

11 February 2025 - Norway Life Science 2025

The coming era of predictive medicine

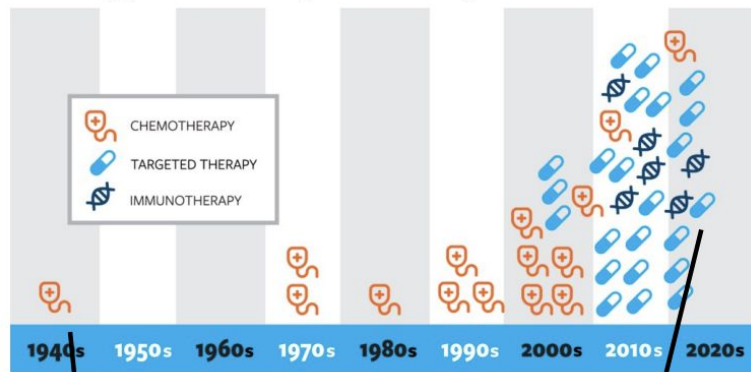
Patrick Schwab, Senior Director and Head of Biomedical AI, GSK

Co-authors: Frederik Träuble, Lachlan Stuart, Andreas Georgiou, Pascal Notin (Harvard), Arash Mehrjou, Ron Schwessinger, Mathieu Chevalley, Kim Branson, Bernhard Schölkopf (Max Planck Institute), Cornelia van Duijn (Oxford), Debora Marks (Harvard)

The good: drug development works for society



FDA approved therapies for lung cancer over time



Source: Lung Cancer Research Foundation

Mechlorethamine Hydrochloride

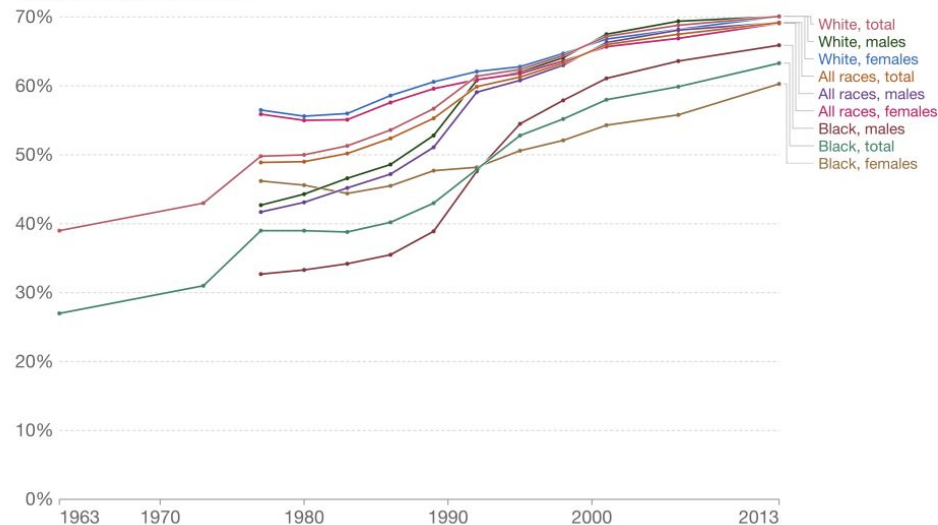
Candidate chemical warfare agent research

Mobocertinib

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

Five-year cancer survival rates by sex and race, 1963 to 2013

Percentage of cancer patients surviving at least five years following diagnosis of any cancer type. This is shown by sex and race in the United States.

