‘Are lower HDL-c levels among the South Indian diabetic cohort compared with the Scottish diabetic population genetically driven?? A follow-up work’

Supervisors
UK
Prof Colin NA Palmer
Prof Ewan R Pearson
India
Dr Guha R Pradeepa
Dr Radha Venkatesan

24/03/2021

PhD Student Symposium 2021

Dr Mehul Kumar Chourasia
PhD Medicine (3rd Year)

Funding Acknowledgement

"This research was funded by the National Institute for Health Research (NIHR) (INSPIRED 16/136/102) using UK aid from the UK Government to support global health research. The views expressed in this publication are those of the author(s) and not necessarily those of the NIHR or the UK Department of Health and Social Care"

Background

South Indian

HDL-c description (n =4,315); Male = 2615 (60.6%)					
Variable	Mean	Std Dev	Median	Minimum	Maximum
HDL (Baseline)	1.08	0.25	1.06	0.13	2.69
HDL (After)	1.21	0.29	1.16	0.16	3.88

Difference between Before and After HDL value (Paired T test)						
N	Mean	Std Dev	Min	Max	t Value	P value
4,315	-0.14	0.20	-3.10	1.40	-43.83	<.0001

Scottish

HDL-c description (n= 10,633), Male = 5958(56.03%)					
Variable	Mean	Std Dev	Median	Minimum	Maximum
Hdl (Baseline)	1.33	0.409	1.26	0.31	3.93
Hdl (After)	1.60	0.48	1.51	0.58	8.11

Difference between Before and After HDL value (Paired T test)						
N	Mean	Std Dev	Min	Max	t Value	P value
10,633	-0.27	0.32	-5.58	1.46	-87.97	<.0001

Preliminary data suggested that HDL-c profile between the two populations [Scottish (1.33±0.41) and South Indian (1.08±0.25)] were significantly lower by 20 % (p value <0.001)

'Are lower HDL-c levels among the South Indian diabetic cohort compared with the Scottish diabetic population genetically driven??

*Unit mmol/l

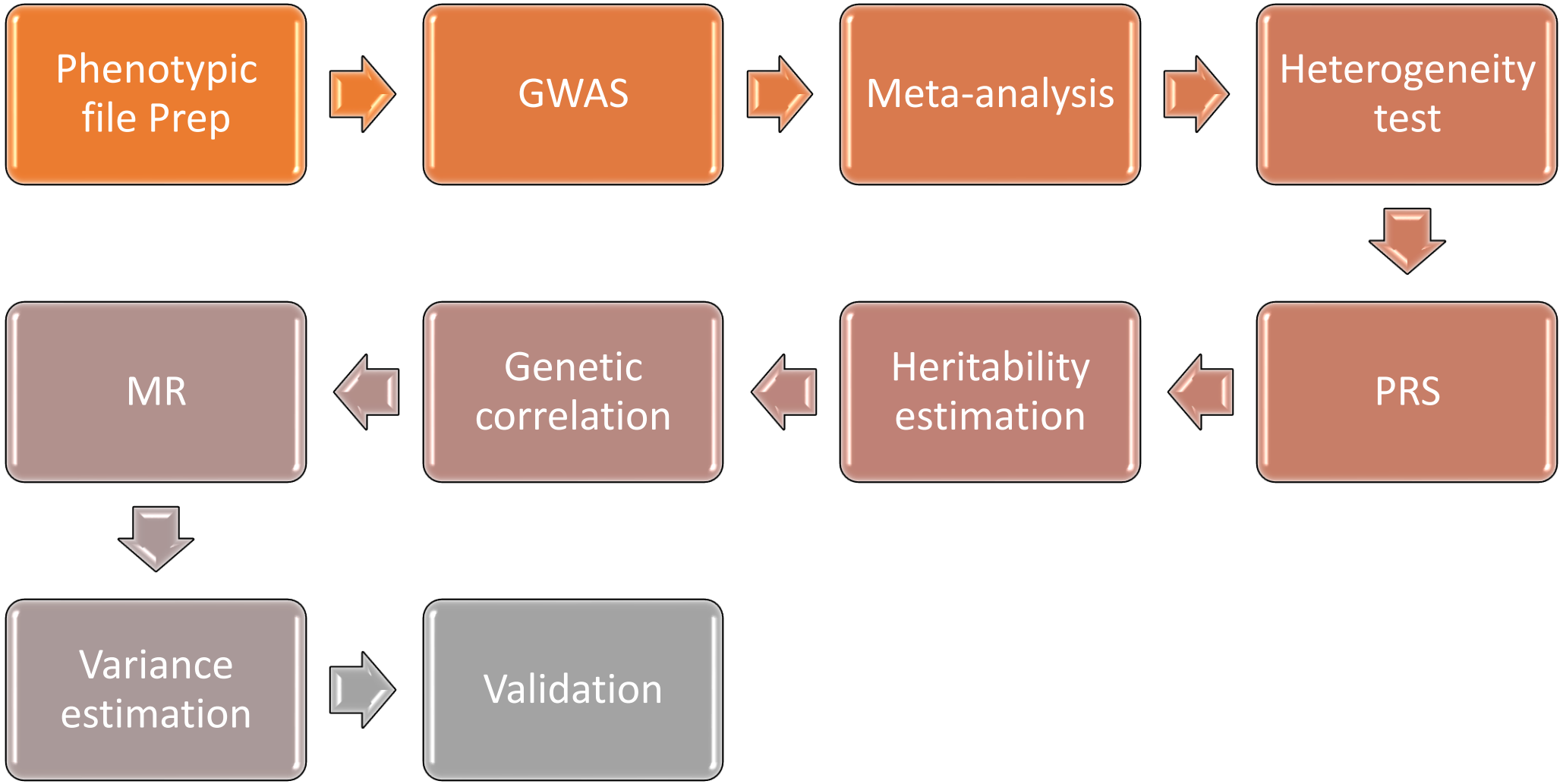
Study Questions

1. What are the genetic variants associated with HDL-c among South Indian and Scottish T2D Population ?
2. Is their any heterogeneity among genetic variants exist between two study population ?
3. Can PRS for HDL-c behave similarly between two study population?
4. How much heritability for HDL-c exist between two study population ?
5. How much genetic correlation present between the study population?
6. Are causal estimate by GLGC PRS (IV) similar or varied between the study group and can we quantify it?
7. How much difference in HDL-c levels caused by genetic variation between the study population ?

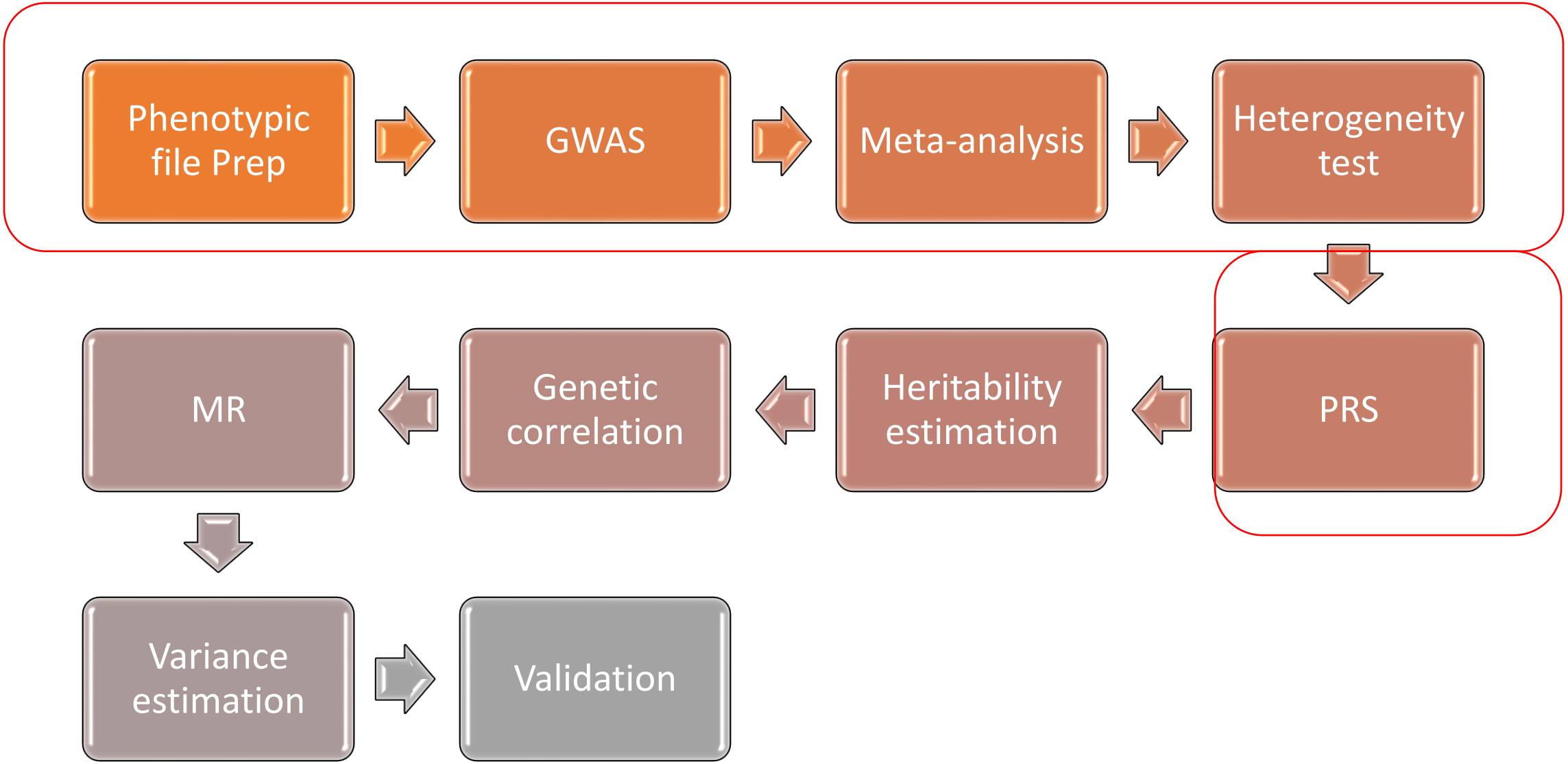
Study tools

Order	Analysis	Software used
1	GWAS	snptest , BOLT-LMM
2	Conditional analysis	snptest, GCTA
3	Meta analysis	GWAMA
4	Heterogeneity Analysis (Q , I^2)	GWAMA, MANTRA
5	Annotation, Visualisation, and functional consequences of genes	FUMA, LocusZoom
6	Gene based test / gene set analysis	MAGMA (provided by FUMA)
7	Polygenic risk score (PRS)	PRSice-2
8	Heritability	SumHer
9	Genetic Correlation	SumHer, Popcorn
10	MR	MRbase

