Local genetic correlation analysis reveals heterogeneous etiologic sharing of complex traits

Yiliang Zhang[1,#], **Qiongshi Lu**[2,3,#], Yixuan Ye[4], Kunling Huang[3], Wei Liu[4], Yuchang Wu[2], Xiaoyuan Zhong[2], Boyang Li[1], Zhaolong Yu[4], Brittany G. Travers[5], Donna M. Werling[6], James J. Li[7], Hongyu Zhao[1,4]

1 Department of Biostatistics, Yale School of Public Health, New Haven, CT, USA 06510
2 Department of Biostatistics and Medical Informatics, University of Wisconsin-Madison, Madison, WI, USA 53706
3 Department of Statistics, University of Wisconsin-Madison, Madison, WI, USA 53706
4 Program of Computational Biology and Bioinformatics, Yale University, New Haven, CT, USA 06510
5 Occupational Therapy Program in the Department of Kinesiology, University of Wisconsin-Madison, Madison, WI, USA 53706
6 Laboratory of Genetics, University of Wisconsin-Madison, Madison, WI, USA 53706
7 Department of Psychology, University of Wisconsin-Madison, Madison, WI, USA 53706
# These authors contributed equally to this work

Background

- Local genetic correlation quantifies the genetic similarity of complex traits in specific genomic regions, which could shed unique light on etiologic sharing and provide additional mechanistic insights into the genetic basis of complex traits compared to global genetic correlation.
- Accurate estimation of local genetic correlation remains challenging, in part due to extensive linkage disequilibrium in local genomic regions and pervasive sample overlap across studies.

Key contribution

- We introduce SUPERGNOVA, a unified framework to estimate both global and local genetic correlations using GWAS summary statistics.

Summary of results

- SUPERGNOVA substantially outperforms other methods in estimation accuracy and power.
- SUPERGNOVA uses GWAS summary statistics as the input and is robust to arbitrary sample overlap between GWAS datasets.
- Applied to 30 complex traits, SUPERGNOVA identified 150 trait pairs with significant local genetic covariance, including 86 pairs without a significant global correlation.
- The literature on ASD genetics seemed paradoxical – WES reveals that genetics contribute to ASD and intellectual disability while GWAS suggests that genetics contribute to ASD and higher IQ.
- We show that the positive, consistently-identified, yet paradoxical genetic correlation between autism spectrum disorder and cognitive performance could be explained by two etiologically-distinct genetic signatures with bidirectional local genetic correlations.

SUPERGNOVA provides robust estimation for local genetic covariance while accounting for LD and sample overlap

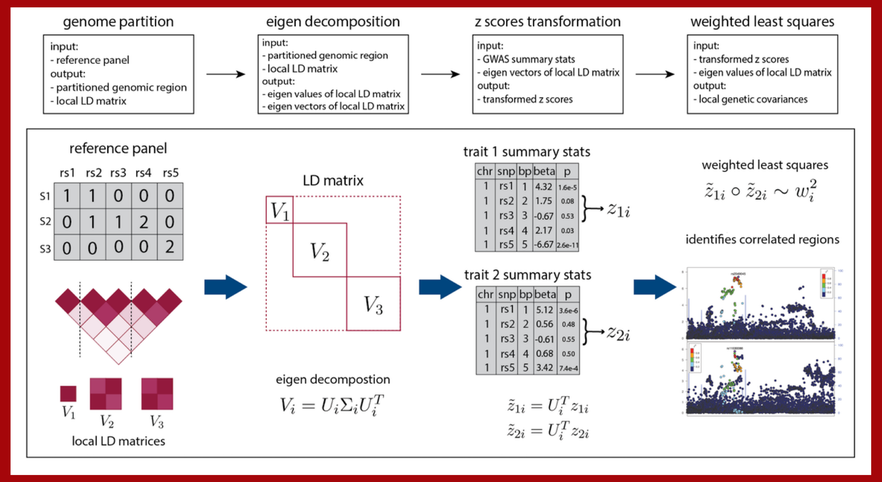

Figure 1. SUPERGNOVA workflow. Details on the statistical framework areAgain, METAL does not report A1/A2 based on allele frequency. I believe it's based on described in the Methods section. w_i denotes the diagonal elements of Σ_i, which are also the eigenvalues of each local LD matrix. Notation, ∘ in the last step indicates the element-wise product.

