Neuron

Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk

Highlights

- We test risk gradients in gene co-expression networks based on the omnigenic model
- Across psychiatric disorders, schizophrenia risk best fits the model's predictions
- Findings hold in networks from excitatory neurons and in iPSC-derived neurons
- Prioritizing genes most connected to schizophrenia risk reveals potential targets

Authors

Christopher Borcuk, Madhur Parihar, Leonardo Sportelli, ..., Alessandro Bertolino, Daniel R. Weinberger, Giulio Pergola

Correspondence

drweinberger@libd.org (D.R.W.), giulio.pergola@libd.org (G.P.)

In brief

Genetic risk for schizophrenia spans the entire genome. Borcuk et al. show that network neighbors of risk genes also harbor risk, most evidently in excitatory neurons. They leverage this property to propose potential risk genes that are "guilty by association." Stem cell data support the link between these genes and schizophrenia risk.

Borcuk et al., 2024, Neuron *112*, 1–16
November 6, 2024
https://doi.org/10.1016/j.neuron.2024.08.005

CellPress

Neuron

CellPress
OPEN ACCESS

Article

Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk

Christopher Borcuk,[1,2] Madhur Parihar,[1] Leonardo Sportelli,[1,2] Joel E. Kleinman,[1,3] Joo Heon Shin,[1] Thomas M. Hyde,[1,3,4] Alessandro Bertolino,[2,5] Daniel R. Weinberger,[1,3,4,6,7,*] and Giulio Pergola[1,2,3,8,*]

[1]Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD, USA
[2]Group of Psychiatric Neuroscience, Department of Translational Biomedicine and Neuroscience, University of Bari Aldo Moro, Bari, Italy
[3]Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[4]Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[5]Azienda Ospedaliero-Universitaria Consorziale Policlinico, Bari, Italy
[6]Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[7]Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD, USA
[8]Lead contact
*Correspondence: drweinberger@libd.org (D.R.W.), giulio.pergola@libd.org (G.P.)
https://doi.org/10.1016/j.neuron.2024.08.005

SUMMARY

The omnigenic model posits that genetic risk for traits with complex heritability involves cumulative effects of peripheral genes on mechanistic "core genes," suggesting that in a network of genes, those closer to clusters including core genes should have higher GWAS signals. In gene co-expression networks, we confirmed that GWAS signals accumulate in genes more connected to risk-enriched gene clusters, highlighting across-network risk convergence. This was strongest in adult psychiatric disorders, especially schizophrenia (SCZ), spanning 70% of network genes, suggestive of super-polygenic architecture. In snRNA-seq cell type networks, SCZ risk convergence was strongest in L2/L3 excitatory neurons. We prioritized genes most connected to SCZ-GWAS genes, which showed robust association to a CRISPRa measure of PGC3 regulation and were consistently identified across several brain regions. Several genes, including dopamine-associated ones, were prioritized specifically in the striatum. This strategy thus retrieves current drug targets and can be used to prioritize other potential drug targets.

INTRODUCTION

Schizophrenia (SCZ) is a major psychiatric disorder with a prominent genetic risk component. Before the genome-wide era, some models attempting to explain SCZ's complex heritability proposed oligogenic views of risk.[1,2] Current genome-wide association studies (GWASs) have established that hundreds or perhaps thousands of relatively common variants set the stage for illness. GWASs highlight the likely involvement of synaptic functions, a feature shared by other brain disorders.[3,4] An important caveat is that GWASs identify risk variants and not risk genes. Often, we do not know which genes the risk variants affect, hindering the discovery of mechanisms of illness and novel therapeutics. Thus, "risk genes" have been commonly prioritized as those in loci harboring GWAS significant variants or those harboring a high density of trait-associated variants, even below strict statistical significance. In the following, we define "GWAS genes" as those found in loci above significance after correction for multiple comparisons and "risk genes" as those with a relatively high density of nominally associated variants in their genomic proximity within or outside GWAS significant loci.

Boyle et al.[5] proposed that the genetic architecture of SCZ may be so widespread in the genome that it involves virtually every gene related to brain cell biology. Their "omnigenic model" posits the existence of a relatively few "core genes" whose products directly affect the cellular processes contributing to a phenotype; in contrast, many "peripheral genes" should be more prevalent, indirectly contributing to a phenotype through their effect on core genes. Due to the extensive interactions between genes in multiple signaling pathways, virtually every gene in a relevant cell type will contribute a small increment in risk (i.e., showing non-zero signal in GWASs) by impacting core genes. *In silico* tests have supported an omnigenic model in contrast to a polygenic model of SCZ genome-wide risk by showing that independent of the pathways aggregating risk genes into sets, the predictive power of cumulative scores is above zero even outside any pathway.[6] Rammos et al.[6] also showed that master regulators such as miR-137 were enriched for polygenic signals, suggesting that even in an omnigenic context, genetic risk may

Please cite this article in press as: Borcuk et al., Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.08.005

be denser in gene clusters associated with potential master regulators. This evidence is supported by other studies on miR-137[7,8] and on other risk factors such as C4.[9,10]

How are core genes reflected in the summary statistics from GWASs? The mechanisms of core genes and the variants affecting them could impact multiple biological layers, from DNA structure to transcription to gene product function at the protein level. Here, we consider specifically the role of gene expression. Genetic variants may modulate gene expression and therefore link it to GWAS results via expression quantitative trait loci (eQTLs). In this regard, any direct regulation of expression of hypothetical core genes, for instance, associated with variants within or adjacent to a core gene itself (*cis*) would likely result in a relatively larger GWAS weight compared with loci without these properties. In this scenario, sets of GWAS genes are likely enriched for core genes. Following this model, fine-mapping efforts and transcriptome-wide association study (TWAS) approaches have combined *cis*-eQTLs with GWAS summary statistics to identify potential core genes.[11,12]

In contrast, the *trans* regulation of transcription supports the effect of genetic variation on peripheral genes potentially interacting with core genes and is, in fact, a centerpiece of the proposed omnigenic model.[13] Approximately 70% of gene expression heritability is estimated to depend on *trans* regulation, with up to only 30% on *cis* regulation.[13] While the *cis* signal is local, the *trans* signal is blurred across the entire genome—hence, more likely omnigenic. To appreciate this point, suppose a variant mapping in a peripheral gene is statistically associated with the expression of a core gene and is, e.g., a *cis*-eQTL for the peripheral gene and therefore a *trans*-eQTL for the core gene. In such a case, the peripheral gene would gain a relatively higher GWAS statistic compared with peripheral genes lacking this *trans*-eQTL property. This example clarifies why, in the presence of *trans* heritability, GWAS loci genes are not necessarily the core genes. If 70% of gene expression depends on *trans* heritability, it will be rather frequent that GWAS genes are not core genes.

To quantify the GWAS signal attributed to variants at the gene level, multi-marker analysis of genomic annotation (MAGMA) assigns a risk value to genes based on the combination of GWAS effect sizes of their *cis*-mapping SNP variants.[14] Notably, these variants are proximal to their physically closest gene but not necessarily eQTLs; if they are functional, they may act via mechanisms other than transcription regulation. In the context of the omnigenic model, we may expect core genes to have a relatively high risk signal as assigned by MAGMA, because their *cis* variants have the potential to directly affect a trait and should be prominent in GWASs.[13] However, in addition, the variants found to be *cis* to peripheral genes, if they are highly co-expressed with core genes, may have a relatively high risk signal. Critically, such risk signal, if based on *trans*-eQTL effects, will be proportional to the biological proximity of the peripheral gene to a core gene. Importantly, because of the three-dimensional properties of DNA, biological proximity and physical proximity may be unrelated.[15] When biological proximity is defined in terms of gene networks, this implies a gradient of risk, at the center of which lie core genes. *Trans*-acting peripheral genes will obtain relatively high risk statistics because they are connected to a core gene, the more so if they are connected to many core genes. In other words, the transcription-regulation-dependent fraction of GWAS signals can be modeled as consisting of a local (*cis*) and a network-dependent (*trans*) component.

Here, we interrogate how GWAS signals are embedded into the architecture of gene co-expression networks. Weighted gene co-expression network analysis (WGCNA)[16] has been used to identify sets of co-expressed genes (modules) that may be enriched for cell type, biological pathways, and traits, such as for SCZ.[17–24] We hypothesize that clusters of co-expressed genes, called modules in the context of WGCNA, when enriched for GWAS or risk genes (as defined above) likely contain core genes. The rationale is that the *cis* component of the GWAS signal favors potential core genes, and if a module is enriched for the GWAS signal, it likely includes core genes. We further hypothesize that genes outside the module have risk signals proportional to the strength of their connectivity with risk-enriched modules (Figure 1A). If modules enriched for genes associated with a trait indeed contain core genes, then the peripheral genes partnering with the core are in the best position to yield relevant *trans* effects. These effects are proportional to the connectivity to core genes. Therefore, a gene's connectivity to a risk-enriched module will be correlated with the risk signal strength of that gene, as assessed via MAGMA (Figure 1B, left). If the module is not enriched for genes associated with a trait, then the relationship between gene-level connectivity in that module and gene-level risk signal should be negligible (Figure 1B, right).

If a trait is omnigenic, then across multiple modules, the gene-level correlation between connectivity and risk signal will be high for risk-enriched modules and low for low-risk modules that are presumably depleted of potential core genes (Figure 1C). The strength of this across-module linear regression between module risk strength and across-network risk convergence is termed "across-network convergence of risk" (ANCR). The first part of this study investigates ANCR in co-expression networks using GWAS statistics for different disorders. High ANCR indicates that the network-driven, *trans*-heritability component of genetic risk is significant for a trait, such that connectivity to risk-enriched modules well represents the distribution of risk signal in the network. Once the presence of high ANCR is established for a certain trait(s), we may attempt to pinpoint genes central to the convergence of risk as those most connected to a set of prioritized genes. These "GWAS-connected" hub genes should lie proximal to the center of ANCR and should therefore have a higher probability of being core genes. At the very least, they are well poised to influence core modules.

In this study we demonstrate that ANCR present in brain co-expression networks is strongest with SCZ and spans most of the network. We then identify "SCZ-connected" genes as those most connected to SCZ GWAS loci[4] and show their robust association to a CRISPR activation (CRISPRa) measure of regulation by SCZ risk genes, with high loss-of-function intolerance. These genes are found consistently across several brain regions or specifically in the striatum (STR), are biologically coherent, and may serve to help us further unravel SCZ molecular mechanisms.

Please cite this article in press as: Borcuk et al., Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.08.005

CellPress
OPEN ACCESS

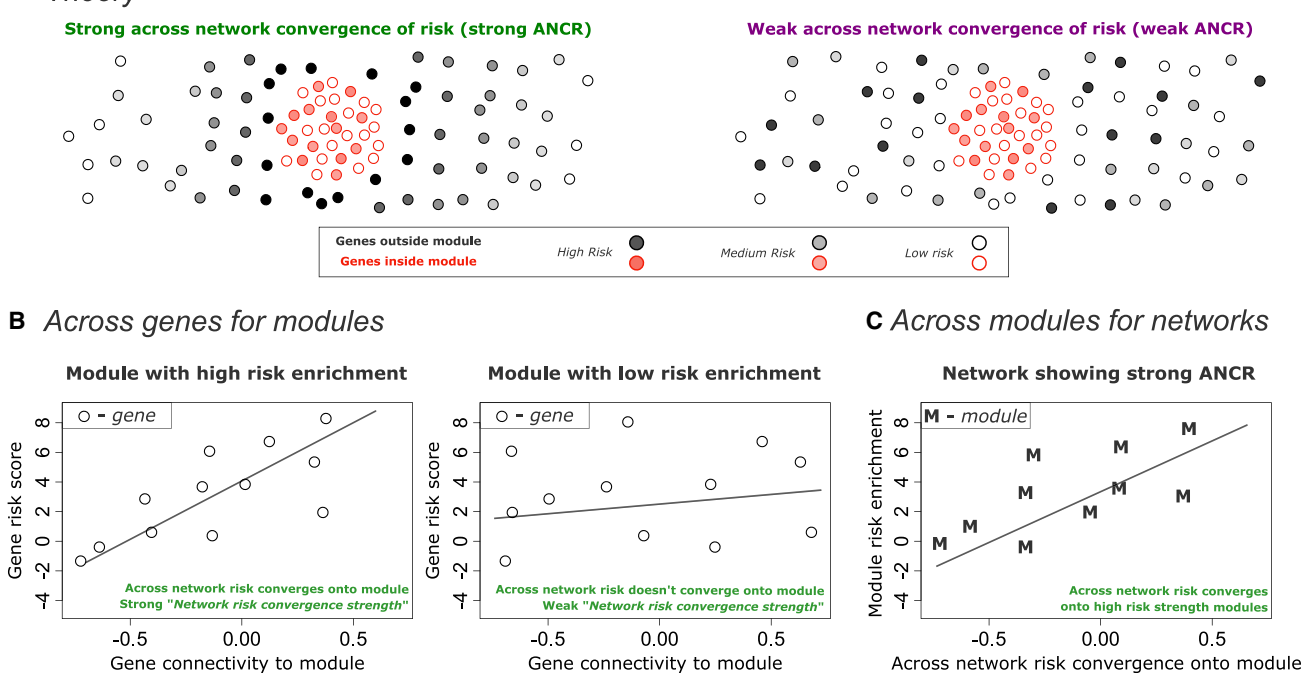

Figure 1. Across-network convergence of risk

(A) Module genes are in red and other network genes in black. Color intensity depicts gene risk scores. The closer, i.e., more connected, a peripheral gene is to the core module, the greater its risk signal.

(B) If module risk enrichment reflects "coreness," modules with higher risk enrichment should have a strong correlation across all genes in the network of connectivity to module vs. gene risk signal. This across-gene correlation strength is thus a per-module index termed here as "network risk convergence strength."

(C) A network whose risk-enriched modules lie at the convergence of an across-network risk gradient should have a strong positive relationship between module risk enrichment and network risk convergence strength. The strength of this across-module association is the basis behind the per-network ANCR index.

ANCR, across-network convergence of risk.

