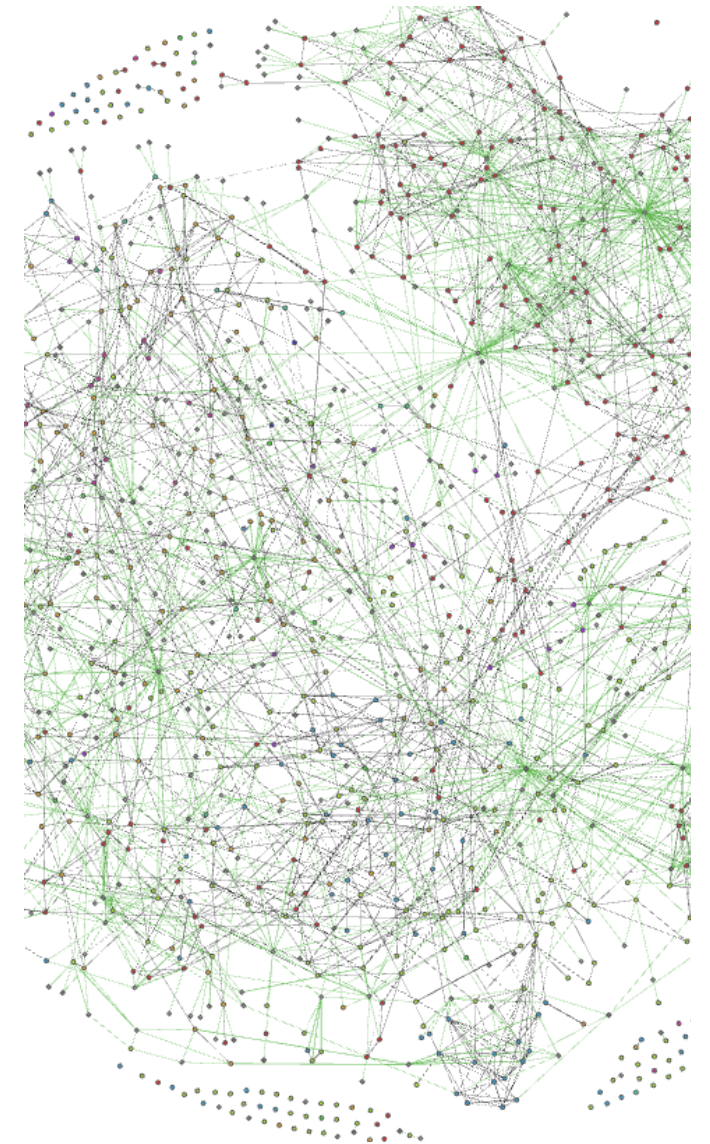# Linking metabolomics to diseases using human genetics: a CLSA study

Yiheng Chen, MSc

PhD Student in Human Genetics

Lady Davis Institute for Medical Research, Jewish General Hospital,

McGill University
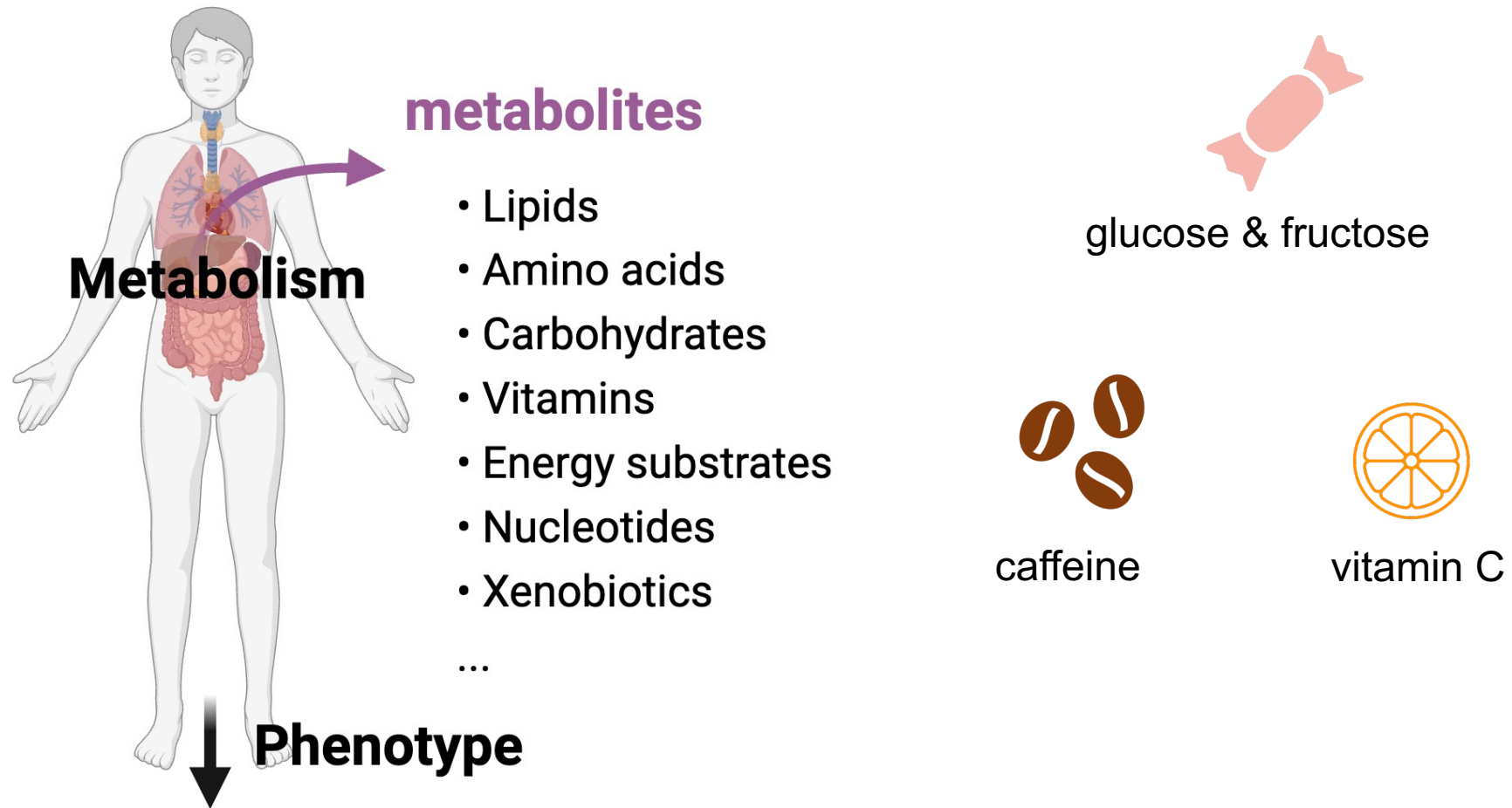
Yiheng Chen[1,2], Tianyuan Lu [1,3,4], Ulrika Pettersson-Kymmer[5], Isobel D. Stewart[6], Guillaume Butler-Laporte [1,7], Tomoko Nakanishi [1,2,8,9], Agustin Cerani[1,7], Kevin Y. H. Liang[1,3], Satoshi Yoshiji[1,2,8,9], Julian Daniel Sunday Willett[1,3,10], Chen-Yang Su [1,11], Parminder Raina[12,13], Celia M. T. Greenwood [1,3,7,14], Yossi Farjoun[1,4,15,16], Vincenzo Forgetta[1,4], Claudia Langenberg[17,18,6], Sirui Zhou[1,2], Claes Ohlsson [19,20] & J. Brent Richards [1,2,4,7,21,22] ✉
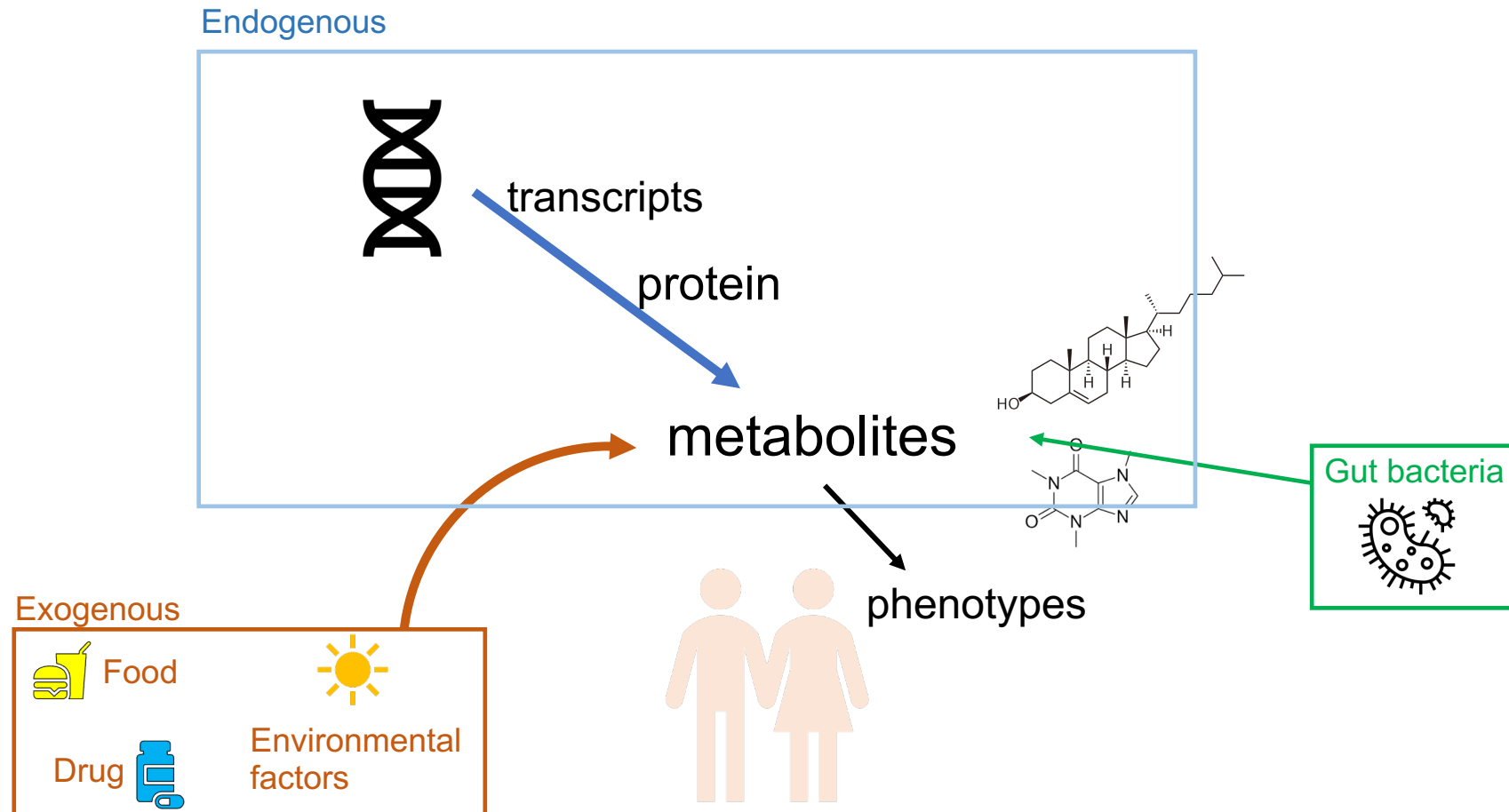
Table of contents

- Background
- Methods
- Data
- Study
  - Research aims
  - Design
  - Results
  - Conclusion

What are metabolites?

metabolites

- Lipids
- Amino acids
- Carbohydrates
- Vitamins
- Energy substrates
- Nucleotides
- Xenobiotics

...

Metabolism

Phenotype

glucose & fructose

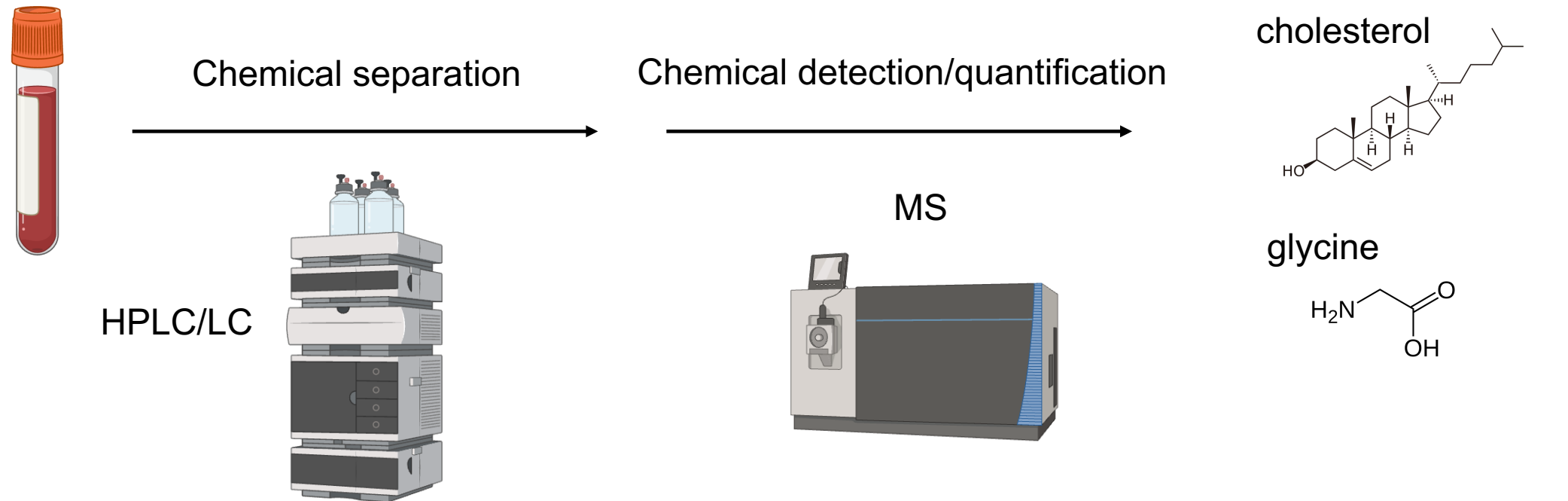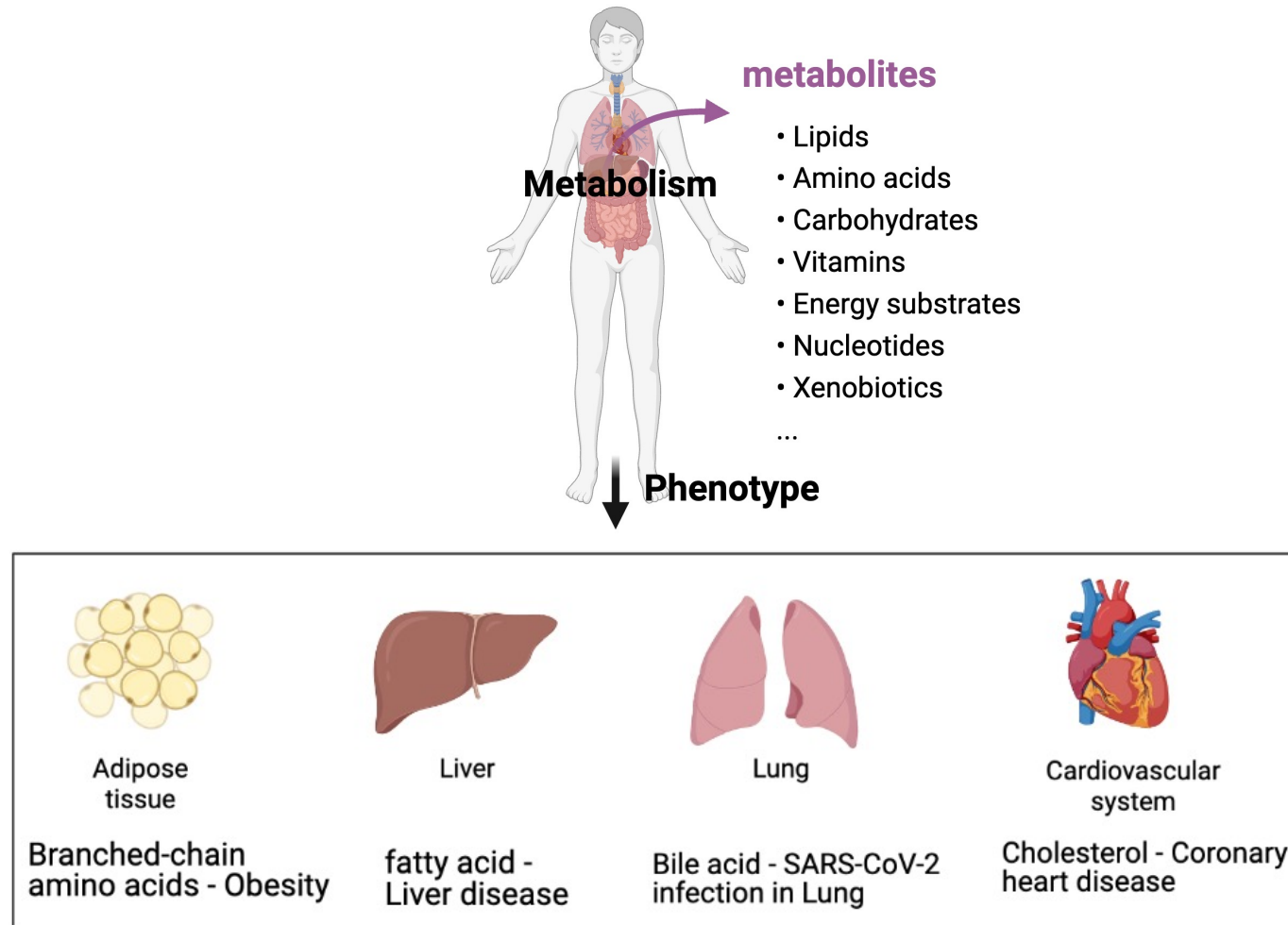caffeine

vitamin C

(Image generated using Biorender.com)

What factors can influence metabolites?

