People-Centred AI
UNIVERSITY OF SURREY

UNIVERSITY OF SURREY

Polygenic Scores for Type 2 Diabetes and High Blood pressure comorbidity prediction

Полігенні шкали для прогнозування коморбідності діабету 2 типу та високого кров'яного тиску

Prof Inga Prokopenko, PhD

Professor of e-One Health, Vice-Chancellor’s Distinguished Chair

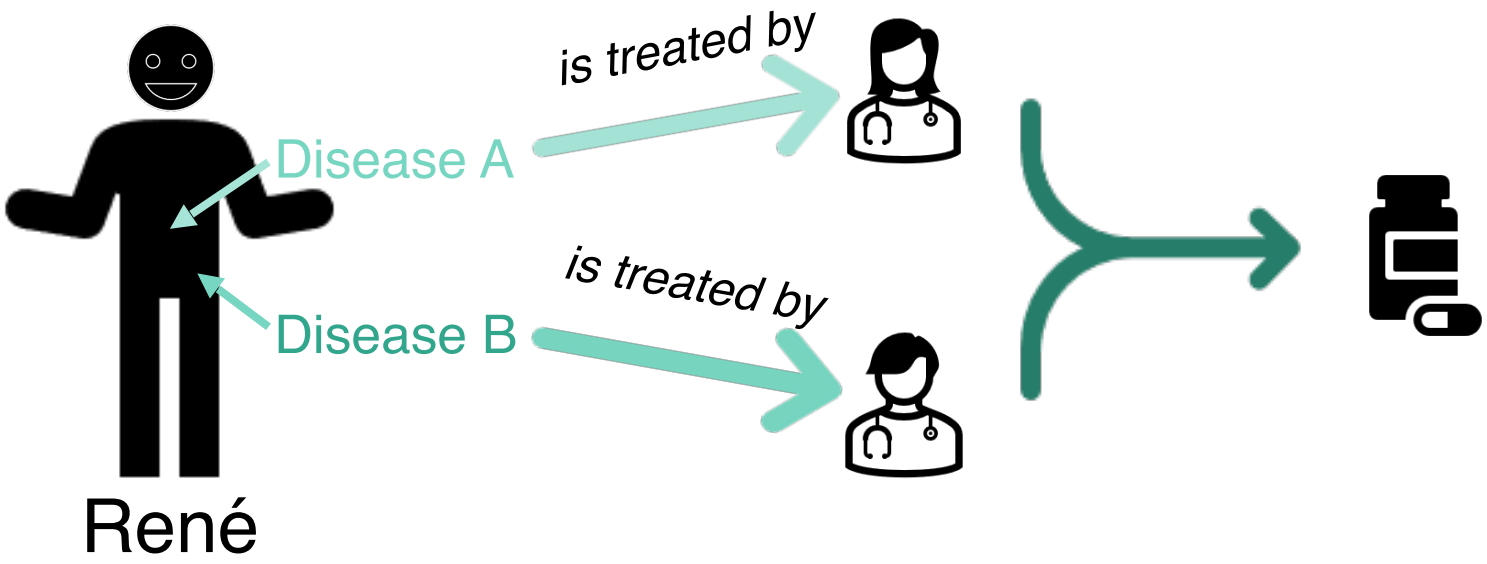
Head, Section of Statistical Multi-Omics

Co-Director, Surrey People-Centred AI Institute

University of Surrey, UK

Marzieiev Readings, 23-24 October, 2025, Kyiv

Precision medicine as a solution

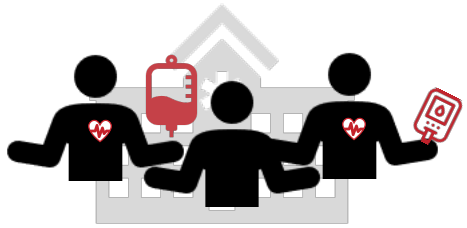


Precision medicine

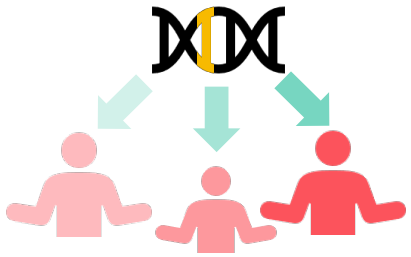
+ better understanding of the biological mechanisms

Genetics as an additional data to apply precision medicine

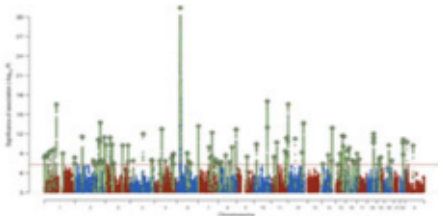
Problem statement



Comorbidity and complication are **major threats** on our healthcare systems



Using data available to do **stratification of risks and tailored treatment** can improve it, and **genetic data** offers additional and novel information



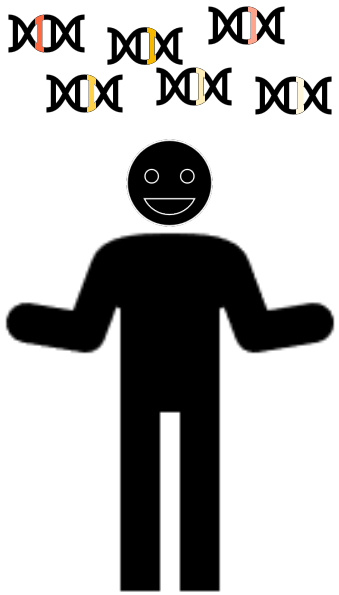
Genome-wide investigation can illuminate underlying **biological mechanisms of complex conditions**

Polygenic score (PGS), or genetic information as the additive sum of all genetic variants

$$PGS_{René} = \sum \beta_{ij} \times SNV_i$$

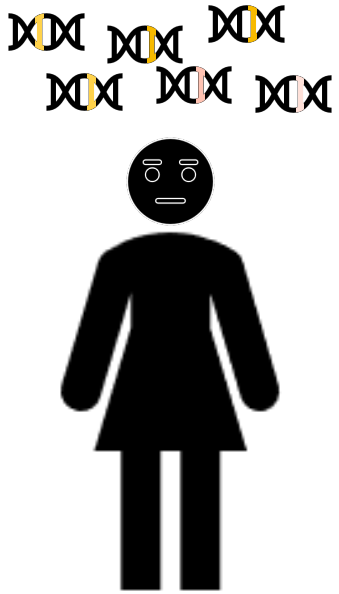
PGS: sum of **all known (common) variants**, usually Single Nucleotide Variants (SNVs) to calculate a **genetic risk of getting a particular disease, j**

