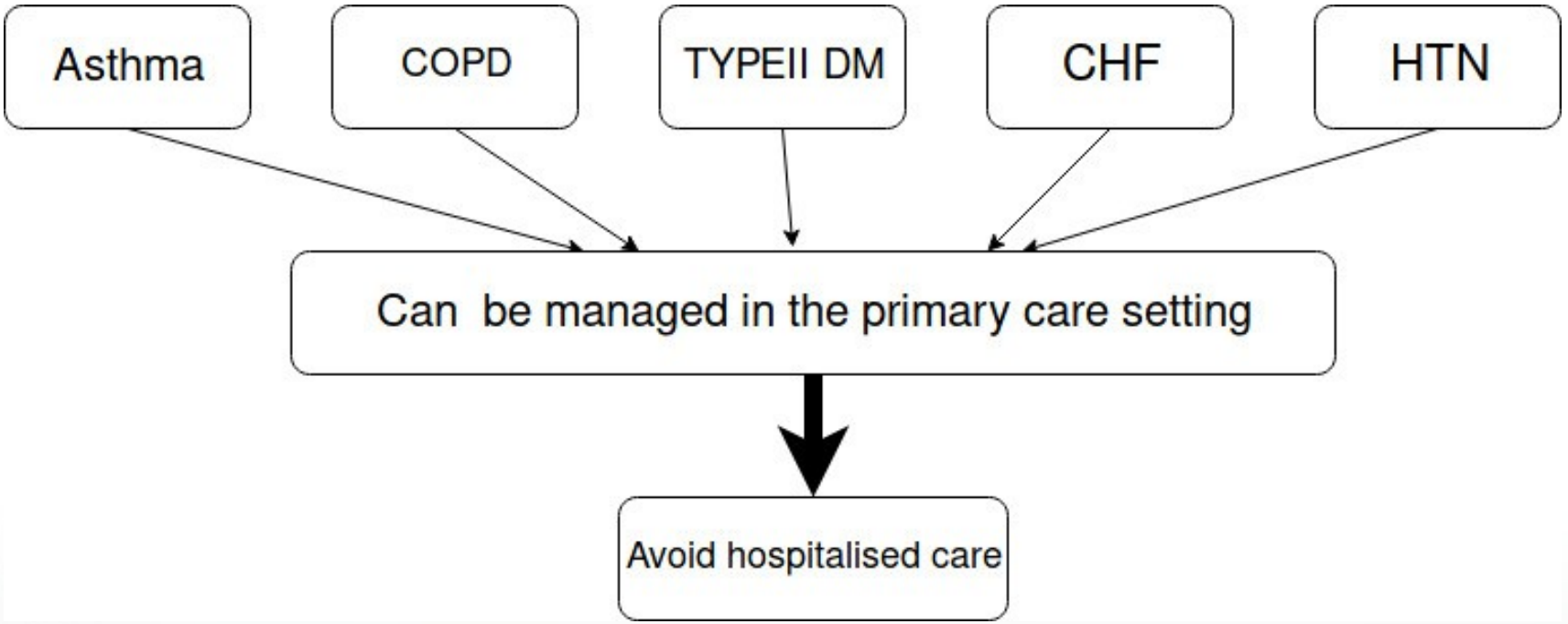
Estimation of a polygenic risk score for ambulatory care sensitive conditions

Arindam Basu
School of Health Sciences,
University of Canterbury
arindam.basu@canterbury.ac.nz