How are metabolites measured?

- High performance liquid chromatography–mass spectrometry (HPLC/LC-MS)

Chemical separation

Chemical detection/quantification

HPLC/LC

MS

>1000 metabolites

cholesterol

glycine

Why do we need to study metabolites?

metabolites

- Lipids
- Amino acids
- Carbohydrates
- Vitamins
- Energy substrates
- Nucleotides
- Xenobiotics
...

Metabolism

Phenotype

Adipose tissue	Liver	Lung	Cardiovascular system
Branched-chain amino acids - Obesity	fatty acid - Liver disease	Bile acid - SARS-CoV-2 infection in Lung	Cholesterol - Coronary heart disease

(Lynch & Adams et al., 2021; Siddik, & Shin 2019; Chen et al., 2021; Image generated using Biorender.com)

Challenges?

- The metabolite-disease associations are susceptible to
    - **Reverse causation** (e.g., obesity vs metabolites)
    - **Confounding** (e.g., cardiovascular disease vs metabolites -- confounding factor: obesity)
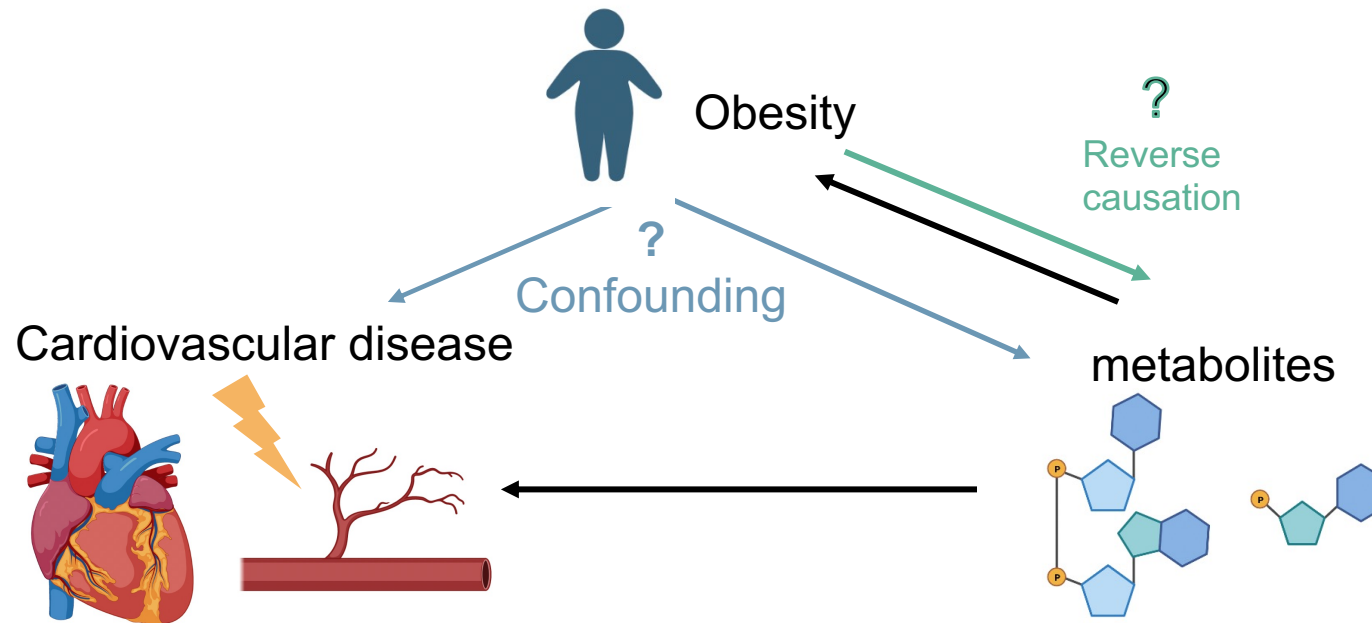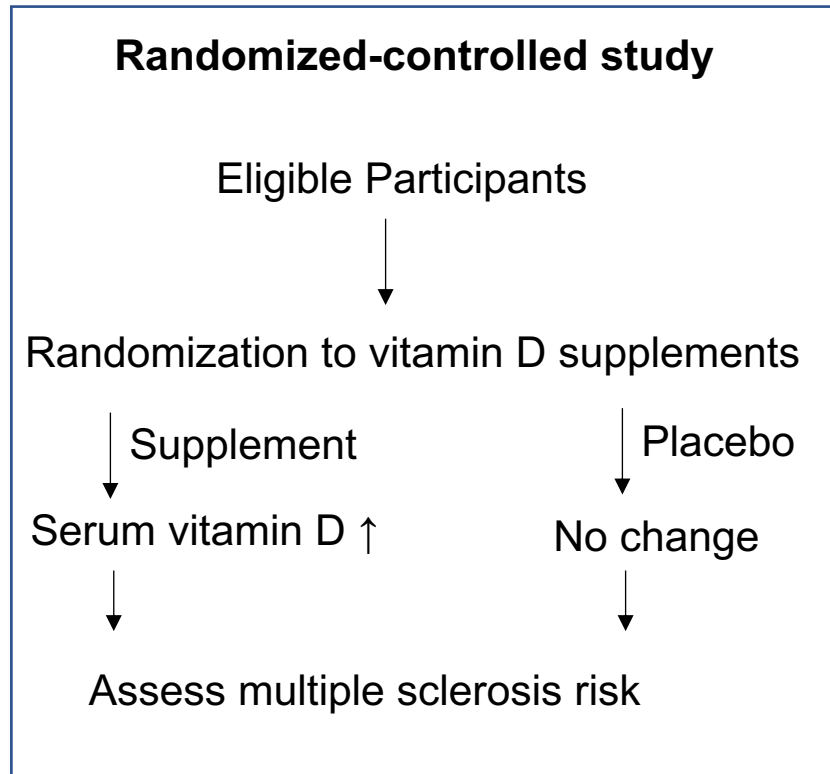
Table of contents

- Background
- Methods
- Data
- Study
  - Research aims
  - Design
  - Results
  - Conclusion

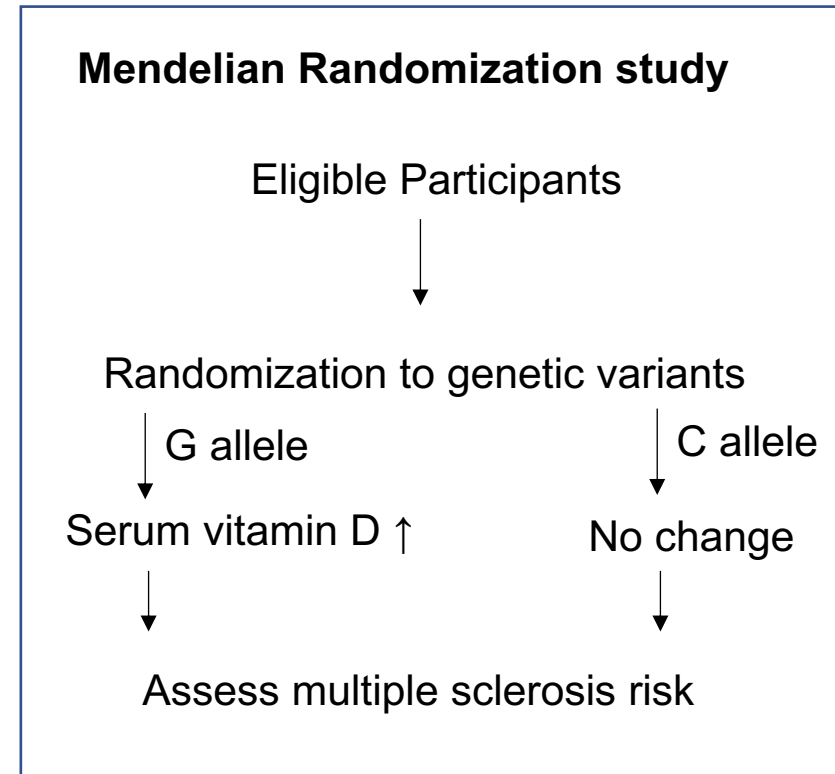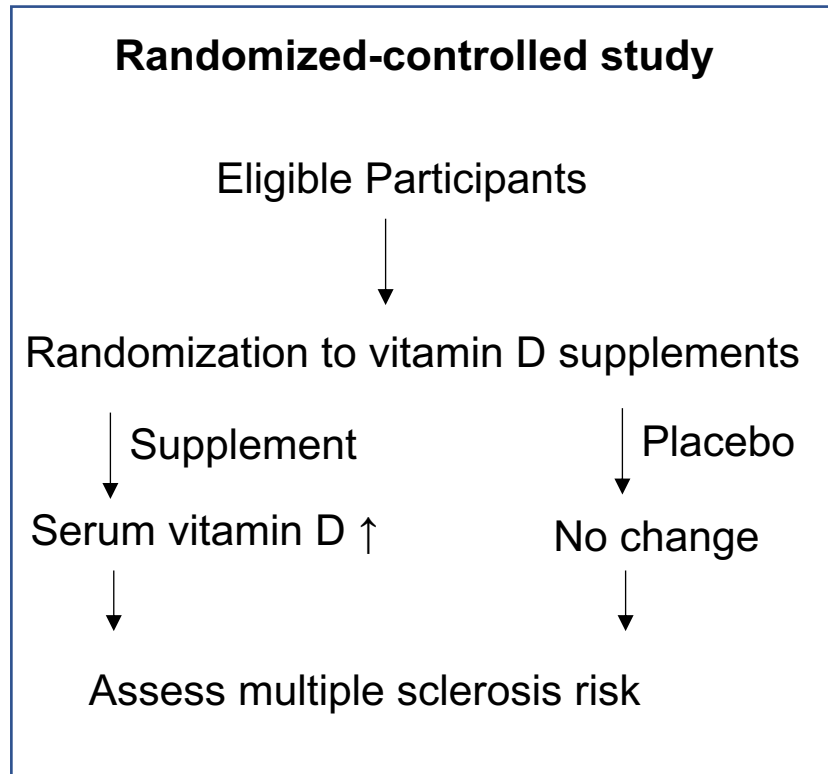# Randomized controlled study to investigate metabolite and diseases

Randomized-controlled study

Eligible Participants

↓

Randomization to vitamin D supplements

Supplement ↓   Placebo ↓

Serum vitamin D ↑   No change

↓   ↓

Assess multiple sclerosis risk

Advantages
1. Reduce the bias related to confounding by the randomization process
2. Prospective design to avoid reverse causation

Disadvantages
1. Costly
2. Long duration
3. Unethical for potentially harmful interventions

Mendelian Randomization for metabolite and diseases

Randomized-controlled study

Eligible Participants

↓

Randomization to vitamin D supplements

Supplement ↓   ↓ Placebo

Serum vitamin D ↑   No change

↓   ↓

Assess multiple sclerosis risk

Mendelian Randomization study

Eligible Participants

↓

Randomization to genetic variants

G allele ↓   ↓ C allele

Serum vitamin D ↑   No change

↓   ↓

Assess multiple sclerosis risk

(Davey Smith, G., & Ebrahim, S. 2003;
Mokry, Lauren E et al., 2015)

Vitamin D and Risk of Multiple Sclerosis: A Mendelian Randomization Study

Lauren E. Mokry, Stephanie Ross, Omar S. Ahmad, Vincenzo Forgetta, George Davey Smith, Aaron Leong,

