Pleiotropy testing identifies novel loci associated with lipid traits in multiple ethnicities
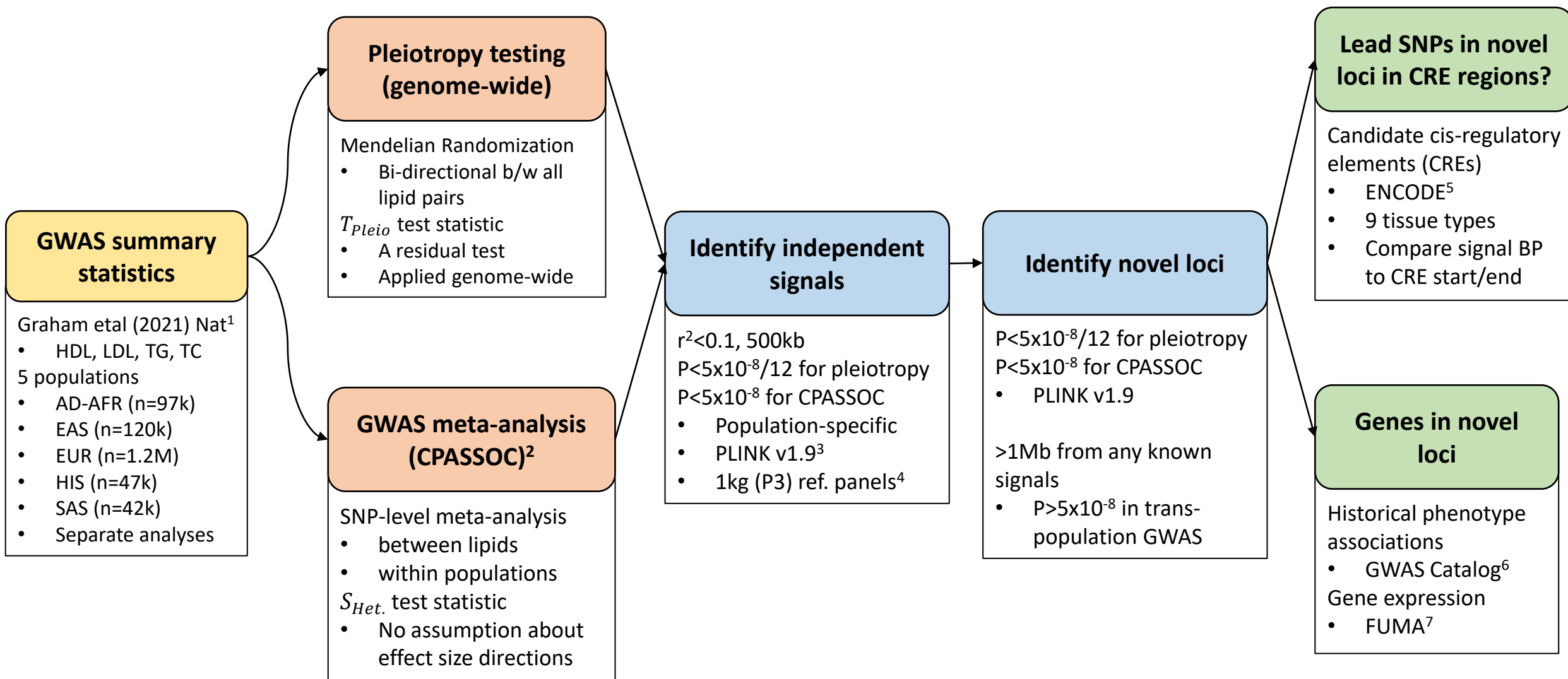
Noah Lorincz-Comi, Gen Li, Yihe Yang, Xiaofeng Zhu
Department of Population & Quantitative Health Sciences
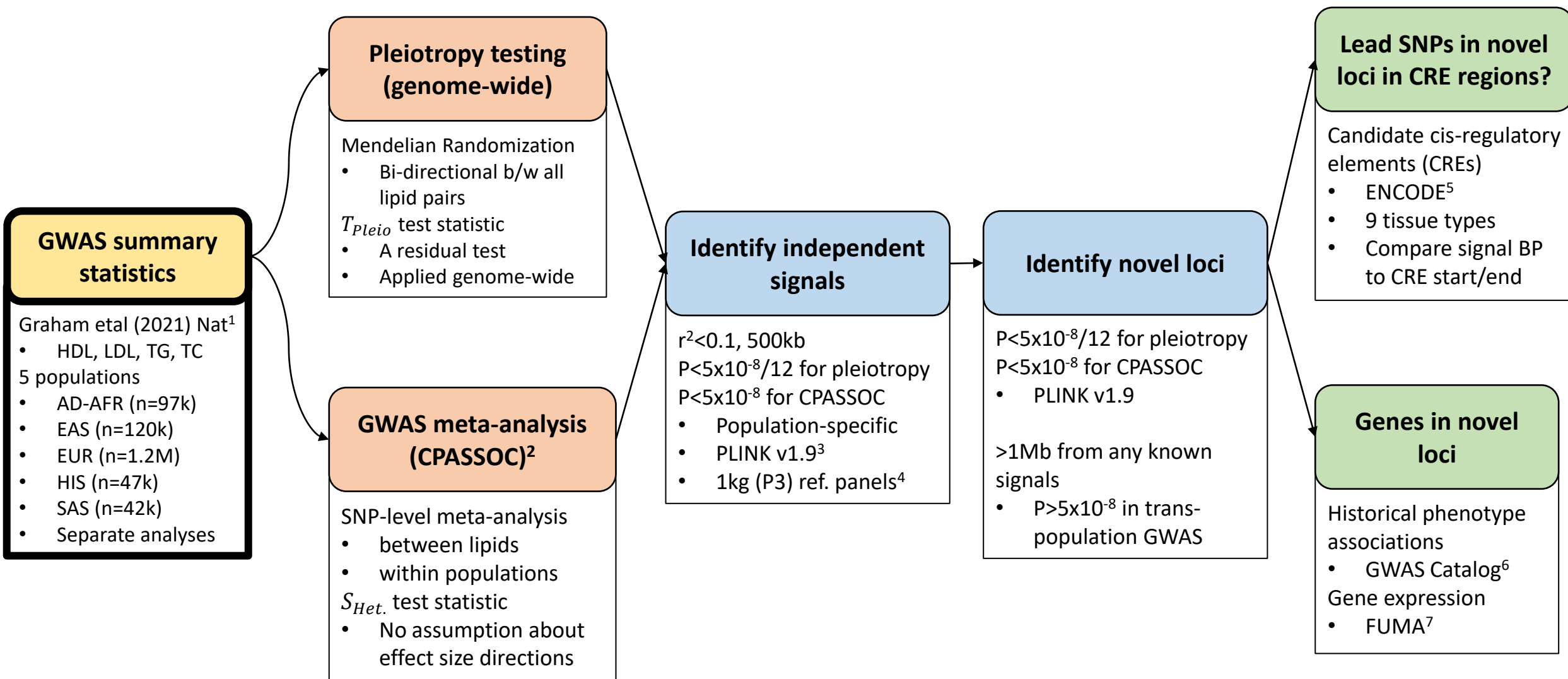Case Western Reserve University
October 26th, 2022

Disclosures

- I do not have any relationships to report within the last 24 months with ACCME defined ineligible companies.

- I will not be discussing unlabeled/investigational uses of medical devices or pharmaceuticals during this presentation.

Goal

- Identify novel loci associated with serum lipids
- Standard GWAS testing identifies many loci, but we believe some are still missed and can be found without increasing sample size
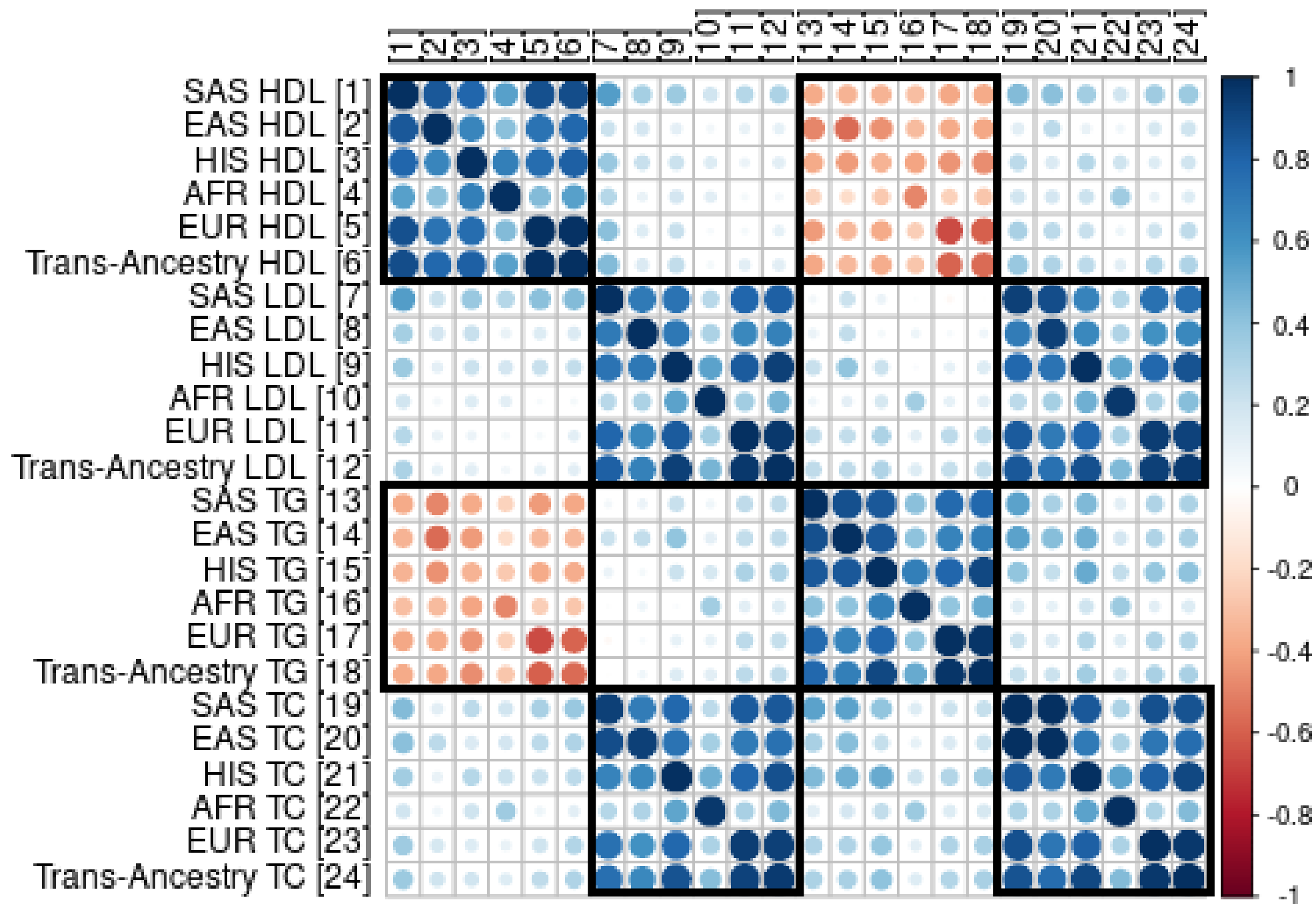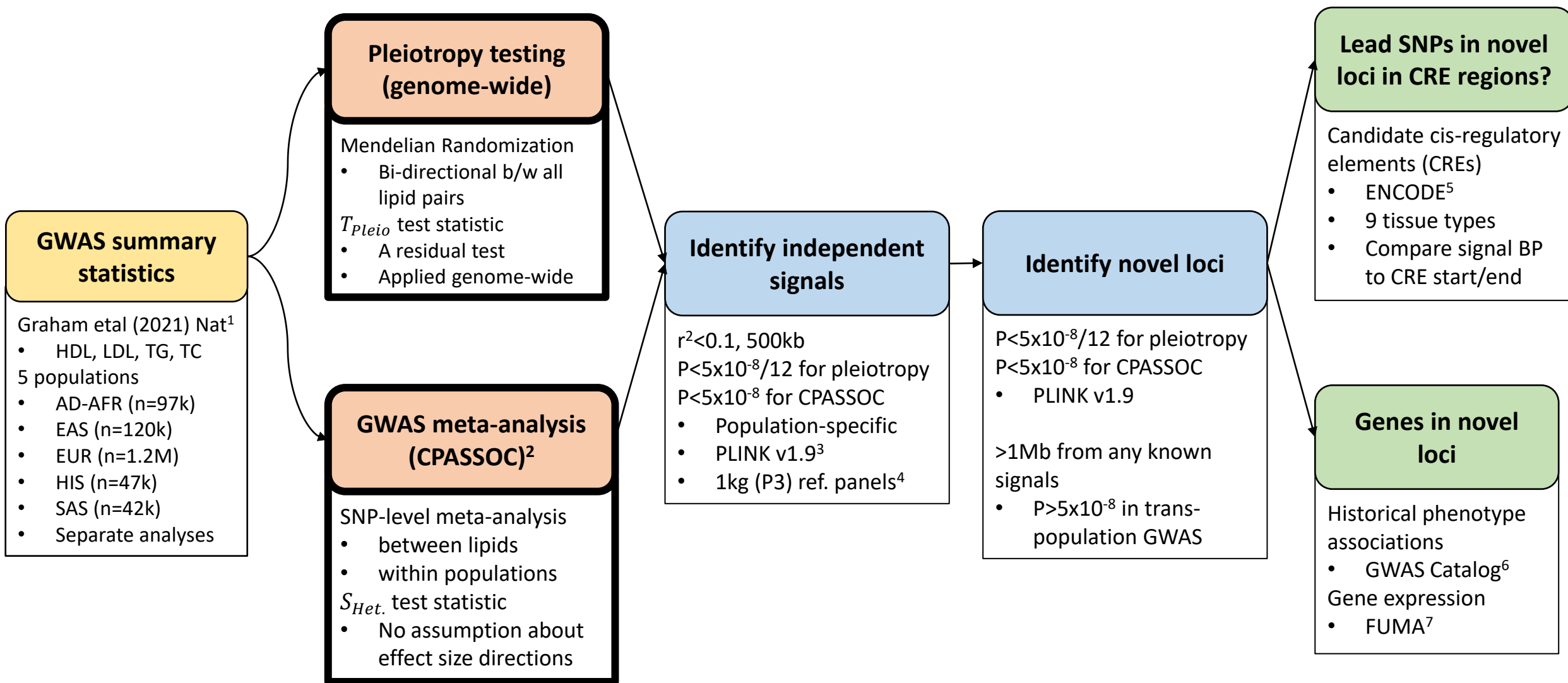
GWAS summary statistics

