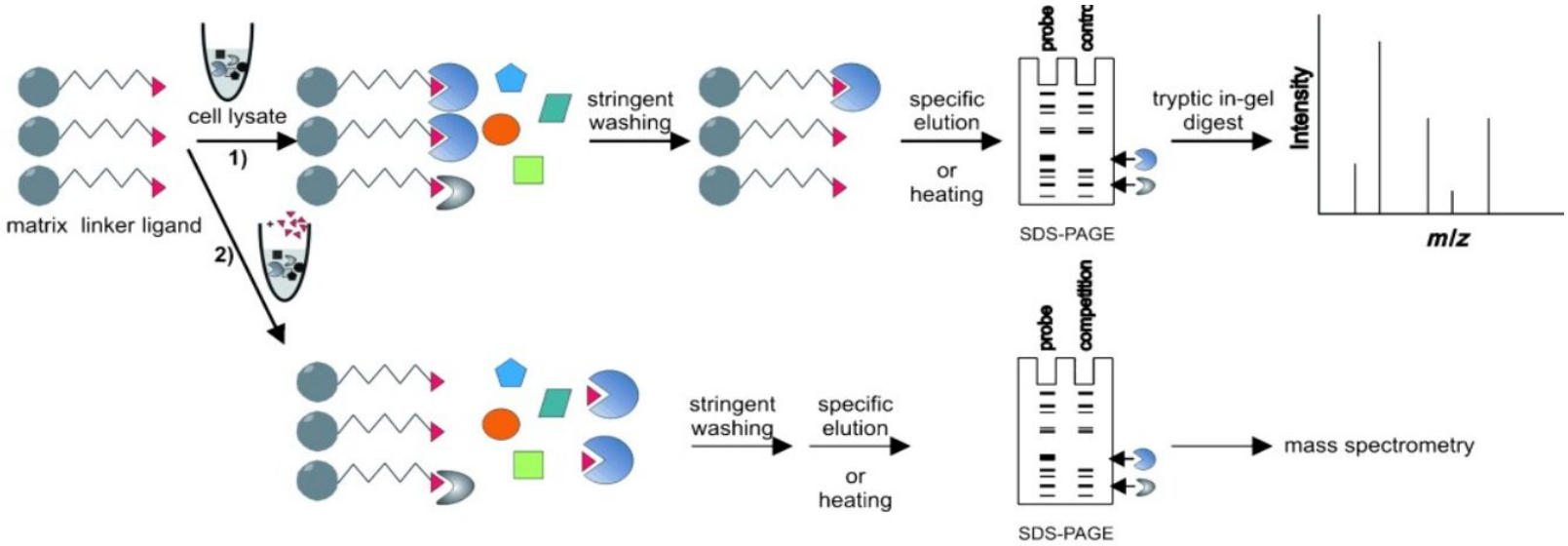
Thermal proteome profiling



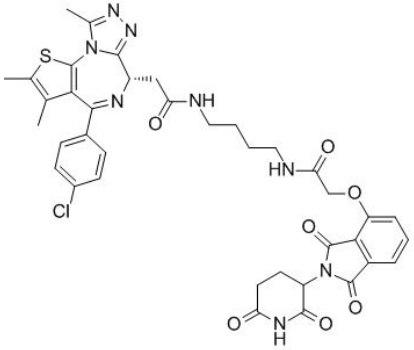
Example 1: Chemoproteomics for target ID



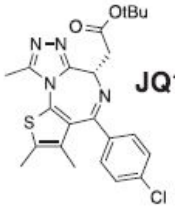
- Chemoproteomics methods are based on two principles: (1) **bait/prey** and (2) **competition**.
- Commonly used methods include affinity-based profiling (shown below), activity-based profiling, SILAC, *etc.*



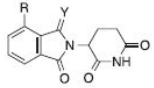
Example 2: Confirmation of selective degradation of protein target *in vivo*



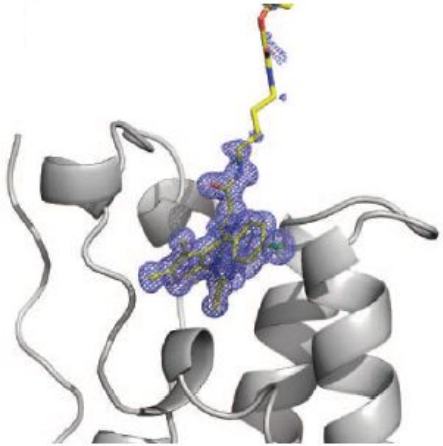
dBET1



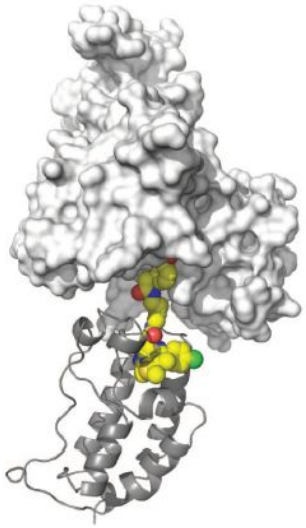
JQ1(S)



