**SCHIZOPHRENIA**

The *DGCR5* long noncoding RNA may regulate expression of several schizophrenia-related genes

Qingtuan Meng¹, Kangli Wang¹, Tonya Brunetti²,³, Yan Xia¹, Chuan Jiao¹, Rujia Dai¹, Dominic Fitzgerald², Amber Thomas², Lindsey Jay², Heather Eckart², Kay Grennan⁴, Yuka Imamura-Kawasawa⁵,⁶, Mingfeng Li⁶, Nenad Sestan⁶, Kevin P. White²,³, Chao Chen¹,⁸, Chunyu Liu¹,⁴,⁹*

A number of studies indicate that rare copy number variations (CNVs) contribute to the risk of schizophrenia (SCZ). Most of these studies have focused on protein-coding genes residing in the CNVs. Here, we investigated long noncoding RNAs (lncRNAs) within 10 SCZ risk–associated CNV deletion regions (CNV-lncRNAs) and examined their potential contribution to SCZ risk. We used RNA sequencing transcriptome data derived from postmortem brain tissue from control individuals without psychiatric disease as part of the PsychENCODE BrainGVEX and Developmental Capstone projects. We carried out weighted gene coexpression network analysis to identify protein-coding genes coexpressed with CNV-lncRNAs in the human brain. We identified one neuronal function–related coexpression module shared by both datasets. This module contained an lncRNA called *DGCR5* within the 22q11.2 CNV region, which was identified as a hub gene. Protein-coding genes associated with SCZ genome-wide association study signals, de novo mutations, or differential expression were also contained in this neuronal module. Using *DGCR5* knockdown and overexpression experiments in human neural progenitor cells derived from human induced pluripotent stem cells, we identified a potential role for *DGCR5* in regulating certain SCZ-related genes.

**INTRODUCTION**

Schizophrenia (SCZ) is a severe psychiatric disorder with a complex genetic basis and high heritability (1, 2). It affects about 1% of the world’s population and causes considerable social and economic burden (3). Studies have indicated that rare copy number variations (CNVs) have strong effects on SCZ risk (4, 5). Recently, the largest SCZ CNV study ever, based on 21,094 SCZ cases and 20,227 control samples, identified more than 10 SCZ risk–associated CNVs (6). However, the mechanisms by which these CNVs contribute to SCZ risk are not clear. CNVs may increase SCZ risk by changing gene dosage or disrupting the sequences of genes within the deleted regions. Some of these genes encode proteins involved in neurodevelopmental pathways including synaptic long-term potentiation and neuregulin signaling (7). For example, CNV deletions at 2p16.3 and at 22q11 associated with SCZ risk (*NRXN1* and *COMT*, respectively) each contain a protein-coding gene related to synaptic function (8, 9).

Protein-coding genes have been the focus of studies investigating SCZ-associated CNVs. For example, protein-coding genes *COMT*, *PRODH*, and *ZDHHC8* have been regarded as potential SCZ candidate genes because they are located within the 22q11.2 CNV region (10). However, CNV regions also harbor noncoding RNAs. One common type is long noncoding RNA (lncRNA), defined as RNA of 200 nucleotides or longer in size with little or no protein-coding potential (11, 12). Most lncRNAs show spatial-temporal, brain region–specific, or cell type–specific expression patterns in the brain (13, 14). LncRNAs can regulate gene expression via both cis and trans regulation (15) and may contribute to synaptic plasticity as well as neuronal development and differentiation (13, 16). This is of particular importance because SCZ is generally considered to be a neurodevelopmental disorder (17). Dysregulation of lncRNAs that are involved in neuronal functions may contribute to the etiology of SCZ. The expression of lncRNAs located in CNV deletion regions (CNV-lncRNAs) could be disrupted by the deletion, which, in turn, may alter the expression of their regulatory targets (18, 19). The potential functions of lncRNAs located within CNV deletion regions associated with SCZ risk are intriguing and likely to be important in the etiology of SCZ. However, the roles of CNV-lncRNAs in SCZ have not been investigated systematically.

To study the functions of lncRNAs, coexpression analysis among well-annotated protein-coding genes and lncRNAs has been used (20), because coexpression implies some degree of coregulation (21, 22). However, it is important to recognize that coexpression does not necessarily translate into a causal regulatory relationship. Experimental validation of the predicted regulatory relationships between a lncRNA and its coexpressed genes becomes critically important. Thus, identifying lncRNAs that are coexpressed with protein-coding genes implicated in SCZ suggests that these lncRNAs may contribute to SCZ through their regulation of SCZ-related genes.

