



@omicscience

# From Molecules to Medical Records – Insights From Large-Scale, Multi-Omic Studies

BIH Lecture Frontiers in Translational Medicine – Scientific and Structural  
Challenges

11th October 2021

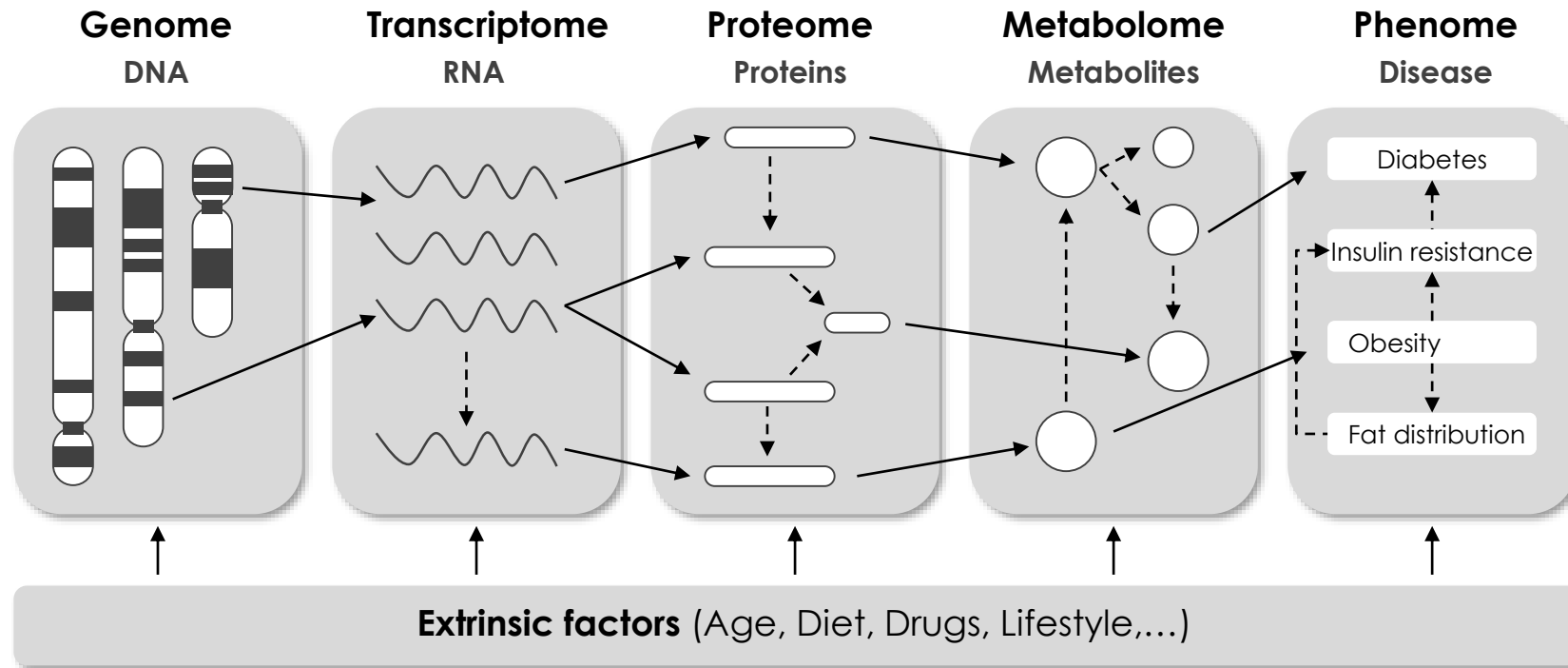
**Claudia Langenberg**

Computational Medicine, Berlin Institute of Health at Charité – Universitätsmedizin  
Berlin

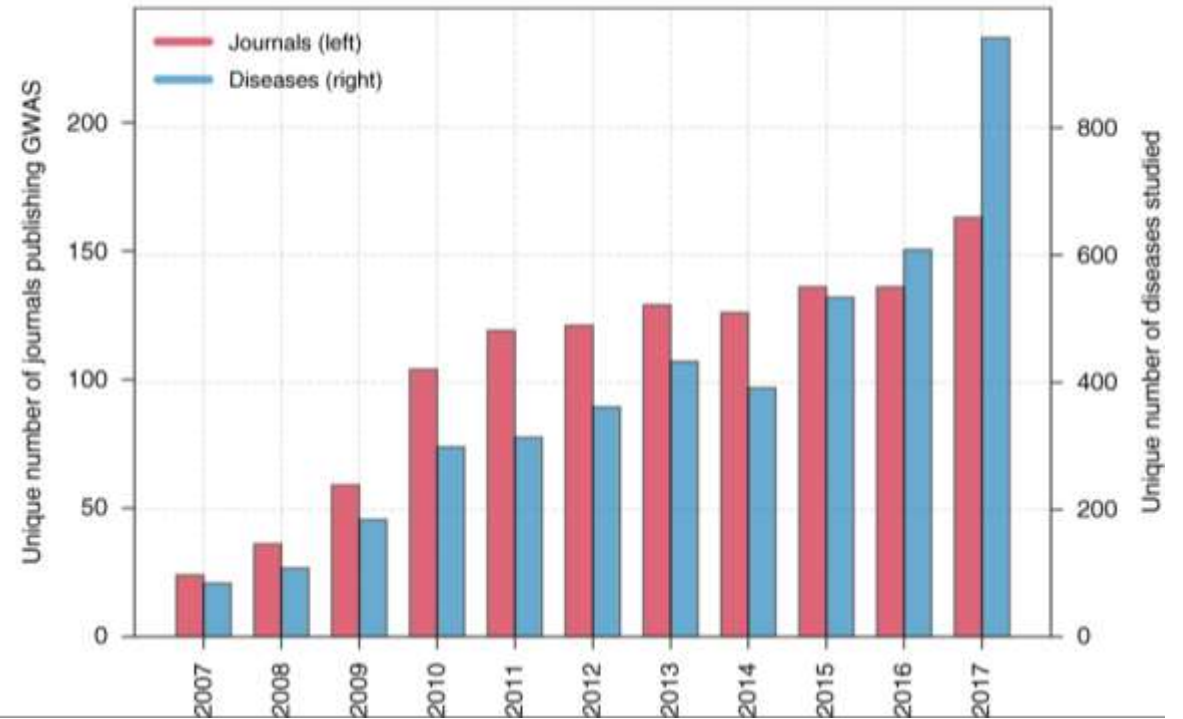
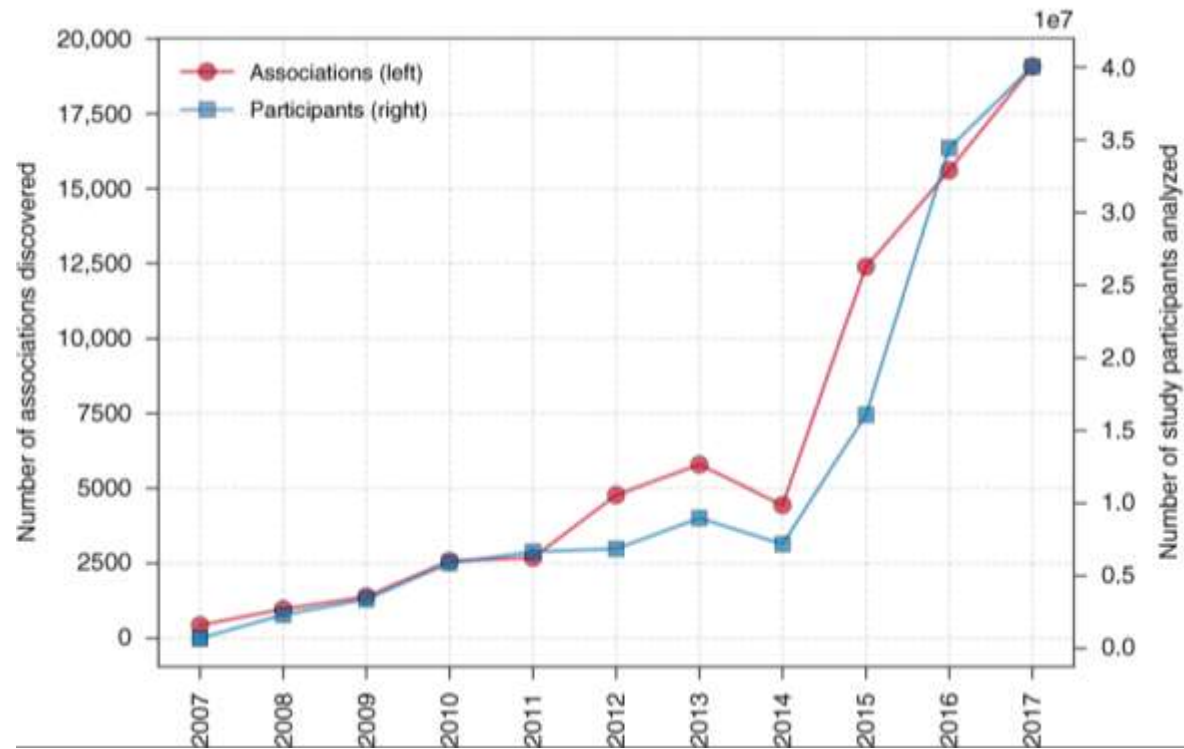
MRC Epidemiology Unit, University of Cambridge, UK  
Honorary Consultant Physician, Public Health England



# 'OMICS?



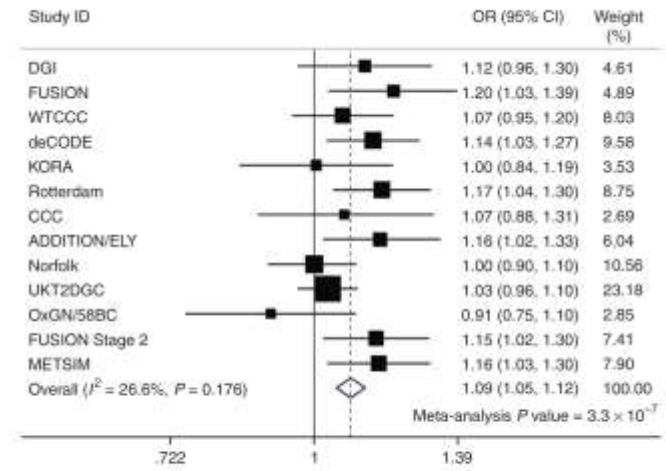
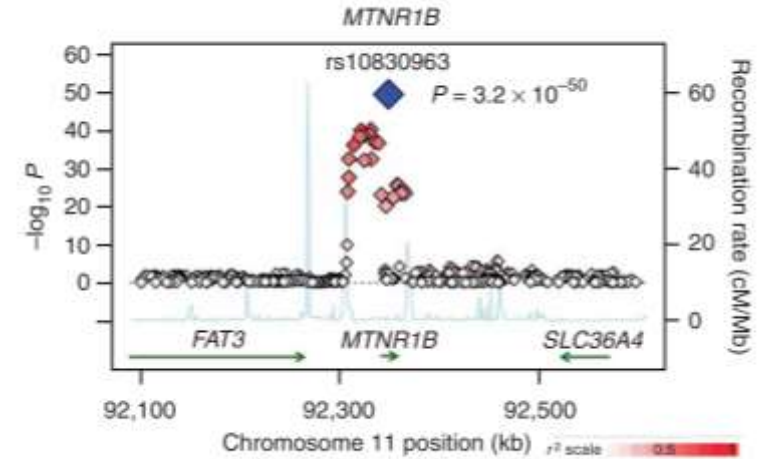
# The growth of GWAS, 2007–2017



# The beginning

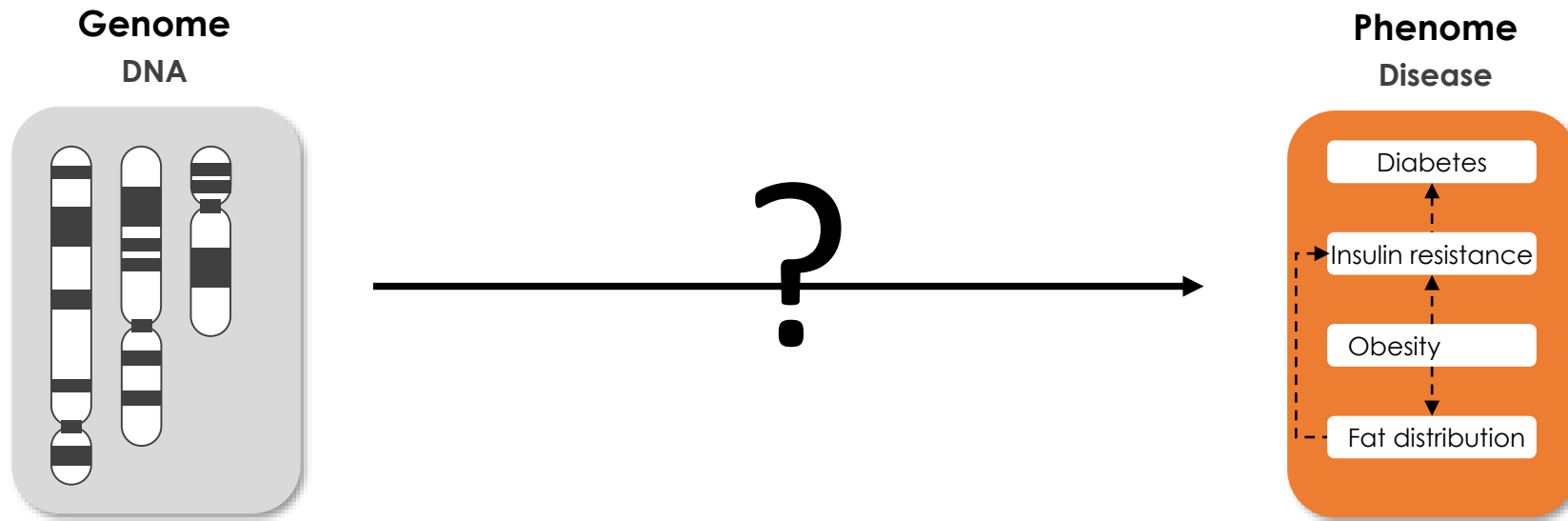
LETTERS

nature  
genetics

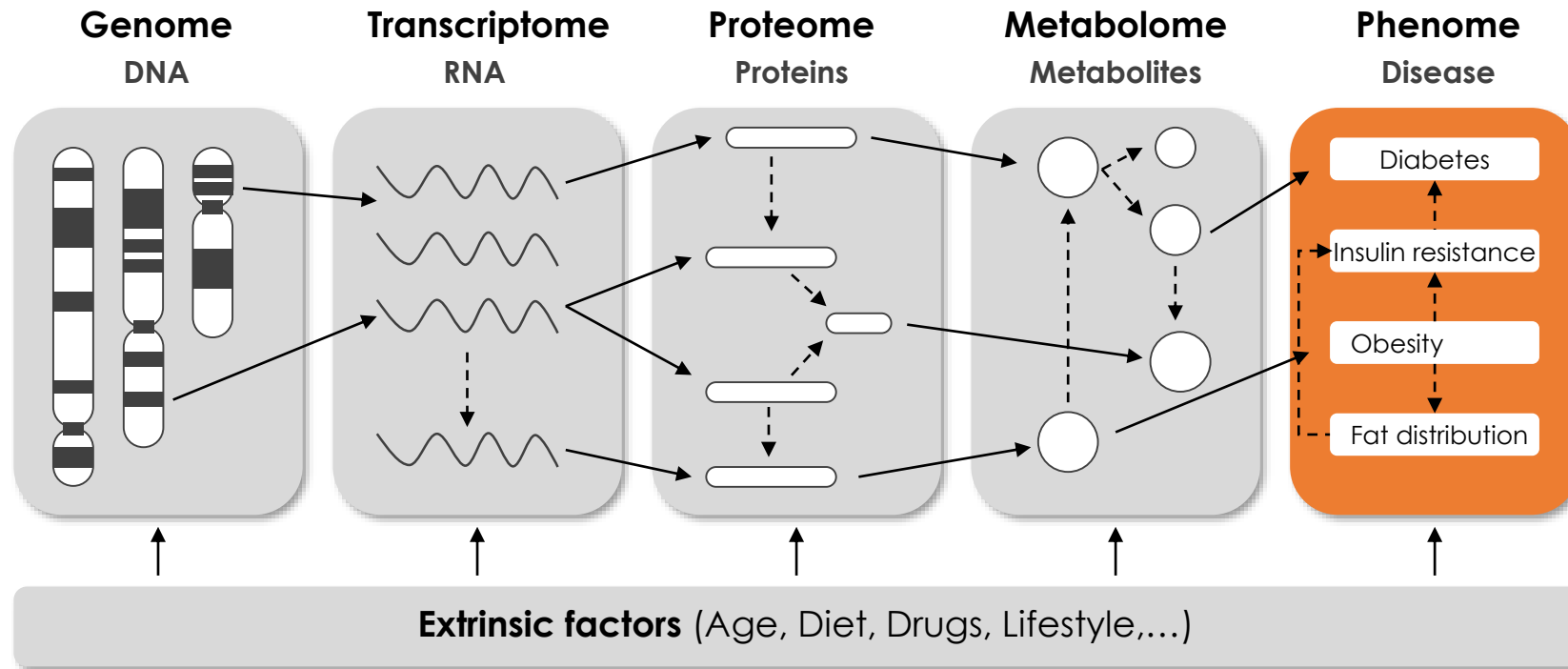


**Figure 2** Association of rs10830963 with type 2 diabetes (T2D) in 13 case-control studies.

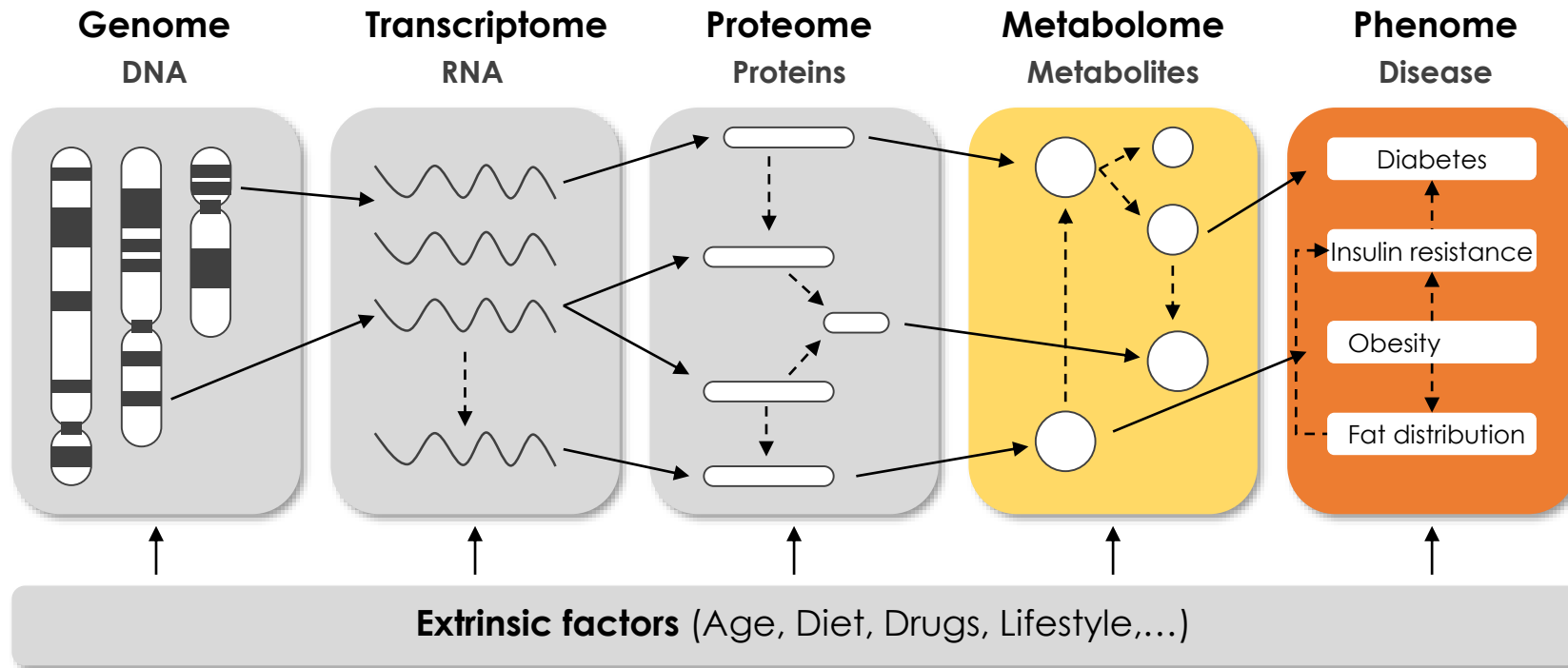
# From association to mechanism



# Genome to phenome via 'omics'

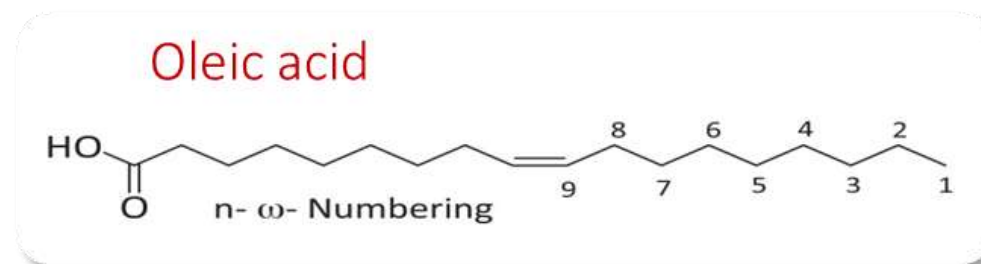
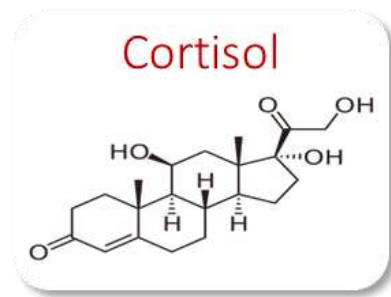
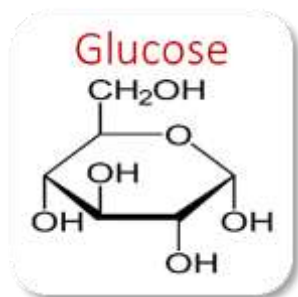


