Polygenic risk scores

Sarah Medland
with thanks to Lucia Colodro Conde & Baptiste Couvy Duchesne
What are Polygenic risk scores (PRS)?

• PRS are a **quantitative measure of the cumulative genetic risk** or vulnerability that an individual possesses for a trait.

• The traditional approach to calculating PRS is to **construct a weighted sum of the betas** (or other effect size measure) for a set of **independent loci** thresholded at different significance levels.
  • Typically the independence is LD based (LD r2 <= .2) via clumping.
The classics


• Evans DM, Visscher PM., Wray NR. Harnessing the information contained within genome-wide association studies to improve individual prediction of complex disease risk. Human Molecular Genetics. 2009; 18(18): 3525-3531.


Further reading


Traditional approach

1. GWAS “Discovery Sample”
2. GWAS “Target Sample”
3-5. Select top SNPs and identify risk alleles

Variance of target sample phenotype explained by predictor

- 6. Genomic profiles weighted sum of risk alleles
- 7. Evaluate

Association analysis summary statistics

Discovery & Target samples could be:
A. Same Disorder
B. Different disorders
C. Disorder subtypes

Wray et al (2014) *J Child Psychol Psychiatry*
Traditional approach

1. GWAS “Discovery Sample”
   Association analysis summary statistics
   Discovery & Target samples could be:
   A. Same Disorder
   B. Different disorders
   C. Disorder subtypes

2. GWAS “Target Sample”
   3-5. Select top SNPs and identify risk alleles
   Apply
   MUST BE INDEPENDENT

6. Genomic profiles weighted sum of risk alleles

Variance of target sample phenotype explained by predictor

7. Evaluate

Association analysis summary statistics

1. GWAS “Discovery Sample”

3-5. Select top SNPs and identify risk alleles

Discovery & Target samples could be:
A. Same Disorder
B. Different disorder
C. Disorder subtypes

2. GWAS “Target Sample”

Apply

5. Genomic profiles weighted sum of risk alleles

6. Variance of target sample phenotype explained by predictor

\[ 1 \times \beta_C - 2 \times \beta_G + 1 \times \beta_A + 0 \times \beta_T = 2 \times \beta_C + 2 \times \beta_T = 0.052 \]

Effect size from GWAS

Polygenic score:

\[ \beta_C = -.02, \beta_G = .01, \beta_A = .002, \beta_T = .025 \]
Main uses of PRS

1) Single disorder analyses

2) Cross-disorder analysis

3) Sub-type analysis
Single trait analyses

PGC-MDD July 2016

% variance explained of depression score

<5e-8 p=0.012
p=0.00042
p=1.4e-06
p=6.6e-07
p=6.2e-08
p=4.3e-08

OPEN

Molecular Psychiatry (2017) 00, 1–7
www.nature.com/mp

ORIGINAL ARTICLE

A direct test of the diathesis–stress model for depression

L Colodro-Conde, B Couvy-Duchesne, G Zhu, WL Coventry, EM Byrne, S Gordon, MJ Wright, GW Montgomery, PAF Madden, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, S Ripke, LI Eaves, AC Heath, NR Wray, SE Medland and NG Martin

The diathesis–stress theory for depression states that the effects of stress on the depression risk are dependent on the diathesis or vulnerability, implying multiplicative interactive effects on the liability scale. We used polygenic risk scores for major depressive disorder (MDD) calculated from the results of the most recent analysis from the Psychiatric Genomics Consortium as a direct measure of the vulnerability for depression in a sample of 5221 individuals from 3083 families. In the same we also had measures of...
Moderated single trait analyses

% variance explained of depression score

PSLE * PRS interaction

p=0.0076

A direct test of the diathesis–stress model for depression

L Colodro-Conde¹,², B Couvy-Duchesne¹,², G Zhu³, WL Coventry⁴, EM Byrne⁵, S Gordon⁶, MJ Wright⁷,⁸, GW Montgomery⁹, RAF Madden⁴, MA Thompson⁴, S Ripke⁴,⁵,⁶,⁷,⁸,⁹,¹⁰, LJ Eaves¹¹, AC Heath⁵, NR Wray⁴,⁵, SE Medland⁴ and NG Martin¹

