2020-11-12

Measuring Stability of 3D Chromatin Conformations and Identifying Neuron Specific Chromatin Loops Associated with Schizophrenia Risk

Tyler M. Borrman
University of Massachusetts Medical School

MEASURING STABILITY OF 3D CHROMATIN CONFORMATIONS AND IDENTIFYING NEURON SPECIFIC CHROMATIN LOOPS ASSOCIATED WITH SCHIZOPHRENIA RISK

A Dissertation Presented

By

TYLER BORRMAN

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

NOVEMBER 12th, 2020

BIOINFORMATICS & COMPUTATIONAL BIOLOGY

MEASURING STABILITY OF 3D CHROMATIN CONFORMATIONS AND IDENTIFYING NEURON SPECIFIC CHROMATIN LOOPS ASSOCIATED WITH SCHIZOPHRENIA RISK

A Dissertation Presented

By

TYLER BORRMAN

This work was undertaken in the Graduate School of Biomedical Sciences

Bioinformatics and Integrative Biology

Under the mentorship of

Zhiping Weng, Ph.D., Thesis Advisor

Job Dekker, Ph.D., Member of Committee

Nicholas Rhind, Ph.D., Member of Committee

Amir Mitchell, Ph.D., Member of Committee

Wei Wang, Ph.D., External Member of Committee

Celia Schiffer, Ph.D., Chair of Committee

Mary Ellen Lane, Ph.D.,

Dean of the Graduate School of Biomedical Sciences

November 12th, 2020

This thesis is dedicated to my Mom and Dad for always encouraging me to pursue my dreams and providing me with the tools required to achieve them.

ACKNOWLEDGEMENTS

The results presented in this thesis are a culmination of tireless effort from colleagues and collaborators, guidance from advisors, and support from family and friends. I thank my advisor, Zhiping Weng, for being a constant inspiration throughout my graduate studies. Zhiping has reflected in me a passion for problem solving, critical thinking, and determination. I am incredibly grateful for the opportunity to work with Zhiping and call myself a member of ‘Zlab’.

I owe a great deal of thanks to all my collaborators, without whom none of this work would be possible. Specifically, I would like to thank Houda Belaghzal, Prashanth Rajarajan, and Will Liao. Together we made some extraordinary discoveries and I feel so fortunate to have worked alongside such brilliant scientists. The projects in this thesis would also cease to exist without the innovation and guidance of the following professors, Job Dekker, Schahram Akbarian, Brian M. Baker, and Brian G. Pierce. I thank all the members of the Weng lab and Dekker lab, for helpful contributions, discussions, and advice. I would like to acknowledge my TRAC committee: Celia Schiffer, Job Dekker, and Nick Rhind for their supervision and direction throughout the many years of my doctoral research. I also thank my undergraduate advisor, Javier Arsuaga, who sparked my interest in applying mathematics and computation to biological problems.

I would never have pursued a career in science without the love and support from my family and friends. Thank you Mom for teaching me everything

and always being there to listen or provide words of comfort. Thank you Dad for convincing me nothing is ever out of reach. I would also like to thank my siblings: Kristina, you paved the way for so many things and I am so grateful I could watch you bravely take the first steps toward many of life's challenges. Katie, I probably would have lost my mind long ago without you. Thanks for lifting me up when I am down and always being someone I can relate and confide in. Travis, my Jurassic Brother, I am so lucky and proud to have you in my life. Thanks for your constant encouragement, inspiring enthusiasm, and love. I would also like to acknowledge my two closest friends: Anthony Antonaros, thanks for your unwavering friendship and for the encouragement to pursue graduate studies even at the expense of our routine morning trips to Golden Coffee. Zoey Espinoza, thanks for pushing me to get out of my comfort zones, move with you to the TL, and always chase my ambitions. I would also like to thank my Worcester family of friends. Our movie nights and trips to Café Neo were an essential remedy for difficult days in lab. Lastly, I would like to thank my loving partner, Hannah MacMillan, for all her care, dedication and support. During the last weeks of thesis writing, spent together under isolation from a global pandemic, I thank you so much Hannah for being my world.

ABSTRACT

The 23 pairs of chromosomes comprising the human genome are intricately folded within the nucleus of each cell in a manner that promotes efficient gene regulation and cell function. Consequently, active gene rich regions are compartmentally segregated from inactive gene poor regions of the genome. To better understand the mechanisms driving compartmentalization we investigated what would occur if this system was disrupted. By digesting the genome to varying sizes and analyzing the fragmented 3D structure over time, our work revealed essential laws governing nuclear compartmentalization.

At a finer resolution within compartments, chromatin forms loop structures capable of regulating gene expression. Genome wide association studies have identified numerous single nucleotide polymorphisms (SNPs) associated with the neuropsychiatric disease schizophrenia. When these SNPs are not located within a gene it is difficult to gain insight into disease pathology; however, in some cases chromatin loops may link these noncoding schizophrenia risk variants to their pathological gene targets. By generating 3D genome maps, we identified and analyzed loops of glial cells, neural progenitor cells, and neurons thereby expanding the set of genes conferring schizophrenia risk.

The binding of T-cell receptors (TCRs) to foreign peptides on the surface of diseased cells triggers an immune response against the foreign invader. Utilizing available structural information of the TCR antigen interface, we developed computational methods for successful prediction of TCR-antigen

binding. As this binding is a prerequisite for immune response, such improvements in binding prediction could lead to important advancements in the fields of autoimmunity and TCR design for cancer therapeutics.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF COPYRIGHTED MATERIALS PRODUCED BY THE AUTHOR

Chapter II contains unpublished text/figures. This paper is currently under revision at *Nature Genetics*.

Chapter III is adapted from a manuscript published in *Science*:

Rajarajan, P., Borrman, T., Liao, W., Schrode, N., Flaherty, E., Casiño, C., … Akbarian, S. (2018). Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk. Science, 362(6420). https://doi.org/10.1126/science.aat4311

No permission was required to use the material from the publication

Chapter IV is adapted from a manuscript published in *Proteins: Structure, Function and Bioinformatics*:

Borrman, T., Cimons, J., Cosiano, M., Purcaro, M., Pierce, B. G., Baker, B. M., & Weng, Z. (2017). ATLAS: A database linking binding affinities with structures for wild-type and mutant TCR-pMHC complexes. Proteins: Structure, Function and Bioinformatics, (85), 908–916. https://doi.org/10.1002/prot.25260

No permission was required to use the material from this publication

Chapter V contains unpublished test/figures. This paper is currently under revision at *Bioinformatics*.

Because the history of evolution is that life escapes all barriers. Life breaks free. Life expands to new territories. Painfully, perhaps even dangerously. But life finds a way.

-Michael Crichton (1990), *Jurassic Park*

CHAPTER I: INTRODUCTION

Computational Biology

The origins of computational biology can be traced back to the early 1960s. Amino acid sequence arrangements of proteins were being determined for the first time. However, method limitations required larger protein sequences to be assembled by their smaller peptide sequence fragments. For large proteins composed of hundreds or thousands of amino acids, sequence assembly proved to be a combinatorial nightmare. To address this challenge Margaret Dayhoff and colleagues developed COMPROTEIN: a computer program to aid primary protein structure determination (Dayhoff and Ledley, 1962; Gauthier *et al.*, 2019). COMPROTEIN was one of the first programs described as bioinformatics software and was coded entirely on punch-cards.

As the sequences of more proteins were determined and technology advanced, computational biology expanded from its use in determining protein primary structure to predicting tertiary structure, the three-dimensional (3D) shape of a protein. Since its establishment in 1994, the Critical Assessment of protein Structure Prediction (CASP) initiative has held a prediction experiment every two years where dozens of groups across the globe implement computational methods with the goal of predicting target protein 3D structure from sequence (Moult *et al.*, 1995).

Beyond protein studies, computational biology has become a fundamental component to the analysis of nucleic acids. The increase in computational biology pioneers and methods was driven largely by the publication of the human genome at the start of the 21st century and the advancements to next generation sequencing technologies which followed (Craig Venter *et al.*, 2001; Lander *et al.*, 2001). As sequencing costs decreased and data increased a need emerged for the computational biologist, capable of drawing connections in such vast and complex data.

Today, computational biology describes a field fundamental to the study of proteomics, transcriptomics, genomics, and epigenomics. From the processing of billions of paired-end DNA sequences to machine learning strategies for prediction of protein binding affinity, this thesis describes in detail the development of computational methods, models, and algorithms, to explain biological phenomena.

Genome structure

In the field of biology and beyond, the ability to visually represent what is hidden from the naked eye has been key to understanding how our world works.

The discovery that DNA forms a 3D double helix structure led to many critical scientific breakthroughs over the last sixty-seven years (James Watson and Crick, 1953). The knowledge that a single strand of the double helix can act as a template to code for a partner strand allowed us to answer the question of how DNA is replicated from cell to cell, and how inherited traits can be passed

down from generation to generation (J Watson and Crick, 1953; Meselson and Stahl, 1958). The central dogma of molecular biology, that information is transferred from DNA and RNA to proteins but information cannot be transferred from a protein to DNA, was also predicated upon solving the 3D structure of DNA (F. H. Crick, 1958). Although the atomic structure of DNA has been known for some time, the larger picture illustrating how exactly the double helix is folded inside the cell nucleus still remains a puzzle.

Only within the last decade, with the aid of deep sequencing technologies, have we been able to begin unraveling the full 3D architecture of DNA in the context of how markedly long molecules of double helix DNA spatially fit inside the cell nucleus. The sheer size of a genome can make this question difficult to answer. The human genome is roughly three billion base pairs of DNA in length. Fitting all of that DNA inside a nucleus roughly 10 μm in diameter is analogous to fitting 30 miles of string inside a basketball. The nucleus is packed with DNA; however, we now know that this packing is not an arbitrary random entanglement of the DNA polymer. From human to mouse to fly and even in single cell organisms, the 3D architecture of DNA inside the nucleus is highly organized in a manner to promote cell function (Lieberman-Aiden *et al.*, 2009; Duan *et al.*, 2010; Sexton *et al.*, 2012; Zhang *et al.*, 2012).

Although some general features of this 3D architecture are conserved across varying cell types and even across different species (Dixon *et al.*, 2012), changes in the 3D architecture of DNA have been associated with gene

regulation, cell type specificity, and cell fate (Dixon *et al.*, 2015; Tang *et al.*, 2015; Schmitt *et al.*, 2016). It follows that the mechanisms controlling the 3D architecture must be properly regulated and impairment of this architecture can lead to improper cell function and subsequent disease (Lupiáñez *et al.*, 2015).

A DNA molecule does not solely fold itself into its proper 3D conformation, but is aided by a myriad of DNA-binding proteins to form chromatin. In eukaryotes, the basic structural unit of chromatin is generated by the wrapping of 146 base pairs of DNA around eight histone proteins forming the nucleosome (Kornberg, 1974; Morse and Simpson, 1988). This wrapping of DNA into nucleosome structures not only helps to compact DNA molecules to fit inside the nucleus, but furthermore can act as method of gene regulation. Nucleosomes can inhibit transcription initiation and inversely, loss of nucleosomes can lead to transcriptional activation (Lorch, LaPointe and Kornberg, 1987; Han and Grunstein, 1988). Zooming out from the nucleosome level and examining full chromosomes, condensin complex proteins condense DNA to a further degree when preparing the cell for mitotic division (Hirano and Mitchison, 1994). The 3D architecture of chromatin inside the nucleus is never static, but instead a dynamic system that must replicate chromosomes followed by radical condensation and spatial chromosome segregation every cell cycle.

Visualization of mitotic cells through microscopy show chromosomes as condensed sausage like structures (Rieder and Khodjakov, 2003); however, the spatial occupancy of chromosomes during interphase is more difficult to resolve.

Early microbeam radiation experiments, fluorescence microscopy results, and more recent deep sequencing data have all supported the now unequivocal conclusion that individual chromosomes again occupy separate territories during interphase of the cell cycle (Cremer *et al.*, 1982; Bolzer *et al.*, 2005; Lieberman-Aiden *et al.*, 2009).

While the existence of chromosome territories is clear, how certain features of the chromosome relate to the chromosome's spatial localization is still under debate. For instance, there is evidence for both random and non-random proximity of homologous chromosomes in somatic cells dependent upon cell type and species (Henikoff, 1997; Bolzer *et al.*, 2005). In humans, chromosome size has been associated with nuclear positioning with smaller chromosomes primarily located in the interior and larger chromosomes residing around the nuclear periphery (Sun, Shen and Yokota, 2000). However, this positioning may not be a result of chromosome size, but instead due to the clustering of gene rich active regions of chromosomes versus inactive gene poor regions. In agreement with this notion, the similar sized human chromosomes, 18 and 19, were shown to occupy different nuclear neighborhoods in lymphoblastoid cells. Specifically, gene rich Chromosome 19 was positioned near the nuclear center while the gene poor Chromosome 18 was closer to the periphery (Boyle, 2001).

Keeping active and inactive regions of the genome spatially segregated makes logical sense. Just as a city may cluster its financial district into a space

separated from parks and housing to promote commerce, the cell may separate gene rich from gene poor regions to promote transcriptional activity (Figure 1.1).

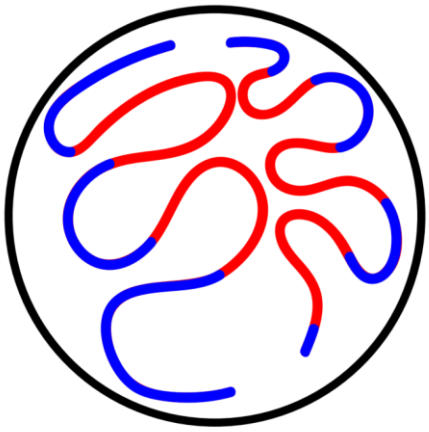


Figure 1.1 | Illustration of compartmentalization.
Nucleus (black) housing chromosomes with spatially separated active regions (red) and inactive regions (blue) of the genome.

Experimental evidence from the last couple of decades has made this hypothesis of nuclear compartmentalization undoubtable (Fraser and Bickmore, 2007; Lieberman-Aiden *et al.*, 2009). However, information regarding the stability of these compartments and the mechanisms governing compartmentalization is still lacking. In the next section, I review the tools and experiments used to identify compartments and the current hypotheses proposed regarding compartment formation and its properties.

Compartments

Fluorescence *in situ* hybridization (FISH) and microscopy techniques provided some of the first evidence for the spatial segregation of active and inactive chromatin domains into compartments. In FISH, hybridization of a fluorescent probe to a genomic region of interest facilitates visual localization of the genomic region within the nucleus. FISH studies in Drosophila showed that the eye color gene *brown* can relocate near inactive regions after an insertion of inactive chromatin in the *brown* gene itself. Silencing of wildtype *brown* due to its proximity to inactive chromatin provided further evidence supporting a functional role for compartmentalization (Dernburg *et al.*, 1996). Similar FISH experiments in mouse indicate that active genes migrate toward specific nuclear regions promoting transcription, which may act as ‘transcription factories’ (Osborne *et al.*, 2004; Shopland *et al.*, 2006). However, this model has been challenged by more recent studies which provide evidence for highly dynamic and transient transcription clusters (Cisse *et al.*, 2013; Furlong and Levine, 2018). While FISH studies laid the groundwork for visualizing interphase compartmentalization patterns in eukaryotic cells, they were typically restricted to investigating a single locus or only a handful of loci. Later chromosome conformation capture experiments extended these clues from individual loci to validating a genome wide phenomenon.

