生物資訊與分析

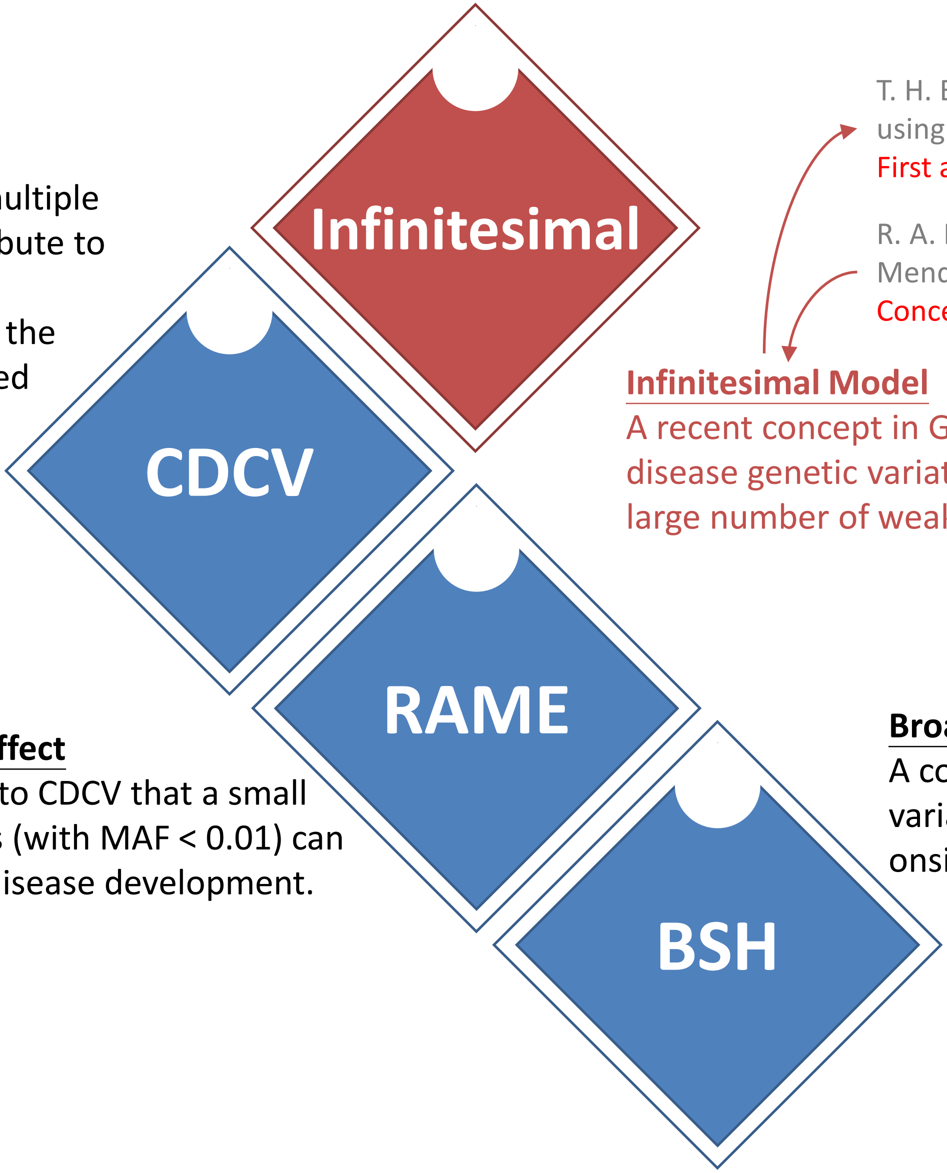
PRS

陳佳煒 (Jia-Wei Chen)

2024/04/16

Genetic architecture of complex diseases

Common Disease Common Variant
An early hypothesis in GWAS that multiple common variants collectively contribute to disease susceptibility.
However, CDCV cannot fully explain the missing heritability - the unaccounted genetic contribution to disease risk.



T. H. E. Meuwissen, B. J. Hayes, M. E. Goddard, Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819–1829 (2001)
First application of PRS

R. A. Fisher, The correlation between relatives on the supposition of Mendelian inheritance. *Trans. R. Soc. Edinb.* **52**, 399–433 (1918).
Concept of infinitesimal model

Infinitesimal Model
A recent concept in GWAS that complex disease genetic variation results from a large number of weak-effect variants

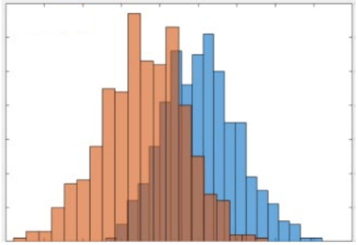
Rare Alleles of Major Effect
An alternative concept to CDCV that a small number of rare variants (with MAF < 0.01) can significantly influence disease development.

Broad Sense Heritability Model
A concept that neither common nor rare variants alone explain the missing heritability. It onsiders about GxG, GxE, and epigenetic effects

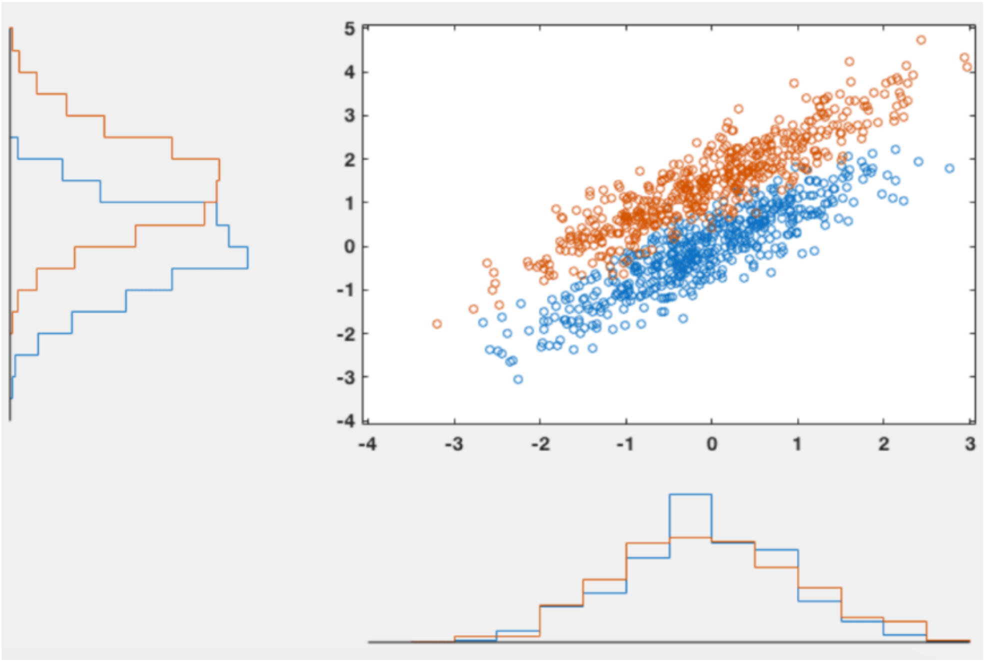
Association and prediction

Association
Group (population)-based concept focuses on statistical relationships between variables at the group level, which informs us about broader patterns and relationships within populations.