The diathesis–stress theory for depression states that the effects of stress on the depression risk are dependent on the diathesis or vulnerability, implying multiplicative interactive effects on the liability scale. We used polygenic risk scores for major depressive disorder (MDD) calculated from the results of the most recent analysis from the Psychiatric Genomics Consortium as a direct measure of the vulnerability for depression in a sample of 5221 individuals from 3083 families. Within the same we also had measures of...
Cross-trait analysis

Original Investigation
September 2018

Association Between Population Density and Genetic Risk for Schizophrenia

Lucía Colodro-Conde, PhD1; Baptiste Couvy-Duchesne, PhD2,2; John B. Whitfield, PhD1; et al

Author Affiliations

Single and cross-trait analyses

Correlations: Genome-wide Polygenic Scores (pT = 0.3) and phenotypes

Krapohl et al (2016)
Molecular Psychiatry
Sub-type analysis

Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression

Naomi R. Wray, Stephan Ripke, [...] the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium

Nature Genetics 50, 668–681 (2018) | Download Citation
The power of the predictor is a function of the power of the GWAS in the **discovery sample** (due to its impact on the accuracy of the estimation of the betas).

“I show that discouraging results in some previous studies were due to the low number of subjects studied, but a modest increase in study size would allow more successful analysis. However, I also show that, for genetics to become useful for predicting individual risk of disease, hundreds of thousands of subjects may be needed to estimate the gene effects.”

(Dudbridge, 2013)
PRS and power

For simple power calculations you can use a regression power calculator (for $r^2$ of up to 0.5%).

As a general rule of thumb you usually want 2,000+ people in the target dataset.

→ R AVENGEME (https://github.com/DudbridgeLab/avengeme)

Power calculator for discovery (GWAS) sample needed to achieve prediction of $r^2$ in target sample

```r
sampleSizeForGeneScore(targetQuantity, targetValue, nsnp, n2 = NA, vg1 = 0,
cov12 = vg1, pi0 = 0, weighted = TRUE, binary = FALSE,
prevalence = 0.1, sampling = prevalence, lambdaS = NA,
shrinkage = FALSE, logrisk = FALSE, alpha = 0.05, r2gx = 0,
corgx = 0, r2xy = 0, adjustedEffects = FALSE)
```
Power of PRS analysis increases with GWAS sample size

PGC-MDD1: N=18k
max variance explained = 0.08%, p=0.018

PGC-MDD2: N=163k
max variance explained = 0.46%, p= 5.01e-08

Molecular Psychiatry
General steps of processing
(1) GWAS summary statistics

→ From PGC results, other public domain GWAS, unpublished GWAS

SNP identifier (rs number, Chr:BP)

Both Alleles (effect/reference, A1/A2)

Effect

• Beta from association with continuous trait
• OR from an ordinal trait - convert to log(OR)
• Z-score, MAF and N (from an N weighted meta-analysis)

p-value

(frequency of A1)
GWAS summary statistics

- From PGC results, other public domain GWAS, unpublished GWAS

SNP identifier (rs number, Chr:BP)

Both Alleles (effect/reference, A1/A2)

Effect

- Beta from association with continuous trait
- OR from an ordinal trait - convert to log(OR)
- Z-score, MAF and N (from an N weighted meta-analysis)

p-value

(frequency of A1)

Make sure that your target genotypes are named the same way as your discovery data!

- Imputation reference and genomic build
(2) Find SNPs in common with your local sample and QC

- Imputed data
- QC
  - $R^2 \geq 0.6$
  - MAF $\geq 0.01$
  - No indels
  - No ambiguous strands (*) - A/T or T/A or G/C or C/G

```bash
for ((i=1;i<=22;i++))
do
  awk '{ if ($5<=.01 & $6<=.99 & $6>=.6) print $1}' file"$i".info >> available.snps
done
```
(*) On ambiguous strands

GWAS chip results are expressed relative to the + or – strand of the genome reference.
(3) Clumping

- Select most associated SNP per LD region (pruning)
- Plink1.9 --bfile bfileReferencePanelForLD
  --extract QCedListofSNPs
  --clump gwasFileWithPvalue
  --clump-p1 (#Significance threshold for index SNPs)
  --clump-p2 (#Secondary significance threshold for clumped SNPs)
  --clump-r2 (#LD threshold for clumping)
  --clump-kb (#Physical distance threshold for clumping)
  --out OutputName
# Clump data in 2 rounds using plink2