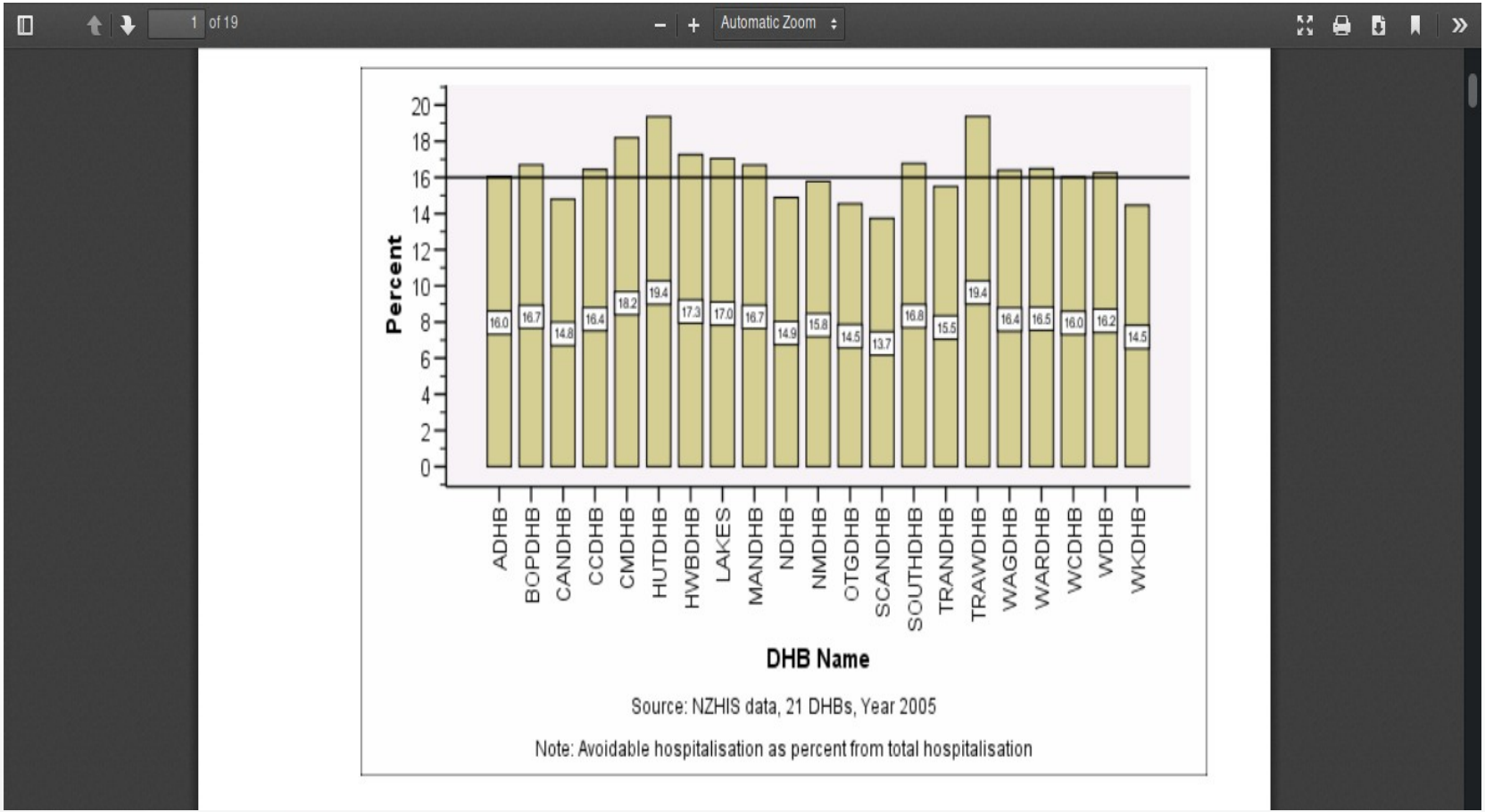
Outline of the prsentation

- Ambulatory sensitive hospitalisations
- Common genetic variants that can explain ASH
- What do GWAS studies tell us about ASH
- Estimation of PRS from GWAS studies
- Next steps

