The American Journal of Human Genetics, Volume *104*

Supplemental Data

Genes with High Network Connectivity Are Enriched for Disease Heritability

Samuel S. Kim, Chengzhen Dai, Farhad Hormozdiari, Bryce van de Geijn, Steven Gazal, Yongjin Park, Luke O'Connor, Tiffany Amariuta, Po-Ru Loh, Hilary Finucane, Soumya Raychaudhuri, and Alkes L. Price

Supplementary figures

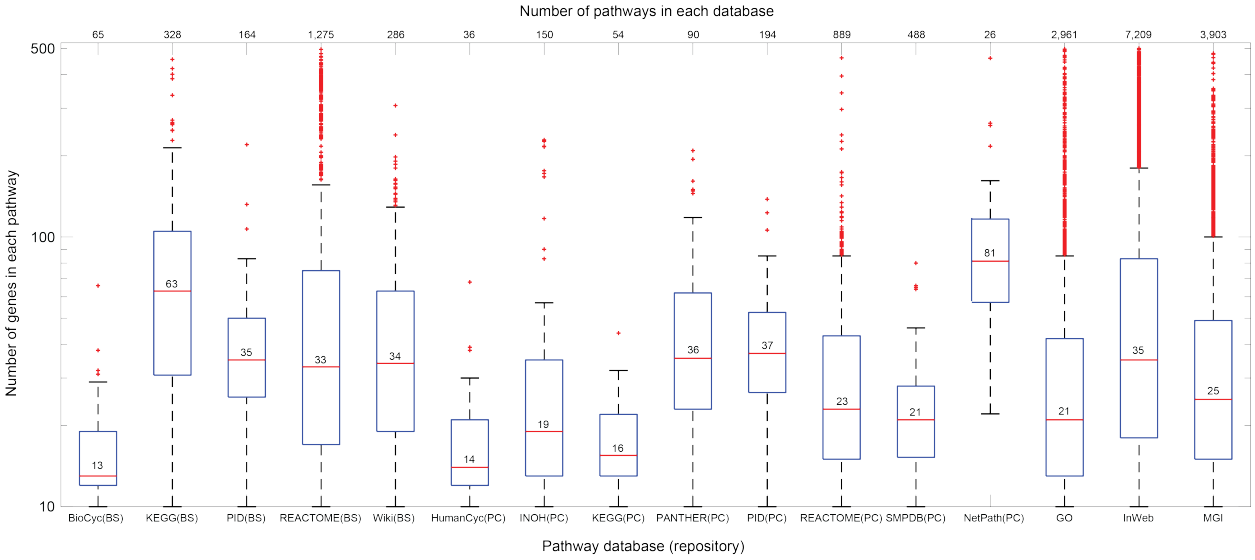

Figure S1. Summary of pathways analyzed. We analyzed 18,119 pathways, each consisting of at least 10 and at most 500 coding genes. Number in the boxplot represents the median number of genes for each database. Number of pathways in each database is shown on the top of the figure. BS: NCBI BioSystems. PC: PathwayCommons. MGI: Mouse Genome Informatics. GO: Genome Ontology. InWeb: InWeb protein-protein interactions. See Table S1 for a description of 18,119 pathways analyzed.

