Computational Biology For Novel Targets and Novel Therapeutics: A Multiscale-Modeling Approach

Jitao David Zhang, Computational Biologist, Roche Innovation Center Basel
Department of Pharmacy, University of Basel, March 2023

Pre-read, not shown

Pre-read of the lecture

- *Multiscale modeling of drug mechanism and safety*, Drug Discovery Today (2020)
- *An introduction to machine learning*, Clinical Pharmacology & Therapeutics (2020)
- *Causal inference in drug discovery and development*, *arXiv*, 2022

Acknowledgements

- **Predictive Modeling and Data Analysis (PMDA) team members,** especially for the work introduced here performed by Fabian Birzele, Iakov Davydov, Petra Schwalie, Daniel Marbach, Guido Steiner, Roland Schmucki, Laura Badi, Klas Hatje, Martin Ebeling, Marco Berrera, Nikolaos Bertenis, Tony Kam-Thong, Alex Seitz, Sophia Maedler, Luis Wyss, Rafael Mueller, Miroslav Phan, Andrea Ciuprina, and Anthony Sonrel.
- **Colleagues of Pharma Research and Early Development**, including Erich Küng, Claudia Bossen, Megana Prasad, Filip Roudnicky, Lisa Sach-Peltason, Manuela Jaklin, Stefan Kustermann, Michael Reutlinger, Christian Kramer, Michael Cannarile, Sabine Hoves, Marina Bacac, Ramona Schlenker, Sylvia Herter, Stephan Dettling, Marisa Mariani, Tamara Huesser, Vinko Tosevski, and Alison Ribeiro.

Computational biology transforms drug discovery and development

Integrating public & internal knowledge and data

Optimizing molecules by data- and model-driven approaches

Understanding mechanism and safety profiles

Translating clinical findings to better targets, preclinical models, and molecules