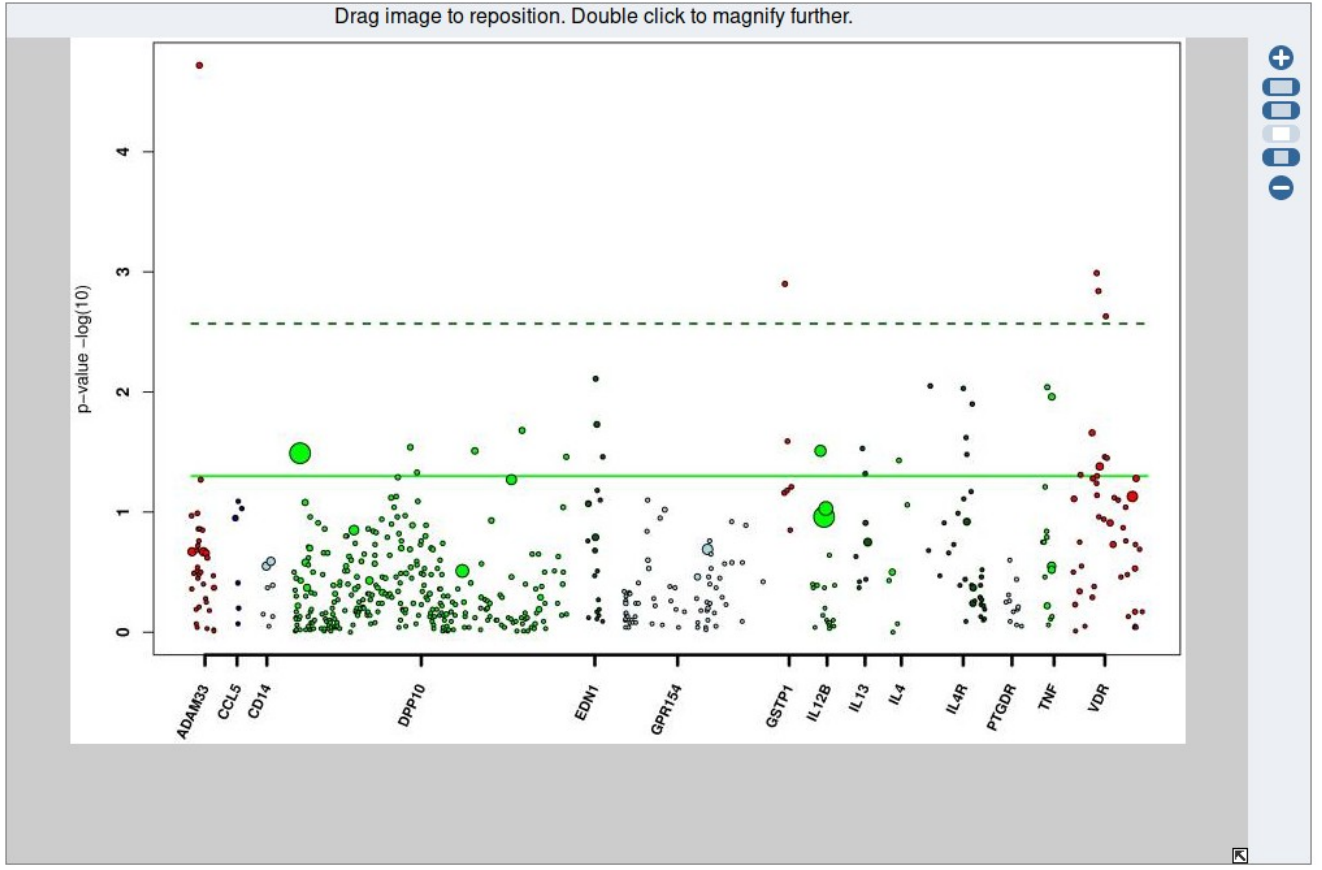
Ambulatory sensitive hospitalisation



Hospitalisation due to ASH in NZ



One sample of GWAS results due to Asthma



List of GWAS studies on Asthma

TABLE 2. ASTHMA GENETIC LOCI IDENTIFIED BY GENOME-WIDE ASSOCIATION STUDIES (GWAS)

Reference	Primary Cohort Size	Replication Sample Size	Gene/Region	Novel Gene/Pathway
Moffatt *et al.* (14)	994 subjects with asthma, 1,243 subjects without asthma	5,621 subjects	*17q21* (*ORMDL3*)*	Yes
Himes *et al.* (9)	359 subjects with asthma, 846 control subjects	18,891 subjects	*PDE4D*	No
Li *et al.* (11)	473 subjects with asthma, 1,892 control subjects		6p21 (*HLA-DR, HLA-DQ*), 5q31 (*IL13, RAD50*)	Yes (*RAD50*)
Sleiman *et al.* (16)	793 subjects with asthma, 1,988 control subjectss	917 subjects with asthma, 1,546 control subjects	17q21*, *DENND1B**	Yes (*DENND1B*)
Hancock *et al.* (17)	492 Mexican trios	177 Mexican trios	*TLE4*	*Yes*
Choudhry S *et al.* (18)	96 cases, 88 controls (Puerto Rican)	284 Puerto Rican trios	5q23.3	Yes