Calculating PGS within a population…



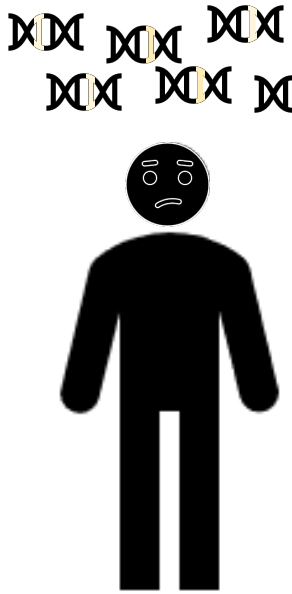
René

$$PGS_{René,\,j} = \sum_{i \in René} \beta_{ij} \times SNV_i$$



Françoise

$$PGS_{Françoise,\,j} = \sum_{i \in Françoise} \beta_{ij} \times SNV_i$$

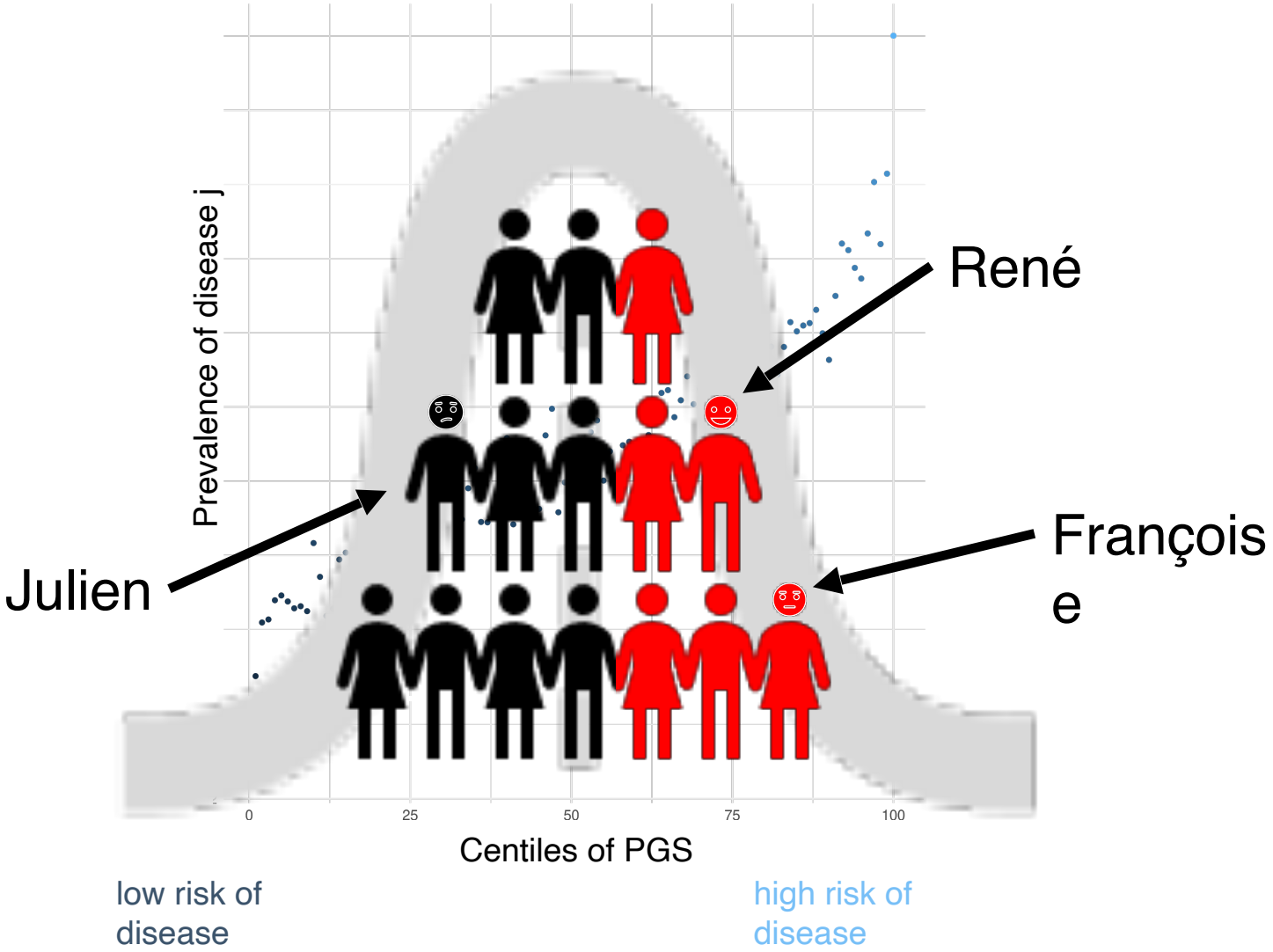


Julien

$$PGS_{Julien,\,j} = \sum_{i \in Julien} \beta_{ij} \times SNV_i$$

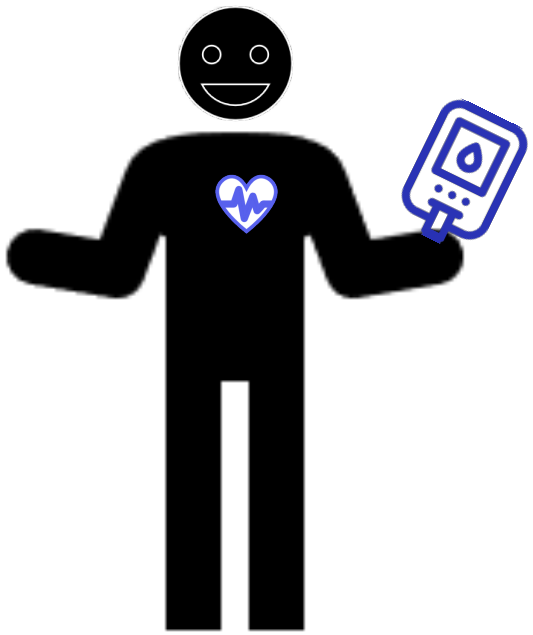
PGS is a relative number that needs to be compared within a **genetically closed population**

… allows to identify individuals at risk



The PGS distribution helps to identify **individuals** at **higher risk** of a **particular disease j**

The T2D – high BP project aims with René



René

René has **T2D and hypertension**

Like others in his situation, he is **struggling to reach optimal levels of glucose and BP**

Could we have **predicted René's comorbidity problem**, and **tailored a medical monitoring + treatment**?