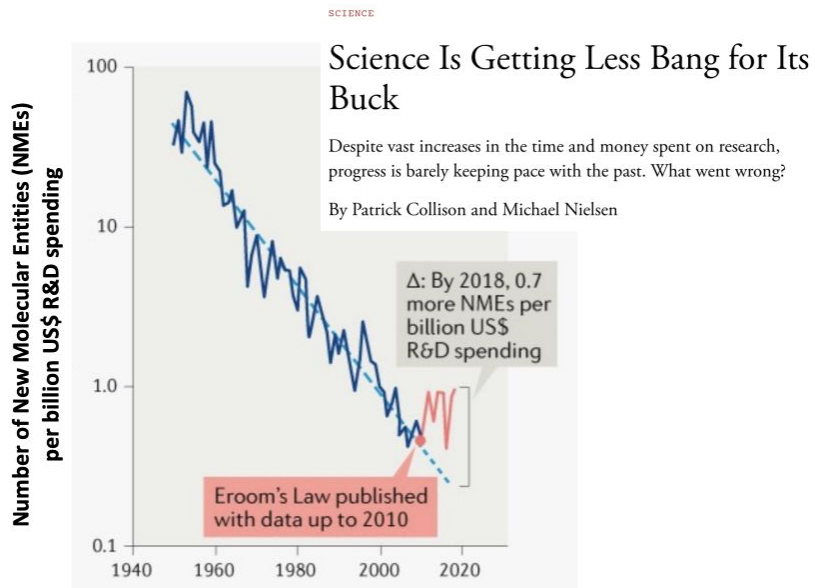


Source: National Cancer Institute

CC BY

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.**

What do we know works?

1

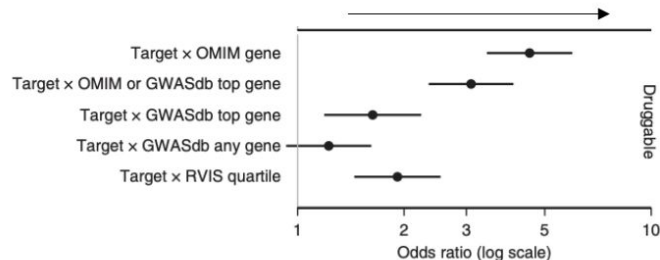
**Population-scale Biobanks –
Variation in Genetic Background
(>100'000 people)**



biobank^{uk}

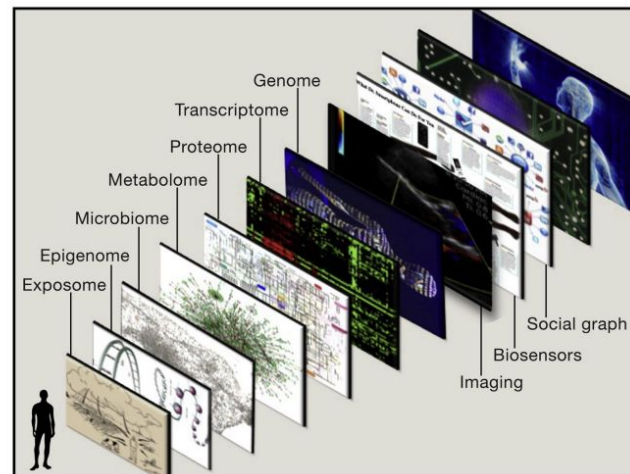


Higher probability of approval (EU/US)