Drug Discovery and Development

Target Identification

Lead Discovery

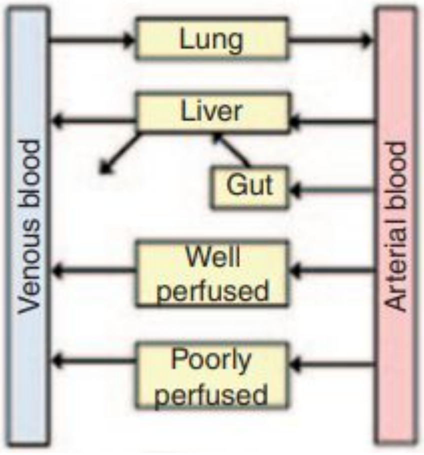
Medicinal Chemistry

In vitro studies

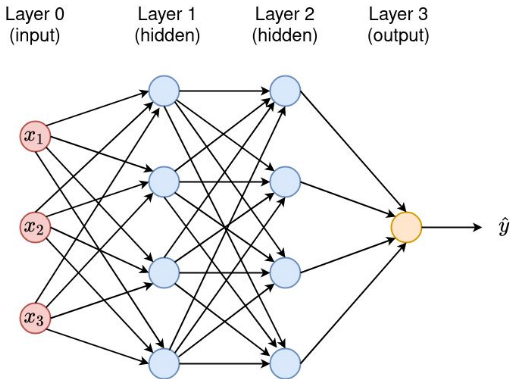
In vivo studies

Clinical Trials

Three types of computational models

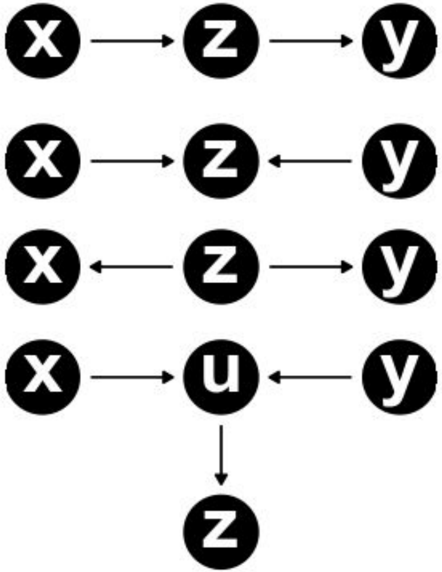


Mechanistic models



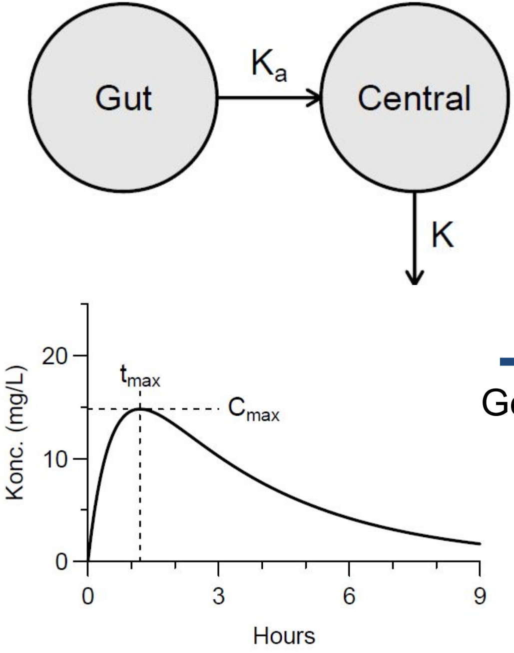
$$y \sim f(x)$$

Statistical and machine-learning models

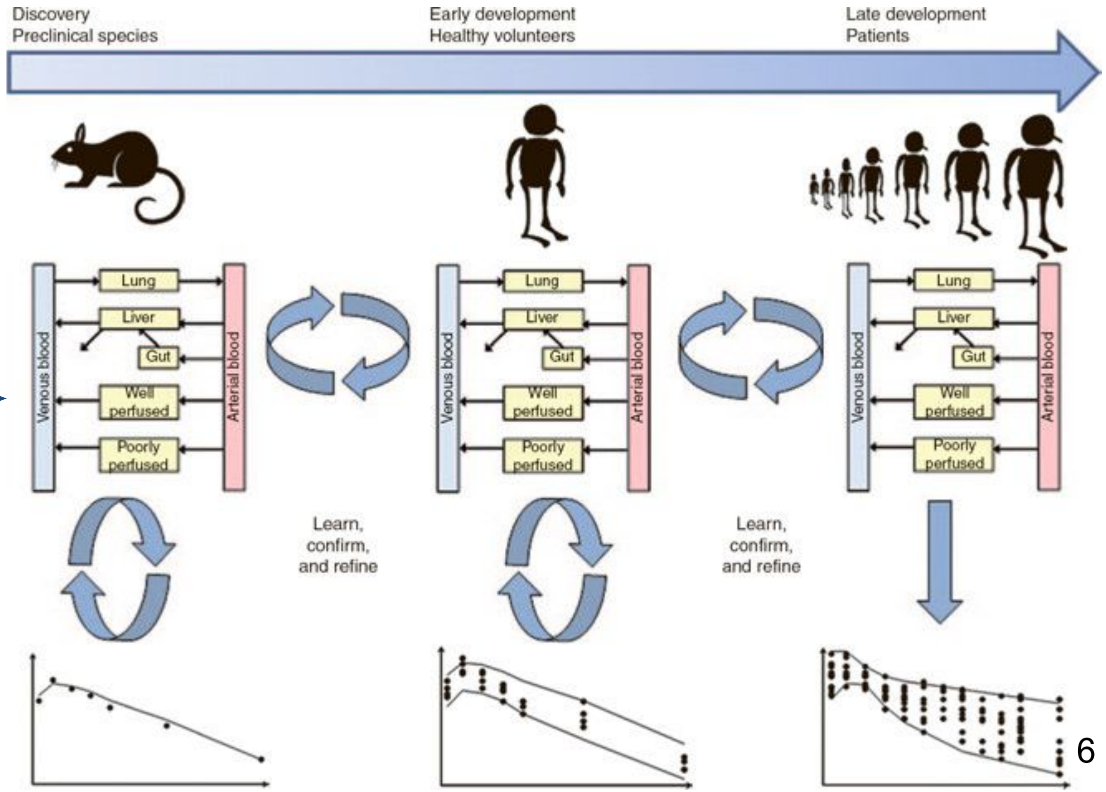


