



Proteomic prediction of common and rare diseases

Julia Carrasco Zanini Sanchez, PhD

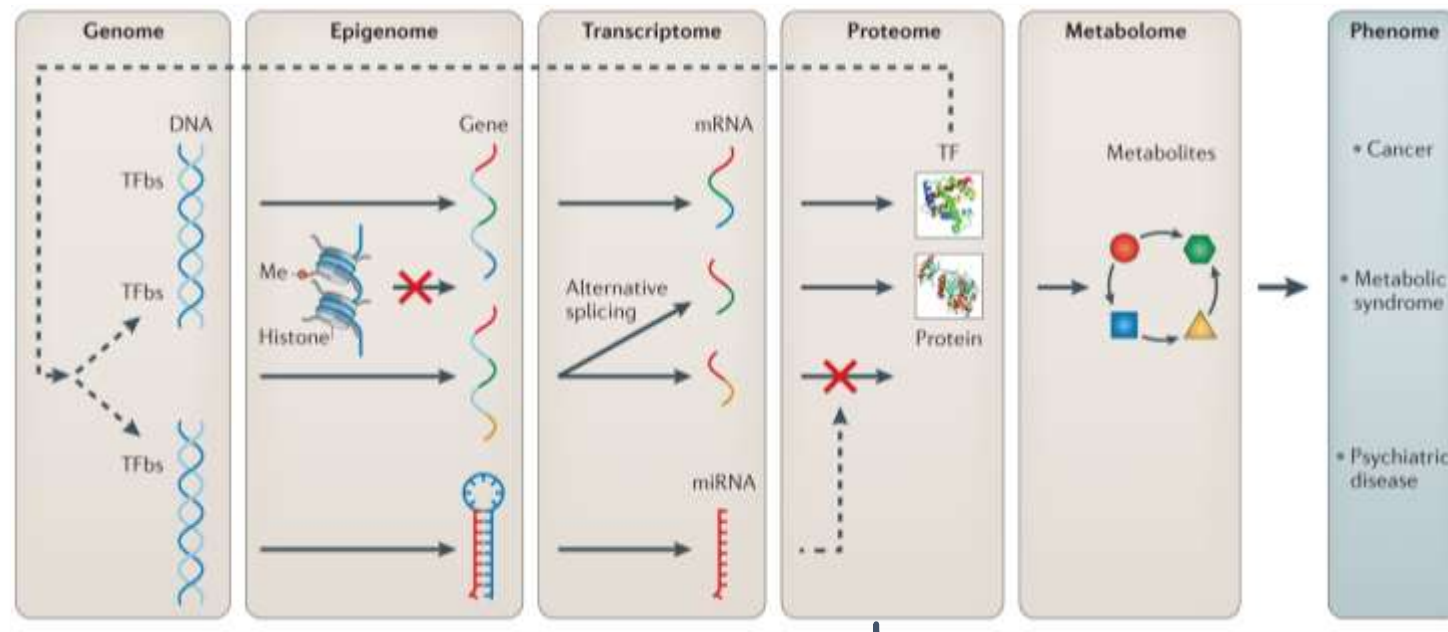
GSK, Genomic Sciences | GSK Research and Development | Stevenage, UK

MRC Epidemiology Unit | University of Cambridge

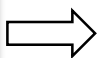
PDRA in Computational Genomics and Multi-omics | Precision Health University Research
Institute | QMUL

j.carrasco-zanini-sanchez@qmul.ac.uk

Studying the plasma proteome at scale



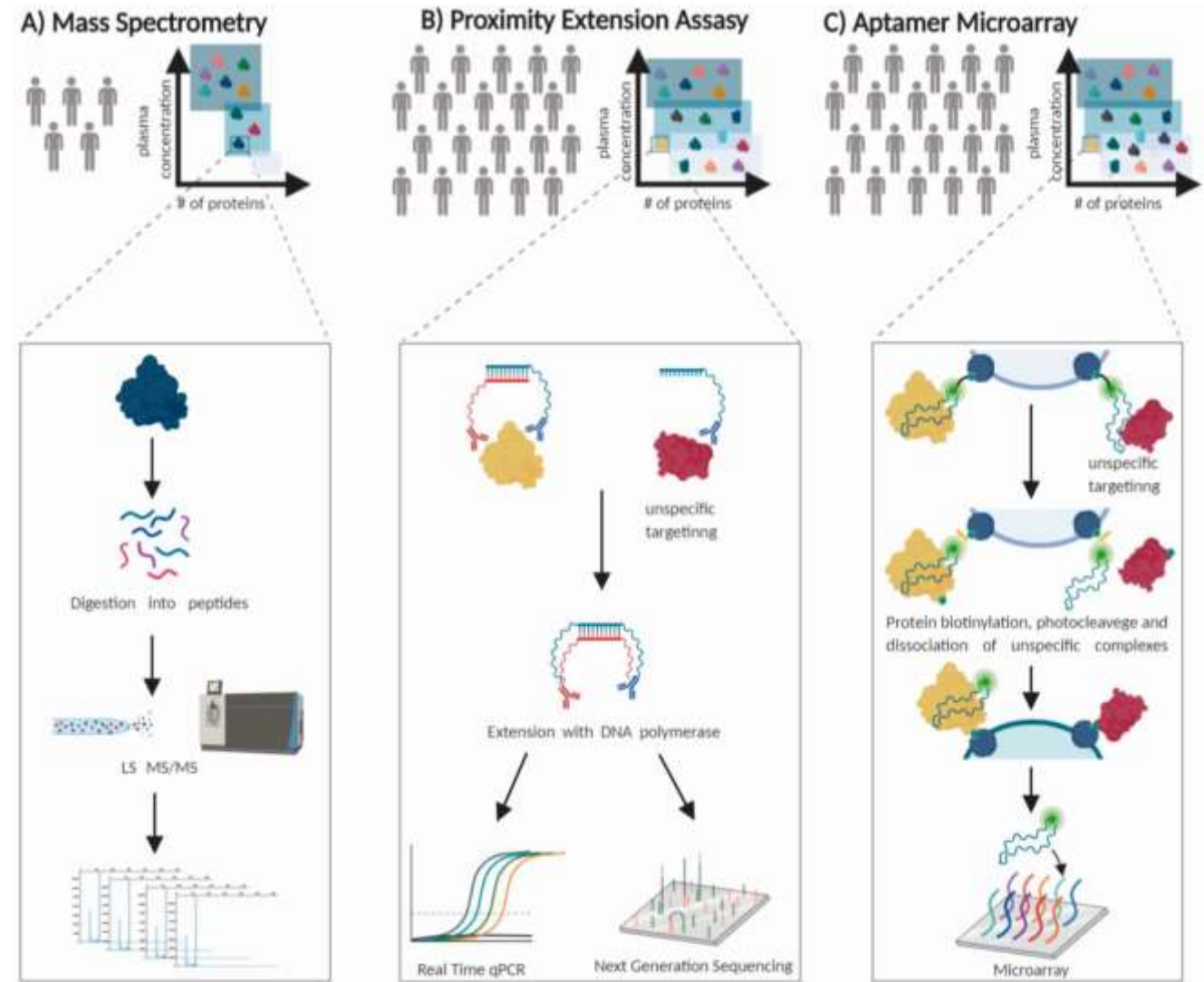
- Central layer of information transfer and are the main effector molecules on cellular function.
- Capture genetic predisposition, lifestyle and environmental factors.
- Largest class of pharmaceutical drug targets, FDA-approved laboratory tests and biomarkers.