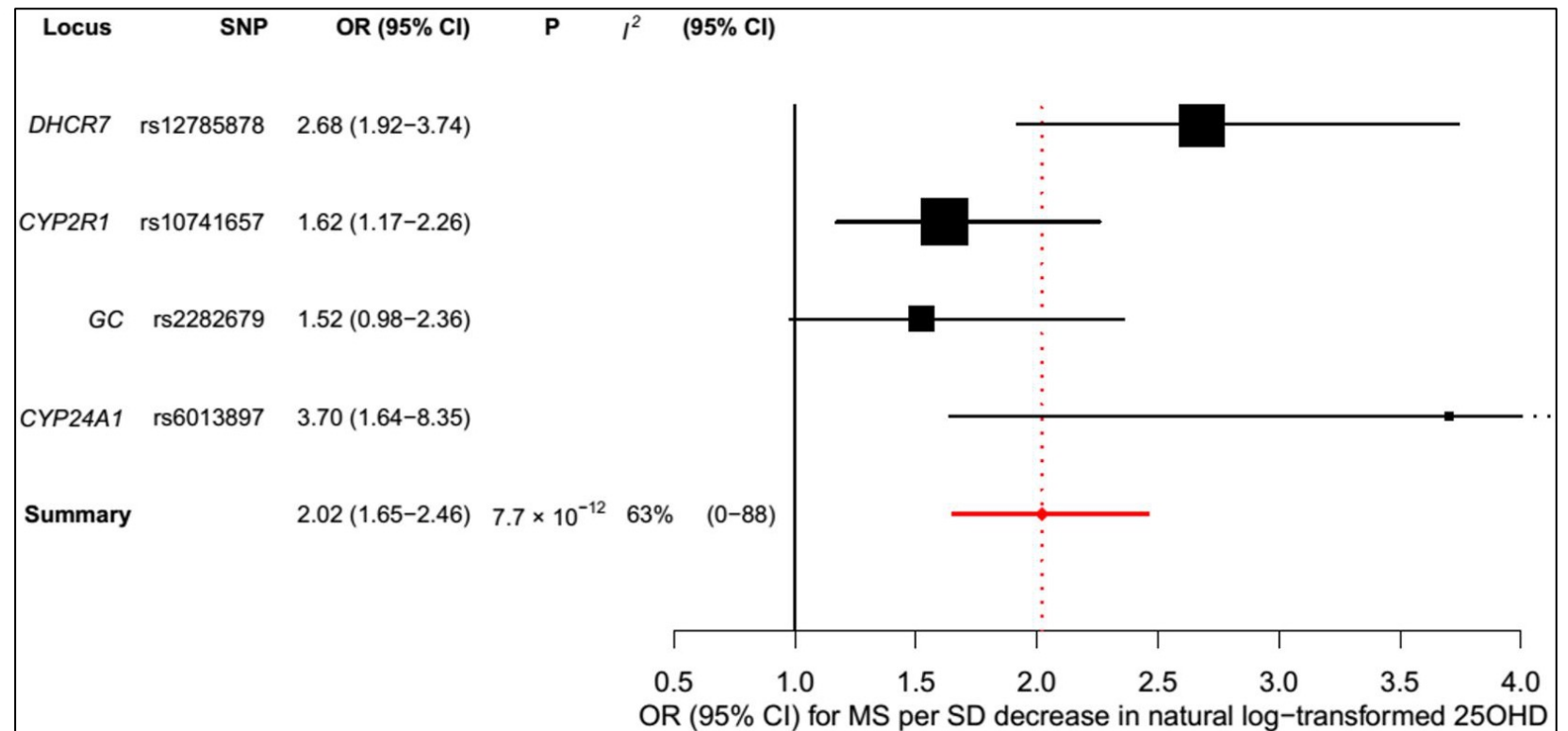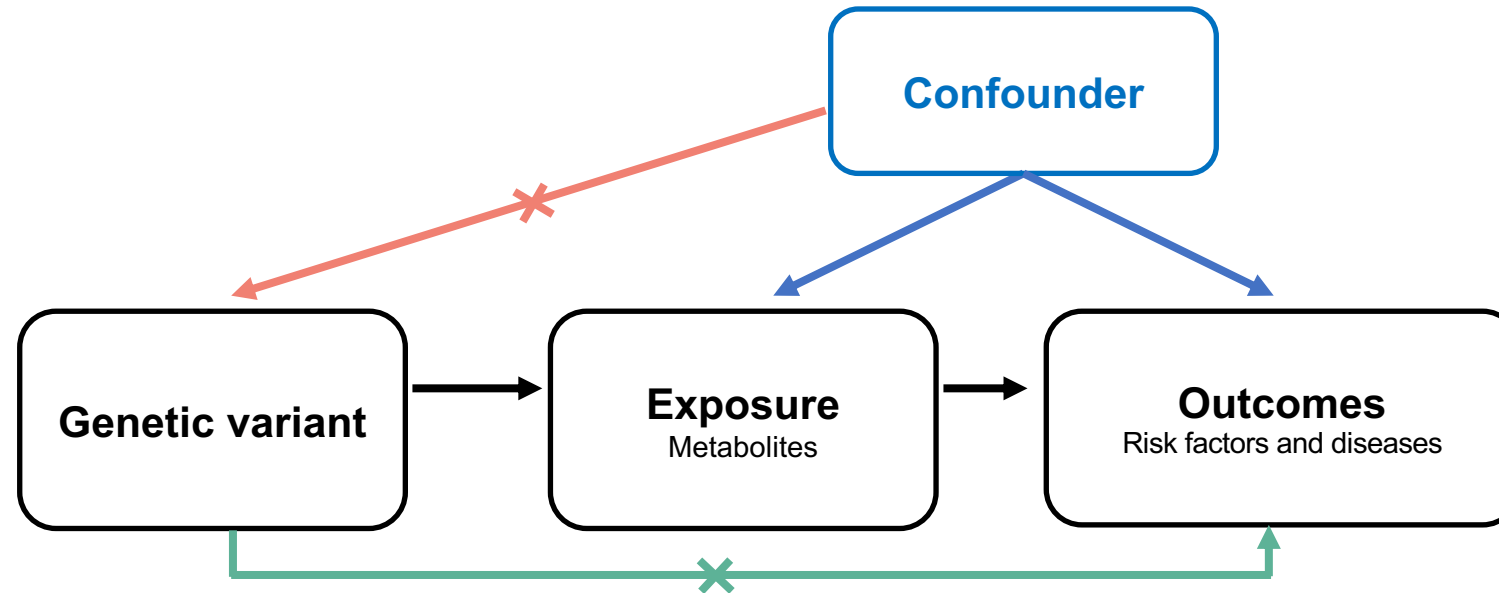Celia M. T. Greenwood, George Thanassoulis, J. Brent Richards ✉

Fig 4. Mendelian randomization estimate of the association of 25OHD level with risk of multiple sclerosis.

Mendelian Randomization assumptions

Assumptions:
1. Genetic variants are associated with exposure.
2. Genetic variants are not associated with confounders of exposure and outcome.
3. Genetics variants only associated with the outcome through the exposure.

(Skrivankova et al., 2021)

Mendelian Randomization

- **Definition**: use measured variation in genes of known phenotypes to examine the causal effect of a modifiable exposure (e.g., vitamin D supplementation) on disease (e.g., multiples sclerosis)

- **Advantages**:
    - Avoid confounding by simulating randomized control trials: According to Mendel's second law - Independent assortment: <u>alleles sort into gametes independently and randomly</u>
    - Avoid reverse causality: <u>genes function before the occurrence of disease outcomes</u>.

Mendelian Randomization for metabolite and diseases
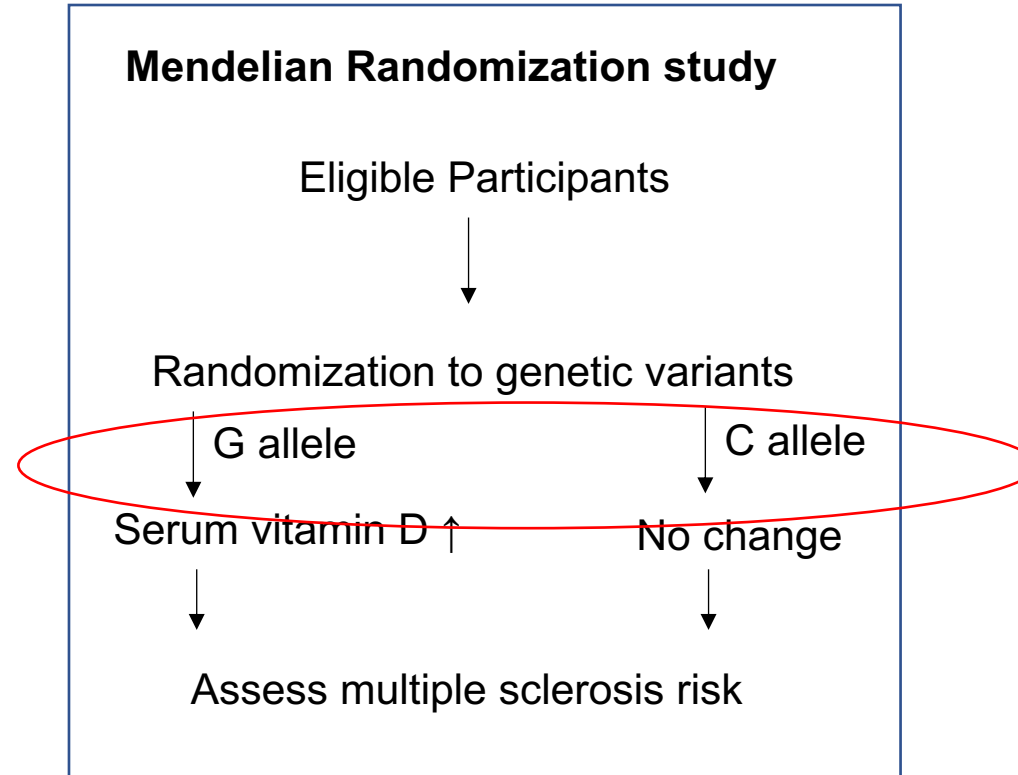
Genome-wide association study (GWAS) to identify genetic variants that are associated with vitamin D levels
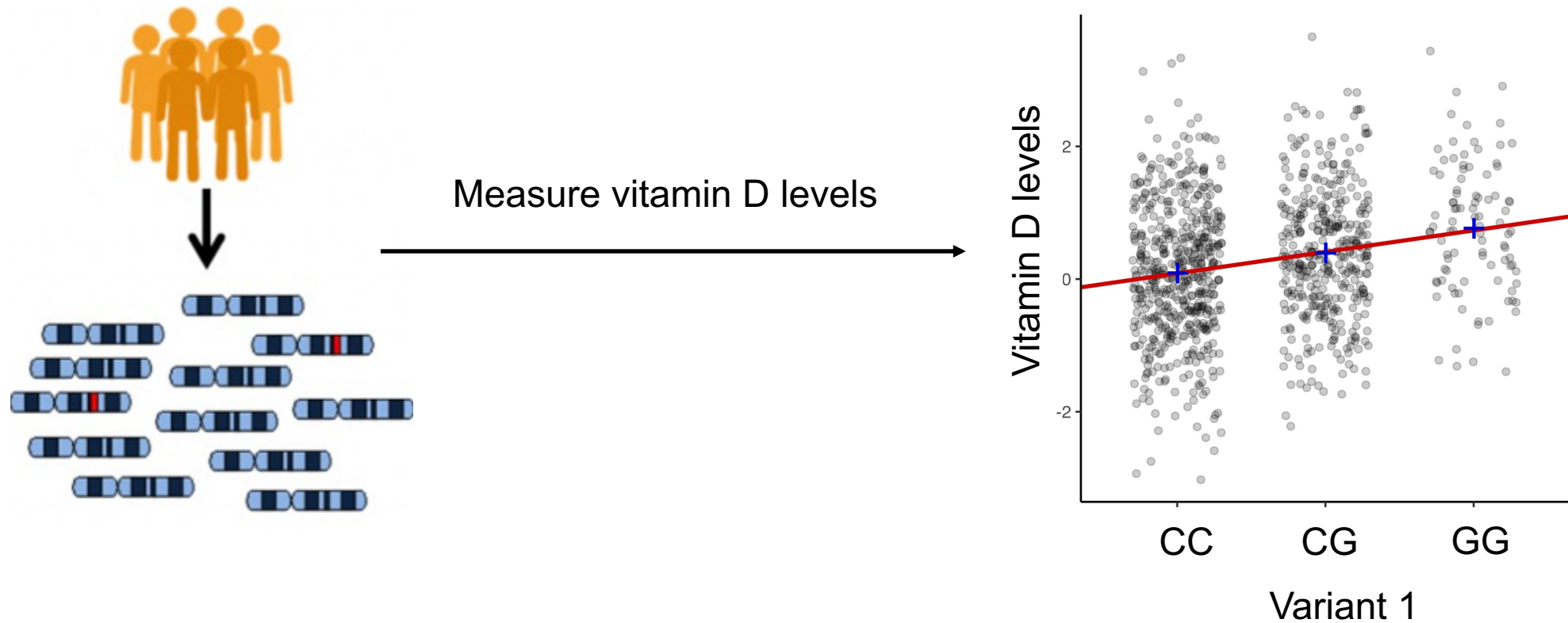
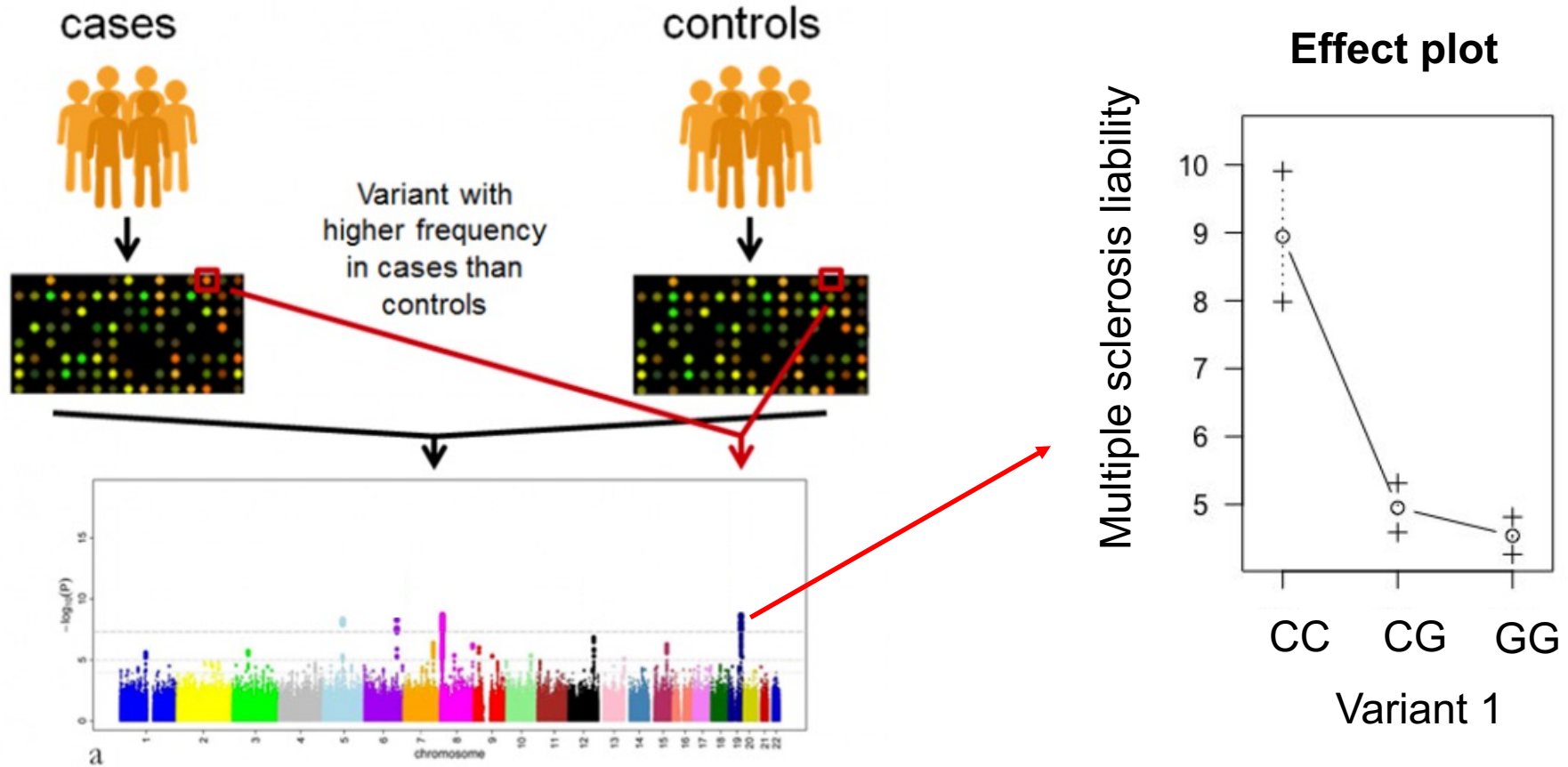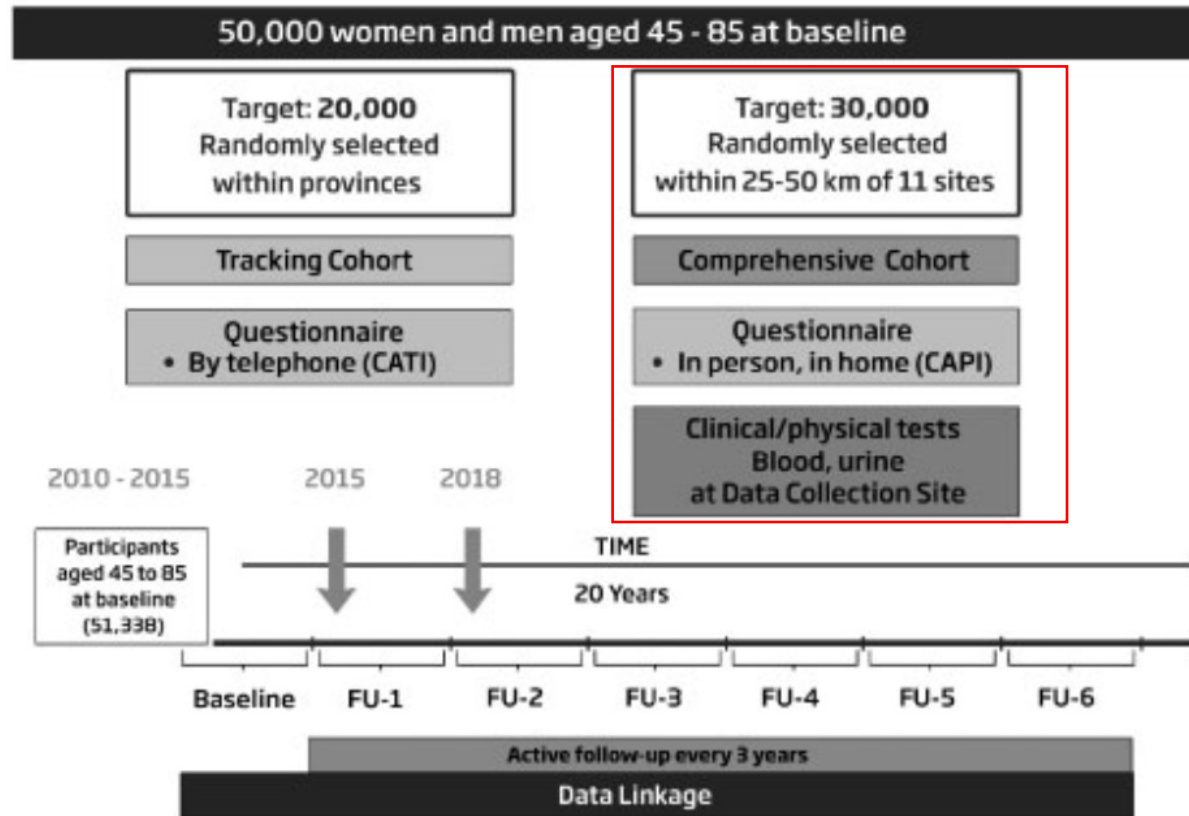# The effect of the same variant on multiple sclerosis assessed in another GWAS