RESULTS

We tested our omnigenic hypothesis using previously published networks for which we had the required statistics available,[8,17–26] as well as novel networks from the subgenual anterior cingulate cortex (sACC) and amygdala (AMY).[27] To compute gene risk scores, we used MAGMA, i.e., a weighted sum of GWAS association with SCZ of the variants harbored in *cis* with a gene.[4,14] We also calculated MAGMA scores for other neuropsychiatric,[28–31] neurological,[32–34] and immune[35,36] disorders. Adult neuropsychiatric disorders, especially SCZ, showed generally increased MAGMA scores across the genome compared to other disorders (Figure S1A). GWAS sample size did not appear as a strong confounder for MAGMA score distributions (Figure S1B). In quantile-quantile (QQ) plots comparing the distribution of MAGMA score quartiles vs. theoretical quartiles, we found that all traits generally have higher MAGMA scores than theoretically predicted at the upper ends of their distributions (Figure S1C). Henceforth, "risk score" refers to gene-level MAGMA scores.

Evaluating across-network convergence onto modules enriched for risk

If gene-level risk is influenced by *trans* relationships to core genes, we expect risk across a co-expression network to converge most strongly onto gene modules containing putative core genes or core pathways, which we operationalized as modules enriched for GWAS risk signals. Thus, we expect the risk enrichment of a module to be associated with the convergence of the risk gradient onto that module. To test this expectation, we used the median risk score of genes within a module to characterize module risk enrichment. This was computed in progressively smaller sets of top risk score genes within a module. We thus estimated whether sets with more or fewer genes best predicted risk gradient convergence. The convergence is computed, for each module, as the association between gene risk and connectivity to that module across all genes in the network. For example, some disorders may require only a few high risk score genes per module to drive the relationship between module risk enrichment and the risk gradient convergence. Other disorders may instead require many genes or show no significant relationship at all (for example, there would be no relationship for a disorder featuring exclusively *cis* heritability). For a schematic understanding, see Figure 2A.

To index gene connectivity to a module, we used the WGCNA measure *kme*, which is the correlation of each gene in the co-expression network (including gray genes) with any module's first principal component. The across-module relationship was computed using a robust linear model controlling for module

Please cite this article in press as: Borcuk et al., Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.08.005

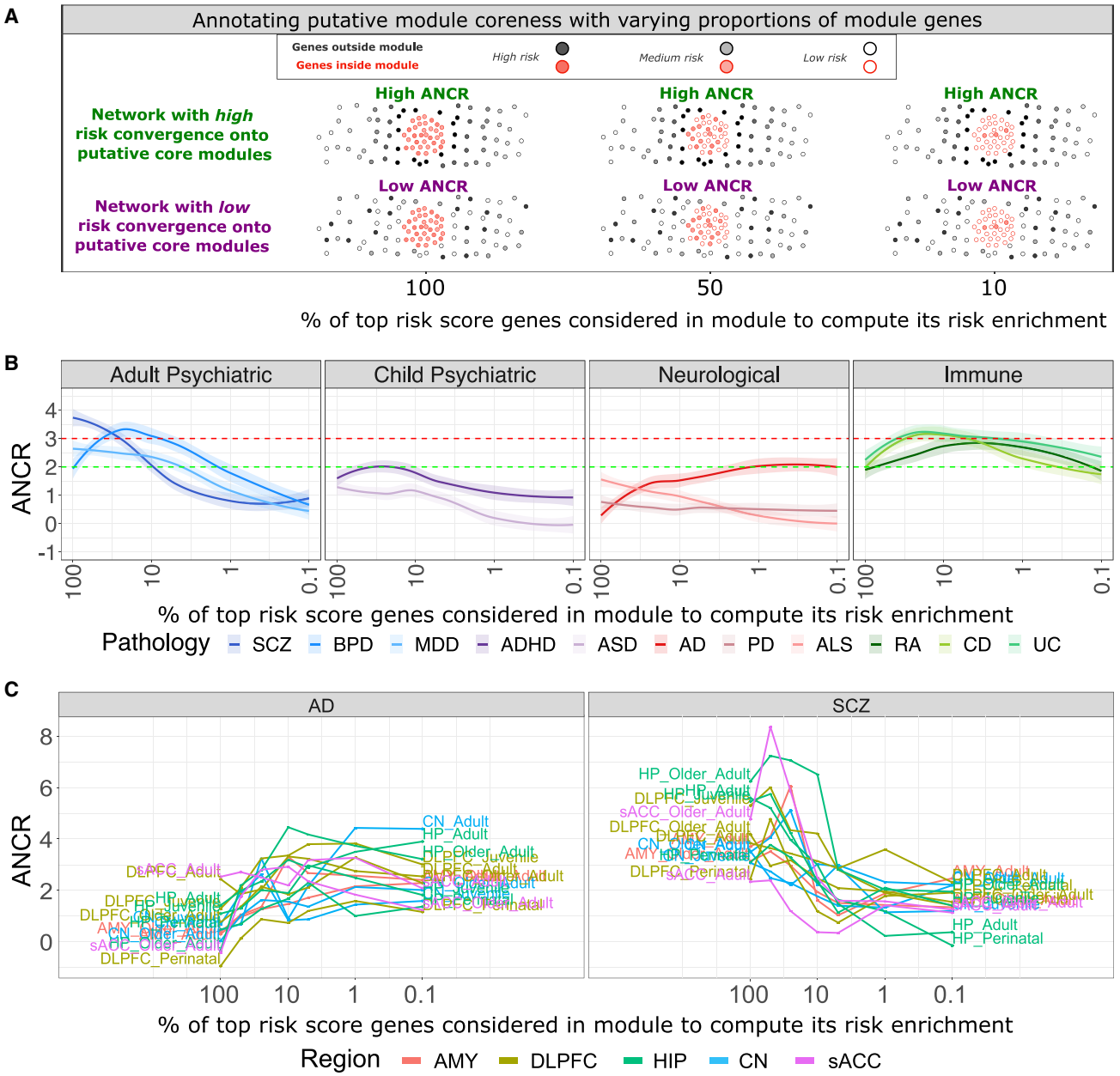

Figure 2. Across-network convergence of risk in gene co-expression networks

(A) The green-labeled network shows high ANCR (color intensity of genes increase with proximity to module) for modules whose risk score strength is spread throughout module genes (left), found within half of module genes (middle), or concentrated in the top few risk genes (right). Bottom violet-labeled network has low or absent ANCR (color intensity is diffuse outside module). See Figure 1A legend for additional details.

(B) Evaluating the proportion of in-module risk genes needed to drive across-network risk convergence. ANCR was re-computed, each time considering only a certain percentage of within-module top risk score genes for the computation of module risk enrichment; this percentage laid across the x axis. Red line: ANCR of 3, nominal significance $p = 0.01$. Green line: ANCR of 2, approximating nominal significance $p = 0.05$. Data were pooled across all networks into smooth plots.

(C) Line plots show network-wise signatures for SCZ and AD.

Abbreviations: AD, Alzheimer's disease; ADHD, attention-deficit/hyperactivity disorder; ALS, amyotrophic lateral sclerosis; AMY, amygdala; ANCR, across-network convergence of risk; ASD, autism spectrum disorder; BPD, bipolar disorder; CD, Crohn's disease; CN, caudate nucleus; DLPFC, dorsolateral prefrontal cortex; HP, hippocampus; MDD, major depressive disorder; PD, Parkinson's disease; RA, rheumatoid arthritis; sACC, subgenual anterior cingulate cortex; SCZ, schizophrenia; UC, ulcerative colitis.

See also Figures S1 and S2.

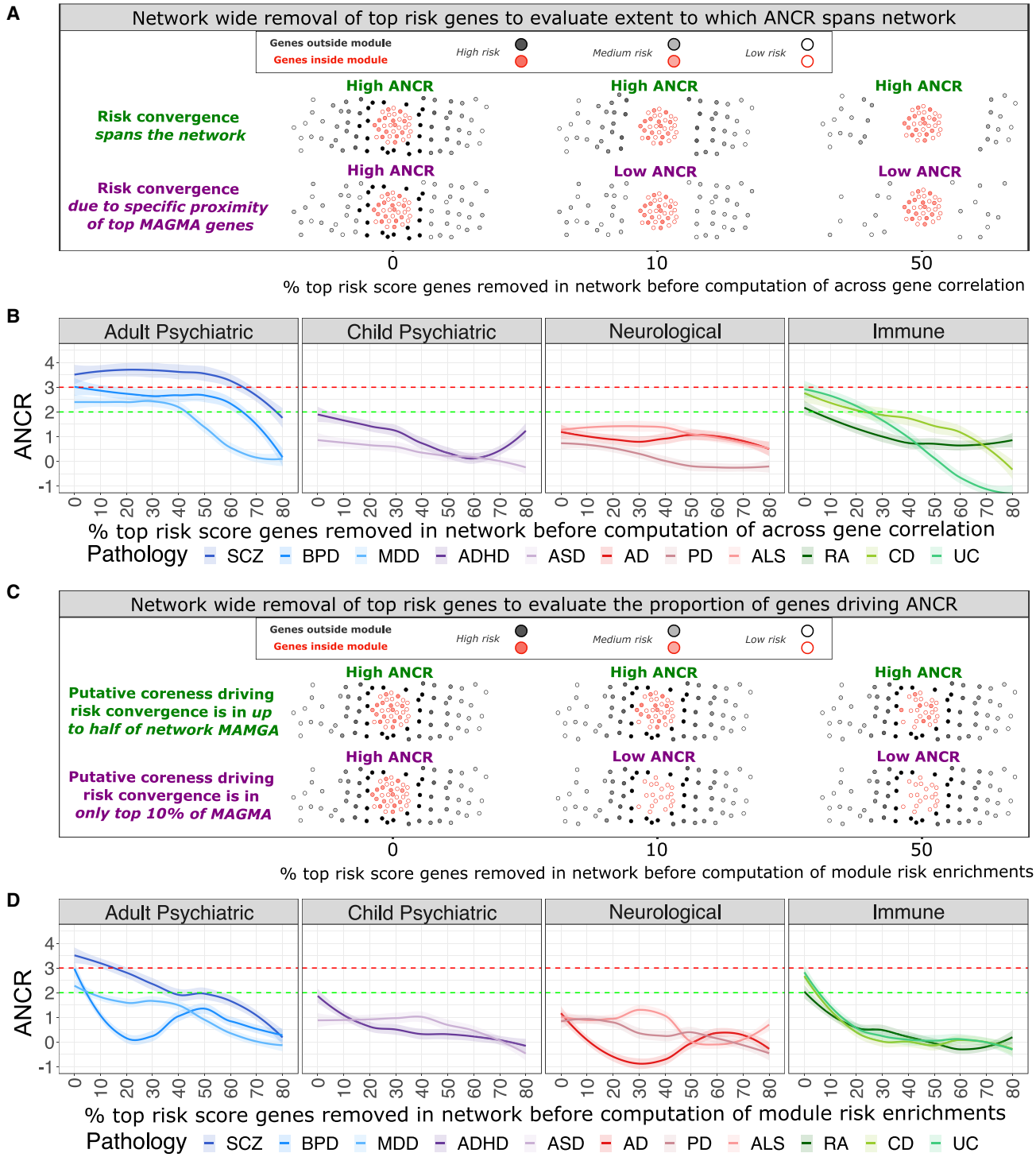

Figure 3. Effects of network-wide top risk score gene removal on properties of ANCR

(A) The green-labeled network shows risk convergence spanning the network, where across-network risk convergence onto the risk module is still present after the removal of top risk score genes. Instead, the violet-labeled network is characterized by stronger connectivity to a top proportion of risk score genes. See Figure 1A legend for additional details.

(B) ANCR was re-computed, each time removing a certain percentage of network-wide top risk score genes from the computation of the across-gene correlation; this percentage laid across the x axis. See Figure 2B legend for additional details.

(C) In the green-labeled network, within-module high risk score genes associating a module to risk convergence are found even outside the top 50% of risk score genes network-wide (some high-intensity red genes still present in right side). In the violet-labeled network, module risk genes associating a module to risk

(legend continued on next page)

Please cite this article in press as: Borcuk et al., Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.08.005

size. The resulting across-module *Z* score is the per-network index of how much co-expression network-wide risk scores converge on highly risk enriched modules, an index we refer to as ANCR.

We found that adult neuropsychiatric disorders show significant ANCR (Figure 2B). Immune disorders show an ANCR of similar strength to adult psychiatric disorders. SCZ showed the strongest ANCR (Figure 2B, left, dark blue), with risk convergence onto a module best predicted when using 100% of module genes to compute module risk enrichment (median ANCR > 3.00, median $p < 0.01$). A strong ANCR was replicated in SCZ when using linkage disequilibrium (LD) score regression-based heritability (Figure S2A) or risk enrichment of Psychiatric Genomics Consortium (PGC3) fine-mapped genes (Figure S2B) as alternative measures of module risk score. Additional PGC3 gene sets and SCHEMA genes mapping to rare coding variants[37] are also generally enriched in modules that show the strongest network connectivity to risk (Figures S2B and S2C). While SCZ and major depressive disorder (MDD) module risk scores showed the strongest ANCR when 100% of the genes were considered, bipolar disorder (BPD) ANCR was strongest when considering the top 10%–50% of module genes (all ANCR > 3.00, all $p < 0.01$). Alzheimer's disease (AD) showed an opposite relationship, with ANCR increasing as the module gene inclusion decreases. These results suggest that the *trans* component of heritability identifiable through networks is larger in SCZ than AD, as genetic signal spreads out much more widely. This phenomenon could be associated with different fractions of *cis* vs. *trans* heritability, a greater number of core genes or core pathways, or less overlap between core genes and GWAS genes. Fittingly, the most significant co-expression networks for AD at 1% module inclusion were predominantly found in the hippocampus at adult stages (HP Adult, Older_Adult; all ANCR > 3.5, all $p < 0.01$, Figure 2C, left).

The supplemental information reports additional analyses evaluating how different factors affect the computation of ANCR. Immune disorders, as well as BPD, are also found to have high ANCR in co-expression networks derived from immune and irritable-bowel-syndrome-related tissues (Figure S2D). Using mean MAGMA and gene set MAGMA approaches as module risk enrichment parameters in the computation of ANCR yields nearly identical results to the main analysis reported here (Figures S2E and S2F). Moreover, results persisted when using gene-level risk scores derived from Hi-C-coupled MAGMA (H-MAGMA) (Figure S2G) or when predicting the convergence of risk derived from published TWAS *Z* scores (Figure S2H). ANCR did not drastically change between pure-European-only and mixed-ancestry networks (Figure S2I) and remained unchanged when using different ancestry panels for computing MAGMA for each gene (Figure S2J). Ancestry of the summary statistic used may have an effect, though this is difficult to disentangle from the large variation in GWAS sample sizes. Disorders that have increased MAGMA scores, or more MAGMA significant genes, appear to have a relatively higher ANCR, consistent with a high polygenicity. However, there is no relationship between ANCR and SNP-based heritability per se (Figure S2K). Finally, we report an additional analysis showing that for adult psychiatric disorders, genes with greater risk are generally more connected with each other in co-expression networks and a protein-protein interaction (PPI) network, as would be expected by a strong network-wide risk gradient. (Figure S2L).

