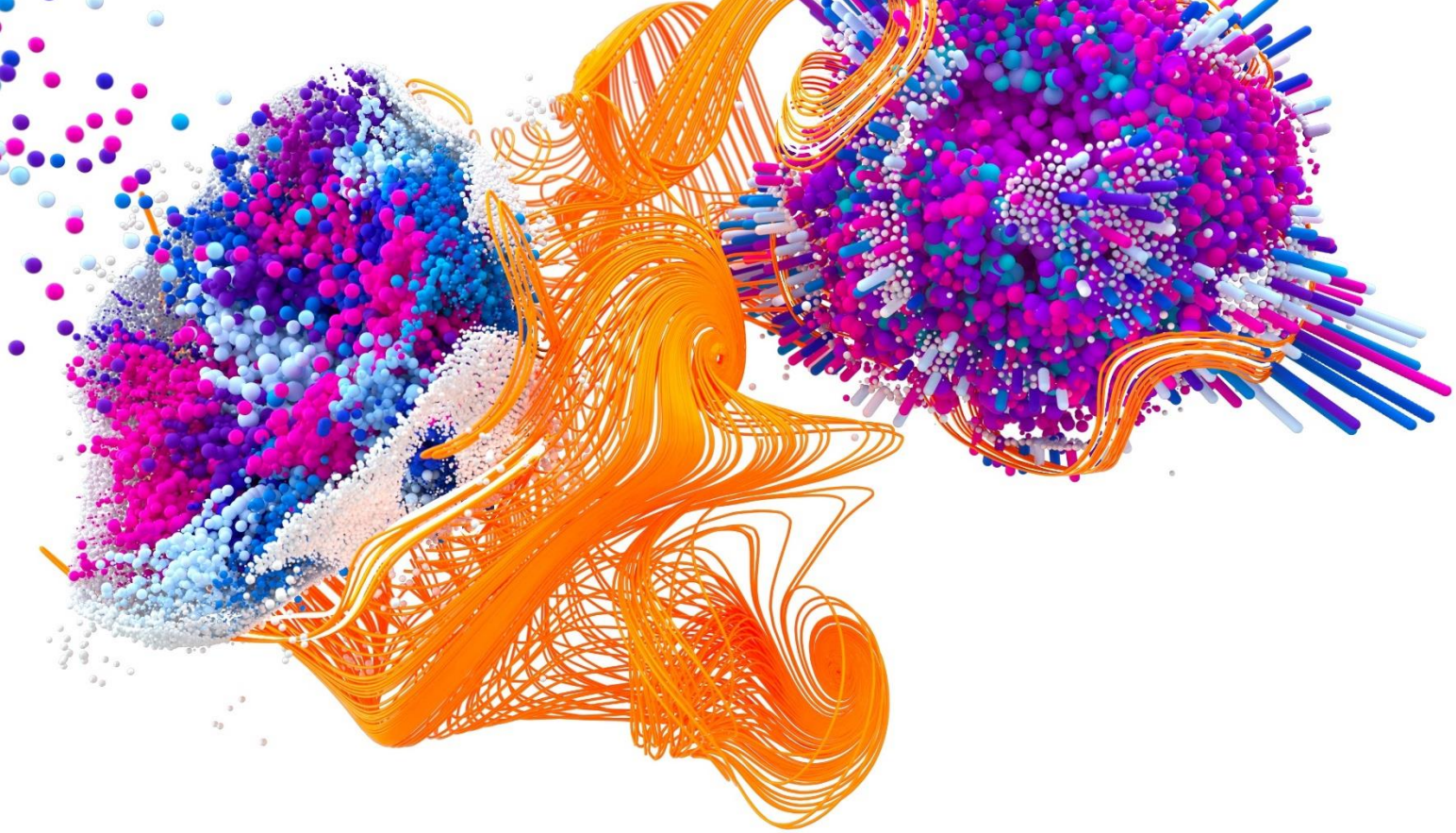


gsk.com

Learning the many languages of biomedicine

Dr Patrick Schwab (Head of Biomedical AI) on behalf of the team. EFGCP Better Medicines for Children Conference 2024

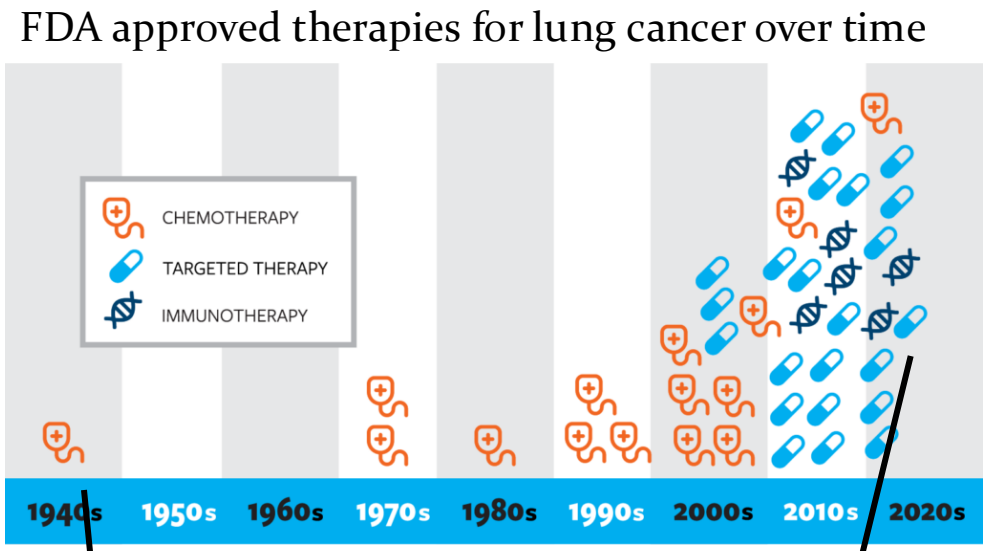
GSK.ai Biomedical AI group

- AI for Health and Biology, Software, and Technology
- Based in Heidelberg/Germany, Zug/Switzerland and London/UK

- We **help identify, monitor, and treat disease** with Clinical AI
- We **create a map of the immune system** using AI-guided experimentation
- We **advance the science of AI for Health** in partnership with leading European/Swiss institutions (e.g., ETH Zurich, Oxford, Cambridge, King's College)

