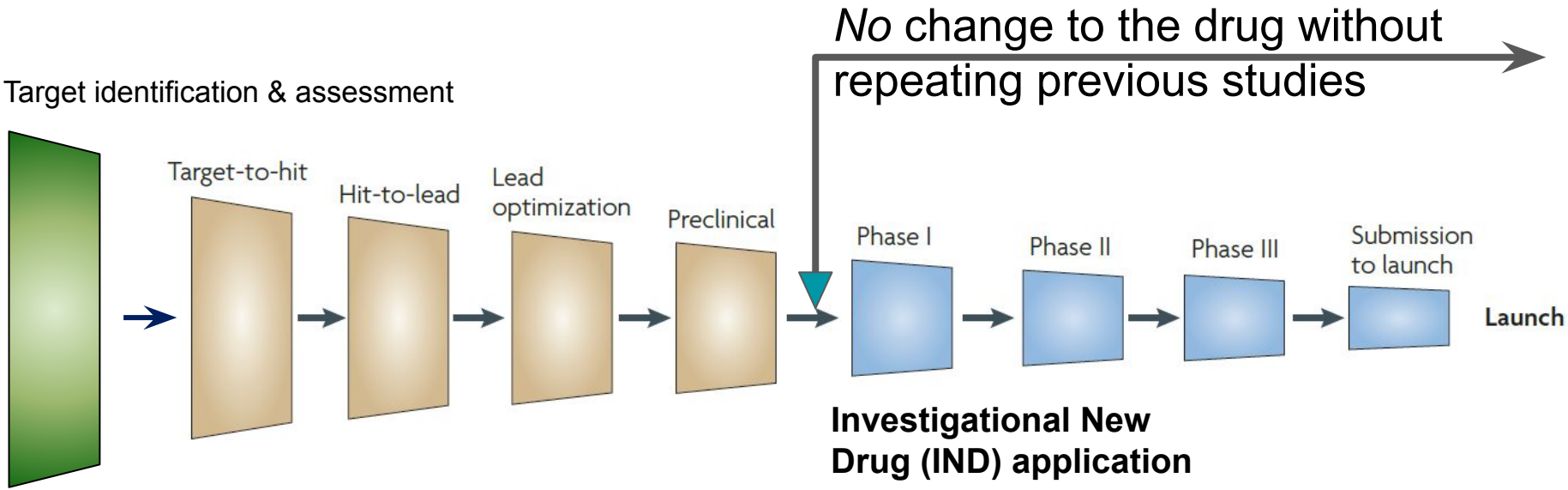
For which patients shall the drug work, and how?

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

Dr. Jitao David Zhang
May 2025

From drug discovery to drug development



Outline of Module V

- Lecture 11
  - Biomarker for dose prediction
  - Biomarker for patient-stratification and biology understanding: Merck/Genentech
  - Challenges and caveats
- Lecture 12
  - Integrating statistical and mechanistic modelling: Griffiths *et al.*
  - Mechanistic modelling of biological systems: from Boolean network to Agent-based modelling
  - Causal inference

Phases of clinical trials

Investigational New Drug (IND) application

New Drug Application (NDA)

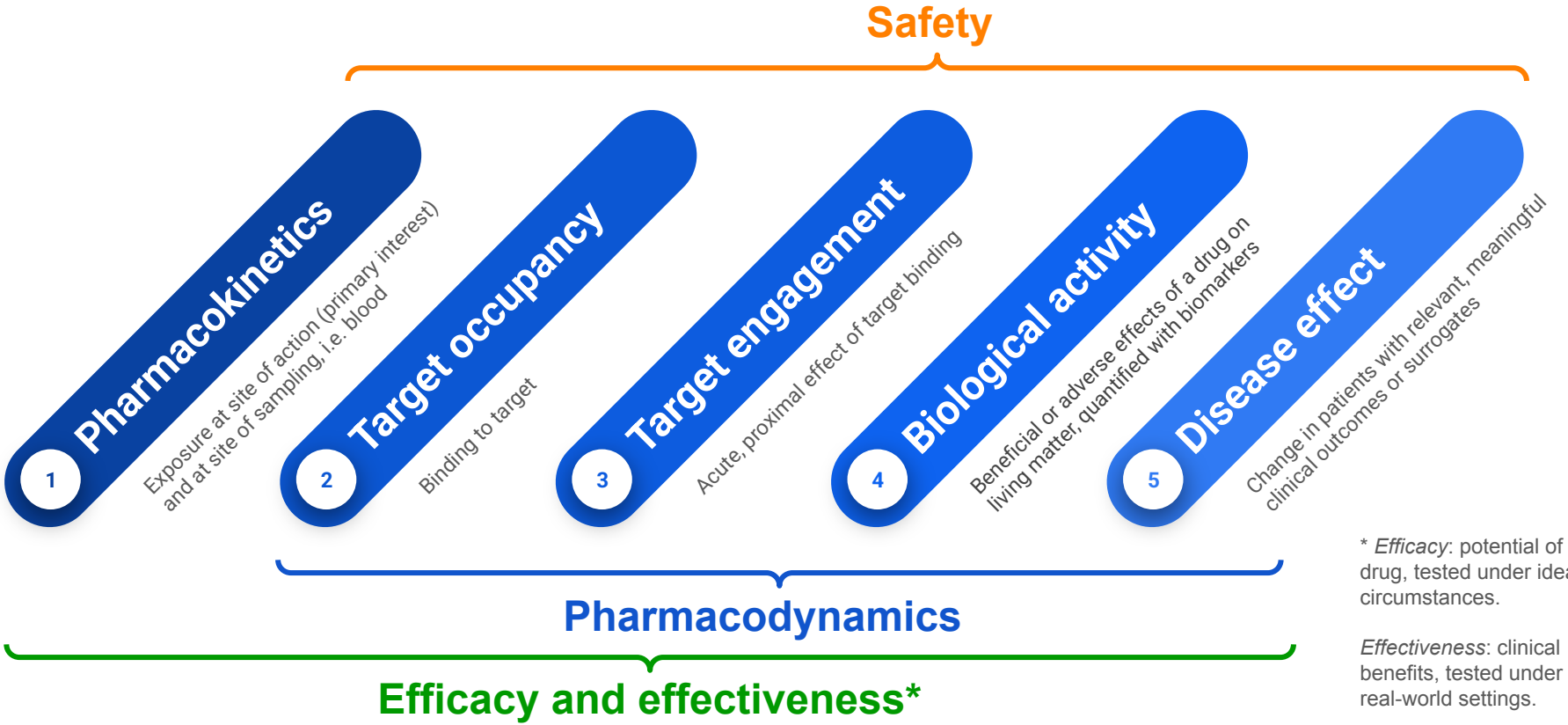
- **Aim:** Getting PK/PD data to verify the drug behaves as expected.
- **Dose:** *Microdosing*, e.g. *1% of predicted dose*.
- **Subjects:** <15 healthy subjects
- **Time:** A few weeks

- Finding safe dose ranges and optimal dosing regimens with further PK/PD data.
- Sub-therapeutic single and multiple ascending doses
- 20-100 healthy subjects (patients)
- A few months

- Assessing efficacy and safety profiles of the drug, and determining the dosing regimen.
- Therapeutic dose
- Usually 100-300 patients with a specific disease
- A year or longer

- Comparing efficacy, effectiveness, and safety profiles with the standard-of-care treatment option.
- Therapeutic dose
- Usually 300-3000 patients
- Usually several years

The chain of translation

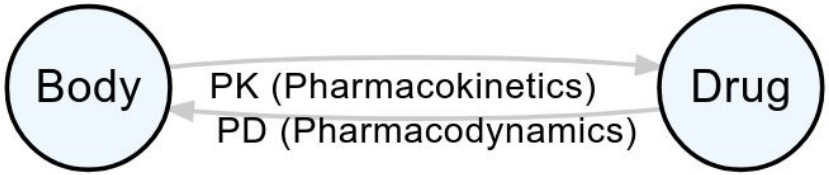


* *Efficacy*: potential of a drug, tested under ideal circumstances.

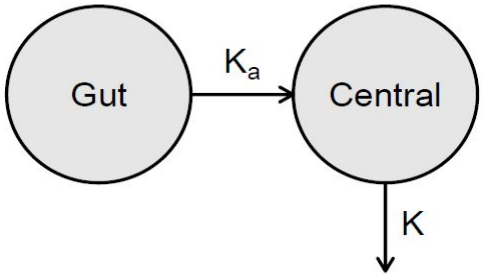
Effectiveness: clinical benefits, tested under real-world settings.

A refresher of PK/PD Modelling

(A)



(B)

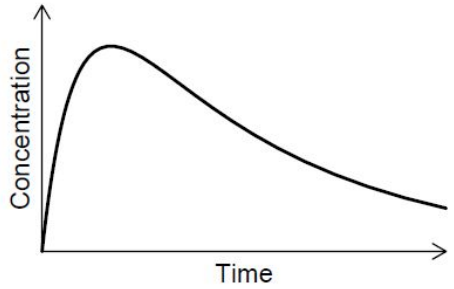


$$\frac{dA_{gut}}{dt} = -K_a \cdot A_{gut}$$

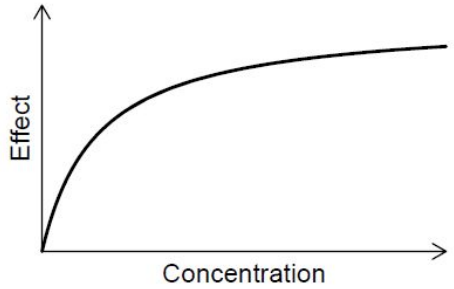
$$\frac{dA}{dt} = \overbrace{F \cdot K_a \cdot A_{gut}}^{\text{from gut}} - \overbrace{K \cdot A}^{\text{elimination}}$$

$$A_{oral}(t) = \frac{K_a F A_0}{K_a - K}\left(\exp(-K \cdot t) - \exp(-K_a \cdot t)\right)$$

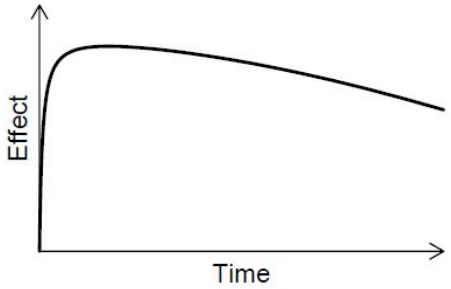
(C)



(a) PK model

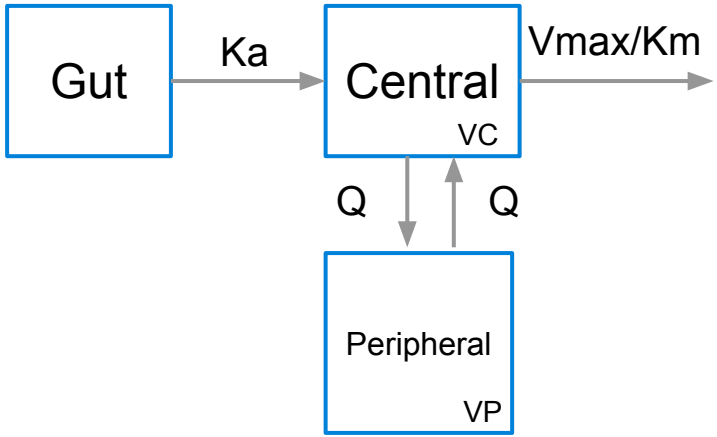


(b) PD model



(c) Combined PK/PD model

An example of a two-compartment PK model



Ordinary Differential Equation (ODE) based model of *in vivo* PK, assuming two compartments (central and peripheral), and the Michaelis-Menten model of drug elimination.

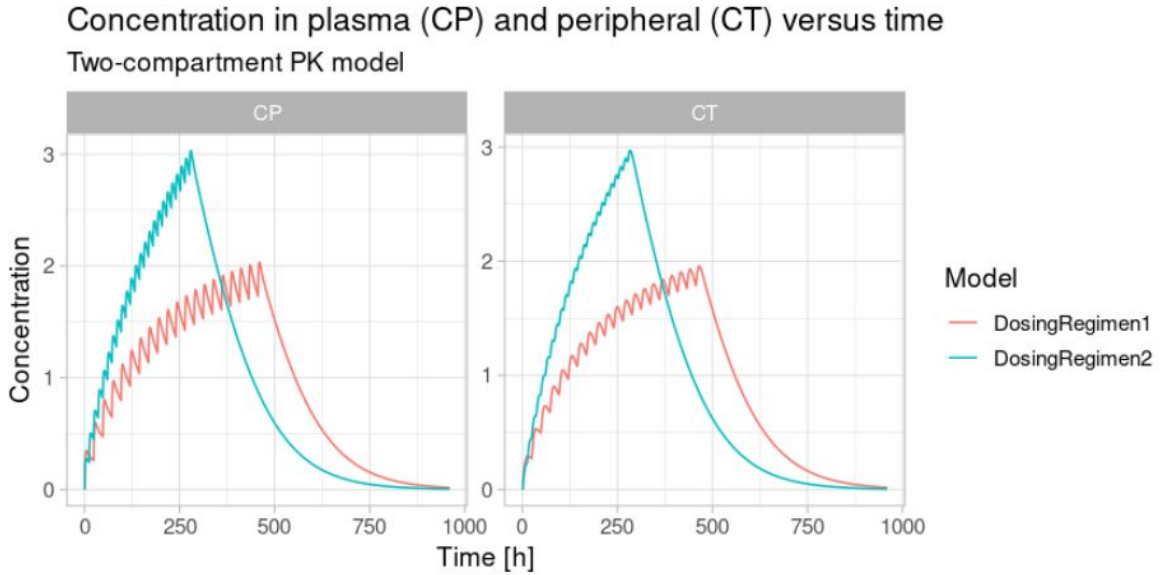
- **VC** (20) Central volume (volume)
- **Q** (2) Inter-compartmental clearance (volume/time)
- **VP** (10) Peripheral volume of distribution (volume)
- **Ka** (Ka, 0.5) Absorption rate constant (1/time)
- **Vmax** (1) Maximum velocity of elimination (mass/time)
- **Km** (3) Michaelis constant for elimination (mass/volume)

Values of the parameters derive from *in vitro* assays (for instance *Vmax* and *Km*), previous *in vivo* studies, or predictions (for instance with machine learning).

Simulating two dosing regimens

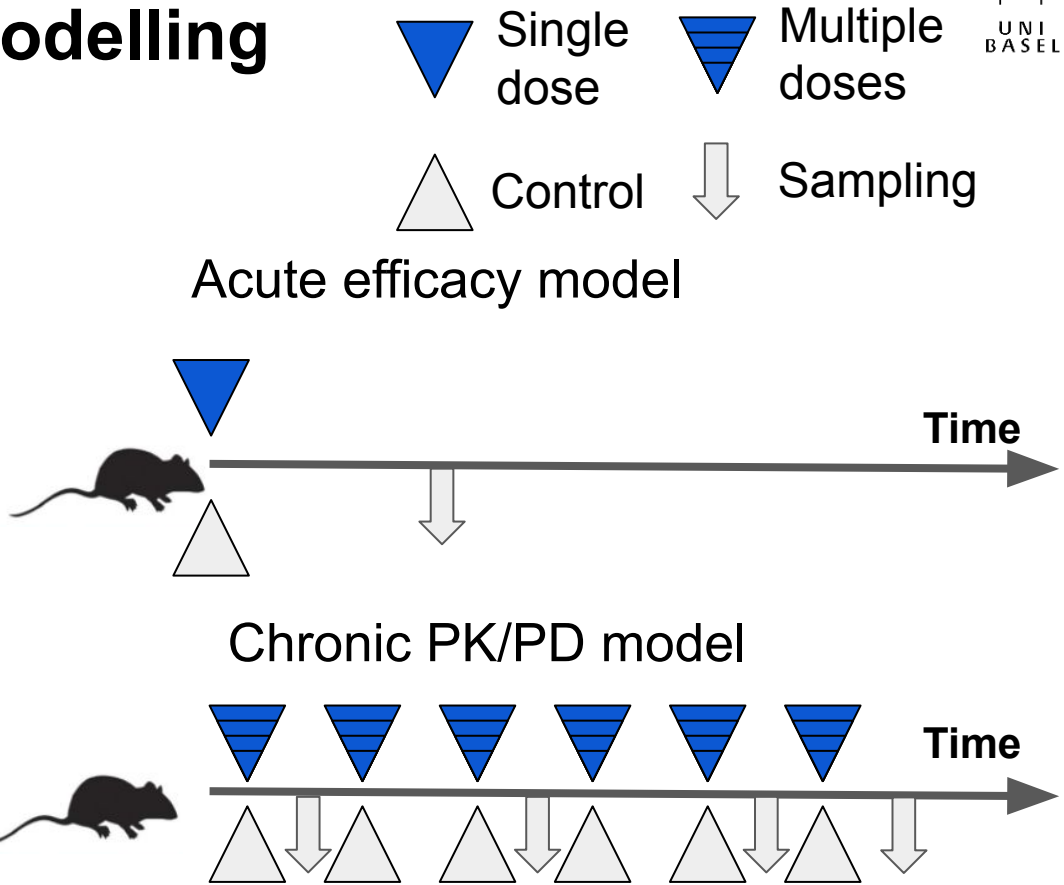
Dosing regimen 1: dosing **8** units per **12** hours, for **24** doses

Dosing regimen 2: dosing **10** units per **24** hours, for **20** doses.

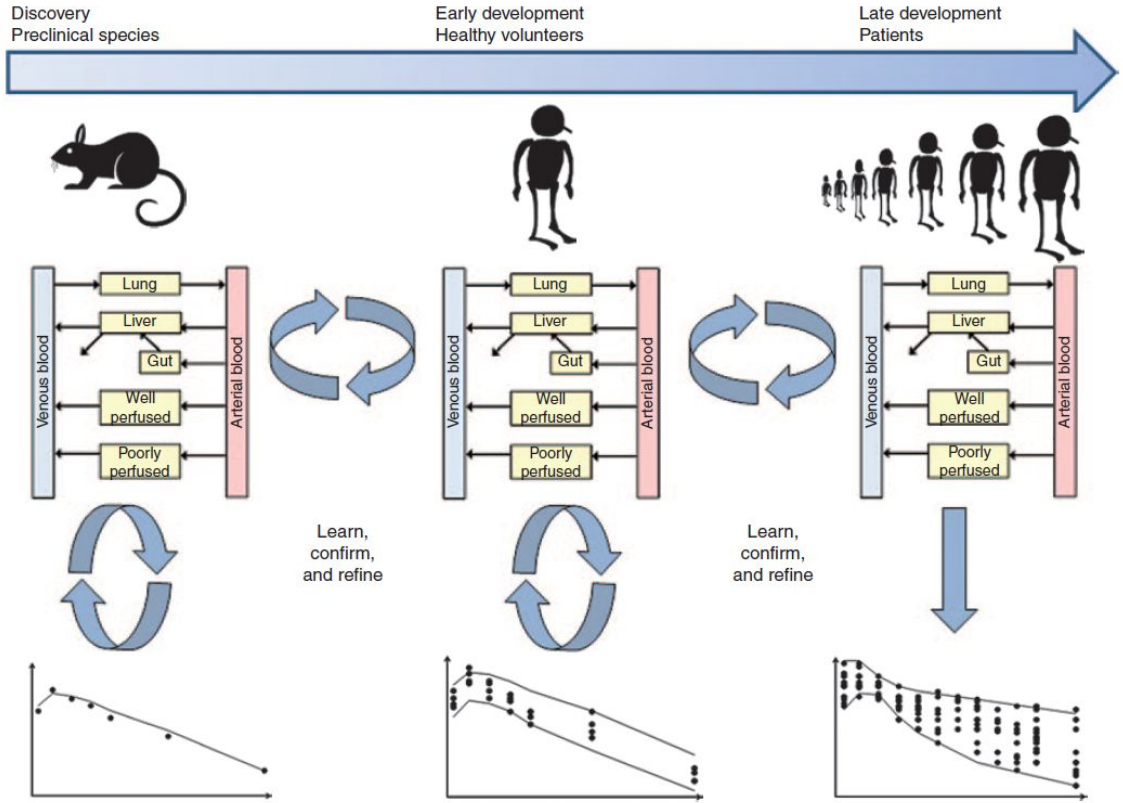


Translational PK/PD Modelling

Samples from blood and tissues or organs organ can be analysed for pharmacokinetics, pharmacodynamics, and dose-exposure-response relationships.



Physiologically-based pharmacokinetic modelling (PBPK) is a natural extension of PK modelling



Empirical, stratified, and individualized medicine

Empirical medicine

- Vaccines
- Non-steroid anti-inflammatory drugs (NSAIDs)

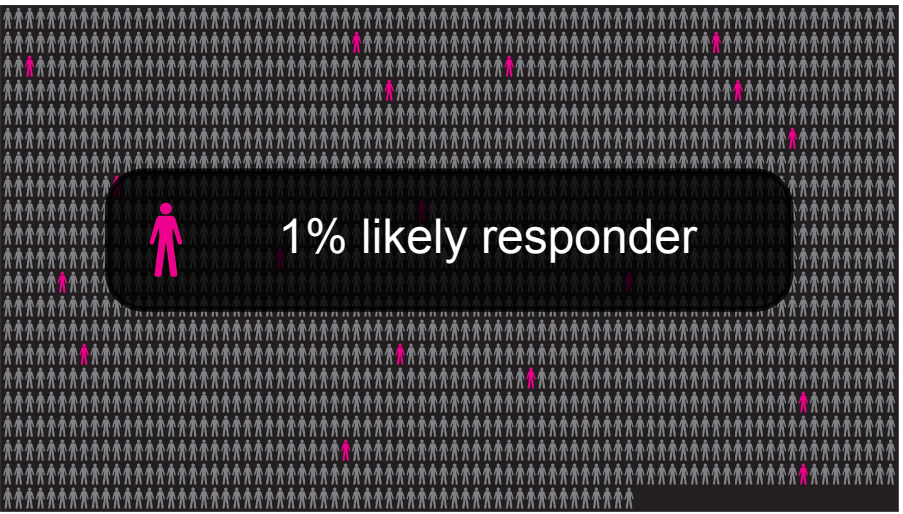
Stratified Medicine

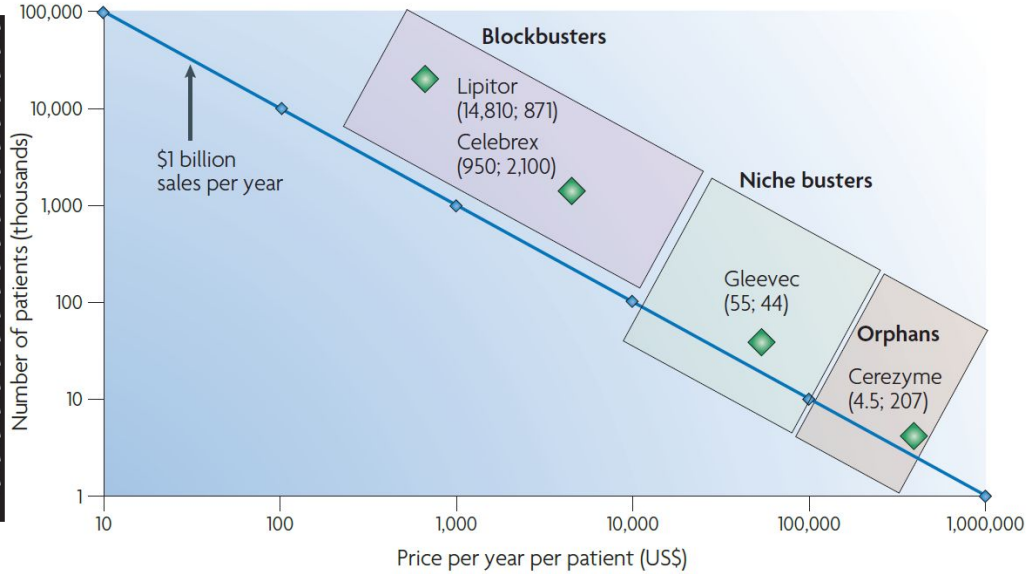
- Vemurafenib (Zelboraf)
- Trastuzumab (Herceptin)

Individualized medicine

- CAR-T therapy

Why stratified medicines are becoming popular?