2

**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.

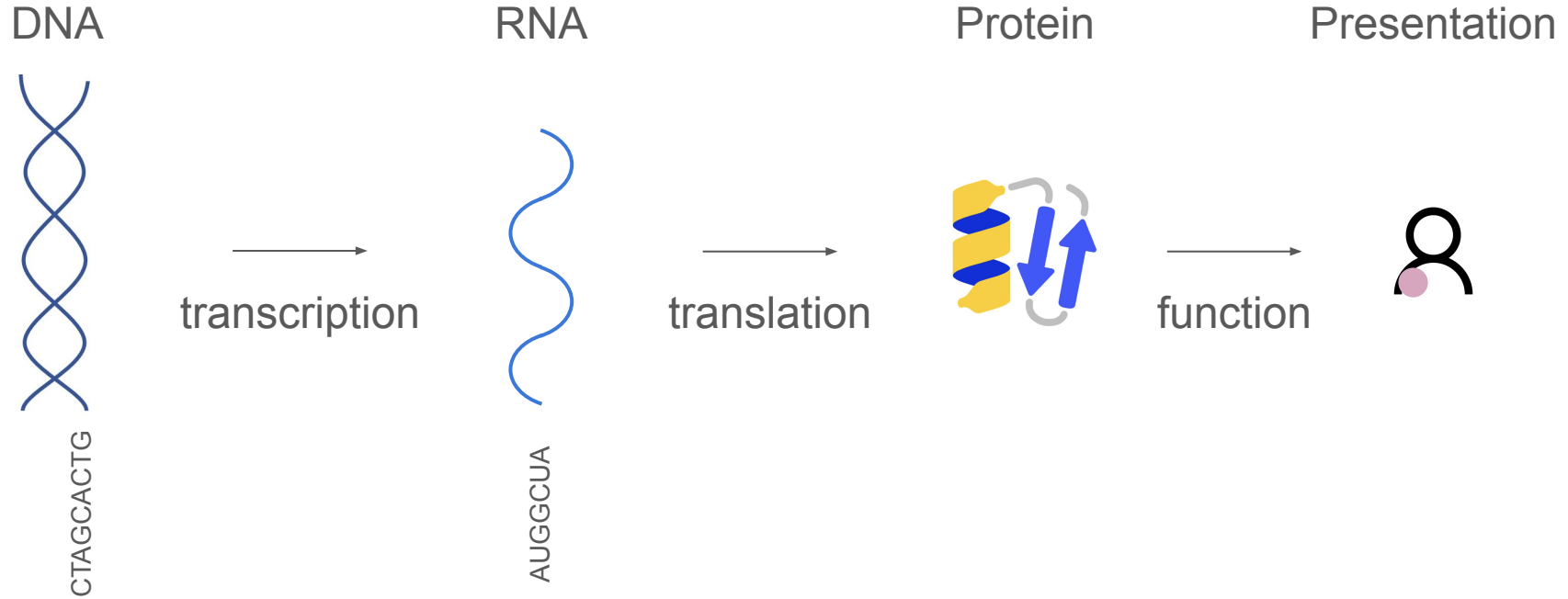


***What if* we could predict whole-organism mechanisms based on individual biological backgrounds?**

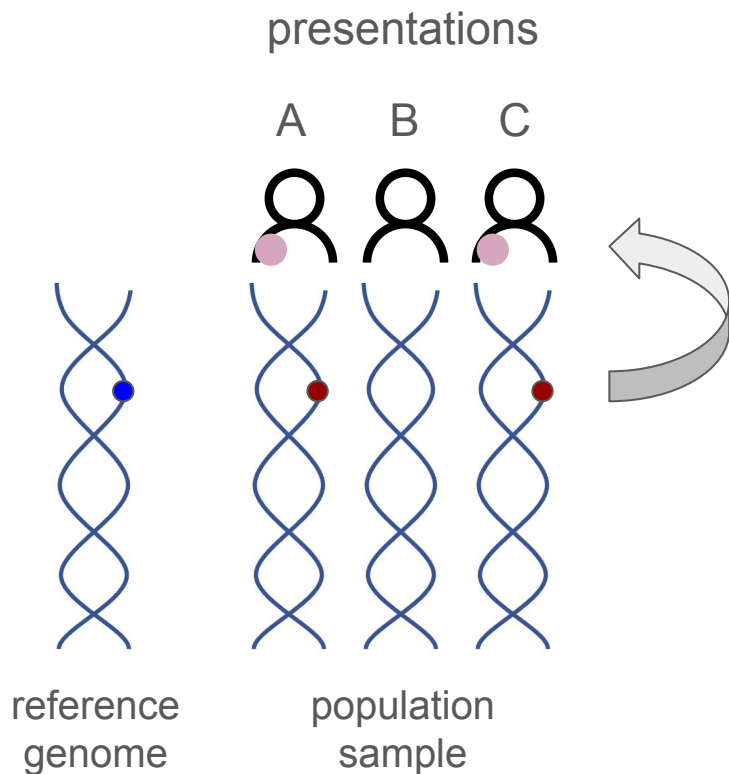
Towards a generative model of biomedical mechanisms