Graham etal (2021) Nat[1]
- HDL, LDL, TG, TC

5 populations
- AD-AFR (n=97k)
- EAS (n=120k)
- EUR (n=1.2M)
- HIS (n=47k)
- SAS (n=42k)
- Separate analyses

Pleiotropy testing (genome-wide)

Mendelian Randomization
- Bi-directional b/w all lipid pairs

T_{Pleio} test statistic
- A residual test
- Applied genome-wide

GWAS meta-analysis (CPASSOC)[2]

SNP-level meta-analysis
- between lipids
- within populations

$S_{Het.}$ test statistic
- No assumption about effect size directions

Identify independent signals

$r^2<0.1$, 500kb
$P<5 \times 10^{-8}/12$ for pleiotropy
$P<5 \times 10^{-8}$ for CPASSOC
- Population-specific
- PLINK v1.9[3]
- 1kg (P3) ref. panels[4]

Identify novel loci

$P<5 \times 10^{-8}/12$ for pleiotropy
$P<5 \times 10^{-8}$ for CPASSOC
- PLINK v1.9

>1Mb from any known signals
- $P>5 \times 10^{-8}$ in trans-population GWAS

Lead SNPs in novel loci in CRE regions?

Candidate cis-regulatory elements (CREs)
- ENCODE[5]
- 9 tissue types
- Compare signal BP to CRE start/end

Genes in novel loci

Historical phenotype associations
- GWAS Catalog[6]

Gene expression
- FUMA[7]

[1]: Graham (2021), [2] Zhu (2015), [3] Purcell (2007), [4] IGSR (2015); [5] ENCODE Cnsrtm. (2012-2020), [6] Buniello (2019), [7] Watanabe (2017)

GWAS summary statistics

Graham etal (2021) Nat[1]
- HDL, LDL, TG, TC
5 populations
- AD-AFR (n=97k)
- EAS (n=120k)
- EUR (n=1.2M)
- HIS (n=47k)
- SAS (n=42k)
- Separate analyses

Pleiotropy testing (genome-wide)

Mendelian Randomization
- Bi-directional b/w all lipid pairs
T_{Pleio} test statistic
- A residual test
- Applied genome-wide

GWAS meta-analysis (CPASSOC)[2]

SNP-level meta-analysis
- between lipids
- within populations
$S_{Het.}$ test statistic
- No assumption about effect size directions

Identify independent signals

$r^2 < 0.1$, 500kb
$P < 5 \times 10^{-8}/12$ for pleiotropy
$P < 5 \times 10^{-8}$ for CPASSOC
- Population-specific
- PLINK v1.9[3]
- 1kg (P3) ref. panels[4]

Identify novel loci

$P < 5 \times 10^{-8}/12$ for pleiotropy
$P < 5 \times 10^{-8}$ for CPASSOC
- PLINK v1.9

>1Mb from any known signals
- $P > 5 \times 10^{-8}$ in trans-population GWAS

Lead SNPs in novel loci in CRE regions?

Candidate cis-regulatory elements (CREs)
- ENCODE[5]
- 9 tissue types
- Compare signal BP to CRE start/end

Genes in novel loci

Historical phenotype associations
- GWAS Catalog[6]
Gene expression
- FUMA[7]

[1]: Graham (2021), [2] Zhu (2015), [3] Purcell (2007), [4] IGSR (2015); [5] ENCODE Cnsrtm. (2012-2020), [6] Buniello (2019), [7] Watanabe (2017)

1. Data

Mean value (mg/dL)						
	Trans-ancestry	EUR	AFR	HIS	EAS	SAS
HDL	49.5	54.2	53.1	47.1	51.1	42.1
LDL	123.2	129.2	120.2	134.8	109.8	122.0
TG	147.8	138.5	114.7	188.4	138.1	159.1
TC	201.1	208.0	197.1	209.3	195.3	195.9

SNP heritability (SE) (LDSC)[8]						
	Trans-ancestry	EUR	AFR	EAS	SAS	HIS
HDL	0.08 (0.01)	0.10 (0.01)	0.06 (0.01)	0.08 (0.01)	0.07 (0.02)	0.08 (0.02)
LDL	0.06 (0.01)	0.08 (0.01)	0.06 (0.02)	0.07 (0.02)	0.06 (0.01)	0.07 (0.02)
TG	0.07 (0.01)	0.09 (0.01)	0.05 (0.01)	0.15 (0.04)	0.12 (0.04)	0.12 (0.04)
TC	0.07 (0.01)	0.09 (0.01)	0.07 (0.02)	0.09 (0.02)	0.05 (0.02)	0.13 (0.03)

GWAS summary statistics

Graham etal (2021) Nat[1]
- HDL, LDL, TG, TC

5 populations
- AD-AFR (n=97k)
- EAS (n=120k)
- EUR (n=1.2M)
- HIS (n=47k)
- SAS (n=42k)
- Separate analyses

Pleiotropy testing (genome-wide)

Mendelian Randomization
- Bi-directional b/w all lipid pairs

T_{Pleio} test statistic
- A residual test
- Applied genome-wide

GWAS meta-analysis (CPASSOC)[2]

SNP-level meta-analysis
- between lipids
- within populations

$S_{Het.}$ test statistic
- No assumption about effect size directions

Identify independent signals

r^2<0.1, 500kb
P<5x10^{-8}/12 for pleiotropy
P<5x10^{-8} for CPASSOC
- Population-specific
- PLINK v1.9[3]
- 1kg (P3) ref. panels[4]

Identify novel loci

P<5x10^{-8}/12 for pleiotropy
P<5x10^{-8} for CPASSOC
- PLINK v1.9

>1Mb from any known signals
- P>5x10^{-8} in trans-population GWAS

Lead SNPs in novel loci in CRE regions?

Candidate cis-regulatory elements (CREs)
- ENCODE[5]
- 9 tissue types
- Compare signal BP to CRE start/end

Genes in novel loci

Historical phenotype associations
- GWAS Catalog[6]

Gene expression
- FUMA[7]

[1]: Graham (2021), [2] Zhu (2015), [3] Purcell (2007), [4] IGSR (2015); [5] ENCODE Cnsrtm. (2012-2020), [6] Buniello (2019), [7] Watanabe (2017)

CPASSOC

- **SNP-wise GWAS meta-analysis between correlated traits**

- H_0 for SNP j:
  - $\beta_j^{HDL} = \beta_j^{LDL} = \beta_j^{TG} = \beta_j^{TC} = 0$

- Inference with respect to the set

- Can identify signals missed in single-trait GWAS
  - In at least two scenarios
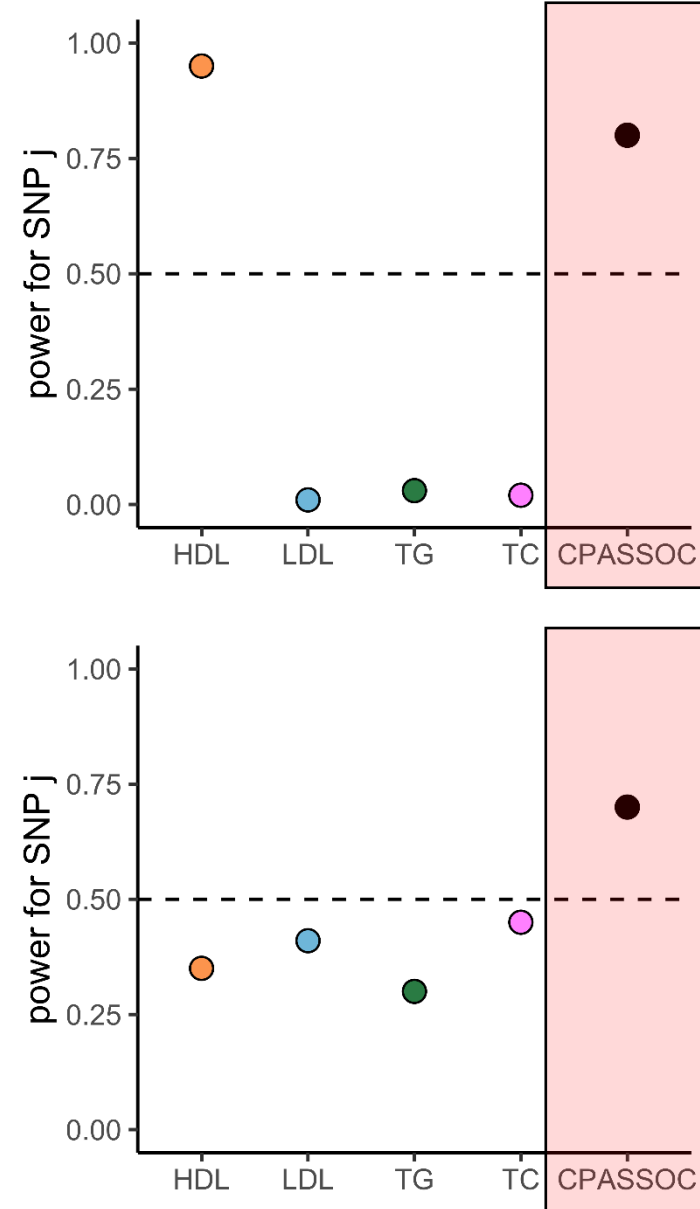
Horizontal pleiotropy testing

- Asks: **Is SNP *j* associated with Lipid 2 independent of Lipid 1?**

- H_0 for SNP *j*:
  - $\widetilde{\alpha}_j = \alpha_j - \beta_j \theta = 0$

- Requires causal effect θ
  - Mendelian Randomization
    - Single exposure
    - 12 lipid pairs x 6 populations=72 MR models
    - Minimum P(pleio.) between exposures used for inference
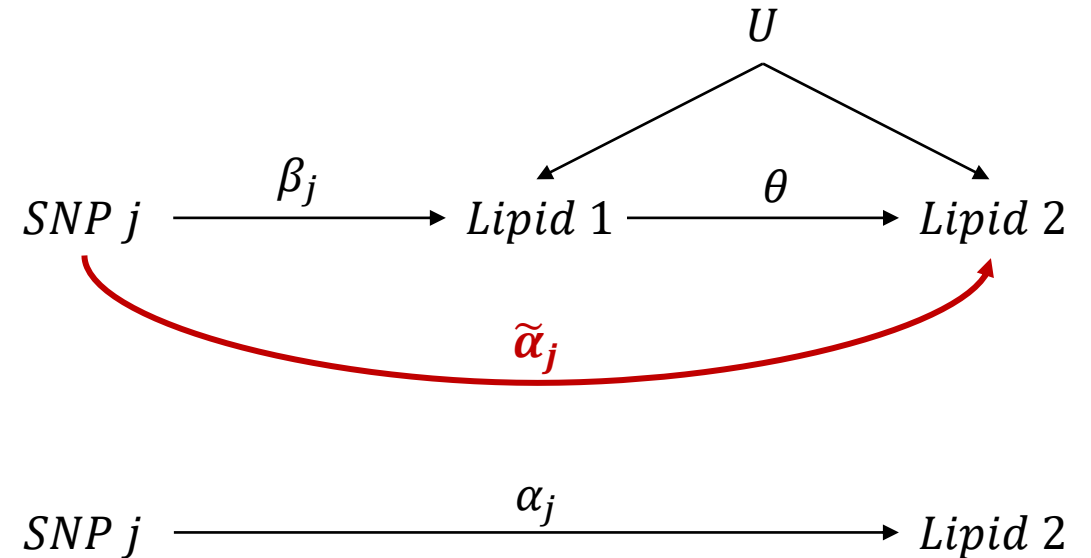
Consider $SNP_j \rightarrow LDL \rightarrow HDL$ and we want to test $\widetilde{\alpha}_j = 0$