Effect of removing top risk score genes network-wide

When characterizing ANCR in the context of an omnigenic architecture, we should expect an across-gene relationship of "connectivity to module" and risk score that is continuous across all genes in the co-expression network, i.e., not characterized by a select group of top risk score genes having stronger connectivity to the risk module. To evaluate this characteristic, we iteratively removed top risk score genes throughout the whole co-expression network and recomputed ANCR each time (Figure 3B). We used 50% module gene inclusion for computing module risk enrichment, since this showed strong ANCR for all adult psychiatric disorders. In adult psychiatric disorders, we see a plateau that dips after removing a large percentage of genes, while other disorders exhibit a more immediate decrease. This dip is seen around 60%–70% in SCZ and BPD and at 40% for MDD. This indicates that the relationship between risk score and connectivity to highly risk-enriched modules is seen across up to 70% of network genes. We find similar results using H-MAGMA to compute gene-level risk scores (Figure S3A). Risk scores derived from TWAS *Z* scores appear to span less of the co-expression network in SCZ, perhaps due to the fewer genes available from the start (Figure S3B). Immune disorders do not exhibit the same plateau as adult neuropsychiatric disorders but rather an immediate decrease, indicating that the across-gene relationship is more strongly characterized by the top few deciles of risk genes specifically being more strongly connected to risk-enriched modules. This profile is also present in networks derived from immune and irritable-bowel-syndrome-related tissues (Figure S3C).

In a similar vein, we were interested in evaluating the proportion of top risk score genes network-wide that were essential for correctly ranking module risk enrichment values for their association with network-wide risk convergence onto modules. To test this, after removing a certain proportion of network-wide top risk score genes, we re-computed module risk-enrichment values while keeping 100% of other network genes for the across gene computation of the connectivity to module and risk score. If strong ANCR disappears after removing a certain proportion of network-wide top risk score genes, it suggests that genes driving risk convergence may be densely found within that proportion. In SCZ, there is a gradual decrease in ANCR as top

convergence are no longer present after removing the top 10% of risk score genes network-wide (no more high-intensity red genes in middle and right side). See Figure 1A legend for additional details.

(D) ANCR was re-computed, each time removing a certain percentage of network-wide top risk score genes from the computation of module risk enrichment; this percentage laid across the x axis. See Figure 2B legend for additional details.

Abbreviations: see Figure 2.

See also Figure S3.

risk score genes are removed, indicating that genes driving ANCR are rather dispersed across the risk spectrum, though they are more likely to harbor more GWAS signals (Figure 3D). In BPD, instead, there is an immediate dropoff after removing only 10% of top risk score genes, suggesting that fewer genes concentrated in the top decile of risk are sufficient to drive co-expression network risk convergence. The contrast between BPD and SCZ is at odds with their purportedly strong genetic correlation.[28,38] To understand this result, we investigated the similarity between BPD and SCZ gene-level risk scores within different quantiles of risk and found that the top 10%–20% of top risk score genes showed the strongest similarity in terms of across-gene correlation strength (Figure S3D), gene overlap (Figure S3E), and similarity of module risk enrichment (Figure S3F). Immune disorders show a similar pattern to BPD.

In summary, from the examination of the risk score distribution within the co-expression networks, we found that psychiatric disorders, in particular SCZ, most resemble the architecture expected from the omnigenic model. In psychiatric disorders in general, our analysis detects a highly polygenic architecture, where core genes are likely graded across risk in SCZ but more compactly found in the top decile of risk in BPD.

ANCR in cell-type-specific networks

We evaluated ANCR in specific cell types of the dorsolateral prefrontal cortex (DLPFC), using two public DLPFC 10×x single-nucleus RNA sequencing (snRNA-seq) datasets with pre-defined cell type clusters.[39,40] WGCNA co-expression networks were computed from pseudo-bulked expression data for each cell type cluster. We computed ANCR in these cell-type-specific co-expression networks just as described in Figure 2. SCZ was the only disorder to show consistently strong ANCR. All neuronal, in contrast to glial, cell types showed a similar profile to what was seen in bulk brain tissue co-expression networks (Figure 4A). The strongest ANCR is found in layer 2 and layer 3 excitatory cell types. Immune disorders, which showed strong ANCR in bulk tissue co-expression networks, were now much weaker in all cell types described here, hinting to an influence of microglia in bulk signatures (microglial clusters either were not present in our single-cell datasets or had too few genes to compute co-expression networks).

Next, we looked at how removal of top risk score genes network-wide influences ANCR in these snRNA-seq datasets, and we performed an identical analysis as was described for Figures 3A and 3B. We found that in excitatory neuronal cell types, SCZ ANCR again spanned up to 70% of the network (Figure 4B). Conversely, in inhibitory cell types, ANCR strength decreases immediately as top risk score genes are removed from the co-expression network, indicating that risk-enriched modules appear to be more specifically connected to the upper deciles of risk score genes.

In summary, neuronal rather than glial cell types recapitulate the results seen in bulk brain co-expression networks, with the strongest ANCRs seen in excitatory neurons of layers 2 and 3. Risk convergence onto risk-enriched modules in excitatory networks is a continuous property across the whole co-expression network, while risk convergence onto risk-enriched modules in inhibitory networks is more indicative of connectivity specifically in the top deciles of risk genes. These results serve both to validate the results seen in bulk co-expression networks within independent datasets and to parse the properties of ANCR to a higher degree of biological granularity.

Identification of SCZ-connected candidate core genes based on network evidence

Our results show consistently high ANCR for SCZ, providing evidence for the presence of SCZ core genes that are driving risk gradient convergence. We hypothesized that genes found closest to the center of a network-wide SCZ risk gradient should be enriched for SCZ core genes. To pinpoint genes proximal to the center of the SCZ risk gradient, we computed an index of how close genes lie to SCZ GWAS loci genes with respect to all other genes of a co-expression network. This way, we obtained a ranking of genes most disproportionately connected with SCZ GWAS loci genes relative to the background for each network. Three gene lists were used: the set of 120 SCZ genes prioritized by PGC3,[4] a set of genes overlapping with a 200 kbp window off of GWAS-significant SNPs, and a set of negative control genes overlapping with a 200 kbp window off of SCZ GWAS least-associated SNPs (kbp_200_negative).

We found that, specifically in brain tissues, genes most connected to SCZ GWAS loci genes are loss-of-function intolerant (Figure 5A). This finding suggests a brain-specific sensitivity of SCZ genes to deleterious mutation. When we examined the enrichment of gene ontologies in the top 300 connected genes, we saw further brain specificity and even brain region specificity. The gene ontologies of GABAergic synapse and postsynaptic density were shown to have the highest median fold change in brain co-expression networks in the top 300 SCZ-connected genes, and this appeared to be brain specific (Figure 5B). The ontology with the highest median fold change in STR co-expression networks was regulation of dopaminergic synaptic transmission, which did not have a similarly high fold change in any other region set (Figure 5B).

We aimed to validate this measure of connectivity to SCZ GWAS loci genes against an *in vitro* indication of SCZ GWAS gene interaction. To evaluate this, we used a public dataset describing how gene expression is regulated by the CRISPRa activation of PGC3 eGenes, eGenes being those found to have strongest eQTL effects in the CommonMind Consortium (CMC) data.[41] We found that genes more strongly connected to PGC3 genes also have a greater expression in response to CRISPRa activation of PGC3 eGenes ($p < 0.001$; Figure 5C, left), while no association was found with connectivity to a null set ($p = 0.551$; Figure 5C, right). This association was also significant when compared to connectivity to null sets matched to the same guanine-cytosine (GC) content, gene length and DLPFC median expression (Figure 5D, empirical $p = 0.04$). Importantly, the SCZ risk score by itself showed no significant association with expression response to eGene CRISPRa activation ($p = 0.919$; Figure 5E), meaning that a gene's contribution to SCZ risk can only be understood when putting GWAS statistics in the context of gene-gene interactions—in this case, at the co-expression level. These results highlight the capacity of this connectivity approach to detect brain- and region-specific genes

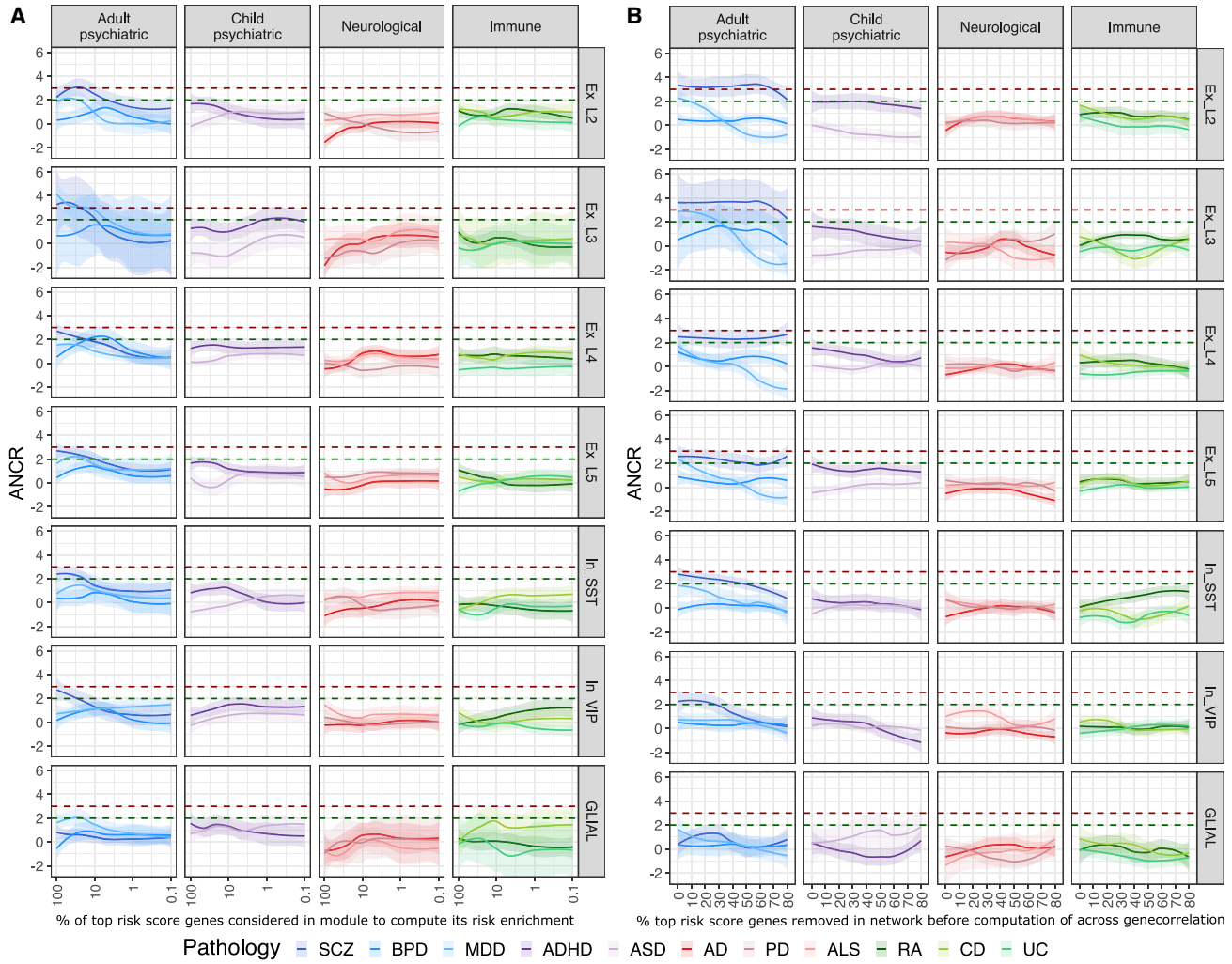

Figure 4. Across-network convergence of risk in cell-type-specific networks

(A) For each cell type group (vertically laid panels), we pooled data from all relevant cell-type-cluster networks to obtain smooth plots depicting ANCR. See Figures 2A and 2B legends for additional details.

(B) Removing network-wide top risk score genes to evaluate the extent to which ANCR extends across the network in cell-type-specific networks. See Figures 3A and 3B legends for additional details.

Abbreviations: see Figure 2.

driving SCZ risk and validate that this measure is indicative of true *in vitro* interactions with prioritized SCZ genes.

We aimed to prioritize SCZ-connected genes as those consistently connected to SCZ GWAS loci genes across co-expression networks. We created consensus matrices displaying all possible pairwise intersections between the top 300 connected genes from Genotype-Tissue Expression (GTEx)-derived co-expression networks across brain regions (Figures S4A and S4B). Hierarchical clustering of the consensus matrices revealed a clear grouping of cerebrum regions, with the three striatal regions consistently constituting a sub-branch (Figures S4A and S4B). In the negative bin, we found grouping of non-cerebellar CNS regions but no striatal sub-branch (Figure S5).