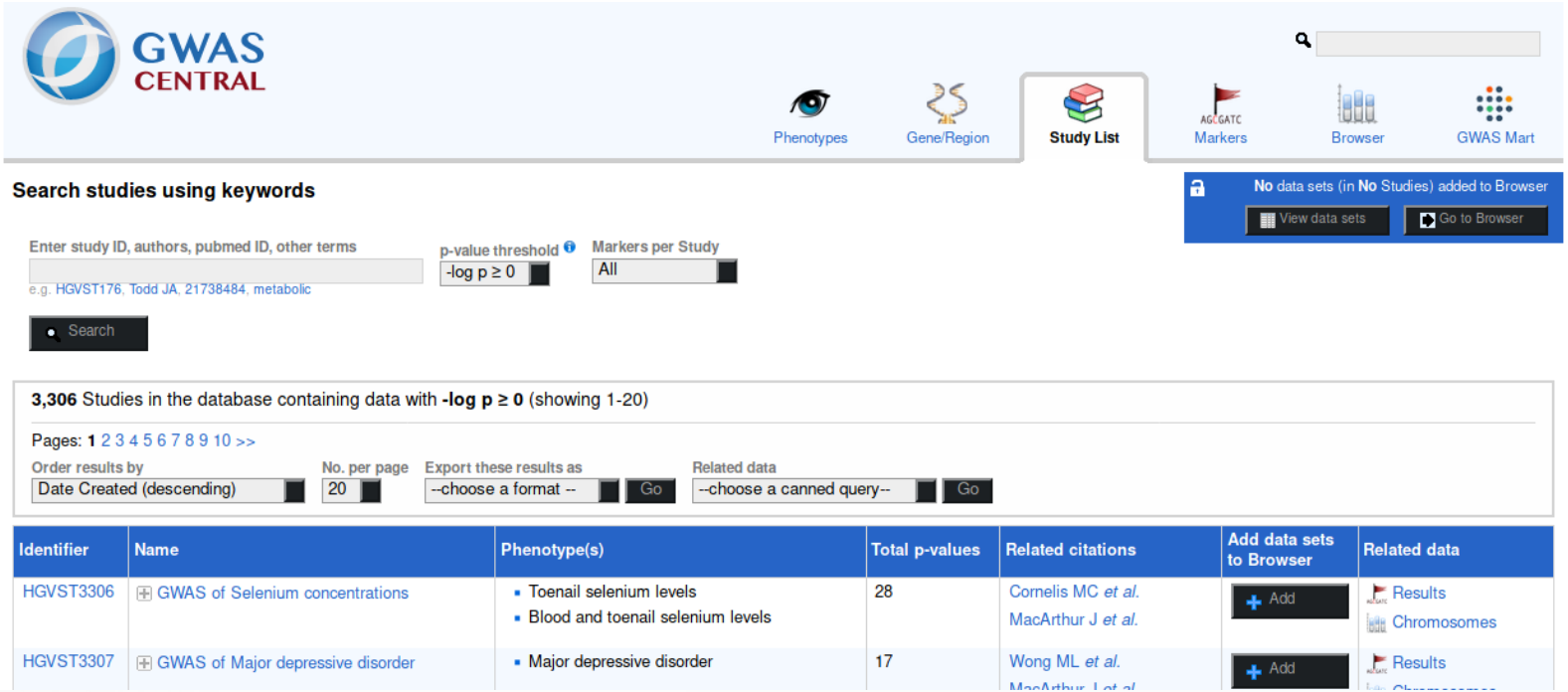
Large consortium meta-analysis of GWAS studies on Hypertension (partial list)