$\widetilde{\alpha}_j$

HDL

LDL

Horizontal pleiotropy
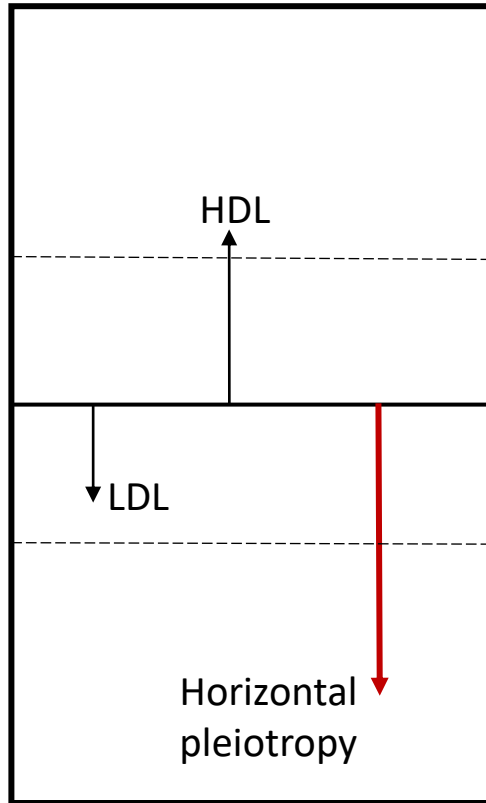
Pleiotropy test uses power from exposure phenotype

HDL

LDL

Horizontal pleiotropy
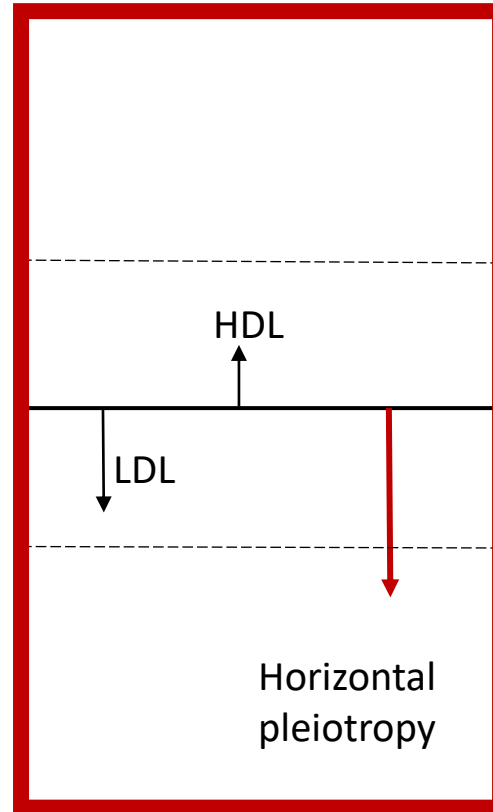
Pleiotropy test magnifies small effects

LDL

HDL

Horizontal pleiotropy
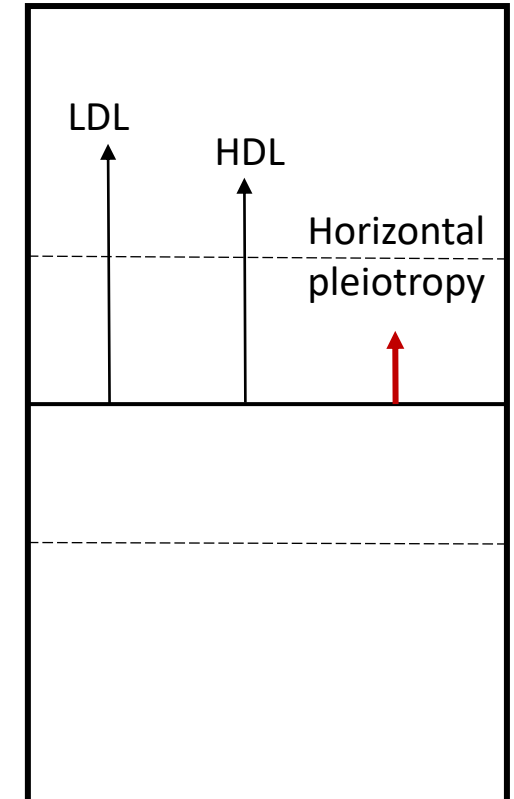
Pleiotropy test only identifies *non-mediated* associations

LDL→ HDL Horizontal Pleiotropy Test

Original HDL GWAS
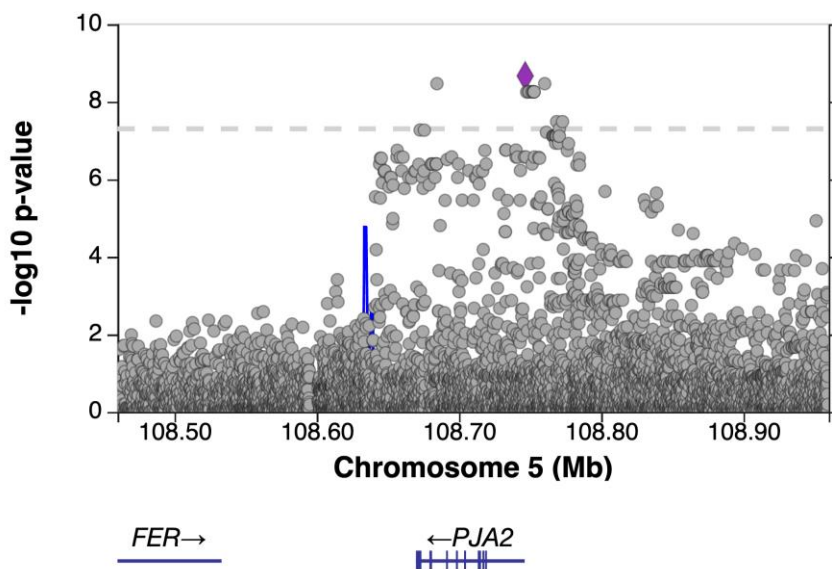
Original LDL GWAS

Locus detected in pleiotropy testing

Smallest P=2.2x10⁻⁹

Lead pleiotropy test SNP: **rs2963006**
$T_{horizontal\ Pleiotropy}$ = 5.98
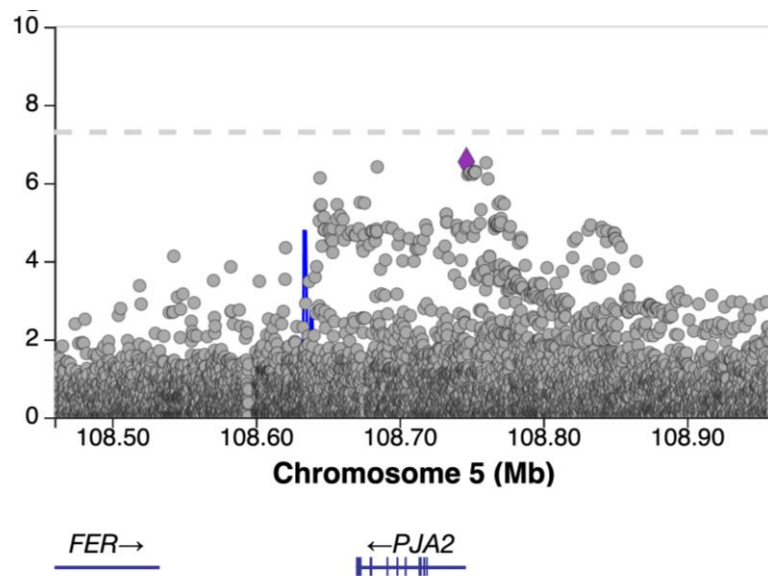
Locus not detected in HDL GWAS

Smallest P=2.9x10⁻⁷

rs58542926 effect (SE): 0.008 (0.0013)
$T_{HDL\ GWAS}$ = 5.00
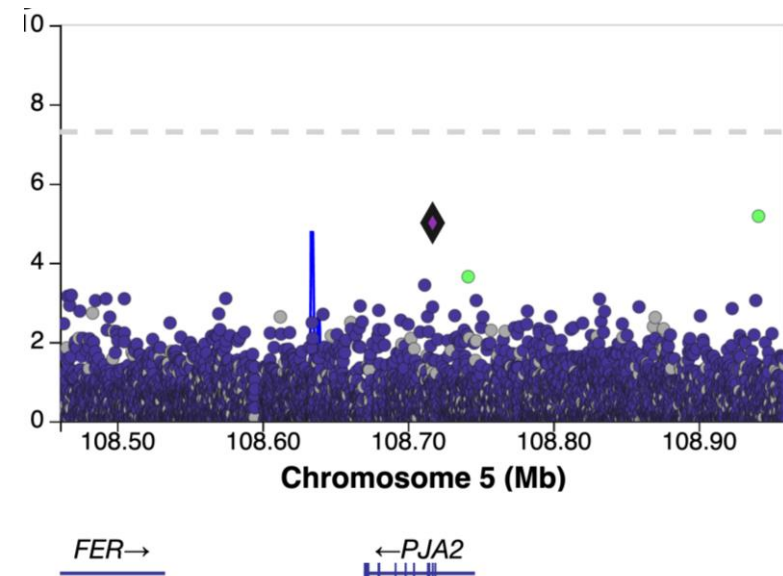
Locus highly sig. in TG GWAS

Smallest P= 6.7x10⁻⁷

rs58542926 effect(SE): -0.001 (0.001)
$T_{LDL\ GWAS}$ = -4.8

Exposure and outcome have opposite
sign at lead horizontal pleiotropy SNP

GWAS summary statistics

Graham etal (2021) Nat[1]
- HDL, LDL, TG, TC

5 populations
- AD-AFR (n=97k)
- EAS (n=120k)
- EUR (n=1.2M)
- HIS (n=47k)
- SAS (n=42k)
- Separate analyses

Pleiotropy testing (genome-wide)

Mendelian Randomization
- Bi-directional b/w all lipid pairs

T_{Pleio} test statistic
- A residual test
- Applied genome-wide

GWAS meta-analysis (CPASSOC)[2]

SNP-level meta-analysis
- between lipids
- within populations

$S_{Het.}$ test statistic
- No assumption about effect size directions

Identify independent signals

r^2<0.1, 500kb
P<5x10^{-8}/12 for pleiotropy
P<5x10^{-8} for CPASSOC
- Population-specific
- PLINK v1.9[3]
- 1kg (P3) ref. panels[4]

Identify novel loci

P<5x10^{-8}/12 for pleiotropy
P<5x10^{-8} for CPASSOC
- PLINK v1.9

>1Mb from any known signals
- P>5x10^{-8} in trans-population GWAS

Lead SNPs in novel loci in CRE regions?

Candidate cis-regulatory elements (CREs)
- ENCODE[5]
- 9 tissue types
- Compare signal BP to CRE start/end

Genes in novel loci

Historical phenotype associations
- GWAS Catalog[6]

Gene expression
- FUMA[7]

[1] Graham (2021), [2] Zhu (2015), [3] Purcell (2007), [4] IGSR (2015); [5] ENCODE Cnsrtm. (2012-2020), [6] Buniello (2019), [7] Watanabe (2017)