16

Table of contents

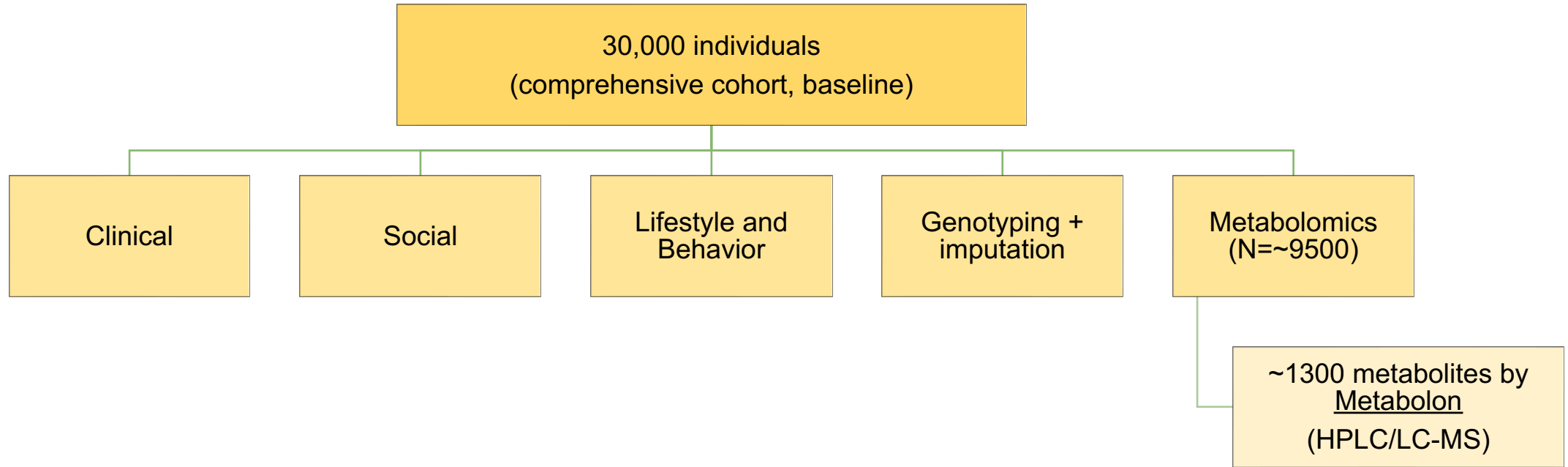# Canadian Longitudinal study of aging (CLSA)

Figure 1. CLSA data collection timeline.

The CLSA Comprehensive cohort are:
- more educated,
- have higher household income,
- more Canadian born,
- better general health.

(Raina et al., 2019)

Data

```
                    ┌─────────────────────────────────┐
                    │      30,000 individuals         │
                    │ (comprehensive cohort, baseline)│
                    └─────────────────────────────────┘
```

| Clinical | Social | Lifestyle and Behavior | Genotyping + imputation | Metabolomics (N=~9500) |

~1300 metabolites by Metabolon
(HPLC/LC-MS)

Genomics data in CLSA

- 26,622 individuals
  - 50% are females
  - 93% were identified as European ancestry
- 794,409 genotyped variants
- ~308 million imputed genetic variants (using the TopMed reference panel)

(Forgetta et al., 2022)

Metabolomics data in CLSA

- EDTA plasma

- HPLC/LC-MS (by Metabolon Inc.)

- 1,314 biochemicals, with 1,071 compounds of known identity (named biochemicals) and 243 compounds of unknown structural identity (unnamed biochemicals).

- Metabolomics data after different normalization and imputation steps are provided
  - Data after batch normalization or QC matrix normalization
  - Data without or with imputed values with minimum value detected for a given metabolite

Table of contents

- Study
  - Research aims
  - Design
  - Results
  - Conclusion

Genomic atlas of the plasma metabolome prioritizes metabolites implicated in human diseases
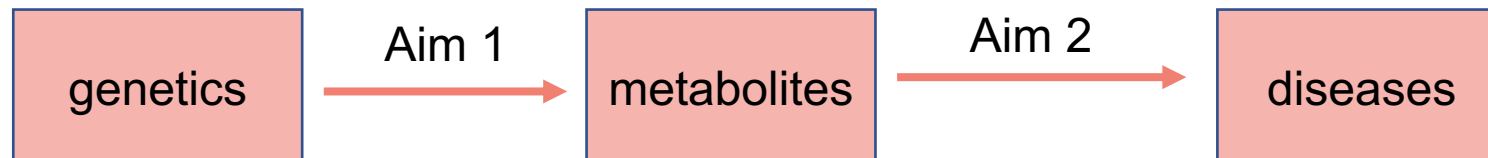
Yiheng Chen[1,2], Tianyuan Lu[1,3,4], Ulrika Pettersson-Kymmer[5], Isobel D. Stewart[6], Guillaume Butler-Laporte[1,7], Tomoko Nakanishi[1,2,8,9], Agustin Cerani[1,7], Kevin Y. H. Liang[1,3], Satoshi Yoshiji[1,2,8,9], Julian Daniel Sunday Willett[1,3,10], Chen-Yang Su[1,11], Parminder Raina[12,13], Celia M. T. Greenwood[1,3,7,14], Yossi Farjoun[1,4,15,16], Vincenzo Forgetta[1,4], Claudia Langenberg[17,18,6], Sirui Zhou[1,2], Claes Ohlsson[19,20] & J. Brent Richards[1,2,4,7,21,22]

Project aims

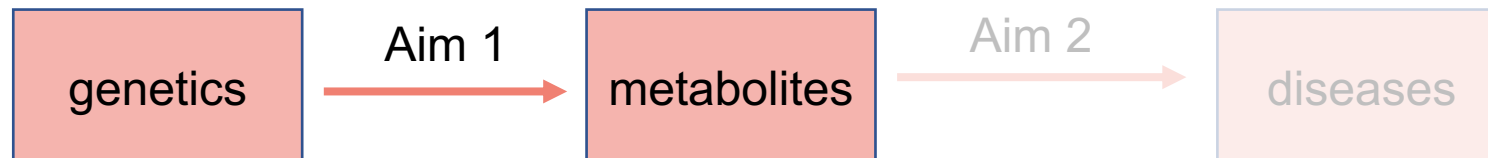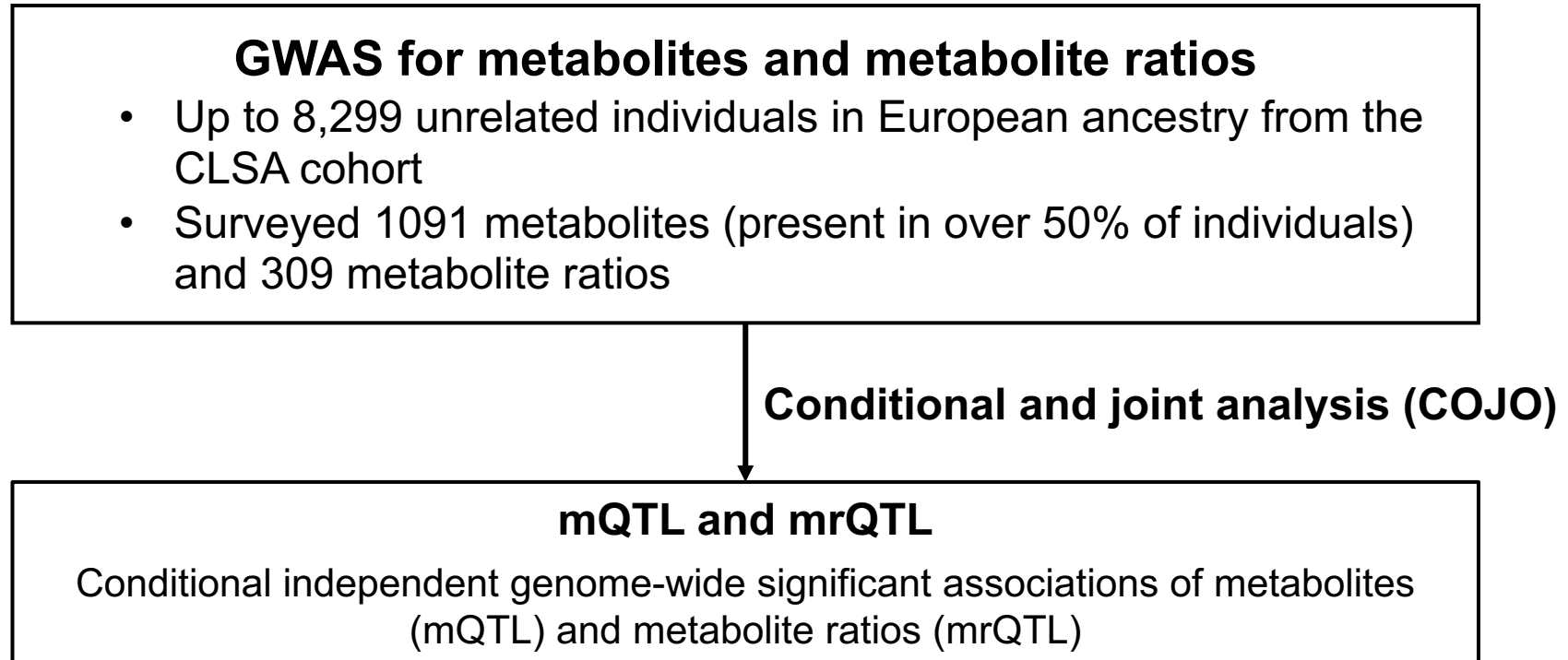Aim 1: Identify the genetic determinants of circulating metabolites

Aim 2: Identify potentially causal metabolites for 12 traits and diseases using Mendelian Randomization

Project aims

Aim 1: Identify the genetic determinants of circulating metabolites

Aim 2: Identify potentially causal metabolites for 12 traits and diseases using Mendelian Randomization

| genetics | → Aim 1 → | metabolites | → Aim 2 → | diseases |

Analysis design – GWAS

GWAS for metabolites and metabolite ratios
- Up to 8,299 unrelated individuals in European ancestry from the CLSA cohort
- Surveyed 1091 metabolites (present in over 50% of individuals) and 309 metabolite ratios

Conditional and joint analysis (COJO)

mQTL and mrQTL

Conditional independent genome-wide significant associations of metabolites (mQTL) and metabolite ratios (mrQTL)

mQTL: metabolite quantitative trait loci
mrQTL: metabolite ratio quantitative trait loci