# Blood based 'omics' at scale



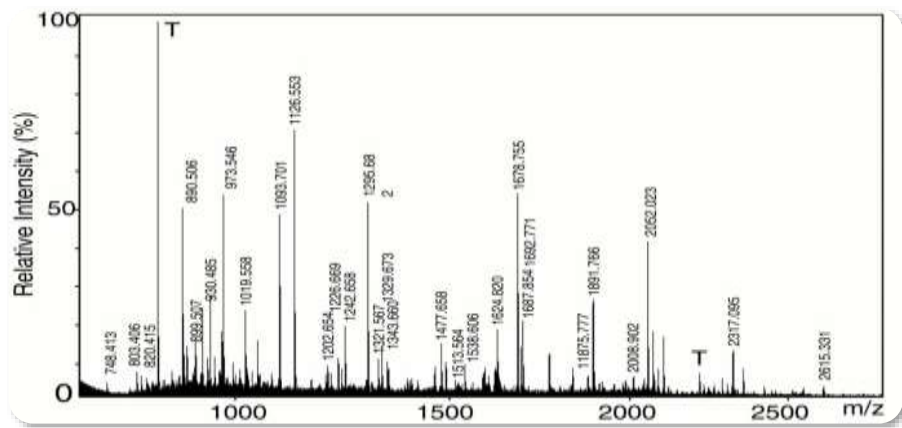
# Metabolome

Entirety of small molecules (<1kDa) in biospecimens like blood, urine, cells or tissues.

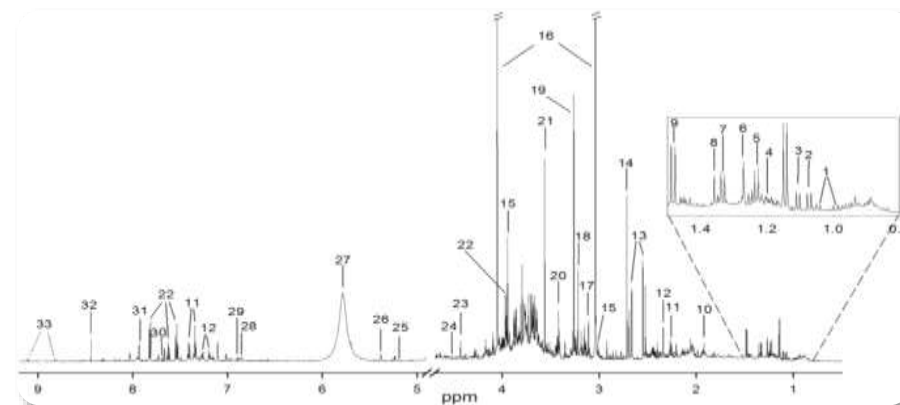


## Measurement techniques

Mass spectrometry  
(high-sensitivity, 1000s of molecules)



Nuclear magnetic resonance spectroscopy  
(high reproducibility)





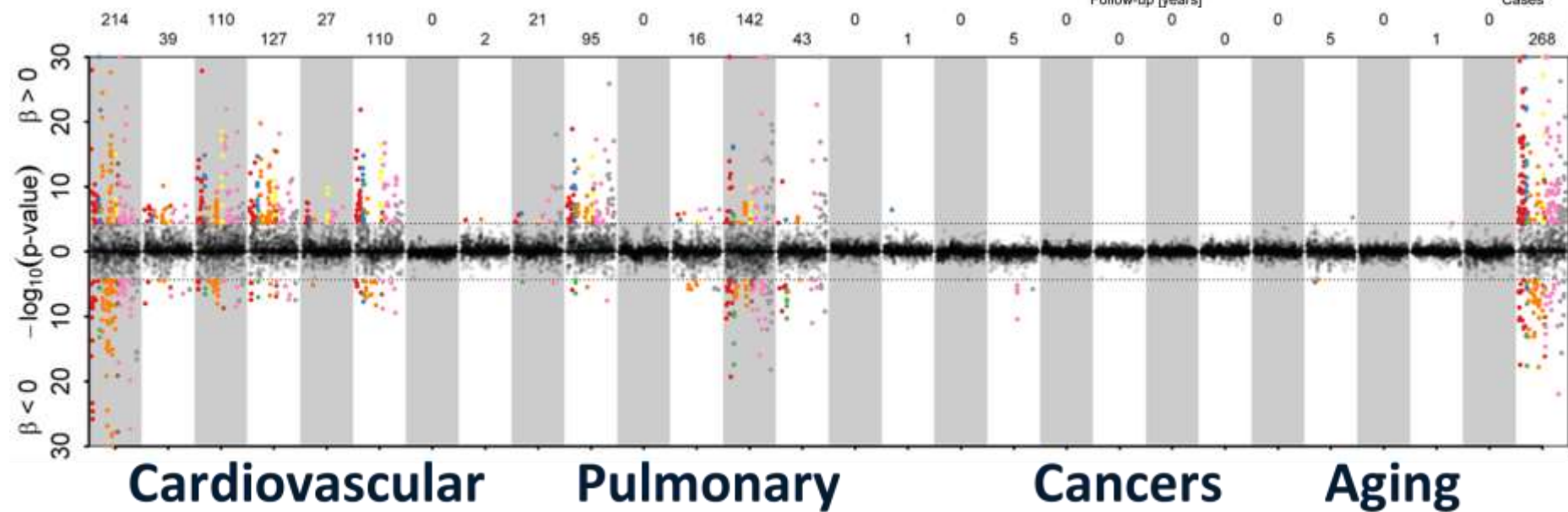
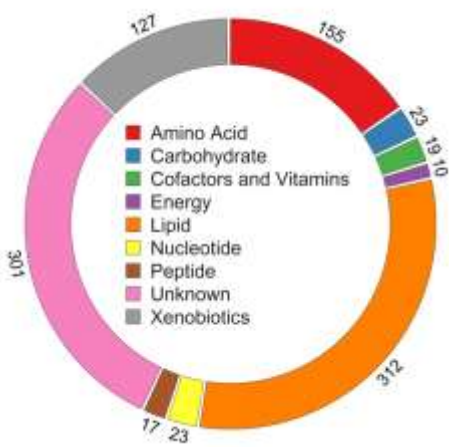
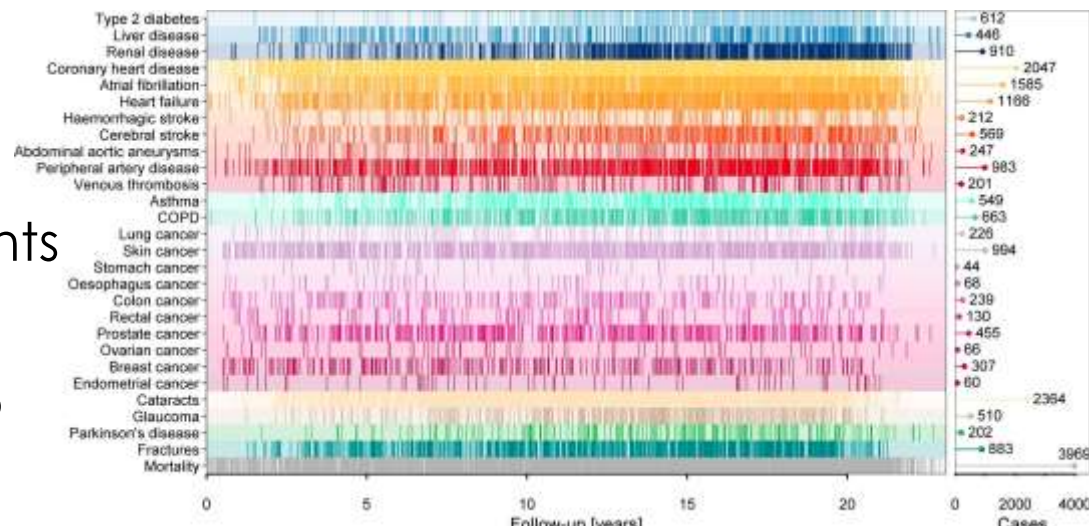
# Metabolome-wide disease associations

EPIC-Norfolk Cohort (n=25,639)

Baseline 1993-97, mean age 60 yrs, 54% women

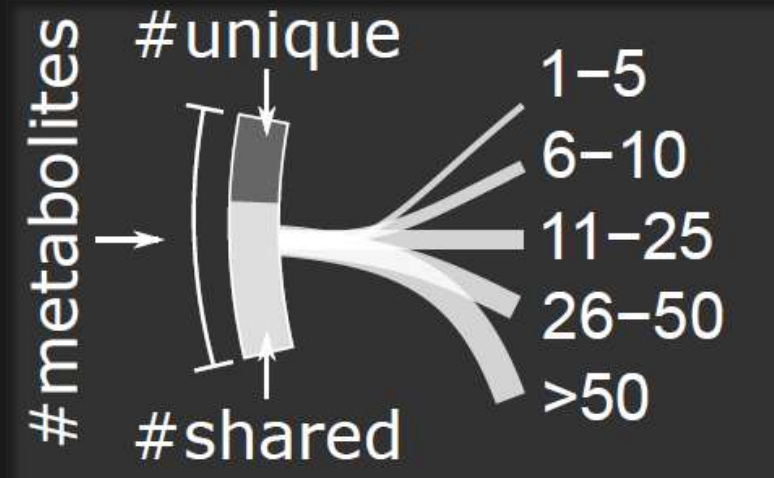
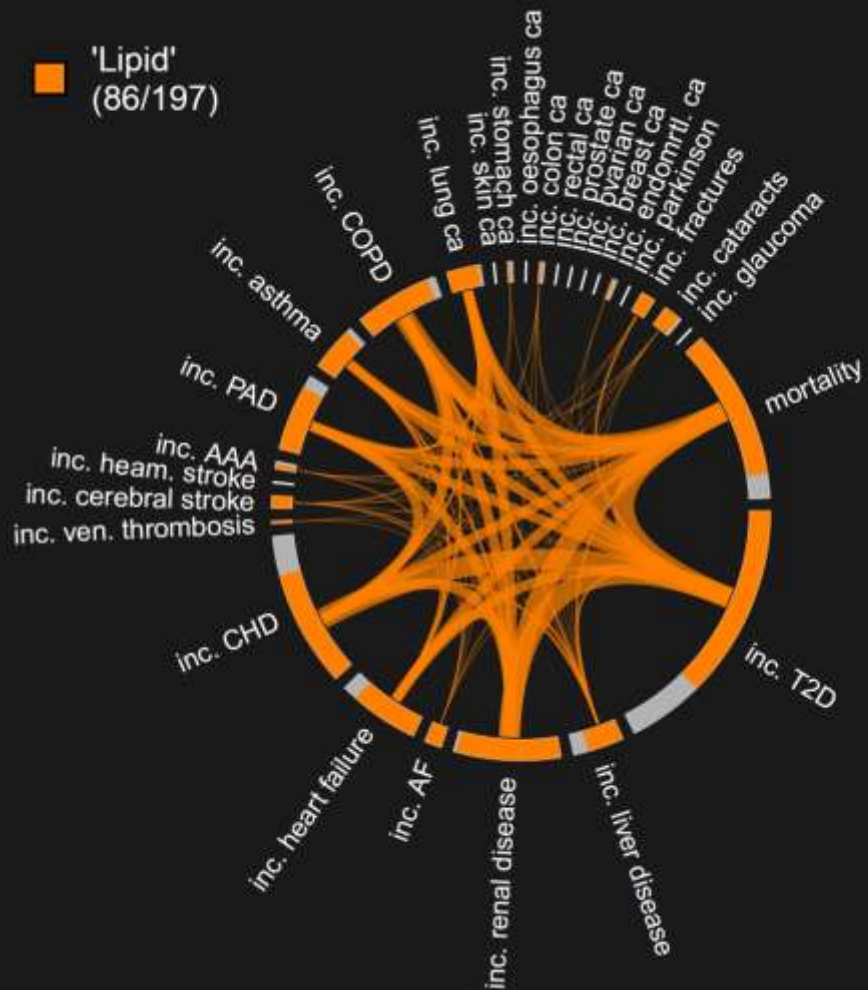
Untargeted metabolomics (n=987) > 11k participants

Incidence of 27 diseases using record linkage  
(hospitalisations): 219,415 person years of follow-up

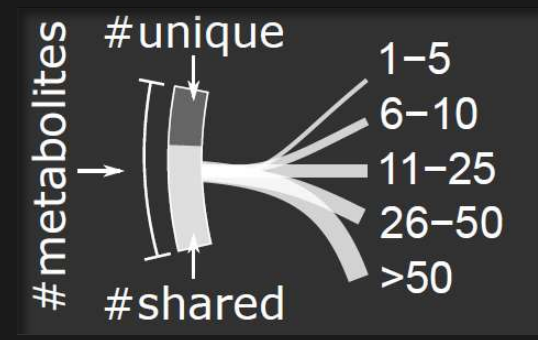
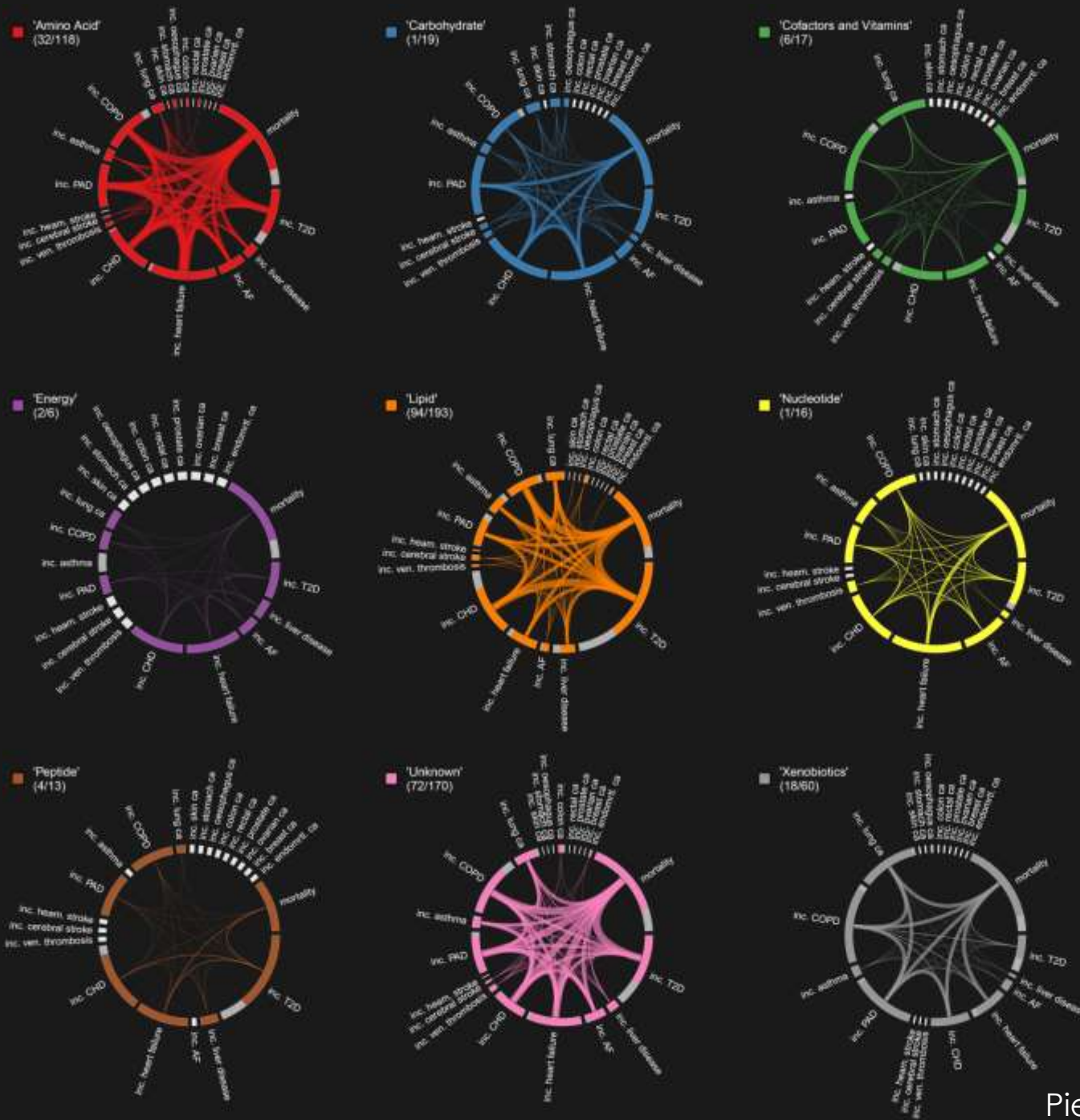


