Cis-Regulatory Hubs: a Relevant 3D Model to Study the Genetics of Complex Diseases with an Application to Schizophrenia

Authors

Loïc Mangnier [1,2,3,7], Charles Joly-Beauparlant [3,4], Arnaud Droit [3,4,7], Steve Bilodeau [5,6,7,8], Alexandre Bureau [1,2,7]

Corresponding Authors

steve.bilodeau@crchudequebec.ulaval.ca, alexandre.bureau@fmed.ulaval.ca

Affiliations

1 Centre de Recherche CERVO, Québec, Canada
2 Département de Médecine Sociale et Préventive, Université Laval, Québec, Canada
3 Centre de Recherche du CHU de Québec - Université Laval, Québec, Québec, Canada
4 Département de Médecine Moléculaire, Université Laval, Québec, Canada
5 Centre de recherche du CHU de Québec – Université Laval, Axe Oncologie, Québec, Québec, Canada, G1V 4G2
6 Centre de Recherche sur le Cancer de l'Université Laval, Québec, Québec, Canada, G1R3S3
7 Centre de recherche en données massives de l'Université Laval, Québec, Québec, Canada, G1V 0A6
8 Département de biologie moléculaire, biochimie médicale et pathologie, Faculté de Médecine, Université Laval, Québec, Québec, Canada, G1V 0A6

Abstract

Background: The 3-dimensional (3D) conformation of the chromatin creates complex networks of noncoding regulatory regions (distal elements) and genes with important implications in gene regulation. Despite the importance of the role of noncoding regions in complex traits, little is known about their interplay within regulatory hubs and the implication in multigenic diseases like schizophrenia.

Results: Here we show that cis-regulatory hubs (CRHs) in neurons highlight functional interactions between distal elements and promoters, providing a model to explain the epigenetic mechanisms involved in complex diseases. CRHs represent a new 3D model, where several distal elements interact to create a complex network of active genes. Indeed, we found that CRHs represent functional structures, showing higher transcriptional activity. In a disease context, CRHs highlighted strong enrichments in schizophrenia-associated genes, schizophrenia-associated SNPs and schizophrenia heritability compared to equivalent tissue- and non-tissue-specific structures. Also, genes, by sharing the same distal elements, converge to common biological processes associated with schizophrenia. Finally, the results showed that in a complex disease etiology, small CRHs by linking fewer distal elements to promoters constitute a more informative structure than larger hubs.

Conclusion: CRHs are a new 3D model of the chromatin interactions between gene promoters and their distal elements highlighting causal regulatory processes and providing a better understanding of complex disease etiology such as schizophrenia. Indeed, by providing a finer scale chromosome architecture, we have genetic and statistical evidence that CRHs represent a major advancement in 3D models to study the epigenetic underlying processes involved in complex traits.

Introduction

The etiology of complex diseases involves a large range of causal factors, both genetic and environmental, leading to gene expression changes (Vliet et al., 2007; Do et al., 2017). Models currently used in the etiology of complex diseases suggest that the majority of risk variants are located in noncoding regions explaining a large portion of the heritability (Maurano et al., 2012). Indeed, most risk variants are enriched in distal noncoding regions, disturbing the tissue-specific transcriptional program and therefore playing a key role in the disease etiology (Zhang & Lupski, 2015). The difficulty to assign distal regulatory elements to genes hampered the ability to discover the underlying molecular mechanisms. Since the role of noncoding regions in complex phenotypes has been well-described, there is also strong evidence on the 3-Dimensional (3D) organization of the genome in the gene regulation. The addition of the 3D genome organization, captured by chromosome conformation assays (van Berkum et al., 2010), revealed the physical proximity between regulatory elements and genes and their involvement in gene regulation. In addition to chromatin loops connecting promoters to distal noncoding regions (Gorkin et al., 2014; Bouwman & de Laat, 2015; Dekker & Mirny, 2016), the genome is parsed into larger domains including topologically associating domains (TADs) (Dixon et al., 2012) and A/B compartments (Lieberman-Aiden et al., 2009). Interestingly, DNA sequence variations influencing the 3D genome organization are associated with complex disease risks (Gorkin et al., 2019). For example, structural variants disrupting TADs, which are enriched in enhancer-promoter interactions, lead to fused-TADs promoting ectopic promoter-enhancer connections and disruption of the normal transcriptional program (Fudenberg & Pollard, 2019; Melo et al., 2020). However, precisely identifying which genes are affected by a risk variant remains a challenge.

The combination of chromatin interactions and microscopy-based techniques established that groups of genes share the same physical environment (Cardozo Gizzi et al., 2019). In fact, promoters interact with enhancers inside complex organizations, forming regulatory hub structures (Oudelaar et al., 2019; Campigli et al., 2020). These hubs exhibit distinct organization from known 3D features, encompassing in the majority of cases a few numbers of promoters, strongly involved in biological processes (Espinola et al., 2021). On the one hand, highly interconnected enhancers converge to genes with crucial phenotypic implications, with dynamic enhancer cross-talk at genome-wide level occurring more frequently during differentiation (Madsen et al., 2020). Interestingly, promoters are often regulated by several enhancers (Espinola et al., 2021). This model supports that distal enhancers regulate a single or multiple gene transcription by sharing resources (Espinola et al. 2021). At the molecular level, enhancers increase the gene activity through modulation of transcriptional bursting (Fukaya et al., 2016) or indirectly influencing transcription activation (Benabdallah et al., 2019). Furthermore, super interacting promoters are enriched in lineage-specific genes (Song et al., 2020), known to play a crucial role in diseases, while multiple enhancers connected to a promoter provide phenotypic robustness in environmental or genetic perturbations (Tsai et al., 2019). Interestingly, the organisation of genes and noncoding regulatory regions may be pre-established, present in different cells, highly dynamic during differentiation (Rubin et al., 2017; Espinola et al., 2021). However, whether and how promoters and enhancers interacting in hubs are involved in the etiology of complex diseases are still open questions.

Schizophrenia is a complex chronic brain disorder associated with perturbations in the transcriptional programs of neurons. Indeed, schizophrenia is characterized by long-standing desillusions and hallucinations strongly reducing life-expectancy (Sullivan et al., 2012). Recent findings suggest that schizophrenia is explained by polygenic architecture (Smeland et al., 2020), where it has been shown that most of the risk variants are located in noncoding regions. Schizophrenia risk loci have been shown to be enriched in active enhancers or promoters in neurons from the adult human frontal lobe (Roussos et al., 2014; Fullard et al., 2018; Girdhar et al., 2018; Hauberg et al., 2020). Also, recent studies have demonstrated the involvement of 3D organization in the trait. For example, chromatin loops have shown to be enriched in expression quantitative trait loci (eQTLs) or schizophrenia-risk variants impacting the proximal gene regulation (Rajarajan et al., 2018). In addition, ultra-rare structural variants in TAD borders lead to gene dysregulations increasing the risk of schizophrenia (Halvorsen et al., 2020). However, the implication of regulatory hubs in the schizophrenia etiology remains to be addressed.

In the present study, we are defining cis-regulatory hubs (CRHs) as 3D structures linking one or more gene promoters to networks of distal elements which capture complex patterns of gene regulation. In neurons, CRHs are strongly enriched in schizophrenia-associated genes, SNPs, and heritability compared to equivalent structures. Therefore, CRHs are consistent with current models suggesting that promoters interact more than distal elements. We provide evidence that smaller CRHs, due to fewer connections between promoters and distal elements, are more impacted by risk variants than large CRHs, supporting their relevance in disease etiology.

Results

Promoters and distal elements interactions create cis-regulatory hubs in neurons

To understand regulatory processes involved in complex phenotypes, we built cis-regulatory hubs (CRHs) as bipartite graphs using the 3D contacts between the promoter of genes and their distal elements. We defined CRHs using chromatin contacts provided by Hi-C data with and without additional epigenetic features defining classes of distal regulatory elements. to evaluate their role in schizophrenia etiology (See Supplementary Methods). Since open chromatin regions in the prefrontal cortex of schizophrenia individuals have shown to be enriched in risk variants (Bryois et al., 2018) and that H3K27ac regions are strongly associated with schizophrenia-risk variants (Girdhar et al., 2018), we focused our attention on the activity-by-contact (ABC) model (Fulco et al., 2019) of enhancer-promoter interactions. The approach integrates the frequency of physical contacts between distal elements and promoters (500bp from an annotated TSS) with the activity defined by the DNAse accessibility and the occupancy of H3K27ac, taking into account the background activity (See Methods). The ABC model is a good predictor of differential gene expression (Fulco et al., 2019) and a useful tool to link noncoding variants to their target genes (Nasser et al., 2021).

Using available datasets in neurons derived from induced pluripotent stem cells (iPSC) (Rajarajan et al., 2018), the most relevant cell type to study schizophrenia (Sey et al., 2020), we identified 62,658 functional pairs of distal elements and promoters. To build CRHs, the networks surrounding each distal element and promoter were expanded to include directly connected regions. We identified 1,633 CRHs, ranging between 2 and 506 nodes (median of 6 elements). For example, genes involved in glutamatergic transmission or synaptic plasticity pathways (*GRIA1*, *GRIN2A*, and *GRM3*), known to be associated with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), exhibited strong differences regarding CRH complexity (**Figure 1A**). Among the 1,633 CRHs, 15% were pairs of 2 nodes and therefore constituted monogamous relationships, while 85% had 3 elements or more (**Figure 1B**). On average, CRHs contained a significantly higher number of distal elements than promoters, up to 2-fold more (median of 5 distal elements against 2 promoters, Wilcoxon signed-rank test: p-value<=2e-16) (**Figure 1C**). Accordingly, promoters were more connected than distal elements as the 80% least connected promoters had at least twice as many connections as the corresponding 80% distal elements (**Figure 1D**). As expected, the proportion of distal elements was positively correlated with the number of nodes (or complexity) of the CRHs (Spearman tau=0.37, p-value<=2e-16), revealing that complex CRHs are significantly associated with a higher proportion of distal elements. These results suggest that distal regulatory elements and gene promoter regions are organized into complex regulatory structures in neurons.

