bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Cis-Regulatory Hubs: a new 3D Model of Complex Disease Genetics with an Application to Schizophrenia

**Running Title:** Cis-Regulatory Hubs: a model for Complex Disease Genetics


## 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: steve.bilodeau@crchudequebec.ulaval.ca

Alexandre Bureau: 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


**Keywords:** Hi-C, Networks, Epigenetics, Complex Diseases

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# Summary Blurb:

Genes and their regulatory elements are organized within 3D networks in neurons which model functional structures and explain schizophrenia genetic etiology.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.


# Abstract

The 3-dimensional (3D) conformation of the chromatin creates complex networks of noncoding regulatory regions (distal elements) and promoters impacting gene regulation. Despite the importance of the role of noncoding regions in complex diseases, little is known about their interplay within regulatory hubs and implication in multigenic diseases like schizophrenia. Here we show that cis-regulatory hubs (CRHs) in neurons highlight functional interactions between distal elements and promoters, providing a model to explain epigenetic mechanisms involved in complex diseases. CRHs represent a new 3D model, where distal elements interact to create a complex network of active genes. In a disease context, CRHs highlighted strong enrichments in schizophrenia-associated genes, schizophrenia-associated SNPs and schizophrenia heritability compared to equivalent structures. Finally, CRHs exhibit larger proportions of genes differentially expressed in schizophrenia compared to promoter-distal element pairs or TADs. CRHs thus capture causal regulatory processes improving the understanding of complex disease etiology such as schizophrenia. These multiple lines of genetic and statistical evidence support CRHs as 3D models to study dysregulation of gene expression in complex diseases more generally.


# Introduction

The etiology of complex diseases involves a broad 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 most risk variants are located within 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 disease etiology (Zhang and Lupski 2015). The difficulty to assign distal regulatory elements to genes hampered the ability to discover the underlying molecular mechanisms. Consistent with a role of noncoding regions in complex phenotypes, there is also strong evidence on the

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

involvement of 3-Dimensional (3D) chromatin conformation in gene regulation. The 3D genome organization, captured by chromosome conformation assays (van Berkum et al. 2010), revealed the physical proximity between regulatory elements and putative target genes. In addition to chromatin loops connecting promoters to distal noncoding regions (Gorkin et al. 2014; Bouwman and de Laat 2015; Dekker and 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 and 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 most cases few promoters, strongly involved in biological processes (Espinola et al. 2021). In fact, highly interconnected enhancers converge to genes with crucial phenotypic implications, with dynamic enhancer crosstalk at the genome-wide level, occurring more frequently during differentiation (Madsen et al. 2020). 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 lead to phenotypic robustness in environmental or genetic perturbations (Tsai et al. 2019). 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). Interestingly, the organization 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.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Schizophrenia is a complex chronic brain disorder associated with perturbations in the transcriptional programs of neurons (Ruzicka et al. 2020). Indeed, schizophrenia is characterized by long-standing delusions and hallucinations strongly reducing life-expectancy (Sullivan et al. 2012). Recent findings suggest that schizophrenia is explained by a polygenic architecture (Smeland et al. 2020), where most of the risk variants are located within noncoding regions. Schizophrenia risk loci are 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, multiple studies have demonstrated the involvement of 3D organization in the disorder. For example, chromatin loops are 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.

# 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, a natural structure for the 3D contacts between two classes of elements: the promoter of genes and their distal elements. To evaluate their role in schizophrenia etiology, we defined CRHs using chromatin contacts provided by Hi-C data with and without additional epigenetic features defining classes of distal regulatory elements (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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(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 (**Fig 1**, 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), a relevant cell type to study schizophrenia (Sey et al. 2020), we identified 62,658 functional pairs of distal elements and promoters where the ABC score exceeded a threshold of 0.012. The value of the threshold was chosen so that we had, on average, 4.51 distal elements per genes, as recommended by Fulco et al. 2019. CRH were built from these connections between promoters and distal elements (**Fig 1A-1C**). We identified 1,633 CRHs, ranging between 2 and 506 nodes (median of 6 elements). Post-mortem brains are an alternative source of neurons to study 3D contacts. In three samples from post-mortem brains: dopaminergic neurons (mentioned as Dopa_1 and Dopa_2 in the supplementary) and general neuron population (mentioned as Neu), respectively (Espeso-Gil et al. 2020, See **Supplementary Methods**), we observed a strong overlap of the pairs of promoters and distal elements detected by the ABC approach with those found in iPSC-derived neurons. Indeed, we found that over 75% of pairs were either strictly found in iPSC-derived neurons (e.g., identical pair) or in an indirect connection within the same CRH (**Fig S1**), supporting the reproducibility of the proposed method.

To start investigating the complexity of CRHs in iPSC-derived neurons, we surveyed genes associated with schizophrenia (Schizophrenia Working Group of the Psychiatric Genomics Consortium 2014). For example, genes involved in glutamatergic transmission or synaptic plasticity pathways (*GRIA1*, *GRIN2A*, and *GRM3*) exhibited strong differences regarding CRH complexity (**Fig 2A**). Indeed, *GRIN2A* and *GRM3* were found in relatively simple CRHs of 2 (1 promoter, 1 enhancer) and 3 (1 promoter, 2 enhancers) nodes respectively, while *GRIA1* was found in a complex network with 24 genes

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

and 72 regulatory regions. Among the 1,633 CRHs, 15% were pairs of 2 nodes and therefore constituted monogamous relationships, while 85% had 3 elements or more (**Fig 2B**). 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, Two-sided Wilcoxon signed-rank test, p-value<=2e-16) (**Fig 2C**). 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 (**Fig 2D**). Comparable results were observed in post-mortem brain (**Fig S2-S4**). As expected, the proportion of distal elements was positively correlated with the connections between promoters and distal elements (or complexity) within CRHs (Spearman tau=0.37, p-value<=2e-16), revealing that complex CRHs are significantly associated with a higher proportion of distal elements. The above 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 (**Fig 2A**): 1) the proportion of instances where one distal element connects one promoter with at-least one other distal element or the reverse: one promoter connects one distal element with at least one other promoter (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 (63%) were connected to a single promoter while 90% of promoters showed interactions with multiple distal elements. We also observed that 1% of promoters are within 1-1-N relationships versus 5% of distal elements (**Fig 2E**). This result suggests that distal elements share a gene more frequently than genes share a distal element, in accordance with previous