Chromosome conformation capture (3C) experiments and their derivatives allow for the detection of proximal DNA loci within a population of cells (Dekker *et al.*, 2002). In brief, these experiments apply the following three steps:

(1) Cells or isolated nuclei are crosslinked with formaldehyde, linking protein to protein and DNA to protein interactions throughout the nucleus. (2) DNA is digested with a restriction enzyme creating free ends of the cut DNA. (3) Free ends of cross-linked DNA fragments are ligated together.

The resulting DNA ligation products indicate a spatial proximity between the DNA loci on either end of the ligation junction. These loci can then be detected using specific primers in PCR reactions, microarrays, or through deep sequencing (Dekker *et al.*, 2002; Simonis *et al.*, 2006; Lieberman-Aiden *et al.*, 2009).

Chromosome conformation capture-on-chip (4C) extended the 3C method to interrogate the interaction profile of a single locus with the rest of the genome. This was accomplished by hybridizing PCR products from circular ligations on a microarray containing probes for loci scattered throughout the genome (Simonis *et al.*, 2006). 4C provided evidence that the active β-globin genes in fetal liver interact with other actively transcribed regions, and conversely the inactive β-globin genes in fetal brain interact with other transcriptionally silent regions (Simonis *et al.*, 2006). The advent of the 3C/4C methods along with advances in microarray technology allowed for validation of the compartmentalization features

seen in FISH by probing the interactions of one locus versus the rest of the genome. A few years later, a further derivative of the 3C technology, known as Hi-C, would harness deep sequencing technology to simultaneously interrogate all interactions between all genomic loci.

By adding a biotin incorporation step before ligation and subsequent pull down of ligation junctions followed by deep sequencing, the Hi-C method led to the production of the first chromatin contact map of the human genome at one megabase (Mb) resolution (Lieberman-Aiden *et al.*, 2009). Chromosomal territories reported for interphase cells by previous microscopy studies were well represented in the genome wide contact maps, as loci within each chromosome (intra-chromosomal) were found to interact with much greater frequency than loci on different chromosomes (inter-chromosomal). Perhaps the most interesting observations came from chromatin contacts maps generated for intra-chromosomal interactions. For example, examining the contact map of Chromosome 14, a striking checkerboard pattern emerged from the Hi-C data. Overlapping the checkerboard pattern with gene density, histone modification, and open chromatin tracks, it was evident the checkerboard pattern reflected the spatial preferences of compartmentalization. Dark squares of the checkerboard represented either preferential interactions between active genomic regions or the preferential interactions between inactive regions, while light squares represented the depletion of interactions between inactive and active regions. A principle component analysis on the Hi-C contact matrix was used to demarcate

boundaries between the spatially segregated active and inactive genomic regions, defining them as A and B compartments, respectively. (Lieberman-Aiden *et al.*, 2009).

Over the next decade, Hi-C contact maps have demonstrated that compartmentalization occurs across a wide array of human cell types (Schmitt *et al.*, 2016), in other mammals (Zhang *et al.*, 2012; Bonev *et al.*, 2017), in chicken (Gibcus *et al.*, 2018), in fly (Rowley *et al.*, 2017), and even in some plants (Dong *et al.*, 2017). Clearly, A/B compartmentalization of interphase nuclei is a conserved organizational principle (Jost, Bertulat and Cardoso, 2012) for many eukaryotes. Next, I will review the genetic and epigenetic properties of A/B compartment regions and introduce the current theoretical mechanisms proposed for compartment formation.

While direct evidence for genome wide segregation of active and inactive chromatin during interphase was not available until recently, many results from previous chromatin studies fall in line with this concept of compartmentalization and aid in distinguishing A compartments from their B counterparts. As early as 1929, Emil Heitz coined the terms *heterochromatin* and *euchromatin* to describe chromatin that appeared darker and lighter, respectively, on mitotic chromosomes stained with DNA dyes (Jost, Bertulat and Cardoso, 2012). Sedimentation experiments provided further evidence for two structurally distinct types of chromatin: condensed/closed heterochromatin and decondensed/open euchromatin (Gilbert *et al.*, 2004). These two structural states of chromatin are

also in line with gene positioning along chromosomes with euchromatin enriched in genes and heterochromatin comprised primarily of gene deserts. Accordingly, A compartments are predominately composed of euchromatic regions and B compartments are predominately heterochromatin. It is important to note that the A/B compartment designations used here are defined strictly by the spatial segregation illustrated in Hi-C contact maps and although A compartments typically contain active genes there are exceptions to the rule (Vieux-Rochas *et al.*, 2015; Bonev *et al.*, 2017).

While genetic sequence alone may not be enough to accurately predict compartments regions, correlations can be made between sequence composition and A/B compartment designation. B compartments contain many transposable elements and are enriched in adenine (A) and thymine (T) nucleotides, while A compartments are generally more gene dense and enriched in guanine (G) and cytosine (C) (Slotkin and Martienssen, 2007; Lieberman-Aiden *et al.*, 2009; Imakaev *et al.*, 2012). Beyond sequence characteristics, epigenetic marks have shown high correlations with A/B compartments and provide further clues into compartment formation and function (Lieberman-Aiden *et al.*, 2009; Allis and Jenuwein, 2016; Allshire and Madhani, 2018).

Chemical modifications to the lysines of nucleosome histone tails can regulate gene expression by changing the local chromatin environment (Bannister and Kouzarides, 2011). Typically, acetylation of histone tails is associated with active transcription and an open chromatin environment, while

methylation is associated with either transcriptional activation or repression dependent upon the methylation site. For example, the methylated histone marks, H3K4me3 and H3K36me3, are located near active regions (Bannister *et al.*, 2005; Barski *et al.*, 2007), while the marks, H3K27me3 and H3K9me3, are found in repressed regions (Becker, Nicetto and Zaret, 2016; Wiles and Selker, 2017). In some of the first 1 and 0.5 Mb scale contact maps, compartments were shown to correlate with active and repressive histone marks (Lieberman-Aiden *et al.*, 2009; Kalhor *et al.*, 2012). As technology increased and sequencing costs decreased, higher resolution Hi-C maps emerged further refining the compartmentalization phenomenon and its relation to the epigenetic chromatin landscape (Rao *et al.*, 2014; Rowley *et al.*, 2017).

Upon examining kilobase resolution contact maps comprised of billions of Hi-C contacts in human lymphoblastoid cells, spatial preferences appeared within and across the original A/B compartments (Rao *et al.*, 2014). The positions of these subcompartments (A1, A2, B1, B2, and B3) were found to correlate with specific histone modifications and proximity to different sub-nuclear structures revealing a higher order level of 3D chromatin organization.

A1 and A2 subcompartments were found to be enriched in the original A compartment and are both correlated with active histone markers such as H3K27ac, H3K9ac, and H3K36me3 (Rao *et al.*, 2014). However, in contrast to A2, the A1 subcompartment was found to specifically co-localize with nuclear speckles, which are hubs for pre-mRNA splicing factors (Chen *et al.*, 2018). The

co-localization of A1 chromatin with nuclear speckles would in effect bring actively transcribed regions to the machinery which processes transcribed products for nuclear export (Misteli, Cáceres and Spector, 1997), or conversely speckles may form at active genes to promote efficient processing.

B1, B2, and B3 subcompartments all correlate with inactive heterochromatic regions of the genome. B2 and B3 subcompartments are both depleted of active histone marks and enriched for lamin A/C, a protein localized at the nuclear periphery (Rao *et al.*, 2014). In contrast to B3, the B2 subcompartment was found to additionally and specifically co-localize with the nucleolus, where ribosome biogenesis occurs (Boisvert *et al.*, 2007). Unique from all other subcompartments, the B1 subcompartment is specifically enriched in the repressive histone mark H3K27me3 (Rao *et al.*, 2014). Polycomb group proteins trimethylate H3K27 to silence genes, a mechanism that is essential for proper vertebrate and invertebrate development (Aranda, Mas and Croce, 2015). Genomic loci bound by polycomb are shown to cluster inside the nucleus into visible polycomb bodies (Saurin *et al.*, 1998), validating preferential B1 interactions and suggesting a functional role for their subcompartmentalization.

Low (Mb) resolution interaction maps illustrated the spatial segregation of the genome into two compartments highly correlated with active and inactive regions. Later, higher (kb) resolution maps revealed chromatin state dictates a finer level of compartmentalization by the co-localization of regions with similar epigenetic marks or co-localization of regions interacting with specific sub-

nuclear structures. While spatial compartmentalization of the genome is evident, the mechanisms driving this compartmentalization remain ambiguous.

Evidence suggests transcription may be a driver for compartmentalization. Specifically, experiments inhibiting transcription have led to a reduction in compartmentalization (Rowley *et al.*, 2017). However, it is still debatable whether such reductions are a consequence of inhibiting transcription processes such as initiation and elongation or a consequence of loss in RNA polymerase II (Pol II) binding to DNA (Rowley and Corces, 2018).

Applying theories from polymer physics has led to some notable insights into compartment formation. Modeling chromosomes as polymers, a chain of elementary units called monomers, simplifies the complex conformational structure of chromatin. Specifically, the conformation a polymer can adopt relies on three characteristics: (1) flexibility of the chain, (2) interactions between monomers on the chain and (3) interactions with the polymer's surroundings (Rubinstein and Colby, 2003). A chromosome with alternating blocks of A and B compartments where each block may vary in size, can be modeled as a block copolymer. In a block copolymer, each block is comprised of consecutive A or consecutive B monomers. Simulations using block copolymers are capable of reproducing the checkerboard pattern displayed in Hi-C interaction maps (Haddad, Jost and Vaillant, 2017; Falk *et al.*, 2019). Importantly, such studies can provide estimations of attractive forces between A/B monomer types which aid in solving the puzzle of compartment formation. In contradiction to results that

support transcription driving compartment formation, a recent study utilizing block copolymer modeling to investigate compartmentalization in mouse described euchromatic A-A interactions as weak and more or less dispensable for compartmentalization. This study rather suggests that strong heterochromatic B-B interactions are responsible for genome compartmentalization via a phase separation mechanism (Falk *et al.*, 2019).

Similar to the phase separation of oil and water, recent studies have suggested that compartmentalization may be due to phase separation of compartment regions with the surrounding nucleoplasm (Hnisz *et al.*, 2017; Larson *et al.*, 2017; Strom *et al.*, 2017). Liquid-liquid phase separation may occur through the oligomerization of soluble multivalent proteins (Erdel and Rippe, 2018). Self-associating oligomerization of these proteins can lead to a liquid-like droplet forming a phase separate from its surroundings. When these self-associating proteins bind to chromatin they may induce compartmentalization by incorporating one compartment into the liquid-like droplet phase while excluding the unbound chromatin of the opposite compartment. For example, the multivalent protein HP1α binds the H3K9me3 mark enriched in heterochromatin. *In vitro* binding of HP1α with DNA can form liquid-like droplets (Larson *et al.*, 2017). Furthermore, DNA curtain experiments reveal HP1α proteins are capable of compacting DNA (Larson *et al.*, 2017). Accordingly, B compartmentalization may be due in part to the liquid-liquid phase separation of heterochromatic regions bound by HP1α.

Similar to blocking transcription, perturbing other elements intrinsic to a functional cell nucleus has led to changes in compartmentalization. For instance, chromatin is known to form loops spanning more than a megabase in genomic distance (Rao *et al.*, 2014). Degrading proteins vital to loop formation leads to an elimination of loops and a subsequent enhancement of compartmentalization (Rao *et al.*, 2017). These results suggest chromatin loops can act antagonistically to the compartmentalization phenomenon bringing A and B loci in closer proximity than attractive and repulsive forces between them would ordinarily allow.

In terms of understanding how compartments are formed and maintained, there is still much work to be done. Chapter II of this thesis focuses on answering the unknowns of compartmentalization and leads to the discovery of new principles and characteristics pertaining to compartmentalization of interphase nuclei. In the next section, I further introduce another 3D structure, the chromatin loop, and expand on its role in gene regulation, specifically in mediating promoter-enhancer interactions (Figure 1.2).

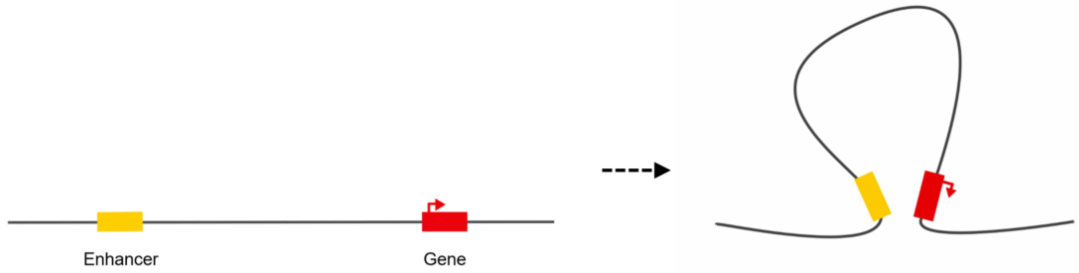


Figure 1.2 | Illustration of chromatin loop mediated promoter-enhancer interaction.
Chromatin loop formation allows a distal enhancer region (yellow) to regulate its target gene (red) by placing the enhancer region in close physical proximity to its target gene promoter.

Promoter-Enhancer Interactions

In eukaryotes, gene activation is defined by its transcription conducted by the Pol II multiprotein enzyme. As a prerequisite to gene activation, Pol II must first bind a promoter region proximal to the transcription start site (TSS) of a gene. Binding of Pol II and other transcription factors (TFs) to promoter regions leads to the initiation of transcription. Hence, occupancy of Pol II and TFs at promoters regulates gene activity (Andersson and Sandelin, 2020). In a similar fashion, enhancer regions distal from a gene in terms of linear sequence can activate a gene through recruitment of regulatory TFs. Enhancer function was identified as early as 1981 by transfection assays which showed a 72 bp repeat of viral DNA could increase activity of the β-globin gene across a distance of over 3 kb (Banerji, Rusconi and Schaffner, 1981). Later studies provided evidence

enhancers can act on gene targets as far as a megabase away (Carter *et al.*, 2002; Lettice *et al.*, 2003). To reconcile how enhancers could regulate such distant targets, the 3D structure of chromatin at these regions was investigated.

3C studies at the β-globin locus revealed chromatin is capable of forming a loop structure linking a distal enhancer to its target promoter and activating β-globin gene expression (Tolhuis *et al.*, 2002; Deng *et al.*, 2012). Formation of a chromatin loop could reconcile how distal enhancers can regulate their target genes. However, the relationship between chromatin loop structures and transcriptional activity is a topic still under much deliberation.