The connectivity between promoters and enhancers is strongly associated with the emergence of tissue-specific phenotypes as they control the transcriptional program (Tsai et al., 2019). Recent studies have shown that highly connected enhancers converge to genes with strong phenotypic impacts (Madsen et al., 2020), while promoters enriched in connections are more tissue-specific (Song et al., 2020). Since we expected that connections of genes or distal elements may play a role in disease emergence, we

investigated in more detail the organization of genes and distal elements in CRHs of 3 nodes or more. Thus, we defined two metrics aiming to characterize genes and distal elements involved in these complex relationships (**Figure 1A**): 1) the proportion of two or more distal elements sharing a promoter or two or more promoters sharing a distal element (i.e., 1-1-N with N>0) and 2) the proportion of polygamous elements (i.e., which are not in a monogamous pair or 1-1-N, forming complex shared interactions by promoters or distal elements). Interestingly, most distal elements were connected to a single promoter while promoters showed interactions with multiple distal elements. Indeed,63% of distal elements were connected to a single promoter while 90% of promoters were linked to 2 or more distal elements. In fact, 1% of promoters are within 1-1-N relationships, while 5% of distal elements are within this kind of relationship (**Figure 1E**). This result suggests that distal elements more frequently share a gene than genes sharing distal elements, in accordance with previous findings in model organisms (Espinola et al., 2021). Controlling with other CRH definitions, we found that promoters were also more connected than distal elements. However, promoters showed fewer connections in our other CRH definitions than the activity-by-contact approach (**Figure S1**). Therefore, the proposed definition of CRHs aligns with previous models suggesting that distal elements interact more specifically (Madsen et al., 2020), whereas promoters are more frequent inside complex relationships.

Next, we wanted to determine the relationship between CRHs and known 3D structures. We focused our analysis on A/B compartments, TADs, and frequently interacting regions (FIREs), respectively segmenting the genome into open and close chromatin, domains of frequent interactions between distal elements and genes, and hotspots for chromatin contacts. The majority of CRHs (76%) shared compartments of the same type, with 46% and 29% for active and inactive compartments, respectively (**Figure 1F**). On the other hand, only a minor portion (8%) of CRHs overlapped several compartments of different types or were in genomic regions not assigned to a compartment (17%). Since it has been shown that A compartments correlated strongly with presence of genes, accessible chromatin, activating and repressive histone marks (Lieberman-Aiden et al., 2009), we argue that CRHs are consistent with the open-chromatin characteristic associated with functional elements. Moreover, most of CRHs (64%) overlapped a single TAD (**Figure 1G**). Interestingly, 26% of TADs included 2 or more CRHs. These observations were confirmed by testing multiple TAD detection algorithms (**Figure S2**). Lastly, CRHs were enriched in FIREs compared to candidate CRHs, tissue-specific regions non integrating 3D contacts (see Methods) (Odds ratio=1.41, p-value<=2.2e-16), although only a minor portion of distal elements or promoters overlapped with FIREs (11% and 13%, respectively). The presence of CRHs within compartments and TADs in addition to the enrichment in FIREs was confirmed using different CRH definitions (**Figure S3**). Collectively, our results support that CRHs are networks of interacting regulatory regions and genes at a finer scale compared to previously defined chromosome structures.

CRHs are defined by active chromatin and the presence of schizophrenia-relevant genes

Genes and regulatory elements sharing the same nuclear environment often show coherent transcriptional states and related molecular functions (Campigli et al., 2020). To further characterize the transcriptional activity of CRHs, we overlayed the chromatin states defined by the Roadmap Epigenomics (Roadmap Epigenomics Consortium et al., 2015). The 18-states model in neurons was subdivided as follows: 1) Active (1_TssA, 2_TssFlnk, 3_TssFlnkU, 4_TssFlnkD, 5_Tx, 7_EnhG1, 8_EnhG2, 9_EnhA1, 10_EnhA2, 12_ZNF/Rpts), 2) Weakly Active (6_TxWk, 11_EnhWk, 14_TssBiv, 15_EnhBiv), and 3) Inactive/Repressor (13_Het, 16_ReprPC, 17_ReprPCWk, 18_Quies). We found that the majority of CRH elements (72%) overlapped at least one Inactive or Repressor region against 53% and 39% of CRH elements for Active or Weakly Active regions, respectively (**Figure 2A left**). Although we observed that most of the CRH elements included the Quiescent state (**Figure 2A right**), CRHs were enriched 2.35-fold in active states and depleted in inactive states (Odds ratio=0.49, p-value<=2e-16) compared to candidate CRHs (**Figure 2B**). To confirm the enrichment of CRHs in functional elements, we used ENCODE candidate elements in neurons (The ENCODE Project Consortium et al., 2020). ENCODE candidate elements are regions exhibiting significant signals in H3K4me3, H3K27ac, DNAse or CTCF. CRHs were strongly associated with H3K4me3 (Odds ratio=1.82, p-value<=2e-16), DNAse (Odds ratio=1.67, p-value<=2e-16) and H3K27ac (Odds ratio=1.44, p-value<=2e-16), while CTCF was not. Our results were supported by other CRH definitions (**Figure S4**). Then, to extract the global pattern of

chromatin states within a CRH, we kept chromatin states representing up to 80% of the total chromatin state signal and observed a striking difference in CRH organizations. Indeed, 35% of CRHs exhibit a unique combination of chromatin states (e.g., found only once in CRHs), whereas about half of CRHs were characterized mainly by up to 3 chromatin states (**Table S1**). Also, CRHs characterized mainly by active states were more complex than those strongly defined by quiescent states (18_Quies) (**Figure 2C**). Considering the above findings, CRHs are enriched in active distal elements and exhibit a variety of chromatin state combinations, suggesting they are important for the control of the transcriptional program of neurons.

If CRHs are enriched in active elements in neurons, we postulated that they would be enriched in schizophrenia-relevant genes. First, we identified 8,075 genes associated with schizophrenia (FDR<=0.05) using H-Magma (Sey et al., 2020), a statistical approach using 3D noncoding regions with genetic data from genome-wide association study for schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2020). We found that 35% of genes significantly associated with schizophrenia are within CRHs compared to 23% for all other genes (1.81-fold enrichment right-tailed exact Fisher test, p-value<=2.2e-16). Moreover, 42% (687/1633) of CRHs include at least 1 schizophrenia-associated gene with 23% (376/1633) harboring several schizophrenia-related genes (mean=1.76, max=70) (**Figure 2D**). Finally, we found that CRHs were enriched in GO biological processes associated with schizophrenia (**Figure 2E**). Taken together, these results suggest that CRHs are associated with the pathoetiology of schizophrenia, constituting an interesting model for understanding gene regulation and the emergence of complex phenotypes.

CRHs containing schizophrenia-associated genes are small and highly expressed

To further characterize CRHs including schizophrenia-associated genes, we included complexity measures. Interestingly, CRHs encompassing schizophrenia-associated genes showed larger distances between elements than CRHs not harboring schizophrenia-associated genes (**Figure 3A**). The number of connections was slightly lower with distal elements than non-associated genes (mean associated genes= 4.39, mean non-associated genes=4.55, Two-tailed t-test p-value=0.004). In addition, schizophrenia-associated genes showed higher expression levels than non-associated genes (median associated=6.18, median non-associated=5.87, Wilcoxon rank-sum test p-value<=2.2e-16) (**Figure 3B**) and were enriched in active distal elements (Odds ratio=1.7902, p-value<=2.2e-16). Moreover, schizophrenia-associated genes were more often monogamous genes compared to non-associated ones, showing a 1.33-fold enrichment. Indeed, 25% of monogamous genes are schizophrenia-associated genes against 20% for non-monogamous ones (**Figure 3C**). The number of distal elements in CRHs harboring schizophrenia-associated genes were correlated negatively with the proportion of associated genes (Spearman tau=-0.47, p-value<=2.2e-16). These results suggest that schizophrenia-associated genes are, in most cases, within small hubs, less connected to distal elements, but expressed at higher levels than non-associated genes.

To confirm that schizophrenia-associated genes are expressed at higher levels, less connected to distal elements, and enriched in active distal elements, we fitted a logistic regression. The status of genes (associated vs. non-associated with schizophrenia), adjusted on gene expression, the number of connections to distal elements, the 90th percentile of the proportion of active distal elements per gene, and the information regarding monogamy were included. As expected, the gene status regarding its association with schizophrenia was positively associated with expression, the 90th percentile of the proportion of active distal elements, and the monogamy status, while negatively associated with the number of connections, confirming our results found with univariate analyses (**Figure 3D**). Collectively, our results suggest that schizophrenia-associated genes are within small hubs characterized by fewer connections to distal elements and higher transcriptional activity.

