Leveraging interindividual variability of regulatory activity refines genetic regulation of gene expression in schizophrenia

Maris Alver (✉ marisalver@gmail.com)

University of Geneva

Nikolaos Lykoskoulis

University of Geneva

Anna Ramisch

University of Geneva

Halit Ongen

University of Geneva <https://orcid.org/0000-0002-4197-5790>

Emmanouil Dermitzakis

University of Geneva

Article

Keywords: schizophrenia, gene expression, regulatory activity

Posted Date: October 29th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-989712/v1>

License: This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Schizophrenia is a polygenic psychiatric disorder with limited understanding about the mechanistic changes in gene expression regulation. To elucidate on this, we integrate interindividual variability of regulatory activity with gene expression and genotype data captured from the prefrontal cortex of 272 cases and controls. We show that regulatory element activity is structured into 10,936 and 10,376 cis-regulatory domains in cases and controls, respectively, which display distinct regulatory element coordination structures in both states. By studying the interplay among genetic variants, gene expression and cis-regulatory domains, we ascertain that changes in coordinated regulatory activity tag alterations in gene expression levels ($p=8.62e-06$, $OR=1.60$), unveil case-specific QTL effects, and identify regulatory machinery changes for genes affecting synaptic function and dendritic spine morphology in schizophrenia. Altogether, we show that accounting for coordinated regulatory activity provides a novel mechanistic approach to reduce the search space for unveiling genetically perturbed regulation of gene expression in schizophrenia.

Introduction

Schizophrenia (SCZ) is a severe mental illness that oftentimes leads to a lifetime of chronic disability. While several lines of evidence converge on the neurodevelopmental origin for the disorder and augment the impact of both genetic and environmental factors in disease aetiology, the pathophysiology of SCZ still remains incompletely understood with little progress in novel treatment development^{1,2}.

The last decades of extensive research has yielded valuable insights into the genomic and molecular underpinnings of SCZ. Large-scale genomic analyses triggered by prior studies of strong heritability estimates for SCZ (60-80%)^{3,4} have unveiled its highly polygenic architecture⁵⁻⁷. Complemented by gene expression and chromatin profiling analyses from dorsolateral prefrontal cortex (DLPFC), the SCZ risk variants have been shown to localize in functional regulatory genomic elements with approximately half of these displaying brain tissue-specific expression quantitative trait locus (eQTL) effects^{5,8-10}, and to be enriched for open chromatin and evolutionary conserved regions^{11,12}. Genes identified by differential expression and genome-wide association analyses (GWAS) are associated with brain developmental pathways, synaptic function, and immune response¹³⁻¹⁵. While immense efforts have resulted in comprehensive SCZ-specific resources for perturbed gene expression patterns and for fine-mapping and annotating discovered disease-associated genomic associations, little is known about the regulatory machinery changes and how genetic effects are propagated onto gene expression that drive molecular abnormalities in SCZ development.

Gene transcription profiles are defined by the activity of regulatory elements (REs) that overlap with open chromatin. Since the activity of REs are modulated by genetic variation, changes in chromatin accessibility result in gene expression variability and hence constitute as an intermediate phenotype for profiling eQTL effects on gene expression¹⁶⁻²⁰. Systematic measurement of interindividual correlation between chromatin activity levels has revealed that the variability of nearby regulatory activity is

structured into well-delimited sets of cis-regulatory domains (CRDs)²¹. The coordinated activity of REs within CRDs are under tight genetic control, mediate cis and trans effects of genetic variants onto gene expression and provide a higher-order structural resolution of functional regulatory associations²¹. Accounting for the three-dimensional (3D) genome organization in cis that captures concerted effect of regulatory activity could thereby facilitate a more robust signal detection for identifying disruption in regulatory function and for delineating deviations in gene expression cascades specific to disease. To build on this concept, we set out to analyse the interplay among genetic variants, coordinated regulatory activity and gene expression to characterize genetically perturbed regulatory machinery changes specific to SCZ. To this end, we integrated genome-wide genotyping data with RE activity levels (chromatin immunoprecipitation sequencing (ChIP-sequencing profiled for histone mark H3K27ac) and transcriptomic profiles (bulk RNA-sequencing) obtained from the DLPFC of SCZ cases and control subjects (dbGaP phs000979.v3.p2). At least two levels of molecular data were available for 272 individuals: 98 SCZ cases and 174 controls (188 males and 84 females, 164 African Americans and 108 Europeans; **Supplementary Fig. 1**).

Results

Distinct regulatory element coordination in SCZ

To study the coordination of REs, we systematically measured interindividual correlation between nearby chromatin peaks. We identified 10,938 CRDs in SCZ cases, 10,376 CRDs in controls and 11,374 CRDs in the combined set (i.e., across SCZ cases and controls), regrouping 28.9% (n=40,819), 31.4% (n=44,391) and 38.4% (n=54,278) of the peaks, respectively (**Table 1**), and capturing a higher-order structural resolution of regulatory activity. The majority of the CRDs contained two REs, while some captured correlated activity among >80 REs (mean number of REs per CRD 4.7; mean CRD length 138 kb in the combined set; **Supplementary Fig. 6**). As expected, we identified a high concordance of CRDs between SCZ cases and controls (42%; **Supplementary Fig. 7**), which represent a uniform coordination of regulatory activity in DLPFC in SCZ cases and controls. However, 58% of the CRDs showed distinctive RE coordination in SCZ cases and in controls, implying notable state-dependent differences in the regulatory machinery.