GWAS vs Pleiotropy test signals
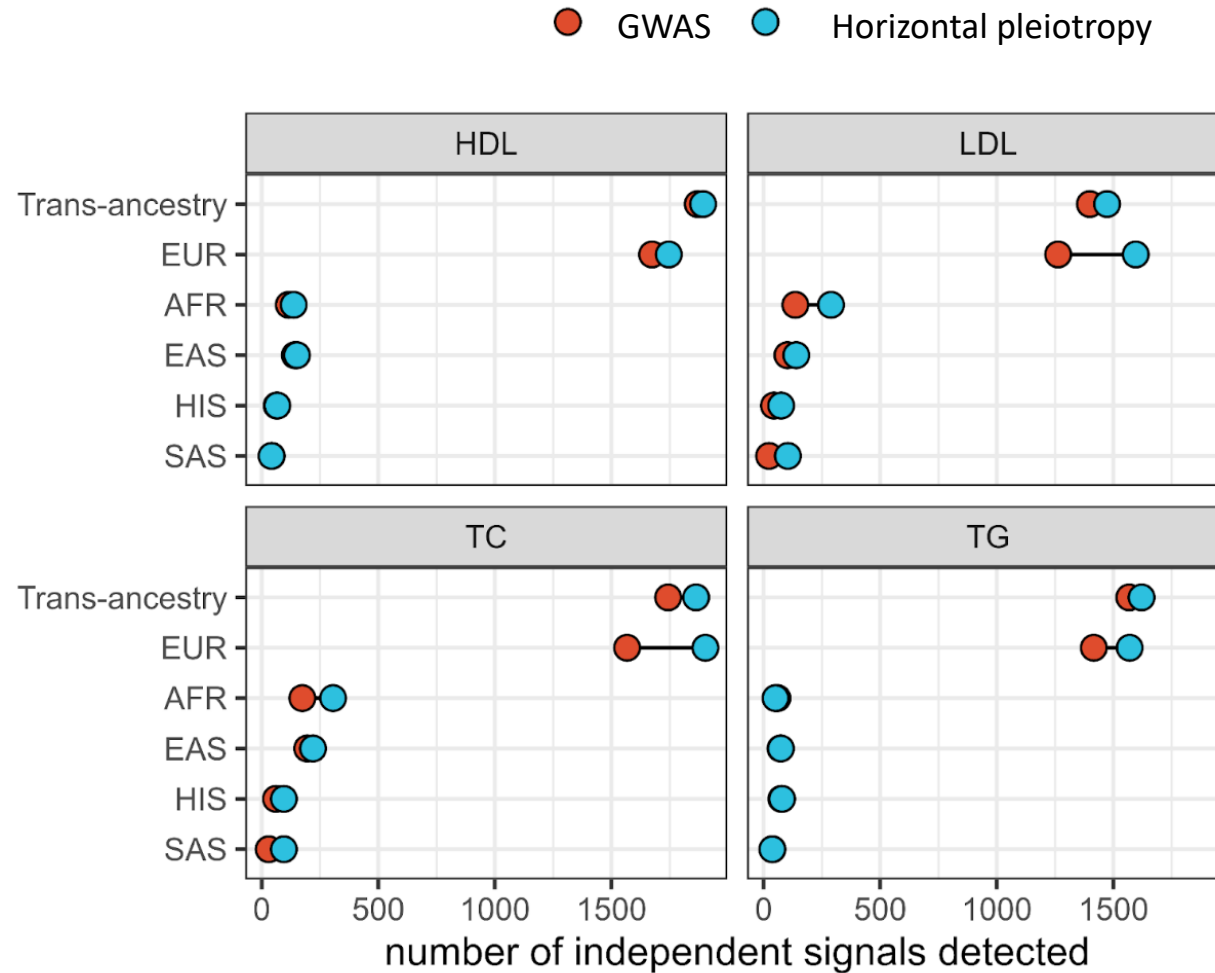
Pleiotropy testing identifies novel signals and uses evidence from multiple exposures

# Unique Novel Loci		
	CPASSOC	H. Pleio. Test
AFR	34 (14%)	15 (20%)
EAS	4 (2%)	6 (7.5%)
EUR	102 (41%)	21 (28%)
HIS	10 (4%)	2 (3%)
SAS	25 (10%)	6 (7.5%)
Trans.	48 (19%)	21 (28%)
>1 Pop. Group	24 (10%)	5 (6%)
Sum	247	76

+ 72 identified by both = **395 novel loci**

Lipids for all novel loci detected using pleiotropy testing

GWAS summary statistics

Graham etal (2021) Nat[1]
- HDL, LDL, TG, TC

5 populations
- AD-AFR (n=97k)
- EAS (n=120k)
- EUR (n=1.2M)
- HIS (n=47k)
- SAS (n=42k)
- Separate analyses

Pleiotropy testing (genome-wide)

Mendelian Randomization
- Bi-directional b/w all lipid pairs

T_{Pleio} test statistic
- A residual test
- Applied genome-wide

GWAS meta-analysis (CPASSOC)[2]

SNP-level meta-analysis
- between lipids
- within populations

$S_{Het.}$ test statistic
- No assumption about effect size directions

Identify independent signals

$r^2<0.1$, 500kb
$P<5\times10^{-8}/12$ for pleiotropy
$P<5\times10^{-8}$ for CPASSOC
- Population-specific
- PLINK v1.9[3]
- 1kg (P3) ref. panels[4]

Identify novel loci

$P<5\times10^{-8}/12$ for pleiotropy
$P<5\times10^{-8}$ for CPASSOC
- PLINK v1.9

>1Mb from any known signals
- $P>5\times10^{-8}$ in trans-population GWAS

Lead SNPs in novel loci in CRE regions?

Candidate cis-regulatory elements (CREs)
- ENCODE[5]
- 9 tissue types
- Compare signal BP to CRE start/end

Genes in novel loci

Historical phenotype associations
- GWAS Catalog[6]

Gene expression
- FUMA[7]

[1]: Graham (2021), [2] Zhu (2015), [3] Purcell (2007), [4] IGSR (2015); [5] ENCODE Cnsrtm. (2012-2020), [6] Buniello (2019), [7] Watanabe (2017)

Example genes

- *FER*
  - Blood pressure [EUR; 30578418]

- *MIR3662, MYBI*
  - **Hemoglobin** [EUR; PMID 26366553]

- *ABCG8*
  - **Gallstones** [AFR, EUR, HIS; 27094239]

- *SLC27A5*
  - **Insomnia** [EUR; 35835914]

Loci detected using the horizontal pleiotropy test are enriched in CRE's, but not as much as original GWAS signals
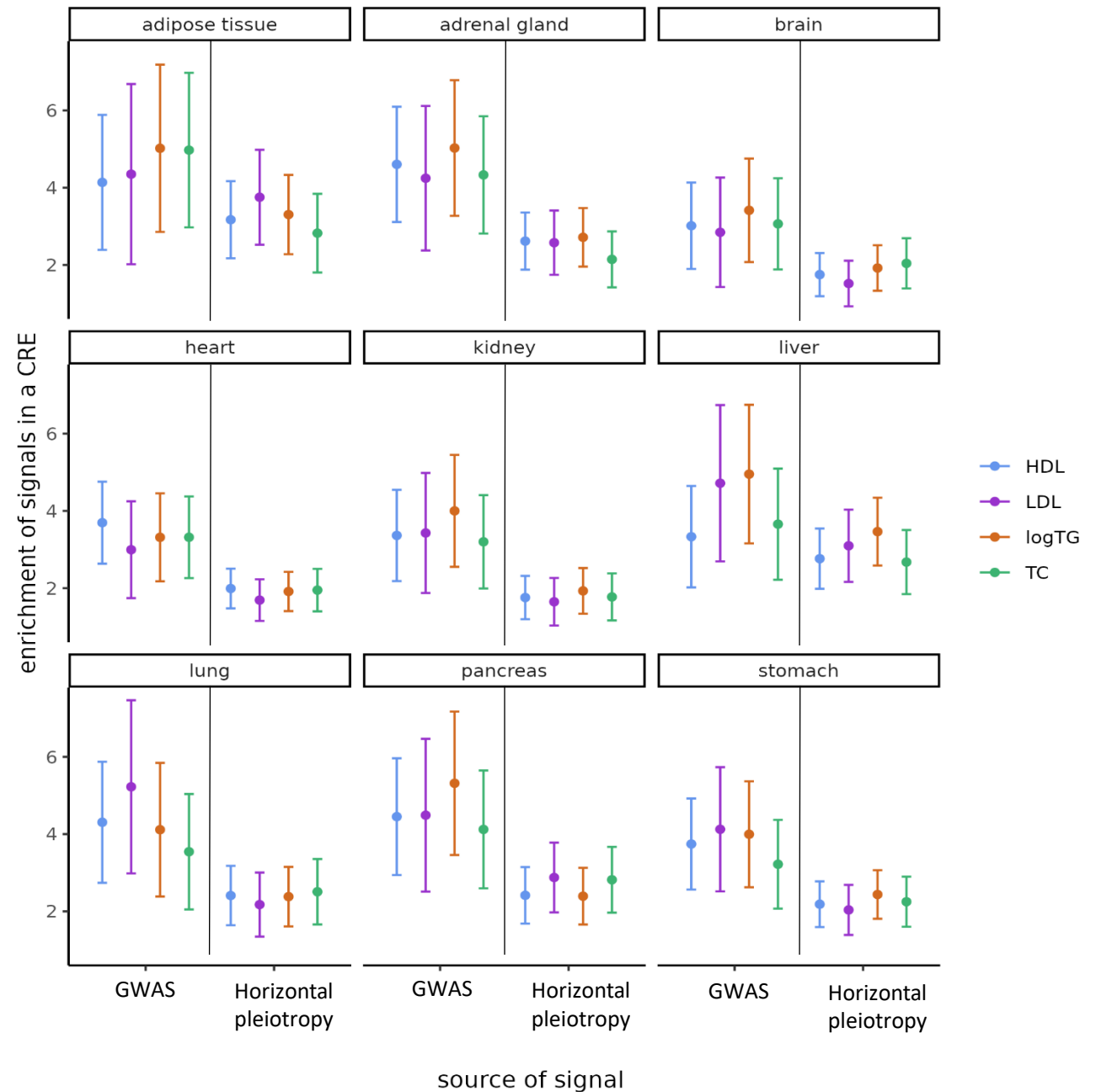
Conclusions

- CPASSOC more powerful method
- Horizontal pleiotropy test can identify novel loci missed by GWAS
  - May provide additional inference

Future Directions

- Apply pleiotropy test to complex disease
- Understand *why* some loci are significant in horizontal pleiotropy testing

Thank you

Mentor Dr. Xiaofeng Zhu

Colleagues Gen Li & Dr. Yihe Yang
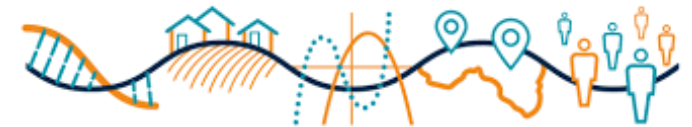
References

[1] Graham, S. E., Clarke, S. L., Wu, K. H. H., Kanoni, S., Zajac, G. J., Ramdas, S., ... & Sim, X. (2021). The power of genetic diversity in genome-wide association studies of lipids. *Nature*, *600*(7890), 675-679.

[2] Zhu, X., Feng, T., Tayo, B. O., Liang, J., Young, J. H., Franceschini, N., ... & COGENT BP Consortium. (2015). Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension. The American Journal of Human Genetics, 96(1), 21-36.

[3] Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ & Sham PC (2007) PLINK: a toolset for whole-genome association and population-based linkage analysis. American Journal of Human Genetics, 81.

[4] A global reference for human genetic variation, The 1000 Genomes Project Consortium, Nature 526, 68-74 (01 October 2015)

[5] Luo, Y., Hitz, B. C., Gabdank, I., Hilton, J. A., Kagda, M. S., Lam, B., ... & Cherry, J. M. (2020). New developments on the Encyclopedia of DNA Elements (ENCODE) data portal. *Nucleic acids research*, *48*(D1), D882-D889.

[6] *Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, McMahon A, Morales J, Mountjoy E, Sollis E, Suveges D, Vrousgou O, Whetzel PL, Amode R, Guillen JA, Riat HS, Trevanion SJ, Hall P, Junkins H, Flicek P, Burdett T, Hindorff LA, Cunningham F and Parkinson H.*
The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019.
Nucleic Acids Research, 2019, Vol. 47 (Database issue): D1005-D1012.

[7] K. Watanabe, E. Taskesen, A. van Bochoven and D. Posthuma. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**:1826. (2017)