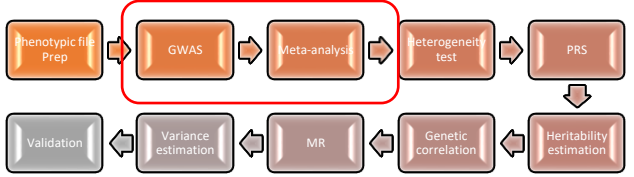
Study method…

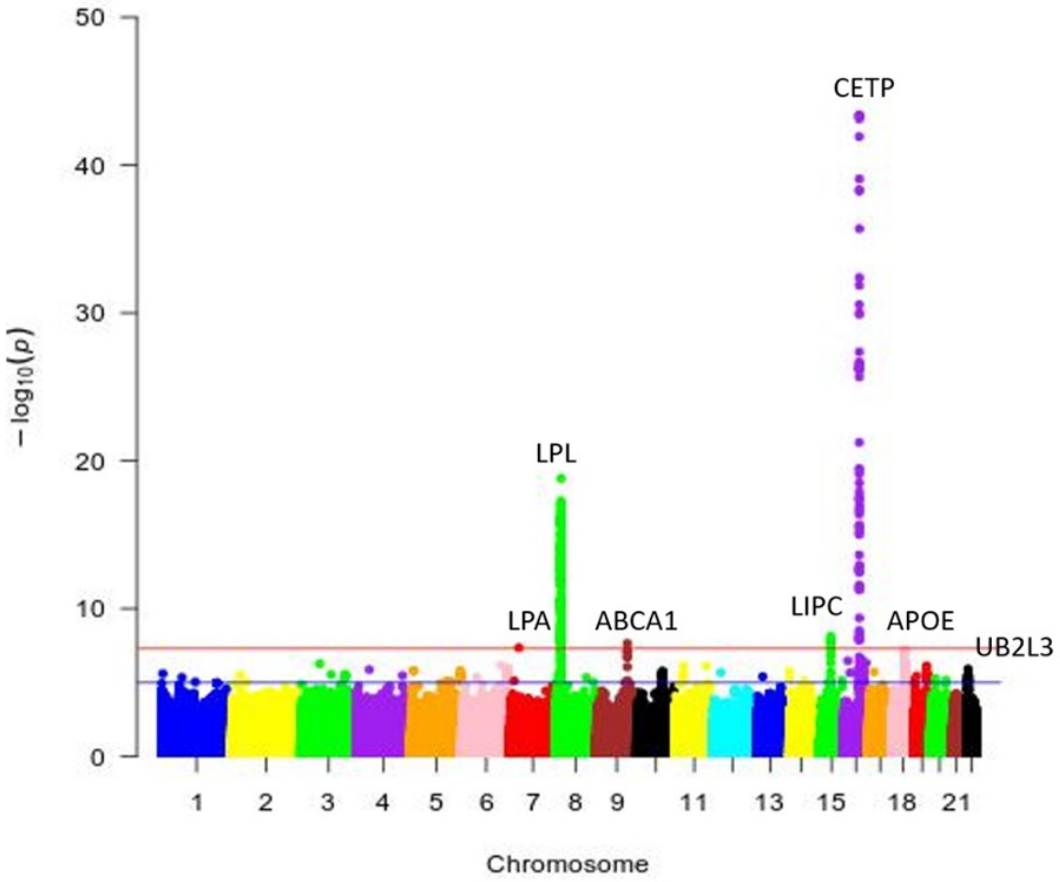


Study method…

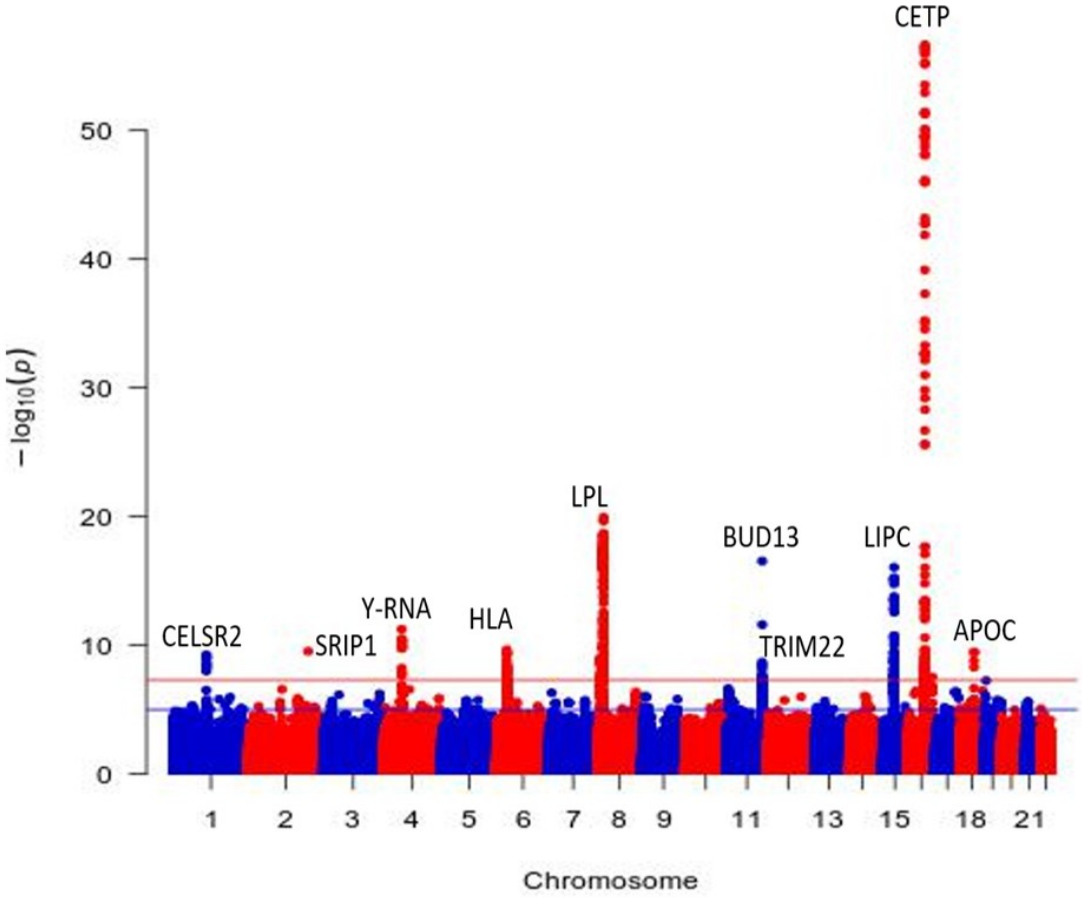


Manhattan plots for meta-analysis of GWAS from both the cohorts





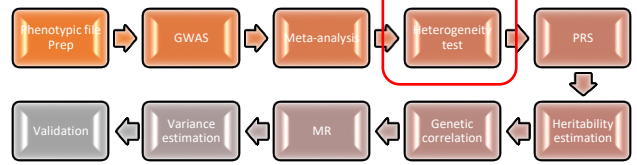
a. South Indian, India

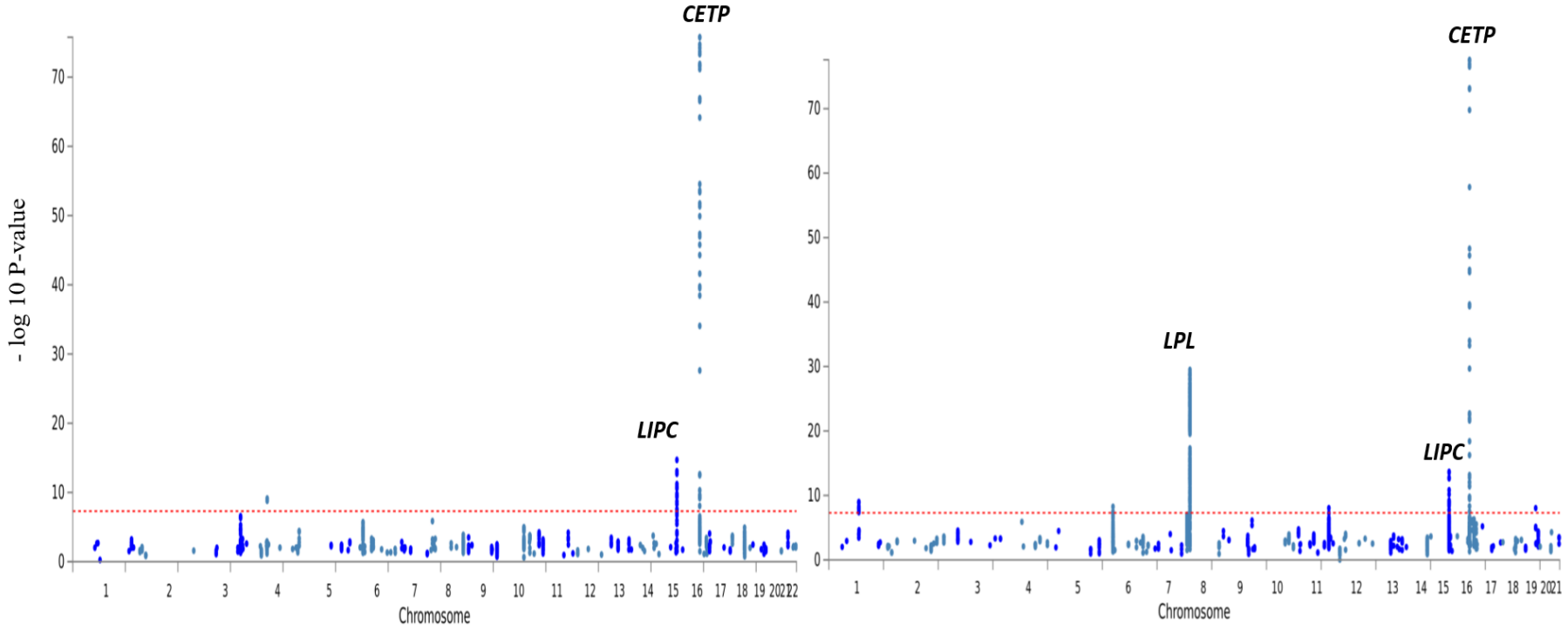


b. Scottish, UK

Adjusted for Age and Sex

Manhattan plot of trans-ethnic meta-analysis of selected SNPs

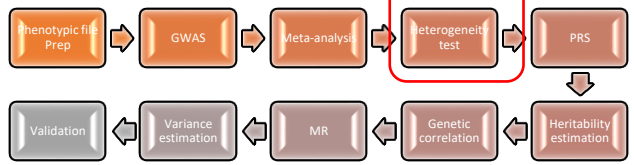




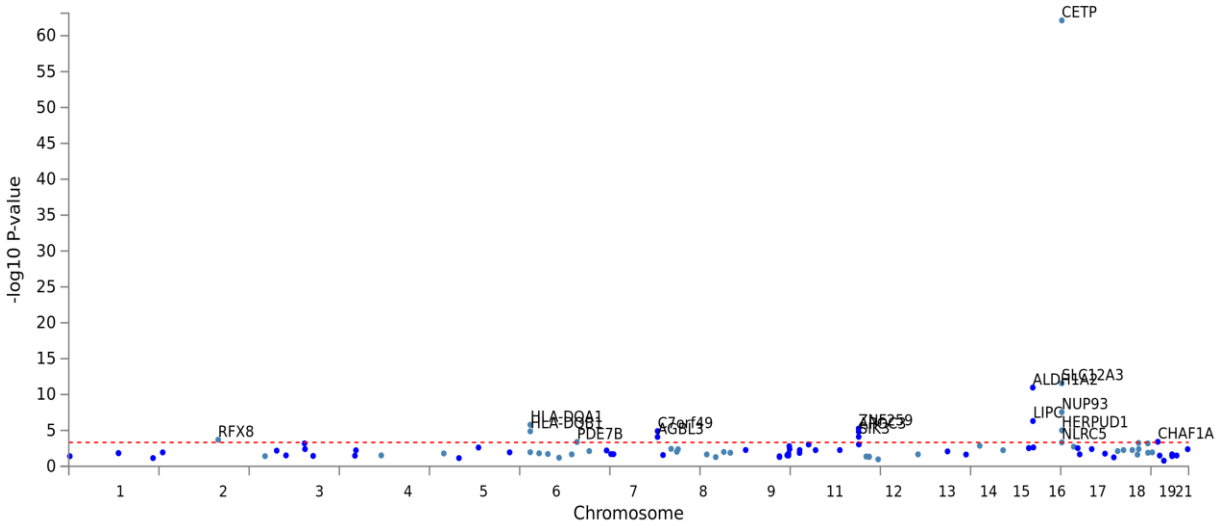
a. SNPs with Posterior probability > 0.5 (n= 890)

b. SNPs with Posterior probability ≤ 0.5 (n = 1,714)