Given unique RE coordination in SCZ cases and controls, we sought to investigate the mechanism for SCZ-specific CRD formation. Specifically, we asked whether the chromatin peaks within SCZ-specific CRDs were differentially active or whether these had larger variance in activity in controls compared to SCZ cases. To this end, we considered CRDs in SCZ cases composed of peaks not part of any CRD in controls and compared single peak activities and mean correlation estimates among peaks per CRD between the two groups. We discovered that 53% of the peaks within SCZ-specific CRDs (3,540 peaks in 2,212 CRDs) were differentially active in SCZ cases at false discovery rate (FDR) 5% (**Supplementary Fig. 8a**). While the majority of the peaks (71%) showed lower activity in SCZ cases (**Fig 1a**), only a third of SCZ-specific CRDs had all underlying peaks differentially active between SCZ cases and controls (**Supplementary Fig. 8b**), implying that the regulatory activity originating from those genomic regions

likely results in inhibition of downstream molecular cascades. The peaks of SCZ-specific CRDs displayed significantly higher mean correlation in SCZ cases compared to controls (Mann-Whitney U test $p = 5.02 \times 10^{-47}$; **Fig. 1bc**), indicating that changes in the 3D structure of the genome, rather than differential activity, were responsible for SCZ-specific CRD formation. At FDR 5%, we identified eleven SCZ-specific CRDs to be associated with the expression of proximal genes, for example *POU3F1* (also known as *OCT6*, transcriptional repressor for myelin-specific genes²²), *KIF5A* (neuronal-specific vesicular transporter²³), *NECAB1* (Ca^{2+} -binding in neurons²⁴), and *PDCD1LG2* (immune checkpoint receptor ligand²⁵) (**Supplementary Table 1**). These associations exemplify coordinated regulatory changes specific to the disease state that affect or are affected by gene expression perturbations.

Changes in CRD activity track alternations in gene expression in SCZ

Having identified several distinct RE coordination structures in DLPDC in SCZ cases that were absent in controls, we focused next on CRDs that had the same structure of RE coordination across SCZ cases and controls (i.e., CRDs identified in the combined set). We set out to determine differences in regulatory activity between SCZ cases and controls and investigated their relation to genes that were differentially expressed (DEGs) between the two groups. At FDR 5%, we identified 1,141 CRDs (599 lower activity, 542 higher activity) and 1,363 genes (937 up-regulated, 426 down-regulated) to be differentially active and expressed in SCZ cases, respectively (**Table 1, Supplementary Fig. 9, Supplementary Table 2, Supplementary Table 3**). The differences in effect size for CRDs were subtle, reflecting a narrow variability range in regulatory activity (**Supplementary Fig. 9a**). The determined DEGs were in concordance with those previously reported in SCZ pathogenesis (**Supplementary Fig. 10**)^{9,26} and were significantly enriched for gene ontology (GO) terms related to sex-hormone and interferon- γ -mediated signalling, glucocorticoid receptor and glutamate receptor binding, axonogenesis and synapse assembly (**Supplementary Table 4**). The DEGs were significantly enriched for differentially active CRDs (Fisher's exact test $p = 8.62 \times 10^{-6}$, odds ratio 1.60) (**Fig. 2a**) with the majority of the genes (86%) showing the same direction of effect as the CRD in which the gene transcription start site (TSS) lied, indicating that deviations in gene expression track alterations in the regulatory machinery.

In addition to ascertaining that differentially active CRDs localized within the genomic proximity of DEGs, we additionally tested for direct association between gene expression and CRD activity (i.e., CRDs identified in the combined set). At FDR 5%, we identified in cis 95, 634 and 1,197 CRD-gene associations in SCZ cases, in controls, and in the combined set, respectively (**Table 1, Supplementary Fig. 11, Supplementary Table 6**). The majority of the genes were associated with a single CRD and the majority of the CRDs with a single gene with only a handful of CRDs being associated with up to ten different genes (**Supplementary Fig. 12**). Most gene TSSs clustered at CRD boundaries (**Fig. 2bc**), corroborating the proximal role of coordinated regulatory activity in gene transcription.

Genetic regulation of CRD activity and gene expression in SCZ

We next sought to study the genetic regulation of CRD activity and gene expression, search for SCZ-specific QTL effects and interrogate whether QTL effects colocalize with SCZ risk variants. At 5% FDR and in cis, we discovered 796 and 2,929 functionally independent CRD activity QTLs (aCRD-QTLs), and 867 and 6,166 functionally independent eQTLs in SCZ cases and controls, respectively (**Table 1, Supplementary Table 6, Supplementary Table 7, Supplementary Fig. 13**). The strength of the association was correlated with the genomic distance from the molecular phenotype (**Supplementary Fig. 14**). While almost all SCZ-identified QTL effects replicated in controls (**Supplementary Fig. 15**), 5% of aCRD-QTLs (n=40) and 15% of eQTLs (n=130) showed SCZ-specificity, i.e., these affected CRD activity or gene expression only in SCZ cases or displayed significant change in effect size compared to controls (**Table 1, Fig. 3ab**). The SCZ-specific genotype-dependent variability in CRD activity and in gene expression imply context-dependent and pathway-activated gain in regulatory capacity. Results of gene enrichment analysis for genes associated with SCZ-specific eQTLs conform with posed hypotheses linking dysregulation of retinoid binding and adenosine deaminase activity with SCZ^{27,28} (**Supplementary Table 8**). Colocalization analyses for SCZ risk variants with aCRD-QTLs and with eQTLs showed modest yet proportionally similar enrichment for shared functional effects (1.6% for SCZ-identified aCRD-QTLs and 1.8% for SCZ-identified eQTLs). Interestingly, the aCRD-QTLs colocalized with different GWAS variants compared to eQTLs that shared a functional effect with SCZ risk variants (**Supplementary Table 9**).