Results
-- Associations of metabolite levels and genetic loci

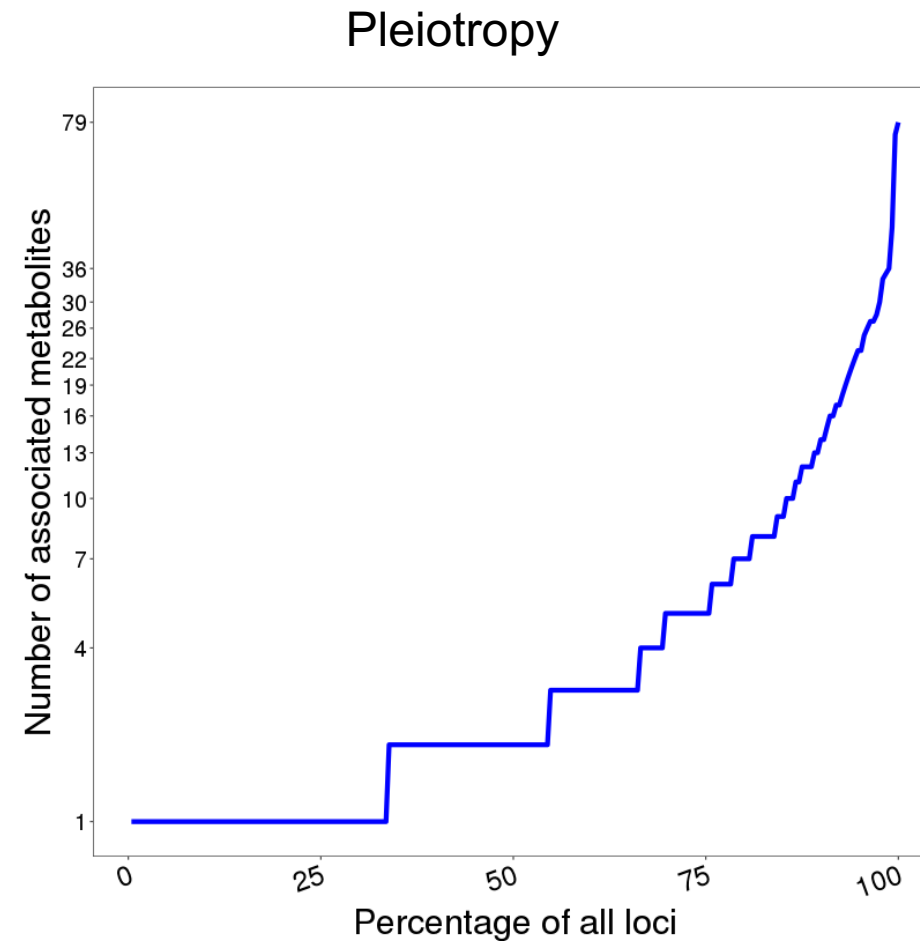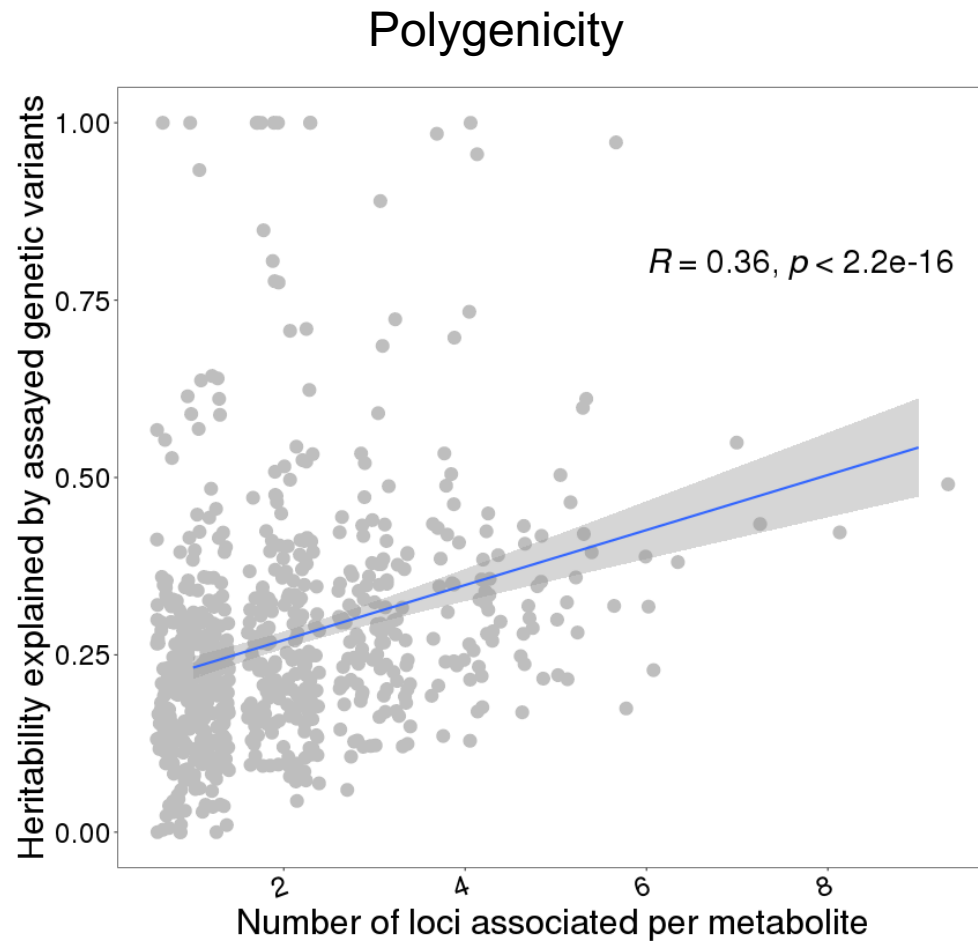1,702 independent variant–metabolite associations from 690 metabolites
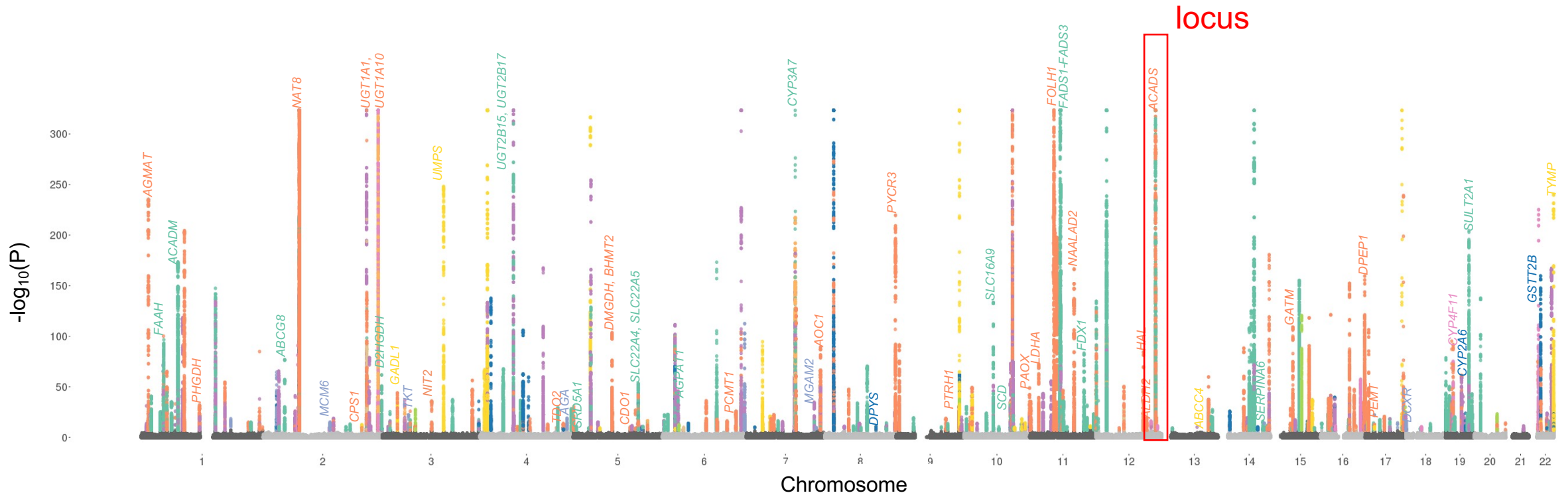
Results
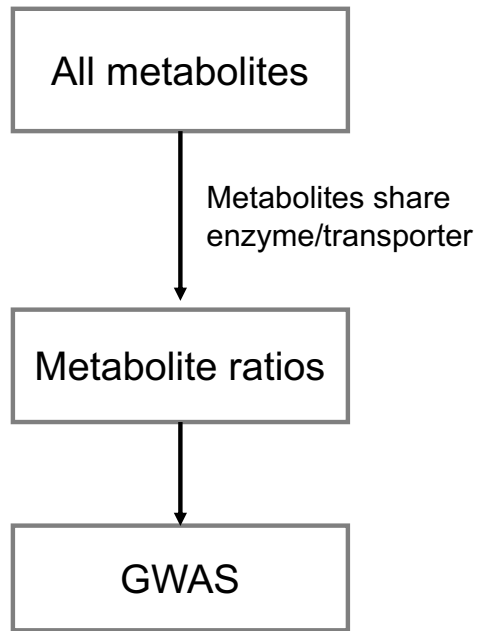## -- Novelty of associations and heritability of metabolite levels
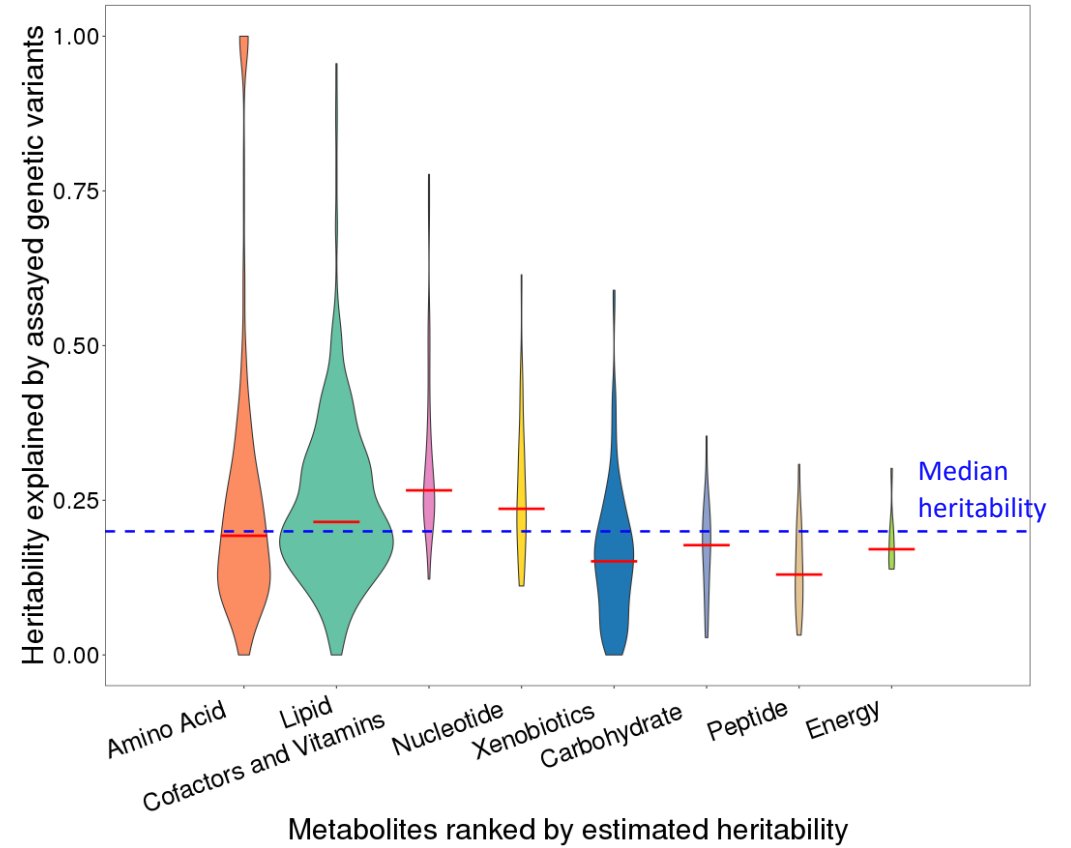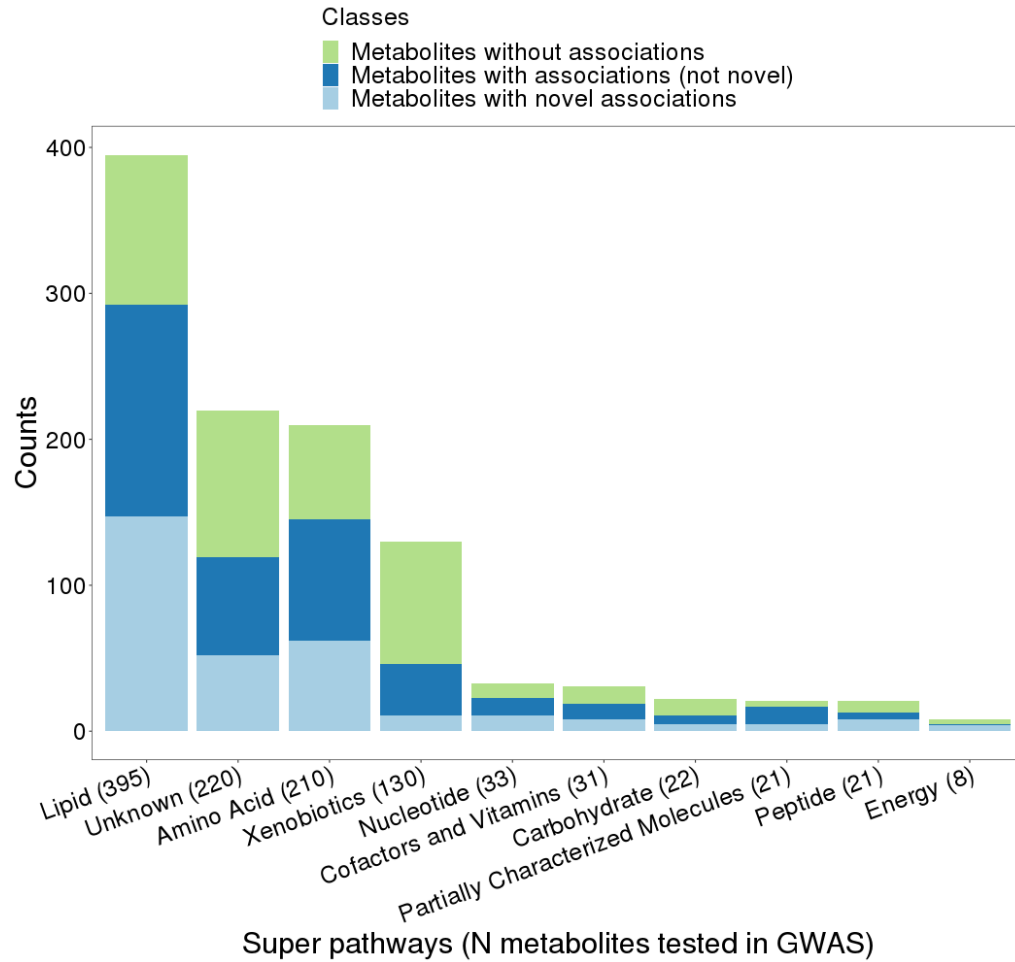
Results
-- Genetic architecture of metabolite levels

Polygenicity

Pleiotropy

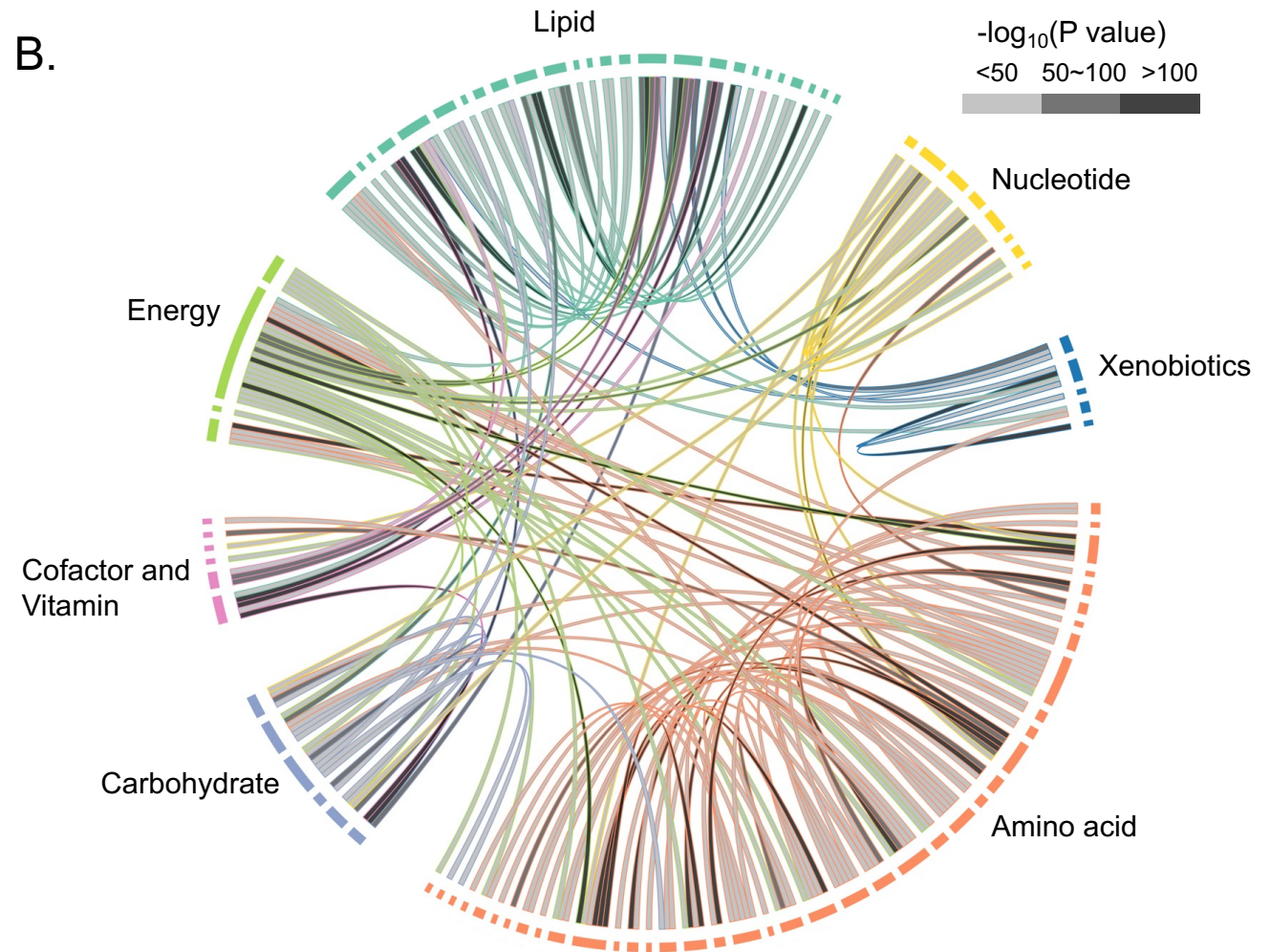# Results
-- Associations of metabolite ratios and genetic loci