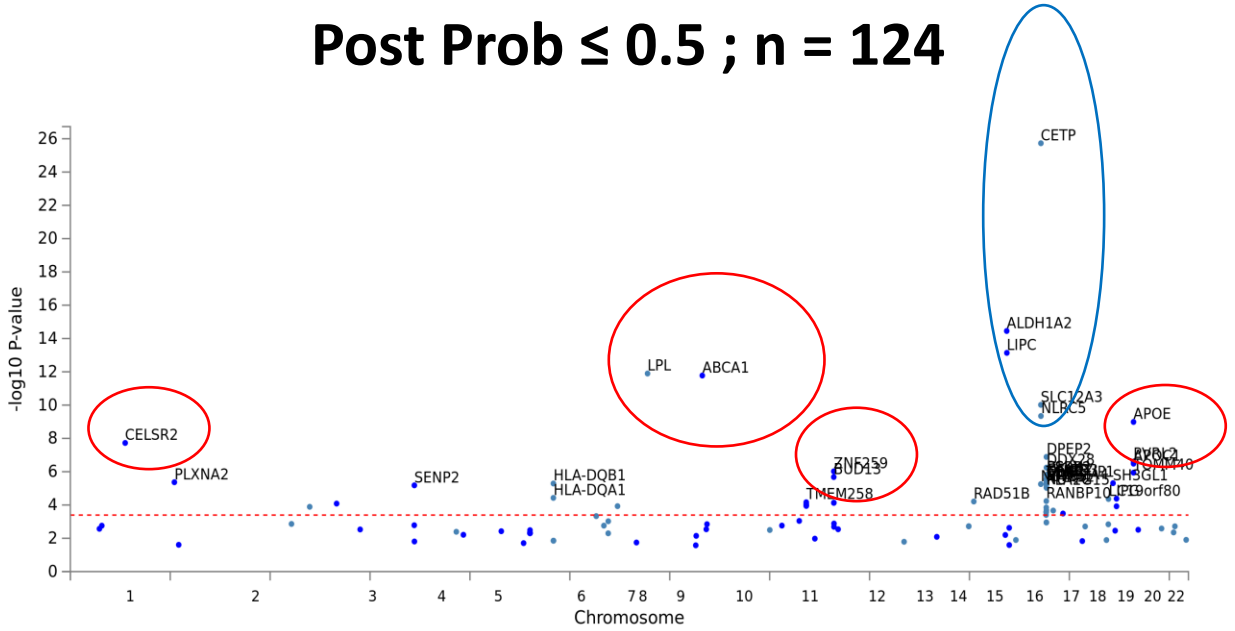
Comparison of genes Manhattan plot heterogenous Vs non heterogenous

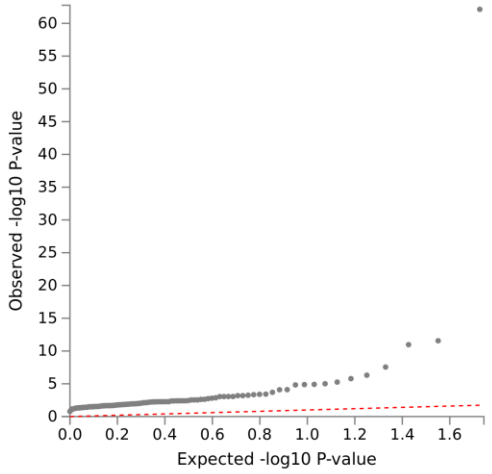


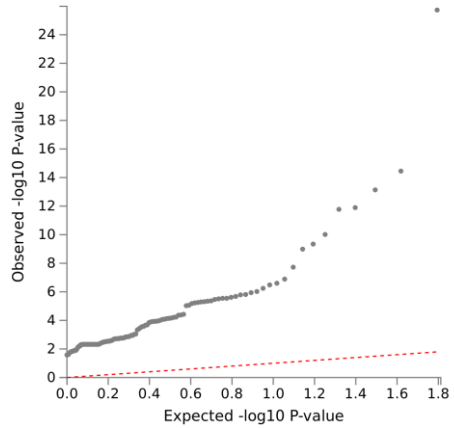
Post Prob > 0.5, n= 106



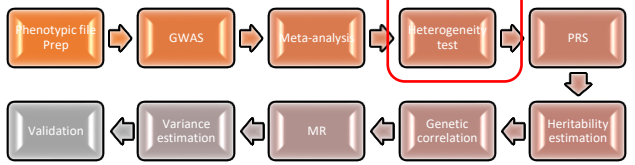
Post Prob ≤ 0.5 ; n = 124



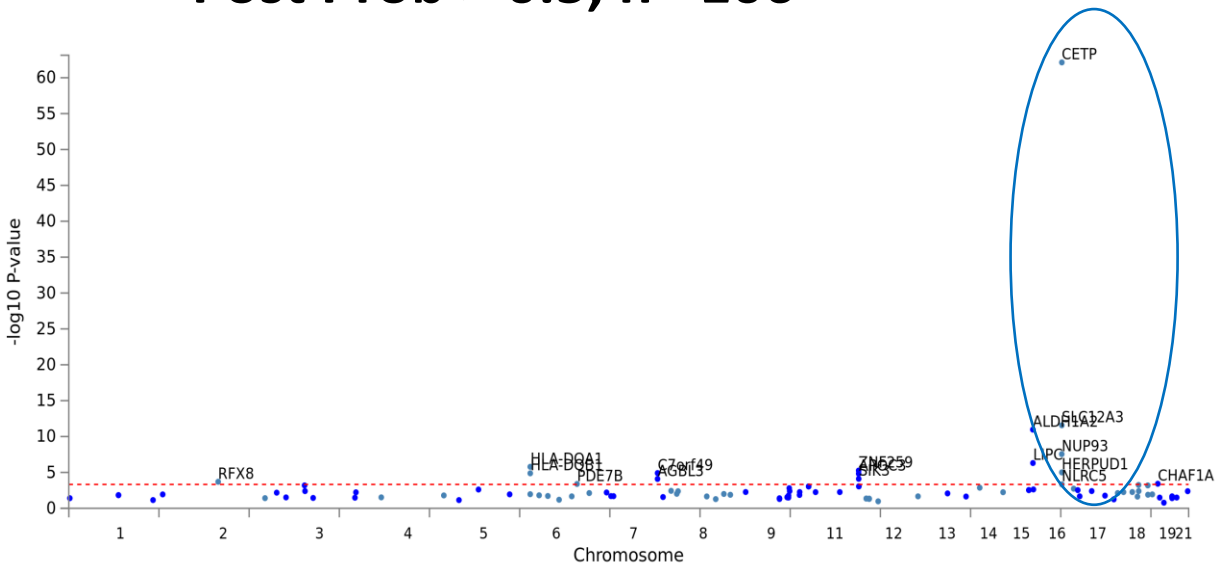




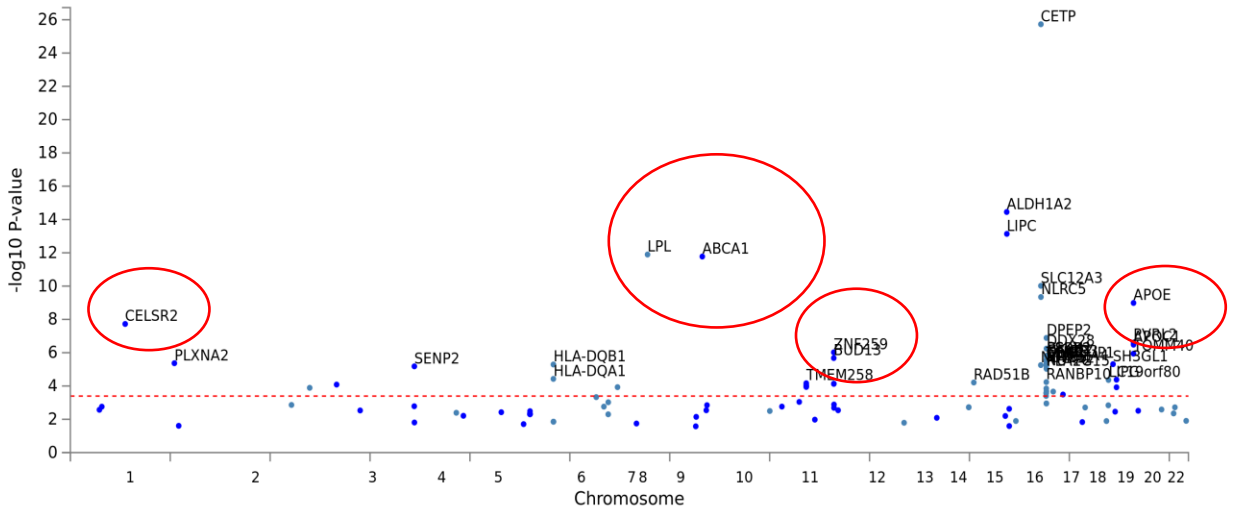
Comparison of genes Manhattan plot heterogenous Vs non heterogenous