# 1st clumping & extract tops snps for 2nd round
for ((i=1;i<=22;i++))
do
    plink2 --bfile reference --chr "$i" --extract available.snps --clump GWAS.noambig
    --clump-p1 1 --clump-p2 1 --clump-r2 .5 --clump-kb 250 --out traitX"$i".round1
    awk '{print $3, $5}' traitX"$i".round1.clumped > traitX"$i".round2.input
    awk '{print $3}' traitX"$i".round1.clumped > traitX"$i".extract2
done

# 2nd clumping & extract tops snps for profile
for ((i=1;i<=22;i++))
do
    plink2 --bfile reference --chr "$i" --extract traitX"$i".extract2 --clump traitX"$i".round2.input --
    clump-p1 1 --clump-p2 1 --clump-r2 .2 --clump-kb 5000 --out traitX"$i".round2
    awk '{print $3}' traitX"$i".round2.clumped > traitX"$i".selected
done
(4) Calculate risk scores

The traitX"$i".selected files will contain the lists of top independent snps. Merge the alleles, effect & P values from the discovery data onto these files.

To do a final strand check merge the alleles of the target set onto these files. If any SNPs are flagged as mismatched you will have to manual update the merged file - flip the strands (ie an A/G snp would become a T/C snp) but leave the effect as is.

Create Score files (SNP EffectAllele Effect) and P files contain (SNP Pvalue).

for ((i=1;i<=22;i++))
do
awk '{ if ($6==$8 || $6==$9  ) print $0, "match" ; if ($6!=$8 && $6!=$9 ) print $0, "mismatch"}'
traitX."$j".merged > strandcheck.traitX."$i"
grep mismatch strandcheck.traitX* done
(4) Calculate risk scores

for ((i=1;i<=22;i++))
do
plink --noweb --dosage Your_chr"$i".plink.dosage.gz format=1 Z --fam
Your_chr"$i".plink.fam --score traitX."$i".score --q-score-file traitX."$i".P --q-score-range p.ranges --out Your_chr"$i".PRS
done

p.ranges
S1  0.00 0.000001
S2  0.00 0.01
S3  0.00 0.10
S4  0.00 0.50
S5  0.00 1.00
(5) Run association analysis – unrelated individuals

```r
base <- lm (ICV ~ age + sex + PC1 + PC2 + PC3 + PC4 + other-covariates, data = mydata)
score1 <- lm (ICV ~ S1 + age + sex + PC1 + PC2 + PC3 + PC4 + other-covariates, data = mydata)
score2 <- lm (ICV ~ S2 + age + sex + PC1 + PC2 + PC3 + PC4 + other-covariates, data = mydata)
model_base <- summary(base)
model_score1 <- summary(score1)
model_score2 <- summary(score2)
model_base$r.squared
model_score1$r.squared
model_score2$r.squared
anova(base, score1)
anova(base, score2)
```
(5) Run association analysis, controlling for relatedness

```
gcta --reml
--mgrm-bin GRM
--pheno phenotypeToPredict.txt
--covar discreteCovariates.txt
--qcovar quantitativeCovariates.txt
--out Output
--reml-est-fix
--reml-no-constrain
```
Other Methods
Genetic Best Linear Unbiased Predictor

Application to genetic data (animal breeding) HENDERSON, C. R. (1950). Estimation of genetic parameters

Review of method and example:

Charles Roy Henderson
1911-1989
BLUP in context of linear models

GWAS estimates: marginal SNP effect

\[ y = Xb + e \]

N individuals

\[ Y_{Nx1} \] phenotype centered

\[ X_{Nx1} \] SNP centered

Joint and conditional SNP effect

\[ y = Xb + e \]

N individuals

\[ Y_{Nx1} \] phenotype centered

\[ X_{Nxm} \] SNPs centered

Yang et al., 2012

BLUP effect

\[ y = Zs + e, \]

N individuals

\[ Y_{Nx1} \] phenotype centered

\[ Z_{Nxm} \] SNPs centered

\[ s_{mx1} \] vector of SNP effects assumed \( \sim N(0, \sigma^2_s) \)

Goddard et al., 2009

\[ \hat{b} = (X'X)^{-1}X'y \]

\[ \hat{b} = (X'X)^{-1}X'y \]

\[ \hat{s} = Z'(ZZ'\sigma^2_s + I\sigma^2_e)^{-1}y \]
Calculating BLUP effect sizes

\[ y = Zs + e, \]
\[ \hat{s} = Z'(ZZ'\sigma_s^2 + I\sigma_e^2)^{-1}y \]