What are the shared pathogenetic mechanisms between T2D and high BP?

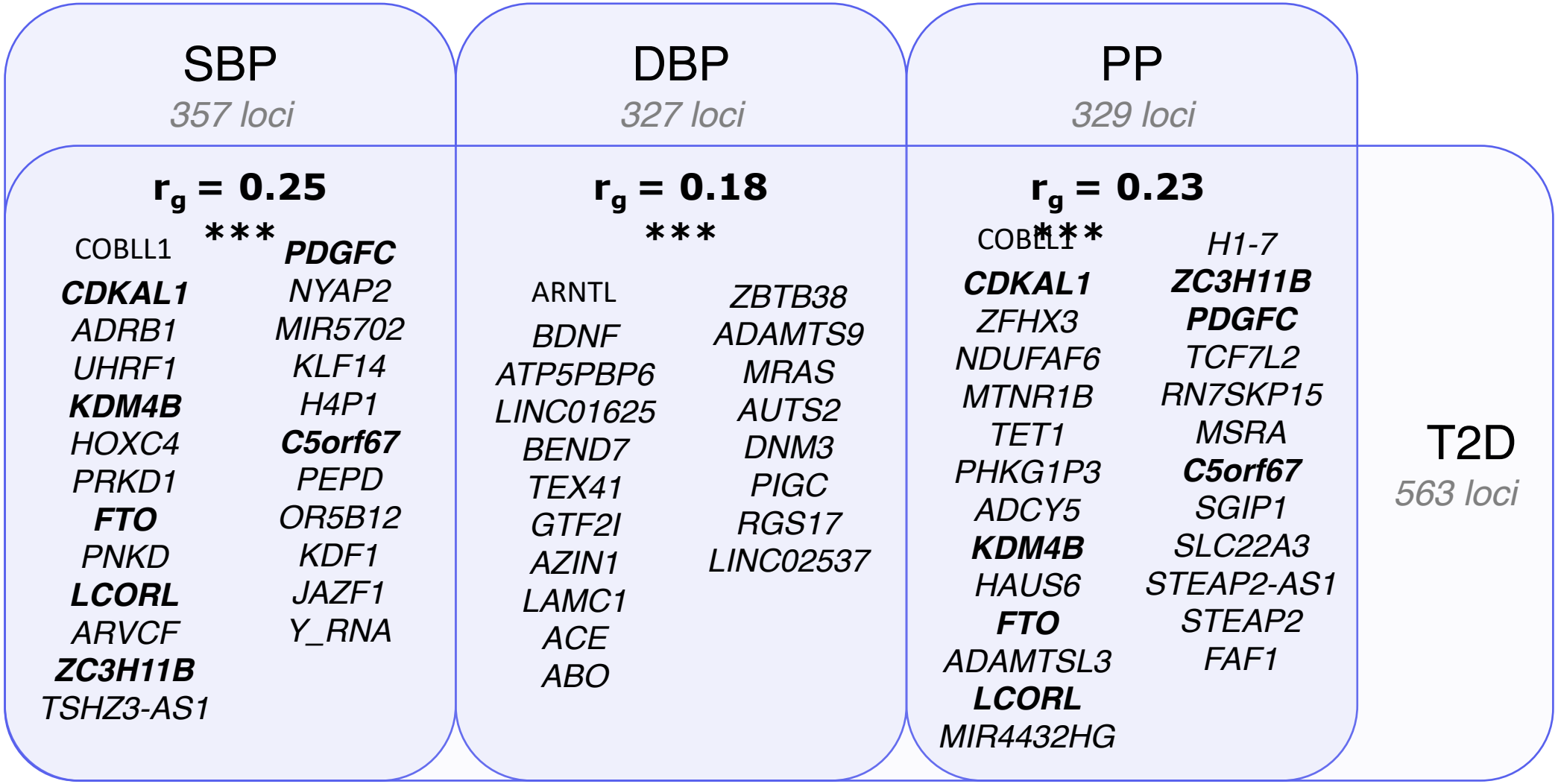
Material and methods

- A large-scale multiomic cohort of 460,000 individuals, the biobank uk

- GWAS of
  - T2D
  - High BP: SBP, DBP, PP

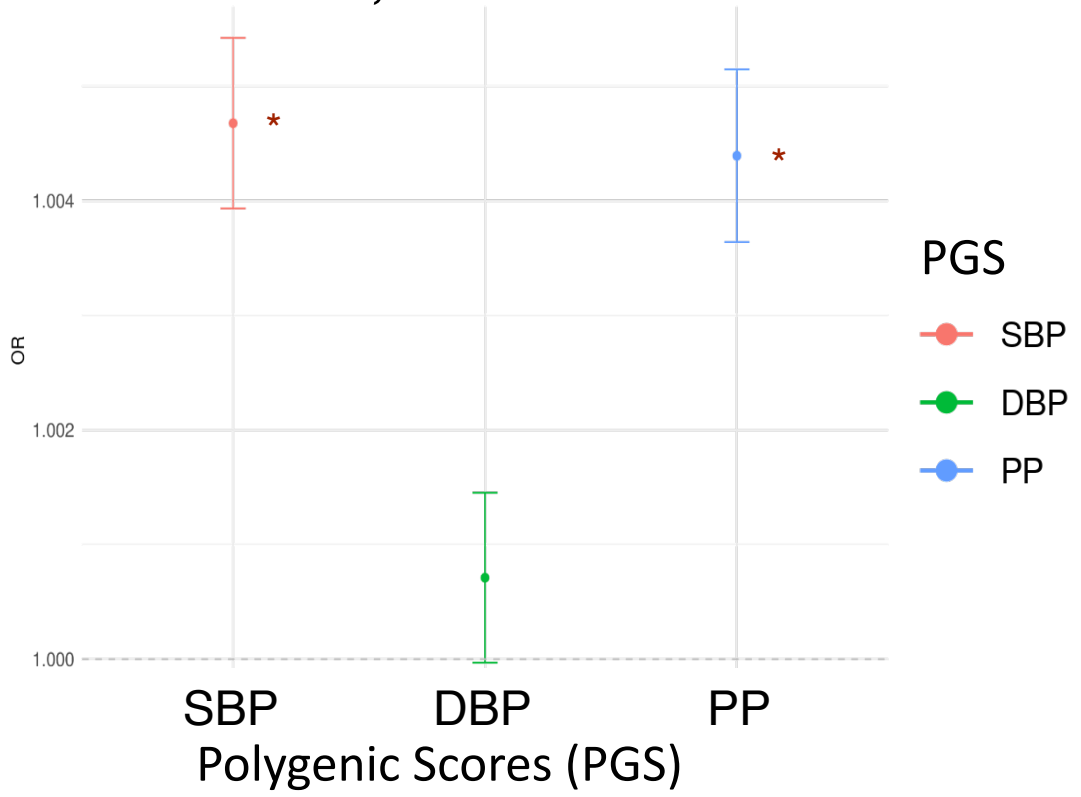
- Latest tools for genomic investigation
  - PGS, via *comorbidPGS*
  - Mendelian Randomization (MR)
  - Colocalisation
  - single-cell ATAC sequencing (scATAC-seq)