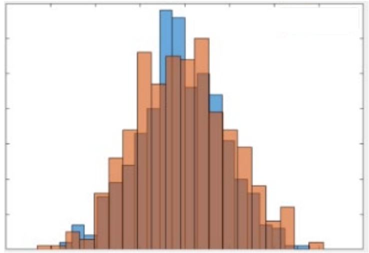
Prediction
Individual-based concept focus on personalized outcomes for specific individuals, and considers unique characteristics, medical history, and relevant factors



It’s an associated variable but is not predictive if it’s solely in a prediction model



It’s not an associated variable but improves the prediction performance when adding to the prediction model



https://robertoivega.com/association-prediction-studies/

	w_1	w_2	…	w_j	…	w_S		
	SNP 1	SNP 2	…	SNP *j*	…	SNP *S*	→	PRS
Ind 1								PRS_1
Ind 2								PRS_2
⋮								⋮
ind *i*				g_{ij}				PRS_i
⋮								⋮
ind *N*								PRS_N

$$PRS_i = \sum_{j=1}^{S} w_j \cdot g_{ij}$$

GWAS summary statistic

GWAS Catalog ⬇	https://www.ebi.ac.uk/gwas/
HuGeAMP ⬇	https://kp4cd.org/

Biobank GWAS summary statistic

China	http://www.mulinlab.org/pheweb/
Finnish ⬇	https://pheweb.sph.umich.edu/FinMetSeq/
Japan ⬇	https://pheweb.jp/ (hum0197.v18, hum0014.v32)
Korean ⬇	https://koges.leelabsg.org/
UK ⬇	https://github.com/Nealelab/UK_Biobank_GWAS
Taiwan	https://taiwanview.twbiobank.org.tw/pheweb.php

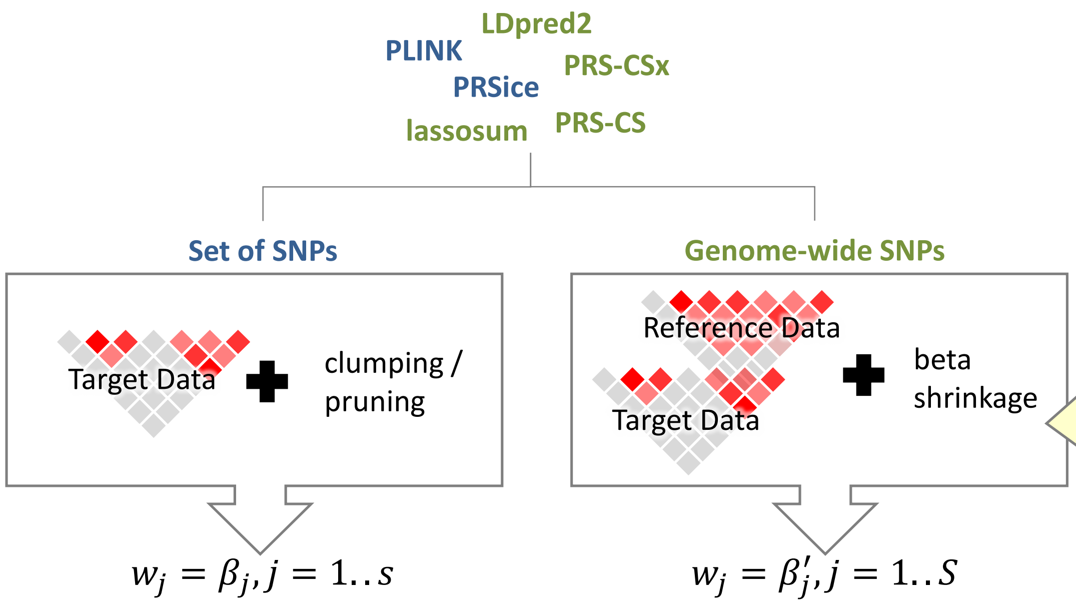
Base Data
GWAS summary statistic
(OR, beta, p-value)

Independent!

Target Data
Individual genotype data
(array, imputation)

SNP weights (PGS)

PGS Catalog ⬇	https://www.pgscatalog.org/
Cancer-PRSweb ⬇	https://prsweb.sph.umich.edu:8443/



$w_j = \beta_j, j = 1..s$

$w_j = \beta'_j, j = 1..S$

PRS, a composite measure derived from multiple genetic variants, can be simply treated as a linear combination of genotypes.

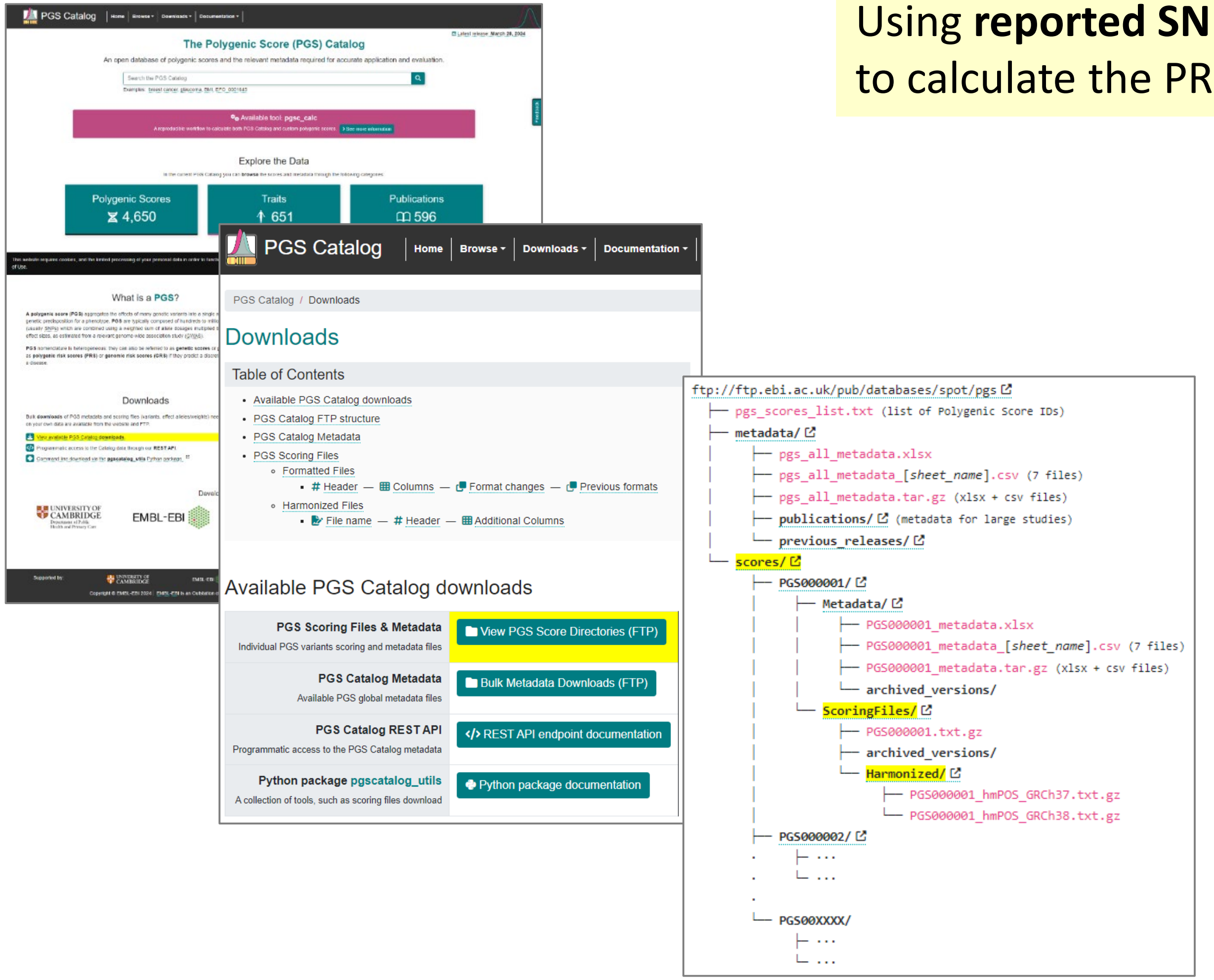
In regression modeling (linear combination), incorporating more predictors may enhance the R^2 (the proportion of variation explained by the model). However, it may cause overfitting due to the of model complexity, noise, or large betas (sensitive to minor changes).