Table 1

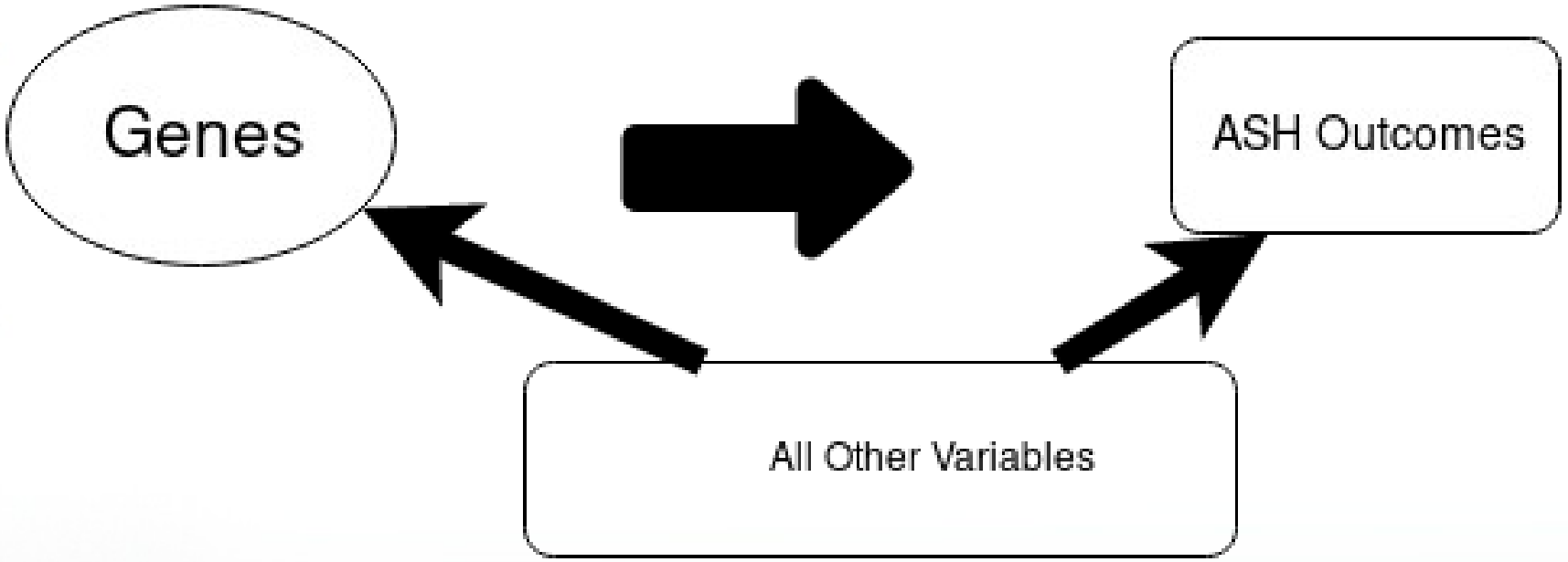
Genome-wide Association Results for Systolic Blood Pressure SNPs with P Value $<1\times10^{-6}$ Sorted by Systolic Blood Pressure Meta-analysis P Value

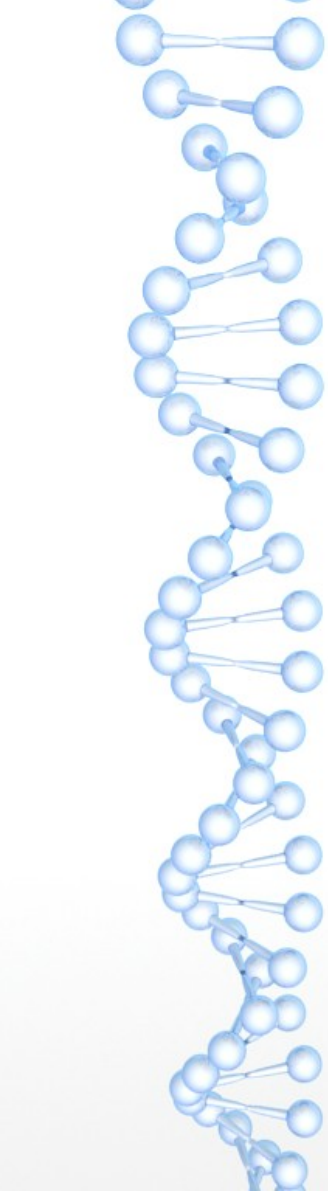
SNP identifier	Chr	Position	Gene	MAF	CHARGE Meta-analysis SBP			CHARGE Meta-analysis DBP			CHARGE Meta-analysis Hypertension		
					Beta	SE	p	Beta	SE	p	Beta	SE	p
rs2681492	12	88537220	*ATP2B1*	0.20	**-1.26**	**0.19**	**3.0E-11**	-0.62	0.11	4.6E-08	-0.14	0.03	8.4E-08
rs2681472	12	88533090	*ATP2B1*	0.18	**-1.29**	**0.19**	**3.5E-11**	-0.64	0.11	3.7E-08	-0.16	0.03	1.7E-08
rs11105354	12	88550654	*ATP2B1*	0.18	**-1.30**	**0.20**	**3.7E-11**	-0.63	0.11	5.8E-08	-0.16	0.03	1.8E-08
rs11105364	12	88593407		0.18	**-1.30**	**0.20**	**4.8E-11**	-0.63	0.12	1.2E-07	-0.16	0.03	2.1E-08
rs17249754	12	88584717		0.18	**-1.30**	**0.20**	**5.2E-11**	-0.63	0.12	1.0E-07	-0.16	0.03	2.2E-08
rs11105368	12	88598572		0.18	**-1.30**	**0.20**	**5.3E-11**	-0.63	0.12	1.3E-07	-0.16	0.03	2.2E-08
rs12579302	12	88574634		0.18	**-1.29**	**0.20**	**6.2E-11**	-0.62	0.12	1.3E-07	-0.16	0.03	2.2E-08
rs12230074	12	88614998		0.17	**-1.31**	**0.20**	**9.1E-11**	-0.62	0.12	3.4E-07	-0.17	0.03	2.9E-08
rs11105378	12	88614872		0.17	**-1.31**	**0.20**	**9.1E-11**	-0.62	0.12	3.1E-07	-0.17	0.03	2.8E-08
rs4842666	12	88465680		0.17	**-1.20**	**0.21**	**6.5E-09**	-0.62	0.12	4.5E-07	-0.15	0.03	3.4E-07
rs8096897	18	13428905	*C18orf1*	0.01	**-12.87**	**2.33**	**3.2E-08**	-4.07	1.33	2.9E-03	-0.73	0.35	0.04
rs11105328	12	88466521		0.18	**-1.11**	**0.20**	**4.2E-08**	-0.61	0.12	5.1E-07	-0.15	0.03	7.1E-07
rs880315	1	10719453	*CASZ1*	0.35	**0.89**	**0.17**	**2.1E-07**	0.30	0.10	2.9E-03	0.09	0.02	6.2E-05
rs3184504	12	110368991	*SH2B3*	0.48	**0.75**	**0.15**	**5.7E-07**	0.50	0.09	1.7E-08	0.07	0.02	7.4E-04
rs381815	11	16858844	*PLEKHA7*	0.26	**0.84**	**0.17**	**5.8E-07**	0.51	0.10	4.3E-07	0.09	0.02	1.7E-04