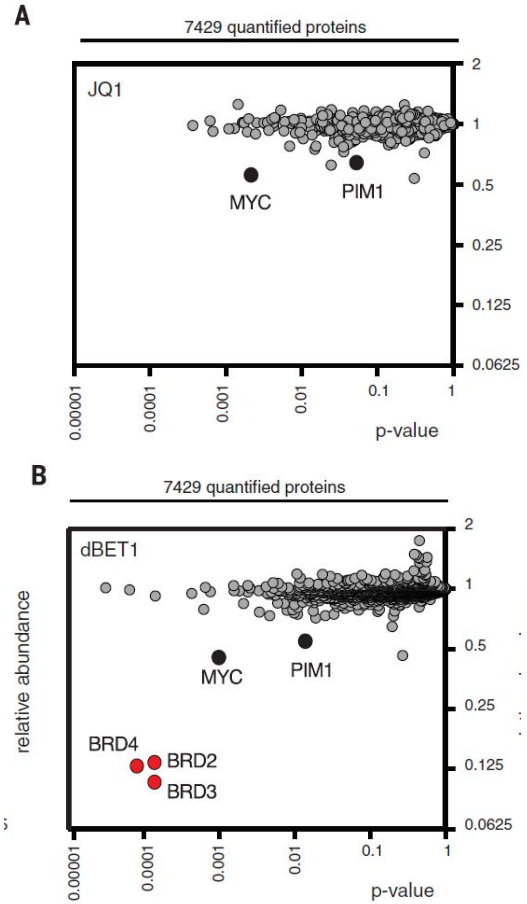
thalidomide: Y=O, R=H
lenalidomide: Y=CH_2, R=NH_2
pomalidomide: Y=O, R=NH_2



Crystal structure of dBET1 binding to its target BRD4



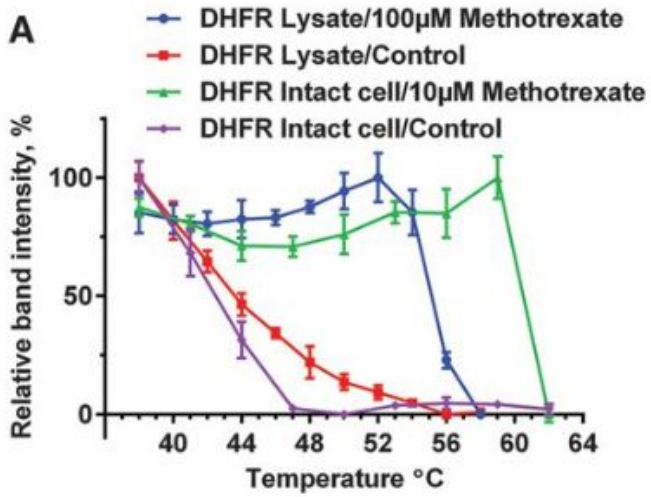
Docking of dBET1-BRD4 to DDB1-CRBN structure



Example 3: thermal proteome profiling identifies drug binding targets

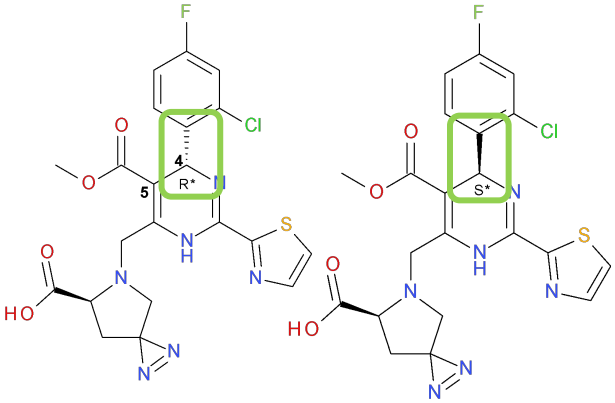
DON'T EAT, NOT EVEN COOKED!
The *death cap* contains *amatoxin*, a thermal stable toxin.

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)



Results of Cellular Thermal Shift Assay (CETSA) to verify DHFR as a target of methotrexate.