Causal models

A mechanistic model: physiologically based pharmacokinetic modeling



Generalization

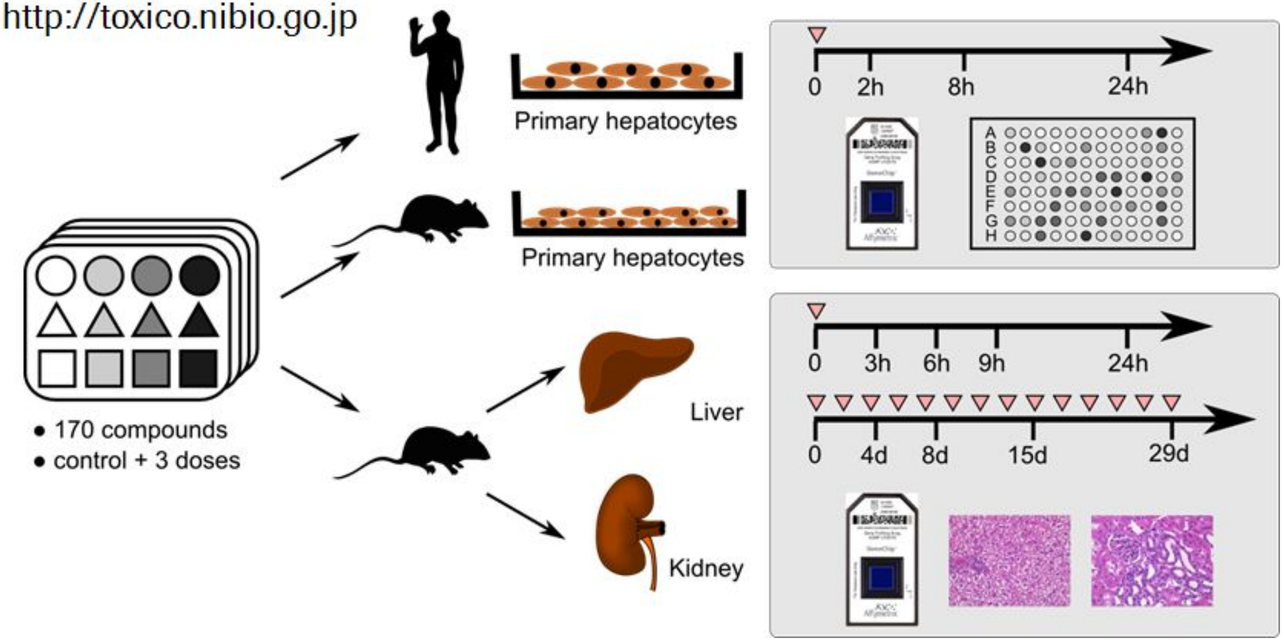


Jones and Rowland-Yeo, 2013

No mechanistic models are available for huge amount of data

TG-GATEs

http://toxico.nibio.go.jp



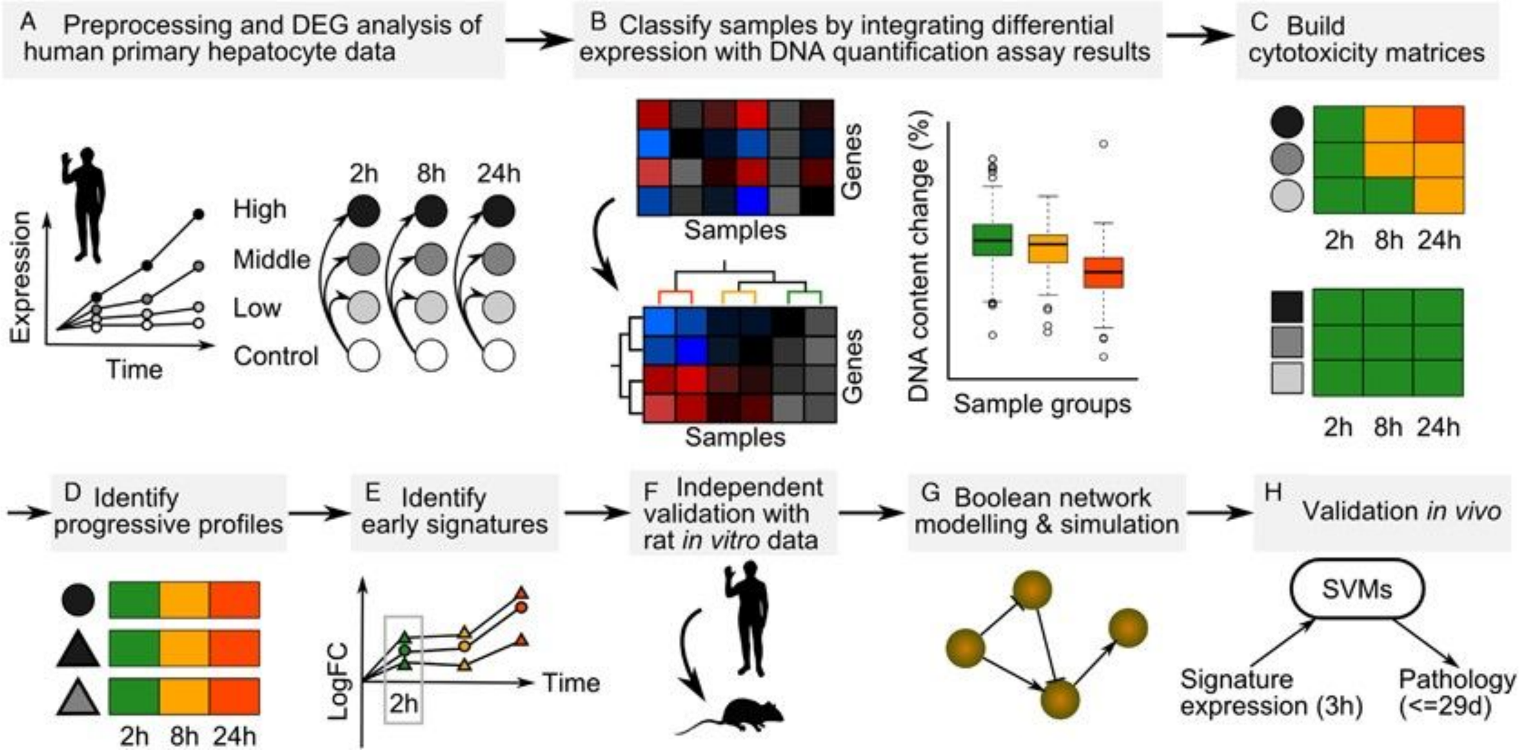
170
Compounds

>2000
Cellular assays

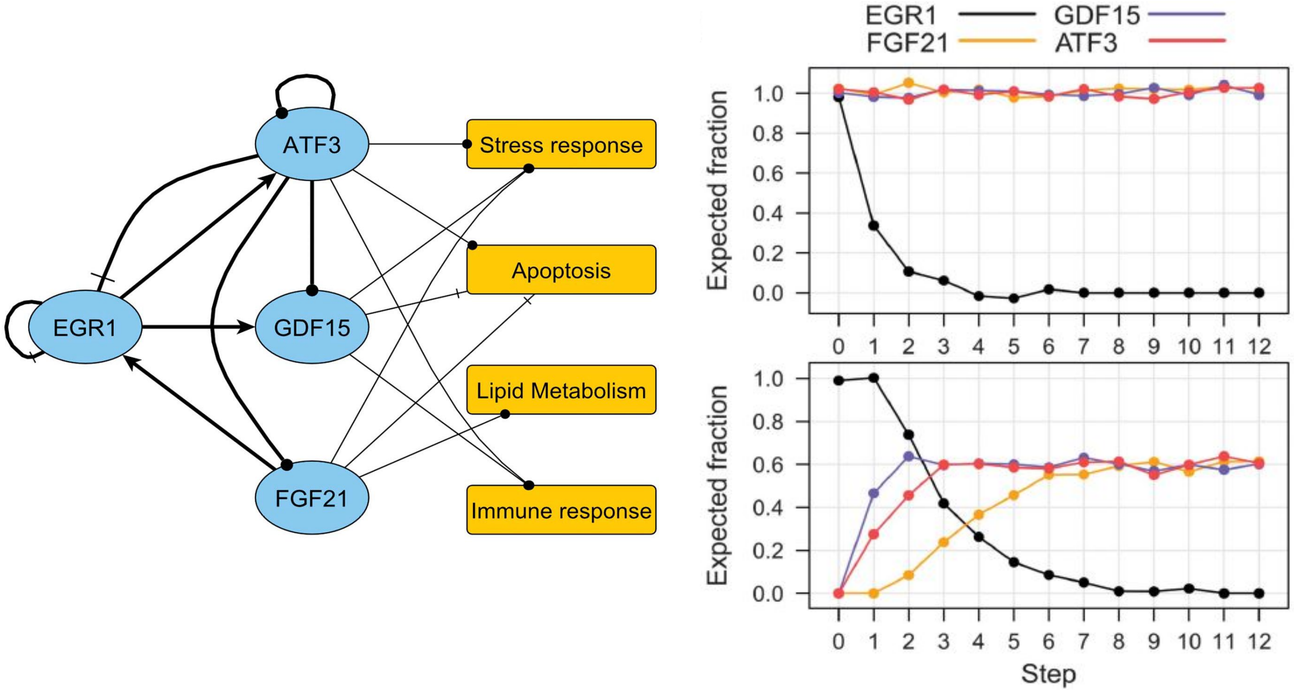
>12000
Pathology records

>24000
Expression profiles