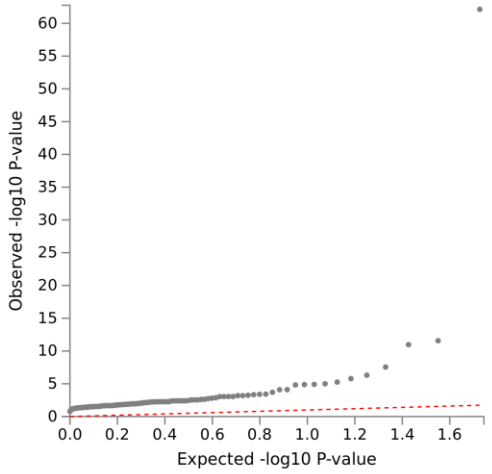


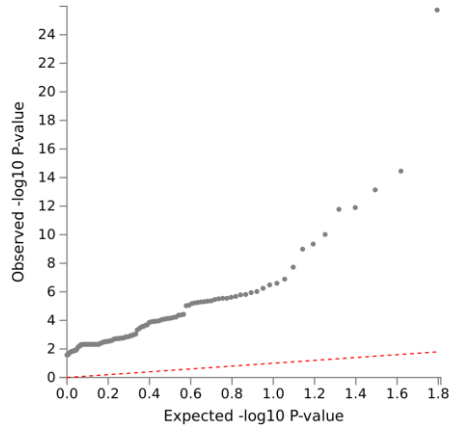
Post Prob > 0.5, n= 106



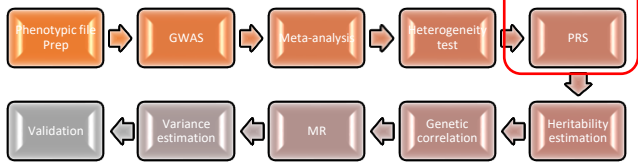
Post Prob ≤ 0.5 ; n = 124

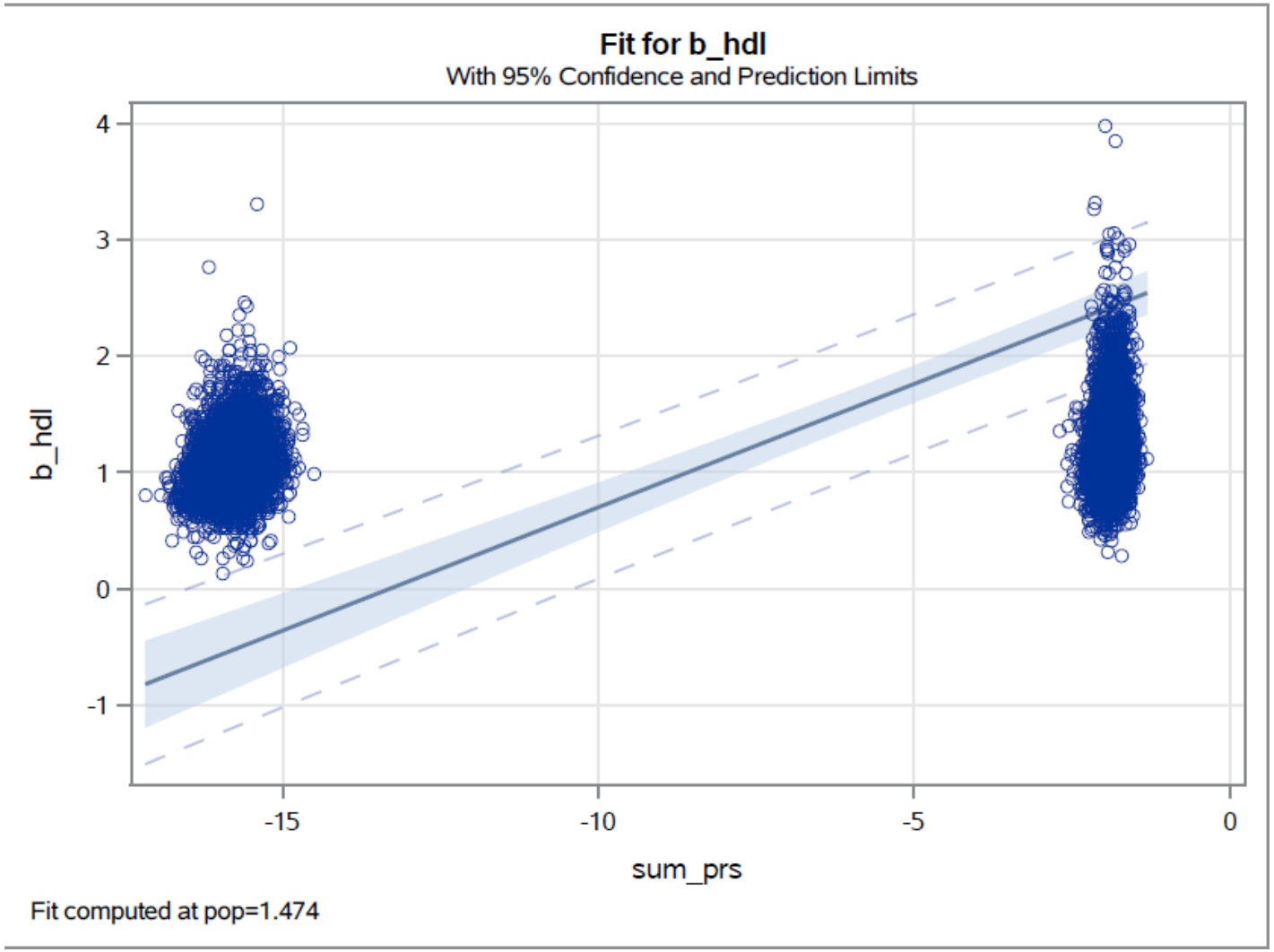






Effect of PRS on baseline HDL-c between the population





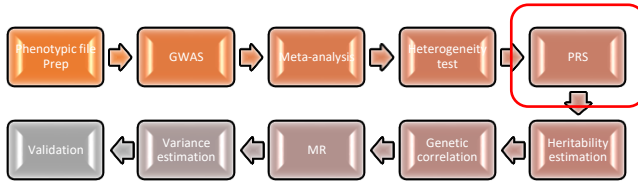
Dependent Variable: b_hdl

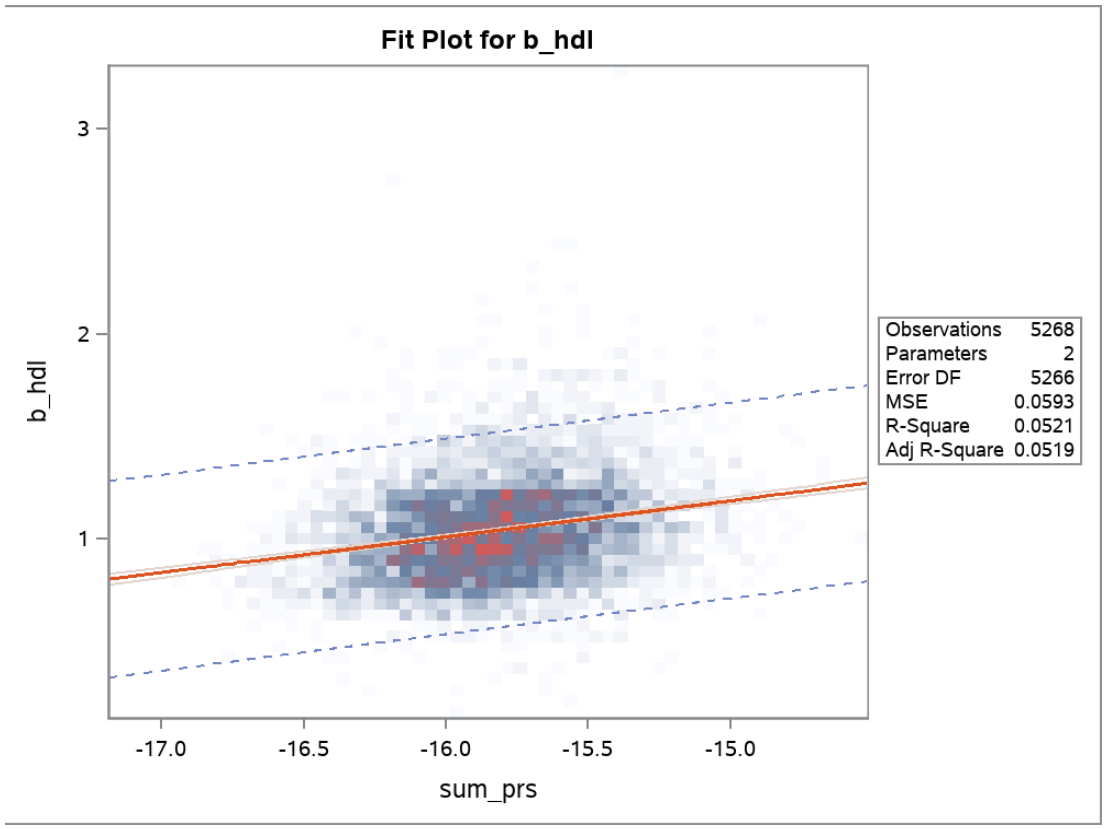
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	3	109.3230082	36.44100274	419.91	<.0001
Error	10002	868.0059319	0.0867832365		
Corrected Total	10005	977.3289401			

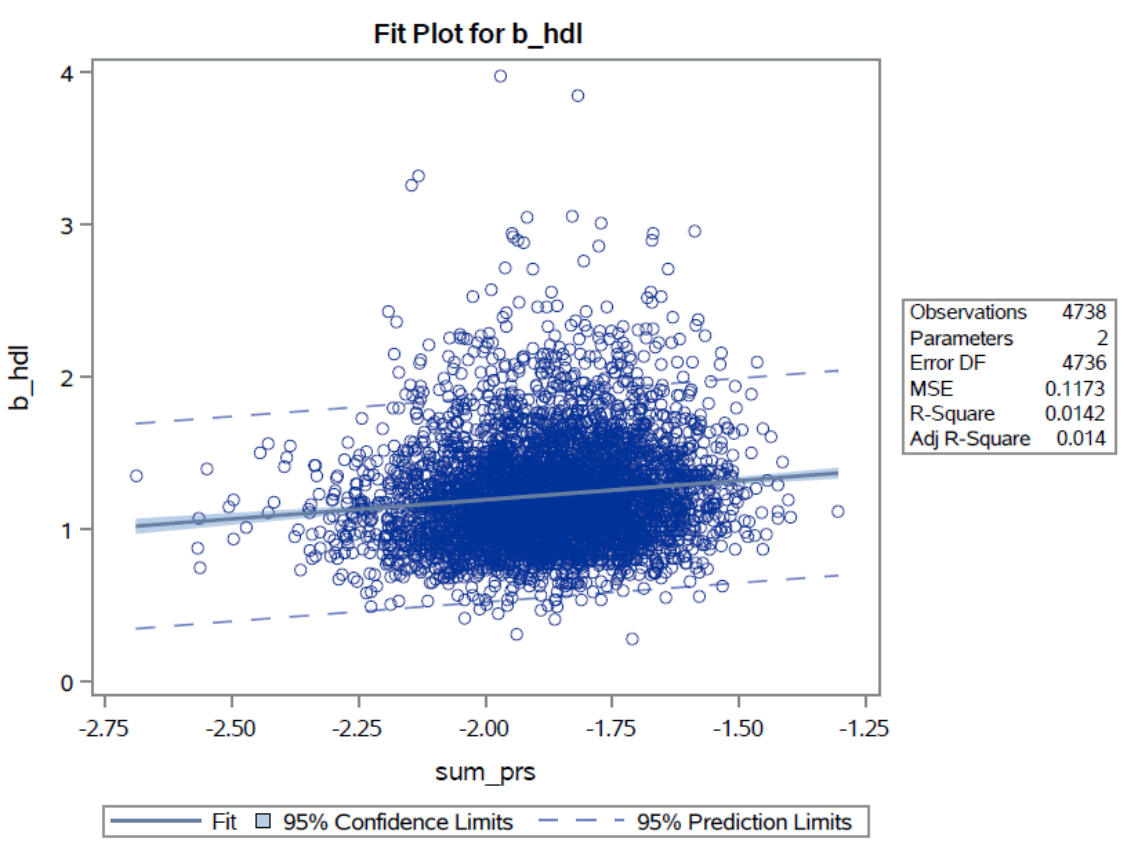
Root MSE	0.2945899464
R-Square	0.1118589696

Parameter	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	5.951009305069	0.398903217	14.92	<.0001
PRS	0.100067184963	0.0362155763	2.76	0.0057
Population	-2.126337842499	0.203973295	-10.42	<.0001

Regression of Baseline HDL-c (a) with Polygenic risk score ($PRS_{GLGC\text{-}new}$)





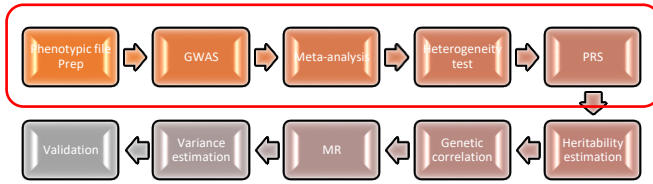


Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	3.82467	0.16361	23.38	<.0001
sum_prs	sum_prs	1	0.17592	0.01034	17.01	<.0001

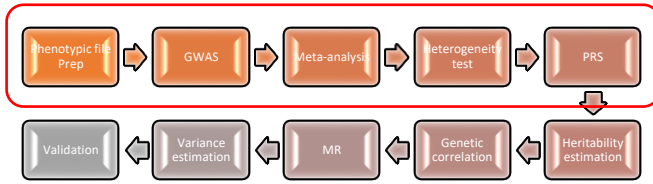
Parameter Estimates

Variable	Label	DF	Parameter Estimate	Standard Error	t Value	Pr > |t|
Intercept	Intercept	1	1.69833	0.05735	29.62	<.0001
sum_prs	sum_prs	1	0.25178	0.03044	8.27	<.0001

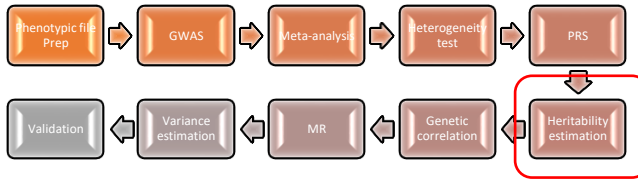


Conclusion

- Heterogeneity in genetic architecture for HDL- Cholesterol exist such as *CETP* region between the study population.
- PRS suggests that genetic loci for HDL-c in Scottish populations are present at both similar and varied frequency compared with South Indian populations, which modulate the HDL-c at certain levels
- Higher HDL-c levels was observed among females with increase in PRS within the population shows an effect of gender on PRS and HDL-c levels.