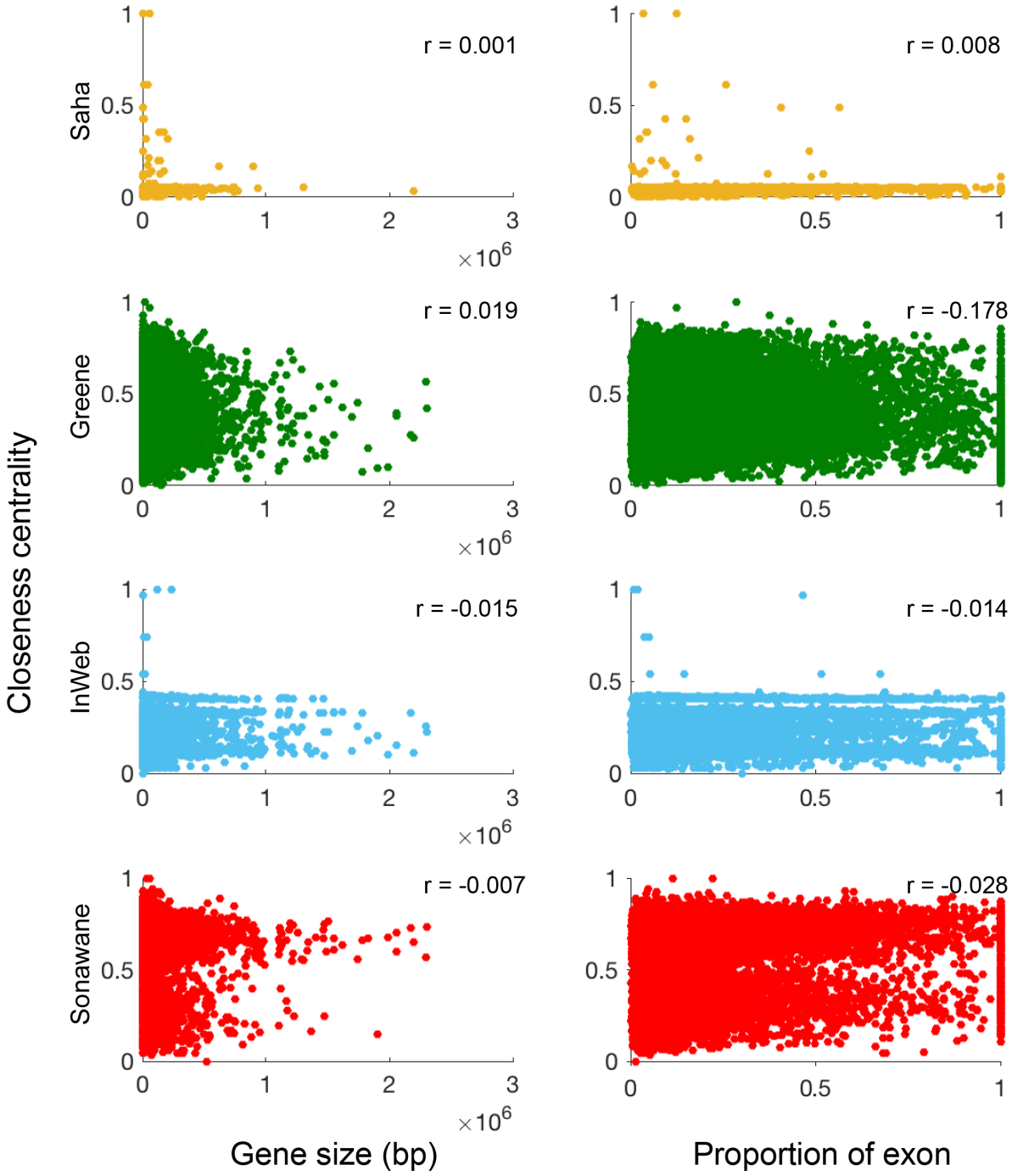

Figure S2. Closeness centrality is independent from the gene size and the proportion of exon. For each of four network annotations, we computed a Pearson correlation between probabilistic annotation values and (1) gene size and (2) proportion of exon. We calculated the proportion of exon as the size of coding regions (bp) that lie inside the gene divided by the size of the gene, defined as (transcription stop - transcription start).