[8] Bulik-Sullivan, B. K., Loh, P. R., Finucane, H. K., Ripke, S., Yang, J., Patterson, N., ... & Neale, B. M. (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nature genetics*, *47*(3), 291-295.
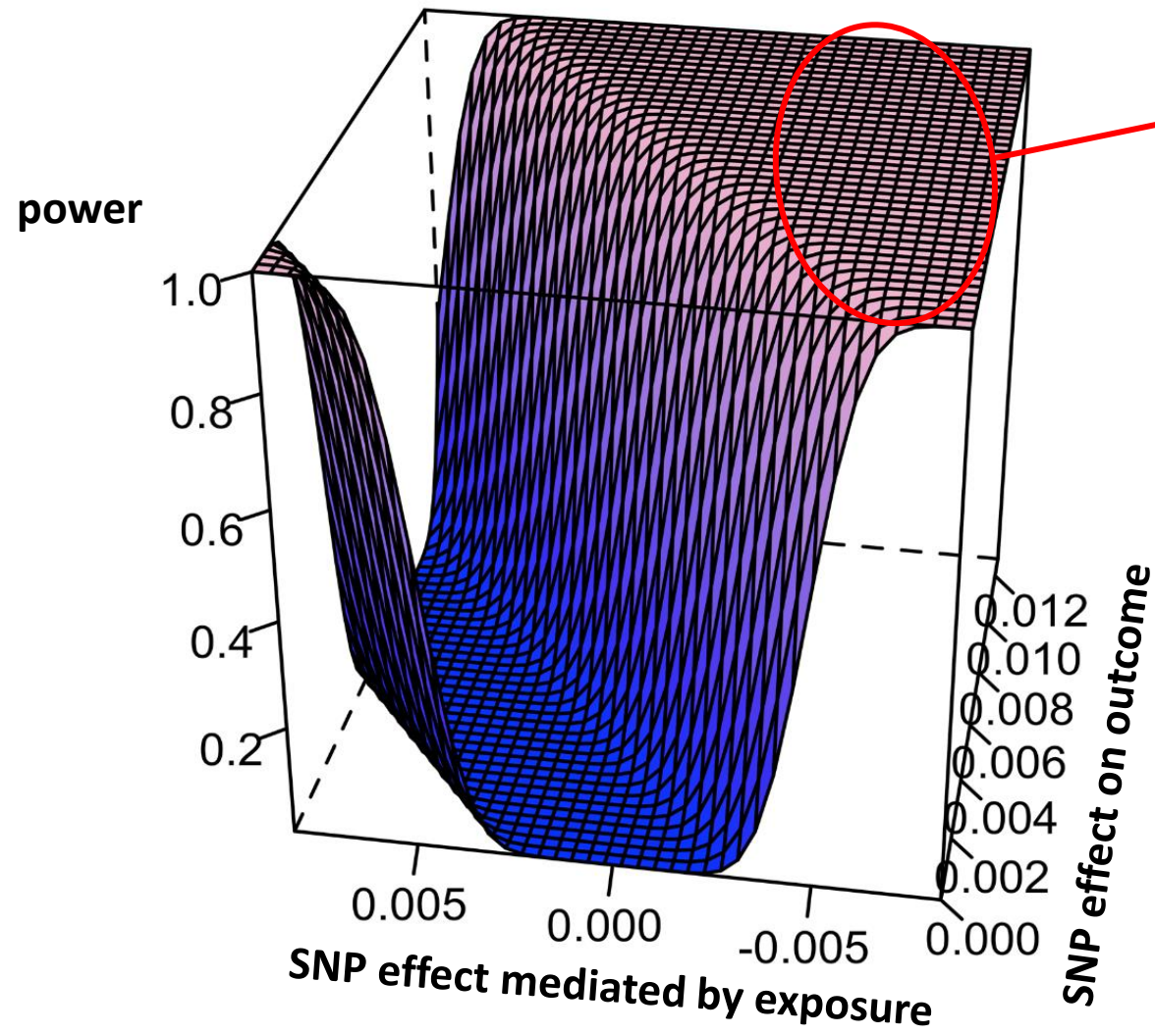
- Large effect on outcome
- **Large effect on exposure in other direction**

19p13.11 locus

TG→HDL causal effect

SNP associations with HDL ($\widehat{\boldsymbol{\alpha}}_j$)
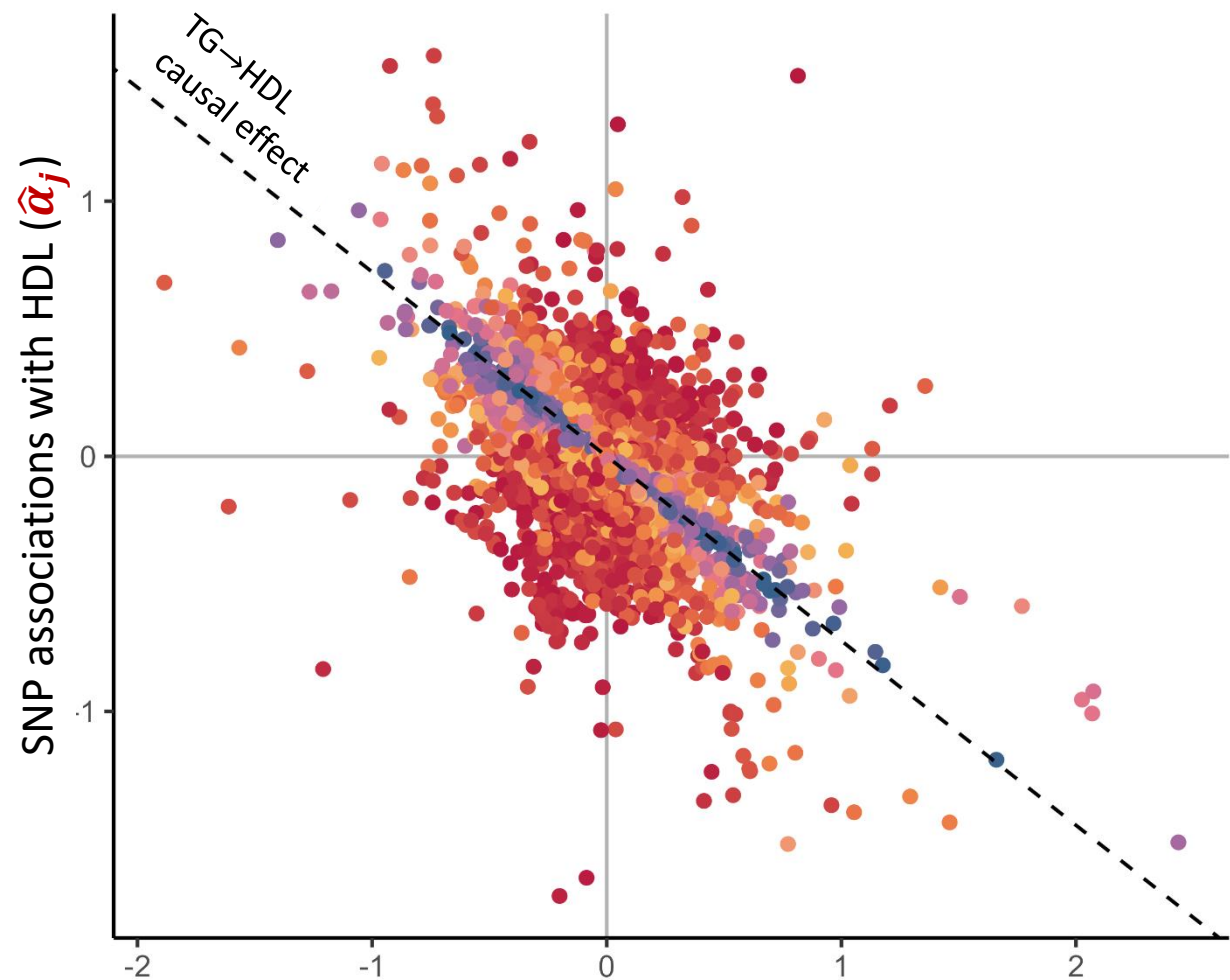
SNP associations with triglycerides ($\widehat{\boldsymbol{\beta}}_j$)

Horizontal pleiotropy test P-value

$<1\times10^{-100}$ 1.00

U

$SNP\ j$ $\boldsymbol{\beta_j}$ → Triglycerides θ → HDL

$\tilde{\alpha}$

$SNP\ j$ $\boldsymbol{\alpha_j}$ → HDL

19p13.11 locus

diagonal y=x

best fit
r=0.45

Associations with HDL through TG ($\widehat{\beta_j}\widehat{\theta}_{TG \to HDL}$)

Total associations with HDL ($\widehat{\alpha_j}$)