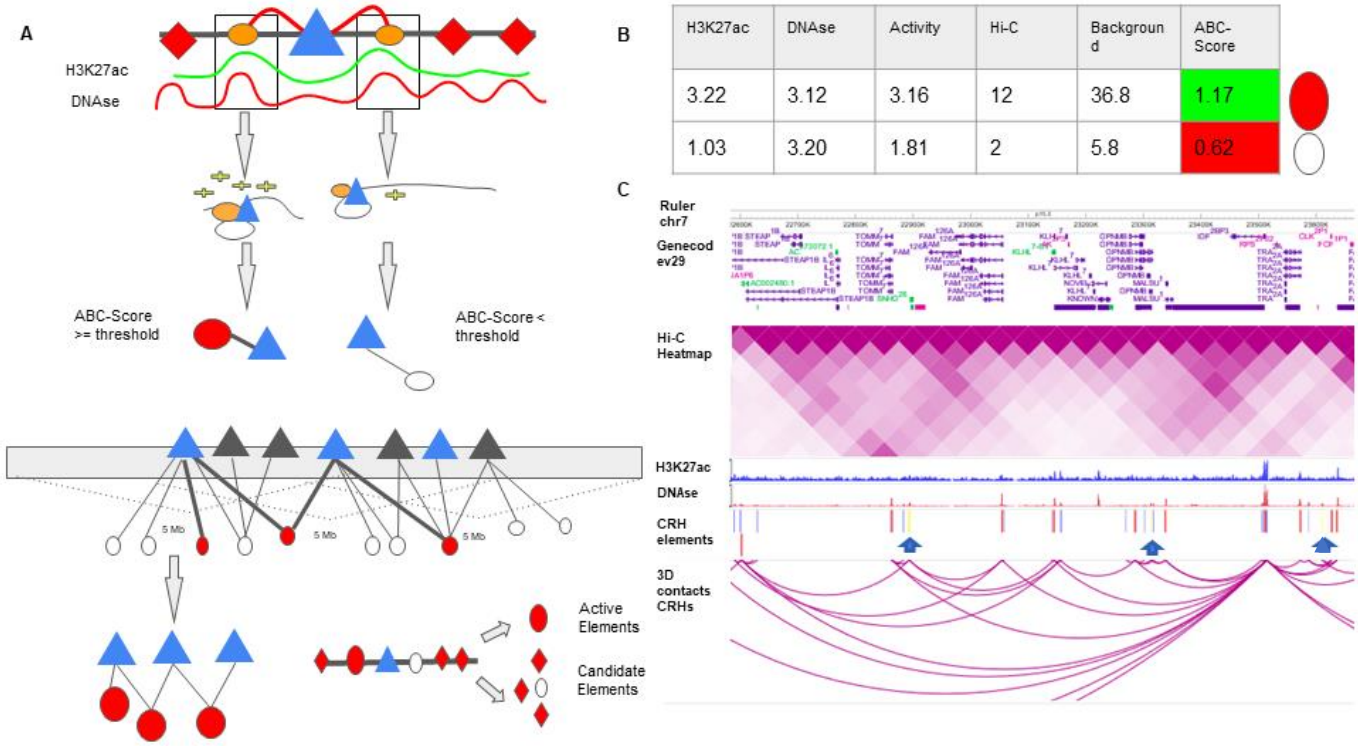
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

findings in model organisms (Espinola et al. 2021). To confirm these results, we compared CRHs built using the ABC approach with other CRH definitions or across post-mortem brain tissues. We found that promoters were also more connected than distal elements (**Fig S5-S6**). However, promoters showed fewer connections in our other CRH definitions than the ABC approach. 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 (Lieberman-Aiden et al. 2009), TADs (Dixon et al. 2012), and frequently interacting regions (FIREs) (Schmitt et al 2016), respectively segmenting the genome into open and close chromatin, domains of frequent interactions between distal elements and genes, and hotspots for chromatin contacts. In iPSC-derived neurons the majority of CRHs (76%) shared compartments of the same type, with 46% and 29% for active and inactive compartments, respectively (**Fig 2F**) while only a minor portion (8%) of CRHs overlapped several compartments of different types or were in genomic regions not assigned to a compartment (17%). This result was confirmed in post-mortem brains where 47%, and 53% of CRHs overlapped active compartments for general neuronal populations and dopaminergic neuronal nuclei (**Fig S7**). 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 (**Fig 2G**). Interestingly, 26% of TADs included 2 or more CRHs. These observations were confirmed in post-mortem brain tissues (**Fig S8**) and by testing multiple TAD detection algorithms (**Fig S9**). Lastly, CRHs were enriched in FIREs compared to candidate CRHs (tissue-specific regions non integrating 3D contacts, see **Methods**) (Two-sided Fisher exact test, 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 the different CRH definitions (**Fig S10**). Collectively, our results support

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

that CRHs are networks of interacting regulatory regions and genes at a finer scale compared to previously defined chromosome structures. Given the similarity between CRHs in neurons from iPSC and from post-mortem brain tissue, from now on results are restricted to neurons from iPSC.




| H3K27ac | DNAse | Activity | Hi-C | Backgroun d | ABC-Score |
|---|---|---|---|---|---|
| 3.22 | 3.12 | 3.16 | 12 | 36.8 | 1.17 |
| 1.03 | 3.20 | 1.81 | 2 | 5.8 | 0.62 |

**Figure 1**

**CRHs are built from ABC-Score methodology**

(**A**) Diagram showing ABC-Score methodology to build functional pairs of promoter (triangles) and distal element (circles). H3K27ac and DNAse signals are shown on an arbitrary scale. Among all distal elements (orange circles), active elements (red circles) are discriminated from candidate elements (white circles) based on the value for the ABC-Score. Candidate elements are DNAse accessible regions without H3K27ac signal and non-overlapping active elements. (**B**) Example for the computation of the ABC-Score. Activity is calculated with geometric mean of H3K27ac and DNAse signals. Finally, the ABC-Score is the product of Activity by Hi-C signal divided by the background activity, within a 5Mb window. Here we used a threshold of 1.12 to determine functional connection. All values shown in the table are

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

arbitrary. (**C**) Representation of the physical contacts of a CRH subset on chromosome 7 from the WashU Epigenome Browser (Zhou et al. 2011). Hi-C data in iPSC-derived neurons is represented. In the CRH element track, we represented distal elements belonging to the CRH (red), promoters (blue) and, elements encompassing noncoding SNPs (yellow bars and blue arrows).

