Introduction to Polygenic Risk Score

Lin, Chien-Wei Taiwan Biobank



Human genomic variation



Single-Nucleotide Variant (SNV)



Linkage disequilibrium (LD)

LD, usually expressed as r^2 , is the nonrandom association of alleles at different loci in a given population Homologous Chromosomes

- $r^2 = 1$: the alleles are completely correlated
- $r^2 = 0$: the alleles are in linkage equilibrium





YRI

CEU

Genome-wide association study (GWAS)

GWAS involves scanning markers across the genomes of many people to find genetic variations associated with a particular disease



Beyond GWAS



Polygenic Risk Score (PRS)

- PRS uses information from external GWAS results to predict risk in our own study sample
- PRS is the sum of trait-associated risk alleles across many genetic loci, typically weighted by effect sizes from a GWAS



PRS construction



ID	Alcohol	Betel nut	Cigarette	Score
Α	0	0	0	0
В	1	0	1	2
С	1	1	1	3

Score	_
Score	

ID	Alcohol	Betel nut	Cigarette	Score
А	0	0	0	$0 \times \beta 1 + 0 \times \beta 2 + 0 \times \beta 3$
В	1	0	1	$1 \times \beta 1 + 0 \times \beta 2 + 1 \times \beta 3$
С	1	1	1	$1 \times \beta 1 + 1 \times \beta 2 + 1 \times \beta 3$

PRS construction



(2) Genotype data

	SNP1	SNP2	SNP3	SNP4
Individual 1	AT	CG	TT	CC
Individual 2	TA	GG	GT	CA
Individual 3	TT	CC	GT	CA
Individual 4	TT	CC	GG	AA

3 Polygenic risk score

Individual 1	1.5	-	0.5	+	4.0	-	0.0	= 5.0
Individual 2	1.5	_	0.0	+	2.0	_	1.5	= 2.0
Individual 3	0.0	-	1.0	+	2.0	-	1.5	= -0.5
Individual 4	0.0	-	1.0	+	0.0	-	3.0	= -4.0

4 PRS distribution



Nature Reviews Methods Primers volume 1, Article number: 59 (2021)

PRS analysis workflow

- Summary statistics
 - Estimated association coefficients for every variant analyzed in a GWAS
- Base data
 - Large GWAS providing information on association coefficients
- Target data
 - Relatively smaller studies used to determine tuning parameters or/and to report final model performance



Challenges in PRS development

Variant selection

- Determine which variants to include
- Weight calculation
 - The trait-associated weights to assign to the selected variants



Med Rev (2021). 2022 Feb 14;1(2):129-149.

Clumping and thresholding (C+T)



Development of PRS



Circ Res. 2020 Apr 24;126(9):1159-1177.

Evaluating PRS

- Accuracy of the PRS can be assessed by various metrics
 - Quantitative traits: R²
 - Binary traits: AUC
- Comparing the difference in phenotypic variance explained by the PRS in two models



 $reduced \ model: \ Phenotype \sim Covariates$



Nature. 2021 Mar;591(7849):211-219.

Open resource

- The Polygenic Score (PGS) Catalog (https://www.pgscatalog.org)
- Scoring info and relevant metadata necessary to apply PRS