Novel biomarkers
Aetiological mechanisms

Techniques to measure the plasma proteome

- **Mass spectrometry** of protein fragments (peptides)
- **Antibody-based** (similar to an ELISA used in clinical chemistry)
- Short oligonucleotides – **aptamers** – which match the 3D-conformation of the target protein

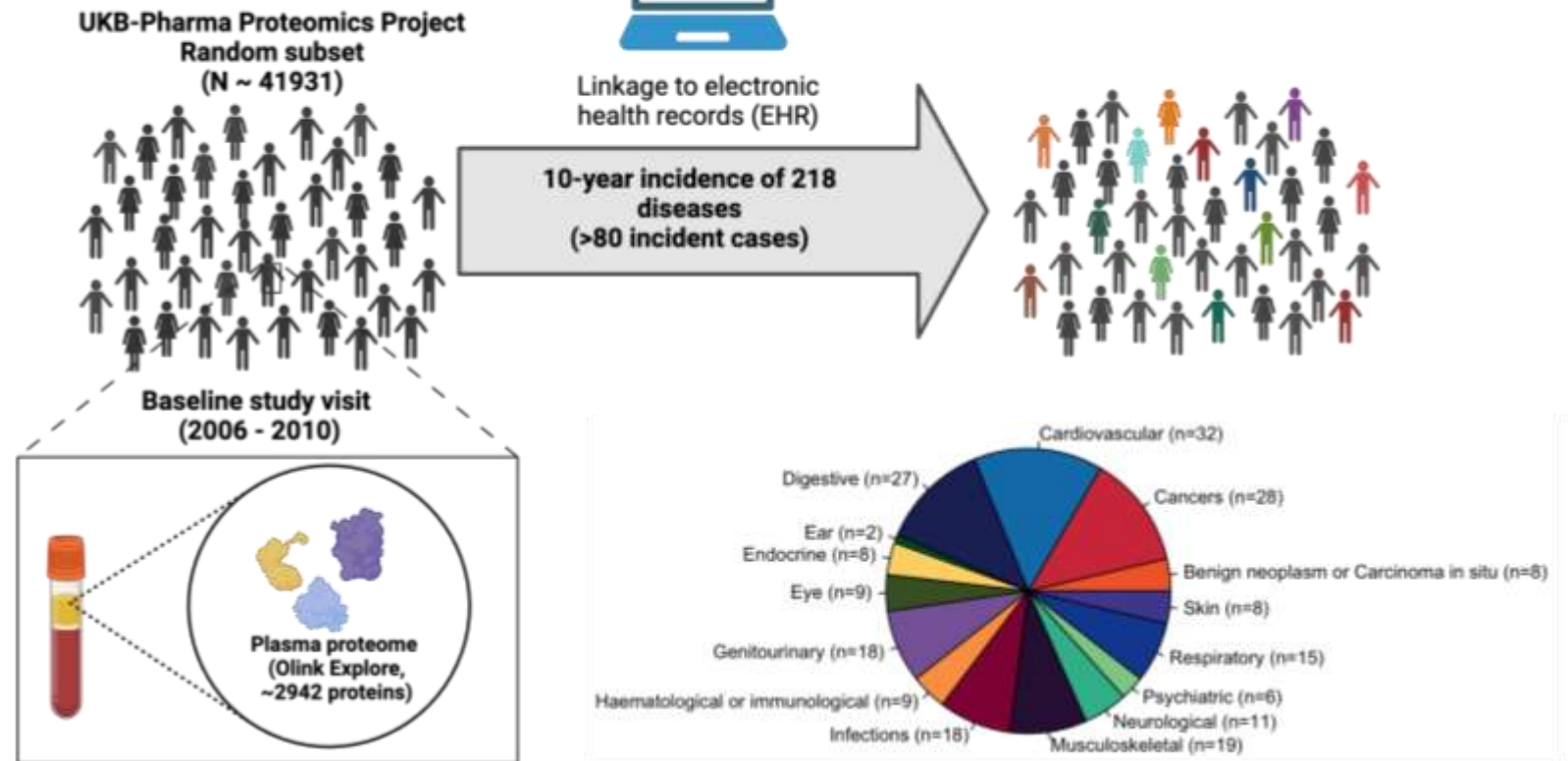
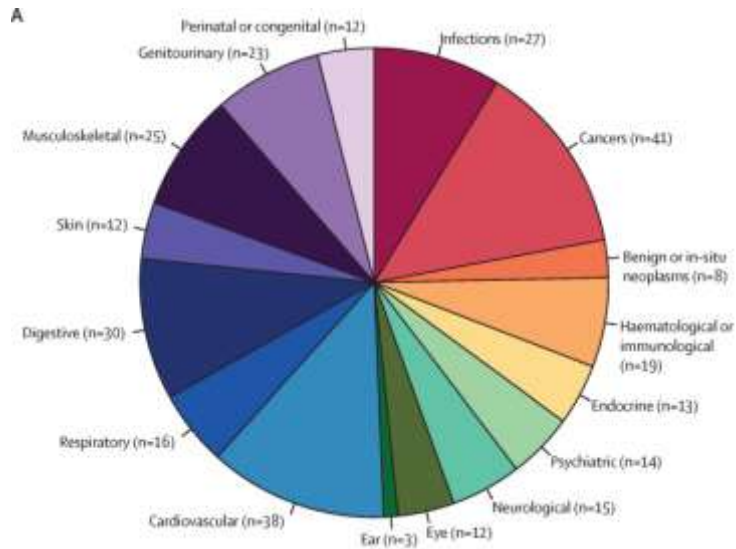


Can we leverage broad-capture plasma proteomics to identify people at high-risk of developing diseases in the future?