Here, we hypothesized that CNV-lncRNAs may be involved in the etiology of SCZ through regulation of coexpressed SCZ-related genes. To test this hypothesis, we retrieved annotated lncRNAs mapped to 10 SCZ risk–associated CNV deletions. We identified protein-coding genes that were coexpressed with CNV-lncRNAs in postmortem control brain tissue samples from individuals without psychiatric disease using weighted gene coexpression network analysis (WGCNA) (23) based on brain transcriptome data from the PsychENCODE...
RESULTS

After sample preparation and RNA sequencing (RNA-seq), we used WGCNA to interrogate 259 BrainGVEX postmortem control adult human brain samples and 37 Developmental Capstone postmortem developing human brain samples. We retrieved lncRNAs that mapped to 10 SCZ risk–associated CNV deletions (Fig. 1), as reported in a recent SCZ CNV study (6). A total of 124 lncRNAs annotated in GENCODE v19 were retrieved from eight SCZ risk–associated CNV deletion regions: 1q21.1, 3q29, 7p36.3, 8q22.2, 15q11.2, 15q13.3, distal 16p11.2, and 22q11.2 (2p16.3 and 9p24.3 were omitted because they did not harbor annotated lncRNAs; data file S1). Eighty-eight CNV-lncRNAs and 17,717 protein-coding genes from the BrainGVEX dataset passed filtering and were included in the WGCNA. Eighty-seven CNV-lncRNAs and 17,472 protein-coding genes from the Developmental Capstone dataset were included in WGCNA after data filtering.

A coexpression module with a CNV-lncRNA as a hub gene in control adult human brain samples

To investigate the regulatory roles of CNV-lncRNAs in the control adult human brain, we performed WGCNA on the RNA-seq data from 259 BrainGVEX control adult human brain tissue samples and identified a total of 45 coexpression modules. Fourteen of these contained 32 of the 88 CNV-lncRNAs (data file S2). The remaining 56 CNV-lncRNAs were not clustered into any module. Protein-coding genes in the CNV-lncRNA coexpression modules were used as input for Gene Ontology (GO) and pathway analysis. We then limited our focus to those neuronal function–related modules containing at least one hub CNV-lncRNA (module membership (MM) ≥ 0.8; P < 0.05), in which the hub CNV-lncRNA was predicted to exert important regulatory functions. Of the 14 modules containing CNV-lncRNAs, the largest two modules contained protein-coding genes enriched for neuronal functions (Fig. 2A, turquoise and dark blue, indicated by orange boxes). The turquoise module contained 2453 protein-coding genes and three nonhub CNV-lncRNAs located within the 15q11.2 and 15q13.3 deletion regions. The dark blue module included two lncRNAs: RP11-701H24.10 from 15q11.2 and a hub CNV-lncRNA DGC5 (MM = 0.80; P = 3.82 × 10^{−60}) from 22q11.2, as well as 1039 protein-coding genes. Protein-coding genes in the dark blue module were related to modulation of synaptic transmission [false discovery rate (FDR) = 1.50 × 10^{−5}], regulation of neuronal projections during development (FDR = 3.83 × 10^{−5}), and axon development (FDR = 5.60 × 10^{−4}). However, these genes were not enriched in any KEGG (Kyoto Encyclopedia of Genes and Genomes) pathway. A module preservation test, which tests for module similarity, showed that all of the CNV-lncRNA coexpression modules were preserved in the brain transcriptome data from the Genotype-Tissue Expression (GTEx) project (26) with Z_{summary} scores > 2 (a score between 2 and 10 indicates that a module is moderately preserved, whereas a score of 10 or above indicates that a module is highly preserved) (23). The two neuronal modules mentioned above were highly preserved with Z_{summary} = 73.8 for the turquoise module and Z_{summary} = 22.6 for the dark blue module (Fig. 3A).

A coexpression module with a hub CNV-lncRNA in control developing human brain samples

