Multi-omics data integration to study human diseases and physiology

Oleg Borisov, PhD
oleg.borisov@uniklinik-freiburg.de
Institute of Genetic Epidemiology, Medical Center - University of Freiburg

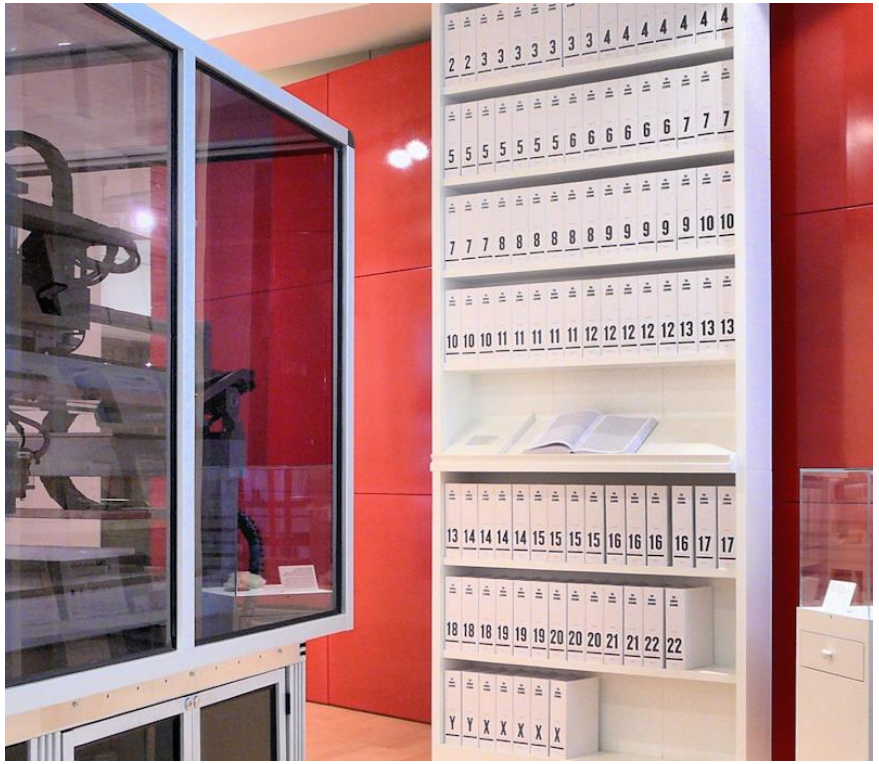
Invited presentation at Leibniz Research Institute for Environmental Medicine

29. Oktober 2024

Human genome and its variation

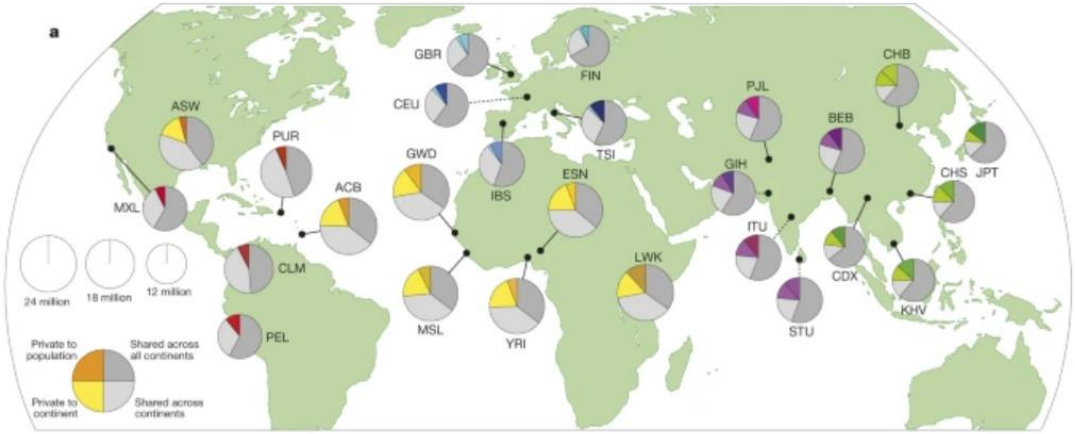
Human Genome Project (2003)

The first printout of the human genome



By Russ London at English Wikipedia, CC BY-SA 3.0, https://commons.wikimedia.org/w/index.php?curid=9923576

1000 Genomes Project
N≈88 million variants in ~2500 genomes (2015)



NHLBI TOPMed Program
N>400 million variants in ~54000 genomes (2021)

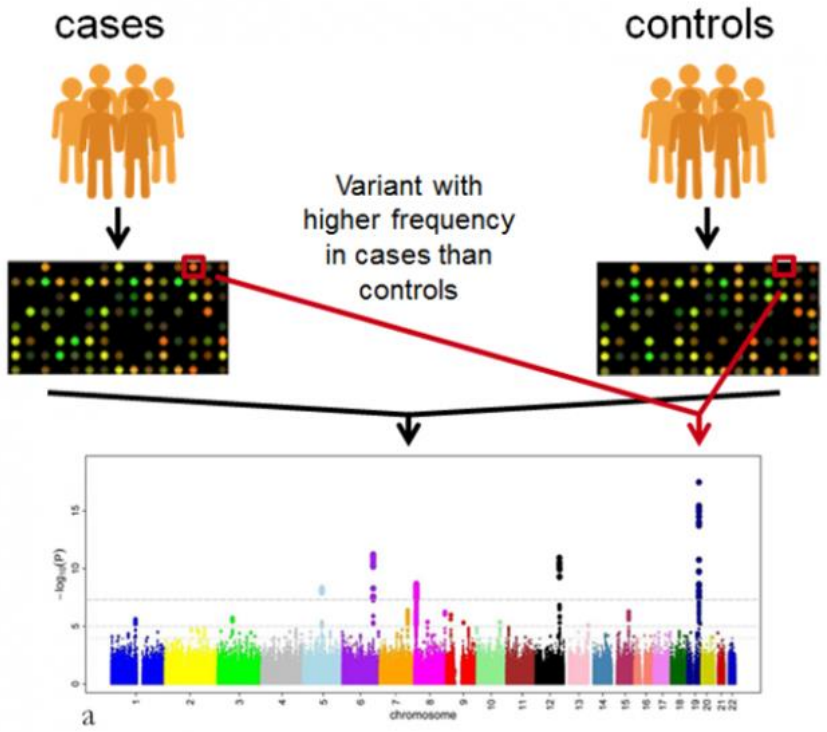
NIH National Heart, Lung, and Blood Institute | BioData CATALYST
TOPMed Imputation Server