UK biobank – integrating EHR with plasma proteomics

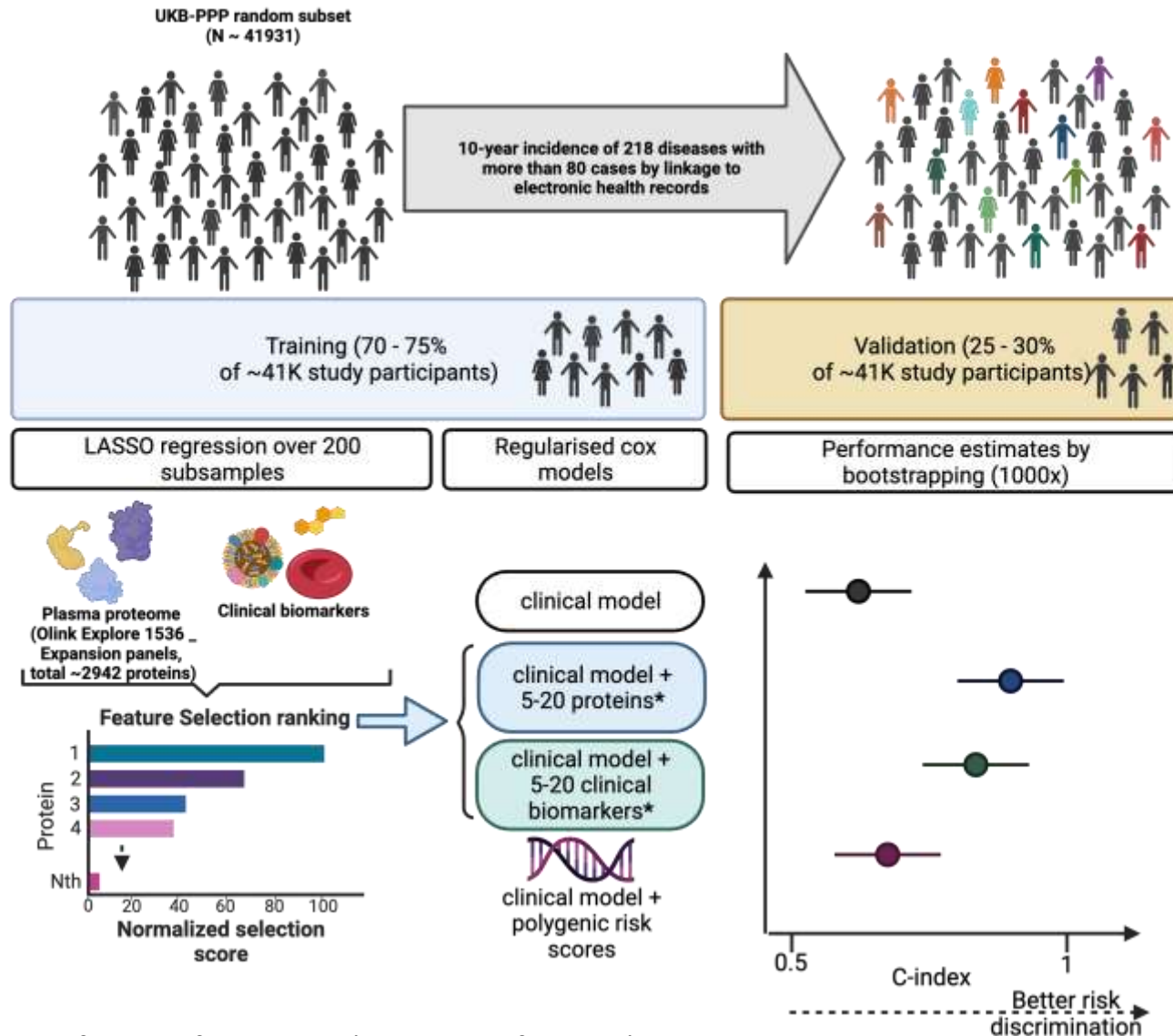
Anna Torralbo
Spiros Denaxas
Cai Ytsma
Natalie Fitzpatrick
Harry Hemingway