Pleiotropy test P-value

$<1\times10^{-100}$ 1.00

U

$SNP\ j$ $\xrightarrow{\beta_j}$ Triglycerides $\xrightarrow{\theta}$ HDL

$\tilde{\alpha}$

$SNP\ j$ $\xrightarrow{\alpha_j}$ HDL
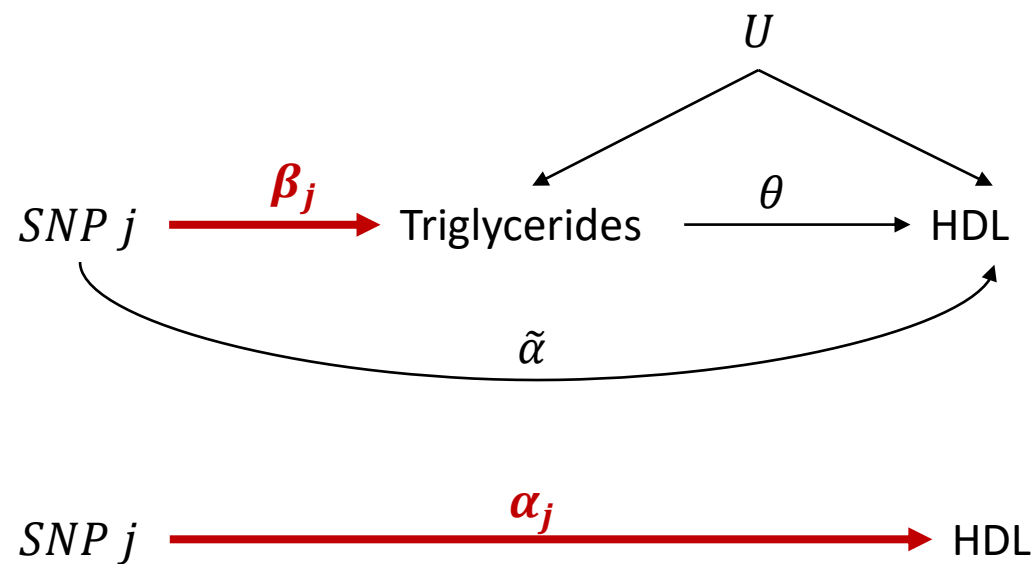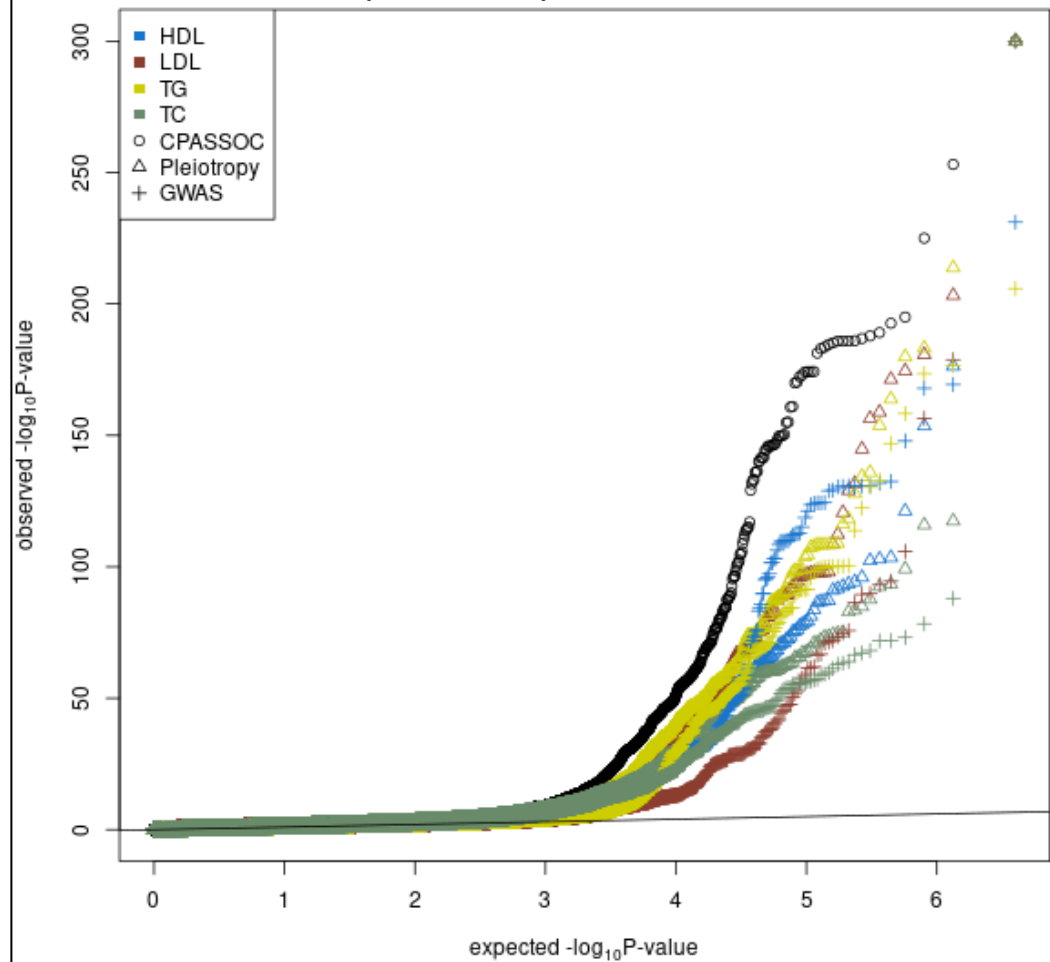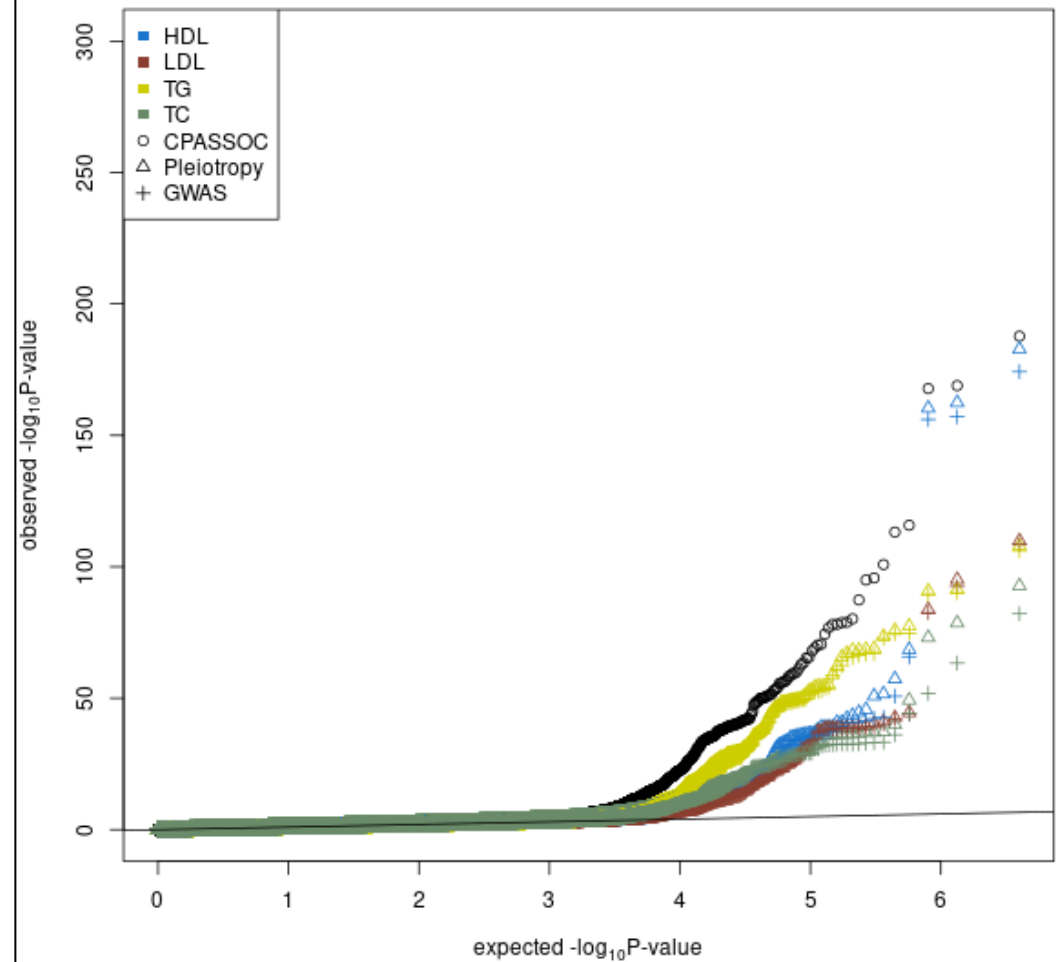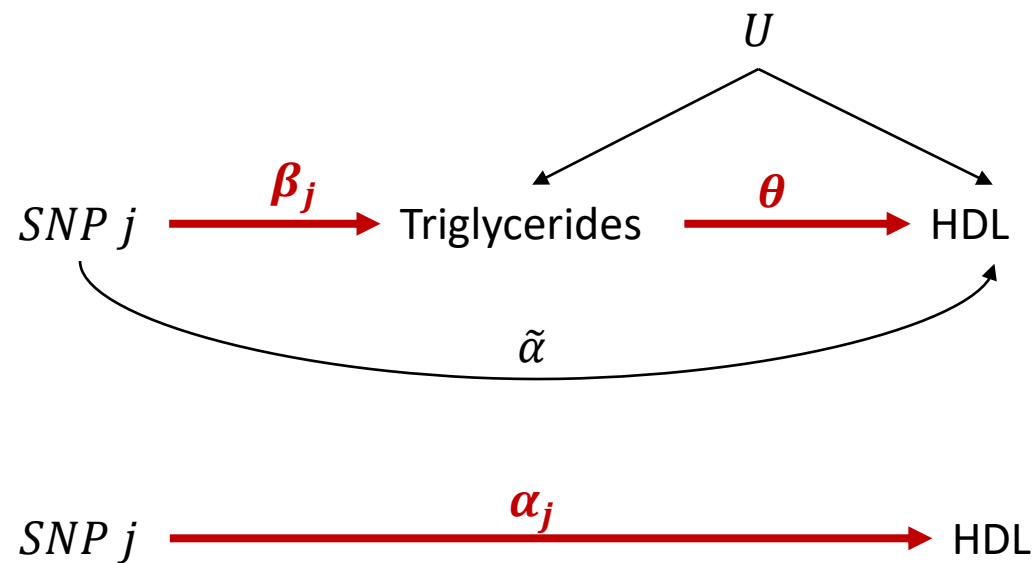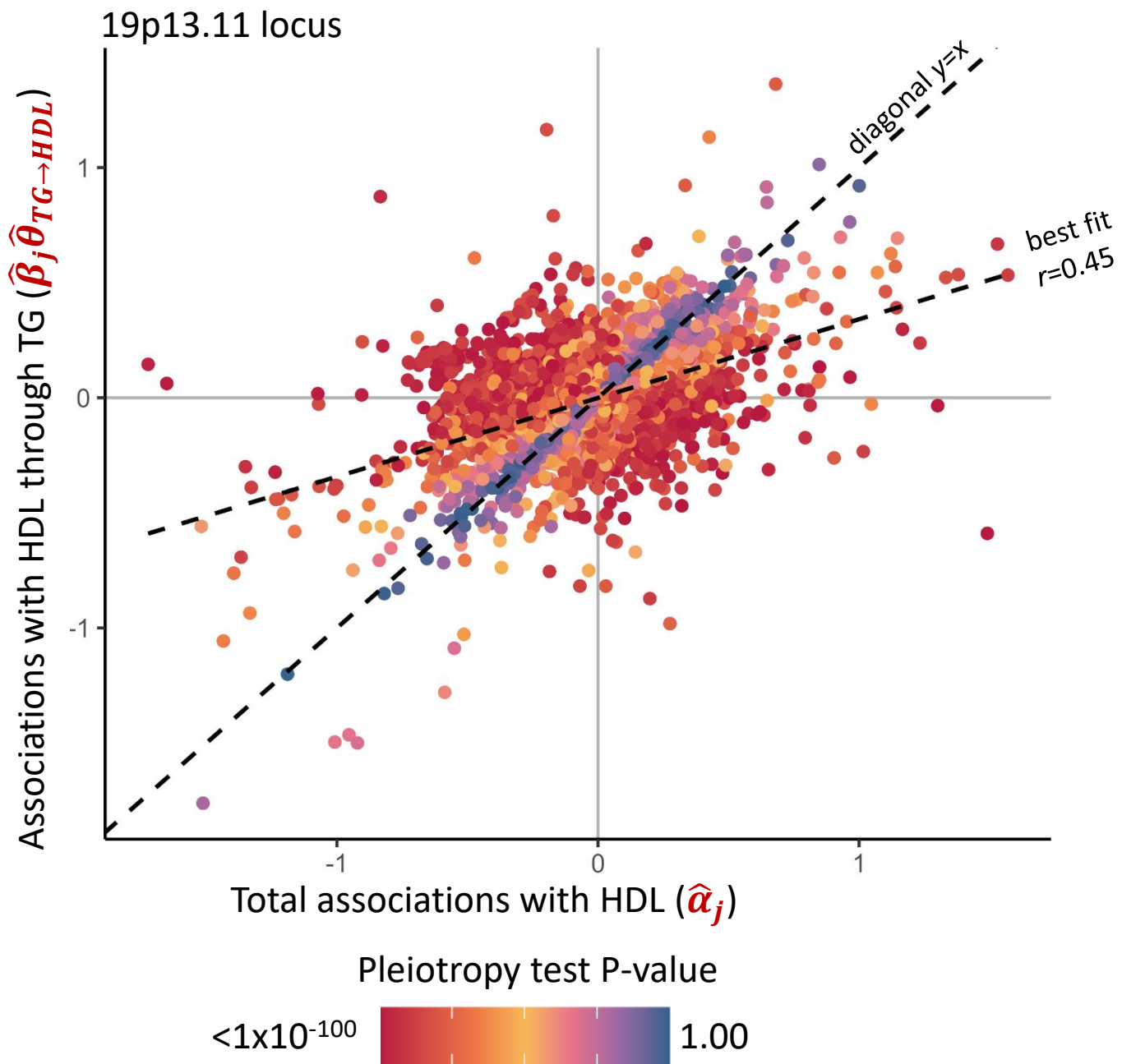
Example of pleiotropy test results in EUR for HDL
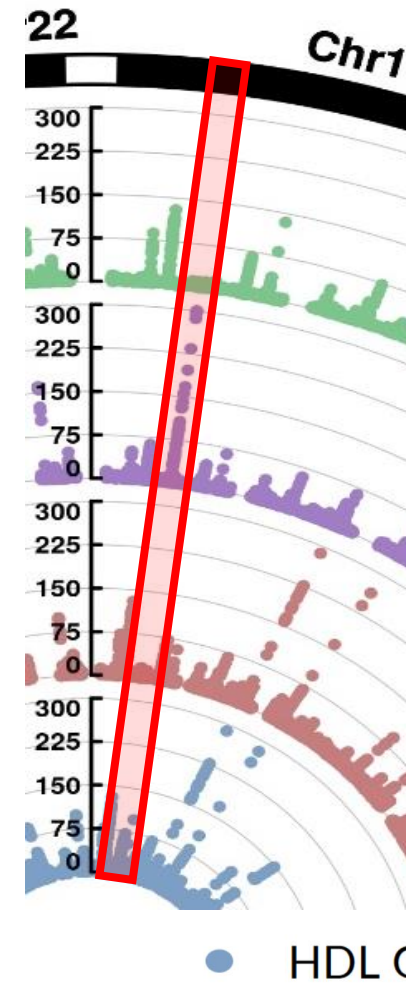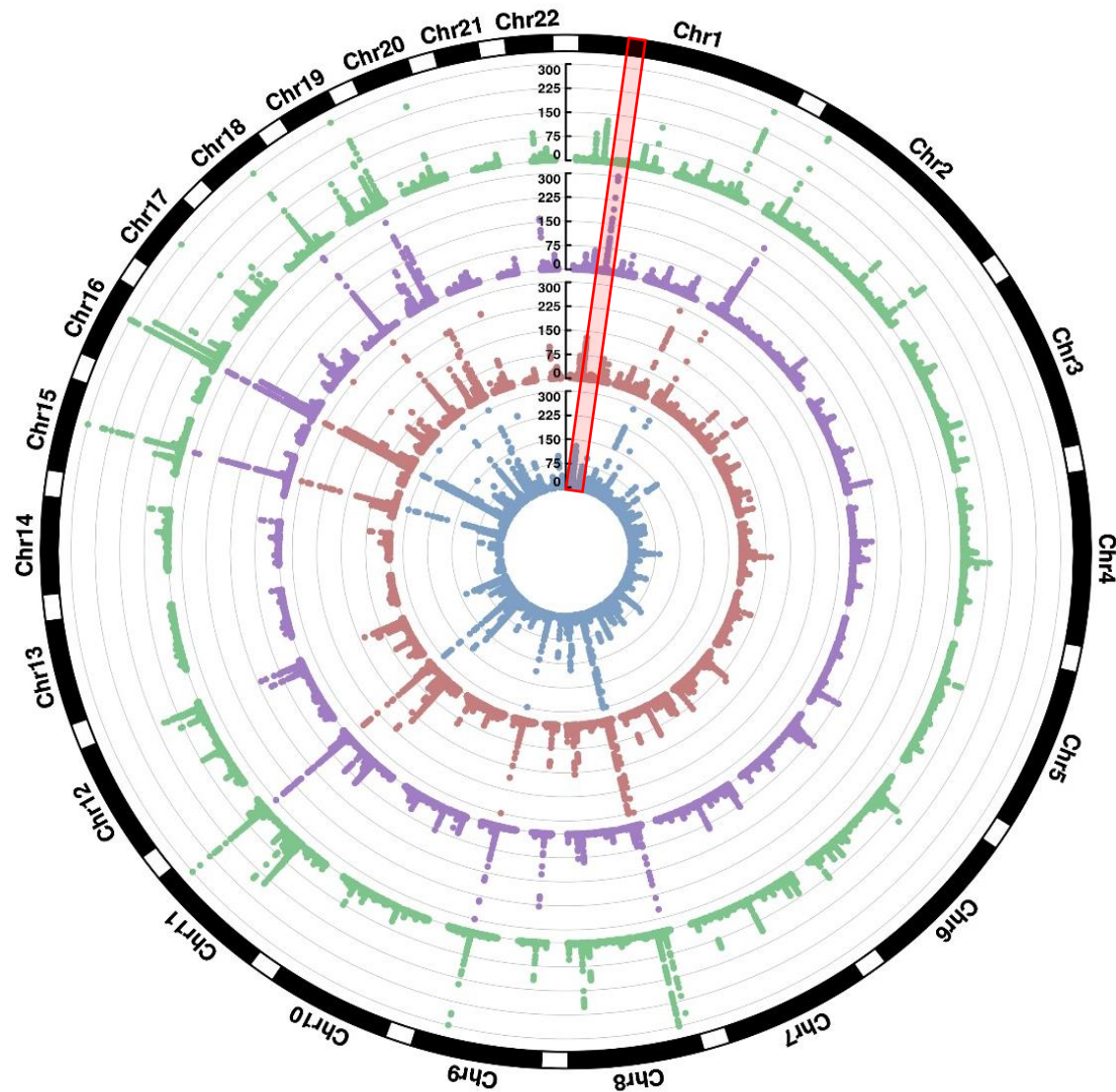
TC → HDL Pleiotropy test