PMID 26432245, 33568819

UNIVERSITÄTS KLINIKUM FREIBURG

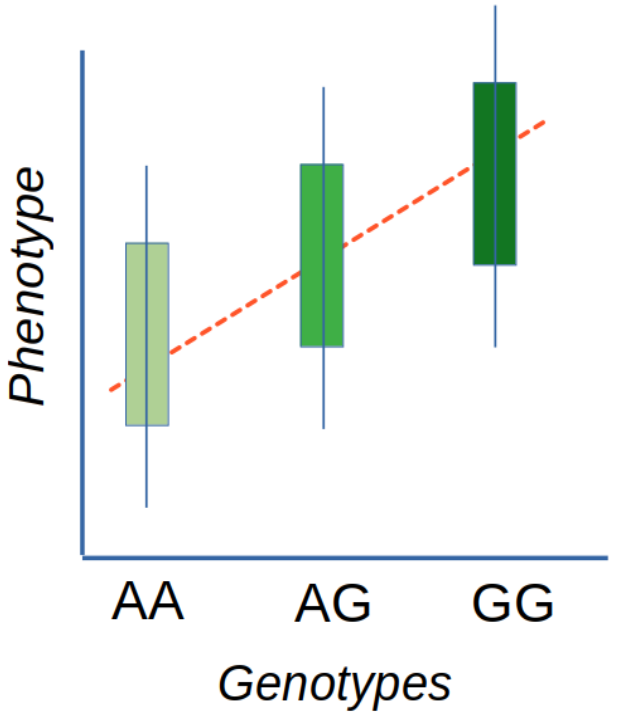
Genotype-phenotype associations

Binary phenotypes



Example: Cases with Diabetes vs Healthy Controls

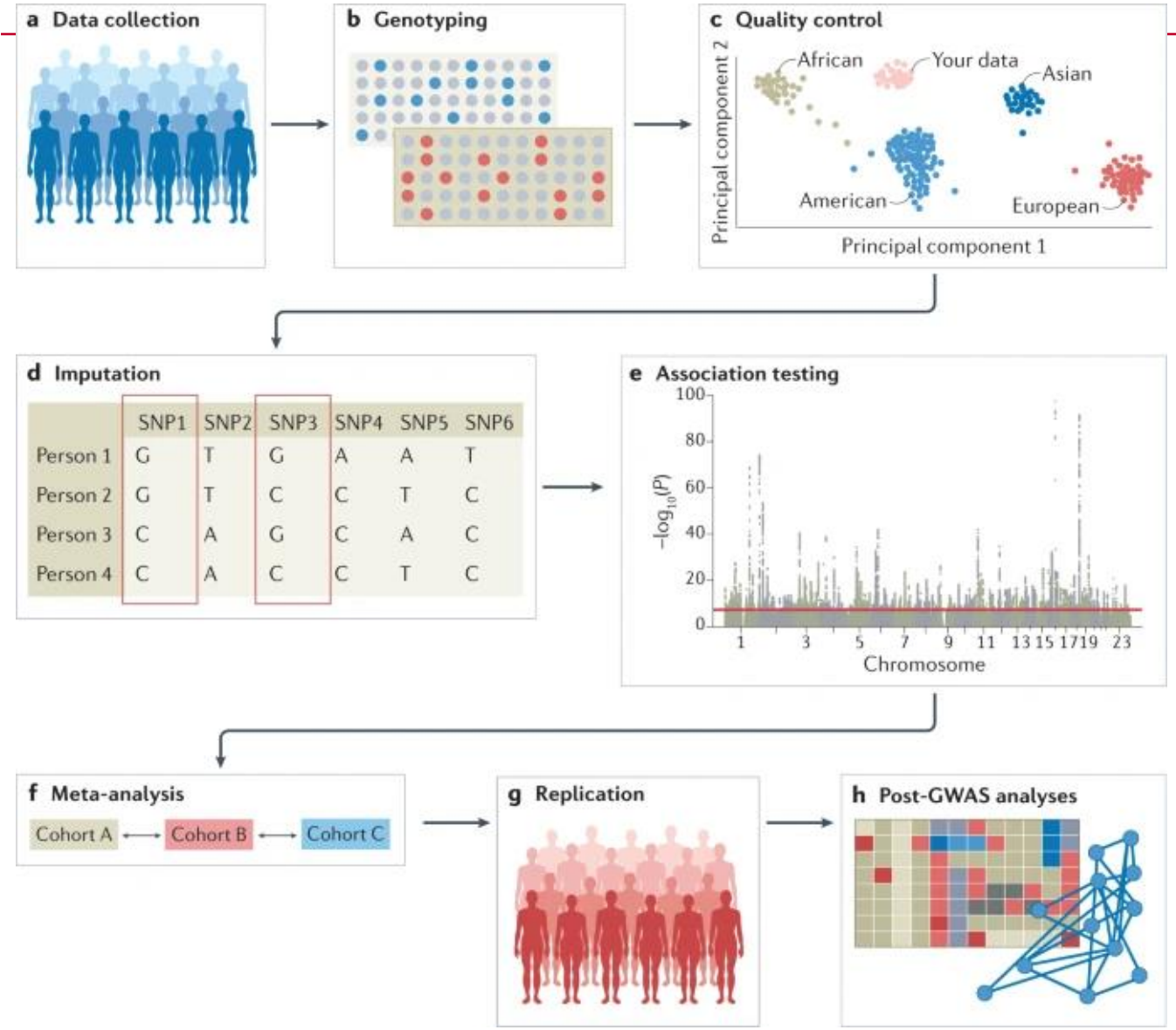
Quantitative phenotypes

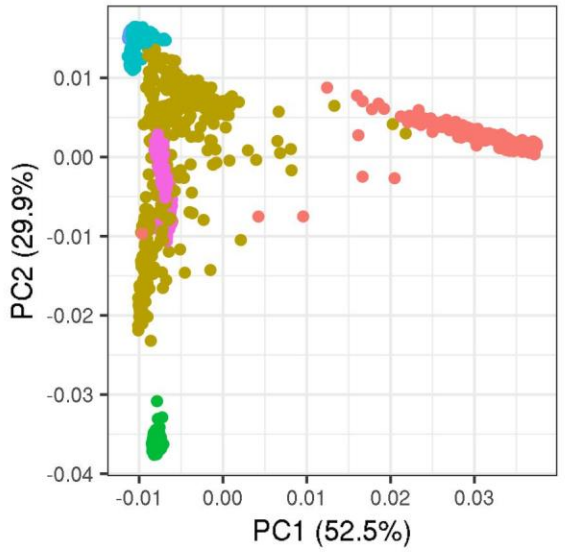


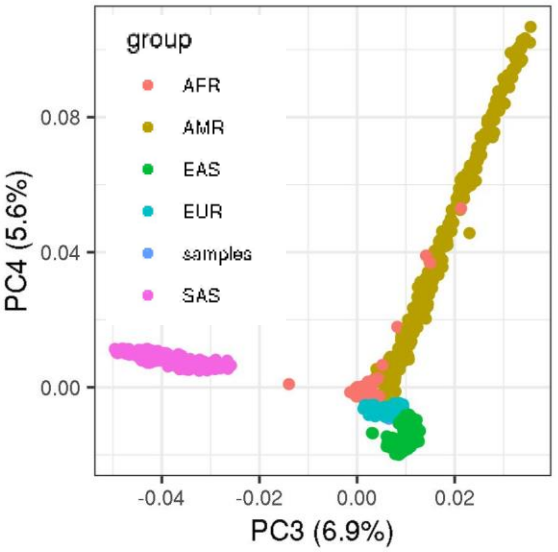