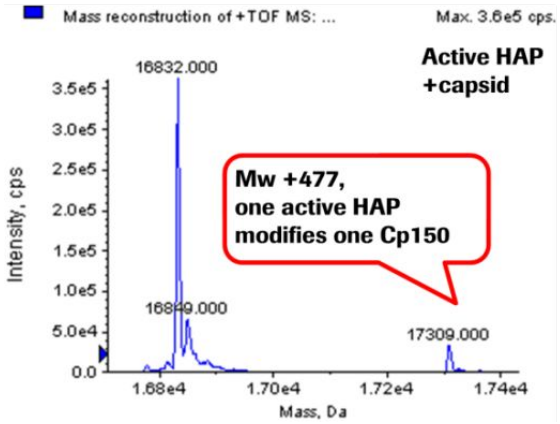
Example 4: photoaffinity labelling confirmed HBV capsid binding and mapped the small molecule binding pocket

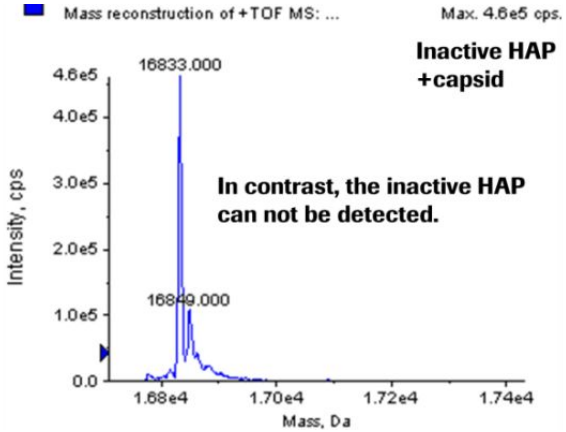


RO-A
EC_{50}: **0.040 µM**
IC_{50}: **0.47 µM**

RO-B
EC_{50}: **>1 µM**
IC_{50}: **>100 µM**

+Cp150, UV, MS



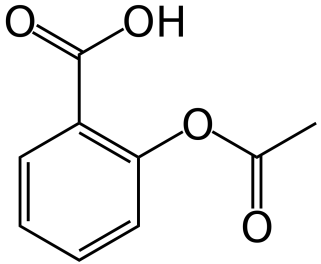


Proteolytic digestion/LC-MS/MS identified labelling site Y118 (Y=Tyrosine) of HBV capsid protein. More photoaffinity probes identified labelling sites at R127 (R=Arginine) and Y38.

Conclusions

- We predict efficacy and safety profiles of drugs by studying the mechanism and mode of action (MoA).
- Molecular modelling and (single-cell) RNA sequencing analysis are essential tools for understanding MoA of nucleotide-based modalities.
- Molecular modelling, RNA sequencing, and proteomics based on mass spectrometry (MS) are essential tools for understanding MoA of small molecules and antibodies.

The road towards MoA 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

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

It takes time and effort and luck to understand why it does so.

References

1. Figures: Lumen Learning, Exploring Nature, National Geographics, Platelet cells (Graham Beards, CC-BY-SA 4.0), Lymphocytes (NicolasGrandjean, CC-BY-SA 3.0), Adipocytes (Public Domain), Hepatocytes (CC-BY-NC 2.0), Neurons and Glia (Public Domain), Blood (CC 3.0), Blood Cells (By A. Rad and M. Häggström. CC-BY-SA 3.0 license), A selective JAK3 inhibitor (London Lab/Weizmann institute)
2. Sender, Ron, Shai Fuchs, and Ron Milo. 2016. “Revised Estimates for the Number of Human and Bacteria Cells in the Body.” PLoS Biology 14 (8). https://doi.org/10.1371/journal.pbio.1002533.
3. www.evocell-itn.eu;
4. Macaulay, Iain C., and Thierry Voet. 2014. “Single Cell Genomics: Advances and Future Perspectives.” PLOS Genetics 10 (1): e1004126. https://doi.org/10.1371/journal.pgen.1004126.
5. Pryor, Rosina, Povilas Norvaisas, Georgios Marinos, Lena Best, Louise B. Thingholm, Leonor M. Quintaneiro, Wouter De Haes, et al. 2019. “Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy.” Cell 178 (6): 1299-1312.e29. https://doi.org/10.1016/j.cell.2019.08.003.
6. Cully, Megan. 2019. “Microbiome Therapeutics Go Small Molecule.” *Nature Reviews Drug Discovery* 18 (July): 569. https://doi.org/10.1038/d41573-019-00122-8.
7. Duscha, Alexander, Barbara Gisevius, Sarah Hirschberg, Nissan Yissachar, Gabriele I. Stangl, Eva Eilers, Verian Bader, et al. 2020. “Propionic Acid Shapes the Multiple Sclerosis Disease Course by an Immunomodulatory Mechanism.” *Cell* 180 (6): 1067-1080.e16. https://doi.org/10.1016/j.cell.2020.02.035.
8. Pryor, Rosina, Povilas Norvaisas, Georgios Marinos, Lena Best, Louise B. Thingholm, Leonor M. Quintaneiro, Wouter De Haes, et al. 2019. “Host-Microbe-Drug-Nutrient Screen Identifies Bacterial Effectors of Metformin Therapy.” *Cell* 178 (6): 1299-1312.e29. https://doi.org/10.1016/j.cell.2019.08.003.
9. Zimmermann, Michael, Maria Zimmermann-Kogadeeva, Rebekka Wegmann, and Andrew L. Goodman. 2019. “Mapping Human Microbiome Drug Metabolism by Gut Bacteria and Their Genes.” *Nature* 570 (7762): 462. https://doi.org/10.1038/s41586-019-1291-3.
10. Shin, Hyun Kil, Young-Mook Kang, and Kyoung Tai No. 2016. “Predicting ADME Properties of Chemicals.” In Handbook of Computational Chemistry, edited by Jerzy Leszczynski, 1–37. Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-007-6169-8_59-1.