logTG → HDL Pleiotropy test

LDL → HDL Pleiotropy test

HDL GWAS

DOCK7 locus

TG→HDL causal effect

SNP associations with HDL

SNP associations with triglycerides

Pleiotropy test P-value

$<1 \times 10^{-100}$ 1.00

U

$DOCK7 \xrightarrow{\text{min. GWAS } P < 10^{-500}} \text{Triglycerides} \xrightarrow{-0.74} \text{HDL}$

min. pleiotropy $P < 10^{-300}$

$DOCK7 \xrightarrow{\text{min. GWAS } P = 1.5 \times 10^{-16}} \text{HDL}$
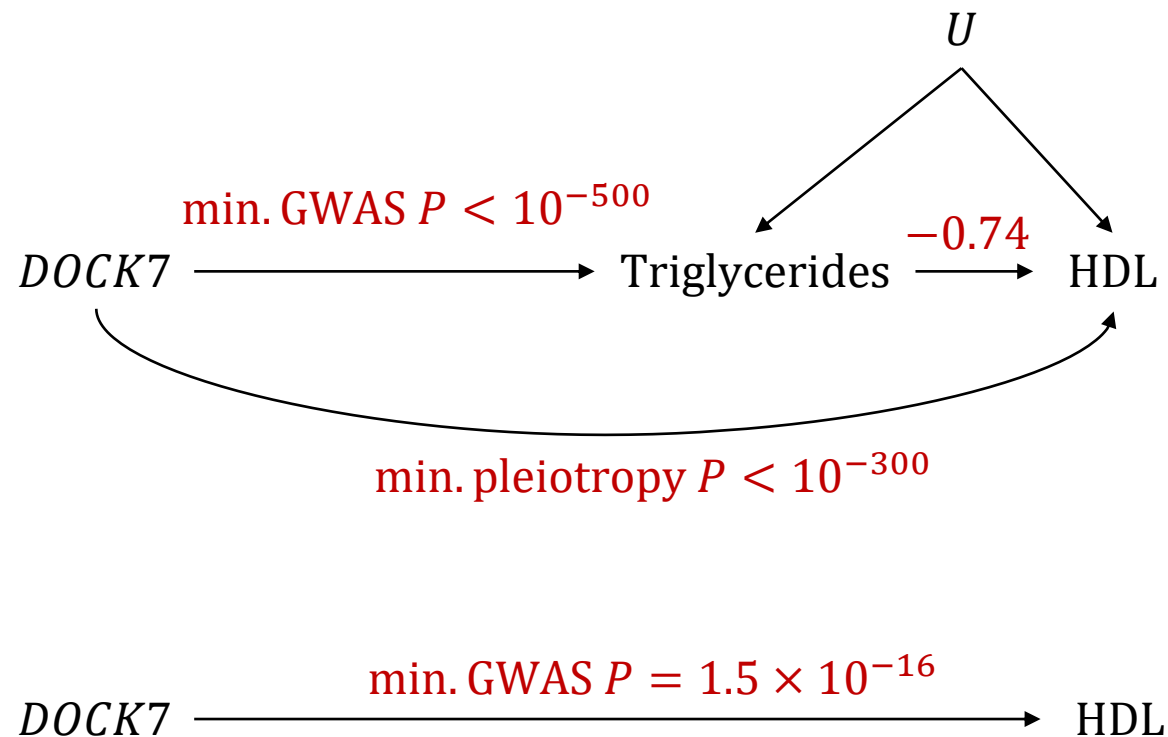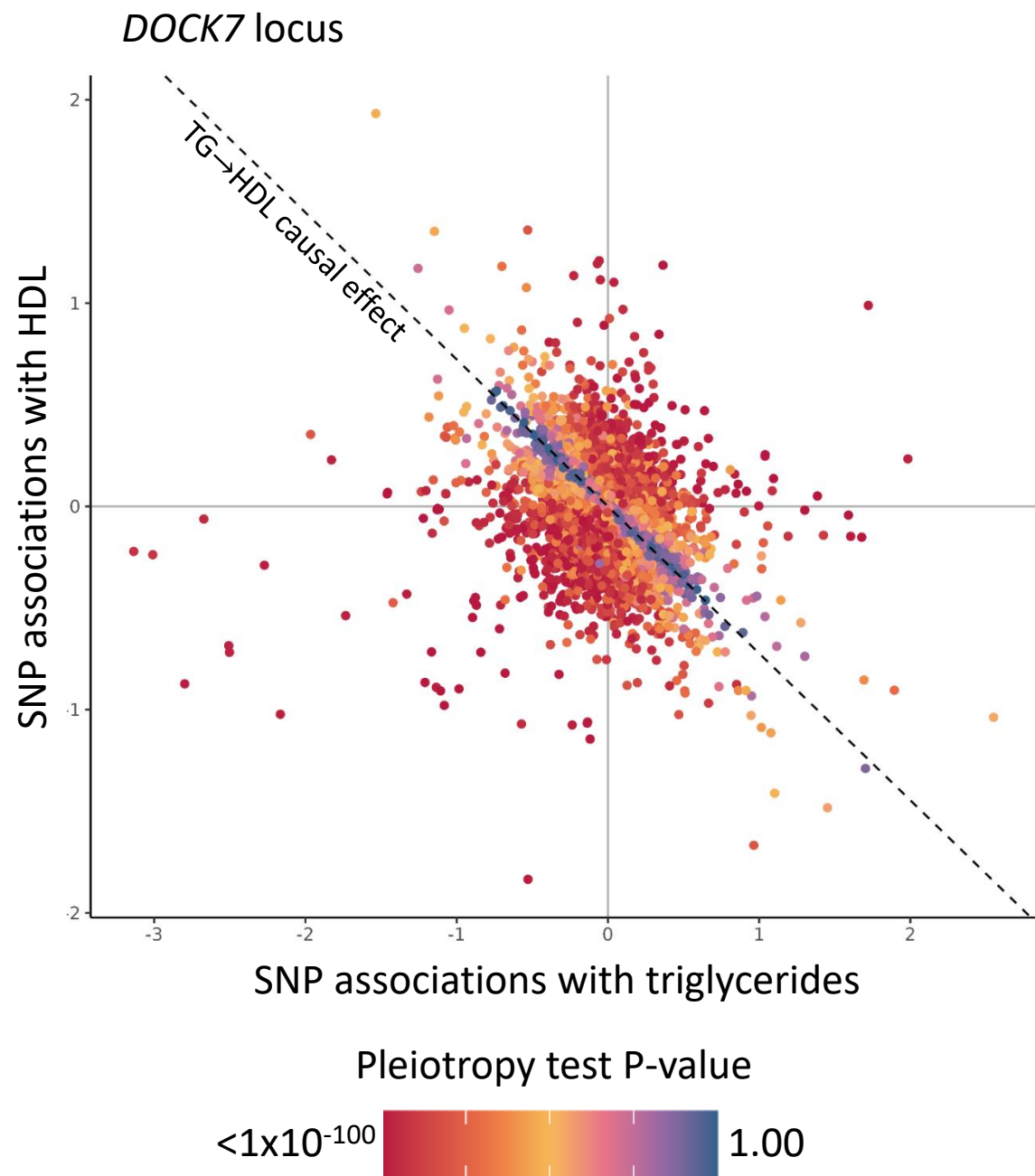
Interpretation: *DOCK7* locus is pleiotropic for HDL and triglycerides