Statistical models identify hidden patterns and discover correlations

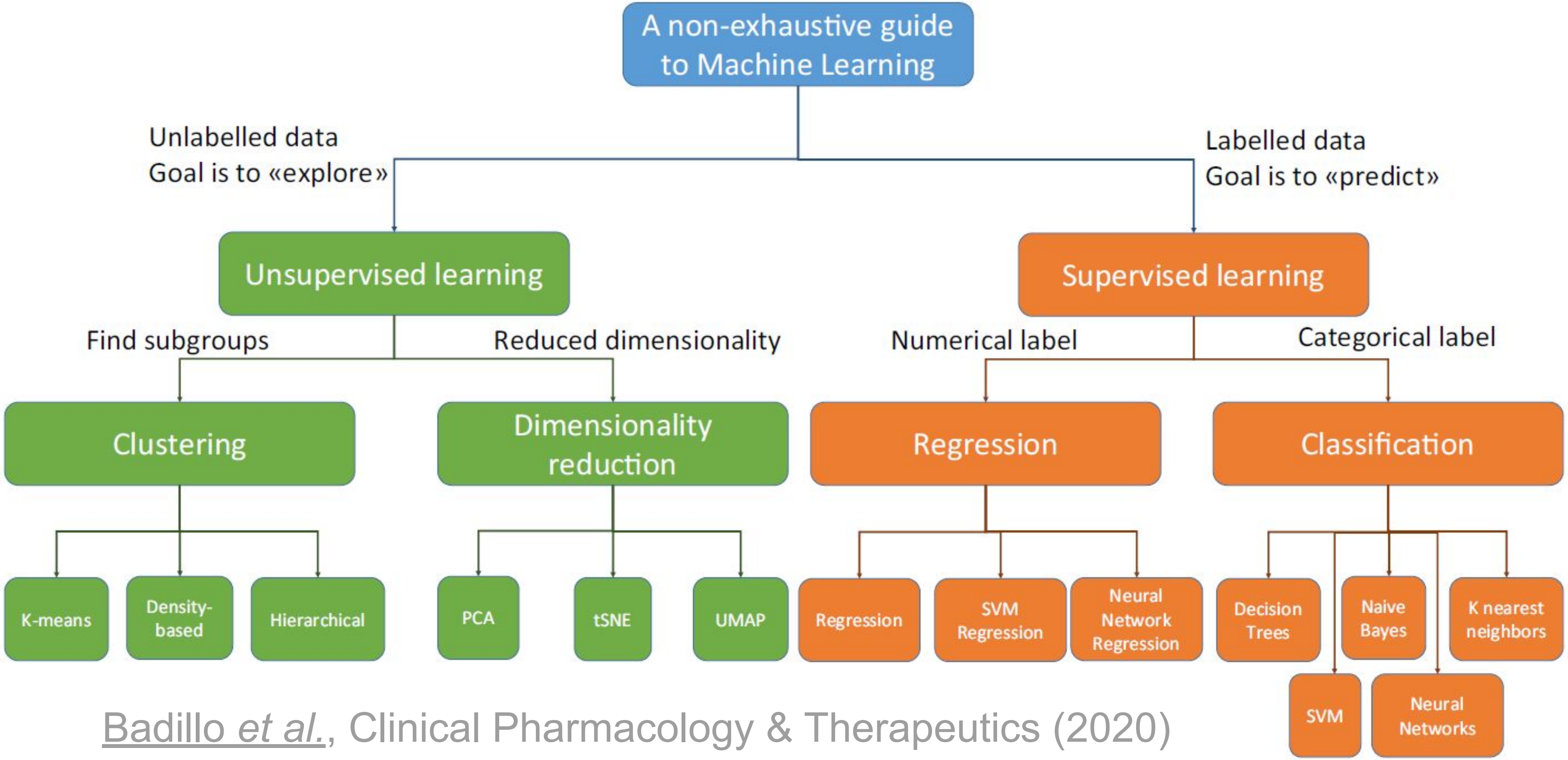


Statistical models and prior knowledge can give rise to mechanistic/causal models



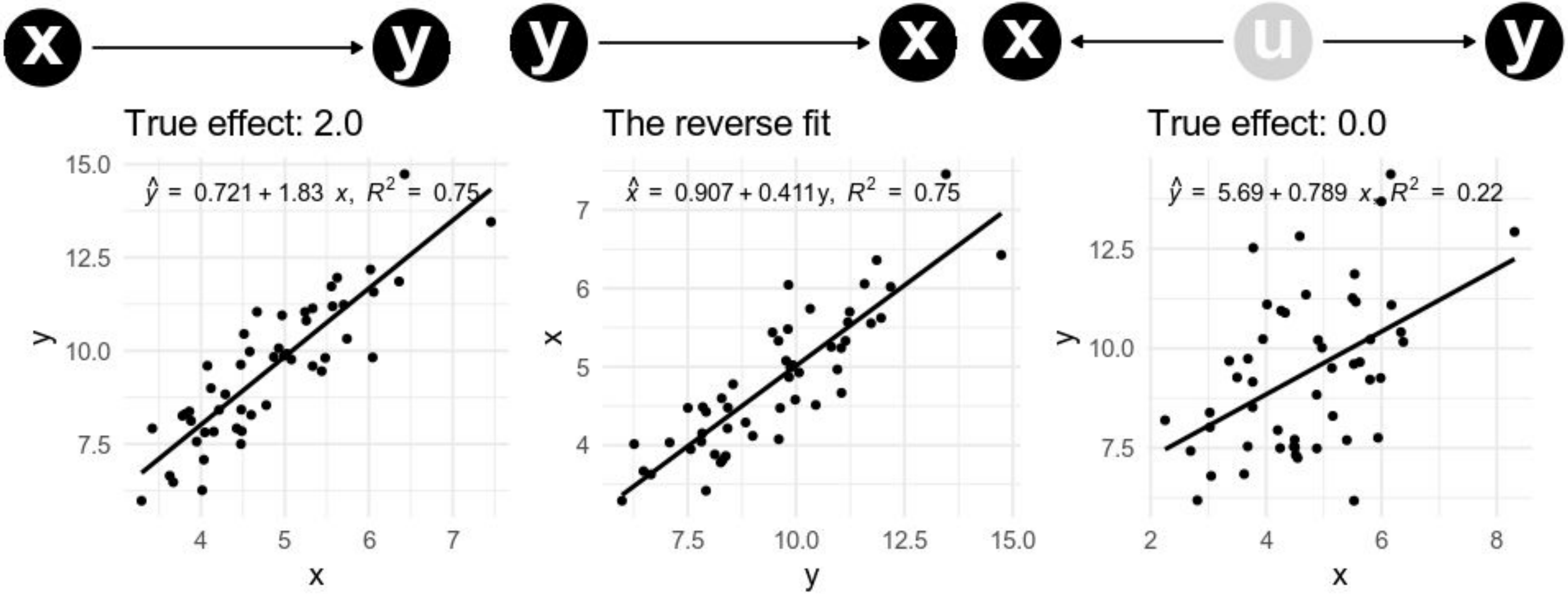
Zhang *et al.*, The Pharmacogenomics Journal (2014)

Machine learning expands and enhances statistical models



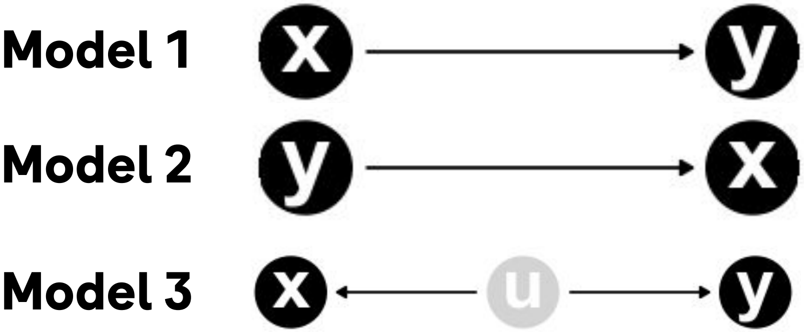
Badillo *et al.*, Clinical Pharmacology & Therapeutics (2020)

