Transcriptomics and proteomics of projection neurons in a circuit linking hippocampus with dorsolateral prefrontal cortex in human brain.

Rahul A. Bharadwaj[1,*], Christopher Borcuk[1,2,*], Gianluca C. Kikidis[1,2,*], Leonardo Sportelli[1,2], Alessandro Bertolino[2,3], Robert N. Cole[4], Lauren DeVine[4], Joel E. Kleinman[1,5], Brady Maher[1], Srinidhi R. Sripathy[1], Vamshi Mallepalli[1], Madhur Parihar[1], Joo Heon Shin[1], Yong Kyu Lee[1], Carly Montoya[1], Amy Deep-Soboslay[1], Thomas M. Hyde[1,5,6], Daniel R. Weinberger[1,7,8,#], Giulio Pergola[1,2,5,#]

[1] Lieber Institute for Brain Development, Johns Hopkins Medical Campus, Baltimore, MD (USA)

[2] Group of Psychiatric Neuroscience, Department of Translational Medicine and Neuroscience, University of Bari Aldo Moro, Bari, Italy

[3] Azienda Ospedaliero-Universitaria Consorziale Policlinico, Bari, Italy

[4] Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD, United States.

[5] Department of Psychiatry and Behavioral Sciences, Johns Hopkins University School of Medicine, Baltimore, Maryland

[6] Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, Maryland

[7] Department of Neuroscience, Johns Hopkins University School of Medicine, Baltimore, Maryland

[8] Department of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland

* Shared first authorship

Corresponding author

Please address correspondence to:

Daniel R. Weinberger

855 N Wolfe St Suite 300, Baltimore, MD, 21205, USA

drweinberger@libd.org

Giulio Pergola

855 N Wolfe St Suite 300, Baltimore, MD, 21205, USA

Giulio.pergola@libd.org

Abstract

RNA-sequencing studies of brain tissue homogenates have shed light on the molecular processes underlying schizophrenia (SCZ) but lack biological granularity at the cell type level. Laser capture microdissection (LCM) can isolate selective cell populations with intact cell bodies to allow complementary gene expression analyses of mRNA and protein. We used LCM to collect excitatory neuron-enriched samples from CA1 and subiculum (SUB) of the hippocampus and layer III of the dorsolateral prefrontal cortex (DLPFC), from which we generated gene, transcript, and peptide level data. In a machine learning framework, LCM-derived expression achieved superior regional identity predictions as compared to bulk tissue, with further improvements when using isoform-level transcript and protein quantifications. LCM-derived co-expression also had increased co-expression strength of neuronal gene sets compared to tissue homogenates. SCZ risk co-expression pathways were identified and replicated across transcript and protein networks and were consistently enriched for glutamate receptor complex and post-synaptic functions. Finally, through inter-regional co-expression analyses, we show that CA1 to SUB transcriptomic connectivity may be altered in SCZ.

Introduction

The latest Psychiatric Genomics Consortium (PGC) study of schizophrenia (SCZ) has identified 287 genetic loci individually conferring small effects on risk for SCZ [1], and strongly enriched for synaptic function. An important current challenge in SCZ research is the question of risk convergence: how do hundreds of spatially distant genetic loci together confer risk for SCZ, and which biological pathways do they converge upon? Studies in the postmortem human brain have expanded our understanding of how SCZ is associated with gene expression by profiling messenger RNA extracted from bulk homogenate tissue. Using relatively large sample sizes, such studies have identified eQTLs and co-expression networks potentially implicated in SCZ risk [2-10]. These insights from bulk tissue data have been further translated into potential pathophysiological mechanisms of SCZ at gene [6,9-17], transcript [4,8,15], and most recently, protein levels [18].

Still, insight is as good as the biological significance of the assays used to derive it. A major caveat of these studies is that information on cell specificity is mixed up across cells and thus difficult to extract from tissue homogenates. Instead, single-cell and single-nuclei approaches (scRNA-seq, snRNA-seq) allow for the identification of many separate cell-type clusters at once [19-22]. However, their high cost currently limits sample sizes so that robust associations with genetic variation, requiring hundreds to thousands of samples, are unavailable. Additionally, snRNA-seq only captures nuclei rather than whole cell bodies. As a consequence, the mRNA synthesized and transported to peripheral cell components such as the synapse cannot be reliably quantified in the nucleus, with decreased validity compared to whole cells. Moreover, protein analysis of nuclei is limited almost exclusively to transcription factors. Further, cell types are inferred only by their genetic profile, while morphological or other anatomical features are not considered; finally, 3' amplification bias, a limitation of the most widely used snRNAseq approach, does not allow

isoform resolution [23]. These caveats are largely circumvented with the use of laser capture microdissection (LCM), a significantly more cost-effective approach that enables the selection and isolation of whole cells prior to RNA-seq or peptide-seq [24].

We have previously used LCM to collect cell-type enriched data from the dentate gyrus of 263 subjects [3,6]. Not surprisingly, DG-LCM data showed a greater proportion of excitatory neurons and greater expression of neuronal genes as compared to bulk hippocampus. Fifteen DG-specific SCZ-risk eQTLs were identified in the cell population-enriched samples not detected in bulk hippocampal tissue. Co-expression networks derived from DG-LCM were more faithful to neuronal gene ontologies and presented a greater SCZ risk gene enrichment than bulk tissue [6], although some of this granular information could be rescued from bulk tissue using statistical models of RNA degradation [25]. The granule cell layer of the DG is nonetheless quite compact and easy to capture morphologically, and an LCM analysis targeting cell populations across a distributed circuit classically linked to SCZ has yet to be tested.

Here, we have utilized LCM to investigate the correspondence of gene expression based on RNAseq and mass spectrometry derived labeled quantitative proteomics (Pep-seq) in a circuit comprised of hippocampal-prefrontal pyramidal neurons. We hypothesized that differences between cases and controls would be apparent at the circuit level not seen when considering one region at a time.

We used LCM to collect cell-specific RNAseq and Pep-seq data from 10 patients with SCZ and 10 healthy controls in a circuit connecting the hippocampal formation with the DLPFC, long implicated in SCZ pathogenesis [26] (Figure 1). We isolated projection neurons in mono-synaptic (CA1 to SUB) and reciprocally (SUB and DLPFC) connected regions [27,28], i.e., from glutamatergic neurons in layer III of the DLPFC [7,12,29,30], in the SUB, and in the CA1 of the hippocampus [31-33].

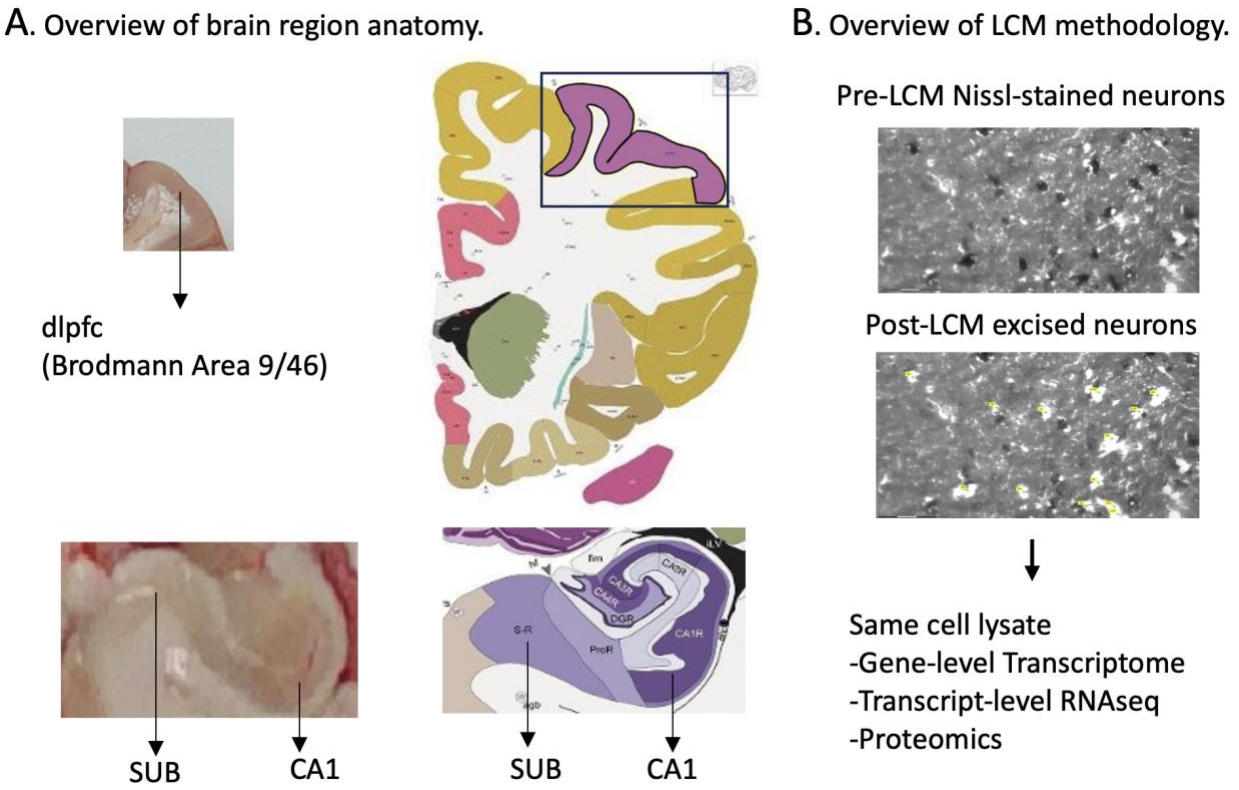

Figure 1. A) Major projection neurons between the hippocampal output centers CA1 and SUB (bottom left), and DLPFC (top left), with anatomically corresponding coronal section atlas images on the right (Allen Brain Atlas, Human Brain Reference Atlas, https://atlas.brain-map.org/)[34] for consistent LCM sampling are depicted (CA1 - cornu ammonis stratum pyramidale; SUB-subiculum large pyramidal layer, dlpfc - dorsolateral prefrontal cortex). Note that the hippocampus close-up image from Allen Brain Atlas has been enlarged to clearly visualize specific cell-types. **B)** Overview of selection and excision of Nissl-stained projection neuron (representative example is from CA1 stratum pyramidale) enrichments using Zeiss Laser Capture Microdissection (LCM) technique with concurrent transcriptomics and proteomics libraries generated from same cell lysate.

Dlpfc URL

https://atlas.brain-map.org/atlas?atlas=265297126#atlas=265297126&plate=111257682&structure=10390&x=40168&y=30802&zoom=-7&resolution=124.49&z=4

Hippocampal formation URL

https://atlas.brain-map.org/atlas?atlas=265297126#atlas=265297126&plate=112360804&structure=10390&x=41600&y=72000&zoom=-6&resolution=62.18&z=3

Results

Advantages of LCM-derived data compared to BULK and 10x snRNAseq

To gauge the quality of the LCM data in comparison to bulk tissue RNAseq, we analyzed the dispersion and intensity of expression data in and across subjects, for both data types. A high-quality dataset that shows consistently strong expression across subjects should show abundant mRNA yield of gene expression while retaining low dispersion. This property should be especially valid for so-called housekeeping genes, whose expression should show generally poor dispersion across subjects. We compared these qualities between DLPFC bulk and DLPFC LCM (Figure 2A, left and middle panels). Gene expression was less dispersed in LCM (across-genes median interquartile range [IQR] = .51, median intensity = 4.3) while retaining similar gene expression intensity as in bulk (median IQR = 1.2, median intensity = 4.4). LCM was also compared to 10x chromium snRNA-seq from an earlier study [19], a popular method for obtaining cell-type enriched expression. We found that pseudo-bulked (aggregating cells to give one value per gene per individual) snRNA-seq gene expression data for layer III specific clusters also shows greater dispersion as compared to LCM (Figure 2A, right panel). In summary, operator-driven cell selection retains an mRNA yield comparable to bulk tissue, while minimizing noise compared with snRNAseq and bulk tissue. We verified that gene expression differed in cell proportions obtained via deconvolution approaches compared to bulk tissue (CIBERSORTx with published 10x DLPFC and HP data [20]). LCM samples showed an increased proportion of specific excitatory neuronal subtype clusters (Figure 2B, subtype comparisons; Tukey $p < .001$). We evaluated how genes were differentially expressed between LCM and Bulk tissues. We found that marker genes for layer III of the DLPFC, according to a spatial transcriptomics study of this region [20], were more strongly expressed in LCM DLPFC compared to bulk (Figure 2C). Genes significantly more co-

expressed in LCM compared to bulk show consistent enrichment for synaptic ontologies (Figure 2D).