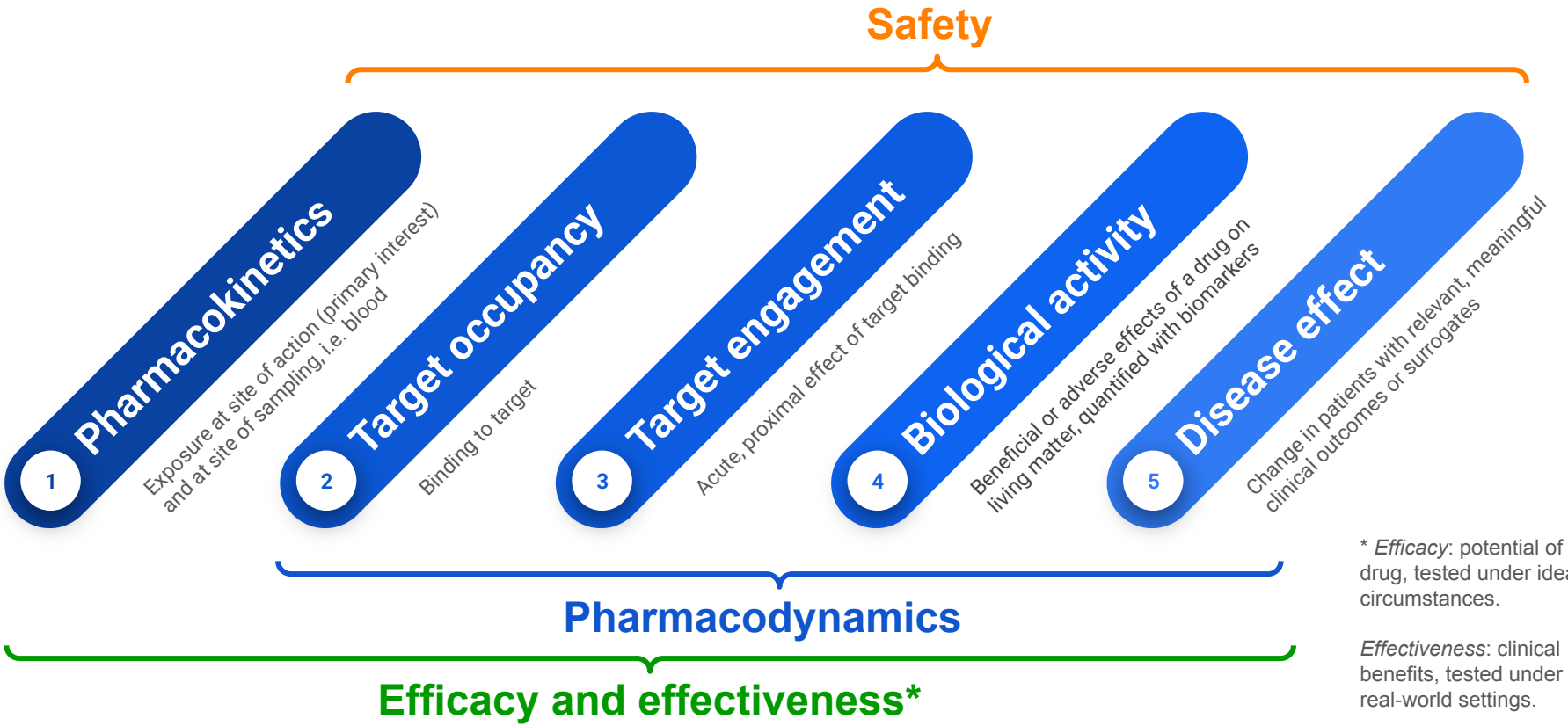


Medical reasons

Commercial reasons

Lecture on 16.05.2025 ends here

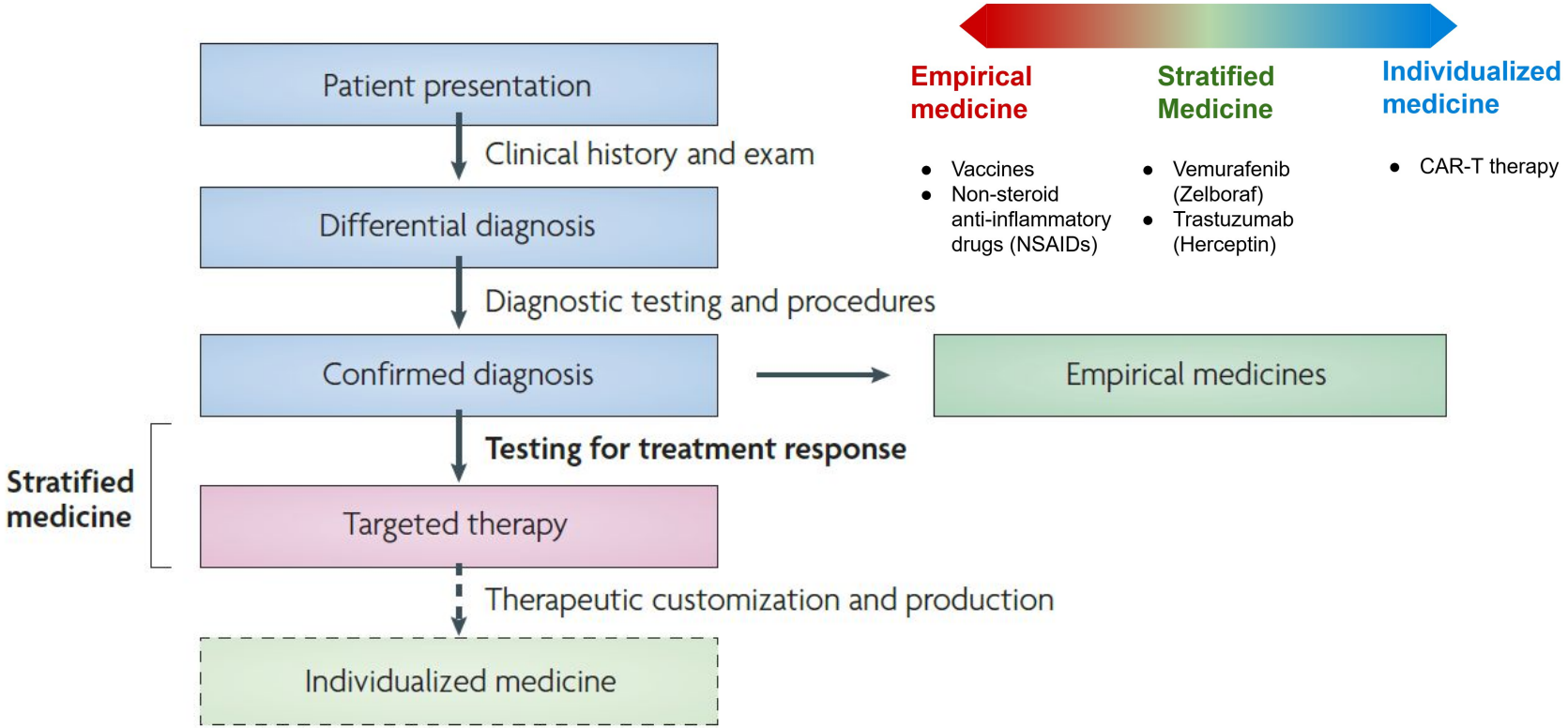
The chain of translation



* *Efficacy*: potential of a drug, tested under ideal circumstances.

Effectiveness: clinical benefits, tested under real-world settings.

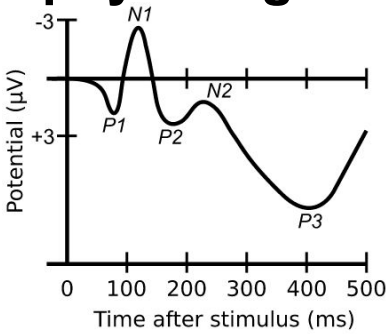
Empirical, stratified, and individualized medicine in the clinical context



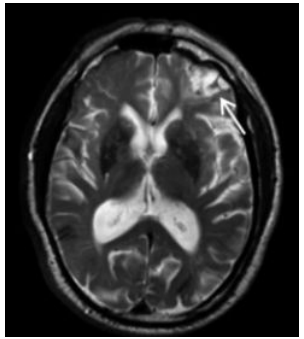
Biomarkers

A objectively measured and evaluated characteristic as an indicator of (1) normal biological process, (2) pathogenic processes, or (3) pharmacological responses to a therapeutic intervention.

Electro-physiological

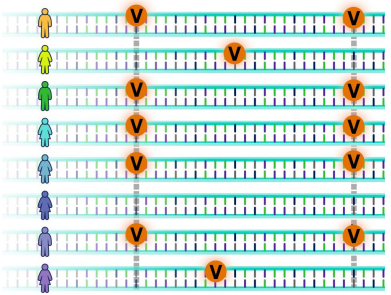


Imaging



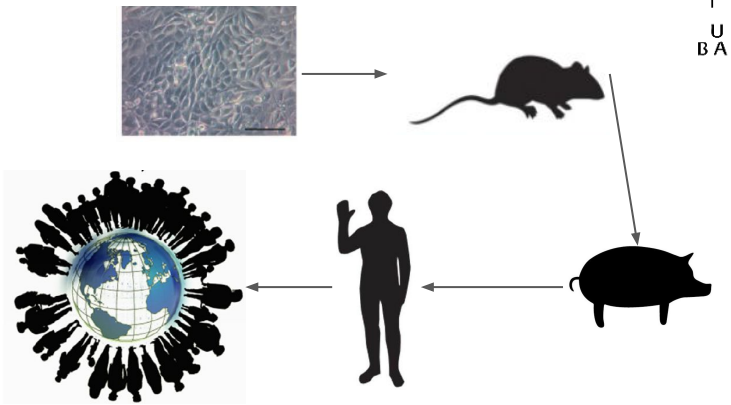
Functional

Molecular



Applications of biomarkers

1. Compound optimization and differentiation from competitors in preclinical study
2. **Human-dose prediction in translational PK/PD modelling**
3. **Patient stratification in clinical studies**

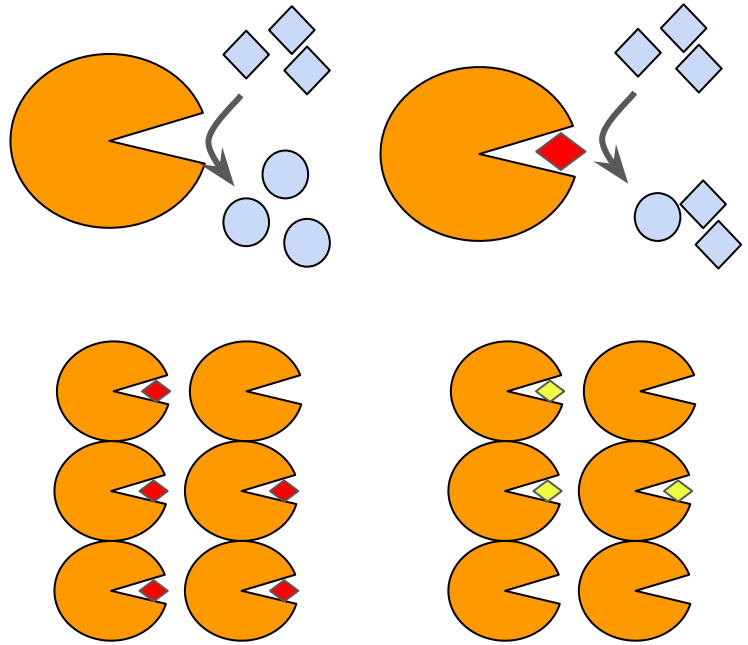


Target Occupancy and Engagement

Target occupancy is the percentage of the protein target occupied by drugs.

Target occupancy affects **target engagement,** which describes the process a drug interacts with its intended protein target in a living system to induce downstream effects. An occupied target is not necessarily engaged: the mode of binding and the physiological context matters.

The mode of binding and the downstream effects are known as the Mechanism of Action (MoA) of the drug.

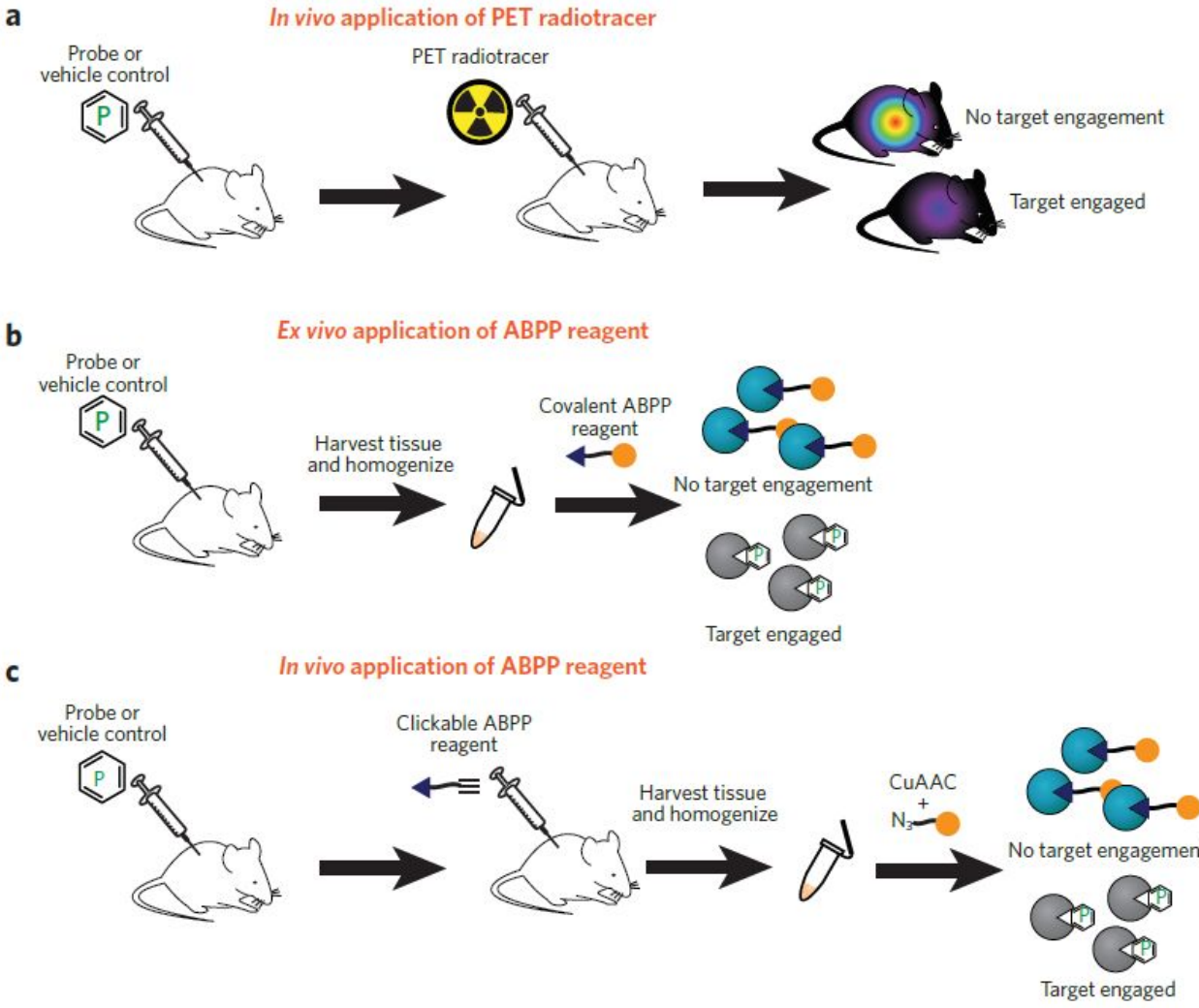


Target occupancy of 83% and 50%, respectively

Target occupancy and engagement profiling in vivo

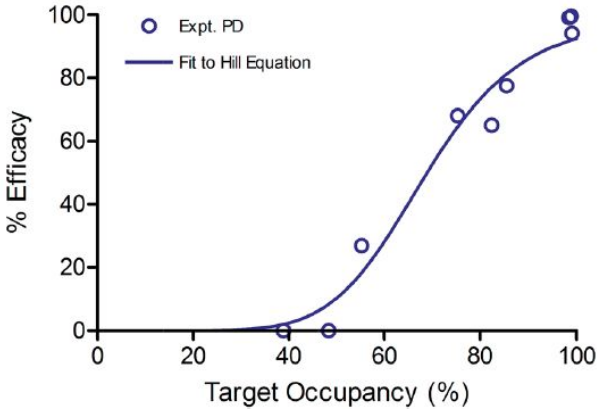
ABPP: Activity-based protein profiling; PET: positron-emission tomography.

Both ABPP reagent and radiotracer binds to the same protein target.

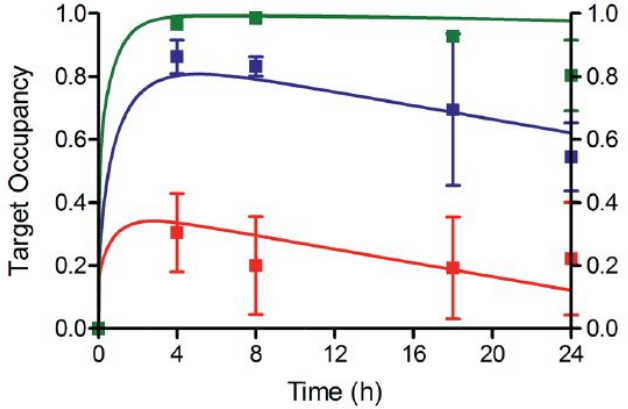


Target occupancy as a biomarker links pharmacokinetics and pharmacodynamics

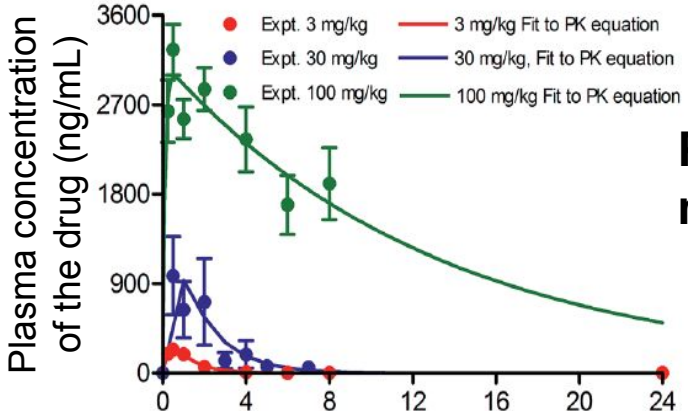
PD modelling



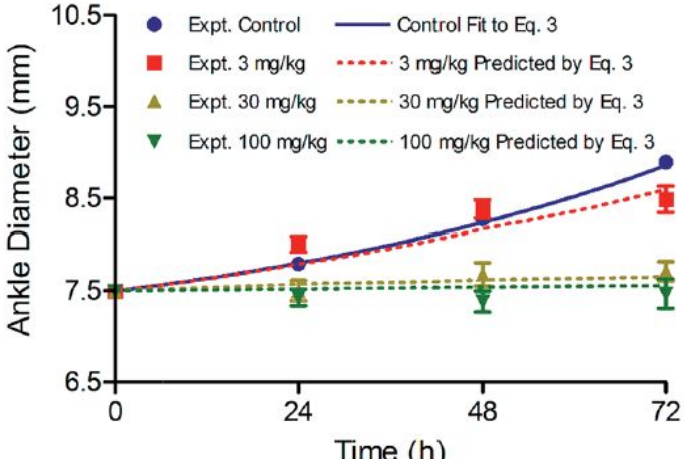
Biomarker modelling



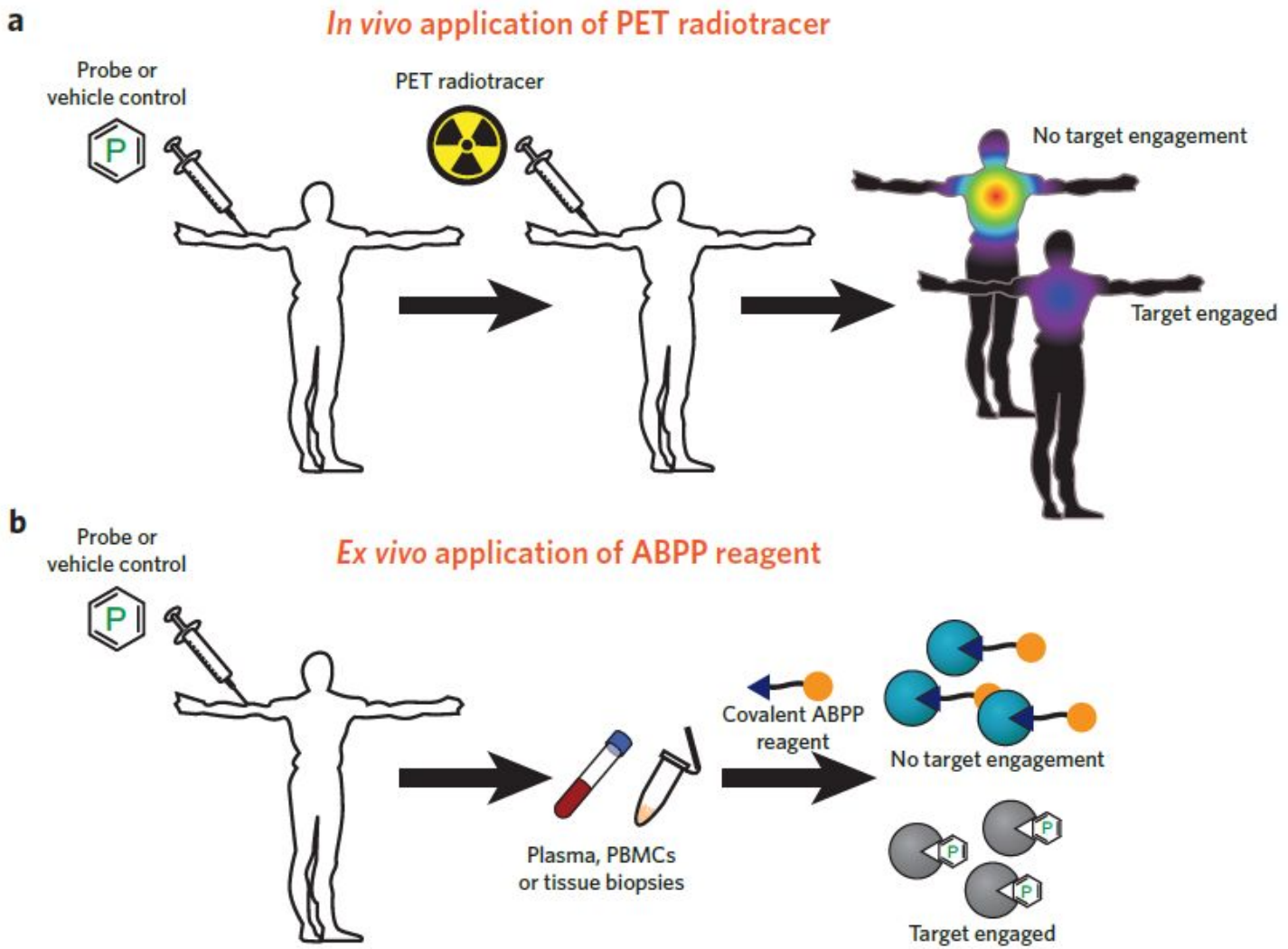
PK modelling



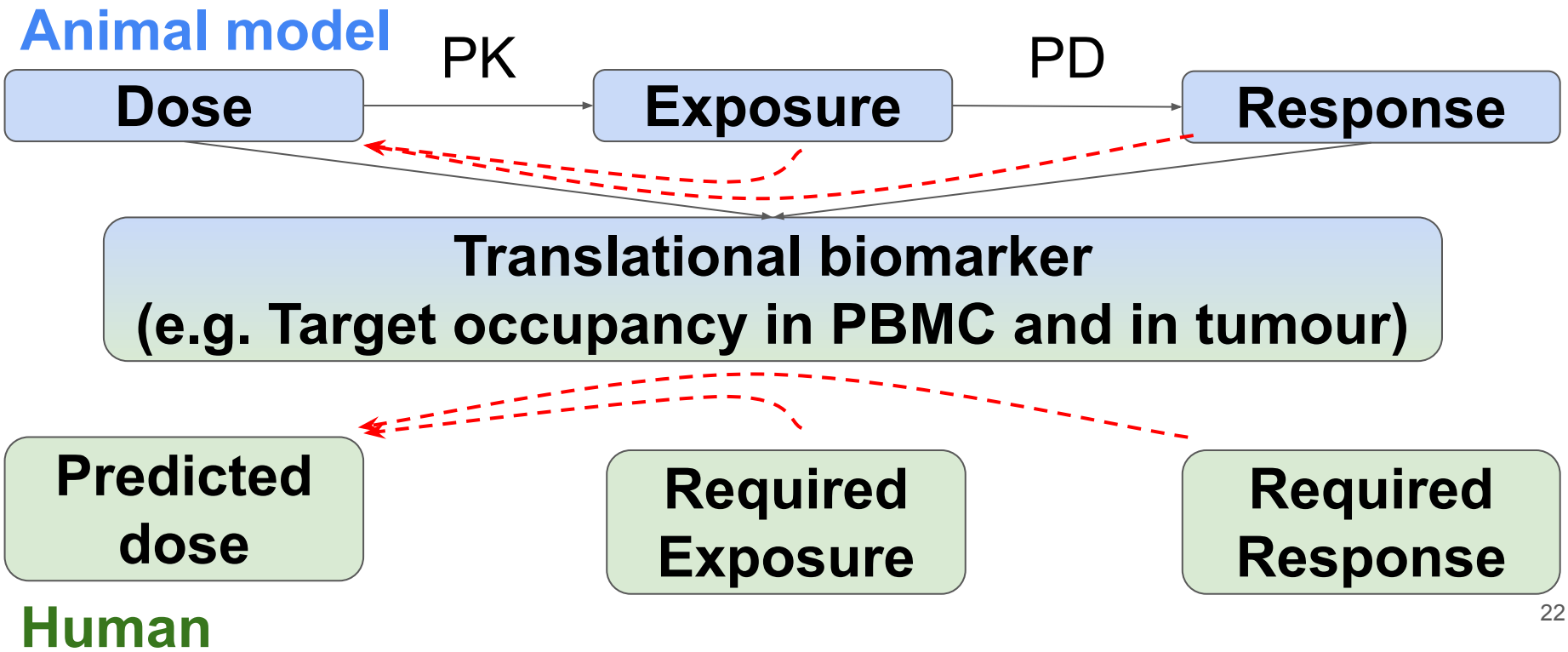
PK/PD modelling



Target occupancy and engagement profiling in human

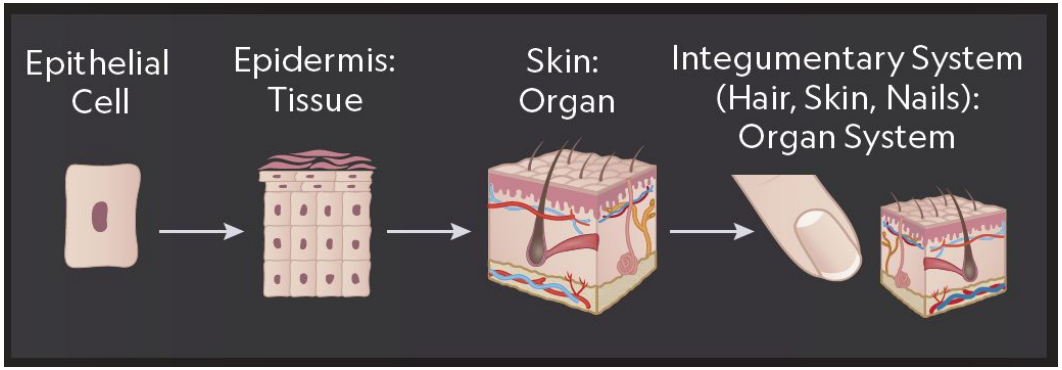


Use of translational biomarker



Once in human, what does the drug do?