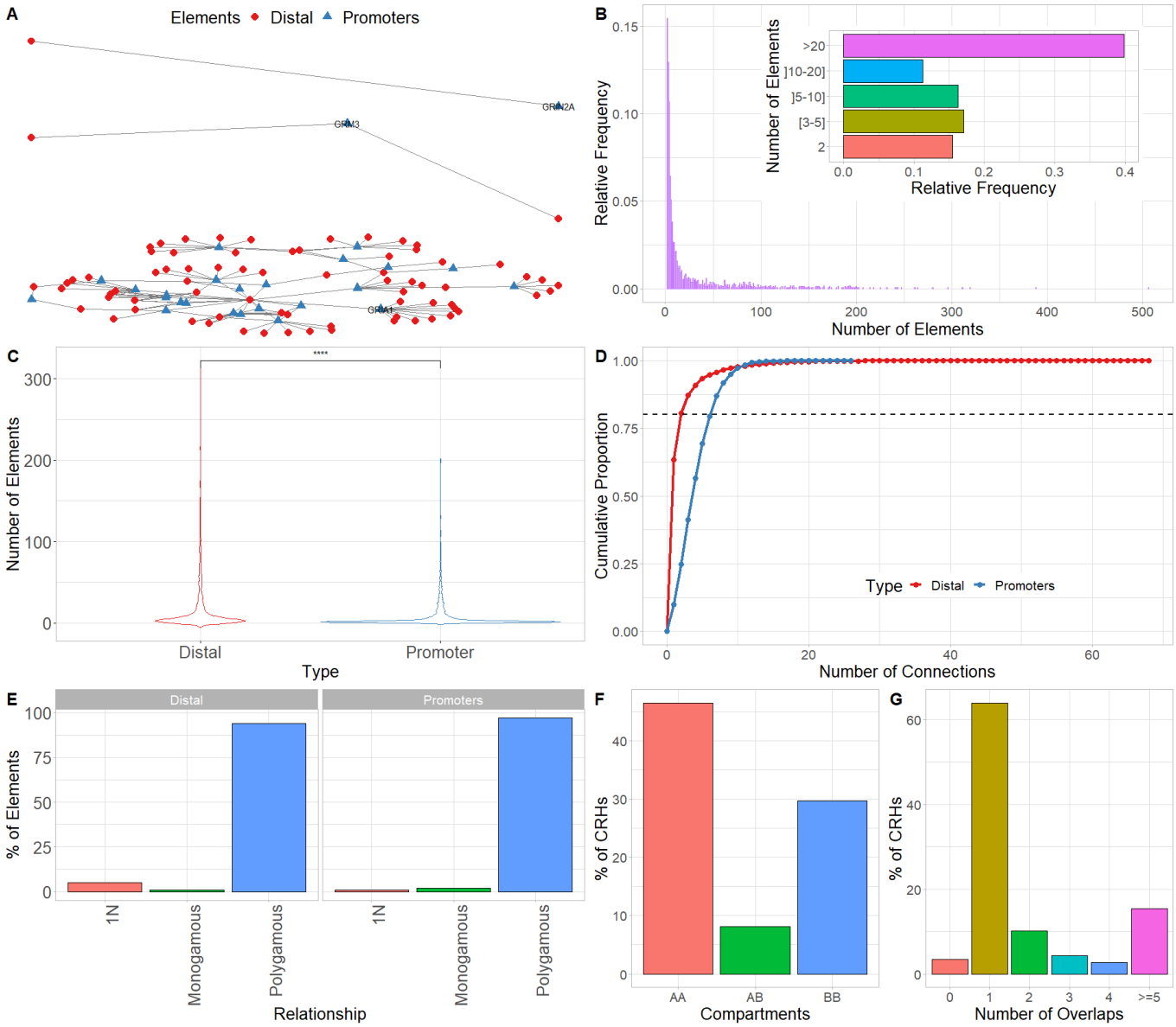



**Figure 2**

**CRH are 3D-based networks mainly constituted by distal elements and more local than high order 3D features**

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(**A**) Cis-regulatory hubs (CRH) connecting promoters (blue) to distal elements (red) for *GRIN2A* (**Top**), *GRM3* (**Middle**), and *GRIA1* (**Bottom**) genes. The genes are within monogamous, 1-1-N and polygamous relationships, respectively. Distal elements are represented by red circles while promoters by blue triangles. (**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) per CRH. We used two-sided Wilcoxon signed-rank test to compare the number of elements. (**D**) Cumulative distribution function of the number of connections for promoters (blue) and distal elements (red). The dotted line shows the 80$^{th}$ percentile of the number of connections. (**E**) Distribution of the kind of relationship for distal elements (**Left**) and promoters (**Right**). (**F**) Distribution of overlap of CRHs with each compartment type (AA: Active-Active; AB: Active-Inactive; BB: Inactive-Inactive). When CRHs overlap several compartments, we restrict our attention to the farthest elements. The CRHs in genomic regions not assigned to compartments (17%) were omitted from the distribution. (**G**) Distribution of the number of TADs overlapped by each CRH, when TADs are detected with the directionality index.

Data Information: In **(C)** **** represent p-value <= 0.0001.

## 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 and their involvement in schizophrenia, we overlayed the chromatin states defined by the Roadmap Epigenomics Consortium (Roadmap Epigenomics Consortium et al. 2015). The 18-states model in neurons was subdivided as follows into three broad categories: 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). At the broad category level, we

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

found that most elements (promoters and distal elements) included in CRHs (58%) overlapped Weakly Active regions against 49% for Inactive or Repressor and 53% for Active regions, respectively (**Fig 3A**). At the individual state level, we observed that 39% of the distal elements included the Quiescent state (**Fig 3B**) but that CRHs were enriched 2.35-fold (Two-sided Fisher exact test, p-value<=2e-16) in active states and depleted in inactive states (Two-sided Fisher exact test, Odds ratio=0.49, p-value<=2e-16) compared to candidate CRHs. 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 (Two-sided Fisher exact test, Odds ratio=1.81, p-value<=2e-16), DNAse (Two-sided Fisher exact test, Odds ratio=1.66, p-value<=2e-16) and H3K27ac (Two-sided Fisher exact test, Odds ratio=1.44, p-value<=2e-16), but not with CTCF. Our results were supported by other CRH definitions (**Fig S11**). 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 across CRHs. Indeed, 35% of CRHs exhibited a unique combination of chromatin states (e.g., a set of states found only once in CRHs) (**Table S1**). Also, CRHs characterized by active states were more complex than those strongly defined by quiescent states (18_Quies) (**Fig 3C**). 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.

As 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 et al. 2020). We found that 35% of genes significantly associated with schizophrenia are within CRHs compared to 23% for all other genes (1.82-fold enrichment, Two-sided Fisher exact test, p-value<=2.2e-16). Moreover, 42% (687/1633) of CRHs include at least 1

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

schizophrenia-associated gene with 23% (376/1633) harboring several schizophrenia-related genes (mean=1.77, max=69) (**Fig 3D**). Finally, we found that CRHs were enriched in GO biological processes associated with schizophrenia (**Fig 3E**). 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.