Follow up work …



nature genetics

ARTICLES

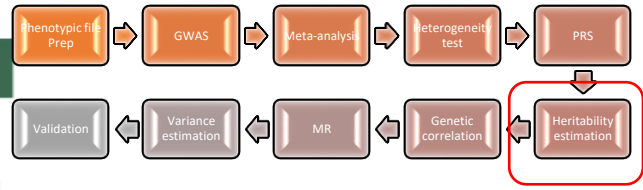
https://doi.org/10.1038/s41588-018-0279-5

SumHer better estimates the SNP heritability of complex traits from summary statistics

Doug Speed[1,2,3*] and David J. Balding[3,4]

LDAK heritability model:

- a SNP with high MAF is expected to contribute more heritability than one with low MAF
- a SNP in a region of low linkage disequilibrium is expected to contribute more than one in a region of high linkage disequilibrium.
- By contrast, LDSC estimates are obtained by setting q_j=1, which corresponds to the assumption that **all SNPs** are expected to **contribute equally** 1



ARTICLES
https://doi.org/10.1038/s41588-018-0279-5
nature genetics

SumHer better estimates the SNP heritability of complex traits from summary statistics

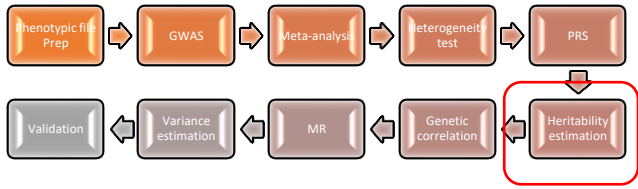
Doug Speed[1,2,3*] and David J. Balding[3,4]

Table 1 | Estimates of h^2_{SNP} and confounding bias for the 24 summary GWAS

			LDSC				SumHer-GC				No. of significant loci after dividing test statistics by			
Trait (disease prevalence, %)	*n*	GIF	h^2_{SNP}	s.d.	1+*A*	s.d.	h^2_{SNP}	s.d.	*C*	s.d.	1	GIF	1+*A*	*C*
Alzheimer's disease[34] (7.5)	54,000	1.09	0.03	0.01	1.07	0.02	0.12	0.03	1.03	0.01	21	19	19	21
Coronary artery[35] (6)	79,000	1.10	0.04	0.01	1.06	0.01	0.15	0.02	0.99	0.01	10	6	7	10
Crohn's disease[36] (0.5)	21,000	1.14	0.15	0.03	1.08	0.01	0.47	0.06	0.97	0.02	64	52	58	64
Ever smoked?[37] (56)	74,000	1.11	0.08	0.01	1.02	0.01	0.19	0.02	0.96	0.01	0	0	0	0
Inflammatory bowel disease[36] (0.7)	35,000	1.17	0.09	0.02	1.13	0.01	0.33	0.03	0.98	0.01	78	59	65	80
Rheumatoid arthritis[38] (0.5)	58,000	1.05	0.05	0.01	1.00	0.01	0.17	0.03	0.90	0.02	109	104	109	123
Schizophrenia[39] (1)	82,000	1.57	0.19	0.01	1.16	0.01	0.42	0.02	0.91	0.01	105	23	63	140
Type 2 diabetes[40] (8)	157,000	1.17	0.08	0.01	1.07	0.01	0.23	0.02	0.95	0.01	38	25	32	42
Ulcerative colitis[36] (0.2)	27,000	1.12	0.06	0.01	1.10	0.01	0.27	0.03	0.99	0.01	38	31	31	38
Bone mineral density[41]	33,000	1.11	0.10	0.02	1.07	0.01	0.28	0.04	1.00	0.01	19	18	18	19
Body mass index[25]	230,000	1.13	0.09	0.01	0.80	0.01	0.33	0.03	0.55	0.02	69	52	135	336
Depressive symptoms[42]	161,000	1.12	0.02	0.00	1.03	0.01	0.07	0.01	0.96	0.01	0	0	0	1
Fasting glucose[43]	58,000	1.08	0.05	0.01	1.04	0.01	0.14	0.03	0.99	0.01	22	20	20	23
Glycated hemoglobin[44]	46,000	1.04	0.02	0.01	1.03	0.01	0.10	0.02	0.99	0.01	10	10	10	10
HDL cholesterol[26]	96,000	1.03	0.07	0.03	1.04	0.07	0.50	0.09	0.68	0.03	130	122	121	216
Height[45]	246,000	2.09	0.20	0.02	1.69	0.06	0.46	0.04	0.98	0.04	720	196	288	754
LDL cholesterol[26]	91,000	1.03	0.08	0.03	1.00	0.04	0.43	0.10	0.73	0.04	101	96	101	155
Menarche age[46]	253,000	1.66	0.15	0.01	1.21	0.02	0.32	0.02	0.89	0.02	289	111	190	354
Menopause age[47]	69,000	1.10	0.06	0.01	1.06	0.02	0.25	0.03	0.92	0.02	49	39	39	55
Neuroticism[42]	171,000	1.26	0.06	0.01	1.06	0.01	0.17	0.02	0.90	0.02	10	4	7	18
Subjective well-being[42]	298,000	1.16	0.02	0.00	1.03	0.01	0.04	0.00	0.97	0.02	0	0	0	0
Triglyceride[26]	92,000	1.02	0.14	0.04	0.92	0.03	0.45	0.11	0.70	0.04	82	82	91	152
Waist-hip ratio[48]	142,000	1.05	0.06	0.01	0.92	0.01	0.20	0.02	0.76	0.01	26	23	33	66
Years of education[49]	329,000	1.54	0.07	0.00	1.11	0.01	0.20	0.01	0.83	0.01	70	13	46	148
Average	**121,000**	**1.21**	**0.04**	**0.00**	**1.04**	**0.00**	**0.12**	**0.00**	**0.93**	**0.00**	**86**	**46**	**62**	**118**
Total											**2,060**	**1,105**	**1,483**	**2,825**

Columns 2 and 3 report the average sample size and the genomic inflation factor (calculated using the published test statistics). Columns 4–11 report estimates of h^2_{SNP} and confounding bias from both LDSC and SumHer-GC (LDSC measures confounding bias via the intercept, 1+*A*, while SumHer-GC uses the scaling factor, *C*). For binary traits, estimates of h^2_{SNP} have been converted to the liability scale, assuming the stated prevalence. Columns 12–15 report the number of significant loci based on the published test statistics, then after correction via genomic control, LDSC and SumHer-GC (dividing test statistics by the GIF, 1+*A* and *C*, respectively).

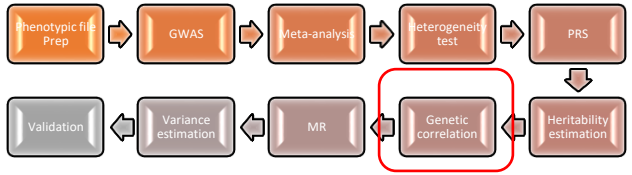
26 . Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. Nat. Genet. 45, 1274–1283 (2013)



SNP heritability (the heritability contributed by all SNPs)

- GLGC $H^2_{SNP} = 0.51$
- Scottish $H^2_{SNP} = 0.43$
- South Indian $H^2_{SNP} = 0.46$

Genetic Correlations between the study cohorts*



Estimating genetic correlation

- Suppose we have summary statistics from two GWAS.
- Instead of $\chi^2(1)$ test statistics, we now use (signed) Z-statistics.
- Let Z_{Aj} and Z_{Bj} denote the two Z-statistics for SNP j, computed using n_{Aj} and n_{Bj} individuals, respectively, of which n_{Cj} were common to both GWAS (if the two GWAS were independent, $n_{Cj} = 0$).
- We assume

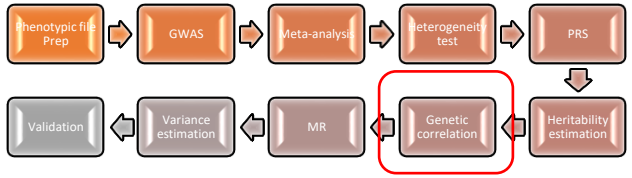
$$E\left[Z_{A_j} Z_{B_j}\right] \approx \frac{c_{AB} n_{S_j}}{\sqrt{n_{A_j} n_{B_j}}} + w\prime_j h^2_{AB} \quad \text{with}$$

$$w\prime_j = \sqrt{n_{A_j} n_{B_j}} \left(q_j + \sum_{l \in N_j} q_l r^2_{jl} \right) / Q$$

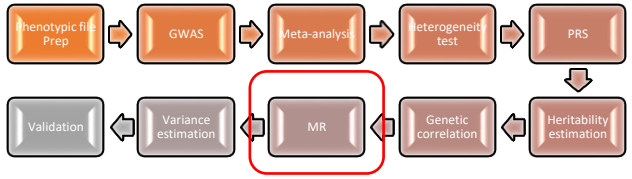
- where c_{AB} is the phenotypic correlation between the two traits and h^2_{AB} is their genetic covariance.
- This equation matches that used by LDSC, except that we have replaced r^2jl/m by qjr^2jl/Q.
- By regressing $Z_{Aj}Z_{Bj}$ on $u_{'j}$, we obtain an estimate of h^2AB, which we then divide by estimates of $\sqrt{h^2_{SNP}}$ for each trait to get an estimate of their genetic correlation.

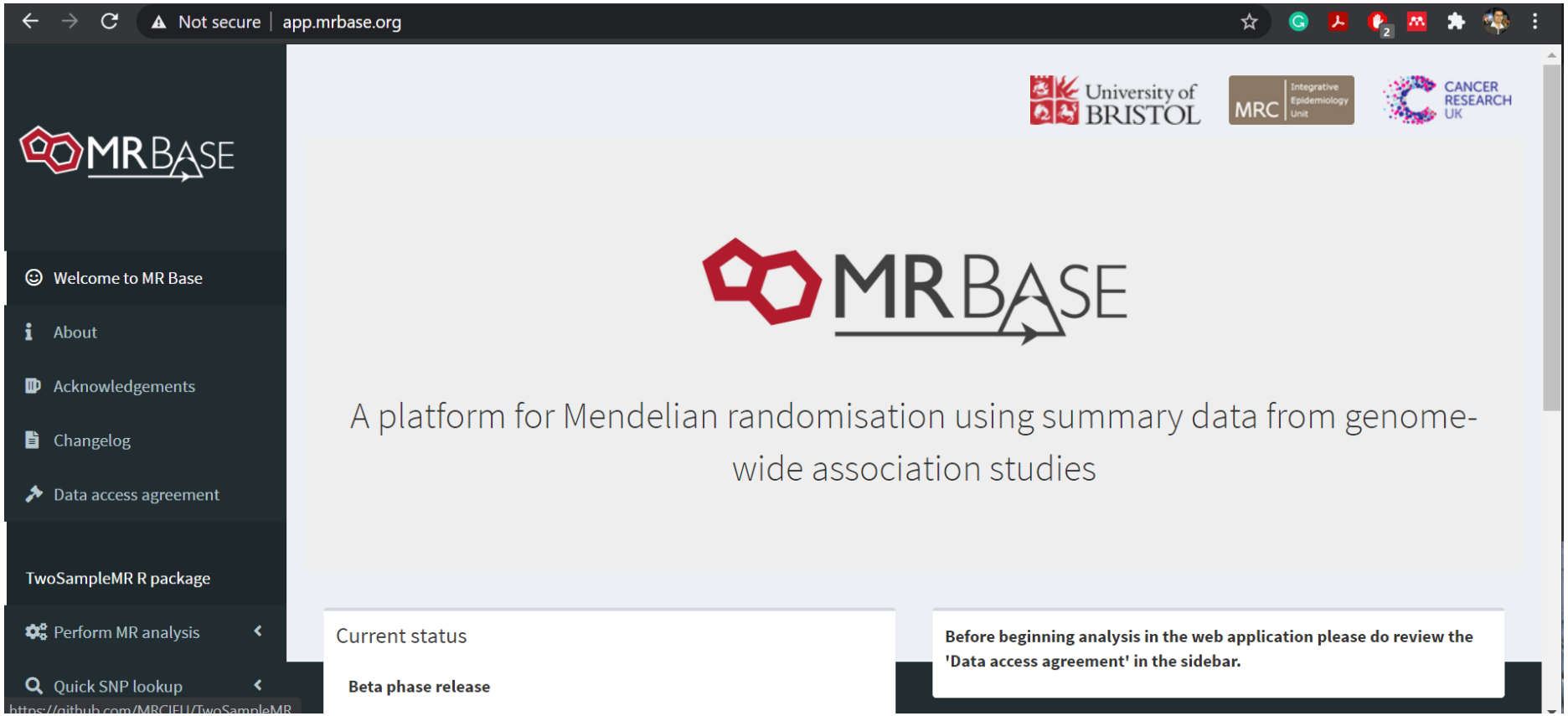
*Speed, D., Balding, D.J. SumHer better estimates the SNP heritability of complex traits from summary statistics. *Nat Genet* **51,** 277–284 (2019). https://doi.org/10.1038/s41588-018-0279-5