CRHs are enriched in schizophrenia-associated SNPs and heritability

Current models suggest that distal regulatory regions explain a great proportion of the schizophrenia etiology (Roussos et al., 2014). In fact, a wide range of genetic variants affecting the gene expression program are involved in the trait (Huo et al., 2019). Since we demonstrated the enrichment in schizophrenia-relevant genes within CRHs, we next assessed the presence of schizophrenia-associated SNPs. We collected 99,194 SNPs with a p-value of association with schizophrenia below various thresholds from genome-wide association studies (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2020). We mapped them to their corresponding CRH and quantified their enrichments using the right-tailed Fisher exact test. For instance, there were 2,058 SNPs with a p-value<=1e-4. At this significance level, we observed enrichments (Odds ratio=1.29, p-value=0.04) in CRHs compared to the candidate CRHs (**Figure 4A**). Then, we used the same methodology as Nasser et al., 2021 to define enrichments in a given functional annotation (proportion of significant SNPs for schizophrenia/proportion of all common SNPs), where the proportion of all common SNPs is the proportion of all common SNPs overlapping a functional annotation, given by the LDSC. Consistent with our previous finding, we observed higher fold enrichments (Enrichment for CRHs / Enrichment for candidate CRHs) for CRHs, becoming stronger with the significance level (**Figure 4B**). This enrichment was stronger with alternative definitions of CRHs (**Figure S5**). Therefore, our results suggest that CRHs are enriched in SNPs for schizophrenia.

After demonstrating the relevance of CRHs with schizophrenia-associated SNPs, we wondered whether they explained schizophrenia heritability. To this end, we leveraged linkage disequilibrium score regression (LDSC) (Finucane et al., 2015) which provides the portion of disease heritability using a functional annotation. First, comparing CRHs to equivalent non-tissue-specific noncoding regions, we ensured to maximize the explained heritability by using tissue-specific elements and integrating 3D contacts. We conditioned on enhancers, promoters, H3K27ac, and DNAse peaks from the LDSC baseline model. In addition, we compared CRHs to equivalent components, defining candidate CRHs as tissue-specific elements with the same characteristics as CRHs without 3D contacts. In doing so, we found higher significant heritability in CRHs than non-tissue-specific noncoding regions and candidate CRHs (**Figure 4C**), with strong associations for CRHs compared to candidate CRHs (Z-Score CRHs=2.41, two-sided p-value = 0.01; Z-Score candidate CRHs=-1.84, two-sided p-value = 0.065). CRHs explained 11-fold more heritability than their respective candidate CRHs or up to 44-fold more than non-tissue-specific elements (**Table S2**). When compared to methods building hubs using only the chromatin contacts without additional epigenetic definition of regulatory elements and with DNAse only, our definition of CRH performed better with schizophrenia heritability, showing enrichments of 3.08 against 0.84 and 2.98, respectively (**Figure S6**). This result demonstrates a better concordance of the CRH including epigenetic features to explain schizophrenia heritability compared to only using chromatin interactions.

Since we observed that schizophrenia-associated genes are enriched in small hubs, highly expressed, and connected to few distal elements, we distinguished strata of CRH number of promoters based on the intraclass correlation (ICC) through a linear mixed model of the gene expression (**Figure S7**). Briefly, ICC is the proportion of total variance explained by CRHs. Supporting our previous findings, we found that small CRHs (<= 3 promoters) are more enriched in schizophrenia heritability than medium (>3 and <=25 promoters) or large ones (>25 promoters) (**Figure 4D**). Overall, these results support that CRHs are a relevant structure to explain the etiology of schizophrenia, where small CRHs are a more informative target to study the trait.

Discussion

Distal elements play a crucial role in complex diseases, such as schizophrenia. Recent studies have characterized relationships between promoters and distal elements interacting in hubs (Madsen et al., 2020, Song et al., 2020, Espinola et al., 2021). However, their precise implications in complex disease etiology remain unclear. In this study, we assessed the role of hubs linking promoters to distal elements

in a complex disease. Thus, we defined CRHs in neurons as complex networks of gene promoters and distal elements in physical proximity (**Figure 1A**) and demonstrated that they constitute a relevant model to study complex diseases. Indeed, CRHs aim to highlight direct and indirect contacts between promoters and distal elements which may not be targeted by other approaches. Our findings recapitulate the interest by integrating 3D contacts with tissue-specific regions to have a deeper understanding of regulatory processes involved in complex diseases. Gene's promoters and distal elements associated with schizophrenia are enriched in CRHs (**Figure 4A**), while these latter explain a larger portion of heritability than candidate CRHs (**Figure 4B**) or other models to characterize networks. Finally, by involving fewer distal elements, small CRHs are more involved in schizophrenia than larger or medium ones, suggesting that CRH structure is crucial in complex disease etiology. Therefore, our results establish that CRHs, by integrating interactions between distal elements and gene promoters, constitute a relevant 3D model to study complex diseases such as schizophrenia.

Previous studies suggest that hubs linking genes to enhancers are involved in the emergence of TADs (Espinola et al., 2021) or highly interconnected enhancers constitute sub-TADs strongly enriched in CTCF (Madsen et al., 2020). Recent studies have either investigated the role of chromatin loops (Rajarajan et al., 2018) or the impact of ultra-rare variants in TAD borders in the emergence of schizophrenia (Halvorsen et al., 2020). In our data, CRHs constitute a more local functional organization than higher-order chromatin features (A/B compartments, TADs) (**Figures 1F-1G**) and are enriched in FIREs. In fact, CRHs are strongly enriched in active regions (**Figure 2B**), defining CRHs as functional hubs with high transcriptional activity. Moreover, CRHs are strongly enriched in schizophrenia-associated genes, which are characterized by higher expression levels (**Figure 3B)**. These results are in line with those of Sey et al., 2020, as they have shown that schizophrenia-associated genes exhibit higher differential expression. However, genes within the same CRH, by sharing one or more distal elements, are functionally close (**Figure 2E**). Finally, we argue that due to complex interplay between several elements and direct implications in gene regulation, focusing on 3D organization such as CRHs may be prioritized in a context of complex phenotypes. Thus, CRHs constitute more functional 3D organizations, aiming to highlight the underlying regulatory processes.

Promoters and distal elements involved in CRHs exhibit different connectivity behaviors. Indeed, CRHs harbor more distal elements than genes (**Figure 1C**), suggesting that within a CRH, genes tend to have more connections compared to distal elements (**Figure 1D**) (Madsen et al., 2020, Espinola et al. 2021). Espinola et al., 2021 have shown that hubs connecting promoters to distal elements do not encompass several promoters, while Madsen et al., 2020 exhibit that enhancers are mostly involved in one-one connections. These results suggest that genes have fewer specific relationships, while enhancers strongly connected to promoters link genes with strong involvement in diseases (Madsen et al., 2020). However, in our data we found that CRH often harbor several genes connected mainly by distal elements, supporting that CRHs can be either promoter hubs, enhancer hubs or multi hubs (**Figure 1A**) (Campigli et al., 2020). Limitations of CRHs defined from Hi-C data are their dependence on Hi-C resolution and the measure of contacts from multiple cells in bulk, which may lead to spurious merging of CRHs with contacts occurring in distinct cell sub-populations. Future studies using single-cell chromosome conformation will be needed to assess the relevance of CRHs at more precise resolution (Nagano et al., 2013).

An important contribution of this study is that CRHs are a relevant model to study complex phenotypes such as schizophrenia. Indeed, the strong enrichments in schizophrenia-associated SNPs and schizophrenia heritability within CRHs (**Figures 4A, 4C**), compared to candidate CRHs, established that linking gene promoters to distal elements is relevant to study complex diseases. Also, we argue that CRHs, by providing a functional environment, offer new ways to study complex traits, by proposing new regions or targeting more precisely distal elements connected to genes known to be associated with a disease. Consistent with this idea, we found that including DNAse hypersensitive sites and H3K27ac-enriched regions to the definition of CRH explains a larger portion of schizophrenia heritability than networks built only from chromatin contacts. Moreover, CRHs by offering complex interplay aim to highlight indirect connections between promoters to distal elements. Thus, when the causal underlying regulatory processes are of interest, CRHs provide an informative model to study complex diseases. Indeed, by constituting a small set of functional regions, offer a larger advantage

than a pair of enhancer-promoter or larger domains, where complex epigenetic interplay can be difficult to capture. Collectively, these results point through the capability of CRHs to capture complex interplay between regulatory regions, aiming to fine-map the functional regions involved in complex diseases, since it remains one of the most important challenges in polygenic diseases.

Moreover, schizophrenia-associated genes show fewer connections than non-associated ones and are enriched in monogamous relationships (**Figure 3C-3D**). These results suggest that schizophrenia-associated genes are more strongly impacted than other active genes by disruptions of their distal elements since they are regulated by fewer connections to distal elements. Interestingly, we found that hubs encompassing a small number of genes highlight stronger schizophrenia heritability enrichments than medium or larger hubs (**Figure 4D**). We expect that small hubs or genes weakly connected to distal elements (monogamous, 1-1-N) will be more impacted by disruptions in their distal elements than large hubs or highly connected genes, supporting the model where weakly connected genes are more involved in disease etiology. We are proposing that small hubs or weakly connected genes should be prioritized when causal regulatory processes involved in schizophrenia and possibly other complex phenotypes are of interest. From this study and others, the emerging model is that a gene with limited connections to distal elements will be more impacted by polymorphisms, while highly connected genes will have stronger environmental or genetic resilience to disruptions in their distal elements (Tsai et al., 2019).

Based on these results, we argue that CRHs capture direct and indirect connections between promoters and distal elements, explaining the underlying regulatory processes involved in complex phenotypes. Future studies will demonstrate whether CRHs as a functional 3D model improve detection power of causal genes or gene pathways to elucidate the underlying causal regulatory processes involved in complex diseases. Indeed, since a large portion of schizophrenia heritability is explained by rare variants with high penetrance, future work will be needed to assess the relevance of CRHs in a rare variant context. CRHs can be integrated as functional annotation in association tests (He et al., 2017) or proposed as new regions to aggregate variants in pathway-based approaches (Wu & Pan, 2018).