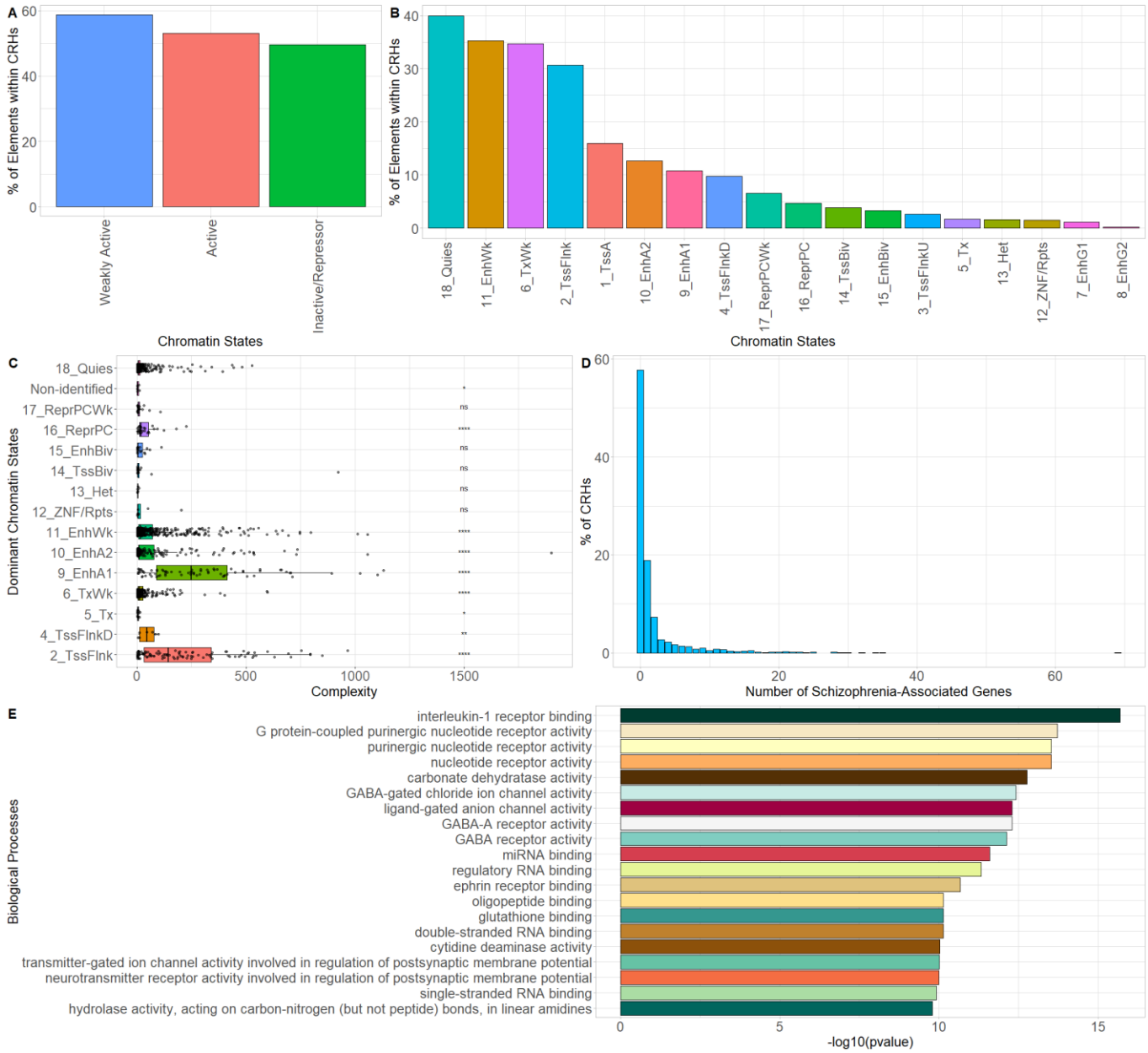



**Figure 3**

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**CRHs are enriched in transcriptionally active elements and genes associated with schizophrenia**

(**A**) Proportion of elements (promoters and distal elements) included within CRHs overlapping chromatin states grouped by activity. (**B**) Proportion of distal elements included within CRHs overlapping individual chromatin states. (**C**) Boxplot representing complexity by most present chromatin state within each CRH. Two-sided Wilcoxon rank-sum test was used to compare complexity for each chromatin state with 18_Quies state. (**D**) Distribution of the number of schizophrenia-associated genes per CRH. (**E**) GO enrichment for all genes found within CRHs. The top-20 biological processes are represented.

Data Information: In (**D**), ns=non-significant, * represent p-value <= 0.05, ** represent p-value <= 0.01, *** represent p-value <= 0.001, while **** represent p-value <= 0.0001.

## CRHs containing schizophrenia-associated genes are small and highly expressed

To further characterize CRHs including schizophrenia-associated genes, we examined their characteristics regarding complexity and gene expression levels. Interestingly, CRHs encompassing schizophrenia-associated genes showed larger distances between elements than CRHs not harboring schizophrenia-associated genes (**Fig 4A**). Also, the number of connections with distal elements was slightly lower for schizophrenia-associated genes than non-associated ones (mean associated genes=4.39, mean non-associated genes=4.55, Two-tailed t-test p-value=0.005). In addition, schizophrenia-associated genes included within CRHs showed higher expression levels than non-associated genes (median associated=6.18, median non-associated=5.87, Two-sided Wilcoxon rank-sum test p-value<=2.2e-16) (**Fig 4B**) and were enriched in active distal elements (Two-sided Fisher exact test, Odds ratio=1.79, p-value<=2.2e-16). Moreover, schizophrenia-associated genes were more often monogamous genes compared to non-associated ones, showing a 1.34-fold enrichment (Two-sided Fisher exact test, p-value=0.04). Indeed, 26% of monogamous genes are schizophrenia-associated genes against 20% for non-monogamous ones (**Fig 4C**). The number of distal elements in CRHs harboring