Results: High T2D-BP genetic correlation and overlap

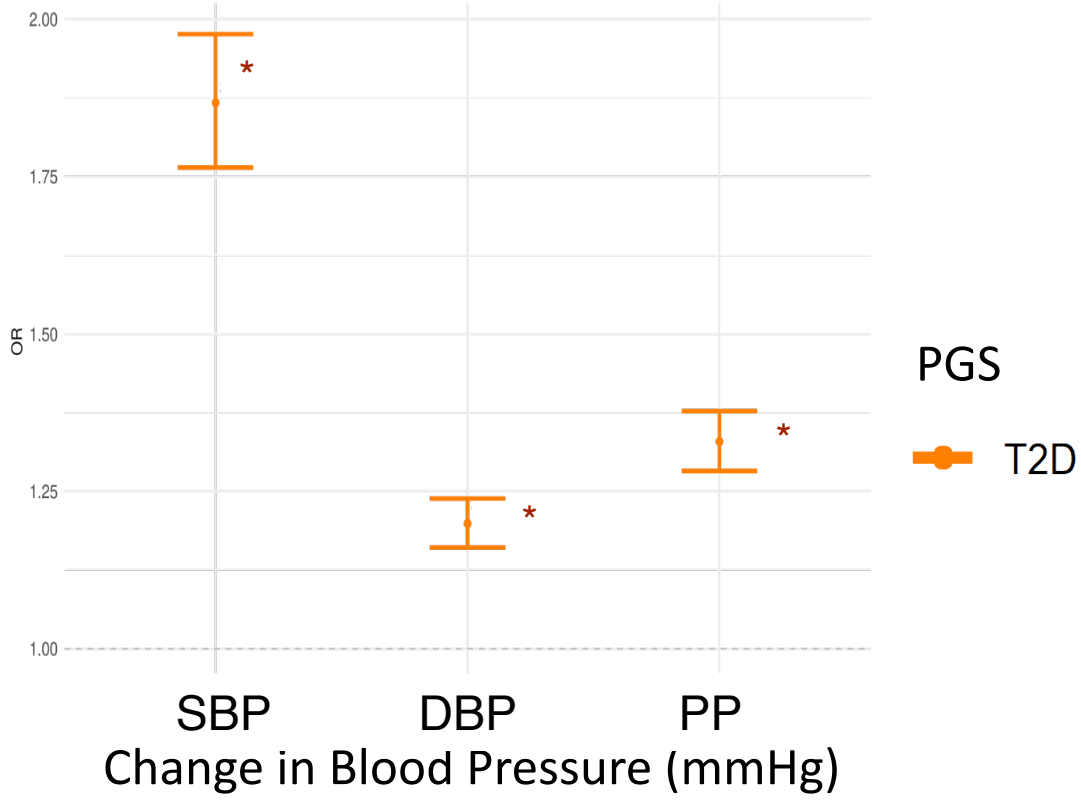


Results: Reciprocal T2D-BP risk prediction using polygenic scores

a. *Association between high BP PGS and risk of T2D, 95% CI*



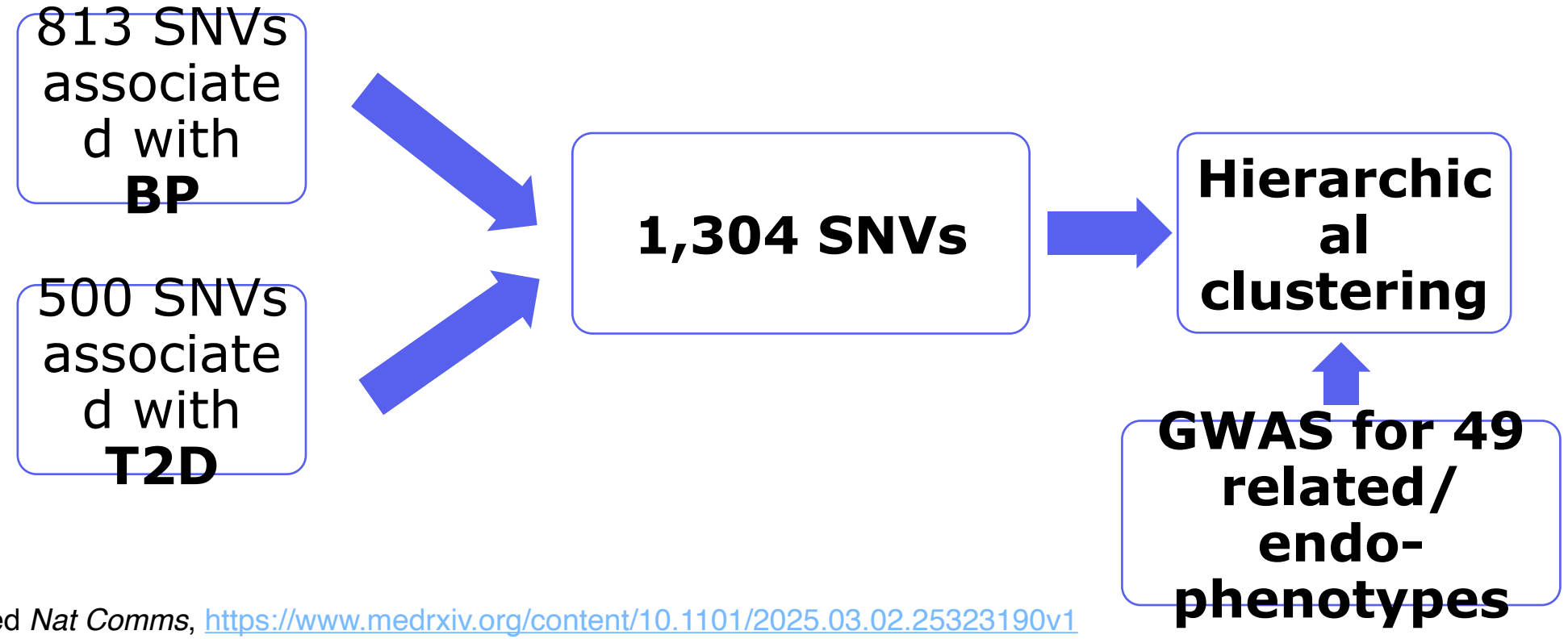
b. *Association between T2D PGS and high BP, 95% CI*