Genetic correlation between the population



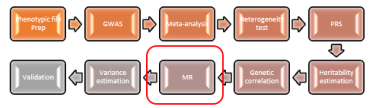
- The estimated genetic correlation = **0.45(SD 0.2)**
- Overlap of samples **0.048 (SD 0.020)**

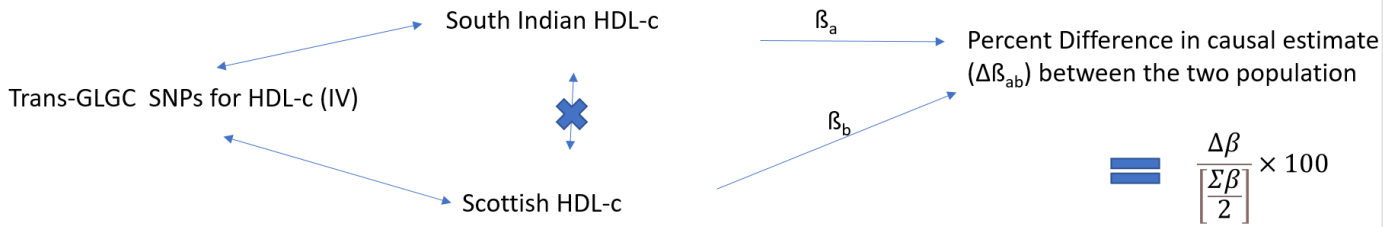




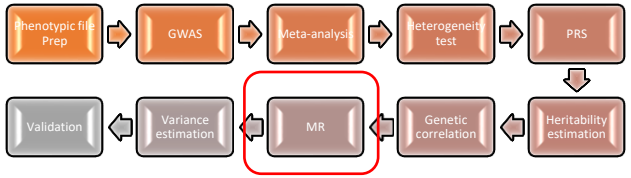
Mendelian randomization (MR) is a method for estimating the causal relationship between an exposure and an outcome using a genetic factor as an instrumental variable (IV) for the exposure

Causal estimate work flow diagram





23



Causal estimate work flow diagram

South Indian HDL-c

Trans-GLGC SNPs for HDL-c (IV)

Scottish HDL-c

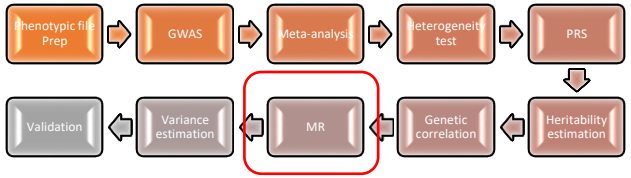
$ß_a$

$ß_b$

Percent Difference in causal estimate ($\Delta ß_{ab}$) between the two population

$$= \frac{\Delta\beta}{\left[\frac{\Sigma\beta}{2}\right]} \times 100$$

Selection of the Instrumental variable (IV) i.e SNPs



SNPs reported for lipid traits; P value < 5*10^-8,
n = 2,399

Duplicate SNPs and SNPs with other lipid traits;
n = 1,837

SNPs specific to HDL-c; n = 562

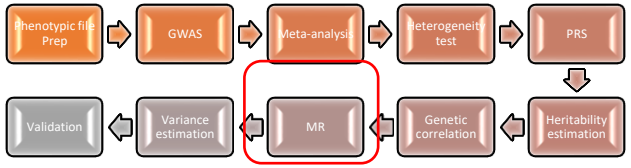
QC failed SNPs, Palindromic SNPs; n =20

SNPs $_{-\text{HDL-c}}$; P value < 5*10^-8 ;
n = 542

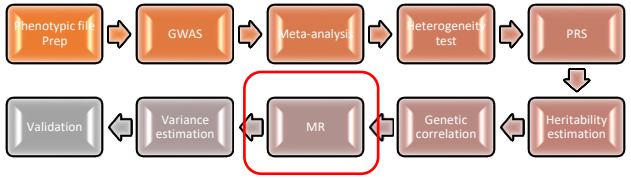
Scottish HDL-c GWAS SNPs
n ~500

South Indian HDL-c GWAS SNPs;
n ~500

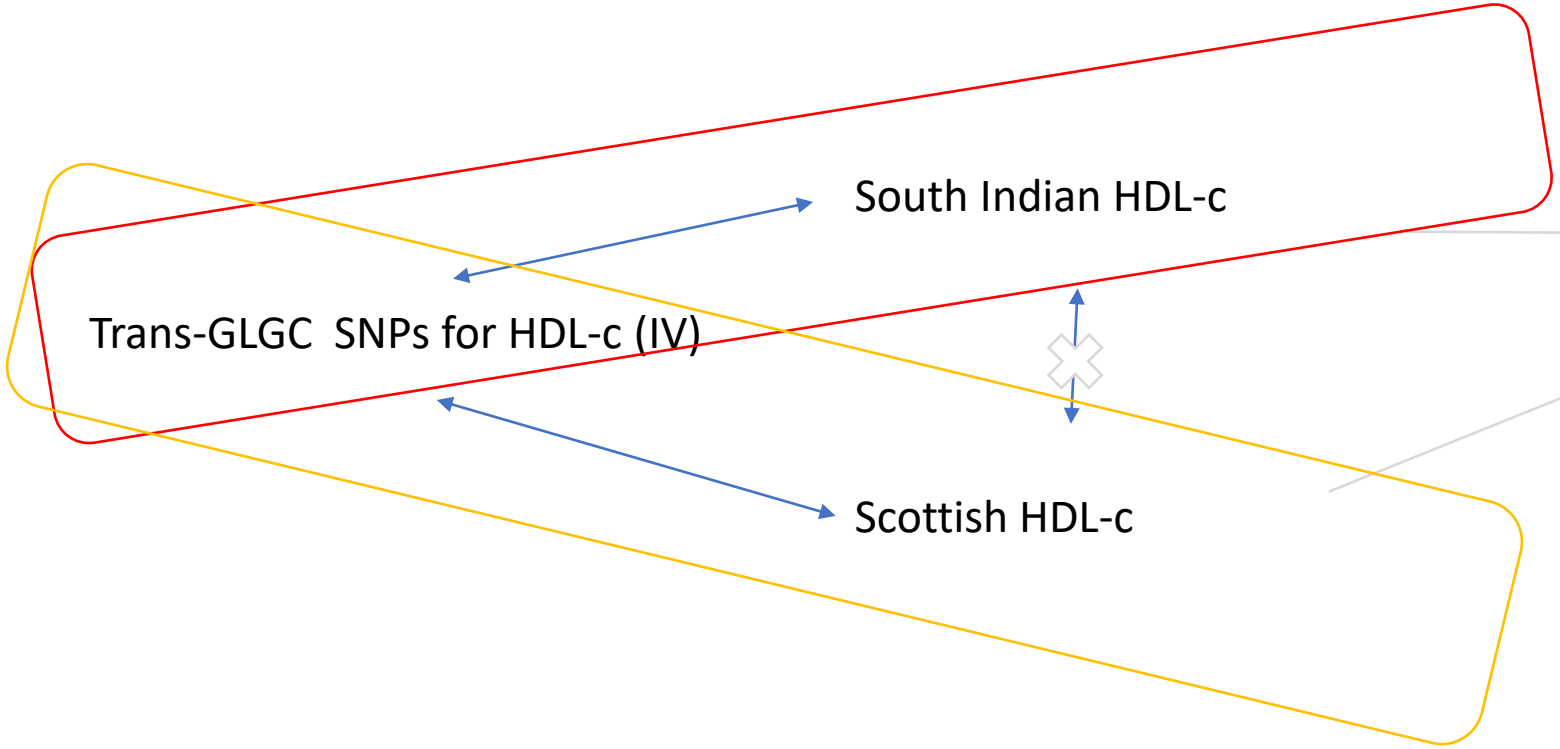
Analysis Performed



	Test	Method used
1.	Causality Estimate	MR Egger
		Weighted median
		Inverse variance weighted
		Simple mode
		Weighted mode
2.	Horizontal Pleiotropy	Egger intercept
3.	Heterogeneity	Q statistics
4.	Outlier detection	Inverse variance weighted