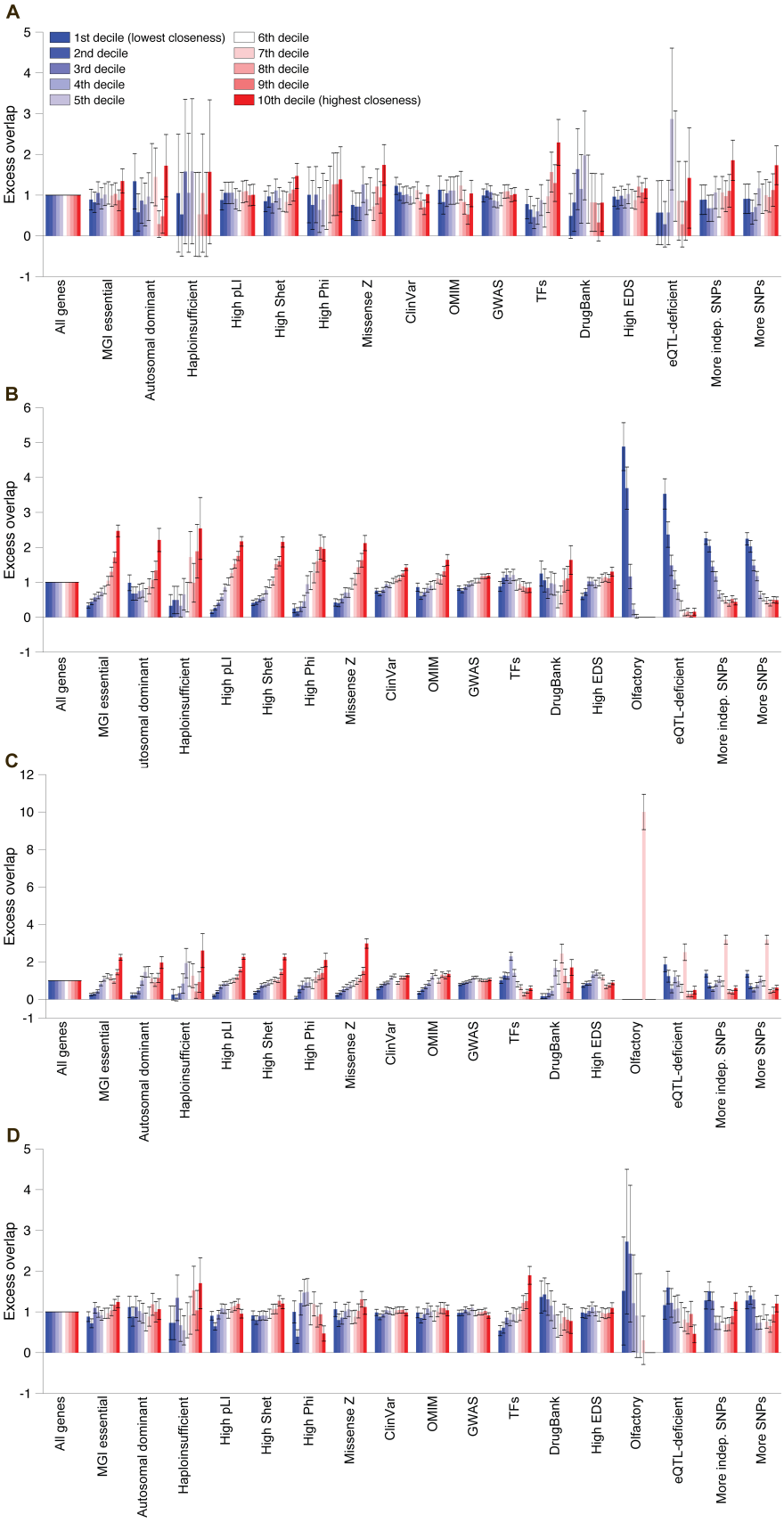

Figure S3. Excess overlap of 18 genes in each decile bin of closeness centrality. For each of 18 gene sets, we report the excess overlap of genes in each decile bin of closeness centrality for (A) Saha skin network, (B) Greene thyroid network, (C) InWeb network, and (D) Sonawane testis network. Error bars represent 95% confidence intervals. There is no olfactory receptor gene in Saha network.