Metabolon

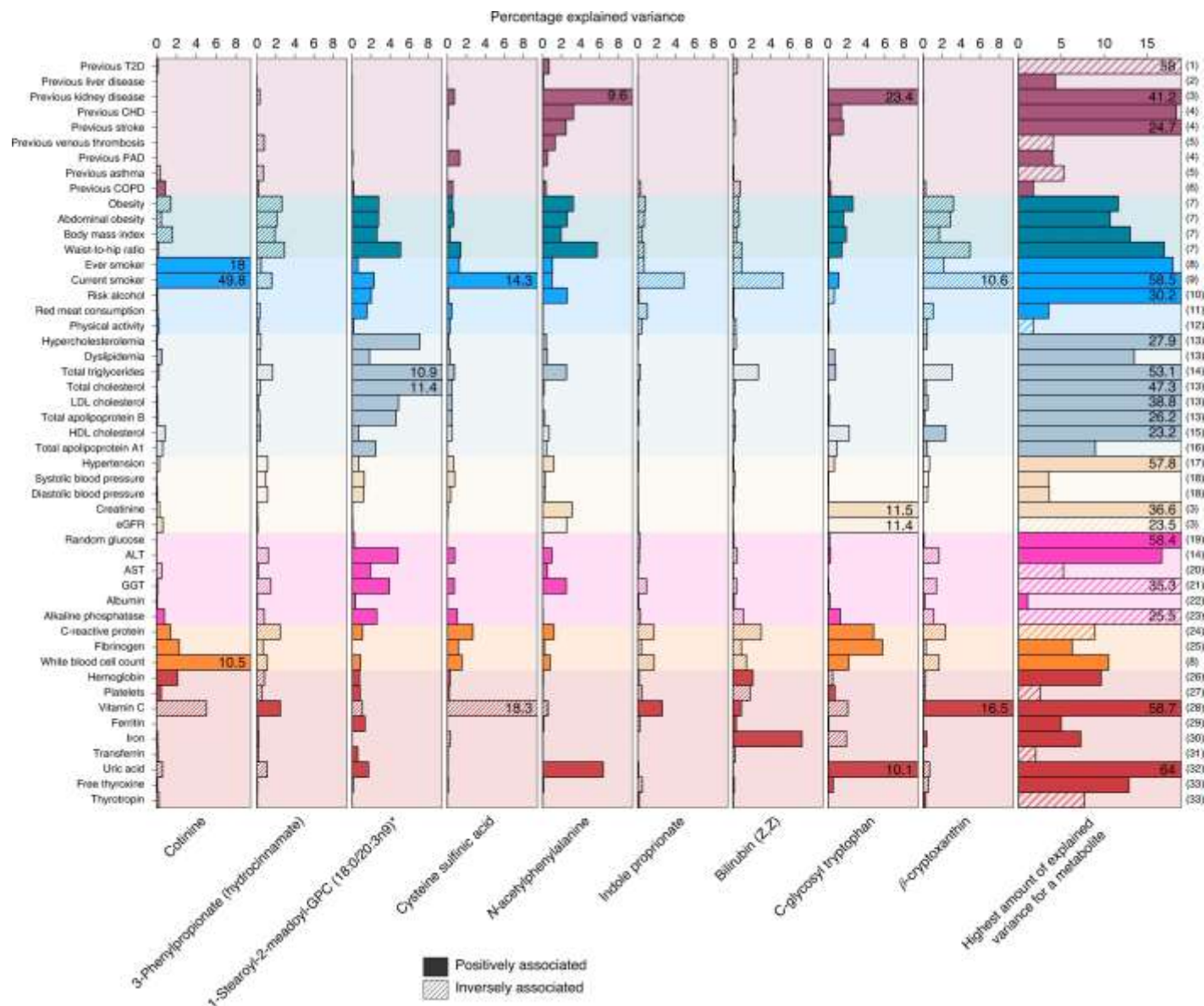
# 'Sharedness' of disease associations



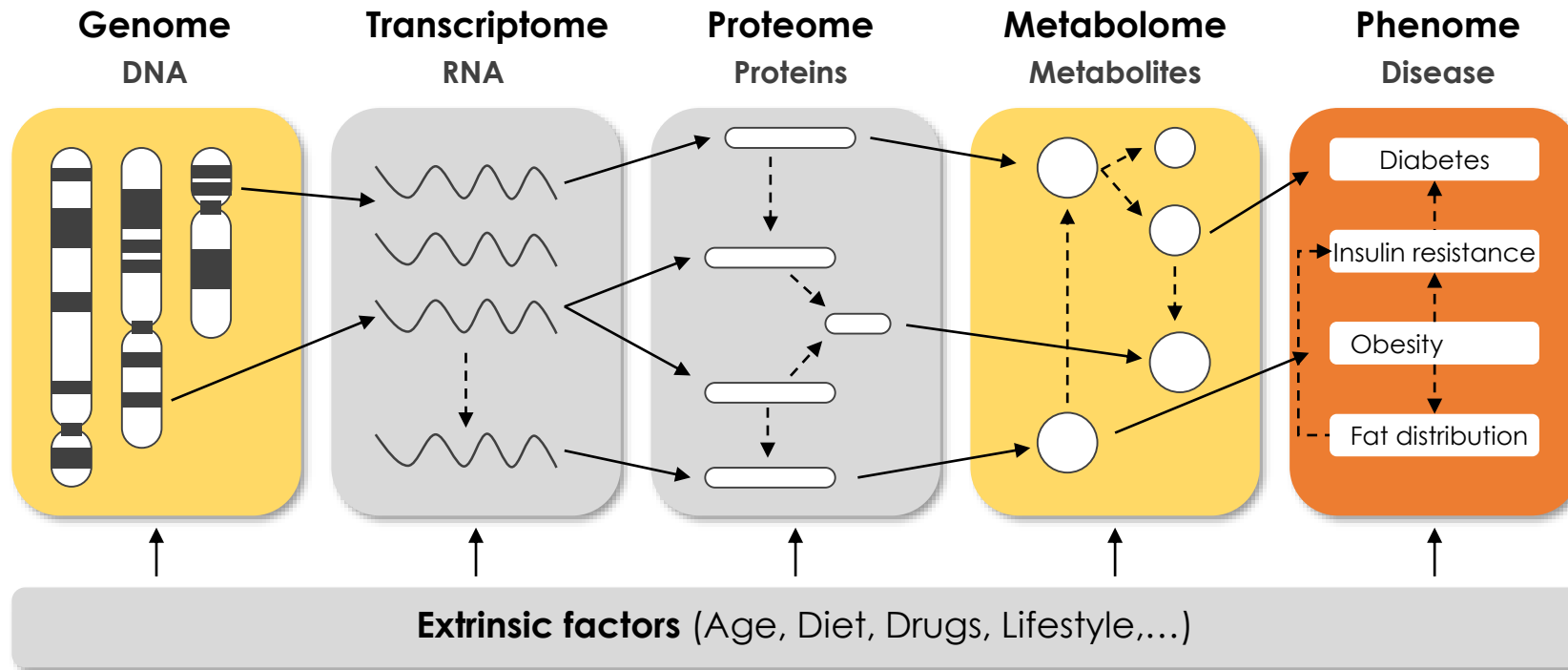
# SHARED ASSOCIATIONS ACROSS 9 METABOLITE CLASSES



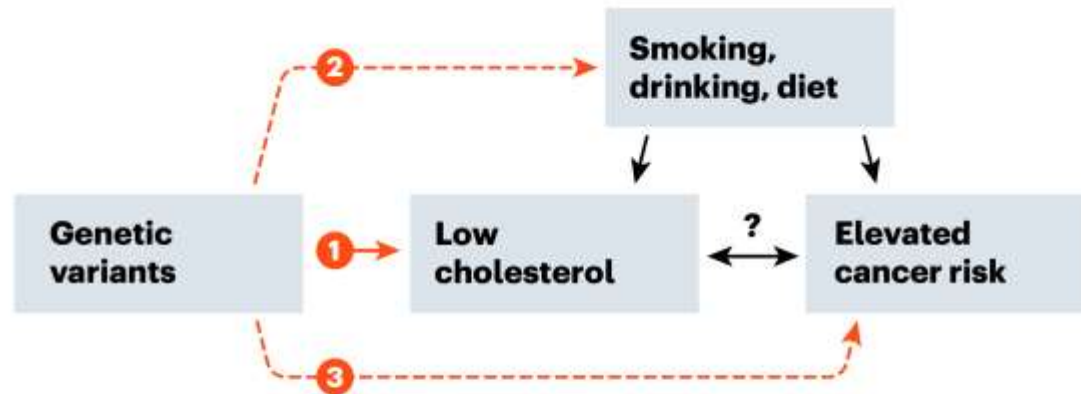
# Antecedents of multimorbidity



# Genome to phenome via 'omics'



# 'Mendelian Randomisation'



## Assumption 1

The link between the variants and having low cholesterol has to be strong and stable over time.

## Assumption 2

The variants must not influence variables that might affect both cholesterol levels and cancer risk.

## Assumption 3

The variants must not be associated with cancer risk in any way other than through the relationship to cholesterol.

Feature

# THE CAUSATION DETECTOR

A technique called Mendelian randomization has become the go-to for drawing lessons from epidemiological data. But are scientists overdoing it? By David Adam



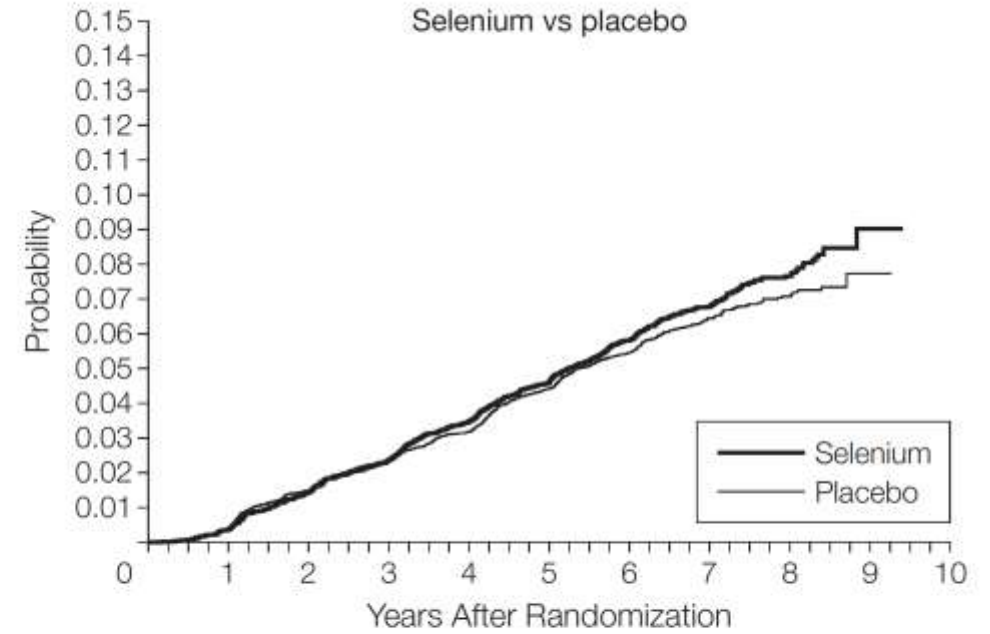
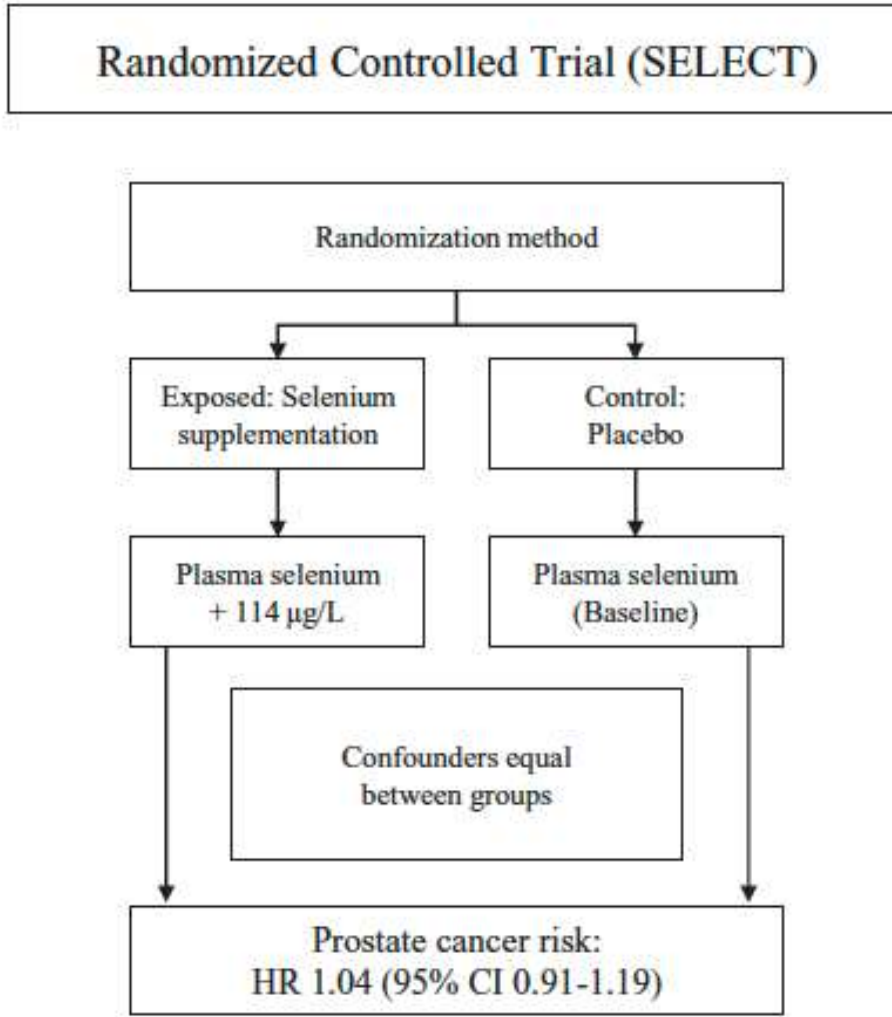
In 1812, the British ophthalmologist James Ware relayed a curious finding to the members of the Royal Society in London. Of thousands of young men recruited to regiments of the British army, only six had been turned away for poor vision in 20 years. But up to one-quarter of students about the same age going to the University of Oxford, UK, relied on a hand glass or spectacles.

Ware didn't draw any conclusions about cause and effect: that poring over books might contribute to poor eyesight, for example, or that the bespectacled are naturally drawn to academic pursuits. And just as well.

Epidemiologists have long been frustrated by observations that link environmental exposures and health. Myopia is a classic example. Decades of studies show that children who spend the most time at school have the worst eyesight. But the data don't reveal whether schooling makes children myopic or whether myopic kids spend more time at school. Or whether something else, such as socio-economic status, drives both.

Fed up with this logical cul-de-sac, by the turn of this century some epidemiologists had begun suggesting that their field should call it a day. Advances in genetics, they said, could do a better job.

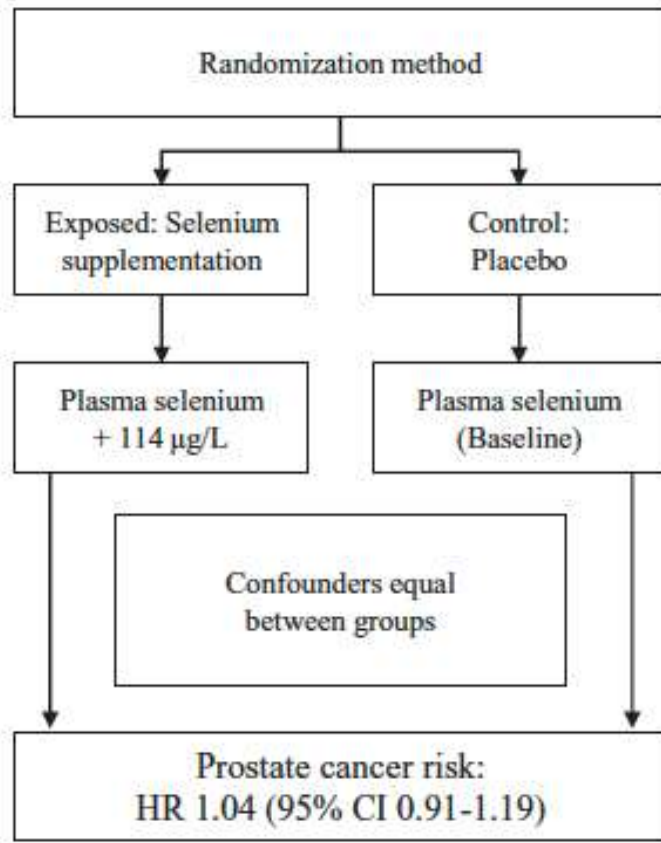
# Application: SELECT Trial



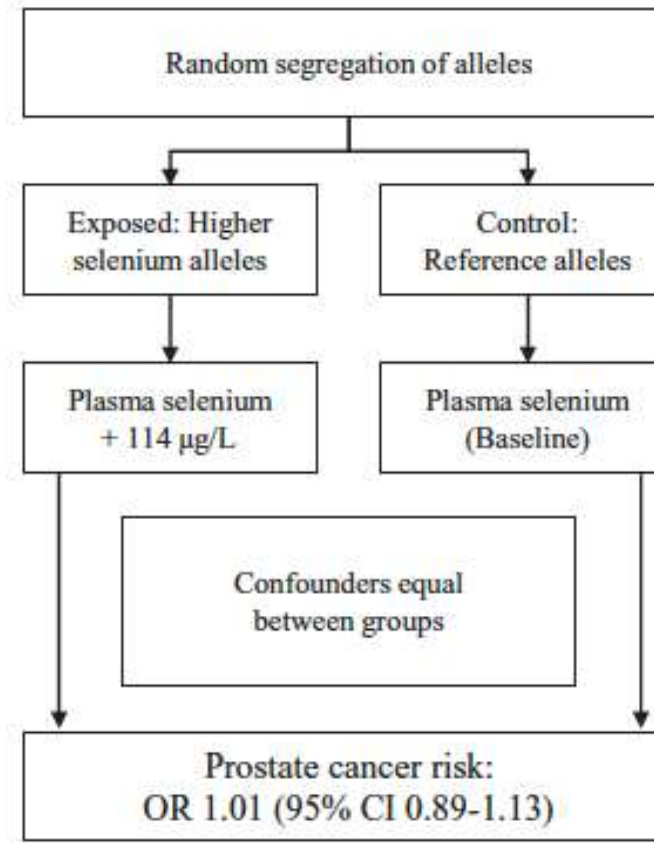
No. at risk	0	1	2	3	4	5	6	7	8	9	10
Placebo	8565	8344	8081	7831	7471	6399	4044	1833	70		
Selenium	8600	8360	8131	7826	7456	6425	4075	1829	66		
Cumulative cases		31	123	202	293	384	474	532	563	575	

# Application: SELECT Trial

## Randomized Controlled Trial (SELECT)



## Mendelian Randomization





# Maximising power: sample size 10-85k

## MRC Fenland Cohort

Baseline 2005-15

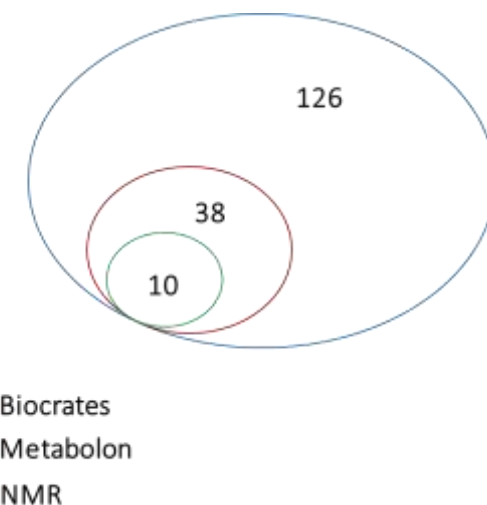
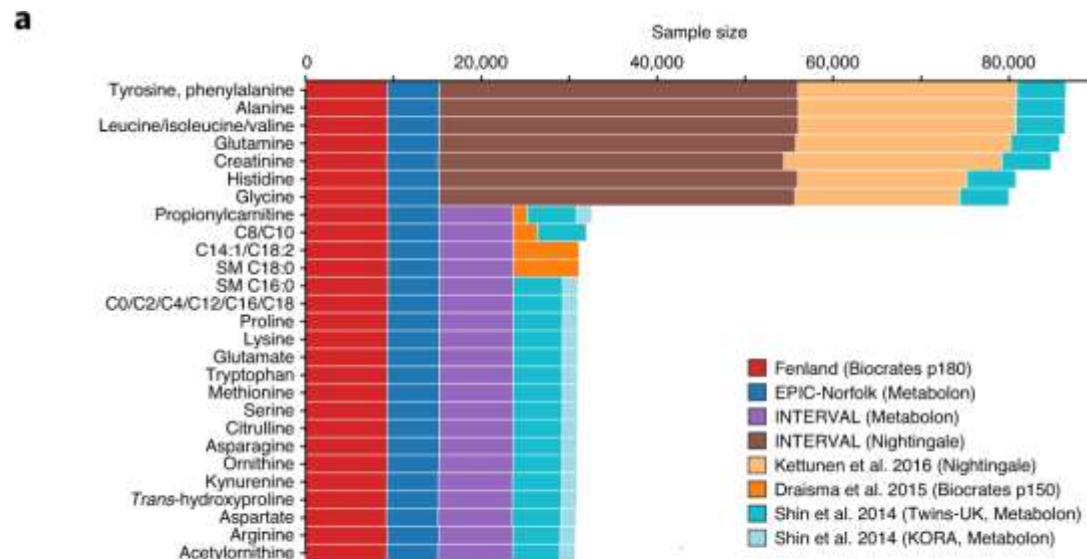
N=12,435, mean age 49 years, 54% women