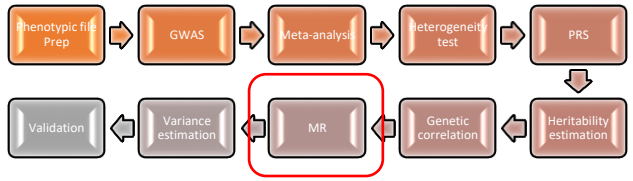
Causal estimate work flow diagram

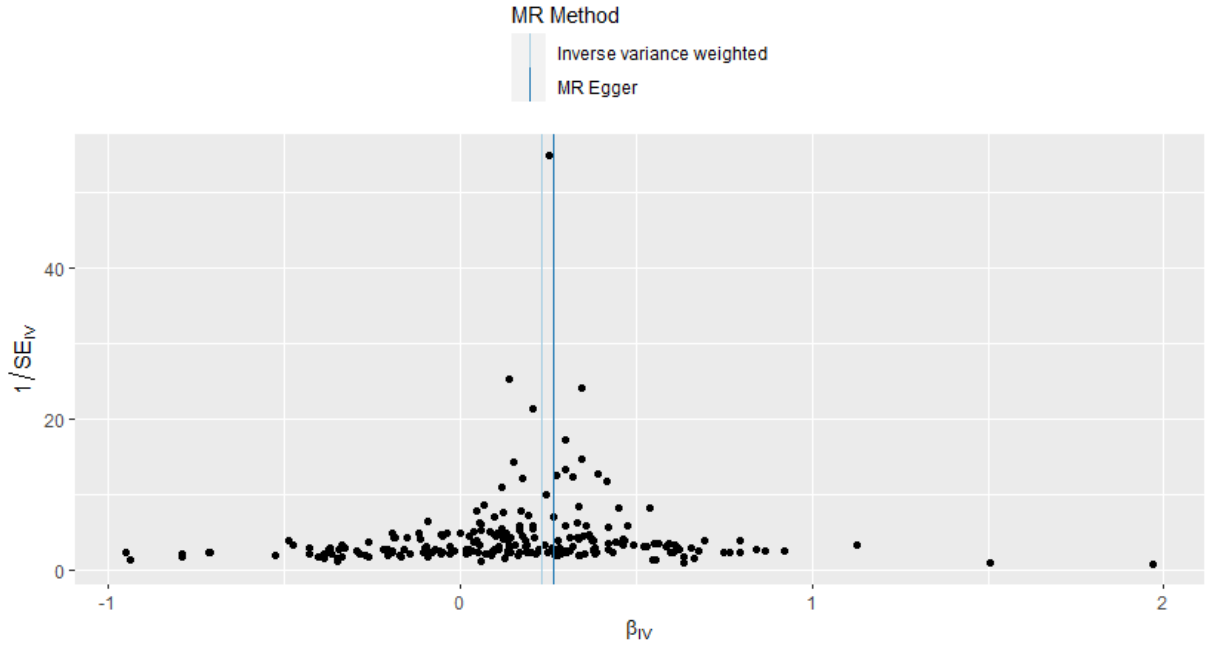


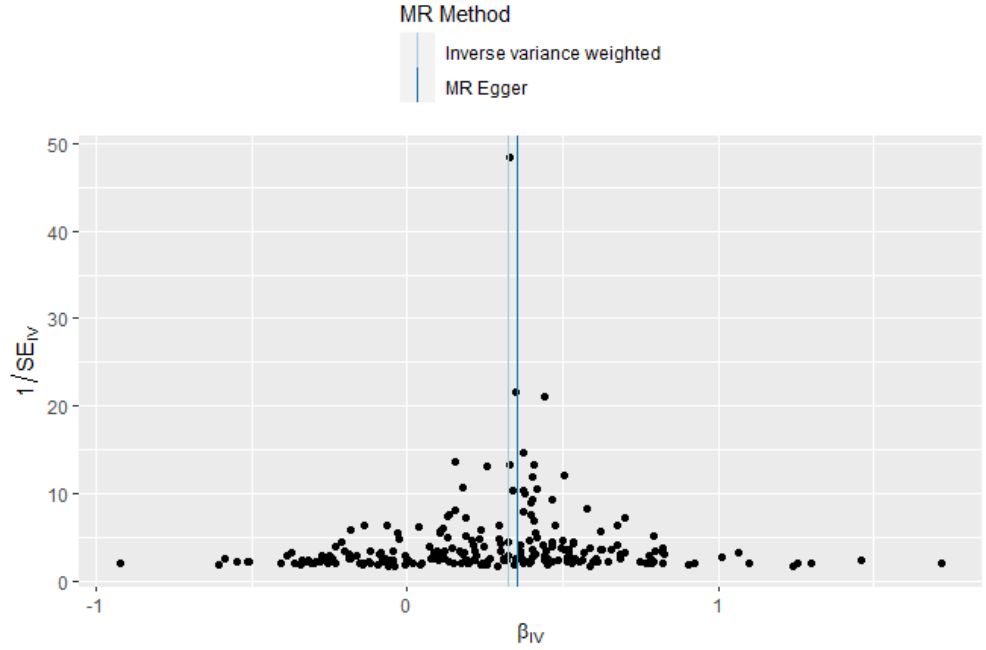
Percent Difference in causal estimate ($\Delta ß_{ab}$) between the two population

$$= \frac{\Delta\beta}{\left[\frac{\Sigma\beta}{2}\right]} \times 100$$

Funnel plot



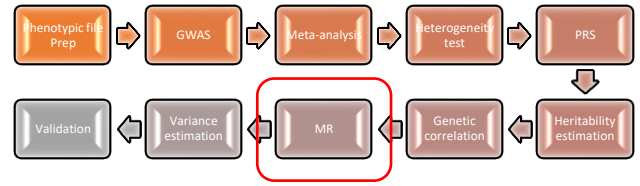


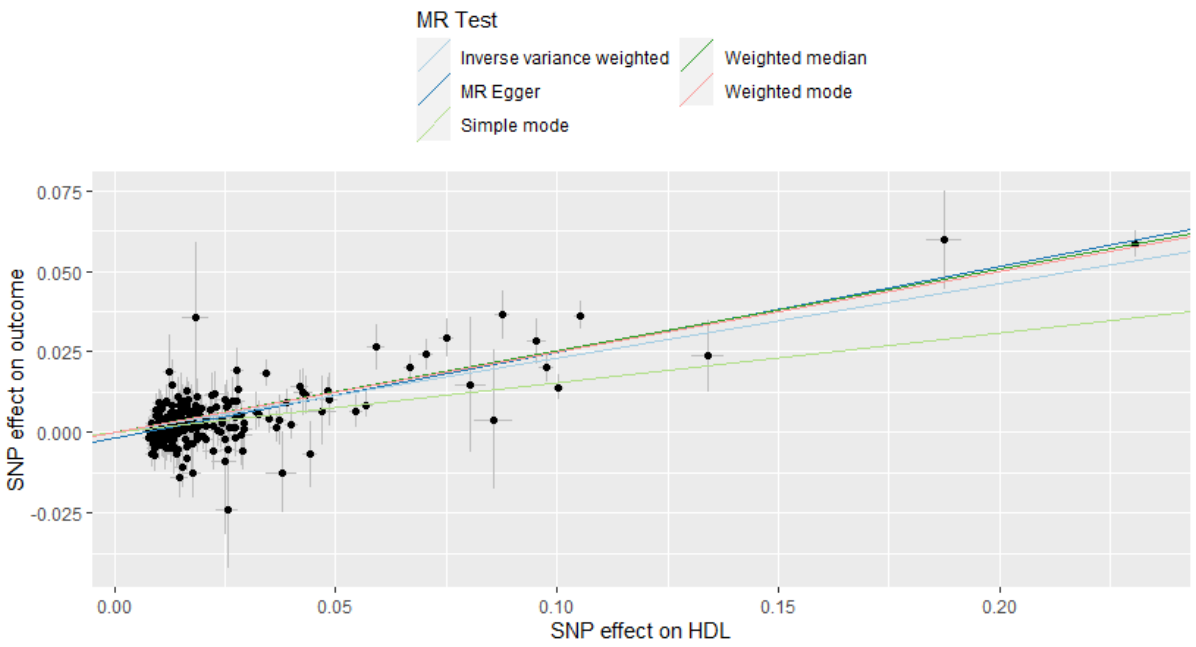


Test	Method	Q statistics	P value	Q statistics	P value
Heterogeneity test	MR Egger	242.98	0.148	291.4	0.0037
	Inverse variance weighted	257.5888	0.050	299.84	0.0015

- Funnel plot symmetry is indicative of no directional pleiotropy

Relationship of the SNP effects on the exposure (HDLc GLGC) against the SNP effects on the outcome(HDL-c South Indian and HDL-c Scottish)

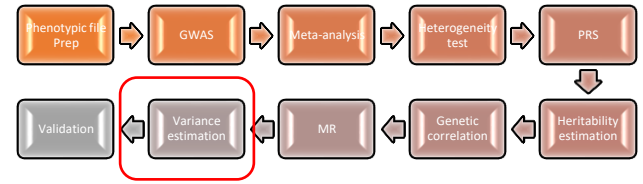


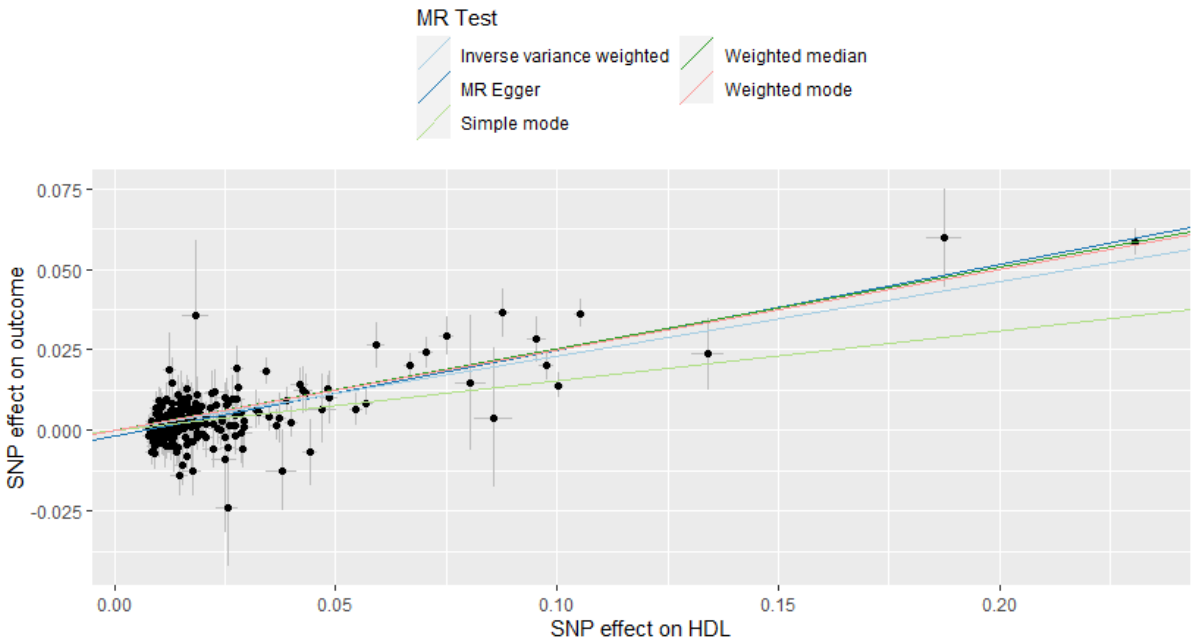


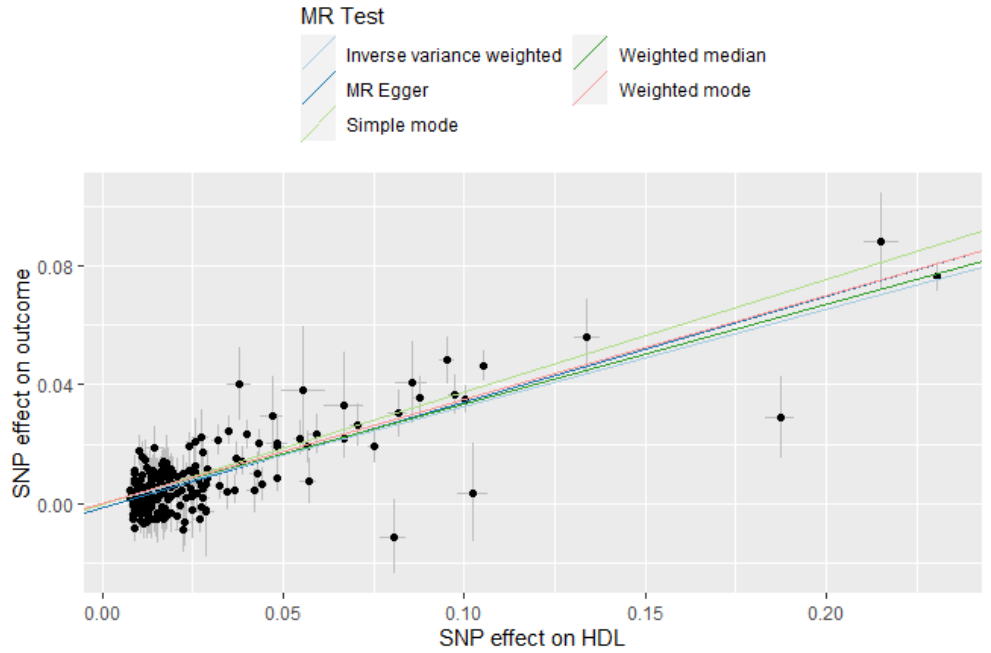
Method	Egger intercept (se); P value	
Pleiotropy test	South Indian	Scottish
	-0.0016(0.0004); 0.0003	0.0013(0.0005); 0.011

Test	snp_r2 outcome	Steiger P value	snp_r2 outcome	Steiger P value
Directionality	0.14	1.49e-40	0.13	1.37e-20

Relationship of the SNP effects on the exposure (HDLc GLGC) against the SNP effects on the outcome(HDL-c MDRF and HDL-c Scottish)

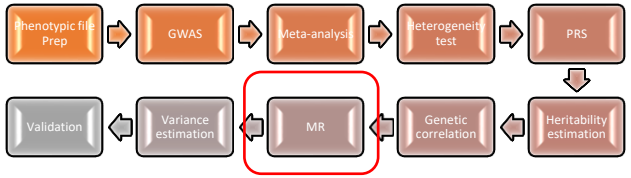






Sl no	Method (n snp =562)	South Indian		Scottish		T test ,P value
		b(se)	P value	b(se)	P value	
1	MR Egger	0.265(0.0143)	1.27e-46	0.35(0.0169)	4.19e-55	0.12; 95%CI (0.12 - 0.11)
2	Weighted median	0.253(0.0179)	4.14e-45	0.33(0.020)	1.29e-60	<0.0001
3	Inverse variance weighted	0.231(0.011)	3.68e-96	0.325(0.012)	3.47e-141	

Given a common exposure, causal estimate is 33 % significantly lower among South Indians for HDL-c as compared with Scottish population.