It is possible to obtain a list of studies on the five conditions from GWAS central



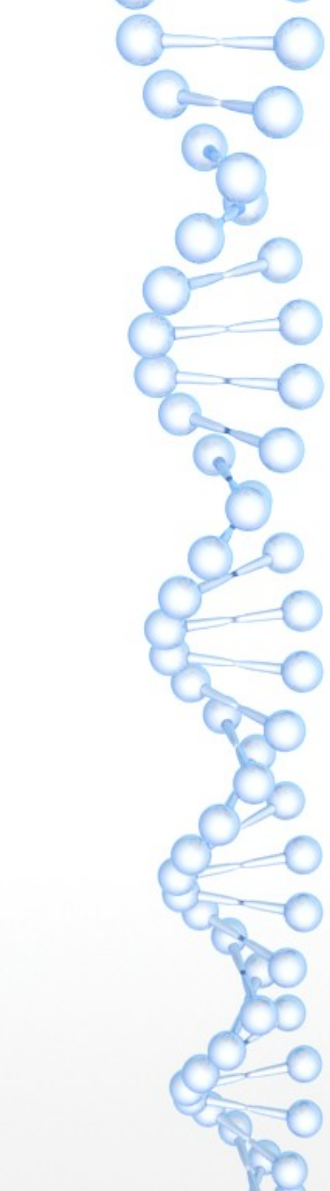
Study question





Polygenic Risk Score

- Single weighted summed score of SNPs from GWAS studies
- Weighted by their respective beta coefficients for continuous outcomes such as blood pressure scores or outcome scores
- (Alternative), weighted by their Odds Ratios for binary outcomes
- Singe PRS is then used in a regression model to predict or study association between genotypic contribution to the phenotype, as in
- Phenotype ~ PRS score + Other variables



Steps in our case

- Identify candidate SNPs through an exploratory meta-analysis
- Quality control of the genotype data
- The candidate SNPs form a base population
- Identify a target population for whom genotype and phenotype data are available (phenotype == “access to primary care” for ASH)
- Construct the PRS in the base population
- Apply to the target population, run models