Identifying loop machinery and characterizing their modes of action in interphase nuclei has provided some insight into the role of chromatin loops in gene regulation. Two key components, CCCTC binding factor (CTCF) and the cohesin complex, have proven to be critical for the formation of many chromatin loop structures. CTCF is a highly conserved transcription factor capable of forming homodimers and known to recognize a sequence specific DNA binding motif (Nichols and Corces, 2015). The cohesin complex is a ring shaped protein complex known to regulate the separation of sister chromatids during cell division (Makrantoni and Marston, 2018). Together, CTCF and cohesin may mediate chromatin loop formation via a loop extrusion mechanism (Fudenberg *et al.*, 2016). In this model, binding of cohesin to chromatin leads to translocation of cohesin along the chromatin fiber causing extrusion of a small chromatin loop away from or out of the cohesin ring. Extrusion continues until cohesin is stalled

by CTCF anchors bound to chromatin at opposite loop ends. Hence, in this model cohesin would bring the CTCF anchors of a loop close together in 3D space and loop size would be dependent on the 1D genomic distance between CTCF anchor sites. In support of the loop extrusion model, recent *in vitro* microscopy experiments show human cohesin complexes can form loops at up to 2.1 kb per second likely in a pseudo-topological or non-topological manner (Davidson *et al.*, 2019; Kim *et al.*, 2019).

Although CTCF/cohesin mediated loops may appear as a promising mechanistic candidate for the linking of promoters with enhancers, the relationship between CTCF/cohesin mediated loops and promoter-enhancer interactions is not as clear cut. Upon identification of loop structures in a high resolution Hi-C map of the human cell line GM12878, a majority of loop anchors are bound by CTCF (~86%) and cohesin (~86%) (Rao *et al.*, 2014). However, far fewer of these loops (~30%) actually linked together known promoters and enhancers. It follows, the functional role of CTCF/cohesin loops likely extends beyond the direct facilitation of promoter-enhancer interactions.

ChIA-PET experiments provide 3D chromatin interaction information between DNA loci bound by a protein of interest (Fullwood *et al.*, 2009). ChIA-PET experiments targeting CTCF have shown CTCF loop anchors are enriched for active epigenomic markers, Pol II occupancy, and TSSs of genes (Tang *et al.*, 2015). Furthermore, constitutively expressed genes were found to be enriched at CTCF/cohesin loop anchors, while tissue specific genes primarily resided within

the body of loops suggesting a functional role for loops in mediating cell type specificity (Tang *et al.*, 2015). The binding motif of CTCF is non-palindromic and thus directional. The vast majority of CTCF loops identified in the human cell line GM12878 contained anchors with CTCF binding motifs in a convergent orientation (Rao *et al.*, 2014). Surprisingly, the directionality of transcription for genes with a TSS located in a CTCF loop anchor is in harmony with the CTCF motif orientation, again suggesting a regulatory role for CTCF/cohesin loops (Tang *et al.*, 2015).

Beyond associations between CTCF/cohesin loops and gene positioning or activity, mutagenesis studies have shown changes in gene expression can occur upon induced loop disruptions. Regulation of the *Shh* gene in mouse by the ZRS enhancer (located ~850 kb away from the *Shh* promoter) is vital to proper limb development. Disrupting CTCF/cohesin loop formation by inversion of a region encompassing the ZRS enhancer leads to downregulation of *Shh* in limb buds and malformation of limb structures (Symmons *et al.*, 2016). Interestingly, limb formation is partially rescued upon placing the inverted ZRS region in closer genomic distance to *Shh* (Symmons *et al.*, 2016). Similarly, inversion and deletion experiments disrupting CTCF loops at the *EPHA4* locus in mice causes ectopic interactions and expression patterns resulting in pathogenic phenotypes (Lupiáñez *et al.*, 2015). Hence, CTCF/cohesin loops may function to provide insulated neighborhoods within which dynamic movement of chromatin can link promoters and enhancers with a higher probability of interaction

occurring based on linear genomic distance. In further support of this model, inversion of CTCF binding-sites away from the convergent orientation seen in loop formations can result in an altered 3D architecture at loop loci, loss of contacts between promoters and enhancers, and changes in gene expression (Guo *et al.*, 2015).

To further dissect the functionality of CTCF/cohesin mediated loops, recent studies have utilized the auxin-inducible degron system to degrade CTCF or a specific subunit of the cohesin complex (Nora *et al.*, 2017; Rao *et al.*, 2017). Depletion of CTCF or a subunit of the cohesin complex led to a loss of loop structures present in 3C based conformation maps (Nora *et al.*, 2017; Rao *et al.*, 2017).

Degradation of CTCF was shown to result in the differential expression of nearly 5,000 genes in mouse which was roughly evenly split between upregulated and downregulated genes (Nora *et al.*, 2017). Downregulated genes due to CTCF degradation were enriched in CTCF near promoters which suggested a role for CTCF/cohesin loop formations in promoting the activity of these genes. However, few of these genes overlap with loop structures that would connect their promoter to an active enhancer region. Hence, CTCF degradation is likely not directly disrupting promoter-enhancer loops formations. Instead of CTCF functioning to actively link promoter with enhancer, its role in gene regulation may be more often inhibitory. Indeed, when analyzing upregulated genes after CTCF degradation, a higher majority of upregulated

genes were genomically closer to active enhancers than down-or non-regulated genes (Nora *et al.*, 2017). Such a result would support CTCF/cohesin loops as forming insulated neighborhoods preventing ectopic interactions between promoters and enhancers which reside in different loop domains.

Contrary to a model where CTCF/cohesin loops function to prevent ectopic promoter-enhancer interactions, degradation of a cohesin subunit resulted in a limited number of gene activations. Only roughly 1% of the unexpressed genes prior to cohesin subunit depletion were activated after subunit depletion (Rao *et al.*, 2017). Upon examining expressed genes before cohesin subunit depletion, the large majority of these genes presented similar expression levels after cohesin subunit depletion and subsequent loss in loop formations (Rao *et al.*, 2017). Hence, while a small subset of genes may be affected by a loss in CTCF/cohesin mediated loops, results from degradation studies imply CTCF/cohesin looping structures may play a limited role in gene regulation.

Beyond CTCF and cohesin, promoter-enhancer interactions may be mediated by other factors. The transcription factor Yin Yang 1 (YY1) is known to bind both enhancer and promoter elements and can form dimers. Deletion of YY1 binding sites or depletion of YY1 has been shown to disrupt chromatin loop formations and alter gene expression (Weintraub *et al.*, 2017). Another factor with evidence for facilitating promoter-enhancer interactions is the transcriptional coactivator mediator. Mediator forms a complex with cohesin capable of

connecting distal genomic loci. 3C studies show occupancy of mediator and cohesin at promoter and enhancer elements predicts chromatin loop formation linking enhancers to the promoters of *Nanog*, *Oct4*, *Phc1*, and *Lefty1* in mouse (Kagey *et al.*, 2010). Aside from protein complexes, there is also evidence non-coding RNAs may play a role in stabilizing promoter-enhancer interactions (Lai *et al.*, 2013; Hsieh *et al.*, 2014). Many questions still remain regarding how these factors coincide with CTCF/cohesin loops and their effect on transcription.

Identification of promoter-enhancer interactions is fundamental to the mapping of regulatory pathways and critical to the dissection of abnormal regulation in diseased states. High resolution Hi-C interaction maps are a powerful resource for the detection of promoter-enhancer looping interactions (Rao *et al.*, 2014). A chromatin loop in a Hi-C map appears as a ‘dot’ of significant interaction intensity above its background neighborhood. These dots represent the high probability of interactions between anchors of loop loci, commonly CTCF sites brought together in close proximity by the cohesin complex in the loop extrusion model (Fudenberg *et al.*, 2016). From a computational standpoint, genome wide identification of these loci presents a significant challenge. Dots representing chromatin loops appear at a resolution of ~10 kb in a Hi-C interaction map. At this resolution, investigating every 3D interaction in a map of the human genome would require the investigation of over a hundred billion pairs of loci.

Several computational methods have been developed for the identification of looping interactions in Hi-C interaction maps. These methods typically come in two flavors: (1) significant enrichment of looping interaction above global background (all interactions at equivalent genomic distance), or significant enrichment of looping interaction above a local background (interactions occurring in an ~50 kb radius surrounding looping interaction). In both methods, significant enrichment is determined using binomial or Poisson statistics and resultant *p*-values are corrected for multiple testing (Ay, Bailey and Noble, 2014; Rao *et al.*, 2014; Mifsud *et al.*, 2017). While these methods have been successful in identification of loops and promoter-enhancer contacts, the reproducibility of significant interactions between replicates was shown to be low for all methods (Forcato *et al.*, 2017). Furthermore, the number of looping interactions identified was shown to be dependent upon sequencing depth, which could convolute results made from samples of differing coverage (Forcato *et al.*, 2017). It is also important to note that identified significant interactions represent only 3D looping interactions which may or may not link a promoter and enhancer. As Hi-C interaction maps represent a population average, some dynamic promoter-enhancer interactions may also fail to be identified as significant via such loop calling methods. It follows, while 3D conformations assays provide evidence for promoter-enhancer interactions, they may be most suitable when investigating candidate promoter and enhancer elements identified previously by other genomic assays.

As less than 2% of the human genome is comprised of protein coding sequences, annotating non-coding enhancer elements and their target promoters could lead to important discoveries in disease pathologies, particularly when disease causing genetic variations lie outside of coding genes (Alexander *et al.*, 2010). It follows, the annotation of enhancer elements is a primary goal for many scientific groups including the Encyclopedia of DNA elements (ENCODE) consortium, a community consisting of dozens of research groups specializing in various fields of genomics (Dunham *et al.*, 2012). Numerous computational methods have utilized assays that identify open chromatin regions (e.g. DNase-seq) or specific histone marks associated with enhancer activity (e.g. H3K27ac or H3K4me1 ChIP-seq) to identify enhancer elements (Ernst *et al.*, 2011; Hoffman *et al.*, 2012; Thurman *et al.*, 2012; Rajagopal *et al.*, 2013; Kundaje *et al.*, 2015). In some cases, enhancers do not always activate their nearest gene in terms of linear sequence space complicating the mapping of enhancers to their gene targets (Lettice *et al.*, 2003). This difficulty in linking promoters with enhancers has been resolved via chromosome conformation assays similar to methods described earlier (Li *et al.*, 2012; Mifsud *et al.*, 2015) or by correlating genomic and epigenomic signals at enhancers and promoters across a range of varying biosamples (Ernst *et al.*, 2011; Thurman *et al.*, 2012). Such efforts have led to lists of candidate regulatory enhancer elements in a variety of different tissue types. However, while genetic or epigenetic marks along with 3D proximity information may be predictive of enhancer activity such assays do not provide a

ground truth for enhancer identification. To validate an enhancer one must show its presence definitively leads to the expression of its target gene.

The canonical method for validating enhancer activity is to clone an enhancer sequence upstream of a minimal promoter of a reporter gene (e.g. *LacZ*) and analyze reporter gene expression. Using such methods *in vivo*, the VISTA enhancer browser contains results from the interrogation of hundreds of enhancer sequences (Visel *et al.*, 2007). Here, the enhancer reporter construct is microinjected into a fertilized mouse egg followed by implantation into a female mouse. *LacZ* staining of the harvested mouse embryo reveals tissue specific enhancer activity (Visel *et al.*, 2007). A drawback to such transgenic mouse assays, is that transfection of the enhancer reporter construct leads to random integration in the mouse genome. Hence, while results from reporter assays validate enhancer function they do not take into account the enhancer's native chromatin context.

The CRISPR/Cas9 genome editing system can be utilized to delete genomic sequences inside the nucleus of the cell (Cong *et al.*, 2013; Yao *et al.*, 2014; Won *et al.*, 2016). This is accomplished by the design of single-guide RNAs (sgRNAs) complementary to regions flanking the DNA sequence selected for deletion. The binding of sgRNA to a catalytically active Cas9 protein guides Cas9 to a site flanking the selected DNA sequence for subsequent Cas9 mediated cleavage. To validate enhancer function, this method can be adapted to deletion of a candidate enhancer element. Quantitative measuring of gene

expression for the candidate enhancer target before and after CRISPR/Cas9 mediated deletion of the enhancer sequence provides validation of enhancer function within the enhancer's original chromatin environment (Yao *et al.*, 2014; Won *et al.*, 2016).

Even when applying current methods and technologies, identification and validation of promoter-enhancer interactions remains a significant challenge. Chapter III of this thesis utilizes Hi-C experiments to assay the 3D genome architecture in human glial cells, neural progenitor cells and neurons. The resultant Hi-C interaction maps are used to predict cell-type specific promoter-enhancer interactions occurring at or near schizophrenia risk variants. Upon identification of risk variants in close spatial proximity to potential gene targets, CRISPR/Cas9 strategies are applied to test the regulatory functions of risk variant loci.

Schizophrenia, Neurons, and 3D organization

Symptoms of schizophrenia include hallucinations, delusions, disorganized speech, and social withdrawal, along with a broad set of cognitive dysfunctions (Kahn *et al.*, 2015). Schizophrenia is not considered a rare disorder. Studies of disease prevalence estimate about seven individuals per one thousand will develop schizophrenia during their lifetime (McGrath *et al.*, 2008). The disorder has also been shown to have an effect on average life span. Based on the standardized mortality ratio, individuals diagnosed with schizophrenia

have a two to threefold increased risk of dying with suicide as a main contributing factor (McGrath *et al.*, 2008).

Current treatments of schizophrenia rely on the action of antipsychotic drugs such as chlorpromazine or clozapine which block receptors in the dopamine pathway (Kane and Correll, 2010; Miyamoto *et al.*, 2012). While such antipsychotic drugs help to manage patient symptoms, in many cases, patients symptoms are either resistant to treatment or successful symptom treatment is accompanied by adverse side effects (Lally and MacCabe, 2015). It follows, specialized treatments taking into account patient specific characteristics may result in healthier patient outcomes. However, the lack of knowledge in regard to the etiology of schizophrenia presents a major challenge to the advent of improved or specialized therapeutics to treat the disorder.

While the etiology of schizophrenia is largely unknown, evidence suggests that along with environmental factors the disorder carries a heritable genetic component. In a study of over 30,000 twin pairs born in Denmark with 448 twin pairs affected by schizophrenia, heritability of the disorder was estimated to be 79% in accordance with previous studies (Sullivan, Kendler and Neale, 2003; Hilker *et al.*, 2018). Both rare and common allele variants have been implicated in the pathology of schizophrenia (McClellan, Susser and King, 2007; Purcell *et al.*, 2009). To further explore this landscape of genetic variants, large scale genome-wide association studies (GWAS) have been employed to search for SNPs associated with schizophrenia. A recent GWAS encompassing genotypes

from 36,989 schizophrenia cases and 113,075 controls reported 108 risk loci harboring risk variants significantly associated with the disorder (Ripke *et al.*, 2014). Expanding upon this study, Pardiñas and colleagues performed GWAS analysis on an additional 11,260 cases and 24,542 controls, which resulted in the identification of 50 additional novel risk loci (Pardiñas *et al.*, 2018). When risk loci were overlapped with epigenetic marks for active enhancers, risk SNPs were found to be enriched in enhancers active in brain (Ripke *et al.*, 2014).