Jonathan Davitte
Damien Croteau-Chonka
Meg Ehm



Kuan V. et al. 2022. Lancet Digital Health

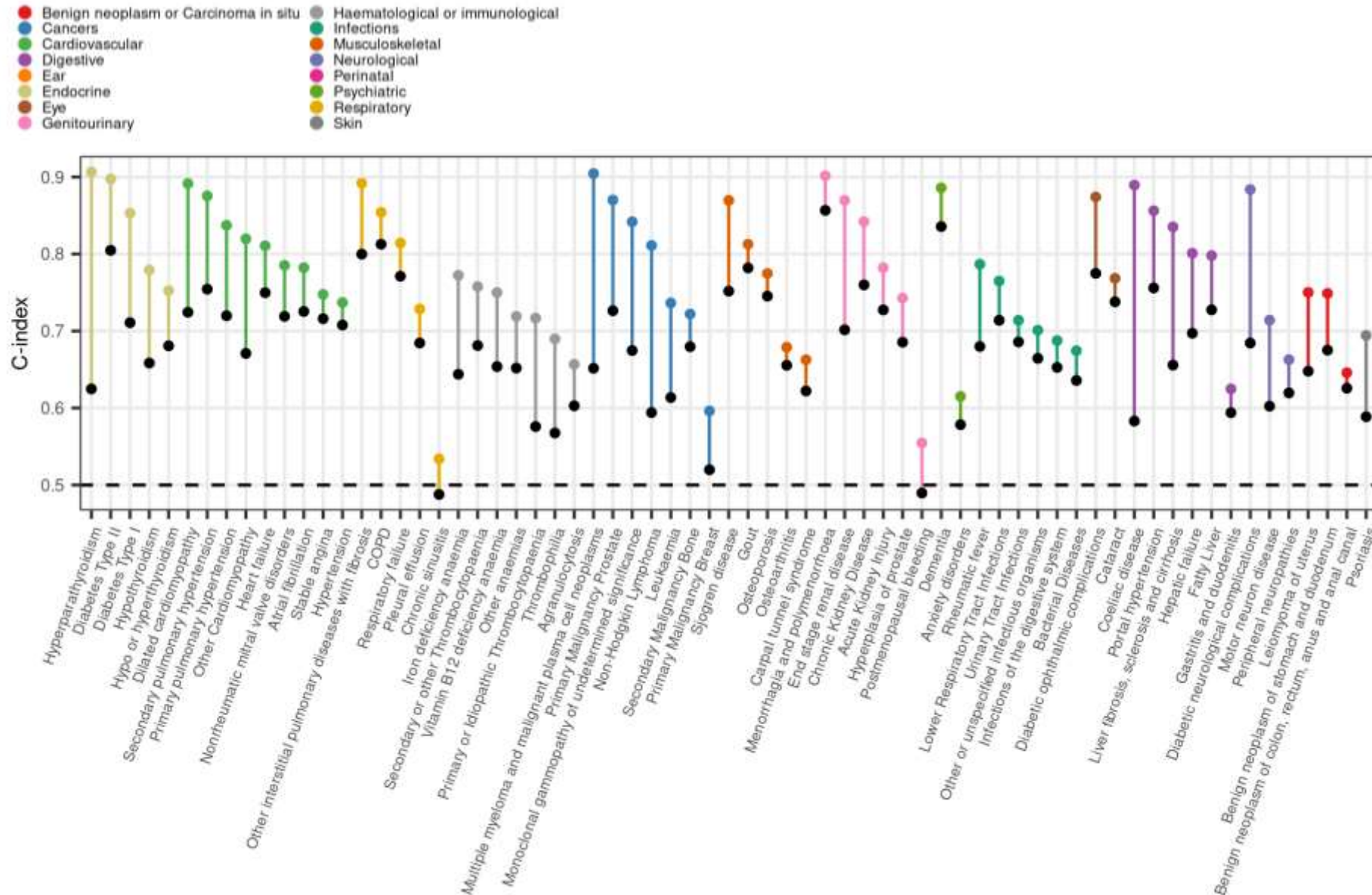
Study design



- Exclusion of prevalent cases and incident cases within the first 6 months.
- Clinical model included age, sex, BMI, ethnicity, smoking status, alcohol consumption, family history.
- **2942 proteins** (Olink Explore 1536 + Expansion panels).
- **37 clinical biomarkers** (standard laboratory assays and blood cell traits)
- **Polygenic risk scores** from Genomics PLC

* Best performing feature set (5, 10 or 20 features)

Proteins improve predictive performance over and above basic clinical models for 67 disease



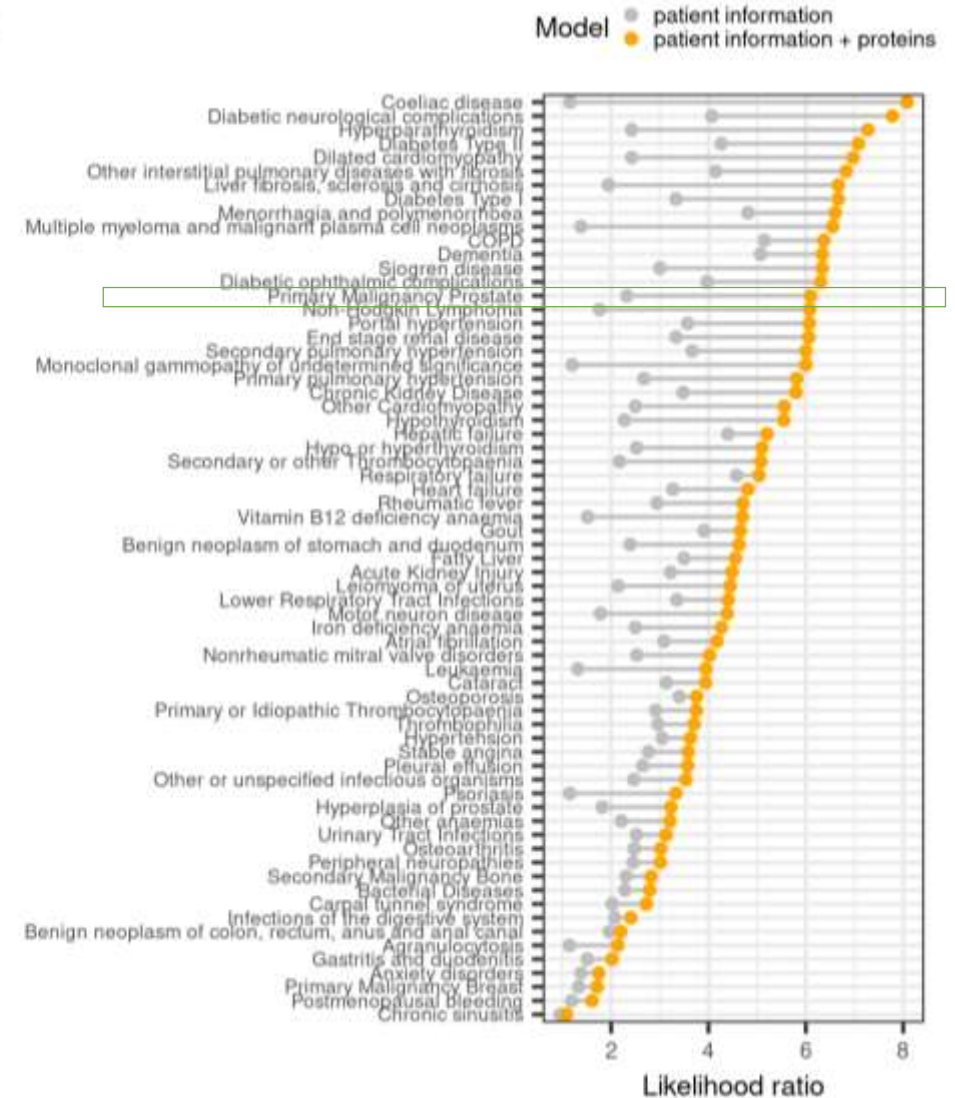
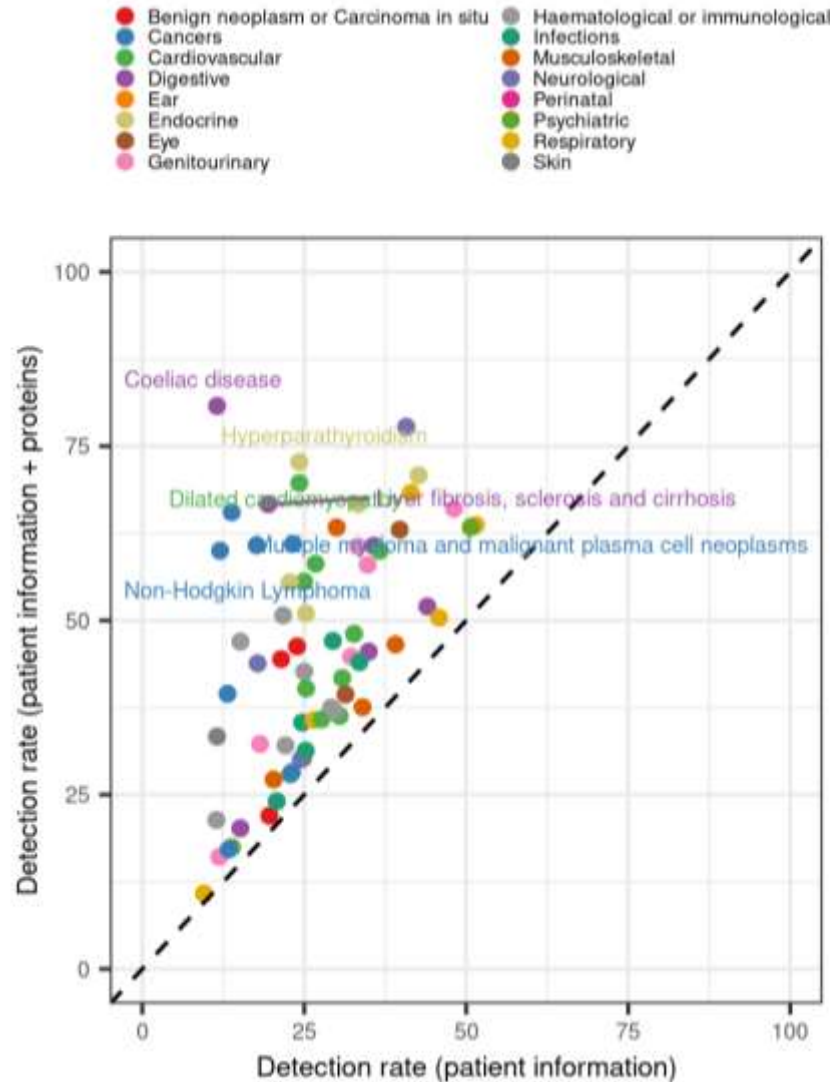
- Improvements in C-index ranging from 0.02 – 0.31.