The central dogma - (simplified) information flow in biology



Natural experiments offer unique view on mechanisms



Single Nucleotide Polymorphism (SNP)
Natural variation in a single location connects presentation to a potential source

... CTAGCACTGCTAGCACTGCTAGCACTG  CTAGCACTGCTAGCACTGCTAGCACTG ...

... however, the human genome spans 3 billion base pairs ...

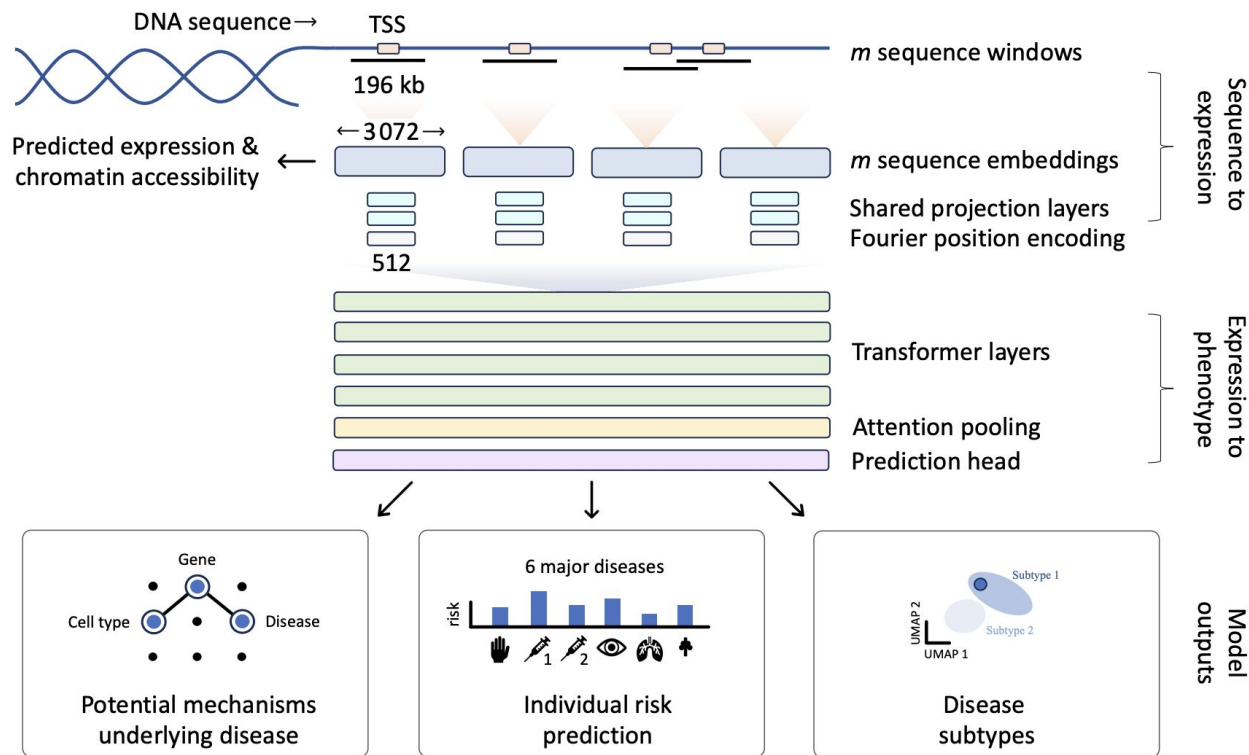
If your genome were a library, it would encompass