Complexity Increases Through a System

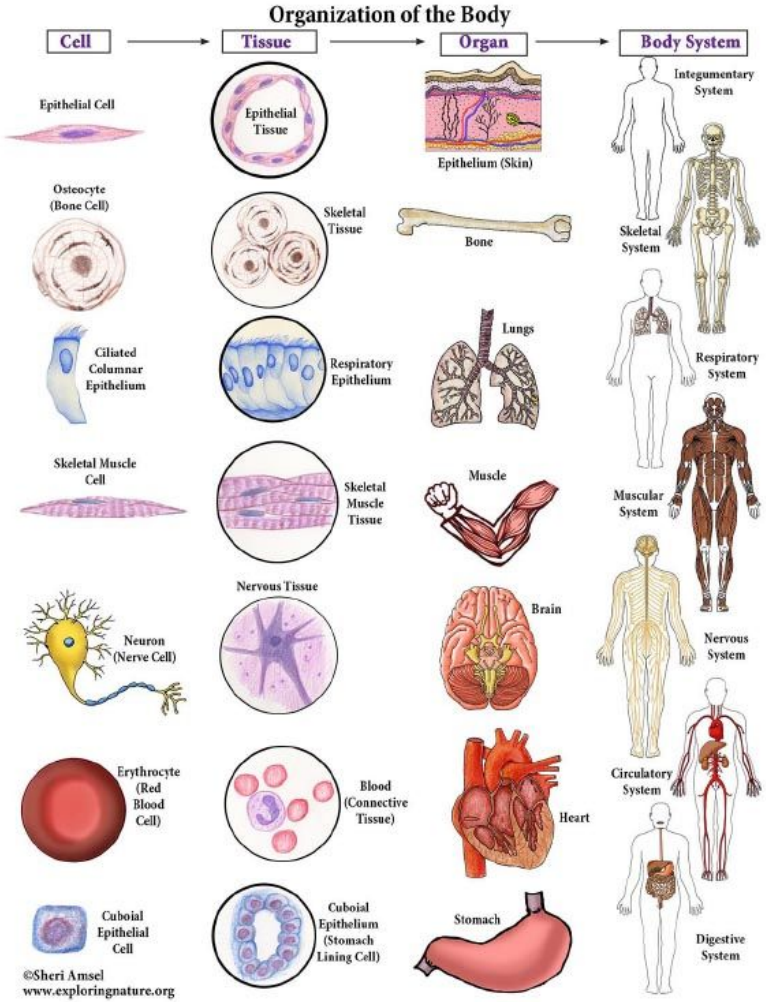


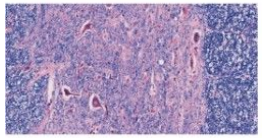
Cells: basic building blocks, variable morphologies and functions

Tissues: groups of specialized cells that communicate and collaborate

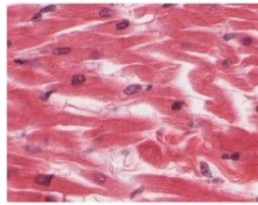
Organ: group of tissues to perform specific functions

Organ systems: group of organs and tissues



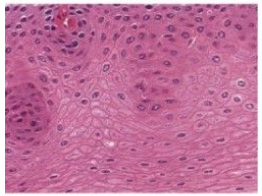


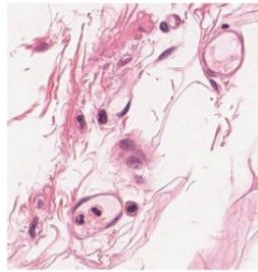
Nervous tissue
Brain
Spinal cord
Nerves



Muscle tissue
Cardiac muscle
Smooth muscle
Skeletal muscle

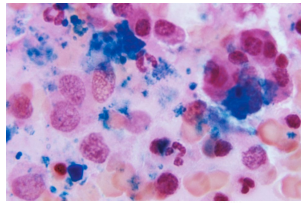
Epithelial tissue
Lining of GI tract organs and other hollow organs
Skin surface (epidermis)

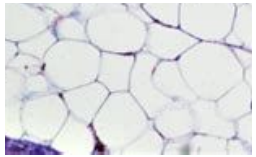


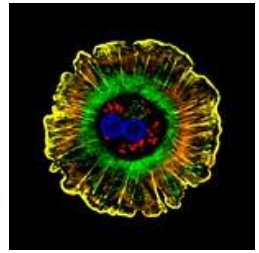


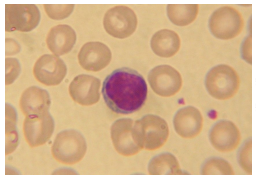
Connective tissue
Fat and other soft padding tissue
Bone
Tendon

Four major tissue types







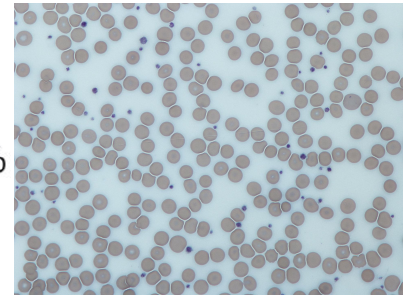


adipocytes 0.2%

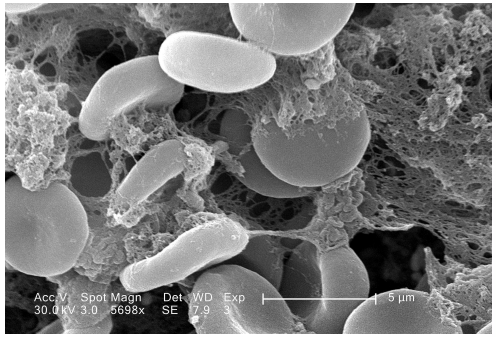
lymphocytes 1.5%

bone marrow cells 2.5%

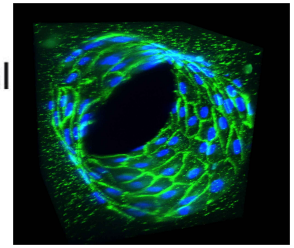
hepatocytes 0.8%



platelets 4.9%

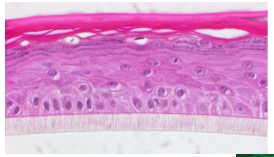


erythrocytes, 84%
25×10^{12} of 30×10^{12}
total human cell count



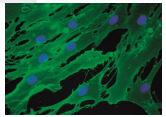
bronchial endothelial cells 0.5%

vascular endothelial cells 2.1%



epidermal cells 0.5%

respiratory interstitial cells 0.5%

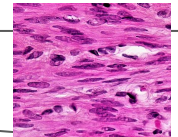


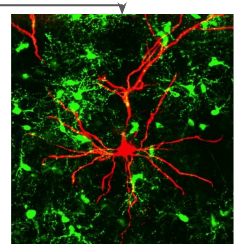
dermal fibroblasts 0.1%

muscle cells 0.001%

other 2.0%

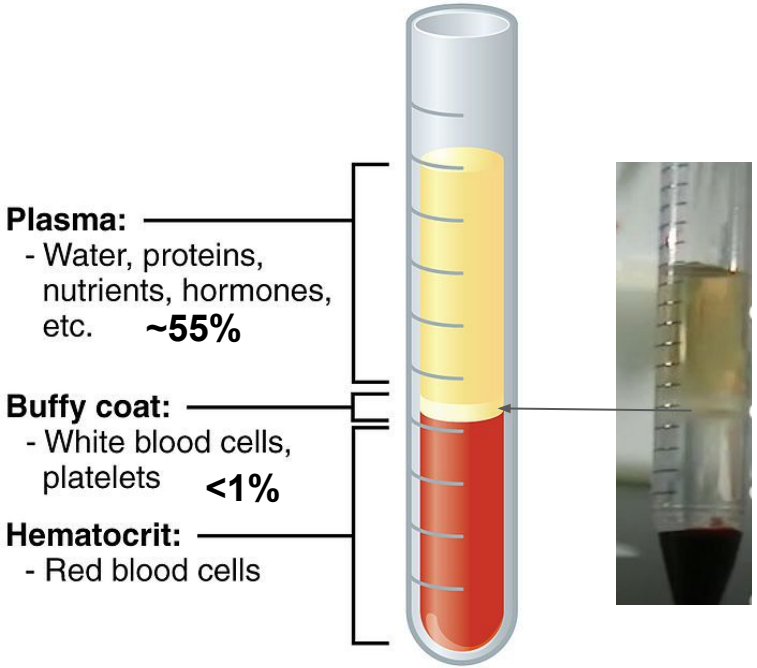
neurons and glia 0.6%

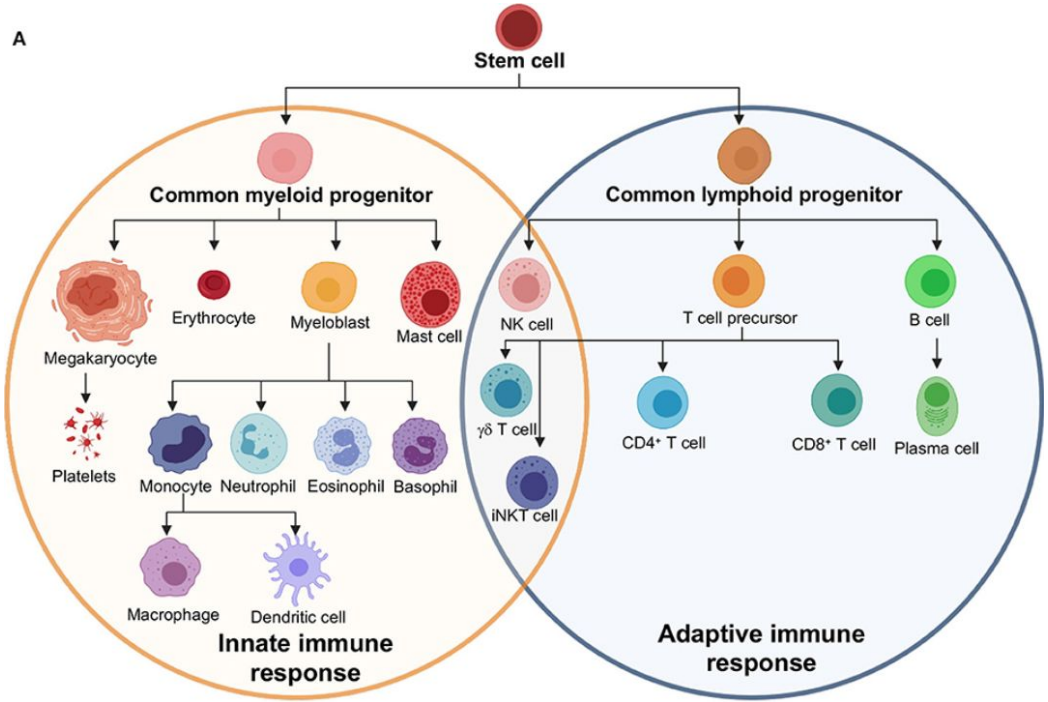




UNI BASEL

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

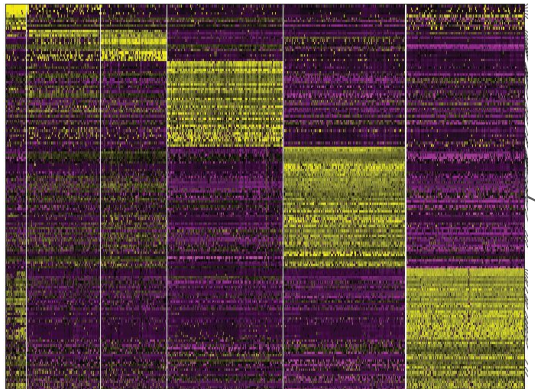




What's in a drop of blood? Count the genes!

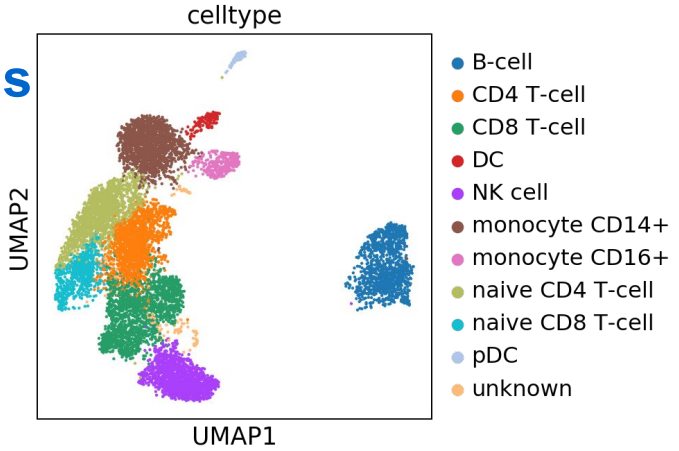
Sequencing

Genes



Cells

Data analysis



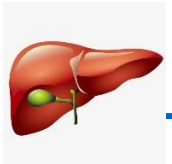
Low Expression

High Expression

Single-cell sequencing (scSeq) workflow

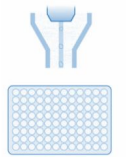
Tissue dissociation

Single cell capture and transcriptome sequencing



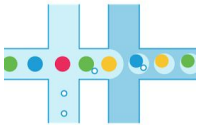
Solid tissue

Dissociated cells/nuclei



Flow cytometry 96-384 cells

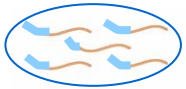
Cell/nuclei suspension



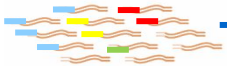
Droplet microfluidics 1000s of cells

Throughput

Single cells or nuclei



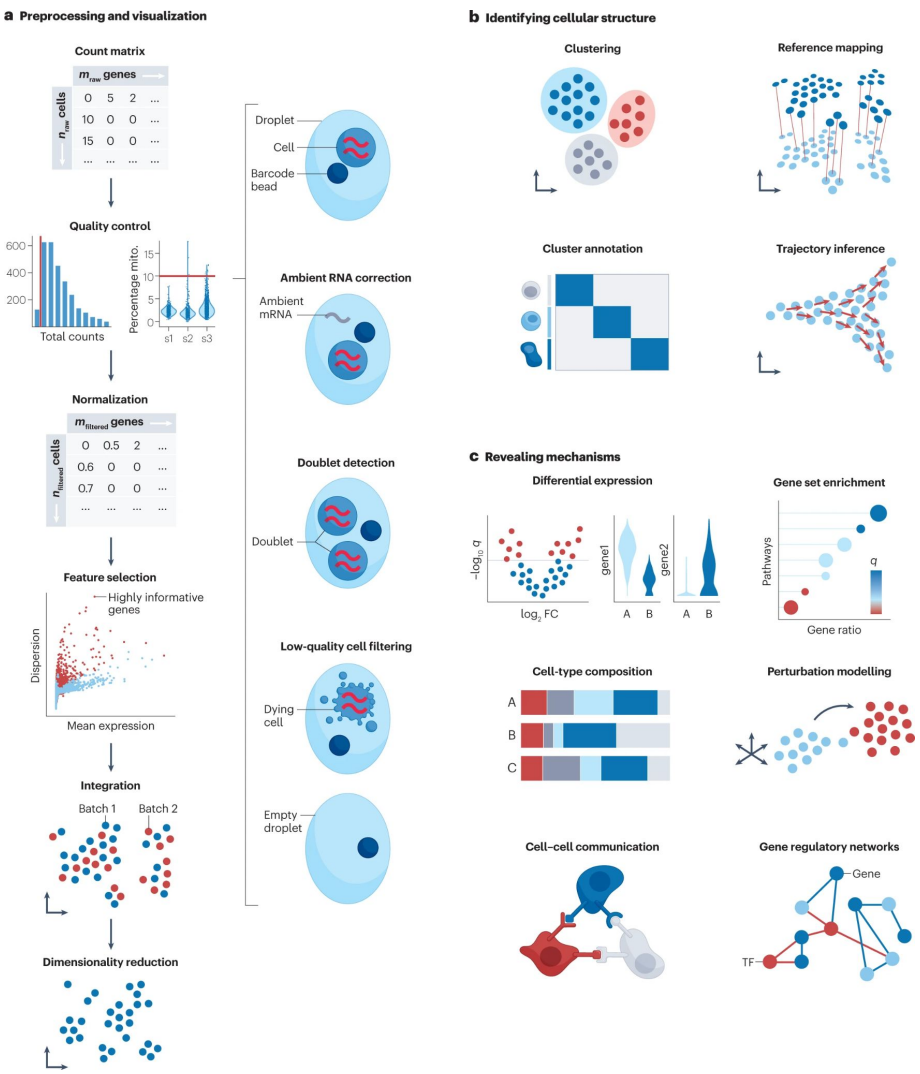
cDNA



Pooled cDNA libraries

Next-Generation Sequencing

Overview of the computational workflow

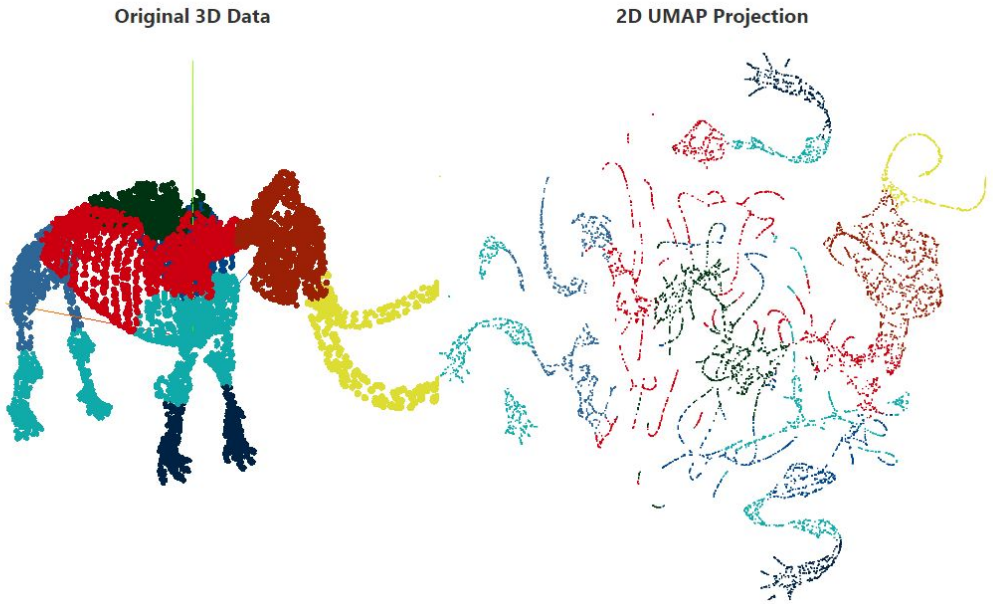


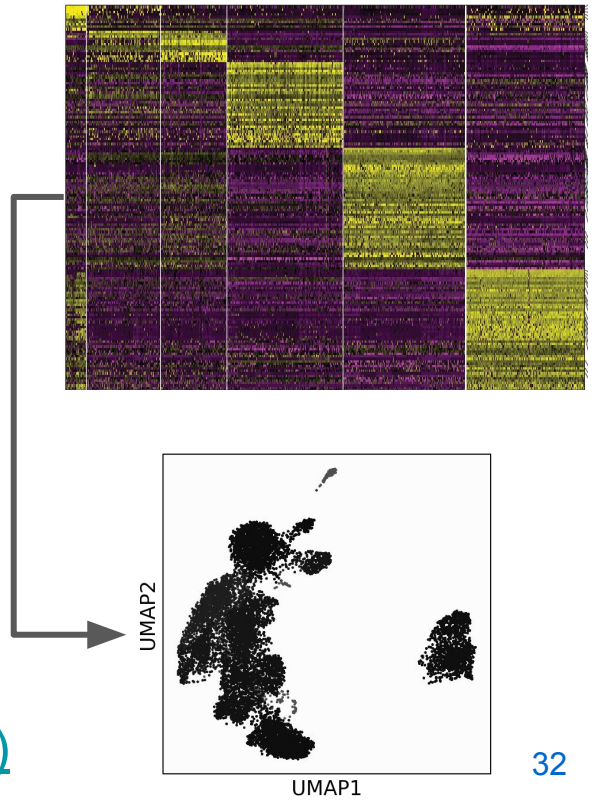
Heumos et al., Nature Reviews Genetics, 2023

How to represent voxels with pixels?

The elephant bull *Tusker (1992-2023)* at Zolli Basel plays with a tree trunk on a post (2022)

Uniform Manifold Approximation and Projection (UMAP) for dimension reduction

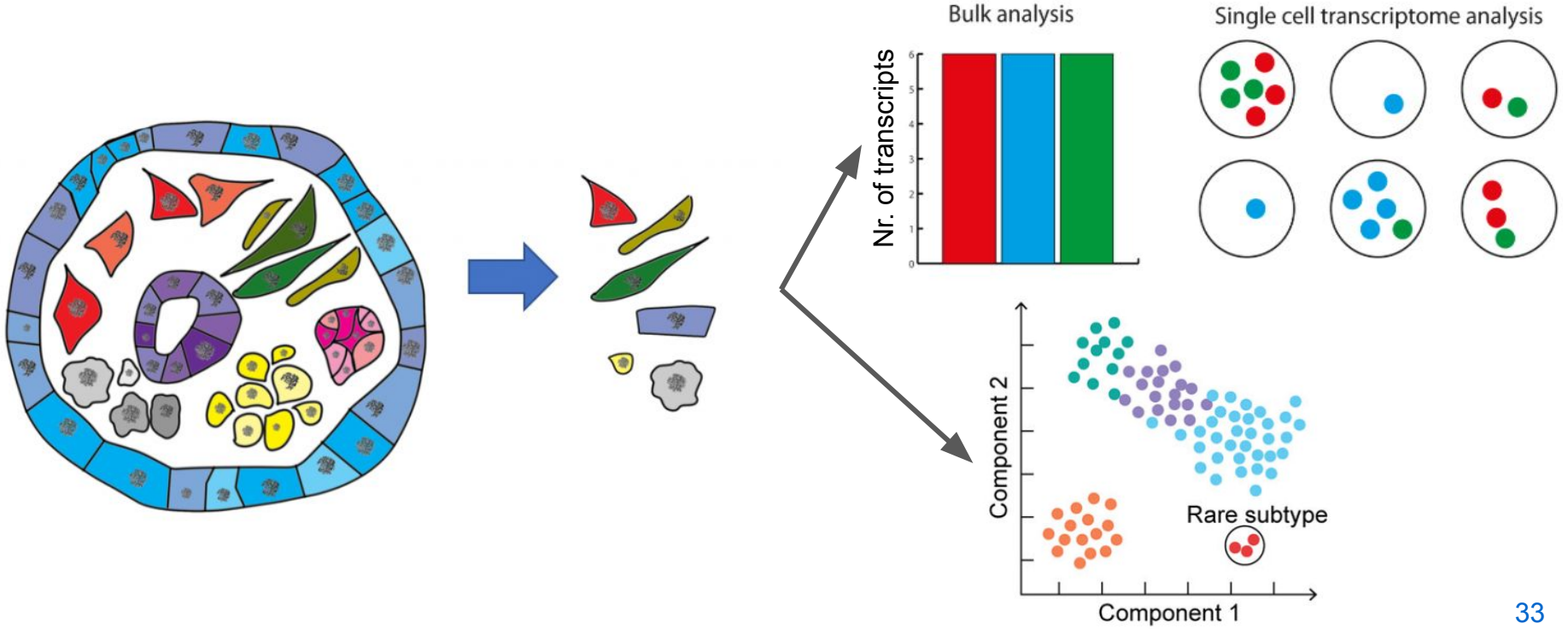




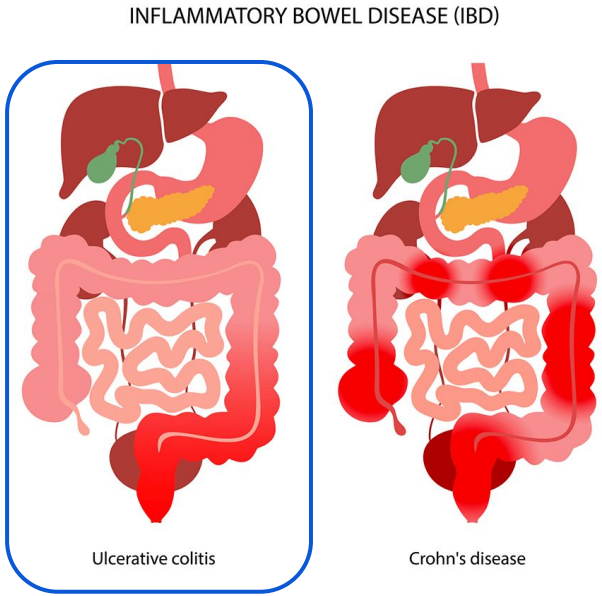
Understanding UMAP by A. Coenen and A. Pearce