We intersected the lists of top 300 SCZ-connected genes, across region-specific co-expression networks from various consortia (Lieber Institute for Brain Development [LIBD], CMC, GTEx), for the DLPFC (4 lists), the hippocampus (HP, 3 lists), the STR (5 lists), the sACC (3 lists), and the AMY (3 lists). See Table 1 for more details. The resulting SCZ-connected genes can be found in Table S3. The juvenile DLPFC replicated 91% of the adult-derived DLFPC genes, and perinatal DLPFC replicated 61%. This is greater conservation than seen from adult-to-juvenile caudate nucleus (CN; 42%) or adult-to-juvenile (37%) or -perinatal (20%) HP. Connectivity to kbp_200_negative bin genes least associated with SCZ revealed no genes in the DLPFC, 38 genes in the STR, and 92 genes in the HP, with enrichments in gliosis and immune response, respectively (Table S4). The majority of these genes are strongly connected to significant TWAS genes as well (Figure S6A). In summary, we find that genes most connected to SCZ GWAS loci genes

Please cite this article in press as: Borcuk et al., Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.08.005

CellPress
OPEN ACCESS

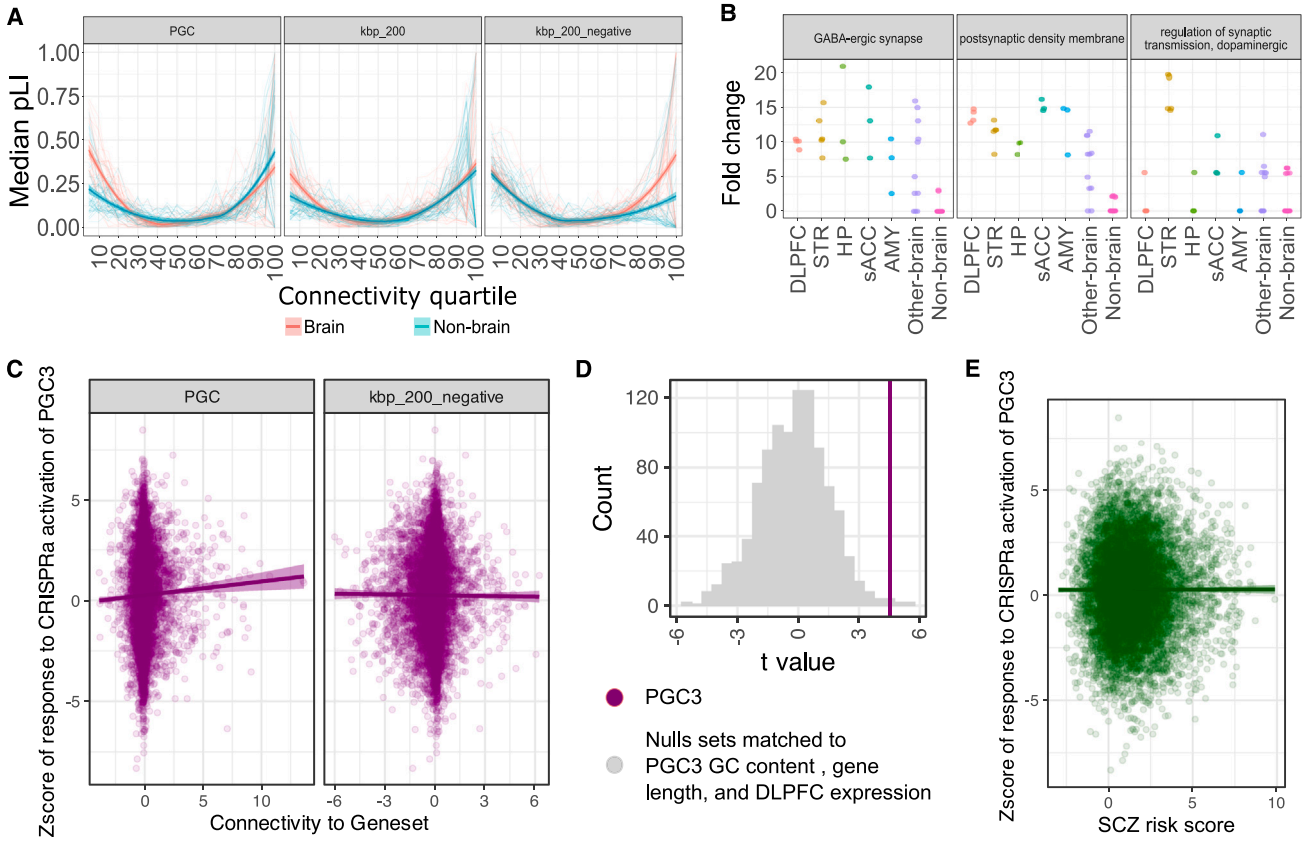

Figure 5. Genes most connected to SCZ GWAS loci genes are loss-of-function intolerant, show coherent region-specific ontologies, and are affected by CRISPRa of PGC3 genes

(A) In co-expression networks from brain and non-brain tissues, we calculated the median pLI (y axis) for various quartiles of PGC3 connectivity (x axis).

(B) Fold change for gene ontologies in the top 300 genes connected to PGC3. Points represent co-expression networks, which are colored by region set. The first 2 leftward panels show the ontologies with the strongest fold changes brain-wide. The right panel shows the regulation of synaptic transmission of dopamine secretion, which is the ontology with the highest fold change in the STR.

(C) Association between gene connectivity to PGC3 genes in adult DLPFC co-expression networks and expression response to CRISPRa of PGC3 eGenes. Gene connectivity to PGC3 genes (left) or a null set kbp_200_negative (right) was considered. The x axis shows the median gene set connectivity value across adult DLPFC co-expression networks, and the y axis shows the *Z* score indexing a gene's response to CRISPRa of PGC3 genes. Points represent genes.

(D) Comparison to null sets matched for same gene length, GC content, and median expression as PGC3-prioritized SCZ GWAS loci genes. Gray histogram indicates the distribution of t values for linear models associating connectivity to null sets and response to CRISPRa of PGC3 eGenes. The purple line indicates the t value when using connectivity to actual PGC3-prioritized SCZ GWAS loci genes in the model.

(E) Association between SCZ risk score and expression response to CRISPRa of selected PGC3 genes.

Abbreviations: pLI, probability of loss of function; see also Figure 2.

See also Figures S4–S6.

show coherent region-specific ontology profiles, and adult-derived region-specific SCZ-connected genes show strongest early-life conservation in the DLPFC.

Cerebrum-wide and region-specific SCZ-connected genes

There was remarkable overlap between region sets, so we prioritized genes that appeared in at least four region sets as cerebrum-wide SCZ-connected genes (permuted probability of obtaining a random SCZ-connected gene set larger than 0 was set to $\alpha < 0.001$). This procedure identified 14 genes (Table 2) strongly enriched for anterograde transsynaptic signaling (10 genes, false discover rate [FDR] $q = 2.45 \times 10^{-9}$). Most of these genes were subunits for various synaptic receptors and channels, e.g., NMDA was represented by *GRIN1* and *GRIN2B*. These genes were not highly ranked in non-brain co-expression networks. Most of these genes were also among the top 300 connected in the juvenile and perinatal co-expression networks of the DLPFC and the HP, while overlap with the juvenile caudate was lower.

To validate these genes in published co-expression networks (which do not have gene-wise connectivity data), we evaluated the gene-wise across-module relationship between module connectivity and median risk enrichment. Cerebrum-wide SCZ-connected genes were strongly connected to risk-enriched modules consistently across published co-expression networks (Figure 6A). We found that this measure of connectivity to risk also showed an association with regulation by

Please cite this article in press as: Borcuk et al., Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.08.005

Table 1. Region- and age-defined intersections for determining SCZ-connected genes

Region set	*n* lists	Consensus across …		
		LIBD	CMC	GTEx
DLPFC	4	DLPFC_adult, DLPFC_older_adult	DLPFC	frontal cortex (BA9)
STR	5	CN_adult, CN_older_adult	–	caudate, putamen, nucleus accumbens
HP	3	HP_adult, HP_older_adult	–	hippocampus
AMY	3	AMY_adult, AMY_older_adult	–	amygdala
sACC	3	sACC_adult, sACC_older_adult	–	subgenual anterior cingulate cortex (BA24)

Abbreviations: LIBD, Lieber Institute for Brain Development; CMC, CommonMind Consortium; GTEx, Genotype-Tissue Expression. See also Table 2.

CRISPRa of PGC3 eGenes (Figure S6B). SCHEMA-prioritized genes[37] and previously identified molecular consensus genes co-expressed with SCZ risk genes[26] showed similarly high connectivity to risk modules. The cerebrum-wide SCZ-connected genes identified here were also found to be strongly connected to higher risk score genes in the string PPI network and postnatal co-expression networks (Figures 6B–6E, DLPFC shown), reflecting the preservation of these insights from transcriptomics into protein-level networks. Again, SCHEMA and molecular consensus genes were strongly connected to risk in the postnatal DLPFC.

In order to detect region specificity in a strict manner, we removed all genes that appeared in any other of the 46+ GTEx lists, brain or otherwise (Table 2, shaded) (permuted probability of obtaining a random SCZ-connected gene set larger than 0 < 0.001). The DLPFC, sACC, and AMY did not show strong regional specificity. By contrast, 17 striatal-specific genes were detected. No genes were detected in the negative bin. These genes were enriched for potassium ion transport (3 genes, FDR q = 0.0016) and response to dopamine (3 genes, FDR q = 0.0015). *DRD1* was also present, and given the classical importance of dopamine signaling in SCZ, we were prompted to take a closer look at other dopamine receptors. We found that *DRD2* was ranked within the top 300 specifically in the dorsal STR (the caudate and putamen) and not in the nucleus accumbens, while *DRD3* was highly ranked only in the nucleus accumbens (Figure S6C). In the juvenile LIBD caudate co-expression network, 11 of these 17 genes are conserved (64%), including 6 of 7 genes related to enrichment of potassium ion transport and response to dopamine. Finally, we identified 7 HP-specific genes. In the negative bin, 3 genes were identified. There were no significant Gene Ontology enrichment profiles. In perinatal and juvenile co-expression networks, 3 genes—*MIR137HG*, *PCLO*, and *MBOAT7*—were conserved. The finding of *MIR137HG* is encouraging, as it has been previously identified as a master regulator[8,42] in the context of an omnigenic model.[6]

DISCUSSION

The purpose of this study was to evaluate the potential *trans* heritability of SCZ and an omnigenic architecture of SCZ risk funneled through gene co-expression networks in relation to other disorders. We found that for adult psychiatric disorders, especially SCZ, GWAS-signal-enriched co-expression modules likely drive the convergence of risk across their co-expression network, a concept we term ANCR. For adult psychiatric disorders, stronger ANCR was found when assigning module risk enrichments considering a larger percentage of top risk score genes (25%–100%) within a module, while AD ANCR scores were strongest when using 1%. This suggests that mechanistic drivers of risk convergence (putative core) for adult psychiatric disorders may be more generally related to molecular pathways (if we consider modules as representations of these), in contrast to AD, in which they are more related to individual gene effects, e.g., *cis* effects compared to network-associated *trans* effects. We also re-assigned module risk enrichments after removing top risk score genes network-wide, giving a more general idea of the proportion of the risk spectrum that contains putative core genes or pathways. In SCZ, network-wide convergence of risk onto a module is predicted by genes within the top 40% of top risk score genes network-wide, suggesting that putative core genes are found diffusely across the risk spectrum. In contrast, for BPD, risk convergence is substantially predicted by the top 10% of BPD risk score genes, suggesting an enrichment of core genes to this top decile. Interestingly, this decile of risk genes is also the most correlated with SCZ risk, raising the question of whether network-wide convergence is a characteristic of BPD or rather a characteristic of the top decile of risk genes that are largely common to SCZ. Similar profiles to BPD were seen for immune disorders, suggesting that they are more susceptible to *cis*-mapping variants and are thus more detectable through GWAS signals.

For the evaluation of *trans* heritability, we found that the relationship between connectivity to risk enriched modules and risk score spans 40%–70% of network genes, supporting at least a super-polygenic landscape. These results are consistent with evidence that approximately 70% of 1 MB windows in the genome contribute to heritability for SCZ.[43] Considering that heritability and GWAS sample size do not appear to explain these differences and that AD and immune disorders are starkly different from adult psychiatric disorders, we interpret these results as a reflection of the *trans* heritability of a trait as it is funneled through co-expression networks. We speculate that greater polygenicity in psychiatric disorders supports the distribution of genetic signals throughout the network.[13]

The intriguing idea of our omnigenic-derived prediction of core-gene-driven risk convergence is that SCZ gene importance assessed via GWAS signal *does not* reflect a hierarchy of mechanistic importance. GWAS signals do not only reflect direct pathogenic involvement but can be confounded by other variables such as density of genetic variation within its locus and pleiotropic pathogenicity. Harrison and Weinberger (2005) postulated that various combinations of susceptibility genes converge to effect a common pattern of synaptic dysfunction and emergent symptoms, a proposal echoed in the latest report from the

Table 2. SCZ-connected genes prioritized by either cerebrum-wide replicability or strict regional specificity

EnsemblID	Symbol	Adult					Juvenile			Perinatal		In set
		sACC	DLP	STR	HIP	AMY	DLP	STR	HIP	DLP	HIP	
ENSG00000176884	*GRIN1*	X	X	X	X	X	X	–	X	X	X	–
ENSG00000164061	*BSN*	X	X	X	X	–	X	X	X	X	X	–
ENSG00000196876	*SCN8A*	X	X	X	X	–	X	X	X	X	X	–
ENSG00000124140	*SLC12A5*	X	X	X	X	–	X	–	X	X	X	–
ENSG00000130477	*UNC13A*	X	X	X	X	–	X	–	X	X	X	–
ENSG00000186111	*PIP5K1C*	X	X	–	X	X	X	–	X	X	X	–
ENSG00000113763	*UNC5A*	X	X	–	X	X	X	–	X	X	–	–
ENSG00000273079	*GRIN2B*	X	X	X	X	–	X	–	–	X	–	–
ENSG00000158445	*KCNB1*	X	X	X	–	X	X	–	–	X	–	kbp_200
ENSG00000106976	*DNM1*	X	X	X	X	–	X	–	X	–	X	–
ENSG00000165238	*WNK2*	X	X	–	X	X	X	–	X	–	X	kbp_200
ENSG00000072071	*ADGRL1*	X	–	X	X	X	–	–	X	–	X	–
ENSG00000198216	*CACNA1E*	X	X	X	X	X	X	–	–	–	–	–
ENSG00000166206	*GABRB3*	X	X	X	X	–	X	X	–	–	–	–
ENSG00000113749[a]	*HRH2*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000119547[a]	*ONECUT2*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000123360[a]	*PDE1B*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000100302[a]	*RASD2*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000122012[a]	*SV2C*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000180720[a]	*CHRM4*	–	–	X	–	–	–	X	–	–	–	kbp_200
ENSG00000077522[a]	*ACTN2*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000141404[a]	*GNAL*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000166342[a]	*NETO1*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000066248[a]	*NGEF*	–	–	X	–	–	–	X	–	–	–	kbp_200
ENSG00000186642[a]	*PDE2A*	–	–	X	–	–	–	X	–	–	–	–
ENSG00000111262[a]	*KCNA1*	–	–	X	–	–	–	–	–	–	–	–
ENSG00000184845[a]	*DRD1*	–	–	X	–	–	–	–	–	–	–	–
ENSG00000102780[a]	*DGKH*	–	–	X	–	–	–	–	–	–	–	–
ENSG00000078328[a]	*RBFOX1*	–	–	X	–	–	–	–	–	–	–	kbp_200
ENSG00000134343[a]	*ANO3*	–	–	X	–	–	–	–	–	–	–	kbp_200
ENSG00000170921[a]	*TANC2*	–	–	X	–	–	–	–	–	–	–	kbp_200
ENSG00000125505[a]	*MBOAT7*	–	–	–	X	–	–	–	X	–	X	–
ENSG00000186472[a]	*PCLO*	–	–	–	X	–	–	–	X	–	X	–
ENSG00000225206[a]	*MIR137HG*	–	–	–	X	–	–	–	X	–	X	kbp_200
ENSG00000196220[a]	*SRGAP3*	–	–	–	X	–	–	–	–	–	–	–
ENSG00000130368[a]	*MAS1*	–	–	–	X	–	–	–	–	–	–	–
ENSG00000165802[a]	*NSMF*	–	–	–	X	–	–	–	–	–	–	–
ENSG00000173581[a]	*CCDC106*	–	–	–	X	–	–	–	–	–	–	–
ENSG00000101695[a]	*RNF125*	–	–	–	neg	–	–	–	–	–	–	–
ENSG00000158352[a]	*SHROOM4*	–	–	–	neg	–	–	–	–	–	–	–

Adult columns: X indicates consensus across lists. sACC, 3 adult sACC lists; DLP, 4 adult DLPFC lists; STR, 5 adult STR lists; HIP, 3 adult HP lists; AMY, 3 adult AMY lists; "neg" indicates that these genes are connected to negative control genes (mapping to least significant PGC3 SNPs and hence not related to schizophrenia). Juvenile columns contain an X if the gene is found in the respective juvenile list. Perinatal columns contain an X if the gene is found in the respective perinatal list (there is no STR perinatal list). In set, if the gene is already part of the PGC3 or kbp_200 gene set. Abbreviations: see Figure 2.