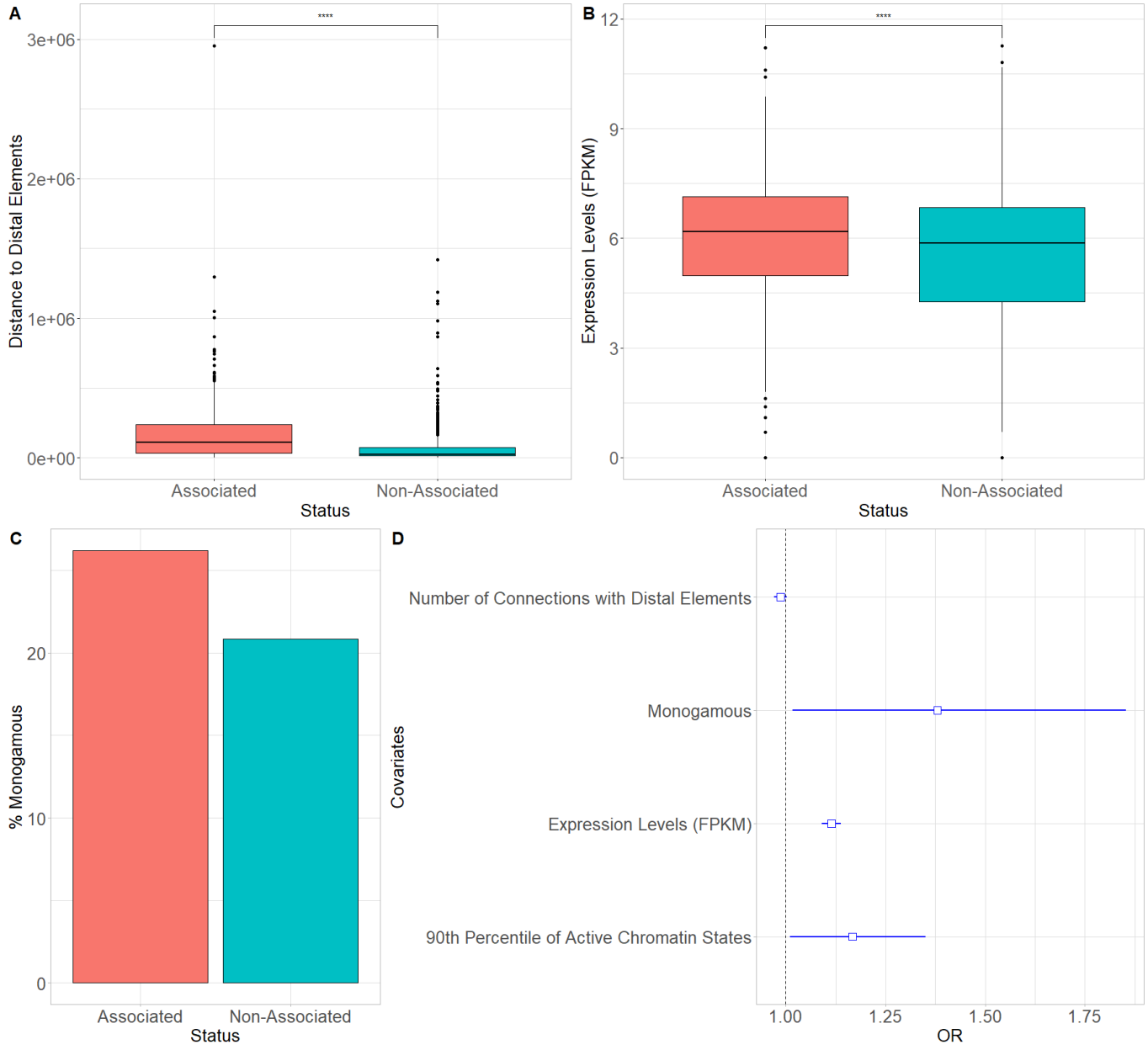
bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

schizophrenia-associated genes was 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.

## Multivariate analysis of CRH features with respect to schizophrenia-associated genes

To examine the mutually adjusted influence of the factors examined in the previous sections on schizophrenia-associated genes, we fitted a logistic regression of the status of genes (associated vs. non-associated with schizophrenia) on RNA level, the number of connections to distal elements, the 90th percentile of the proportion of active distal elements per gene, and the information regarding monogamy. As expected, the gene status regarding its association with schizophrenia was positively associated with RNA level, the 90th percentile of the proportion of active distal elements, and the monogamy status, while it was negatively associated with the number of connections, confirming our results found with univariate analyses (**Fig 4D**). Collectively, our results suggest that schizophrenia-associated genes are within small hubs characterized by fewer connections to distal elements and higher transcriptional activity.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.




## Figure 4

### Features of schizophrenia-associated genes

(**A**) Boxplot of mean distance between elements for CRHs encompassing schizophrenia-associated genes and CRHs not harboring ones. Differences were assessed using two-sided Wilcoxon rank-sum test. (**B**) Boxplot of RNA levels for schizophrenia-associated genes and non-associated ones. Differences were assessed using two-sided Wilcoxon rank-sum test. (**C**) Percentage of monogamous genes which are associated with schizophrenia or non-associated. (**D**) Odds ratio (OR) and their 95% confidence interval

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

for a logistic regression of the association status of genes with schizophrenia (yes/no). The dotted line represents the null value.

Data Information: In (**A**) and (**B**) **** represent p-values <=2e-16.

## 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 disorder (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 (after clumping, See **Methods**) from genome-wide association studies (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. 2020). We mapped them to their corresponding CRH and quantified their enrichments at various association p-value thresholds using the two-sided 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 (**Fig 5A**). Then, we used the same methodology as Nasser et al. 2021 to define enrichment in common SNPs overlapping a given functional annotation (proportion of significant SNPs for schizophrenia/proportion of all common SNPs). Consistent with our previous finding, we observed higher fold enrichments (Enrichment for elements of interest / Enrichment for candidates) for CRHs than for distal elements, becoming stronger with the significance level (**Fig 5B**). This enrichment was stronger with alternative definitions of CRHs (**Fig S12**). 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 explained by a functional annotation. First, comparing CRHs to equivalent non-tissue-specific noncoding regions, we

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

ensured to maximize the explained heritability by using tissue-specific elements and integrating 3D contacts by conditioning 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 equivalent to those found in CRHs but without 3D contacts. Among functional annotations with significant heritability, the heritability enrichment was higher for CRHs than for non-tissue-specific noncoding regions and candidate CRHs (**Fig 5C**), with strong enrichment signal 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 and using DNAse, CRHs built using the ABC approach performed better regarding schizophrenia heritability, showing enrichments of 3.08 against 0.84 and 2.98, respectively (**Fig S13**). 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 (**Tables S2-S3**). This result demonstrates a better concordance of the CRH including epigenetic features to explain schizophrenia heritability compared to only using chromatin interactions or combining chromatin interactions with chromatin accessibility.

Since we observed that schizophrenia-associated genes are highly expressed, enriched in small hubs, and connected to few distal elements, we defined strata of CRH number of promoters based on the proportion of total variance explained by CRHs (intraclass correlation) through a linear mixed model of the gene expression (**Fig S14**). 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) (**Fig 5D**). Overall, these results support that CRHs, especially small ones, are a relevant structure to explain the etiology of schizophrenia.