Triglycerides GWAS
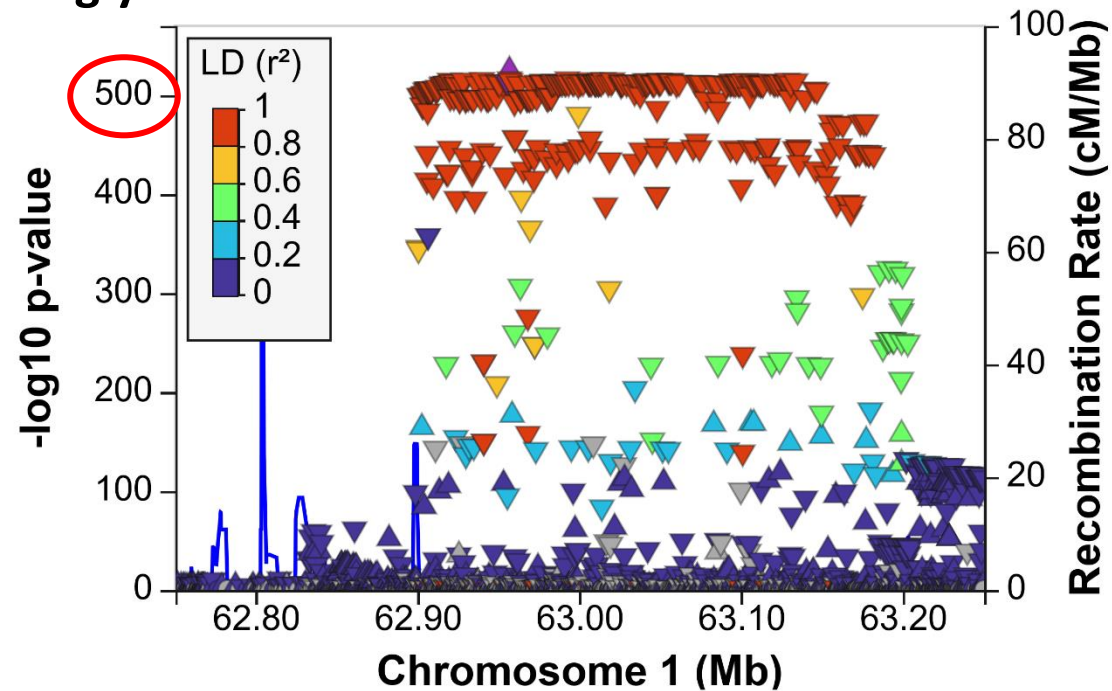
HDL GWAS

GWAS Catalog hits for tg gwas
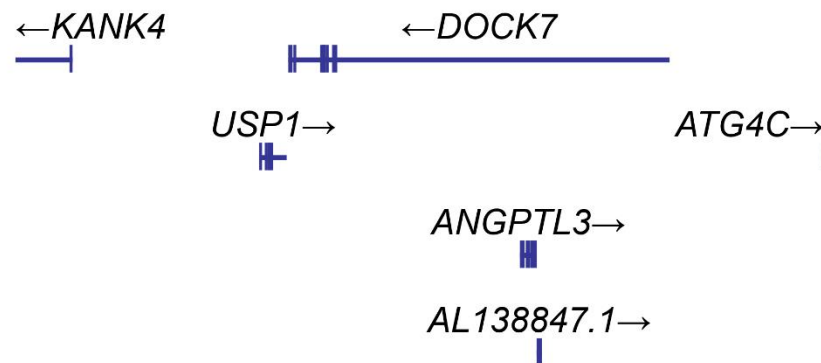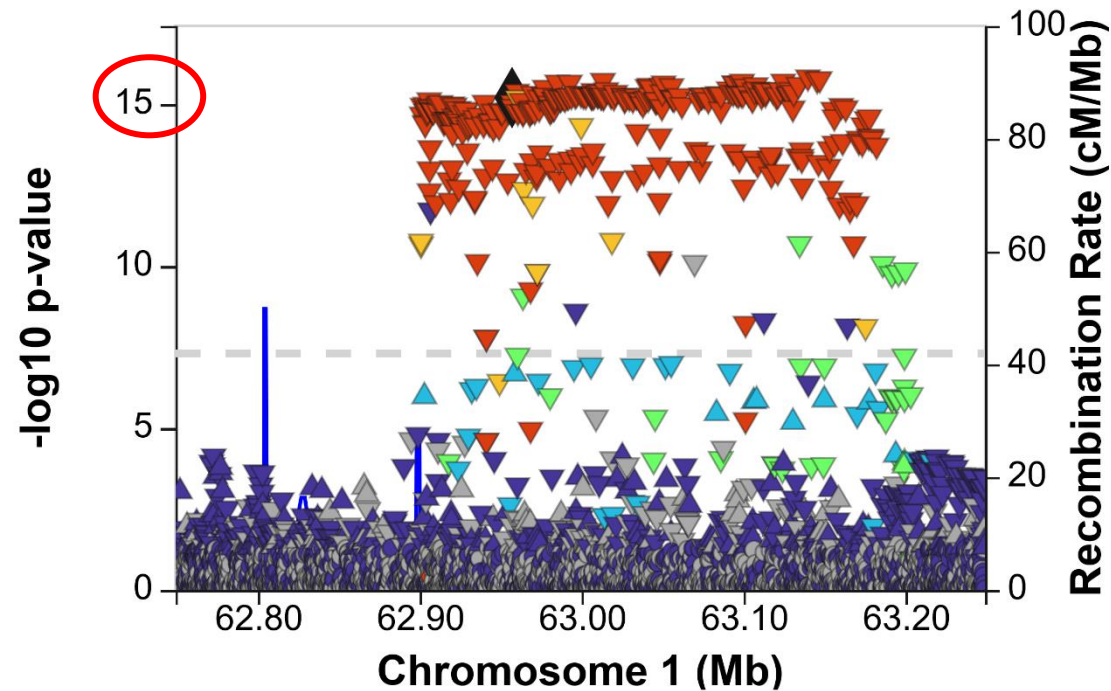
GWAS Catalog hits for hdl gwas
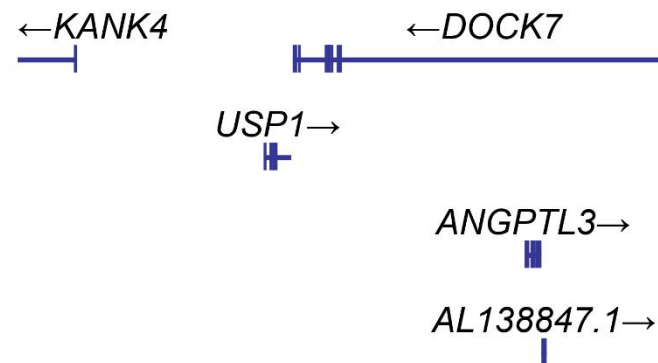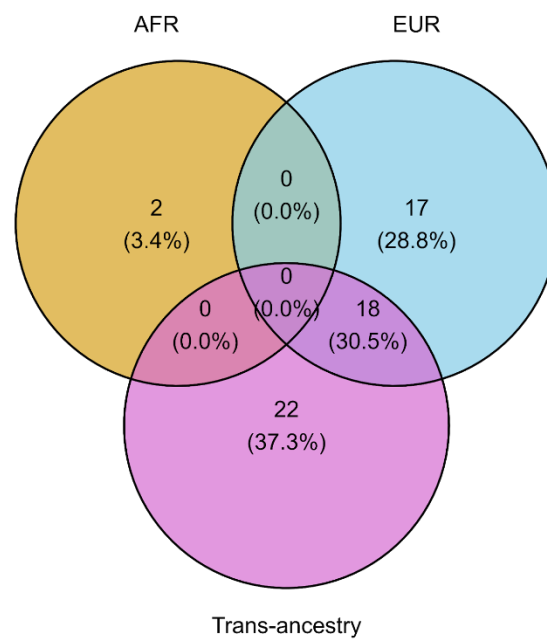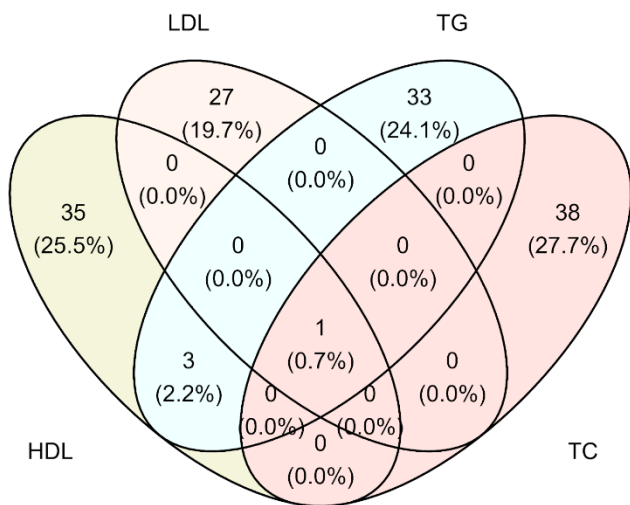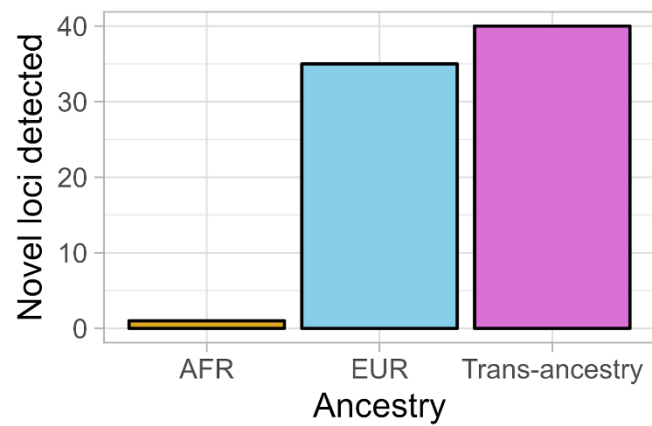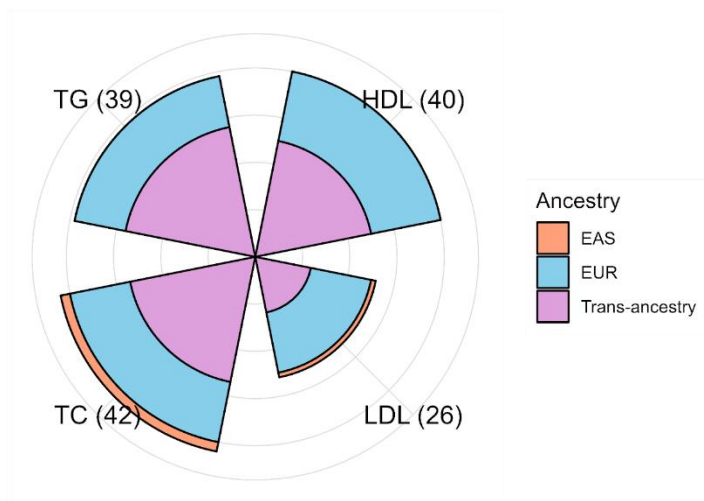
(Example)

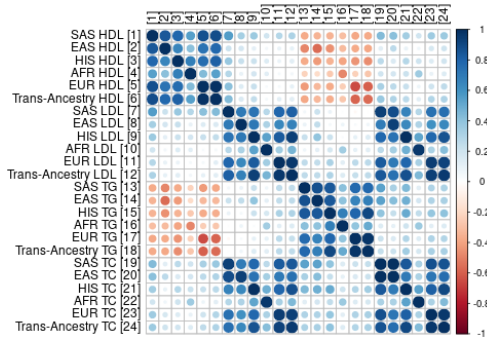Let X_k have causal effect θ on X_s

It can be shown that

$$\theta = \text{GenCorr}(X_k, X_s)\sqrt{h_s^2/}$$

Lipids off-diagonal may b
pleiotropic

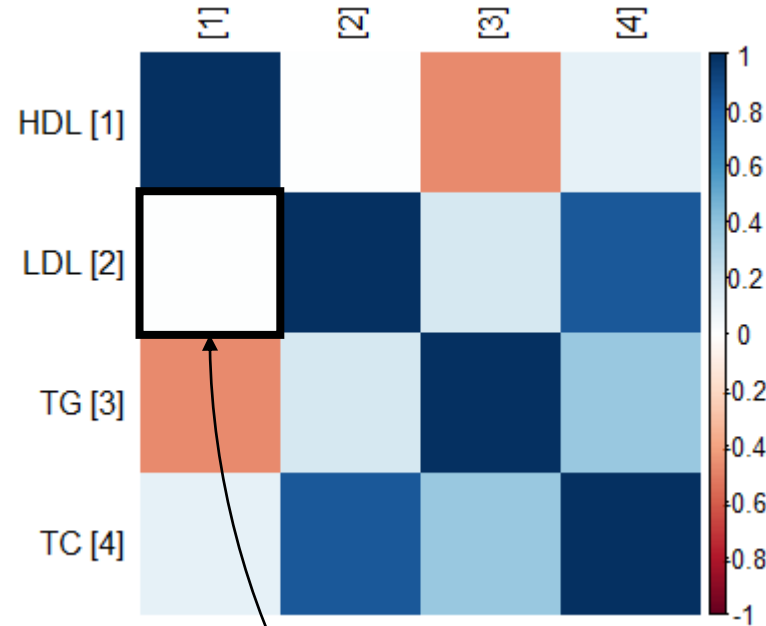Genetic correlations vs Causal effects b/w all lipid pairs
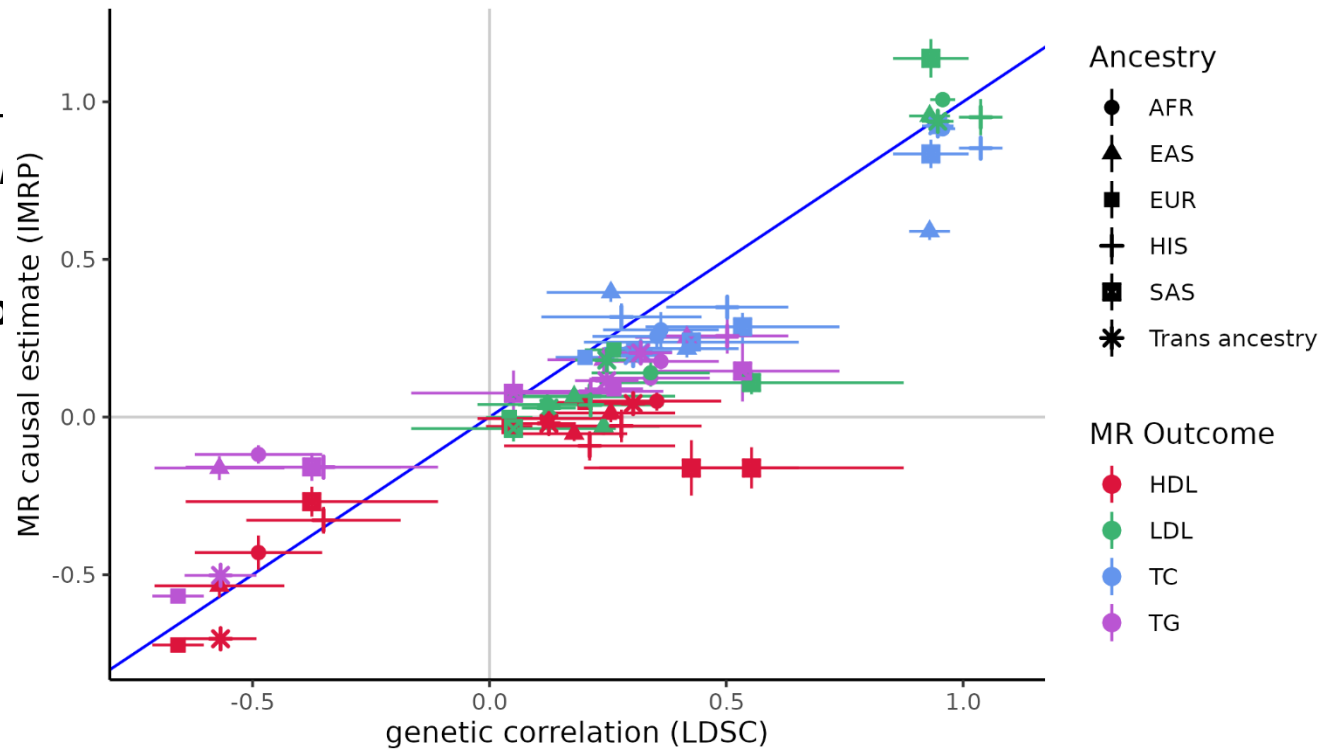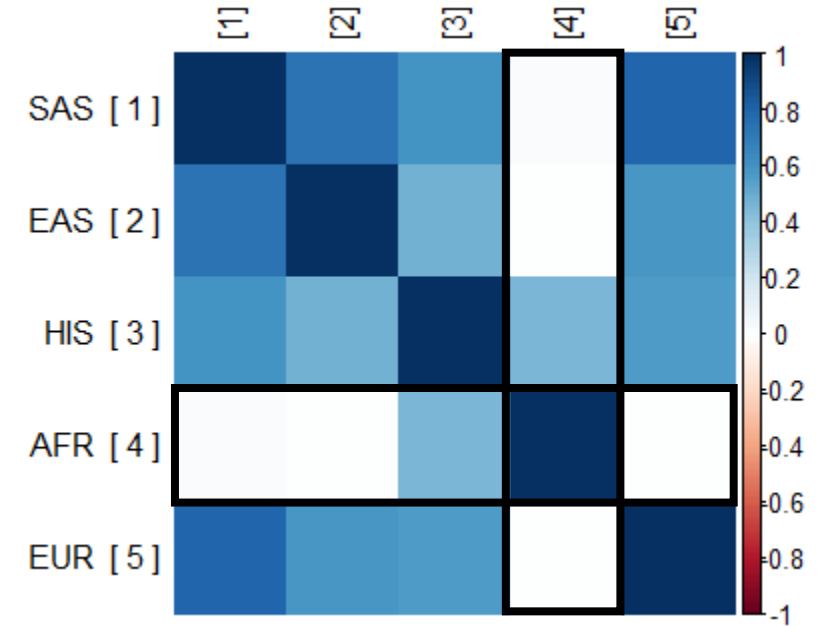
Decomposing genetic correlation into two parts

Between-lipid genetic correlations

Between-population genetic correlations

Suggests HDL and LDL have no causal relationship