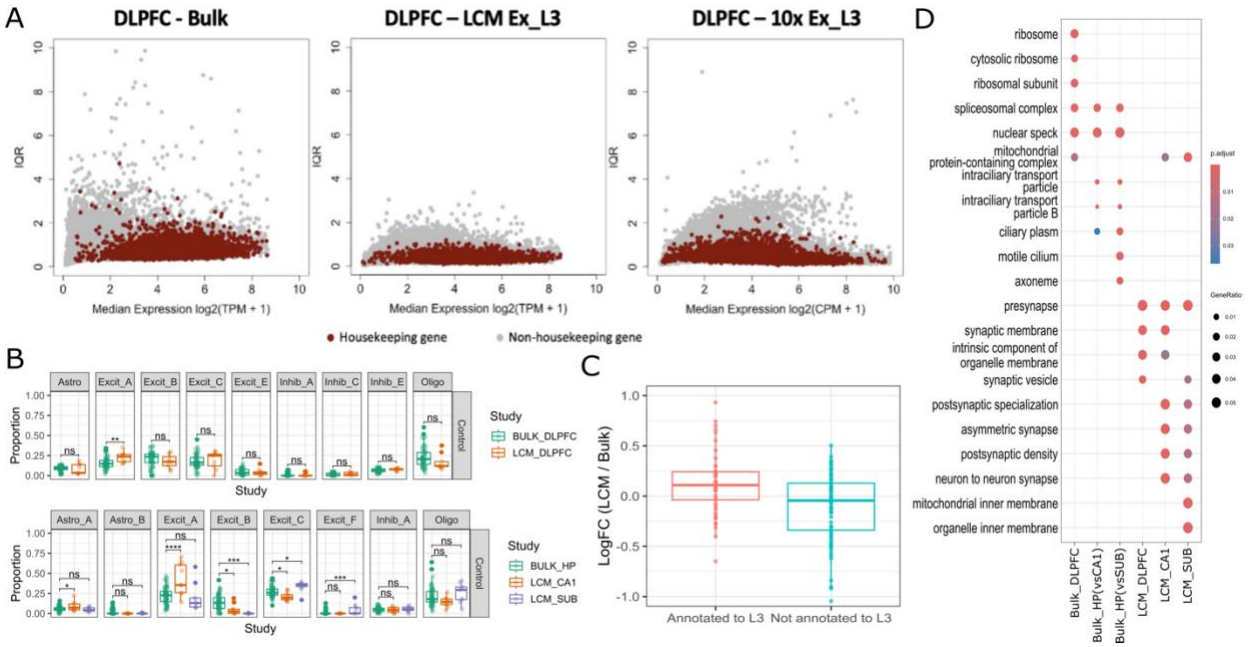

Figure 2. Cell type proportion and quality of gene expression in bulk, LCM, and pseudobulk 10x snRNAseq data. A) To gauge expression intensity, we computed median expression across subjects (shown on the x-axis). To gauge expression dispersion across subjects, we computed the interquartile range (IQR, shown on the y axis). Points are genes. In the LCM we see that gene expression of housekeeping genes as well as non-housekeeping genes is less dispersed while retaining the same intensity of Bulk tissue and pseudobulk 10x data. **B)** Cell-type proportions were estimated using CIBERSORTx in hippocampal tissues **(bottom panels)** and DLPFC **(top panels).** LCM tissues show a specific increase in the proportion of certain excitatory neuronal subtypes (from HP snRNA signature matrix CA1 = Excit_A, SUB = Excit_F and Excit C; from DLPFC snRNA signature matrix DLPFC = Excit_A) as compared to bulk tissue. **C)** We evaluated if marker genes for L3 were more differentially expressed in LCM versus Bulk as compared to marker genes for other layers. We used markers published from a spatial transcriptomics analysis of the DLPFC [3] (Maynard et al., 2021). D) We checked Gene ontology (cellular compartment) enrichments for significant differentially expressed genes (DEGs) between Bulk and LCM samples. The y-axis shows enriched terms for DGE in each brain region (x-axis).

Machine learning prediction of regional identity

We determined differentially expressed genes between regions in a leave-one-out cross-validation framework and used a Random Forest classifier to predict the region of the left-out sample. We computed 100 iterations to generate reliable distributions of expression profiles between brain regions. We then compared prediction accuracies within the same samples between the different omics data. This design minimizes confounders and allows assaying an interconnected circuit of brain regions implicated in SCZ.

Figure 3A shows the gene ontology enrichment for the most predictive genes for each region pair comparison. Figure 3B shows the prediction results. At the total gene level, bulk tissue gene reads distinguished DLPFC/HP with 69% accuracy, 95% confidence interval (CI) [0.657, 0.722]. LCM-derived mRNA reads discriminated DLPFC/SUB with an average of 88% accuracy, CI [0.876, 0.887]; in contrast, the DLPFC/CA1 discrimination was comparable with bulk tissue with an average of 72.5% accuracy, CI [0.720, 0.731] (Figure 3B). We tested LCM SUB/CA1, to assess LCM region distinction in two subdivisions of the hippocampal formation. LCM SUB/CA1 prediction achieved 67% accuracy, CI [0.63, 0.71]. At the transcript level, LCM enabled remarkably superior predictions: DLPFC/SUB achieved 92.5% accuracy, CI [0.919, 0.931], and DLPFC/CA1 94.6%, CI [0.942, 0.951], whereas bulk DLPFC/HP transcripts obtained 84.2%, CI [0.81, 0.87]. LCM SUB/CA1 prediction achieved 73% accuracy, CI [0.70, 0.75]. The superiority of LCM RNA-seq prediction with respect to bulk was significant (two sample t-tests: gene bulk DLPFC/HP vs LCM DLPFC/SUB; $t(79) = 5.53$, $p < .001$; DLPFC/HP vs DLPFC/CA1; $t(79) = 3.83$, $p < 0.01$; transcript bulk DLPFC/HP vs LCM DLPFC/SUB; $t(79) = 5.50$, $p < 0.01$; DLPFC/HP vs DLPFC/CA1; $t(79) = 6.8$, $p < .001$).

We next calculated the predictive region accuracy on the same subjects using the peptide-Seq data. The LCM Pep-seq data distinguished DLPFC/SUB with a 97.5% accuracy, CI [0.97, 0.98], DLPFC/CA1 with a 97.4% accuracy, CI [0.96, 0.982], and SUB/CA1 with a 94.3% accuracy, CI [0.937, 0.948]. LCM Pep-seq data outperformed LCM RNA-seq data significantly (paired t-tests: LCM transcripts DLPFC/SUB vs LCM Pep-seq DLPFC/SUB; $t (39) = 44.6$, $p < .001$; LCM transcript DLPFC/CA1 vs Pep-seq DLPFC/CA1; $t (38) = 11.5$, $p < .001$). In summary, using LCM over bulk tissue and transcripts over genes allows for more precise evaluations of regional

expression characteristics. The gain obtained in the leap from transcriptomics to proteomics is remarkable.

Machine learning prediction of diagnosis

Once we established that LCM transcript and protein data grant more precise regional identity, we tested whether this additional biological information might help in better distinguishing SCZ patients from healthy controls. We used the same machine learning approach and determined differentially expressed genes between the presence/absence of a diagnosis, ultimately predicting diagnosis status in the left-out individual within each brain region. LCM transcript data predicted diagnosis status with an average of 52%, CI [0.50, 0.53] in DLPFC, 39%, CI [0.38, 0.40] in SUB, and 35%, CI [0.34, 0.36] in CA1. LCM protein data predicted diagnosis with 55.6% accuracy, CI [0.55, 0.56] in DLPFC, 46.9% accuracy CI [0.45, 0.48] in SUB, and 40% accuracy, CI [0.39, 0.41] in CA1 (Figure 3C). Although the confidence intervals of Pep-seq predictions were uniformly superior to transcript data, only for the DLPFC were they above chance. We performed the same analysis, including all three brain regions at once for diagnosis status prediction, and found a 44% accuracy, CI [0.43, 0.45], for transcripts, and 52%, CI [0.52, 0.53] for proteins.

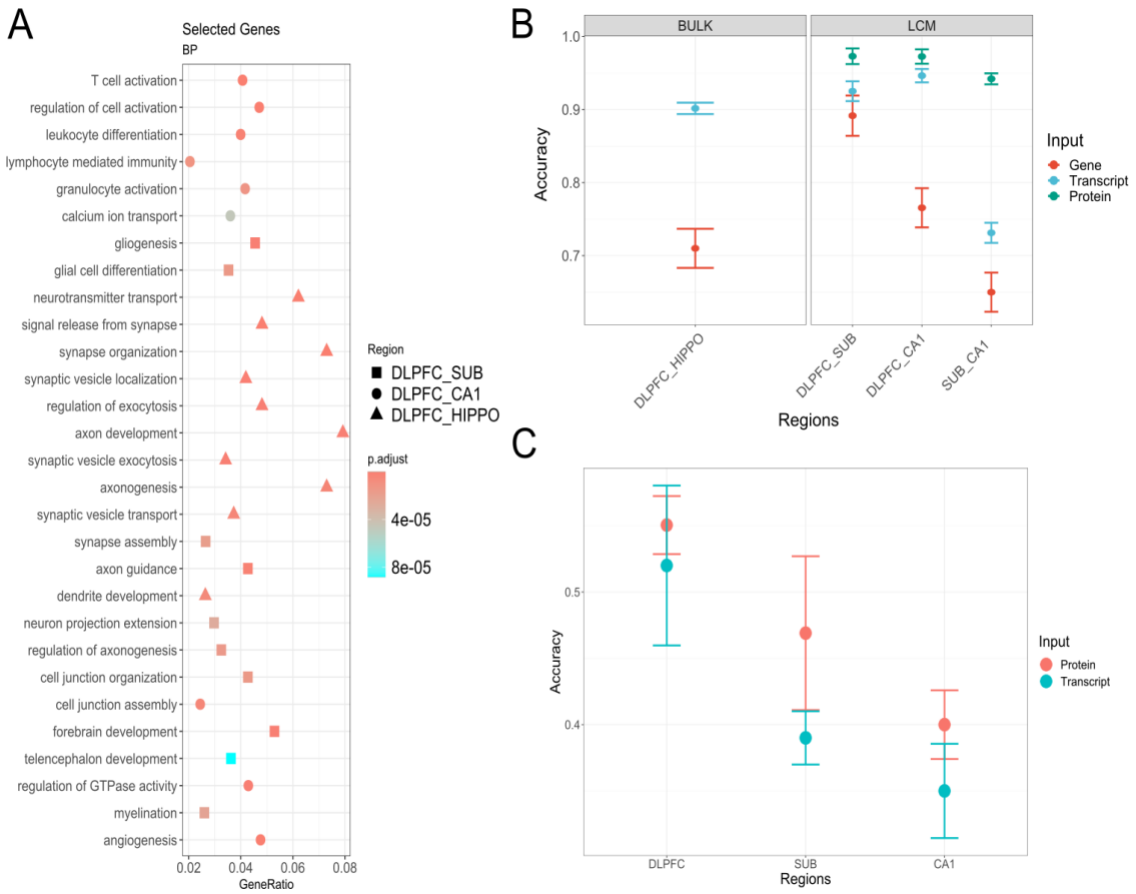

Figure 3. Machine learning prediction of regional identity and diagnosis. We established differentially expressed genes between each brain region to assess regional identity through a random forest leave-one-out cross-validation analysis. **A)** The gene ontology enrichment for the most predictive genes. **B)** shows average prediction accuracy and error bars between DLPFC/HIPPO (Bulk tissue) and DLPFC/SUB, DLPFC/CA1, and SUB/CA1 (LCM data). Red dots represent gene-level data, blue dots transcript data, and green dots protein data. **C)** Accuracies and error bars for diagnosis prediction with LCM protein (red) and transcript (blue) data.

LCM co-expression profiles show increased co-expression strength in neuronal gene sets

Techniques for parsing out molecular pathways present in expression data commonly use gene co-expression across subjects. We hypothesized that LCM data would have superior resolution to define the co-expression of synaptic and neuronal gene sets compared to bulk tissue data. We used the CLustering by Inferred Co-expression (CLIC) database to extract gene sets consistently co-expressed across published RNA expression datasets. A Z-score was computed for each gene set, comparing its co-expression to a null distribution of random gene sets of equal size with similar

gene length and GC content. The same analysis was performed on LCM expression data as well as on bulk tissue data from the LIBD brain repository [5]. The median co-expression strength value was computed within LCM and bulk tissue networks separately. To visualize how co-expression strength compares between the two techniques we plotted the strength values of LCM against bulk tissue in a scatter plot (Figure 4A). We found that neuronal sets (e.g., synaptic and dendritic sets) lay more in the top left above the best-fit line, i.e., they had stronger co-expression strength in LCM than would be expected by a linear relationship. To obtain a quantitative assessment, we scaled co-expression strength values within each dataset (within LCM and within bulk tissue) and compared the scaled values of neuronally relevant sets. The scaled values of neuronal gene sets were significantly increased in LCM (Figure 4B, statistic for synaptic terms: $t(18) = 4.2$, $p < .001$). In summary, verified neuronal gene sets, representing a ground truth model for neuronal co-regulation of gene expression, showed more prevalent co-expression strength in LCM vs bulk data, demonstrating that LCM neuronally enriched samples have a greater potential to precisely parse out neuronally relevant pathways related to SCZ risk.