Given that SCZ is thought to be a neurodevelopmental disorder (17), we investigated temporal regulatory roles of CNV-lncRNAs. The ventrolateral frontal cortex (VFC) of the brain, a part of the prefrontal cortex, has been implicated in the etiology of SCZ (27). Thus, we investigated the temporal regulation of CNV-lncRNAs by applying WGCNA to RNA-seq data from 37 VFC brain tissue samples (across nine stages of development, BrainGVEX (24) and Developmental Capstone (25) projects. The BrainGVEX dataset was used to study the regulatory roles of CNV-lncRNAs in postmortem brain tissue from adult individuals without psychiatric disorders. The Developmental Capstone dataset was used to identify the temporal regulation of CNV-lncRNAs during normal human brain development. We then focused on hub CNV-lncRNAs in neuronal modules that were the most interconnected lncRNAs based on their correlations with the eigengene (the first principal component of a module). Last, we used knockdown and overexpression experiments in human neural progenitor cells (hNPCs) to validate the predicted regulatory relationships between hub CNV-lncRNAs and SCZ-related protein-coding genes. Through this genome-wide search, we identified one lncRNA, DGC5, in the 22q11.2 CNV region that served as a hub regulator for the expression of several SCZ-associated protein-coding genes, including those related to common single-nucleotide polymorphisms (SNPs).
These three neuronal modules were enriched for neuronal activities (Fig. 2B). A preservation test showed that all of the CNV-lncRNA coexpression modules, except the pale turquoise module \(Z_{\text{summary}} = 0.11\), were preserved in the BrainCloud \(Z\) transcrip-
tome dataset \(Z_{\text{summary}} > 2\). These three neuronal modules were highly preserved with \(Z_{\text{summary}} > 10\) (brown module = 37.2; yellow module = 33.8; green-yellow module = 11.3) (Fig. 3B). However, only the brown module contained a hub CNV-lncRNA. This module consisted of 10 nonhub CNV-lncRNAs from CNVs at 1q21.1, 3q29, 15q11.2, and distal 16p11.2, as well as the hub CNV-lncRNA DGCR5 (MM = 0.87; \(P = 1.93 \times 10^{-12}\)) and 2687 protein-coding genes. Protein-
coding genes in this module were enriched for those regulating neuro-
transmitter concentrations \((FDR = 1.69 \times 10^{-17})\), presynaptic processes involved in chemical synaptic transmission \((FDR = 5.05 \times 10^{-11})\), and potassium ion transport \((FDR = 3.21 \times 10^{-10})\). KEGG pathway analysis demonstrated that the genes in this module were involved in the synaptic vesicle cycle pathway \((FDR = 2.50 \times 10^{-7})\), the neuro-
active ligand-receptor interaction pathway \((FDR = 5.77 \times 10^{-7})\), and the calcium signaling pathway \((FDR = 5.17 \times 10^{-7})\).

**Enrichment of protein-coding genes coexpressed with DGCR5 for SCZ-related genes**

The coexpression modules with a hub CNV-lncRNA identified in both the BrainGVEX and Developmental Capstone datasets differed to some extent, given that they captured different spatiotemporal dynamics. Only 187 protein-coding genes and one hub CNV-lncRNA, DGCR5, overlapped between these two modules. DGCR5 was found to have the highest MM value (BrainGVEX dataset, MM = 0.80; Developmental Capstone dataset, MM = 0.87) among the CNV-lncRNAs in these two modules, suggesting that it may have an important regulatory function.

To investigate a potential association between DGCR5 and SCZ, we assessed the enrichment of 500 module genes, the expression of which highly correlated with DGCR5 expression (Spearman correlation \(|r| > 0.5\); \(P < 0.05\); data file S3) against four lists of SCZ-related genes. The first SCZ-related gene set included 343 protein-coding genes adjacent to the 108 SCZ loci identified in a genome-wide association study (GWAS; denoted GWAS genes) by the Psychiatric Genomics Consortium (PGC) (29). The second list comprised 291 protein-coding genes containing SCZ de novo mutations (DNMs) (30). The third list comprised 693 SCZ differentially expressed genes (DEGs) identified from the CommonMind dataset (31). We also included 747 SCZ DEGs identified from the BrainGVEX dataset (fig. S1 and data file S4) in the enrichment analysis.

In the BrainGVEX dataset, we found that 19 of the top 500 genes coexpressed with DGCR5 were significantly enriched in the SCZ GWAS genes \(P_{\text{H}} = 1.86 \times 10^{-3}\); \(P_{\text{H}}\) represents hypergeometric \(P\) value) and that 22 of the top 500 genes were enriched in the SCZ DNM genes \(P_{\text{H}} = 9.34 \times 10^{-6}\), fig. S2A). Considering that gene length can affect enrichment (32), we performed permutation tests \(n = 1.00 \times 10^{5}\) by controlling for gene length and found that the enrichment was not due to the length of SCZ GWAS and DNMs but reflected their coexpression with DGCR5 (fig. S3). However, the top 500 genes coexpressed with DGCR5 were not enriched with the DEGs from either the BrainGVEX or CommonMind datasets. In the Developmental Capstone dataset, 28 of the top 500 genes coexpressed with DGCR5 overlapped with the BrainGVEX DEGs \(P_{\text{H}} = 1.27 \times 10^{-2}\), and 47 of the 500 genes were enriched in the
CommonMind DEGs ($P_H = 5.66 \times 10^{-10}$) (fig. S2B). However, enrichment was not significant for the SCZ GWAS genes or DNM genes.