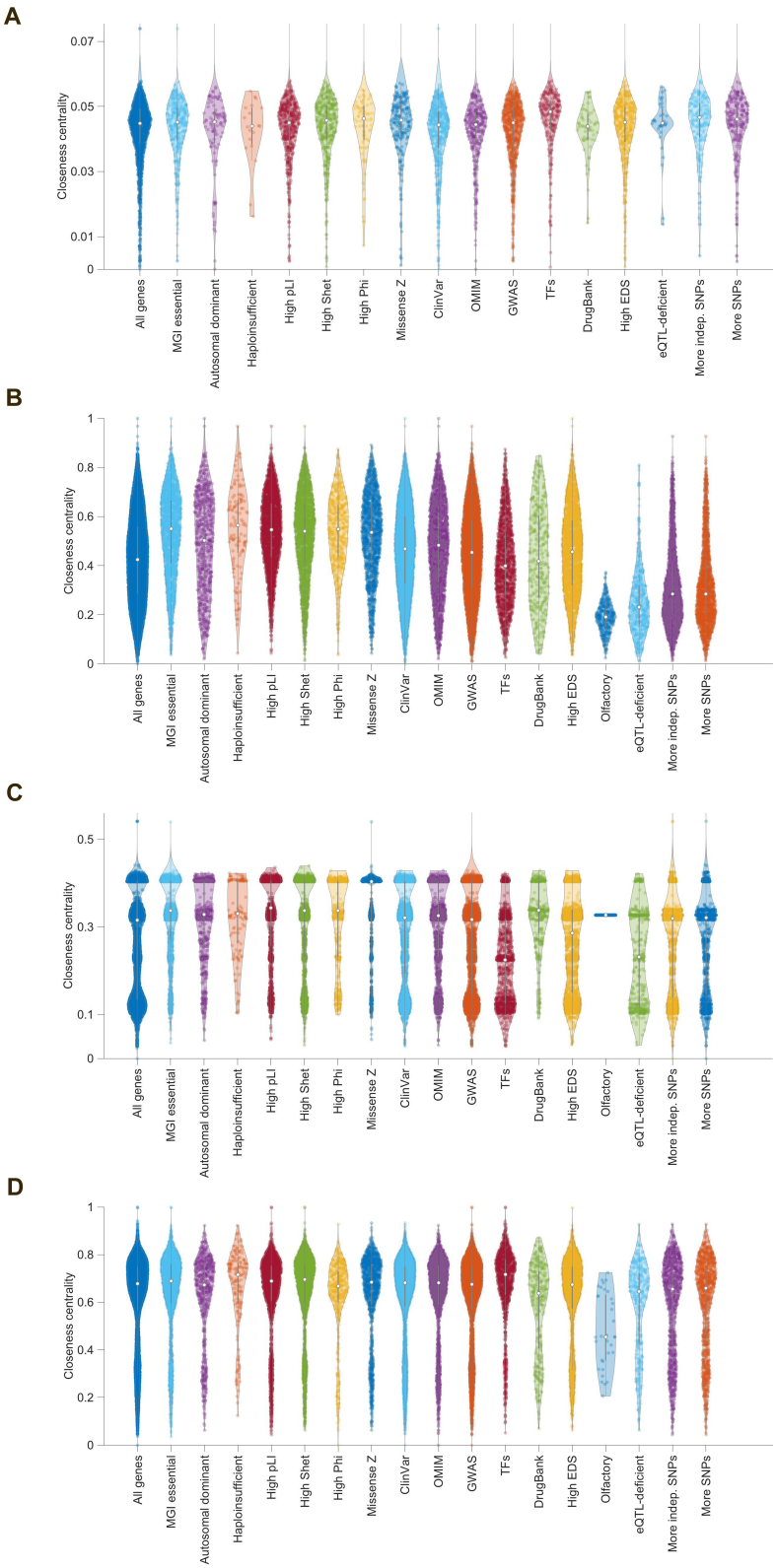

Figure S4. Distribution of closeness centrality in 18 gene sets. For each of 18 gene sets, we show the distribution of closeness for (A) Saha skin network, (B) Greene thyroid network, (C) InWeb network, and (D) Sonawane testis network. Box plot is shown as a grey line inside the violin plot where white dot represents median. Each colored dot represents a gene. Outliers are not displayed for Saha and InWeb (see Web Resources for closeness scores)

Supplementary tables

See Excel file for all supplementary tables. Titles and captions are provided below.