To assess common genetic regulation of coordinated regulatory activity and gene expression, we investigated the association of aCRD-QTL and eQTL effects on the other molecular phenotype. Specifically, we correlated aCRD-QTLs with gene expression and eQTLs with CRD activity over gene-CRD associations detected across all samples at nominal significance. We identified that up to 46% of the eQTL and aCRD-QTL variants had an effect on CRD activity and gene expression, respectively (**Supplementary Fig. 16**). The considerable overlap between aCRD-QTL and eQTL effects for relevant CRD-gene pairs corroborates the functional interplay among genetic variants, CRDs and genes.

Refining eQTL perturbations reveals regulatory machinery changes specific to SCZ

Given the established interplay among genetic variants, genes and CRDs, we interrogated the functional directionality between them. We reasoned that the effect of a genetic variant on gene expression could either be mediated by or propagated to the changes in coordinated RE activity and that deviations in the regulatory machinery in SCZ cases compared to controls would imply molecular dysregulation specific to disease. To test this, we considered the previously discovered 1,197 CRD-gene pairs ascertained across SCZ cases and controls at FDR 5% and identified the same genetic variant (eCRD-QTL) that affected both molecular phenotypes and by that determined eCRDQTL-CRD-gene triplets for causal inference (**Methods**). Using Bayesian Networks (BN), we tested three relationship patterns: i) causal model in which the genetic variant affects first the CRD activity which then regulates the gene expression, ii) reactive model in which the genetic variant affects the gene expression which modulates the CRD activity, and iii)

independent model in which the genetic variant affects the gene and the CRD independently; and studied these relationships separately in SCZ cases (n=59) and controls (n=105) (**Supplementary Fig. 17**). We discovered that at FDR 5%, 91.9% of the CRD-gene pairs had a cis-QTL effect (n=1,100; **Table 1**), indicating that the simultaneous change in CRD activity and gene expression was affected by the same nearby genetic driver. We observed more causal models in controls than in SCZ cases (**Supplementary Fig. 18ab**), which were likely driven by smaller SCZ sample size as reflected by the distribution of the probabilities for the most likely model for each triplet (**Supplementary Fig. 18cd**). The probability of the causal model increased the further the gene TSS was from the eCRD-QTL in both SCZ cases and controls (**Supplementary Fig. 18ef**), denoting the role of CRDs mediating the genetic effect onto distal genes.

To study the proportion of differential regulatory mechanisms between SCZ cases and controls and ascertain which molecular functions were affected by these changes, we first estimated the accuracy for BN results using bootstrapping to provide confidence for retrieved probabilities and next carried out gene enrichment analyses for genes associated with different regulatory mechanisms. The accuracy for inferring the most likely causal relationship for triplets was lower in SCZ cases (mean accuracy estimation 69.0%, sd = 16.6) than in controls (mean accuracy estimation 75.0%, sd = 18.0) (**Supplementary Fig. 19ab**). To exclude ambiguous signals, we proceeded with triplets that surpassed the accuracy estimation of 55% (748 triplets, 68% of studied triplets; **Methods, Supplementary Fig. 19c, Supplementary Table 10**). While two-thirds of the triplets displayed the same regulatory mechanism in SCZ cases and controls (**Fig. 4a**), one-third of studied triplets showed a change in directional effect from QTL variant onto molecular phenotype (**Fig. 4b, Supplementary Fig. 19d**). These deviations in regulatory mechanism in SCZ reflect gain or loss in the regulatory capacity that could either be driven by context-dependent or genetically predisposed developmental derailment of gene expression, or affected by external stimuli (e.g., treatment). The genes associated with change-associated triplets were significantly enriched for GO terms related to small GTPase binding, filopodium assembly and cellular lipid catabolic process (**Supplementary Table 11**), highlighting alterations in the regulatory machinery for gene expression affecting synaptic function and plasticity, and dendritic spine morphology in SCZ.

Perturbation in the regulatory mechanism of gene expression in SCZ is exemplified by a triplet consisting of an eCRD-QTL 7:5849993:C:G associated to gene *FSCN1* and to a CRD composed of 5 REs (chr7:5623132-5705414) (**Supplementary Fig. 20**). *FSCN1* is an actin-binding protein that is required for the formation of actin-based cellular protrusions and affects dendritic spine morphology²⁹⁻³². Based on BN, the genetic variant affected first the activity of the CRD and then the expression of the gene in controls (causal model probability 0.99), whereas in SCZ cases the change in CRD activity was a reaction to gene expression (reactive model probability 0.77), indicating that the eCRD-QTL effect on *FSCN1* expression was not mediated via the associated CRD activity in SCZ cases as seen for controls (**Fig. 4d**). This was further supported by significant downregulation of *FSCN1* expression and CRD activity in SCZ cases compared to controls (FDR 5% p-value 0.02 and 0.005, respectively; **Supplementary Fig. 20ab, Supplementary Table 2, Supplementary Table 3**). Moreover, we identified an opposite direction of eCRD-QTL effect on *FSCN1* expression in SCZ cases compared to controls (**Fig. 4c**), but not on CRD activity (**Supplementary Fig. 20c**). Interestingly, this association was identified only in individuals with African

American ancestry (**Supplementary Fig. 20de**) as the genetic variant was completely monomorphic in HBCC Europeans. The MAF spectrum of 12% in HBCC African Americans and 0% in HBCC Europeans is in concordance with population frequencies estimated in larger datasets (MAF 13% in Africans/African Americans and 0.4% in non-Finnish Europeans)³³. These results indicate that the downregulation of *FSCN1* expression in SCZ cases was driven by a different intermediary regulatory mechanism that deviated from one seen in controls and represents a dysregulated step within an abnormal molecular cascade affecting dendritic spine morphology in SCZ.