Beta shrinkage (small beta) can help make beta toward zero that decreases model complexity and sensitivity to minor changes.

Method – External PRS

PGS Catalog

Using **reported SNP weights** (e.g., PGS catalog and Cancer-PRSweb) to calculate the PRS on the target data



PGS Catalog | Home | Browse ▾ | Downloads ▾ | Documentation ▾

Latest release: March 28, 2024

The Polygenic Score (PGS) Catalog

An open database of polygenic scores and the relevant metadata required for accurate application and evaluation.

Search the PGS Catalog

Examples: breast cancer, glaucoma, BMI, EFO_0001645

Available tool: pgsc_calc

A reproducible workflow to calculate both PGS Catalog and custom polygenic scores. > See more information

Explore the Data

In the current PGS Catalog you can **browse** the scores and metadata through the following categories:

Polygenic Scores ⧗ 4,650 | Traits 651 | Publications 596

This website requires cookies, and the limited processing of your personal data in order to functi
of Use.

What is a PGS?

A polygenic score (PGS) aggregates the effects of many genetic variants into a single n
genetic predisposition for a phenotype. **PGS** are typically composed of hundreds-to-millio
(usually SNPs) which are combined using a weighted sum of **allele dosages** multiplied b
effect sizes, as estimated from a relevant genome-wide association study (GWAS).

PGS nomenclature is heterogeneous: they can also be referred to as **genetic scores** or
as **polygenic risk scores (PRS)** or **genomic risk scores (GRS)** if they predict a discret
a disease.

Downloads

Bulk **downloads** of PGS metadata and scoring files (variants, effect alleles/weights) nee
on your own data are available from the website and FTP.

- View available PGS Catalog downloads
- Programmatic access to the Catalog data through our **REST API**
- Command line download via the **pgscatalog_utils** Python package

Develo

UNIVERSITY OF CAMBRIDGE Department of Public Health and Primary Care — EMBL-EBI

Supported by: UNIVERSITY OF CAMBRIDGE — EMBL-EBI

Copyright © EMBL-EBI 2024 | EMBL-EBI is an Outstation o

PGS Catalog | Home | Browse ▾ | Downloads ▾ | Documentation ▾

PGS Catalog / Downloads

Downloads

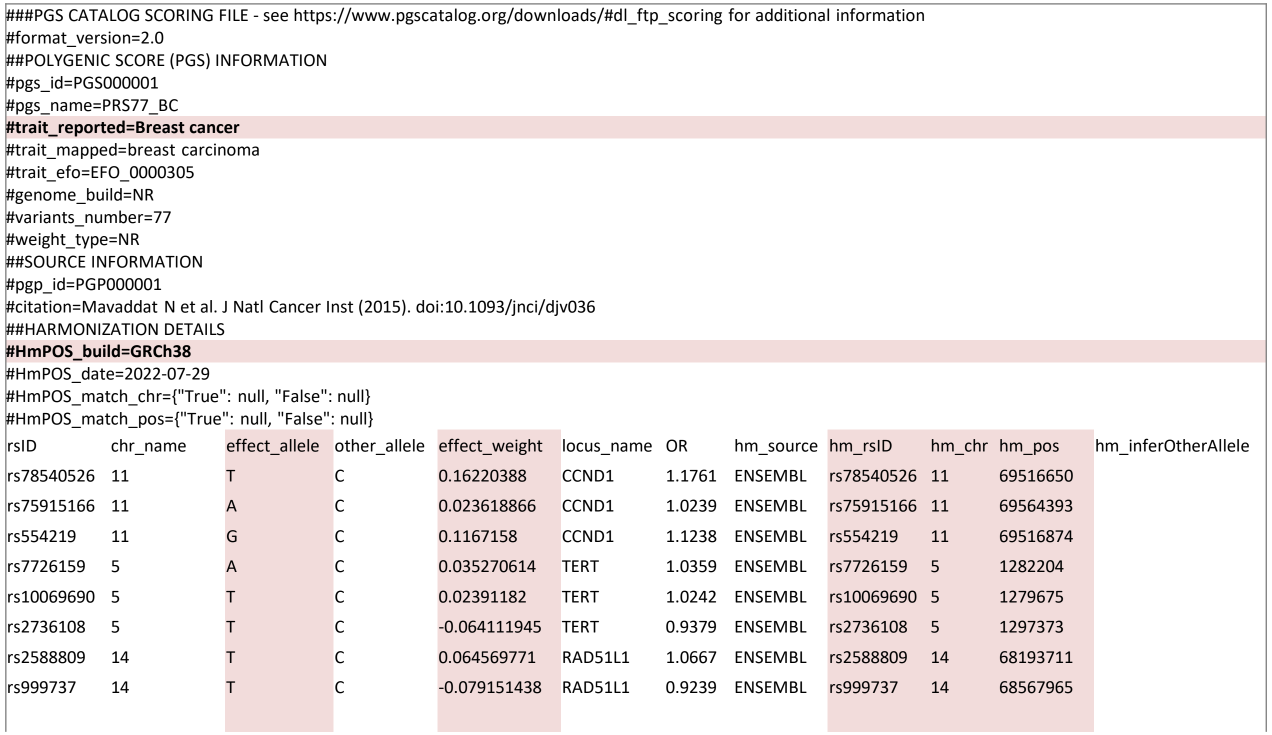
Table of Contents

- Available PGS Catalog downloads
- PGS Catalog FTP structure
- PGS Catalog Metadata
- PGS Scoring Files
  - Formatted Files
    - # Header — ▦ Columns — Format changes — Previous formats
  - Harmonized Files
    - File name — # Header — ▦ Additional Columns

Available PGS Catalog downloads

PGS Scoring Files & Metadata Individual PGS variants scoring and metadata files	View PGS Score Directories (FTP)
PGS Catalog Metadata Available PGS global metadata files	Bulk Metadata Downloads (FTP)
PGS Catalog REST API Programmatic access to the PGS Catalog metadata	</> REST API endpoint documentation
Python package pgscatalog_utils A collection of tools, such as scoring files download	Python package documentation

```
ftp://ftp.ebi.ac.uk/pub/databases/spot/pgs
  ├── pgs_scores_list.txt (list of Polygenic Score IDs)
  ├── metadata/
  |     ├── pgs_all_metadata.xlsx
  |     ├── pgs_all_metadata_[sheet_name].csv (7 files)
  |     ├── pgs_all_metadata.tar.gz (xlsx + csv files)
  |     ├── publications/ (metadata for large studies)
  |     └── previous_releases/
  └── scores/
        ├── PGS000001/
        |     ├── Metadata/
        |     |     ├── PGS000001_metadata.xlsx
        |     |     ├── PGS000001_metadata_[sheet_name].csv (7 files)
        |     |     ├── PGS000001_metadata.tar.gz (xlsx + csv files)
        |     |     └── archived_versions/
        |     └── ScoringFiles/
        |           ├── PGS000001.txt.gz
        |           ├── archived_versions/
        |           └── Harmonized/
        |                 ├── PGS000001_hmPOS_GRCh37.txt.gz
        |                 └── PGS000001_hmPOS_GRCh38.txt.gz
        ├── PGS000002/
        .     ├ ...
        .     └ ...
        .
        └── PGS00XXXX/
              ├ ...
              └ ...
```



```
###PGS CATALOG SCORING FILE - see https://www.pgscatalog.org/downloads/#dl_ftp_scoring for additional information
#format_version=2.0
##POLYGENIC SCORE (PGS) INFORMATION
#pgs_id=PGS000001
#pgs_name=PRS77_BC
#trait_reported=Breast cancer
#trait_mapped=breast carcinoma
#trait_efo=EFO_0000305
#genome_build=NR
#variants_number=77
#weight_type=NR
##SOURCE INFORMATION
#pgp_id=PGP000001
#citation=Mavaddat N et al. J Natl Cancer Inst (2015). doi:10.1093/jnci/djv036
##HARMONIZATION DETAILS
#HmPOS_build=GRCh38
#HmPOS_date=2022-07-29
#HmPOS_match_chr={"True": null, "False": null}
#HmPOS_match_pos={"True": null, "False": null}
```