Table S1. List of 18,119 pathways analyzed. For each pathway, we report a pathway ID, pathway description, database, Entrez IDs for genes, and the number of protein-coding genes.

Table S2. List of 42 independent diseases and traits analyzed. For each trait, we report a trait identifier, trait description, reference, sample size, and heritability z-score. We selected these 42 traits based on a heritability z-score > 6 (see Material and Methods). We further indicated brain or blood-related traits.

Table S3. Lists of genes in 18 gene sets compared with closeness centrality. We compiled lists of genes for 18 metrics (gene sets) that we compared with closeness centrality. We report Entrez IDs for genes in the 18 gene sets.

Table S4. S-LDSC results of all pathway-trait pairs We applied S-LDSC to 760,869 pathway-trait pairs, conditioning on all-genes annotation and the baseline-LD model. For each pathway-trait pair, we report a proportion of SNPs, enrichment, and τ.

Table S5. S-LDSC results for 156 significantly enriched pathway-trait pairs. For each significantly enriched pathway-trait pair, we report a proportion of SNPs, enrichment, and a τ^*. The 8 significant pathway-trait pairs were reported in previous genetic studies: "pathways in cancer" for height[109]; "neuropeptide hormone activity" for BMI[110]; "immune response" for both Crohn's disease and ulcerative colitis[89]; "T-cell receptor," "abnormal T-cell physiology," and "cytokine-mediated signaling pathway" for rheumatoid arthritis[90]; "absent corpus callosum" for years of education[92].

Table S6. Average gene size of annotations. For all-genes annotation, all pathways, pathway, network, and pathway+network, and Quack annotations, we report an average size of genes (and its standard deviation) and an average number of genes.

Table S7. Heritability enrichment of enriched pathway-trait pairs. We meta-analyzed (A) 156 enriched pathway-trait pairs; (B) 13 enriched pathway-trait pairs for ExAC, Cassa, and Samocha gene sets; (C) 169 enriched pathway-trait pairs (a and b combined). In each case, we report meta-analyzed enrichments and τ^*.

Table S8. S-LDSC results of 156 enriched pathway-trait pairs excluding genes implicated by GWAS. We removed genes implicated by previous GWAS studies (see Material and Methods; average of 5% of genes (2 genes) removed) and applied S-LDSC conditional on all-genes and baseline-LD model annotations. For each pathway-trait pair, we report a proportion of SNPs, enrichment, τ^*, and the number of genes in a pathway excluding GWAS significant genes.

Table S9. S-LDSC results of 195 pathway-trait pairs from previous pathway enrichment studies. We applied S-LDSC to 95 pathway-trait pairs from five previous genetic studies[2,33–36] and 100 from a recent study[37] (A) conditioning on the baseline-LD model and all-genes annotation and (B) conditioning on all-genes annotation only. In each pathway-trait pairs for each case, we report a proportion of SNPs, enrichment, and τ. We assessed the statistical significance based on global FDR $< 5\%$ across 18,119 pathways tested ($\tau^* < 0.000989$).

Table S10. S-LDSC results of 13 enriched pathway-trait pairs for ExAC, Cassa, Samocha gene sets. For each of 13 enriched pathway-trait pairs, we report a proportion of SNPs, enrichment, and τ.

Table S11. Correlation of network annotations with baseline-LD model annotations. We report the Pearson correlation between baseline-LD model annotations and (A) Saha network annotations of different centralities, (B) Saha network annotations of different tissues, (C) Greene network annotations of different centralities, (D) Greene network annotations of different tissue, (E) InWeb network annotations of different centralities, (F) Sonawane network annotations of different centralities, and (G) Sonawane network annotations of different tissue.

Table S12. Summary of gene networks analyzed. We report the number of genes, the number of edges, and the distribution of edge weights for each of four networks (InWeb, Saha, Sonawane, Greene).

Table S13. Excess overlap of 18 genes in each decile bin of closeness centrality. For each of 18 gene sets, we report the excess overlap (and standard error) of genes in each decile bin of closeness centrality for (A) Saha skin network, (B) Greene thyroid network, (C) InWeb network, and (D) Sonawane testis network.