Example: Height

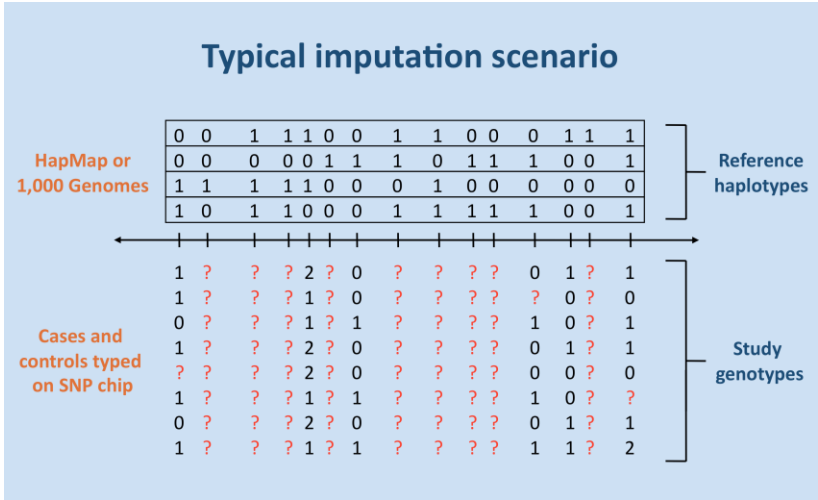
https://www.ebi.ac.uk/training/online/course/gwas-catalog-exploring-snp-trait-associations-2019/what-gwas-catalog/what-are-genome-wide

Genome-wide association study (GWAS)









DOI 10.1038/s43586-021-00056-9

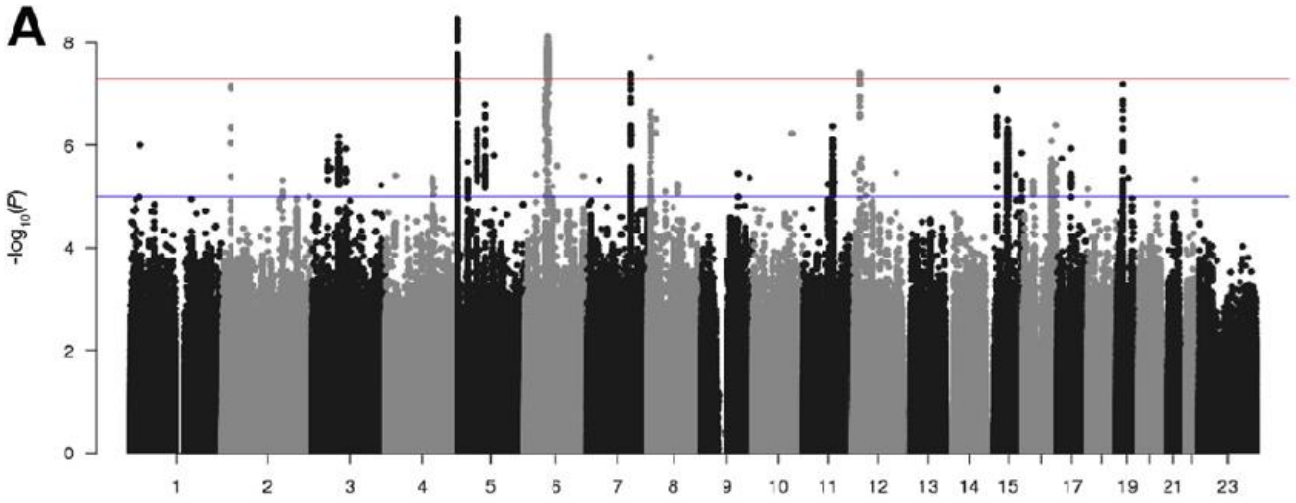
https://mathgen.stats.ox.ac.uk/impute/impute_v2.html