Results

-- Associations of metabolite ratios and genetic loci

- 16 additional associations that were not captured by single metabolite GWAS

- E.g., rs2472297-caffeine/paraxanthine
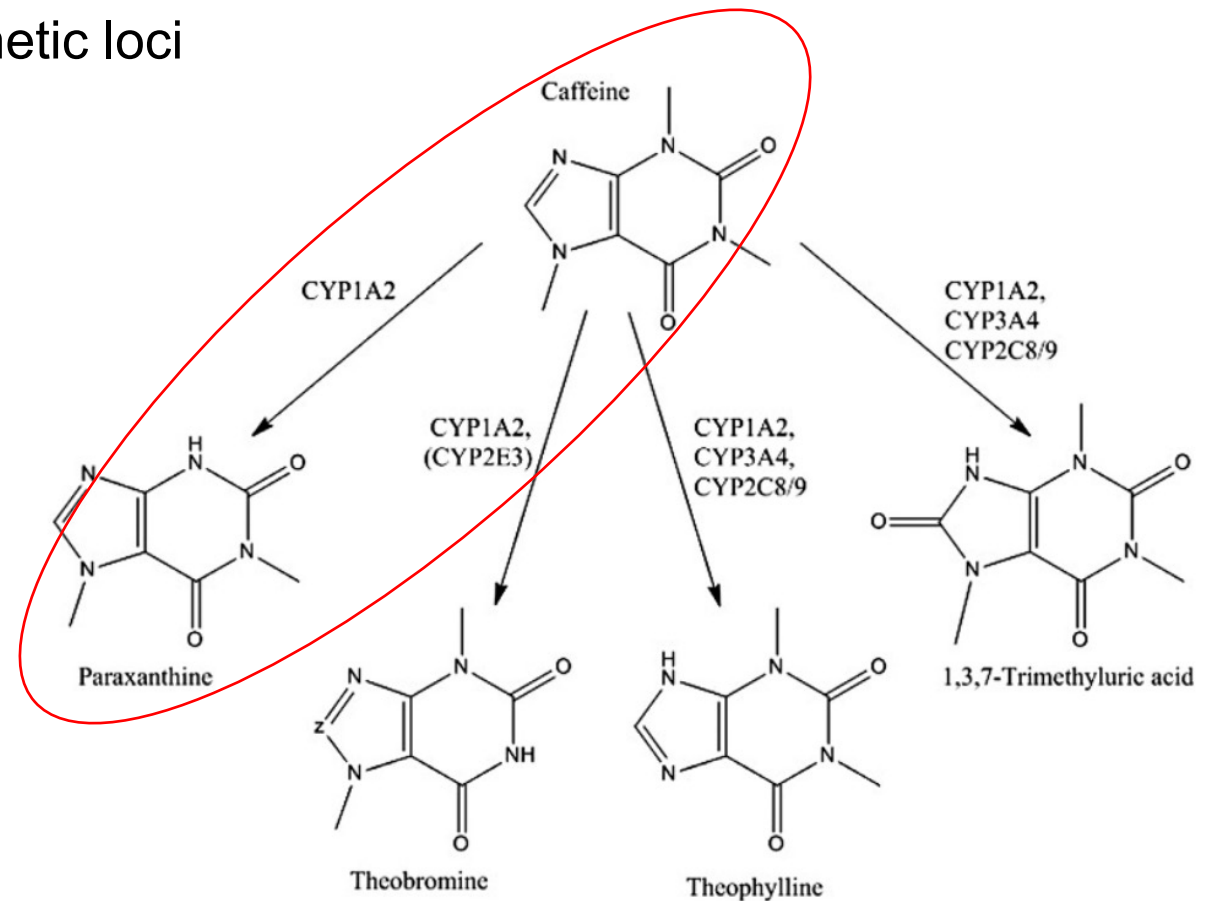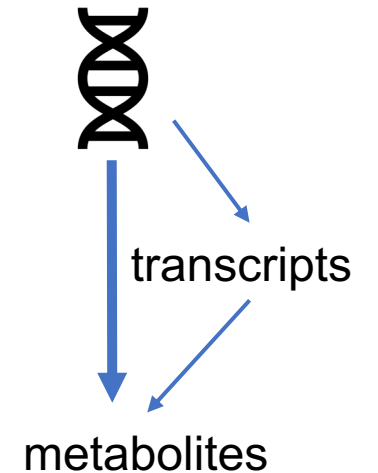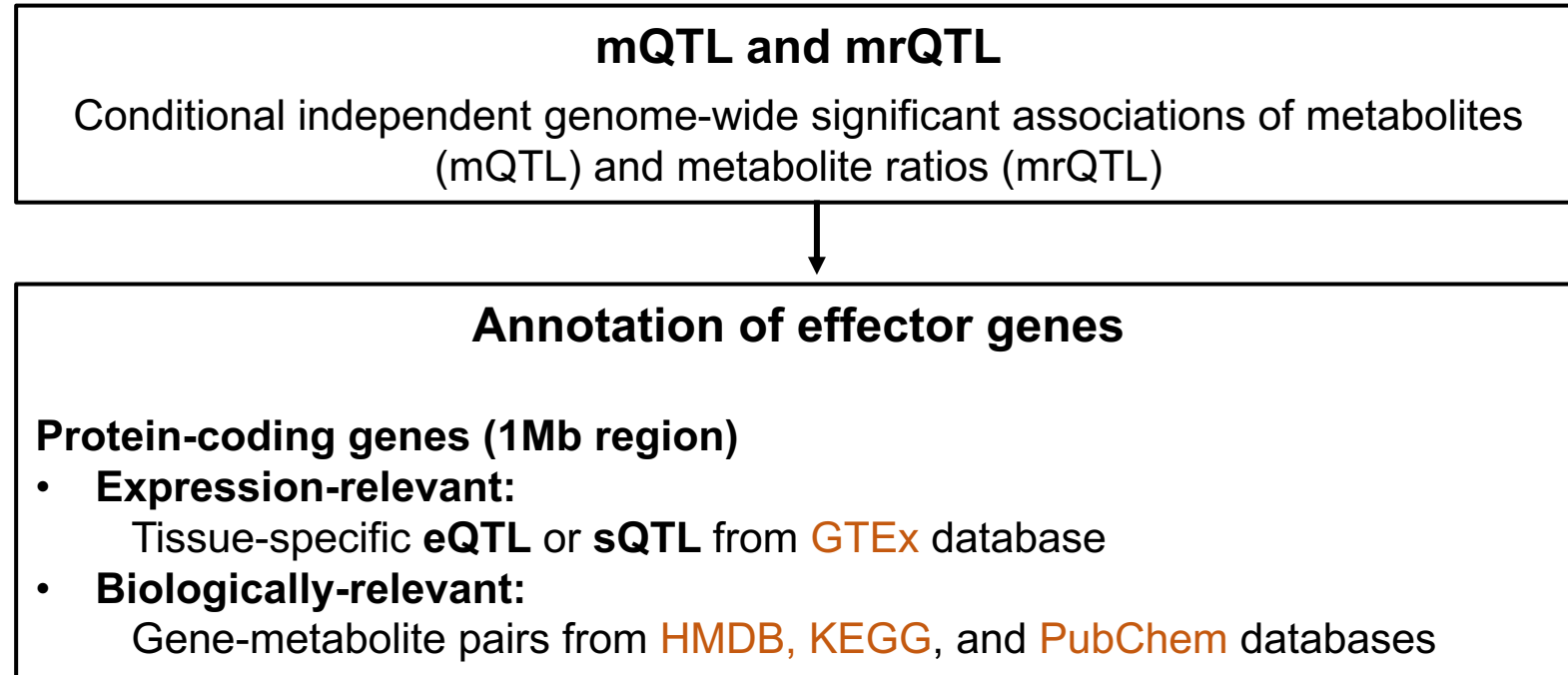
(the closest gene is *CYP1A2*)

FIGURE 4.2

Caffeine metabolism.

30

Analysis design – identification of effector genes

---

mQTL and mrQTL

Conditional independent genome-wide significant associations of metabolites (mQTL) and metabolite ratios (mrQTL)

Annotation of effector genes

Protein-coding genes (1Mb region)
- **Expression-relevant:**
  Tissue-specific **eQTL** or **sQTL** from GTEx database
- **Biologically-relevant:**
  Gene-metabolite pairs from HMDB, KEGG, and PubChem databases

---

transcripts

metabolites

eQTL: expression quantitative trait loci
sQTL: splicing quantitative trait loci

31

Results
-- Assignment of effector genes

a

Annotation of genes per locus
Protein-coding genes (1Mb region)
- **Expression-relevant:**
  - Tissue-specific *cis*-eQTL or *cis*-sQTL
- **Biologically-relevant:**
  - Genes encode proteins that metabolize the metabolites

Number of genes prioritized

Expression-relevant genes

Biologically-relevant genes

451   94   168

b

6.4%

7.4%

86.2%

Protein Types
- Enzyme
- Transporter
- Others

8

11

2

14

8
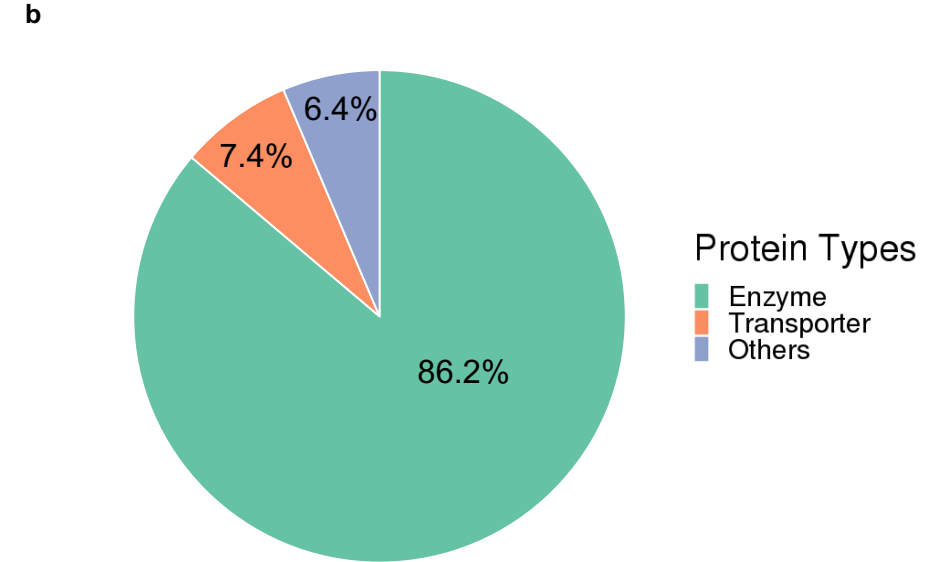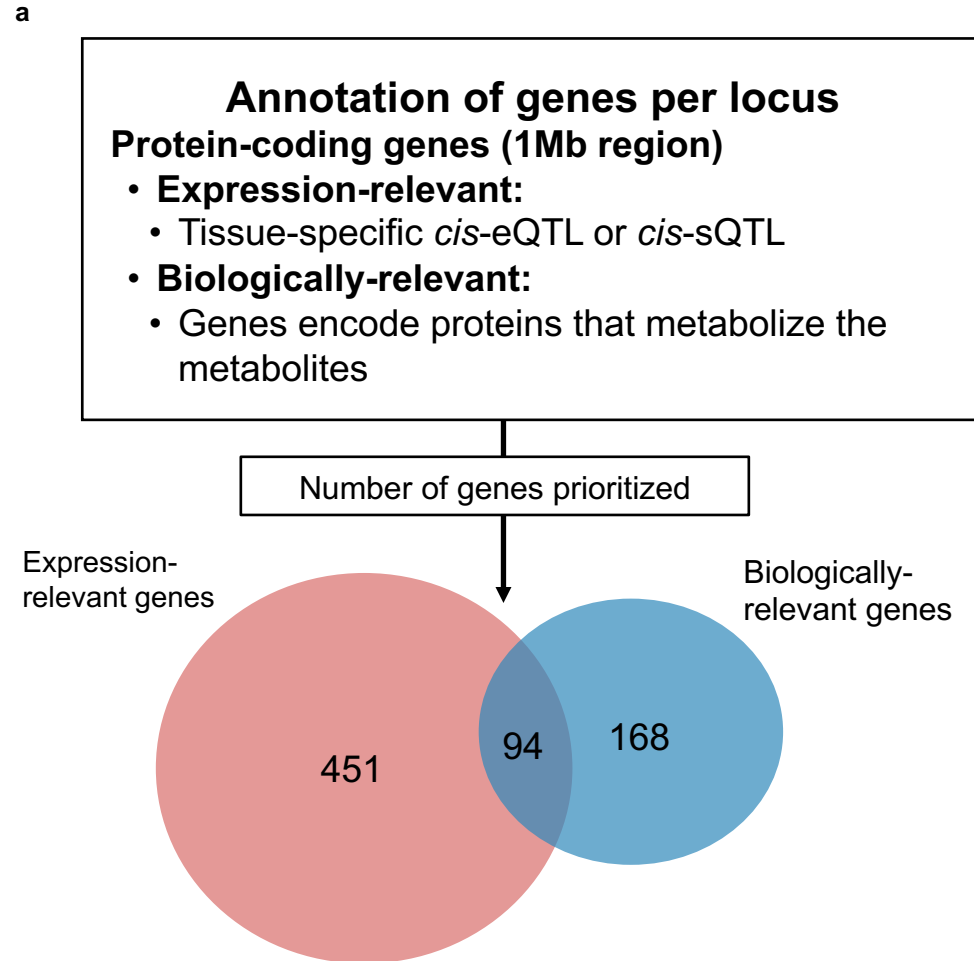
Results
-- Explored the linkage of the effector genes with drug and phenotype information