10,708 genotyped using 3 different arrays

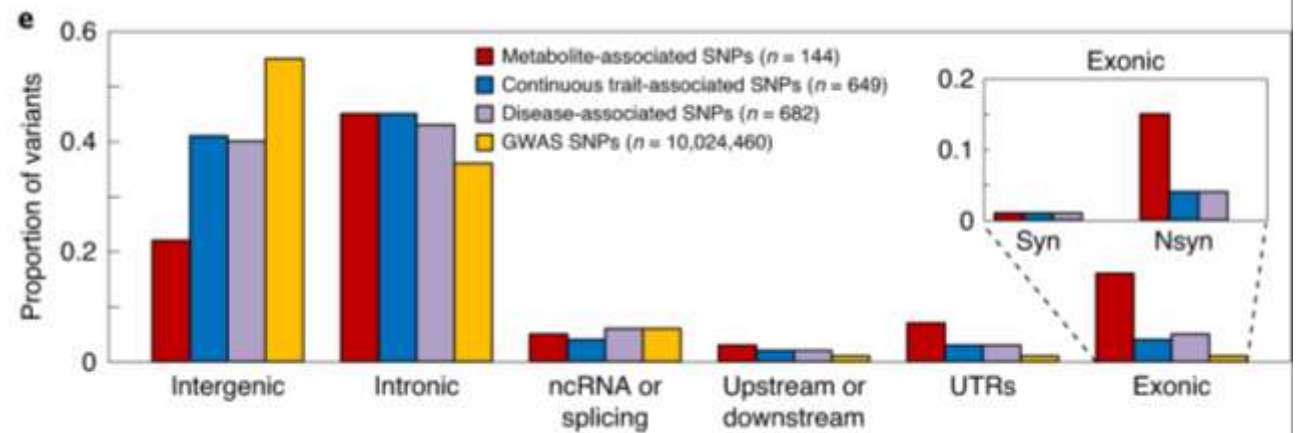
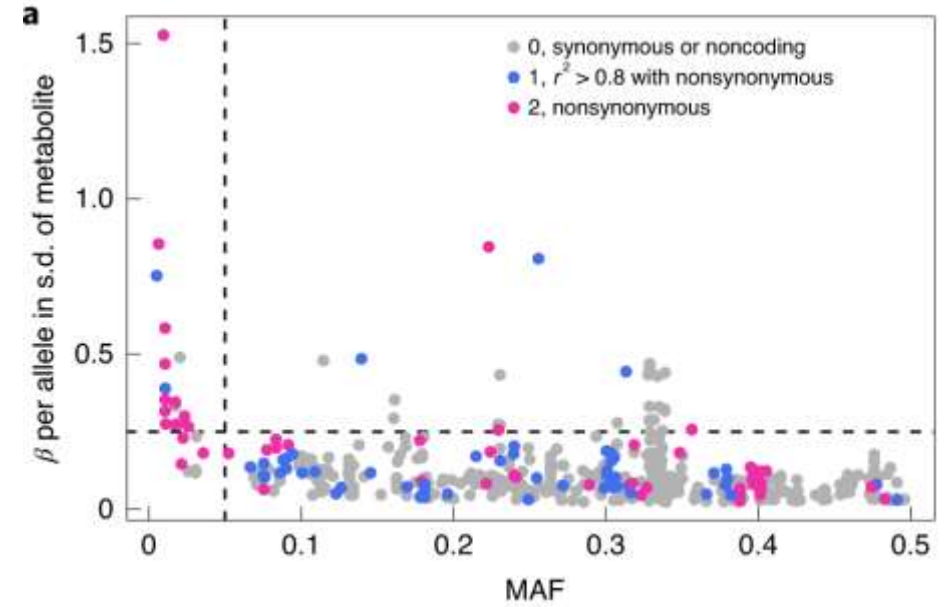
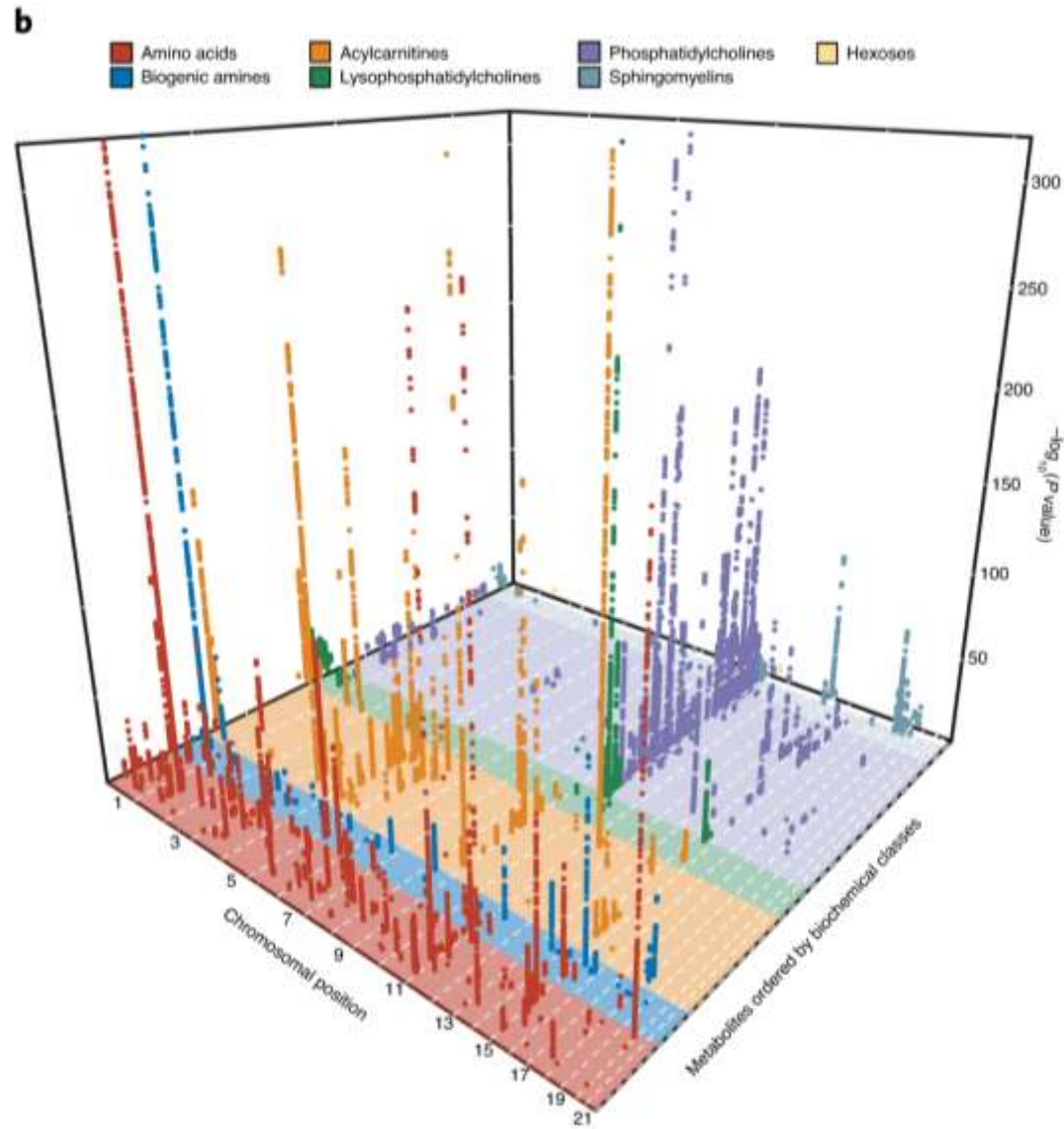
Biocrates (AbsoluteIDQ p180)

174 targeted metabolites

- Hexoses
- Amino Acids
- Biogenic Amines
- Acylcarnitines
- Glycerophospholipids
- Sphingolipids



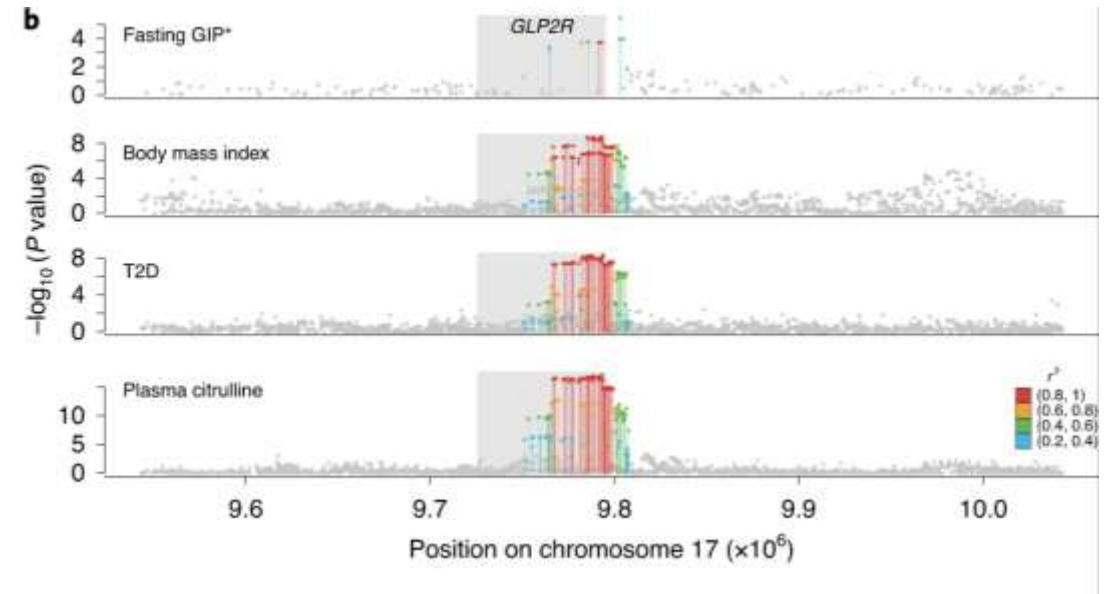
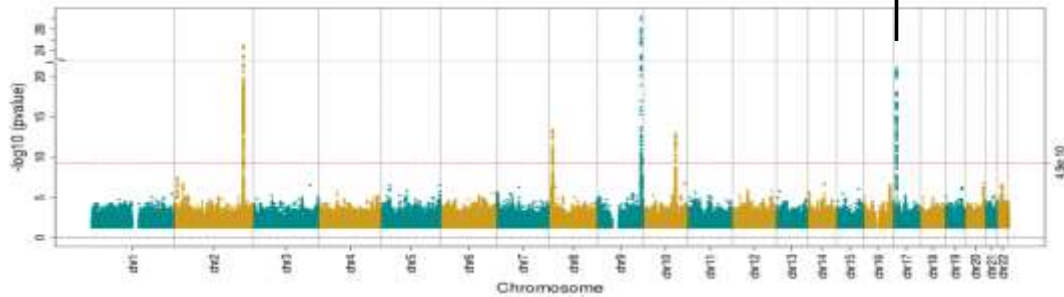
# 144 regions, 499 locus-metabolite associations



# GLP2R and citrulline levels

GLP2R - Glucagon-like peptide 2 receptor  
 - Expressed in the gut

Glucagon receptor  
 GLP1R - insulin secretion

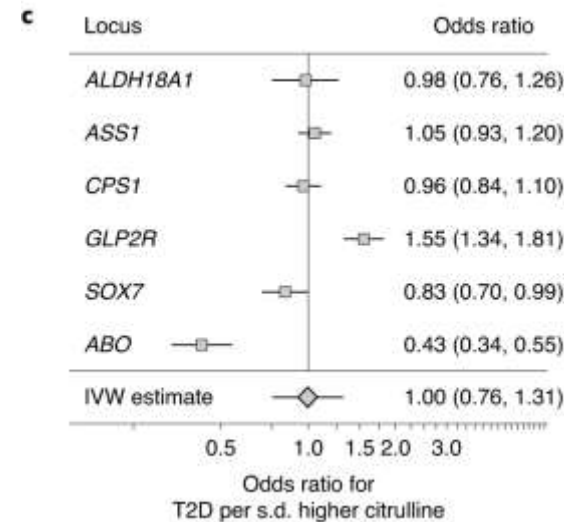


- GLP2 stimulates intestinal growth > analogues used to treat short bowel syndrome
- Citrulline is a biomarker of intestinal function and target engagement

Citation: Clinical and Translational Gastroenterology (2015) 6, e93; doi:10.1038/ctg.2015.15  
 © 2015 the American College of Gastroenterology. All rights reserved. 2255-384X/15  
 www.nature.com/ctg

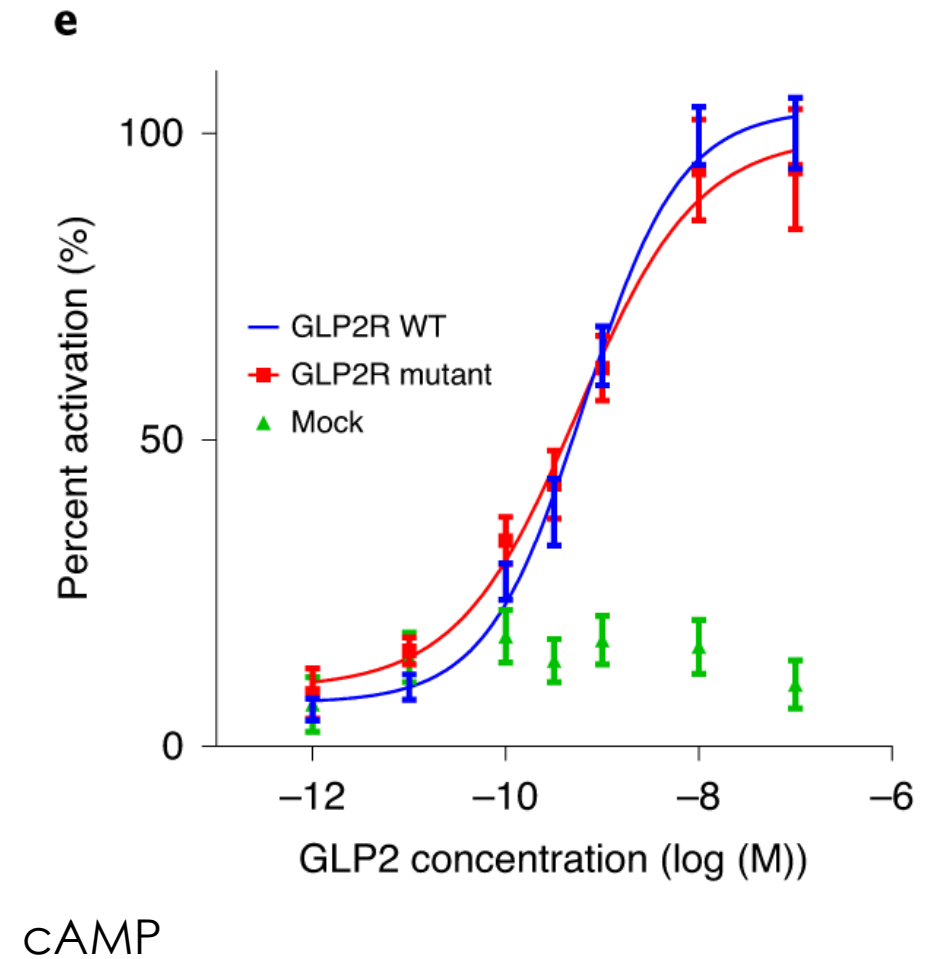
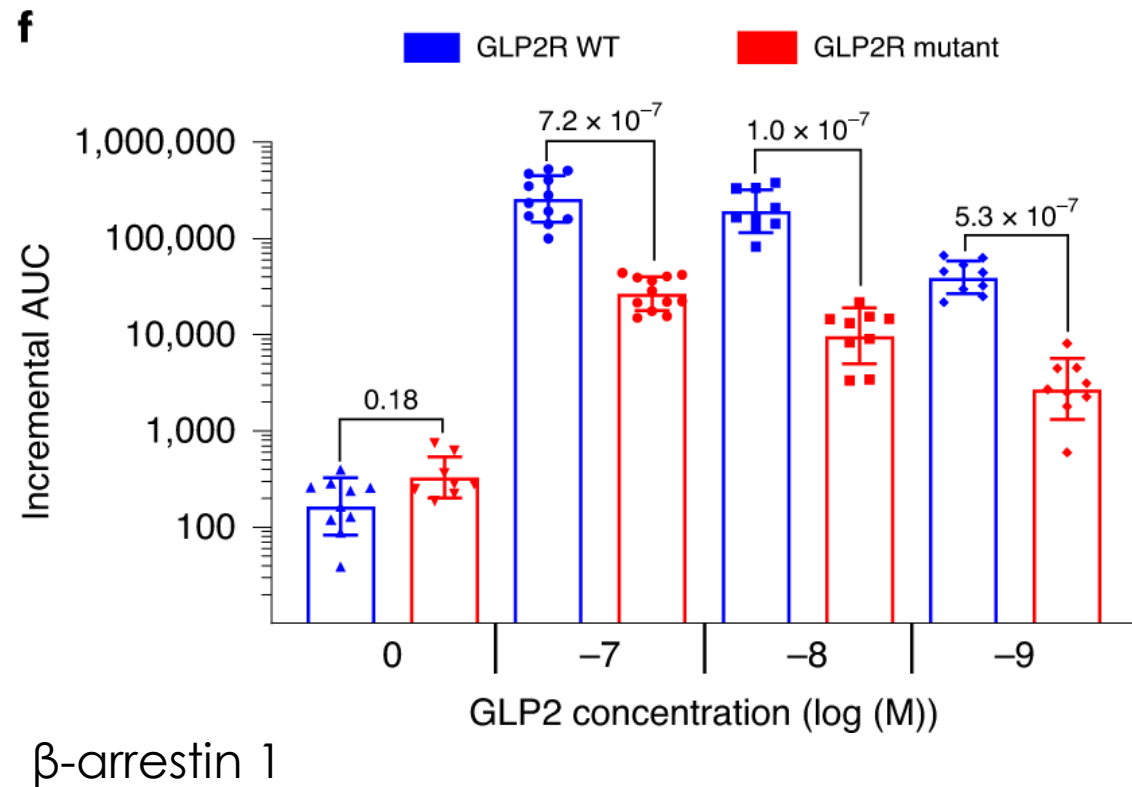
## Effect of Teduglutide, a Glucagon-like Peptide 2 Analog, on Citrulline Levels in Patients With Short Bowel Syndrome in Two Phase III Randomized Trials

Douglas L. Seidner, MD, AGAF, FACP, CNSC<sup>1,2</sup>, Francisca Joly, MD, PhD<sup>3</sup> and Nader N. Youssef, MD<sup>4</sup>



GLP2R signaling ↑  
 GIP levels (chronic) ↑  
 GIP receptor signaling (beta-cells) ↓  
 Reduced insulin secretion ↓ T2D ↑