## CRHs predict the association between schizophrenia-associated noncoding SNPs and differentially expressed genes

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Since we observed that CRHs are enriched in schizophrenia-associated SNPs and schizophrenia heritability, we wondered whether they represent a useful structure to link noncoding SNPs to genes differentially expressed in schizophrenia. To conduct this investigation, we leveraged information on 8,413 up- and down-regulated genes in all the available cell-types from a large set of schizophrenia patient brain tissues compared to controls (differentially expressed genes or DEGs) from SZBDMulti-Seq (Ruzicka et al. 2020) and schizophrenia-associated SNPs from genome-wide association studies (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. 2020). First, CRHs were strongly enriched in DEGs compared to candidate genes, not included in CRHs (Two-sided Fisher exact test, Odds ratio=5.33, p-value<=2e-16). CRHs encompassing at least one DEG exhibited a slightly larger proportion of distal elements compared to CRHs without DEGs (median of 68% compared to 66%, Two-sided Wilcoxon rank-sum test, p-value=8.01e-06). These results suggest that CRHs may capture the links between SNPs in regulatory regions and DEGs. We tested the hypothesis that links between noncoding SNPs and DEGs are better captured by CRHs than by promoter-distal element pairs and TADs, respectively the simplest form of CRHs and one of the most studied 3D-feature in disease etiology (Fudenberg & Pollard 2019; Bryois et al. 2020), by measuring the proportions of DEGs linked with noncoding SNPs. We first assigned schizophrenia-associated noncoding SNPs to each kind of structure (See **Methods**) and observed that 30% of CRHs exhibited at least one such assigned SNP in their regulatory regions, compared to 4%, and 95% for pairs and TADs, respectively. In the subset of elements where we observed assigned SNPs, CRHs exhibited a larger proportion of DEGs, exhibiting median proportion of 25% compared to 0% for regulatory regions directly connected with gene promoters and 18% for TADs, respectively (**Fig 5E**). Therefore, as intermediate structures compared to promoter-distal element pairs and TADs, CRHs better capture links between noncoding SNPs to gene expression variation possibly involved in schizophrenia.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

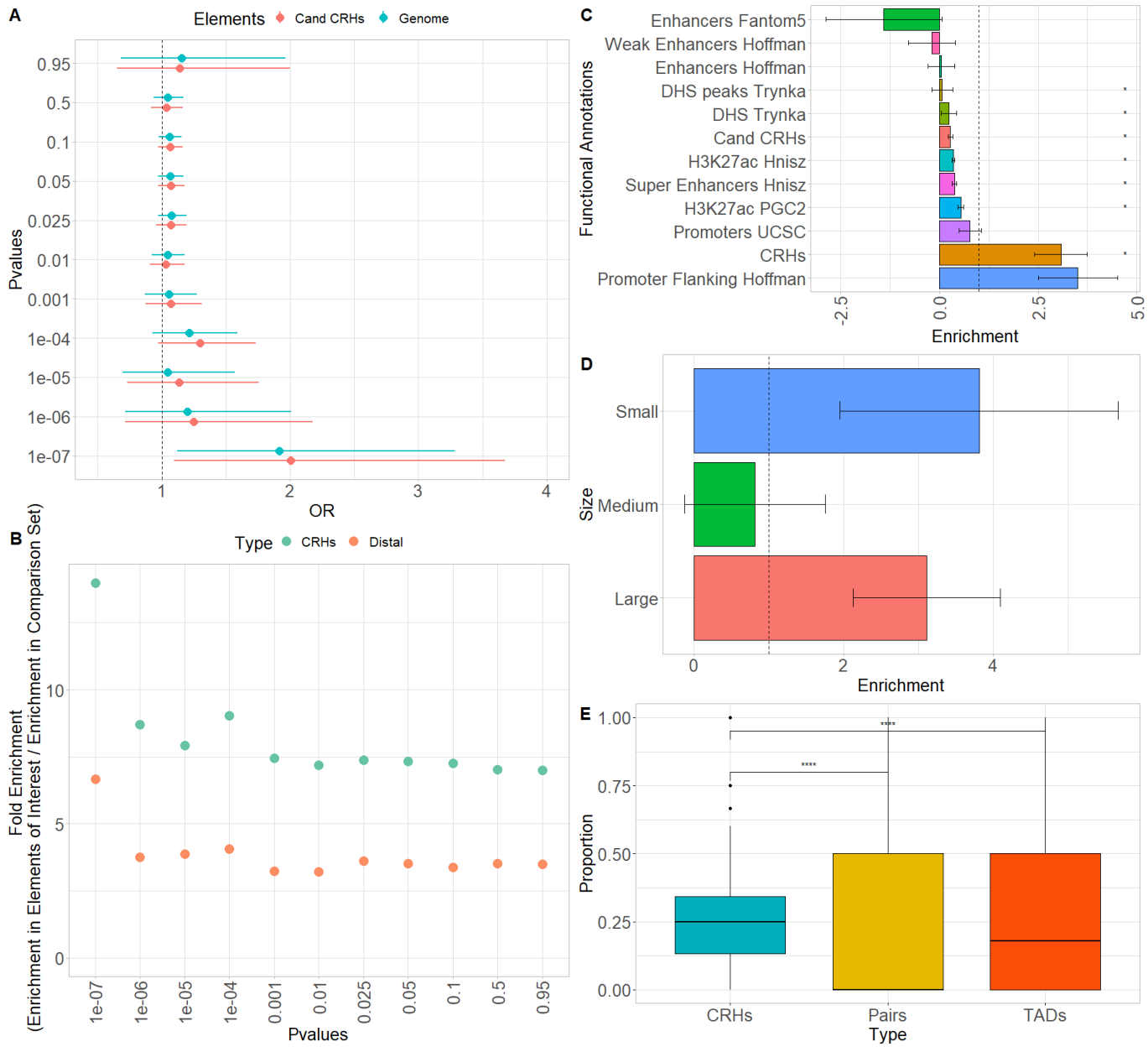



**Figure 5**

**CRHs are enriched in schizophrenia-associated SNPs, schizophrenia heritability, and capture links between noncoding SNPs and genes differentially expressed in schizophrenia**

(**A**) SNP enrichment analysis measured through odds ratio (OR) and their 95% confidence intervals for CRHs compared to candidate CRHs and the rest of the genome at different significance levels. The dotted line represents the null value. (**B**) Fold enrichment of distal elements and CRHs compared to their

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

respective candidate sets for different significance levels. The dotted line represents the null value. (**C**) Schizophrenia heritability enrichment measured with LDSC with error bars for CRHs, candidate CRHs, and non-tissue noncoding elements. The dotted line represents the null value. Errors bars represent the standard errors around the estimates of enrichment. (**D**) Schizophrenia heritability enrichment (measured with LDSC) with error bars for CRHs, considering the number of genes within CRHs. The dotted line represents the null value. (**E**) Boxplot of DEG proportions within CRHs, promoter distal elements pairs, and TADs. We considered only elements where we observed noncoding schizophrenia-associated SNPs. Differences were assessed using two-sided Wilcoxon rank-sum test.