Table S14. Per-gene closeness centrality scores and gene membership in 18 gene sets We report the closeness centrality for all protein-coding genes that exist in each of (A) Saha skin network, (B) Greene thyroid network, (C) InWeb network, and (D) Sonawane testis network. We indicate gene membership in each of 18 gene sets, marking '1' if in the corresponding gene set.

Table S15. Correlation between closeness centrality and 18 gene sets. We report Pearson correlations between closeness centrality and 18 gene sets analyzed for (A) Saha skin network, (B) Greene thyroid network, (C) InWeb network, and (D) Sonawane testis network.

Table S16. Excess fold overlap of network and pathway+network annotations with baseline-LD model annotations. We report the excess fold overlap between baseline-LD model annotations and network, pathway+network, Quack, and all-genes annotations.

Table S17. Correlation of network and pathway+network annotations with baseline-LD model annotations. We report the Pearson correlation between baseline-LD model annotations and network, pathway+network, Quack, and all-genes annotations.

Table S18. Excess fold overlap / correlation among functional annotations from the baseline-LD model. We report (A) excess fold overlap and (B) correlation among functional annotations from the baseline-LD model.

Table S19. TFs enriched in high closeness centrality genes. For (A) Saha, (B) Greene, (C) InWeb, (D) Sonawane networks, for high closeness centrality genes (top decile), we report significantly enriched TFs (Benjamini-Hochberg adjusted p-value < 0.05). The description of TFs is provided in the ENCODE Chip-Seq Significance Tool (see Web Resources).

Table S20. Correlation of deciles of closeness centrality with baseline-LD model annotations. For four networks, we constructed binarized network annotations based on deciles of closeness centrality. We report the Pearson correlation between these annotations and baseline-LD model annotations.

Table S21. Enriched GO terms of high closeness centrality genes. For (A) Saha, (B) Greene, (C) InWeb, (D) Sonawane networks, we report significantly enriched GO terms in the following GO categories: biological process (BP), cellular component (CC), and molecular function (MP) (Benjamini-Hochberg adjusted p-value < 0.05).

Table S22. Heritability enrichment of network annotations. For each of 4 network annotations, we report meta-analyzed enrichments and τ^* across 42 independent traits. We highlight the network attaining highest enrichment for each trait. For the three tissue-specific networks (Saha, Greene, Sonawane), we also report meta-analyzed enrichments and τ^* of network annotations constructed using the tissue that maximized the excess overlap with the High pLI (ExAC) gene set.

Table S23. Heritability enrichment of network annotations conditioning on one annotation from the baseline-LD at a time. For (A) Saha, (B) Greene, (C) InWeb, (D) Sonawane networks, we meta-analyzed network annotations across 42 independent traits, conditioning on one annotation from the baseline-LD model at a time. We report meta-analyzed enrichments and τ^* across 42 independent traits. We highlighted annotations that significantly reduced τ^* (using Bonferroni-corrected p-val).

Table S24. Heritability enrichment of deciles of closeness centrality. For four networks, we constructed binarized network annotations based on deciles of closeness centrality. We applied S-LDSC and meta-analyzed results across 42 independent traits. We report meta-analyzed enrichments and τ^*. For the three tissue-specific networks (Saha, Greene, Sonawane), we also report meta-analyzed enrichments and τ^* of binarized network annotations constructed using the tissue that maximized the excess overlap with the High pLI (ExAC) gene set.

Table S25. Heritability enrichment of network annotations from network perturbation analysis. We randomly removed 10% to 90% of edges from the original networks and computed closeness centrality on networks with edges removed; we performed five separate perturbation analyses for each value of the proportion of edges removed. We applied S-LDSC and meta-analyzed results across 42 independent traits. We report meta-analyzed enrichments and τ^*.

Table S26. Heritability enrichment of DSD-network annotations. We applied the diffusion state distance (DSD) algorithm[84] to transform gene networks' edge weights with a random walk (k = 5). Then, we constructed network annotations by re-computing closeness on DSD-transformed networks and meta-analyzed results across 42 independent traits. We report meta-analyzed enrichments and τ^*.