\(Z'Z\): nxn variance-covariance matrix of genotypes
Often not available from GWAS
Can be estimated from the GWAS allele frequencies and LD from a reference panel
(assumed same population)
Yang et al., 2012

\texttt{gcta64} \hspace{1cm} \texttt{--bfile ReferencePanelForLD} \hspace{1cm} \texttt{--cojo-file GWAS\_sumstat.ma} \hspace{1cm} \texttt{--cojo-sblup} \texttt{1.33e6} \hspace{1cm} \texttt{--cojo-wind 1000} \hspace{1cm} \texttt{--thread-num 20} \hspace{1cm} \texttt{--cojo-sblup = m * (1 / h_{SNP}^2 - 1)}

With \(m\) the number of SNPs
BLUP limitations and perspective

\[ y = Zs + e, \]
\[ \hat{s} = Z'(ZZ'\sigma_s^2 + I\sigma_e^2)^{-1}y \]

Requires to inverse \( (ZZ'\sigma_s^2 + I\sigma_e^2) \)
Which can be computationally intensive for large sample sizes

Open field of prediction models

• BLUP “shrinks” the estimates: hypothesis of normally distributed effect sizes “infinitesimal model”
• Other shrinkage methods include LASSO: hypothesis of mixture of effect sizes (double exponential...)

• Non-additive models? That may include epistasis, dominance
• Semi-parametric models

see Goddard et al., 2009 for review
LDpred

Bayesian estimation of the BLUP effect sizes: “posterior mean effect size of each marker by using a prior on effect sizes and LD information from an external reference panel”

Vilhjalmsson et al., 2015
LDpred

Application to real data
Vilhjalmsson et al., 2015

BLUP marginally better than Pruning + Thresholding
Multiple testing due to the high resolution in p-value threshold.

Authors suggest $p<0.001$ if using the best fit PRS.

Significance threshold dependent on LD in the target sample and distribution of the phenotype predicted.

Unclear if it holds for phenotypes with skewed distributions and for non UK samples.

Euesden et al., 2014
<table>
<thead>
<tr>
<th></th>
<th>Classic / OLS</th>
<th>BLUP</th>
<th>PRSice</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dosage or best guess</td>
<td>Best guess</td>
<td>Best guess</td>
<td>Dosage or best guess</td>
</tr>
<tr>
<td>clumping</td>
<td></td>
<td>BLUP effects summed over all SNPs</td>
<td>clumping</td>
</tr>
<tr>
<td>Multiple PRS by p-value thresholds</td>
<td>Unique PRS</td>
<td>All p-value threshold tested</td>
<td></td>
</tr>
<tr>
<td>Bonferroni correction</td>
<td></td>
<td></td>
<td>Unclear significance threshold for association</td>
</tr>
<tr>
<td>Hypothesis: effect sizes of SNPs normally distributed</td>
<td></td>
<td>Hypothesis: effect sizes of SNPs normally distributed</td>
<td></td>
</tr>
<tr>
<td>Fast (can be parallelized)</td>
<td>Fast (can be parallelized)</td>
<td>Matrix inversion, can be long for large N</td>
<td>Slower and harder to parallelize (R package)</td>
</tr>
<tr>
<td>PLINK</td>
<td>GCTA, PLINK</td>
<td>R (PLINK)</td>
<td></td>
</tr>
</tbody>
</table>
PRS-on-Spark (PRSoS): a novel, efficient and flexible approach for generating polygenic risk scores

A couple of “worked” examples
Original Investigation

September 2018

Association Between Population Density and Genetic Risk for Schizophrenia

Lucía Colodro-Conde, PhD¹; Baptiste Couvy-Duchesne, PhD¹,²,³; John B. Whitfield, PhD¹; et al

Author Affiliations

Background

• The prevalence of schizophrenia is higher in urban areas than in rural areas \( \rightarrow \) O.R. = 2.39 (1.62–3.51), (Vassos et al 2012, Schizophrenia Bulletin).
• Two major hypotheses have been proposed to explain this phenomenon:
  
  (1) *causation hypothesis*: the stress of city life and undefined factors in the urban environment increase the risk of this disease.
  
