Estimating genetic nurture using genome-wide association study summary statistics

Yuchang Wu[1], Xiaoyuan Zhong[2], Yunong Lin[3], Zijie Zhao[1], Jiawen Chen[3], Boyan Zheng[4], James J. Li[5,6], Jan S. Greenberg[6,7], Marsha R. Mailick[6,7], Jason M. Fletcher[4,8,9], Qiongshi Lu[1,3,9,#]

1 Department of Biostatistics and Medical Informatics, University of Wisconsin–Madison
2 University of Wisconsin–Madison
3 Department of Statistics, University of Wisconsin–Madison
4 Department of Sociology, University of Wisconsin–Madison
5 Department of Psychology, University of Wisconsin-Madison
6 Waisman Center, University of Wisconsin-Madison
7 School of Social Work, University of Wisconsin–Madison
8 La Follette School of Public Affairs, University of Wisconsin–Madison
9 Center for Demography of Health and Aging, University of Wisconsin–Madison

Background

- Marginal effect estimates in GWAS are mixtures of the direct and indirect genetic effects. Existing methods to dissect these effects require individual-level genetic and phenotypic data within families, which is difficult to obtain in practice.
- Here, we propose a novel statistical framework to estimate the direct and indirect genetic effects using only GWAS summary statistics conducted on self and offspring phenotypes.

Key idea

- The direct and indirect effects are given when both parents and offspring's genotype are jointly in the regression model.
- If instead using summary statistics from marginal regressions of own phenotype on own genotype and offspring phenotype on parental genotype, we could also estimate the direct and indirect effects.

Applications to birth weight

- Using own birth weight ~ own genotype GWAS and offspring birth weight ~ maternal genotype GWAS summary statistics[1] as input
- Our method gave consistent results with those obtained using individual-level data[1]

Applications to educational attainment (EA)

- Using EA3 summary statistics that excludes 23andMe samples[2] and offspring EA ~ parental genotype GWAS summary statistics from UKB, WLS, and HRS cohorts as inputs
- Direct and indirect effects showed distinct patterns of genetic correlations with 45 complex traits
- Using autism spectrum disorder (ASD) cohorts, we found that SNPs associated with higher direct EA are over transmitted from healthy parents to their ASD offspring.

Practical significance

- A traditional GWAS approach in conjunction with phenotypic data on participants' children could greatly benefit studies on the effects genetic nurture on human complex traits.

WISCONSIN UNIVERSITY OF WISCONSIN-MADISON

LU LABORATORY
statistical genetics & genome informatics

GWAS summary statistics with a multi-generational design to estimate the direct and indirect effects

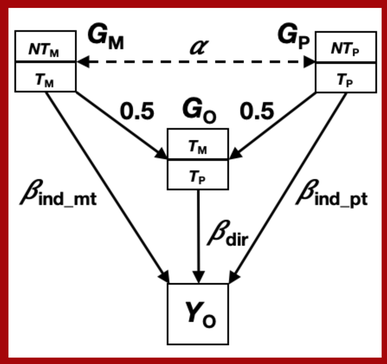

Figure 1. Schematic diagram of direct and indirect genetic effects. $G_{M,P,O}$ represents the maternal, paternal, and offspring genotype, respectively. α is the correlation between spousal genotypes at a locus. Y_O is the offspring phenotype. *T* and *NT* represent transmitted and non-transmitted alleles from a parent to the offspring, respectively. $\beta_{dir,ind\_pt,ind\_mt}$ are direct, indirect paternal, and indirect maternal effect sizes on the offspring's phenotype. Using summary statistics from marginal regressions of $Y_O \sim G_{O,M,P}$ we could also estimate the direct and indirect effects.