UNIVERSITÄTS KLINIKUM FREIBURG

Statistical analysis and visualization

Example 1

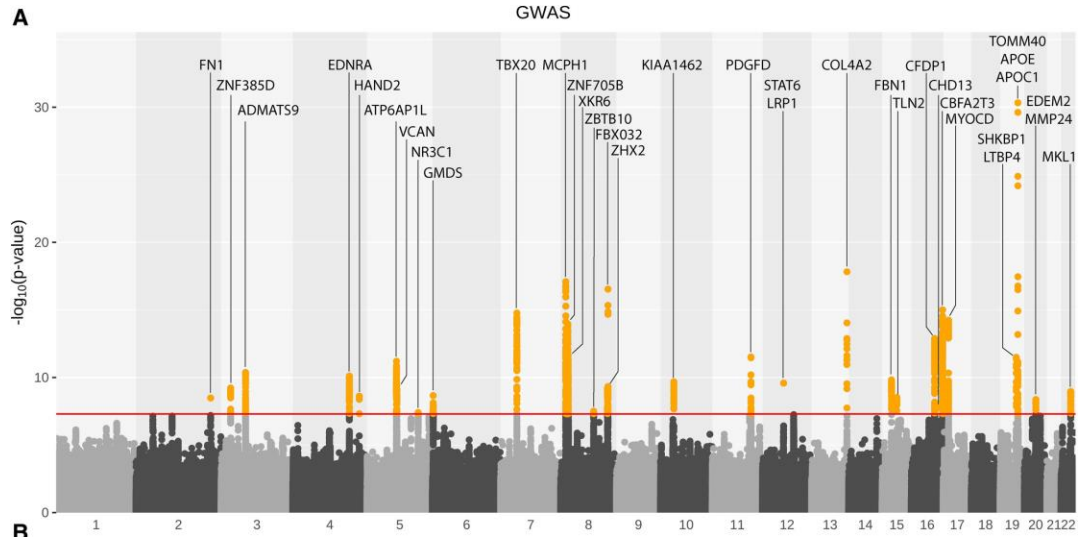
- GWAS of esophageal adenocarcinoma
- 7 000 cases and 7 000 controls



Dong, J., Maj, C., Tsavachidis, S., ..., **Borisov, O.**, et al. (2020). Sex-Specific Genetic Associations for Barrett's Esophagus and Esophageal Adenocarcinoma. Gastroenterology.

Example 2

- GWAS of carotid artery intima-media thickness
- 45 185 participants

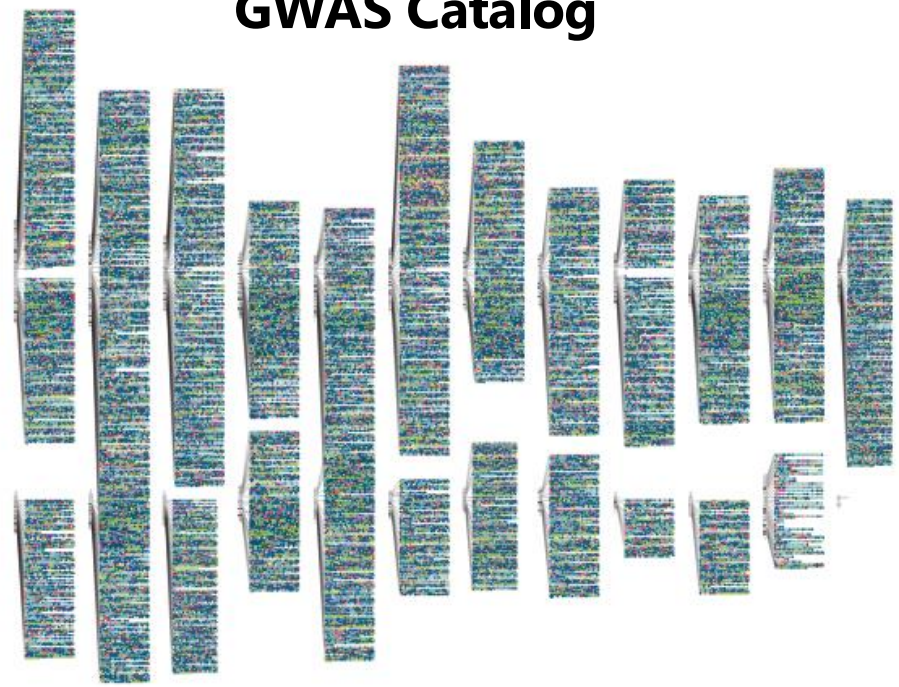


Yeung, M. W., Wang, S., van de Vegte, Y. J., **Borisov, O.**, et al. (2022). Twenty-Five Novel Loci for Carotid Intima-Media Thickness: A Genome-Wide Association Study in >45 000 Individuals and Meta-Analysis of >100 000 Individuals. Arteriosclerosis, Thrombosis, and Vascular Biology.

UNIVERSITÄTS KLINIKUM FREIBURG

Molecular biobanks

GWAS Catalog



As of 2024-10-09, the GWAS Catalog contains 7027 publications, 686504 top associations and 97285 full summary statistics.

https://www.ebi.ac.uk/gwas/home

Open Targets Genetics

Search for a gene, variant, study, or trait...

PCSK9 1_154453788_C_T rs4129267 LDL cholesterol (Willer CJ et al. 2013)