Pascat et al, accepted *Nat Comms*, https://www.medrxiv.org/content/10.1101/2025.03.02.25323190v1

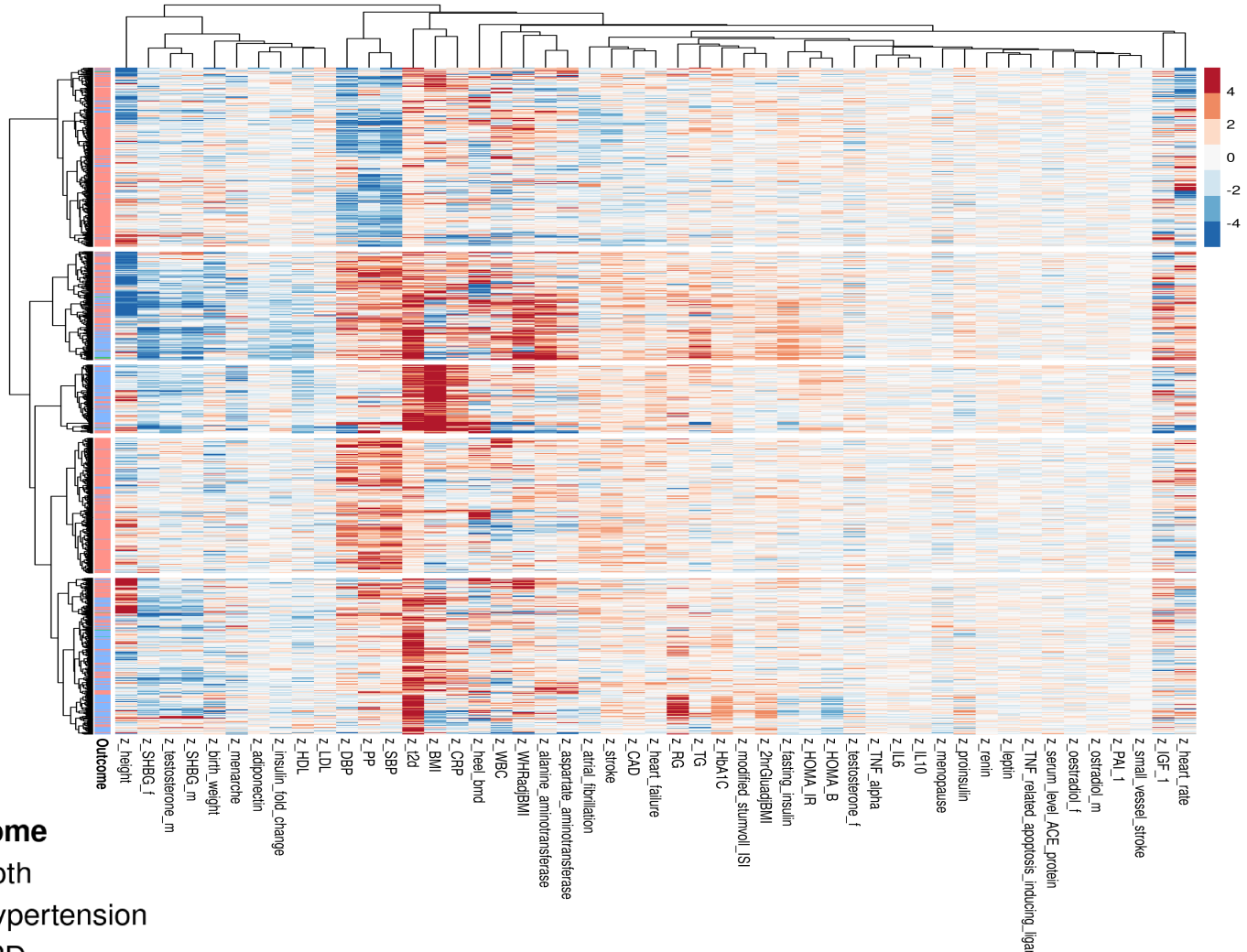
Clustering of T2D-BP SNVs into groups of different pathogenetic processes

We grouped **1,304 SNVs of T2D and high BP** using **49 endophenotype GWAS traits**



Pascat et al, accepted *Nat Comms*, https://www.medrxiv.org/content/10.1101/2025.03.02.25323190v1