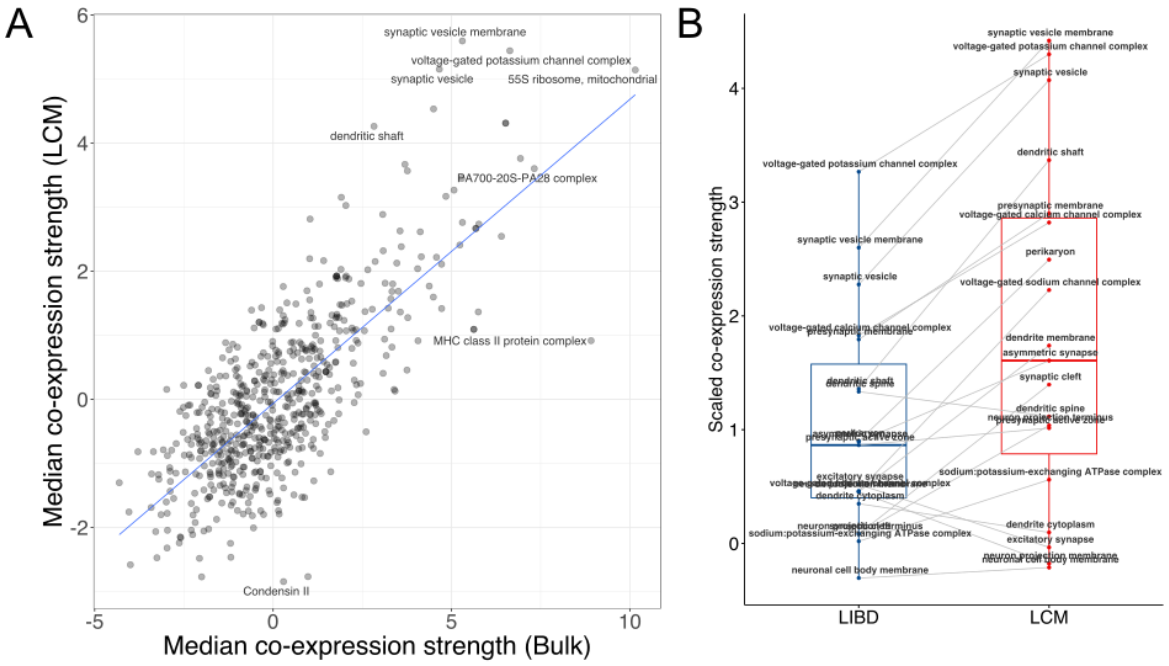

Figure 4. Co-expression strength evaluation of CLIC gene sets. A) LCM appears to have greater relative co-expression strength of neuronal gene sets, which are found on the top of the scatter plot well above the best-fit line. **B)** This difference in neuronal co-expression strength is significant when comparing the scaled values between LCM and LIBD.

Evaluating signatures of genes consistently expressed across the transcriptome and proteome in brain and in iPSC-derived neurons

We aimed to identify and characterize genes that are consistently expressed between transcriptomic and proteomic states. To this end, we computed a translatability score per gene as the across-subject association between RNA-seq and protein expression. We identified separate sets of genes with nominal significance for the linear RNA-protein association for each brain region and assessed gene ontology enrichments. Genes with significant RNA-protein translatability were most strongly enriched for synaptic terms across both hippocampal tissues and for cell surface in the DLPFC (Figure 5A). Given that transcription and translation vary over time depending on cellular states and activity, we hypothesized that stronger translatability would be related to increased gene expression stability over time.

To assign a stability measure per gene, we used published iPSC RNA-seq expression data [35]. We considered a gene to have stronger stability if it had similar expression profiles between two mature neuronal states, occurring at days in vitro (DIV) 56 and 70 in this case. The iPSC gene stability measure was thus computed as the across-subjects Kendall's tau correlation coefficient between gene expression at DIV56 vs DIV70. Across all tissues, the relationship between gene translatability and stability (Figure 5B) was strongly described using a spline regression with a knot at iPSC gene stability = 0.59 (CA1 $F(2,3181) = 18.3$, DLPFC $F(2,3158) = 29.7$, SUB $F(2,3166) = 25.8$, all adjusted $R^2 > 0.01$,_all $p < .001$; Figure 5C). Figure 5B shows no significant relationship between transcriptomics and proteomics quantifications with low stability in iPSCs (<.59 iPSC stability; CA1 $F(1,2073) = 2.41$, DLPFC $F(1,2073) = 1.95$, SUB $F(1,2061) = 1.20$, all $p > .1$), and a strong linear relationship for high-stability genes (>.59 iPSC stability; CA1 $F(1,1107) = 8.87$, $p = .002$, DLPFC $F(1,1084) = 15.1$, $p < .001$, SUB $F(1,1104) = 14.1$, $p < .001$).

To evaluate diagnosis-dependent effects, iPSC gene stability was also computed using iPSC-derived neuronal samples only from SCZ or only from CTR subjects, and gene ontology cellular compartment enrichment was evaluated for significantly stable genes (genes with iPSC stability FDR-adjusted $p < .05$). Genes found significant using both SCZ- and CTR-derived stability measures were enriched for synaptic and cell adhesion terms. Genes uniquely significant to CTR-derived stability were again enriched for synaptic terms. In contrast, those uniquely significant to SCZ-derived stability were not enriched for synaptic terms but rather for cell adhesion terms with additional unique enrichments in nuclear processes (spindle, nuclear speck) and mitochondrial matrix. These results suggest that prioritizing findings that are consistent between transcriptomic and proteomic states may permit a more precise identification of stable synaptic mechanisms related to SCZ.

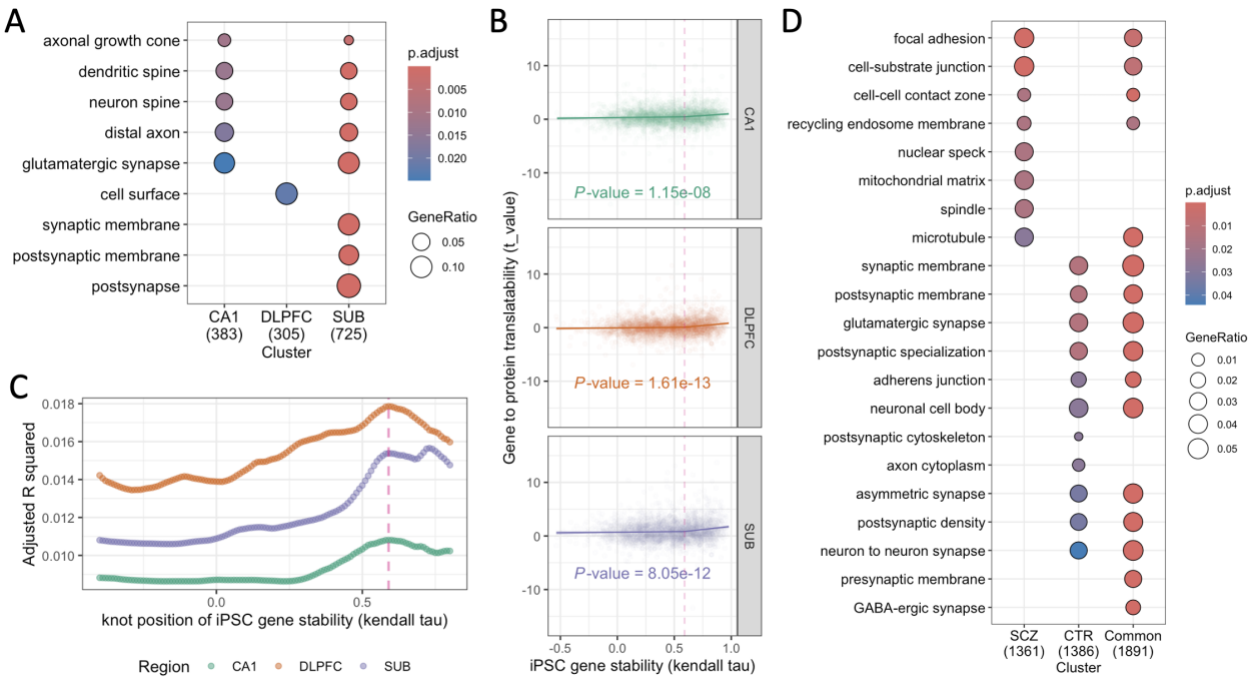

Figure 5. Gene translatability between transcriptome and proteome, and relationship to gene stability in iPSCs. A) The strongest gene ontology terms (biological process) enriched in sets of genes whose expression profiles were significantly similar between transcriptomic and proteomic states. The color scale depicts the adjusted p-value of the ontology enrichment, and the size of the dot depicts the ratio of gene ontology genes in the set. **B)** The relationship between the translatability of a gene between transcriptomic and proteomic states and the stability of a gene in iPSCs. On the y-axis, "Gene to protein translatability" is computed as the t-value arising from a linear model between gene expression and protein expression across subjects, including relevant confounders. On the x-axis, iPSC gene stability is computed as the Kendall tau correlation coefficient between DIV56 expression and DIV70 expression across iPSC samples. Points represent genes. The line represents the best-fit line based on a spline function with a knot at iPSC gene stability = 0.59. **C)** This knot value was shown to result in the highest adjusted R2 for the CA1 and DLPFC and was the second highest local maxima in SUB after knot = 0.72. The y-axis shows the adjusted R2 value when using a specific knot value for iPSC gene stability (found on the x-axis) for the spline regression of "Gene to protein translatability" vs. "iPSC gene stability." **D)** Enrichment for genes found significantly stable from DIV52 to DIV70 (Kendall fdr adjusted $p < 0.05$) using only CTR subjects or only SCZ subjects. Genes found uniquely significant using only CTR subjects-derived stability are labeled "CTR," genes found uniquely significant using only SCZ subjects-derived stability are labeled "SCZ," and genes with significant stability in both are labeled "Common." Abbreviations: DIV. days in vitro; iPSC, induced pluripotent stem cells; SCZ, schizophrenia; CTR, control

Gene co-expression networks identify SCZ risk pathways consistent across transcriptome and proteome

Weighted gene correlation network analysis (WGCNA) was used to identify networks in each region using gene, transcript, and protein data separately. For each network we prioritized modules

that were enriched for SCZ risk, using a method similar to that implemented by Pergola et al., 2023 [6], shown to robustly re-identify previously prioritized SCZ modules from published data. We computed overrepresentation statistics per module for six separate lists of SCZ risk genes, as well as enrichment for SCZ risk using the Multi-marker Analysis of GenoMic Annotation (MAGMA) tool [36], leaving seven separate SCZ risk enrichment measures per module. If a module was significantly enriched in at least three of these measures (Figure 6A, green tiles), we labeled it a SCZ risk module. We only prioritized modules that had significant overlap across at least two expression measures within the same region, using the Jaccard index (JI). For example, a SCZ risk gene module in DLPFC was only considered further if it had a significant JI with another SCZ risk module in the transcript network or the protein network of DLPFC. These modules were generally enriched for excitatory cell types matching the region of origin (Figure 5A, purple tiles), published differentially expressed genes between SCZ patients and controls (DEGs) found in the DLPFC and Caudate [4,5], differentially methylated genes, and loss of function genes (Figure 6A, orange tiles).

We tracked the flow of significant intersections across expression types for each region and identified ontologies for genes shared between transcript and protein modules. This procedure served to profile only the most robustly co-expressed proteins, as well as to parse the protein modules further into more specific isoform resolutions that are better resolved using our transcript isoform modules. We identified consistent ontologies related to the ionotropic glutamate receptor complex across the DLPFC and SUB (Figure 6B). Specific to the subiculum were ontologies generally related to the synapse, specifically the glutamatergic synapse. This suggests that the excitatory synapses may be specifically central to SCZ genetic risk in the glutamatergic neurons

of the mono-synaptic circuit we analyzed here. Five genes were replicated across at least 2 regions; *SLC12A5* between the CA1 and the DLPFC, which is critically implicated in the GABA-A channel switch from excitatory to inhibitory function [37], and *AP3B2*, *GRIA2*, *PLCL2*, *QSOX1* between the DLPFC and SUB.