Direct and indirect effects on EA showed distinct patterns of genetic correlations with 45 complex traits

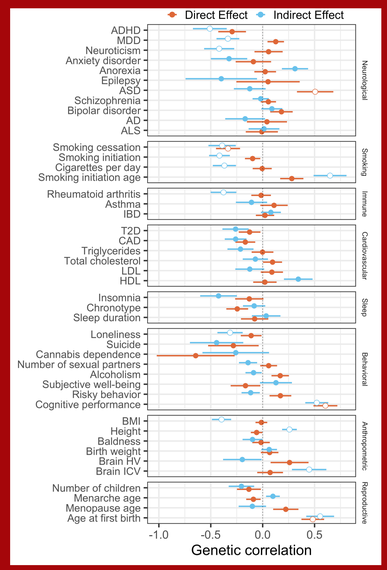

Figure 2. Genetic correlations of EA (direct and indirect effects) with 45 complex traits using LDSC[3]. Dots and intervals indicate point estimates and standard errors, respectively. Significant correlations at false discovery rate (FDR) cutoff of 0.05 are highlighted with white circles. ADHD: attention deficit/hyperactivity disorder; MDD: major depressive disorder; ASD: autism spectrum disorder; AD: Alzheimer's disease; ALS: amyotrophic lateral sclerosis; IBD: inflammatory bowel disease; T2D: type-2 diabetes; CAD: coronary artery disease; LDL and HDL: low and high-density lipoprotein; BMI: body-mass index; HV: hippocampal volume; ICV: intracranial volume

18 significant genetic correlations in total, 4 of which were with direct effect on EA while 14 were with the indirect effects, which highlighted the substantial contribution of genetic nurture on the etiologic sharing among complex traits.

SNPs associated with higher direct EA are over transmitted from parents to autism spectrum disorder offspring

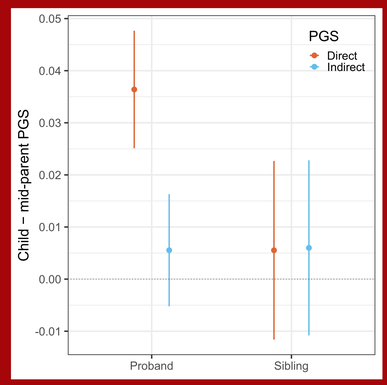

Figure 3. Polygenic transmission disequilibrium test (pTDT) for direct and indirect EA PGSs. 7,804 probands, 3,242 siblings and their parents were used from the Autism Genome Project (AGP; 2,188 probands), Simons Simplex Collection (SSC; 1,794 probands and 1,430 siblings), and Simons Foundation Powering Autism Research for Knowledge (SPARK; 3,822 probands and 1,812 siblings) cohorts. All probands are ASD cases while their siblings and parents do not have ASD. Dots are the mean differences between children and mid-parent PGSs and intervals indicate one standard error. $p = 1.25\times10^{-3}$ for the direct PGS of probands.

Both direct and indirect EA PGSs are predictive between families, direct PGS is more predictive within families

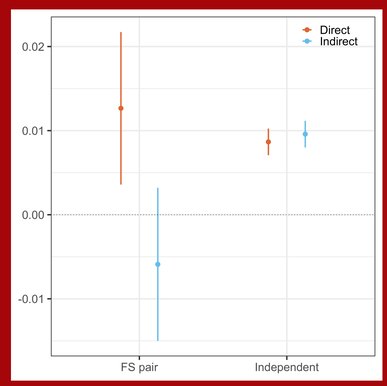

Figure 4. Predictive performances of direct and indirect polygenic scores for EA on European ancestry full siblings (FS; N = 16,580 pairs) and independent samples (N = 370,308) in the UKB. We used PRSice-2 to compute the PGS with a p-value cutoff given by PUMAS. y-axis is the PGS effect size obtained by regressing EA on direct and indirect PGSs. Dots and intervals indicate point estimates and one standard error, respectively.

Both direct and indirect PGSs were significantly associated with EA in independent samples ($p = 4.63\times10^{-8}$ and $p = 1.46\times10^{-9}$) with similar effect sizes. Direct effect PGS for EA was positively associated with EA phenotype in full sibling pairs with an effect size comparable to that in the population. The indirect PGS was negatively correlated with EA in full siblings. Due to limited sample size, neither direct nor indirect PGSs reached statistical significance in FS ($p = 0.16$ and $p = 0.52$).