rsID	chr_name	effect_allele	other_allele	effect_weight	locus_name	OR	hm_source	hm_rsID	hm_chr	hm_pos	hm_inferOtherAllele
rs78540526	11	T	C	0.16220388	CCND1	1.1761	ENSEMBL	rs78540526	11	69516650	
rs75915166	11	A	C	0.023618866	CCND1	1.0239	ENSEMBL	rs75915166	11	69564393	
rs554219	11	G	C	0.1167158	CCND1	1.1238	ENSEMBL	rs554219	11	69516874	
rs7726159	5	A	C	0.035270614	TERT	1.0359	ENSEMBL	rs7726159	5	1282204	
rs10069690	5	T	C	0.02391182	TERT	1.0242	ENSEMBL	rs10069690	5	1279675	
rs2736108	5	T	C	-0.064111945	TERT	0.9379	ENSEMBL	rs2736108	5	1297373	
rs2588809	14	T	C	0.064569771	RAD51L1	1.0667	ENSEMBL	rs2588809	14	68193711	
rs999737	14	T	C	-0.079151438	RAD51L1	0.9239	ENSEMBL	rs999737	14	68567965	

Method – External PRS

```
$  plink --bfile dat_auto_qc \
   --score PGS000001_hmPOS_GRCh38.txt 9 3 5 \
   --out dat_auto_qc

$  plink --bfile dat_auto_qc \
   --score PGS000001_hmPOS_GRCh38.txt 9 3 5 sum \
   --out dat_auto_qc
```

PGS000001_hmPOS_GRCh38.txt

```
###PGS CATALOG SCORING FILE - see https://www.pgscatalog.org/downloads/#dl_ftp_scoring for additional information
#format_version=2.0
##POLYGENIC SCORE (PGS) INFORMATION
#pgs_id=PGS000001
#pgs_name=PRS77_BC
#trait_reported=Breast cancer
#trait_mapped=breast carcinoma
#trait_efo=EFO_0000305
#genome_build=NR
#variants_number=77
#weight_type=NR
##SOURCE INFORMATION
#pgp_id=PGP000001
#citation=Mavaddat N et al. J Natl Cancer Inst (2015). doi:10.1093/jnci/djv036
##HARMONIZATION DETAILS
#HmPOS_build=GRCh38
#HmPOS_date=2022-07-29
#HmPOS_match_chr={"True": null, "False": null}
#HmPOS_match_pos={"True": null, "False": null}
```

rsID	chr_name	effect_allele	other_allele	effect_weight	locus_name	OR	hm_source	hm_rsID	hm_chr	hm_pos	hm_inferOtherAllele
rs78540526	11	T	C	0.16220388	CCND1	1.1761	ENSEMBL	rs78540526	11	69516650	
rs75915166	11	A	C	0.023618866	CCND1	1.0239	ENSEMBL	rs75915166	11	69564393	
rs554219	11	G	C	0.1167158	CCND1	1.1238	ENSEMBL	rs554219	11	69516874	
rs7726159	5	A	C	0.035270614	TERT	1.0359	ENSEMBL	rs7726159	5	1282204	
rs10069690	5	T	C	0.02391182	TERT	1.0242	ENSEMBL	rs10069690	5	1279675	
rs2736108	5	T	C	-0.064111945	TERT	0.9379	ENSEMBL	rs2736108	5	1297373	
rs2588809	14	T	C	0.064569771	RAD51L1	1.0667	ENSEMBL	rs2588809	14	68193711	
rs999737	14	T	C	-0.079151438	RAD51L1	0.9239	ENSEMBL	rs999737	14	68567965	
		3		5				9			

dat_auto_qc.**profile**

FID	IID	PHENO	CNT	CNT2	SCORE

FID	IID	PHENO	CNT	CNT2	SCORSUM

$$PRS_i = \sum_{j=1}^{S} \frac{w_j \cdot g_{ij}}{2 \cdot N_i}$$

N_i = non-missing SNPs in sample i

$$PRS_i = \sum_{j=1}^{S} w_j \cdot g_{ij}$$

Preparation

- **Base data**

If possible, try to have the following information

#id	**ID**	SNP ID, **same representation** as in target data
#ch	CHR	chromosome, **same genome build** as in target data
#bp	BP	physical position, **same genome build** as in target data
#a1	**A1**	effect allele
	A2	other alleles
#b	**OR/BETA**	estimates
#s	SE	standard error of BETA
#p	**P**	p-value
#n	**N**	sample size

- **Target data**
  - plink-formatted files (.bed, .bim, .fam)
  - phenotype and covariate files

- **LD information**
  - **LDpred2**

    HapMap3 LD blocks and LD matrix
  - **lassosum**

    1000 genomes project Phase I LD blocks

    (automatically download when installing lassosum)
  - **PRS-CS/PRS-CSx**

    1000 genomes project LD

    UK Biobank LD

Preparation

Target data plink-formatted files (.bed, .bim, .fam)

Extended map file
(dat_auto_qc.bim**)**

Major allele	G	A	G	
Minor allele	T	G	C	
Physical position	565433	752566	753541	
Genetic distance	0	0	0	
Marker ID	marker1	marker2	marker3	
Chr	1	1	1	…

Family ID	Individual ID	Paternal ID	Maternal ID	Sex	Phenotype
FAM001	ind1	0	0	1	2
FAM001	ind2	0	0	1	2
FAM001	ind3	0	0	2	1
⋮					

Family information file
(dat_auto_qc.fam**)**

marker1	marker2	marker3	
11	11	10	…
11	10	01	
10	10	11	
⋮			

Binary-coded genotype file
(dat_auto_qc.bed**)**

Target data phenotype and covariate files

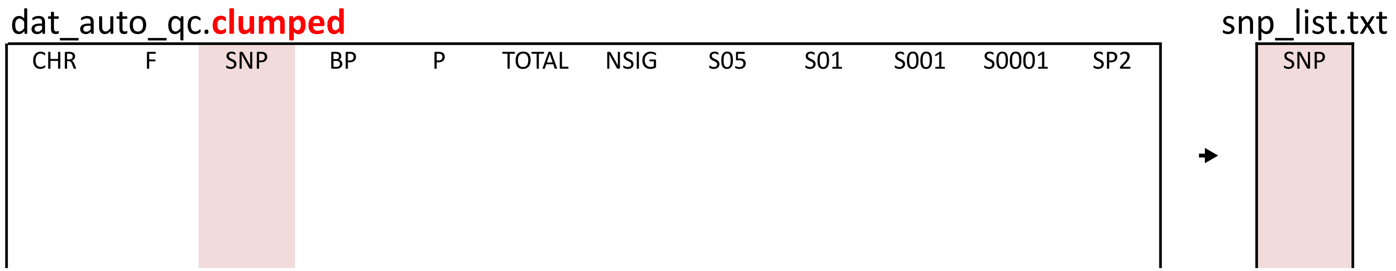
pheCov.txt

		phenotypes				covariates		subpopulation structure		
FID	IID	bt_1	bt_2	qt_1	qt_2	age	sex	pc1	pc2	…
FAM001	ind1									
FAM001	Ind2									
FAM001	ind3									
⋮										