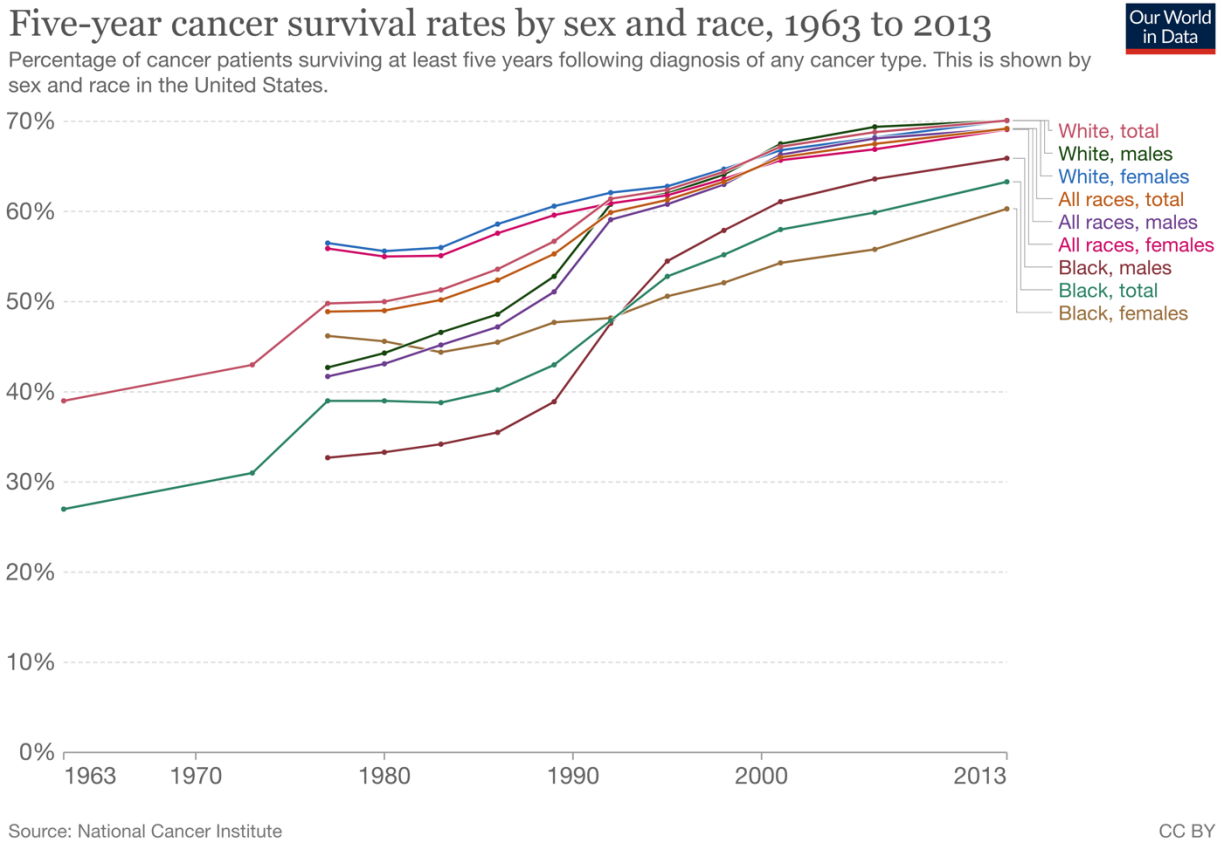
the team

The good: drug development works for society



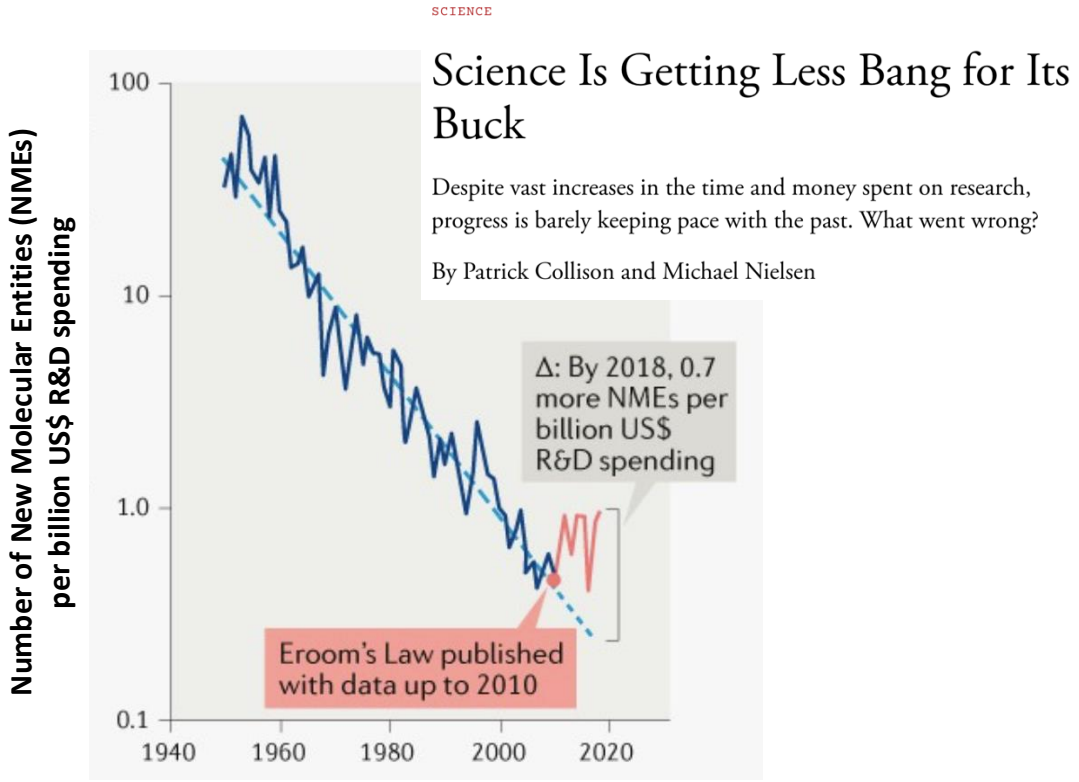
Mechlorethamine Hydrochloride
Candidate chemical warfare agent research

Mobocertinib
NSCLC with EGFR exon 20 insertion w/ progression after platinum therapy



Innovative medicines play a crucial role in enabling humans to live longer, healthier lives.

The bad: we are doing less with (exponentially) more



"Eroom's law":
Exponential drop in R&D productivity.

- **Failure is the norm**: Probability of success for a new medicine is **~5.5%**
- **Median cost** of a new drug is **$1.1 billion**
- Trend stable for decades with recent reversal (potentially) due to **increasing personalization** and **molecular/genetic evidence**.

Ringel, Michael S., et al. "Breaking Eroom's Law." Nature reviews. Drug discovery (2020).
Wouters et al. "Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018". JAMA (2020)
Ringel et al. "Breaking Eroom's law." Nature reviews Drug Discovery (2020)

What do we know works?

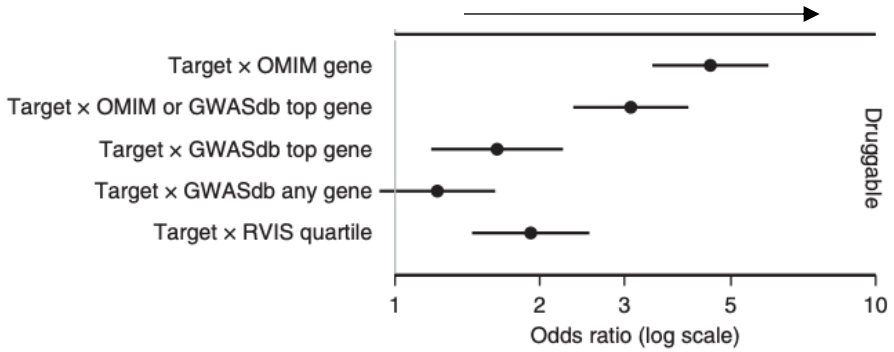
Population-scale Biobanks –
Variation in Genetic Background
(>100’000 people)

FINNGEN

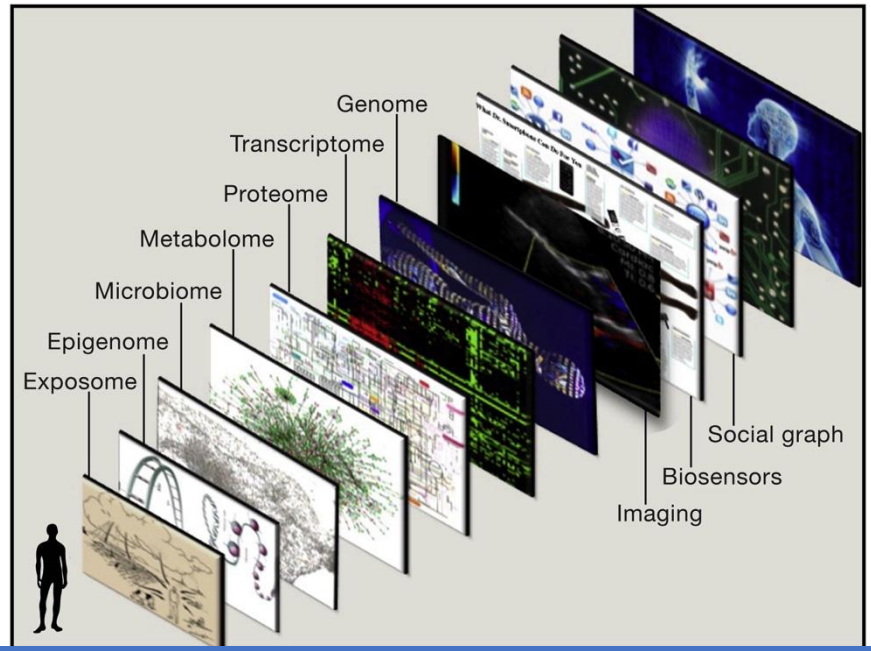
biobank uk

All of Us RESEARCH PROGRAM

Higher probability of approval (EU/US)



Targeted Biomarkers supporting Therapy
2x (immune) to 8x (oncology) higher probability of trial success



A causal & targeted link between disease of interest and a molecular mechanism substantially improves our chances of successful translation.