Discussion

Deciphering regulatory mechanisms of gene expression that reflect molecular perturbation in SCZ are under extensive scrutiny yet are hindered by the complexity of the SCZ phenotype and scarcity of relevant molecular data. Studying regulatory activity that tracks changes in gene expression requires a higher order analysis approach for signal discernment due to narrow variability range in regulatory activity and extensive multiple testing burden¹². Here we show that taking account of interindividual correlation between regulatory activity allows to refine changes in gene expression specific to disease, asserting that disease manifestation stems from dysregulated gene expression cascades that are steered by and propagated to the concerted action of REs. Interrogation of common genetic regulation of gene expression and CRD activity corroborated that correlated changes in gene expression and CRD activity are affected by the same genetic driver. Our results agree with findings showing considerable overlap between QTL effects on chromatin accessibility and gene expression¹², that a single genetic variant drives the association between multiple chromatin peaks and a single gene²⁰, and on the convergence of deviations detected in different molecular layers as seen for gene expression and methylation in SCZ³⁴. Applying causal inference to study the causal relationships among genetic variants, genes and CRDs revealed regulatory machinery changes affecting synaptic function and dendritic spine morphology in SCZ which are in line with established molecular abnormalities identified for the disorder^{35–39}. The deviations in regulatory mechanism reflect gain or loss in the regulatory capacity that could either stem from genetic predisposition, are acquired in disease progression or result from chronic pharmacology. Clear discernment of the proposed origins of effect was hampered due to small sample size and unavailability of relevant data yet allowed to draw the following conclusions.

First, while we found modest colocalization for detected QTLs with SCZ-predisposing genetic variants, the results reflect direct correlation between sample size and QTL signal detection⁴⁰ and hence are in proportion to colocalization signals ascertained in previous findings^{9,12,26,41}. Previous studies have highlighted the concordance of SCZ heritability enrichment for open chromatin regions in fetal and in SCZ DLPFC samples and the stability of methylation and expression feature deviations in fetal brain development that persist into adulthood for those affected by the disorder^{9,12,13,42}, implying that a considerable proportion of signals detected in the current analysis do reflect brain development derailment due to genetic predisposition for SCZ. Second, while molecular data used in this study was extracted from bulk tissue, consistent comparison with signals identified in control samples provides

confidence that the identified deviations in SCZ cases captured the most notable disease-specific molecular abnormalities in DLPFC. This is further supported by investigations revealing that DLPFC transcriptomic profiles are generally biased toward neuronal populations³⁴ and that SCZ risk variants are overrepresented in neuronal vs non-neuronal open chromatin regions⁴³. Third, studying QTL effects on gene expression and CRD activity separately in SCZ cases and controls allowed to discriminate context-dependent genetic effects on both molecular phenotypes, indicating gain in regulatory capacity that translated into gene expression and coordinated regulatory activity variability only in SCZ cases or showed significantly different or even discordant effect direction, as seen for *FSCN1* gene, between the two groups. Fourth, a previous study that showed concordant DEG signals with those found in the current analysis, identified that differential gene expression in SCZ was not driven by antipsychotic intervention²⁶, ensuring that treatment effect was not the main trigger for DEG results between SCZ cases and controls in the current analysis. Lastly, inclusion of individuals of European and African American ancestry augmented signal identification and corroborates that the genetic basis of SCZ and its biology are broadly shared across populations^{7,44,45}.

Altogether, we have outlined that leveraging higher-order structural resolution of regulatory activity allows to reduce the search space for unveiling genetically perturbed regulation of gene expression specific to SCZ. We anticipate that cell-type specific gene expression and open chromatin exposure profiles in larger sample sets would allow better delimitation of CRD chromatin peak content, facilitate the identification of trans-regulatory hubs across different chromosomes as well as enhance more robust detection of origin effect for gene expression deviation, thereby increasing our understanding of perturbed functional pathways underlying schizophrenia and for prioritizing targets for experimental investigation and novel treatment development.

Methods

Molecular and phenotype data

Molecular and phenotype data for the Human Brain Collection Core (HBCC) was accessed through dbGaP (study accession phs000979.v3.p2; request #88083-1 approved by NIH on January 31st, 2020). All patients met DSM-IV criteria for a lifetime Axis I diagnosis of psychiatric disorders including schizophrenia or schizoaffective disorder. Controls had no history of psychiatric diagnoses or addictions. At least two levels of molecular data (i.e., whole-genome genotype, RNA-sequencing or ChIP-sequencing data) were available for 272 individuals. For genotype data, we determined the intersect of single nucleotide variant (SNV) content across three Illumina genotyping arrays (HumanHap650Y, Human1M-Duov3 and HumanOmni5M-Quad) after filtering the SNVs using standard procedure with PLINK v2.0⁴⁶ and imputed the derived genotype matrix using the Haplotype Reference Consortium reference panel⁴⁷. Next, we applied post-imputation quality control filters by European and African American ancestral group separately and considered the union of filtered SNVs retrieved in both ancestry sets for the final SNV set. This yielded 8,245,179 biallelic SNVs in 272 individuals. Sequence data was mapped onto the human

genome (hg19) with either BWA-MEM v0.7.16⁴⁸ for ChIP-seq data or STAR⁴⁹ for RNA-seq data. Gene expression was quantified using QTLtools⁵⁰ and filtered for protein-coding and lincRNA genes and for the union quantifications detected in $\geq 50\%$ in SCZ cases and in $\geq 50\%$ in controls. This yielded 21,988 genes for 243 individuals. ChIP-seq peak calling and quantification was carried out with HOMER v14.11.1⁵¹. We first determined ChIP-seq peak coordinates across SCZ cases and controls to get a population scale call set of ChIP-seq peaks and then quantified the peaks for each individual according to the identified peak coordinates. This yielded 141,219 ChIP-seq peaks for 193 individuals. To account for confounding factors in gene expression and ChIP-seq peak data, we regressed out ancestry (captured by principal component (PC) analysis on the genotype data), and technical variables (captured by PC analysis on the molecular phenotype data). For the latter, we used the number of PCs that maximized the number of QTLs discovery. Both gene expression and ChIP-seq data were normalized such that these matched a normal distribution with mean 0 and standard deviation 1.