Data Information: In (**C**) * represent p-value <=0.05 after Bonferroni correction. In (**E**) **** represent p-values <=2e-16.

# 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 contributions to 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 (**Fig 1A**). CRHs aim to highlight direct and indirect contacts between promoters and distal elements which may not be targeted by other approaches. Our findings confirm the interest in integrating 3D contacts with tissue-specific regions to gain a deeper understanding of regulatory processes involved in complex diseases, where genetic disruptions may impact the transcription program of several genes (**Fig 1C**). CRHs are enriched in gene promoters and distal elements associated with schizophrenia (**Fig 5A**) and explain a larger portion of heritability than candidate CRHs (**Fig 5C**) or other definitions to characterize CRHs (**Fig S13**). Also, assessing the functional interest of CRHs in schizophrenia etiology, we found that CRHs are more efficient to

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

capture the links between noncoding SNPs to genes differentially expressed in schizophrenia compared to TADs and promoter-distal element pairs. Thus, through CRHs, impacts of polymorphisms on gene expression variation can be better targeted. 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 that 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). CRHs constitute a more local functional organization than higher-order chromatin features (A/B compartments, TADs) (**Fig 2F-2G**) and are enriched in FIREs. In fact, CRHs are strongly enriched in active regions (**Fig 3C**), 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 (**Fig 4B**) and active regulatory regions (**Fig 4D**). These results are in line with those of Sey et al. 2020, as they have shown that schizophrenia-associated genes exhibit higher differential expression in schizophrenia. Based on the above lines of evidence, we argue that focusing on CRHs should be prioritized over other levels of 3D organization in a context of complex phenotypes. Thus, through CRHs, impacts of polymorphisms on gene expression variation can be better targeted, aiming to highlight underlying regulatory processes.

Promoters and distal elements involved in CRHs exhibit different connectivity behaviors. Indeed, CRHs harbor more distal elements than genes (**Fig 2C**), suggesting that within a CRH, genes tend to have more connections compared to distal elements (**Fig 2D**) (Madsen et al. 2020, Espinola et al. 2021). Espinola et al. 2021 have shown that hubs connecting promoters to distal elements encompass a single promoter, while Madsen et al. 2020 exhibited that enhancers are mostly involved in one-to-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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

et al. 2020). However, in our data we found that CRH often harbor several genes connected by distal elements, supporting that CRHs can be either promoter hubs, enhancer hubs or multi hubs (**Fig 2A**) (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 higher resolution (Nagano et al. 2013).

An important contribution of this study is to establish CRHs as a relevant model to study complex diseases such as schizophrenia. Indeed, we found strong enrichments in schizophrenia-associated SNPs, schizophrenia heritability within CRHs (**Fig 5A, 5C**), compared to candidate CRHs. Also 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 aim to highlight indirect connections between promoters and distal elements and our results show they offer an advantage over a pair of enhancer-promoter or larger domains to efficiently link noncoding SNPs to DEGs in schizophrenia. Collectively, these results point to the capability of CRHs to capture complex interplay between regulatory regions, which can help to fine map the functional regions involved in complex diseases, 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 (**Fig 4C-4D**). 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 (**Fig 5D**). 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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

connected genes are more involved in disease etiology. 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 pathways to elucidate the underlying causal regulatory processes involved in complex diseases. Indeed, since a substantial portion of schizophrenia heritability remains to be explained, future work will be needed to assess the relevance of CRHs to help detect the rare variants which may be involved. 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 and 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**).

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**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 and 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 et al. 2021 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.

## Summary Statistics for Schizophrenia

The original SCZ3 GWAS summary statistics (Schizophrenia Working Group of the Psychiatric Genomics Consortium et al. 2020) used in SNP and heritability enrichments were downloaded from the PGC site https://www.med.unc.edu/ pgc/results-and-downloads. To assess independent SNPs in enrichment analyses, we used the clumped SNP file keeping the SNPs with the highest association signal with schizophrenia for a given genomic window.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## Schizophrenia-associated genes

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 (Schizophrenia Working Group of the Psychiatric Genomics Consortium 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.

## 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 non-neuron-specific annotations corresponding to regulatory regions included within CRHs: promoters, H3K27ac histone marks, DNase I hypersensitive sites, ChromHMM/Segway predictions, super-enhancers, and FANTOM5 enhancers. By proceeding this way, we sought an unbiased comparison of SNP heritability in neuron CRHs 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 consider the background activity and avoid inflating the enrichment signal, as suggested by Finucane et al. 2015.

## Linking noncoding SNPs to DEGs in schizophrenia

To link noncoding SNPs to differentially expressed genes, we only considered the clumped SNPs, since they represent the genetic variations the most associated with schizophrenia within genomic windows, without applying a p-value threshold. Then, the proportion of DEGs was calculated for the subset of CRHs, promoter distal-element pairs and TADs where we observed at least one SNP in one distal element.

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

## 3D Features and other analyses

All analyses related to 3D features (A/B compartments, TADs, FIREs), functional enrichments, peak calling and other technical details are presented in supplemental method section.

## Availability of data and materials

Datasets analyzed in this study are publicly available from: PsychENCODE Knowledge Portal (https://www.synapse.org, syn13363580, syn20833047) for Hi-C in iPSC, and post-mortem brains, respectively. 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.

## Ethics approval and consent to participate

Not applicable

## Consent for publication

Not applicable

## Competing interests

None declared.

## Funding

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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.

## 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, whose contributors and funding are listed in the Supplement.

## Authors' contributions

Conceptualization: LM and AB

Data Curation: LM and AB

Formal Analysis: LM and AB

Investigation: LM and AB

Methodology: LM, AB, CJB, SB

Supervision: AB

Validation: AB, CJB, SB

Visualization: LM

Writing- original draft: LM, AB, SB, CJB & AD

Writing- review & editing: LM, AB, SB, CJB & AD

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

# References

Abascal, F., Acosta, R., Addleman, N. J., Adrian, J., Afzal, V., Aken, B., . . . Zimmerman, J. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature, 583*, 699–710. doi:10.1038/s41586-020-2493-4

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

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

Bryois, J., Garrett, M. E., Song, L., Safi, A., Giusti-Rodriguez, P., Johnson, G. D., . . . Crawford, G. E. (2018, 12). Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nature Communications, 9*, 3121. doi:10.1038/s41467-018-05379-y

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

Cardozo Gizzi, A. M., Cattoni, D. I., Fiche, J. B., Espinola, S. M., Gurgo, J., Messina, O., . . . Nollmann, M. (2019). Microscopy-Based Chromosome Conformation Capture Enables Simultaneous Visualization of Genome Organization and Transcription in Intact Organisms. *Molecular Cell, 74*, 212–222.e5. doi:10.1016/j.molcel.2019.01.011

Consortium, R. E., Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., . . . Kellis, M. (2015, 2). Integrative analysis of 111 reference human epigenomes. *Nature, 518*, 317–330. doi:10.1038/nature14248

Csardi, Gabor & Nepusz, Tamas. (2005). The Igraph Software Package for Complex Network Research. InterJournal. Complex Systems. 1695.