Wong et al. “Estimation of clinical trial success rates and related parameters ” Biostatistics (2018)

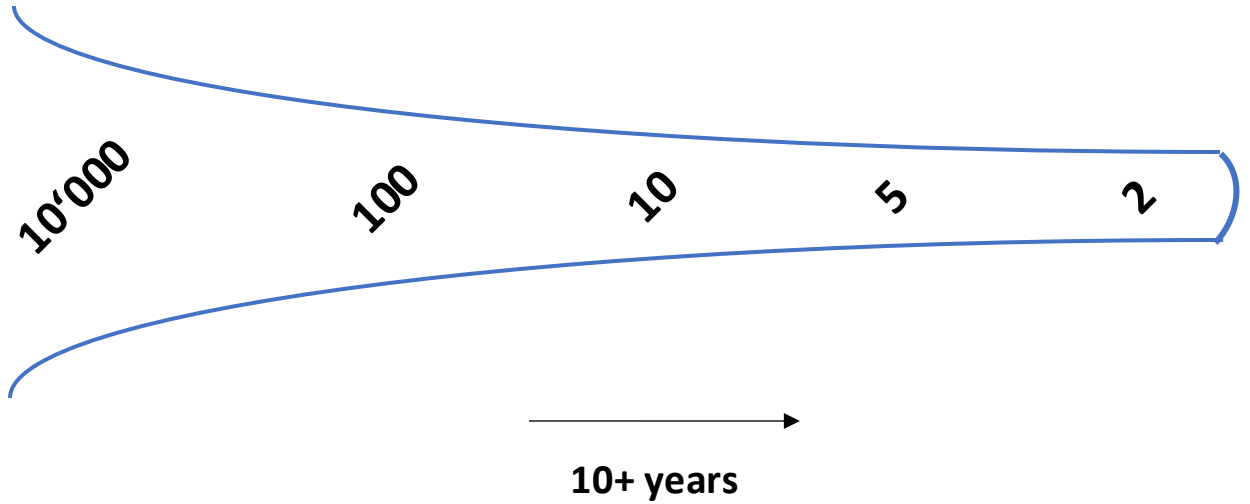
Topol, Eric J. "Individualized medicine from prewomb to tomb." Cell 157.1 (2014): 241-253.

Nelson, Matthew R., et al. "The support of human genetic evidence for approved drug indications." Nature genetics 47.8 (2015): 856-860.

Ringel, Michael S., et al. "Breaking Eroom's Law." Nature reviews. Drug discovery (2020).

Drug development process

	Discovery	Preclinical	Phase I	Phase II	Phase III
Probability of success		60%	50%	30%	60%



1 approved medicine

Drug discovery is an information science mainly concerned with making optimal decisions under uncertainty at every step of the development process.

Possibly the most complex human endeavor.

Cook, David, et al. "Lessons learned from the fate of AstraZeneca's drug pipeline: a five-dimensional framework." Nature reviews Drug discovery 13.6 (2014): 419-431.

What would the ideal patient journey look like?

Prevention
we reduce risk of future disease

Diagnosis
we identify at risk patients early for testing

Treatment
we provide access to precise medicines with high success rates

Monitoring
we understand individual long-term risk at any time

Optimal long-term health and well-being

Enormous potential for improvements in clinical care

- Patients often face diagnostic delays (e.g., 3+ years in attention-deficit/hyperactivity and Autism spectrum disorders[1,2])
- Patients are undertreated (e.g., 30% untreated in ADHD – some 3m in US[3])
- When treated, they often do not respond well (e.g., response rates as low as 20% to some ASD treatments)
- Frequently poorly understood and limited long-term monitoring[4]

1 Diagnostic delay in children with neurodevelopmental conditions attending a publicly funded developmental assessment service: findings from the Sydney Child Neurodevelopment Research Registry BMJ Open 2022
2 Age at diagnosis and diagnostic delay across attention-deficit hyperactivity and autism spectrums Aust N Z J Psychiatry. 2023 Oct 27;58(2):142–151.
3 https://www.cdc.gov/adhd/data/index.html
4 Peterson, B. S., Trampush, J., Maglione, M., Bolshakova, M., Brown, M., Rozelle, M., ... & Hempel, S. (2024). ADHD Diagnosis and Treatment in Children and Adolescents.

How can AI help us do better for patients?

Prevention
we reduce risk of future disease

1 **Precision screening and testing**

Diagnosis
we identify at risk patients early for testing

Treatment
we provide access to precise medicines with high success rates

2 **Long-term risk Monitoring and prediction**

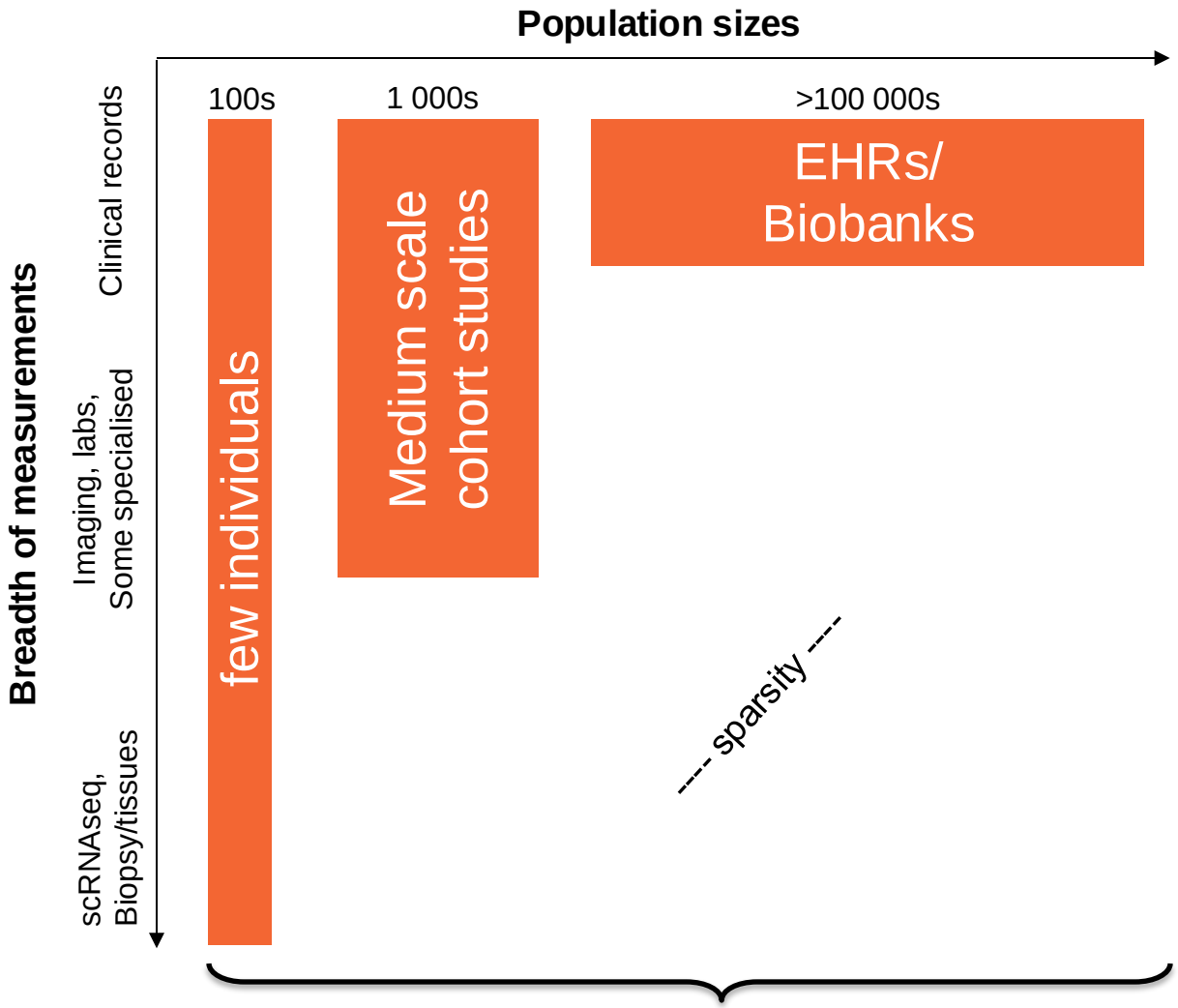
Monitoring
we understand individual long-term risk at any time