Method – C+T (Clumping + Thresholding)

```
$  plink --bfile dat_auto_qc \
   --clump sumStat.txt \
   --clump-snp-field ID \
   --clump-field P \
   --clump-p1 1 --clump-p2 1 --clump-r2 0.2 --clump-kb 500 \
   --out dat_auto_qc


$  awk '{print $3}' dat_auto_qc.clumped > snp_list.txt
```



dat_auto_qc.**clumped**

CHR	F	SNP	BP	P	TOTAL	NSIG	S05	S01	S001	S0001	SP2

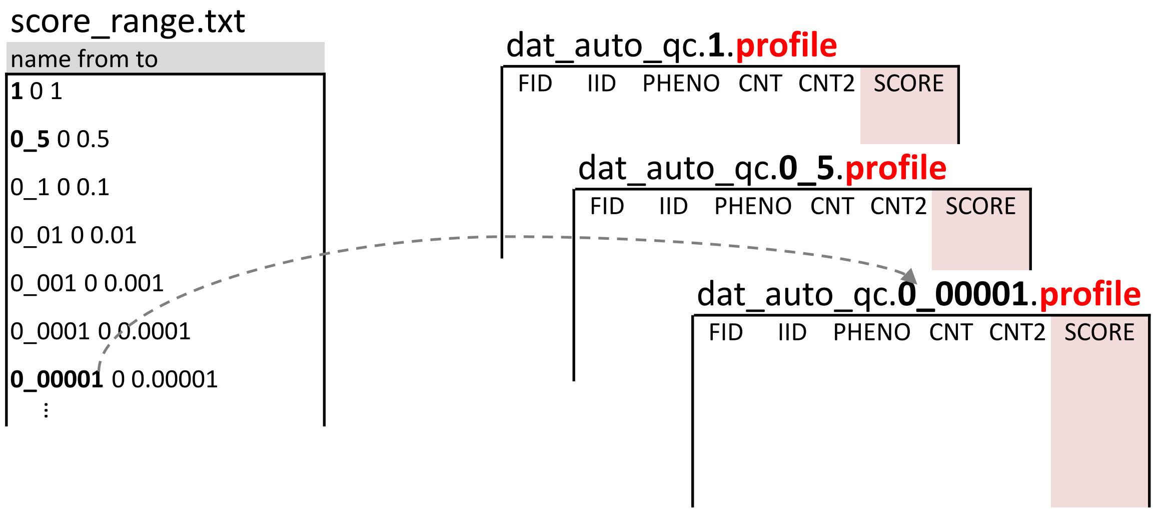
snp_list.txt

SNP

Each row is a clump of markers
indexed by the 'SNP' column (smallest p-value)

Method – C+T (Clumping + Thresholding)

```
$  plink --bfile dat_auto_qc \
   --extract snp_list.txt \
   --score sumStat.txt #id #a1 #b \
   --q-score-range score_range.txt sumStat.txt #id #p \
   --out dat_auto_qc
```



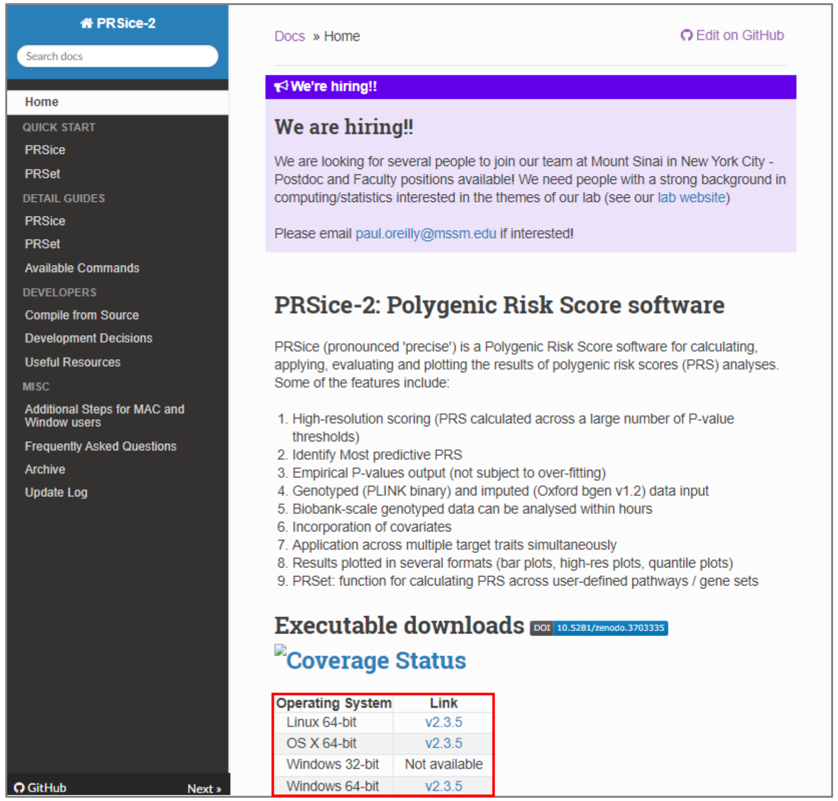
score_range.txt

name from to
1 0 1
0_5 0 0.5
0_1 0 0.1
0_01 0 0.01
0_001 0 0.001
0_0001 0 0.0001
0_00001 0 0.00001
⋮

dat_auto_qc.**1**.**profile**
FID IID PHENO CNT CNT2 SCORE
dat_auto_qc.**0_5**.**profile**
FID IID PHENO CNT CNT2 SCORE
dat_auto_qc.**0_00001**.**profile**
FID IID PHENO CNT CNT2 SCORE

Method – PRSice2

- https://choishingwan.github.io/PRSice/



We're hiring!!

We are hiring!!

We are looking for several people to join our team at Mount Sinai in New York City - Postdoc and Faculty positions available! We need people with a strong background in computing/statistics interested in the themes of our lab (see our lab website)

Please email paul.oreilly@mssm.edu if interested!

PRSice-2: Polygenic Risk Score software

PRSice (pronounced 'precise') is a Polygenic Risk Score software for calculating, applying, evaluating and plotting the results of polygenic risk scores (PRS) analyses. Some of the features include:

1. High-resolution scoring (PRS calculated across a large number of P-value thresholds)
2. Identify Most predictive PRS
3. Empirical P-values output (not subject to over-fitting)
4. Genotyped (PLINK binary) and imputed (Oxford bgen v1.2) data input
5. Biobank-scale genotyped data can be analysed within hours
6. Incorporation of covariates
7. Application across multiple target traits simultaneously
8. Results plotted in several formats (bar plots, high-res plots, quantile plots)
9. PRSet: function for calculating PRS across user-defined pathways / gene sets

Executable downloads DOI 10.5281/zenodo.3703335

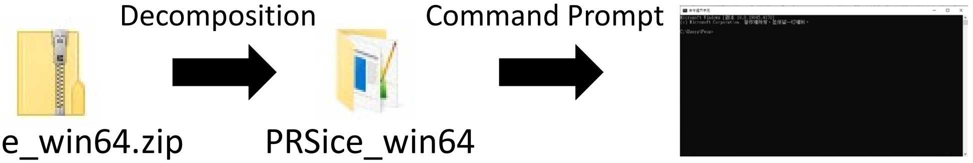
Coverage Status

Operating System	Link
Linux 64-bit	v2.3.5
OS X 64-bit	v2.3.5
Windows 32-bit	Not available
Windows 64-bit	v2.3.5