Methods

Hi-C data and pre-processing

Hi-C data for neurons from induced pluripotent stem cells at 10Kb resolution were obtained from PsychENCODE Synapse platform (.hic format, intra-chromosomal). In the present study, we refer to these data as the neuron Hi-C dataset. Except for Score-FIRE calculation and ABC score, we applied KR-normalization with the Juicer toolbox (Durand et al., 2016) to obtain either a sparse or dense matrix.

Cis-Regulatory hubs

Cis-Regulatory hubs (CRHs) were built based on the ABC model (Fulco et al., 2019) to capture active regulatory processes between distal elements and gene promoters. To validate analyses shown in the paper, two other methods to build CRHs were also proposed (**See Supplementary Methods**).

ABC-Score: The Activity-by-Contact model (Fulco et al., 2019) defines active enhancers based on a quantitative score of DNAse (ENCSR278FVO), H3K27ac (ENCSR331CCW), and normalized Hi-C contact number. This score is computed relative to a background activity over a 5Mb window around a candidate element. Here candidate element refers to DNAse peaks on

which enhancers are defined (Fulco et al., 2019). Then, we set the threshold to 0.012; beyond which a candidate element is considered as a distal element. This value was selected to ensure that the mean number of distal enhancers per promoter is between 2 and 5 in the neuron Hi-C dataset (Fulco et al., 2019).

As an extension of the ABC-Score, CRHs were defined as bipartite networks (igraph R package; Csardi & Nepusz, 2005) between promoters and distal elements. Due to the nature of the methodology of the ABC-Score, contacts between distal elements and promoters were restricted. In proposing CRHs based on the ABC-Score, active regulatory phenomena occurring in our tissue were captured.

CRHs are conceived to capture regulatory phenomena based on Hi-C. For the purpose of enrichment analysis for different external validation sources, SNPs or disease heritability, equivalent sets with elements having the same characteristics but in no 3D contacts with promoters were proposed. These elements were referred to as candidate CRHs. Thus, the same approach as Nasser and colleagues was applied, where candidate distal elements are all DNAse peaks which do not overlap ABC distal elements. Also, candidate promoters were all promoters for known hg19 genes not included in CRHs.

3D Features

A/B Compartments: A/B compartments were defined at a 500 kb resolution of the contact matrix (using a 100 kb resolution had little impact on the results). The first principal component (PC) of a suitably normalized Hi-C contact matrix over a chromosome arm captures the plaid pattern of A/B compartments (Lieberman-Aiden et al., 2009). As GC content is higher in A compartments than in B compartments, the correlation of the first PC with GC content was used to orient the first PC so that positive values correspond to the A compartments and negative values to B compartments (Imakaev et al., 2012). The transformation applied to the ratio observed/expected (O/E) contact matrices was selected that 1) maximized the number of autosomal chromosome arms where the first PC had the highest correlation with GC content over the first three PCs and 2) had the highest correlation of the first PC with GC content in these chromosome arms. The transformation was selected among the following three: O/E - 1 with clipping of values below percentile 1 and above percentile 99, log (O/E) and log (O/E) with clipping of values below percentile 1 and above percentile 99. The last transformation was selected based on the above criteria, with 30/40 autosomal chromosome arms where the first PC correlated the most strongly with GC content and correlations between 0.38 and 0.87 within these chromosome arms. The A/B compartment for the remaining 10 chromosome arms (6q, 8q, 9p, 10q, 12p, 18p, 18q, 19q, 20p and 21p) were set to missing.

TADs calling: TADs were called using the directionality index (Dixon et al., 2012), insulation score (Crane et al., 2015) or with Arrowhead algorithm from Juicer software (Rao et al., 2014; Durand et al., 2016)
Directionality Index (DI) was computed as presented by Dixon et al., 2012 at 10 Kb resolution. Briefly, for each 10 Kb bins the number of upstream and downstream contacts were calculated. A bias toward upstream regions at the end of a TAD was expected and conversely, a bias toward downstream regions, at the beginning of a TAD was expected. As mentioned by Gorkin et al., 2019, the original approach to computing the DI using a 200 Kb window size was applied to capture more local features. DI values for each 10Kb bins were used to build a Hidden Markov Model and predict upstream bias, downstream bias, and no bias states, respectively. Regions switching from upstream bias to downstream bias were called topological boundaries.
Insulation Score (INS) was computed as presented by Crane et al., 2015. Simply, for each 10 Kb bin, the average number of contacts in 400Kb windows upstream and downstream on O/E matrices was computed. A local minimum for INS at TADs borders was expected. INS was normalized at the chromosome level to take account of differences between chromosomes. Then INS was scaled between 0 and 1, where 0 is complete insulation and 1 is no insulation respectively.

Arrowhead TADs were annotated using Arrowhead (Rao et al., 2014; Durand et al., 2016) at 10 Kb resolution.

Frequently Interacting Regions (FIREs) (Schmitt et al., 2016) was computed with FIREcaller R package (Crowley et al., 2021) at 10Kb resolution, with minor adjustments to fit our data format.

Summary Statistics for Schizophrenia

The original SCZ3 GWAS summary statistics (Ripke et al., 2020) used in SNP and heritability enrichments were downloaded from the PGC site https://www.med.unc.edu/ pgc/results-and-downloads.

Schizophrenia-associated genes

In order to assess schizophrenia-associated genes, H-Magma (Sey et al., 2020) was used on iPSC-derived Hi-C neurons (Rajarajan et al., 2018) with schizophrenia SNP summary statistics (Ripke et al., 2020) to link noncoding SNPs to their target gene. To determine significant schizophrenia-associated genes, all genes with a p-value lower or equal to 0.05 after Benjamini and Hochberg correction were selected.

Functional Enrichment analysis

Genes inside CRHs were used for GO enrichment analysis at the CRH or gene level. To do so, at the CRH level the clusterProfiler R package (Wang et al., 2012) was used with the compareCluster function to perform over-representation tests.

Partitioning Heritability for Schizophrenia

The LDSC regression (Finucane et al., 2015) was used to partition SNP heritability for schizophrenia integrating CRHs. For the main analysis, a modified version of the LDSC baseline model was used, only considering annotations corresponding to non-coding regions consistent with elements included within CRHs, promoters, H3K27ac histone marks, DNase I hypersensitive sites, ChromHMM/Segway predictions, super-enhancers, and FANTOM5 enhancers that are not specific to neurons. By proceeding this way, the relevance of CRHs in the neuron Hi-C dataset regarding SNP heritability enrichment compared to candidate CRHs or equivalent sets of genomic features widely used for this purpose. CRHs and candidate CRHs were extended by 500bp upstream and downstream to take into account the background activity and not inflated the enrichment signal, as suggested by Finucane et al., 2015. All heritability enrichment results at the individual level and CRH level for the complete baseline model and his modified version are given for the ABC and control methods in Supplementary Tables (**Supplementary Tables 2-3**).

Peak calling

Peak calling for ChIP-Seq data was performed with MACS2 (Zhang et al., 2008) software through the following command:

```
macs2 callpeak \
-t bamfile \
-n alias \
-f BAM \
-g hs \
-p .1 \
--call-summits \
--outdir outputdirectory
```

Genome Build

All coordinates in the human genome are reported using build hg19.

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

Datasets analyzed in this study are publicly available from: PsychENCODE Knowledge Portal (https://www.synapse.org, syn13363580) for Hi-C, PGC3 (https://www.med.unc.edu/ pgc/results-and-downloads) for summary statistics, SCREEN (https://screen.encodeproject.org) for Encode candidate regulatory elements (CREs) in neural progenitor cell originated from H9, Roadmap Epigenomics data Portal (https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/core_K27ac/jointModel/final) for 18-states model in E007 , ENCODE data portal (ENCSR539JGB) and GEO (GSE142670) for reference epigenome and RNA-Seq in neurons.

Various analyzes show in this paper, as well additional documentation and CRHs in neurons are available at: https://github.com/lmangnier/Hi-C_analysis.

Competing interests

None declared.

Funding

This work was partly funded by the Canadian Statistical Sciences Institute through a Collaborative Research Team grant and by a National Science and Engineering Research Council of Canada Discovery grant to AB. Some data analyses were performed on computing resources from Compute Canada.

Authors' contributions

Study was conceived and overseen by AB and LM. Data analysis was performed by LM and AB with support of CJB and SB. The manuscript was written by LM, AB, and SB with input and comments from all authors. All authors read and approved the final manuscript.

Acknowledgements

We would like to acknowledge Antoine Bodein, Julien Prunier and Christophe Khun-Narom Tav of the Arnaud Droit Lab for important discussions and feedback during the preparation of this manuscript. We also thank Katie Pollard, Geoff Fudenberg and Luis Chumpitaz-Diaz from the Pollard Lab for their insightful advice. Data were generated as part of the PsychENCODE Consortium, supported by: U01MH103392, U01MH103365, U01MH103346, U01MH103340, U01MH103339, R21MH109956, R21MH105881, R21MH105853, R21MH103877, R21MH102791, R01MH111721, R01MH110928, R01MH110927, R01MH110926, R01MH110921, R01MH110920, R01MH110905, R01MH109715, R01MH109677, R01MH105898, R01MH105898, R01MH094714, P50MH106934, U01MH116488, U01MH116487, U01MH116492, U01MH116489, U01MH116438, U01MH116441, U01MH116442, R01MH114911, R01MH114899, R01MH114901, R01MH117293, R01MH117291, R01MH117292 awarded to: Schahram Akbarian (Icahn School of Medicine at Mount Sinai), Gregory Crawford (Duke University), Stella Dracheva (Icahn School of Medicine at Mount Sinai), Peggy Farnham (University of Southern California), Mark Gerstein (Yale University), Daniel Geschwind (University of California, Los Angeles), Fernando Goes (Johns Hopkins University), Thomas M. Hyde (Lieber Institute for Brain Development), Andrew Jaffe (Lieber Institute for Brain Development), James A. Knowles (University of Southern California), Chunyu Liu (SUNY Upstate Medical University), Dalila Pinto (Icahn School of Medicine at Mount Sinai), Panos Roussos (Icahn School of Medicine at Mount Sinai), Stephan Sanders (University of California, San Francisco), Nenad Sestan (Yale University), Pamela Sklar (Icahn School of Medicine at Mount Sinai), Matthew State (University of California, San Francisco), Patrick Sullivan (University of North Carolina), Flora Vaccarino (Yale University), Daniel Weinberger (Lieber Institute for Brain Development), Sherman Weissman (Yale University), Kevin White (University of Chicago), Jeremy Willsey (University of California, San Francisco), and Peter Zandi (Johns Hopkins University).