3 ***In-silico*** **treatment prediction**

Optimal long-term health and well-being

Challenge 1: The disease tensor

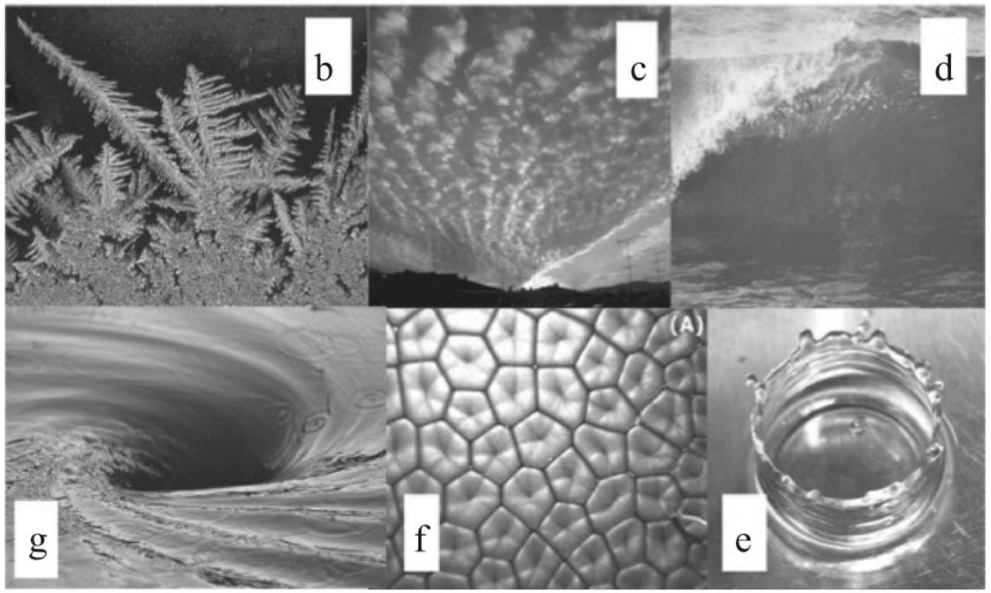
Sparse multiscale and multimodal data informs disease understanding and causal biomarkers

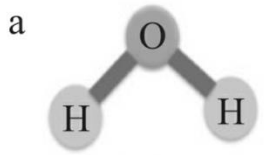


- The lower right of the disease tensor is **sparse**
- The sparsity is **associated with measurement complexity**
- The **sparsity alters over time**: new tech extends dimensionality, grows outwards as tech becomes scaled
- Different people at different scales, bias towards severity and social/economic in complex bucket
- AI/ML is well suited for continuously learning from this (growing) resource

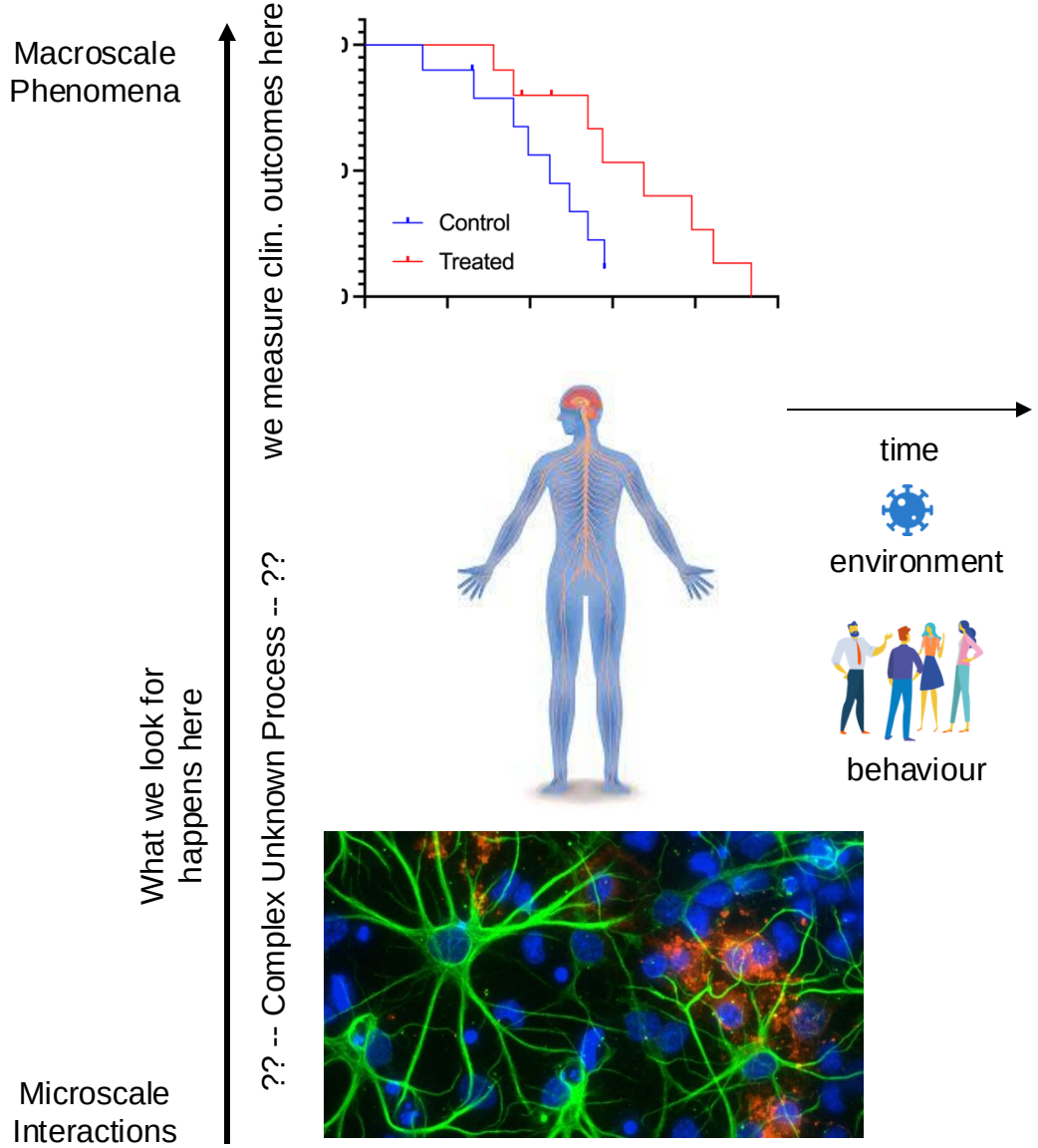
With data generation progressing on all scales, **making the most of the disease tensor** to derive (potentially) causal insights is quickly emerging as a **central challenge for computational medicine**.

Challenge 2: Disease is a system level function

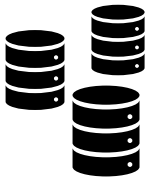




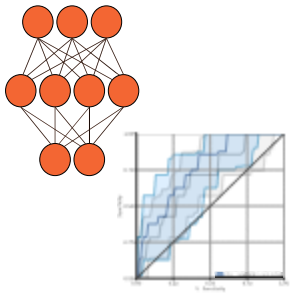
Our world is full of examples of (apparently) **simple components and underlying governing rules** giving rise to **enormous macroscale complexity and diversity** (emergence/self-organisation).



A flowchart for unravelling patient heterogeneity from patient data



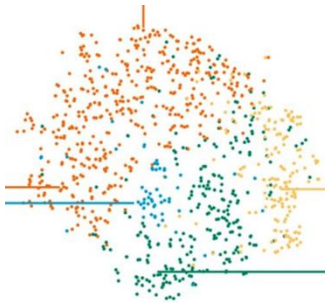
Collect diverse & representative data



Build response predictors

How predictable are outcomes and for which patients?

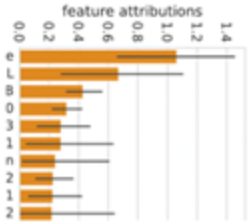
What data predicts outcomes better?



Derive patient subtypes

What are the groups of patients with similar trajectories?

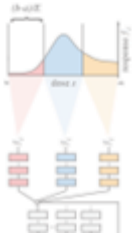
Are there any features that define their differences?