- **Linux**

```
$  wget https://github.com/choishingwan/PRSice/releases/download/2.3.5/PRSice_linux.zip
$  unzip PRSice_linux.zip -d PRSice
$  cd PRSice
$  Rscript PRSice.R --prsice PRSice_linux
```

- **Windows**



Decomposition

Command Prompt

PRSice_win64.zip

PRSice_win64

```
>  cd PRSice_win64
>  Rscript PRSice.R --prsice PRSice_win64.exe
```

Method – PRSice2

```
$  Rscript PRSice.R \
   --prsice PRSice_linux \
   --target dat_auto_qc \
   --base sumStat.txt \
   --binary-targe T \
   --snp ID \
   --A1 A1 \
   --stat BETA \
   --pvalue P \
   --beta \
   --pheno pheCov.txt \
   --pheno-col bt_1 \
   --cov pheCov.txt \
   --cov-col age,sex,@pc[1-10] \
   --out dat_auto_qc
```

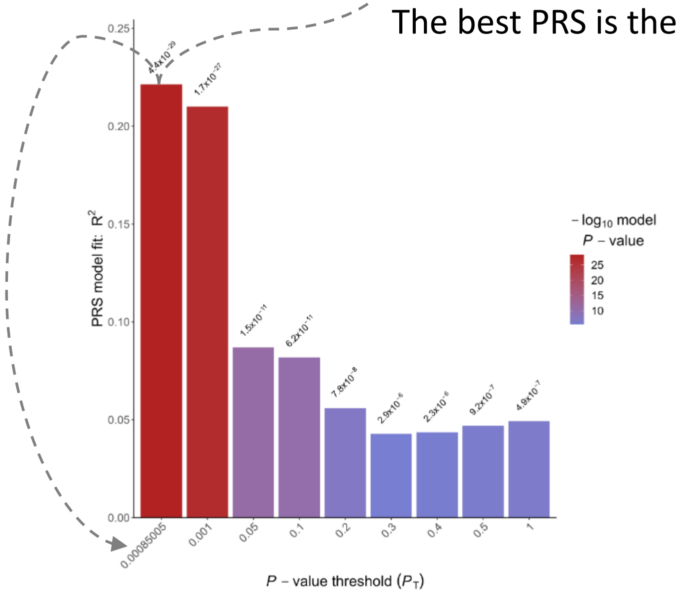
dat_auto_qc.**summary**

Phenotype	Set	Threshold	PRS.R2	Full.R2	Null.R2	Prevalence	Coefficient	Standard.Error	P	Num_SNP

dat_auto_qc.**best**

FID	IID	In_Regression	PRS

The best PRS is the one exhibits the highest R^2 with the phenotype



Method – lassosum

```
>  install.packages(c("devtools","RcppArmadillo", "data.table", "Matrix"), dependencies=TRUE)
>  devtools::install_github("tshmak/lassosum")

>  library(lassosum); library(data.table)
>  phecov <- fread("pheCov.txt")
>  ss <- fread("sumStat.txt")
>  cor <- p2cor(p = ss$P, n = N, sign = log(ss$OR)) # sign = ss$BETA
>  out <- lassosum.pipeline(cor = cor, chr = ss$CHR, pos = ss$BP, A1 = ss$A1, A2 = ss$A2, ref.bfile = "dat_auto_qc", test.bfile =
   "dat_auto_qc", LDblocks = "ASN.hg19")
```

Method – lassosum

```
> result <- validate(out, pheno = phecov$qt_1, covar = phecov[, c("age","sex",paste0("pc",1:10))])
> ss_ <- data.table(ss[out$sumstats$order][,sbeta:= result$best.beta]
```

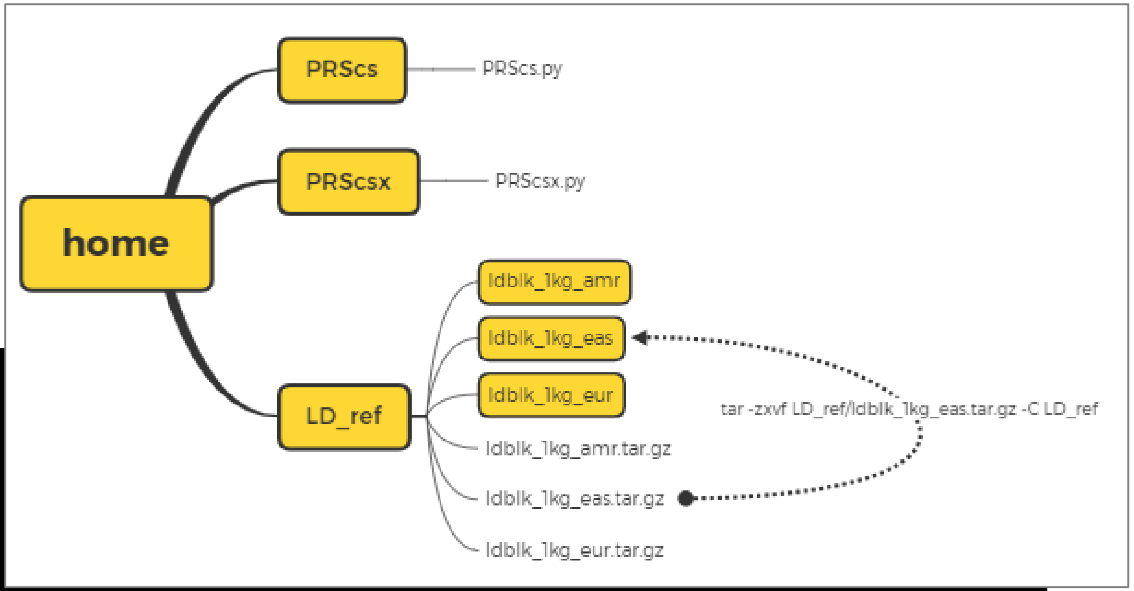
ID	CHR	BP	A1	A2	OR	SE	P	N	sbeta

shrinkage beta

```
> result$results.table
```

FID	IID	pheno	best.prs

Method – PRS-CS / PRS-CSx



```
tar -zxvf LD_ref/ldblk_1kg_eas.tar.gz -C LD_ref
```

```
$  git clone https://github.com/getian107/PRScs.git
$  git clone https://github.com/getian107/PRScsx.git

$  mkdir LD_ref
$  wget -O LD_ref/ldblk_1kg_amr.tar.gz https://www.dropbox.com/s/uv5ydr4uv528lca/ldblk_1kg_amr.tar.gz?dl=0
$  wget -O LD_ref/ldblk_1kg_eas.tar.gz https://www.dropbox.com/s/7ek4lwwf2b7f749/ldblk_1kg_eas.tar.gz?dl=0
$  wget -O LD_ref/ldblk_1kg_eur.tar.gz https://www.dropbox.com/s/mt6var0z96vb6fv/ldblk_1kg_eur.tar.gz?dl=0

$  tar -zxvf LD_ref/ldblk_1kg_amr.tar.gz -C LD_ref
$  tar -zxvf LD_ref/ldblk_1kg_eas.tar.gz -C LD_ref
$  tar -zxvf LD_ref/ldblk_1kg_eur.tar.gz -C LD_ref
```