Correlation is caused by causation or confounding



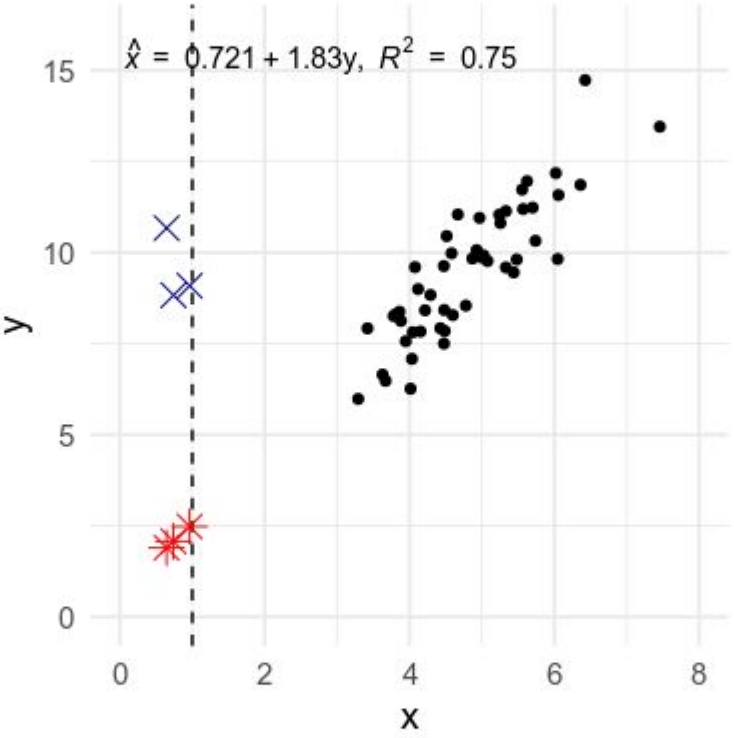
Statistical models alone cannot derive causality from correlation

We learn causality by (1) listing models explicitly and (2) manipulating a variable and observe the outcomes



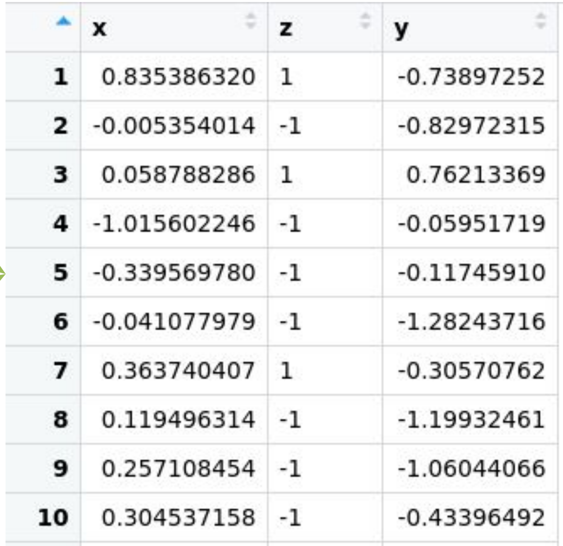
Assume that the data is generated by either Model 1, or Model 2, or Model 3. And assume that we can manipulate the value of X by setting it to 1.0 (the dash line).

Question: which outcomes (red stars or blue crosses) would support which models? Why?



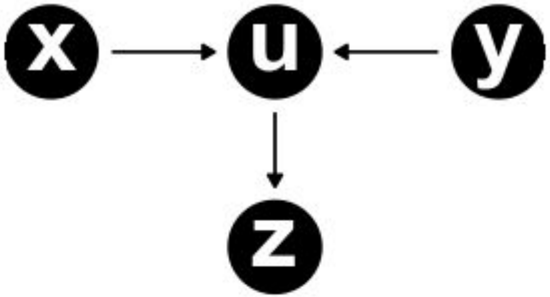
Causality is crucial for drug discovery

Biomarker, tox study, pathology, omics data, real-world data, EHR, …



	x	z	y
1	0.835386320	1	-0.73897252
2	-0.005354014	-1	-0.82972315
3	0.058788286	1	0.76213369
4	-1.015602246	-1	-0.05951719
5	-0.339569780	-1	-0.11745910
6	-0.041077979	-1	-1.28243716
7	0.363740407	1	-0.30570762
8	0.119496314	-1	-1.19932461
9	0.257108454	-1	-1.06044066
10	0.304537158	-1	-0.43396492

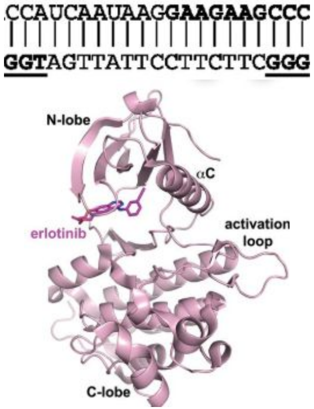
The descendant



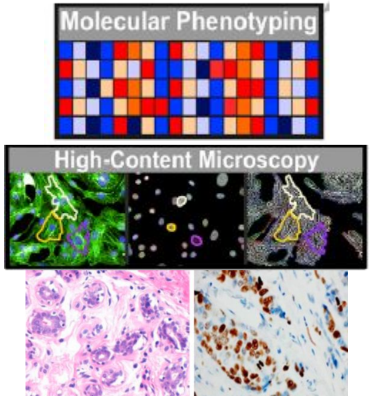
We need both models (knowledge + assumptions) and data to infer causality.