  (2) *selection hypothesis*: individuals with genetic liability for schizophrenia move into urban areas.
Twin models have shown genetic factors have a higher impact on the country vs. city living as people grow older, while the impact of family background decreases.

Whitfield et al. 2005, *Twin Research and Human Genetics*
Adults with higher genetic risk for schizophrenia are more likely to live in urbanised and populated areas than those with lower risk.
Methods

- 15,544 individuals in 7,015 families (65.6% females, age mean: 54.4, SD: 13.2) living in Australia.

- Participants were genotyped genome-wide and imputed to 1000G v.3.

- Reported their **postcode** as part of the protocols of several studies on health and wellbeing conducted from QIMR.
The map and bar chart represent the population density per square kilometre in Australia. Sydney has the highest population density, followed by Melbourne, Brisbane, Perth, Adelaide, Canberra, Hobart, and Darwin, with the least population density. The map uses color coding to indicate different density levels, while the bar chart provides a visual comparison of population density among the listed cities.
Measures of urbanicity:

- Population density
- Remoteness
- Socio economic status (SES)

(data from the Australian Bureau of Statistics)
phenotype = intercept + beta0*covariates + beta1*g + e with g ~ N(0, GRM)

phenotype: population density or remotedness
covariates: PRS-SCZ, age, sex, (SES),
        4 first genetic principal components, imputation chip
e: error
GRM: Genetic correlation matrix

We calculated p-values using the t-statistic calculated on the basis of the Fix_eff and SE from the GCTA output.
We then applied Bonferroni correction (Sidak method) for multiple testing yielding a significant threshold of 0.004.

Genome-wide Complex Trait Analysis v. 1.22
(Yang J et al 2011, Am J Hum Genet )
(a) % variance explained in population density

% variance explained in remoteness

with SES as covariate
Conclusions

- People with a higher genetic risk for schizophrenia may prefer to live in more urban and populated areas.
  - Importantly, this study does not use a case-control sample but an unselected population sample where the genetic risk for schizophrenia was estimated.
- Greater genetic predisposition to schizophrenia is at least one mechanism explaining why this illness is more prevalent in city environments.
- Future research should test if this effect is replicated in another countries, analyse migration effects and identify what aspects of urbanised life correlate with SCZ genetic risk.
Original Article

A direct test of the diathesis–stress model for depression
OPEN

L Colodro-Conde¹², B Couvy-Duchesne¹², G Zhu¹, W L Coventry⁴, E M Byrne⁵, S Gordon¹, M J Wright³, G W Montgomery⁵, P A F Madden⁷, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium¹³, S Ripke⁸, L J Eaves¹¹, A C Heath⁷, N R Wray³, S E Medland¹ and N G Martin¹
Diathesis-Stress model in depression

Depression = Diathesis + Stress + Diathesis*Stress

Diathesis: (Predisposition, Vulnerability)
Stress: (Disruption of psychological equilibrium)

Hypothesised contribution to risk

Depression = Polygenic risk scores (PRS) + Personal stressful life events (PSLE) + Network stressful life events (NSLE) + lack of social support (SS) + PRS*PSLE + PRS*NSLE + PRS*SS
Sample and data

- 5,221 twins from 3,083 twin families
- European ancestry (<6SD from PC1/PC2 centroid)
- Mean age 35.7, range 17-85, 66% females
- Depression, Personal & Network stressful life events, Perceived social support
- GWAS arrays, imputed to 1000G reference
Measure of Depression

12 depression items

Selected from:

Delusion Symptoms Scale Inventory (DSSI) + Symptoms Check List (SCL)

All scores were estimated using Item Response Theory (IRT) – improves distribution, deals with missingness
REALITY CHECK - Increased odds of DSM-IV MDD diagnosis per decile of depression IRT score

- Association between Depression score and lifetime DSM-IV MDD diagnoses from telephone interview studies conducted 4-12 years after DSSI/SCL)
- \( p\)-value = 3.0e−108

→ Disease odds >6x in top decile of depression score compared to first decile
## Measures of Stress

### Personal stressful life events (PSLE)
(adapted from List of threatening experiences (Brugha et al. 1985,))

**Serious problem** with spouse, family member, friend, neighbour, workmate  
**Event:** Divorce, separation, illness, injury, accident, burgled, robbed, lost job, financial problems, legal troubles...