Method – PRS-CS / PRS-CSx

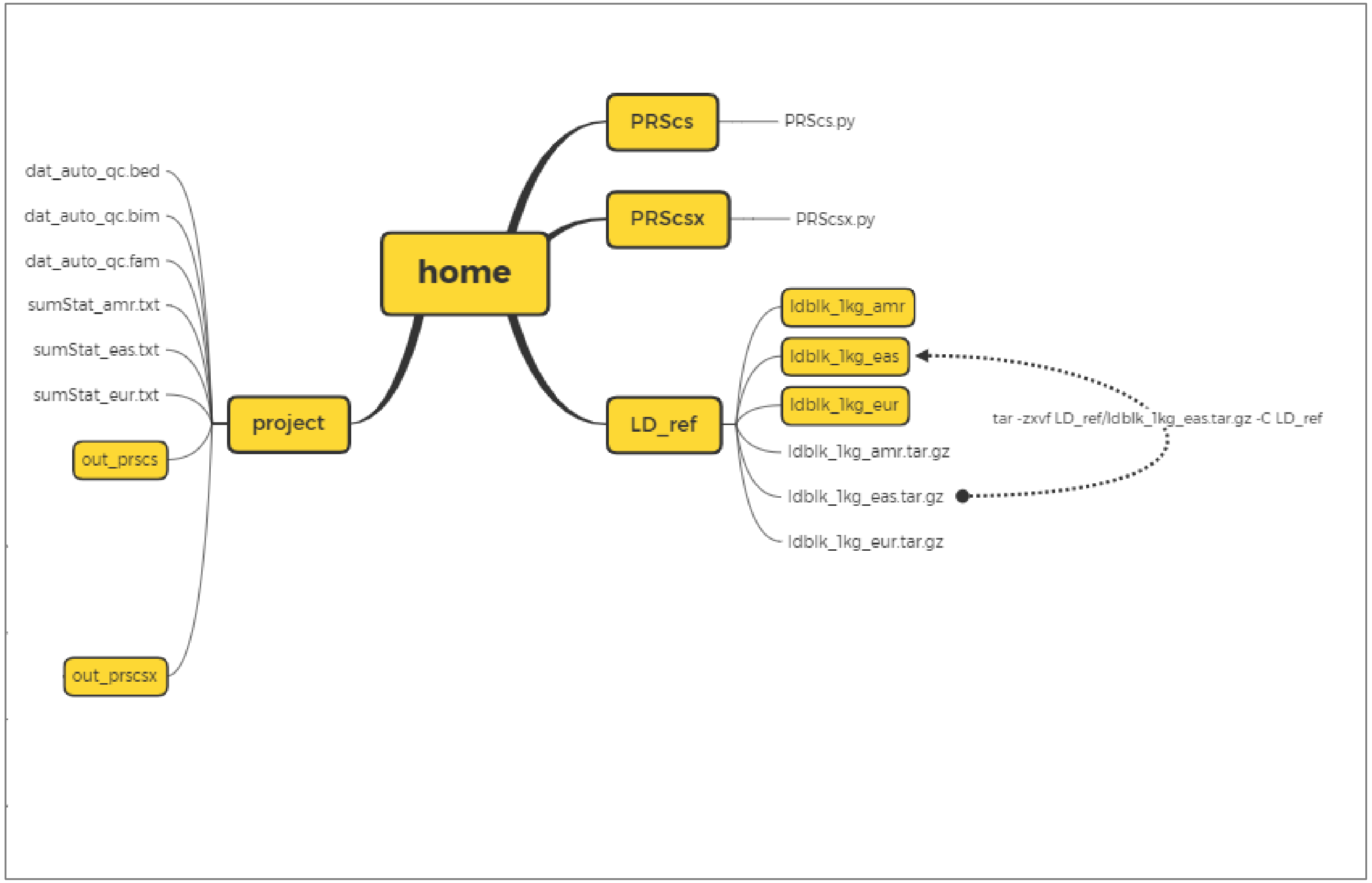
sumStat_*.txt recommended

SNP	A1	A2	BETA/OR	SE

sumStat_*.txt

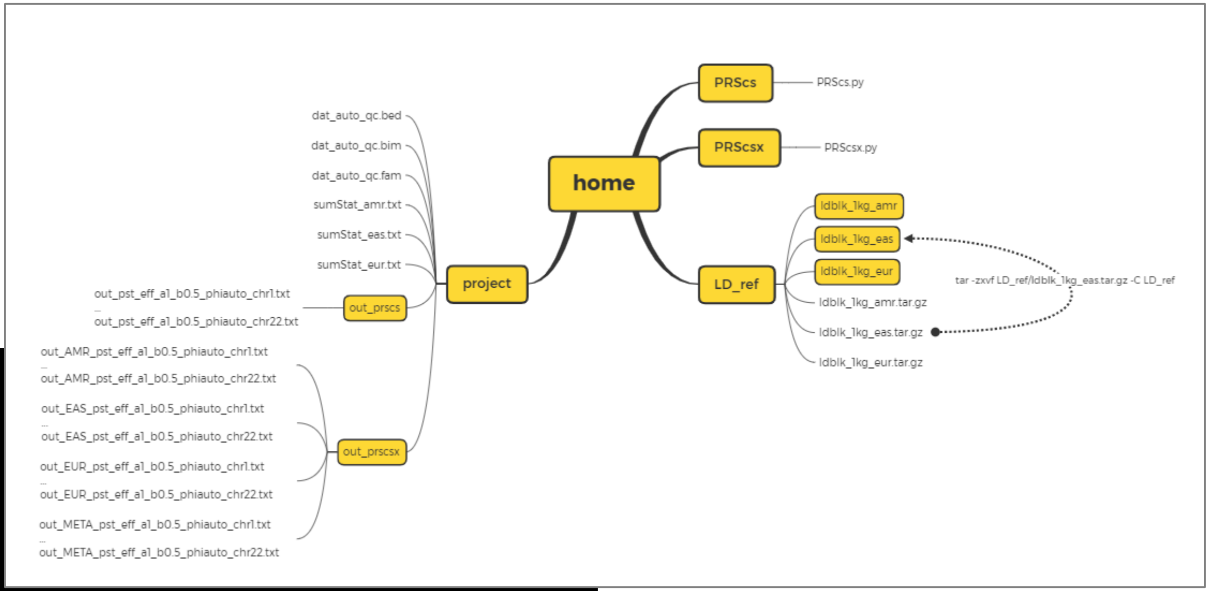
SNP	A1	A2	BETA/OR	P

SNP column is rsID, because the representation for SNP in provided LD information is rsID



```
tar -zxvf LD_ref/ldblk_1kg_eas.tar.gz -C LD_ref
```

Method – PRS-CS / PRS-CSx


home
PRScs
PRScs.py
PRScsx
PRScsx.py
LD_ref
ldblk_1kg_amr
ldblk_1kg_eas
ldblk_1kg_eur
ldblk_1kg_amr.tar.gz
ldblk_1kg_eas.tar.gz
ldblk_1kg_eur.tar.gz
project
dat_auto_qc.bed
dat_auto_qc.bim
dat_auto_qc.fam
sumStat_amr.txt
sumStat_eas.txt
sumStat_eur.txt
out_prscs
out_pst_eff_a1_b0.5_phiauto_chr1.txt
...
out_pst_eff_a1_b0.5_phiauto_chr22.txt
out_prscsx
out_AMR_pst_eff_a1_b0.5_phiauto_chr1.txt
...
out_AMR_pst_eff_a1_b0.5_phiauto_chr22.txt
out_EAS_pst_eff_a1_b0.5_phiauto_chr1.txt
...
out_EAS_pst_eff_a1_b0.5_phiauto_chr22.txt
out_EUR_pst_eff_a1_b0.5_phiauto_chr1.txt
...
out_EUR_pst_eff_a1_b0.5_phiauto_chr22.txt
out_META_pst_eff_a1_b0.5_phiauto_chr1.txt
...
out_META_pst_eff_a1_b0.5_phiauto_chr22.txt

```
tar -zxvf LD_ref/ldblk_1kg_eas.tar.gz -C LD_ref
```

```
$  cd PRScs
$  python3 PRScs.py  --ref_dir=../LD_ref/ldblk_1kg_eas \
   --bim_prefix=../project/dat_auto_qc \
   --sst_file=../project/sumStat_eas.txt \
   --n_gwas=N --seed=1 --out_dir=../project/out_prscs/out

$  mkdir dis_prscsx
$  cd PRScsx
$  python3 PRScsx.py --ref_dir=../LD_ref \
   --bim_prefix=../project/dat_auto_qc \
   --sst_file=../project/sumStat_amr.txt,../project/sumStat_eas.txt,../project/sumStat_eur.txt \
   --n_gwas=N_amr,N_eas,N_eur --pop=AMR,EAS,EUR \
   --seed=1 --meta=TRUE --out_dir=../project/out_prscsx --out_name=out
```