AHR aryl hydrocarbon receptor

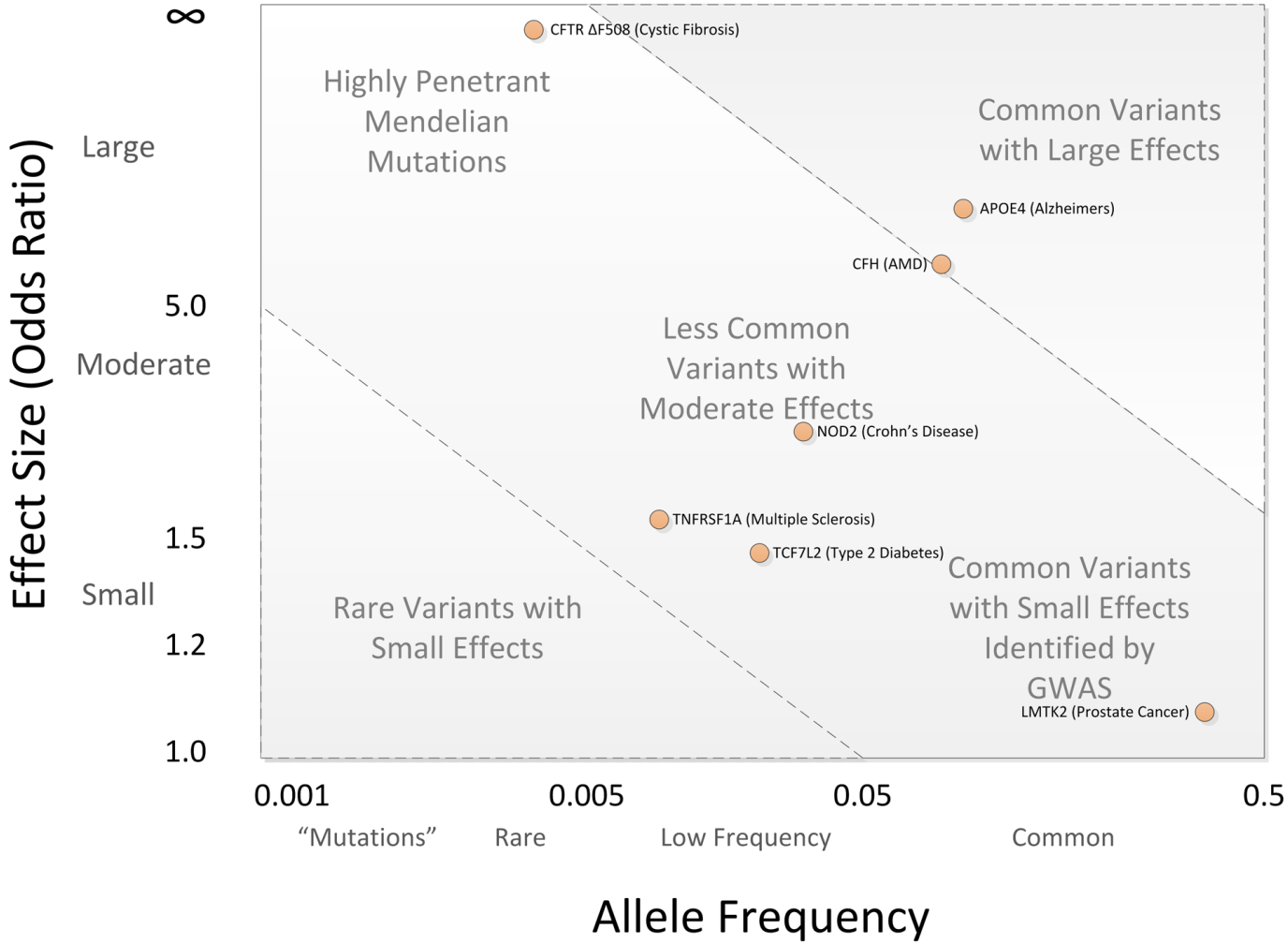
Ensembl: ENSG00000106546 | gnomAD

Study ID	Trait	Publication	N Initial	Lead Variant	P-value	Beta
NEALE2_1498	Coffee intake	UKB Neale v2 (2018)	334,659	7_17244953_T_C	3.2e-95	0.039
NEALE2_1727	Ease of skin tanning	UKB Neale v2 (2018)	353,697	7_17095084_G_A	1.5e-83	0.14
GCST008524	Bitter non-alcoholic beverage consumption	Zhong VW (2019)	85,852	7_17244953_T_C	3.0e-75	0.030