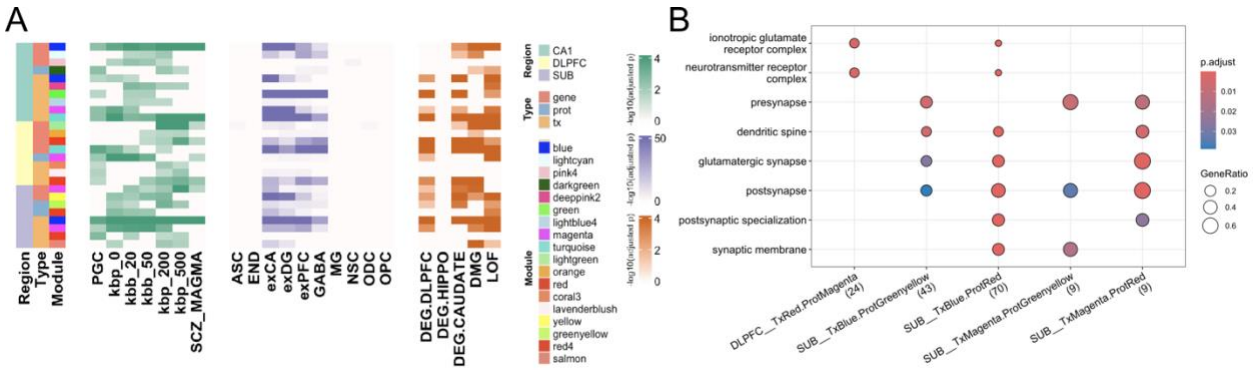

Figure 6. Identification of replicable SCZ co-expression partners. A) Gene enrichment analysis results for modules prioritized by SCZ risk and cross-expression type replicability. Green grids show enrichment results for SCZ risk genes and SCZ MAGMA. Purple grids show enrichment results for cell-type specificity using a human single-cell atlas. Orange grids show enrichment testing results for differentially expressed genes in SCZ-CTR contrasts, differentially methylated genes, and loss of function variant intolerant genes in the green grid. **B)** Gene ontologies of each significant intersection between transcript and protein prioritized modules. In parentheses are the number of genes present in the intersection. The FDR-adjusted p value is depicted with a color scale and the ratio of present ontology genes is depicted by the size of the dot.

Interregional circuit-level effects per subject

As a subject-wise phenotype, we investigated transcriptomic coupling as the across-gene correlation strength between pairs of regions and assessed associations with diagnosis. Transcriptomic coupling was strongly significant for all region pairs in all subjects (all Pearson R > .78, all p < .001) (Figure 7A). We used a two-way ANOVA to interrogate the combined effects of region-pair and diagnosis on transcriptomic coupling, considering relevant confounders. We found no significant diagnosis effect or interaction between diagnosis and region pair. The region pair of SUB with CA1 had significantly higher transcriptomic coupling than the CA1 with DLPFC

($F(2)=7.714$, $p < 0.01$, SUB_CA1-CA1_DLPFC Tukey $p < .001$), consistent with the monosynaptic nature of the relationships of CA1 to SUB and their higher cytoarchitecture similarity.

We then considered transcriptomic coupling at the level of specific subsets of genes. Neuronal CLIC gene sets (e.g., synaptic and dendritic sets) showed lower across gene correlation strength than all other CLIC sets on average (Figure 7B), possibly in association with the inter-regional variability of neurons compared to glia. We interrogated if transcriptomic certain gene sets, such as neuronal ones, showed diagnosis effects, e.g., related with neuronal functional alterations. Figure 7C shows the top three strongest differentially correlated gene sets considering all region pairs. They were all neuronal, arising from the SUB to CA1 connection and nominally significant (all $p < .05$, all $t(20) > 2.6$). Neuronal gene sets in general were descriptively more strongly correlated in CTR subjects. The effect was not significant after multiple comparisons (Figure 7D), therefore we consider it exploratory.

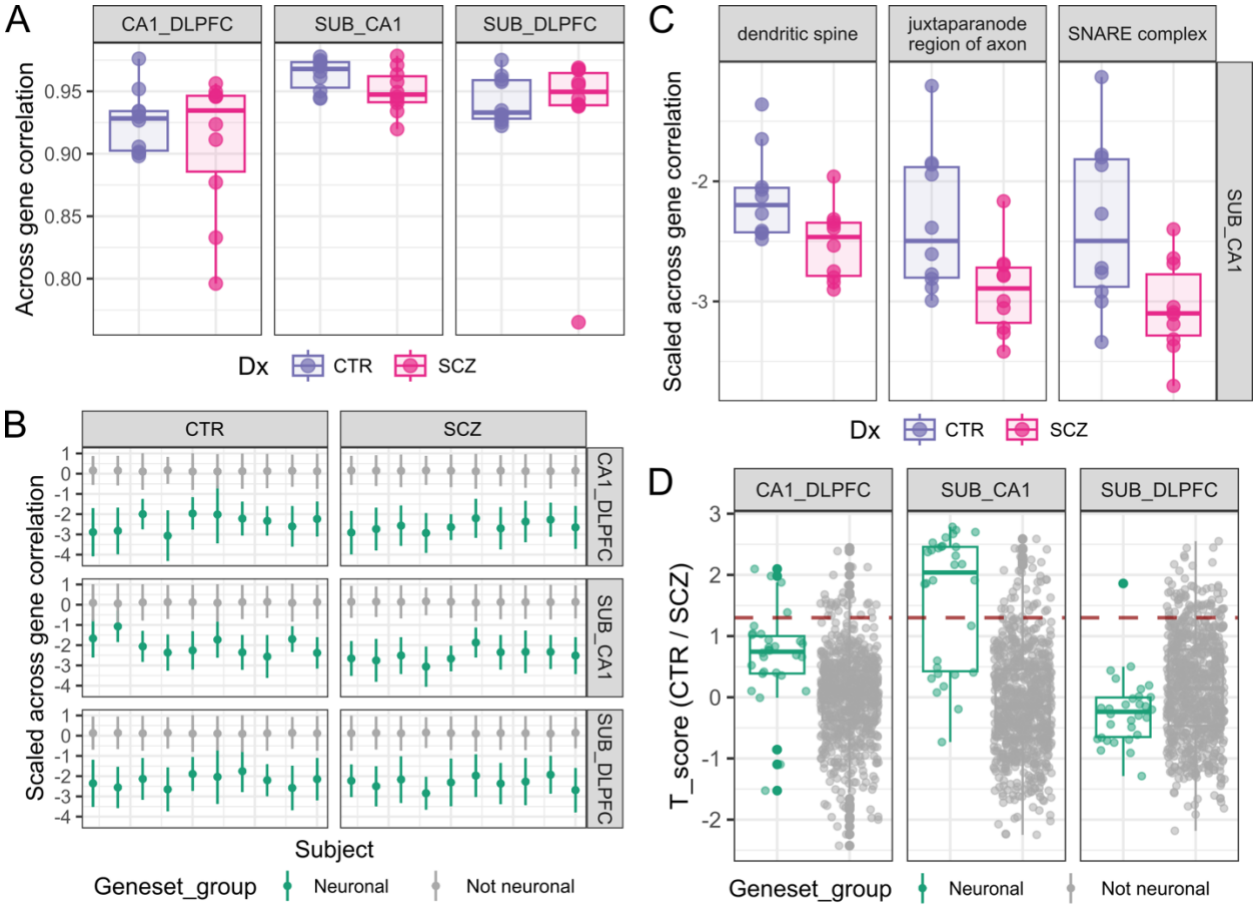

Figure 7. Interregional connectivity per subject. A) To evaluate transcriptomic coupling between region pairs, we computed the across-gene correlation strength (Pearson R) per subject per region pair, computing across the union of all CLIC genes. On the y-axis is the across gene correlation strength. Points are subjects, colored and spread on the x-axis by diagnosis. **B) We computed the** across-gene correlation strength for each CLIC gene set, each time across only the genes within the set. **The values scaled within each subject, found on the y-axis.** Points and error bars representing neuronal gene sets containing neuronally relevant key terms are colored in green, and all other gene sets are colored in grey. The points depict the median, and the bars depict the sd. **C)** For each gene set, we evaluated between subject diagnosis effects on scaled across gene correlations, using a linear model account for subject-wise confounders. The three gene sets with the strongest diagnosis effects are shown here, each showing decreased transcriptomic coupling between the SUB and CA1 in SCZ patients with nominal significance. Points are subjects. **D)** The diagnosis effect t statistic arising from a linear model are shown on the y-axis. The dashed red line reflects nominal significance. Points represent gene sets jittered along the x-axis; neuronal gene sets containing neuronally relevant key terms are colored in green, and all other gene sets are colored in grey.

Directionality analysis of interregional circuit-level effects

To gain more insight into the directionality of circuit effects and their relation to SCZ, we implemented an exploratory analysis evaluating the directionality of inter-regional transcriptomic coupling through between-region across-subject gene co-expression. We tested whether interregional connectivity to excitatory synapses revealed coherent properties about the

directionality of the circuits we were evaluating. The predominant connection in the CA1-SUB hippocampal circuit is from the CA1 to the SUB, though back projections are also present [38]. One can thus hypothesize that CA1 genes influencing excitatory synaptic expression in the SUB should be mainly pre-synaptic (based on predominant projections from the CA1 to SUB), while SUB genes receiving excitatory synaptic expression in CA1 should be mainly dendritic (based on SUB dendrites receiving those CA1 projections). To test this hypothesis, we took the CLIC gene sets for "presynaptic membrane" (representing the pre-synapse of a projecting region) and "dendrite membrane" (representing the post-synapse of a receiving region) of one region and evaluated their connectivity to "excitatory synapse" genes of another region. In the CTR-derived networks, we find our hypothesis was supported by gene set comparison results. The results show that CA1 genes from the "presynaptic membrane" gene set were significantly connected to the "excitatory synapse" gene set in the SUB (Figure 5C, upper fourth panel from left, green vs purple), while "dendrite membrane" genes from the SUB were significantly connected to "excitatory synapse" genes in the CA1 (Figure 8A, upper fifth panel from left, orange vs purple). In SCZ-derived networks, these findings are observed with also a significant but weaker association of dendritic membrane genes from SUB to CA1. We also find strong presynaptic membrane connectivity to excitatory synapses in both directions for the DLPFC-SUB circuit, consistent with reciprocal glutamatergic projections. Connectivity to excitatory synaptic genes was then related to the SCZ risk strength of a gene through MAGMA. SCZ risk at the gene level was associated with connectivity to "excitatory synapse" for every within-region comparison (Figure 8B, 3 leftmost panels). An inter-regional association of SCZ risk to excitatory synapse connectivity was found from the CA1 to the SUB, meaning higher SCZ risk genes in the CA1 are more connected to excitatory synaptic genes in the SUB (Figure 8B, 4th panel from left), but the same was not found

for the reverse direction (Figure 8B, 5th panel from left). In summary, we found consistent SCZ partners strongly enriched for glutamate receptor complex, coupled with reduced SUB_CA1 similarity of excitatory synaptic genes, and a positive association in the CA1 between SCZ risk and connectivity to the SUB excitatory synaptic genes.

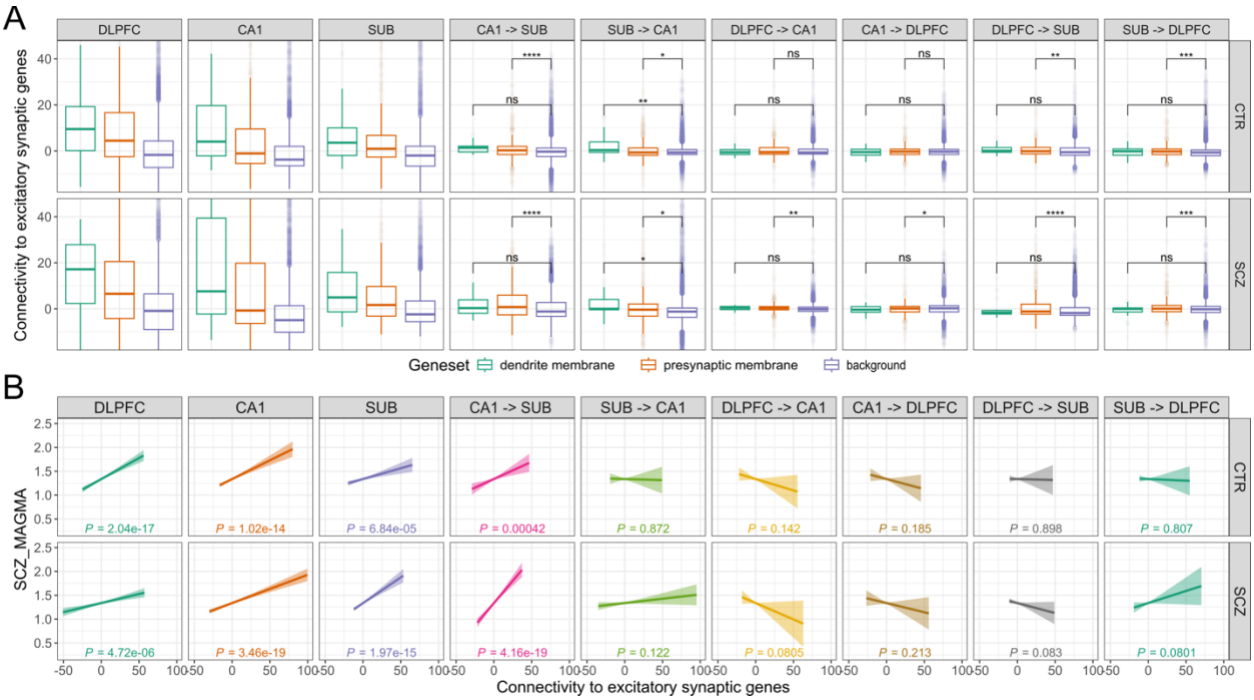

Figure 8. Directionality of interregional connectivity associations to SCZ. A) Comparison of "Connectivity to excitatory synaptic genes" in dendritic membrane and presynaptic membrane by pairwise comparison to background genes in the network. The y-axis depicts the sum of connectivity to excitatory synaptic genes residualized to total connectivity across all genes. Boxplots are representative of populations of genes from each gene set, gene set identity indicated by color. Asterisks indicate p-values from pair-wise comparisons against "Other", using Welch two-sample t-test, FDR corrected (n comparisons = 36, including all comparisons against "Other" in interregional and within-regional networks from both diagnoses); *:pFDR<.05, **:pFDR<.01, ***:pFDR<.001. All within-region comparisons to "Other" are significant pFDR<.001, these are not shown in set. The y-axis is scaled to best visualize gene populations arising from interregional networks, hence some whisker extremities from within-regional networks and some outlier points throughout are not visualized. Facet headings are in format *"Region1 -> Region2". Region1* is where the connectivity per gene was computed to the excitatory synaptic genes of *Region2*. **B)** Association of "Connectivity to excitatory synaptic genes" with "SCZ risk". The y-axis shows SCZ_MAMGA. The x-axis depicts the sum of connectivity to excitatory synaptic genes residualized to total connectivity across all genes. Geom_smooth(method = lm()) was used to visualize the across gene relationship. In set are p_values computed from *lm(SCZ_MAGMA ~ Connectivity to excitatory synaptic genes)* . The leftmost 3 panels show gene connectivity to excitatory synaptic genes within the same region. Facet headings are in format *"Region1 -> Region2". Region1* is where the connectivity per gene was computed to the excitatory synaptic genes of *Region2*.