**DGCR5 may be a potential regulator of SCZ-related genes in the coexpression module**

The significant enrichment of genes coexpressed with DGCR5 for SCZ-related genes implied that DGCR5 may be a regulator of these SCZ-related genes. We performed key driver analysis (KDA) for the DGCR5 coexpression module in both the BrainGVEX dataset (blue module) and Developmental Capstone dataset (brown module) using a previously published method (33). Key drivers are those genes whose neighbors are significantly enriched in the input gene list in contrast to other random genes from the same network. Using the four SCZ-related gene sets mentioned above as input genes in KDA, we identified 135 key driver genes in the DGCR5 coexpression module of the BrainGVEX dataset and 321 key driver genes for the Developmental Capstone dataset (data file S5). DGCR5 was one of the 21 key drivers shared in the coexpression module between the BrainGVEX dataset ($P = 2.04 \times 10^{-9}$) and the Developmental Capstone dataset ($P = 4.05 \times 10^{-164}$) (Fig. 4). These results suggested that DGCR5 may act as a regulator of SCZ-related genes in the coexpression module.

We decided to validate the predicted regulation by DGCR5 of SCZ-related genes in the list of the top 500 DGCR5-coexpressed genes using hNPCs derived from normal human induced pluripotent stem cells (hiPSCs; fig. S4). Nineteen SCZ GWAS genes and 22 SCZ DNM genes from the BrainGVEX coexpression module, as well as the top 20 BrainGVEX DEGs and the top 20 CommonMind DEGs ranked by their correlation with DGCR5 in the Developmental Capstone coexpression module, were selected for quantitative polymerase chain reaction (qPCR) analysis (data file S3). Of the selected genes, two genes (ABCC12 and LAMB3) were not detected in previous RNA-seq studies of normal control hNPCs derived from hiPSCs or human embryonic stem cells (34, 35), but both of these genes showed low expression in hNPCs from our qPCR analysis. We then observed the effects of DGCR5 knockdown or overexpression in the hNPCs on the expression of these selected genes. Eight of the 19 SCZ GWAS genes and 4 of the 22 SCZ DNM genes were down-regulated in response to DGCR5 knockdown (72% efficiency, $P_{\text{adj}} = 8.67 \times 10^{-3}$; Fig. 5A). Four BrainGVEX DEGs and three CommonMind DEGs were down-regulated in response to DGCR5 knockdown (Fig. 5B). Knocking down LINC01637, a CNV-IncRNA selected from a nonneuronal module that served as a negative control for DGCR5 knockdown, did not perturb expression of any of the genes coexpressed with DGCR5 (Fig. 5C).

**DISCUSSION**

CNVs have been implicated in SCZ etiology (36, 37). Previous investigations have focused on how protein-coding genes inside the CNVs rather than noncoding genes might contribute to SCZ risk. Although there have been some investigations into noncoding RNAs [lncRNAs and microRNAs (miRNAs)] at CNVs (19, 38, 39), the potential importance of CNV-IncRNAs in SCZ etiology has not been well studied. Here, we investigated the potential regulatory roles for CNV-IncRNAs in the human brain and their possible contributions to SCZ. We found that several CNV-IncRNAs were coexpressed with...
DGCR5 had significantly lower expression in postmortem brain tissue from SCZ cases (fold change = 0.73; FDR = 3.21 × 10⁻³).

Among the CNV-lncRNAs in the neuronal modules, the hub CNV-lncRNA DGCR5 was shared by both the BrainGVEX and Developmental Capstone datasets. We experimentally validated that the CNV-lncRNA DGCR5 could regulate some of the selected coexpressed genes (SCZ GWAS genes, DMN genes, or DEGs) in hNPCs derived from hiPSCs. Some of these genes, such as IGSP9B (40) and NRGN (41), have been linked to SCZ in animal models or brain imaging studies. DGCR5 has been found to be down-regulated in brain tissue from individuals with SCZ (fold change = 0.73; FDR = 1.35 × 10⁻²) in the CommonMind dataset (31). In the latest PsychENCODE Capstone RNA-seq data from 601 SCZ cases and 1253 controls, Gandal et al. (42) found that the DGCR5 coexpression module was associated with SCZ, and that DGCR5 had significantly lower expression in postmortem brain tissue from SCZ cases compared to control brain tissue (fold change = 0.94; FDR = 3.21 × 10⁻³).

In a recently published paper (43), DGCR5 was found to be the only CNV-lncRNA in the neuronal coexpression module (CD1) that was down-regulated in postmortem brain tissue from individuals with SCZ, autism, and bipolar disorder. Together, these findings implicate DGCR5 in an increased risk for SCZ potentially through its regulatory effects on coexpressed SCZ-related genes.

Our study provides evidence for a potential regulatory relationship between the CNV-lncRNA DGCR5 and genes outside of the CNV region. Our study also reveals intriguing connections between SCZ-associated rare CNVs and common SNP variants. Through the CNV-lncRNA DGCR5, rare CNVs and common SNP variants may converge onto common pathways regulating gene expression. Our study suggests that noncoding RNAs within the CNV regions may have regulatory roles that implicate them in SCZ etiology.