More details on the T2D-BP clustering of SNVs



*Cluster*1 – **Inverse T2D-BP risk**

*Cluster*2 - **Metabolic Syndrome (WHR, IR, shorter stature)**

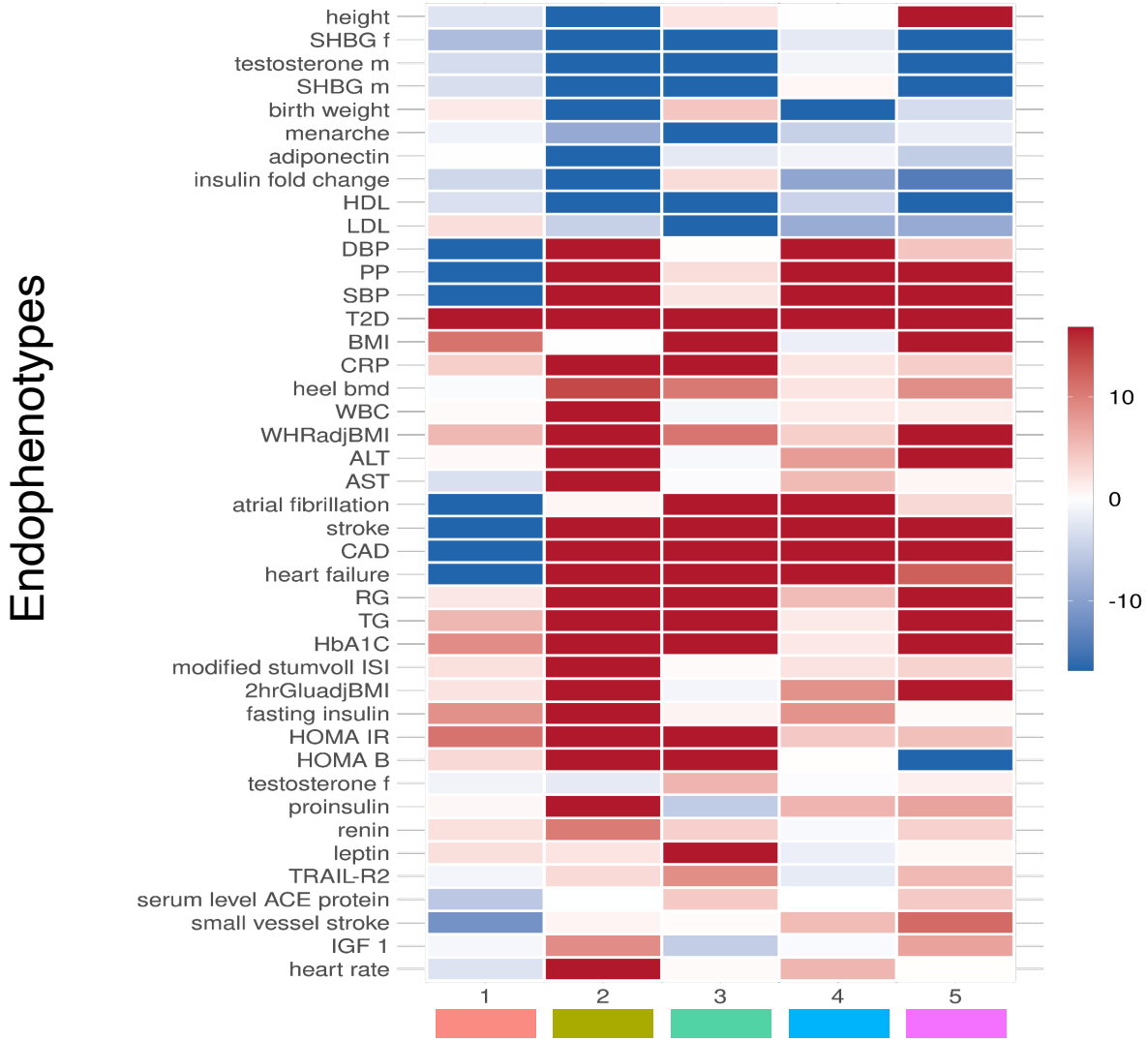
*Cluster*3 – High adiposity

*Group*4 – Vascular dysfunction

*Group*5 – Reduced beta-cell function

Pascat et al, accepted *Nat Comms*, https://www.medrxiv.org/content/10.1101/2025.03.02.25323190v1

Results: Clustering of T2D-BP SNVs into groups of different pathogenetic processes

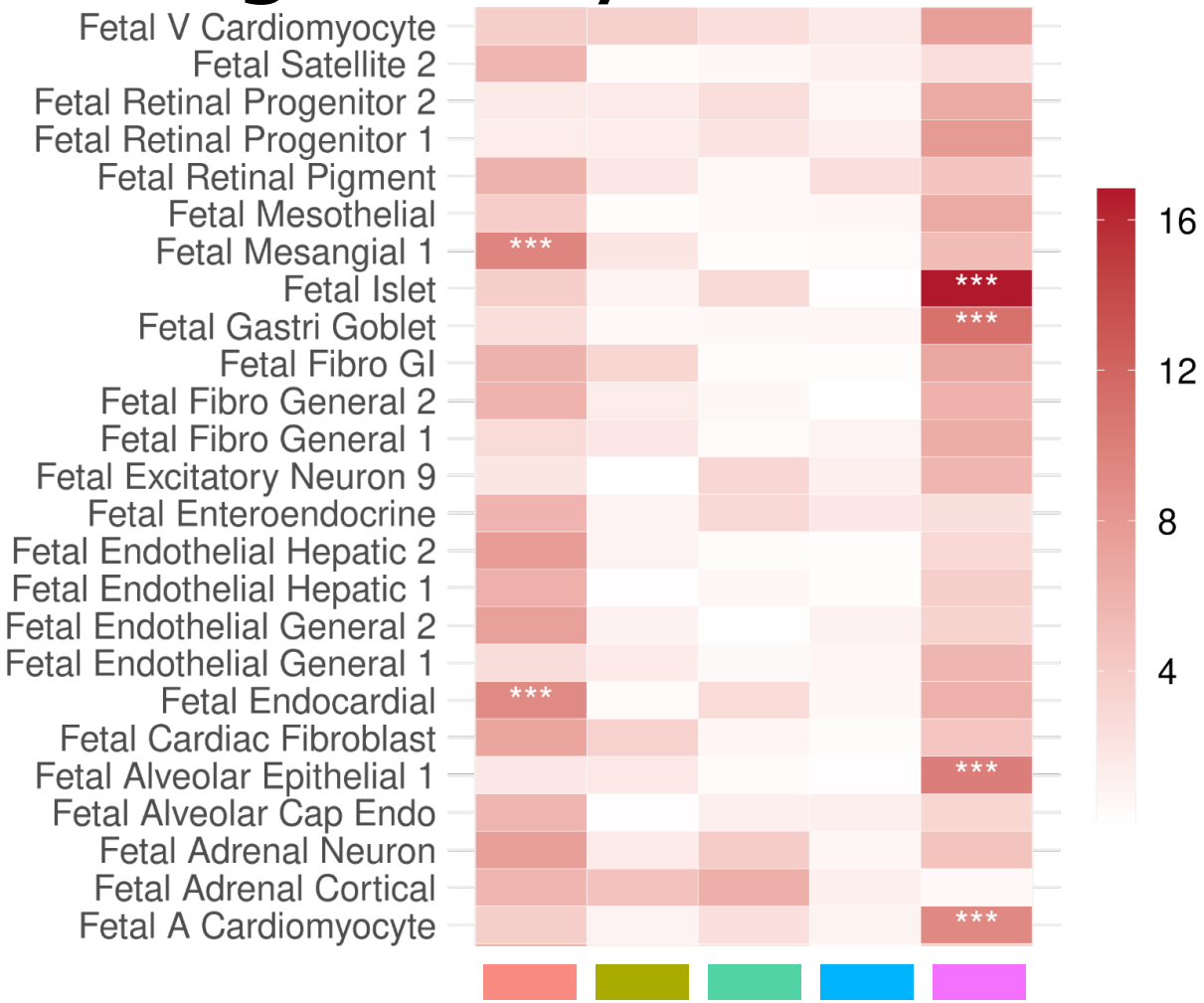


[More details]