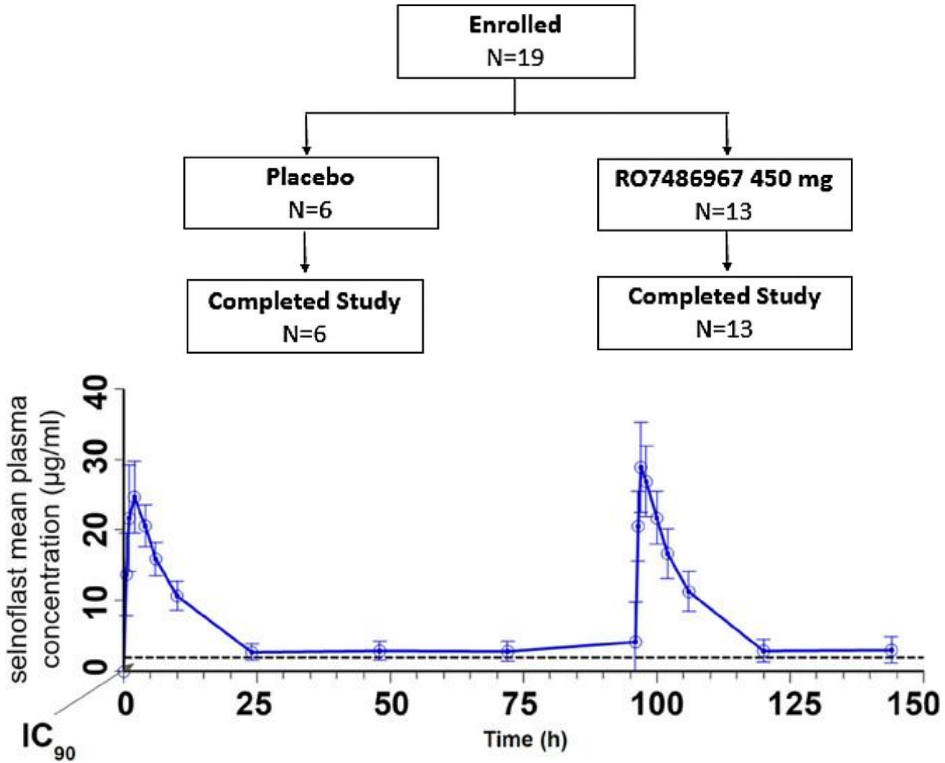
UMAP by Leland McInnes on SciPy 2018 (YouTube)

Single-cell biology benefits both disease understanding and drug discovery



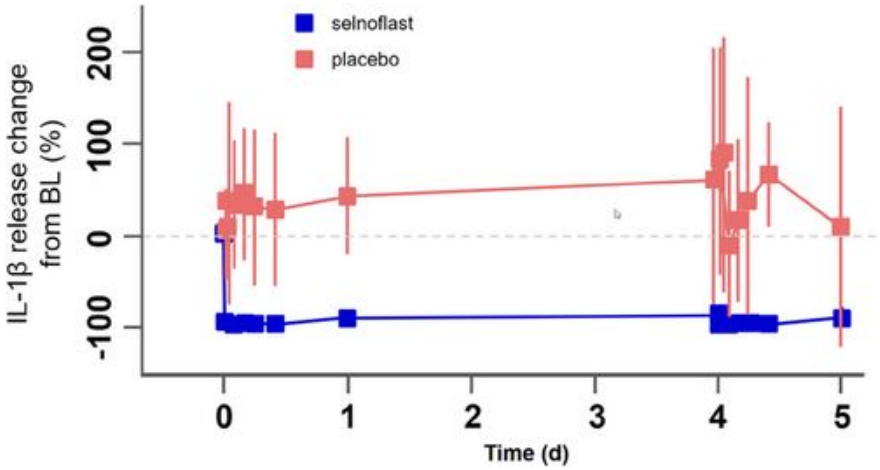
Phase 1b study of Selnoflast in UC

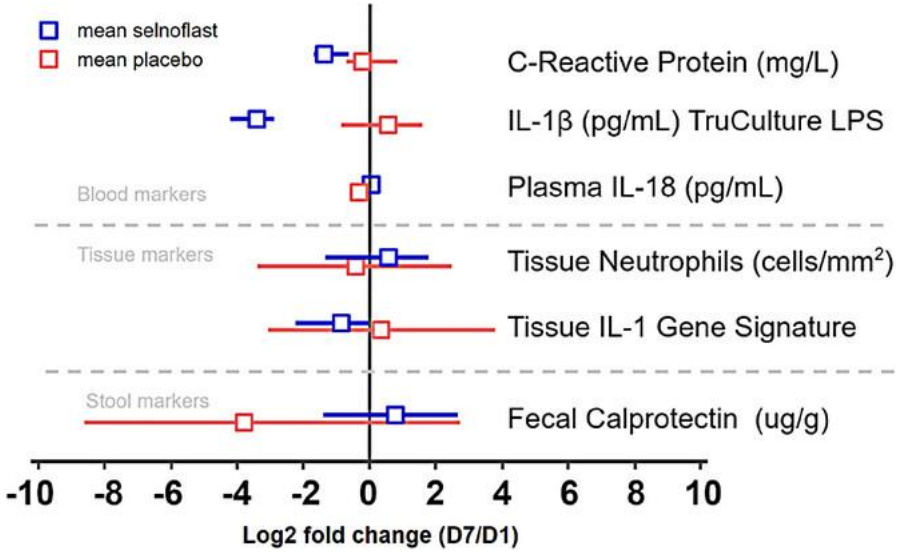




Klughammer, B. et al. A randomized, double-blind phase 1b study evaluating the safety, tolerability, pharmacokinetics and pharmacodynamics of the NLRP3 inhibitor selnoflast in patients with moderate to severe active ulcerative colitis. Clinical and Translational Medicine 13, e1471 (2023). IC90 was calculated from *in vitro* studies (2.0 ug/mL or 1.94 ug/g).

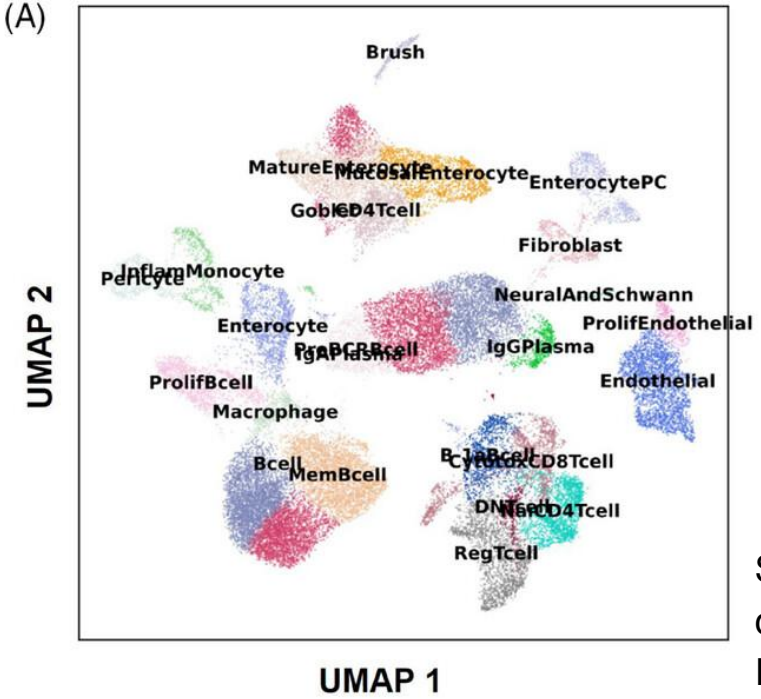
Biological assay and omics readout as biomarkers

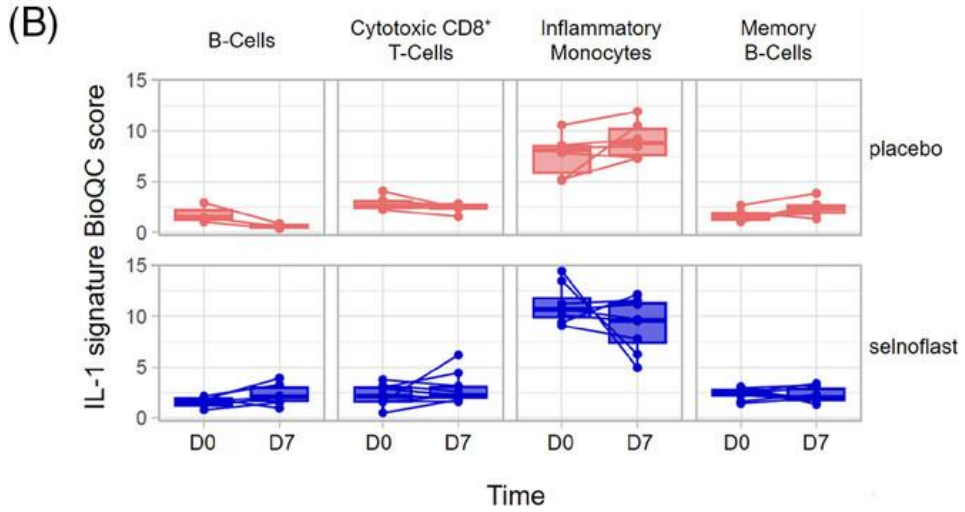




Selnoflast is a specific antagonist of NLRP3, a protein component of the *inflammasome*. Activation of inflammasome induces interleukin 1 beta (IL-1b), which in turn induces expression of downstream genes (IL-1b gene signatures).

Single-cell RNA-seq revealed that selnoflast failed to induce the changes that we had hoped for



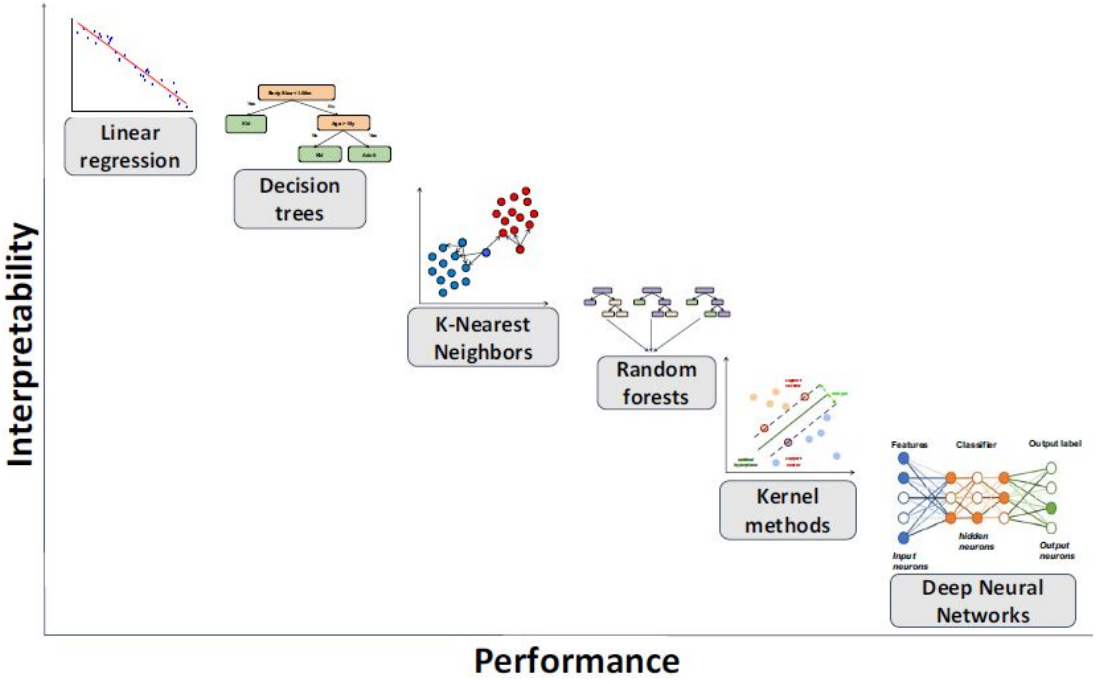


Selnoflast was safe and well-tolerated. Selnoflast 450 mg once a day achieved plasma and tissue exposure predicted to maintain IL-1β IC_{90} over the dosing interval. However, PD biomarker results showed no robust differences between treatment arms, suggesting no major therapeutic effects are to be expected in UC.

Beware of curse of dimensionality when studying high-dimensional biomarker data

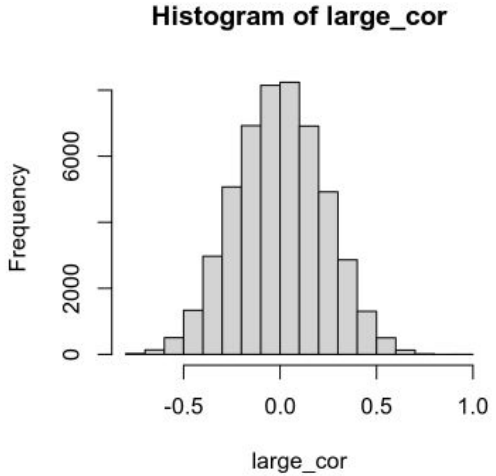
Non-causal statistical models are useful for hypothesis generation and exploratory analysis.

Particular caution is required for *high-dimensional data*, for extrapolation, and for designing interventions.

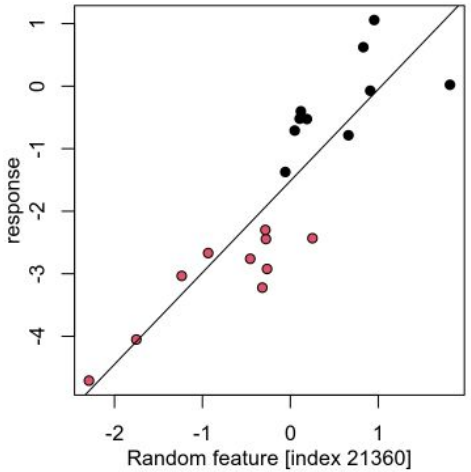


Simulating the curse of dimensionality

```r
set.seed(1887)
patient_group <- gl(2,10)
response <- c(rnorm(10, 0), rnorm(10, -3))
random_features_large <- matrix(rnorm(20*50000), nrow=20)
large_cor <- cor(response, random_features_large, method="spearman")
hist(large_cor)
```




```r
largest_cor_ind <- which.max(large_cor)
{
  compactPar()
  plot(random_features_large[, largest_cor_ind],
       response,
       bg=patient_group,pch=21,
       xlab=sprintf("Random feature [index %d]", largest_cor_ind))
  abline(lm(response ~ random_features_large[, largest_cor_ind]))
}
```



Conclusions

- Biomarkers (1) guide compound optimization and differentiation in preclinical studies, (2) support human dose prediction in translational PK/PD studies, and (3) allow patient stratification in clinical trials;
- Mathematical and computational biology is indispensable for biomarker identification;
- Beware of curse of dimensionality when using high-dimensional data for biomarker identification. It is probably beneficial to integrate mechanistic, statistical, and causal thinking and modeling.

That was it, MCBDD 2025. THANK YOU!

UNI BASEL

9. Kommentare zur Vorlesung mit Übungen

9.1) 47. Ich finde an der Vorlesung besonders gut:

- - interesting combination of different scientific fields
  - engaged (and entertaining!) lecturer
  - biology background knowledge is explained well
  - course website makes accessing information easy
  - regular emails giving updates and details about the course/exercises
- Always very interesting lecture, independent on the topic of the days lecture.
- Die Vorlesungen ist sehr Praxis orientiert
- Good lecturer
- I enjoy the very concrete connections to the real world and industry use. I also really like how the lecturer mixes in math, biology, chemistry and informatics into every topic in a very natural way that makes sense
- I especially liked when had David shared new AI/ML tools and how he or his collegues use them in practice
- It is really practical orientated, it's nothing I would think in my free time about, but very good explained
- Overall, I really enjoy this class. David is one of the best teachers I've had this semester - he's highly knowledgeable and genuinely passionate about the subject. The course content is also excellent. It's structured in a logical way that helps connect and reinforce everything we learn. While some parts are a bit advanced for my current level, I'm still able to follow along and grasp the general idea, even if I don't understand every detail.
- Really engaging, connects all the time with real life scenarios and examples, both historical and current. Gives a great stepping stone to dive deeper into topics one is interested in. Real projects that were discussed were extremely interesting and excellent to learn many things from.
- really interesting, and make the compliacted knowledge easy to understand
  used clear and engaging examples to simplify complex topics

- The biology examples are always relevant and make me like biology even more because they're so interesting.
- Vorlesungen sind interessant zuzuhören und sehr praxis nah.

9.2) 48. Ich finde, an der Vorlesung könnte verbessert werden (Erläuterungen/Verbesserungsvorschläge):

- I think it would be great to encourage students who sit in the back to sit closer to the front so we aren't so spread out.
- maybe improving the slides, but that may take time from preparing content, and i would not prefer less quality content over better slides
- Nothing major comes to mind. However, it would be helpful if some of the technical terms or acronyms on the slides were explained. That way, students who aren't familiar with them can follow along more easily without needing to look things up during the lecture.

9.3) 49. Ich finde an den Übungen besonders gut:

- - clear expectations
  - personally, I enjoyed the coding exercises
- API
- group works, small exercises on a focussed topic
- Homework assignment covered both practical programming exercises and reading of current papers
- I like that the exercises were very varied and also practical to what I assume I will need later working in the industry
- I really appreciated that we got to do some programming this semester. All the exercises were interesting and engaging. They provided a great way to revisit and reinforce the concepts we learned in class from a more practical perspective.
- I really like the emphasis on learning rather than on grades. It is inspiring.
- Praxisnähe
- Sind gut auf den Lernstoff abgestimmt

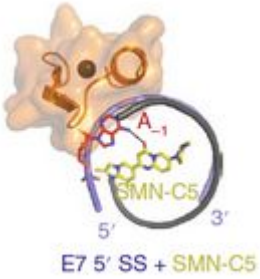
9.4) 50. Ich finde, an den Übungen könnte verbessert werden (Erläuterungen/Verbesserungsvorschläge):

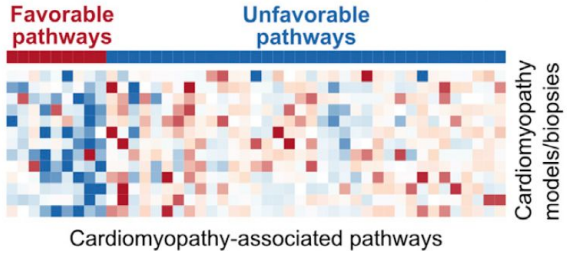
- N/A
- Nothing to add.
- There can be a mini quiz created after every paper/topic/manuscript given to read. It could be during the class and also discuss the topics.

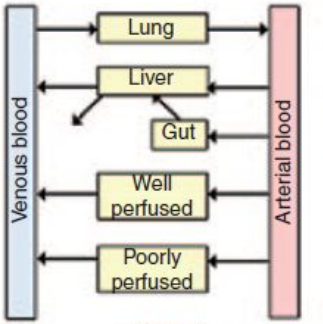
Outline of lecture 12

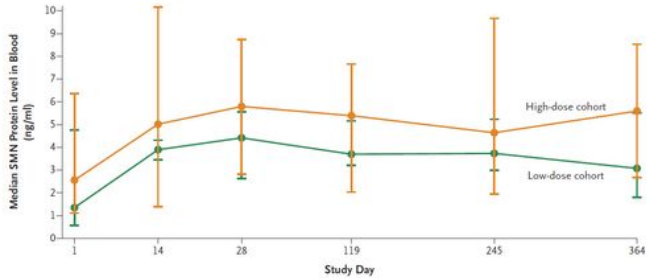
- An example of integrating statistical and mechanistic modelling: Griffiths *et al.*
- Mechanistic modelling of biological systems: from Boolean network to Agent-based modelling
- Causal inference
- Where can we go from here

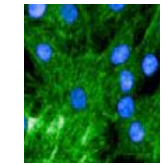
Drug discovery relies on *in vitro*, *in vivo*, and computational models across scales

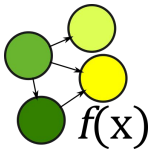


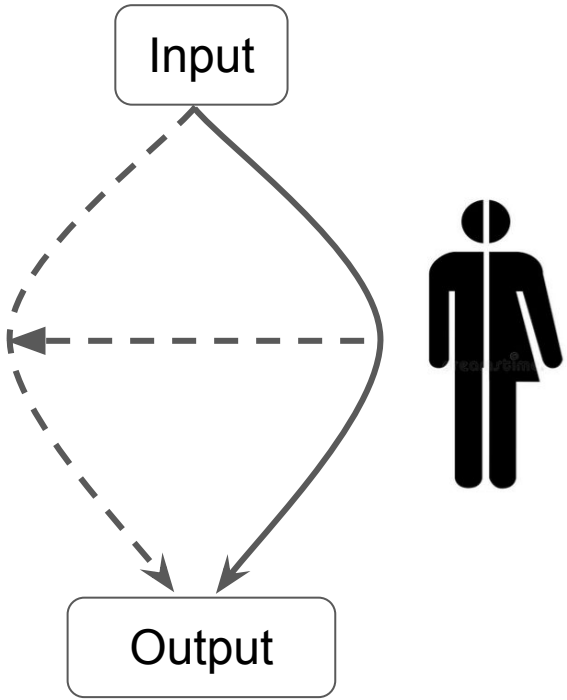












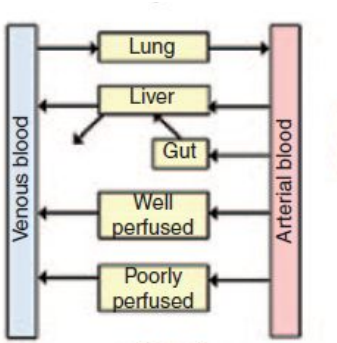
Examples of molecular, omics and cellular, organ and system, and population modelling

Mechanistic and computational models *explain*

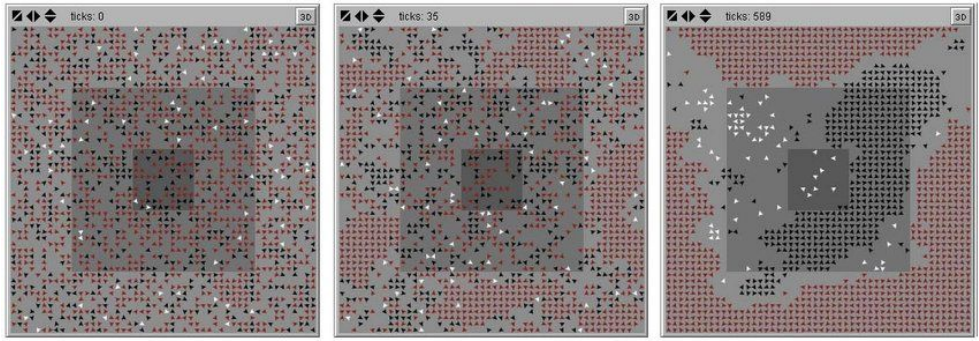
Compartment models

$$\frac{d[LR]}{dt} = k_1[L][R] - k_2[LR]$$

Kinetics of ligand-target interaction

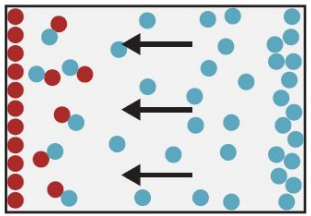


Particle models



Transport models

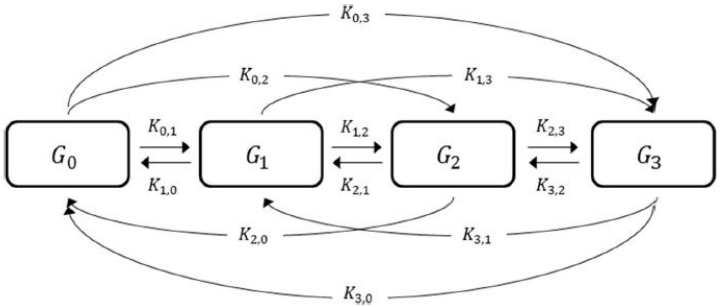
Reaction-Diffusion System



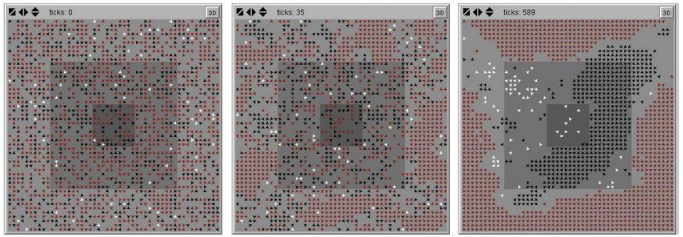
Diffusion

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

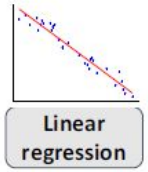
Finite state models



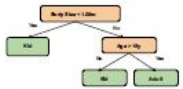
Integration of knowledge, assumptions and data across scales is key for drug discovery



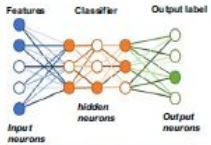
$$\frac{d[LR]}{dt} = k_1[L][R] - k_2[LR]$$



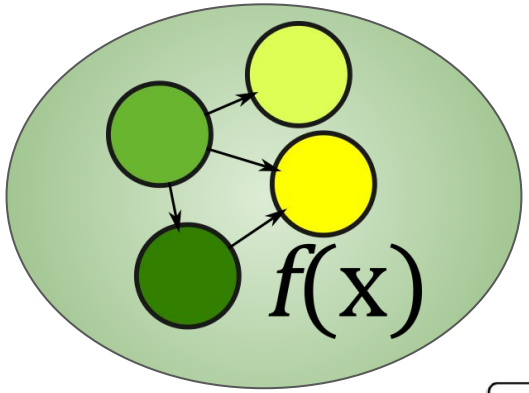
Linear regression

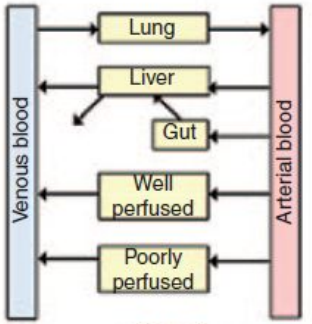


Decision trees

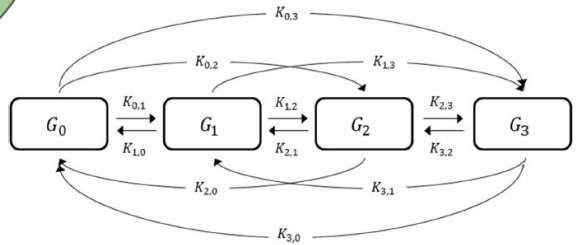


Deep Neural Networks



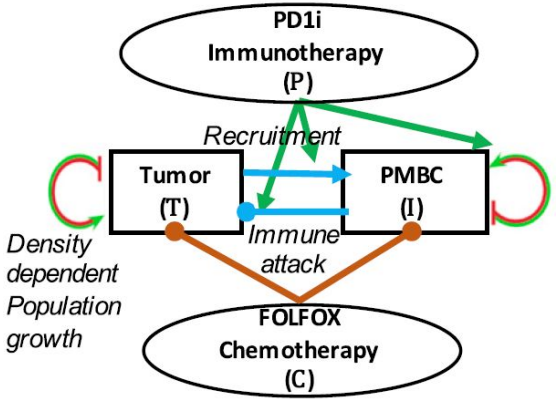


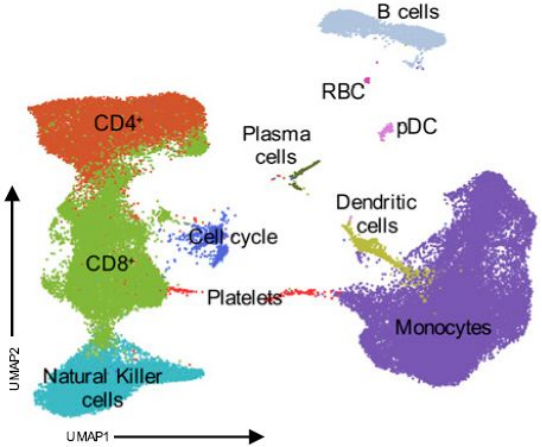
$$\frac{\partial u}{\partial t} = \underbrace{D\frac{\partial^2 u}{\partial x^2}}_{\text{Diffusion}} + \underbrace{ku}_{\text{Binding}}$$



An example of integrated modelling

- Griffiths et al. (PNAS 2020) profiled peripheral immune cell abundance in time series following treatment of Gastrointestinal (GI) tumours with immunotherapy in a small clinical trial.
- The authors used **compartment models** to characterize cell-cell interactions and analysed **single-cell omics data** to reveal immune cell abundance, pathway activity patterns, and differentiation status.