Discussion

We used LCM to collect excitatory neuron-enriched samples from mono-synaptically linked nodes of a SCZ-associated neural circuit, i.e., from glutamatergic neurons in layer III of DLPFC, in the subiculum, and in the CA1 hippocampal field. Genes with higher expression in LCM compared to bulk were found to be enriched for synaptic gene ontologies. Also, differentially expressed genes (DEGs) in LCM compared to bulk DLPFC were enriched for layer III-specific markers [39]. LCM had a more distinct expression as compared to bulk tissue and pseudo-bulked 10x snRNASeq samples, with reduced dispersion across subjects while retaining similar expression intensity. Using a machine learning approach, LCM predicted regional identity better than bulk tissue expression. Protein and transcript expression predicted regional identity better than gene expression. Although our sample size is too small to compute SCZ-related DEG statistics on a per-gene basis, we used the same prediction paradigm to show that an ensemble of nominal DEGs predicts diagnosis above chance level in the DLPFC when using proteomics but not in the CA1 or SUB. Co-expression analyses showed that LCM is an ideal conduit for the precise evaluation of neuronal pathways in co-expression, with increased co-expression strength of neuronal pathways in LCM compared to bulk tissue samples. We identified SCZ risk modules consistently co-expressed across gene, transcript and protein for each region. Across all regions, consensus SCZ risk partners showed consistent ontologies related to ionotropic glutamate receptor complexes, thus prioritizing altered ionotropic glutamate receptor complex function in excitatory neurons as a likely mechanism driving SCZ risk at least within this interconnected circuit enriched for and these specific projection neurons. We attempted to tie these findings to an expression-based measure of cross-regional similarity of excitatory synaptic genes and found potentially reduced similarity

between the CA1 and subiculum in SCZ patients. We interpret this finding as consistent with decreased signaling within this circuit in patients.

LCM demonstrated greater proportions of specific neuronal excitatory subtypes as assessed via cell type deconvolution. The LCM CA1 was enriched for the HP Excit A snRNA cell type cluster, which Tran and colleagues [20] found to be their HP cell type cluster most similar to a previously described CA1 cluster by Habib et al 2019 [40]. The LCM SUB was enriched for Excit D and Excit F, which could not be referenced against any subiculum cluster, but did show higher enrichment of CA1 and CA3 compared with dentate gyrus neurons, indicating at the very least an enrichment of pyramidal neurons over granule cells. The LCM DLPFC was enriched for the DLPFC Excit A snRNA cell type cluster, which Tran and colleagues found to be their DLPFC cell type cluster most similar to a previously described L4 cluster by Velmeshev et al 2019 [20,41]. Mapping to L4 rather than L3 may be explained by only a mixed L2/L3 cluster having been identified by Velmeshev et al 2019. There appears to be a certain ambiguity in the genetic signature of L3 DLPFC neurons. For instance, Maynard et al. found that layer 3 had the fewest specific genes compared to other layers in their spatial transcriptomic study [3]. As a paper recently showed that the projection targets of a DLPFC subfield greatly influence its transcriptome [42], snRNA-seq clusters of the DLPFC within a layer may be more diverse than hippocampal clusters and more challenging to resolve via computational deconvolution. The uncertain nature of how transcriptomic signatures are derived is one reason one may opt to use techniques such as LCM to parse out specific cell types via visual inspection.

Transcript and protein level analyses outperformed expression at the gene level, showcasing greater precision in discerning distinct brain regions such as the DLPFC, SUB, and CA1. This heightened accuracy also revealed different region-specific pathways that were not apparent with

sole gene expression data, offering insights into the nuanced biological processes and potential applications for targeted interventions. Notably, we did not look at protein pathways due to a lack of GO accuracy when mapping proteins back to the gene level with current reference panels. Furthermore, our machine learning results preliminarily hint that the LCM data in the DLPFC may show diagnosis effects of SCZ, although the sample size in this study does not warrant any conclusion. These findings underscore the importance of considering regional specificity in neurobiological studies.

Our interregional circuit level analysis reflects how different genes are coordinated between regions, a metric shown to relate to functional connectivity [43]. Neuronal gene sets displayed lower transcriptomic coupling than other types of sets within the same individual, which may be explained by the higher similarity between glial types across tissues in comparison to neuronal types, and the greater density of glial cells as compared to neurons. Regardless, our exploratory analysis found that neuronal pathways show nominally significant increased transcriptomic coupling between the SUB and CA1 in CTR subjects as compared with SCZ. We probed the directionality of circuit-level perturbations and found that CA1 genes most connected to excitatory synapses in the SUB are coherent at the circuit level (presynaptic rather than post-synaptic) and more enriched for SCZ risk. Our results suggest that gene-level perturbations in the CA1 and SUB may cause circuit-level perturbations between them and may help explain perturbations of hippocampal circuitry in SCZ [44].

Given the predictive power of protein and transcript over gene in defining regional identity, we decided to use strict criteria of only prioritizing SCZ partner genes that appear in transcript and protein risk modules. We also studied the characteristics of genes in which RNAseq shows high-fidelity for protein approximation and found that genes with higher temporal stability of the mRNA

provide better fits for protein prediction. Therefore, RNA degradation appears a relevant factor in this context, and the use of Pep-seq can help refine gene sets relevant to SCZ. Prioritized genes were enriched in the ionotropic glutamate receptor complex across the DLPFC and SUB subfields, with *GRIA2*, encoding the GluR-2 subunit of the AMPA receptor, identified across the DLPFC and SUB. This finding resonates with a recent review that highlighted consistent decreased expression of *GRIA2* in SCZ across 6/8 published studies of the hippocampus [45].

Other genes replicated as SCZ partners across the DLPFC and the SUB include *QSOX1*, *AP3B2*, and *PLCL2*. *AP3B2* is involved in the formation of clathrin-coated synaptic vesicles, hence reflecting fundamental presynaptic processes. *QSOX1* is a transcriptional regulator with splicing specific alterations in SCZ. Its variant one transcript (ENST00000367602) has been previously shown to have altered expression in both the BA10 and caudate [46], and is the same isoform specifically found as a SCZ partner in both our DLPFC and the SUB networks.

SLC12A5 was found to be a SCZ partner gene in both the DLPFC and the CA1. This gene encodes for the K+-Cl− cotransporter 2 (KCC2), the main Cl^- exporter in neurons [47]. Expression of *SLC12A5* has previously been reported to be affected in SCZ patients, with decreased expression of the full length *SLC12A5* in the hippocampus [37], and increased expression of the *EXON6B* isoform in the DLPFC [48]. *SLC12A5* is linked to the regulation of inhibitory function and is a principal driver of switching excitatory-to-inhibitory GABA polarity through neurodevelopment, a phenomenon associated with SCZ [49]. However, *SLC12A5* also plays a direct role in post-synaptic glutamatergic function, where it has been shown to regulate the formation of glutamatergic synapses through cytoskeletal coupling in dendritic spines [50]. This literature matches our findings, given that “dendritic spine” is enriched in SCZ partner genes found in the SUB. Reduced synapse maturity or reduced synaptic puncta density are consistently reported properties of SCZ-derived

iPSCs [51], and may be the cause or an effect of the reduced number of stable synaptic genes that we identified in SCZ-derived iPSCs. Our study thus suggests that altered regulation of glutamatergic synapse formation by *SLC12A5* is a potential mechanism of SCZ risk.

The small sample size of our study limited our ability to generate per gene or per module diagnosis-driven statistics, such as DEGs or detecting modules diagnosis-associated module eigengenes as is typically done in WGCNA studies and to test for interactions of diagnosis and regional circuitry. Regardless, preliminary results are encouraging for further investigations of interconnected circuits involved in SCZ. As LCM is not a 'single cell' technique at present but depends on pools of similar cells identified either by morphology or specific cell markers, caution is warranted when interpreting our results as reflecting a glutamatergic neuronal circuit. Another limitation is the use of short-read sequencing and computational isoform transcript reconstruction. Yet, our results show that transcript reconstruction still increases information content for regional identity discrimination.

In conclusion, we show that LCM data allow for high accuracy in constructing predictive expression profiles for specific brain regions, potentially revealing disease-relevant pathways otherwise overshadowed by the homogeneous nature of bulk tissue preparation. LCM also allows for high-resolution transcript characterization, which is not available with most snRNAseq technology. Transcript and peptide data give near-perfect region prediction accuracies, adding biological information that gene expression level data lacks. LCM also had stronger co-expression strength of neuronally related gene sets as compared to bulk or 10x snRNAseq, indicating higher potential in interrogating the roles that these pathways play in SCZ. Finally, by computing gene

co-expression networks, we identified SCZ partner genes co-expressed across transcript isoforms and protein modalities that are consistent with other data implicating perturbations in glutamatergic receptor complexes. These SCZ partners may cause circuit-level perturbations, such as between the CA1 and the SUB, where we found decreased transcriptomic coupling in SCZ patients as compared to CTR. We also found that SCZ risk is overrepresented in CA1 genes that are strongly connected to excitatory synapse genes in the SUB, suggesting that the CA1 to SUB glutamatergic circuit may be more affected by risk than the opposite direction in the same path. These genes hint at the perturbation of glutamatergic synapses as a driving mechanism behind SCZ risk and may be of use in developing future treatments.

Materials and Methods

Table 1 below provides detailed demographic information on postmortem human brain subjects and samples used.

Diagnosis	Region	Age	Sex	RIN	PMI	Manner of Death	History of Smoking	Antipsychotic Toxicology detection at Death	Comorbid Substance Use Disorder (type)	Nicotine/ Cotinine Toxicology at death
SCZ	CA1, SUB, DLPFC	29.9	M	8.5	23	Suicide	No	None	No	Negative
SCZ	CA1, SUB, DLPFC	46	M	6.7	22.5	Natural	No	Clozapine	No	Negative
SCZ	CA1, SUB, DLPFC	65.3	M	8.8	15.5	Natural	Yes	Chlorpromazine	No	Positive
SCZ	CA1, SUB, DLPFC	51.2	M	6.7	23	Undetermined	Yes	Olanzapine	Alcohol	Positive
SCZ	CA1, SUB, DLPFC	26.2	M	7.7	28.5	Suicide	Yes	Haloperidol, Olanzapine	Cannabis, Opioid	Positive
SCZ	CA1, SUB, DLPFC	67	M	7.2	36.5	Natural	Yes	Risperidone	Alcohol	Positive
SCZ	CA1, SUB, DLPFC	23.1	M	8.3	32.5	Suicide	No	None	Cannabis, Hallucinogen	Negative
SCZ	CA1, SUB, DLPFC	60.7	M	7.2	28	Natural	No	Lurasidone	Alcohol, Sedative	Negative
SCZ*	CA1, SUB, DLPFC	49	M	6.9	41	Natural	No	Risperidone, Paliperidone	No	Negative
SCZ	CA1, SUB, DLPFC	60.1	M	7.3	38.5	Natural	No	None	No	Negative
CTR	CA1, SUB, DLPFC	49.3	M	8.4	29.5	Accident	No	Negative	No	Negative

CTR	CA1, SUB, DLPFC	42	M	7.5	28.5	Natural	No	Negative	No	Negative
CTR	CA1, SUB, DLPFC	62	M	6.5	29	Natural	No	Negative	No	Negative
CTR	CA1, SUB, DLPFC	51.6	M	8.5	38.5	Accident	Yes	Negative	No	Positive
CTR	CA1, SUB, DLPFC	47.8	M	7.1	25.5	Natural	No	Negative	No	Negative
CTR	CA1, SUB, DLPFC	72.4	M	7.4	27.5	Accident	Yes	Negative	No	Negative
CTR	CA1, SUB, DLPFC	61.8	M	7	22.5	Natural	Yes	Negative	No	Positive
CTR	CA1, SUB, DLPFC	59.9	M	7.4	26	Natural	No	Negative	No	Negative
CTR	CA1, SUB, DLPFC	34.6	M	8.3	36	Accident	No	Negative	No	Negative
CTR	CA1, SUB, DLPFC	52.7	M	8	28	Accident	No	Negative	No	Negative

Table 1. Human brain subject demographics. SCZ=schizophrenia. CTR=neurotypical control. All SCZ subjects were positive and CTR subjects were negative for lifetime antipsychotic use. M=male. Brain regions: DLPFC=dorsolateral prefrontal cortex, CA1=cornu ammonis of hippocampal formation pyramidal CA1 neurons, SUB=large pyramidal neuronal layer of subiculum. RIN= RNA Integrity Number determined from DLPFC, PMI= postmortem interval in hours. AGE is shown in years.