Comparison with published clusterings

Pascat et al, accepted *Nat Comms*, https://www.medrxiv.org/content/10.1101/2025.03.02.25323190v1

Results: Cluster characterization: regulatory elements using scATAC-seq atlas



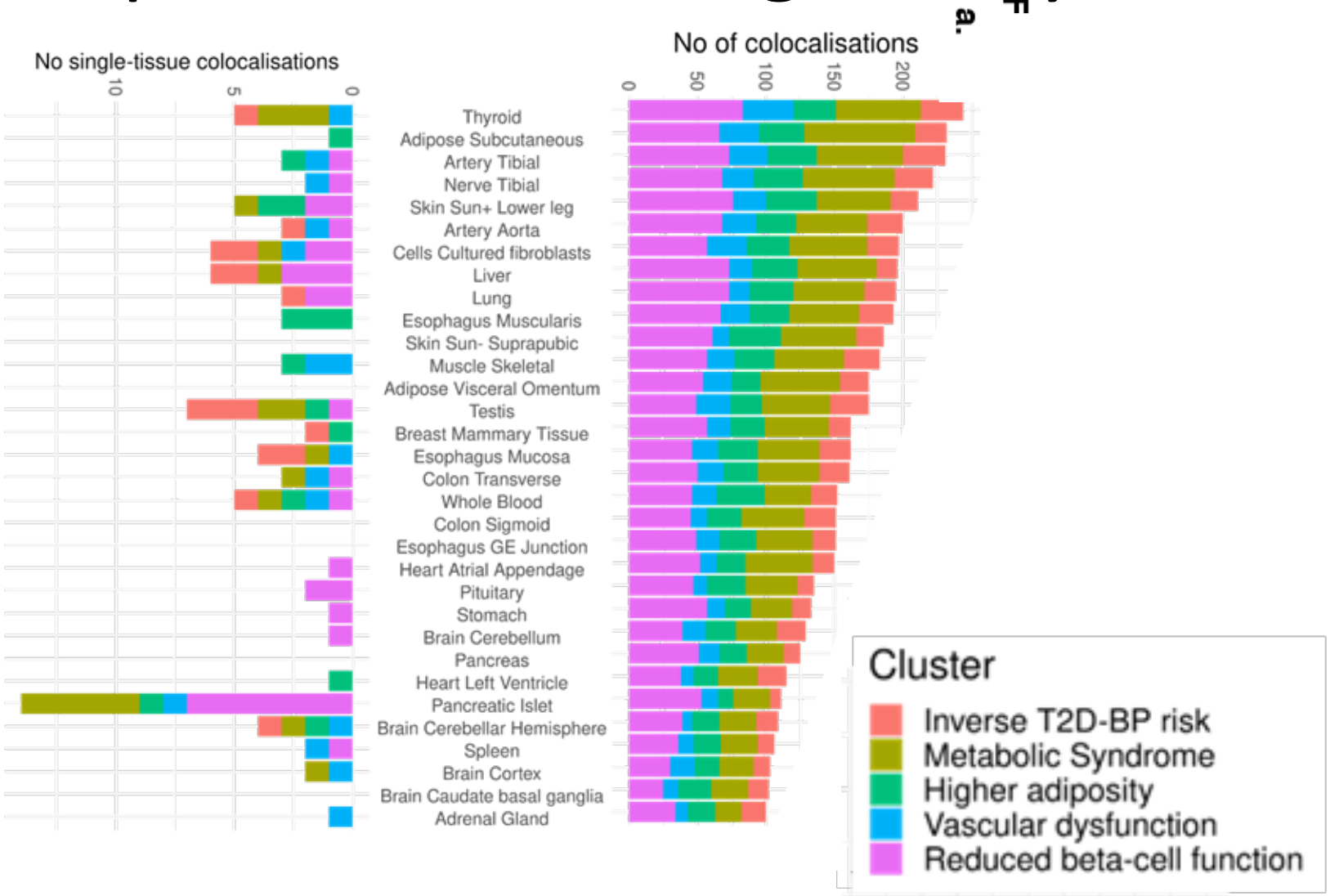
Atlases of single cell of open chromatin region (**CATLAS**) in 222 cell types derived from 30 adult tissues and 15 fetal tissues

Enrichment between the clusters and the open chromatin regions of **the fetal cells**

Methods

Pascat et al, accepted *Nat Comms*, https://www.medrxiv.org/content/10.1101/2025.03.02.25323190v1