c
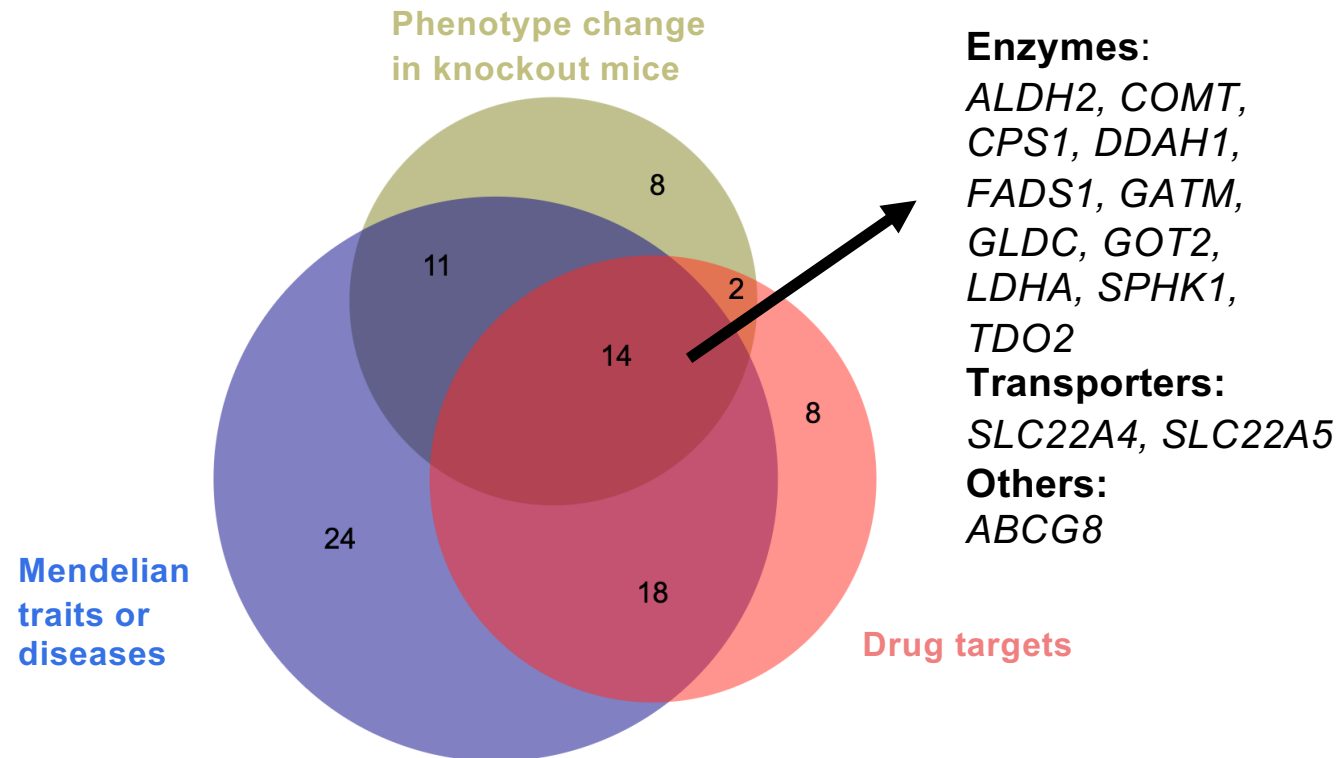
Phenotype change in knockout mice

Mendelian traits or diseases

Drug targets

Enzymes:
ALDH2, COMT, CPS1, DDAH1, FADS1, GATM, GLDC, GOT2, LDHA, SPHK1, TDO2
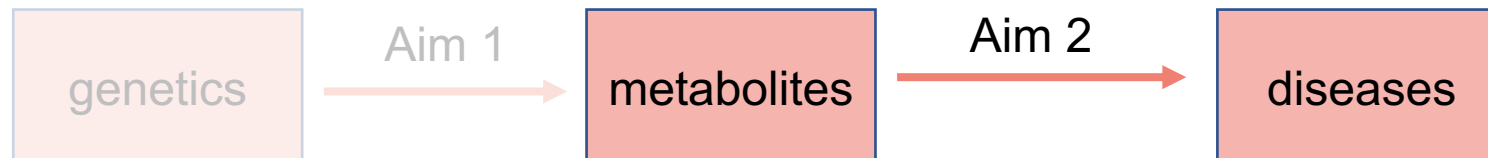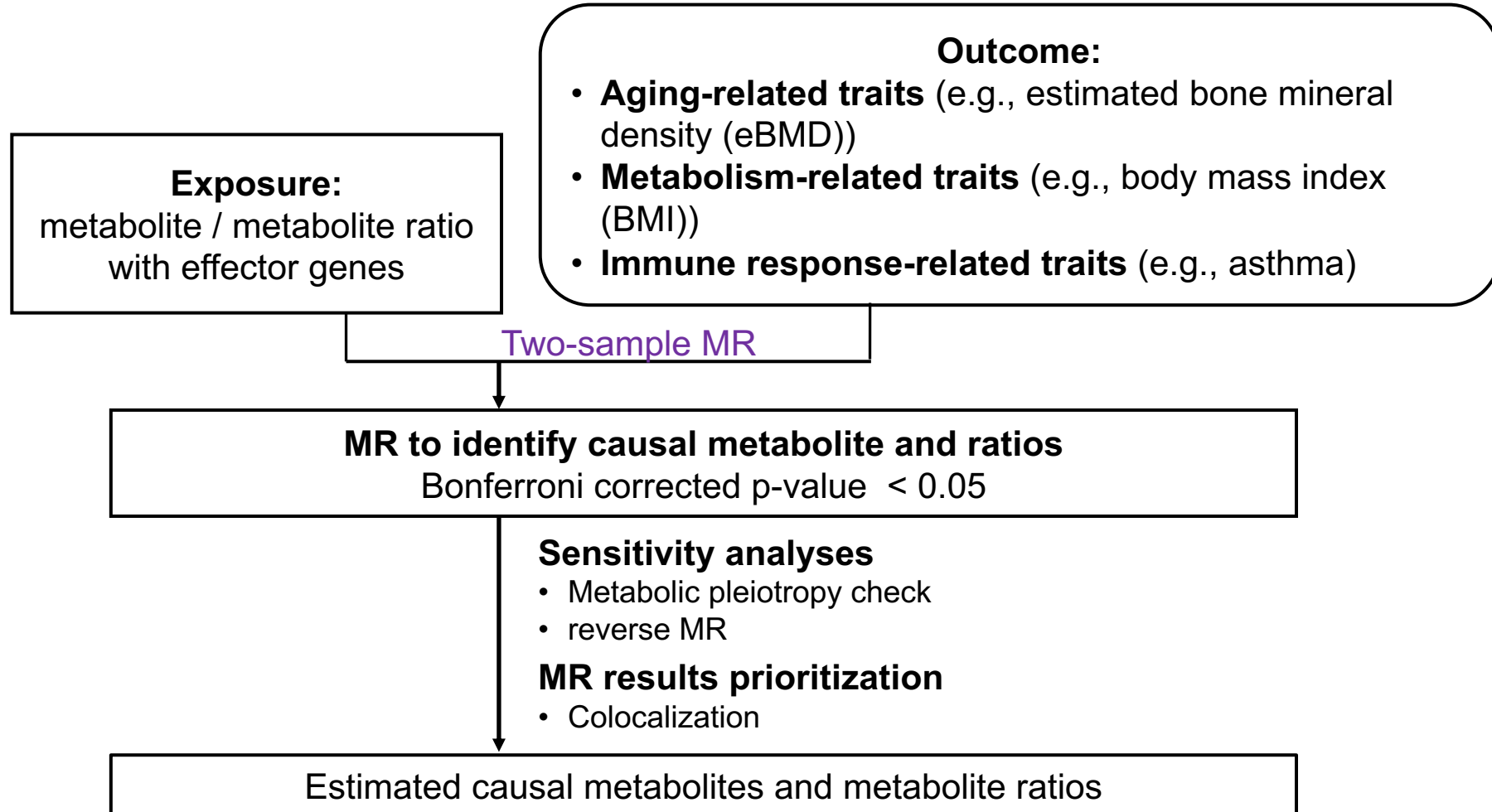**Transporters**:
SLC22A4, SLC22A5
Others:
ABCG8

Project aims

Aim 1: Identify the genetic determinants of circulating metabolites

Aim 2: Identify potentially causal metabolites for 12 traits and diseases using Mendelian Randomization

# **Analysis design** – mendelian randomization

Exposure:
metabolite / metabolite ratio
with effector genes

Outcome:
- **Aging-related traits** (e.g., estimated bone mineral density (eBMD))
- **Metabolism-related traits** (e.g., body mass index (BMI))
- **Immune response-related traits** (e.g., asthma)