Identify biomarkers

What predictive markers predict different on-treatment trajectories? (clinical, molecular, …)

Combination Treatments



Response guided therapy

Data example: A patient's journey in clinical records

Patient demographics

Age, sex, location, …

Diagnoses

Diabetes, Type 1

Fatigue

Cirrhosis

Hepatitis B

Hepatocellular carcinoma

Prescriptions

Insulin

NSAIDs

Antivirals

Lab tests

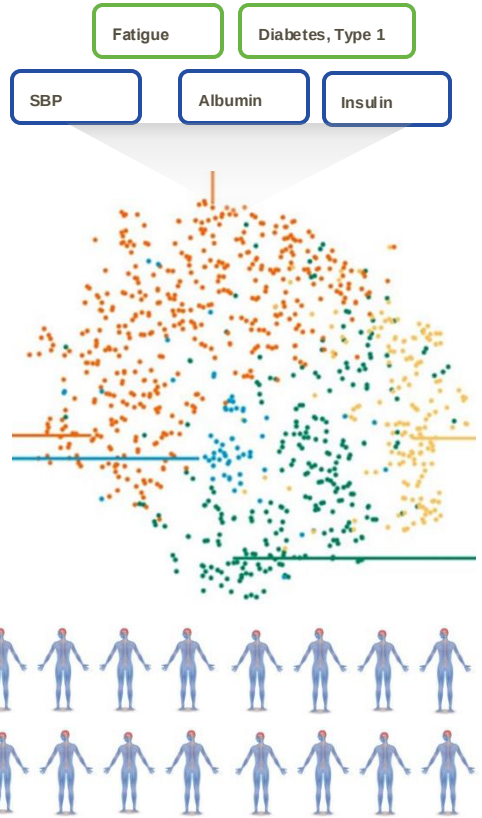
Albumin

Bilirubin ↑

SBP

Alpha feto-protein (AFP) ↑

Could we learn to understand the “language” of clinical records?



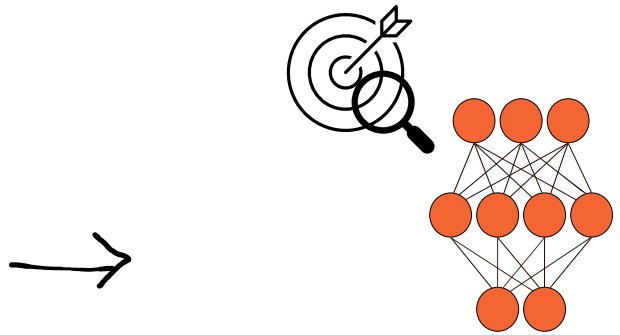
70’000 individuals from 3’000+ clinics

Who might carry, unknown to them, **Hepatitis B**?

Who is at risk of **Cirrhosis** or **Hepatocellular Carcinoma**?

AI/ML models uncover the hidden patterns of Hepatitis B infection …

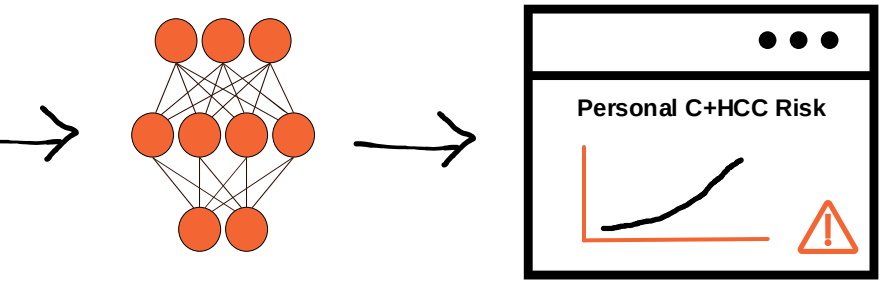
- Analgesics use and abdominal pain
- Male (2x risk)
- Body weight
- Low white blood cell count
- Increased bilirubin
- estimated glomerular filtration rate (eGFR) reduced

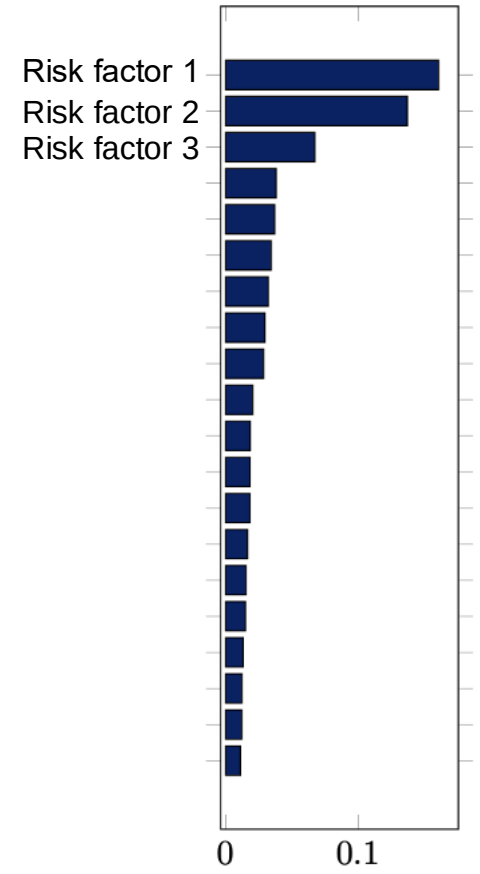


May enable prioritization for **targeted HepB screening.**

GSK

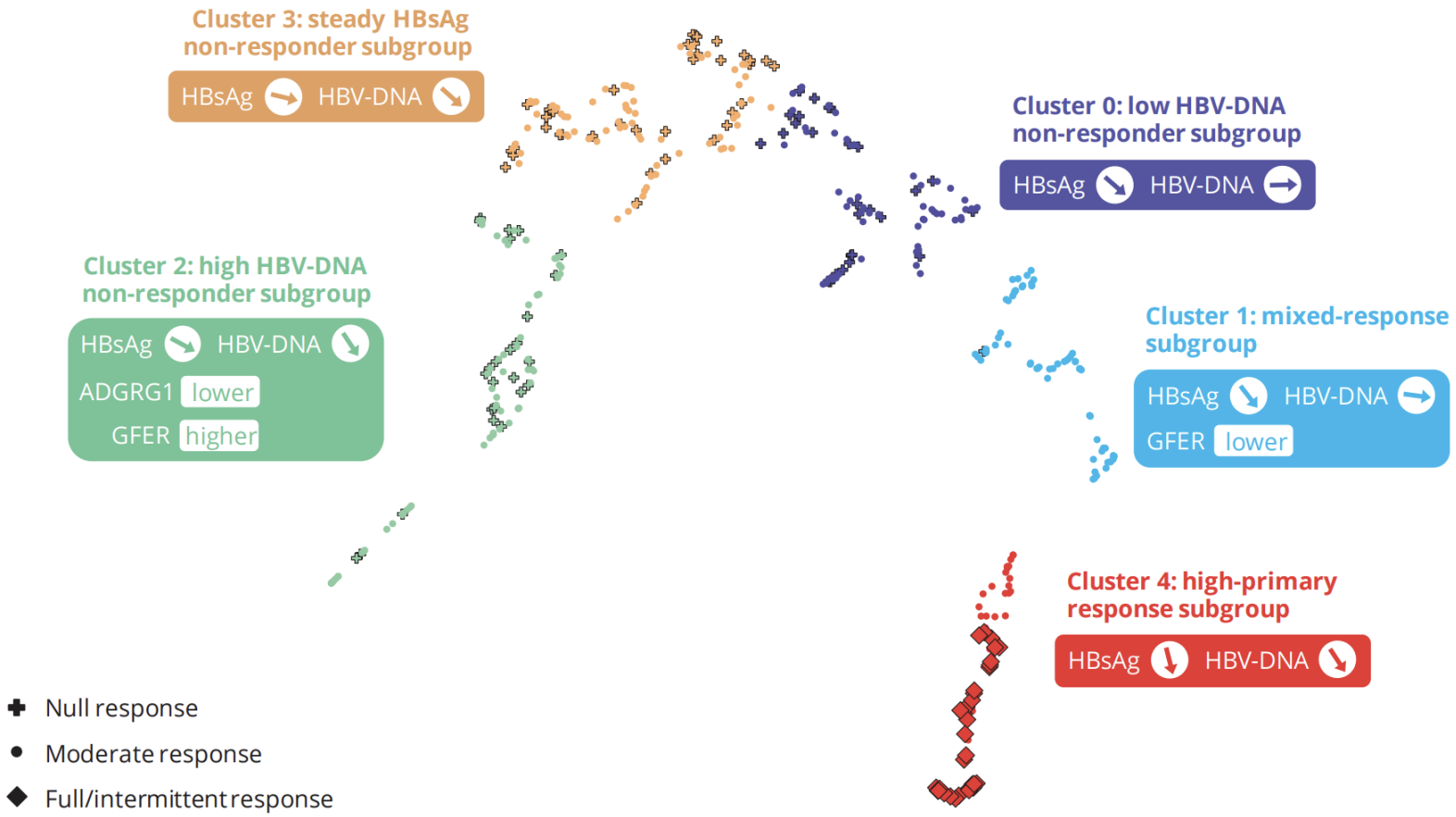
… tell us about *why* a patient may be at particularly high risk

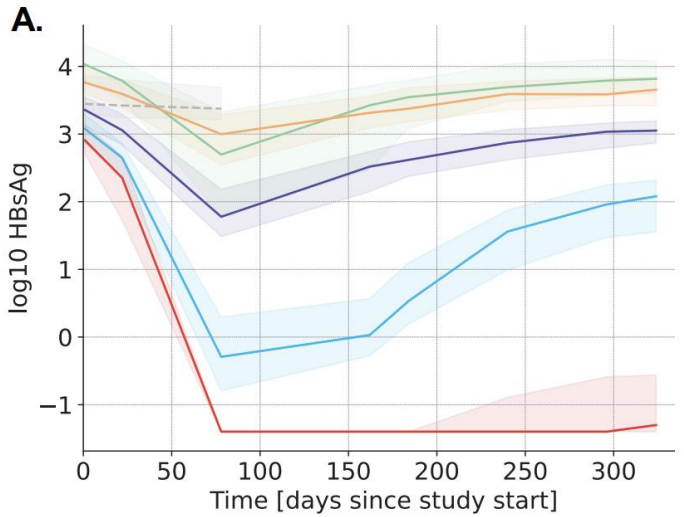




Quantifying disease risk by learning from 70’000+ individuals.