https://genetics.opentargets.org/

UNIVERSITÄTS KLINIKUM FREIBURG

Allele frequency spectrum

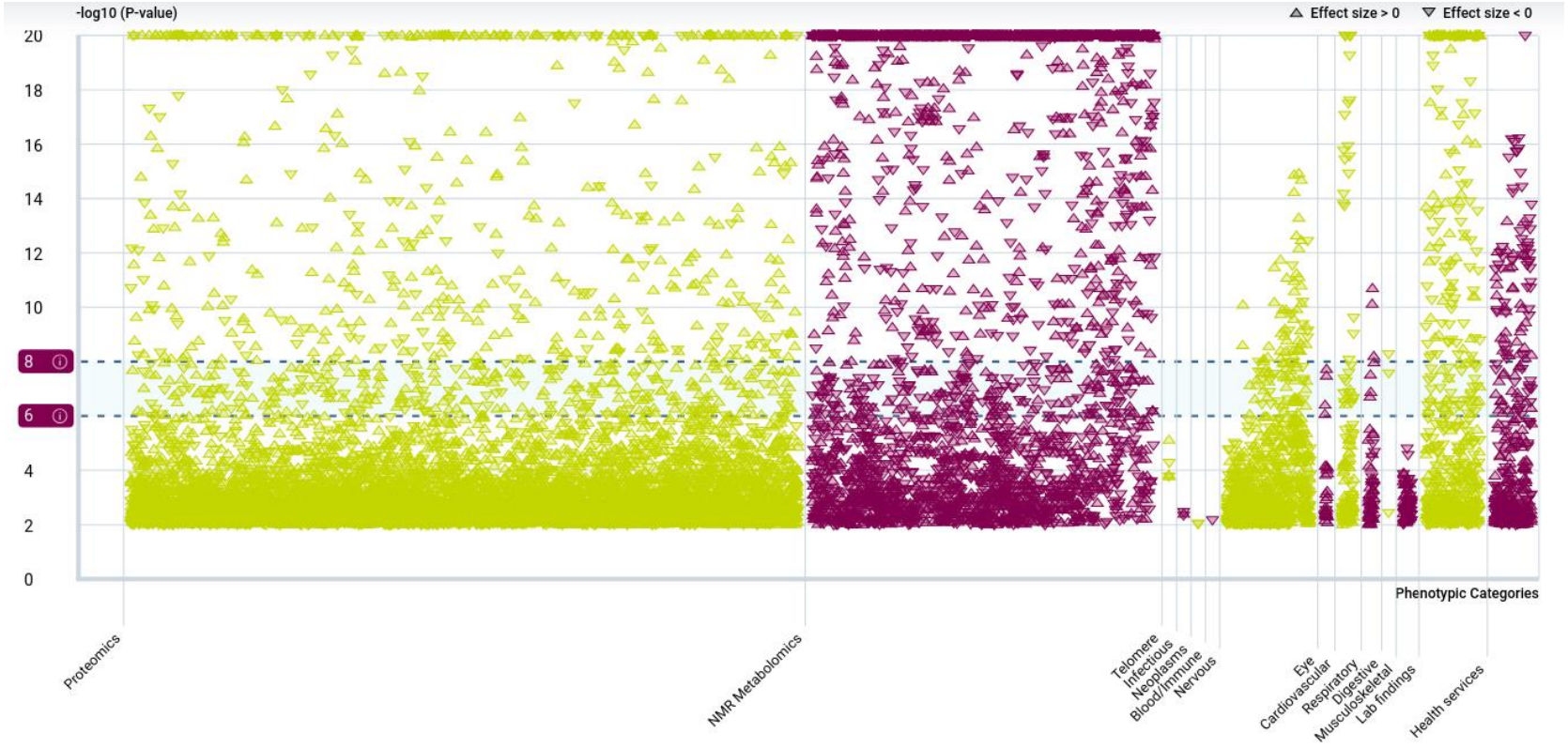


PMID 23300413

UNIVERSITÄTS KLINIKUM FREIBURG

Rare variants: sequencing studies

- Variant-level and gene-level associations with thousands of human traits
  - Based on the UK Biobank (500,000 individuals)
- Example: APOE gene (phenome-wide associations)



AstraZeneca PheWAS Portal

https://azphewas.com/

UNIVERSITÄTS KLINIKUM FREIBURG

Epidemiological approach for multi-omics data analysis

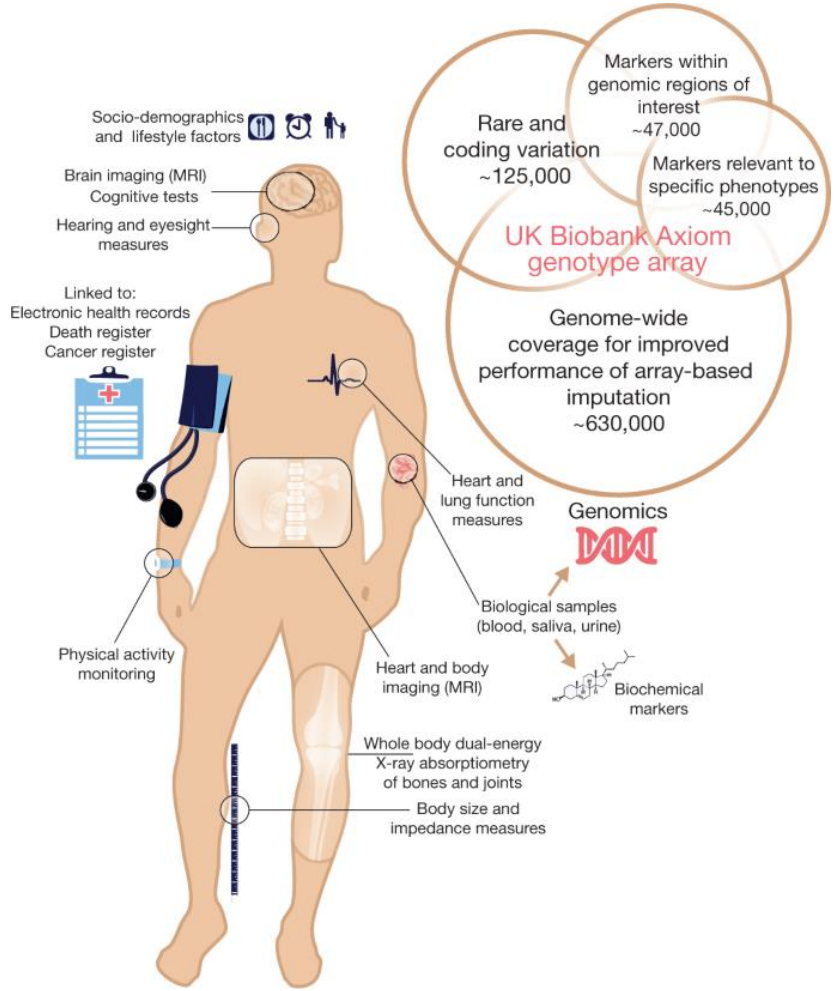
– The data from large-scale biobanks and population based cohorts enable efficient target identification and drug discovery at scale.

Cohort / Database	Number of participants	Available Data
UK Biobank	500,000	Whole-genome sequencing, proteomics, metabolomics, imaging
All of Us	250,000 (ongoing, up to 1 million)	Genotyping, sequencing, medical endpoints
FinnGen	500,000	Genotyping, National health registries

https://www.ukbiobank.ac.uk/
https://allofus.nih.gov/
https://www.finngen.fi/en

UK Biobank: multi-omics and deep phenotyping data

- 500,000 volunteer participants.
- Age: 40-69 years at the time of recruitment.
- Genotyping and whole-genome sequencing.
- Proteomics (Olink, 3000 proteins).
- Metabolomics, Radiomics (Imaging).
- Phenotyping data:
- ICD-10 codes: (>7 million codes)
- Self-reported medical conditions.
- …



https://www.ukbiobank.ac.uk (PMID 30305743)

UNIVERSITÄTS KLINIKUM FREIBURG

Proteogenomics study in the UK Biobank

Article | Open access | Published: 04 October 2023

Plasma proteomic associations with genetics and health in the UK Biobank

Benjamin B. Sun, Joshua Chiou, Matthew Traylor, Christian Benner, Yi-Hsiang Hsu, Tom G. Richardson, Praveen Surendran, Anubha Mahajan, Chloe Robins, Steven G. Vasquez-Grinnell, Liping Hou, Erika M. Kvikstad, Oliver S. Burren, Jonathan Davitte, Kyle L. Ferber, Christopher E. Gillies, Åsa K. Hedman, Sile Hu, Tinchi Lin, Rajesh Mikkilineni, Rion K. Pendergrass, Corran Pickering, Bram Prins, Denis Baird, Alnylam Human Genetics, AstraZeneca Genomics Initiative, Biogen Biobank Team, Bristol Myers Squibb, Genentech Human Genetics, GlaxoSmithKline Genomic Sciences, Pfizer Integrative Biology, Population Analytics of Janssen Data Sciences, Regeneron Genetics Center, ... Christopher D. Whelan + Show authors