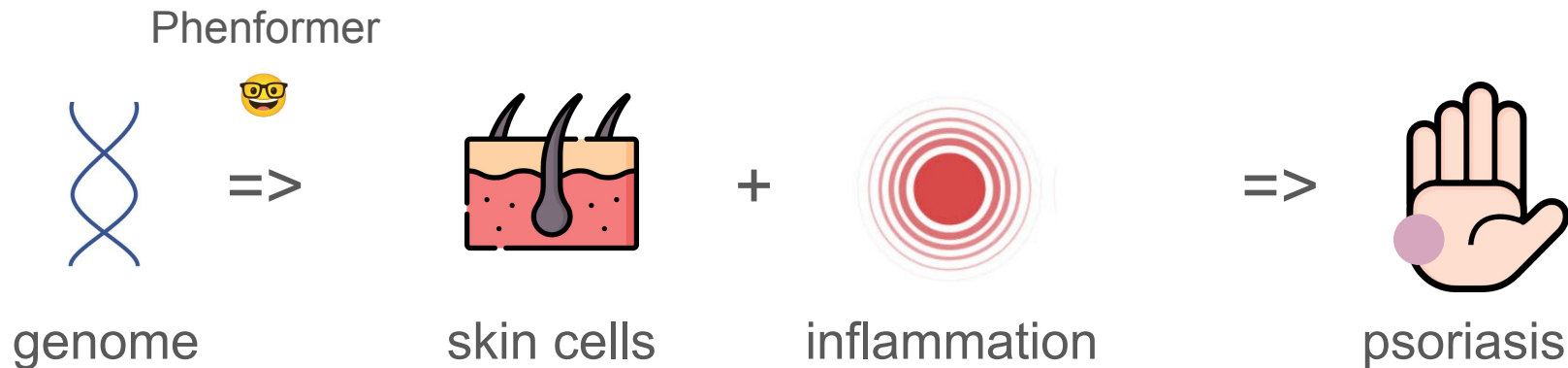
**6'000 books with 250 pages
each.**

a single letter misses the whole (*your!*) story

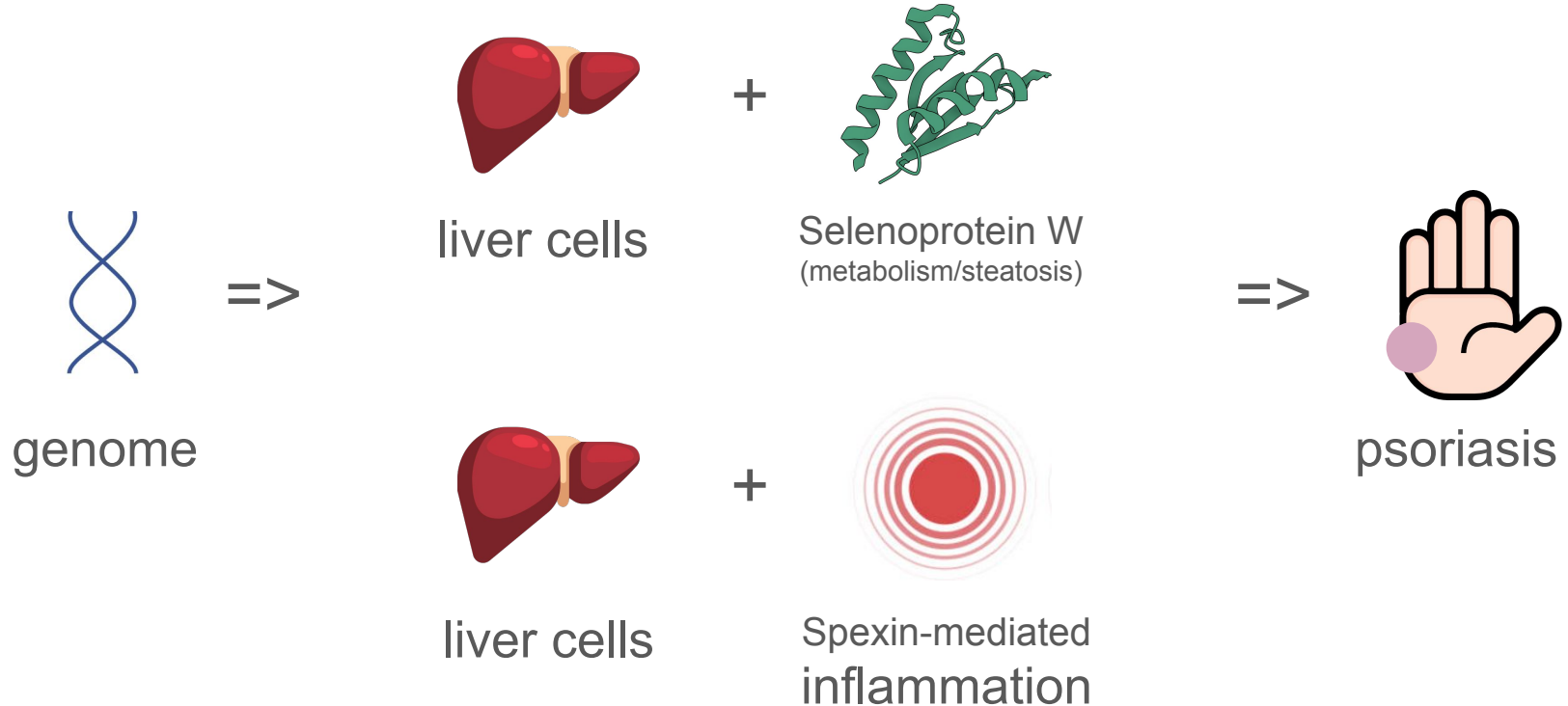
Phenformer - a multi-scale genetic language model to read whole genomes



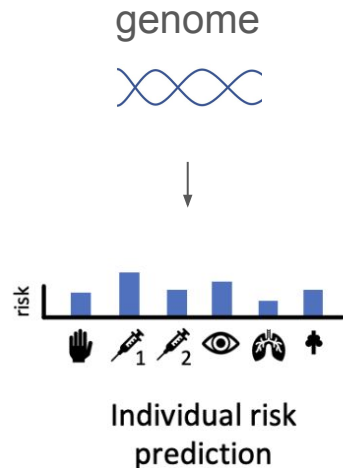