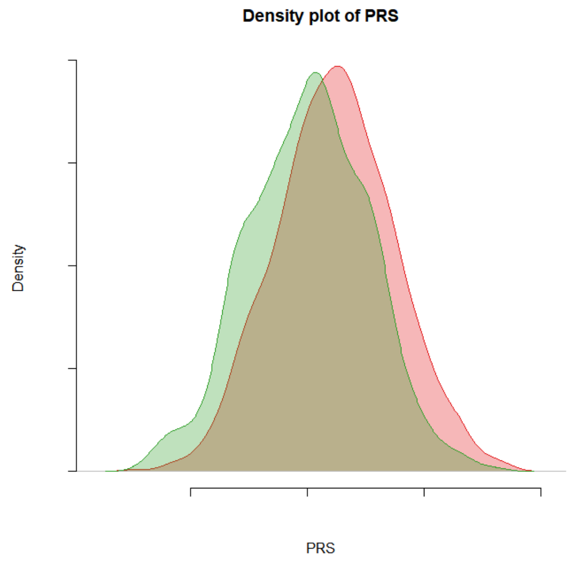
dat_auto_qc_pst_eff_a[1]_b[0.5]_phiauto_chr*.txt

CHR	RSID	BP	A1	A2	Weight

PRS – Figures

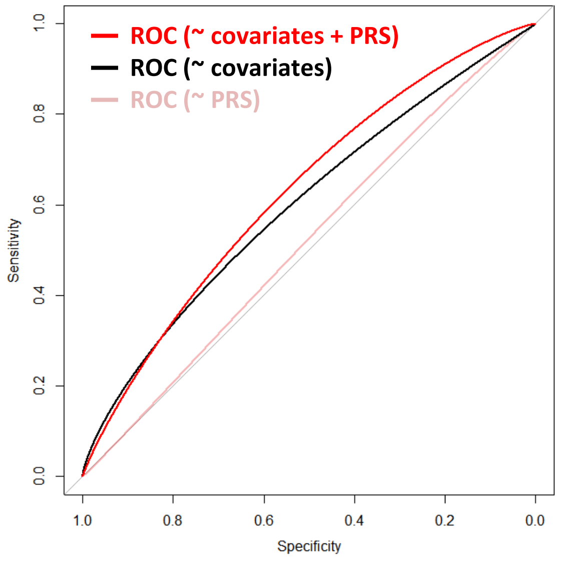
Density plot

Use it to understand patterns, trends, and the underlying structure of numeric data among groups.



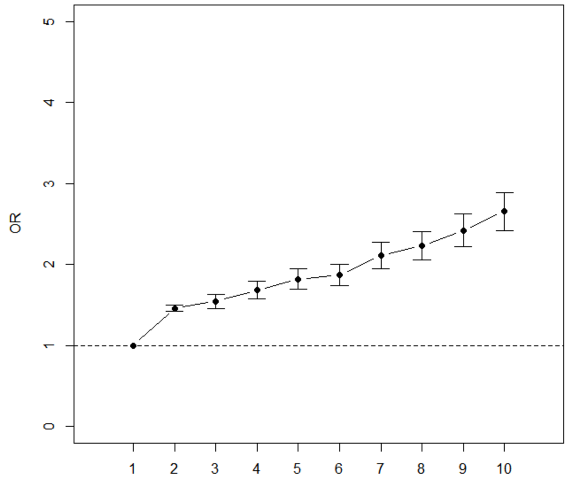
ROC curve

Use it to evaluate binary classifiers and understand their discrimination ability.



OR decile plot

It categorizes large data sets into 10 equally sized subsections (deciles) based on a given metric (e.g.,OR), and then fit a logistic regression model with a binary trait and a categorized *PRS* as the predictor



95% $\mathrm{CI}_{\mathrm{delta}}$ of OR = [OR $\pm$ 1.96 $\times$ OR $\times$ se(BETA)]
95% $\mathrm{CI}_{\mathrm{MLE}}$ of OR = [exp(BETA $\pm$ 1.96 $\times$ se(BRTA))]

log(OR) = BETA

Note

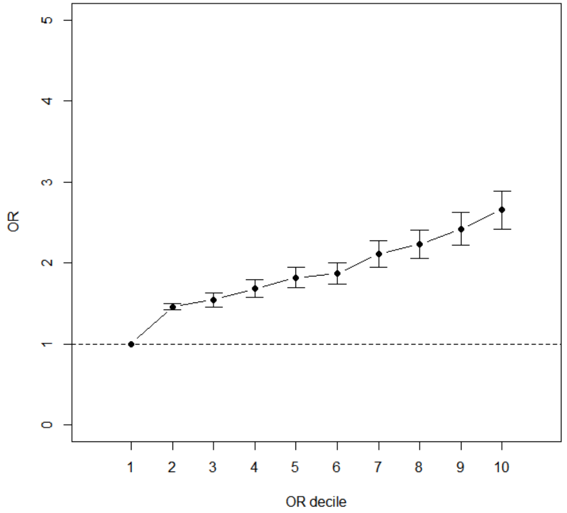
- **QC for base data**
  - Duplicated SNPs: it occurs an error when using plink **--score** to calculate PR
  - Ambiguous SNPs: if there is no information about strands of base and target data, exclude them!
  - Mismatch SNPs: when using plink **–score** to calculate PRS, it treats flipped alleles of a SNP as distinct
- **QC for target data**
  - Array data: GWAS QC
  - Imputation data: Sample QC (based on array sample QC) + Variant QC (infoscore, CR, MAF)
- **Software usage**
  - Consistence of genome builds: base data, target data, and reference data
  - Practice https://choishingwan.github.io/PRS-Tutorial/

    https://privefl.github.io/bigsnpr/articles/

Software	LD resource	LD genome build
PLINK	target data	
PRSice2	target data	
LDpred2	HapMap3	hg18, GRCh37 (hg19), GRCh38 (hg38)
lassosum	1000 genomes project Phase I	GRCh37 (hg19), GRCh38 (hg38)
PRS-CS/PRS-CSx	1000 genomes project Phase 3 UK Biobank	GRCh37 (hg19)

Note

- P**R**S is based on effect (risky) alleles of SNPs and these alleles have either positive or negative effects. In large-scale studies, the cumulative effect tends to be dominated by the positive effects. Therefore, higher PRS values are often linked to an increased risk of diseases. In an OR decile plot, we may expect to observe a upward trend between OR and PRS.



- As we know, DNA is relatively stable throughout an individual's life. Therefore, relying solely on a PRS prediction model is insufficient. To more accurately assess disease risk, we must consider additional factors or covariates: demographic variables (age and gender), environmental variables (abc-covariates), etc.
- Hingorani, Aroon D., et al. (2023) Performance of polygenic risk scores in screening, prediction, and risk stratification: secondary analysis of data in the Polygenic Score Catalog. BMJ medicine 2.1 – Disapproval of PRS

Thanks for your attention!!
<(_ _)>