*Sampled from the right hemisphere.

Human postmortem brain tissue acquisition and dissection

Postmortem human brain tissue was obtained as previously described [5]. All brain samples were collected at the LIBD via audiotaped witnessed informed consent with legal next-of-kin at the time of autopsy, at the Office of the Chief Medical Examiner of the state of Maryland under Maryland Department of Health IRB protocol #12-24, and at the Department of Pathology at Western Michigan University Homer Stryker MD School of Medicine under WCG protocol ##20111080. Audiotaped informed consent to study brain tissue was obtained from the legal next of kin on every case collected at LIBD. Details of the donation process and specimen handling have been described previously [5]. After next of kin provided audiotaped informed consent to brain donation, a standardized 36-item telephone screening interview was conducted (the LIBD autopsy questionnaire) to gather additional demographic, clinical, psychiatric, substance abuse, treatment, medical and social history. A psychiatric narrative summary was written for every donor to include data from multiple sources, including the autopsy questionnaire, medical examiner documents (investigative reports, autopsy reports and toxicology testing), macroscopic and microscopic neuropathological examinations of the brain and extensive psychiatric, detoxification and medical record reviews and/or supplemental family informant interviews using the mini-international neuropsychiatric interview. Two board-certified psychiatrists independently reviewed every case to arrive at DSM-V lifetime psychiatric and substance use disorder diagnoses, including SCZ and bipolar disorder, as well as substance use disorders, and if for any reason agreement was not reached between the two reviewers, a third board-certified psychiatrist was consulted.

All donors were free from significant neuropathology, including cerebrovascular accidents and neurodegenerative diseases. Available postmortem samples were selected based on RNA quality ($RIN \geq 6.5$).

A toxicological analysis was performed in each case. The non-psychiatric non-neurological neurotypical individuals had no known history of significant psychiatric or neurological illnesses, including substance abuse. Positive toxicology was exclusionary for neurotypical individuals but not for individuals with psychiatric disorders.

Human postmortem brain processing, dissections, and donor subject details

Postmortem fresh human brain dissections and freezing were performed as described previously[5]. Anatomically, a total of two frozen blocks containing – (i) the dorsolateral prefrontal cortex (Brodmann areas 9/46) and (ii) the subiculum and (iii) cornu ammonis CA1 of the human hippocampal formation at the level of the lateral geniculate nucleus – were dissected out from each of the 20 subject brains. The blocks were dissected out from previously fresh frozen coronal human brain slices using a hand-held dental drill (Cat# UP500-UG33, Brasseler, Savannah, Georgia) as described before [5]. Specific gross anatomical landmarks above were matched to corresponding coronal brain sections from the Allen Human Brain Reference Atlas (https://atlas.brain-map.org) for each subject to ensure consistency.

The demographic data are summarized in Table 2 with demographic information, including sex, ancestry, age, postmortem interval, manner of death and treatment history for both donor diagnostic groups.

Table 2: Summary Table of human brain subjects.

Cohort Description	Schizophrenia	Control
n	10	10
Age at death	47.9 ± 16.3	53.4 ± 11.0
Sex (% male)	100%	100%
Ancestry (% European)	100%	100%
PMI (hours)	28.9 ± 8.2	29.1 ± 4.8
Manner of death (% suicide)	20%	0%
Comorbid Substance Use Disorder	50%	0%
Smoking at Death (Tox)	40%	20%
Lifetime Antipsychotic Use	100%	0%
Tox Positive Antipsychotics at Death	70%	0%
Estimated age of Onset (Schizophrenia)	22.0 ± 7.4	n/a
History of IP Psych Treatment	100%	n/a

Laser capture microdissection

All brain sections were collected using the Leica 350 cryostat precooled to -20C at the level of the lateral geniculate nucleus of the thalamus for the hippocampal formation (including dentate gyrus, CA3/2/1, and subiculum) and BA9/46 DLPFC for every subject (Table 1). Every section was obtained at a thickness ranging between 25-40 microns and transferred onto glass slides coated with precharged PEN membrane (Zeiss # 415190-9081-000). An average of 10 sections corresponding to ~5000 layer 3 pyramidal neuron enrichments, ~5000-7000 pyramidal neurons of the CA1, and ~5000-7000 large pyramidal neuron enrichments of subiculum were harvested from every subject. All sections were uniformly Nissl stained using the ThermoFisher Arcturus Histogene staining kit (KIT0401). Nissl stained sections were compared to anatomical Nissl sections of the human Allen Brain Atlas for DLPFC (BA9/46), CA1 stratum pyramidale, and

subicular formation for the major pyramidal layer (https://atlas.brain-map.org) on every subject to ensure anatomical consistency before neuron enrichments were cut using laser capture microdissection as described previously [3]. Nissl-stained sections were used to determine the boundaries of layer 3 pyramidal neurons. Starting from the apical surface of the neocortex (layer 1), layer 3 pyramidal neurons were distinctly identified by lower-density larger neuronal cell bodies relative to both layers 2 and 4. This approach has been successfully implemented with the Leica laser capture system to cut out the entire layer 3 cells [52]. Here, the Zeiss PALM system additionally offers the feature to selectively highlight neuronal cells within neocortical layers for microdissection, which facilitated selective enrichment for layer 3 pyramidal neurons. Specific pyramidal neurons were highlighted manually and entered into the Zeiss PALM Robo v4.9 LCM system. The Zeiss cool UV laser (355nm) uses photon excitation energy to catapult selected neurons onto Zeiss LCM tube caps (Carl Zeiss™ 415190-9211-000) and therefore never comes in direct contact with the tissue. The no-contact high accuracy Zeiss PALM microbeam features allows for selective enrichment of morphologically defined neurons with corresponding high extraction yields for both RNA and protein from the same neuron lysate. Once the neurons were highlighted, cut, and catapulted onto LCM adhesive caps, the caps were immediately stored on dry ice until lysis and RNA extraction by RNeasy Micro kit (Qiagen). Over 250 ng of total RNA was isolated from the pooled neuron enrichments for each donor. This relatively high output quantity from LCM enables more accurate steady-state mRNA quantification and avoids the amplification bias and computational confounders associated with RNA preamplification. The quantity and integrity of the RNA were determined by NanoDrop and BioAnalyzer (Agilent).

RNA sequencing

Briefly, 200 ng RNA from every sample was utilized for RNA-sequencing without any pre-amplification as described previously [2]. Sequencing libraries were prepared using the TruSeq Stranded Total RNA Library Preparation kit with RiboZero Gold rRNA depletion. For quality control, synthetic External RNA Controls Consortium (ERCC) RNA Mix 1 was spiked into each sample. Samples were balanced for diagnosis within each batch. The resulting paired-end, strand-specific libraries were sequenced on an Illumina HiSeq3000 at the LIBD Genomics Sequencing Facility. FASTQ files were generated using the Illumina Real-Time Analysis module by performing image analysis, base calling, and the BCL Converter (CASAVA v1.8.2). Alignment of reads was performed to the hg38/GRCh38 human genome (Gencode release 25, GRCh38.p7, chromosome only) using HISAT2 (v2.04) and Salmon (v0.7.2) using the reference transcriptome to initially guide alignment based on annotated transcripts. The synthetic ERCC transcripts were quantified with Kallisto (v0.43.0).

RNA Data processing

Counts were generated as described previously. The sorted BAM files from HISAT2 alignments were generated and indexed using SAMtools (v1.6; HTSlibv1.6). The quality of alignment was assessed using RSeQC (v2.6.4). Transcriptomes were characterized for genes and transcripts.

Gene counts were generated using the SubRead utility featureCounts (v1.5.0-p3) for paired-end, reverse-stranded read counting. Transcript counts and transcripts per million (TPM) estimates were generated using Salmon. These estimated transcript count Salmon files were used for downstream differential expression analysis. Quality control of samples was determined based on ERCC concentrations, genome alignment rate, gene assignment rate (>20%), and mitochondrial mapping rate (<6%) as described previously [3].

Direct Comparison of protein profiles by Tandem Mass Tag analysis

Sample preparation. A total of 60 neuronal lysates were processed for proteomics extractions from the organic phase (note that same neuronal lysate supernatant phase were used for RNA extractions and RNA sequencing as described above). Isopropanol precipitated organic phase pellets were washed in absolute ethanol, then resuspended by pipetting in SDS (10%) and DTT (200mM) (Sigma, #43819) at 60C for 1 hour to reduce disulfide bonds. Lysate volumes were made up to 250µl Tri-ethyl ammonium bicarbonate buffer (100mM) (TEAB buffer, Sigma, #T7408) followed by gentle mixing, and incubation at -80C for 4 hours. Thawed lysates were then alkylated by iodoacetamide (250 mM) (Sigma #I6125) in the dark for 15 mins followed by centrifugation and protein quantification via BCA Assay (Pierce). Post-quantification, a total of 50 µg protein per sample was taken for tandem mass tag (TMT) labeling.

TMT Labeling. All 60 protein samples were TCA/Acetone precipitated and pellets resolubilized in TEAB buffer (100mM). Trypsin digestion was performed by adding 10µg Trypsin/ LysC mixture (V5071, Promega) overnight at 37C. Post-digestion, individual samples were then labeled (TMT 10-plex, ThermoFisher Scientific) so every sample had a unique isobaric mass tag, as per kit instructions. All samples within each of the 6 batches (10 samples per batch) were randomly paired and labeled. Each 10-plex batch had 5 control and 5 SCZ samples of similar brain regions to minimize variability and for better analysis design. Briefly, the TMT-10 plex reagents (0.8µg/vial) were brought to room temperature, and 41µl of anhydrous acetonitrile was added to each vial, vortexed, and centrifuged. An entire vial was added to each sample of 100ug tryptic peptide and left at room temperature for 1 hour. Next, 8µl of hydroxylamine (5%) was added to quench the reaction, followed by multiplex combining of batches (10 samples/batch) and dry

vacuuming them by centrifugation. The combined TMT labeled peptides were resuspended in 10mM TEAB and separated using an acetonitrile gradient (0-90% in 10mM TEAB on a 5µm C18 Waters XBridge column using Agilent 1200 capillary HPLC in normal flow mode and Agilent 1260 micro-fraction collector) into 84 fractions, further concentrated into 24 fractions by combining all odd rows and all even rows into 12 fractions each.

Tandem Mass Spectrometry – Peptide fractions were resuspended in 200µl of 0.1% trifluoroacetic acid loaded and desalted on a Waters Oasis plate C18. Next, peptides were eluted in basic (10 mM TEAB) in steps at 5, 10, 25, and 75% acetonitrile. Fractions were dried and rehydrated/ resuspended in 20 µL loading buffer (2% acetonitrile in 0.1% formic acid) and analyzed by reverse phase liquid chromatography interfaced with tandem mass spectrometry (LC/MSMS) using an Easy-LC 1200 HPLC system1 interfaced with an Orbitrap Fusion Lumos Tribrid Mass Spectrometer (ThermoFisher Scientific). Peptides (20% each fraction) were loaded onto a C18 trap (S-10 µM, 120 Å, 75 µm × 2 cm, YMC, Japan) and subsequently separated on an in-house packed PicoFrit column (75 µm × 200 mm, 15 µ, +/−1 µm tip, New Objective) with C18 phase (ReproSil-Pur C18-AQ, 3 µm, 120 Å2) using 2–90% acetonitrile gradient at 300 nl/min over 120 min. Eluting peptides were sprayed at 2.0 kV directly into the Lumos. Survey scans (full ms) were acquired from 300 to 1700 m/z with data-dependent monitoring of up to 15 peptide masses (precursor ions), each individually isolated in a 1.2 Da window and fragmented using HCD activation collision energy 32 and 30 s dynamic exclusion. Precursor and the fragment ions were analyzed at resolutions 70 000 and 35 000, respectively, with automatic gain control (AGC) target values at 3e6 with 50 ms maximum injection time (IT) and 1e5 with 200 ms maximum IT, respectively.