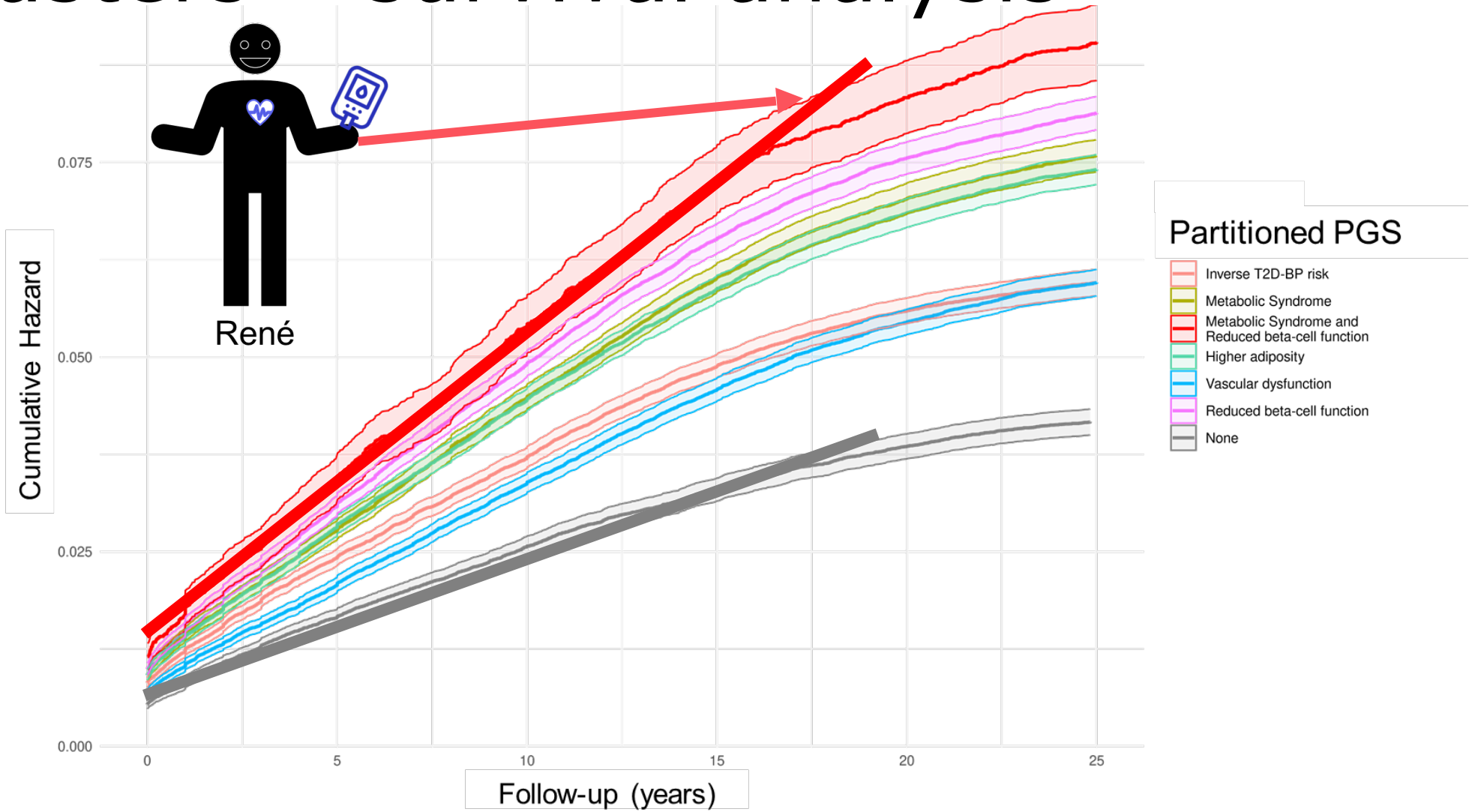
Results: Cluster characterisation by gene expressions and regulatory mechanisms



Distribution of colocalised loci across clusters in 50 human adult tissues.
Bars – number of colocalised signals per tissue
Colours – contribution from the five clusters

Pascat et al, accepted *Nat Comms*, https://www.medrxiv.org/content/10.1101/2025.03.02.25323190v1

Results: Partitioned PGS based on clusters – survival analysis



Pascat et al, accepted *Nat Comms*, https://www.medrxiv.org/content/10.1101/2025.03.02.25323190v1

Results: Partitioned PGS based on clusters: how can we predict the risk of T2D-BP clustering?

Individuals in the top 10% of the PGS distribution based on…	**Relative risk of comorbidity (T2D and hypertension)**
cluster	1.04
cluster	1.14
cluster	1.36
cluster	1.44
cluster	1.55
clusters	2.13

PGS distribution

Pascat et al, accepted *Nat Comms*, https://www.medrxiv.org/content/10.1101/2025.03.02.25323190v1

Discussion

- T2D and BP regulation genetics are **highly intertwined**

- We identified **five clusters of pathogenetic processes** underlying the T2D-BP relationship, including an **inverse T2D-BP risk cluster**

- We bring forward a property of **partitioned PGS to identify high-risk individuals for complex phenotype** (and complications) at a **very young age**

Vincent Pascat[1,2], Liudmila Zudina[2,3], Lucas Maurin[1], Anna Ulrich[2], Jared G. Maina[1], Ayse Demirkan[2,3,4], Zhanna Balkhiyarova[2,3,4], Amélie Bonnefond[1,2], Igor Pupko[3], Yevheniya Sharhorodska[3], François Pattou[1,5], Bart Staels[6], Marika Kaakinen[2,3,4,7], Amna Khamis[1,2], Amélie Bonnefond[1,2], Patricia Munroe[8,9], Philippe Froguel[1,2*], Inga Prokopenko[1,3,4*]

[1] Université de Lille, Inserm UMR1283, CNRS UMR8199, European Genomic Institute for Diabetes (EGID), Institut Pasteur de Lille, Lille University Hospital, Lille, France; [2] Department of Metabolism, Digestion, and Reproduction, Imperial College London, London, UK; [3] Section of Statistical Multi-omics, Department of Clinical and Experimental Medicine, University of Surrey, Guildford, UK; [4] People-Centred Artificial Intelligence Institute, University of Surrey, Guildford, UK; [5] Department of General and Endocrine Surgery, CHU Lille, F-59000, Lille, France; [6] Univ. Lille, Inserm, CHU Lille, Institut Pasteur de Lille, U1011- EGID, F-59000 Lille, France; [7] Institute for Molecular Medicine Finland, University of Helsinki, Helsinki, Finland; [8] William Harvey Research Institute, Barts and the London Faculty of Medicine and Dentistry, Queen Mary University of London, EC1M 6BQ London, UK; [9] National Institute of Health and Care Research, Barts Cardiovascular Biomedical Research Centre, Queen Mary University of London, EC1M 6BQ London, UK

This research has been conducted using the UK Biobank Resource under application number 236. This project was in part funded by the Agence Nationale de la Recherche under the Programme d'Investissement d'Avenir (PreciDIAB, ANR-18-IBHU-0001 and RHU PreciNASH ANR-16-RHUS-0006), by the European Union through the "Fonds Européen de Développement Regional" (FEDER), by the "Conseil Régional des Hauts-de-France" (Hauts-de-France Regional Council), by the "Métropole Européenne de Lille" (MEL, European Metropolis of Lille), and by the European Research Council (ERC OpiO – 101043671, to AB)

The authors would like to thank all the investigators from different consortia that built and shared the GWAS meta-analysis, eQTLs, and scATAC-seq atlases used in this study, as well as the UK Biobank participants and dedicated staff.