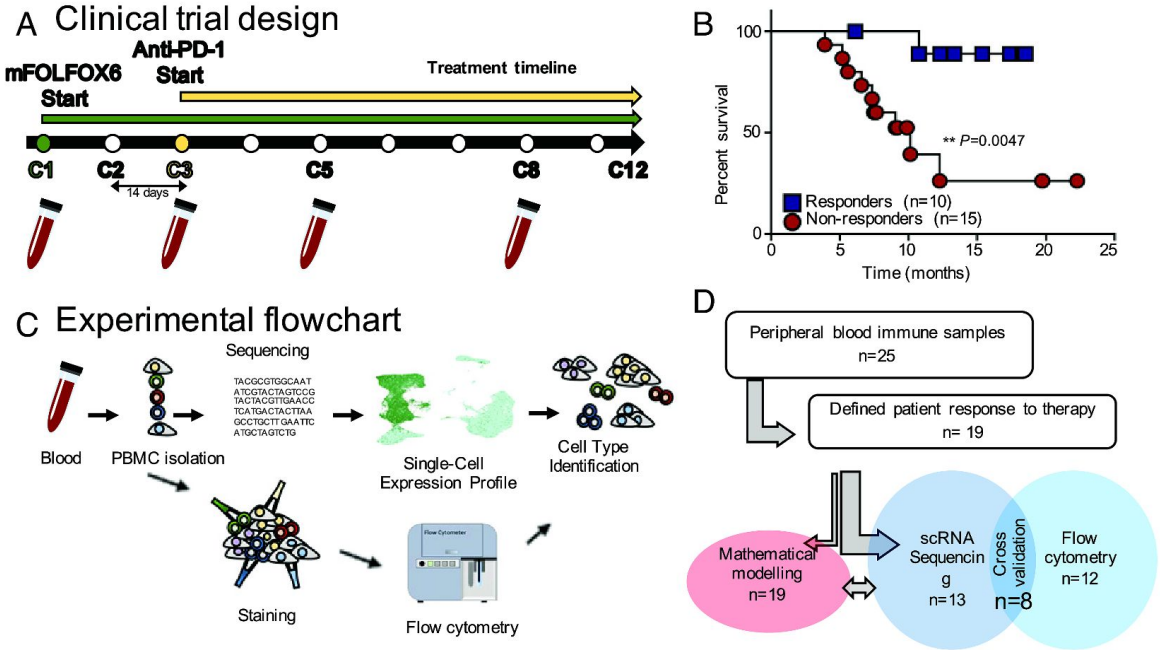


How this study enriches our knowledge

- **Facts**
  - Clinical response to immune checkpoint inhibitors varies substantially.
  - Possible contributing factors correlate only weakly with patient response, including (1) tumor cell mutational load and antigen production, (2) immune-cell infiltration and signalling status, (3) Cross-talk between tumour and immune cells.
  - It is challenging to obtain tumour tissue samples.
- **Questions**: Can circulating immune cells serve as a surrogate measurement of a tumour's interaction with the host immune cells and reflect response to therapy early in the course of treatment?
- **Conclusions**: It is possible to predict patient response with the evolution of peripheral immune cell abundance and signalling over time, as well as how immune cell interact with the tumor.

Design of the clinical trial

- mFOLFOX6 (modified FOLFOX6): a chemotherapy regiment.
- Patient response was assessed by RECIST (Response Evaluation Criteria in Solid Tumors) 1.1 guidelines, using computer tomography (CT).

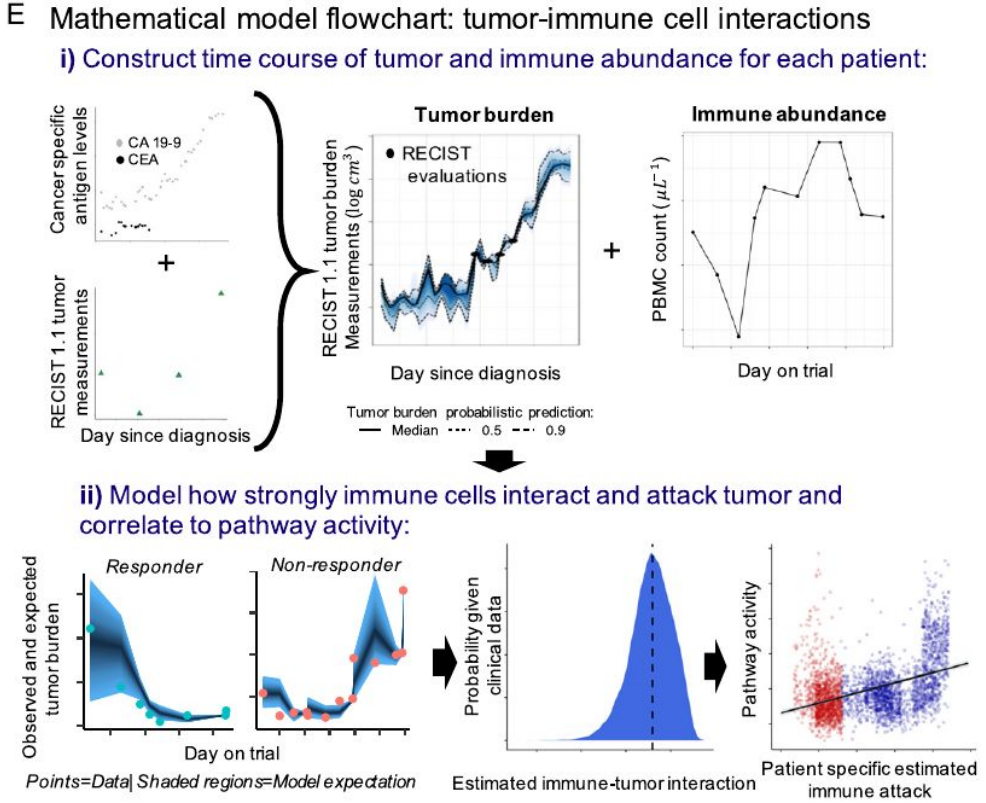


Mathematical modelling of tumour-immune cell interactions

Model inputs (all in time series):

- Tumour burden, inferred by combining antigen values and RECIST evaluation with a *Gaussian process* latent variabl model.
- Abundance of PBMCs

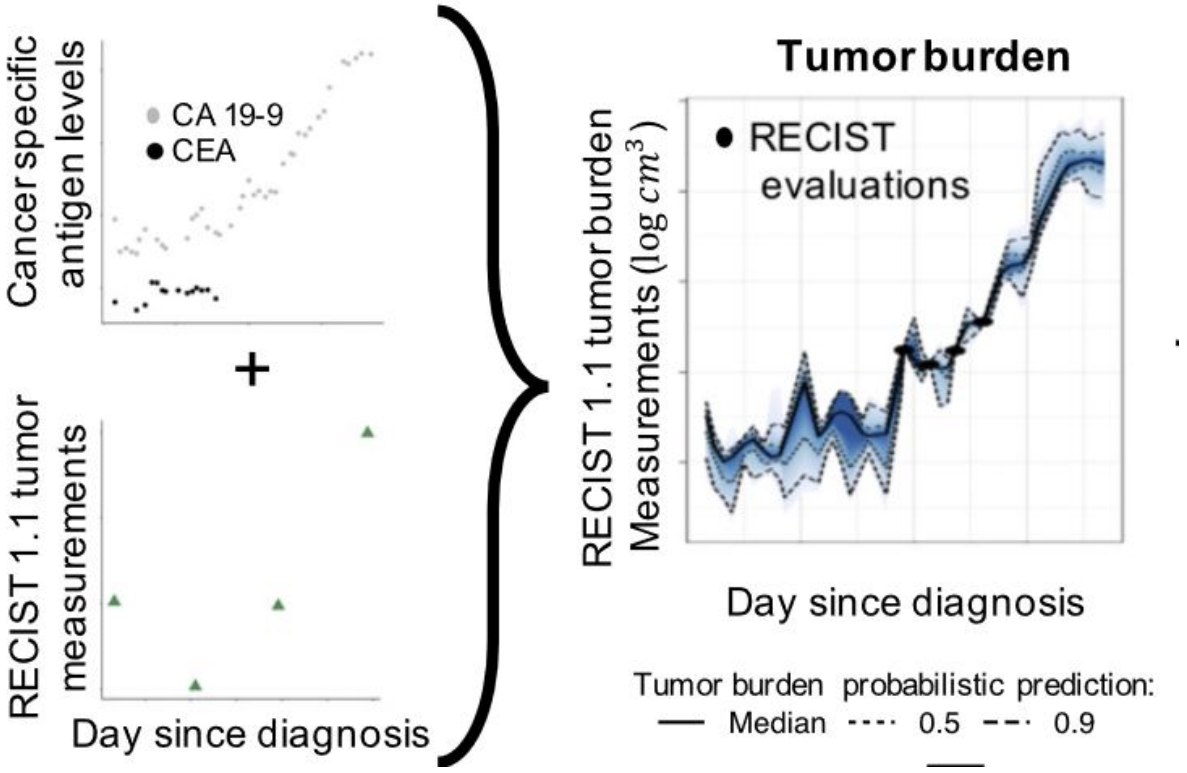
Model output: estimated ability of immune cells to kill tumour cells



Modelling time-series data with Gaussian Process

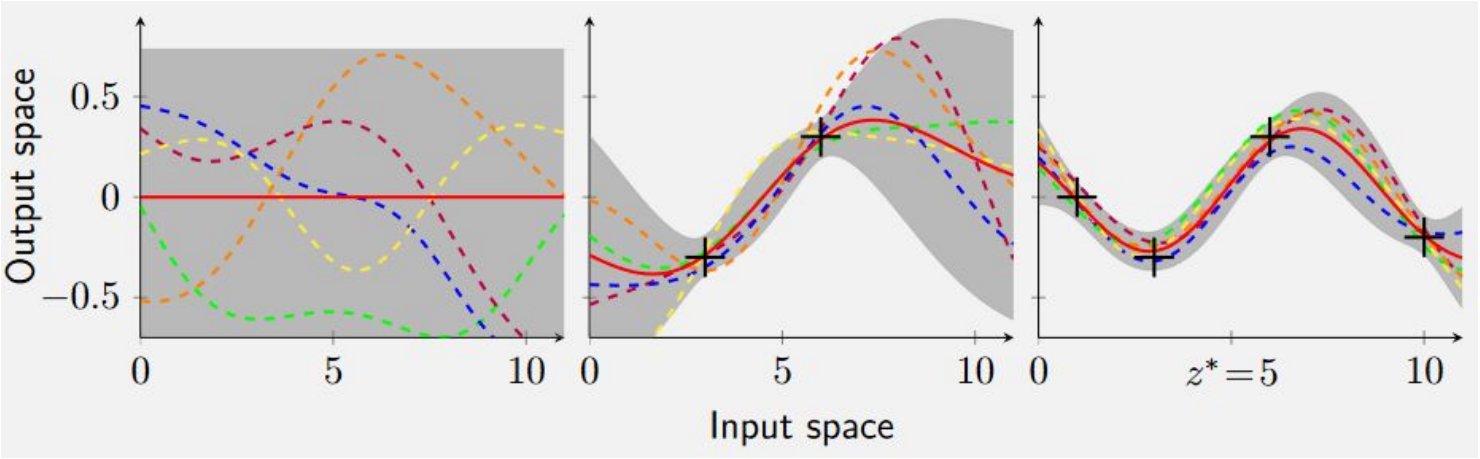
Challenge: tumor growth assessment (RECIST) involves much labour and costly.

Solution: assuming that the data is generated by a Gaussian process, and that cancer-specific antigens is correlated with the tumor growth, we can ‘impute’ the missing data with other biomarker data with *Gaussian Process*.



Intuitions about Gaussian Process

1. Observe y values as if they are generated from a multivariate Gaussian distribution with *indefinite* dimensions, and time-dependent correlations.
2. We can infer the autocorrelation (i.e. kernel function) by data: the more data, the better we can infer.



The Lotka-Volterra model of predator-prey relationships

- The Lotka-Volterra equations modelling predator-prey relationships.

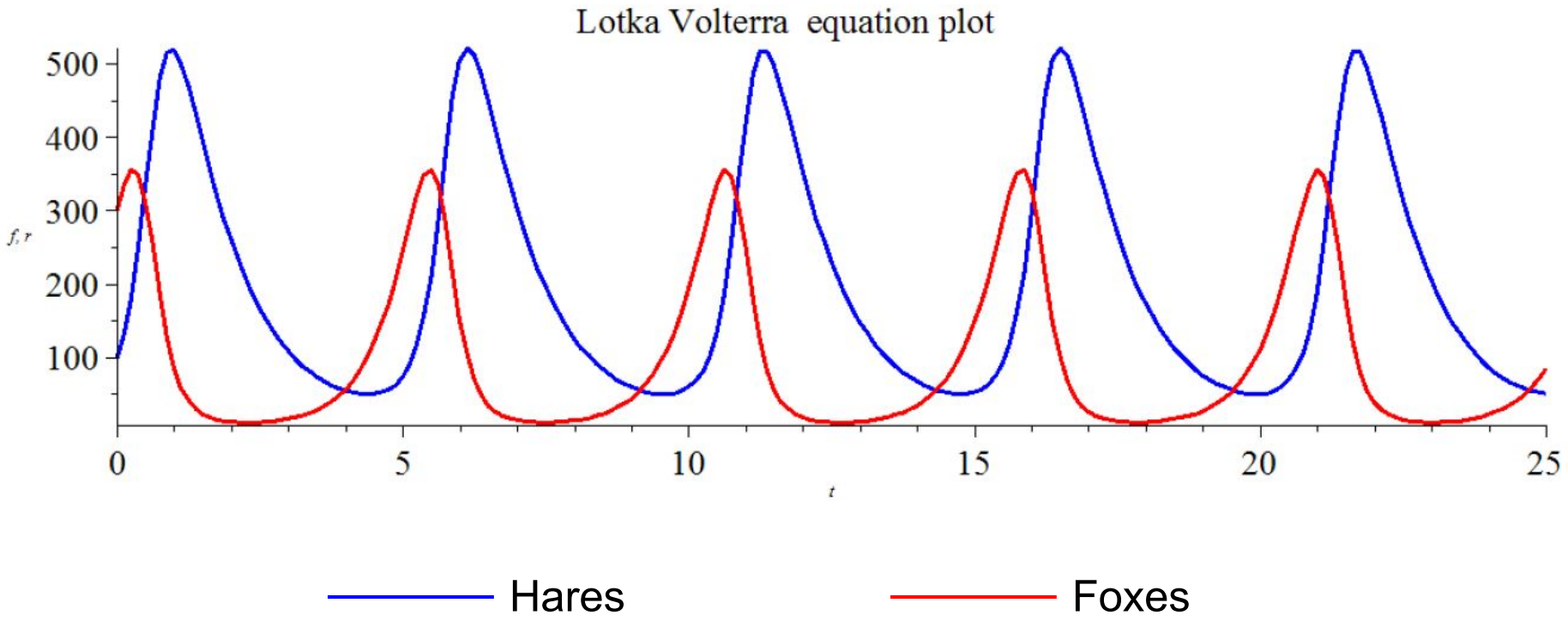
$$\frac{dx}{dt} = \alpha x - \beta xy, \tag{1}$$

$$\frac{dy}{dt} = -\gamma y + \delta xy, \tag{2}$$

where

- x is the number of prey (*e.g.* rabbits),
- y is the number of predator (*e.g.* foxes),
- $\frac{dx}{dt}$ and $\frac{dy}{dt}$ represent growth rates of the two populations,
- t represents time,
- α, β, γ, and δ are real parameters specifying the interaction of the two species.

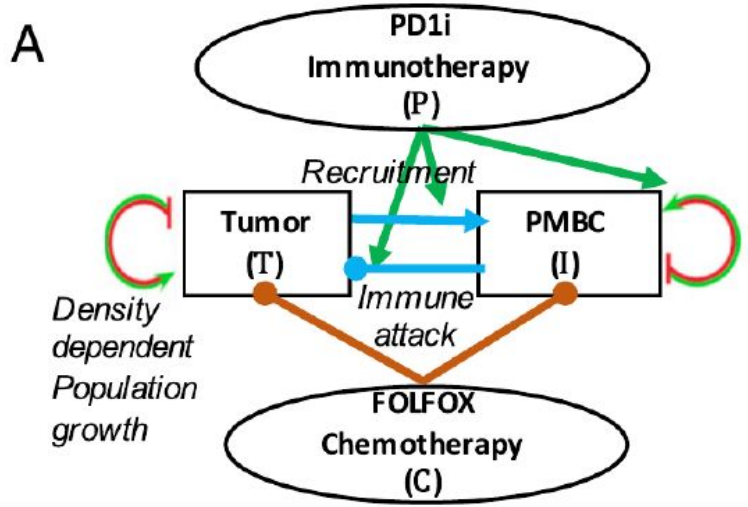
The Lotka-Volterra equations, visualized



Modelling of interactions between tumour and immune cells

Modelling assumptions:

- Tumor cells are attacked by immune cells
- Tumor cells recruit immune cells
- Chemotherapy kills both tumour and immune cells
- Anti-PD1 inhibitor immunotherapy impacts immune proliferation, recruitment, and cytotoxic tumor activity.



$$RGR_T = \frac{1}{T}\frac{dT}{dt} = r_T(1-\gamma_T T) - (\alpha+\beta_\varphi P)I - \sum_i \overrightarrow{\mu_T}[i]\,C_i$$

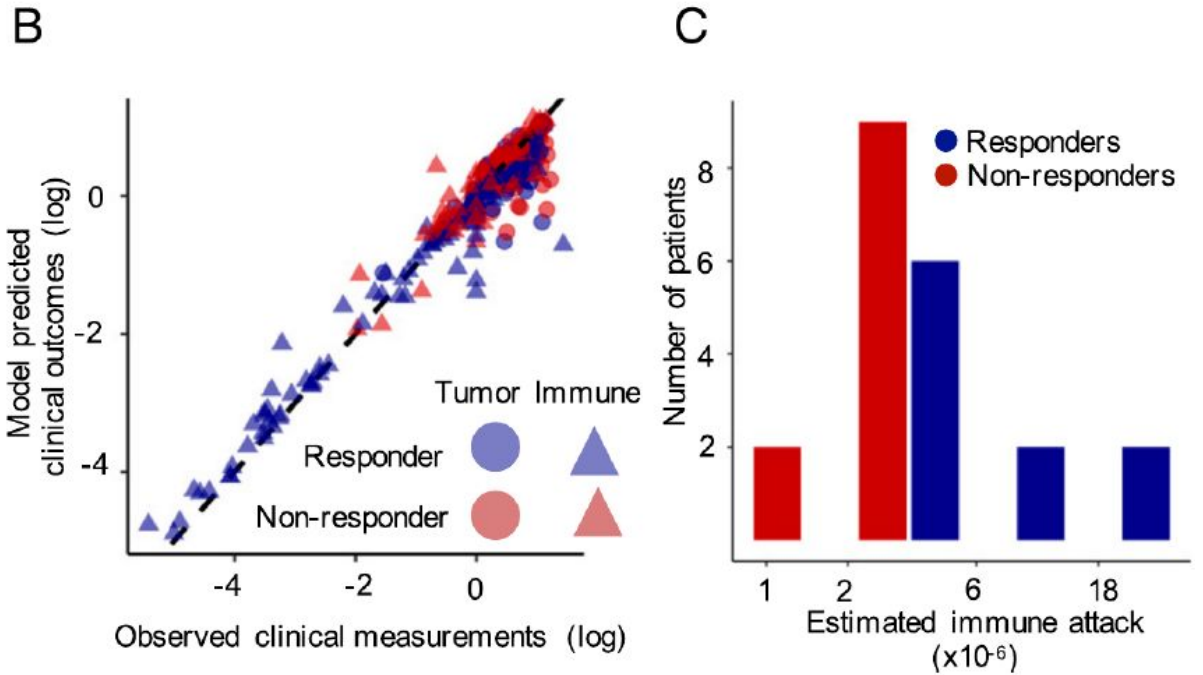
Tumor growth; Immune attack; Chemotherapy

$$RGR_I = \frac{1}{T}\frac{dI}{dt} = (r_I+\beta_r P)(1-\gamma_I I) + (\lambda+\beta_\lambda P)T - \sum_i \overrightarrow{\mu_I}[i]C_i$$

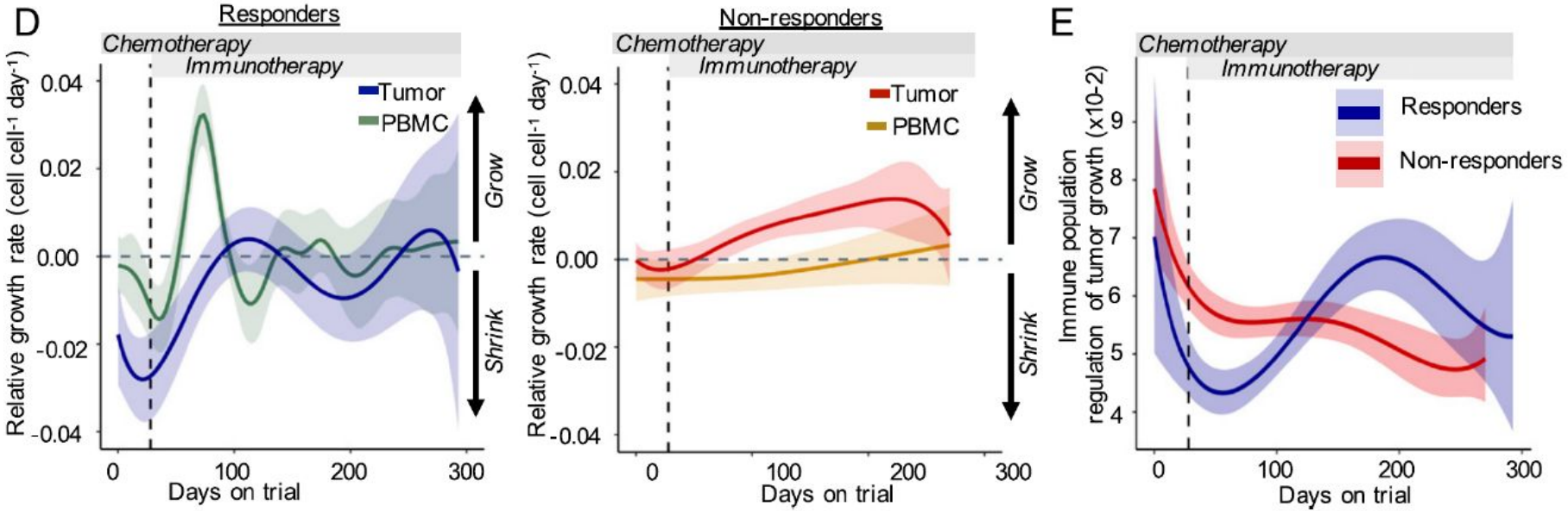
Immune growth; Recruitment; Chemotherapy

Model prediction and performance

- The strength of immune-tumor interaction is estimated by statistically fitting the growth rate of immune cells and tumor size to model predictions.
- Changes in tumor burden and immune cell abundance are described by data fitting, using a Bayesian hierarchical model.