				8 0				PGS003886.txt		以「文字編輯」打開
PGS Catalo	IG Home Browse - Download	Is • Documentation •	###PGS CATALOG #format_versio	SCORING	FILE – see http	s://www.	pgscatalog.org/downloads/	#dl_ftp_scoring fo	r additional information	
An	The F	Polygenic Score (PGS) Car res and the relevant metadata required for a	Latest release: Feb. 20, 2 Latest release: Feb. 20, 2 ccurate application and evaluation.	<pre>##POLYGENIC SC #pgs_id=PGS003 #pgs_name=BMI_ #trait_reporte #trait_mapped= #trait_efo=EF0 #genome_build=</pre>	DRE (PGS) 886 PRScsx_ARI d=Body mas body mass _0004340 ng19	INFORMATION B_EASweights ss index (BMI) index				
	Search the PGS Catalog		Q	<pre>#variants_numb #weight type=b</pre>	er=941406 eta					
	Examples: breast cancer, glauco	ma, BMI, EFO_0001645		##SOURCE INFOR	MATION					
	A reproducible workflow t	Available tool: pgsc_calc	> See more information	#citation=Shim rsID chr_na rs1048488 rs6689308 rs6687776 rs6678318 rs6671356 rs6604968 rs4970405 rs12726255	I et al. ne 1 1 1 1 1 1 1 1	Nat Commun (20 chr_position 760912 C 1029805 G 1030565 T 1030633 A 1040026 C 1041700 G 1048955 G 1048955 G	23). doi effect T A C G T A A A A	:10.1038/s41467-023-41985 allele other_allele 3.977365e-05 1.361395e-05 1.361395e-04 1.177362e-05 2.712557e-05 -4.050493e-07 3.790887e-05 -2.532721e-05	-1 effect_weight	
	In the surrent DCC C		fellouing antopping.	rs17160824	1	1053452 A 1060608 A	G	4.643302e-05		
	In the current PGS C	atalog you can browse the scores and metadata through the	tollowing categories:	rs12757754 rs1571150 rs7290 1	1 1 1477244	1061262 G 1474304 A C T	A C 3.7736	8.627162e-06 -2.011896e-05 20e-05		
Po	olygenic Scores	Traits	Publications	rs3766180	1	1478153 C	T	1.488039e-05		
		A 004	~ = = = =	rs7517401	1	1483010 A	G	2.194888e-07		
	Z 4,223	1 624	LL 583	rs3820075 rs3766170	1 1	1490559 G 1497008 C	A T	5.615336e-05 3.047732e-05		
				rs3766169	1	1497201 C	A	-5.221937e-06		
				rs9439468	1	1499298 G	A	1.071501e-05		
				rs6603791	1	1500941 G 1505255 T	A C	4.2819196-05		
				rs7519837	1	1510801 T	c	9.804223e-06		

Open resource

Cancer-PRSweb (<u>https://prsweb.sph.umich.edu:8443</u>)

	Center for precision Health data science UNIVERSITY OF MICHIGAN																			
Home Q	Search P	hecode	e 📲 Meth	nod	E Co	ntact	🖽 New	s 💷 PR	S Overview											
											Overvie	w								
					Association	Overall Perfor	mance ¢		Breast cano	cer [female]						## PRSweb reference PRSW ## PRSweb LD reference MGI	EB_PHECODE174.1_0	nco-1COGS-Overall-BRCA_LASSOSU	M_MGI_20200211	
GWAS Source o	Phenotype Description	• Method •	Tuning Parameter	# SNPs	¢ P-value ¢	Pseudo- R2	Brier ¢ Score	100 -	Malignant n Breast cano	eoplasm of	female breast					## PRSweb date 20200211 ## GWAS source 29059683 ## GWAS reference PUBM	ED] (m-1)		
] ====						60 -								## GWAS id Onco_iCOGS_0 ## GWAS URL http://bcac.	verall_BRCA ccae.medschl.cam	ac.uk/bcacdata/oncoarrav/awas-	icoas-and-oncoarray-summar	v-results/
Michailidou K et al. 2017	Breast cancer (female), Overal	Lassosum	s=0.5;Lambda=0.00545	6 44815	5.5E-105	0.0565412	0.137 (Acquired al	bsence of b	reast			Benigr		<pre>## PRS method Lassosum 0.4 ## PRS tuning parameter s0.5</pre>	.4 (MAF >= 1%; au _Lambda0.0054555	itosomal variants only) 1478116852		
Michailidou K et al. 2017	Breast cancer [female]; ER pos	Lassosum	s=0.5;Lambda=0.00885	9 61635	4.7E-92	0.049783	0.138 (20 -	Radioth	erapy				ı		## PRS evaluation in MGI ## Genome build GRCh37/hg19	Fi i	-		
Michailidou K et al. 2017	Breast cancer [female], Overal	P&T	P<=0.000316	2723	3.9E-91	0.0463534	0.138 (16 -	Maligna	nt neoplasr	n, other			Mammogra		1 868418 C T 1 1120431 G A	T C A G	-1.63701239816651e-05 0.000458393745540083		
Michailidou K et al. 2017	Breast cancer [female], Overal	P&G	P<=5e-8	326	4.4E-81	0.0410092	0.139 (12 -	Second	nerapy ary maligna	ncy of lymph nodes			Other spec		1 1256608 G A 1 1276077 G A	A G A G	-0.000413492163926478 0.000696343975624821		
Michailidou K et al. 2017	Breast cancer [female], Overal	Lassosum	s=0.5;Lambda=0.00885	9 18356	3.2E-78	0.039308	0.139 0	8 -	Second	ary maligna	nt neoplasm		D	Inflamn		1 1315065 G T 1 1525437 G A 1 1973608 A G	T G A G	-0.000346618630582058 0.00177341550848111 -0.00091288340755925		
Michailidou K et al. 2017	Breast cancer [female]; ER pos	P&T	P<=7.94e-05	1142	4.1E-78	0.0389762	0.139 (4		enign neor	lasm of lymph nodes	sites	_ Breast condition	ns, congenitale		1 2205581 C A 1 2339139 G A	A C A G	-0.000473772396560399 -0.000284476830040966		
Michailidou K et al. 2017	Breast cancer [female]; ER pos	Lassosum	s=0.9;Lambda=0.00885	9 98026 I Table	1.6E-66	0.0340499	0.139 (0			PRS PheW	AS				1 23972434 A G 1 2397283 G A 1 3097111 G A 1 3148263 C A	G A A G A G A C	-0.00137683254334791 -0.00137257261430263 -0.000223645711854932 -0.000728083919434533 PRS Weight File		
Integrating p repository fo thresholds) o population-b construct do For more info	egrating published and freely available genome-wide association studies (GWAS) summary statistics from multiple sources (published GWAS, the NHGRI-EBI GWAS Catalog, or UKB-based GWAS), we created an online pository for polygenic risk scores (PRS) for common cancer traits. Our framework condenses these summary statistics into PRS using linkage disequilibrium pruning and p-value thresholding (fixed or data-adaptively optimized esholds) or penalized, genome-wide effect size weighting. We evaluate them in the cancer-enriched cohort of the Michigan Genomics Initiative (MGI), a longitudinal biorepository effort at Michigan Medicine, and in the pulation-based UK Biobank Study (UKB). For each PRS construct, measures on performance, calibration, and discrimination are provided. Beyond the cancer PRS evaluation in MGI and UKB, the PRSweb platform features nstruct downloads, risk evaluation in the top percentiles, and phenome-wide PRS association studies (PRS-PheWAS) for a subset of PRS that are predictive for the primary cancer.											nized ures								
Cancer Trait											Evaluation C	Cohort	Odds Rat	іо Тор						
Select Car	ncer Site/	Trait								•	Select Cor	nort 💌	1% vs F	Rest 💌	Sh	now table with PRS				