Cis-regulatory domain calling and quantification

For CRD calling, we used the pipeline developed in Delaneau et al. 2019²¹. First, we built a correlation map from chromatin data by systematically measuring interindividual correlation (i.e., Pearson correlation coefficient) between all possible pairs of ChIP-seq peak quantifications located on the same chromosome (within a 250-peak sliding window). Next, we applied hierarchical clustering on the data on a per chromosome basis to get a binary tree that regroups chromatin peaks for each chromosome depending on the correlation levels they exhibited. We relied on three empirical criteria for CRD calling: (i) overall correlation that requires the mean level of correlation within a CRD to be at least twice the background, (ii) edge correlation that requires the mean level of correlation of the peaks at the CRD boundaries to be at least twice the background, and (iii) a requirement that the CRD covers at least two nonoverlapping regulatory elements. We quantified CRD activity on a per-individual basis by enumerating all ChIP-seq peaks per CRD and taking the mean of all single peak quantifications per individual to retrieve a single quantification value for each individual. CRDs called in SCZ cases were used for characterizing SCZ-specific CRDs. CRDs identified in the combined set were used for the rest of the downstream analyses.

CRD structure analysis

For determining CRD sharing between SCZ cases and controls, we compared ChIP-seq peak correlation maps between SCZ cases and controls and called a CRD shared if $\geq 50\%$ of the peaks overlapped. To assess the features of SCZ-specific CRDs, we considered only CRDs in SCZ cases composed of peaks not regrouping into any CRD in controls. For peak activity estimation, we used ChIP-seq peak quantifications uncorrected for covariates and applied a Mann-Whitney U test per peak activity between SCZ cases and controls. Significant differences between SCZ cases and controls were determined at FDR 5% using *R/qvalue* package⁵². For estimating peak activity correlation per SCZ-specific CRD between SCZ cases and controls, we used ChIP-seq peak quantifications corrected for biological and technical covariates, calculated the mean Pearson correlation estimate between peak

activities per CRD and used Mann-Whitney U test for comparing these correlation estimates between SCZ cases and controls.

Differential CRD activity and differential gene expression analysis

Differential CRD activity analysis and differential gene expression analysis were carried out using DESeq2⁵⁵. Significant associations were determined at FDR 5%⁵². For differential CRD activity analysis, we used unnormalized ChIP-sequencing peak read counts obtained with HOMER⁵¹ and summed these up per CRD using the correlation map identified in the combined set. For differential gene expression analysis, we used RNA-sequencing read counts. To identify covariates for correction, we carried out an association testing i) between all available biological and technical covariates and disease status (SCZ/CTL) (Mann-Whitney U test), and ii) between all available biological and technical covariates and individual ChIP-sequencing peak activity or gene expression quantifications (linear regression) and calculated π_1 estimate⁵² to identify the proportion of true associations. R/*clusterProfiler* package⁵⁶ was used for gene set enrichment analysis.

Mapping molecular quantitative trait loci (QTLs)

QTL mapping was carried out using the standard procedure implemented in the QTLtools software package⁵⁰. Specifically, we performed 1,000 permutations to correct for the number of genetic variants being tested in cis per molecular phenotype (+/- 1 Mb window) and corrected for the number of molecular phenotypes being tested genome-wide using false discovery rate (FDR)⁵². To identify multiple QTLs with independent effects on a molecular phenotype, we used the conditional analysis approach based on a forward-backward scan implemented in QTLtools⁵⁰. For SCZ-specific QTL discovery, we considered QTL effects identified in SCZ cases and for each variant-phenotype pair ran a linear regression including genotype, disease status (SCZ/CTL), and covariates, and tested for significance of a genotype * disease status interaction on molecular phenotype (gene expression or CRD activity). This was followed with FDR 5% correction for the number of QTLs tested. We assessed the likelihood of a shared functional effect between SCZ risk variants from four GWAS studies^{5-7,45} and SCZ-identified QTLs using regulatory trait concordance (RTC)^{53,54}.

CRD and gene association

We used QTLtools cis permutation pass⁵⁰ to identify CRDs associated with a gene in a +/-1 Mb window from a transcription start site of a gene. We performed these analyses i) to identify genes associated with SCZ-specific CRDs using CRDs identified in SCZ cases, and ii) to capture comparable associations for SCZ cases and controls using CRDs detected in the combined set. We tested QTL effects for association with the other molecular phenotype (i.e., aCRD-QTLs with gene expression eQTLs with CRD activity) via CRD-gene associations detected across all samples at nominal significance.

Causal inference

To quantify gene-CRD pairs identified as significant at FDR 5% across SCZ cases and controls, we used PC analysis-based dimensionality reduction. For each gene-CRD pair, we aggregated gene expression with CRD activity and used the coordinates on PC1 as new pseudo-phenotypes for QTL mapping in a cis window. This effectively gave us eCRDQTL-CRD-gene triplets consisting of a genetic variant, a CRD and a gene, all associated with each other.

We applied a Bayesian Network approach to infer the most likely causal relationship for eCRDQTL-CRD-gene triplets common to SCZ cases and controls and conducted the analyses separately in SCZ cases and in controls. This approach allowed to estimate the most likely network from which the observed data originates by calculating the posterior probabilities for the three possible causal models⁵⁷: i) causal model in which the genetic variant affects first the CRD and then the gene, ii) reactive model in which the genetic variant affects the gene and then the CRD, iii) independent model in which the genetic variant affects the gene and the CRD independently.