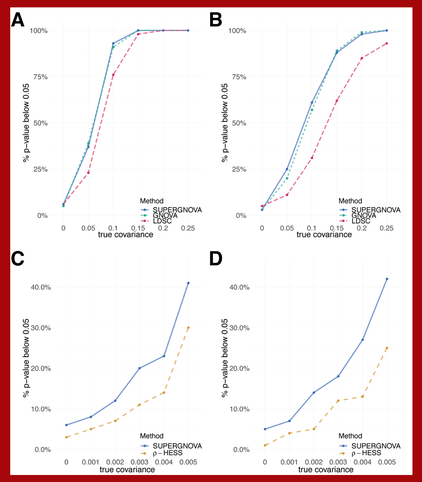

Figure 2. Simulation results. (A-B) Global genetic covariance estimation. (C-D) Local genetic covariance estimation. GWASs had a full sample overlap in panels B and D.

Global and local genetic correlations of 30 complex traits

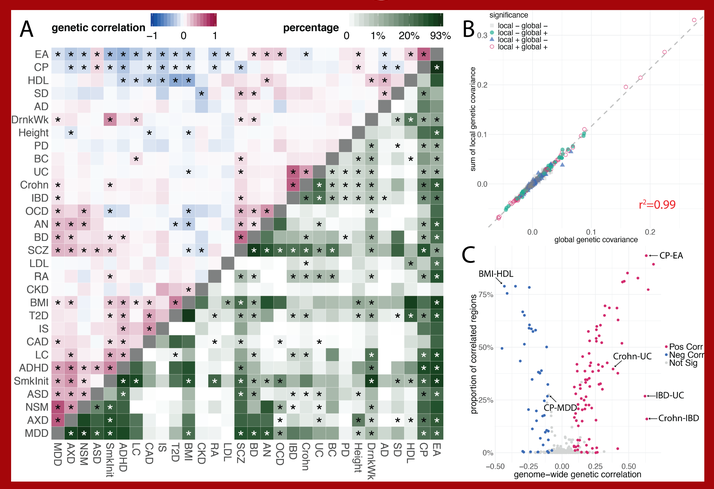

Figure 3. Global and local genetic correlations among 30 complex traits. (A) Estimates of global genetic correlations (upper triangle) and estimated proportions of correlated regions among 435 trait pairs (lower triangle). Asterisks in the upper triangle highlight significant genetic correlations after Bonferroni correction for 435 pairs. Asterisks in the lower triangle indicate at least one significantly correlated region between the traits after Bonferroni correction for all 1,006,072 regions in 435 trait pairs. (B) Global genetic covariance estimates were highly concordant with the sums of local genetic covariance. (C) Volcano plot comparing the global genetic correlation and proportion of correlated local regions. Color of each data point represents the significance and direction of global correlation.

The shared genetic basis of ASD and cognitive ability

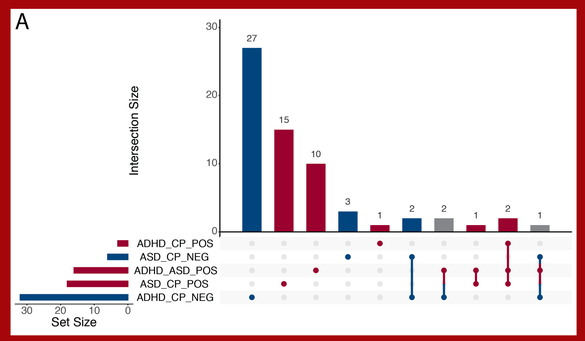

Figure 4. Bidirectional local genetic covariance between ASD and CP. (A) Regions with significant local genetic covariance among ADHD, ASD, and CP (FDR < 0.1). We use different colors (red, blue, and gray) to annotate region categories of positive, negative, and mixed covariance directions.

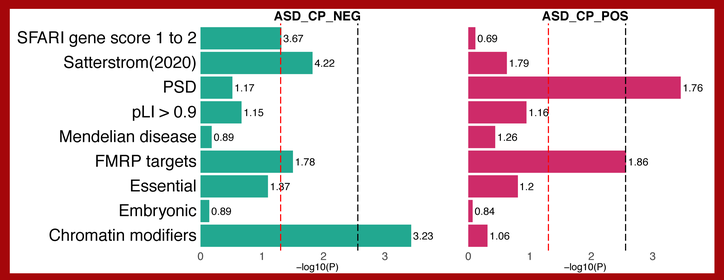

Figure 5. Enrichment for gene sets in correlated regions between ASD and CP. Regions with opposite correlations between ASD and CP were enriched for different mechanistic pathways. Fold enrichment values are labeled next to each bar. The red dashed lines mark the p-value cutoff of 0.05 and the black dashed lines denote the p-value thresholds after Bonferroni correction (p=2.8e-3).