PRS tutorial

Basic Tutorial for Polygenic Risk Score Analyses

Search docs

⊟ Home

Overview

Datasets

Requirements

Citation

1. QC of Base Data

2. QC of Target Data

3. Calculating and analysing PRS

PLINK

PRSice-2

LDpred-2

lassosum

4. Visualizing PRS Results

Home

Overview

Docs » Home

This tutorial provides a step-by (PRS) analyses and accompany provide a simple introduction of existing users with a better und "underneath the hood" of popul

The tutorial is separated into fo paper: the first two sections on constitute a 'QC checklist' for P (here with examples using PLIN Section 3 of the paper, while the protocols

REVIEW ARTICLE https://doi.org/10.1038/s41596-020-0353-1

Check for updates

Tutorial: a guide to performing polygenic risk score analyses

Shing Wan Choi^{1,2}, Timothy Shin-Heng Mak^{®3} and Paul F. O'Reilly^{1,2}⊠

A polygenic score (PGS) or polygenic risk score (PRS) is an estimate of an individual's genetic liability to a trait or disease, calculated according to their genotype profile and relevant genome-wide association study (GWAS) data. While present PRSs typically explain only a small fraction of trait variance, their correlation with the single largest contributor to phenotypic variation—genetic liability—has led to the routine application of PRSs across biomedical research. Among a range of applications, PRSs are exploited to assess shared etiology between phenotypes, to evaluate the clinical utility of genetic data for complex disease and as part of experimental studies in which, for example, experiments are performed that compare outcomes (e.g., gene expression and cellular response to treatment) between individuals with low and high PRS values. As GWAS sample sizes increase and PRSs become more powerful, PRSs are set to play a key role in research and stratified medicine. However, despite the importance and growing application of PRSs, there are limited guidelines for performing PRS analyses, which can lead to inconsistency between studies and misinterpretation of results. Here, we provide detailed guidelines for performing and interpreting PRS analyses. We outline standard quality control steps, discuss different methods for the calculation of PRSs, provide an introductory online tutorial, highlight common misconceptions relating to PRS results, offer recommendations for best practice and discuss future challenges.

visualising PRS results, accompanies Section 4 of the paper.

- 1. Quality Control (QC) of Base Data
- 2. Quality Control (QC) of Target Data
- 3. Calculating and analysing PRS
- 4. Visualising PRS Results

We will be referring to our guide paper in each section and so you may find it helpful to have the paper open as you go through the tutorial.