References (continued)

11. Mädler, Sophia Clara, Alice Julien-Laferriere, Luis Wyss, Miroslav Phan, Albert S. W. Kang, Eric Ulrich, Roland Schmucki, et al. 2020. “Besca, a Single-Cell Transcriptomics Analysis Toolkit to Accelerate Translational Research.” BioRxiv, September, 2020.08.11.245795. https://doi.org/10.1101/2020.08.11.245795.
12. Andrews, Tallulah S., Vladimir Yu Kiselev, Davis McCarthy, and Martin Hemberg. 2021. “Tutorial: Guidelines for the Computational Analysis of Single-Cell RNA Sequencing Data.” Nature Protocols 16 (1): 1–9. https://doi.org/10.1038/s41596-020-00409-w.
13. Sturm, Gregor, Francesca Finotello, Florent Petitprez, Jitao David Zhang, Jan Baumbach, Wolf H. Fridman, Markus List, and Tatsiana Aneichyk. 2019. “Comprehensive Evaluation of Transcriptome-Based Cell-Type Quantification Methods for Immuno-Oncology.” Bioinformatics 35 (14): i436–45. https://doi.org/10.1093/bioinformatics/btz363.
14. Villani, Alexandra-Chloé, Rahul Satija, Gary Reynolds, Siranush Sarkizova, Karthik Shekhar, James Fletcher, Morgane Griesbeck, et al. 2017. “Single-Cell RNA-Seq Reveals New Types of Human Blood Dendritic Cells, Monocytes, and Progenitors.” Science 356 (6335): eaah4573. https://doi.org/10.1126/science.aah4573.
15. Finotello, Francesca, Clemens Mayer, Christina Plattner, Gerhard Laschober, Dietmar Rieder, Hubert Hackl, Anne Krogsdam, et al. 2019. “Molecular and Pharmacological Modulators of the Tumor Immune Contexture Revealed by Deconvolution of RNA-Seq Data.” *Genome Medicine* 11 (1): 34. https://doi.org/10.1186/s13073-019-0638-6.
16. Fridman, Wolf H., Laurence Zitvogel, Catherine Sautès–Fridman, and Guido Kroemer. 2017. “The Immune Contexture in Cancer Prognosis and Treatment.” *Nature Reviews Clinical Oncology* 14 (12): 717–34. https://doi.org/10.1038/nrclinonc.2017.101.
17. Moisan, Annie, Marcel Gubler, Jitao David Zhang, Yann Tessier, Kamille Dumong Erichsen, Sabine Sewing, Régine Gérard, et al. 2017. “Inhibition of EGF Uptake by Nephrotoxic Antisense Drugs In Vitro and Implications for Preclinical Safety Profiling.” Molecular Therapy - Nucleic Acids 6 (March): 89–105. https://doi.org/10.1016/j.omtn.2016.11.006.
18. Chang, Chia-Yu, Hsiao-Chien Ting, Ching-Ann Liu, Hong-Lin Su, Tzyy-Wen Chiou, Horng-Jyh Harn, and Shinn-Zong Lin. 2018. “Induced Pluripotent Stem Cells: A Powerful Neurodegenerative Disease Modeling Tool for Mechanism Study and Drug Discovery.” Cell Transplantation 27 (June): 096368971877540. https://doi.org/10.1177/0963689718775406.

References (continued)