Causal estimate work flow diagram

Trans-GLGC SNPs for HDL-c (IV)

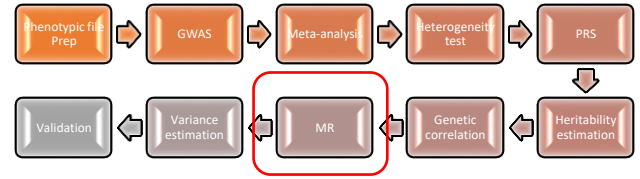
South Indian HDL-c

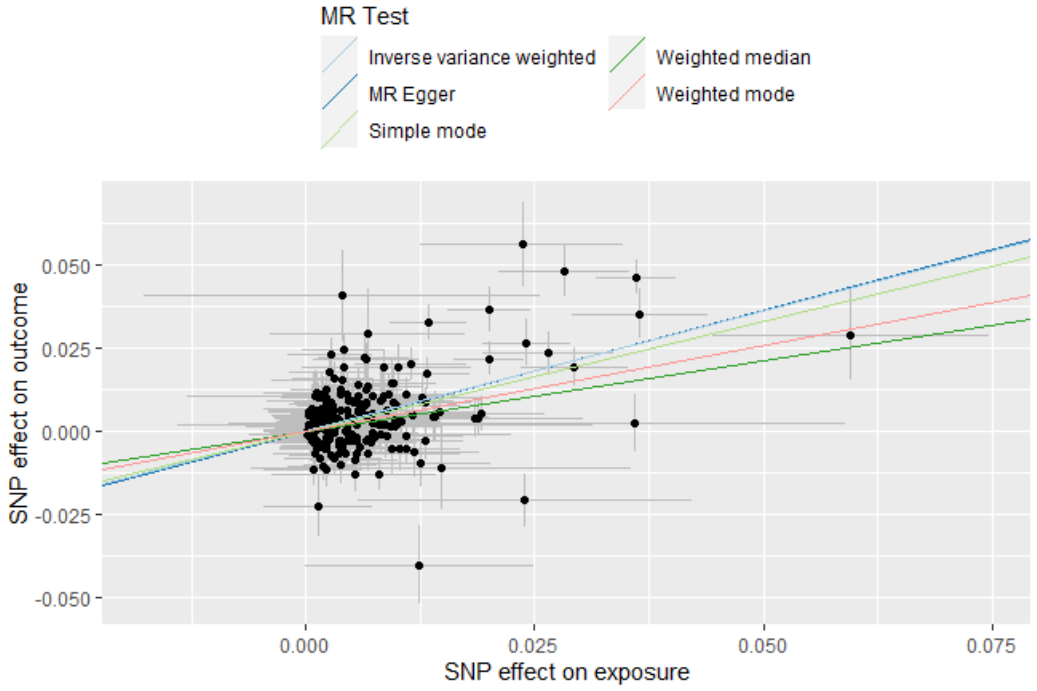
Scottish HDL-c

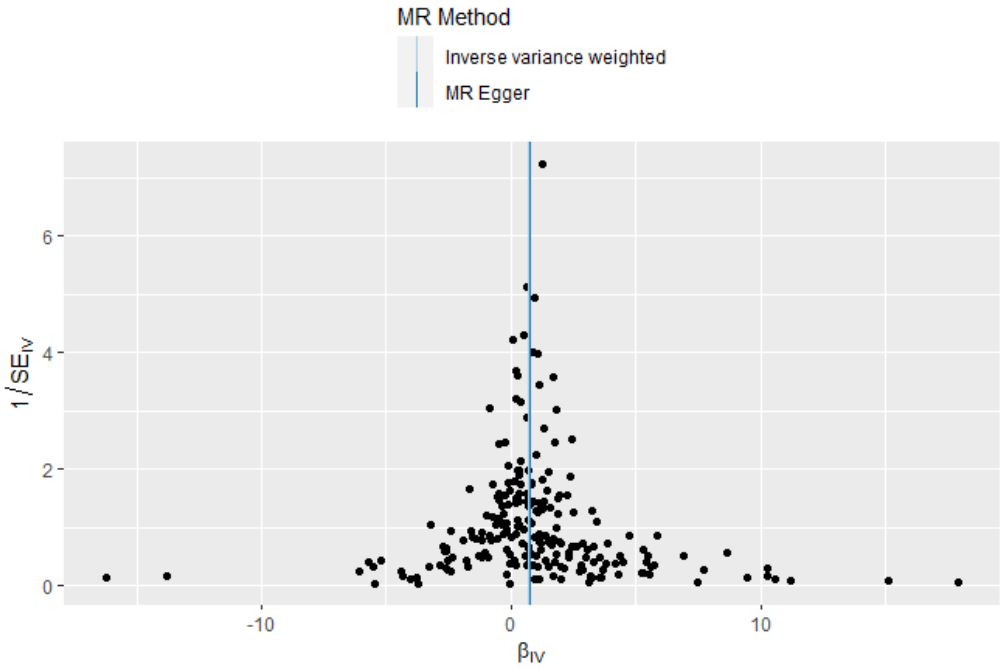
Percent Difference in causal estimate ($\Delta ß_{ab}$) between the two population

$$= \frac{\Delta\beta}{\left[\frac{\Sigma\beta}{2}\right]} \times 100$$

Relationship of the SNP effects on the exposure (HDL-c South Indian) against the SNP effects on the outcome(HDL-cScottish)



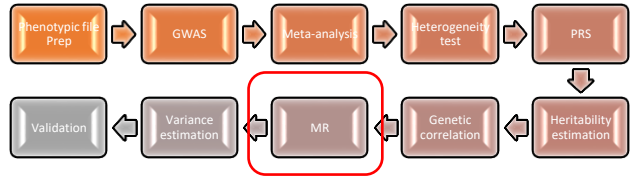




Test	Method	Q statistics	P value
Heterogeneity test	MR Egger	914.66	5.98e-37
	Inverse variance weighted	915.37	7.22e-37

Method	Egger intercept (se); P value
Pleiotropy test	-0.0003 (0.0005); 0.560

Test	Steiger P value
Directionality	0.01

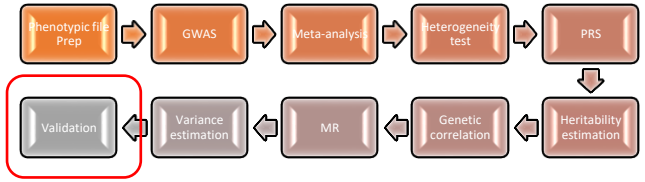


Outliers

SNP	rsid	GENE	Q_statistic	P value	SI EAF	SC EAF	SI Beta	SC Beta	Delta EAF %	Delta Beta %
15:58680178:C:T	**rs261291**	ALDH1A2	21.29650	3.93e-06	0.38	0.35	0.0134	0.035	-8	184
16:56991363:T:C	**rs183130**	CETP upstream	35.75394	2.24e-09	0.28	0.32	0.058	0.076	13	33
16:81534790:C:T	**rs2925979**	CMIP	16.49763	4.87e-05	0.76	0.69	0.004	0.024	-9	193

$$\text{Percentage Difference} = \frac{|\Delta V|}{\left[\frac{\Sigma V}{2}\right]} \times 100$$

$$= \frac{|V_1 - V_2|}{\left[\frac{(V_1 + V_2)}{2}\right]} \times 100$$



Cross-validation of findings

Analysis	Current	Re Analysis
PRS	PRSice, additive model	SumHer, using Priors
Heritability	SumHer, LDAK model	Popcorn, LDSC model
Genetic correlation	SumHer, LDAK model	Popcorn, LDSC model

Final Observations

1. HDL-c trait is highly heritable (~50%) in both the study population.
2. Heterogeneity of genetic variants exist among the study group.
3. Genetic correlation is nearly half between the study population.
4. PRS generated by Trans ethnic GLGC GWAS explains equal variance (~13-14 %) but causal estimate is 33% lower among south Indians compared with Scottish population.

Summary

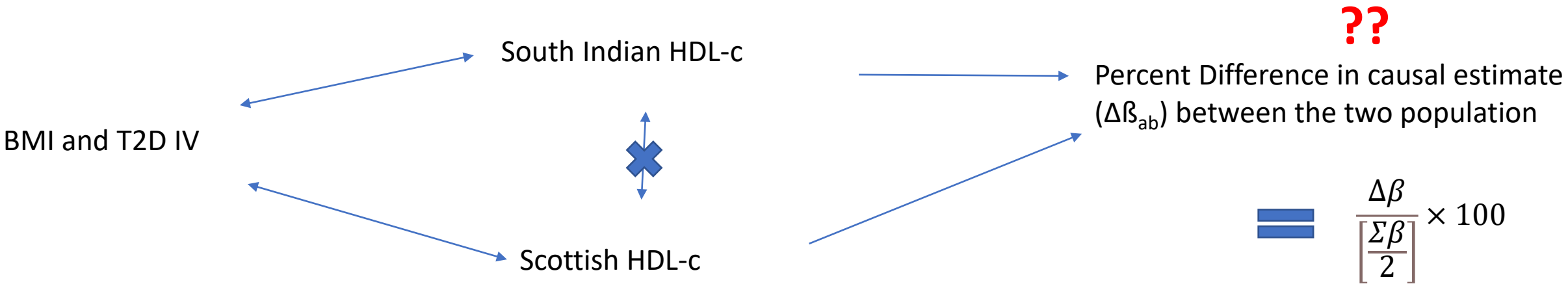
- Genetic loci for HDL-c in Scottish populations are present at both similar and varied frequency compared with South Indian populations, which modulate the HDL-c at different levels.
- These findings suggests that '**most likely**' lower HDL-c levels among the South Indian diabetic cohort compared with the Scottish diabetic population are genetically driven.

Limitation

- Different estimates may have observed with different statistical tool/Model used.
- There is chance of selection bias due to difference in health care system.
- Validation of findings are required.

Way Forward

- With Dr George's work on PheGWAS, suggest that BMI and T2D have different level of genetic correlation on HDL-c trait.
- In dendrogram tree, HDL-c is more closely associated with BMI among South Indian cohort compared with Scottish population



Acknowledgement

My supervisors

Dundee

- Prof Colin NA Palmer
- Prof Ewan R Pearson

MDRF

- Dr Guha Pradeepa
- Dr Radha Venkatesan

INSPIRED team

Dr Fred, Ms Isobel, Dr Simona, Dr Sundar, Dr Moneeza,

Anand, Aravind, Charvi, and all other INSPIRD colleagues

&

My family, my wife ‘Aradhana’ and my son ‘**Shrihaan**’

Thank you!

Any Questions?