# Reduced recruitment of $\beta$ -arrestin to *GLP2R*



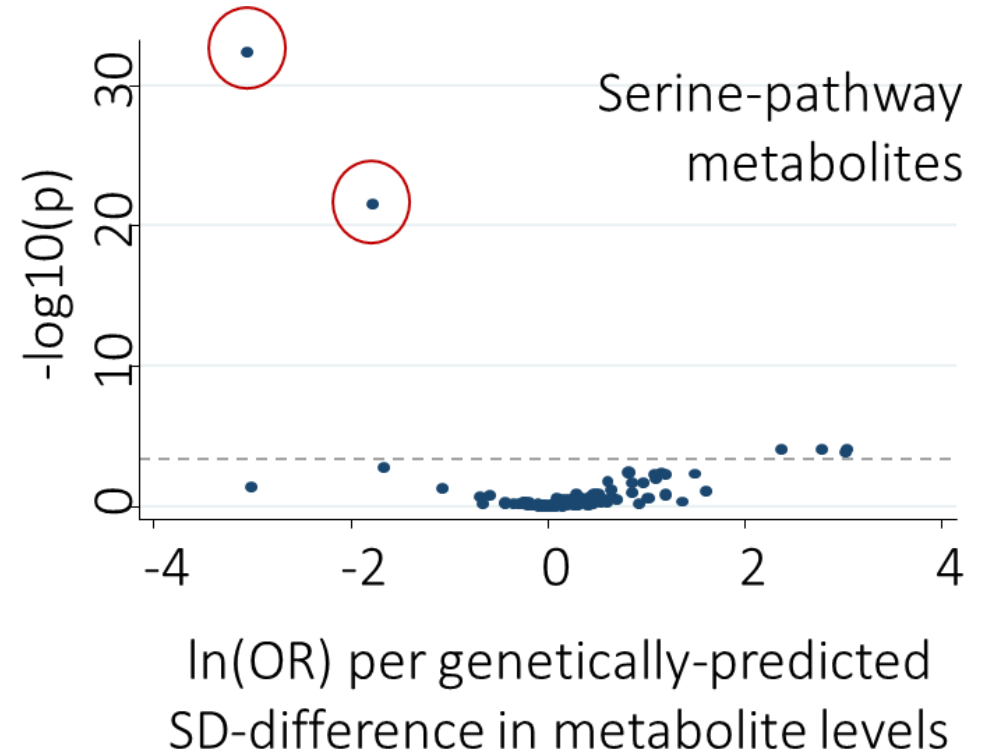
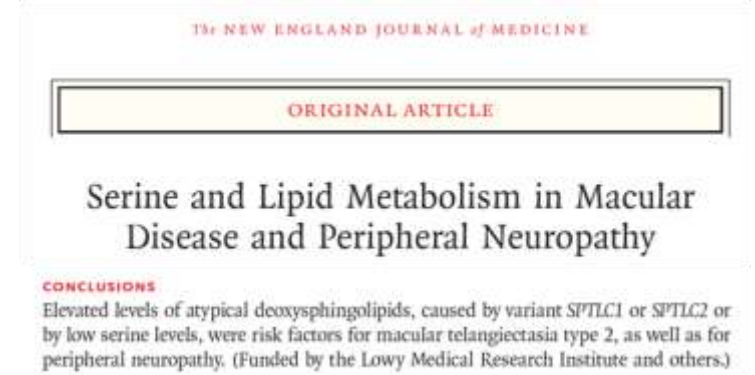
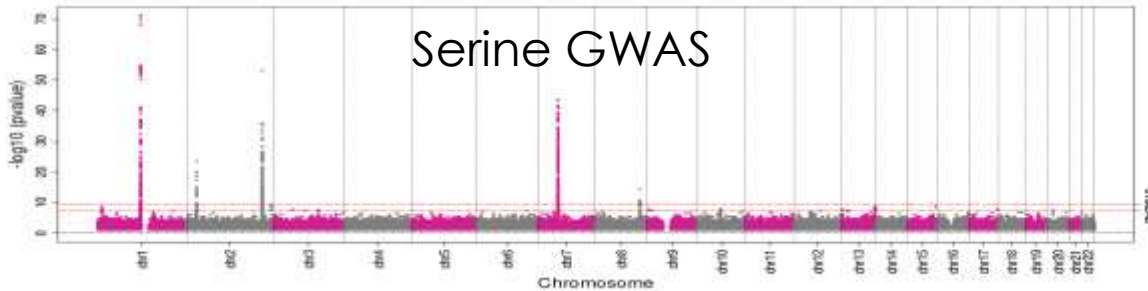
# Understanding locus-disease associations



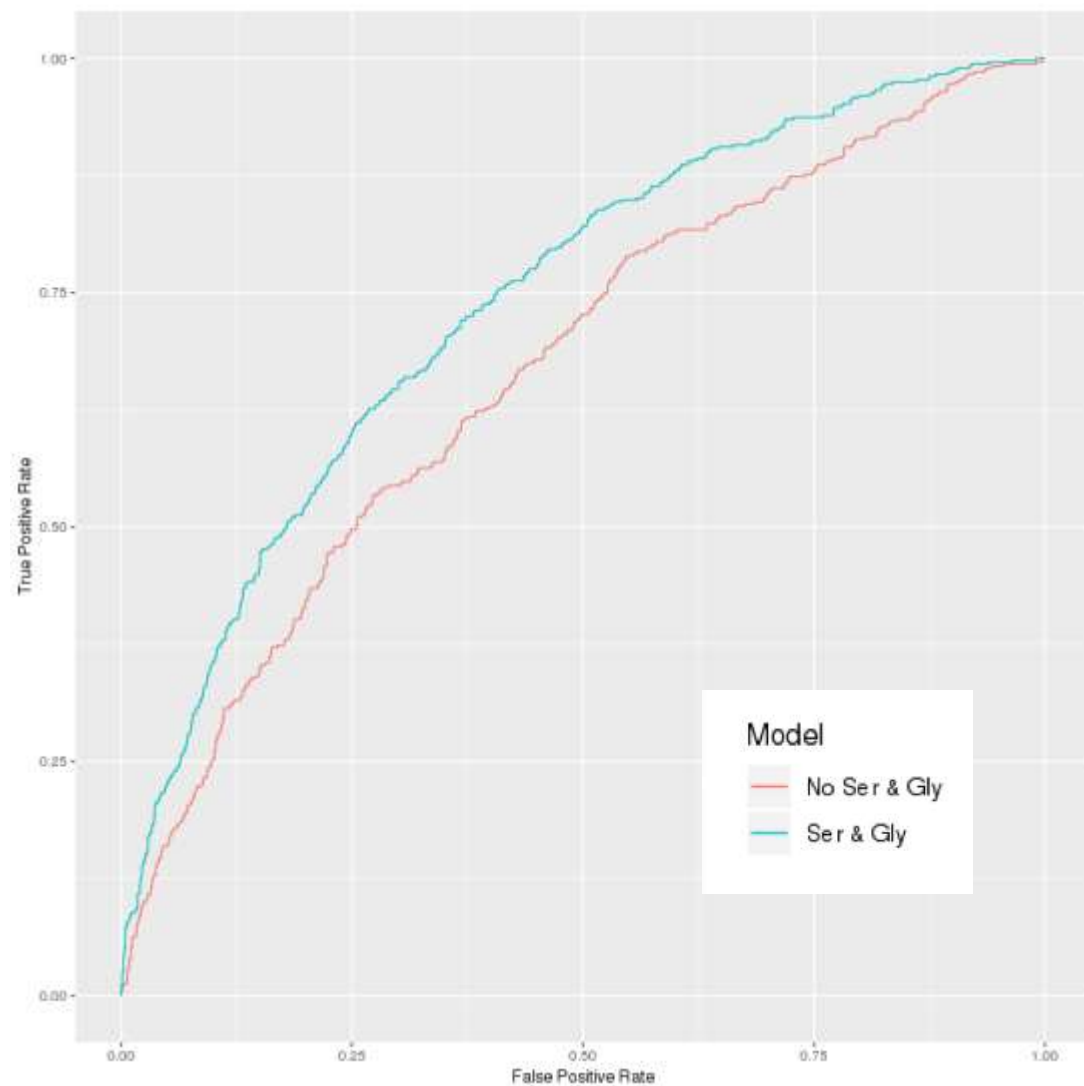
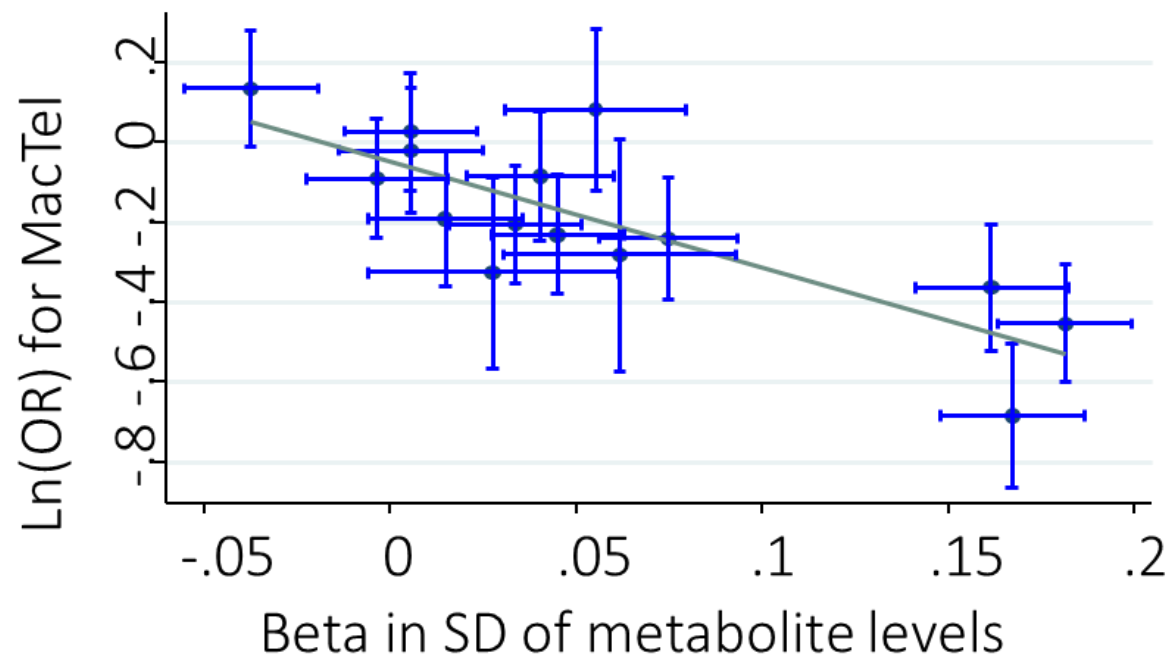
Genome-wide analyses identify common variants associated with macular telangiectasia type 2

Thomas S Scerri<sup>1,2</sup>, Anna Quagliari<sup>1,2</sup>, Carolyn Cai<sup>3</sup>, Jana Zernant<sup>3</sup>, Nori Matsunami<sup>4</sup>, Lisa Baird<sup>4</sup>, Lea Schepke<sup>5</sup>, Roberto Bonelli<sup>1,2</sup>, Lawrence A Yannuzzi<sup>3,6</sup>, Martin Friedlander<sup>5,7</sup>, MacTel Project Consortium<sup>8</sup>, Catherine A Egan<sup>9</sup>, Marcus Fruttiger<sup>10</sup>, Mark Leppert<sup>4</sup>, Rando Allikmets<sup>3,11</sup> & Melanie Bahlo<sup>1,2,12</sup>

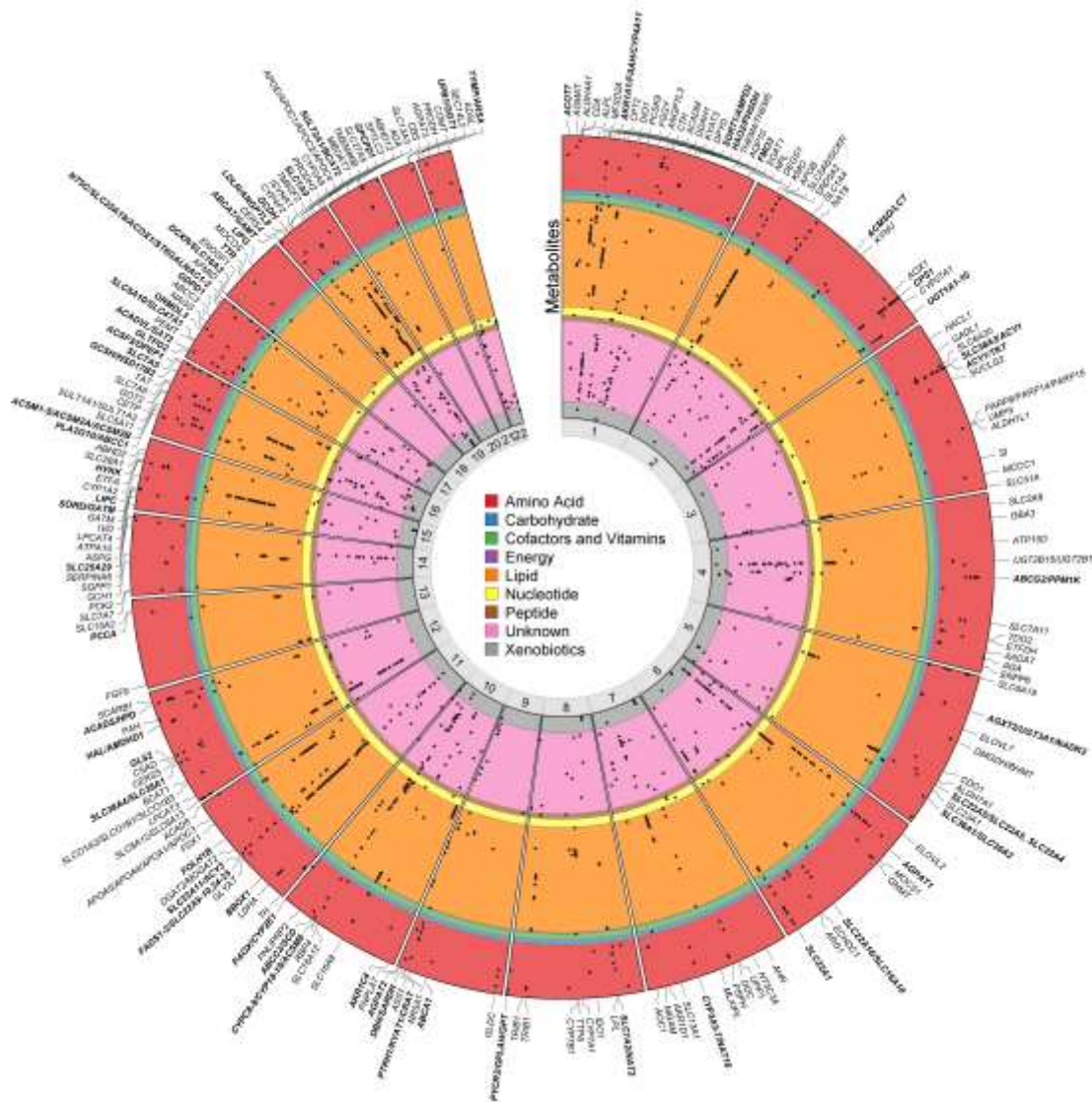
“loci associated with glycine and serine metabolism”



# Dose-response and MacTel2 prediction



# Broadening scope: untargeted metabolomics



Metabolon

# Summary (I of II)

- Increased scale and greater variant diversity increased the number of identified variants
- Cross-platform feasibility
- Specific mQTL characteristics
- Clinical utility and improved understanding of disease mechanisms
- Webserver: a resource for the scientific community (Helmholtz Centre Munich)

**OMICSCIENCE** Cross-platform genetic discovery of small molecule products of metabolism and application to clinical outcomes  
Lotta Peltner et al. Nature Genetics, 2020 (accepted); preprint available at bioRxiv.org  
Results from this publication are freely available for academic use only (see License terms). Contact the corresponding author for non-academic access. Responsibility for the results lies with the corresponding author of this publication. By using these results in your research, you agree to cite our publication. Our legal notice and data protection statement apply. Developed by Johannes Kolter.

Range SNP Gene Export data Show/hide columns Search:

ENSMBL gene identifier:  
  
Hint: perform a gene id lookup by entering a gene symbol (e.g. FADS1).

p-value threshold (meta-analysis)

Metabolites  
Search available: Amino acids, Alanine, Arginine, Asparagine, Aspartate, Glutamate, Glutamine  
Search selected: Amino acids, Citrulline

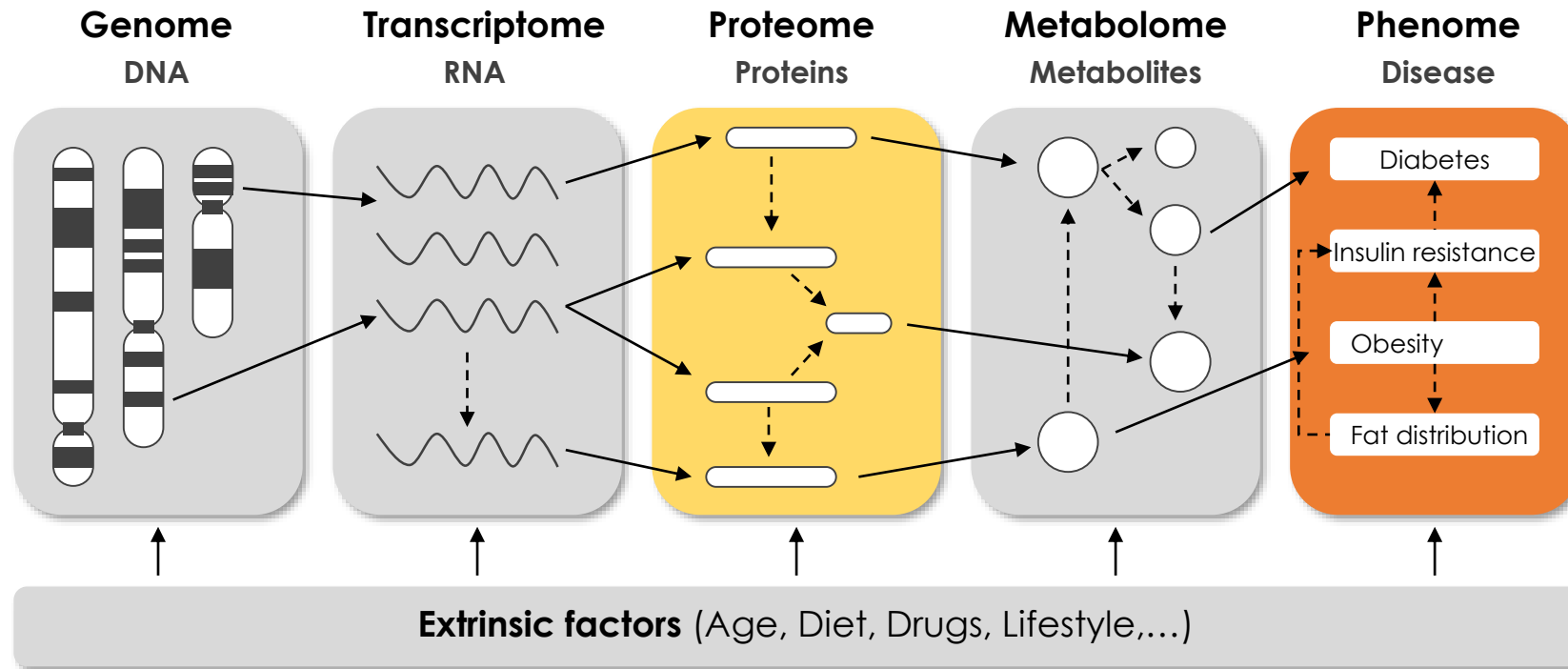
Metabolite	Pathway	Chr	Pos	rs ID	rs ID alias	A1	A2	Freq. A1	Weight	Z-score	p-
Citrulline	Amino acids	17	9,726,539	rs139594098		t	c	0.02	23,615	2.320	0.02
Citrulline	Amino acids	17	9,727,553	rs78235307		c	g	0.03	23,615	2.820	0.00
Citrulline	Amino acids	17	9,733,449	rs7207982	rs8146978	t	c	0.21	30,940	2.430	0.01
Citrulline	Amino acids	17	9,743,073	rs74739416		a	c	0.02	23,615	-2.010	0.04
Citrulline	Amino acids	17	9,743,127	rs77044028		t	c	0.02	23,615	-1.992	0.04
Citrulline	Amino acids	17	9,746,361	rs7202006		a	g	0.39	23,615	2.402	0.01
Citrulline	Amino acids	17	9,746,499	rs77856432		a	g	0.04	23,615	-2.204	0.02
Citrulline	Amino acids	17	9,746,503	rs7775577		a	g	0.04	23,615	-2.118	0.02
Citrulline	Amino acids	17	9,748,738	rs12950647		t	c	0.19	23,615	-2.417	0.01
Citrulline	Amino acids	17	9,747,679	rs59968881	rs82066030	t	c	0.40	23,615	2.448	0.01
Citrulline	Amino acids	17	9,748,647	rs82066031		t	c	0.40	23,615	2.549	0.01