… and also enable mapping of response to investigational medicines



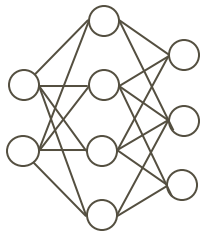


Weis et al. Deep learning cluster analysis reveals subtypes in response to antisense oligonucleotide therapy in chronic hepatitis B. EASL 2023

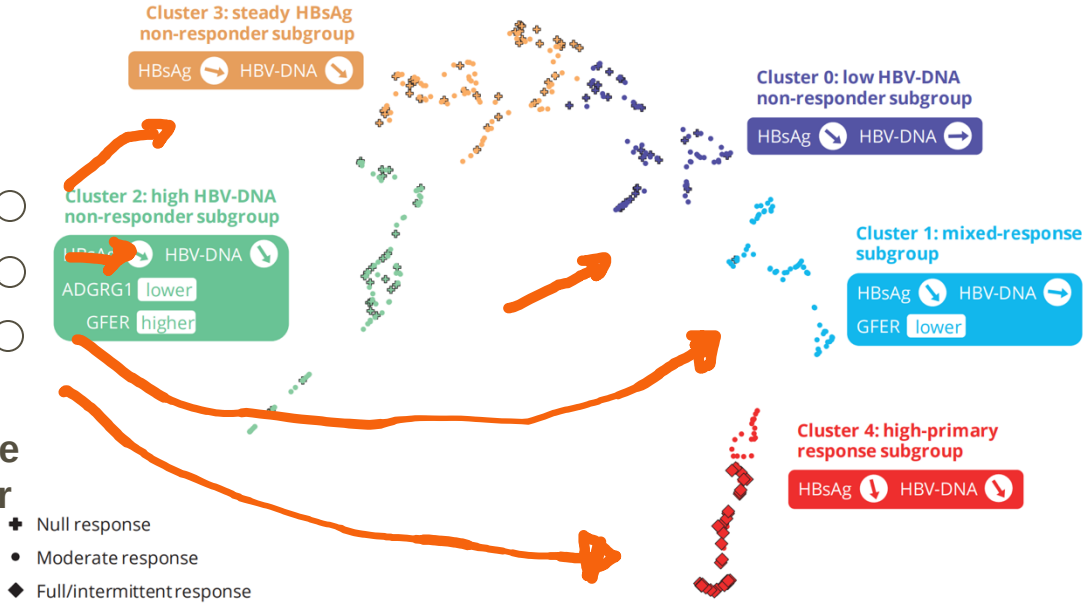
Response maps facilitate multi-modal prediction of on-treatment trajectories

Pre-treatment individual measurements

Proteomics *	Olink
Transcriptomics (Microarray)	Epistem
IgM	PPD
Enhanced liver fibrosis (ELF)	Nordic
N-terminal pro-peptide of type III collagen (PRO-C3)	
TLR8 mutations	Virology
Clinical & subject-level markers	various