[a]Region-specific genes not found in any GTEx list outside of the region set.

Please cite this article in press as: Borcuk et al., Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.08.005

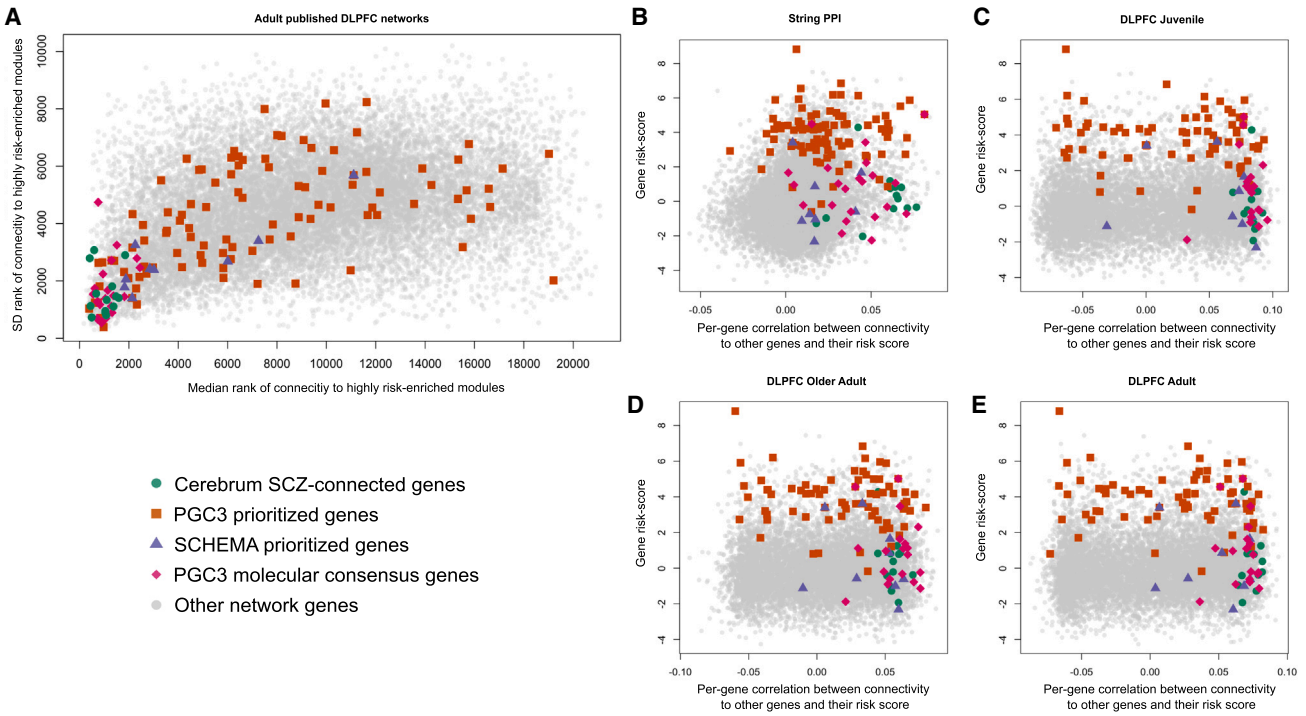

Figure 6. Cerebrum-wide SCZ-connected genes are strongly connected to SCZ risk in published adult DLPFC co-expression networks and the string PPI network

(A) In each co-expression network, each gene was ranked based on its across-module association between kme and median SCZ risk score. The x axis shows the median of a gene's rank, and the y axis shows the standard deviation of a gene's rank across co-expression networks. We highlighted the cerebrum-wide SCZ-connected genes, PGC3-prioritized genes,[4] SCHEMA-prioritized genes,[37] and PGC3 consensus molecular genes.[26]

(B–E) We evaluated the connectivity to SCZ risk of cerebrum-wide consensus genes at the proteomic level in the string PPI network (B) and in gene co-expression networks (C–E). For each gene, a "connectivity to risk" index was computed as the across-gene Spearman rho correlation between the connectivity to other genes and their risk score. This scatterplot displays genes as points, where the spearman rho "connectivity to risk" index is plotted on the x axis, with the gene-level risk score on the y axis.

See also Figures S4–S6.

PGC3.[44] Therefore, rather than in the modulation of some genes, the etiopathogenetic unit of SCZ may lie in multivariate patterns of gene ensembles. The functional underpinnings of GWAS signals are thus better understood when placed in the context of these complex gene interactions. Network approaches provide a better approximation of the hierarchy of mechanistic importance, which is not reflected by GWAS signal itself but by proximity to GWAS signal in a biological space derived from these gene interactions.[5,13]

Following this logic, we proposed lists of genes most strongly connected with SCZ risk genes based on GWAS. This operation is independent of GWAS risk scores but offers orthogonal support to GWAS candidates.[45] We show that the SCZ connectivity index is indicative of real interactions with PGC3 genes *in vitro* and revealed cerebrum-wide and tissue-specific gene sets with biologically coherent profiles. This approach was further legitimized by the presence of the previously described *MIR137HG* and the classical antipsychotic drug target *DRD2* (as well as *DRD1* and *DRD3* in the nucleus accumbens), which suggests the potential to uncover effective drug targets.[8,42] Previous work examining the co-expression context of these dopamine receptors revealed that system-level correlates of dopaminergic signaling can be predicted via genetic variation in *DRD1* and *DRD2* co-expression partners, thus supporting the idea that *DRD1*/*DRD2* co-expression connections play a functional role.[18,46–51] Many SCZ-connected genes did not have strong GWAS signals themselves. Although genes with greater GWAS signals are more likely to be core, genes are not equal regarding the density of nearby genetic variants, or eQTLs (i.e., low *cis* heritability is taken over by *trans* heritability); moreover, highly loss-of-function-intolerant genes may be especially difficult to find through GWASs.[3,11,24,52] Such genes may also have disruptive *cis* variants or bear the potential to contribute to other disorders representing exclusion criteria for SCZ GWASs. Some cerebrum-wide SCZ-connected genes have been shown to be clearly affected in SCZ patients, such as with differential transcript expression of *SLC12A5* in the DLPFC.[53] The gene lists provided in the current report represent a starting point to identify potentially mechanistic genes beyond GWAS significance.

SCZ-connected genes found in the adult DLPFC are almost completely conserved in the juvenile DLPFC and largely conserved in perinatal DLPFC, highlighting that risk convergence may emerge earliest here.[26] This conservation of SCZ-connected

genes from adult to juvenile and perinatal stages is partially shared by the HP as well. In stark contrast, the juvenile CN shows low conservation of cerebrum-wide SCZ-connected genes (30%) but stronger conservation of adult striatal-specific genes (64%). These findings are consistent with the idea that dopaminergic striatal systems mature relatively late in brain development, such that the dopamine surge hits a life stage close to that of SCZ clinical onset.

Our study comes with limitations. The absolute cohort sizes are rather limited ($n < 50$) for some co-expression networks. The conservation of DLPFC SCZ-connected genes in the juvenile DLPFC co-expression network may be weakened by the age difference between LIBD and CMC datasets (CMC subjects had a minimum age of 17 while LIBD subjects in the juvenile network were aged between 6 and 25). However, the mean age of the CMC is relatively high at 61.5 ± 19.9 years, meaning that these few young subjects should have minimal effect. Some of the intersections that are made (between region sets, between GTEx STR) have overlapping subjects, reducing the power of the consensus. Moreover, in terms of oligogenic disorders, if core genes are only a handful or mainly driven by *cis*-eQTLs, our method may not find sufficient variation between the majority of modules. In such cases, only one or two modules per network may contain core genes at all, and thus, the relationship across modules might not be captured adequately by linear models, suggesting that our model may be best suited for polygenic disorders.

In conclusion, we show that genes highly connected to SCZ risk-enriched modules accumulate relatively high SCZ-GWAS signals, meaning high SCZ gene risk infers connectivity to potential core genes and vice versa. This phenomenon was similar in other adult psychiatric disorders and much weaker in presumed oligogenic disorders. We contend that co-expression network strategies can effectively complement GWAS strategies to identify genes otherwise not associated with SCZ by GWASs, although they sit in the best network position to affect the largest possible number of SCZ risk genes. In line with this perspective, we identified sets of cerebrum-wide and region-specific potential core genes, well supported by the putative biology of SCZ (synaptic, dopamine-related genes). The genes in our SCZ connectivity analyses not associated with SCZ risk by GWASs may represent valuable entry points into core pathways that would not otherwise be identified.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, Giulio Pergola (giulio.pergola@libd.org).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- Data including gene network information for published networks, novel networks, and cell-type specific networks computed on published snRNA-Seq data, as well as per gene values and ranks for connectivity to SCZ_PGC3 gene sets will be deposited at GitHub as of the date of publication (https://doi.org/10.5281/zenodo.13381309).
- This paper analyzes existing, publicly available data. These are listed in the key resources table.
- Code to reproduce the analyses will be deposited at GitHub as of the date of publication (https://github.com/LieberInstitute/ANCR).
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

ACKNOWLEDGMENTS

This project has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Marie Skłodowska-Curie grant agreement no. 798181 awarded to G.P.; from the project "Unraveling new neural network activities for the treatment of negative symptoms and socio-cognitive abilities in schizophrenia" (PRIN: Progetti di Ricerca di Rilevante Interesse Nazionale – Bando 2022 PNRR Prot. P2022HNBJX) awarded to G.P.; from the European Union funding within the MUR PNRR Extended Partnership Initiative on Neuroscience and Neuropharmacology (project no. PE00000006 CUP H93C22000660006 "MNESYS, A multiscale integrated approach to the study of the nervous system in health and disease") to A.B. and G.P.; from the Apulian regional government for the project "Early Identification of Psychosis Risk" to A.B. The Lieber Institute for Brain Development supported tissue collection and maintenance, analysis, infrastructure, and personnel.

This paper reflects only the authors' views, and the European Union and Research Executive Agency are not liable for any use that may be made of the information contained therein. No funding body intervened in the manuscript preparation.

Prof. Antonio Rampino (University of Bari Aldo Moro, Italy) has given our team precious and continuous support. We are grateful for the comments and feedback received by Dr. Loredana Bellantuono, Gianluca Christos Kikidis (University of Bari Aldo Moro, Italy), Nicola Pedreschi (Oxford University, UK), and Fabiana Rossi (Campus Biomedico di Roma, Italy).

AUTHOR CONTRIBUTIONS

Conceptualization, G.P. and D.R.W.; data curation, C.B., M.P., L.S., and J.H.S.; formal analysis, G.P., C.B., M.P., and L.S.; funding acquisition, G.P., D.R.W., and A.B.; investigation, G.P., C.B., and D.R.W.; methodology, G.P., C.B., and M.P.; project administration, G.P.; resources, J.E.K., T.M.H., J.H.S., A.B., and D.R.W.; software, C.B., M.P., and L.S.; supervision, G.P., A.B., and D.R.W.; validation, G.P.; visualization, C.B. and M.P.; writing – original draft, G.P., C.B., M.P., and D.R.W.; writing – review & editing, all authors.

DECLARATION OF INTERESTS

A.B. received consulting fees by Biogen and lecture fees by Otsuka, Janssen, and Lundbeck. D.R.W. serves on the scientific advisory boards of Sage Therapeutics and Pasithea Therapeutics. G.P. received lecture fees by Lundbeck.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS
  - Tissue samples: LIBD
  - Tissue samples: GTEx
- METHOD DETAILS
  - Data preprocessing
  - Published co-expression networks
  - Gene expression pre-processing: LIBD
  - Network identification: LIBD
  - Gene expression pre-processing: GTEx
  - Network identification: GTEx
  - Gene-level MAGMA associations

- Across-network convergence of risk (ANCR)
- Heritability enrichment
- Computing fold enrichment of modules or combined modules
- Computing risk gradient strength
- MAGMA gene set
- TWAS *Z* scores
- Ancestry specific MAGMA
- Quantile-wise genetic similarity between SCZ and other traits
- Processing of 10x snRNA sequencing data
- Leveraging networks to boost case-control prediction based on genetic risk
- Comparing *in vitro* regulation by CRISPRa modification of PGC3 eGenes
- Gene-wise across-module kme vs. median MAGMA in published networks

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.neuron.2024.08.005.