Black dots : Basic clinical model

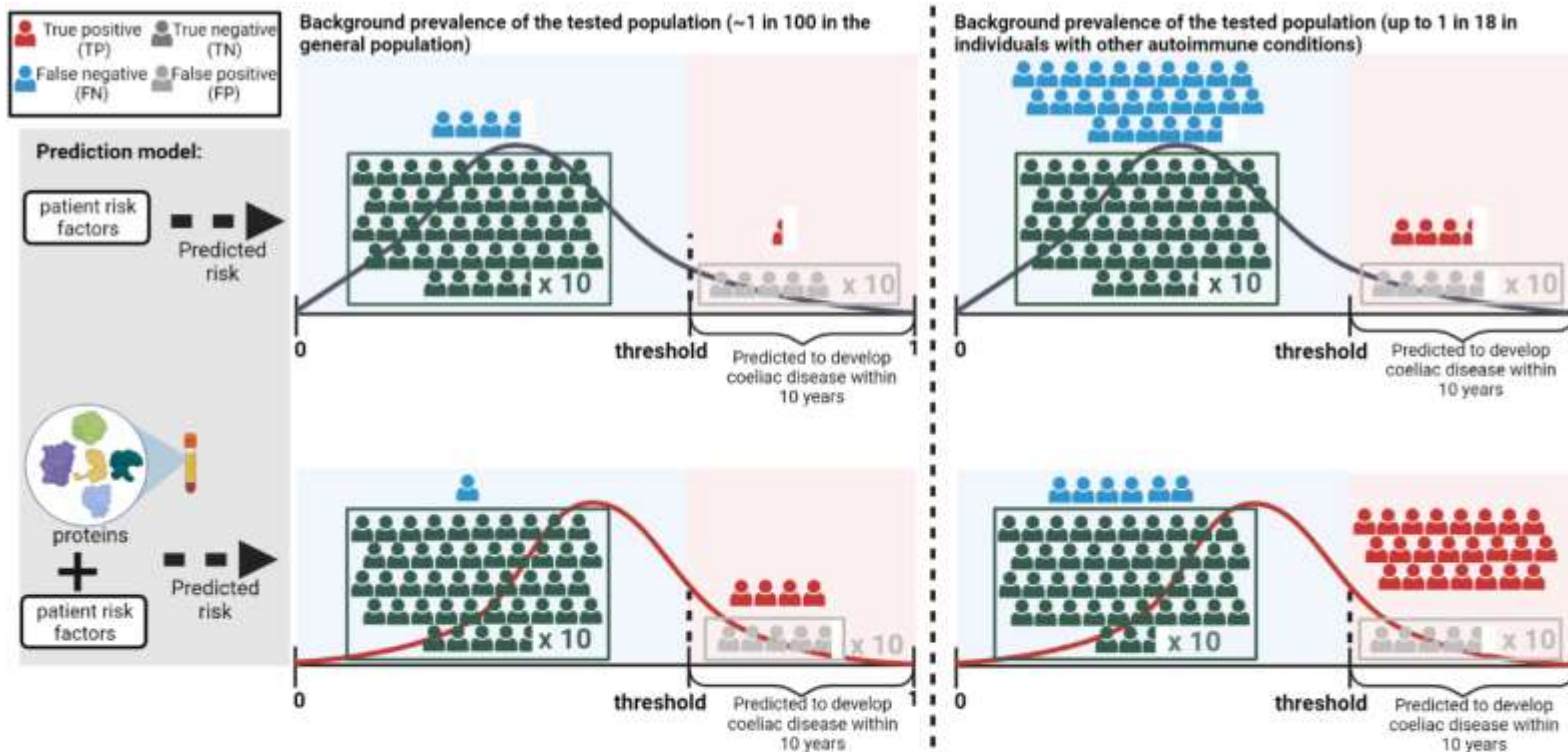
Colored dots : Basic clinical model + 5 – 20 proteins

Improvement in detection rates and likelihood ratios

- Performance metrics relevant for screening.
- Detection rate at a 10% False positive rate (FPR).
- LR: Likelihood of seeing a high “proteomic risk” in an individual that will develop the disease within 10 years compared to an individual who won’t.