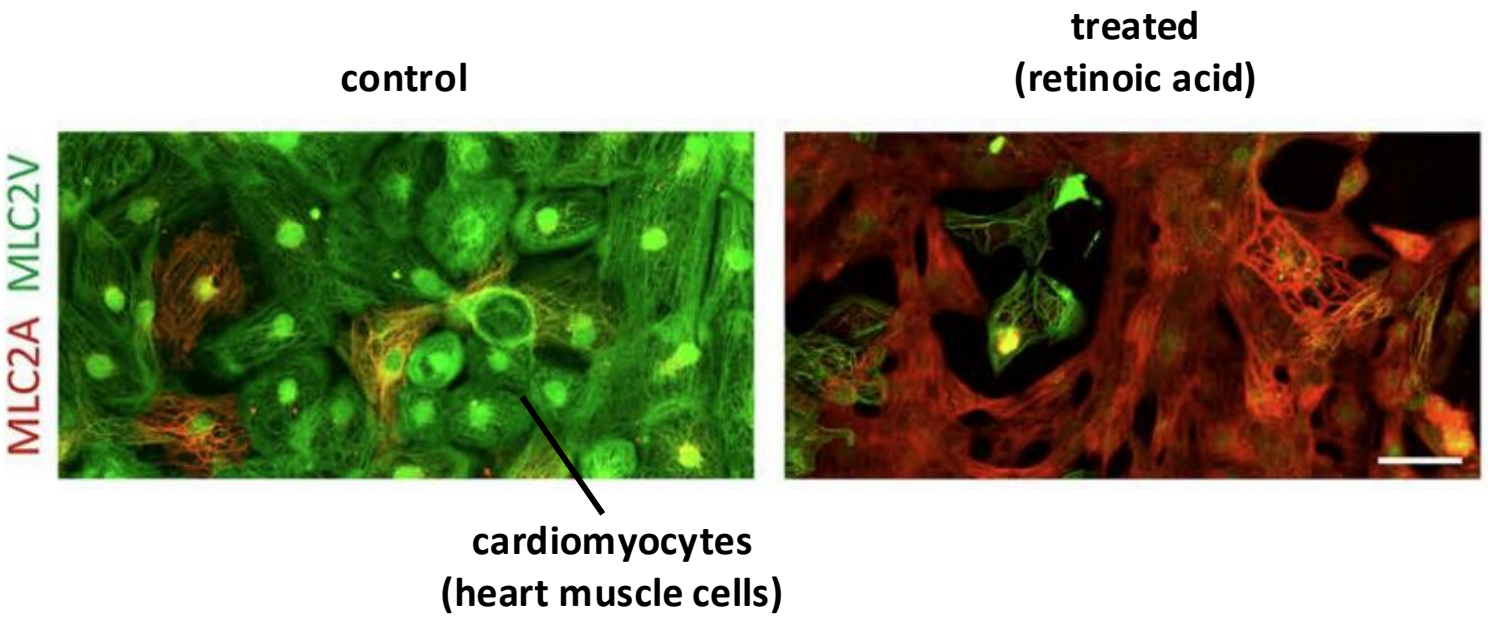
Response Predictor

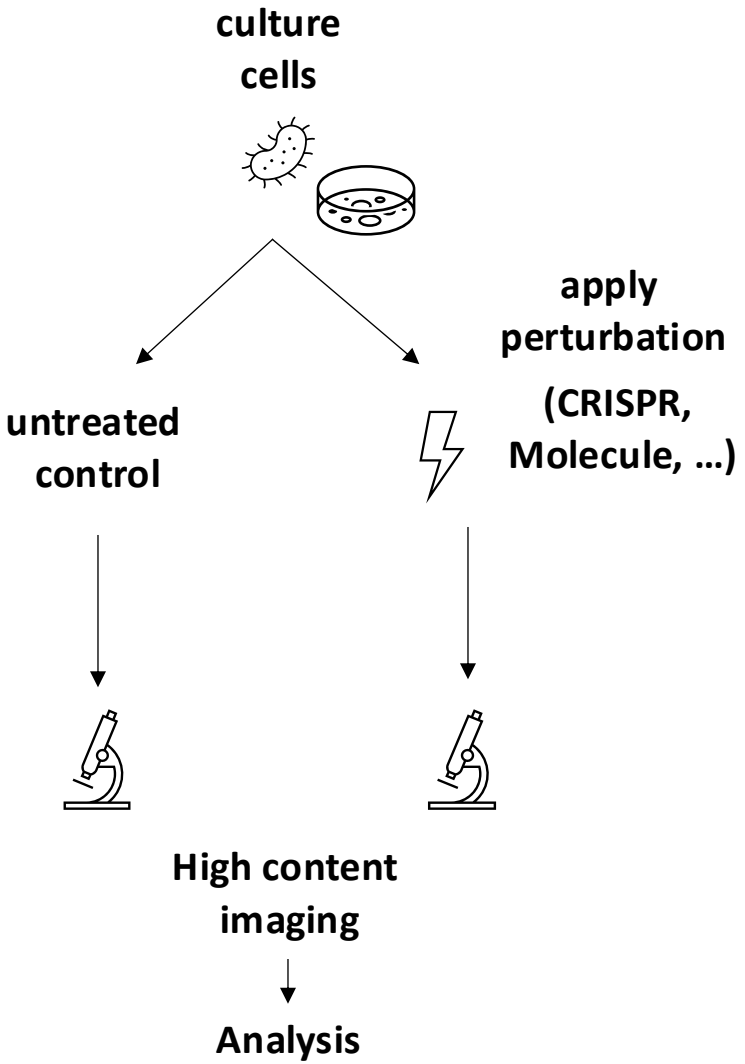


3 ***In-silico*** **treatment prediction**

Weis et al. Deep learning cluster analysis reveals subtypes in response to antisense oligonucleotide therapy in chronic hepatitis B. EASL 2023

Decoding the language of the cell with perturbation experiments

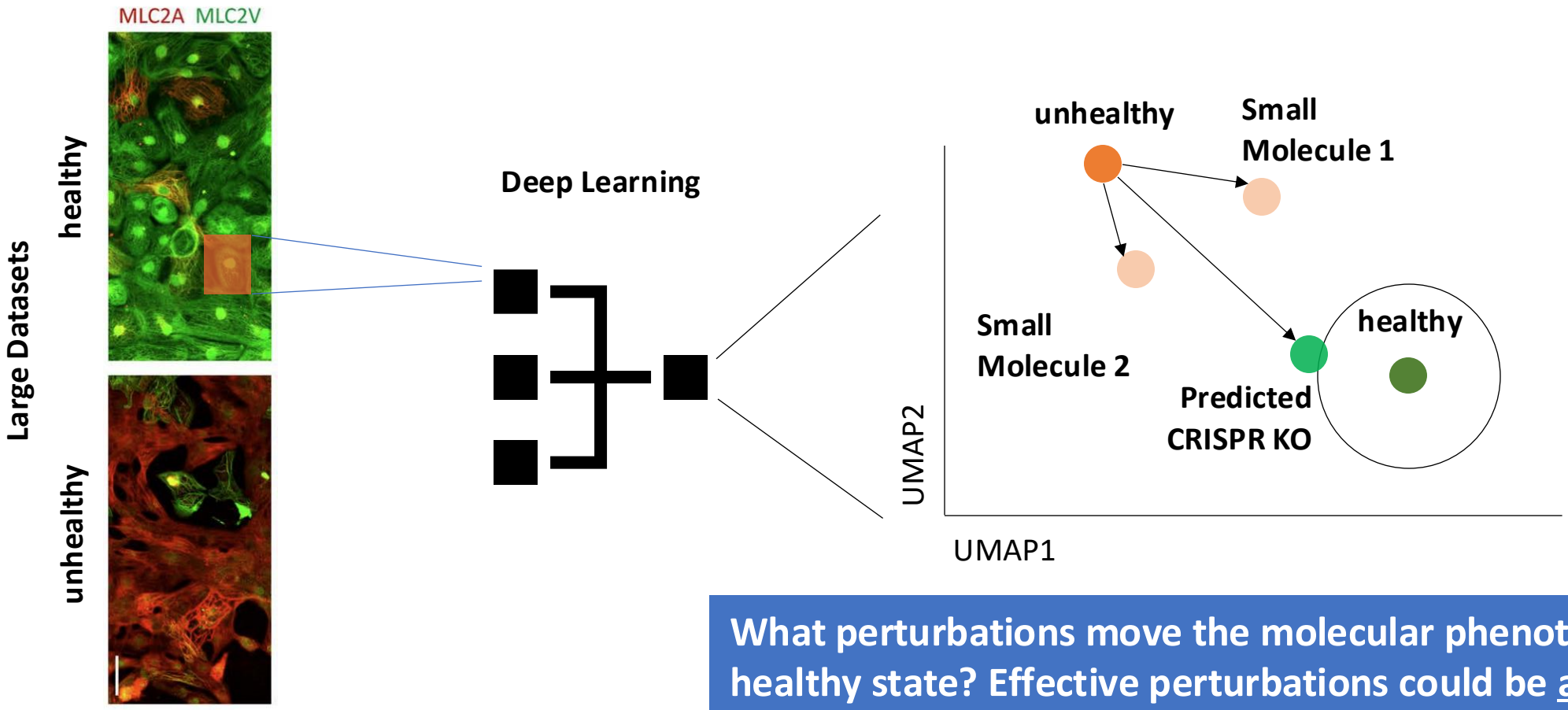




Comparison between experimental readouts in control and intervention settings enables us to understand causal effects of perturbations.

Cyganek, Lukas, et al. "Deep phenotyping of human induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes." JCI insight 3.12 (2018).

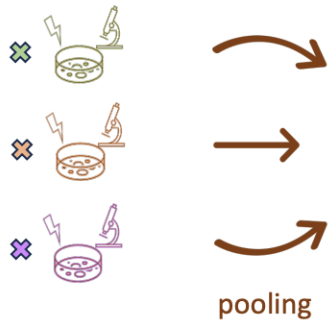
Decoding the language of the cell with perturbation experiments



What perturbations move the molecular phenotype towards the healthy state? Effective perturbations could be attractive treatments for further development.

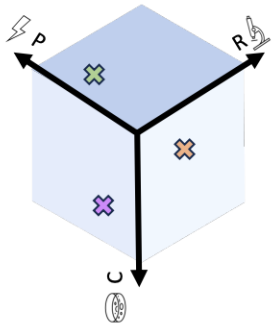
Cyganek, Lukas, et al. "Deep phenotyping of human induced pluripotent stem cell-derived atrial and ventricular cardiomyocytes." JCI insight 3.12 (2018).

Joint embedding of readout (R), context (C), and perturbation (P) achieves broad understanding of biology across environments



perturbation experiments

pooling



training

Large Perturbation Model (LPM)

Query:

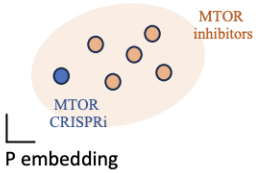
- **P:** STAT1 CRISPRi
- **R:** Gene expression of PSMA1
- **C:** LINCS protocol using A549 cells

Model outputs:

- Perturbation (P) embedding → Z_p
- Readout (R) embedding → Z_r
- Context (C) embedding → Z_c
- Predicted observation → Y

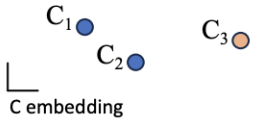
Tasks:

What would be the outcome of unperformed experiments?

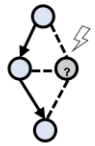


What molecular mechanisms is a compound acting on?

What molecular functions is a gene associated with?



Which contexts act similarly when perturbed?

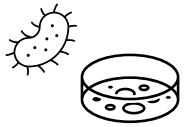


What gene-gene interactions exist for unseen perturbations?