The enrichment of schizophrenia risk SNPs at active enhancers in brain suggests disruption of regulatory mechanisms may underlie disorder pathology. Such a model would assign a possibility of risk to noncoding variants which, unlike coding variants, are incapable of directly altering the amino acid composition of a protein product. Supporting this model, schizophrenia risk variants were shown to be enriched for alleles that affect gene expression and lie within promoter or enhancers (Roussos *et al.*, 2014). Expression quantitative trait loci (eQTLs) are defined as loci containing variants which alter gene expression (Nica and Dermitzakis, 2013). A schizophrenia associated eQTL within the intron of its target gene was found to be in closer spatial proximity to its target gene TSS compared with other sequences closer in genomic space (Roussos *et al.*, 2014). Furthermore, this eQTL variant locus resided in an enhancer region suggesting potential promoter-enhancer looping mechanisms may be at play.

It is plausible that eQTL variants located in enhancer regions could alter their gene target’s expression by disrupting binding of transcription factors to

enhancer loci, disrupting the local chromatin state of the enhancer neighborhood, or by disrupting promoter-enhancer loop formation, none of which are necessarily mutually exclusive events. In further support of risk variants modulating gene expression, complex trait associated SNPs are significantly more likely to be eQTLs than minor-allele-frequency matched SNPs (Nicolae *et al.*, 2010). Additionally, when overlapping GWAS identified schizophrenia risk loci with an independent eQTL dataset generated from RNA sequencing of dorsolateral prefrontal cortex from both schizophrenia and control cases, results indicated ~20% of schizophrenia risk loci have variants that could contribute to altered gene expression (Fromer *et al.*, 2016).

Genetic regulatory pathways contributing to schizophrenia pathology may have cell type specific characteristics. Neurons and glial cells make up two broad classes of cells in the nervous system. Neurons are electrically excitable capable of relaying information through electrical impulses, while glial cells provide mechanical and metabolic support for neurons among other functions (Squire *et al.*, 2008). Recent studies have shown a preferential link between schizophrenia risk architecture and the neuron cell type (Genovese *et al.*, 2016; Skene *et al.*, 2018). Correlations between cell type specificity and enrichment for schizophrenia SNP heritability across genes, associated neurons with schizophrenia risk architecture over embryonic, progenitor or glial cells (Skene *et al.*, 2018). Similarly, an analysis of ultra-rare variants, reported individuals with schizophrenia were enriched for ultra-rare variants in genes that were neuronally

expressed (Genovese *et al.*, 2016). It follows that an analysis of cell type specific gene regulation systems may lead to pivotal observations concerning schizophrenia etiology.

Human induced pluripotent stem cell (hiPSC) technology has advanced the study of human disease mechanisms in vitro. Over a decade ago it was discovered adult mice fibroblasts can be reprogrammed into pluripotent stem cells by the introduction of four transcription factors: *Oct3/4*, *Sox2*, *c-Myc*, and *Klf4* (Takahashi and Yamanaka, 2006). Building upon this, Brennand and colleagues reprogrammed fibroblasts from schizophrenia patients into hiPSCs followed by subsequent differentiation into both neural progenitor cells (NPCs) and neurons (Brennand *et al.*, 2011). RNA analysis of control and schizophrenia hiPSC neurons showed differential expression in several genes characterizing potential regulatory mechanisms of the disorder (Brennand *et al.*, 2011). Further advancements in hiPSC technology has led to the differentiation of hiPSC derived NPCs by overexpression of the neuronal transcription factor, *NGN2*, producing near pure populations of postmitotic neurons capable of forming mature pre and post-synaptic formations (Ho *et al.*, 2016). Alternatively, hiPSC derived NPCs may also be differentiated into astrocytes of the glial lineage capable of responding to inflammatory stimulants and displaying phagocytic capacity (TCW *et al.*, 2017). Such methods provide new systems for testing the effects of differentiation and cell type specificity on regulatory gene networks and 3D genome organization.

Gene expression changes influence determination of cell type and cell fate (Dunham *et al.*, 2012; Wang *et al.*, 2018). However, much is still unknown regarding the mechanisms of cell type specific regulation and the related role of 3D genome architecture. Many 3D genome conformations are conserved across a variety of cell types, tissues, and even syntenic regions between species (Dixon *et al.*, 2012; Schmitt *et al.*, 2016). While 3D genome architecture is highly conserved, cell type specific changes in 3D architecture such as compartment switching and altered looping contacts have been reported (Rao *et al.*, 2014; Schmitt *et al.*, 2016; Won *et al.*, 2016). Such changes have also been described throughout the course of development. A/B compartment switching was observed upon differentiation of human embryonic stem cells (ESCs) into four other ESC-cell-derived lineages (Dixon *et al.*, 2015). Differentiation of mouse ESCs to neural stem cells resulted in a widespread gain of chromatin loop formations (Pękowska *et al.*, 2018). High resolution Hi-C maps of mouse ESCs, derived NPCs and further derived cortical neurons revealed dynamic chromatin looping formations occur near genes encoding neural transcription factors in a cell type specific manner which correlated with gene expression (Bonev *et al.*, 2017). Further experiments associating cell type specific regulatory programs with changing 3D architecture will help unravel the mechanisms driving cell fate and likely elucidate discoveries related to diseases driven by malfunctions in specific cell types.

Won and colleagues, characterized 3D genome architectures in human brain development in the context of associating the 3D genome with

schizophrenia risk variants (Won *et al.*, 2016). This study generated Hi-C maps of bulk tissue from the germinal zone (primarily neural progenitor cells) and cortical plate (primarily adult neurons). In Chapter III, we expand upon this study by applying hiPSC technology to produce high resolution Hi-C maps of hiPSC derived NPCs, glial cells and neurons. Such contact maps allowed for analyses revealing cell type specific chromatin conformation changes which we further studied in the context of schizophrenia risk.

In the next section, I move away from the 3D genome and introduce background relevant to Chapters IV and V, which describe computational methods for prediction of T cell receptor antigens via 3D structural information.

TCR-pMHC Interactions

In humans and other vertebrates, T cells and B cells cooperate in the adaptive immune system to protect the body from invading pathogens. Immunological protection is facilitated via recognition, elimination and memory of such pathogens (Owen *et al.*, 2013). While B cells produce antibodies which can directly bind pathogen antigens and signal for their destruction, most T cells recognize antigens displayed by specific proteins termed major histocompatibility complex (MHC) molecules, on the surface of antigen presenting cells (Zinkernagel and Doherty, 1974). Upon recognition of pathogenic antigens displayed on infected cells, T cells may initiate the direct killing of such infected cells or trigger other immune cells to attack the identified pathogen.

Before a T cell is equipped to successfully and specifically target infected cells, each T cell must be tested for competence during development. T cell precursors originating in the bone marrow travel to the thymus via the bloodstream to undergo critical developmental processes including: T cell receptor formation, positive selection and negative selection (Koch and Radtke, 2011).

T cells use T-cell receptors (TCRs) to recognize antigens displayed by MHC molecules on infected cells. The TCR is made up of two chains linked by disulfide bonds on the cell surface. These chain pairs are typically α and β, or γ and δ, which are encoded by the four genetic loci TRA, TRB, TRG, TRD, respectively. αβ T cells represent the dominant participants in adaptive immune functions and will be the focus of work presented in this thesis. In order to recognize a diverse and broad range of hazardous antigens, αβ T cells in the blood encompass a large repertoire of variable αβ TCRs, estimated to be on the order of 10^7 unique αβ TCRs per human (Arstila *et al.*, 1999). This large assortment of variable TCRs is accomplished by V(D)J recombination events during thymic development.

Within each TRA and TRB locus, multiple genes exist to encode single domains of the TCR chain. For example, in humans, the TRB locus contains over 60 genes encoding the variable domain of the TCR β chain (Lefranc, 2011). V(D)J recombination of these genes leads to the translation of a functional chain encoded by a random subset of these genes (Hozumi and Tonegawa, 1976).

Nucleotide deletions and insertions during the recombination process, and combinatorial αβ pairing, leads to further diversification of the TCR repertoire. Hence, before successful exit from the thymus, each T cell will express a distinct TCR, such that a population of mature T cells will be adequate for recognition of variable pathogenic antigens.

Once a T cell has formed a TCR, functional testing of the TCR takes place during positive selection. Here, further maturation of the T cell is dependent upon the affinity between the TCR and MHC molecules of an antigen presenting cell displaying self (non-foreign) antigens. If binding between TCR and self-antigen-MHC is not strong enough, the developing T cell will not be positively selected and instead undergo apoptosis (Boehmer and Kisielow, 1990; Starr, Jameson and Hogquist, 2003). Therefore, positive selection functions to ensure the population of mature T cells are capable of recognizing antigen-MHC targets. However, strong recognition of self-derived antigens is hazardous, as this could lead to an immune response and attack against healthy cells and tissues. To ensure the mature T cell engages the correct targets, negative selection deletes T cells with too strong an affinity for self-antigen-MHC complexes (Boehmer and Kisielow, 1990; Starr, Jameson and Hogquist, 2003). Finally, if a T cell survives both positive and negative selections in the thymus, it is then recruited to secondary lymphoid organs such as the lymph node to patrol for harmful invaders.

The first step toward an immune response occurs via recognition of antigen-MHC by a patrolling TCR. These MHC molecules are highly polymorphic with over 25,000 recorded human MHC alleles (Robinson *et al.*, 2020). The high degree of polymorphism has been attributed to evolutionary causes, whereby infectious diseases may drive MHC polymorphism (Radwan *et al.*, 2020). Antigen presenting MHC molecules are categorized into distinct classes. Class I MHC molecules are comprised of a single α chain noncovalently bound to a β_2-microglobulin protein. These molecules are expressed on the surface of the majority of nucleated cells in the body, and they display peptide fragments from endogenous antigens degraded in the cytosol. In contrast, class II MHC molecules are comprised of an α chain and a noncovalently associated β chain, are expressed in a limited set of cell types, and they display peptide fragments from exogenous antigens endocytosed into the cell (Owen *et al.*, 2013).

During thymic development T cells also commit to a specific cell lineage by expression of either cluster of differentiation 4 (CD4) or cluster of differentiation 8 (CD8) coreceptor molecules. These coreceptors maintain specific binding preferences, such that CD4+ T cells interact with MHC class II molecules and CD8+ T cells interact with MHC molecules of class I (Germain, 2002). Upon recognition of antigen, these T cell lineages have varying effector functions. CD4+ T cells (helper T cells) help to activate an immune response by stimulating responses in other cells; for example, helper T cells can stimulate B cells to produce antibodies to opsonize pathogenic targets (Crotty, 2015).

Alternatively, CD8+ T cells (cytotoxic T cells) are capable of directly killing infected cells; for example, cytotoxic T cells may release serine proteases into the target cell triggering caspase activation and cell death (Barry and Bleackley, 2002). Hence, the MHC class preference of T cell coreceptors also functions to guide T cell specific effector functions toward their appropriate target cell type. Given the larger setting of how T cells mount an immune response, the remainder of this introduction will focus on a specific step in the immune response, namely the interaction between TCR and peptide-MHC (pMHC) (Figure 1.3).

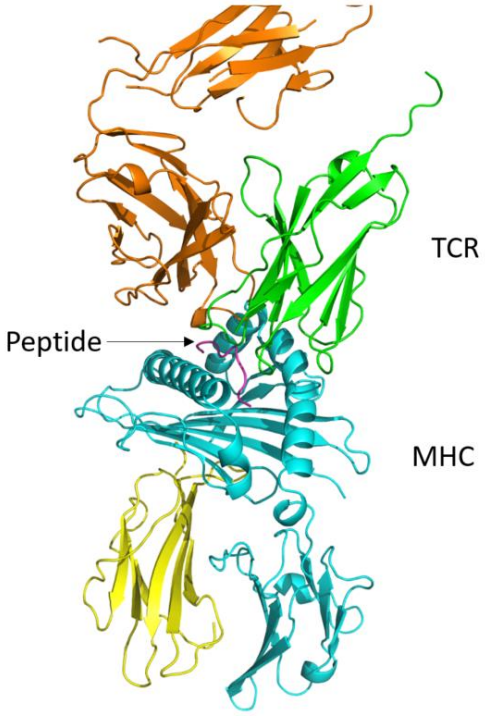


Figure 1.3 | Crystal structure of TCR-pMHC complex.
Crystal structure of the human A6 TCR (α chain:green, β chain: orange) recognizing the Tax peptide (magenta) and MHC allele HLA-A*02:01 (cyan). β_2-microglobulin is shown in yellow (Garboczi *et al.*, 1996).

In 1987, Bjorkman and colleagues successfully determined the first MHC structure, the human HLA-A2 molecule. The crystal structure revealed membrane distal domains formed a β-sheet platform topped by α-helices with a long groove between the helices, likely representing the location for bound peptide antigen (Strominger *et al.*, 1987). Determination of the class I MHC structure led to many hypotheses concerning the docking orientation of an interacting TCR. It was not until nearly a decade later that human and mouse TCR-pMHC complex structures would be solved, and these provided answers to previous hypotheses (Garboczi *et al.*, 1996; Garcia *et al.*, 1996). In both reports, the TCR was found to dock in a conserved diagonal orientation over the pMHC molecule surface (Figure 1.3).

The variable domain of each TCR α and β chain contains three complementarity-determining region (CDR) loops. CDR1 and CDR2 loops are encoded by the variable gene segment, while the CDR3 loops are encoded by the junction of variable and joining (or variable, diversity, and joining) gene segments, leading to a greater diversity in CDR3 (Hughes *et al.*, 2003). The solved TCR-pMHC structures showed that the CDR3 loops were primarily positioned to contact the peptide antigen, whereas the less diverse CDR1 and CDR2 loops were positioned closer to the MHC helices. Placing the more diverse TCR region over the peptide makes logical sense, as this could help improve discrimination against the vast space of foreign peptide antigens. Future studies

have shown a consensus of diagonal docking topologies for TCR interactions between class I and class II pMHC complexes, although atypical docking topologies do exist (Rossjohn *et al.*, 2015).

Due to negative selection, TCRs bind pMHC molecules with weak affinities and short half-lives in comparison with antibody-antigen affinities. Dissociation constants, K_Ds, from TCR-pMHC binding assays are in the micromolar range (K_D = 0.1 µM – 500 µM) and half-lives are on the order of seconds (Cole *et al.*, 2007, 2013; Bridgeman *et al.*, 2012).