Received: November 8, 2023
Revised: April 3, 2024
Accepted: August 7, 2024
Published: September 4, 2024

REFERENCES

1. Risch, N. (1990). Linkage strategies for genetically complex traits. I. Multilocus models. Am. J. Hum. Genet. *46*, 222–228.
2. Gershon, E.S. (2000). Bipolar illness and schizophrenia as oligogenic diseases: implications for the future. Biol. Psychiatry *47*, 240–244. https://doi.org/10.1016/s0006-3223(99)00299-1.
3. Pardiñas, A., Holmans, P., Pocklington, A., Escott-Price, V., Ripke, S., Carrera, N., Legge, S., Bishop, S., Cameron, D., Hamshere, M., et al. (2018). Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. Nat. Genet. *50*, 381–389. https://doi.org/10.1038/s41588-018-0059-2.
4. Trubetskoy, V., Pardiñas, A.F., Qi, T., Panagiotaropoulou, G., Awasthi, S., Bigdeli, T.B., Bryois, J., Chen, C.Y., Dennison, C.A., Hall, L.S., et al. (2022). Mapping genomic loci implicates genes and synaptic biology in schizophrenia. Nature *604*, 502–508. https://doi.org/10.1038/s41586-022-04434-5.
5. Boyle, E.A., Li, Y.I., and Pritchard, J.K. (2017). An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell *169*, 1177–1186. https://doi.org/10.1016/j.cell.2017.05.038.
6. Rammos, A., Gonzalez, L.A.N., The Schizophrenia Working Group of the Psychiatric Genomics Consortium 2, Weinberger, D.R., Mitchell, K.J., and Nicodemus, K.K. (2019). The role of polygenic risk score gene-set analysis in the context of the omnigenic model of schizophrenia. Neuropsychopharmacology *44*, 1562–1569. https://doi.org/10.1038/s41386-019-0410-z.
7. Cosgrove, D., Harold, D., Mothersill, O., Anney, R., Hill, M.J., Bray, N.J., Blokland, G., Petryshen, T.; The Wellcome Trust Case Control Consortium, and Richards, A., et al. (2017). MiR-137-derived polygenic risk: effects on cognitive performance in patients with schizophrenia and controls. Transl. Psychiatry *7*, e1012. https://doi.org/10.1038/tp.2016.286.
8. Pergola, G., Rampino, A., Sportelli, L., Borcuk, C.J., Passiatore, R., Di Carlo, P., Marakhovskaia, A., Fazio, L., Amoroso, N., Castro, M.N., et al. (2024). A miR-137-Related Biological Pathway of Risk for Schizophrenia Is Associated With Human Brain Emotion Processing. Biol. Psychiatry. Cogn. Neurosci. Neuroimaging *9*, 356–366. https://doi.org/10.1016/j.bpsc.2023.11.001.
9. Sekar, A., Bialas, A.R., de Rivera, H., Davis, A., Hammond, T.R., Kamitaki, N., Tooley, K., Presumey, J., Baum, M., Van Doren, V., et al. (2016). Schizophrenia risk from complex variation of complement component 4. Nature *530*, 177–183. https://doi.org/10.1038/nature16549.
10. Kim, M., Haney, J.R., Zhang, P., Hernandez, L.M., Wang, L.K., Perez-Cano, L., Loohuis, L.M.O., de la Torre-Ubieta, L., and Gandal, M.J. (2021). Brain gene co-expression networks link complement signaling with convergent synaptic pathology in schizophrenia. Nat. Neurosci. *24*, 799–809. https://doi.org/10.1038/s41593-021-00847-z.
11. Gusev, A., Mancuso, N., Won, H., Kousi, M., Finucane, H.K., Reshef, Y., Song, L., Safi, A.; Schizophrenia Working Group of the Psychiatric Genomics Consortium, and McCarroll, S., et al. (2018). Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. Nat. Genet. *50*, 538–548. https://doi.org/10.1038/s41588-018-0092-1.
12. Huckins, L.M., Dobbyn, A., Ruderfer, D.M., Hoffman, G., Wang, W., Pardiñas, A.F., Rajagopal, V.M., Als, T.D., Nguyen, H.T., Girdhar, K., et al. (2019). Gene expression imputation across multiple brain regions provides insights into schizophrenia risk. Nat. Genet. *51*, 659–674. https://doi.org/10.1038/s41588-019-0364-4.
13. Liu, X., Li, Y.I., and Pritchard, J.K. (2019). Trans Effects on Gene Expression Can Drive Omnigenic Inheritance. Cell *177*, 1022–1034.e6. https://doi.org/10.1016/j.cell.2019.04.014.
14. de Leeuw, C.A., Mooij, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219.
15. Dekker, J. (2008). Gene regulation in the third dimension. Science *319*, 1793–1794. https://doi.org/10.1126/science.1152850.
16. Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. Stat. Appl. Genet. Mol. Biol. *4*, Article17. https://doi.org/10.2202/1544-6115.1128.
17. Fromer, M., Roussos, P., Sieberts, S.K., Johnson, J.S., Kavanagh, D.H., Perumal, T.M., Ruderfer, D.M., Oh, E.C., Topol, A., Shah, H.R., et al. (2016). Gene expression elucidates functional impact of polygenic risk for schizophrenia. Nat. Neurosci. *19*, 1442–1453. https://doi.org/10.1038/nn.4399.
18. Pergola, G., Di Carlo, P., D'Ambrosio, E., Gelao, B., Fazio, L., Papalino, M., Monda, A., Scozia, G., Pietrangelo, B., Attrotto, M., et al. (2017). DRD2 co-expression network and a related polygenic index predict imaging, behavioral and clinical phenotypes linked to schizophrenia. Transl. Psychiatry *7*, e1006. https://doi.org/10.1038/tp.2016.253.
19. Gandal, M.J., Haney, J.R., Parikshak, N.N., Leppa, V., Ramaswami, G., Hartl, C., Schork, A.J., Appadurai, V., Buil, A., Werge, T.M., et al. (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science *359*, 693–697. https://doi.org/10.1126/science.aad6469.
20. Gandal, M.J., Zhang, P., Hadjimichael, E., Walker, R.L., Chen, C., Liu, S., Won, H., van Bakel, H., Varghese, M., Wang, Y., et al. (2018). Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. Science *362*, eaat8127. https://doi.org/10.1126/science.aat8127.
21. Pergola, G., Di Carlo, P., Jaffe, A.E., Papalino, M., Chen, Q., Hyde, T.M., Kleinman, J.E., Shin, J.H., Rampino, A., Blasi, G., et al. (2019). Prefrontal Coexpression of Schizophrenia Risk Genes Is Associated With Treatment Response in Patients. Biol. Psychiatry *86*, 45–55. https://doi.org/10.1016/j.biopsych.2019.03.981.
22. Radulescu, E., Jaffe, A.E., Straub, R.E., Chen, Q., Shin, J.H., Hyde, T.M., Kleinman, J.E., and Weinberger, D.R. (2020). Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain. Mol. Psychiatry *25*, 791–804. https://doi.org/10.1038/s41380-018-0304-1.
23. Werling, D.M., Pochareddy, S., Choi, J., An, J.Y., Sheppard, B., Peng, M., Li, Z., Dastmalchi, C., Santpere, G., Sousa, A.M.M., et al. (2020). Whole-Genome and RNA Sequencing Reveal Variation and Transcriptomic

Coordination in the Developing Human Prefrontal Cortex. Cell Rep. *31*, 107489. https://doi.org/10.1016/j.celrep.2020.03.053.

24. Hartl, C.L., Ramaswami, G., Pembroke, W.G., Muller, S., Pintacuda, G., Saha, A., Parsana, P., Battle, A., Lage, K., and Geschwind, D.H. (2021). Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. Nat. Neurosci. *24*, 1313–1323. https://doi.org/10.1038/s41593-021-00887-5.

25. Walker, R.L., Ramaswami, G., Hartl, C., Mancuso, N., Gandal, M.J., de la Torre-Ubieta, L., Pasaniuc, B., Stein, J.L., and Geschwind, D.H. (2019). Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. Cell *179*, 750–771.e22. https://doi.org/10.1016/j.cell.2019.09.021.

26. Pergola, G., Parihar, M., Sportelli, L., Bharadwaj, R., Borcuk, C., Radulescu, E., Bellantuono, L., Blasi, G., Chen, Q., Kleinman, J.E., et al. (2023). Consensus molecular environment of schizophrenia risk genes in coexpression networks shifting across age and brain regions. Sci. Adv. *9*, eade2812. https://doi.org/10.1126/sciadv.ade2812.

27. Zandi, P.P., Jaffe, A.E., Goes, F.S., Burke, E.E., Collado-Torres, L., Huuki-Myers, L., Seyedian, A., Lin, Y., Seifuddin, F., Pirooznia, M., et al. (2022). Amygdala and anterior cingulate transcriptomes from individuals with bipolar disorder reveal downregulated neuroimmune and synaptic pathways. Nat. Neurosci. *25*, 381–389. https://doi.org/10.1038/s41593-022-01024-6.

28. Stahl, E.A., Breen, G., Forstner, A.J., McQuillin, A., Ripke, S., Trubetskoy, V., Mattheisen, M., Wang, Y., Coleman, J.R.I., Gaspar, H.A., et al. (2019). Genome-wide association study identifies 30 loci associated with bipolar disorder. Nat. Genet. *51*, 793–803. https://doi.org/10.1038/s41588-019-0397-8.

29. Wray, N.R., Ripke, S., Mattheisen, M., Trzaskowski, M., Byrne, E.M., Abdellaoui, A., Adams, M.J., Agerbo, E., Air, T.M., Andlauer, T.M.F., et al. (2018). Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. Nat. Genet. *50*, 668–681. https://doi.org/10.1038/s41588-018-0090-3.

30. Demontis, D., Walters, R.K., Martin, J., Mattheisen, M., Als, T.D., Agerbo, E., Baldursson, G., Belliveau, R., Bybjerg-Grauholm, J., Bækvad-Hansen, M., et al. (2019). Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat. Genet. *51*, 63–75. https://doi.org/10.1038/s41588-018-0269-7.

31. Grove, J., Ripke, S., Als, T.D., Mattheisen, M., Walters, R.K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O.A., Anney, R., et al. (2019). Identification of common genetic risk variants for autism spectrum disorder. Nat. Genet. *51*, 431–444. https://doi.org/10.1038/s41588-019-0344-8.

32. Nalls, M.A., Blauwendraat, C., Vallerga, C.L., Heilbron, K., Bandres-Ciga, S., Chang, D., Tan, M., Kia, D.A., Noyce, A.J., Xue, A., et al. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol. *18*, 1091–1102. https://doi.org/10.1016/s1474-4422(19)30320-5.

33. van Rheenen, W., Shatunov, A., Dekker, A.M., McLaughlin, R.L., Diekstra, F.P., Pulit, S.L., van der Spek, R.A.A., Võsa, U., de Jong, S., Robinson, M.R., et al. (2016). Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. Nat. Genet. *48*, 1043–1048. https://doi.org/10.1038/ng.3622.

34. Jansen, I.E., Savage, J.E., Watanabe, K., Bryois, J., Williams, D.M., Steinberg, S., Sealock, J., Karlsson, I.K., Hägg, S., Athanasiu, L., et al. (2019). Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. Nat. Genet. *51*, 404–413. https://doi.org/10.1038/s41588-018-0311-9.

35. Okada, Y., Wu, D., Trynka, G., Raj, T., Terao, C., Ikari, K., Kochi, Y., Ohmura, K., Suzuki, A., Yoshida, S., et al. (2014). Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature *506*, 376–381. https://doi.org/10.1038/nature12873.

36. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat. Genet. *47*, 979–986. https://doi.org/10.1038/ng.3359.

37. Singh, T., Poterba, T., Curtis, D., Akil, H., Al Eissa, M., Barchas, J.D., Bass, N., Bigdeli, T.B., Breen, G., Bromet, E.J., et al. (2022). Rare coding variants in ten genes confer substantial risk for schizophrenia. Nature *604*, 509–516. https://doi.org/10.1038/s41586-022-04556-w.

38. Mullins, N., Forstner, A.J., O'Connell, K.S., Coombes, B., Coleman, J.R.I., Qiao, Z., Als, T.D., Bigdeli, T.B., Børte, S., Bryois, J., et al. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. Nat. Genet. *53*, 817–829. https://doi.org/10.1038/s41588-021-00857-4.

39. Batiuk, M.Y., Tyler, T., Dragicevic, K., Mei, S., Rydbirk, R., Petukhov, V., Deviatiiarov, R., Sedmak, D., Frank, E., Feher, V., et al. (2022). Upper cortical layer-driven network impairment in schizophrenia. Sci. Adv. *8*, eabn8367. https://doi.org/10.1126/sciadv.abn8367.

40. Ruzicka, W.B., Mohammadi, S., Fullard, J.F., Davila-Velderrain, J., Subburaju, S., Tso, D.R., Hourihan, M., Jiang, S., Lee, H.-C., Bendl, J., et al. (2024). Single-cell multi-cohort dissection of the schizophrenia transcriptome. Science *384*, eadg5136. https://doi.org/10.1126/science.adg5136.

41. Townsley, K.G., Li, A., Deans, P.M., Fullard, J.F., Yu, A., Cartwright, S., Zhang, W., Wang, M., Voloudakis, G., Girdhar, K., et al. (2023). Convergent impact of schizophrenia risk genes. Preprint at bioRxiv. https://doi.org/10.1101/2022.03.29.486286.

42. Sakamoto, K., and Crowley, J.J. (2018). A comprehensive review of the genetic and biological evidence supports a role for MicroRNA-137 in the etiology of schizophrenia. Am. J. Med. Genet. B Neuropsychiatr. Genet. *177*, 242–256. https://doi.org/10.1002/ajmg.b.32554.

43. Loh, P.R., Bhatia, G., Gusev, A., Finucane, H.K., Bulik-Sullivan, B.K., Pollack, S.J., de Candia, T.R., Lee, S.H., Wray, N.R., Kendler, K.S., et al. (2015). Contrasting genetic architectures of schizophrenia and other complex diseases using fast variance-components analysis. Nat. Genet. *47*, 1385–1392. https://doi.org/10.1038/ng.3431.

44. Schrode, N., Ho, S.M., Yamamuro, K., Dobbyn, A., Huckins, L., Matos, M.R., Cheng, E., Deans, P.J.M., Flaherty, E., Barretto, N., et al. (2019). Synergistic effects of common schizophrenia risk variants. Nat. Genet. *51*, 1475–1485. https://doi.org/10.1038/s41588-019-0497-5.

45. Pergola, G., Penzel, N., Sportelli, L., and Bertolino, A. (2023). Lessons Learned From Parsing Genetic Risk for Schizophrenia Into Biological Pathways. Biol. Psychiatry *94*, 121–130. https://doi.org/10.1016/j.biopsych.2022.10.009.

46. Braun, U., Harneit, A., Pergola, G., Menara, T., Schäfer, A., Betzel, R.F., Zang, Z., Schweiger, J.I., Zhang, X., Schwarz, K., et al. (2021). Brain network dynamics during working memory are modulated by dopamine and diminished in schizophrenia. Nat. Commun. *12*, 3478. https://doi.org/10.1038/s41467-021-23694-9.

