Predicting Polygenic Risk of Psychiatric Disorders

Supplemental Information

Table S1. Genetic and statistical concepts for genetic risk prediction

Single nucleotide polymorphism (SNP)	A polymorphic mutation in the genome. Many SNPs in an associated GWAS region often 'tag' causal variants because of LD structure in the genome but are not themselves causal.
Haplotype	A segment of DNA inherited identity-by-descent from a common ancestor that is characterized by a set of co-inherited SNPs.
Linkage disequilibrium (LD)	The correlation among proximal loci due to less historical recombination relative to distal/interchromosomal variants. Elevated association significance among nearby non-causal but non-independent loci are typically driven by LD.
Pleiotropy	The apparently unrelated effects of the same genes or variants influencing different biological pathways.
Sum of squares	Used in regression analysis to find the function that best fits the data by summing the squared difference of fitted values from the mean
Regression to the mean	The phenomenon in which a variable with an extreme value in an initial measure tends towards average upon a second measure. Galton described this phenomenon with the observation that extremely tall or short parents tend to have children of more average height than expected by chance.
Genome-wide association study (GWAS)	GWASs test the marginal associations of a single genotyped or imputed variant at a time with a phenotype of interest. These typically correct for potential confounders such as age, sex, population structure, etc.
GWAS summary statistics	Test statistics for the association between each SNP and phenotype, typically including the SNP identifier, risk allele, p value, and effect size estimate.
Effect size	A quantity measuring the effect of an allele on a phenotype, such as an odds ratio for binary traits, or a regression coefficient for continuous traits.
Biometrical model	Decomposes the variation in a phenotype into its genetic and environmental components.
Heritability	The fraction of trait variation that can be explained by genetic as opposed to environmental factors. Family-based and population-based estimates typically differ considerably (1).
Broad-sense (H^2)	All genetic factors contributing to phenotypic variation including additive, epistatic, and dominance effects, etc. It is typically studied in families, such as in twin studies.

Narrow-sense (h^2)	Strictly defined as the additive genetic factors contributing to a trait. This can be estimated in families (e.g., parent-offspring regression) or with SNPs in unrelated populations (e.g. h_g^2) (2).
Genetic architecture/ polygenicity	The complexity measured by number of independent causal loci and distribution of corresponding effect sizes contributing to a phenotype across the genome
Genetic correlation	The additive genome-wide relationship between two or more traits.
Mendelian randomization	An epidemiological tool that uses genetic variants to assess causal relationships between exposures and outcomes.

Supplemental Note. Considerations for evaluating the validity of PRS applications in published literature and clinical models

- The traits under investigation must show evidence of significant heritability.
- PRS must be generated from well-powered, independent GWAS performed with careful quality control.
- The fraction of phenotypic variation explained must be significant and meaningful. For PRS to be clinically meaningful, they should improve prediction accuracy in translational models when evaluated jointly with other clinical factors.
- The phenotype on which PRS are trained (i.e. the GWAS cohorts) should be similar to the phenotype evaluated in the test set. Appropriate conclusions should not extrapolate far beyond the original GWAS.
- Ancestry differences comprise a PRS challenge, in that prediction will perform best when the training and prediction cohorts are genetically similar.
- Individual-level risk should be evaluated against a stationary and consistent measurement (e.g., the average 50% of the population) rather than arbitrary comparisons (e.g., to the rest of the population).
- The methodological choice should be rigorous with consideration to LD, ancestry, and the genetic architecture of the trait.

Supplemental References

1. Zuk O, Hechter E, Sunyaev SR, Lander ES (2012): The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci USA. 109: 1193–1198.
2. Visscher PM, Hill WG, Wray NR (2008): Heritability in the genomics era — concepts and misconceptions. Nat Rev Genet. 9: 255–266.