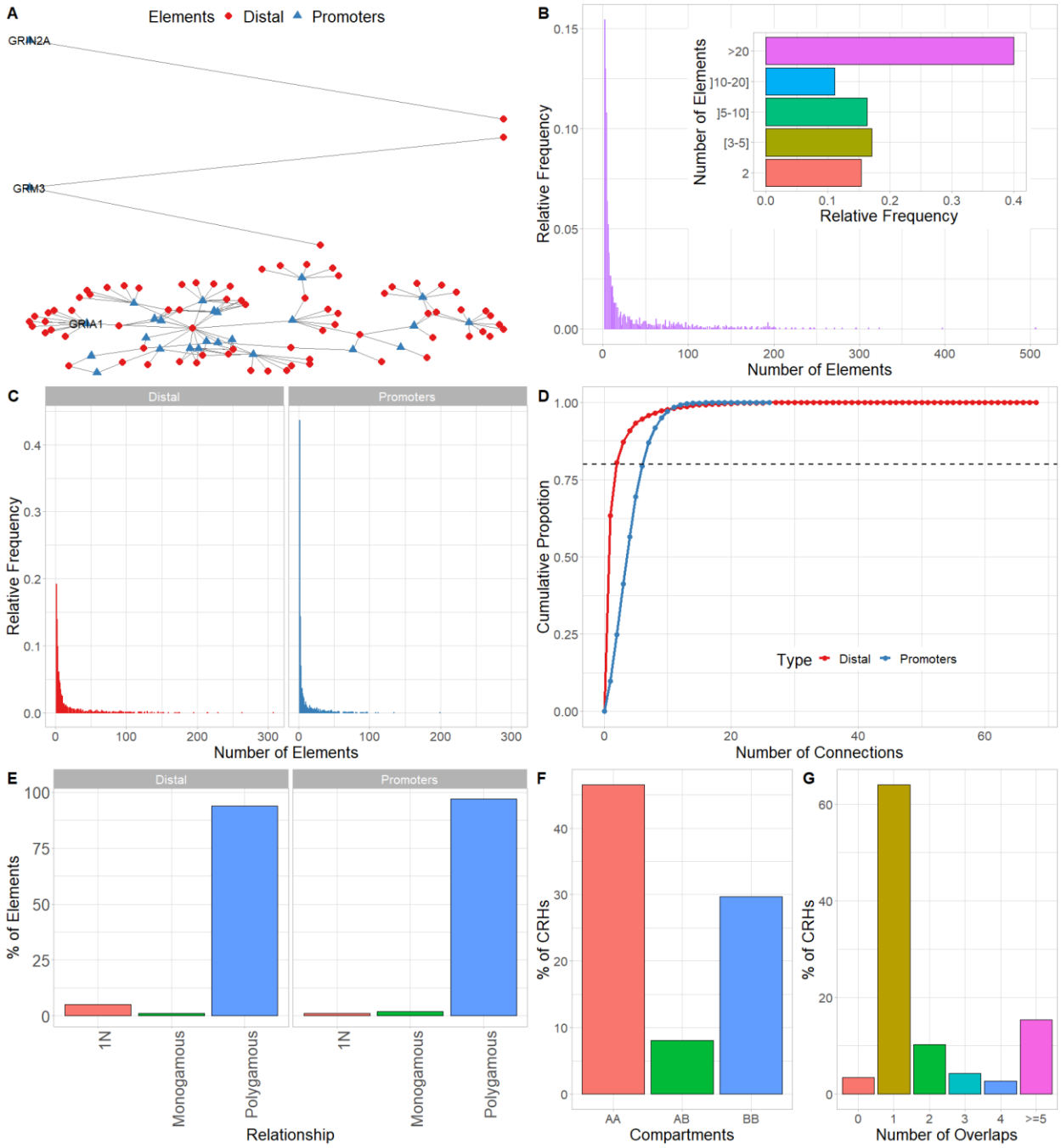

Figure 1 CRH structure, complexity and link to 3D features. **A** Cis-regulatory hub (CRH) connecting promoters (blue) to distal elements (red) for *GRIN2A* (**Top**), *GRM3* (**Middle**), and *GRIA1* (**Bottom**) genes, associated with relevant biological processes in schizophrenia (glutamatergic transmission or synaptic plasticity pathways). The genes are within monogamous, 1-1-N and polygamous relationships, respectively. **B** Distribution of the number of elements (promoters and distal elements) within CRHs. The subpanel shows the number of CRH elements by aggregated categories . **C** Distribution of the number of promoters (blue) and distal elements (red). **D** Cumulative distribution function of the number of connections for promoters (blue) and distal elements (red). The dotted line shows the number of connections where 80% are less or equal to this value. **E** Distribution of the kind of relationship for distal elements (**Left**) and promoters (**Right**). **F** Overlapping of CRHs per compartment type (AA: Active-Active; AB: Active-Inactive; BB: Inactive-Inactive). When CRHs overlap several compartments, we restrict

our attention to farthest elements. The CRHs in genomic regions not assigned to compartments (16%) were omitted from the analysis. **G** Distribution of the number of TADs overlapped by CRHs, when TADs are detected with the directionality index.

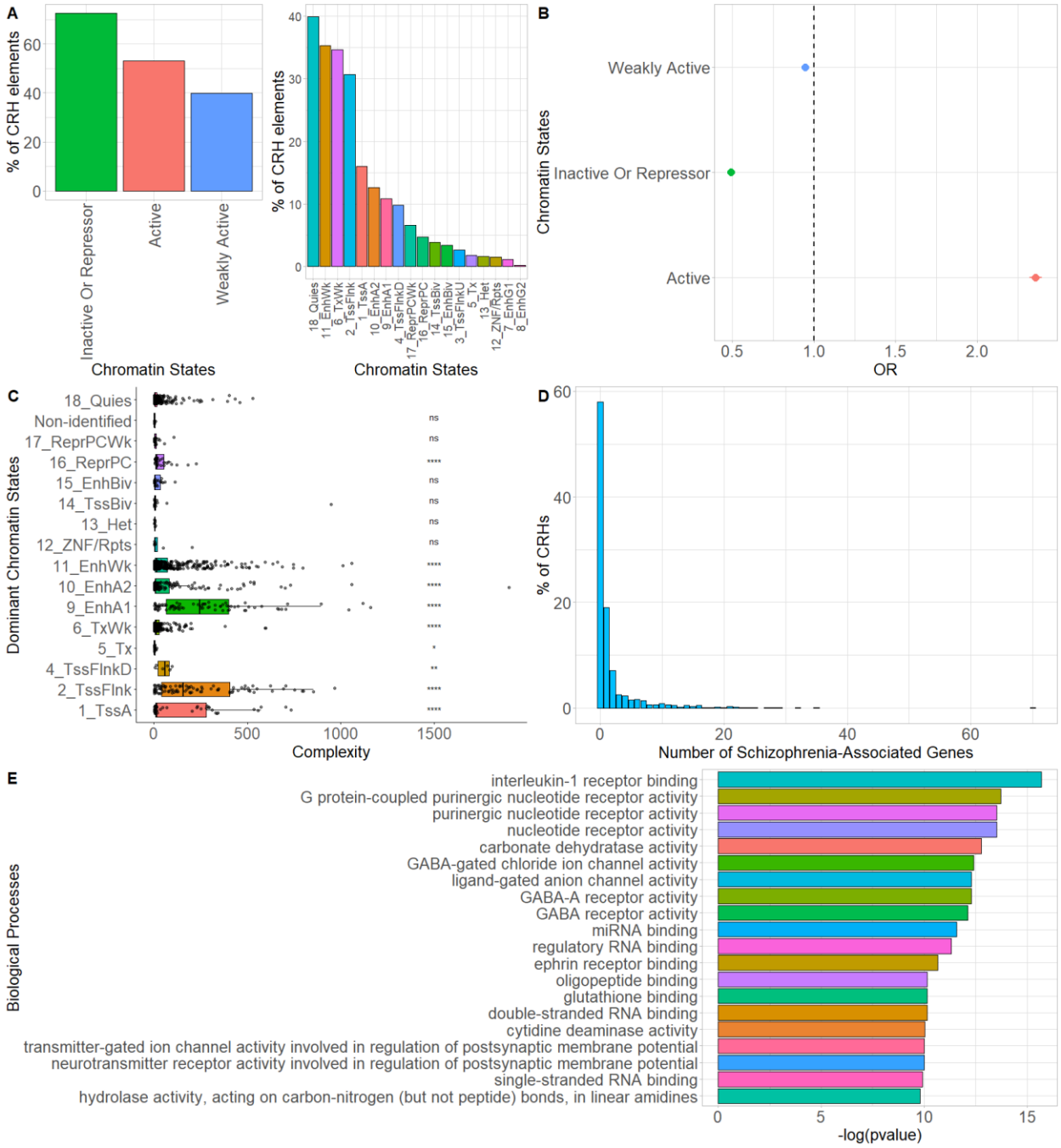

Figure 2 CRHs as active functional environment. **A (Left)** Proportion of CRH elements (promoters and distal elements) overlapping chromatin states grouped by activity **(Right)** Proportion of CRH elements (promoters and distal elements) overlapping individual chromatin states. **B** Enrichments in chromatin states as measured with odds ratio (OR) and their 95% confidence intervals for CRHs compared to candidate CRHs in chromatin states. The dotted line represents the null value. **C** Boxplot of complexity by most present chromatin state within the CRH. Stars represent significance at 0.05 level with right-tailed Wilcoxon rank-sum test compared to 18_Quies state. **D** Distribution of the number of schizophrenia-associated genes within CRHs. **E** Go enrichment for CRHs. We consider the top-20 biological processes most enriched.