Multiscale Modelling of Drug Mechanism and Safety

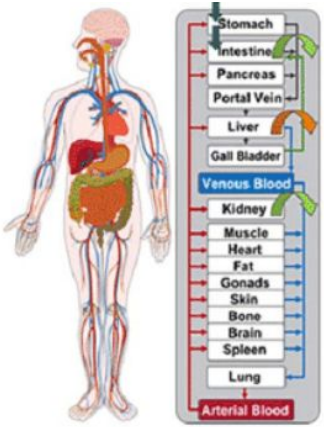
Forward translation



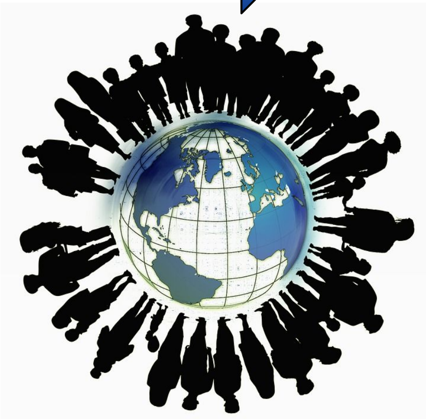
Molecular modeling



Omics & cellular modeling



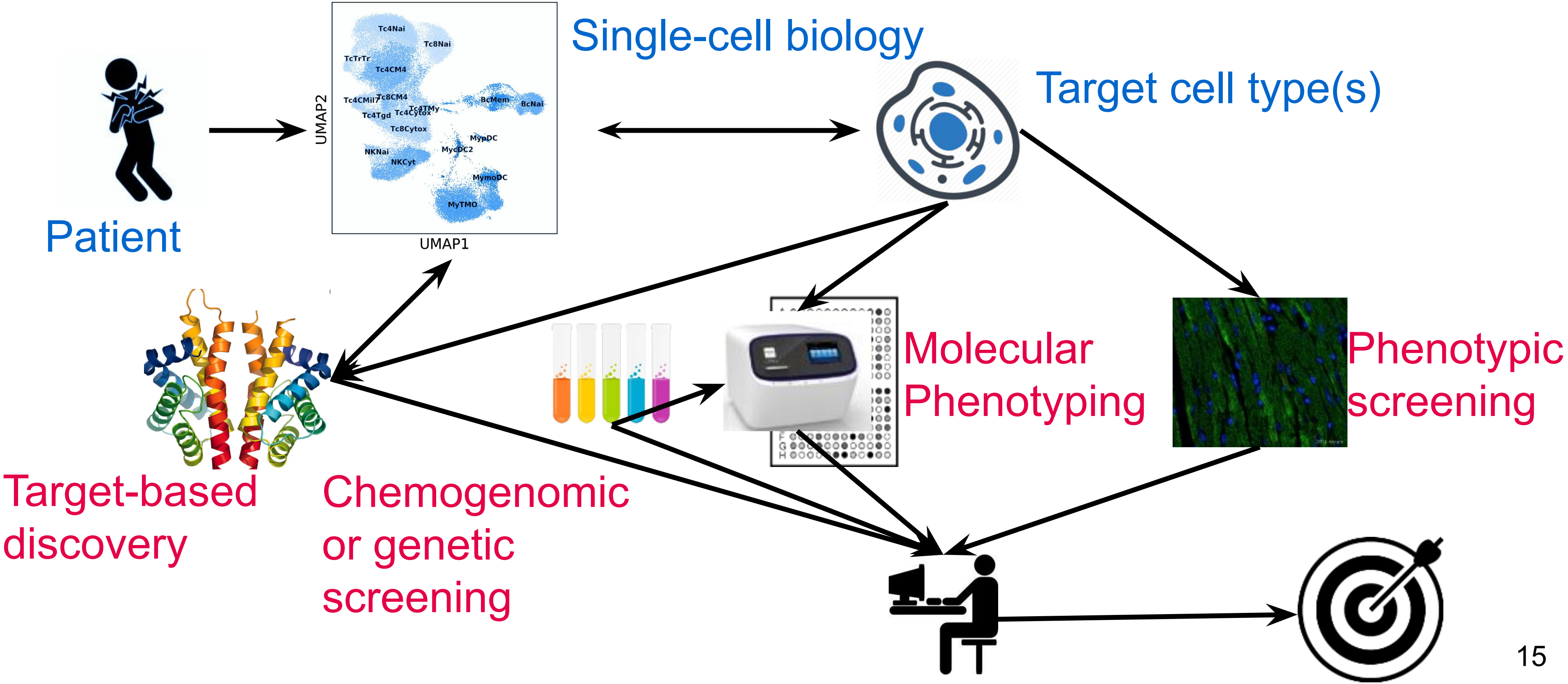
Organ & system modeling



Populational modeling

Reverse translation

Computational biologists work with experimentalists to empower drug discovery



Patient

Single-cell biology

Target cell type(s)

Molecular Phenotyping

Phenotypic screening

Target-based discovery

Chemogenomic or genetic screening

Conclusions

- Computational biologists in drug discovery build and use mechanistic, causal, and statistical/machine-learning models.
- Model and data offer insights into disease biology, as well as into pharmacology and safety profiles of drug candidates.
- We experiment with a multiscale-modeling approach to drug discovery by integrating data and models at molecular, cellular, organ and system, and population levels.

Interested in learning more about quantitative aspects of drug discovery?

- Fall Semesters: *Applied Mathematics and Informatics In Drug Discovery* (AMIDD): an introductory course.
  - See more information at http://AMIDD.ch.
- Spring Semesters: *Mathematical and Computational Biology in Drug Discovery* (MCBDD), an advanced course.
  - See more information at http://MCBDD.ch.

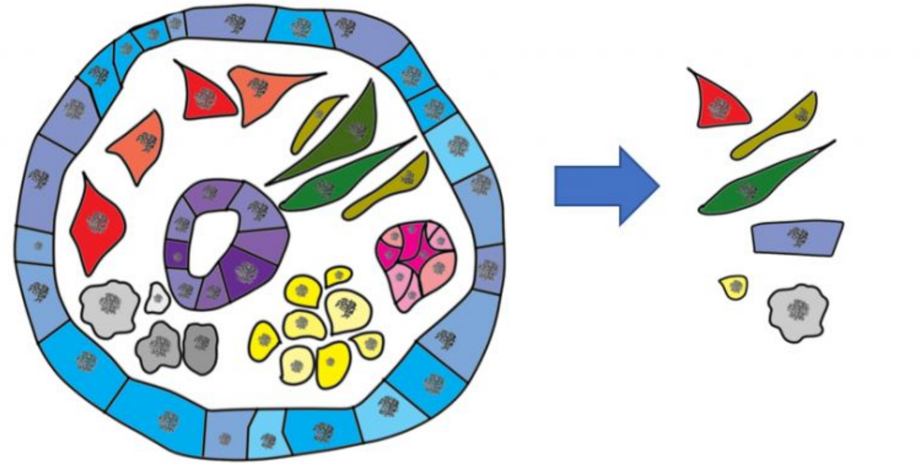
Doing now what patients need next

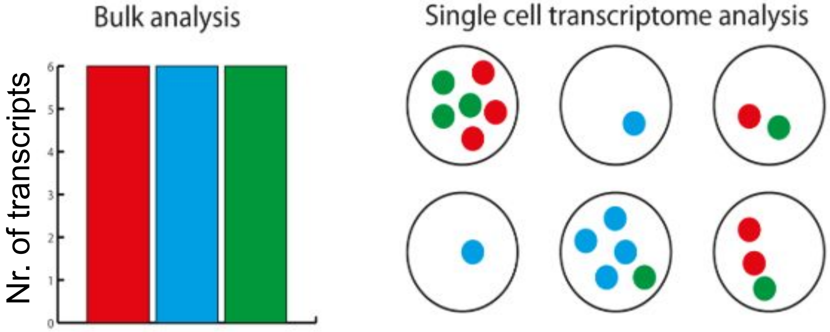
Backup material

Types of computational models

	Prediction for independent and identically distributed samples	Prediction under changing distributions or intervention	Answer counterfactual questions	Obtain physical insight	Learn from data (data-driven discovery)
Mechanistic models	yes	yes	yes	yes	maybe
Causal models	yes	yes	maybe	maybe	maybe
Statistical models	yes	no	no	no	yes