Nature **622**, 329–338 (2023) | Cite this article

110k Accesses | **165** Citations | **324** Altmetric | Metrics

Article | Open access | Published: 04 October 2023

Rare variant associations with plasma protein levels in the UK Biobank

Ryan S. Dhindsa, Oliver S. Burren, Benjamin B. Sun, Bram P. Prins, Dorota Matelska, Eleanor Wheeler, Jonathan Mitchell, Erin Oerton, Ventzislava A. Hristova, Katherine R. Smith, Keren Carss, Sebastian Wasilewski, Andrew R. Harper, Dirk S. Paul, Margarete A. Fabre, Heiko Runz, Coralie Viollet, Benjamin Challis, Adam Platt, AstraZeneca Genomics Initiative, Dimitrios Vitsios, Euan A. Ashley, Christopher D. Whelan, Menelas N. Pangalos, ... Slavé Petrovski + Show authors

Nature **622**, 339–347 (2023) | Cite this article

39k Accesses | **25** Citations | **141** Altmetric | Metrics

Common variants

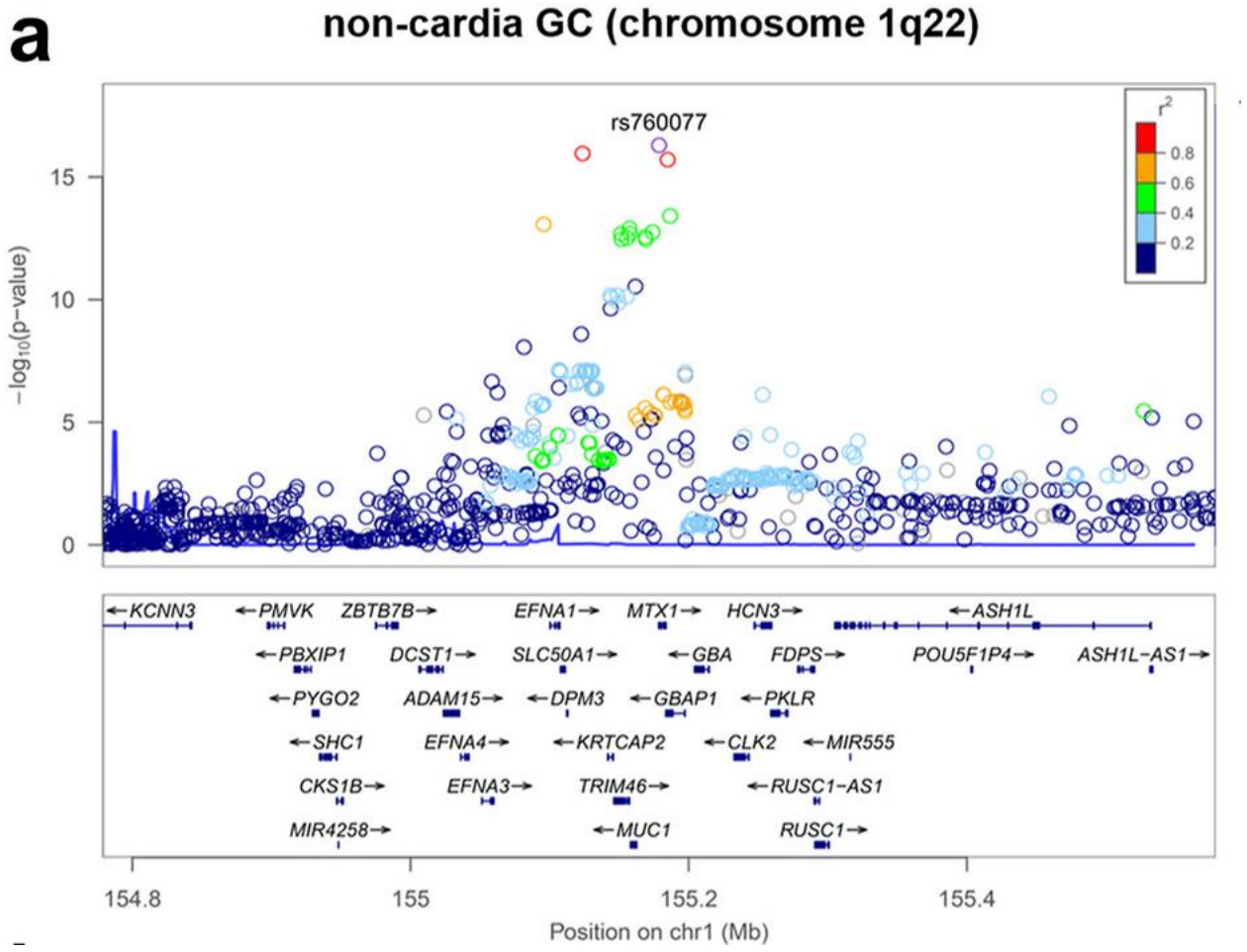
- N=54,219 UK Biobank participants.
- protein quantitative trait locus (pQTL) mapping of 2,923.
- N=14,287 primary genetic association (81% previously undescribed).

Rare variants

- N=5,433 rare genotype–protein associations (81% previously undetected in the study of common variants).
- Aggregated signals: 1,962 gene–protein associations.
- Of the 691 gene-level signals from protein-truncating variants, 99.4% were associated with decreased protein levels.

Mapping genetic associations to causal genes

- Genetic variants identified in GWAS are often located in non-coding regions.
- These variants usually play regulatory role (e.g., modulation of gene expression).