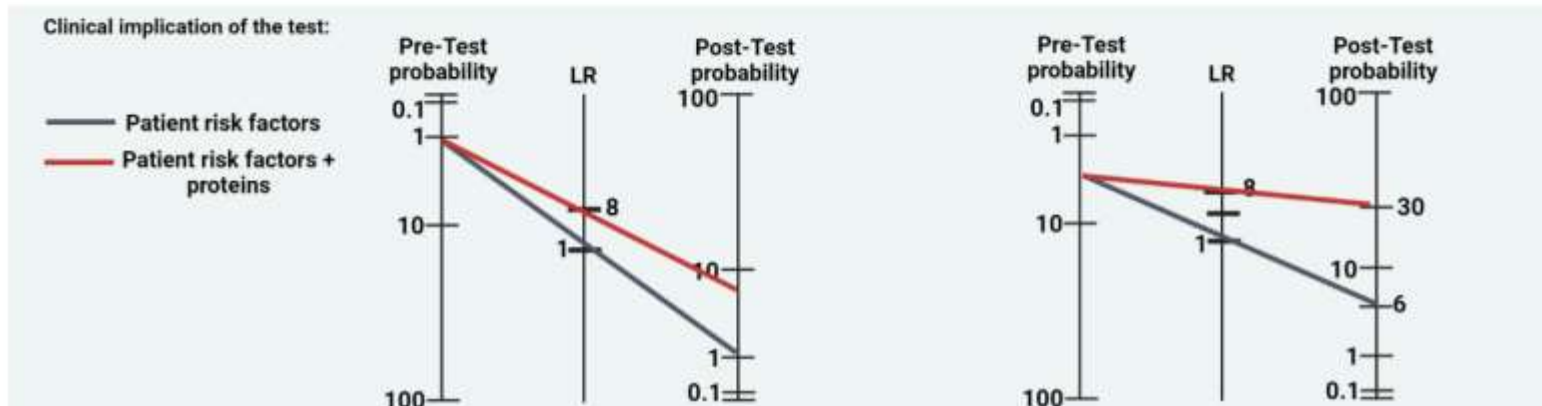
An example of the theoretical benefit of proteomic screening in coeliac disease



$$\text{Detection rate (DR)} = \frac{TP}{FN + TP}$$

$$\text{False positive rate} = \frac{FP}{FP + TN}$$

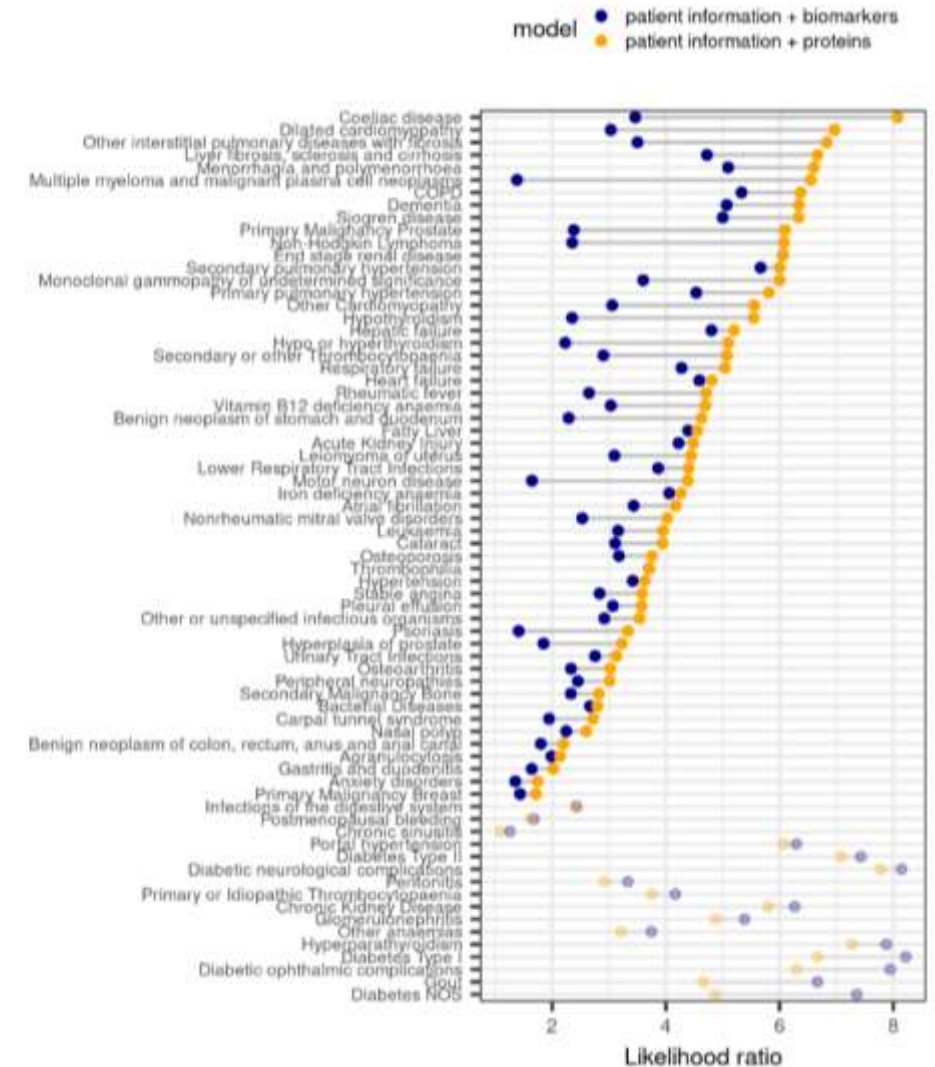
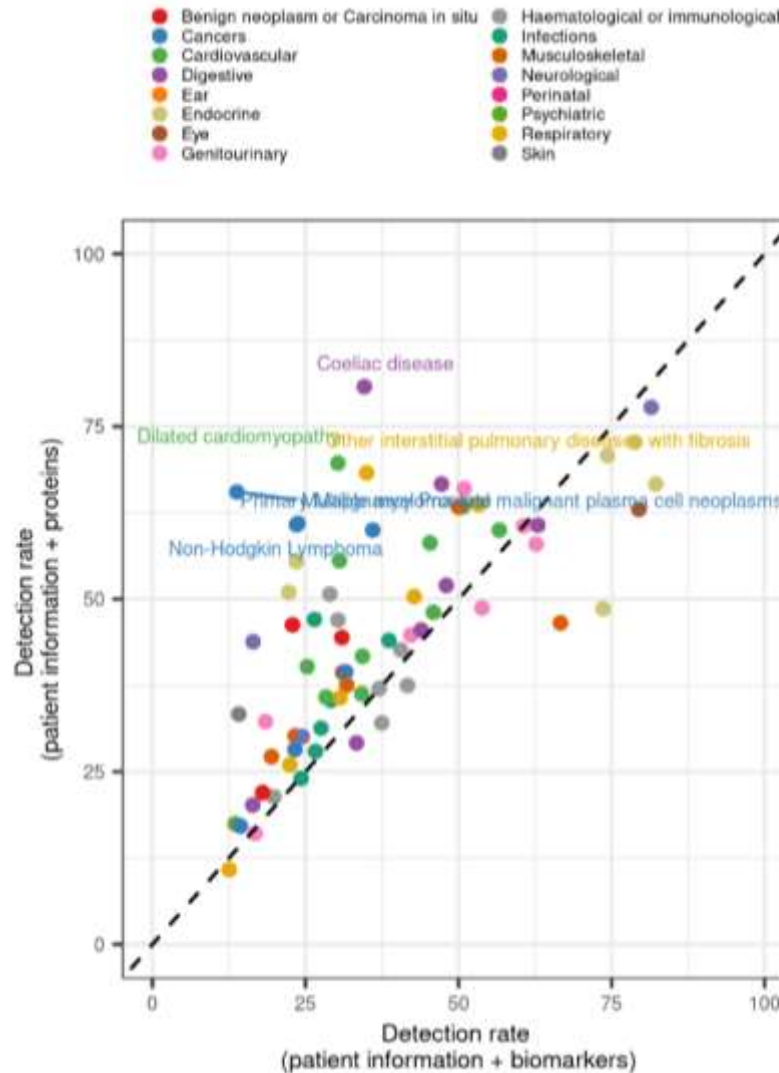
$$\text{Likelihood ratio (LR)} = \frac{DR}{FPR}$$