To provide confidence for retrieved probabilities, we carried out 100 bootstrapping runs for each tested triplet separately for SCZ cases and controls using sampling with replacement. We estimated how many times the most probable model across bootstrapping runs for each triplet was the same as in the original Bayesian Network results and filtered out all triplets that fell below a confidence threshold of 55%: this corresponds to the lower quartile value in SCZ cases. *R/clusterProfiler* package⁵⁶ was used for gene set enrichment analysis for genes that belonged to triplets showing directional change from eCRD-QTL onto gene expression/CRD activity between SCZ cases and controls. We considered two scenarios: i) causal model in controls, but reactive/independent in SCZ cases; ii) reactive/independent in controls, but causal in SCZ cases.

References

- 1 Sullivan, P. F. & Geschwind, D. H. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. *Cell* **177**, 162-183, doi:10.1016/j.cell.2019.01.015 (2019).
- 2 Rapoport, J. L., Giedd, J. N. & Gogtay, N. Neurodevelopmental model of schizophrenia: update 2012. *Mol Psychiatry* **17**, 1228-1238, doi:10.1038/mp.2012.23 (2012).
- 3 McGuffin, P., Farmer, A. E., Gottesman, I. I., Murray, R. M., Reveley, A. M. Twin concordance for operationally defined schizophrenia. Confirmation of familiarity and heritability. *Arch Gen Psychiatry* **41**, 541-545 (1984).
- 4 Sullivan, P. F., Kendler, K. S. & Neale, M. C. Schizophrenia as a complex trait: evidence from a meta-analysis of twin studies. *Arch Gen Psychiatry* **60**, 1187-1192, doi:10.1001/archpsyc.60.12.1187 (2003).
- 5 Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature* **511**, 421-427, doi:10.1038/nature13595 (2014).

- 6 Pardinas, A. F. *et al.* Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet* **50**, 381-389, doi:10.1038/s41588-018-0059-2 (2018).
- 7 The Schizophrenia Working Group of the Psychiatric Genomics Consortium. Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. Preprint at <https://doi.org/10.1101/2020.09.12.20192922> (2020).
- 8 Bacanu, S. A. *et al.* Functional SNPs are enriched for schizophrenia association signals. *Mol Psychiatry* **19**, 276-277, doi:10.1038/mp.2013.33 (2014).
- 9 Jaffe, A. E. *et al.* Developmental and genetic regulation of the human cortex transcriptome illuminate schizophrenia pathogenesis. *Nat Neurosci* **21**, 1117-1125, doi:10.1038/s41593-018-0197-y (2018).
- 10 Wang, D. F. *et al.* Comprehensive functional genomic resource and integrative model for the human brain. *Science* **362**, doi:10.1126/science.aat8464 (2018).
- 11 Fullard, J. F. *et al.* An atlas of chromatin accessibility in the adult human brain. *Genome Res* **28**, 1243-1252, doi:10.1101/gr.232488.117 (2018).
- 12 Bryois, J. *et al.* Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat Commun* **9**, 3121, doi:10.1038/s41467-018-05379-y (2018).
- 13 Won, H. *et al.* Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature* **538**, 523-527, doi:10.1038/nature19847 (2016).
- 14 Huo, Y., Li, S., Liu, J., Li, X. & Luo, X. J. Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nat Commun* **10**, 670, doi:10.1038/s41467-019-08666-4 (2019).
- 15 Radhakrishnan, R., Kaser, M. & Guloksuz, S. The Link Between the Immune System, Environment, and Psychosis. *Schizophrenia Bull* **43**, 693-697, doi:10.1093/schbul/sbx057 (2017).
- 16 Karlic, R., Chung, H. R., Lasserre, J., Vlahovicek, K. & Vingron, M. Histone modification levels are predictive for gene expression. *P Natl Acad Sci USA* **107**, 2926-2931, doi:10.1073/pnas.0909344107 (2010).
- 17 Kilpinen, H. *et al.* Coordinated effects of sequence variation on DNA binding, chromatin structure, and transcription. *Science* **342**, 744-747, doi:10.1126/science.1242463 (2013).
- 18 Wong, E. S. *et al.* Interplay of cis and trans mechanisms driving transcription factor binding and gene expression evolution. *Nat Commun* **8**, 1092, doi:10.1038/s41467-017-01037-x (2017).