### Network stressful life events (NSLE)
(adapted from List of threatening experiences)

**Illness, Injury, death or personal crisis in close network** (spouse, child, mother, father, twin, sibling, someone else close)

### Social Support (SS)
(Kessler Perceived SS, KPSS)

How much your **close network:** listens to your **worries,** **understands** the way you feel / think, **helps you** if needed, **shares private feelings** with you

All scores were estimated using Item Response Theory (IRT) – improves distribution, deals with missingness
MAIN EFFECTS - POLYGENIC RISK SCORES

N=163k

(max variance explained = 0.46%, p = 4.3e-08)
Main effects - Polygenic Risk Scores

N=163k

Max variance explained = 0.46%, p = 4.3e-08

N=18k

Max variance explained = 0.08%, p = 0.018

Note increased variance accounted for with larger N
MAIN EFFECTS - POLYGENIC RISK SCORES

PRS main effects appear larger in males
MAIN EFFECTS - STRESSORS

Note sex differences for SS

* Blue is negative
Test of Interaction - it’s all coming from females!
the effect of the PSLE-diathesis interaction is visible when comparing the bottom (minimal PSLE) and top (maximal PSLE) edges of the surface.
Practical
Todays Data

http://labs.med.miami.edu/myers/LFuN/LFuN.html
post-mortem gene expression in ‘brain’ tissue
N=364
Real data – unfiltered!
https://sites.google.com/broadinstitute.org/ukbbgwasresults/

This site contains the results of the GWAS and heritability analyses conducted by the Neale Lab. Please refer to the description of these analyses here for details.
for i in {1..22}
  do
  echo $i
  rm chr"$i".pass
  zcat chr"$i".info.gz | awk '{ if ($5>=.01 && $7 >=.6) print $1}' > chr"$i".pass
  done

for i in {1..22}
  do
  echo $i
  ~/bin/plink2 --vcf chr"$i".dose.vcf.gz --extract chr"$i".pass --make-bed --out QCchr"$i" --threads 5
  done

for i in {2..22}
  do
  echo QC"$i".bed QC"$i".bim QC"$i".fam >> join.list
  done

for i in 20160 20161 20162 2887 3466 3476
  do
  echo SNP P A1 A2 Beta > "$i".4clumping
  zless "$i".assoc.tsv.gz | awk '{print $1, $6, $9}' | sed 's/:/ /g' | awk '{ if (NR>1) print $1 :: $2, $6, $3, $4, $5}'
  >> "$i".4clumping
  done

~/bin/plink1.9 --bfile QC1 --merge-list join.list --make-bed --out gwide
for i in 20160 20161 20162 2887 3466 3476
do
~/bin/plink1.9 --bfile ../imputed/gwide --clump "$i".4clumping --clump-p1 1 --clump-p2 1 --clump-r2 .2 --clump-kb 2000 --clump-verbose --clump-annotate A1 A2 Beta --out ind"$i"
done

for i in {1..22}
do
echo $i
~/bin/plink2 --bfile QCchr"$i" --exclude 3alleles --make-bed --out QC"$i" --threads 5
done

for i in 20160 20161 20162 2887 3466 3476
do
grep INDEX ind"$i".clumped | awk '{print $2, $7, $8, $9, $10}' | sed 's/,//g' >> ind"$i".scores
done

~/bin/plink1.9 --bfile ../imputed/gwide --score ind"$i".scores 1 4 5 sum no-mean-imputation include-cnt --out tobacco"$i"

for i in 20160 20161 20162 2887 3466 3476
do
awk '{ if ($2 <= .01) print $0 }' ind"$i".scores > ind"$i".b
awk '{ if ($2 <= .0001) print $0 }' ind"$i".scores > ind"$i".c
awk '{ if ($2 <= .000001) print $0 }' ind"$i".scores > ind"$i".d
cp ind"$i".scores ind"$i".a
done

for i in 20160 20161 20162 2887 3466 3476
do
for j in a b c d
do
~/bin/plink1.9 --bfile ../imputed/gwide --score ind"$i"."$j" 1 4 5 sum include-cnt --out "$j"."$i"
done
done

done
Today's data

- PRS for Ever Smoked and Pack Years
  - a no threshold
  - b \leq .01
  - c \leq .0001
  - d \leq .000001

- Phenotypes expression of BDNF, CHRNA5 & HTR2A in Cortex

- Covariate AD status & Ancestry MDS