Miladinovic et al. In-silico biological discovery with large perturbation models. (under review; 2024)

LPMs permit a query-answer interaction mode to specify and simulate context, readout and perturbation of interest

Cell culture + **CRISPRi perturbation** + **Measure readout** =



Query:

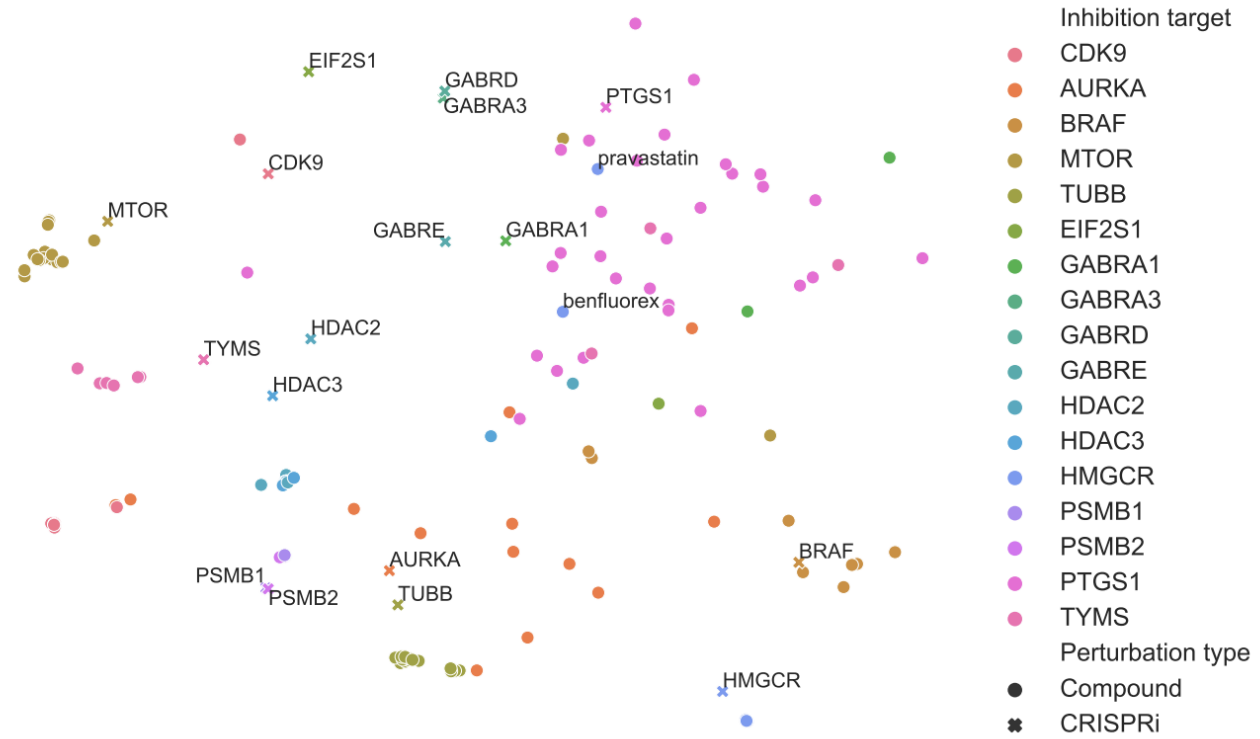
- **P:** STAT1 CRISPRi
- **R:** Gene expression of PSMA1
- **C:** LINCS protocol using A549 cells

Model outputs:

Perturbation (P) embedding	Z_p
Readout (R) embedding	Z_r
Context (C) embedding	Z_c
Predicted observation	Y

Miladinovic et al. In-silico biological discovery with large perturbation models. (under review; 2024)

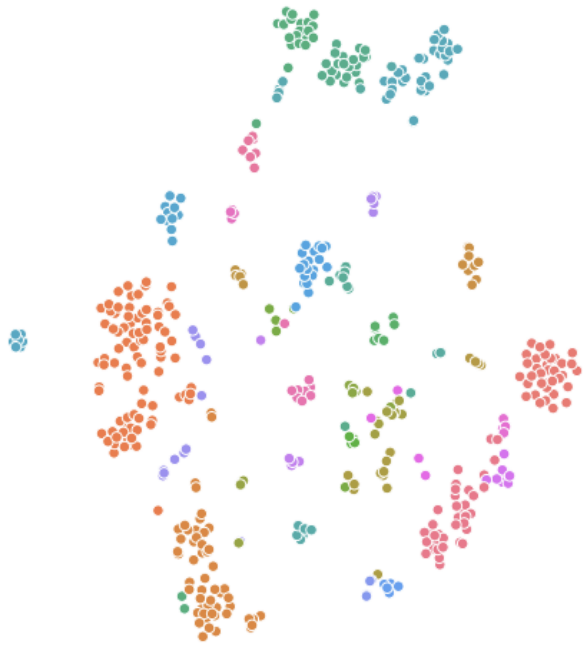
LPMs internally create a map of perturbations by their similarity



We find that LPMs internally organize perturbations by their mechanistic similarity, and this even across different types of perturbation (e.g. chemical versus genetic perturbation).

Miladinovic et al. In-silico biological discovery with large perturbation models. (under review; 2024)

LPMs internal map also reflects functional properties of perturbations



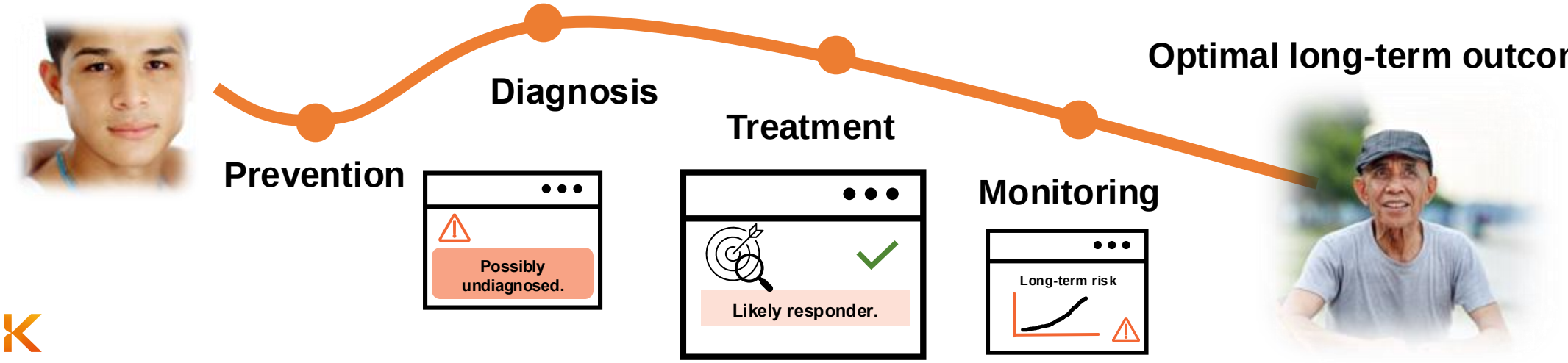
- mitochondrial protein translocation, tRNA synthesis, EIF2
- 39S ribosomal subunit, mitochondrial
- 40S ribosomal subunit, cytoplasmic
- 60S ribosomal subunit, cytoplasmic
- histone acetylation 1
- INO80 complex
- glycolysis
- translocon/protein glycosylation
- vesicular trafficking
- nuclear export
- ESCRT complex
- chromosome segregation
- SNARE complex
- NuA4 histone acetyltransferase complex and TTT complex
- Mediator complex
- TFIIH/nucleotide excision repair
- histone/nucleosome synthesis
- m6A mRNA methylation
- Chaperonin TCP-1
- Paf complex
- spliceosome
- COP9 signalosome
- exosome and mRNA turnover
- DNA replication
- EIF3 2
- EIF3 1/Ku antigen complex
- Pol I and rRNA biosynthesis
- Integrator complex 1
- nonsense-mediated decay
- mitochondrial transcriptional regulation
- mTOR signaling
- 28S ribosomal subunit, mitochondrial
- NELF complex
- proteasome
- TFIID complex

… and this also extends to classifying molecular functions associated with different perturbrations.

Miladinovic et al. In-silico biological discovery with large perturbation models. (under review; 2024)

Summary

- Real world care is challenging and leaves tremendous potential for **improving patient journeys using data-driven precision medicine**.
- Rich and diverse data from previously treated patients can help us **make informed and objective decisions** with the individual at the center.
- Machine learning empowers to **leverage these data for better decisions and discovery at scale.**



GSK

Thank you! Questions more than welcome.

- Weis et al. Deep learning cluster analysis reveals subtypes in response to antisense oligonucleotide therapy in chronic hepatitis B. EASL (2023)
- Mehrjou et al. GeneDisco: A Benchmark for Experimental Design in Drug Discovery. ICLR (2022)
- Träuble et al. Multi-megabase scale genome interpretation with genetic language models. In submission. (2024)
- Chevalley et al. CausalBench: A Large-scale Benchmark for Network Inference from Single-cell Perturbation Data. (2022) https://arxiv.org/abs/2210.17283

Patrick Schwab
patrick.x.schwab@gsk.com
@schwabpa

A Goyal, RN Ke & Bengio

Mila

the team

OATML (Yarin Gal)

B. Schölkopf

UNIVERSITY OF OXFORD

Max Planck Institute for Intelligent Systems

CIFAR

Stefan Bauer

UNIVERSITY OF CAMBRIDGE

CCAIM (Pietro Lio)