Two-sample MR

MR to identify causal metabolite and ratios
Bonferroni corrected p-value < 0.05

Sensitivity analyses
- Metabolic pleiotropy check
- reverse MR

MR results prioritization
- Colocalization

Estimated causal metabolites and metabolite ratios
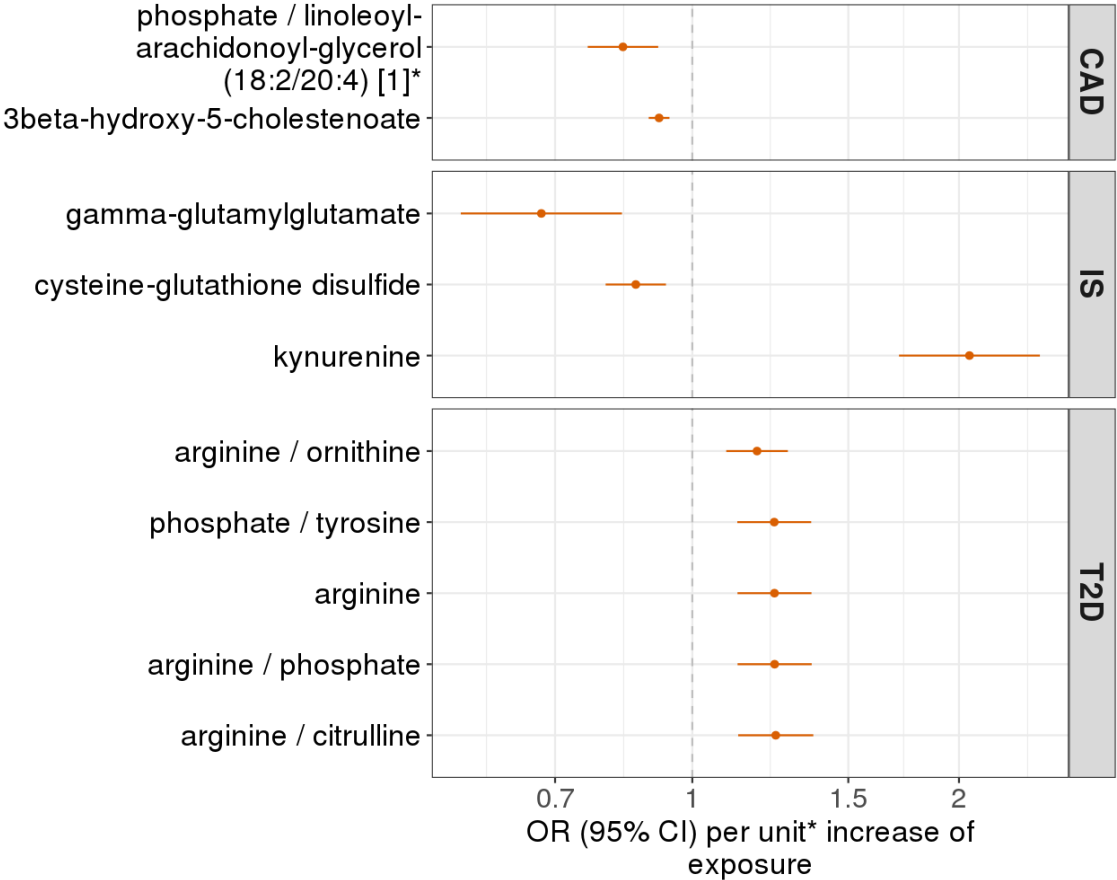
Results
-- Metabolites and metabolite ratios that have estimated causal effect on <u>aging-related</u> traits and diseases
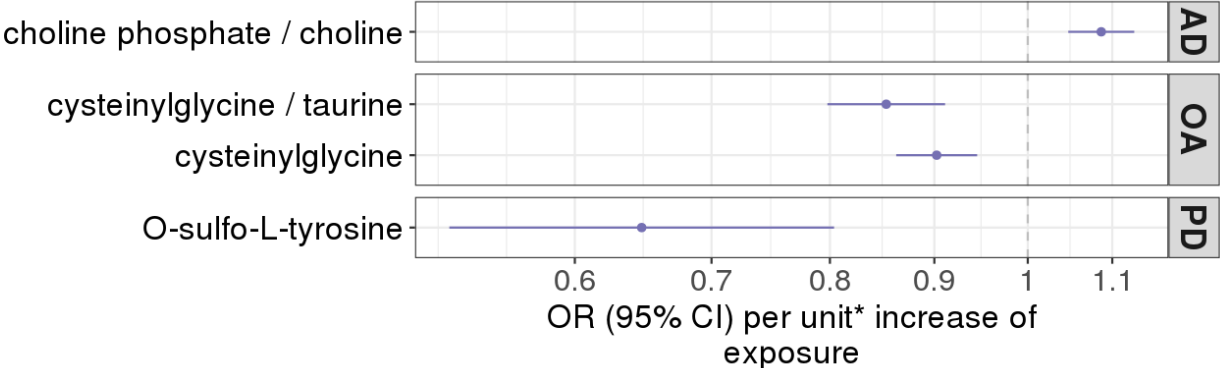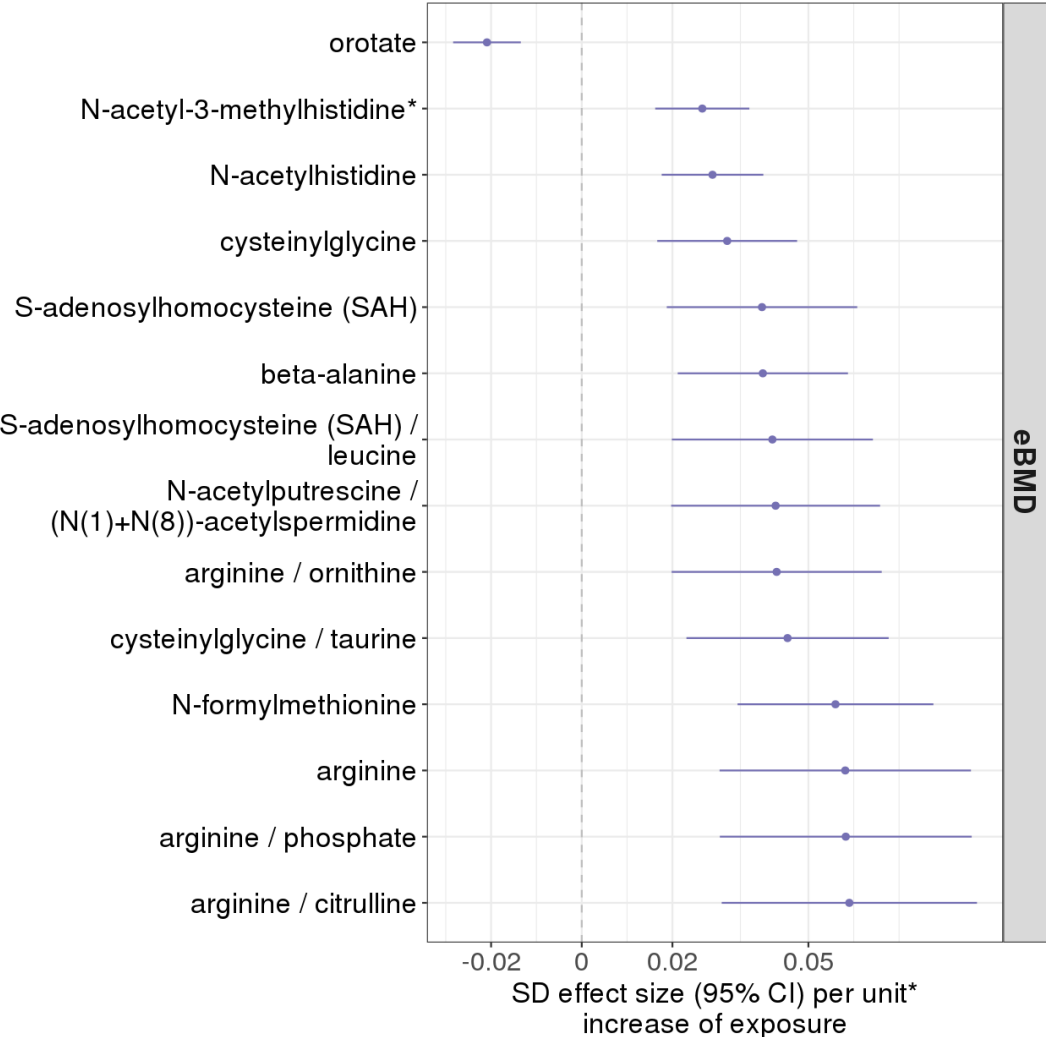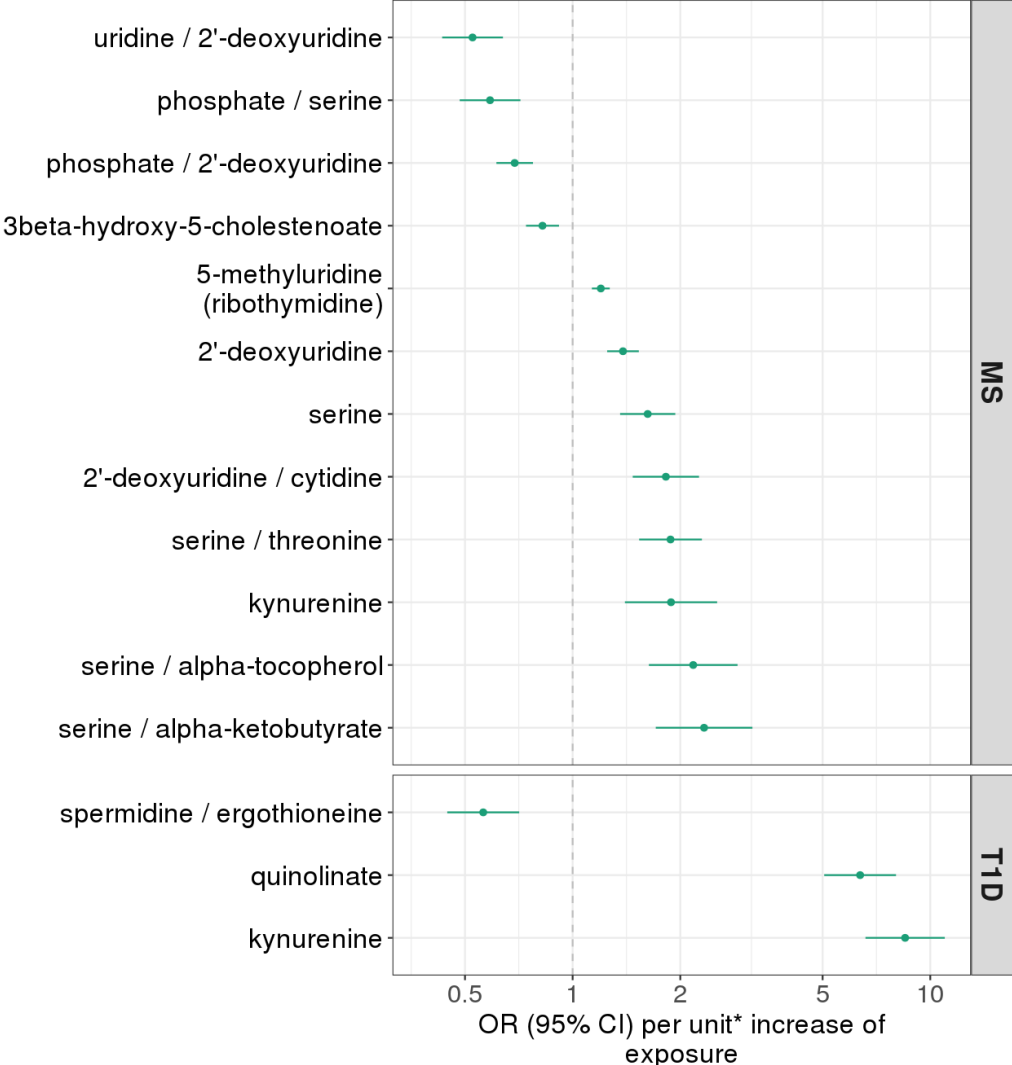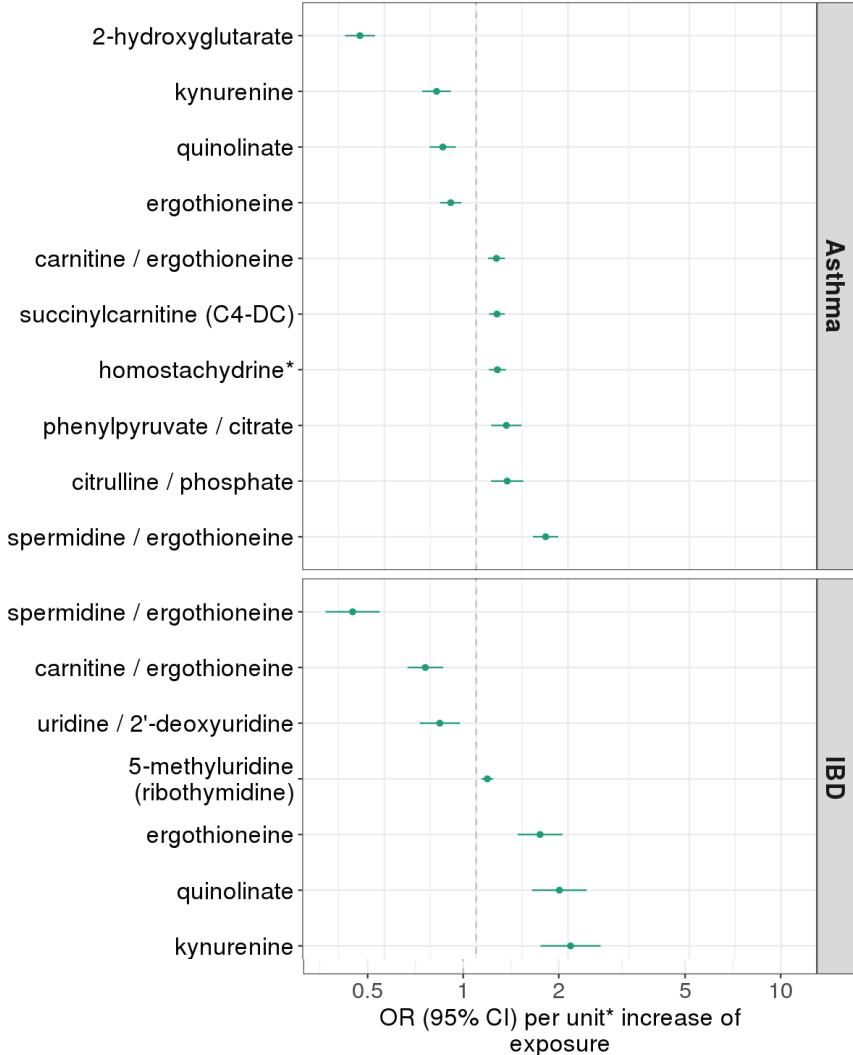
Results
-- Metabolites and metabolite ratios that have estimated causal effect on <u>metabolism-related</u> traits and diseases
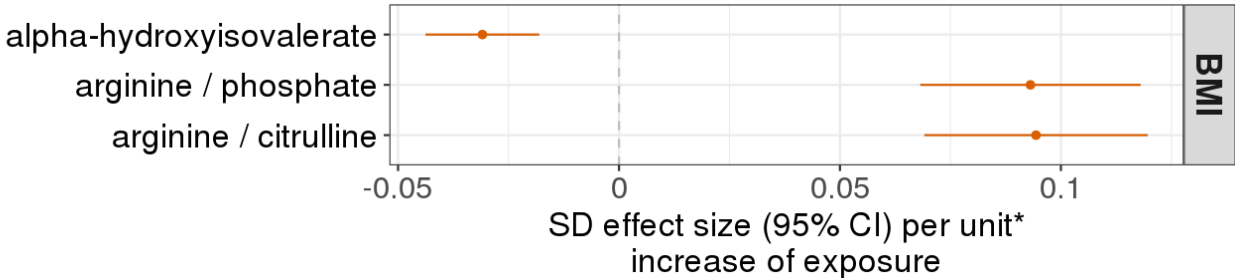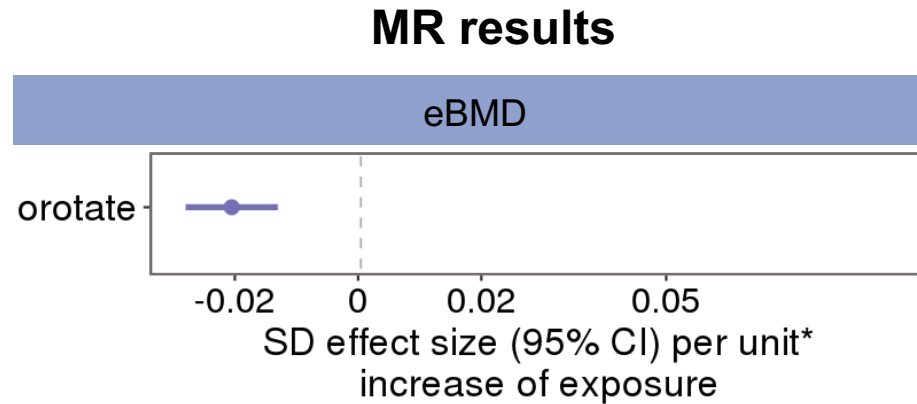
Results
-- Metabolites and metabolite ratios that have estimated causal effect on <u>immune-related</u> traits and diseases

Results
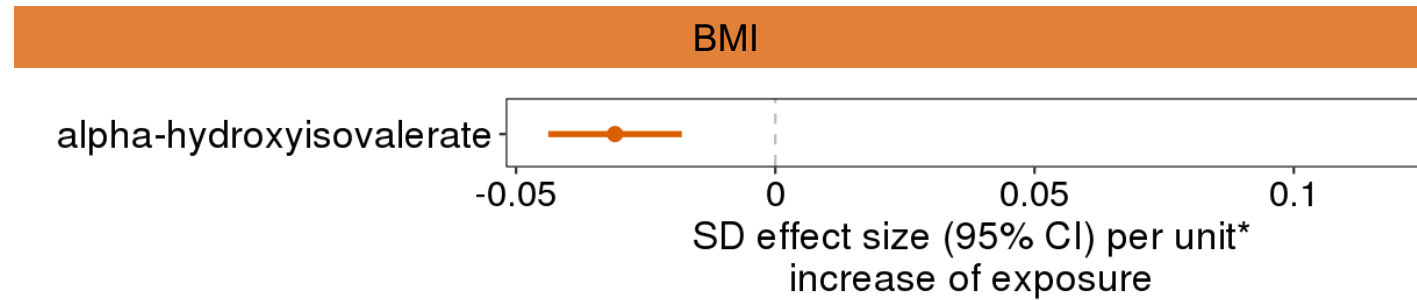## -- metabolite with estimated causal effect on eBMD

MR results

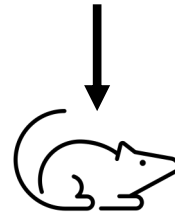**Observational results:** Association between orotate and hip fracture risk in an independent cohort

	OR	95% CI	P value
orotate	1.15	1.08-1.15	1.3×10^{-5}

Results
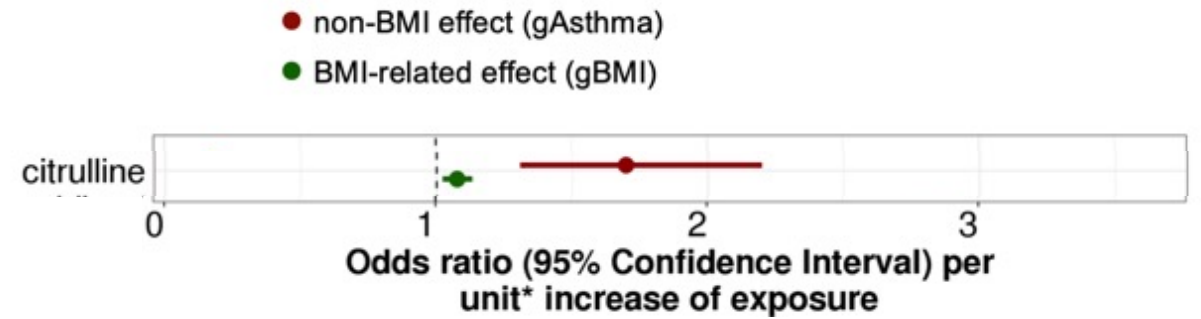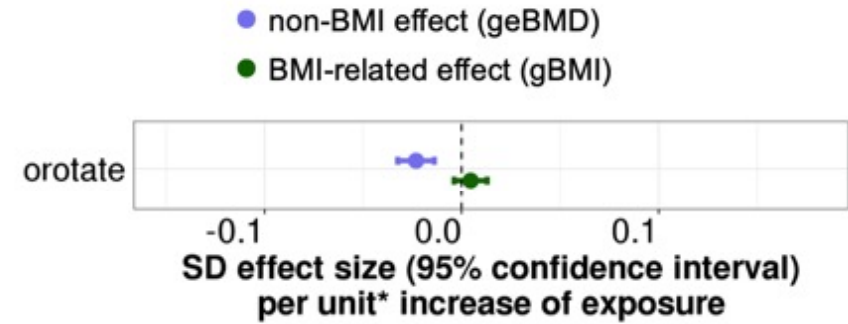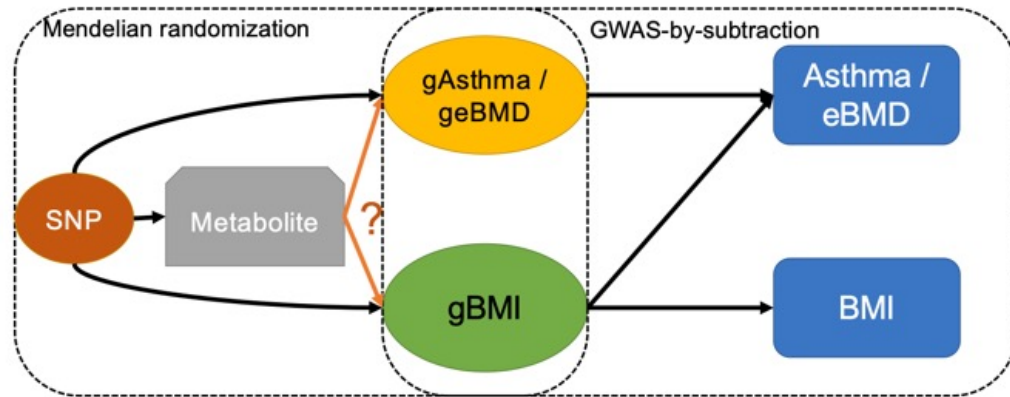## -- metabolite with estimated causal effect on BMI

alpha-hydroxyisovalerate - lactate dehydrogenase (*LDHA)*

Ldha heterozygous mice have decreased total fat amount (p=9.94×10^{-7})
source: IMPC database

Results
--Effect of metabolite on eBMD and asthma that are dependent or independent of BMI

Limitation of the study

- Most available disease and trait GWAS are from individuals in European ancestry

- Metabolomics data are relative measurements

- Low statistical power of MR

Conclusion

- We identified genetic determinants of circulating metabolites.

- We inferred the causal effect of metabolite levels and ratios on twelve traits and diseases that are predominantly influenced by different mechanisms (aging, metabolism, and immune response).

Acknowledgement

- Prof. Brent Richards
  - Yann Ilboudo
  - Tianyuan Lu
  - Satoshi Yoshiji
  - Kevin Liang
  - Takayoshi Sasako
  - Vince Forgetta
  - Julian Willett
  - Guillaume Butler-Laporte
  - Tomoko Nakanishi
  - Darin Adra
  - Joseph Farjoun
  - Laetitia Lauren
  - David Morrison
- Prof. Celia Greenwood
  - Kathleen Klein
  - Lai Jiang

- Yixiao Zeng
- Ting Zhang
- Prof. Danilo Bzdok
- Prof. Sirui Zhou
- Prof. Claudia Langenberg
  - Isobel Stewart
- Prof. Nick Wareham
- Prof. Maik Pietzner

Thank you for listening!