19. Takahashi, Toshio. 2019. “Organoids for Drug Discovery and Personalized Medicine.” Annual Review of Pharmacology and Toxicology 59 (1): 447–62. https://doi.org/10.1146/annurev-pharmtox-010818-021108.
20. Budayeva, Hanna G., and Donald S. Kirkpatrick. 2020. “Monitoring Protein Communities and Their Responses to Therapeutics.” Nature Reviews Drug Discovery 19 (6): 414–26. https://doi.org/10.1038/s41573-020-0063-y.
21. Lukonin, Ilya, Denise Serra, Ludivine Challet Meylan, Katrin Volkmann, Janine Baaten, Rui Zhao, Shelly Meeusen, et al. 2020. “Phenotypic Landscape of Intestinal Organoid Regeneration.” Nature 586 (7828): 275–80. https://doi.org/10.1038/s41586-020-2776-9.
22. Drawnel, Faye M., Stefano Boccardo, Michael Prummer, Frédéric Delobel, Alexandra Graff, Michael Weber, Régine Gérard, et al. 2014. “Disease Modeling and Phenotypic Drug Screening for Diabetic Cardiomyopathy Using Human Induced Pluripotent Stem Cells.” Cell Reports 9 (3): 810–20. https://doi.org/10.1016/j.celrep.2014.09.055.
23. Traag, Vincent, Ludo Waltman, and Nees Jan van Eck. 2019. “From Louvain to Leiden: Guaranteeing Well-Connected Communities.” Scientific Reports 9 (1): 5233. https://doi.org/10.1038/s41598-019-41695-z.
24. *Understanding UMAP,* Andy Coenen and Adam Pearce, https://pair-code.github.io/understanding-umap/
25. How exactly UMAP works, Nikolay Oskolkov, https://towardsdatascience.com/how-exactly-umap-works-13e3040e1668
26. McInnes, Leland, and John Healy. 2018. “UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.” ArXiv:1802.03426 [Cs, Stat], February. http://arxiv.org/abs/1802.03426.
27. Zappia, Luke, Belinda Phipson, and Alicia Oshlack. 2018. “Exploring the Single-Cell RNA-Seq Analysis Landscape with the ScRNA-Tools Database.” PLOS Computational Biology 14 (6): e1006245. https://doi.org/10.1371/journal.pcbi.1006245.
28. Abdelaal, Tamim, Lieke Michielsen, Davy Cats, Dylan Hoogduin, Hailiang Mei, Marcel J. T. Reinders, and Ahmed Mahfouz. 2019. “A Comparison of Automatic Cell Identification Methods for Single-Cell RNA Sequencing Data.” Genome Biology 20 (1): 194. https://doi.org/10.1186/s13059-019-1795-z.
29. Janas, Maja M., Mark K. Schlegel, Carole E. Harbison, Vedat O. Yilmaz, Yongfeng Jiang, Rubina Parmar, Ivan Zlatev, et al. 2018. “Selection of GalNAc-Conjugated SiRNAs with Limited off-Target-Driven Rat Hepatotoxicity.” Nature Communications 9 (1): 723. https://doi.org/10.1038/s41467-018-02989-4.

References (continued)

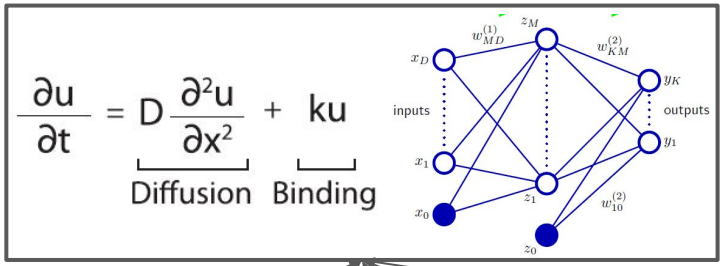
30. Jackson, Aimee L., and Peter S. Linsley. 2010. “Recognizing and Avoiding SiRNA Off-Target Effects for Target Identification and Therapeutic Application.” Nature Reviews Drug Discovery 9 (1): 57–67. https://doi.org/10.1038/nrd3010.
31. Romond, Edward H., Edith A. Perez, John Bryant, Vera J. Suman, Charles E. Geyer, Nancy E. Davidson, Elizabeth Tan-Chiu, et al. 2005. “Trastuzumab plus Adjuvant Chemotherapy for Operable HER2-Positive Breast Cancer.” New England Journal of Medicine 353 (16): 1673–84. https://doi.org/10.1056/NEJMoa052122.
32. Gao, Jinxu, Adelphe Mfuh, Yuka Amako, and Christina M. Woo. 2018. “Small Molecule Interactome Mapping by Photoaffinity Labeling Reveals Binding Site Hotspots for the NSAIDs.” Journal of the American Chemical Society 140 (12): 4259–68. https://doi.org/10.1021/jacs.7b11639.
33. Bollag, Gideon, James Tsai, Jiazhong Zhang, Chao Zhang, Prabha Ibrahim, Keith Nolop, and Peter Hirth. 2012. “Vemurafenib: The First Drug Approved for BRAF -Mutant Cancer.” Nature Reviews Drug Discovery 11 (11): 873–86. https://doi.org/10.1038/nrd3847.
34. Luebker, Stephen A., and Scott A. Koepsell. 2019. “Diverse Mechanisms of BRAF Inhibitor Resistance in Melanoma Identified in Clinical and Preclinical Studies.” Frontiers in Oncology 9. https://doi.org/10.3389/fonc.2019.00268.
35. Kimball’s Biology Page, http://www.biology-pages.info/
36. Molina, Daniel Martinez, Rozbeh Jafari, Marina Ignatushchenko, Takahiro Seki, E. Andreas Larsson, Chen Dan, Lekshmy Sreekumar, Yihai Cao, and Pär Nordlund. 2013. “Monitoring Drug Target Engagement in Cells and Tissues Using the Cellular Thermal Shift Assay.” Science 341 (6141): 84–87. https://doi.org/10.1126/science.1233606.
37. Zhou, Zheng, Taishan Hu, Xue Zhou, Steffen Wildum, Fernando Garcia-Alcalde, Zhiheng Xu, Daitze Wu, et al. 2017. “Heteroaryldihydropyrimidine (HAP) and Sulfamoylbenzamide (SBA) Inhibit Hepatitis B Virus Replication by Different Molecular Mechanisms.” Scientific Reports 7 (1): 42374. https://doi.org/10.1038/srep42374.
38. Dai, Jiang, Yi-Jiao Huang, Xinhua He, Ming Zhao, Xinzheng Wang, Zhao-Shan Liu, Wen Xue, et al. 2019. “Acetylation Blocks CGAS Activity and Inhibits Self-DNA-Induced Autoimmunity.” Cell 176 (6): 1447--1460.E14. https://doi.org/10.1016/j.cell.2019.01.016.