<https://omicscience.org>

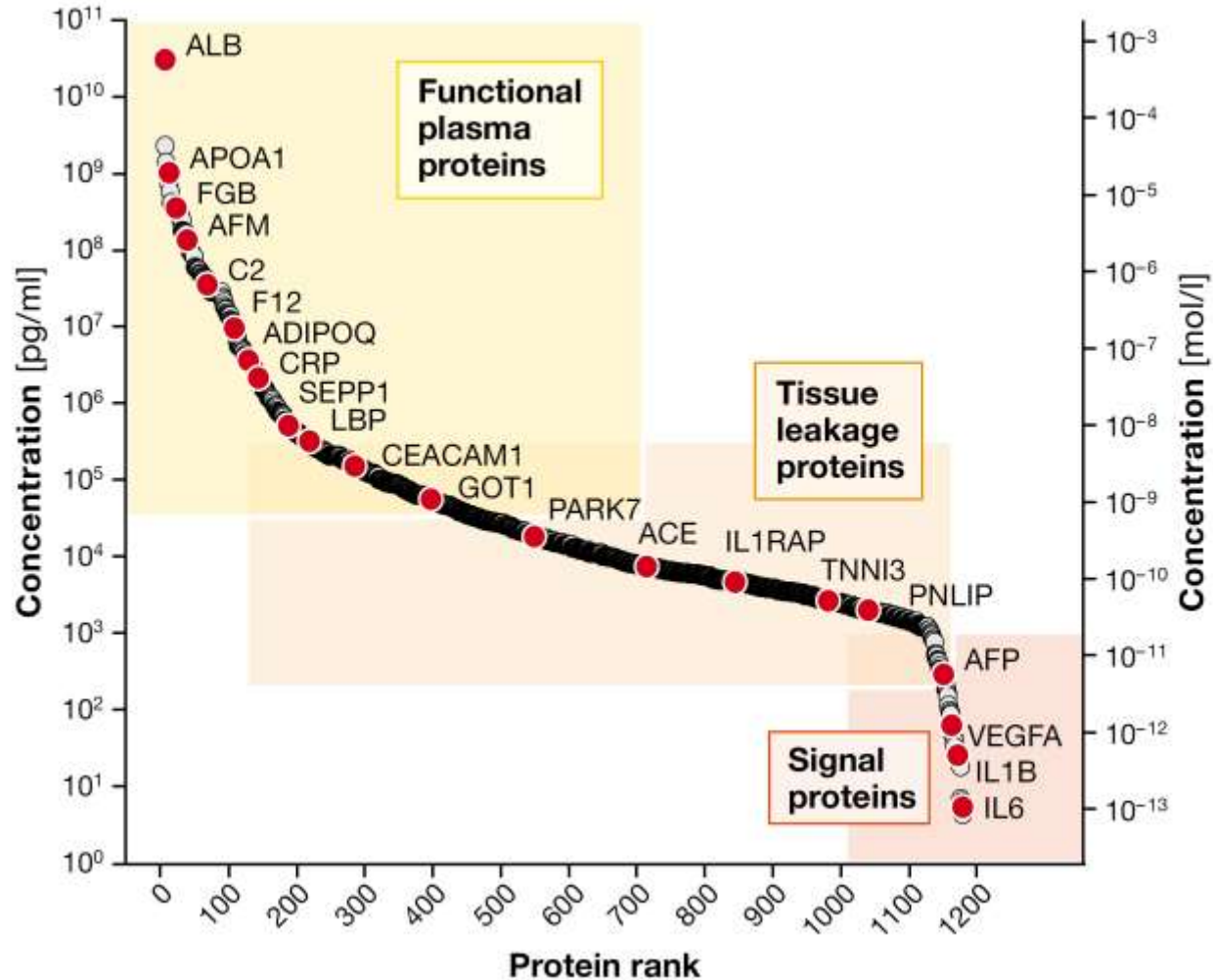
[https://biocrates.com/2021\\_cohort\\_webinar](https://biocrates.com/2021_cohort_webinar)



# Genome to phenome via 'omics'



# The plasma proteome



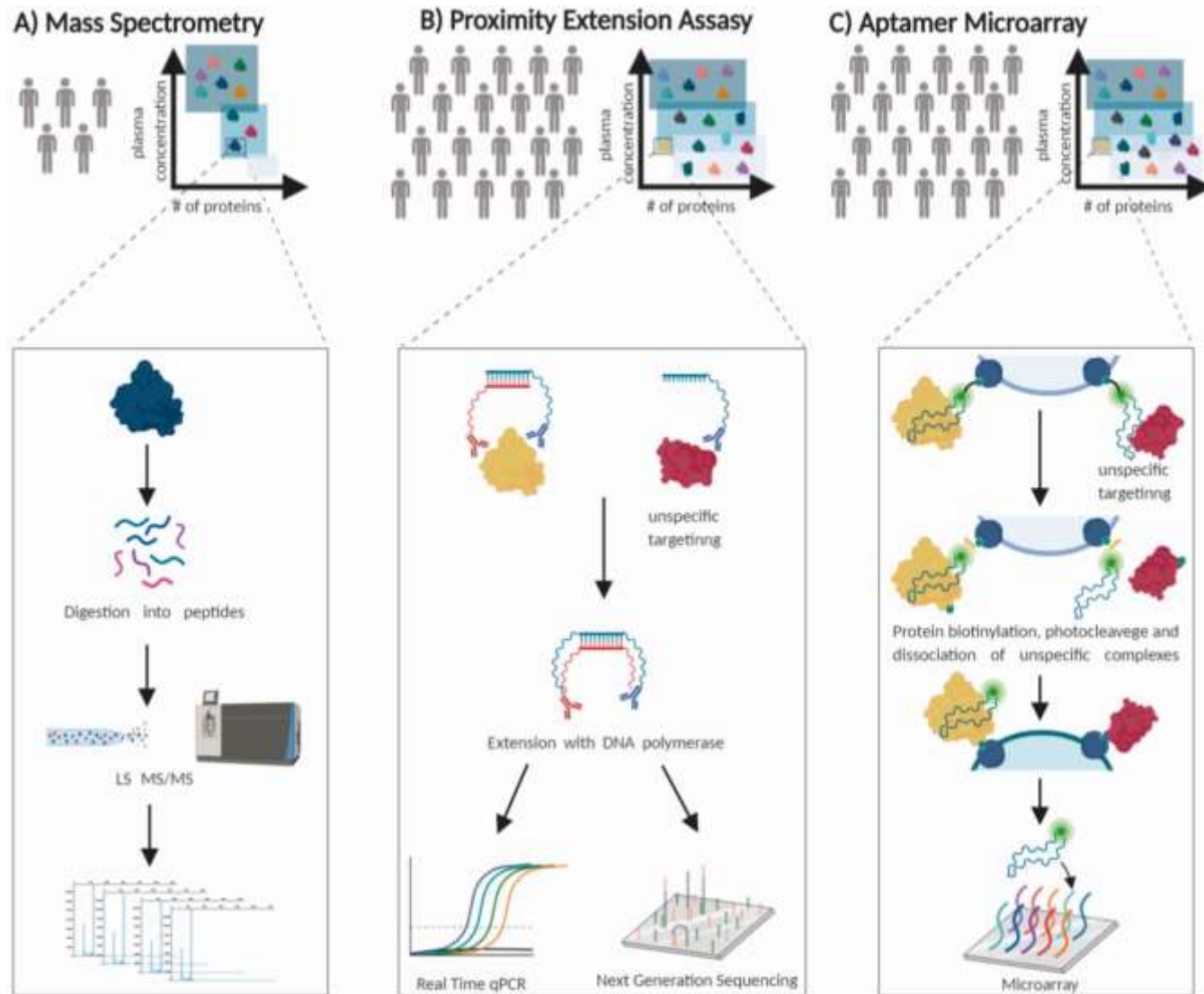
Key roles in diverse biological processes, dysregulated in disease, important drug targets

Concentrations of plasma proteins span almost **10 orders of magnitude**

Huge dynamic range

No single technique is currently able to provide reliable measurements for all proteins

# From proteins to 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

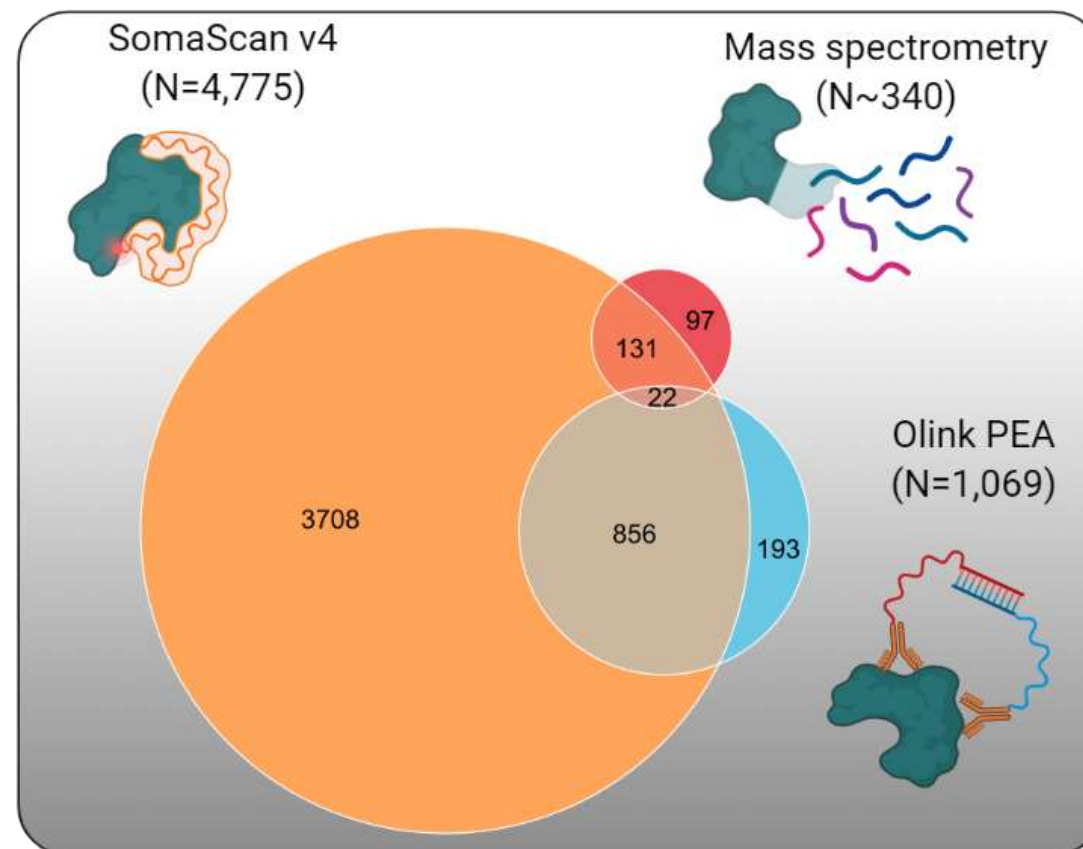
# Population proteomics

**12,435 participants** born 1950-75 and living in Cambridgeshire

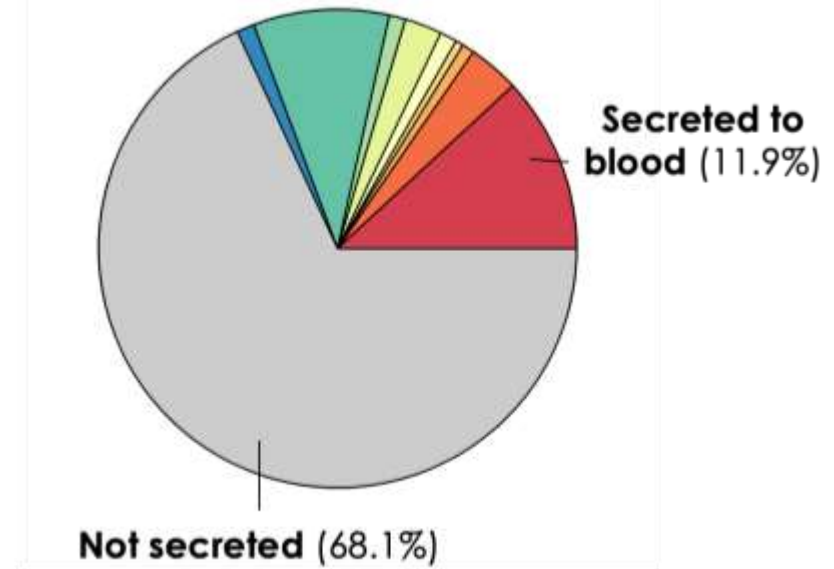
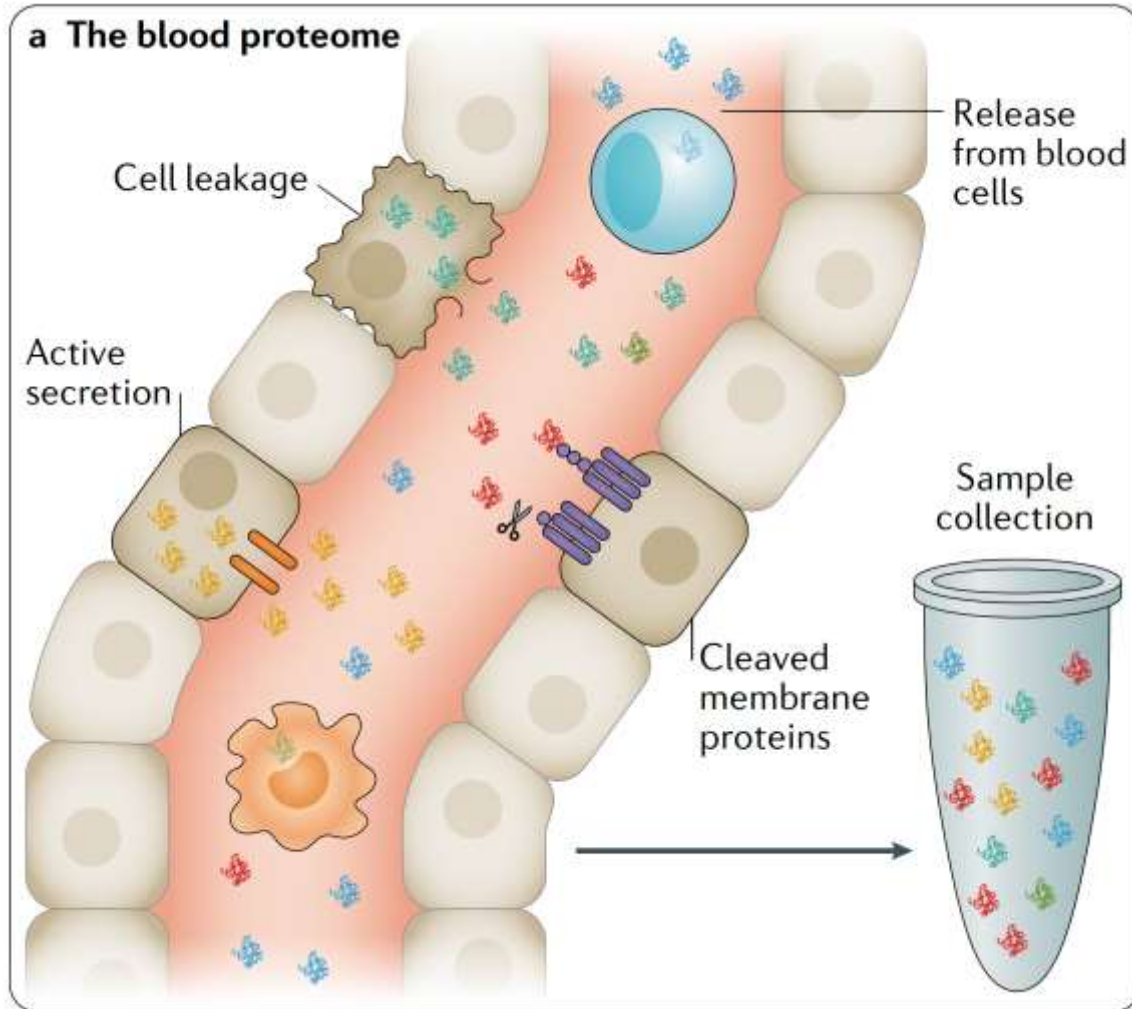
**10,708 genotyped**

Proteomics data

- **SomaScan v4** (~5,000 aptamers, N=12,435)
- **Olink** (12 panels; ~1,100 proteins, N=485)
- **MS-proteomics** (M. Ralser, Scanning SWATH ~340 proteins, N=485, target >12k)



# SomaScan v4: 4,775 protein targets



## Actively secreted proteins

- Coagulation factors
- Cytokines

Products of **cell leakage and turnover**

**Soluble fragments** of membrane proteins

## Plasma protein patterns as comprehensive indicators of health

Stephen A. Williams <sup>1,12\*</sup>, Mika Kivimaki <sup>2</sup>, Claudia Langenberg <sup>3</sup>, Aroon D. Hingorani<sup>4,5,6</sup>, J. P. Casas<sup>7</sup>, Claude Bouchard <sup>8</sup>, Christian Jonasson<sup>9</sup>, Mark A. Sarzynski<sup>10</sup>, Martin J. Shipley<sup>2</sup>, Leigh Alexander<sup>1</sup>, Jessica Ash<sup>1</sup>, Tim Bauer<sup>1</sup>, Jessica Chadwick<sup>1</sup>, Gargi Datta <sup>1</sup>, Robert Kirk DeLisle<sup>1</sup>, Yolanda Hagar<sup>1</sup>, Michael Hinterberg<sup>1</sup>, Rachel Ostroff<sup>1</sup>, Sophie Weiss<sup>1</sup>, Peter Ganz<sup>11,12</sup> and Nicholas J. Wareham<sup>3,12</sup>

# Isolated post-challenge hyperglycaemia

Elevated glucose 2-hours after an oral glucose load (IGT)

Strongly predictive of cardiometabolic diseases

Rarely measured (complexity, logistics, time)

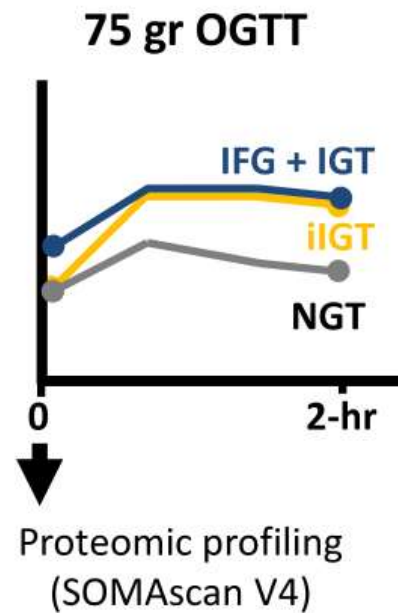
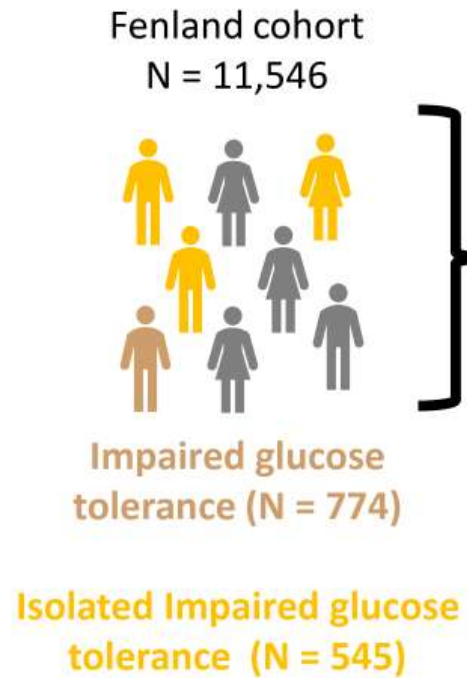
Very common (globally ~7.5% of adults)