Profiles of relative growth rates differ between responders and non-responders

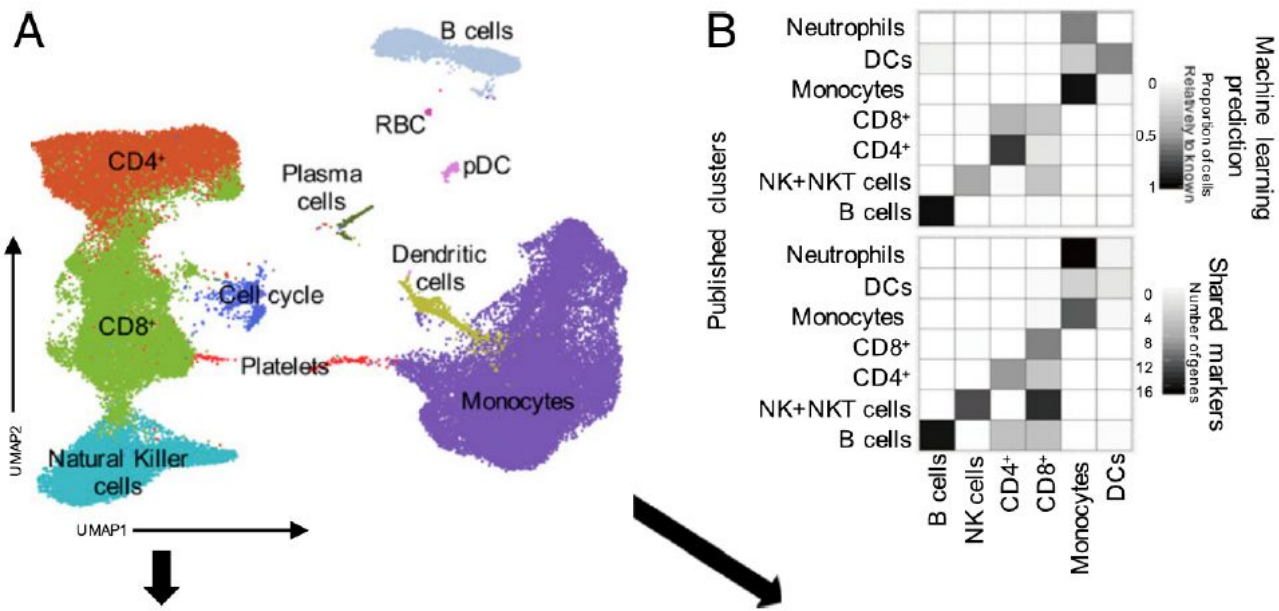


- Neither tumor nor PBMC responds to chemotherapy in non-responders.
- Responders have lower PBMC abundance in general at baseline.

Immune cell population identified by scRNAseq

PBMCs were analysed at three time points:

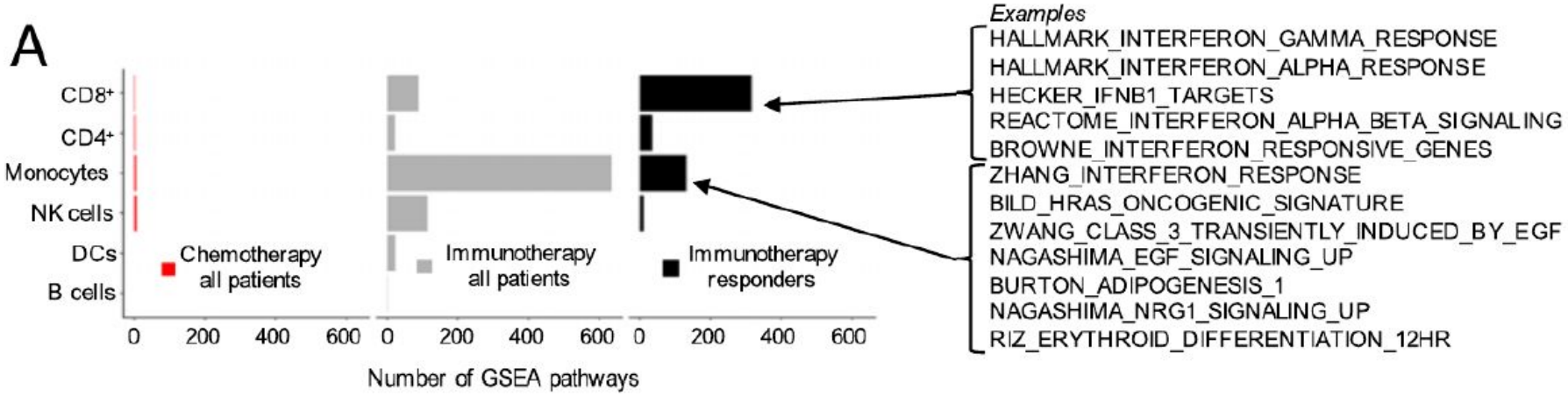
1. Cycle 1 (C1): baseline before treatment;
2. Cycle 3 (C3): chemotherapy alone;
3. Cycle 5 (C5): chemotherapy + anti-PD-1.



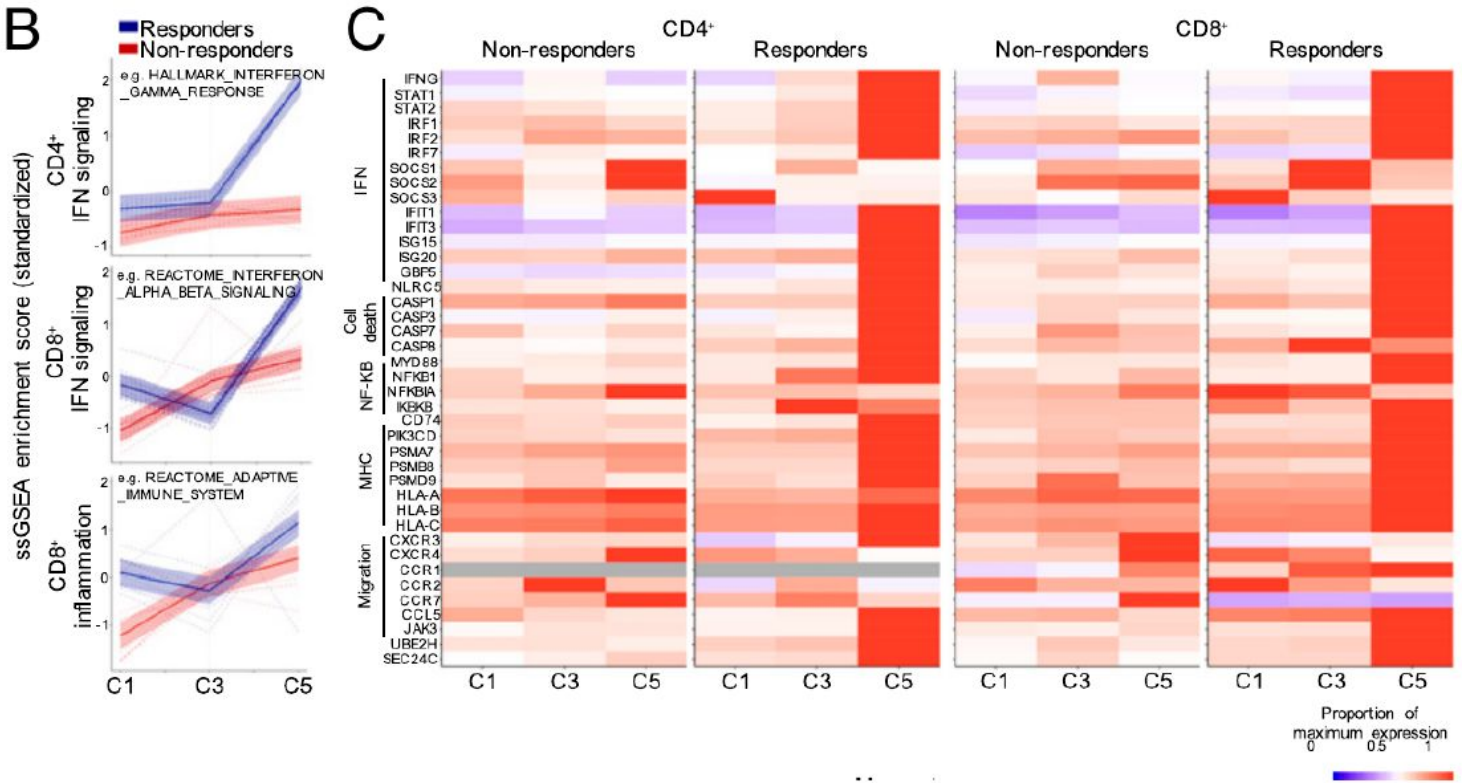
A total number of 70781 cells from 13 patients (7 responders and 6 non-responders) were profiled.

Pathway analysis

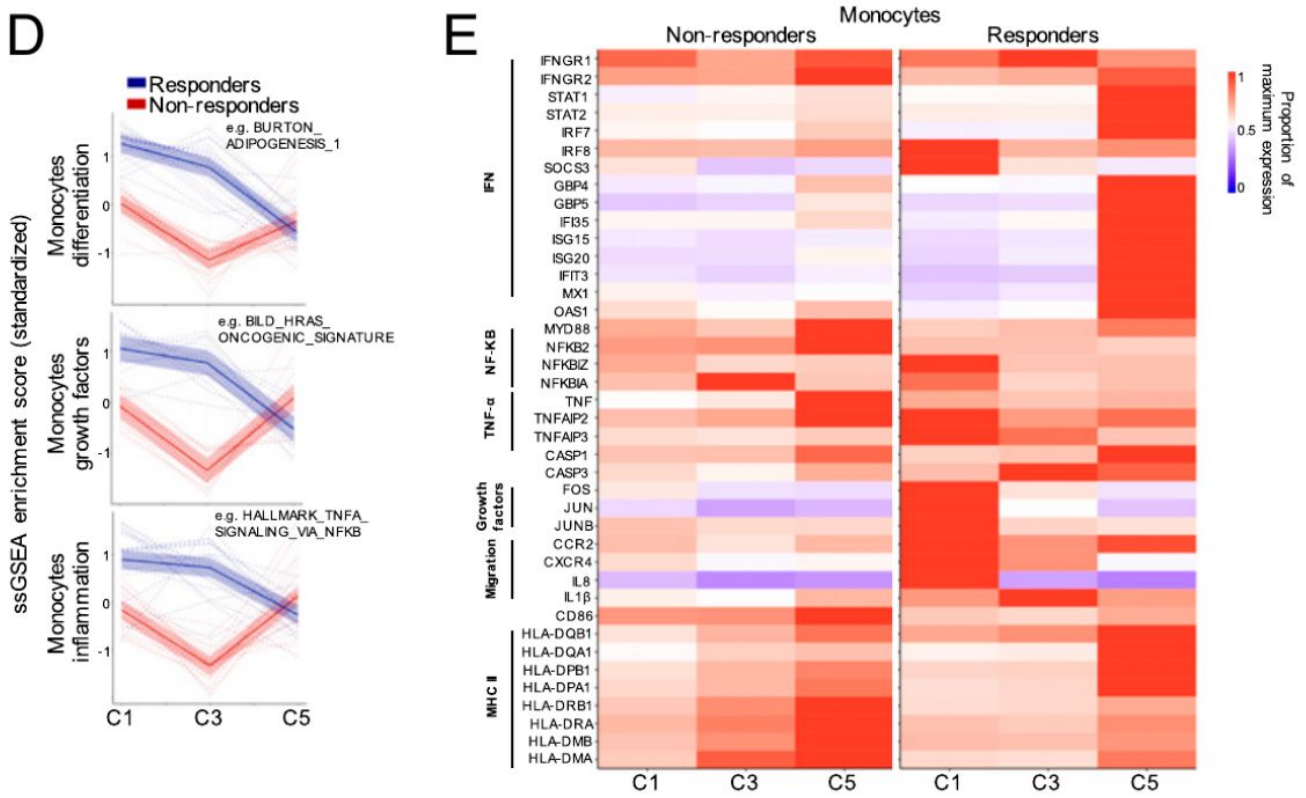
Single-sample gene-set enrichment analysis was performed to identify pathway differences before therapy, during chemotherapy, and during the early combo of chemotherapy and immunotherapy using a *random effects linear model*.



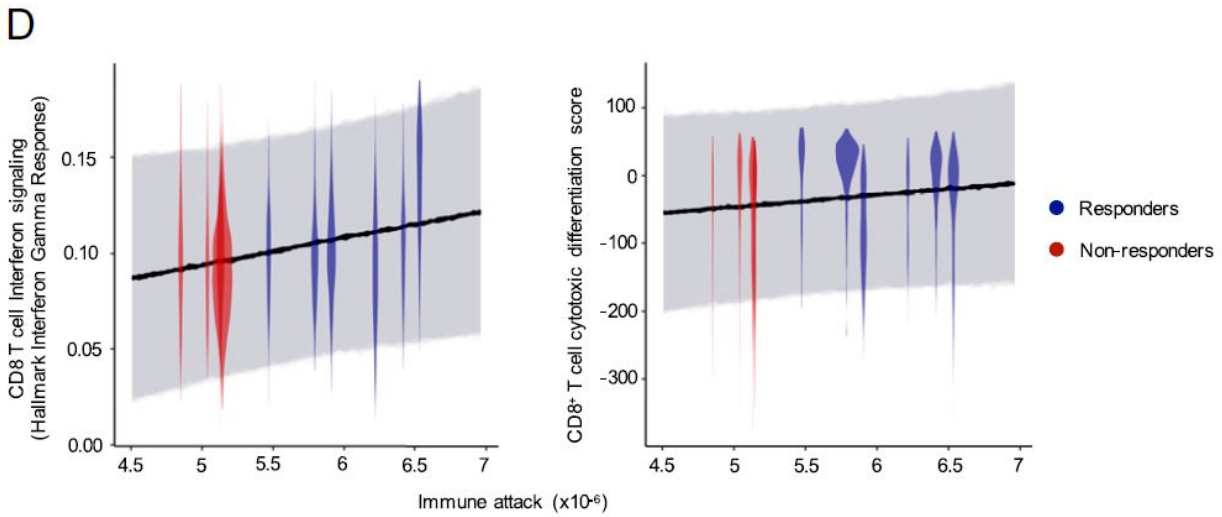
Responders show changes in T-cell signalling during treatment

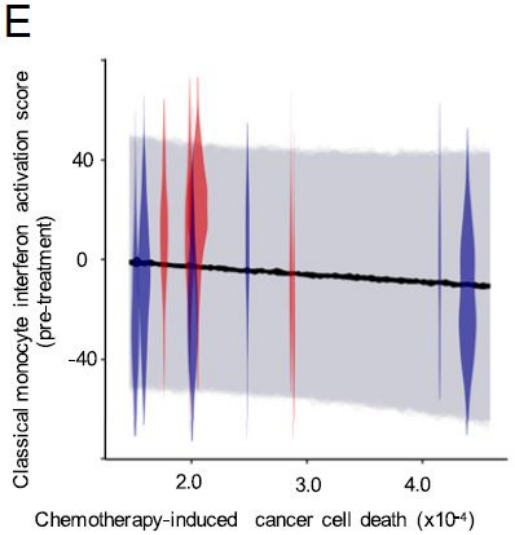


Responders show changes in monocyte signalling during treatment



Associations between omics data and inferred model parameter





Impact and limitation of the study

- *On the biological side*, the results suggest that peripheral blood phenotypes can be used as biomarker of patient responsiveness to therapy. The idea seems to be confirmed the findings by Wu *et al.*, *Peripheral T cell expansion predicts tumour infiltration and clinical response*, Nature 2020.
- *On the modelling side*, the study integrates machine learning, omics data analysis, mathematical modelling techniques to link macroscopic findings, for instance antigen and RECIST scores, with cellular findings, including scRNAseq and flow cytometry. This study exemplifies what we call *multiscale modelling of drug mechanism and safety*.
- We do not know why some patients respond to anti-PD-1 or anti-PDL1 therapies better than other patients based on findings reported in both papers. Nevertheless, both studies suggest that immune cells in peripheral blood may be used as biomarkers in certain settings.

Conclusions

- Understanding how drugs work and how to develop better drugs requires *causal reasoning*, for which there are no scientific consensus yet.
- Integrated mechanistic, computational, and statistical modelling across scales is a viable approach towards causal reasoning.
- Mathematical and computational biology is indispensable to address this grand challenge.

Ways to learn more about mathematical & computational biology in drug discovery

- **People** around you, both with the same and different backgrounds;
- **Reading**, including Journal *Nature Reviews Drug Discovery*, blogs *In the Pipeline*, *CureFFI*, and newsletter *This Week in Mathematical Oncology*;
- **Online courses**: *Statistical Rethinking* by Richard McElreath, with freely available lecture videos on YouTube, *and Information Theory, Inference, and Learning Algorithms* by David MacKay, with freely available lecture videos.

FDA's opinions on Artificial Intelligence and Machine Learning in Drug Development

1. Human-led governance, accountability, and transparency
2. Quality, reliability, and representativeness of data
3. Model development, performance, monitoring, and validation

References

1. Cristescu, Razvan, Robin Mogg, Mark Ayers, Andrew Albright, Erin Murphy, Jennifer Yearley, Xinwei Sher, et al. 2018. “Pan-Tumor Genomic Biomarkers for PD-1 Checkpoint Blockade–Based Immunotherapy.” *Science* 362 (6411): eaar3593. https://doi.org/10.1126/science.aar3593.
2. Litchfield, Kevin, James L. Reading, Clare Puttick, Krupa Thakkar, Chris Abbosh, Robert Bentham, Thomas B. K. Watkins, et al. 2021. “Meta-Analysis of Tumor- and T Cell-Intrinsic Mechanisms of Sensitization to Checkpoint Inhibition.” Cell 184 (3): 596-614.e14. https://doi.org/10.1016/j.cell.2021.01.002.
3. Krishnamoorthy, Vijay, Danny J. N. Wong, Matt Wilson, Karthik Raghunathan, Tetsu Ohnuma, Duncan McLean, S. Ramani Moonesinghe, and Steve K. Harris. 2020. “Causal Inference in Perioperative Medicine Observational Research: Part 1, a Graphical Introduction.” British Journal of Anaesthesia 125 (3): 393–97. https://doi.org/10.1016/j.bja.2020.03.031.
4. Krishnamoorthy, Vijay, Duncan McLean, Tetsu Ohnuma, Steve K. Harris, Danny J. N. Wong, Matt Wilson, Ramani Moonesinghe, and Karthik Raghunathan. 2020. “Causal Inference in Perioperative Medicine Observational Research: Part 2, Advanced Methods.” British Journal of Anaesthesia 125 (3): 398–405. https://doi.org/10.1016/j.bja.2020.03.032.
5. Wright, Sewall. 1934. “The Method of Path Coefficients.” The Annals of Mathematical Statistics 5 (3): 161–215. https://doi.org/10.1214/aoms/1177732676.
6. Emdin, Connor A., Amit V. Khera, and Sekar Kathiresan. 2017. “Mendelian Randomization.” JAMA 318 (19): 1925–26. https://doi.org/10.1001/jama.2017.17219.
7. Lawlor, Debbie A., Roger M. Harbord, Jonathan A. C. Sterne, Nic Timpson, and George Davey Smith. 2008. “Mendelian Randomization: Using Genes as Instruments for Making Causal Inferences in Epidemiology.” Statistics in Medicine 27 (8): 1133–63. https://doi.org/10.1002/sim.3034.
8. Lv, Bo-Min, Yuan Quan, and Hong-Yu Zhang. 2021. “Causal Inference in Microbiome Medicine: Principles and Applications.” Trends in Microbiology, April. https://doi.org/10.1016/j.tim.2021.03.015.
9. Smith, George Davey, and Shah Ebrahim. 2008. Mendelian Randomization: Genetic Variants as Instruments for Strengthening Causal Inference in Observational Studies. Biosocial Surveys. National Academies Press (US). https://www.ncbi.nlm.nih.gov/books/NBK62433/.
10. Williamson, Jon. 2019. “Establishing Causal Claims in Medicine.” International Studies in the Philosophy of Science 32 (1): 33–61. https://doi.org/10.1080/02698595.2019.1630927.

References (continued)

11. Vasudevan, Rama K., Maxim Ziatdinov, Lukas Vlcek, and Sergei V. Kalinin. 2021. “Off-the-Shelf Deep Learning Is Not Enough, and Requires Parsimony, Bayesianity, and Causality.” *Npj Computational Materials* 7 (1): 1–6. https://doi.org/10.1038/s41524-020-00487-0.
12. Chetty, R. K., J. S. Ozer, A. Lanevschi, I. Schuppe-Koistinen, D. McHale, J. S. Pears, J. Vonderscher, F. D. Sistare, and F. Dieterle. 2010. “A Systematic Approach to Preclinical and Clinical Safety Biomarker Qualification Incorporating Bradford Hill’s Principles of Causality Association.” *Clinical Pharmacology & Therapeutics* 88 (2): 260–62. https://doi.org/10.1038/clpt.2010.77.
13. Havel, Jonathan J., Diego Chowell, and Timothy A. Chan. 2019. “The Evolving Landscape of Biomarkers for Checkpoint Inhibitor Immunotherapy.” Nature Reviews Cancer 19 (3): 133–50. https://doi.org/10.1038/s41568-019-0116-x.
14. Litchfield, Kevin, James L. Reading, Clare Puttick, Krupa Thakkar, Chris Abbosh, Robert Bentham, Thomas B. K. Watkins, et al. 2021. “Meta-Analysis of Tumor- and T Cell-Intrinsic Mechanisms of Sensitization to Checkpoint Inhibition.” Cell 184 (3): 596-614.e14. https://doi.org/10.1016/j.cell.2021.01.002.
15. Trusheim, Mark R., Ernst R. Berndt, and Frank L. Douglas. 2007. “Stratified Medicine: Strategic and Economic Implications of Combining Drugs and Clinical Biomarkers.” Nature Reviews Drug Discovery 6 (4): 287–93. https://doi.org/10.1038/nrd2251.
16. Trusheim, Mark R., Ernst R. Berndt, and Frank L. Douglas. 2007. “Stratified Medicine: Strategic and Economic Implications of Combining Drugs and Clinical Biomarkers.” *Nature Reviews Drug Discovery* 6 (4): 287–93. https://doi.org/10.1038/nrd2251.
17. Bender, Andreas, and Isidro Cortés-Ciriano. 2020. “Artificial Intelligence in Drug Discovery: What Is Realistic, What Are Illusions? Part 1: Ways to Make an Impact, and Why We Are Not There Yet.” *Drug Discovery Today*, December. https://doi.org/10.1016/j.drudis.2020.12.009.
18. Bender, Andreas, and Isidro Cortes-Ciriano. 2021. “Artificial Intelligence in Drug Discovery: What Is Realistic, What Are Illusions? Part 2: A Discussion of Chemical and Biological Data.” *Drug Discovery Today* 26 (4): 1040–52. https://doi.org/10.1016/j.drudis.2020.11.037.
19. Robers, M.B., R. Friedman-Ohana, K.V.M. Huber, L. Kilpatrick, J.D. Vasta, B.-T. Berger, C. Chaudhry, et al. 2020. “Quantifying Target Occupancy of Small Molecules Within Living Cells.” Annual Review of Biochemistry 89 (1): 557–81. https://doi.org/10.1146/annurev-biochem-011420-092302.
20. Simon, Gabriel M., Micah J. Niphakis, and Benjamin F. Cravatt. 2013. “Determining Target Engagement in Living Systems.” *Nature Chemical Biology* 9 (4): 200–205. https://doi.org/10.1038/nchembio.1211.
21. Grimwood, Sarah, and Paul R. Hartig. 2009. “Target Site Occupancy: Emerging Generalizations from Clinical and Preclinical Studies.” Pharmacology & Therapeutics 122 (3): 281–301. https://doi.org/10.1016/j.pharmthera.2009.03.002

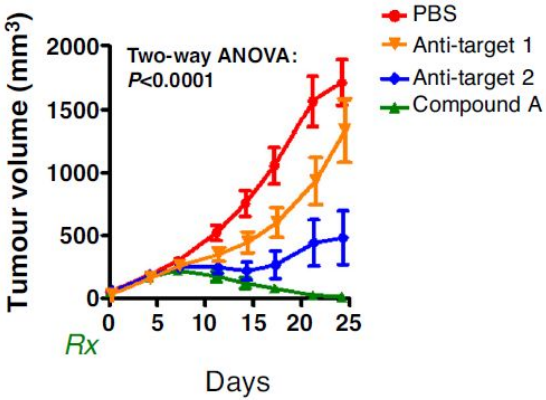
References (continued)

22. Woude, Lieke L. van der, Mark A. J. Gorris, Altuna Halilovic, Carl G. Figdor, and I. Jolanda M. de Vries. 2017. “Migrating into the Tumor: A Roadmap for T Cells.” Trends in Cancer 3 (11): 797–808. https://doi.org/10.1016/j.trecan.2017.09.006.
23. Griffiths, Jason I., Pierre Wallet, Lance T. Pflieger, David Stenehjem, Xuan Liu, Patrick A. Cosgrove, Neena A. Leggett, et al. 2020. “Circulating Immune Cell Phenotype Dynamics Reflect the Strength of Tumor–Immune Cell Interactions in Patients during Immunotherapy.” Proceedings of the National Academy of Sciences, June. https://doi.org/10.1073/pnas.1918937117.
24. Lotka-Volterra equation visualization: https://upload.wikimedia.org/wikipedia/commons/d/d7/Lotka_Volterra_equation_Maple_plot.png, adapted due to possible mislabelling
25. MacKay, David J. C. 2003. Information Theory, Inference, and Learning Algorithms. Cambridge, UK ; New York: Cambridge University Press. http://www.inference.org.uk/mackay/itila/book.html.
26. McElreath, Richard. 2020. Statistical Rethinking: A Bayesian Course with Examples in R and Stan. 2nd ed. CRC Texts in Statistical Science. Boca Raton: Taylor and Francis, CRC Press.
27. Macnamara, Cicely K. 2021. “Biomechanical Modelling of Cancer: Agent-Based Force-Based Models of Solid Tumours within the Context of the Tumour Microenvironment.” Computational and Systems Oncology 1 (2): e1018. https://doi.org/10.1002/cso2.1018.
28. Gündner, Anna Lisa, Gonzalo Duran-Pacheco, Silke Zimmermann, Iris Ruf, Tim Moors, Karlheinz Baumann, Ravi Jagasia, Wilma D. J. van de Berg, and Thomas Kremer. 2019. “Path Mediation Analysis Reveals GBA Impacts Lewy Body Disease Status by Increasing α-Synuclein Levels.” Neurobiology of Disease 121 (January): 205–13. https://doi.org/10.1016/j.nbd.2018.09.015.
29. Wright, Sewall. 1920. “The Relative Importance of Heredity and Environment in Determining the Piebald Pattern of Guinea-Pigs.” Proceedings of the National Academy of Sciences 6 (6): 320–32. https://doi.org/10.1073/pnas.6.6.320.
30. Pearl, Judea. 2009. “Causal Inference in Statistics: An Overview.” Statistics Surveys 3: 96–146. https://doi.org/10.1214/09-SS057.