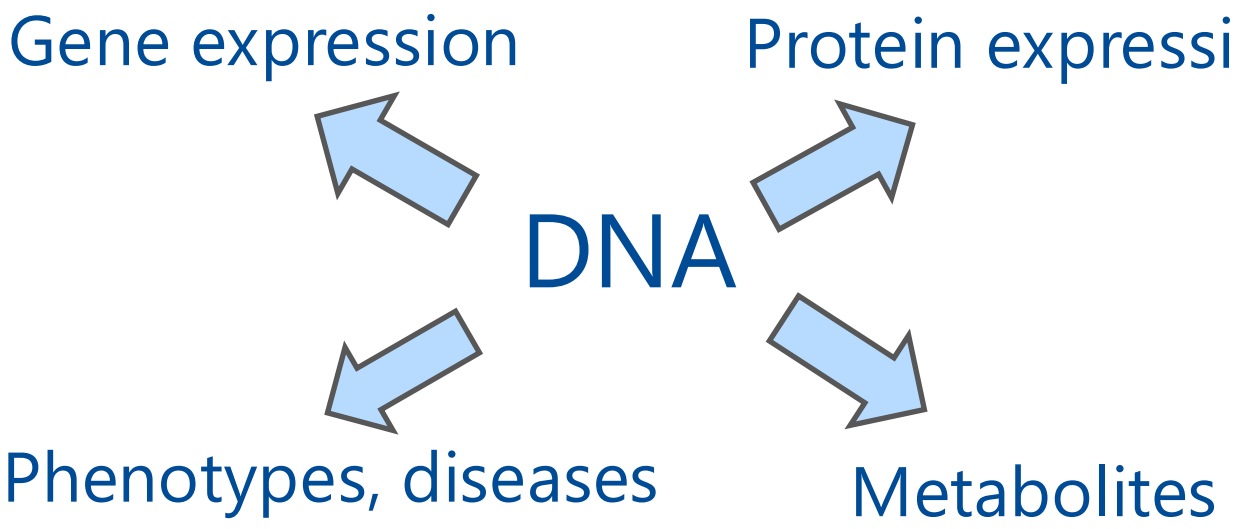


Dissecting the genetic heterogeneity of gastric cancer. T. Hess, C. Maj, **O. Borisov** et al. eBioMedicine, Volume 92, 104616

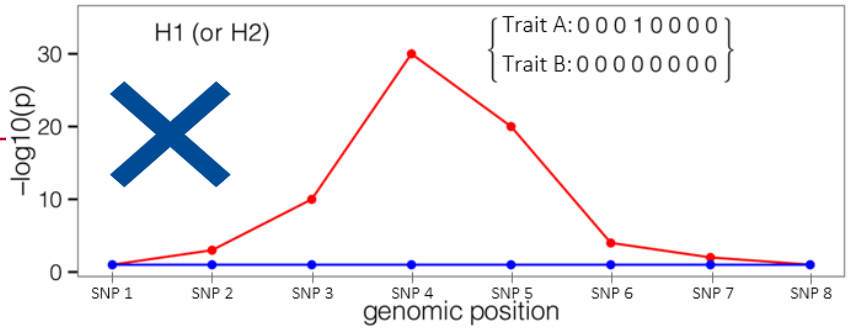
UNIVERSITÄTS KLINIKUM FREIBURG

Genetic colocalization

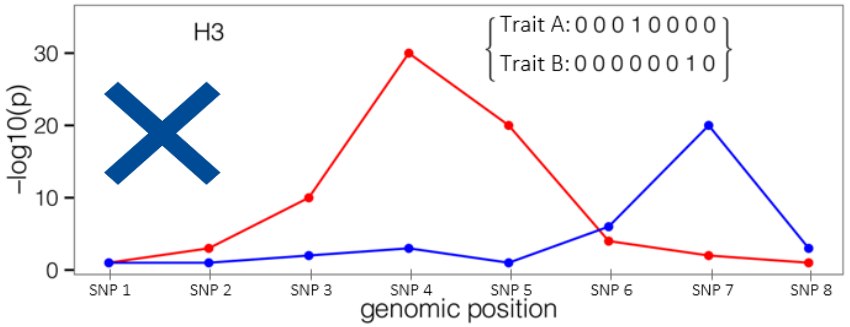
- A statistical method used to assess shared genetic etiology across two phenotypes (e.g. GWAS of complex disease and gene expression).



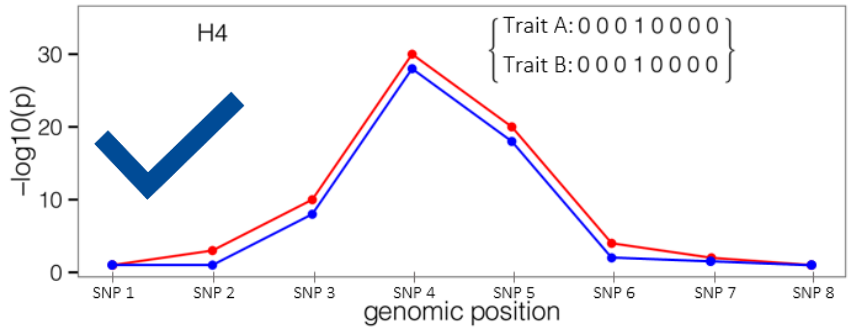
Only one trait has association



Different causal variants



Same causal variant



PMID: 24830394 (Fig. 1, modified)

UNIVERSITÄTS KLINIKUM FREIBURG

Resources available for colocalization

Dataset	Study	Phenotypes	Participants
Gene expression	GTEx v8, Human kidney atlas	N > 10,000 genes	N ≈ 15,000 samples N = 49 tissues
Plasma proteomics	Icelanders and UK Biobank proteomics	N > 4,000 proteins	N > 50,000
Phenome-wide association studies	UK Biobank PheWeb and FinnGen (DF9)	N > 3,500 binary outcomes	N > 500,000
…			

genepicoloc: R package to facilitate colocalization analysis

Input preprocessing
- Perform QC
- Harmonize, liftOver
- Identify significant regions

External studies
- Indexed summary statistics

Colocalization
- Computationally efficient
- 1,000,000+ jobs

https://genepi-freiburg.github.io/genepicoloc

UNIVERSITÄTS KLINIKUM FREIBURG