**Isolated** post-challenge hyperglycaemia is missed by FPG, HbA1c

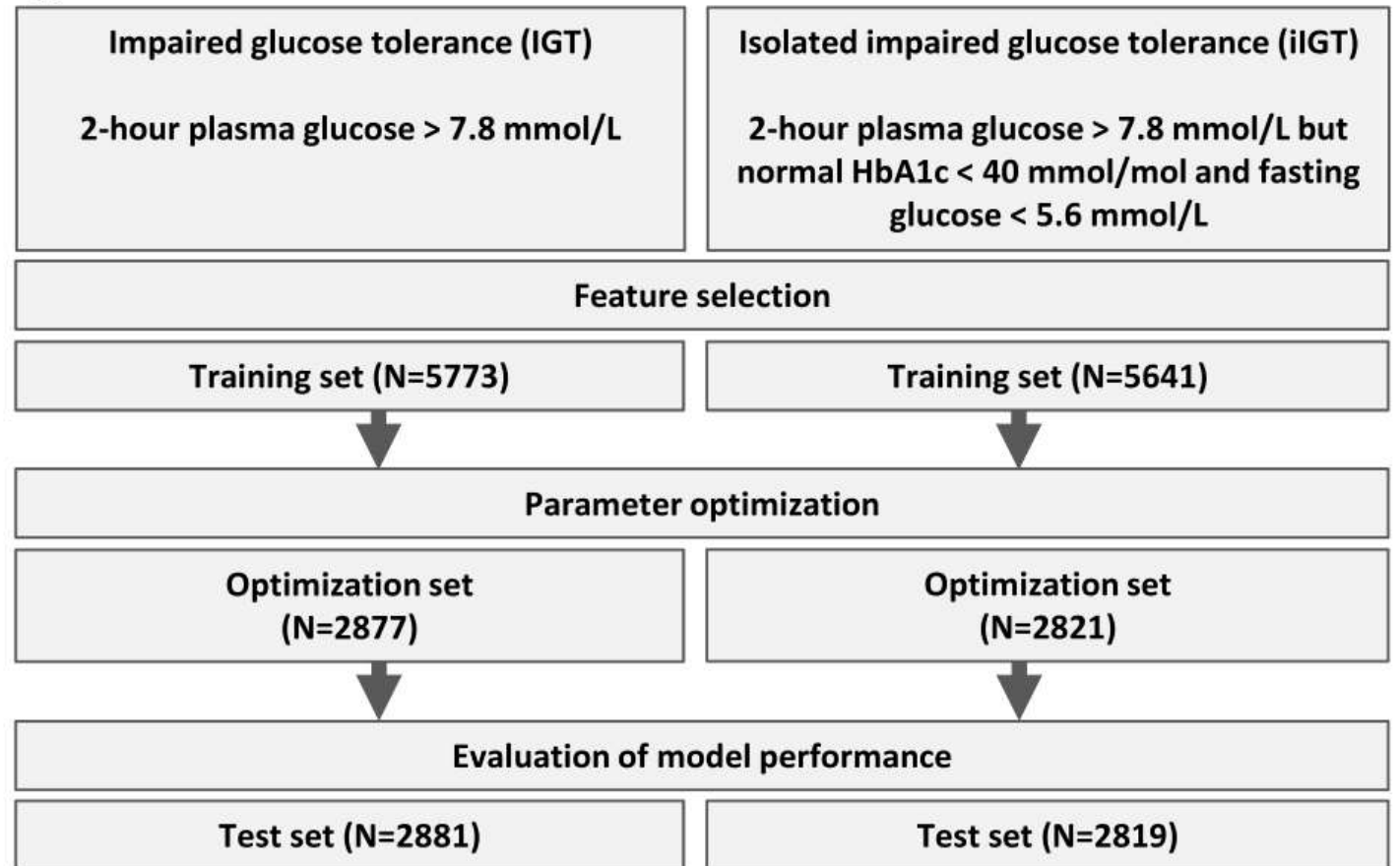
**Question: Is it feasible to identify a (fasting) proteomic signature to design a simple test that predicts isolated IGT?**

# Predicting post-challenge hyperglycaemia

a



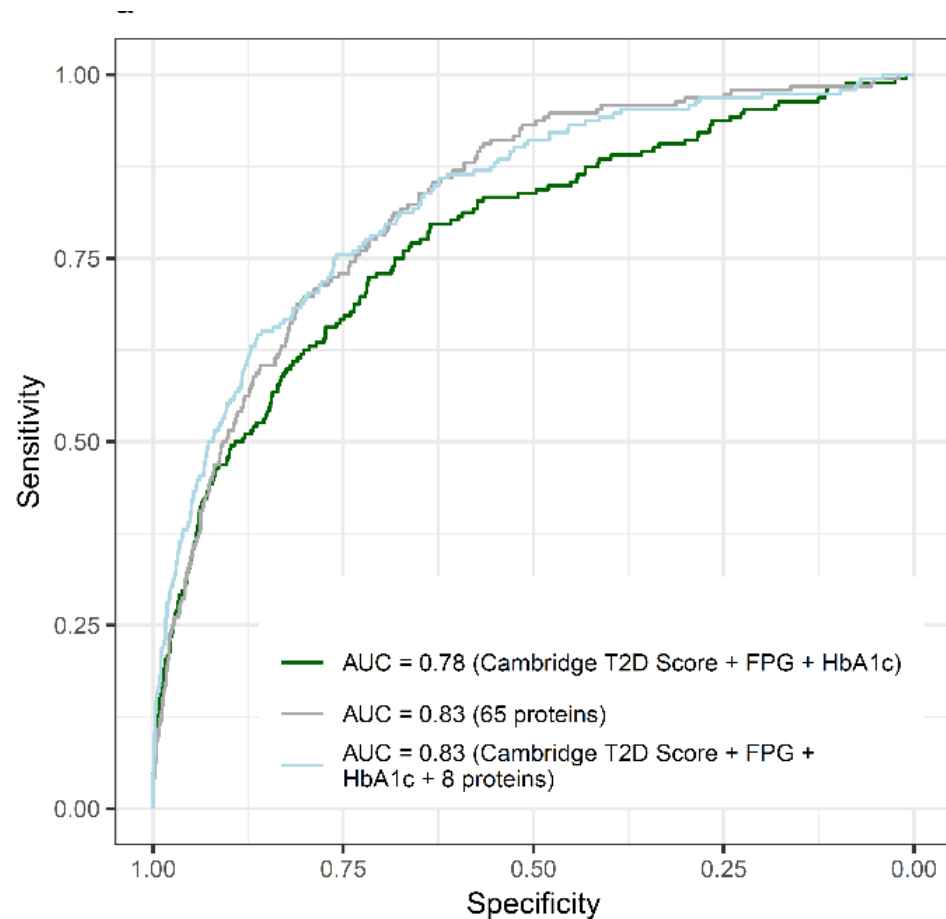
b



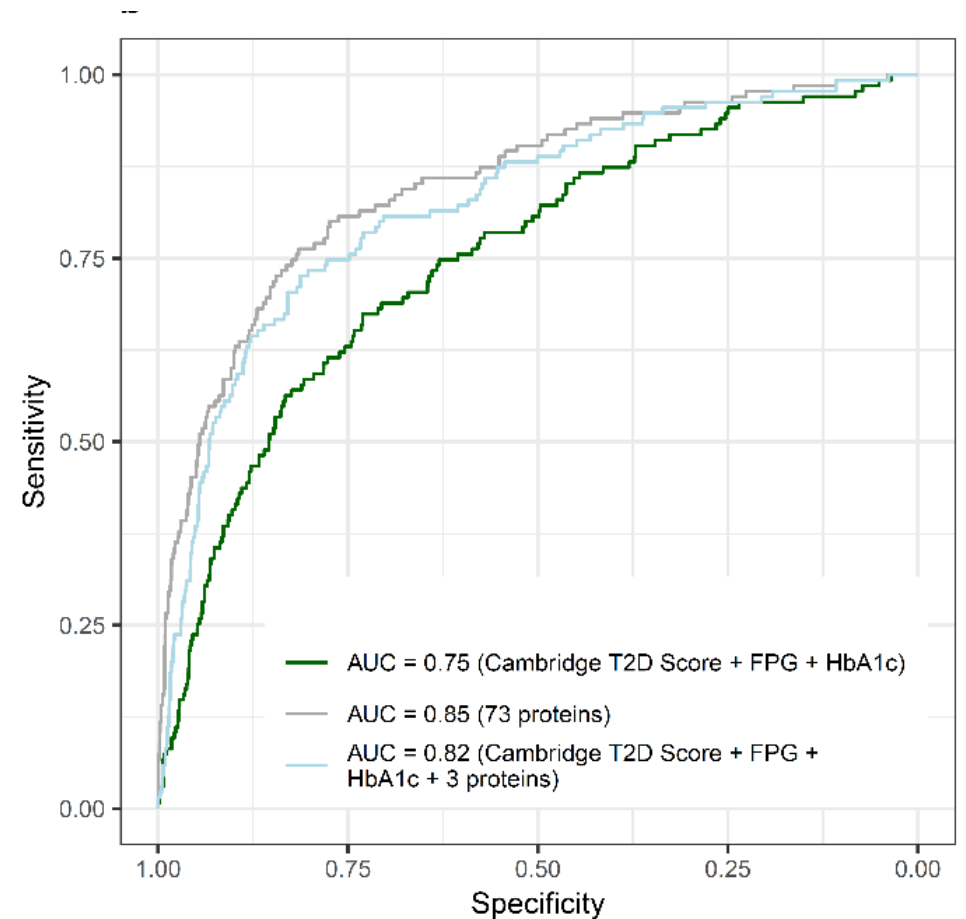


# Prediction performance

## IGT

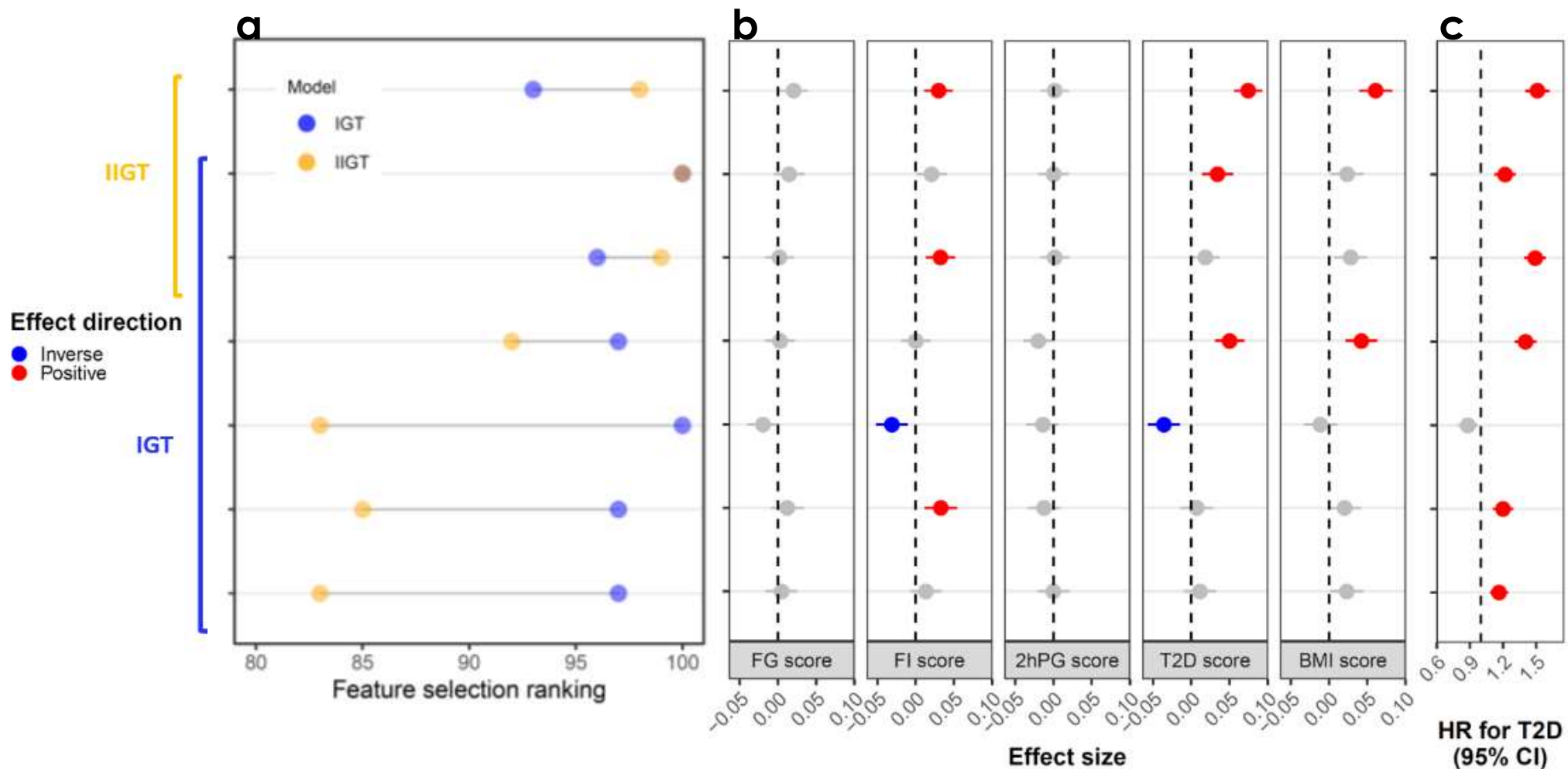


## Isolated IGT

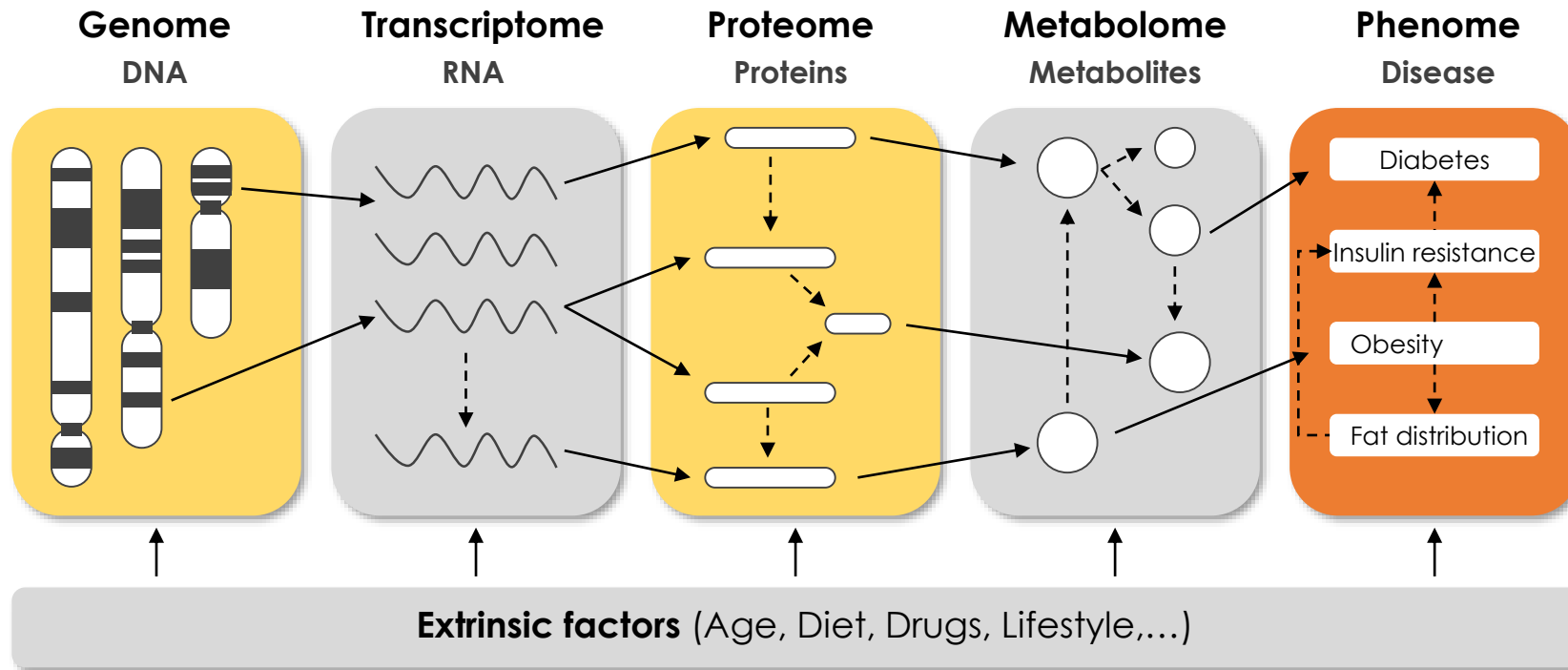


# Top selected proteins from prediction models

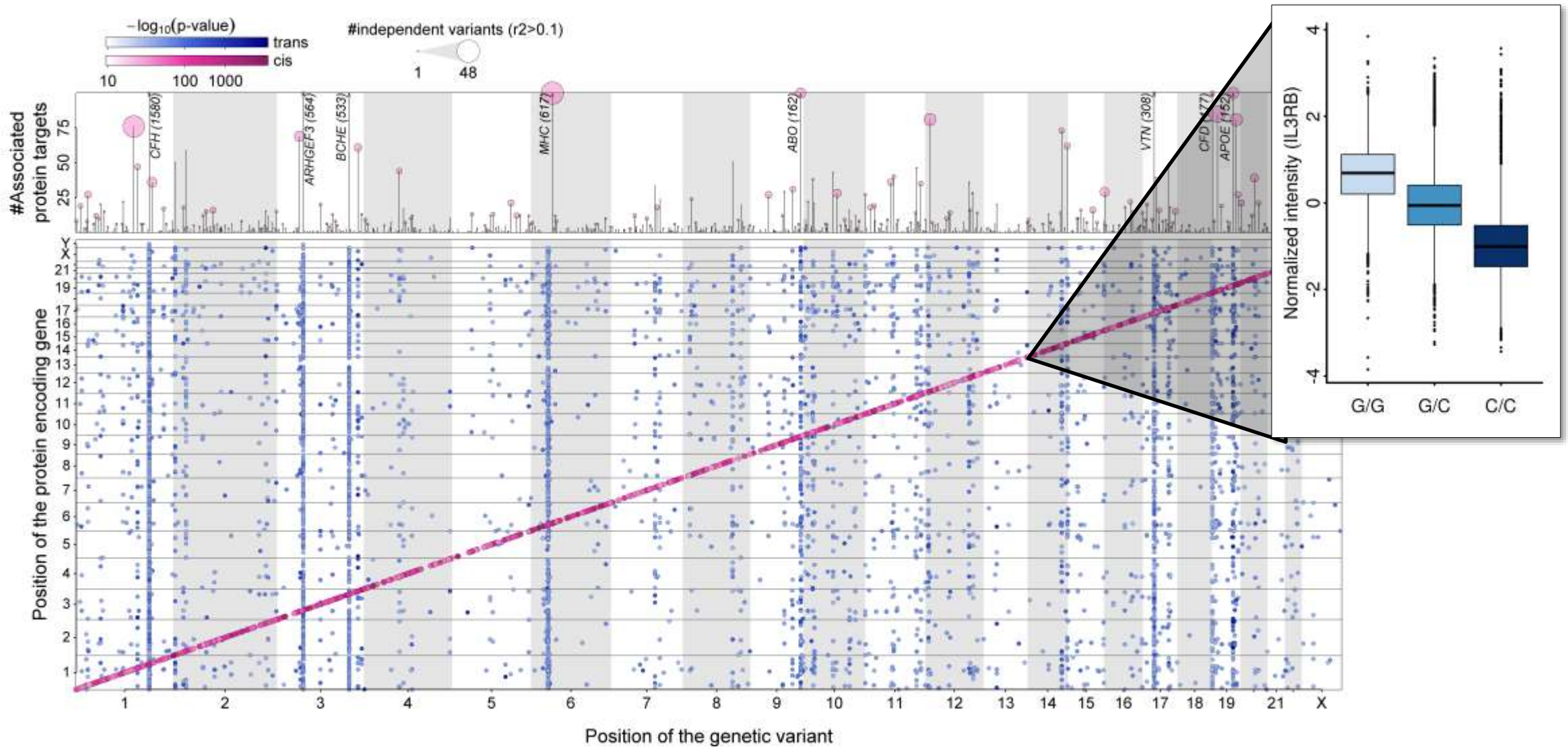
- Associated with genetic susceptibility to impaired glucose homeostasis
- Associated with risk to develop T2D in an independent cohort