There are a number of limitations to our study. First, coexpression does not necessarily mean causality. Not all of the genes coexpressed with DGCR5 were found to be regulated by DGCR5. Future experimental validation of gene regulation by DGCR5 will be critically important. Second, the coexpression modules in this study included some transcription factors and their regulatory targets. It is unlikely that simple pairwise regulator-target analysis can fully describe all of these regulatory relationships. Other important genes could contribute to gene expression as upstream regulators or as part of parallel regulation cascades. Taking advantage of the temporal nature of the Developmental Capstone dataset, which covered multiple stages of human brain development, we could study the potential temporal regulation of CNV-lncRNAs in both human brain and cells used in this study was small, which limited the ability to identify temporal regulatory roles for CNV-lncRNAs. Only a few selected genes from the coexpression module of the Developmental Capstone dataset were found to be regulated by DGCR5. The hNPC cellular model may also be limited in its ability to capture temporal regulation of gene expression by DGCR5 because these cells represent a very short period in human brain development. Further investigations using a greater number of postmortem human brain tissue samples at different stages of development may help to elucidate the exact regulatory roles of DGCR5.

Last, lncRNAs have diverse functional mechanisms in the regulation of gene expression. They could interact with protein targets or bind to miRNAs as competing endogenous RNAs (44). For example, several studies (45–49) have demonstrated a role for DGCR5 in the regulation of miRNAs in cancer cell lines. In the Gandal et al. (43) study, which was based on human brain transcriptome data, the DGCR5 coexpression module (CD1) contained one miRNA.
hsa-miR-1227) and some validated targets of DGCR5. The DGCR5-regulated genes within the CD1 module were also predicted to be targets of hsa-miR-1227 (data file S6), implying a potential mediator effect for DGCR5 on these target genes. In the future, integration of multidimensional datasets will help to reveal the regulatory mechanisms of DGCR5 in SCZ.

In summary, population (PsychENCODE BrainGVEX and GTEx) and developmental (PsychENCODE Developmental Capstone and BrainCloud) brain transcriptome datasets have suggested a potential regulatory relationship between a hub CNV-lncRNA, DGCR5, and SCZ-related genes. DGCR5 potentially could represent a bridge connecting rare CNV deletions and common GWAS findings, given that the 22q11.2 deletion may extend its functional impact beyond the protein-coding genes residing in the deleted regions.

MATERIALS AND METHODS

Study design

The objective of this study was to investigate the potential contribution of CNV-lncRNAs to SCZ risk. First, we retrieved annotated lncRNAs mapped to SCZ risk–associated CNV deletion regions reported in the SCZ CNV study from the PGC (6). We then identified protein-coding genes that were coexpressed with the CNV-lncRNAs in postmortem human brain tissue from individuals without psychiatric disorders (BrainGVEX dataset, n = 259) and across different stages of human brain development (Developmental Capstone dataset, n = 37). Last, after identifying hub CNV-lncRNAs, we performed knockdown and overexpression experiments in hNPCs derived from hiPSCs derived from hiPSCs. Then, we used qPCR analysis to validate the predicted regulation of SCZ-related protein-coding genes by hub CNV-lncRNAs. The knockdown and overexpression experiments were conducted in three biological replicates. Data are means ± SEM. Gene expression in the control group was normalized to 1. Two-tailed t test was used for comparison between two groups. P values were adjusted for multiple testing using the Benjamini-Hochberg method. *P_{adj} < 0.05, **P_{adj} < 0.01.

Brain samples and data collection

PsychENCODE BrainGVEX data

Because frontal cortex has been implicated in the etiology of SCZ (27, 50), we chose frontal cortex samples to study the regulatory roles of CNV-lncRNAs in human adult brains. Human postmortem brain tissues were collected from four collections of the Stanley Medical Research Institute (SMRI): The Neuropathology Consortium, Array Collection, New Collection, and Extra Collection. SCZ diagnosis was based on the Diagnostic and Statistical Manual of Mental Disorders,
Fourth (DSM-IV) criteria and medical records. Diagnosis of unaffected controls was according to the structured interviews with their family member(s). For all brain specimens, the samples were excluded if they met any of the exclusion criteria: (i) notable structural brain pathology on postmortem examination, (ii) history of notable focal neurological signs premortem, (iii) history of a central nervous system disease that could be expected to alter gene expression in a persistent way, (iv) documented intelligence quotient of <70, and (v) poor RNA quality. For unaffected controls, two additional exclusion criteria were used: (i) substance abuse within 1 year of death or notable alcohol-related changes in the liver and (ii) age less than 30 years. Thus, we obtained a total of 96 SCZ cases and 75 control brain samples from SMRI. Additional 185 brain samples of unaffected controls were collected from the Banner Sun Health Research Institute according to the individual medical records or neuropsychological screening assessment. Total RNA was extracted from the collected brain samples and used for RNA-seq (Supplementary Materials). This study generated RNA-seq data from a total of 356 samples. Previous paper published by Gandal et al. (43) used part of these data. Metadata for the 356 samples, including sex, ethnicity, age of death, diagnosis, postmortem interval (PMI), RNA integrity number (RIN), batch, and sequencing quality matrix, are listed in data file S7. The RNA-seq data of SCZ cases and controls were used for differential gene expression analysis (Supplementary Materials), and the RNA-seq data of control samples were used for WGCNA to capture protein-coding genes that were coexpressed with the CNV-lncRNAs in nonpsychiatric individuals.