Structural features of the TCR-pMHC interface have been associated with affinity changes. The α-helices which cradle the peptide in MHC class I molecules are closed at their ends in contrast to an open conformation in MHC class II structures (Brown *et al.*, 1993). This structural feature conforms to MHC class I molecules displaying peptides with lengths typically ranging from eight to ten amino acids, while class II molecules can display peptides with lengths beyond fourteen amino acids, protruding from open ends of the MHC binding groove (Bjorkman, 2015). The different antigen surfaces of pMHC classes are associated with a change in binding affinity by interacting TCRs. Specifically, TCR-pMHC affinities for class I pMHC molecules are stronger, on average, compared with class II pMHC molecules. Changes in affinity were ascribed to significantly greater on-rates for TCRs binding with class I pMHC compared with class II pMHC, while similar off-rates were reported (Cole *et al.*, 2007).

Aside from structural distinctions due to MHC class, weaker affinities for pMHC have also been noted for autoreactive TCRs binding self peptides compared with canonical TCRs engaging foreign peptides (Bridgeman *et al.*, 2012). This observation may be in concordance with thymic selection, whereby the weak binding of such autoreactive TCRs works to prevent harmful autoimmune responses.

Structural conformations of the TCR also play a role in determining binding affinity. Increased contacts made between the TCR CDR3 loops and peptide result in higher affinity TCR-pMHC interactions (Cole *et al.*, 2013). Furthermore, structure based design methods have led to the development of TCR mutants capable of binding pMHC up to 400 times more strongly than wildtype TCR (Haidar *et al.*, 2009; Zoete *et al.*, 2013; Malecek *et al.*, 2014; Pierce *et al.*, 2014). Collectively, such studies support the hypothesis that 3D interactions at the TCR-pMHC interface can be used to inform studies aimed at altering TCR-pMHC binding affinity.

Binding affinity for TCR-pMHC complexes can be determined using surface plasmon resonance (SPR) or isothermal titration calorimetry (ITC) experiments. In SPR experiments, binding affinity is quantified by optically measuring how fast light travels through a sensor surface (i.e. measuring the refractive index of the sensor surface). Here, the sensor surface makes up the floor of a flow cell, through which an aqueous solution can pass under continuous flow. To detect binding, a ligand (e.g. pMHC) is immobilized onto the sensor

surface, while an analyte (e.g. TCR) is flowed over the ligand bound surface. As the analyte binds to the ligand, an accumulation of binding complexes on the surface leads to a change in the refractive index of the surface. Analyzing these changes in refractive index over time and at different concentrations of analyte leads to determination of analyte-ligand binding affinity, K_D (Merwe, 2000). Similarly, ITC experiments can be used to derive K_D values for TCR-pMHC interactions by directly measuring changes in enthalpy. However, ITC is rarely used for analyzing TCR-pMHC binding due to the much larger amounts of protein required for the study in comparison to SPR (Miller *et al.*, 2007). Indeed, the vast majority of affinity values used for training and testing in Chapter IV were obtained from SPR experiments.

Several methods exist for the computational prediction of TCR-pMHC interactions. These methods can be categorized as either sequence based or structure based depending on the information used for prediction. Sequence based methods rely on machine learning algorithms that input sequence features of TCRs, peptides, and MHC molecules during model training. Such methods have been utilized to predict immunogenicity of a peptide or to predict reactivity of a TCR sequence (Tung *et al.*, 2011; Gielis *et al.*, 2019). Structural based methods rely on energy force fields and scoring functions, sometimes in combination with machine learning, to predict TCR-pMHC interactions. Such methods have similarly been utilized to predict immunogenicity of a peptide, but

also to predict TCR-pMHC binding affinity (Pierce and Weng, 2013; Lanzarotti, Marcatili and Nielsen, 2018; Schneidman-Duhovny *et al.*, 2018).

Although TCR-pMHC binding is a prerequisite for an immune response, binding affinity does not directly confer immunogenicity (Stone and Kranz, 2013). Extracellular TCR-pMHC binding causes a cascade of intracellular signaling events which eventually leads to activated transcription factors initiating immune response programs (Owen *et al.*, 2013). Experimental assays measuring T cell proliferation, cytotoxicity, or cytokine secretion are vital toward unraveling the larger picture of antigen immunogenicity.

In Chapters IV and V, I focus on the TCR-pMHC interaction in the context of antigen prediction via structural properties. Chapter IV introduces the development of a publicly available database, Altered TCR Ligand Affinities and Structures (ATLAS), linking 3D TCR-pMHC complexes with their experimentally measured binding affinities. Regression analyses performed with data from ATLAS were used to predict hundreds of TCR-pMHC binding energies. In Chapter V, we repurpose deep sequencing data to expand previous antigen predictions methods by investigating over two million modeled TCR-pMHC interactions.

CHAPTER II: COMPARTMENT-DEPENDENT CHROMATIN INTERACTION DYNAMICS REVEALED BY LIQUID CHROMATIN HI-C

Preface

This chapter is adapted from a manuscript currently under review at Nature Genetics authored by Houda Belaghzal, myself, Andrew D. Stephens, Denis L. Lafontaine, Sergey Venev, Zhiping Weng, John F. Marko, and Job Dekker titled: Compartment-dependent chromatin interaction dynamics revealed by liquid chromatin Hi-C.

The project was conceived by Job Dekker. 3C, 5C, Hi-C and liquid chromatin Hi-C and chromatin fractionation experiments were performed by Houda Belaghzal. Restriction digestion efficiency (DpnII-seq) experiments were performed by Denis L. Lafontaine. Micromechanical studies and their analysis was performed by Andrew D. Stephens. Data analysis was performed by myself and Houda Belaghzal. Specifically, I computationally processed liquid chromatin Hi-C sequencing datasets, developed the liquid-chromatin-Hi-C computational analysis toolkit to calculate LOS and $t_{1/2}$ stability metrics, developed the DpnII-seq computational pipeline, processed and analyzed fragment size assessment sequencing data, and performed the sub-nuclear structure analysis. Tools for liquid chromatin Hi-C analysis were contributed by Sergey Venev. Polymer scaling ideas relevant to data interpretation were provided by John F. Marko. The

paper was written and figures produced by Houda Belaghzal, myself, and Job Dekker with contributions from all coauthors.

Abstract

Nuclear compartmentalization of active and inactive chromatin is thought to occur through microphase separation mediated by interactions between loci of similar type. The nature and dynamics of these interactions are not known. We developed liquid chromatin Hi-C to map the stability of associations between loci. Before fixation and Hi-C, chromosomes are fragmented, removing the strong polymeric constraint to enable detection of intrinsic locus-locus interaction stabilities. Compartmentalization is stable when fragments are over 10-25 kb. Fragmenting chromatin into pieces smaller than 6 kb leads to gradual loss of genome organization. Lamin-associated domains are most stable, while interactions for speckle and polycomb-associated loci are more dynamic. Cohesin-mediated loops dissolve after fragmentation. Liquid chromatin Hi-C provides a genome-wide view of chromosome interaction dynamics.

Introduction

Genomic and imaging approaches are producing high-resolution descriptions of the conformation of chromosomes in cell populations, in single cells, across the cell cycle, and during development (Lieberman-Aiden *et al.*, 2009; Bickmore and Van Steensel, 2013; Nagano *et al.*, 2017; Naumova *et al.*, 2013; Nagano *et al.*, 2013; Rao *et al.*, 2014; Bonev and Cavalli, 2016; Wang *et al.*, 2016; Dekker and Mirny, 2016; Ramani *et al.*, 2017; Dekker *et al.*, 2017; Hug

et al., 2017; Nir *et al.*, 2018; Chen *et al.*, 2018; Gibcus *et al.*, 2018; Kaaij *et al.*, 2018). At the mega-base (Mb) scale chromosomes are compartmentalized and different types of chromosomal domains can be discerned. Hi-C interaction maps display a “plaid” pattern, which reflects the segregation of the genome in two major spatial compartments referred to as A and B compartments that correspond to open, active chromatin and closed, silent chromatin, respectively (Simonis *et al.*, 2006; Lieberman-Aiden *et al.*, 2009). High-resolution (kb) Hi-C maps allowed splitting compartments in 5 subtypes (A1, A2, B1, B2, and B3) that differ in interaction patterns and chromatin state (Rao *et al.*, 2014). At the scale of tens to hundreds of kb, topologically associating domains (TADs) were identified as domains separated by boundaries that are in many cases bound by CTCF. Higher resolution Hi-C (Rao *et al.*, 2014), ChIA-PET (Tang *et al.*, 2015), and 4C data (de Wit *et al.*, 2015; Guo *et al.*, 2015; Vietri Rudan *et al.*, 2015) showed that convergent CTCF sites at boundaries can engage in looping interactions.

Major questions revolve around the molecular and biophysical processes by which different aspects of chromosome conformation form. TADs and loops between CTCF sites form via loop extrusion cohesin (Riggs, 1990; Nasmyth, 2001; Alipour and Marko, 2012; Sanborn *et al.*, 2015; Fudenberg *et al.*, 2016, 2018; Rao *et al.*, 2017). Less is known about the processes that determine compartmentalization. Compartmentalization has been proposed to be the result of polymer phase separation driven by attractions between chromatin domains of

the same or similar status (Lieberman-Aiden *et al.*, 2009; Jost *et al.*, 2014; Di Pierro *et al.*, 2016; Michieletto, Orlandini and Marenduzzo, 2016; Erdel and Rippe, 2018; MacPherson, Beltran and Spakowitz, 2018; Nuebler *et al.*, 2018; Shi *et al.*, 2018; Falk *et al.*, 2019). Polymer models simulating such attractions can reproduce the plaid pattern characteristic of Hi-C interaction maps (Jost *et al.*, 2014; Di Pierro *et al.*, 2016; Nuebler *et al.*, 2018; Falk *et al.*, 2019).

Hi-C interaction maps are steady-state datasets and do not reveal the biophysical nature of the interactions that drive compartment formation or the dynamic mobility of loci within them. Live cell imaging studies have shown that loci are constrained in their motion and that there is variation in the dynamics and mobility of loci, e.g. euchromatic vs. heterochromatic loci and loci tethered to the nuclear periphery vs. loci located in the nuclear interior (Marshall *et al.*, 1997; Hediger *et al.*, 2002; Thakar, Gordon and Csink, 2006; Bronshtein *et al.*, 2009, 2015; Therizolsa *et al.*, 2010; Shinkai *et al.*, 2016; Nagashima *et al.*, 2019). Imaging-based studies have been instrumental in uncovering aspects of chromatin interactions and dynamics, but are limited in scale, i.e. only one or a few specific loci can be studied at one time. In addition, when whole genome dynamics are analyzed microscopically (e.g. (Zidovska, Weitz and Mitchison, 2013)), positions of specific sequences have not yet been determined. Therefore, new approaches are required to identify and quantify the molecular processes and biophysical forces involved in chromatin interactions and nuclear compartmentalization. Here we describe liquid chromatin Hi-C, a Hi-C variant

that quantifies the stability of chromosome conformation and chromatin interactions genome-wide.

Results

Measuring stability of chromatin interactions and nuclear compartmentalization

The formation of spatially segregated heterochromatic and euchromatic domains can be viewed as microphase separation of a polymer composed of different types of monomers (loci). A “block copolymer” is a polymer that contains a series of alternating blocks (e.g., A-type and B-type, or blocks of euchromatin and heterochromatin), each composed of multiple monomers (A monomers and B monomers; Figure 2.1, A). When As attract As and Bs attract Bs, such polymer can fold into spatially segregated domains of As and Bs (Figure 2.1, A, (de Gennes, 1979; Leibler, 1980; Matsen and Schick, 1994)). Applied to chromatin in vivo, microphase separation may underlie the formation of segregated compartments. The biophysical forces and interaction dynamics that determine chromosome compartmentalization are not known.

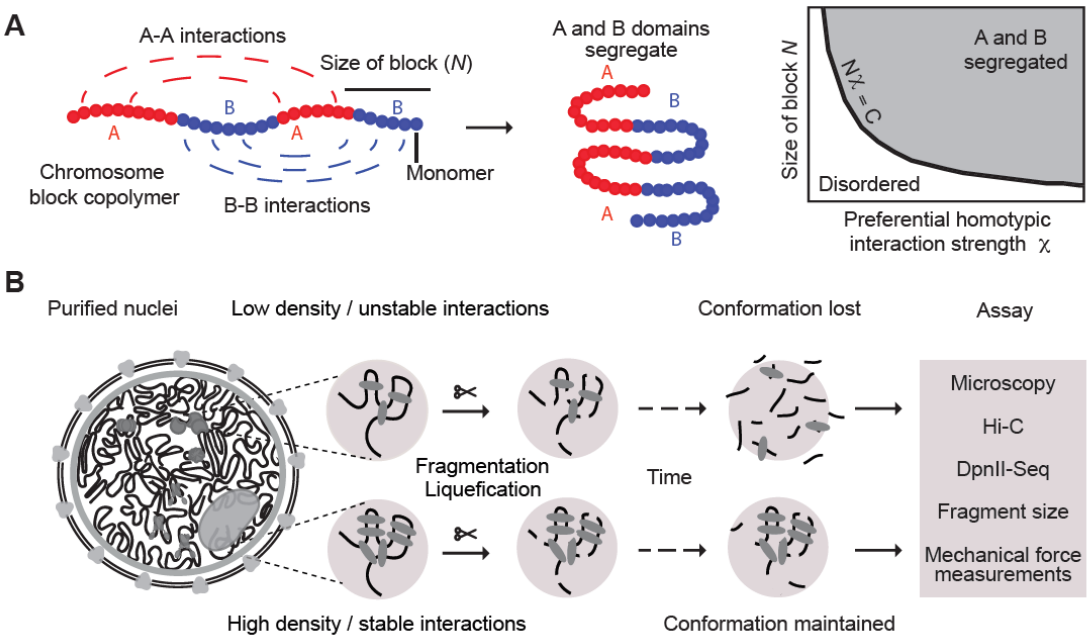


Figure 2.1 | Approach for measuring chromatin interaction stability.

(**A**) Block copolymer composed of a series of alternating A and B blocks, each composed of a number of monomers (left). The polymer can fold into spatially segregated domains of As and Bs (middle). Flory-Huggins polymer theory predicts that spatial segregation will occur when the product of the length of the blocks N (the number of monomers that make up blocks) and their effective preferential homotypic interaction strength χ (difference in the strength of homotypic interactions as compared to heterotypic (A-B) interactions) is larger than a critical value C. (**B**) Workflow to determine the stability of chromatin interactions genome-wide. DNA: black, varying chromatin features or proteins maintaining DNA conformation: grey ovals.).