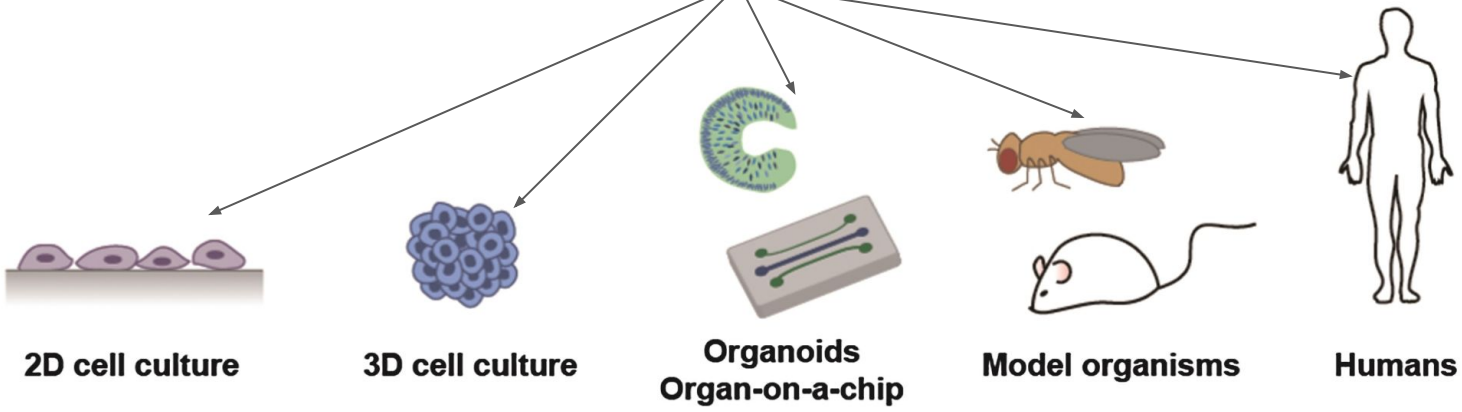
References (continued)

39. Hart, Charles P. 2005. “Finding the Target after Screening the Phenotype.” *Drug Discovery Today* 10 (7): 513–19. https://doi.org/10.1016/S1359-6446(05)03415-X.
40. Ziegler, Slava, Sonja Sievers, and Herbert Waldmann. 2021. “Morphological Profiling of Small Molecules.” Cell Chemical Biology 28 (3): 300–319. https://doi.org/10.1016/j.chembiol.2021.02.012.
41. Winter, Georg E., Dennis L. Buckley, Joshiawa Paulk, Justin M. Roberts, Amanda Souza, Sirano Dhe-Paganon, and James E. Bradner. 2015. “Phthalimide Conjugation as a Strategy for in Vivo Target Protein Degradation.” Science 348 (6241): 1376–81. https://doi.org/10.1126/science.aab1433.
42. Aebersold, Ruedi, and Matthias Mann. 2016. “Mass-Spectrometric Exploration of Proteome Structure and Function.” Nature 537 (7620): 347–55. https://doi.org/10.1038/nature19949.
43. Zhou, Jing C., Bob Feller, Bill Hinsberg, Geeta Sethi, Paul Feldstein, Joshua Hihath, Erkin Seker, Maria Marco, Andre Knoesen, and Robert Miller. 2015. “Immobilization-Mediated Reduction in Melting Temperatures of DNA–DNA and DNA–RNA Hybrids: Immobilized DNA Probe Hybridization Studied by SPR.” Colloids and Surfaces A: Physicochemical and Engineering Aspects 481 (September): 72–79. https://doi.org/10.1016/j.colsurfa.2015.04.046.
44. Hagedorn, Peter H., Malene Pontoppidan, Tina S. Bisgaard, Marco Berrera, Andreas Dieckmann, Martin Ebeling, Marianne R. Møller, et al. 2018. “Identifying and Avoiding Off-Target Effects of RNase H-Dependent Antisense Oligonucleotides in Mice.” Nucleic Acids Research 46 (11): 5366–80. https://doi.org/10.1093/nar/gky397.
45. Rehmsmeier, Marc, Peter Steffen, Matthias Hochsmann, and Robert Giegerich. 2004. “Fast and Effective Prediction of MicroRNA/Target Duplexes.” RNA (New York, N.Y.) 10 (10): 1507–17. https://doi.org/10.1261/rna.5248604.
46. Tyanova, Stefka, Tikira Temu, and Juergen Cox. 2016. “The MaxQuant Computational Platform for Mass Spectrometry-Based Shotgun Proteomics.” Nature Protocols 11 (12): 2301–19. https://doi.org/10.1038/nprot.2016.136.
47. xkcd: https://xkcd.com/1217/