**PsychENCODE Developmental Capstone data**

The PsychENCODE Developmental Capstone project (25) was designed to study transcriptional mechanisms involving human brain development. In this project, RNA-seq data (Illumina GAIIx, polyadenylated selection, 76–base pair single-end read) were generated from a total of 606 nonpsychiatric brain samples, which were collected from 16 brain regions from 41 individuals spanning from eight postconceptional weeks to over 40 years of age. Because VFC has been found being involved in SCZ (27), we included RNA-seq data of 40 VFC samples from the PsychENCODE Developmental Capstone project into WGCNA to study the temporal regulation of CNV-lncRNAs. Three VFC samples with no individual information were removed. The remaining 37 samples (15 prenatal and 22 postnatal samples) covered nine different brain developmental stages from early prenatal to adult (data file S7). More detailed information about the Developmental Capstone data could be found on the PsychENCODE website (http://psychENCODE.org/).

**Validation data (GTEx and BrainCloud)**

To validate the CNV-lncRNA coexpression patterns in the BrainGVEX dataset, we used RNA-seq data of 101 adult frontal cortex samples (RIN > 6) from the GTEx project (26). We also used the BrainCloud transcriptome data (28) to validate the CNV-lncRNA coexpression relationships detected in the Developmental Capstone data. The BrainCloud data are microarray data (Gene Expression Omnibus accession number: GSE30272) of 269 prefrontal cortex samples from fetus to adults. Metadata of the 101 GTEx samples and 269 BrainCloud samples are listed in data file S7.

**CNV-lncRNA retrieval**

Because CNV-lncRNAs could be directly disrupted by CNV deletions, we chose to focus on the SCZ risk CNV deletion regions but not those protective or duplicated CNVs. Thus, 10 SCZ risk–associated CNV deletions identified in the PGC study (6) were included in this study. Eight of the 10 CNV regions (1q21.1, 2p16.3, 3q29, 7p36.3, 15q11.2, 15q13.3, distal 16p11.2, and 22q11.2) have been previously implicated in SCZ, and the other two loci (8q22.2 and 9p24.3) are novel SCZ CNVs reported by PGC (6). We obtained the deletion positions of the CNVs by combining the CNV deletion information from the DECIPHER database (51) (https://decipher.sanger.ac.uk/). Annotated lncRNAs mapped to the aforementioned 10 CNV deletion regions were retrieved from GENCODE v19.

**Weighted gene coexpression network analysis**

Raw expression data of the BrainGVEX and Developmental Capstone were processed (Supplementary Materials) before gene coexpression analysis. lncRNAs and mRNAs that had FPKM (fragments per kilobase per million mapped reads) values of >0.1 in at least 10 samples were selected for coexpression analysis in the BrainGVEX data. RNAs with FPKM values of >0.1 in at least 10% of the samples were used for coexpression analysis in the Developmental Capstone data. WGCNA (v1.51) in R package was used to identify CNV-lncRNA coexpression modules. Biweight midcorrelation (bicor) was used to calculate pairwise gene correlations in the BrainGVEX data, which is recommended as a robust alternative to outlier measurements. For Developmental Capstone data, Spearman correlation was used to calculate the pairwise gene correlations, which is a nonparametric measure of ranked correlations between variables. It does not depend on the data distribution. For coexpression network construction, the power value was set at 4 and 10 for the BrainGVEX and Developmental Capstone data, respectively, using the pickSoftThreshold function. The dynamicTreeCut algorithm was used with the following parameters: deepSplit = 4, pamStage = T, pamRespectsDendro = F, and minClusterSize = 30. CNV-lncRNAs and protein-coding genes with similar expression patterns were clustered into the same module. The CNV-lncRNA coexpression modules were further used for functional annotation and preservation tests (Supplementary Materials). MM is a measurement used to assess the correlation between a gene and corresponding module eigengene.

**Key driver analysis**

To identify key drivers in the CNV-lncRNA coexpression modules, we performed KDA using the KDA R package (v0.1) (33). KDA takes an input gene list of interest (generally a disease-associated gene list) and a gene network as input files. It first generates a subnetwork consisting of nodes that are no more than H-layers away from each node in the target gene list in the network. Then, for each node in the subnetwork, it assesses the enrichment of its H-layer downstream neighbors for target gene list. In this study, a combined SCZ-related gene set (343 SCZ GWAS genes, 291 SCZ DNM genes, 747 BrainGVEX DEGs, and 693 CommonMind DEGs) was used as the input gene list. To identify key drivers in the large coexpression module, a cutoff of weight value >0.08 was used to filter out those gene pairs with low correlations in the network. H = 1 was found to be optimal for both networks of the BrainGVEX and Developmental Capstone data. The KDA results were visualized using Cytoscape (v3.5.1).