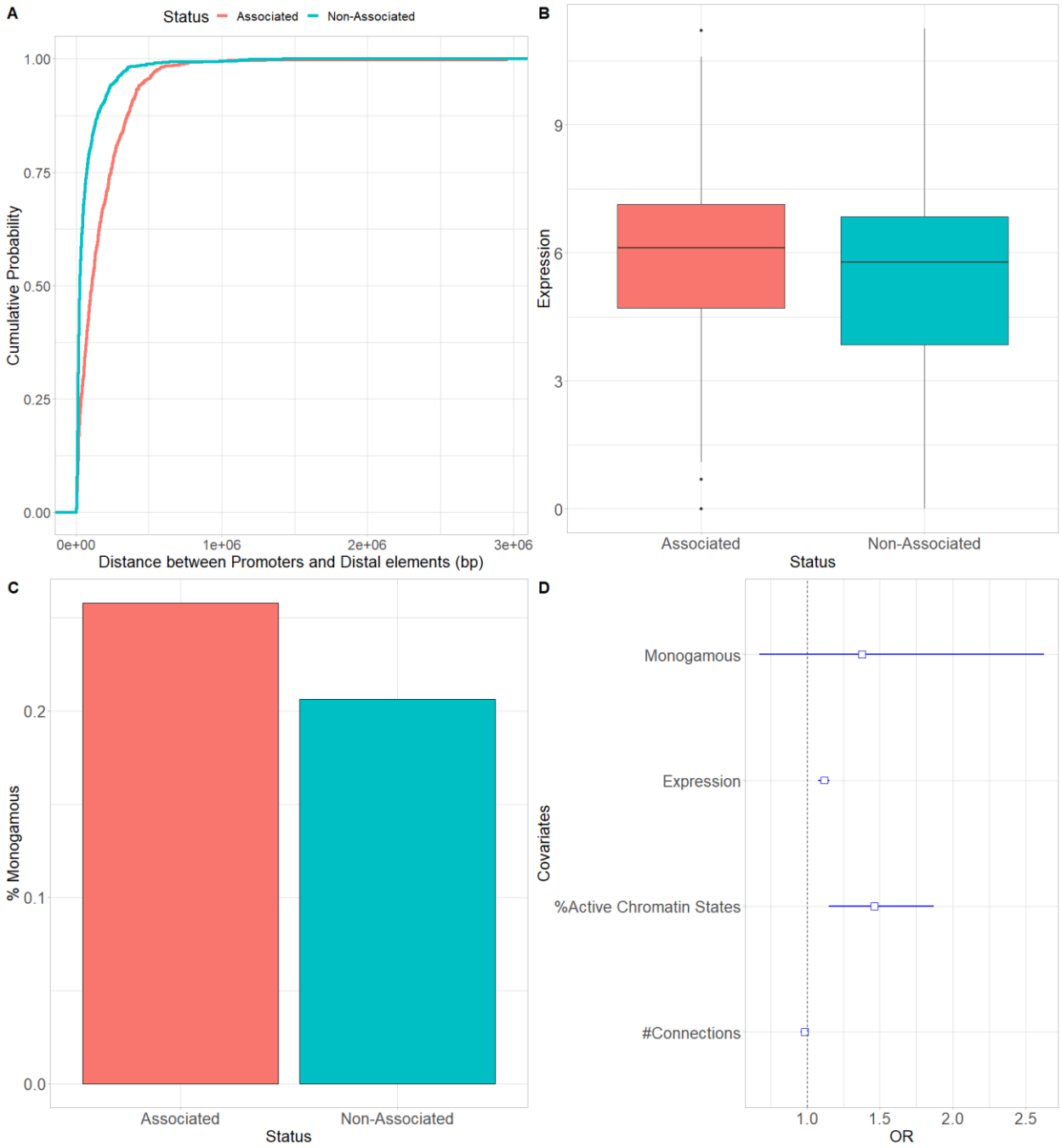

Figure 3 Features of schizophrenia-associated genes. **A** Cumulative distribution of mean distance between elements for CRHs encompassing schizophrenia-associated genes and CRHs not harboring ones. **B** Boxplot of expression for schizophrenia-associated genes and non associated ones. **C** Percentage of Monogamous genes which are associated with schizophrenia or non associated. **D** Odds ratio (OR) and their 95% confidence interval for a logistic regression of the status of genes regarding association with schizophrenia. The dotted line represents the null value.

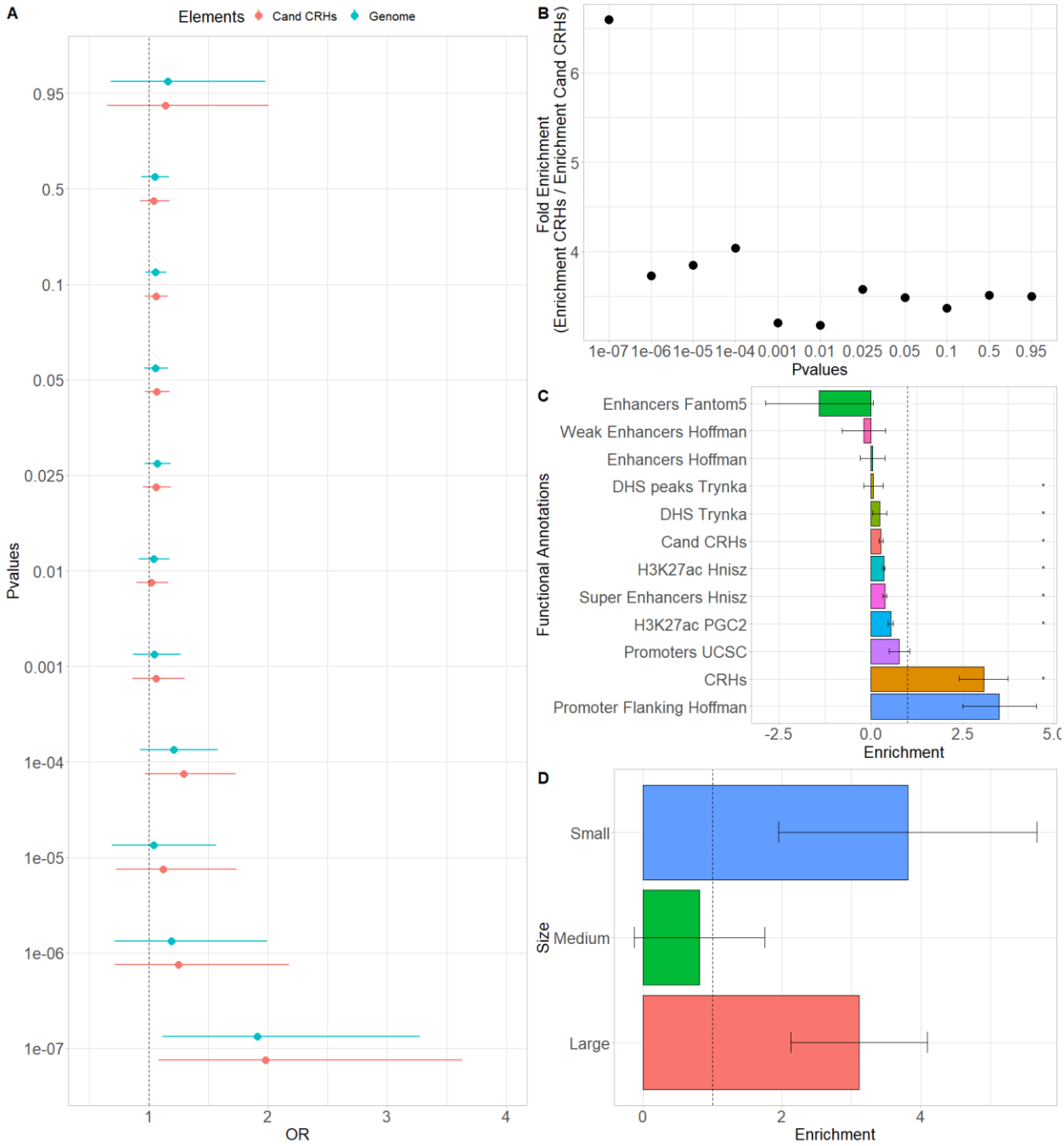

Figure 4 Schizophrenia-associated SNPs and heritability enrichments in CRHs. **A** Enrichment analysis measured through odds ratio (OR) and their 95% confidence intervals for CRHs compared to candidate CRHs and the rest of the genome for different p-values . The dotted line represents the null value. **B** Fold enrichment of CRHs compared to candidate CRHs for different p-values. The dotted line represents the null value. **C** Schizophrenia heritability enrichment (measured with LD-Score regression) with error bars for CRHs, candidate CRHs and non tissue equivalent elements. The dotted line represents the null value, while the star shows statistical significance at 0.05 level after Bonferroni correction. **D** Schizophrenia heritability enrichment (measured with LD-Score regression) with error bars for CRHs, considering the number of genes within CRHs. The dotted line represents the null value.