$$\text{Posttest probability} = \frac{\text{Pretest odds} * LR}{(\text{Pre - test odds} * LR) + 1}$$

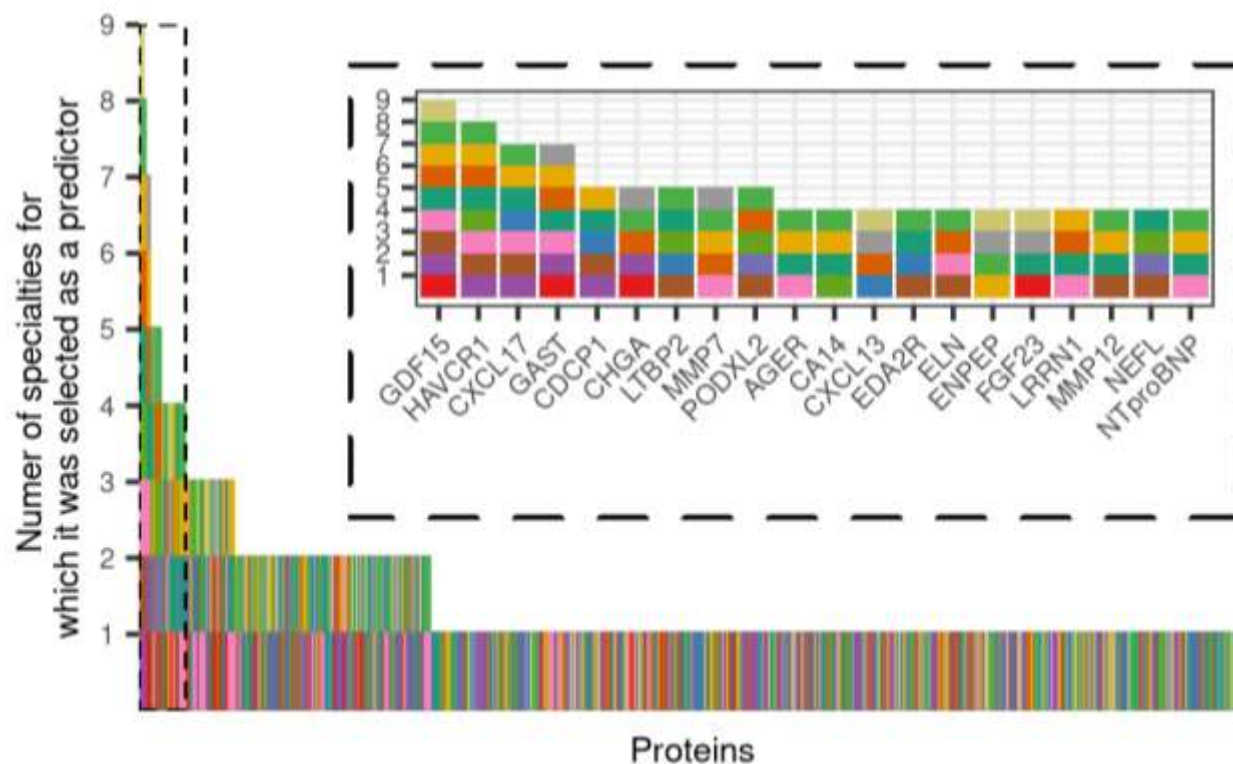
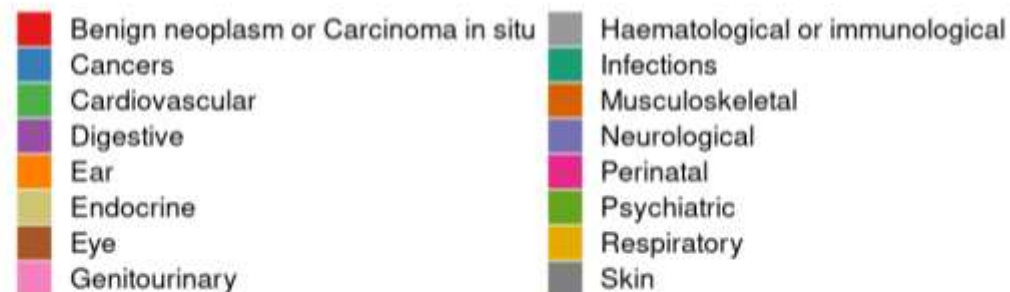
Head-to-head comparison: protein signatures vs clinical biomarkers