47. Benjamin, K.J.M., Chen, Q., Jaffe, A.E., Stolz, J.M., Collado-Torres, L., Huuki-Myers, L.A., Burke, E.E., Arora, R., Feltrin, A.S., Barbosa, A.R., et al. (2022). Analysis of the caudate nucleus transcriptome in individuals with schizophrenia highlights effects of antipsychotics and new risk genes. Nat. Neurosci. *25*, 1559–1568. https://doi.org/10.1038/s41593-022-01182-7.

48. D'Ambrosio, E., Pergola, G., Pardiñas, A.F., Dahoun, T., Veronese, M., Sportelli, L., Taurisano, P., Griffiths, K., Jauhar, S., Rogdaki, M., et al. (2022). A polygenic score indexing a DRD2-related co-expression network is associated with striatal dopamine function. Sci. Rep. *12*, 12610. https://doi.org/10.1038/s41598-022-16442-6.

49. Fazio, L., Pergola, G., Papalino, M., Di Carlo, P., Monda, A., Gelao, B., Amoroso, N., Tangaro, S., Rampino, A., Popolizio, T., et al. (2018). Transcriptomic context of DRD1 is associated with prefrontal activity

and behavior during working memory. Proc. Natl. Acad. Sci. USA *115*, 5582–5587. https://doi.org/10.1073/pnas.1717135115.

50. Selvaggi, P., Pergola, G., Gelao, B., Di Carlo, P., Nettis, M.A., Amico, G., Fazio, L., Rampino, A., Sambataro, F., Blasi, G., and Bertolino, A. (2019). Genetic Variation of a *DRD2* Co-expression Network is Associated with Changes in Prefrontal Function After D_2 Receptors Stimulation. Cereb. Cortex *29*, 1162–1173. https://doi.org/10.1093/cercor/bhy022.
51. Sportelli, L., Eisenberg, D.P., Passiatore, R., D'Ambrosio, E., Antonucci, L.A., Bettina, J.S., Chen, Q., Goldman, A.L., Gregory, M.D., Griffiths, K., et al. (2024). Dopamine signaling enriched striatal gene set predicts striatal dopamine synthesis and physiological activity in vivo. Nat. Commun. *15*, 3342. https://doi.org/10.1038/s41467-024-47456-5.
52. Finucane, H.K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P.R., Anttila, V., Xu, H., Zang, C., Farh, K., et al. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. *47*, 1228–1235. https://doi.org/10.1038/ng.3404.
53. Tao, R., Li, C., Newburn, E.N., Ye, T., Lipska, B.K., Herman, M.M., Weinberger, D.R., Kleinman, J.E., and Hyde, T.M. (2012). Transcript-specific associations of SLC12A5 (KCC2) in human prefrontal cortex with development, schizophrenia, and affective disorders. J. Neurosci. *32*, 5216–5222. https://doi.org/10.1523/jneurosci.4626-11.2012.
54. GTEx Consortium (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. Science *369*, 1318–1330. https://doi.org/10.1126/science.aaz1776.
55. Bhattacharya, A., Vo, D.D., Jops, C., Kim, M., Wen, C., Hervoso, J.L., Pasaniuc, B., and Gandal, M.J. (2023). Isoform-level transcriptome-wide association uncovers genetic risk mechanisms for neuropsychiatric disorders in the human brain. Nat. Genet. *55*, 2117–2128. https://doi.org/10.1038/s41588-023-01560-2.
56. Venables, W.N., and Ripley, B.D. (2002). Modern Applied Statistics with S, Fourth Edition (Springer).
57. Maechler, M. (2024). Sfsmisc: Utilities from 'Seminar Fuer Statistik (ETH Zurich).
58. de Leeuw, C.A., Mooji, J.M., Heskes, T., and Posthuma, D. (2015). MAGMA: generalized gene-set analysis of GWAS data. PLoS Comput. Biol. *11*, e1004219. https://doi.org/10.1371/journal.pcbi.1004219.
59. Yurko, R., Roeder, K., Devlin, B., and G'Sell, M. (2020). H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives. Ann. Hum. Genet. *85*, 97–100. https://doi.org/10.1111/ahg.12412.
60. de Leeuw, C., Sey, N.Y.A., Posthuma, D., and Won, H. (2020). A response to Yurko et al: H-MAGMA, inheriting a shaky statistical foundation, yields excess false positives. Preprint at bioRxiv. https://doi.org/10.1101/2020.09.25.310722.
61. Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. *43*, e47. https://doi.org/10.1093/nar/gkv007.
62. Hoffman, G.E., Bendl, J., Voloudakis, G., Montgomery, K.S., Sloofman, L., Wang, Y.C., Shah, H.R., Hauberg, M.E., Johnson, J.S., Girdhar, K., et al. (2019). CommonMind Consortium provides transcriptomic and epigenomic data for Schizophrenia and Bipolar Disorder. Sci. Data *6*, 180. https://doi.org/10.1038/s41597-019-0183-6.

Please cite this article in press as: Borcuk et al., Network-wide risk convergence in gene co-expression identifies reproducible genetic hubs of schizophrenia risk, Neuron (2024), https://doi.org/10.1016/j.neuron.2024.08.005

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
Postmortem brain tissue: sACC and AMY from the LIBD	Zandi et al.[27]	Synapse: https://www.synapse.org/#!Synapse:syn4921369
Postmortem brain tissue: brain and non-brain from the GTEx Consortium	GTEx Consortium[54]	https://www.gtexportal.org/home/downloads/adult-gtex/bulk_tissue_expression (release V8)
snRNA sequencing of DLPFC – Batiuk et al.	Batiuk et al.[39]	Zenodo: https://doi.org/10.5281/zenodo.6921620
snRNA sequencing of DLPFC – Ruzicka et al.	Ruzicka et al.[40]	Synapse: https://www.synapse.org/Synapse:syn25922167
Deposited data		
WGCNA information from published networks, novel networks, networks derived from published snRNA data. Other data, including ANCR scores, connectivity to SCZ risk genes.	This paper	Zenodo: https://doi.org/10.5281/zenodo.13381309
Gene expression regulation by CRISPRa modification of PGC3 eGenes	Townsley et al.[41]	Supplemental Data 1; Sheet: META_DEGs, column: *Z* score
SCHEMA genes	Singh et al.[37]	Extended Data Table 1
PGC3 SCZ risk gene lists	Trubetskoy et al.[4]	figshare: https://doi.org/10.6084/m9.figshare.19426775
TWAS Z-scores	Bhattacharya et al.[55]	Supplementary Data 4
GWAS – SCZ	Trubetskoy et al.[4]	figshare: https://doi.org/10.6084/m9.figshare.19426775
GWAS – BPD	Stahl et al.[28]	figshare: https://doi.org/10.6084/m9.figshare.14671998
GWAS – MDD	Wray et al.[29]	figshare: https://doi.org/10.6084/m9.figshare.14672085
GWAS – ADHD	Demontis et al.[30]	figshare: https://doi.org/10.6084/m9.figshare.14671965
GWAS – ASD	Grove et al.[31]	figshare: https://doi.org/10.6084/m9.figshare.14671989
GWAS – AD	Jansen et al.[34]	surfsara: https://vu.data.surfsara.nl/index.php/s/l7aiRr1UEgdoJfZ
GWAS – PD	Nalls et al.[32]	No longer available from the original publication (for more information contact, Nalls et al.)
GWAS – ALS	van Rheenen et al.[33]	projectmine: http://databrowser.projectmine.com/
GWAS – UC and CD	Liu et al.[36]	EMBL-EBI: https://www.ebi.ac.uk/gwas/publications/26192919
GWAS – RA	Okada et al.[35]	https://plaza.umin.ac.jp/~yokada/datasource/software.htm (European RA GWAS meta-analysis)
Software and algorithms		
WGCNA	Zhang et al.[16]	https://cran.r-project.org/web/packages/WGCNA/index.html
MAGMA	de Leeuw et al.[14]	https://ctg.cncr.nl/software/magma
LDSC derived heritability enrichment	Finucane et al.[52]	https://github.com/bulik/ldsc
MASS R package	Venables et al.[56]	https://cran.r-project.org/web/packages/MASS/index.html
sfsmisc R package	Maechler[57]	https://cran.r-project.org/web/packages/sfsmisc/index.html
R (v4.3.3)	R Core Team	https://www.r-project.org/
Bioconductor packages	Bioconductor	https://www.bioconductor.org/
Other		
Code for analyses in the present study	This paper	https://github.com/LieberInstitute/ANCR

EXPERIMENTAL MODEL AND STUDY PARTICIPANT DETAILS

Tissue samples: LIBD

We included brain tissue of male and female neurotypical individuals with European or African American ancestry, all with RNA Integrity Number $\geq$6. All included neurotypical controls had minimal age-associated neuropathology (determined from postmortem histopathological examination), no substance or drug use from toxicology and were free from any psychiatric or neurological disorder from clinical histories. Population demographics information, including sample sizes, age ranges, ethnicity, and sex distributions can be found in Table S1.

DLPFC samples were derived from Brodmann Areas 9 and 46; Hippocampus samples from the hippocampus proper (all dissections included the dentate gyrus, CA3, CA2 and CA1) plus the subicular complex; for caudate nucleus, samples were taken from the anterior 'head' portion, which is the caudate part most tightly connected with the prefrontal cortex, the subgenual anterior cingulate cortex, and the amygdala. For all tissues, RNA sequencing was performed via the Illumina Ribozero Kit.

Tissue samples: GTEx

We downloaded brain and non-brain tissue data from GTEx release V8 (see Key Resources Table), keeping only samples with gtex.smafrze = = "RNASEQ", and keeping only tissues with at least 40 samples. Age ranges were provided per sample, the mean age was calculated per individual as the mean between the min and max age values provided. Ancestry information was not provided. Population demographics information, including sample sizes, age ranges, and sex distributions can be found in Table S1.

METHOD DETAILS

Data preprocessing

All downstream analyses have been performed primarily using the R statistical software (version 4.3.3).

Published co-expression networks

WGCNA co-expression networks data (module placement and kme values per gene) were collected from various publications[8,17–26] and used to conduct ANCR analyses (see methods sub-section "Across-network convergence of risk (ANCR)", pertaining to Figures 2 and 3). For co-expression networks from Pergola et al., 2023[26] (derived from LIBD) and Pergola et al., 2024[8] (derived from CMC) we had per-gene connectivity data available allowing their use for evaluating per-gene connectivity to SCZ risk genes (see methods sub-section "Leveraging networks to boost case-control prediction based on genetic risk", pertaining to Figures 5 and 6).

Gene expression pre-processing: LIBD

LIBD data had been preprocessed and WGCNA computed for the DLPFC, HP, and CN in Pergola et al., 2023.[24] The same methodology was used here to preprocess data from the sACC and AMY.[27] In summary, gene-level mRNA expression was quantified as log-transformed RPKM values with an offset of 1, i.e., $\log_2$(RPKM+1). Outlier subjects deviating more than 3 standard deviations from the mean were removed. We regressed out the effect of the following variables from each brain region cohort separately: sex, mitochondrial mapping rate, rRNA rate, gene mapping rate, RNA Integrity Number, ancestry estimated via the first ten genomic eigenvariates, and estimated individual neuronal proportion (estimated in BRETIGEA package), and the resulting residuals were rank-normalized using Blom formula. Population demographics information can be found in Table S1.

Network identification: LIBD

In depth methodology may be found in Pergola et al.,.[24] In summary, blom-normalized residuals obtained using the linear models described above were entered as input in the *blockwiseModules* function from the package WGCNA to construct "signed hybrid" network(s), i.e., negative and zero correlation were set to zero and positive correlations were soft-thresholded. We obtained the similarity matrix using Pearson's R correlation index (*minModuleSize:* 40, *maxBlockSize:* 3000, *deepSplit:* 4, *mergeCutHeight:* 0.15, *pamStage:* TRUE, *reassignThreshold:* 1e-06). The parameter used for soft-thresholding is the exponent β to which the correlation matrix is raised to obtain the adjacency matrix. The standard procedure is to pick the lowest possible β that is high enough to satisfy the scale invariance criterion, defined as the $R^2 > 0.8$ in the correlation between original and log-transformed values. Lower β values are often associated with greater network median connectivity. For each network, we selected the parameter β such that connectivity was matched across all networks (Table S2).

Gene expression pre-processing: GTEx

GTEx tissues were pre-processed for the construction of co-expression networks to obtain per-gene connectivity values needed for the "Leveraging networks to boost case-control prediction based on genetic risk" analysis. Prior to preprocessing, genes were removed if they were not present in any one of the LIBD networks, leaving 22,684 genes. All reprocessing was identical to that of LIBD excepting confounder removal. We regressed out the effect of the following variables from each region cohort separately: PC1 of RNAseq expression matrix, mean age, sex (unless the region was only represented by one sex), rRNA rate, gene mapping rate, RNA Integrity Number, postmortem ischemia, and estimated individual neuronal proportion. We marginalized rather than protect age in the preprocessing, as these samples all came from adult individuals. We did not marginalize mitochondrial mapping rate or ancestry, as we did not have access to the data. Population demographics information can be found in Table S1.

Network identification: GTEx

Network construction was carried out in an identical manner to LIBD, except we selected β by matching connectivity only within the 13 "Brain" GTEx tissues. All of the other 36 GTEx tissues were then given a β matched median connectivity to the brain derived median connectivity (Table S2).

Gene-level MAGMA associations

To calculate gene-based associations, we used the revised version of the MAGMA tool v1.09b,[58–60] SCZ-PGC3,[4] MDD,[29] BIP,[28] AD,[34] ALS,[33] PD,[32] ASD,[31] ADHD,[30] UC,[36] CD,[36] RA[35] summary statistics as SNP *p* value data, and 1000G European as the reference data file to estimate linkage disequilibrium between SNPs. We took the following steps: (i) We mapped 1000G SNPs to genes encompassed in each network used a window of 35 kb upstream and 10 kb downstream of each gene, and (ii) we used the default gene analysis method (snp-wise = mean) to compute for each gene a *Z* score based on the mean GWAS effect size of SNPs in the GWAS loci. For H-MAGMA, we used adult brain Hi-C annotation files already computed in the H-MAGMA publication.[59]

Across-network convergence of risk (ANCR)

A zeroth linear model tested the MAGMA with the following confounders: We considered as confounders chromosome location, gene start (to index location within the chromosome), gene length, GC content, strand, number of isoforms, and gene type. These variables are related with SCZ risk importance, but they are not necessarily reflecting co-expression-related signal, and that is why their biological association with SCZ is considered confounding here. Only confounders valid for the trait MAGMA (removing confounders with NAs for the trait, or not having at least 2 levels as required for the lm function in R) were considered further.