References

Smeland, O. B., Frei, O., Dale, A. M., & Andreassen, O. A. (n.d.). The polygenic architecture of. *Nature Reviews Neurology*, *Box 1*. https://doi.org/10.1038/s41582-020-0364-0

Do, C., Shearer, A., Suzuki, M., Terry, M. B., Gelernter, J., Greally, J. M., & Tycko, B. (2017). *Genetic – epigenetic interactions in cis: a major focus in the post-GWAS era*. 1–22. https://doi.org/10.1186/s13059-017-1250-y

Vliet, J. Van, Oates, N. A., & Whitelaw, E. (2007). *Review Epigenetic mechanisms in the context of complex diseases*. *64*, 1531–1538. https://doi.org/10.1007/s00018-007-6526-z

Working, S., & Consortium, G. (2014). *Biological insights from 108 schizophrenia-associated genetic loci*. https://doi.org/10.1038/nature13595

Fukaya, T., Lim, B., Levine, M., Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting Article Enhancer Control of Transcriptional Bursting. *Cell*, *166*(2), 358–368. https://doi.org/10.1016/j.cell.2016.05.025

Rubin, A. J., Barajas, B. C., Furlan-magaril, M., Lopez-pajares, V., Mumbach, M. R., Howard, I., Kim, D. S., Boxer, L. D., Cairns, J., Spivakov, M., Wingett, S. W., Shi, M., Zhao, Z., Greenleaf, W. J., Kundaje, A., Snyder, M., Chang, H. Y., Fraser, P., & Khavari, P. A. (2017). Lineage-specific dynamic and pre-established enhancer – promoter contacts cooperate in terminal differentiation. *Nature Publishing Group*, *49*(10). https://doi.org/10.1038/ng.3935

Melo, S., Scho, R., Holtgrewe, M., Klever, M., Tu, S., Heinrich, V., Pluym, I. D., Matoso, E., Sousa, B. De, Louro, P., Hu, W., Cohen, M., Dufke, A., Latos-biele, A., Vingron, M., Kalscheuer, V., Quintero-rivera, F., Spielmann, M., & Mundlos, S. (2020). *Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases*. 872–884. https://doi.org/10.1016/j.ajhg.2020.04.016

Hauberg, M. E., Creus-muncunill, J., Bendl, J., Kozlenkov, A., Zeng, B., Corwin, C., Chowdhury, S., Kranz, H., Hurd, Y. L., Wegner, M., Børglum, A. D., Dracheva, S., Ehrlich, M. E., & Fullard, J. F. (2020). Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nature Communications*. https://doi.org/10.1038/s41467-020-19319-2

Dekker, J., & Mirny, L. (2016). The 3D Genome as Moderator of Chromosomal Communication. *Cell*, *164*(6), 1110–1121. https://doi.org/10.1016/j.cell.2016.02.007

Zhang, Y., Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., Nussbaum, C., Myers, R. M., Brown, M., Li, W., & Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*, *9*(9). https://doi.org/10.1186/gb-2008-9-9-r137

Nagano, T., Lubling, Y., Stevens, T. J., Schoenfelder, S., Yaffe, E., Dean, W., Laue, E. D., Tanay, A., & Fraser, P. (2013). Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, *502*(7469), 59–64. https://doi.org/10.1038/nature12593

Oudelaar, A. M., Harrold, C. L., Hanssen, L. L. P., Telenius, J. M., Higgs, D. R., & Hughes, J. R. (2019). A revised model for promoter competition based on multi-way chromatin interactions at the α-globin locus. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-13404-x

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S. P., Huntley, M. H., Lander, E. S., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell Systems*, *3*(1), 95–98. https://doi.org/10.1016/j.cels.2016.07.002

Tsai, A., Alves, M. R. P., & Crocker, J. (2019). Multi-enhancer transcriptional hubs confer phenotypic robustness. *ELife*, *8*, 1–17. https://doi.org/10.7554/eLife.45325

Rao, S. S. P., Huntley, M. H., Durand, N. C., Stamenova, E. K., Bochkov, I. D., Robinson, J. T., Sanborn, A. L., Machol, I., Omer, A. D., Lander, E. S., & Aiden, E. L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell*, *159*(7), 1665–1680. https://doi.org/10.1016/j.cell.2014.11.021

Huo, Y., Li, S., Liu, J., Li, X., & Luo, X. J. (2019). Functional genomics reveal gene regulatory mechanisms underlying schizophrenia risk. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-019-08666-4

Fullard, J. F., Hauberg, M. E., Bendl, J., Egervari, G., Cirnaru, M. D., Reach, S. M., Motl, J., Ehrlich, M. E., Hurd, Y. L., & Roussos, P. (2018). An atlas of chromatin accessibility in the adult human brain. *Genome Research*, *28*(8), 1243–1252. https://doi.org/10.1101/gr.232488.117

Benabdallah, N. S., Williamson, I., Illingworth, R. S., Kane, L., Boyle, S., Sengupta, D., Grimes, G. R., Therizols, P., & Bickmore, W. A. (2019). Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Molecular Cell*, *76*(3), 473-484.e7. https://doi.org/10.1016/j.molcel.2019.07.038

Zhang, F., & Lupski, J. R. (2015). Non-coding genetic variants in human disease. *Human Molecular Genetics*, *24*(R1), R102–R110. https://doi.org/10.1093/hmg/ddv259

Bouwman, B. A. M., & de Laat, W. (2015). Getting the genome in shape: The formation of loops, domains and compartments. *Genome Biology*, *16*(1), 1–9. https://doi.org/10.1186/s13059-015-0730-1

Girdhar, K., Hoffman, G. E., Jiang, Y., Brown, L., Kundakovic, M., Hauberg, M. E., Francoeur, N. J., Wang, Y. chih, Shah, H., Kavanagh, D. H., Zharovsky, E., Jacobov, R., Wiseman, J. R., Park, R., Johnson, J. S., Kassim, B. S., Sloofman, L., Mattei, E., Weng, Z., … Akbarian, S. (2018). Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome. *Nature Neuroscience*, *21*(8), 1126–1136. https://doi.org/10.1038/s41593-018-0187-0

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*, *16*(5), 284–287. https://doi.org/10.1089/omi.2011.0118

Gorkin, D. U., Leung, D., & Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. *Cell Stem Cell*, *14*(6), 762–775. https://doi.org/10.1016/j.stem.2014.05.017

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M. J., Amin, V., Whitaker, J. W., Schultz, M. D., Ward, L. D., Sarkar, A., Quon, G., Sandstrom, R. S., Eaton, M. L., … Kellis, M. (2015). Integrative analysis of 111 reference human epigenomes. *Nature*, *518*(7539), 317–330. https://doi.org/10.1038/nature14248

Rajarajan, P., Borrman, T., Liao, W., Schrode, N., Flaherty, E., Casiño, C., Powell, S., Yashaswini, C., LaMarca, E. A., Kassim, B., Javidfar, B., Espeso-Gil, S., Li, A., Won, H., Geschwind, D. H., Ho, S.-M., MacDonald, M., Hoffman, G. E., Roussos, P., … Akbarian, S. (2018). Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science*, *362*(6420), eaat4311. https://doi.org/10.1126/science.aat4311

Halvorsen, M., Huh, R., Oskolkov, N., Wen, J., Netotea, S., Giusti-Rodriguez, P., Karlsson, R., Bryois, J., Nystedt, B., Ameur, A., Kähler, A. K., Ancalade, N., Farrell, M., Crowley, J. J., Li, Y., Magnusson, P. K. E., Gyllensten, U., Hultman, C. M., Sullivan, P. F., & Szatkiewicz, J. P. (2020). Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. *Nature Communications*, *11*(1), 1842. https://doi.org/10.1038/s41467-020-15707-w

Imakaev, M., Fudenberg, G., McCord, R. P., Naumova, N., Goloborodko, A., Lajoie, B. R., Dekker, J., & Mirny, L. A. (2012). Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature Methods*, *9*(10), 999–1003. https://doi.org/10.1038/nmeth.2148

Crowley, C., Yang, Y., Qiu, Y., Hu, B., Abnousi, A., Lipiński, J., Plewczyński, D., Wu, D., Won, H., Ren, B., Hu, M., & Li, Y. (2021). FIREcaller: Detecting frequently interacting regions from Hi-C data. *Computational and Structural Biotechnology Journal*, *19*, 355–362. https://doi.org/10.1016/j.csbj.2020.12.026

Crane, E., Bian, Q., McCord, R. P., Lajoie, B. R., Wheeler, B. S., Ralston, E. J., Uzawa, S., Dekker, J., & Meyer, B. J. (2015). Condensin-driven remodelling of X chromosome topology during dosage compensation. *Nature*, *523*(7559), 240–244. https://doi.org/10.1038/nature14450

Bryois, J., Garrett, M. E., Song, L., Safi, A., Giusti-Rodriguez, P., Johnson, G. D., Shieh, A. W., Buil, A., Fullard, J. F., Roussos, P., Sklar, P., Akbarian, S., Haroutunian, V., Stockmeier, C. A., Wray, G. A., White, K. P., Liu, C., Reddy, T. E., Ashley-Koch, A., … Crawford, G. E. (2018). Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nature Communications*, *9*(1), 3121. https://doi.org/10.1038/s41467-018-05379-y