Whether microphase separation of a block copolymer occurs depends on the interaction strengths between monomers as well as the lengths of the blocks of monomers of each type (Figure 2.1, A). Flory-Huggins polymer theory predicts that spatial segregation will occur when the product of the length of the blocks (N, the number of monomers that make up blocks) and their effective preferential

homotypic interaction strength (χ, a parameter that represents the difference in the strength of homotypic interactions as compared to heterotypic (A-B) interactions) is larger than a critical value C,(de Gennes, 1979; Leibler, 1980; Matsen and Schick, 1994). Large blocks of a polymer can spatially segregate even when attractive interactions among monomers are weak, while short blocks will only phase separate when interactions are sufficiently strong. The dependence of microdomain formation on the product of block size and interaction strength suggests an experimental approach to quantify the strengths and dynamics of interactions between individual loci that drive chromosome compartmentalization (Figure 2.1 A, B). One can start with a compartmentalized state of the genome and fragment the chromosomes by in situ restriction digestion, and then identify conditions where chromatin fragments become so short that the chromatin interaction strength between the segments is not sufficient to maintain a phase- or microphase-separated (due to restriction of separation by the polymeric constraint) state. As a result, chromosomal domains and compartments will disassemble over time and the chromosomal fragments of different type (e.g., As and Bs) will become mixed, i.e. chromatin becomes liquid-like. The kinetics of this dissolution and mixing process can then be assessed genome-wide by Hi-C at different times after chromatin fragmentation. Domains formed by strong, stable, and abundant interactions will dissociate more slowly than domains formed by weak, unstable, or infrequent interactions (Figure 2.1, A, B). Here we describe such a strategy that we call liquid chromatin Hi-C.

Chromosome conformation in isolated nuclei

To facilitate enzymatic fragmentation of chromosomes, we isolated nuclei from K562 cells. We performed four analyses to demonstrate that chromosome conformation in isolated K562 nuclei was the same as that in intact cells. First, DAPI staining and imaging showed that nuclei were intact with Lamin A as a ring at the nuclear periphery (Figure 2.2, A). Second, using 3C (Dekker *et al.*, 2002) we readily detected known looping interactions in the beta-globin locus (Dostie *et al.*, 2006; Chien *et al.*, 2011) (Figure 2.3). Third, 5C analysis (Dostie *et al.*, 2006) of a 1 Mb region surrounding the beta-globin locus showed that known CTCF-mediated interactions were preserved (Figure 2.3, (Tolhuis *et al.*, 2002; Dostie *et al.*, 2006; Splinter *et al.*, 2006; Kang *et al.*, 2017). Fourth, genome-wide Hi-C analysis (Lieberman-Aiden *et al.*, 2009; Belaghzal, Dekker and Gibcus, 2017) confirmed that chromosome territories, compartments (determined by principle component analysis, with compartments captured by the first principle component (PC1 (Lieberman-Aiden *et al.*, 2009; Belaghzal, Dekker and Gibcus, 2017)), TADs, and CTCF-CTCF loops were intact in isolated nuclei and quantitatively similar to those in intact cells (Figure 2.3, and below).

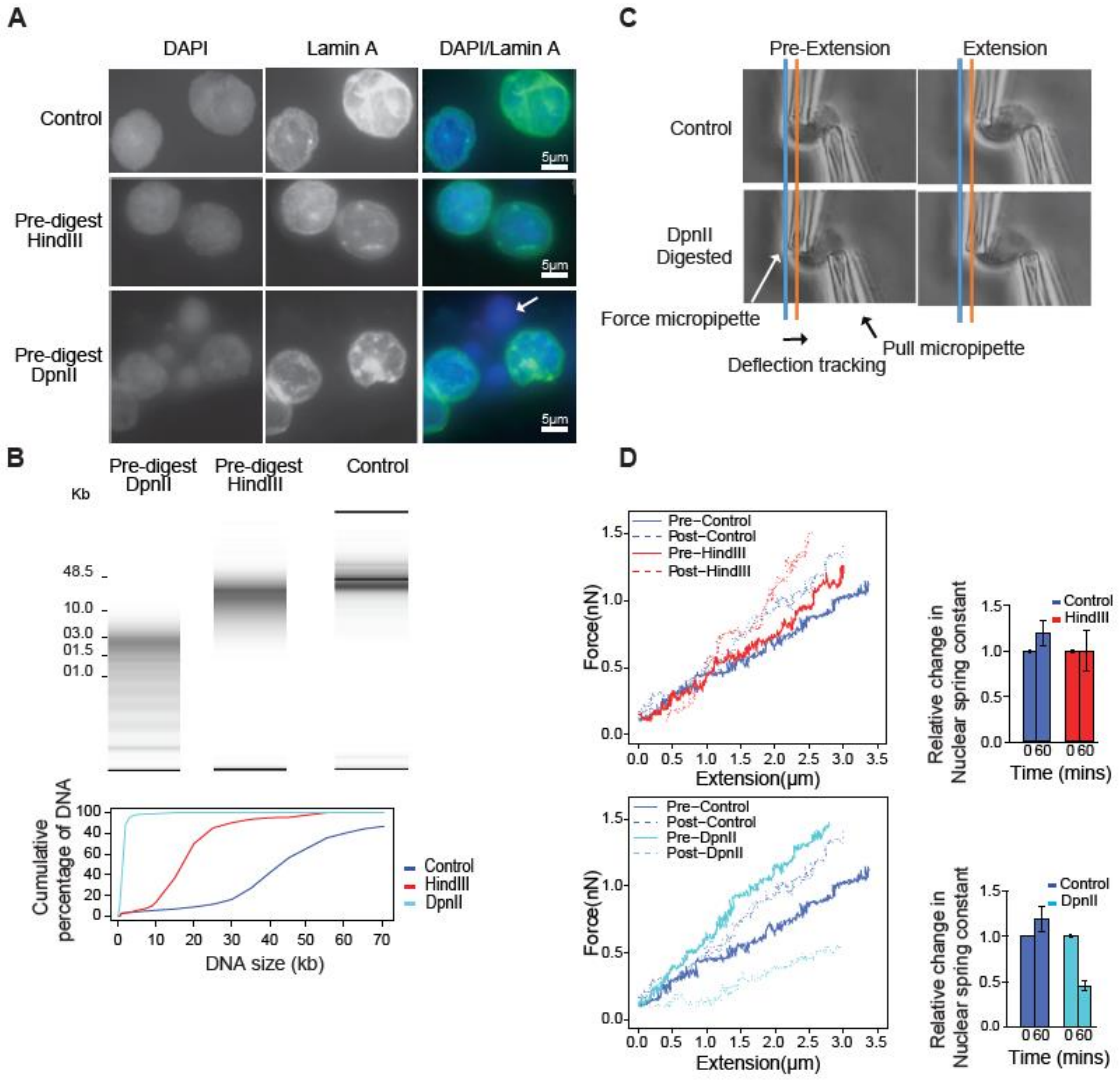


Figure 2.2 | Extensive fragmentation of chromatin leads to liquefied chromatin.

(**A**) Nuclear and chromatin morphology before and after chromatin fragmentation. Top row: control nuclei in restriction buffer, middle row nuclei digested for 4 hours with HindIII. Bottom row: nuclei digested for 4 hours with DpnII. Nuclei were stained with DAPI (left column), with antibodies against Lamin A (middle column). The right column shows the overlay of the DAPI and Lamin A stained images. HindIII digestion did not lead to major alteration in nuclear morphology and chromatin appearance, while DpnII digestion led to the appearance of DAPI stained droplets (arrow) exiting the nuclei. (**B**)Top: DNA purified from undigested nuclei, and nuclei pre-digested with DpnII and HindIII was run on a Fragment Analyzer. Bottom: cumulative DNA length distributions calculated from the

Fragment Analyzer data. (**C**) Micromanipulation of single nuclei. Isolated nuclei were attached to two micropipettes at opposite ends. Nuclei were extended by moving the right micropipette (Extension micropipette) and the force required was calculated from the deflection of the calibrated “force” (left) pipette. Blue and orange lines indicate the position of the force pipette before and after extension for control nuclei. After digestion of nuclei with DpnII (bottom) extension required less force as indicated by the much smaller deflection of the force pipette as compared to control nuclei. (**D**) Force-extension plots (left) for control nuclei before and 60 minutes after incubation in restriction buffer (pre- and post-control), for nuclei before and after digestion with DpnII, and for nuclei before and after HindIII digestion. Right panel: relative change in nuclear spring constants, calculated from the slopes of the force-extension plots shown on the left. Bars indicate standard error of the mean (n = 5 DpnII pre-digested nuclei, and n = 4 HindIII pre-digested nuclei).

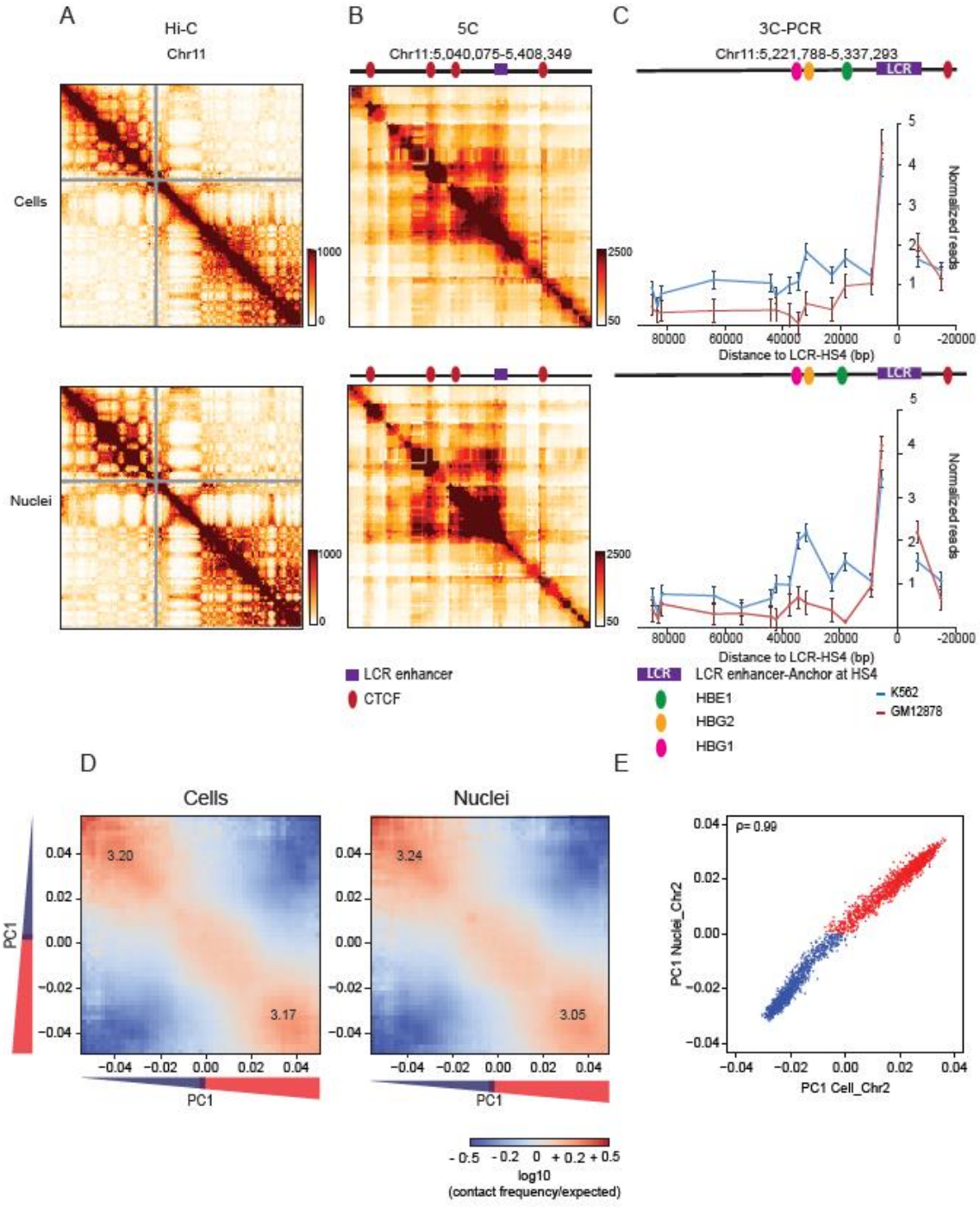


Figure 2.3 | Chromosome conformation in isolated nuclei.
(**A**) Hi-C 2.0 intra-chromosomal interaction maps for K562 cells display

chromosomal compartments and TADs. Top: cells. Bottom: purified nuclei. (**B**) 5C interaction map of 1 Mb region surrounding the beta-globin locus in K562 cells. Top: cells. Bottom: purified nuclei. CTCF-mediated interactions are preserved in purified nuclei. Red circles: positions of CTCF sites, purple square Beta-globin locus control region (LCR). (**C**) 3C-PCR for a 44120 kb region surrounding the beta-globin LCR on chromosome 11, detects at high resolution the known looping interactions between the LCR and the expressed gamma-globin genes (HBE1, HBG2) in K562 cells. Looping interactions are not detected in GM12878 cells that do not express these genes. Top: cells. Bottom: purified nuclei. (**D**) Compartmentalization saddle plots: average intra-chromosomal interaction frequencies between 100 kb bins, normalized by genomic distance. Bins are sorted by their PC1 value derived from Hi-C data obtained with K562 cells. In these plots preferential B-B interactions are in the upper left corner, and preferential A-A interactions are in the lower right corner. Numbers in the corners represent the strength of AA interactions as compared to AB interactions and BB interactions over BA interactions. Left: cells. Right: purified nuclei. (**E**) Spearman correlation of PC1 in cells vs PC1 in nuclei for chromosome 2 at 100kb resolution ($\rho = 0.99$).

Extensive chromatin fragmentation leads to the formation of liquid chromatin

We incubated purified nuclei for four hours with restriction enzymes that digest chromatin with different frequencies. Digestion with HindIII resulted in fragments that ranged in size from ~10-25 kb (Figure 2.2, B). A minority of molecules was over 25 kb (<15%), indicating that most of the genome was fragmented to a similar extent. Digestion with DpnII resulted in fragments that ranged in size between ~1 and ~6 kb, with less than 6% of fragments >6 kb (Figure 2.2, B). Microscopic inspection of nuclear morphology by DAPI and Lamin A immunofluorescence staining showed that fragmentation of chromatin with HindIII had only minor effects on nuclear morphology (Figure 2.2, A). In contrast, fragmentation of chromatin with DpnII led to large-scale alteration of

nuclear morphology as detected by DAPI staining, and large buds of apparently liquid chromatin (not surrounded by Lamin A) protruding from the nuclear periphery (Figure 2.2, A, arrow). After spinning down nuclei we detected no DNA in the supernatant, indicating that the liquified chromatin remains largely within the nuclear envelope.