**DGC5 knockdown in hNPCs**

We used a knockdown experiment in hNPCs to validate the role of DGC5 in regulating other protein-coding genes, because DGC5 is a shared hub CNV-lncRNA (MM ≥ 0.8; P < 0.05) in the neuronal module of both BrainGVEX and Developmental Capstone data and is
specifically and highly expressed in the brain (fig. S5). Because DGCR5 was expressed in both the nucleus and cytoplasm (fig. S6), we used lncRNA Smart Silencer (RiboBio, China) for DGCR5 knockdown. The lncRNA Smart Silencer is a mixture of three antisense oligonucleotides and three small interference RNAs, which could effectively knock down both nuclear and cytoplasmic lncRNAs. The Smart Silencer at the optimal final concentration (50 nM) and Lipofectamine RNAiMAX reagent (13778030, Invitrogen, USA) was used for DGCR5 knockdown. Total RNA was extracted at 24 hours after cell transfection and used for qPCR analysis, because DGCR5 was the most efficiently knocked down at this time point (fig. S7). To validate the specificity of DGCR5 in regulating gene expression, CNV-lncRNA knocked down at this time point (fig. S7). To validate the specificity of DGCR5 coexpressed and checked the effect on detected genes. The sequences of lncRNA Smart Silencer are listed in table S2.

DGCR5 overexpression in hNPCs

For DGCR5 overexpression assay, full length of the major transcript of DGCR5 (transcript ID: NR_002733) was synthesized and cloned into the transient overexpression vector GV144 (GeneChem, China), using the restriction enzymes Xho I and Bam H1 (New England Biolabs, USA). The constructed DGCR5-GV144 overexpression vector or empty GV144 vector was transfected into hNPCs using Lipofectamine LTX and PLUS Reagents (A12621, Invitrogen, USA). Total RNA was extracted at 24 hours after transfection and used for qPCR analysis.

qPCR analysis

Total RNA was extracted with the miRNeasy Mini Kit (217004, Qiagen, USA). Complementary DNA was generated using HiScript II Q RT SuperMix for qPCR (+gDNA wiper) (R223-01, Vazyme, China). Total RNA was extracted with the miRNeasy Mini Kit (217004, Qiagen, USA). The constructed DGCR5-GV144 overexpression vector or empty GV144 vector was transfected into hNPCs using Lipofectamine LTX and PLUS Reagents (A12621, Invitrogen, USA). Total RNA was extracted at 24 hours after transfection and used for qPCR analysis.

Statistical analysis

A threshold FDR of <0.05 and P < 0.01 were used for differential gene expression analysis in the BrainGVEX data. The hypergeometric test was used to assess the significance of enrichment for genes coexpressed with DGCR5 that were SCZ-related genes in both the BrainGVEX and Developmental Capstone datasets. The qPCR data shown in Fig. 5 and fig. S4 are means ± SEM. Statistical analysis was performed with GraphPad Prism (v6.0) and described in each figure legend. The two-tailed t test was used for comparison between two groups (α = 0.05). P values were adjusted for multiple testing using the Benjamini-Hochberg method.

SUPPLEMENTARY MATERIALS

www.sciencetranslationalmedicine.org/cgi/content/full/10/472/eaat6912/DC1

Materials and Methods

Fig. S1. Distribution of fold change of DEGs between SCZ cases and controls in the BrainGVEX dataset.
Fig. S2. Enrichment analysis of top 500 DGCR5-coexpressed genes against SCZ-related gene sets.
Fig. S3. Enrichment of top 500 DGCR5-coexpressed genes with SCZ GWAS and DNMs after controlling for gene length.

REFERENCES AND NOTES


14. Non-coding RNAs in mammalian nervous system development, plasticity, disease, and function and risk for schizophrenia.