Example meta-analysis script file from the metal meta-analysis helper page

```
# VERBOSE ON

# Describe and process the DGI input files
MARKER   SNP
WEIGHT   N
ALLELE   EFFECT_ALLELE NON_EFFECT_ALLELE
FREQ     EFFECT_ALLELE_FREQ
EFFECT   BETA
STDERR   SE
PVAL     P_VAL

PROCESS DGI_three_regions.txt

# Describe and process the FUSION input files
MARKER   SNP
ALLELE   EFFECT_ALLELE NON_EFFECT_ALLELE
FREQ     FREQ_EFFECT
WEIGHT   N
EFFECT   BETA
STDERR   SE
PVAL     PVALUE

PROCESS MAGIC_FUSION_Results.txt.gz

# Describe and process the SardiNIA input files
MARKER   SNP
DEFAULT  4106
ALLELE   AL1 AL2
FREQ     FREQ1
EFFECT   EFFECT
STDERR   SE
PVAL     PVALUE

PROCESS magic_SARDINIA.tbl

# Execute meta-analysis
ANALYZE
```

Exploratory meta-analysis output

Top 10 Meta-Analysis Results

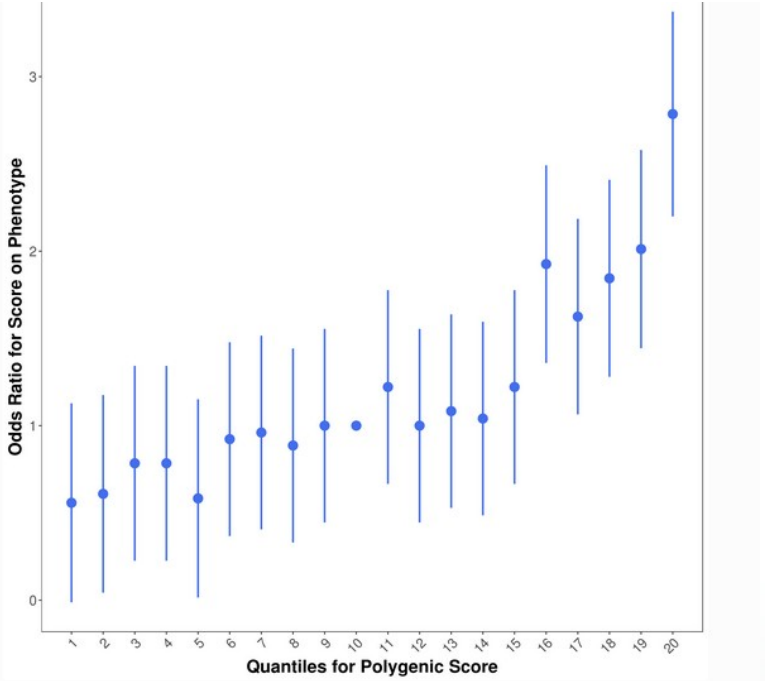
MarkerName	Allele1	Allele2	Weight	Zscore	P-value	Direction
rs560887	t	c	6806	-7.075	$1.491*10^{-12}$	---
rs853787	t	g	6806	6.691	$2.221*10^{-11}$	+++
rs853789	a	g	5339	-6.597	$4.189*10^{-11}$	?--
rs853773	a	g	6806	-6.132	$8.662*10^{-10}$	---
rs537183	t	c	6806	6.007	$1.887*10^{-9}$	+++
rs557462	t	c	6806	6.005	$1.917*10^{-9}$	+++
rs502570	a	g	6806	-6.001	$1.955*10^{-9}$	---
rs563694	a	c	6806	5.975	$2.300*10^{-9}$	+++
rs475612	t	c	6806	-5.867	$4.423*10^{-9}$	---
rs853781	a	g	6806	-5.844	$5.092*10^{-9}$	---

Using PRSice to compute Polygenic Risk Score

Rscript PRSice.R --dir . \

- --prsice ./PRSice \
- --base BASE_GWAS.assoc \
- --target TARGET_DATA \
- --thread 1 \
- --stat OR \
- --binary-target T
- (Script to run a PRS scoring algorithm based on GWAS)

Interpretation of the PRS output for the phenotype (here access to care)



Next steps for this project

- Quality control of the original GWAS data and then conduct a meta-analysis of GWAS studies on a set of defined population for the five conditions together
- Pool together the results of the five conditions and identify a set of candidate genes
- On that basis identify a target population (hardest hurdle to overcome)
- Construct the PRS and fit the PRS to the target population and identify the Odds Ratios

Outcomes and benefits

- A common set of variants for the common conditions will indicate a genetic component for access to care
- It'd be possible to study gene*gene and gene*environment interactions
- This study will extend the scope of genome wide association studies to preventive health
- This is an example of Precision public health as we can now cluster and quantify which population groups based on their genetic profile can benefit most from targeted preventive interventions