We next tested whether different chromatin fragmentation levels had an effect on nuclear stiffness, which reflects the integrity of chromatin interactions inside the nucleus. Nuclei were isolated from K562 cells, attached to two micropipettes at opposite ends, and the whole nucleus was extended by moving an extension micropipette (Methods). The deflection of a force micropipette provides a measure of the force (Figure 2.2, C). This data provides a force vs. extension plot (Figure 2.2, D, plots on the left), in which the slope of the line fitted to the data is the nuclear spring constant in nN/μm (Figure 2.2, D, bar plots on the right). Extension between 0 – 30% strain measures the chromatin-dominated regime of nuclear force response (Banigan, Stephens and Marko, 2017; Stephens *et al.*, 2017, 2018). Isolated single nuclei were measured for their native spring constant before treatment. We find the stiffness can vary somewhat between individual nuclei. We then measured the stiffness of the same nuclei again 60 minutes post-treatment. Nuclei treated with control conditions (only restriction buffer added to the media) showed a slight stiffening of the nucleus (Figure 2.2, D). Treatment of nuclei with HindIII did not significantly decrease the stiffness as compared to their stiffness pre-treatment. In contrast, DpnII-treated

nuclei displayed a significant decrease (~75%) in stiffness, consistent with previous experiments (Stephens *et al.*, 2017). We conclude that chromosome and nuclear organization can tolerate fragmentation to 10-25 kb segments. In contrast, fragmenting the genome to smaller than 6 kb segments results in extensive loss of chromatin-mediated stiffness.

Compartmental segregation requires chromatin fragments larger than 6 kb

To determine how chromosome folding and compartmentalization is altered as a function of chromatin fragmentation level, we applied Hi-C before (conventional Hi-C) and after chromatin liquefication (liquid chromatin Hi-C). Nuclei were digested with either HindIII or DpnII for 4 hours followed by formaldehyde fixation and Hi-C analysis (Figure 2.4, A). Liquid chromatin Hi-C interaction maps obtained from nuclei that were pre-digested with HindIII were remarkably similar to those obtained with nuclei that were not pre-digested (Figure 2.5, A). The relationship between interaction frequency and genomic distance was largely unaffected (Figure 2.5, B). The ratio of intra- vs. inter-chromosomal interactions was also highly similar to that in untreated nuclei (Figure 2.5, B). Compartments (Figure 2.5, D) were readily detectable and captured by the first principle component (PC1). Compartment positions were unaffected (Spearman $\rho = 0.99$).

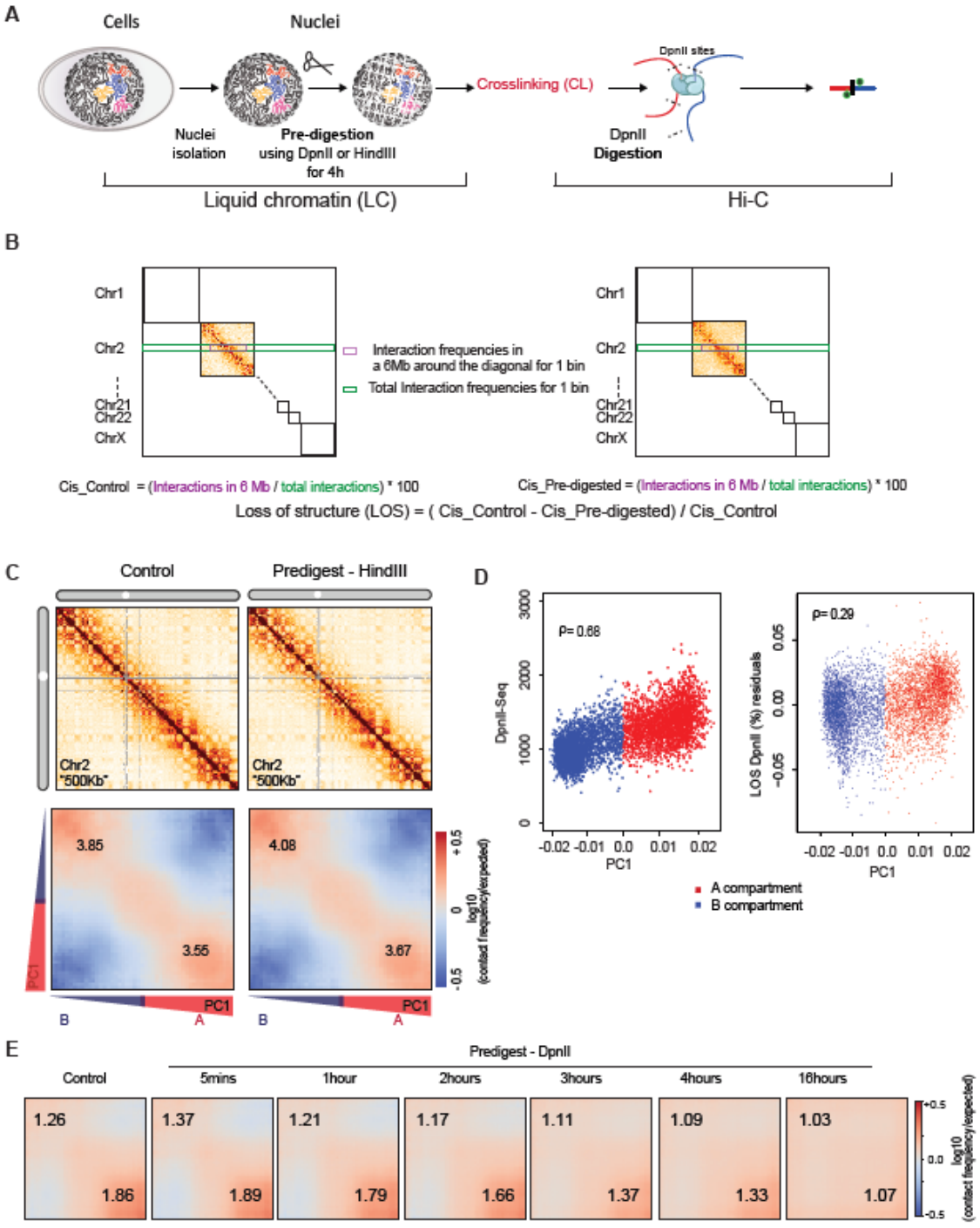


Figure 2.4 | Chromosome conformation dissolution upon chromatin fragmentation.
(**A**) Workflow for Liquid chromatin Hi-C. (**B**) Illustration of loss of structure metric using a pre-digested sample and a control. (**C**) Hi-C interaction maps and

compartmentalization saddle plots for a second replicate of control nuclei (incubated for 4 hours in restriction buffer) and nuclei pre-digested with HindIII for 4 hours. (**D**) Left: Spearman correlation of DpnII restriction digestion efficiency (DpnII-seq) and PC1 for chromosome 2 at 40 kb resolution. Right: Partial correlation of LOS (LOS residuals) with PC1 after controlling for restriction efficiency (DpnII-seq), for chromosome 2 at 40kb resolution. Spearman correlation is indicated. (**E**) compartmentalization saddle plots for the corresponding conditions. Numbers indicate strength of A-A and B-B interactions for inter-chromosomal interactions.

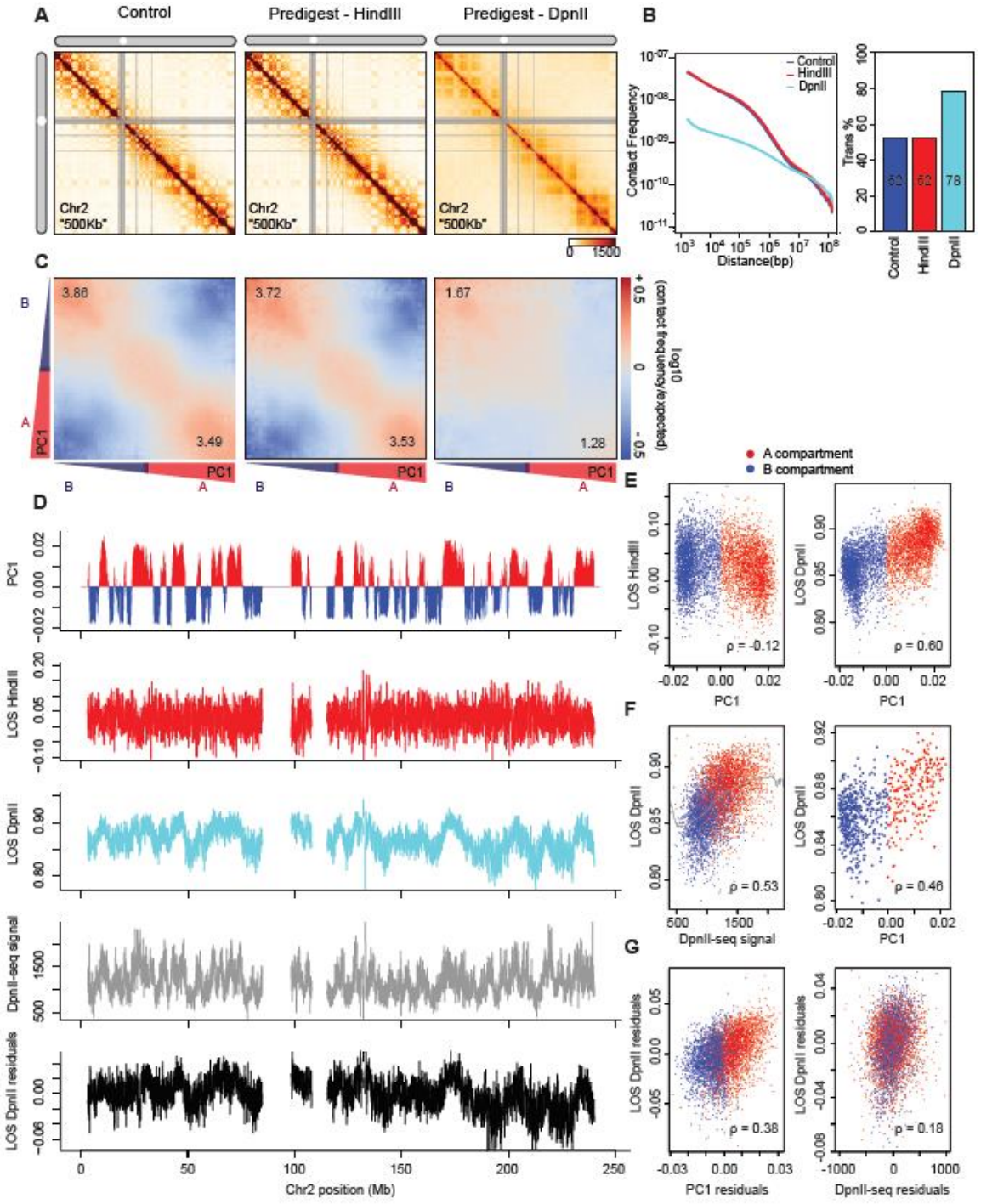

A
Control
Predigest - HindIII
Predigest - DpnII
Chr2 "500Kb"
0 1500
B
Control
HindIII
DpnII
Contact Frequency
Distance(bp)
Trans %
82
82
78
C
3.86
3.72
1.67
3.49
3.53
1.28
PC1
log10 (contact frequency/expected)
D
PC1
LOS HindIII
LOS DpnII
DpnII-seq signal
LOS DpnII residuals
Chr2 position (Mb)
A compartment
B compartment
E
LOS HindIII
LOS DpnII
ρ = -0.12
ρ = 0.60
F
ρ = 0.53
ρ = 0.46
DpnII-seq signal
G
ρ = 0.38
ρ = 0.18
PC1 residuals
DpnII-seq residuals

Figure 2.5 | Hi-C analysis reveals chromosome disassembly upon chromatin liquefication.
(**A**) Hi-C interaction maps of chromosome 2 binned at 500 kb. Left: interaction map for control nuclei in restriction buffer for 4 hours. Middle: nuclei pre-digested for 4 hours with HindIII prior to Hi-C. Right: nuclei digested for 4 hours with DpnII prior to Hi-C (see Figure 2.4, A). (**B**) Left: genome-wide interaction frequency as function of genomic distance for control nuclei (dark blue), nuclei pre-digested with HindIII (red), and nuclei pre-digested with DpnII (cyan). Right: percentage of inter-chromosomal (trans) interaction frequencies. (**C**) Compartmentalization saddle plots: average intra-chromosomal interaction frequencies between 40 kb bins, normalized by expected interaction frequency based on genomic distance. Bins are sorted by their PC1 value derived from Hi-C data obtained with control nuclei. In these plots preferential B-B interactions are in the upper left corner, and preferential A-A interactions are in the lower right corner. Numbers in corners represent the strength of AA interactions as compared to AB interaction and BB interactions over BA interactions (Figure 2.10,B). (**D**) Top plot: Eigenvector 1 values (PC1, 40 kb resolution) across a section of chromosome 2, representing A (red) and B (blue) compartments. Second plot: Loss of pair-wise interactions "LOS" (Methods and Figure 2.4, B) along chromosome 2 at 40 kb resolution for nuclei pre-digested with HindIII. Third plot: LOS for nuclei pre-digested with DpnII. Fourth plot: DpnII-seq signal along chromosome 2 at 40 kb resolution. Bottom plot: LOS-residuals for nuclei pre-digested with DpnII after correction for DpnII signal. (**E**) Correlation between LOS for nuclei pre-digested with HindIII (left) or DpnII (right) and PC1 (for chromosome 2, Spearman correlation values are indicated). (**F**) Left: correlation between LOS for nuclei pre-digested with DpnII and DpnII-seq signal (for chromosome 2). Grey line indicates moving average used for residual calculation. Right: correlation between LOS for nuclei pre-digested with DpnII and PC1 for loci cut to the same extent by DpnII (1000-1100 DpnII-seq reads/ 40 kb bin; for chromosome 2). Spearman correlation values are indicated. (**G**) Left: partial correlation between residuals of LOS for nuclei pre-digested with DpnII and residuals of PC1 after correcting for correlations between LOS and DpnII-seq and PC1 and DpnII-seq signal. Right: partial correlation between residuals of LOS for nuclei pre-digested with DpnII and residuals of DpnII-seq signal after correcting for correlations between LOS and PC1 and DpnII-seq signal and PC1. Spearman correlation values are indicated**.**

Chromosome compartment strength can be visualized and quantified by plotting interaction frequencies between pairs of 40 kb loci arranged by their values along the first eigenvector (PC1) to obtain compartmentalization saddle

plots (Figure 2.5, C). In nuclei pre-digested with HindIII, the strength of preferential A-A and B-B interactions (the ratio of the frequency of A-A and B-B interactions divided by the frequency of A-B interactions) was very similar to untreated nuclei (Figure 2.5, C; see Figure 2.4, C for a replicate).

Much more extensive changes were observed when nuclei were pre-digested for 4 hours with the frequent cutting enzyme DpnII (Figure 2.5, A) followed by formaldehyde fixation and Hi-C analysis. We observed a considerable redistribution of interactions with loss of short range (<10 Mb) intra-chromosomal interactions, and a gain of longer range (>10 Mb) interactions and inter-chromosomal interactions (Figure 2.5, B). The gain in inter-chromosomal interactions appeared to be the result of random mixing of As and Bs from different chromosomes as the preference for interchromosomal A-A and B-B interactions decreased (Figure 2.4, E). Moreover, compartment strength in cis was greatly reduced with a greater relative reduction evident in the A compartment (Figure 2.5, C). These observations show that fragmentation in <6 kb fragments leads to loss of spatial segregation of A and B compartments with more extensive loss of the A compartment.