Method

Full model with trio data:

$$Y_O = G_O\beta_{dir} + G_M\beta_{ind\_mt} + G_P\beta_{ind\_pt} + \varepsilon$$

If instead have 3 GWASs from 3 studies:

- GWAS-O: $Y_O = G_O\beta_O + u_O$
- GWAS-M: $Y_O = G_M\beta_M + u_M$
- GWAS-P: $Y_O = G_P\beta_P + u_P$

Effect sizes from marginal GWASs are linear functions of direct and indirect effect sizes:

$$\hat{\beta}_O = (\tilde{G}_O^T\tilde{G}_O)^{-1}\tilde{G}_O^T\tilde{Y}_O$$

$$= (\tilde{G}_O^T\tilde{G}_O)^{-1}\tilde{G}_O^T(\beta_{dir}\tilde{G}_O + \beta_{ind\_mt}\tilde{G}_M + \beta_{ind\_pt}\tilde{G}_P + \tilde{\varepsilon}_O)$$

$$\xrightarrow{d} \beta_{dir} + \beta_{ind\_mt}\frac{Cov(G_O, G_M)}{Var(G_O)} + \beta_{ind\_pt}\frac{Cov(G_O, G_P)}{Var(G_O)}$$

$$= \beta_{dir} + \beta_{ind\_mt}\frac{1+\alpha}{2+\alpha} + \beta_{ind\_pt}\frac{1+\alpha}{2+\alpha}$$

$$\beta_{dir} = (2+\alpha)\beta_O - \beta_M - \beta_P$$

$$\beta_{ind\_mt} = \frac{3-\alpha^2}{2(1-\alpha^2)}\beta_M + \frac{1-2\alpha-\alpha^2}{2(1-\alpha^2)}\beta_P - \left(1+\frac{\alpha}{2}\right)\beta_O$$

$$\beta_{ind\_pt} = \frac{3-\alpha^2}{2(1-\alpha^2)}\beta_P + \frac{1-2\alpha-\alpha^2}{2(1-\alpha^2)}\beta_M - \left(1+\frac{\alpha}{2}\right)\beta_O$$

$$\beta_{ind} = \frac{\beta_{ind\_mt} + \beta_{ind\_pt}}{2}$$

Our results for birth weigth are consistent with those from individual-level data

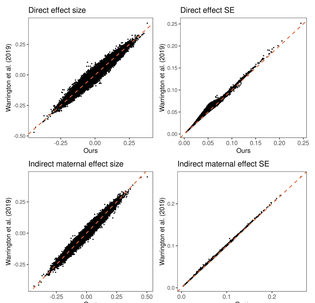

Figure S1. Comparisons of direct and indirect maternal effect sizes and standard errors given by our method vs. those by Warrington et al.[1]

Our method could correct for sample overlap

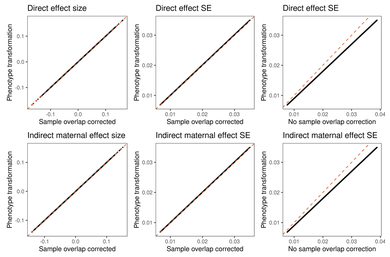

Figure S2. Comparisons of our method vs. phenotype transformation method[1] to dissect direct and indirect maternal genetic effects on birth weight to account for sample overlap. 75,711 independent females of European ancestry in UKB reported both their own and their oldest child's birth weights. They were used in both GWAS-O and GWAS-M creating a complete sample overlap scenario.

References

[1] Warrington, N.M. *et al.* Maternal and fetal genetic effects on birth weight and their relevance to cardio-metabolic risk factors. *Nature Genetics* **51** (2019).
[2] Lee, J.J. *et al.* Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature Genetics* **50** (2018).
[3] Bulik-Sullivan, B. *et al.* An atlas of genetic correlations across human diseases and traits. *Nature Genetics* **47** (2015).

About me

Yuchang Wu, PhD is a Postdoctoral Fellow at University of Wisconsin–Madison. His research interest is statistical analysis in human genetics.

@yy_stat