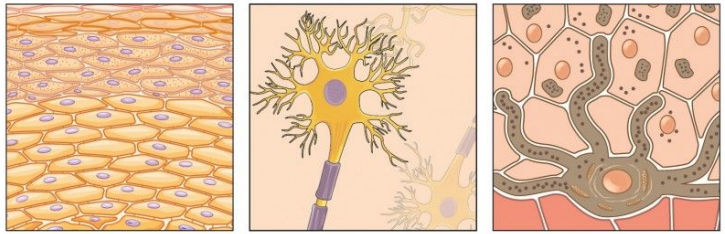
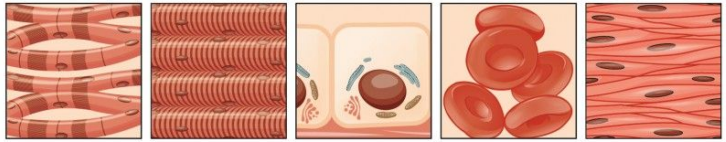
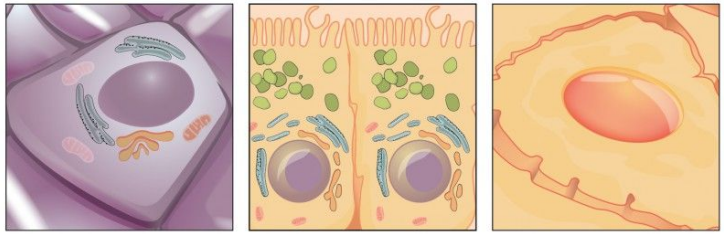
Supplementary Information