Phenformer reads genome to explain how disease comes to be



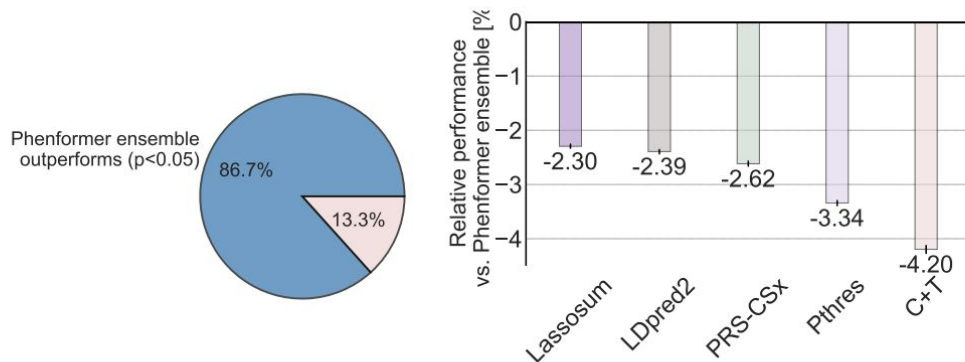
Incredibly, Phenformer explains presentations we do not fully understand
Clinically, psoriasis patients face 3x risk of fatty liver disease. Not known why.