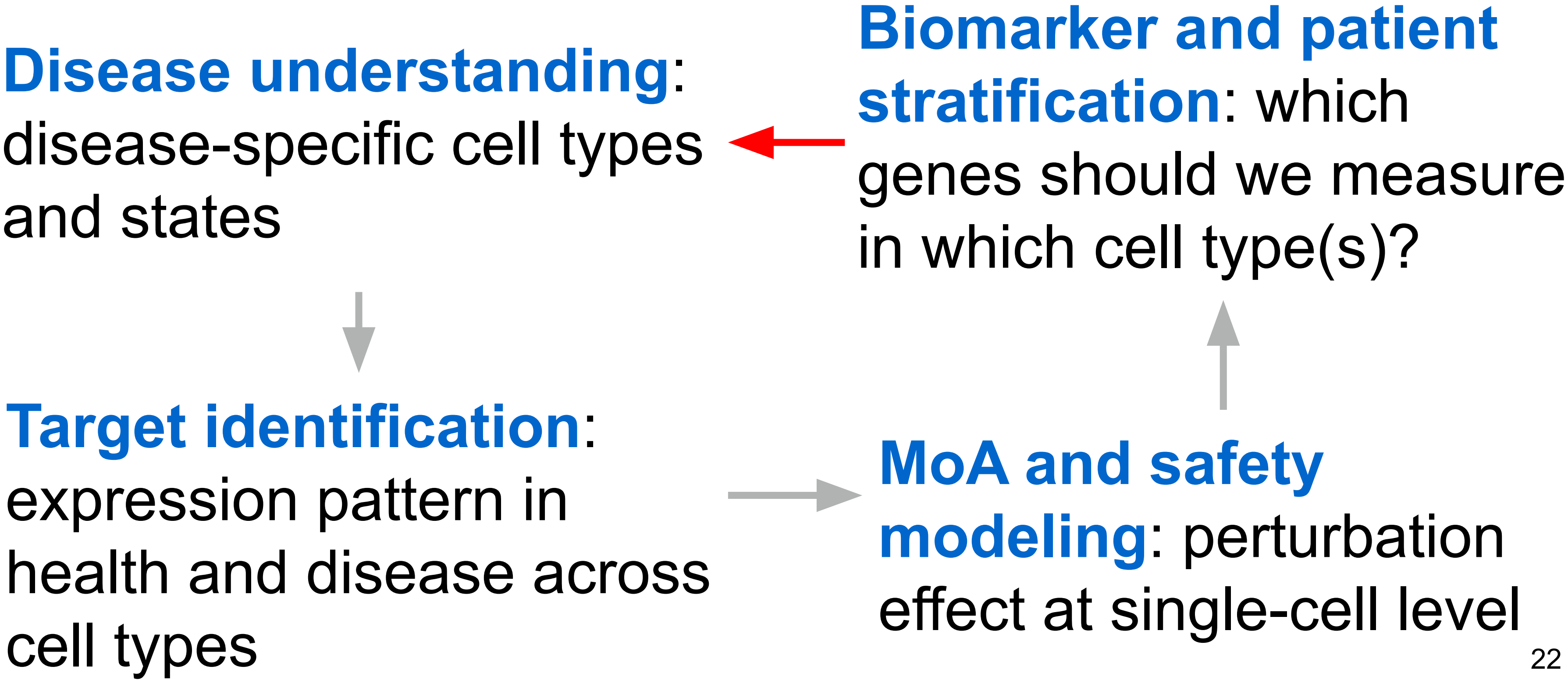
Target identification and mechanism & safety understanding benefits from single-cell biology





www.evocell-itn.eu; Macaulay & Voet, PLoS Genetics 2014

Single-cell biology fuels drug discovery



Disease understanding: disease-specific cell types and states

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

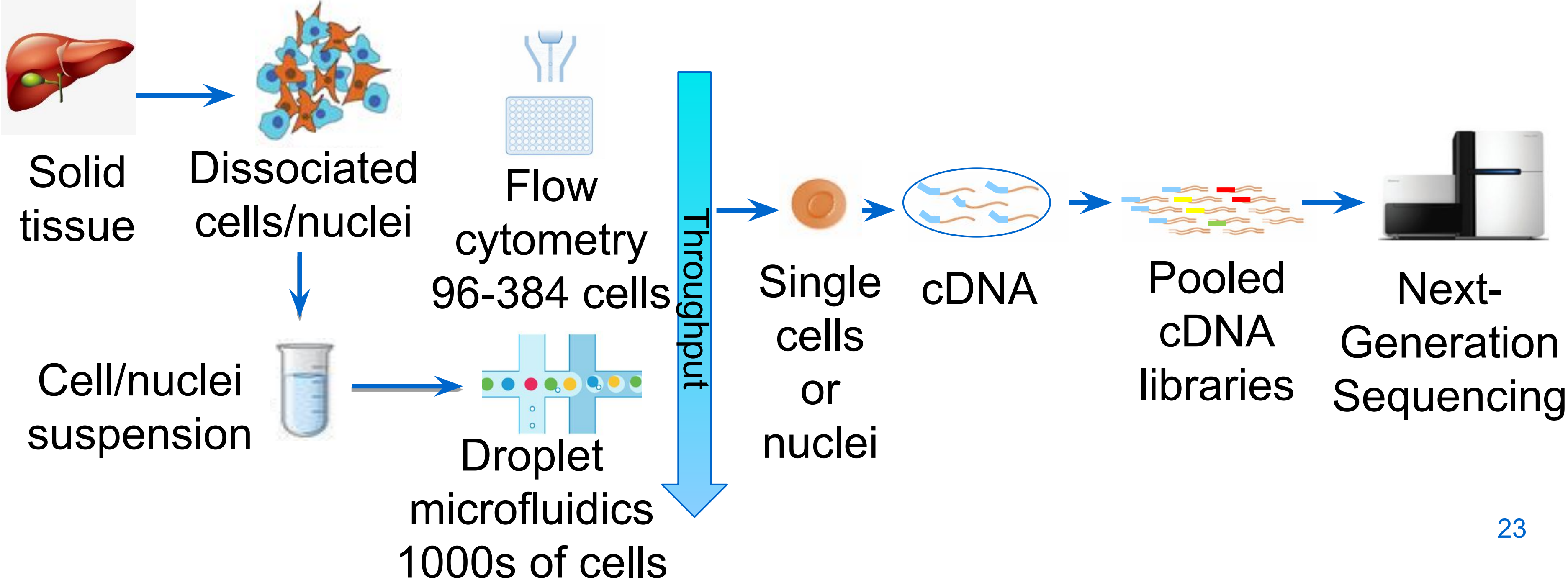
MoA and safety modeling: perturbation effect at single-cell level