Computational and physical models of human biology

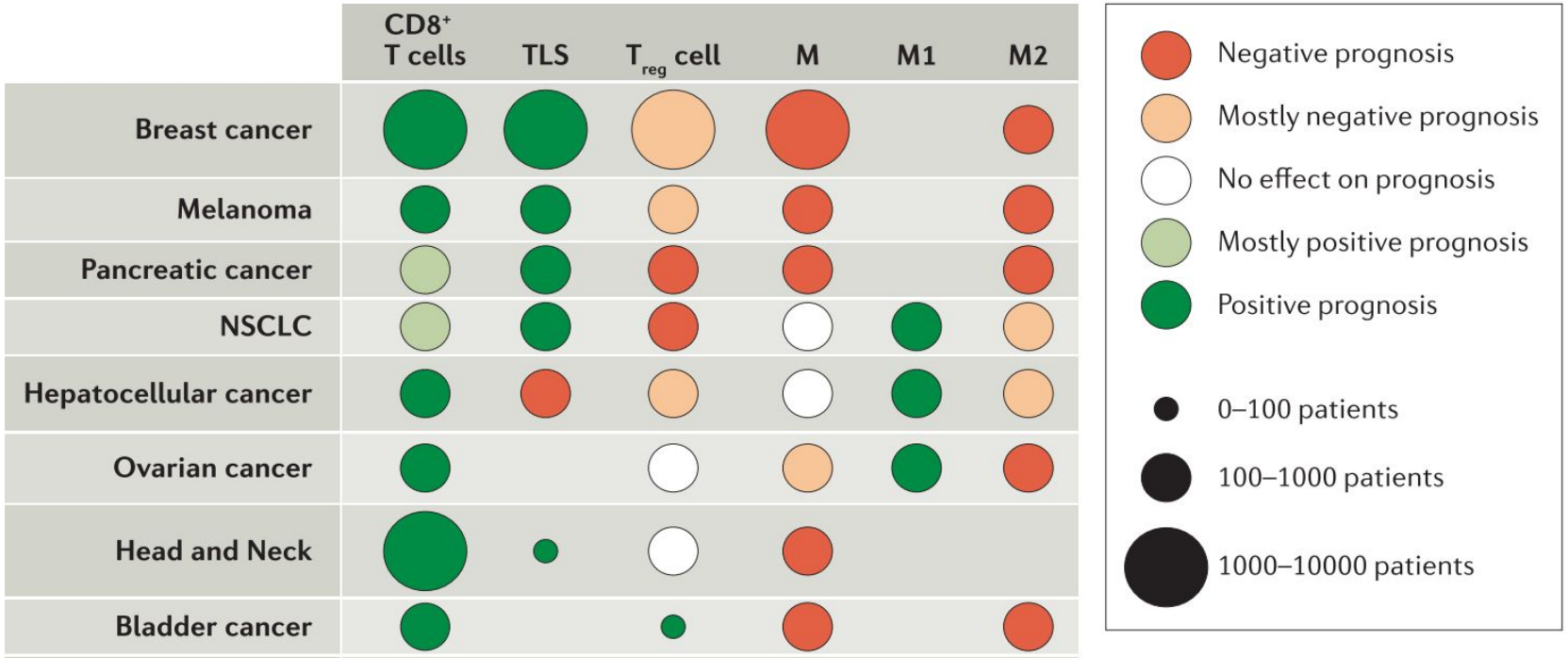




Embryonic origins of tissues

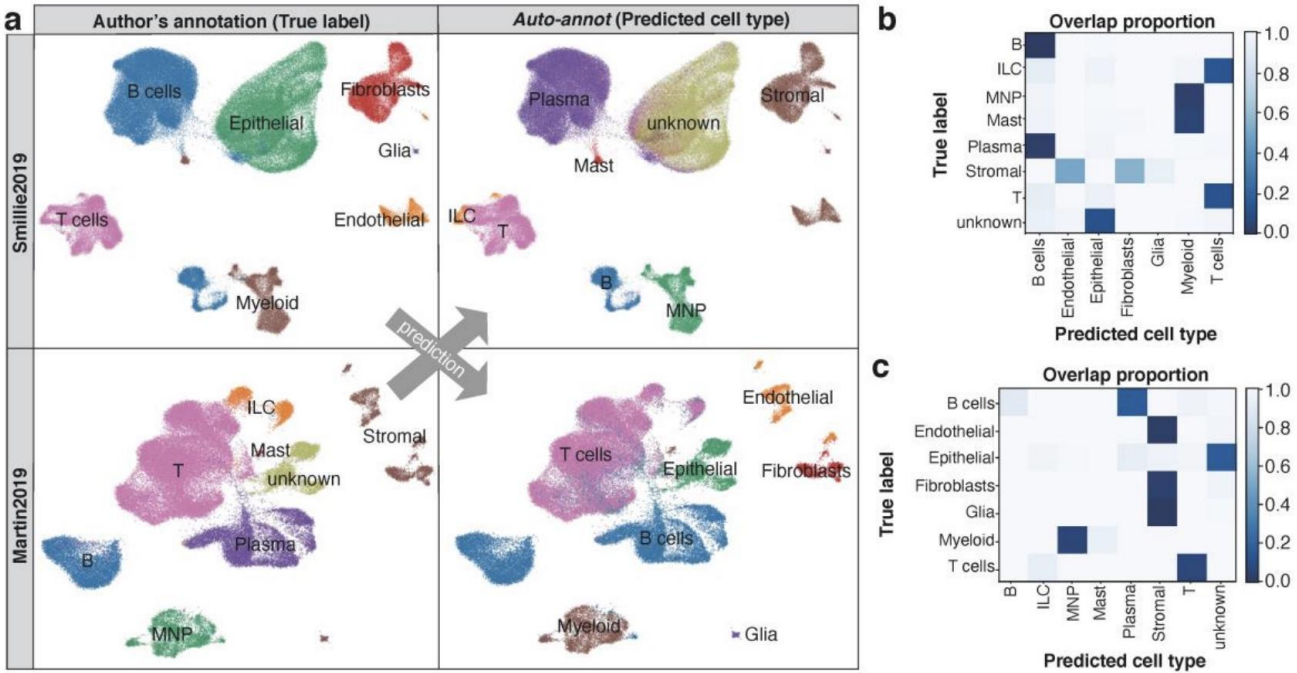
Germ Layer	Gives rise to:
Ectoderm	Epidermis, glands on skin, some cranial bones, pituitary and adrenal medulla, the nervous system, the mouth between cheek and gums, the anus  Skin cells Neurons Pigment cell
Mesoderm	Connective tissues proper, bone, cartilage, blood, endothelium of blood vessels, muscle, synovial membranes, serous membranes lining body cavities, kidneys, lining of gonads  Cardiac muscle Skeletal muscle Tubule cell of kidney Red blood cells Smooth muscle
Endoderm	Lining of airways and digestive system except the mouth and distal part of digestive system (rectum and anal canal); glands (digestive glands, endocrine glands, adrenal cortex)  Lung cell Thyroid cell Pancreatic cell

Abundance of immune cells in tumor microenvironments affect outcome



TLS: tertiary lymphoid structures; T_{reg}: regulatory T cells; M: macrophages; M1/M2: subtypes of macrophages

An example of Inflammatory Bowel Disease (IBD)



We observed Inconsistent cell type nomenclature across studies.
Machine learning allows us compare and integrate multiple studies.