References (continued)

31. Krishnamoorthy, Vijay, Danny J. N. Wong, Matt Wilson, Karthik Raghunathan, Tetsu Ohnuma, Duncan McLean, S. Ramani Moonesinghe, and Steve K. Harris. 2020. “Causal Inference in Perioperative Medicine Observational Research: Part 1, a Graphical Introduction.” British Journal of Anaesthesia 125 (3): 393–97. https://doi.org/10.1016/j.bja.2020.03.031.
32. Krishnamoorthy, Vijay, Duncan McLean, Tetsu Ohnuma, Steve K. Harris, Danny J. N. Wong, Matt Wilson, Ramani Moonesinghe, and Karthik Raghunathan. 2020. “Causal Inference in Perioperative Medicine Observational Research: Part 2, Advanced Methods.” British Journal of Anaesthesia 125 (3): 398–405. https://doi.org/10.1016/j.bja.2020.03.032.
33. Senn, Stephen. 2018. “Statistical Pitfalls of Personalized Medicine.” Nature 563 (7733): 619. https://doi.org/10.1038/d41586-018-07535-2.
34. Xu, Chao, Patanjali Ravva, Jun Steve Dang, Johann Laurent, Céline Adessi, Christine McIntyre, Georgina Meneses-Lorente, and François Mercier. 2018. “A Continuous-Time Multistate Markov Model to Describe the Occurrence and Severity of Diarrhea Events in Metastatic Breast Cancer Patients Treated with Lumretuzumab in Combination with Pertuzumab and Paclitaxel.” Cancer Chemotherapy and Pharmacology 82 (3): 395–406. https://doi.org/10.1007/s00280-018-3621-9.
35. Kather, Jakob Nikolas, Jan Poleszczuk, Meggy Suarez-Carmona, Johannes Krisam, Pornpimol Charoentong, Nektarios A. Valous, Cleo-Aron Weis, et al. 2017. “In Silico Modeling of Immunotherapy and Stroma-Targeting Therapies in Human Colorectal Cancer.” Cancer Research 77 (22): 6442–52. https://doi.org/10.1158/0008-5472.CAN-17-2006.
36. Beckers, Thomas. “An Introduction to Gaussian Process Models.” arXiv, February 10, 2021. https://doi.org/10.48550/arXiv.2102.05497.
37. https://www.fda.gov/media/167973/download
38.

A real-word example with a bispecific antibody



Anti-target 1 TO required for efficacy in tumour and PBMCs

Efficacious concentration

Pharmacokinetics

Intravenous PK

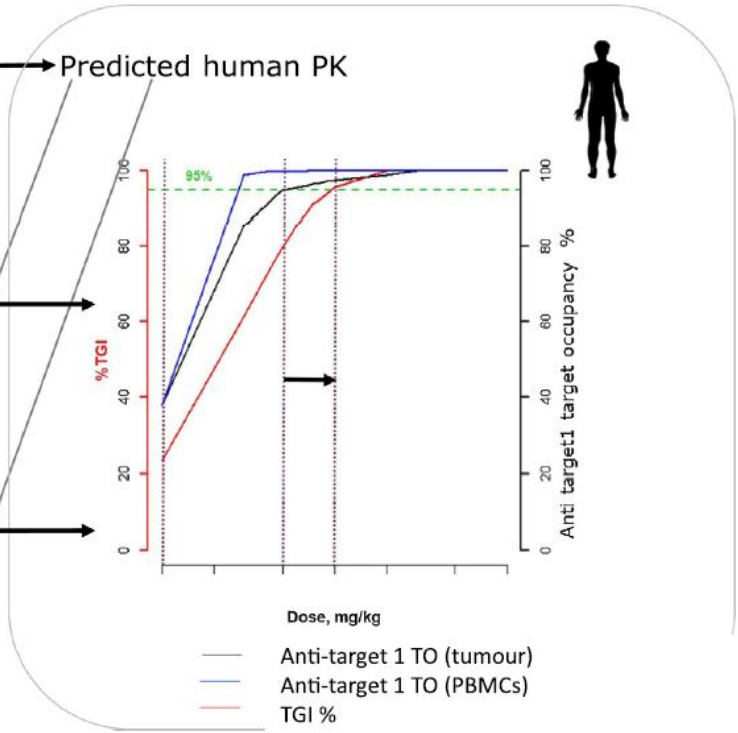
→ Predicted human PK

PK/PD model

Anti-target1 TO in EMT-6 model

PK/efficacy model

Tumour growth model in EMT-6 model



TGI: tumour growth inhibition; TO: target occupancy; PBMC, peripheral blood mononuclear cells

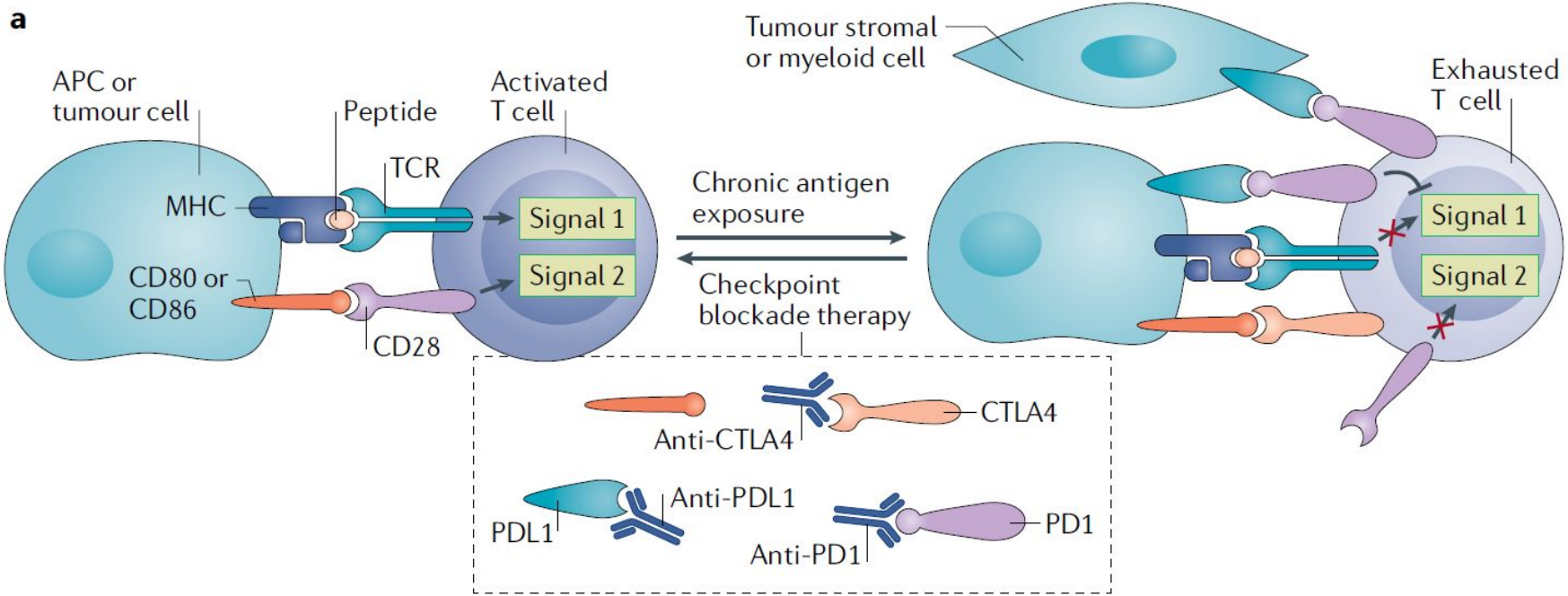
Exposure-response in animal model and translatable biomarkers are essential for dose prediction

TABLE 2

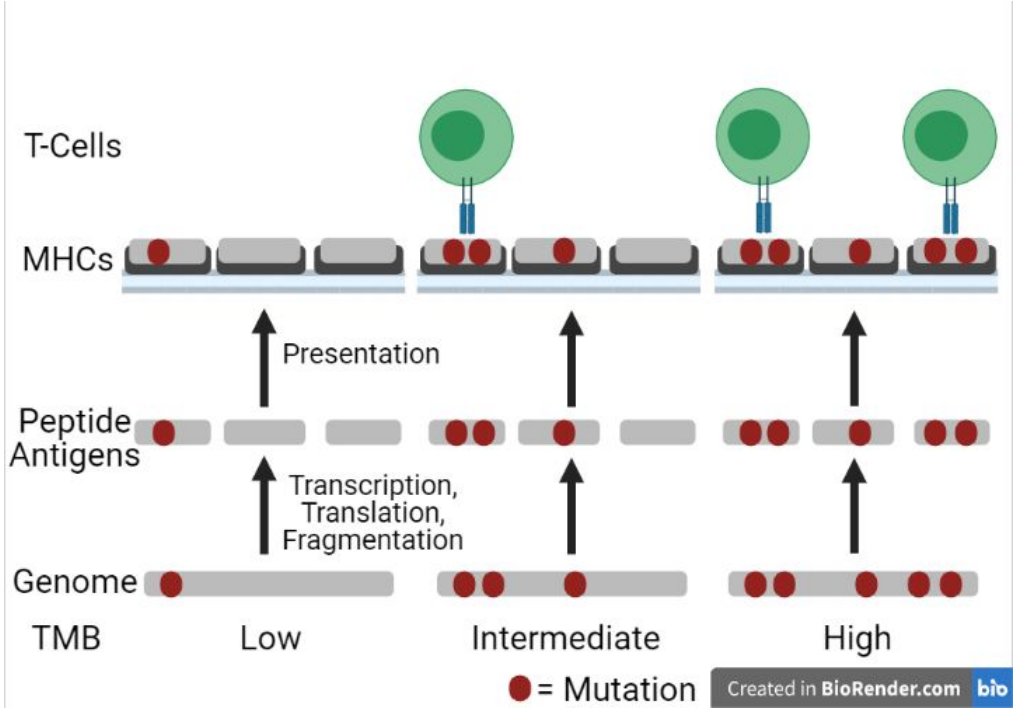
Correlation of responses to dose-related questions (Q) of TmX Guide to dose prediction successes or observation of efficacy in the clinic

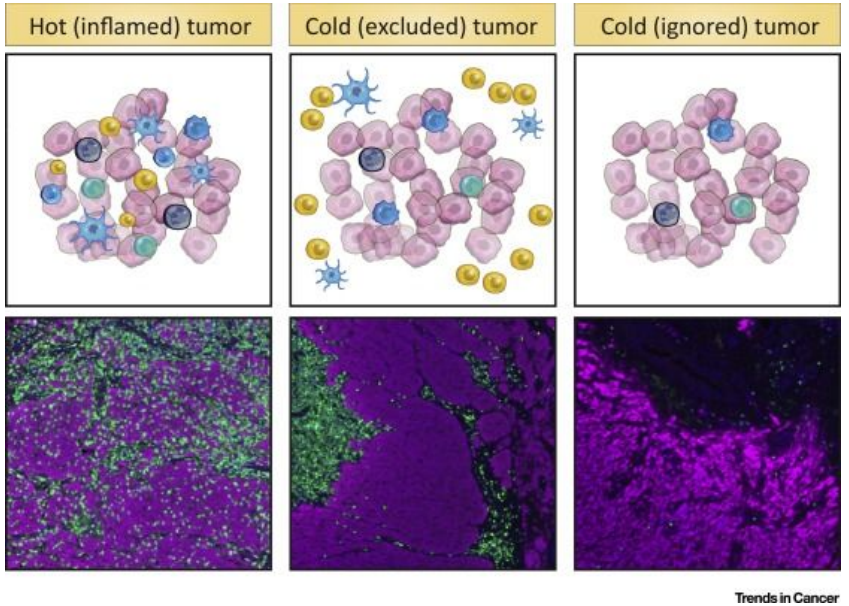
Category	Q1: desired exposure–response in appropriate animal model?	Q2: Translatable biomarkers?	Number of drugs for which model-based active dose prediction is within twofold or clinical efficacy is observed within predicted dose range out of total number in category
1	Yes	Yes	5/6
2[a]	No	No	1/6
3	No	Yes	2/2
4[b]	Yes	No	0/1

Molecular basis of cancer immunotherapy



Tumour mutation burden and immune phenotype may affect the effect of immunotherapy

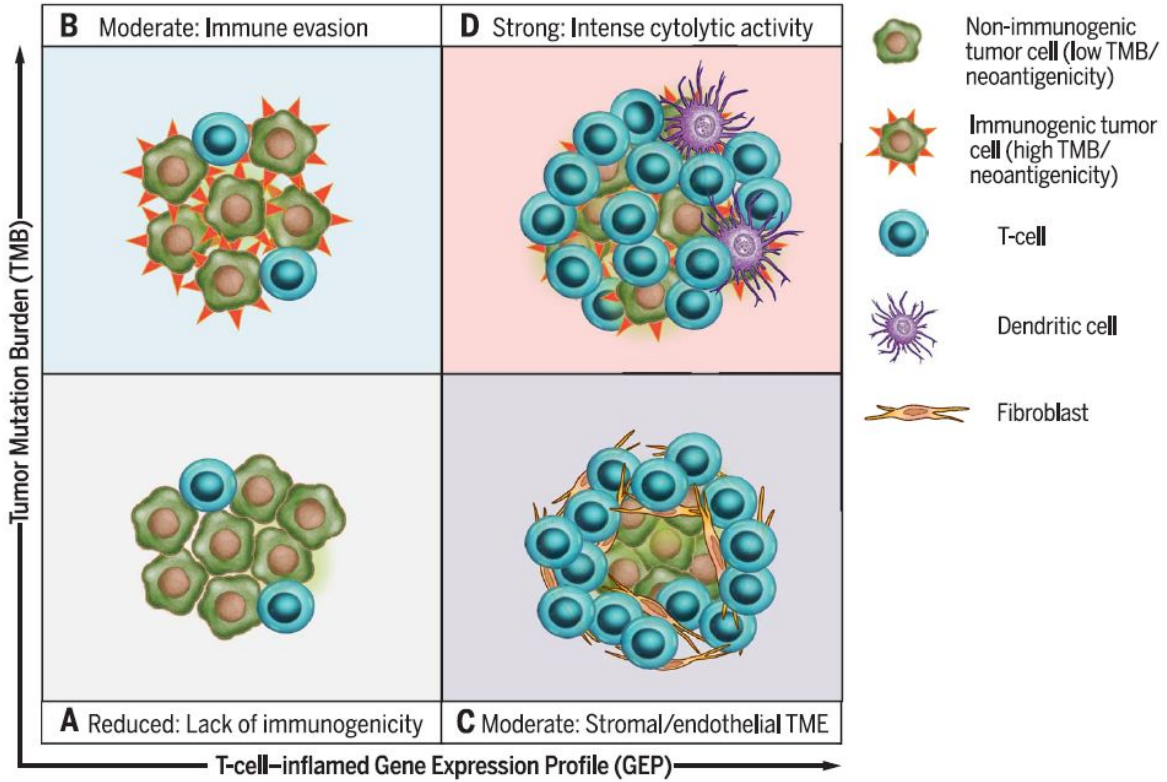




MHC: Major Histocompatibility Complex; TMB:Tumour Mutation Burden.

Cristescu *et al*. established TMB and T-cell-inflamed Gene Expression Profile (GEP) as biomarkers

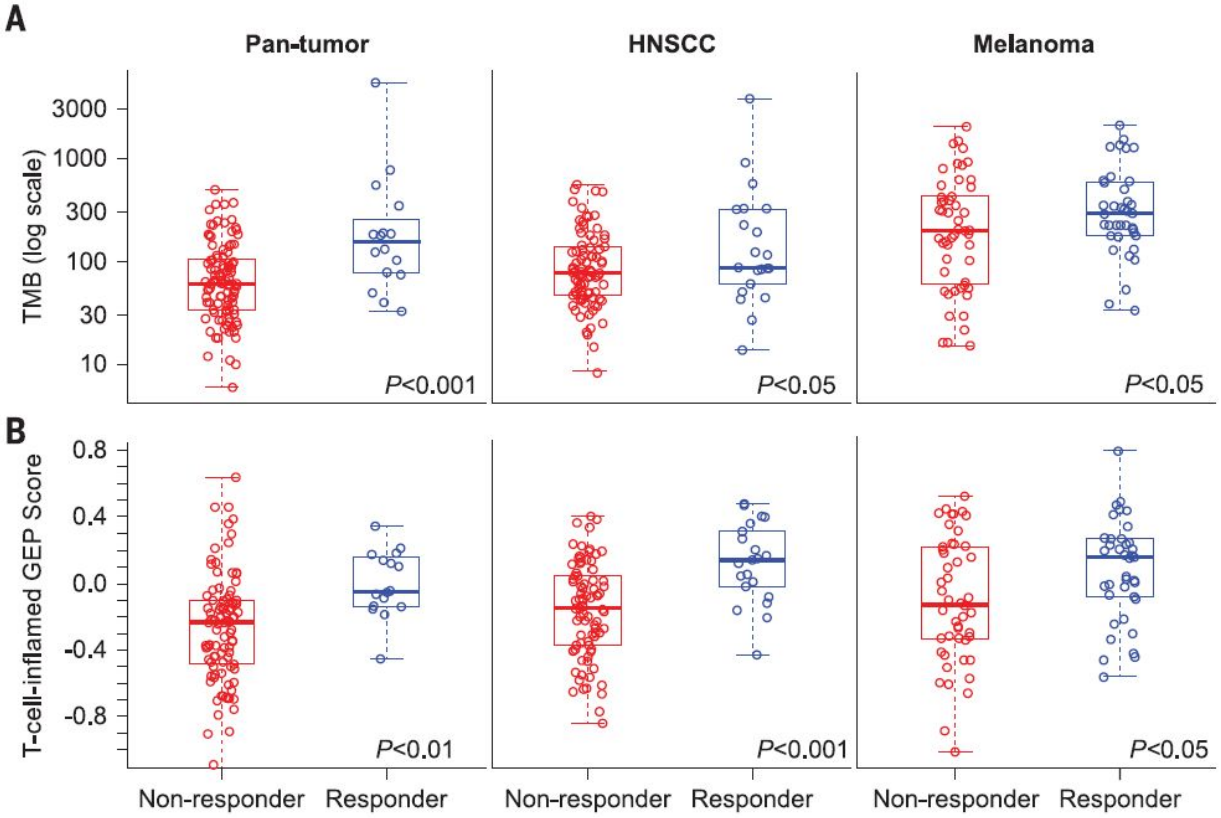
Patients with high tumor mutation burden **AND** a T-cell-inflamed gene expression profile (TME) are more likely to respond to cancer immunotherapy.



Univariate analysis establishes correlation between TMB/GEP and responsiveness

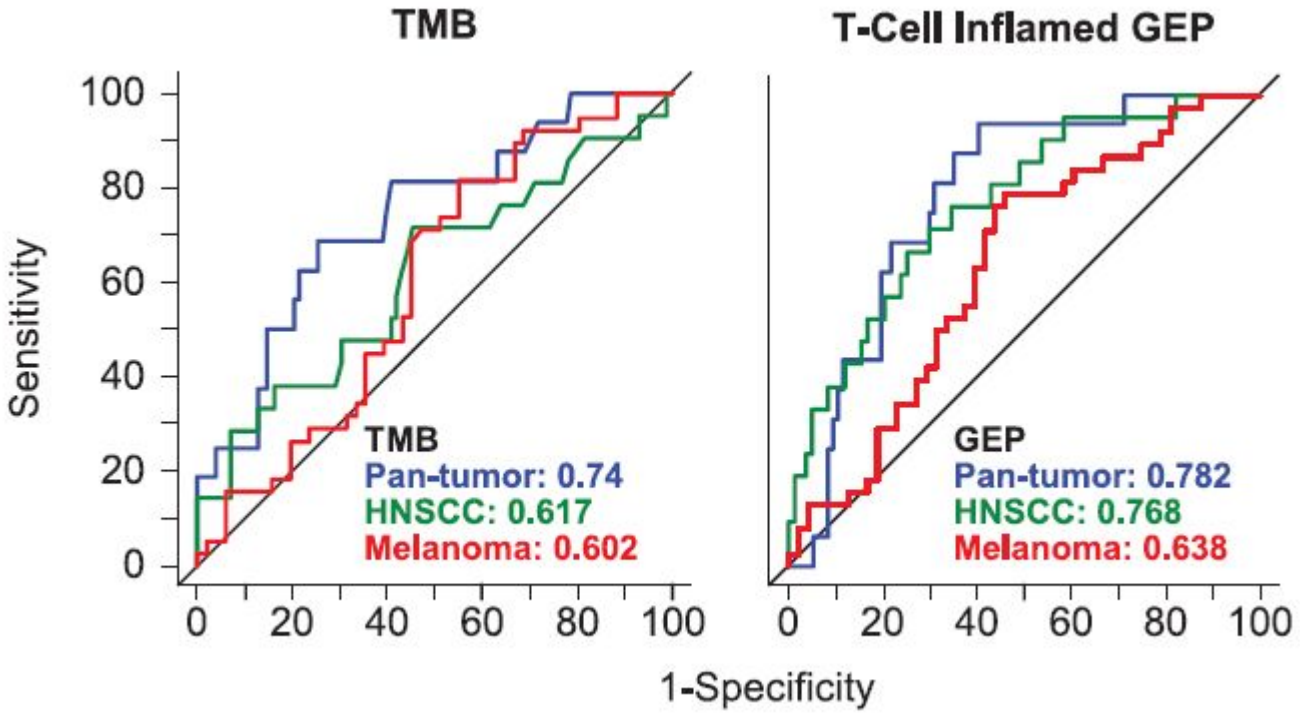
GEP: weighted sum of normalized expression of 18 genes related with immune response (CCL5, CD27, CD274 (PD-L1), CD276 (B7-H3), CD8A, CMKLR1, CXCL9, CXCR6, HLA-DQA1, HLA-DRB1, HLA-E, IDO1, LAG3, NKG7, PDCD1LG2 (PDL2), PSMB10, STAT1, and TIGIT).

HNSCC: head and neck cancer



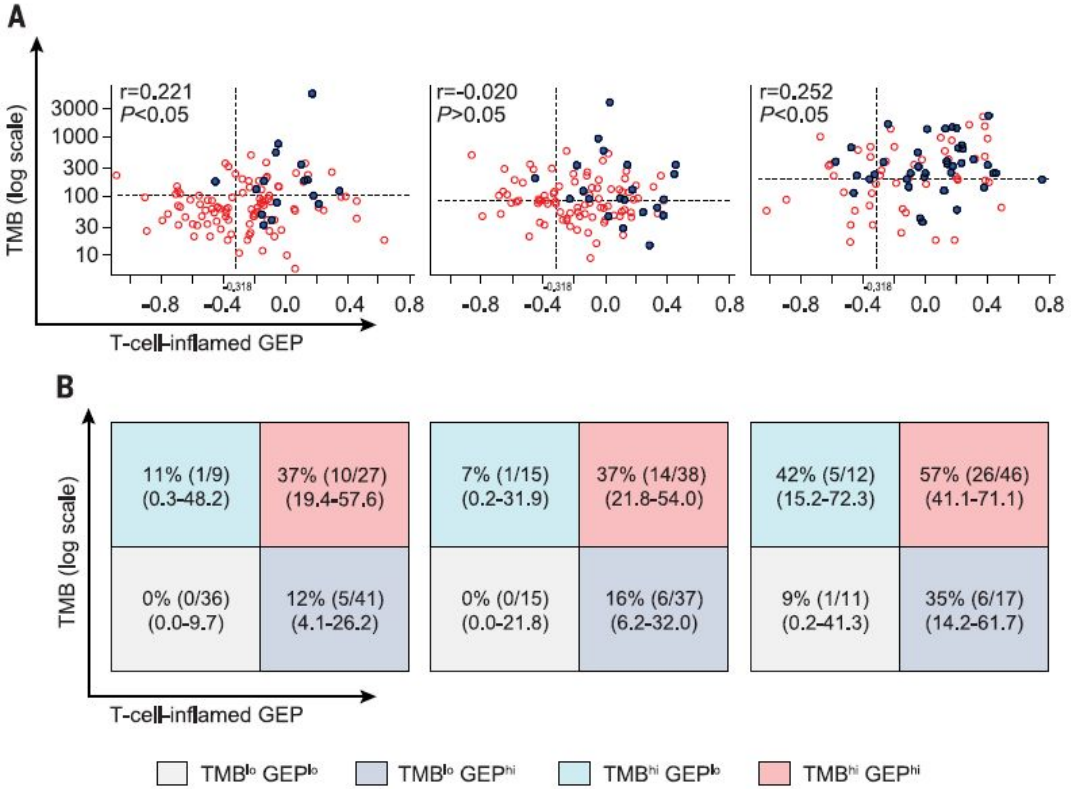
Both TMB and GEP can partially predict responsiveness

Receiver Operating Characteristic (ROC) curves of using either TMB or GEP for binary classification. Metrics: Area Under ROC (AUROC)

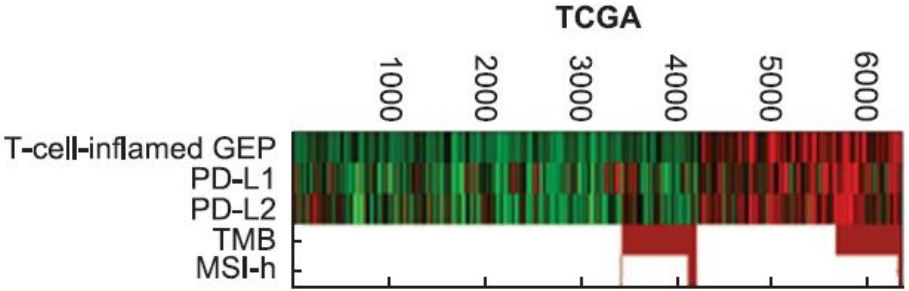


High TMB and high GEP are associated with higher responsiveness to anti-PD1 antibody treatment

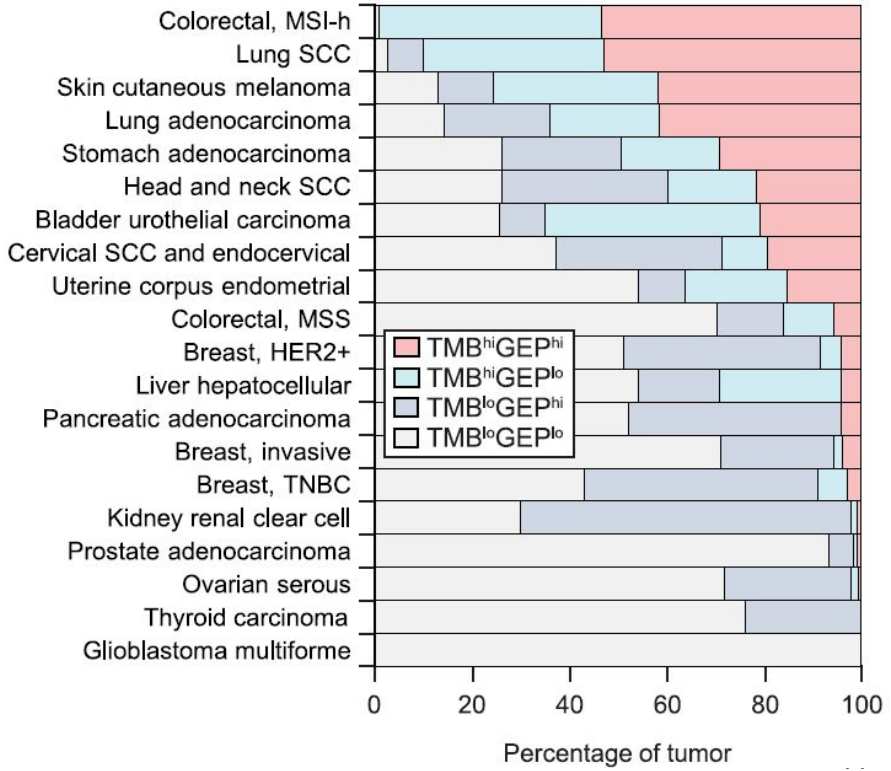
- From left to right: three patient cohorts (pan-cancer; head-and-neck cancer; melanoma)
- Open red circles: non responders; Black dots: responders.