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

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

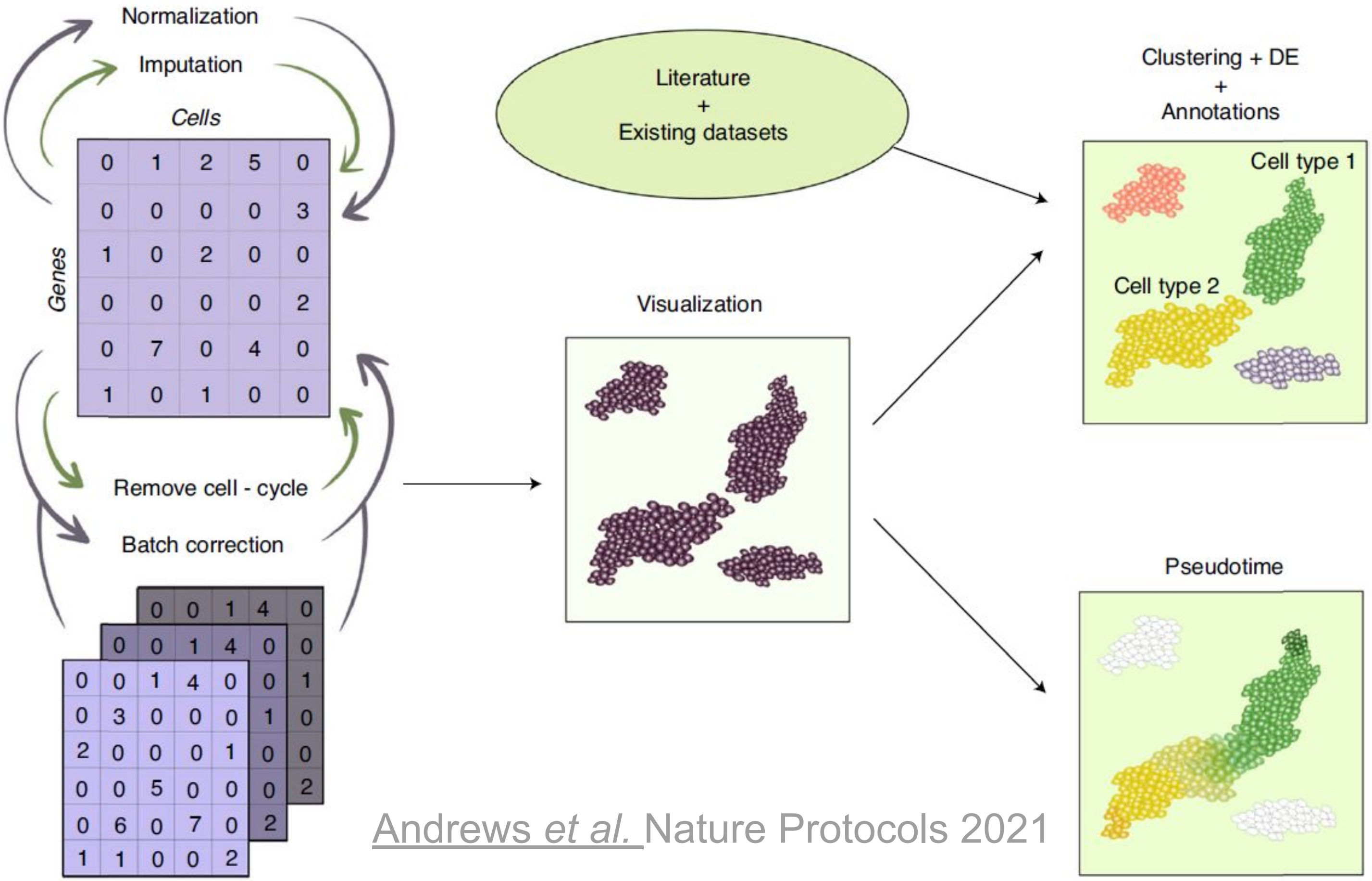
Single cells or nuclei

cDNA

Pooled cDNA libraries

Next-Generation Sequencing

Overview of the computational workflow



0	1	2	5	0
0	0	0	0	3
1	0	2	0	0
0	0	0	0	2
0	7	0	4	0
1	0	1	0	0

0	0	1	4	0
0	3	0	0	0
2	0	0	0	1
0	0	5	0	0
0	6	0	7	0
1	1	0	0	2

Andrews *et al.* Nature Protocols 2021

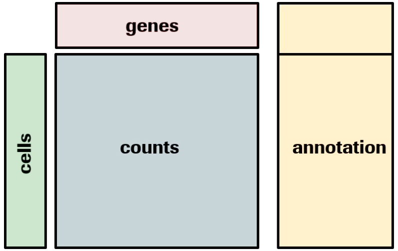
Open-source BESCA (BEDA’s single cell analysis)

An automatized standard workflow

Data loading → Preprocessing → Clustering → Visualization → Analysis

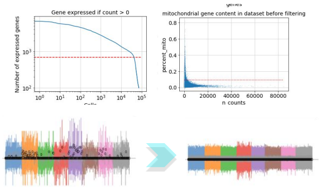
Data loading

From FAIR format



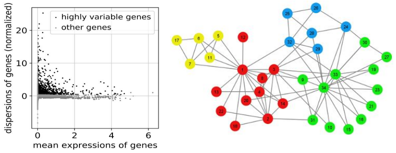
Preprocessing

Quality Control
Filtering,
Normalization



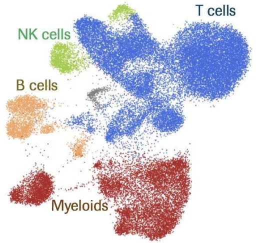
Clustering

Dimensionality reduction
Clustering



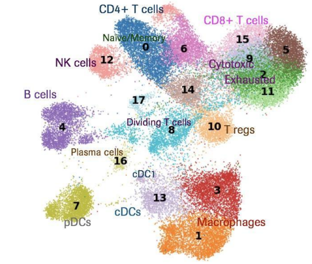
Visualization

UMAP



Analysis

Cell type characterisation



Maedler et al., NAR Genomics and Bioinformatics, 2021
https://github.com/bedapub/besca

Doing now what patients need next