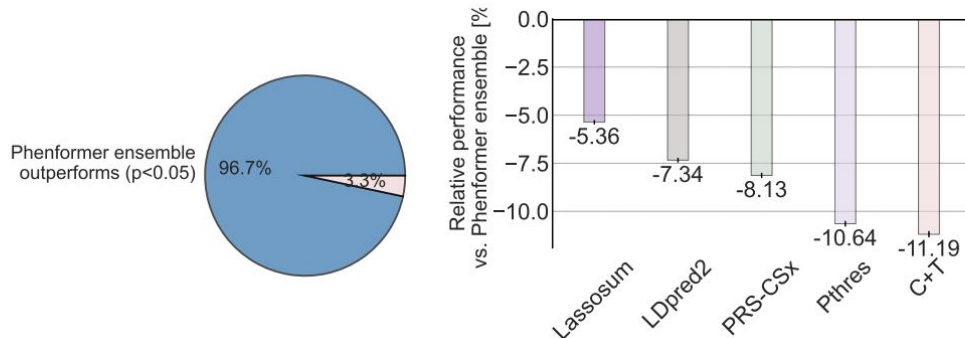
Using Phenformer as a personalised, whole-genome risk predictor considerably enhances predictive accuracy & generalisability



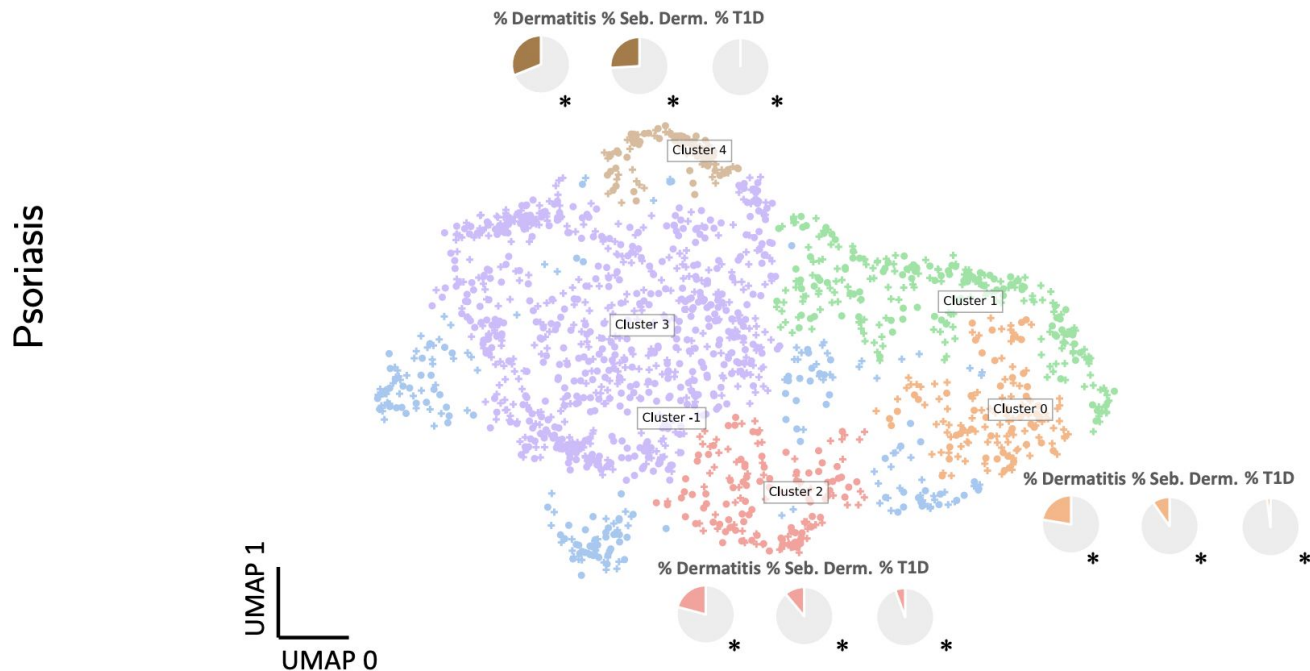
a) Predictive performance (whole genome, mixed ancestry)



b) Predictive performance (whole genome, non-European ancestry)