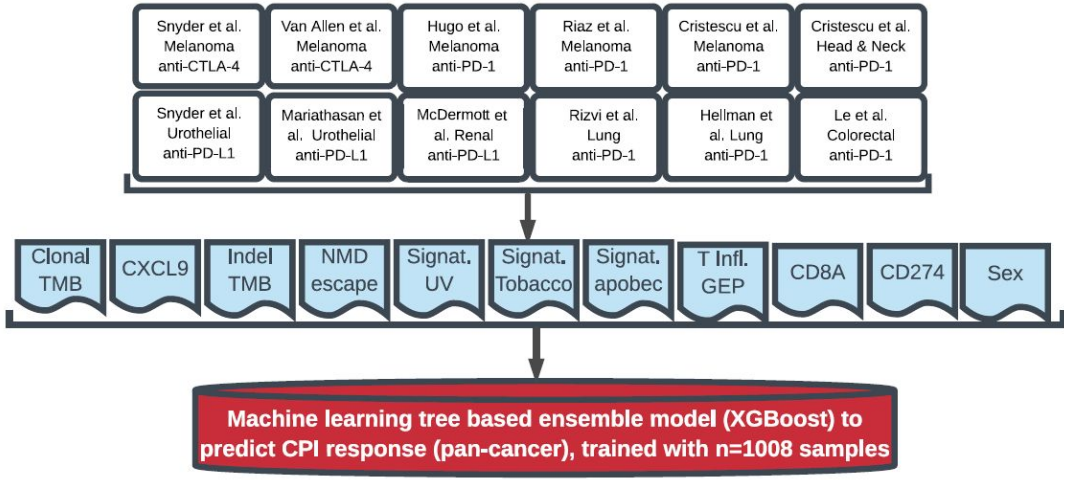
Data mining in public cancer database TCGA suggests potential indications

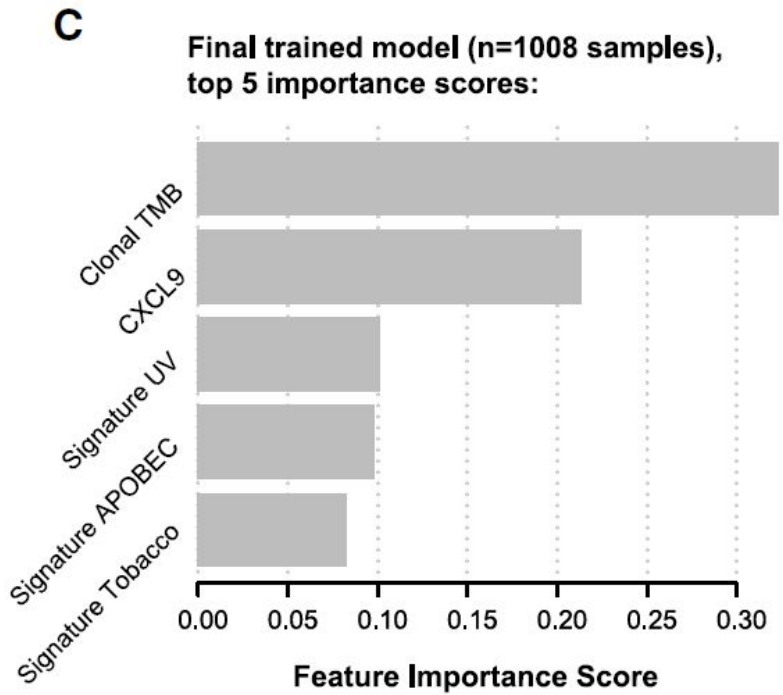


TMB Cut-off (WES)	GEP ≤top tertile (%)	GEP >top tertile (%)
>100	11	12
≤100	55	22



Meta-analysis (Litchfield *et al.* 2021) confirms TMB and T-cell infiltration as predictors of responsiveness



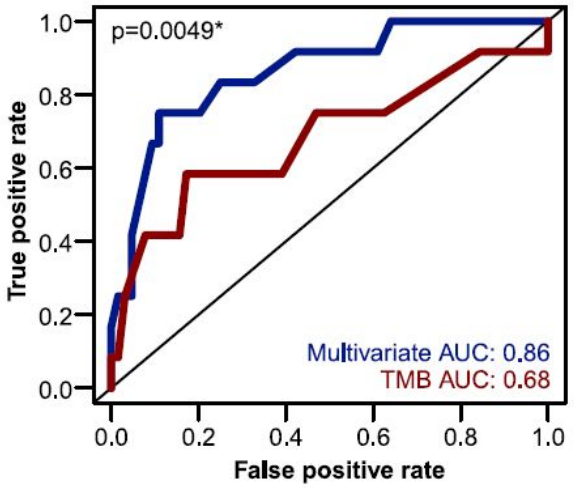


CXCL9 is a chemokine that enhances recruitment of cytotoxic $CD8^+$ T cells into the tumor.

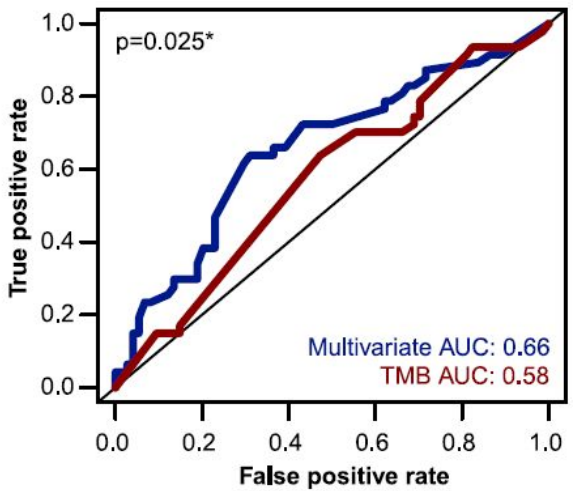
The multivariate classifier improves performance, but to predict responsiveness is an open question

Testing of TMB versus multivariable CPI stratifier performance in three independent test cohorts (total n=341):

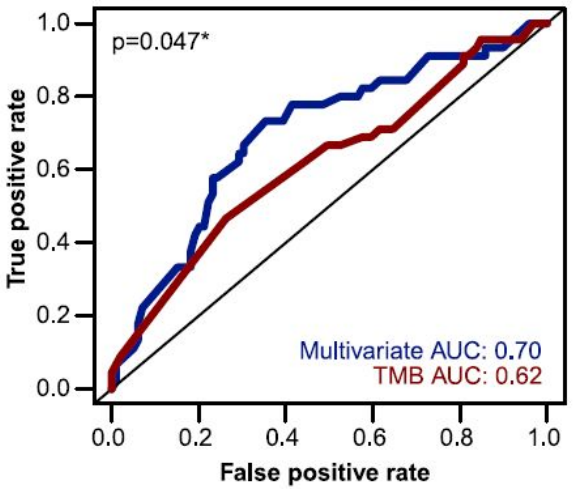
Test cohort 1: KEYNOTE-028 study
(other tumor types, Cristescu et al. 2018, n=76)



Test cohort 2: University Hospital Essen study
(melanoma, Liu et al. 2019, n=121)

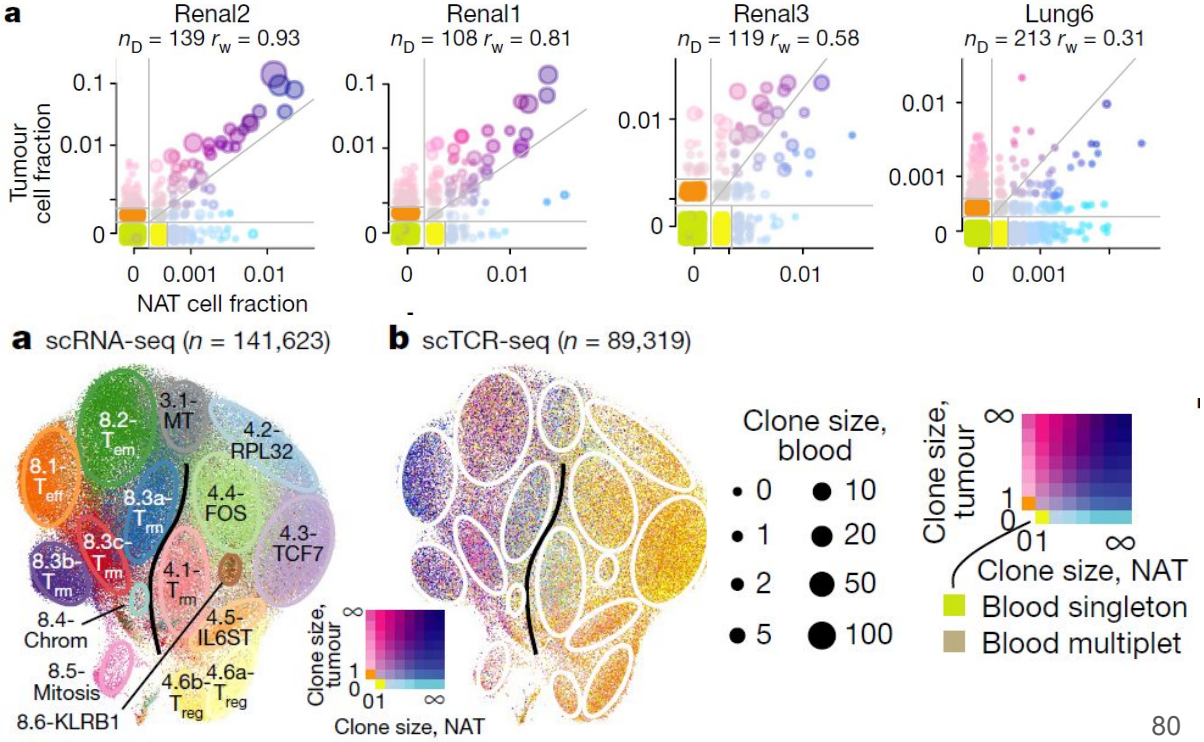


Test cohort 3: Samsung MC study
(NSCLC, Shim et al. 2020, n=144)



Wu *et al.* characterized T cells in tumour, normal adjacent tissue (NAT), and blood using single-cell RNA and TCR sequencing

- Expanded clonotypes (T cells) found in the tumour and normal adjacent tissue can also typically be detected in peripheral blood.
- Intra-tumoural T cells, especially in responsive patients, are replenished with fresh, non-exhausted replacement cells from sites outside the tumour.



Bonus: Mathematical modelling of epidemiology

The SIR (S=susceptible, I=infectious, R=removed) model modelling epidemiology (without viral dynamics, $N = S + I + R$).

$$\frac{dS}{dt} = -\frac{\beta IS}{N}, \tag{3}$$

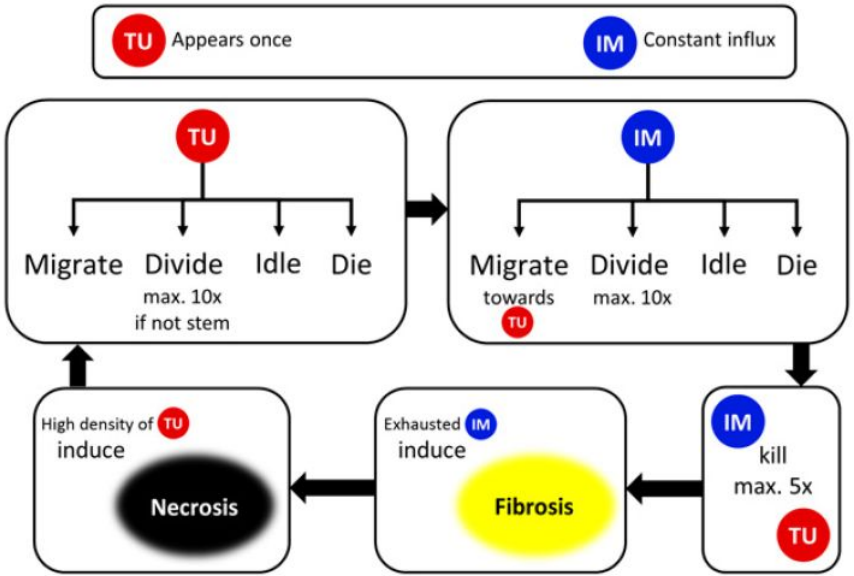
$$\frac{dI}{dt} = \frac{\beta IS}{N} - \gamma I, \tag{4}$$

$$\frac{dR}{dt} = \gamma I \tag{5}$$

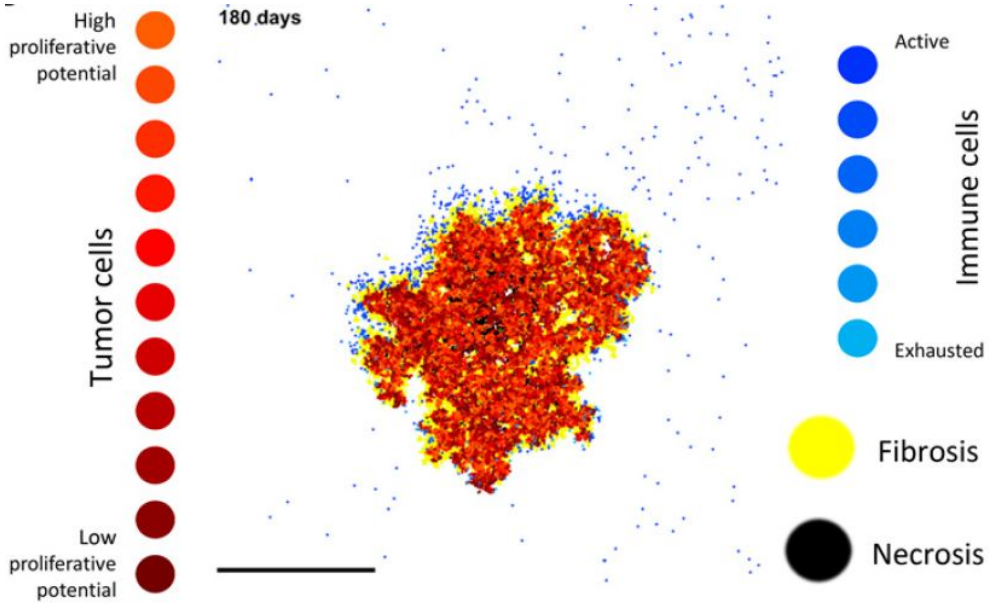
Agent-based modelling of cancer immunotherapy for colorectal cancer without high TMB

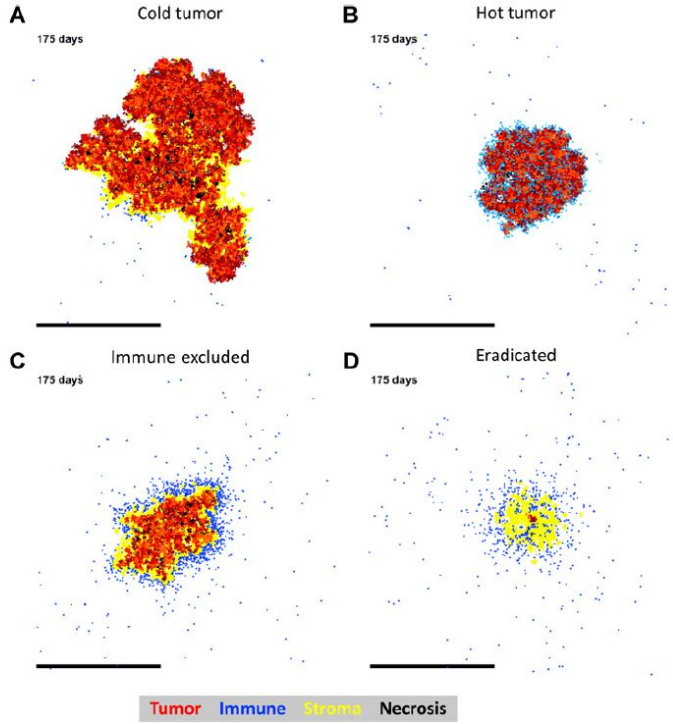
Table 1. Assumptions for the model and references for each assumption

Assumption	Ref.
All cells can migrate, proliferate, and die.	Trivial
Tumor cells are composed of stem cells and non-stem cells. Stem cells can divide symmetrically with a fixed probability.	(14)
Stem cells can proliferate indefinitely, all other cells die after a fixed number of proliferation cycles.	(14)
All cells can spontaneously enter apoptosis.	Own data
Tumor cells can spontaneously enter necrosis.	Own data
Tumor cells that are far from the outer margin have a higher probability of entering necrosis than those cells closer to the margin.	Own data
Immune cells are generated through a steady influx into the domain and proliferation within the domain.	(32), own data
Immune cells move by a "random walk" but have a tendency to migrate toward tumor cells.	(31–33), own data
Immune cells can kill adjacent tumor cells whenever they are close enough. Killing, like other events in the model, occurs stochastically with a fixed probability and is not regulated by other factors.	(23)
Immune cells can kill five times before they become exhausted, which means that they cannot kill anymore but can still proliferate.	(23, 34)
Activated immune cells give rise to stroma through a desmoplastic reaction (stroma reaction). For simplicity, this behavior is restricted to immune cells that have successfully killed five times in the model.	(35, 36)
By default, cells cannot migrate through stroma, but stromal permeability can be increased optionally.	(37)

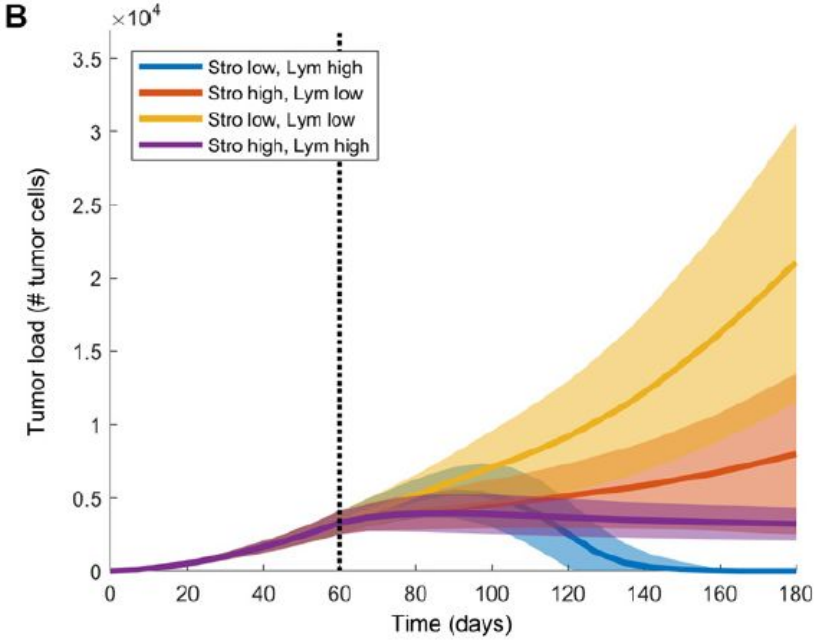


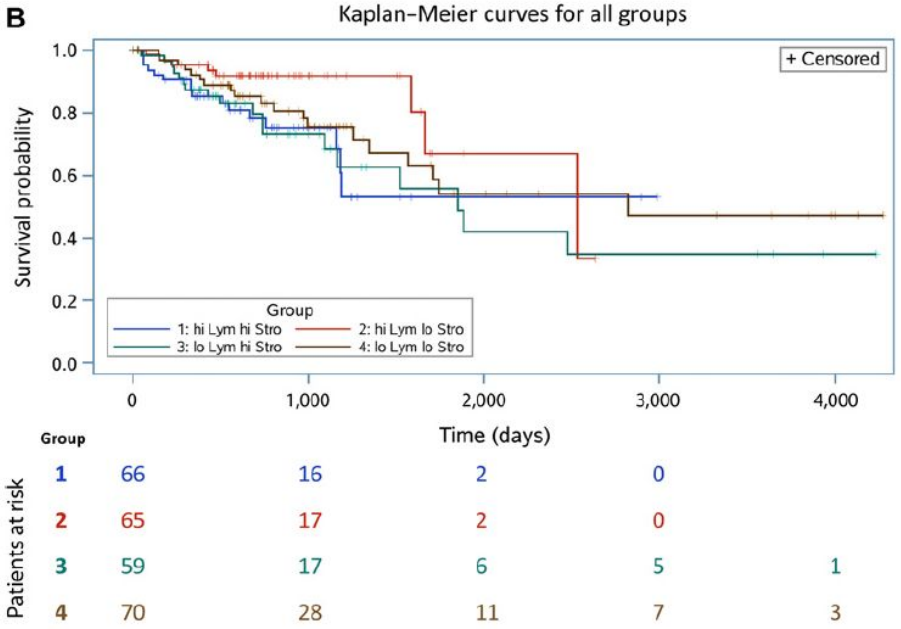
Agent-based modelling of cancer immunotherapy for colorectal cancer without high TMB





Counterfactual and statistical analysis allow us *learn* from the models *confirm* the learnings

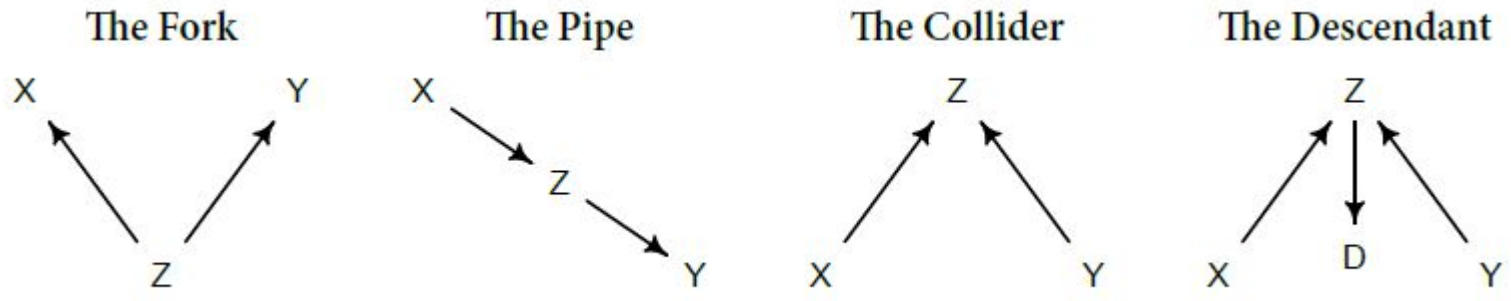




Bradford Hill Criteria for causation

1. **Strength** (effect size)
2. **Consistency** (reproducibility)
3. ***Specificity***
4. **Temporality**
5. **Biological gradient** (dose-response relationship)
6. ***Plausibility***
7. **Coherence**
8. **Experiment**
9. **Analogy** (similarity)
10. **Reversibility (proposed by others)**

Statistical causal inference with Directed Acyclic Graphs (DAGs)



Reading: chapter 1-6 of *Statistical Rethinking* (2nd Edition)