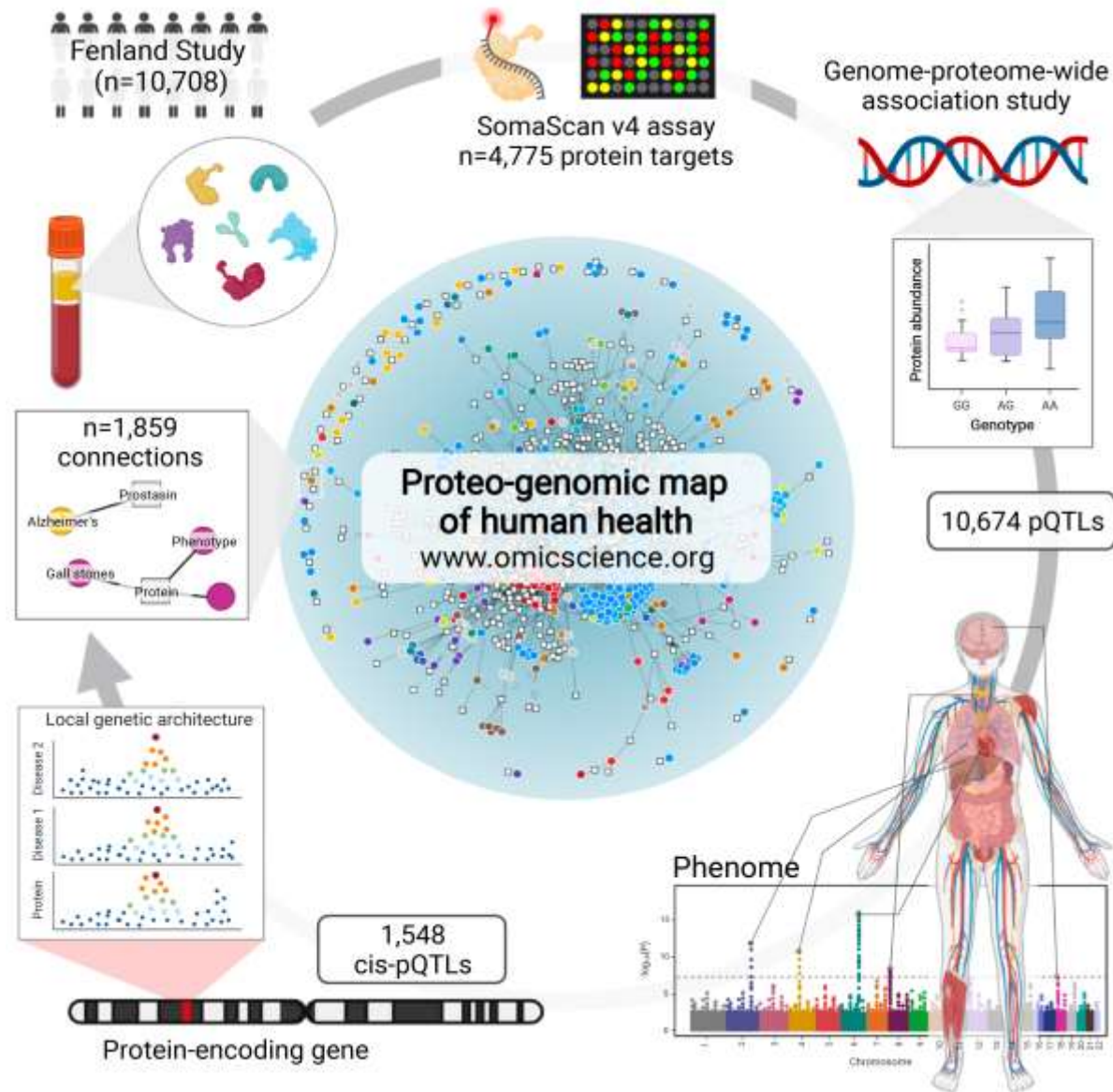
# Genome to phenome via 'omics'



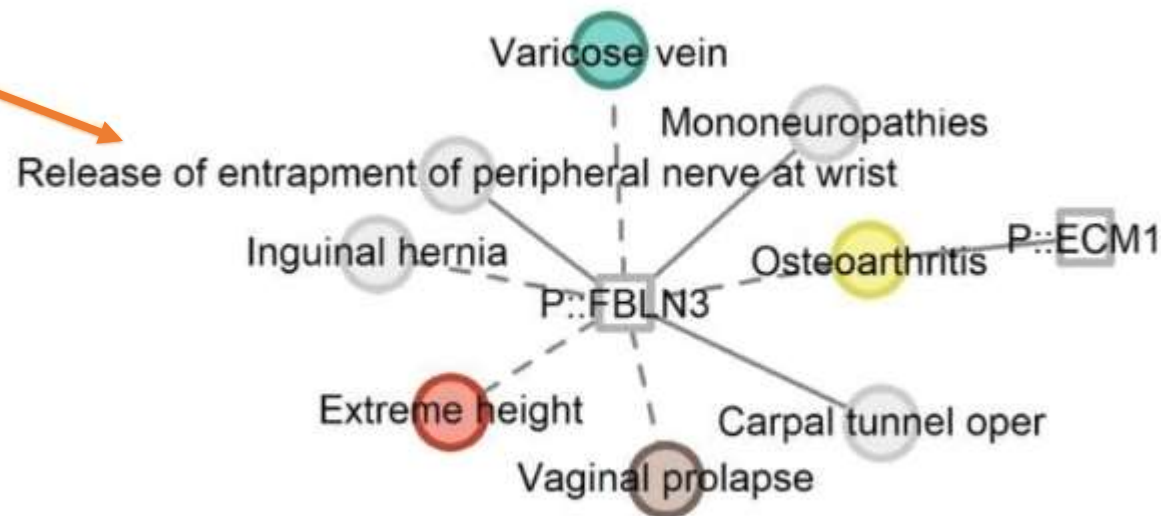
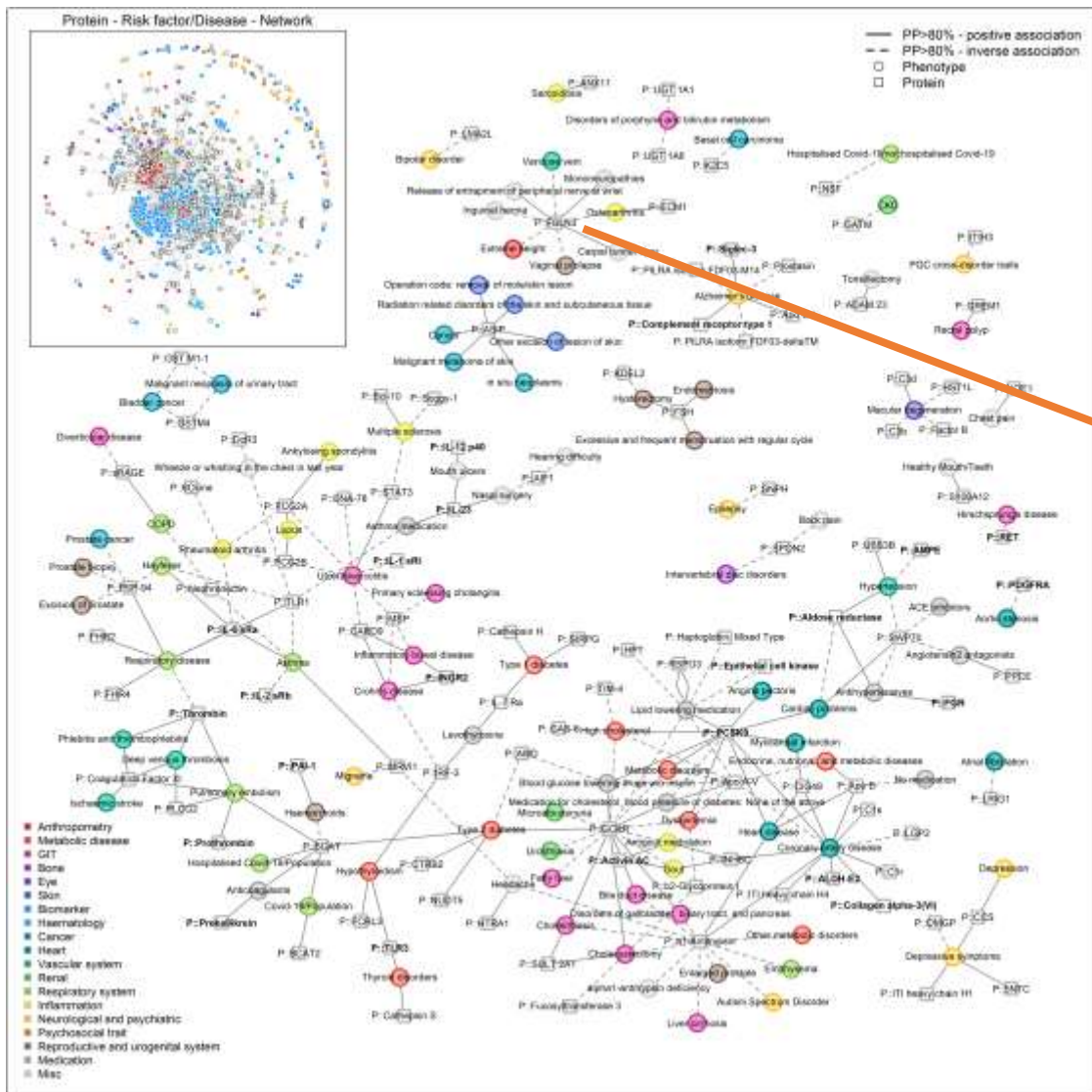
# Protein quantitative trait loci



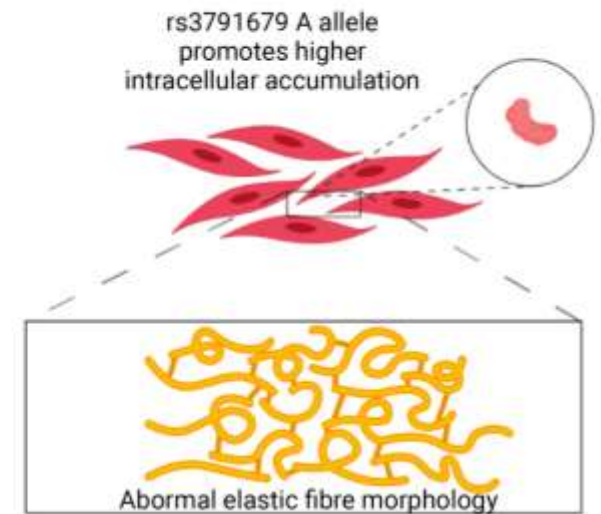
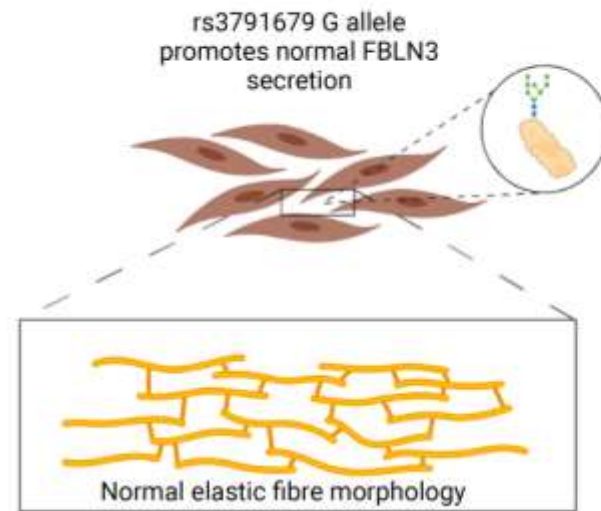
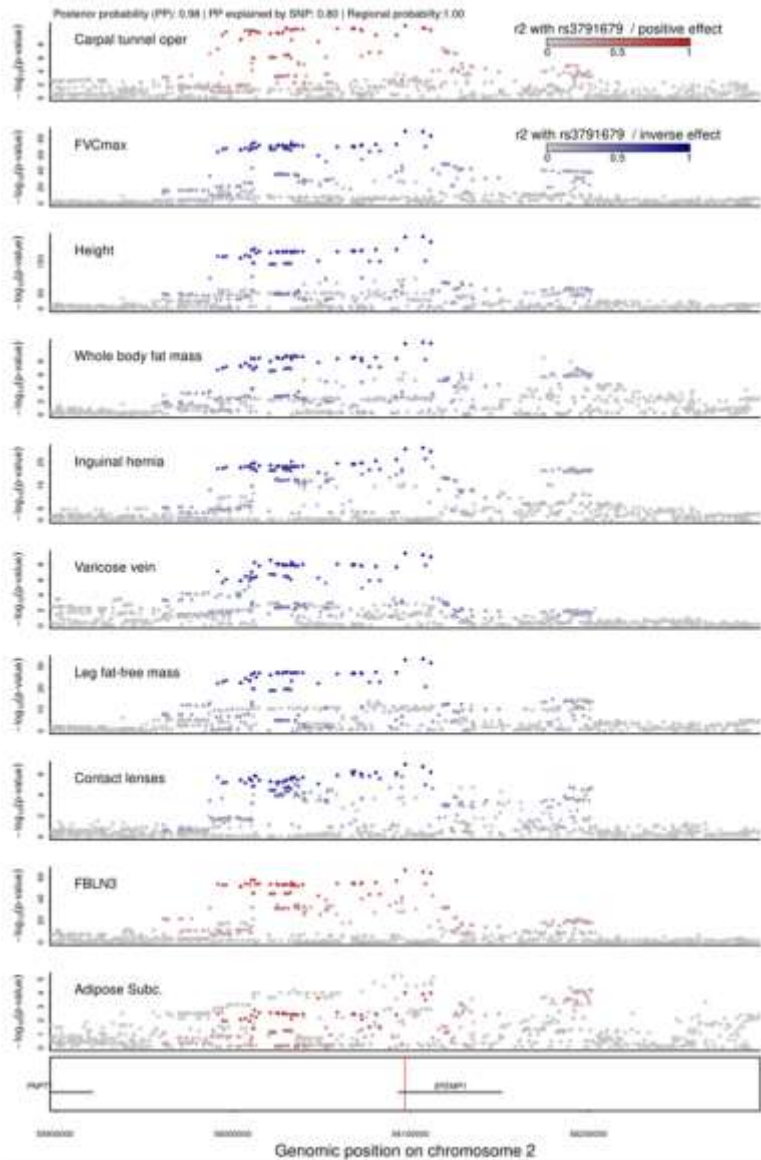
# Proteo-Genomic Map of the Human Phenome



# Convergence of soft tissue disorders: *EFEMP1*



# Convergence of soft tissue disorders: *EFEMP1*



# Summary (II of II)

- Value of the plasma proteome for prediction: isolated IGT
- Value of increased breadth and scale
- Integration with phenomic data: genetically anchored disease map
- Next stages
  - 'Neglected' phenome
  - Covid19 prognosis
  - Clinical curation



**OMICSCIENCE**

Pietzner et al., Nat. Comm., 2020

**Genetic architecture of host proteins involved in SARS-CoV-2 infection**

The summary statistics related to this publication are freely available without restrictions.

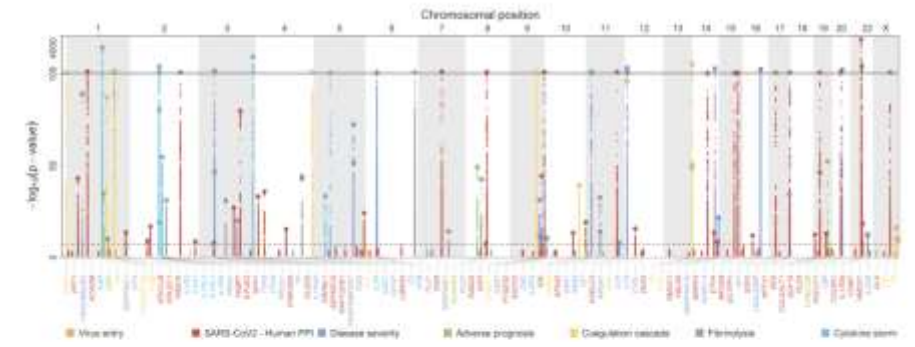
By using these results in your research, you agree to [cite our publication](#).

[Developed by Johannes Raffler](#)

[Legal notice](#)

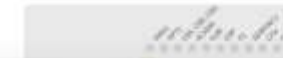
[Data Protection Statement](#)

Website template by [HTML5 UP](#)



Manhattan plot of cis-associations statistics (encoding gene  $\pm 500$ kb) for 179 proteins. The most significant regional sentinel protein quantitative trait loci (pQTL) acting in cis are annotated by larger dots for 104 unique protein targets ( $p < 5 \times 10^{-8}$ ). Starred genes indicate those targeted by multiple aptamers.

## Data access



interactive matrix and annotations for all cis-loci and associated aptamers.

<https://omicscience.org>

<https://lifesciences.somalogic.com/webinar/liquid-health-check/>



**Computational Medicine, BIH:** Maik Pietzner

### **Diabetes Aetiology Group (MRC Epidemiology)**

Nick Wareham, Jian'an Luan, Eleanor Wheeler, Isobel Stewart, Maik Pietzner, Erin Oerton, Rupal Shah, Nicola Kerrison, Julia Carrasco-Zanini Sanchez, Mine Koprulu, (graduates Laura Wittemans, Chen Li, Vicky Au Yeung, Nicholas Bowker, Lina Cai)

### **MRC Epidemiology Colleagues**

Fumiaki Imamura, Tom White, Nita Forouhi, John Perry, Felix Day, Ken Ong, Debora Lucarelli

### **University of Cambridge**

Kay-Tee Khaw, Jules Griffin, Albert Koulman

### **MRC Biostatistics Unit**

Paul Newcombe, Stephen Burgess

### **Institute of Metabolic Science**

Fiona Gribble, Frank Reimann, Stephen O'Rahilly, David B. Savage, Sadaf Farooqui

### **Study PIs and Functional and Laboratory Teams**

Fenland, EPIC-Norfolk, InterAct, Interval, UK Biobank

**Funding:** UK MRC, EU FP7 (EMIF-Metabolic), Pfizer, GSK, NovoNordisk, Somalogic

### **mGAP (metabolome Genetic Architecture Programme) Team**

Isobel Stewart, Luca Lotta, Laura Wittemans, Chen Li, Vicky Au Yeung, Nick Wareham (MRC Epidemiology Unit)

Praveen Surendran, Adam Butterworth, Angela Wood, Joanna Howson, John Danesh (Cardiovascular Epidemiology Unit)

Karsten Suhre (Weill Cornell Medicine, Qatar)

Gabi Kastenmüller, Johannes Raffler, Kieu Do, (Helmholtz, Munich)

Nicole Soranzo, Lorenzo Bomba (Wellcome Sanger Institute)

Cristina Menni, Jonas Zierer, Niccolò Rossi, Alessia Visconti, Mario Falchi, Tim Spector (King's College London)

Eric Gamazon (Vanderbilt University, Nashville)

Eric Fauman (Pfizer, Boston)

### **pGAP (proteome Genetic Architecture Programme) Team**

Aroon Hingorani, Chris Finan (UCL)

### **Collaborators**

Melanie Bahlo, Roberto Bonelli and the MacTel Consortium

Robert Scott, Adrian Cortez (GSK)

Greg Michelotti, Annie Evans (Metabolon)

Steve Williams, Rachel Ostroff (Somalogic)

Thank you!

[www.bihealth.org](http://www.bihealth.org)