- Clinical biomarkers improved prediction for 28 diseases.
 - Of these proteins improved prediction for 24.
- Of the 67 diseases improved by proteins, 52 of those had greater improvements from proteomics compared to clinical biomarkers.

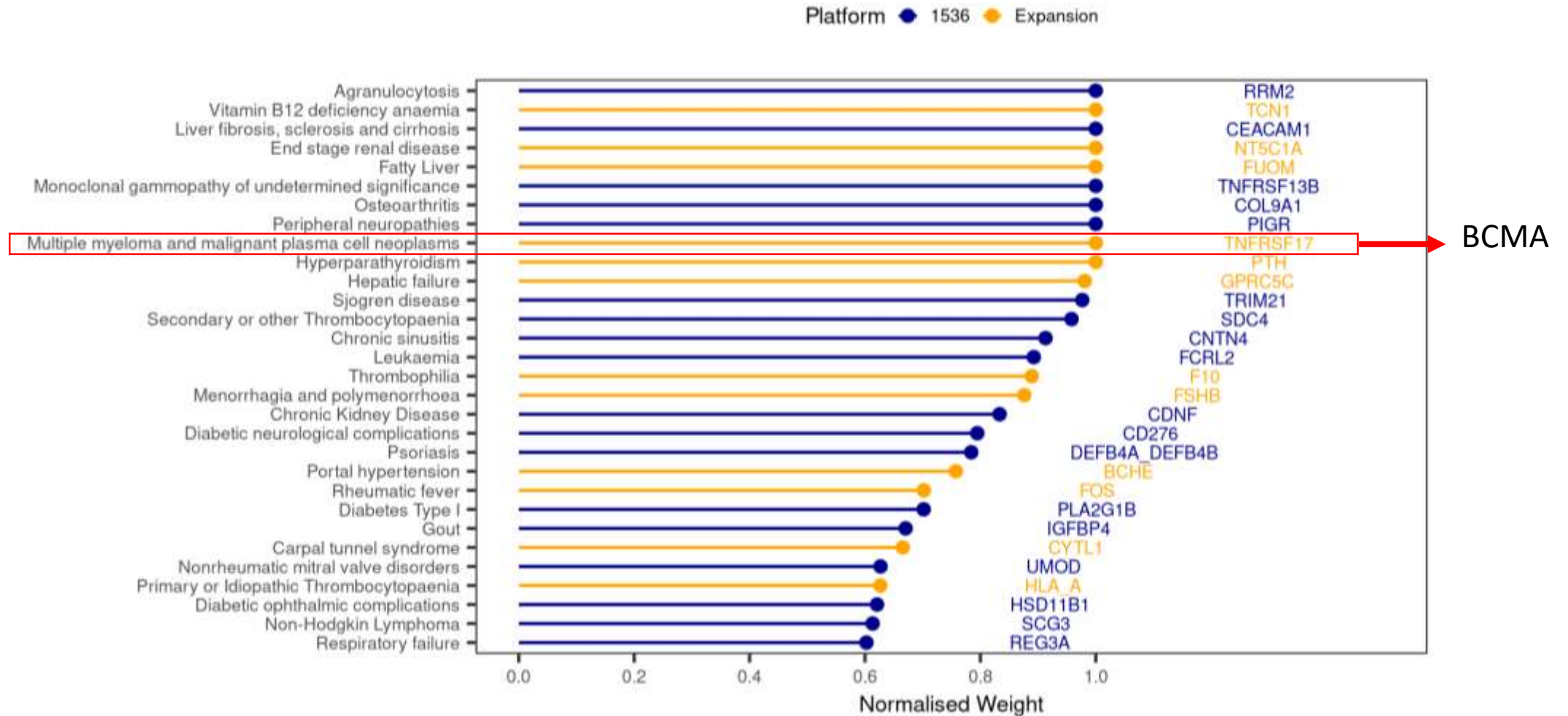


Predictive proteins across more than one disease and clinical specialty

- 501 proteins among signatures for 67 diseases.
- 147 proteins are predictive for more than 1 disease.
 - 89% of those predictive across more than 1 clinical specialty.



Disease-specific predictor proteins



Summary

- Sparse plasma **protein signatures** can improve **identification of people at high-risk of future disease onset**, over and above clinical benchmarks.
 - Achieving screening metrics comparable or higher than current diagnostic markers.
- **Systematic comparison** across diseases highlights **disease-specific biomarkers**, as well as predictive markers across many different diseases.

Limitations and future work

- Benchmarking against disease specific biomarkers (i.e. M-protein for multiple myeloma).
- External validation
 - Alternative proteomic technology
 - Ethnically diverse populations

Acknowledgements



Precision Healthcare University Research Institute (PHURI)



MRC Epidemiology Unit, University of Cambridge



Claudia Langenberg



Maik Pietzner



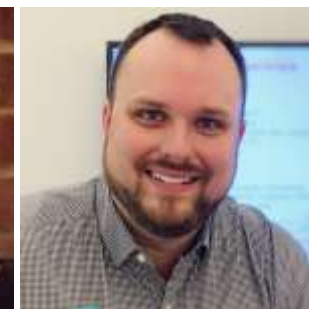
Nick Wareham



Robert A. Scott



Joanna C. Betts



Jonathan Davitte



Praveen Surendran

Health Informatics, UCL, UK



Harry Hemingway



Spiros Denaxas



Ana Torralbo



Christopher Tomlinson



Cai Ytsma

Natalie Fitzpatrick

Tokuwa Kanno
Daniel Freitag



Damien C. Croteau-Chonka



Chloe Robins



Stephan Gade



Frederik Ziebell