Isotopically resolved masses in the precursor (MS) and fragmentation (MS/MS) spectra were analyzed with Proteome Discoverer (PD) software (v2.4, Thermo Scientific). All data were

searched by Mascot (2.6.23) against the Refseq 2017_Complete database (57 479 sequences for taxonomy Homo sapiens) using criteria: sample's species; trypsin digestion allowing one missed cleavage; N-terminal TMT label as fixed modifications; TMT label on lysine, methionine oxidation, asparagine, and glutamine deamidation, and cysteine carbamidomethylation as variable modifications. Peptide identifications were filtered at 1% false discovery rate (FDR) confidence threshold, based on a concatenated decoy database search, using the Proteome Discoverer. Proteome Discoverer uses only the peptide identifications with the highest Mascot score for the same peptide matched spectrum from the different extraction methods.

Protein quantification for analyses. Briefly, peptides identified with a 1% FDR confidence threshold as described above were considered for analysis. Reporter ions from MS/MS spectra with isolation interferences ≥ 30% were excluded from further analysis. Protein relative abundances were determined from the normalized log2 values of the above spectra. The reporter ion intensities were first transformed into log2 notation and for each sample the median value for each peptide was taken to represent that peptide. The values of all samples were then quantile normalized to minimize technical variation, such as differential amounts of material loading, prior to fold change analysis.

Pre-processing

Gene and transcript-level mRNA expression was quantified as Transcripts per Million (TPMs) and annotated as total gene expression separately for each brain region, regardless of alternative transcript quantification using GENCODE release 25 (GRCh38.p7). We selected genes above median TPM of 0.1 and free of floor effects (maximum 0% of zeroes per gene); we log-

transformed TPM values with an offset of 1, i.e., log2(TPM+1). After removing mitochondrial genes, datasets included a variable number of genes for different regions.

For proteins, raw abundance values were divided by the total sum of protein abundance across all genes, then multiplied by a million, giving counts per million (CPM). These values were log-transformed with an offset of 1 to give logCPM.

CIBERSORTx deconvolution

SingleCellExperiment files containing snRNA-seq data for the hippocampus and the DLPFC were taken from Tran et al., 2021 at https://github.com/LieberInstitute/10xPilot_snRNAseq-human, using the links under the "Processed data" header. For each region, the SingleCellExperiment was used to construct a matrix for cell-type expression (nuclei per column, gene per row) with columns annotated for cell-type, that was input into CIBERSORTx to create a signature matrix of cell-type expression profiles, using default parameters (min expression = .75, replicates = 5, sampling = .5). The resulting signature matrix was to compute cell-type proportions from bulk and LCM RPKM data (BULK DLPFC and HP [6], and LCM DG [3]), using 100 permutations, without batch correction.

Differential gene expression analysis and expression dispersion in LCM

To identify significantly differentially expressed genes between LCM and bulk tissue samples, we performed Differential Gene Expression (DGE) analysis by using linear mixed-effects modeling using the *voom* function from the *limma* package [53]. To control technical variations such as differences in library preparation methods, we normalized the data by applying correction factors to several key metrics. These included i) The rate at which reads mapped to the mitochondrial chromosome; ii) The proportion of reads assigned to rRNA (ribosomal RNA); iii) The overall

percentage of reads that aligned to the reference genome and iv) The fraction of reads that mapped to exonic (protein-coding) regions. This normalization step helps to ensure that any biological differences we observe are not confounded by variations introduced during sample processing and sequencing. For each brain region, we compared the gene expression profiles between the corresponding LCM and bulk samples: Bulk DLPFC (n = 25587) vs LCM DLPFC (n = 22782), Bulk HP (n = 27079) vs LCM CA1 (n = 23416), Bulk HP (n = 27079) vs LCM SUB (n = 23454). Genes with an adjusted p-value less than 0.05 were considered statistically significant and were selected for further analysis.

We designated genes as annotated to layer 3 based on a spatial transcriptomics analysis of the DLPFC by Maynard et al 2021, using their Supplementary Table 5 [39]. Their approach identified genes that show significantly increased expression to one or more layers (e.g. L2 and L3) as compared to all other layers (e.g. L1, L4, L5, L6, white matter). We considered genes as annotated to layer 3 as long as there was a significant association to layer 3, even if the significant association includes other layers. Figure 2C compares the LCM/Bulk log2(FC) values between genes annotated to L3 vs genes not annotated to L3.

For each set of differential expression analysis (Bulk DLPFC (deg = 4336) vs LCM DLPFC (deg = 2607), Bulk HP (deg = 4295) vs LCM CA1 (deg = 2200), and Bulk HP (deg = 3512) vs LCM SUB (deg = 1774), the significant genes were split into two groups based on the sign of their log fold change (logFC) values. Genes with positive logFC values indicate higher expression in the LCM samples compared to the bulk samples. Genes with negative logFC values indicate lower expression in the LCM samples compared to the bulk samples. Then, we combined all significant genes from each set and performed GO enrichment analysis to identify over-represented cellular component terms in each group with background universe being all genes from each set. Resulting

p-values from the GO enrichment analysis were adjusted for multiple testing using Benjamini-Hochberg correction to control the false discovery rate (FDR). We aimed to identify genes and cellular components that are differentially expressed between LCM and bulk samples in the DLPFC, CA1, and SUB regions. Figure 2D provides a visual summary of the top enriched terms and their associated gene ratios and significance levels.

Machine Learning Prediction

We studied RNA-seq and Pep-seq data obtained through LCM from post-mortem brain tissues of ten patients with SCZ and ten neurotypical controls (CTR). As a comparison, we included RNA-seq data obtained through bulk tissue from 30 SCZ patients and 30 NC, matched to the demographic characteristics of the LCM sample. After data preprocessing, we employed high-quality gene, transcript, and Pep-seq reads derived from the DLPFC and hippocampal subregions, SUB and CA1, to assess regional identity and compare it to bulk tissue data from the DLPFC and the hippocampus (HP). We determined differentially expressed genes in a leave-one-out cross-validation framework and used a Random Forest classifier to predict the region of the left-out sample in 100 iterations to generate reliable distributions. We also compared prediction accuracies within the same samples between the different omics data. Finally, we determined gene-protein correlation within the most predictive transcripts and peptides across subjects.

Network identification

We regressed out age, mitochondrial mapping rate, rRNA rate, gene mapping rate from gene and transcript level logTPM data, and regressed out age, total_count, and sequencing_batch from

protein level logCPM data. We rank-normalized residuals using Blom formula [54-57] to limit the impact of deviations from normality in expression data [58].

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 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:* 15,000, *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 $R^2 > .8$ in the correlation between original and log-transformed values [59]. Lower β values are often associated with greater network median connectivity. For each network, we selected the parameter β such that median connectivity was equal across all networks [6].

Co-expression strength of CLIC sets

CLIC CEM sets were downloaded from https://gene-clic.org/clic/precomputed, comprising 771 sets in total. Mouse MGI.symbol were converted to ensemble.gene.ids using the *getLDS()* function from the *biomart* package. Only genes in the CLIC CEM set that were also in the original ontology list were considered as part of a "CLIC geneset" (CEM.CEM. == "CEM"). A correlation matrix was computed for each dataset through pairwise Pearson correlation of Blom-normalized residual expression. For each CLIC gene set, we cropped the correlation matrix to only genes in the CLIC gene set and to median value of the cropped matrix's column means to gauge the strength of a CLIC sets connection to itself. This initial co-expression strength value was compared to

connectivity to the background, taking into account gene length, GC content and median expression confounders. To this end, we computed a null distribution of co-expression strength, where the rows of the cropped matrix were resampled with genes having confounder values within the same range of values present in the original CLIC set. A final co-expression strength value was then computed in terms of a z_score in relation to the null:

$$Coexpression\ strength = \frac{initial\ coexpression\ strength - mean\ null\ coexpression\ strength}{sd\ null\ coexpression\ strength}$$

This Co-expression strength value is the one referred to in Figure 3.

Given that our samples are methodologically enriched for glutamatergic neurons and that our SCZ partner genes are ontologically enriched for glutamatergic synapse, we decided to focus on across regional co-expression relative to excitatory synaptic function using the CLIC set of "excitatory synapse". For each gene, we computed the across-subject Kendall's correlation of one region against another, resulting in correlation matrices of interregional connectivity providing pairwise connectivity values of each and every gene in one region to each corresponding gene in another region. Interregional connectivity matrices were computed separately using healthy controls or SCZ patients. From each interregional connectivity matrix, we computed the sum of connectivity of each gene from one region to the genes from the CLIC "excitatory synapse" set in the other region, and vice versa (e.g., CA1 gene's connectivity to excitatory synaptic genes of the SUB). The same was done for within-region connectivity matrices (e.g., the CA1 gene's connectivity to excitatory synaptic genes of the CA1).

Gene to protein translatability

A gene-to-protein translatability was evaluated per gene as a robust linear model using gene logTPM values to predict protein logCPM values, including confounders of age, diagnosis,

mitochondrial_gene_mapping rate (gene level), gene mapping rate, total_protein_count, and protein_sequencing_batch, using the rlm() function from the *MASS* package on R. The translatability score was considered as the t value associated with logCPM in this model.

iPSC stability

The iPSC generation pipeline has been described in detail elsewhere (Page et al., 2021). Ninety-four human neuronal samples with 56 to 70 days in vitro (DIV) belonging to 26 male participants of European ancestry (14 NC and 12 patients with SCZ) were selected after outlier detection and gene filtering as described in (Pergola, Parihar, et al., 2023). We further restricted the neuronal samples to pairs present both DIV 56 and 70 (matching by the 'SAMPLE ID' variable in the summarisedExperiment object), resulting in a final 27 samples (14 NC and 13 SCZ) arising from 12 individuals (6 NC and 6 patients with SCZ). Keeping human genes only, we calculated per gene quantile normalized values from the logRPKM assay. To obtain the gene stability measure, for each gene we computed the across-subjects Kendall's tau correlation coefficient between quantile-normalized gene expression at DIV56 vs DIV70 [cor.test in R; method = "Kendall", exact=FALSE]).

Gene wise MAGMA

To calculate MAGMA, we used the MAGMA tool v1.09b [36], PGC3 summary statistics [60] as SNP p-value data, and 1000G European as the reference data file for a European ancestry population to estimate linkage disequilibrium between SNPs. We took the following steps: i) we mapped 1000G SNPs to genes encompassed in each network module (a window of 35-kb upstream and 10-kb

downstream of each gene ii) we calculated gene MAGMA Z_scores based on PGC3 SNPs p-values.

SCZ risk enrichment

SCZ risk genes were those overlapping with 6 varying extension windows around PGC3 GWAS significant SNPs: PGC3 [120 genes], 0 kbp [178 genes] (meaning, genes that overlapped with the index SNP), 20 kbp [299 genes], 50 kbp [456 genes], 200 kbp [1196 genes], 500 kbp [2475 genes] (500 kbp was the maximum extension where enrichments were found to be significant [61]). We assessed the overrepresentation of SCZ risk genes for each module except grey in each network relative to a universe comprised of all genes. We corrected results for multiple comparisons via Bonferroni (number of non-grey modules in each network). Module enrichment of MAGMA was computed using the *genesettest* () from the *limma* package. We labeled as SCZ risk modules those significantly enriched in at least three of the 7 SCZ risk enrichment statistics.

Jaccard Index analyses

To compute module continuity across expression types, Jaccard Indices (JIs) were computed as the intersection/union of the considered sets. A null distribution for each module pair was computed by resampling each module from their respective universe and recomputing the JI index 10,000 times. A module pair was considered to have significant overlap if the real JI was greater than the max value across the 10,000 permutations, equating to an empirical p_value < .0001.

Gene-transcript-protein flow

We constructed Sankey plots illustrating the flow of expression across different expression types. Intersection with significant JIs were colored in. Only modules significant for SCZ risk and with significant overlap with a module of another expression type within the same region are shown. The universe was set to the intersection between all network genes in tx network and all network genes in protein networks. Significant intersections between transcript and protein were evaluated for gene ontologies using the *compareCluster* (fun = enrichGO, ont = "CC", universe = uni) function from the *clusterProfiler* package, were uni is the intersection between tx and protein universes.

Circuit level effects across regions

We defined transcriptomic coupling as the within-subject across-gene correlation (Pearson R) of log2(TPM + 1) across all genes expressed in both the regions. Transcriptomic coupling was also computed for CLIC gene sets; each set was taken from CLIC using genes with CEM.CEM. == "CEM" and CEM.CEM. == "CEM +" parameters. Diagnosis effects of transcriptomic coupling were evaluated by comparing differences in the transcriptomic coupling value between CTR and SCZ subjects using a *lm ()* taking into account subject-wise confounders of age, mitochondrial mapping rate, rRNA rate, gene mapping rate. The non-scaled Pearson R value was used when comparing transcriptomic coupling that was computed across all genes. Since geneset-wise transcriptomic coupling is biased within each subject towards the value computed across all genes, we first scaled the values across all gene sets within a subject before comparing across subjects to evaluate diagnosis effects.