- 19 McVicker, G. *et al.* Identification of Genetic Variants That Affect Histone Modifications in Human Cells. *Science* **342**, 747-749, doi:10.1126/science.1242429 (2013).
- 20 Gusev, A. *et al.* Transcriptome-wide association study of schizophrenia and chromatin activity yields mechanistic disease insights. *Nature Genetics* **50**, 538-548, doi:10.1038/s41588-018-0092-1 (2018).
- 21 Delaneau, O. *et al.* Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science* **364**, doi:10.1126/science.aat8266 (2019).
- 22 Fulton, D. L., Denarier, E., Friedman, H. C., Wasserman, W. W. & Peterson, A. C. Towards resolving the transcription factor network controlling myelin gene expression. *Nucleic Acids Res* **39**, 7974-7991, doi:10.1093/nar/gkr326 (2011).
- 23 Nakajima, K. *et al.* Molecular Motor KIF5A Is Essential for GABA(A) Receptor Transport, and KIF5A Deletion Causes Epilepsy. *Neuron* **76**, 945-961, doi:10.1016/j.neuron.2012.10.012 (2012).
- 24 Miczan, V. *et al.* NECAB1 and NECAB2 are Prevalent Calcium-Binding Proteins of CB1/CCK-Positive GABAergic Interneurons. *Cereb Cortex* **31**, 1786-1806, doi:10.1093/cercor/bhaa326 (2021).
- 25 Zhao, S. F., Li, F. W., Leak, R. K., Jun, C. & Hu, X. M. Regulation of neuroinflammation through programmed death-1/programmed death ligand signaling in neurological disorders. *Front Cell Neurosci* **8**, doi:10.3389/fncel.2014.00271 (2014).
- 26 Fromer, M. *et al.* Gene expression elucidates functional impact of polygenic risk for schizophrenia. *Nat Neurosci* **19**, 1442-1453, doi:10.1038/nn.4399 (2016).
- 27 Reay, W. R. & Cairns, M. J. The role of the retinoids in schizophrenia: genomic and clinical perspectives. *Mol Psychiatry* **25**, 706-718, doi:10.1038/s41380-019-0566-2 (2020).
- 28 O'Donovan, S. M. *et al.* Cell-subtype-specific changes in adenosine pathways in schizophrenia. *Neuropsychopharmacology* **43**, 1667-1674, doi:10.1038/s41386-018-0028-6 (2018).
- 29 Kraft, R. *et al.* Phenotypes of Drosophila brain neurons in primary culture reveal a role for fascin in neurite shape and trajectory. *J Neurosci* **26**, 8734-8747, doi:10.1523/Jneurosci.2106-06.2006 (2006).
- 30 Bell, S. *et al.* Mutations in ACTL6B Cause Neurodevelopmental Deficits and Epilepsy and Lead to Loss of Dendrites in Human Neurons. *Am J Hum Genet* **104**, 815-834, doi:10.1016/j.ajhg.2019.03.022 (2019).
- 31 Deinhardt, K. *et al.* Neuronal Growth Cone Retraction Relies on Proneurotrophin Receptor Signaling Through Rac. *Sci Signal* **4**, doi:10.1126/scisignal.2002060 (2011).

- 32 Nozumi, M., Nakatsu, F., Katoh, K. & Igarashi, M. Coordinated Movement of Vesicles and Actin Bundles during Nerve Growth Revealed by Superresolution Microscopy. *Cell Rep* **18**, 2203-2216, doi:10.1016/j.celrep.2017.02.008 (2017).
- 33 Karczewski, K. J. *et al.* The mutational constraint spectrum quantified from variation in 141,456 humans (vol 581, pg 434, 2020). *Nature* **590**, E53-E53, doi:10.1038/s41586-020-03174-8 (2021).
- 34 Mendizabal, I. *et al.* Cell type-specific epigenetic links to schizophrenia risk in the brain. *Genome Biology* **20**, doi:10.1186/s13059-019-1747-7 (2019).
- 35 Qu, L. *et al.* The Ras Superfamily of Small GTPases in Non-neoplastic Cerebral Diseases. *Front Mol Neurosci* **12**, doi:10.3389/fnmol.2019.00121 (2019).
- 36 Bhambhani, H. P., Mueller, T. M., Simmons, M. S. & Meador-Woodruff, J. H. Actin polymerization is reduced in the anterior cingulate cortex of elderly patients with schizophrenia. *Transl Psychiat* **7**, doi:10.1038/s41398-017-0045-y (2017).
- 37 Yan, Z., Kim, E., Datta, D., Lewis, D. A. & Soderling, S. H. Synaptic Actin Dysregulation, a Convergent Mechanism of Mental Disorders? *J Neurosci* **36**, 11411-11417, doi:10.1523/Jneurosci.2360-16.2016 (2016).
- 38 Costa, J. F., Dines, M. & Lamprecht, R. The Role of Rac GTPase in Dendritic Spine Morphogenesis and Memory. *Front Synaptic Neuro* **12**, doi:10.3389/fnsyn.2020.00012 (2020).
- 39 Tessier, C. *et al.* Membrane lipidomics in schizophrenia patients: a correlational study with clinical and cognitive manifestations. *Transl Psychiat* **6**, doi:10.1038/tp.2016.142 (2016).
- 40 Aguet, F. *et al.* The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science* **369**, 1318-1330, doi:10.1126/science.aaz1776 (2020).
- 41 Dobbyn, A. *et al.* Landscape of Conditional eQTL in Dorsolateral Prefrontal Cortex and Co-localization with Schizophrenia GWAS. *Am J Hum Genet* **102**, 1169-1184, doi:10.1016/j.ajhg.2018.04.011 (2018).
- 42 de la Torre-Ubieta, L. *et al.* The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis. *Cell* **172**, 289-304, doi:10.1016/j.cell.2017.12.014 (2018).
- 43 Girdhar, K. *et al.* Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nature Neuroscience* **21**, 1126-1136, doi:10.1038/s41593-018-0187-0 (2018).
- 44 de Candia, T. R. *et al.* Additive Genetic Variation in Schizophrenia Risk Is Shared by Populations of African and European Descent. *Am J Hum Genet* **93**, 463-470, doi:10.1016/j.ajhg.2013.07.007 (2013).

- 45 Bigdeli, T. B. *et al.* Contributions of common genetic variants to risk of schizophrenia among individuals of African and Latino ancestry. *Mol Psychiatry* **25**, 2455-2467, doi:10.1038/s41380-019-0517-y (2020).
- 46 Chang, C. C. *et al.* Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* **4**, 7, doi:10.1186/s13742-015-0047-8 (2015).
- 47 McCarthy, S. *et al.* A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279-1283, doi:10.1038/ng.3643 (2016).
- 48 Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. Preprint at *arXiv:1303.3997v2* (2013).
- 49 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 50 Delaneau, O. *et al.* A complete tool set for molecular QTL discovery and analysis. *Nat Commun* **8**, 15452, doi:10.1038/ncomms15452 (2017).
- 51 Heinz, S. *et al.* Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* **38**, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- 52 Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc Natl Acad Sci U S A* **100**, 9440-9445, doi:10.1073/pnas.1530509100 (2003).
- 53 Nica, A. C. *et al.* Candidate Causal Regulatory Effects by Integration of Expression QTLs with Complex Trait Genetic Associations. *Plos Genet* **6**, doi:10.1371/journal.pgen.1000895 (2010).
- 54 Ongen, H. *et al.* Estimating the causal tissues for complex traits and diseases. *Nat Genet* **49**, 1676-1683, doi:10.1038/ng.3981 (2017).
- 55 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 56 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS* **16**, 284-287, doi:10.1089/omi.2011.0118 (2012).
- 57 Scutari, M. Learning Bayesian Networks with the bnlearn R Package. *J Stat Softw* **35**, 1-22, doi:DOI 10.18637/jss.v035.i03 (2010).