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

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

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. doi:10.1186/s13059-017-1250-y

Durand, N. C., Shamim, M. S., Machol, I., Rao, S. S., 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*, 95–98. doi:10.1016/j.cels.2016.07.002

Espeso-Gil, S., Halene, T., Bendl, J. et al. A chromosomal connectome for psychiatric and metabolic risk variants in adult dopaminergic neurons. Genome Med 12, 19 (2020). https://doi.org/10.1186/s13073-020-0715-x

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Espinola, S. M., Götz, M., Bellec, M., Messina, O., Fiche, J. B., Houbron, C., . . . 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*, 477–486. doi:10.1038/s41588-021-00816-z

Finucane, H. K., Bulik-Sullivan, B., Gusev, A., Trynka, G., Reshef, Y., Loh, P. R., . . . Price, A. L. (2015). Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nature Genetics, 47*, 1228–1235. doi:10.1038/ng.3404

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*, 2175–2180. doi:10.1073/pnas.1808631116

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*, 358–368. doi:10.1016/j.cell.2016.05.025

Fulco, C. P., Nasser, J., Jones, T. R., Munson, G., Bergman, D. T., Subramanian, V., . . . Engreitz, J. M. (2019). Activity-by-contact model of enhancer–promoter regulation from thousands of CRISPR perturbations. *Nature Genetics, 51*, 1664–1669. doi:10.1038/s41588-019-0538-0

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

Girdhar, K., Hoffman, G. E., Jiang, Y., Brown, L., Kundakovic, M., Hauberg, M. E., . . . Akbarian, S. (2018). 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

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

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

Halvorsen, M., Huh, R., Oskolkov, N., Wen, J., Netotea, S., Giusti-Rodriguez, P., . . . Szatkiewicz, J. P. (2020, 12). Increased burden of ultra-rare structural variants localizing to boundaries of topologically associated domains in schizophrenia. *Nature Communications, 11*, 1842. doi:10.1038/s41467-020-15707-w

Hauberg, M. E., Creus-muncunill, J., Bendl, J., Kozlenkov, A., Zeng, B., Corwin, C., . . . Fullard, J. F. (2020). Common schizophrenia risk variants are enriched in open chromatin regions of human glutamatergic neurons. *Nature Communications*. doi:10.1038/s41467-020-19319-2

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*, 340–352. doi:10.1016/j.ajhg.2017.07.011

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

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

Lieberman-aiden, E., Berkum, N. L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., . . . Mirny, L. A. (2009). of the Human Genome. *33292*, 289–294.

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

Matthew T. Maurano, E. R., Haugen, A. P., Brody, F. N., Sun1, T. K., Bates1, S. N., Raubitschek6, C. C., . . . Stamatoyannopoulos, J. A. (2012). Systematic Localization of Common Disease-Associated Variation in Regulatory DNA. *Science, 337*, 1190–1195. doi:10.1126/science.1222794.Systematic

Melo, S., Scho, R., Holtgrewe, M., Klever, M.-k., Tu, S., Heinrich, V., . . . Mundlos, S. (2020). Hi-C Identifies Complex Genomic Rearrangements and TAD-Shuffling in Developmental Diseases. 872–884. doi:10.1016/j.ajhg.2020.04.016

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

Nasser, J., Bergman, D. T., Fulco, C. P., Guckelberger, P., Doughty, B. R., Patwardhan, T. A., . . . Engreitz, J. M. (2021, 4). Genome-wide enhancer maps link risk variants to disease genes. *Nature*. doi:10.1038/s41586-021-03446-x

Oudelaar, A. M., Harrold, C. L., Hanssen, L. L., 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*. doi:10.1038/s41467-019-13404-x

Rajarajan, P., Borrman, T., Liao, W., Schrode, N., Flaherty, E., Casiño, C., . . . Akbarian, S. (2018, 12). Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. *Science, 362*, eaat4311. doi:10.1126/science.aat4311

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

Rodriguez-Esteban, R., & Jiang, X. (2017). Differential gene expression in disease: a comparison between high-throughput studies and the literature. *BMC medical genomics*, 10(1), 59. https://doi.org/10.1186/s12920-017-0293-y

Roussos, P., Mitchell, A. C., Voloudakis, G., Fullard, J. F., Pothula, V. M., Tsang, J., . . . Sklar, P. (2014). A Role for Noncoding Variation in Schizophrenia. *Cell Reports, 9*, 1417–1429. doi:10.1016/j.celrep.2014.10.015

Rubin, A. J., Barajas, B. C., Furlan-magaril, M., Lopez-pajares, V., Mumbach, M. R., Howard, I., . . . Khavari, P. A. (2017). Lineage-specific dynamic and pre-established enhancer – promoter contacts cooperate in terminal differentiation. *Nature Publishing Group, 49*. doi:10.1038/ng.3935

Ruzicka, W. B., Mohammadi, S., Davila-Velderrain, J., Subburaju, S., Tso, D. R., Hourihan, M., & Kellis, M. (2020). Single-cell dissection of schizophrenia reveals neurodevelopmental-synaptic axis and transcriptional resilience. *MedRxiv*, 2020.11.06.20225342. https://doi.org/10.1101/2020.11.06.20225342

bioRxiv preprint doi: https://doi.org/10.1101/2021.06.22.449467; this version posted November 26, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

Sey, N. Y., Hu, B., Mah, W., Fauni, H., McAfee, J. C., Rajarajan, P., . . . 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*, 583–593. doi:10.1038/s41593-020-0603-0

Smeland, O. B., Frei, O., Dale, A. M., & Andreassen, O. A. (2020). The polygenic architecture of. *Nature Reviews Neurology*. doi:10.1038/s41582-020-0364-0

Song, M., Pebworth, M. P., Yang, X., Abnousi, A., Fan, C., Wen, J., . . . Shen, Y. (2020). Cell-type-specific 3D epigenomes in the developing human cortex. *Nature, 587*, 644–649. doi:10.1038/s41586-020-2825-4

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/nrg3240Tsai, A., Alves, M. R., & Crocker, J. (2019). Multi-enhancer transcriptional hubs confer phenotypic robustness. *eLife, 8*, 1–17. doi:10.7554/eLife.45325

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

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

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

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

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

Zhou, X., Maricque, B., Xie, M. et al. (2011) The Human Epigenome Browser at Washington University. *Nat Methods* 8, 989–990. https://doi.org/10.1038/nmeth.1772