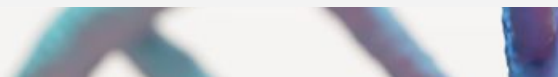
When embedding whole genomes using Phenformer, we obtain a map of mechanistic subtypes with putatively* different underlying mechanisms



* as evidenced by significant ($p<0.05$) differences in comorbidity rates

Summary



- Emerging class of *generative models of mechanisms* hold potential to enable understanding of *how* diseases come to be
 - Such models are capable of interpreting the vast biological background of single individuals
 - Using large compute (18 GPU months), Phenformer processes 3 billion base pairs and predicts for 6'000+ tissues and cell types
 - Individualised understanding of *what* and *how* biological change happens could enable *hyper-targeted therapy* in the future
- 

The coming era of predictive medicine

Frederik Träuble, Lachlan Stuart, Andreas Georgiou, Pascal Notin (Harvard), Arash Mehrjou, Ron Schwessinger, Mathieu Chevalley, Kim Branson, Bernhard Schölkopf (Max Planck Institute), Cornelia van Duijn (Oxford), Debora Marks (Harvard)

“Multi-megabase scale genome interpretation with genetic language models.”
available at: <https://arxiv.org/abs/2501.07737>



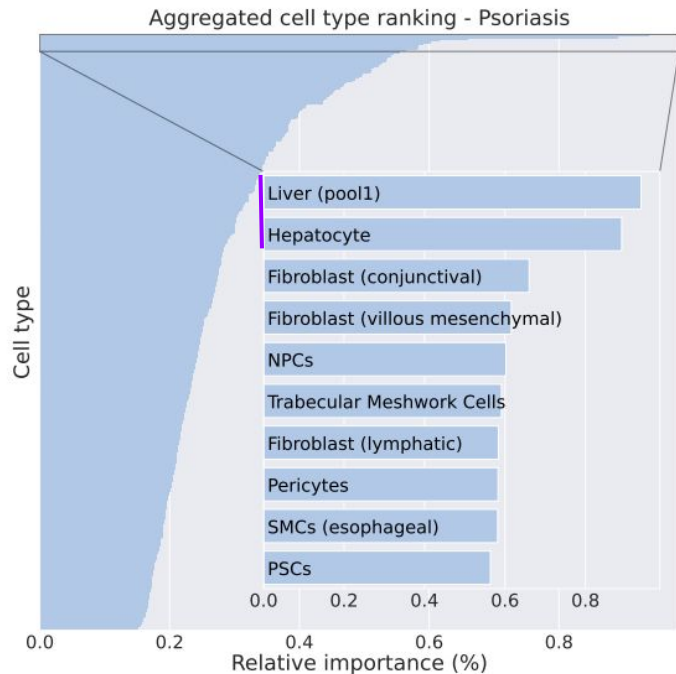
Patrick Schwab
patrick.x.schwab@gsk.com
@schwabpa

We are hiring: software engineers,
AI/ML scientists, internships, ...

the team



Following the <sequence-cell-expression-phenotype> paths enables generation of potential disease-associated mechanisms



- We find many **expected results** - e.g. fibroblasts/psoriasis
- ... but, strikingly, Phenformer also **substantiates many yet-unexplained clinical findings**
 - e.g. **liver-involvement in psoriasis** which to-date lacks mechanistic understanding (psoriasis pts are 2x higher risk for liver disease)
 - ... & many more surprising findings (optic nerve in COPD, appendix/nails in T1D)!

On same test data, Phenformer significantly outperforms existing PRS methods for risk prediction, while maintaining transportability