Nasser, J., Bergman, D. T., Fulco, C. P., Guckelberger, P., Doughty, B. R., Patwardhan, T. A., Jones, T. R., Nguyen, T. H., Ulirsch, J. C., Lekschas, F., Mualim, K., Natri, H. M., Weeks, E. M., Munson, G., Kane, M., Kang, H. Y., Cui, A., Ray, J. P., Eisenhaure, T. M., … Engreitz, J. M. (2021). Genome-wide enhancer maps link risk variants to disease genes. *Nature*. https://doi.org/10.1038/s41586-021-03446-x

Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P. R., Anttila, V., Xu, H., Zang, C., Farh, K., Ripke, S., Day, F. R., Purcell, S., Stahl, E., Lindstrom, S., Perry, J. R. B., Okada, Y., Raychaudhuri, S., Daly, M. J., ... Price, A. L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics*, *47*(11), 1228–1235. https://doi.org/10.1038/ng.3404

Matthew T. Maurano, Richard Humbert, Eric Rynes1, Robert E. Thurman, E., Haugen, Hao Wang, Alex P. Reynolds, Richard Sandstrom, Hongzhu Qu, J., Brody, Anthony Shafer, Fidencio Neri, Kristen Lee, Tanya Kutyavin, S. S.-, Sun1, Audra K. Johnson1, Theresa K. Canfield1, Erika Giste1, Morgan Diegel1, D., Bates1, R. Scott Hansen4, Shane Neph1, Peter J. Sabo1, Shelly Heimfeld5, A., Raubitschek6, Steven Ziegler6, Chris Cotsapas7, 8, Nona Sotoodehnia3, 9, I. G., & Shamil R. Sunyaev11, Rajinder Kaul4, and J. A. S. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science*, *337*(6099), 1190–1195. https://doi.org/10.1126/science.1222794.Systematic

Roussos, P., Mitchell, A. C., Voloudakis, G., Fullard, J. F., Pothula, V. M., Tsang, J., Stahl, E. A., Georgakopoulos, A., Ruderfer, D. M., Charney, A., Okada, Y., Siminovitch, K. A., Worthington, J., Padyukov, L., Klareskog, L., Gregersen, P. K., Plenge, R. M., Raychaudhuri, S., Fromer, M., ... Sklar, P. (2014). A Role for Noncoding Variation in Schizophrenia. *Cell Reports*, *9*(4), 1417–1429. https://doi.org/10.1016/j.celrep.2014.10.015

Abascal, F., Acosta, R., Addleman, N. J., Adrian, J., Afzal, V., Aken, B., Akiyama, J. A., Jammal, O. Al, Amrhein, H., Anderson, S. M., Andrews, G. R., Antoshechkin, I., Ardlie, K. G., Armstrong, J., Astley, M., Banerjee, B., Barkal, A. A., Barnes, I. H. A., Barozzi, I., ... Zimmerman, J. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, *583*(7818), 699–710. https://doi.org/10.1038/s41586-020-2493-4

Wu, C., & Pan, W. (2018). Integration of enhancer-promoter interactions with GWAS summary results identifies novel schizophrenia-associated genes and pathways. *Genetics*, *209*(3), 699–709. https://doi.org/10.1534/genetics.118.300805

Lieberman-aiden, E., Berkum, N. L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., & Mirny, L. A. (2009). *of the Human Genome*. *33292*(October), 289–294.

Schmitt, A. D., Hu, M., Jung, I., Xu, Z., Qiu, Y., Tan, C. L., Li, Y., Lin, S., Lin, Y., Barr, C. L., & Ren, B. (2016). A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports*, *17*(8), 2042–2059. https://doi.org/10.1016/j.celrep.2016.10.061

Ripke, S., O'Dushlaine, C., Chambert, K., Moran, J. L., Kähler, A. K., Akterin, S., Bergen, S. E., Collins, A. L., Crowley, J. J., Fromer, M., Kim, Y., Lee, S. H., Magnusson, P. K. E., Sanchez, N., Stahl, E. A., Williams, S., Wray, N. R., Xia, K., Bettella, F., ... Sullivan, P. F. (2013). Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genetics*, *45*(10), 1150–1159. https://doi.org/10.1038/ng.2742

van Berkum, N. L., Lieberman-Aiden, E., Williams, L., Imakaev, M., Gnirke, A., Mirny, L. A., Dekker, J., & Lander, E. S. (2010). Hi-C: A method to study the three-dimensional architecture of genomes. *Journal of Visualized Experiments*, *39*, 1–7. https://doi.org/10.3791/1869

Dixon, J. R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J. S., & Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, *485*(7398), 376–380. https://doi.org/10.1038/nature11082

He, Z., Xu, B., Lee, S., & Ionita-Laza, I. (2017). Unified Sequence-Based Association Tests Allowing for Multiple Functional Annotations and Meta-analysis of Noncoding Variation in Metabochip Data. *American Journal of Human Genetics*, *101*(3), 340–352. https://doi.org/10.1016/j.ajhg.2017.07.011

Espinola, S. M., Götz, M., Bellec, M., Messina, O., Fiche, J. B., Houbron, C., Dejean, M., Reim, I., Cardozo Gizzi, A. M., Lagha, M., & Nollmann, M. (2021). Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during early Drosophila development. *Nature Genetics*, *53*(4), 477–486. https://doi.org/10.1038/s41588-021-00816-z

Campigli, D., Giammartino, D., Polyzos, A., & Apostolou, E. (2020). Transcription factors : building hubs in the 3D space ABSTRACT. *Cell Cycle*, *19*(19), 2395–2410. https://doi.org/10.1080/15384101.2020.1805238

Lieberman-aiden, E., Berkum, N. L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B. R., Sabo, P. J., Dorschner, M. O., Sandstrom, R., Bernstein, B., Bender, M. A., Groudine, M., Gnirke, A., Stamatoyannopoulos, J., & Mirny, L. A. (2009). *of the Human Genome*. *33292*(October), 289–294.

Gorkin, D. U., Qiu, Y., Hu, M., Fletez-Brant, K., Liu, T., Schmitt, A. D., Noor, A., Chiou, J., Gaulton, K. J., Sebat, J., Li, Y., Hansen, K. D., & Ren, B. (2019). Common DNA sequence variation influences 3-dimensional conformation of the human genome. *Genome Biology*, *20*(1), 1–25. https://doi.org/10.1186/s13059-019-1855-4

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., Grossman, S. R., Anyoha, R., Doughty, B. R., Patwardhan, T. A., Nguyen, T. H., Kane, M., Perez, E. M., Durand, N. C., Lareau, C. A., Stamenova, E. K., Aiden, E. L., Lander, E. S., & Engreitz, J. M. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics*, *51*(12), 1664–1669. https://doi.org/10.1038/s41588-019-0538-0

Sey, N. Y. A., Hu, B., Mah, W., Fauni, H., McAfee, J. C., Rajarajan, P., Brennand, K. J., Akbarian, S., & Won, H. (2020). A computational tool (H-MAGMA) for improved prediction of brain-disorder risk genes by incorporating brain chromatin interaction profiles. *Nature Neuroscience*, *23*(4), 583–593. https://doi.org/10.1038/s41593-020-0603-0

Wu, C., & Pan, W. (2018). Integration of enhancer-promoter interactions with GWAS summary results identifies novel schizophrenia-associated genes and pathways. *Genetics*, *209*(3), 699–709. https://doi.org/10.1534/genetics.118.300805

Song, M., Pebworth, M. P., Yang, X., Abnousi, A., Fan, C., Wen, J., Rosen, J. D., Choudhary, M. N. K., Cui, X., Jones, I. R., Bergenholtz, S., Eze, U. C., Juric, I., Li, B., Maliskova, L., Lee, J., Liu, W., Pollen, A. A., Li, Y., … Shen, Y. (2020). Cell-type-specific 3D epigenomes in the developing human cortex. *Nature*, *587*(7835), 644–649. https://doi.org/10.1038/s41586-020-2825-4

Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J. B., Espinola, S. M., Gurgo, J., Messina, O., Houbron, C., Ogiyama, Y., Papadopoulos, G. L., Cavalli, G., Lagha, M., & Nollmann, M. (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell*, *74*(1), 212-222.e5. https://doi.org/10.1016/j.molcel.2019.01.011

Madsen, J. G. S., Madsen, M. S., Rauch, A., Traynor, S., Van Hauwaert, E. L., Haakonsson, A. K., Javierre, B. M., Hyldahl, M., Fraser, P., & Mandrup, S. (2020). Highly interconnected enhancer communities control lineage-determining genes in human mesenchymal stem cells. *Nature Genetics*, *52*(11), 1227–1238. https://doi.org/10.1038/s41588-020-0709-z

Fudenberg, G., & Pollard, K. S. (2019). Chromatin features constrain structural variation across evolutionary timescales. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(6), 2175–2180. https://doi.org/10.1073/pnas.1808631116

Ripke, S., Walters, J. T. R., & O'Donovan, M. C. (2020). Mapping genomic loci prioritises genes and implicates synaptic biology in schizophrenia. *MedRxiv*. https://doi.org/10.1101/2020.09.12.20192922

Sullivan, P. F., Daly, M. J., & O'Donovan, M. (2012). Genetic architectures of psychiatric disorders: The emerging picture and its implications. *Nature Reviews Genetics*, *13*(8), 537–551. https://doi.org/10.1038/nrg3240

Csardi G, Nepusz T (2006). "The igraph software package for complex network research." *InterJournal*, **Complex Systems**, 1695. https://igraph.org