16. Patterns of long noncoding RNAs.

17. Science

18. J. Neurosci. 316, 316 Science


20. J. Neurosci. 316, 315


22. J. Neurosci. 316, 314

23. J. Neurosci. 316, 314


25. J. Neurosci. 316, 314


27. J. Neurosci. 316, 314


29. J. Neurosci. 316, 314

30. J. Neurosci. 316, 314

31. J. Neurosci. 316, 314

32. J. Neurosci. 316, 314

33. J. Neurosci. 316, 314

34. J. Neurosci. 316, 314

35. J. Neurosci. 316, 314

36. J. Neurosci. 316, 314

37. J. Neurosci. 316, 314

38. J. Neurosci. 316, 314


40. J. Neurosci. 316, 314

41. J. Neurosci. 316, 314

42. J. Neurosci. 316, 314

43. J. Neurosci. 316, 314

44. J. Neurosci. 316, 314

45. J. Neurosci. 316, 314


Acknowledgments: We thank G. Giase for editing the manuscript. Funding: This work was supported by NIH grants 1 U01 MH103340-01 and 1 R01 ES024988 (to C.L.) and National Natural Science Foundation of China (NSFC) grants 81401114 and 31571312, National Key Plan for Scientific Research and Development of China grant 2016YFC1306000, and Innovation-Driven Project of Central South University grants 2015CSX034 and 2016CSX033 (to C.C.). We thank Chicago Biomedical Consortium with support from the Searle Funds at The Chicago Community Trust. Data were generated as part of the PsychENCODE Consortium, supported by U01 MH103392, U01 MH103365, U01 MH103346, U01 MH103340, R01 MH103339, R21 MH109956, R21 MH105881, R21 MH105853, R21 MH103877, R21 MH102791, R01 MH111721, R01 MH110928, R01 MH110927, R01 MH110926, R01 MH110921, R01 MH110920, R01 MH110905, R01 MH109715, R01 MH109677, R01 MH105898, R01 MH105897, R01 MH1094714, and P50 MH106934 awarded to S. Akbarian (Icahn School of Medicine at Mount Sinai), G. Crawford (Duke University), S. Dracheva (Icahn School of Medicine at Mount Sinai), P. Farnham (University of Southern California), M. Gerstein (Yale University), D. Geschwind (University of California, Los Angeles), F. Goes (Johns Hopkins University), T. M. Hyde (Lieber Institute for Brain Development), A. Jaffe (Lieber Institute for Brain Development), J. A. Knowles (University of Southern California), C. Liu (SUNY Upstate Medical University), D. Pinto (Icahn School of Medicine at Mount Sinai), P. Roussos (Icahn School of Medicine at Mount Sinai), S. Sanders (University of California, San Francisco), N. Sestan (Yale University), P. Sklar (Icahn School of Medicine at Mount Sinai), M. State (University of California, San Francisco), P. Sullivan (University of North Carolina), F. Vaccarino (Yale University), D. Weinberger (Lieber Institute for Brain Development), S. Weissman (Yale University), K. White (University of Chicago), J. Willsey (University of California, San Francisco), and P. Zandi (Johns Hopkins University). Author contributions: A.T., D.F., H.E., L.J., and K.P.W. contributed to RNA extraction and RNA-seq. C.J., K.W., R.D., and Y.X. performed WGCNA and KDA. K.G., M.L., F.N.S., and Y.I.-K. helped to revise the manuscript. C.C. and C.L. supervised the overall study design, guided the procedures and analyses, as well as prepared the manuscript. Q.M. initiated the study, did the cell culture and qPCR experiments, and led the manuscript writing. T.B. performed the differential gene expression analysis. Competing interests: K.P.W. is the president and a shareholder of Tempus Labs Inc. The other authors declare that they have no competing interests. Data and materials availability: All data associated with this study are in the paper or the Supplementary Materials. Detailed protocols for brain sample preparation and RNA-seq are available on Synapse (Synapse number: syn4590909). For the PsychENCODE Consortium, data can be requested on a specific page on Synapse (Synapse number: syn4921369).

Submitted 25 March 2018
Accepted 22 October 2018
Published 19 December 2018

10.1126/scitranslmed.aat6912

The DGC5 long noncoding RNA may regulate expression of several schizophrenia-related genes

Qingtuan Meng, Kangli Wang, Tonya Brunetti, Yan Xia, Chuan Jiao, Rujia Dai, Dominic Fitzgerald, Amber Thomas, Lindsey Jay, Heather Eckart, Kay Grennan, Yuka Imamura-Kawasawa, Mingfeng Li, Nenad Sestan, Kevin P. White, Chao Chen and Chunyu Liu

Sci Transl Med 10, eaa6912.
First published 13 December 2018
DOI: 10.1126/scitranslmed.aat6912

Illuminating the genomic mysteries of psychiatric diseases

Schizophrenia and bipolar disorder are complex psychiatric diseases with risks contributed by multiple genes. Studies by the PsychENCODE Consortium, including two in this issue (Chen et al. and Meng et al.), seek to elucidate the genomic elements and regulatory pathways that underpin several psychiatric disorders. Chen et al. analyzed transcriptome data from postmortem brain tissue from patients with schizophrenia or bipolar disorder. They report that the transcription factor POU3F2 is a core regulator of a gene coexpression network associated with these disorders. In a genome-wide analysis of control human brain samples from the adult and developing brain, Meng et al. report that the IncRNA DGC5, which lies within the 22q11.2 deletion associated with schizophrenia risk, regulates expression of several SCZ-associated protein-coding genes.