Declarations

ACKNOWLEDGEMENTS

This research was supported by the Intramural Research Program of the NIMH (NCT00001260, 900142), the Estonian Research Council ETAg grant no PUTJD901 (to M.A.) and the National Centre of Competence in Research (NCCR) grant Synapsy – The Synaptic Bases of Mental Diseases (no 51NF40-158776) from the Swiss National Science Foundation. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

AUTHOR CONTRIBUTIONS

E.T.D, H.O. designed and supervised the study and contributed to data interpretation. M.A. conducted the primary analysis and results interpretation. M.A. and H.O. performed the primary manuscript writing. N.L. and A.R. contributed to data analysis and interpretation and manuscript writing.

COMPETING INTERESTS STATEMENT

Emmanouil Dermitzakis is currently an employee of GSK. The work presented in this manuscript was performed before he joined GSK.

DATA AVAILABILITY

Phenotype and molecular data for the HBCC cohort are deposited at dbGaP (study accession phs000979.v3.p2; https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000979.v3.p2). Genotype data for 1000 Genomes Project are deposited at <https://www.internationalgenome.org/data-portal/data-collection>.

Tables

Table 1. Molecular phenotype associations at FDR 5% in SCZ cases and controls. The columns indicate the molecular phenotype and type of association, and the numbers of identified associations and sample set used. *Gene-CRD associations found using CRDs identified in the combined set or CRDs identified only in SCZ cases.

Molecular phenotype and type of association	Tested sample		
GENE EXPRESSION	SCZ (n=83)	CTL (n=160)	SCZ+CTL (n=243)
Differentially expressed genes	1,363 (937 ↑, 426 ↓)		
eQTL	867	6,166	10,297
SCZ-specific eQTL	130		
CRDs	SCZ (n=74)	CTL (n=119)	SCZ+CTL (n=193)
cQTL	3,239	13,158	26,830
CRDs	10,938	10,376	11,374
Differentially active CRDs	1,141 (542 ↑, 599 ↓)		
aCRD-QTL	796	2,929	4,979
SCZ-specific aCRD-QTL	40		
QTL-CRD-gene	SCZ (n=59)	CTL (n=105)	SCZ+CTL (n=164)
Gene-CRD pairs	95/119*	634	1,197
eCRDQTL-CRD-gene triplets	38	577	1,100

Figures

Figure 1

Features of SCZ-specific CRDs. (a) Difference in medians per peak activity between SCZ cases and controls as a function of the strength of association given in $-\log_{10} p$ -values; purple dots denote peaks that are differentially active between SCZ cases and controls at FDR 5% (3,450 peaks) with light purple and dark purple indicating lower and higher median activity, respectively, in SCZ cases compared to controls. (b) Comparison of differences in per CRD peak activity correlation estimates between SCZ cases and controls for SCZ-specific CRDs (3,078 CRDs). (c) Example region of a correlation structure between 250 peaks on chromosome 5 in SCZ cases and controls, revealing a well-delimited SCZ-specific CRD that is composed of 5 regulatory elements.

Figure 2

Association between CRDs and genes. (a) TSSs of differentially expressed genes (DEG) are localized within differentially active CRDs significantly more often than expected by chance (Fischer's exact test $p = 8.72 \times 10^{-6}$, odds ratio 1.60); coloured dots denote DEGs identified at FDR 5%: purple dots mark DEGs, turquoise dots mark DEGs with TSS within differentially active CRD. (b) Distribution of gene-to-CRD distances for genes localizing outside the associated CRD boundary (545 gene-CRD associations). (c) Distribution of the relative position of gene TSS to the boundary of an associated CRD (652 gene-CRD associations).

Figure 3

Example of a SCZ-specific (a) aCRD-QTL and (b) eQTL effect. Genotype-dependent effect on CRD activity and gene expression identified only in SCZ cases.

Figure 4

Regulatory mechanisms for eCRDQTL-CRD-gene triplets in SCZ cases and controls.: Comparison of the direction of effect from eCRD-QTL onto gene expression and CRD activity for tested triplets between SCZ cases and controls: (a) triplet count for models showing the same regulatory mechanism in SCZ cases and controls, and (b) triplet count for models showing a change in the regulatory mechanism between SCZ cases and controls (term “different” indicates either reactive or independent model; light blue bar indicates triplets, for which the causal model (i.e., mediation via CRD for QTL effect) was not identified in SCZ cases nor in controls). (c) Genotype-dependent effect for eCRD-QTL 7:5849993:C:G on *FSCN1* expression. (d) Distinct regulatory mechanism of genetic regulation on gene expression for SCZ cases and controls for a triplet consisting of an eCRD-QTL 7:584993:C:G, gene *FSCN1* and a CRD composed of 5 REs on chr7:5623132-5705414; the probabilities based on Bayesian Networks for each tested model is given above schematics; shading of the colour for the gene and for the CRD indicates strength in expression and activity, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [SCZCRDsupplementarytablesAlver.xlsx](#)
- [SupplementaryInformationAlver.pdf](#)