***Model 0*;** Cleaning MAGMA of confounder effects.

$$\mathrm{lm}\left(\mathrm{MAGMA} \sim \mathrm{confounders}, \mathrm{data} = \mathrm{Genes}^{\text{Network "all genes"}}\right)$$

Next, for each non-grey module with at least 30 genes, network data was split into two separate datasets: a 'within-module-genes' dataset with all genes assigned to the module, and an 'outside-genes' consisting of all genes other than the module genes (gray genes included). From the zeroth model, median residual was calculated considering only the top 0.1, 1, 5, 10, 24, 50 and 100% of the module genes by residual (minimum 1 gene). Only confounders valid for the module (removing confounders with NAs for the outside module genes, or not having at least 2 levels as required for the lm function in R) were considered further.

***Model 1*;** MAGMA convergence on module (association strength between MAGMA and kme):

$$[\mathrm{z\_kme}] = \mathrm{lm}\left(\mathrm{kme}^{\text{Network module}} \sim \mathrm{MAGMA} + \mathrm{confounders}, \mathrm{data} = \mathrm{Genes}^{\text{Network "genes outside module"}}\right)$$

A normalized t-statistic was calculated for using function *zscoreT* from the *limma*[61] package returning for each module the z_Kme statistic.

Robust linear model (rlm) statistic across modules was calculated evaluating the association between median MAGMA and z_kme confounding for module size, serving as an index of ANCR, using the *rlm()* function from the *MASS*[56] package. This "rlm.medianMAGMA__z_kme" statistic was computed for varying median_MAGMA (at varying percentiles of top MAGMA gene inclusion in the module, or in the network). We chose this rlm approach to minimize bias related to deviations from normality and outliers..

$$[\mathrm{rlm.medianMAGMA\_z\_kme}] = \mathrm{broom} :: \mathrm{tidy}(\mathrm{rlm}(\mathrm{median\_MAGMA} \sim \mathrm{z\_Kme} + \mathrm{module\_Size}) \,\%>\%.[\mathrm{grep}(''\mathrm{z\_Kme}'', .\$\mathrm{term}),] \,\%>\%\, \mathrm{pull}(\mathrm{statistic})$$

Significance for rlm association was computed using the Wald test, through the *f.robftest* function from the *sfsmisc*[57] package.

$$[\mathrm{rlm.medianMAGMA\_z\_kme\_p}] = \mathrm{f.robftest}(\mathrm{rlm}(\mathrm{median\_MAGMA} \sim \mathrm{z\_Kme} + \mathrm{module\_Size}), \mathrm{var} = ''\mathrm{z\_Kme}'')[[''\mathrm{p.value}'']]$$

Heritability enrichment

Heritability enrichment was calculated using the ldsc package with python. LDSC annotation files were made for each module in each network using the make_annot.py function.

```
python make_annot.py \
 --gene-set-file $network/$module.Geneset \
--gene-coord-file ENSG_coord.txt \
--windowsize 100000 \
--bimfile 1000G_EUR_Phase3_plink / 1000G.EUR.QC.$chrom.bim \
--annot-file $network / $module.$chrom.annot.gz
```

Linkage disequilibrium was then computed for each module annotation in each network using the ldsc.py function.

```
python ldsc.py \
--l2 \
 --bfile 1000G_EUR_Phase3_plink/1000G.EUR.QC.$chrom \
 --print-snps hapmap3_snps/hm.$chrom.snp \
--ld-wind-cm 1 \
 --annot $network/$module.$chrom.annot.gz \
 --thin-annot \
 --out $network/$modulet.$chrom
```

The heritability enrichment within a network was then calculated considering all its modules including gray, using the PGC3 SCZ summary statistics considering only europeans. Heritability enrichment values were then used as module risk scores.

```
python ldsc.py \
--h2 $ scz.PGC3.eur.sumstats.gz \
--ref-ld-chr $module_list \
--frqfile-chr 1000G_Phase3_frq/1000G.EUR.QC. \
--w-ld-chr 1000G_Phase3_weights_hm3_no_MHC/weights.hm3_noMHC. \
--overlap-annot --print-cov --print-coefficients --print-delete-vals \
--out . / /$PH_dir/$network_$TRAIT
```

All files were downloaded from https://alkesgroup.broadinstitute.org/LDSCORE.

Computing fold enrichment of modules or combined modules

Using bayesian posterior inclusion probability, PGC3 published a broad set of potentially causal genes (n = 628) that contain at least one SNP that captures 95% of Bayes posterior probability in an LD region *(Supplementary Table 12 from Trubetskoy* et al*., 2022, FINEMAP3.5 == 1)*. We recomputed ANCR where the module risk enrichment parameter was set as the fold enrichment of these genes. Other genes have been prioritized with different criteria, and may represent other potential core sets. PGC3 prioritized genes (n = 120) were chosen based on fine-mapped SNPs (n = 61) using stricter inclusion criteria, as well as with a gene expression derived SMR (n = 59). We also considered the SCHEMA-prioritized genes (n = 10) mapping to rare coding variants. We wanted to identify a property that could be tested for all prioritized gene sets independently of their size, which varied from 10 to 628. To this aim, we ranked modules by the across-gene association of connectivity to that module and risk score (hence termed module-level ANCR). Starting with the top 10% of modules based on their module-level ANCR score, we combined genes within these modules and computed their fold enrichment for published genes sets. We then increased the inclusion percentage by 10% increments and recomputed fold enrichments, to top 20%, top 30%, etc. (For Figure S2C).

Computing risk gradient strength

For Figure S2L, risk gradient strength was evaluated per network by testing if genes with increased MAGMA are located closer to other high MAGMA genes. The string PPI network was downloaded from https://string-db.org/cgi/download, taking only the human specific interactions (9606.protein.links.v11.5.txt.gz). Each protein was assigned its gene alias, and "gene connectivity" was interpreted as the inverse shortest path between genes/proteins in the unweighted PPI network. We also used LIBD gene co-expression adjacency matrixes, where gene connectivity was defined as the β soft-power adjusted connectivity values. MAGMA was first marginalized for the following confounders; chromosome location, gene start (to index location within the chromosome), gene length, GC content, strand, number of isoforms, and gene type. For each gene an across-gene spearman correlation was computed between the "connectivity of that gene to another gene" and "other gene's MAGMA". A high spearman rho correlation was interpreted as a "closeness to MAGMA" value. Finally, for each network, an across-gene kendall correlation was computed between "closeness to MAGMA" and "MAGMA", providing an index of risk gradient strength. (For Figure S2L).

MAGMA gene set

We performed gene set enrichment analysis independently for each module in each network. For the non-competitive test, modules across all networks are appended together and were presented to magma with the "–self-contained" flag. The "MU" output value was extracted and used as a module risk score. The competitive gene set test was run per module per network, with the universe consisting of the module genes + unclustered gray genes from the corresponding network. The "BETA" output was used as a module risk score. Using "BETA_STD" as a risk score showed similar results.

TWAS *Z* scores

TWAS *Z* scores were taken from Battacharya et al.[55] The absolute value was taken and the value was used in place of gene-level magma for the across-gene linear model.

Model 1; TWAS convergence on module (association strength between TWAS_Zscore and kme):
$[z\_kme] = lm(kme^{Network\ module} \sim TWAS\_Zscore + confounders, data = Genes^{Network\ "genes\ outside\ module"})$.

Ancestry specific MAGMA

The latest PGC3 release provides various summary statistics representative of African American (afr, 6,152 cases and 3,198 controls), East Asian (eas, 14,004 cases and 16,757 controls), and Latino (1,234 cases and 3,090 controls) populations in addition to European (eur, 53,386 cases and 77,258 controls). In our main study we used the core summary statistics (eas + eur) matched with 1000G European panel. PGC3 also released summary statistics from the combined populations of all ancestries (all). To address the effect of ancestry specific summary statistics on ANCR results, we computed gene-wise MAGMA for all ancestry summary statistics individually, matching each to their corresponding ancestry 1000G file as an LD reference panel. We also used the MAGMA software to compute a meta-analytical MAGMA *Z* score value across all ancestries each matched to their 1000G LD reference panel (meta). To control for the effect of ancestry on the LD reference panel, we also computed gene-wise MAGMA per 1000G ancestry file, each matched to the summary statistics of combined ancestries. ANCR was recomputed using gene risk scores derived from these combinations.

Quantile-wise genetic similarity between SCZ and other traits

We computed spearman correlations (spearman rho on y axis) of SCZ_MAGMA and other trait MAGMAs within specific 20% proportions of that trait's MAGMA distribution (Figure S3D). We also computed the Jaccard index (JI, on y axis) between gene sets found in corresponding MAGMA deciles between SCZ and other traits. BPD and MDD show the strongest JI with SCZ in the top 10% of risk score genes, again highlighting the stronger similarity with SCZ for the top decile of risk score genes (Figure S3E). Finally, we evaluated similarity to SCZ risk at the co-expression module level, correlating across-modules the median MAGMA of each trait with the median MAGMA of SCZ for each network, using a rlm model accounting for module size. This was also computed after removing a certain percentage of trait MAGMA genes, a similar procedure to Figure S3F.

Processing of 10x snRNA sequencing data

From the Batiuk et al. 2022 study,[39] expression data were downloaded from https://zenodo.org/record/6921620. Expression matrices were provided with UMI gene expression data per nuclei, per cell-type subtype cluster (a total of 33 cell-type clusters were identified). Each cell-type cluster was pseudo-bulked to provide gene expression data per subject, comprising 14 control subjects and 12 subjects diagnosed with schizophrenia. UMI counts were summed up separately for each gene across all cells derived from a specific subtype-sample combination. Subtype-sample combinations with less than 10 cells were filtered out. Pseudo-bulk counts were then normalized to counts per million (CPM): divided by the sum of all counts across all genes in any particular subtype-sample combination and multiplied by 10^6, and log transformed log2(CPM +1). Genes with median log2(CPM +1) less than 0.1 and more than 20% zeros were removed. Subtypes clusters with fewer than 4000 genes after pre-processing were removed. We regressed out the effect of the following variables from each region subtype separately: PC1 of RNAseq expression matrix, age, sex, postmortem ischemia, mean reads per nucleus, and diagnosis.

From the Ruzicka et al. 2021 study,[40] pre-pseudo-bulked expression data, taken as mean count across nuclei per individual, was downloaded at https://www.synapse.org/#!Synapse:syn51032026, from the "pseudobulk_mean_logcounts_filtered.rds" file. Genes with more than 20% zeros were removed. We regressed out the effect of the following variables from each region subtype separately: PC1 of RNAseq expression matrix, age, sex, postmortem ischemia, batch, EUR_Ancestry, EUR_Ancestry, EAS_Ancestry, AMR_Ancestry, SAS_Ancestry, AFR_Ancestry, mito_perc and diagnosis. The resulting residuals were rank-normalized using Blom formula. Networks were computed as described in methods subsection Network identification: LIBD. Networks with fewer than 6 modules were not considered for further analysis.

All cell-type cluster networks with fewer than 6 modules were removed, and only cell-type groups with networks in both snRNA-Seq datasets were considered further. Due to the small number of nuclei mapping to glial cell-types, they were combined into one group.

Leveraging networks to boost case-control prediction based on genetic risk

To define SCZ risk genes, we used the PGC3 definition of 120 priority genes as the list called "PGC3" and then, to obtain genomic regions agnostic of PGC3 fine mapping and reflecting a consensus between *SCZ risk gene* lists, we evaluated enrichment at a

200_kbp distance from the indexed PGC3 GWAS-significant SNPs (n genes = 2725) a method we have previously tested in another dataset.[24]

We generated a negative control gene set by selecting the 329 least significant SNPs in the PGC3 summary statistic table (n genes = 2645, thus matching the number of genome-wide significant SNPs for SCZ) and defining negative loci at the above considered distances from each tagging negative SNP. We excluded from this selection any region located at ±1.5 Mbp from SCZ SNPs to rule out any overlaps between positive and negative gene sets.

We worked directly on the adjacency matrices of each of the co-expression networks we identified, plus other co-expression networks computed on GTEx[54] (including brain and non-brain tissue) and CommonMind Consortium (CMC) data.[8,62] For each adjacency matrix, we calculated the sum of connectivity to PGC3 SCZ genes (defined at PGC3, kbp_200 genomic extension bins from PGC3 significant SNPs, and a kbp_200_negative control window). The vector of PGC3 connectivity values was then marginalized by a vector of ktotal values (sum of connectivity to the entire network). For each network we took the list of 300 most connected genes, and intersected these lists based on region. GTEx datasets are exclusively comprised of adults so only adult LIBD networks were initially intersected. Then to evaluate if these adult SCZ-connected genes were also expressed in juvenile stages, we further intersected with the juvenile lists. The resulting lists were then intersected with perinatal lists to detect genes that are continually expressed from perinatal stages to adulthood. The exact networks intersected to obtain the adult SCZ-connected lists are shown in Table 1.

Comparing *in vitro* regulation by CRISPRa modification of PGC3 eGenes

Zscores detailing expression response to CRISPRa administered activation of PGC3 eGenes were downloaded from Townsley et al.,[41] Supplemental Data 1, sheet name ''META_DEGs'', column ''*Z* score''. The linear association of this score was compared with the median value of PGC3 connectivity (connectivity to PGC3 prioritized genes regressed to background connectivity) across all adult DLPFC networks, the median value of null connectivity (connectivity to kbp_200_negative genes regressed to background connectivity) across all adult DLPFC networks, and to the SCZ MAGMA score, using linear models.

Gene-wise across-module kme vs. median MAGMA in published networks

MAGMA was first marginalized for the following confounders: chromosome location, gene start (to index location within the chromosome), gene length, GC content, strand, number of isoforms, and gene type. All published networks were included in this analysis, excluding Walker et al.[25] and Werling et al.[23] since these did not contain adult subjects. In each a network, for each gene the association between kme and median MAGMA was computed using a robust linear model, accounting for module size (*rlm* function from the MASS package). Each gene was ranked within each network based on the rlm *Z* score. Then, the median rank and standard deviation of rank of each gene was computed and visualized on a scatterplot.