Per gene interregional connectivity to excitatory synaptic genes, association with SCZ risk

For each gene we computed an across-subject Kendall's correlation of one region against another, resulting in correlation matrices of interregional connectivity providing pairwise connectivity values of each and every gene in one region to each and every gene in another region. Kendall's correlation was used on Blom-normalized residuals to compute the matrices. Interregional connectivity matrices were computed separately using healthy controls or SCZ patients. We then computed connectivity to the CLIC gene set of "excitatory synapse" was taken considering both original ontology genes as well as those found by CLIC to be significantly co-expressed with them (CEM.CEM. == "CEM" and CEM.CEM. == "CEM +"), comprising 1,324 genes total. From each interregional connectivity matrix, we computed the sum of the connectivity of each gene from one region to the excitatory synaptic genes in the other region and vice versa (e.g., the CA1 gene's connectivity to excitatory synaptic genes of the SUB). The same was done for within-region connectivity matrices (e.g., the CA1 gene's connectivity to excitatory synaptic genes of the CA1). Comparison of "Connectivity to excitatory synaptic genes" in the dendritic membrane, presynaptic membrane, and excitatory synapse gene sets was achieved by pairwise comparison by all other genes in the network, using Welch two-sample t-test, FDR corrected (n comparisons = 54, including all comparisons against "Other" in interregional and within-regional networks from both diagnoses). Each set was taken from CLIC using genes with CEM.CEM. == "CEM" and CEM.CEM. == "CEM +" parameters, and removing gene overlapping between sets, and not overlapping with "excitatory synapse" genes. The final number of genes in each set was as follows; dendrite membrane = 42, presynaptic membrane = 705. Connectivity to excitatory synaptic genes was then related to SCZ risk strength of a gene through MAGMA using a linear model *lm (SCZ_MAGMA ~ Connectivity to excitatory synaptic genes)*.

References

1. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**(2022).
2. Jaffe, A.E. Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. (2018).
3. Jaffe, A.E. *et al.* Profiling gene expression in the human dentate gyrus granule cell layer reveals insights into schizophrenia and its genetic risk. *Nat Neurosci* **23**, 510-519 (2020).
4. Benjamin, K.J.M. *et al.* Analysis of the caudate nucleus transcriptome in individuals with schizophrenia highlights effects of antipsychotics and new risk genes. *Nature Neuroscience* **25**, 1559-1568 (2022).
5. Collado-Torres, L. Regional Heterogeneity in Gene Expression, Regulation, and Coherence in the Frontal Cortex and Hippocampus across Development and Schizophrenia. (2019).
6. Pergola, G. *et al.* Consensus molecular environment of schizophrenia risk genes in coexpression networks shifting across age and brain regions. *Sci Adv* **9**, eade2812 (2023).
7. Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nature neuroscience* **19**(2016).
8. Gandal, M. *et al.* Transcriptome-wide isoform-level dysregulation in ASD, schizophrenia, and bipolar disorder. *Science (New York, N.Y.)* **362**(2018).
9. Hartl, C.L. *et al.* Coexpression network architecture reveals the brain-wide and multiregional basis of disease susceptibility. *Nat Neurosci* **24**, 1313-1323 (2021).
10. Radulescu, E. *et al.* Identification and prioritization of gene sets associated with schizophrenia risk by co-expression network analysis in human brain. *Molecular psychiatry* **25**(2020).
11. Pergola, G. *et al.* DRD2 co-expression network and a related polygenic index predict imaging, behavioral and clinical phenotypes linked to schizophrenia. *Translational Psychiatry* **7**(2017).
12. Pergola, G. *et al.* Prefrontal Coexpression of Schizophrenia Risk Genes Is Associated With Treatment Response in Patients. *Biological psychiatry* **86**(2019).
13. Antonucci, L. *et al.* O5. Classification of Schizophrenia Using Machine Learning With Multimodal Markers. *Biological Psychiatry* **85**(2019).
14. Selvaggi, P. *et al.* Genetic Variation of a DRD2 Co-expression Network is Associated with Changes in Prefrontal Function After D2 Receptors Stimulation. *Cerebral Cortex* **29**, 1162-1173 (2019).
15. Bhattacharya, A. *et al.* Isoform-level transcriptome-wide association uncovers genetic risk mechanisms for neuropsychiatric disorders in the human brain. *Nat Genet* **55**, 2117-2128 (2023).
16. D'Ambrosio, E. *et al.* A polygenic score indexing a DRD2-related co-expression network is associated with striatal dopamine function. *Scientific Reports* **12**, 1-9 (2022).

17. Taurisano, P. *et al.* The interaction between cannabis use and a CB1-related polygenic co-expression index modulates dorsolateral prefrontal activity during working memory processing. *Brain Imaging Behav* **15**, 288-299 (2021).
18. Aryal, S. *et al.* Deep proteomics identifies shared molecular pathway alterations in synapses of patients with schizophrenia and bipolar disorder and mouse model. *Cell Rep* **42**, 112497 (2023).
19. Batiuk, M.Y. *et al.* Upper cortical layer-driven network impairment in schizophrenia. *Sci Adv* **8**, eabn8367 (2022).
20. Tran, M.N. *et al.* Single-nucleus transcriptome analysis reveals cell-type-specific molecular signatures across reward circuitry in the human brain. *Neuron* **109**, 3088-3103.e5 (2021).
21. Ruzicka, W.B. *et al.* Single-cell multi-cohort dissection of the schizophrenia transcriptome. *medRxiv*, 2022.08.31.22279406 (2022).
22. Nagy, C. *et al.* Single-Cell Genomic Strategies to Understand Psychopathological Processes in Suicide and Associated Psychopathology. *Biological Psychiatry* **89**(2021).
23. Conte, M.I., Fuentes-Trillo, A. & Domínguez Conde, C. Opportunities and tradeoffs in single-cell transcriptomic technologies. *Trends Genet* **40**, 83-93 (2024).
24. Espina, V. *et al.* Laser-capture microdissection. *Nat Protoc* **1**, 586-603 (2006).
25. Jaffe, A.E. *et al.* qSVA framework for RNA quality correction in differential expression analysis. (2017).
26. DR, W., KF, B., R, S. & EF, T. Evidence of dysfunction of a prefrontal-limbic network in schizophrenia: a magnetic resonance imaging and regional cerebral blood flow study of discordant monozygotic twins. *The American journal of psychiatry* **149**(1992).
27. Goldman-Rakic, P.S., Selemon, L.D. & Schwartz, M.L. Dual pathways connecting the dorsolateral prefrontal cortex with the hippocampal formation and parahippocampal cortex in the rhesus monkey. *Neuroscience* **12**, 719-43 (1984).
28. Barbas, H. & Blatt, G.J. Topographically specific hippocampal projections target functionally distinct prefrontal areas in the rhesus monkey. *Hippocampus* **5**, 511-33 (1995).
29. Walker, R.L. *et al.* Genetic Control of Expression and Splicing in Developing Human Brain Informs Disease Mechanisms. *Cell* **179**, 750-771.e22 (2019).
30. Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nat Genet* **50**, 538-548 (2018).
31. Schneider, M. *et al.* Altered DLPFC-Hippocampus Connectivity During Working Memory: Independent Replication and Disorder Specificity of a Putative Genetic Risk Phenotype for Schizophrenia. *Schizophr Bull* **43**, 1114-1122 (2017).
32. Zugman, A. *et al.* Reduced dorso-lateral prefrontal cortex in treatment resistant schizophrenia. *Schizophr Res* **148**, 81-6 (2013).
33. Meyer-Lindenberg, A.S. *et al.* Regionally specific disturbance of dorsolateral prefrontal-hippocampal functional connectivity in schizophrenia. *Arch Gen Psychiatry* **62**, 379-86 (2005).
34. Ding SL, Royall JJ, Sunkin SM, et al. Comprehensive cellular-resolution atlas of the adult human brain [published correction appears in J Comp Neurol. 2017 Feb 1;525(2):407.

doi: 10.1002/cne.24130]. J Comp Neurol. 2016;524(16):3127-3481. doi:10.1002/cne.24080

35. Page, S.C. *et al.* Electrophysiological measures from human iPSC-derived neurons are associated with schizophrenia clinical status and predict individual cognitive performance. *bioRxiv* (2021).
36. de Leeuw CA, M.J., Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data | PLOS Computational Biology. (2015).
37. Hyde, T.M. *et al.* Expression of GABA signaling molecules KCC2, NKCC1, and GAD1 in cortical development and schizophrenia. *J Neurosci* **31**, 11088-95 (2011).
38. Sun, Y. *et al.* CA1-projecting subiculum neurons facilitate object-place learning. *Nat Neurosci* **22**, 1857-1870 (2019).
39. Maynard, K.R. *et al.* Transcriptome-scale spatial gene expression in the human dorsolateral prefrontal cortex. *Nat Neurosci* **24**, 425-436 (2021).
40. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat Methods* **14**, 955-958 (2017).
41. Velmeshev, D. *et al.* Single-cell genomics identifies cell type-specific molecular changes in autism. *Science* **364**, 685-689 (2019).
42. Arion, D., Enwright, J.F., Gonzalez-Burgos, G. & Lewis, D.A. Differential gene expression between callosal and ipsilateral projection neurons in the monkey dorsolateral prefrontal and posterior parietal cortices. *Cereb Cortex* **33**, 1581-1594 (2023).
43. Richiardi, J. *et al.* BRAIN NETWORKS. Correlated gene expression supports synchronous activity in brain networks. *Science* **348**, 1241-4 (2015).
44. Knight, S. *et al.* Hippocampal circuit dysfunction in psychosis. *Transl Psychiatry* **12**, 344 (2022).
45. Yonezawa, K. *et al.* AMPA receptors in schizophrenia: A systematic review of postmortem studies on receptor subunit expression and binding. *Schizophr Res* **243**, 98-109 (2022).
46. Cohen, O.S. *et al.* Transcriptomic analysis of postmortem brain identifies dysregulated splicing events in novel candidate genes for schizophrenia. *Schizophr Res* **142**, 188-99 (2012).
47. Fukuda, A. & Watanabe, M. Pathogenic potential of human SLC12A5 variants causing KCC2 dysfunction. *Brain Res* **1710**, 1-7 (2019).
48. Tao, R. *et al.* Transcript-specific associations of SLC12A5 (KCC2) in human prefrontal cortex with development, schizophrenia, and affective disorders. *J Neurosci* **32**, 5216-22 (2012).
49. Toritsuka, M. *et al.* Developmental dysregulation of excitatory-to-inhibitory GABA-polarity switch may underlie schizophrenia pathology: A monozygotic-twin discordant case analysis in human iPS cell-derived neurons. *Neurochem Int* **150**, 105179 (2021).
50. Llano, O. *et al.* KCC2 regulates actin dynamics in dendritic spines via interaction with β-PIX. *J Cell Biol* **209**, 671-86 (2015).
51. Räsänen, N., Tiihonen, J., Koskuvi, M., Lehtonen, Š. & Koistinaho, J. The iPSC perspective on schizophrenia. *Trends Neurosci* **45**, 8-26 (2022).
52. MacDonald ML, Favo D, Garver M, et al. Laser capture microdissection-targeted mass spectrometry: a method for multiplexed protein quantification within individual layers

of the cerebral cortex. Neuropsychopharmacology. 2019;44(4):743-748. doi:10.1038/s41386-018-0260-0

53. Ritchie, M.E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47 (2015).
54. Pergola, G. *et al.* DRD2 co-expression network and a related polygenic index predict imaging, behavioral and clinical phenotypes linked to schizophrenia. *Transl Psychiatry* **7**, e1006 (2017).
55. Taurisano, P. *et al.* The interaction between cannabis use and a CB1-related polygenic co-expression index modulates dorsolateral prefrontal activity during working memory processing. *Brain Imaging Behav* (2020).
56. Antonucci, L.A. *et al.* Thalamic connectivity measured with fMRI is associated with a polygenic index predicting thalamo-prefrontal gene co-expression. *Brain Struct Funct* **224**, 1331-1344 (2019).
57. Fazio, L. *et al.* Transcriptomic context of DRD1 is associated with prefrontal activity and behavior during working memory. *Proc Natl Acad Sci U S A* **115**, 5582-5587 (2018).
58. Kumari, S. *et al.* Evaluation of gene association methods for coexpression network construction and biological knowledge discovery. *PLoS One* **7**, e50411 (2012).
59. Zhang, B. & Horvath, S. A general framework for weighted gene co-expression network analysis. *Stat Appl Genet Mol Biol* **4**, Article17 (2005).
60. Trubetskoy, V. *et al.* Mapping genomic loci implicates genes and synaptic biology in schizophrenia. *Nature* **604**, 502-508 (2022).
61. Pergola, G. *et al.* Prefrontal Coexpression of Schizophrenia Risk Genes Is Associated With Treatment Response in Patients. *Biol Psychiatry* **86**, 45-55 (2019).