Table S27. Heritability enrichment of networks annotations from consensus networks. We constructed consensus networks and made (A) probabilistic annotations based on closeness centrality and (B) binary annotations based on deciles of closeness centrality. We applied S-LDSC and meta-analyzed results across 42 independent traits. We report meta-analyzed enrichments and τ^*.

Table S28. Correlation between closeness and gene expression. For (A) Saha, (B) Greene, (C) InWeb, (D) Sonawane networks, we report the correlation between closeness centrality and gene expression across 53 GTEx tissues.

Table S29. Heritability enrichment of Saha TSN and TWN gene sets. We constructed gene sets based on membership of genes in Saha tissue-specific networks (TSN) and transcriptome-wide networks (TWN). We applied S-LDSC to 36 TSN and 16 TWN gene sets across 42 independent traits. We report meta-analyzed enrichments and τ^*.

Table S30. Excess overlap of top deciles of closeness centrality of tissue-specific networks with High pLI (ExAC) genes. For each tissue-specific network (Saha, Greene, Sonawane), for each tissue, we report the excess overlap between High pLI (ExAC) genes and the top decile of closeness centrality.

Table S31. Heritability enrichment of network annotations using relevant tissues for brain-related and blood-related traits. For (A) 8 brain-related traits and (B) 10 blood-related traits, we report meta-analyzed enrichments and τ^*.

Table S32. Heritability enrichment of network annotations using 6 other network centrality metrics. For (A) Saha, (B) Greene, (C) InWeb, (D) Sonawane networks, we constructed network annotations based on 6 other network centrality metrics and meta-analyzed results across 42 independent traits. We report meta-analyzed enrichments and τ^*.

Table S33. Heritability enrichment of network annotations with different window sizes. Instead of 100kb windows around genes, we added (A) 10kb or (B) 1Mb windows around genes when constructing network annotations. We report meta-analyzed enrichments and τ^* across 42 independent traits.

Table S34. Heritability enrichment of pathway+network annotations. We meta-analyzed 156 pathway-trait pairs (122 for Saha, which has less pairs as all genes in some pathways do not exist in the Saha network). For each 590 pathway-trait pair, we report a proportion of SNPs, enrichment, and τ. We also report meta-analyzed enrichments and τ^* across 156 pathway-trait pairs (122 for Saha). We highlighted the network attaining highest enrichment for each pathway-trait pairs.

Table S35. Heritability enrichment of pathway+network annotations conditioning on one annotation from the baseline-LD at a time. For (A) Saha, (B) Greene, (C) InWeb, (D) Sonawane networks, we constructed an average annotation across 156 pathway+network annotations and meta-analyzed averaged pathway+network annotations across 42 independent traits, conditioning on one annotation from the baseline-LD model at a time. We report meta-analyzed enrichments and τ^* across 42 independent traits.

Table S36. Network connectivity of enriched pathways and null pathways. Using (A) sum of edge weights or (B) number of edges as network connectivity metrics, we report the number of interacting genes and network connectivity between genes in a pathway and neighboring genes outside the pathway. We constructed null pathways in two ways: (1) each gene sampled from a randomly chosen pathway or (2) each gene randomly sampled from all protein-coding genes.

Table S37. Heritability enrichment of Quack annotations. We applied the Quack random-forest classifier algorithm[38]. We used 18,119 pathways as a training data and applied Quack to four gene networks. We used the output of Quack to construct Quack pathway+network annotations (with 100kb window), applied S-LDSC, and meta-analyzed across 156 pathway-trait pairs (122 for Saha). We report meta-analyzed enrichments and τ^*.

Table S38. Heritability enrichment of 53 pathway+network annotation using pathways excluding genes implicatd by GWAS. From 53 significant pathway-trait pairs after excluding GWAS significant genes, we constructed pathway+network annotations and meta-analyzed results across 53 pathway-trait pairs (40 for Saha). We report meta-analyzed enrichments and τ^*.