Methods

The covariance of z_{1i} and z_{2i} (i.e. z-scores of trait 1 and trait 2 in region i) is

$$Cov(z_{1i}, z_{2i}) = \frac{\sqrt{n_1 n_2}\rho_i}{m_i} V_i^2 + \frac{n_s \rho_t}{\sqrt{n_1 n_2}} V_i$$

where V_i denotes the LD matrix in region i. Assume eigen decomposition of V_i is $V_i = U_i \Sigma_i U_i^T$, then we have

$$Cov(U_i^T z_{1i}, U_i^T z_{2i}) = \frac{\sqrt{n_1 n_2}\rho_i}{m_i} \Sigma_i^2 + \frac{n_s \rho_t}{\sqrt{n_1 n_2}} \Sigma_i$$

where $\Sigma_i = diag(w_{i1}, w_{i2}, \dots, w_{im_i})$ $(w_{i1} \geq w_{i2} \geq \cdots \geq w_{im_i} \geq 0$ are the eigenvalues of $\Sigma_i)$ and U_i is the corresponding orthogonal matrix of eigenvectors. Denote $\tilde{z}_{1i} = U_i^T z_{1i}$ and $\tilde{z}_{2i} = U_i^T z_{2i}$. For $j = 1, 2, \dots, m_i$, the expected value and variance of $\tilde{z}_{1ij}\tilde{z}_{2ij}$ for the jth eigenvalue w_{ij} are

$$E[\tilde{z}_{1ij}\tilde{z}_{2ij}] = \frac{\sqrt{n_1 n_2}\rho_i}{m_i} w_{ij}^2 + \frac{n_s \rho_t}{\sqrt{n_1 n_2}} w_{ij}$$

For each genomic region, we can estimate local genetic covariance and test the significance of $\hat{\rho}_i$ using the weighted regression of $\tilde{z}_{1ij}\tilde{z}_{2ij} - \left(\widehat{n_s \rho_t}/\sqrt{n_1 n_2}\right) w_{ij}$ on the square of eigenvalue weighted by the reciprocal of the variance (**Figure 1**).

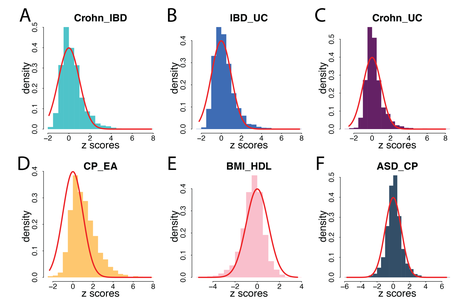

Figure S1. Histograms of z scores of local genetic covariance. The red lines represent the density function of a standard normal distribution. The trait pairs are (A) Crohn-IBD, (B) IBD-UC, (C) Crohn-UC, (D) CP-EA, (E) BMI-HDL, and (F) ASD-CP.

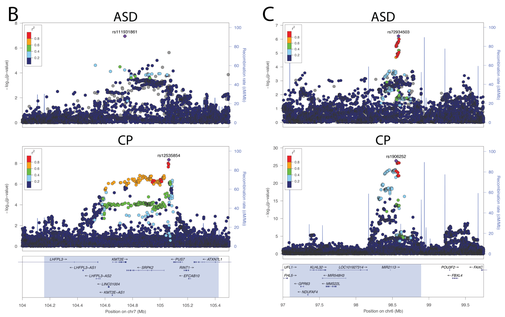

Figure S2. LocusZoom plots for ASD and CP GWAS associations at the *KMT2E* and *POU3F2* loci.

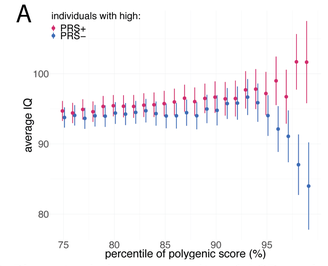

Figure S3. Phenotypic heterogeneity of ASD probands with high PRS+ and PRS-. Average IQ is computed for different groups defined by PRS. Each interval indicates standard error of the estimated mean.

Reference

[1] Zhang et al. (2020). Local genetic correlation analysis reveals heterogeneous etiologic sharing of complex traits. *bioRxiv*