Quantification of chromosome conformation dissolution

Loss of chromosome conformation and dissolution of chromosomal compartments will result in random mixing of previously spatially separated loci both in cis and in trans. In Hi-C, this effect will be apparent by a redistribution of contacts from short-range interactions towards longer range and inter-

chromosomal interactions. We developed a metric which represents the percentage change in short range intra-chromosomal interactions (up to 6 Mb), and concomitant increase in long-range and interchromosomal interactions after fragmentation relative to control nuclei, which we call “loss of structure” (LOS) (Figure 2.4, B).

We first calculated LOS after 4 hours for chromatin fragmented with HindIII. We observe that in general short-range interactions are only somewhat reduced (less than 5%; Figure 2.5, B). When LOS is plotted along chromosomes (Figure 2.5, D), we observed that LOS was very weakly negatively correlated with PC1 (Figure 2.5, D and E left panel).

We performed the same analysis for nuclei pre-digested with DpnII for 4 hours. We find extensive loss of chromosome conformation, with LOS generally >80%. LOS varies along chromosomes and is strongly positively correlated with PC1 with loci in the A compartment displaying the largest loss (Figure 2.5, D and E). These results show that chromatin fragmentation to <6 kb fragments leads to extensive genome-wide dissolution of chromosome conformation, loss of spatial segregation of A and B compartments, with the A compartment affected the most.

After correcting for differential fragmentation LOS remains highly correlated with compartment status

One explanation for the greater effect of fragmentation on chromatin interactions in the A compartment could be that DpnII cuts more frequently in the A compartment producing smaller fragments. We performed two experiments to assess differential fragmentation. First, we determined the cutting frequency of DpnII in isolated nuclei across the genome by sequencing the ends of the DNA fragments (DpnII-seq; Figure 2.6, see Methods). Second, we directly determined the average fragment size along the genome by purifying DNA of different sizes after pre-digestion with DpnII, sequencing the ends and using the data to calculate for each 40 kb bin the average fragment size (see Methods). The average fragment size for most 40 kb bins ranged from 2.7 to 3.7 kb and was on average slightly smaller for A compartments compared to B compartments (3.1 kb and 3.2 respectively). Cutting frequency and average fragment size are both correlated with PC1 and with LOS (Figure 2.5, D and F left panel, Figure 2.4, D, Figure 2.7).

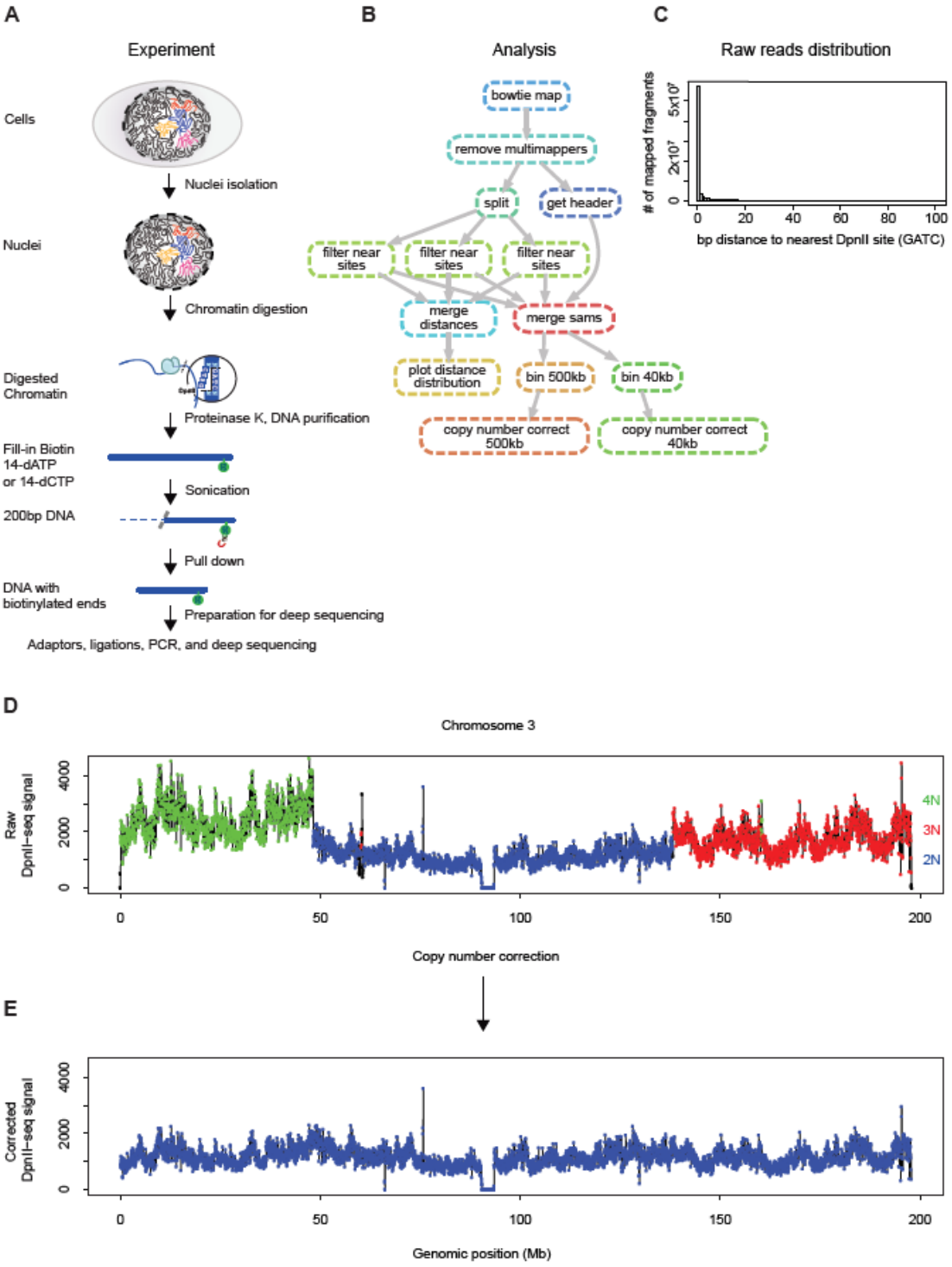

A
Experiment
Cells
Nuclei isolation
Nuclei
Chromatin digestion
Digested Chromatin
Proteinase K, DNA purification
Fill-in Biotin 14-dATP or 14-dCTP
Sonication
200bp DNA
Pull down
DNA with biotinylated ends
Preparation for deep sequencing
Adaptors, ligations, PCR, and deep sequencing
B
Analysis
bowtie map
remove multimappers
split
get header
filter near sites
filter near sites
filter near sites
merge distances
merge sams
plot distance distribution
bin 500kb
bin 40kb
copy number correct 500kb
copy number correct 40kb
C
Raw reads distribution
# of mapped fragments
bp distance to nearest DpnII site (GATC)
D
Chromosome 3
Raw DpnII-seq signal
4N
3N
2N
Copy number correction
E
Corrected DpnII-seq signal
Genomic position (Mb)

Figure 2.6 | Experimental protocol and computational workflow for DpnII-seq.
(**A**) Schematic of DpnII-seq experimental protocol for recovering DNA fragments digested by the restriction enzyme DpnII. (**B**) Directed graph of DpnII-seq computational pipeline (**C**) Histogram of distance to nearest DpnII recognition site for each recovered DpnII digested fragment. (**D**) Raw DpnII-seq signal displaying multiple copy number states (2N, 3N, 4N) within chromosome 3 (data binned at 40 kb). (**E**) Copy number corrected DpnII-seq signal displaying single copy number state (2N) across chromosome 3**.**

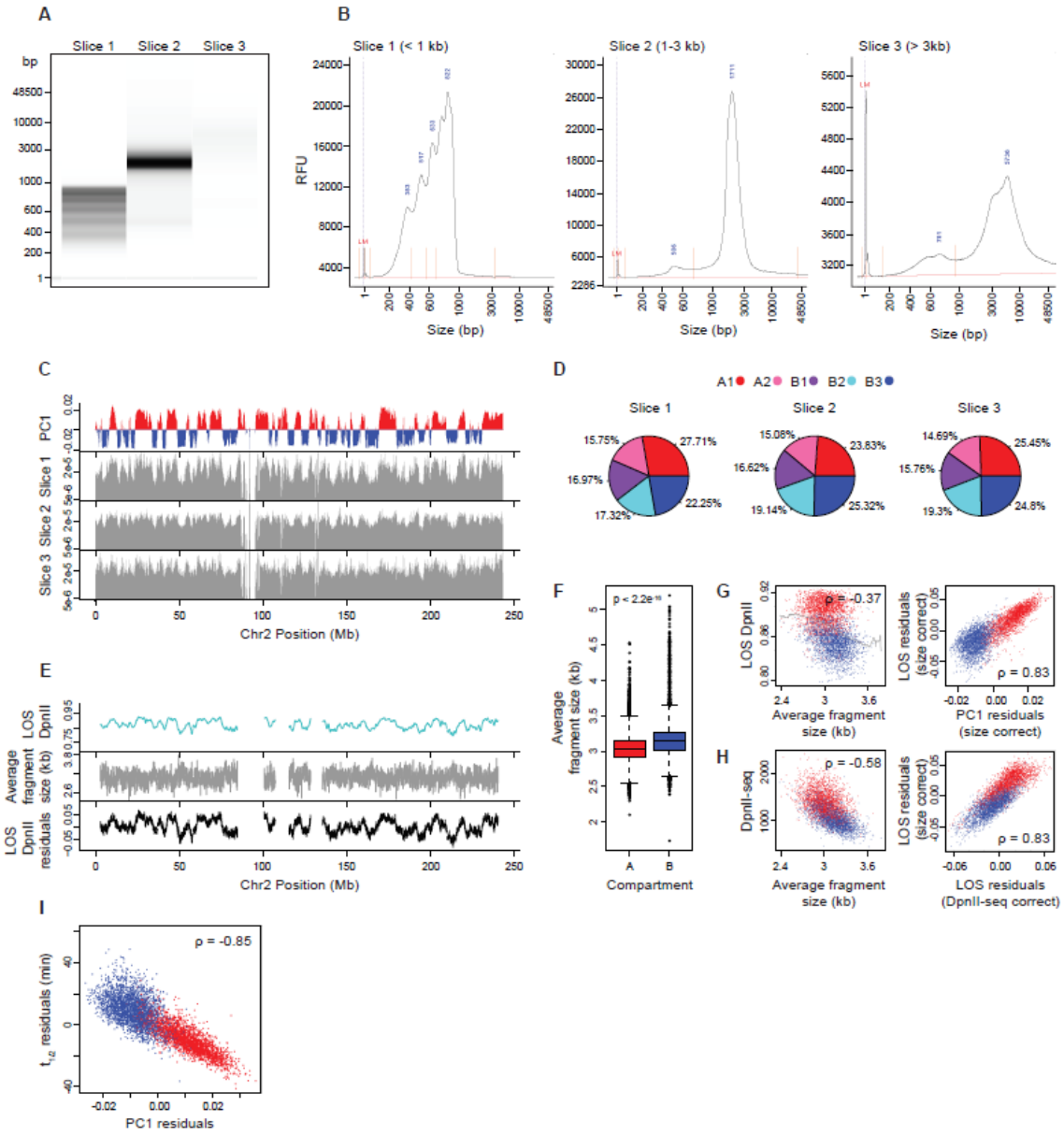


Figure 2.7 | Average fragment size per bin and correlation with chromatin stability.

(**A**) DNA purified from nuclei pre-digested with DpnII for 4 hours were separated into slices of three sizes and run on a Fragment Analyzer. (**B**) Fragment Analyzer distributions of DNA fragment sizes for the three separated slices (RFU: relative fluorescence unit, LM: lower marker, fragment sizes at distribution peaks are given in blue). (**C**) Top plot: Eigenvector 1 values (PC1, 40 kb resolution) across a section of chromosome 2, representing A (red) and B (blue) compartments. Bottom three plots: Normalized coverage of fragments from given slice size across a section of chromosome 2. (**D**) Percentages of fragments mapped to

each subcompartment for given slice size. (**E**) Top plot: LOS along chromosome 2 at 40 kb resolution for nuclei pre-digested with DpnII. Middle plot: Average fragment size estimated for every 40kb bin after pre-digestion with DpnII (Methods). Bottom plot: LOS-residuals for nuclei pre-digested with DpnII after correction for average fragment size. (**F**) Boxplot of average fragment size for A compartment and B compartment bins. Significance determined by two-sample two tailed t-test ($p < 2.2e-16$). (**G**) Left plot: correlation between LOS for nuclei pre-digested with DpnII and average fragment size. Grey line indicates moving average used for residual calculation. Right plot: partial correlation between residuals of LOS for nuclei pre-digested with DpnII and residuals of PC1 after correcting for correlations between LOS and average fragment size and PC1 and average fragment size (for chromosome 2, Spearman correlation values are indicated). (**H**) Left plot: Correlation between DpnII-seq signal and average fragment size. Right plot: correlation between residuals of LOS after correcting for average fragment size and residuals of LOS after correcting for DpnII-seq signal (for chromosome 2, Spearman correlation values are indicated). (**I**) Partial correlation between residuals of $t_{1/2}$ and residuals of PC1 after correcting for correlations between $t_{1/2}$ and average fragment size and PC1 and average fragment size.

Next, we corrected LOS for the differential efficiency of DpnII digestion by calculating the partial correlation between LOS and PC1 after correcting for the correlations of PC1 and LOS with DpnII digestion frequency (see Methods for details). We find that the residuals of PC1 and LOS, are still correlated (Spearman $\rho = 0.38$ for chromosome 2; Figure 2.5, G). Similarly, when we corrected LOS for differences in average fragment size we find that the residuals of LOS remain highly correlated with residuals of PC1 (Spearman $\rho = 0.83$ for chromosome 2, Figure 2.7). To illustrate the correlation between LOS and PC1 independent of fragmentation level directly we selected a set of loci along chromosome 2 that are all cut to the same extent (1000-1100 reads in the DpnII-seq dataset). When we plot LOS vs. PC1 for this set we find a strong correlation (Figure 2.5, F right panel, Spearman $\rho = 0.46$). Finally, we repeated the entire

liquid chromatin Hi-C procedure using a different restriction enzyme, FatI, which has a different pattern of digestion across the genome as compared to DpnII but produces fragments that are similarly small (Figure 2.8). We calculated LOS and corrected for differential FatI digestion along the genome using FatI-seq, exactly as above for DpnII. We again observe a high correlation between residuals of LOS and PC1 (Spearman $\rho = 0.64$ for chromosome 2, Figure 2.8). We conclude that LOS is correlated with compartment status, and that this result is robust for different enzymes and different methods for correcting for digestion efficiency.

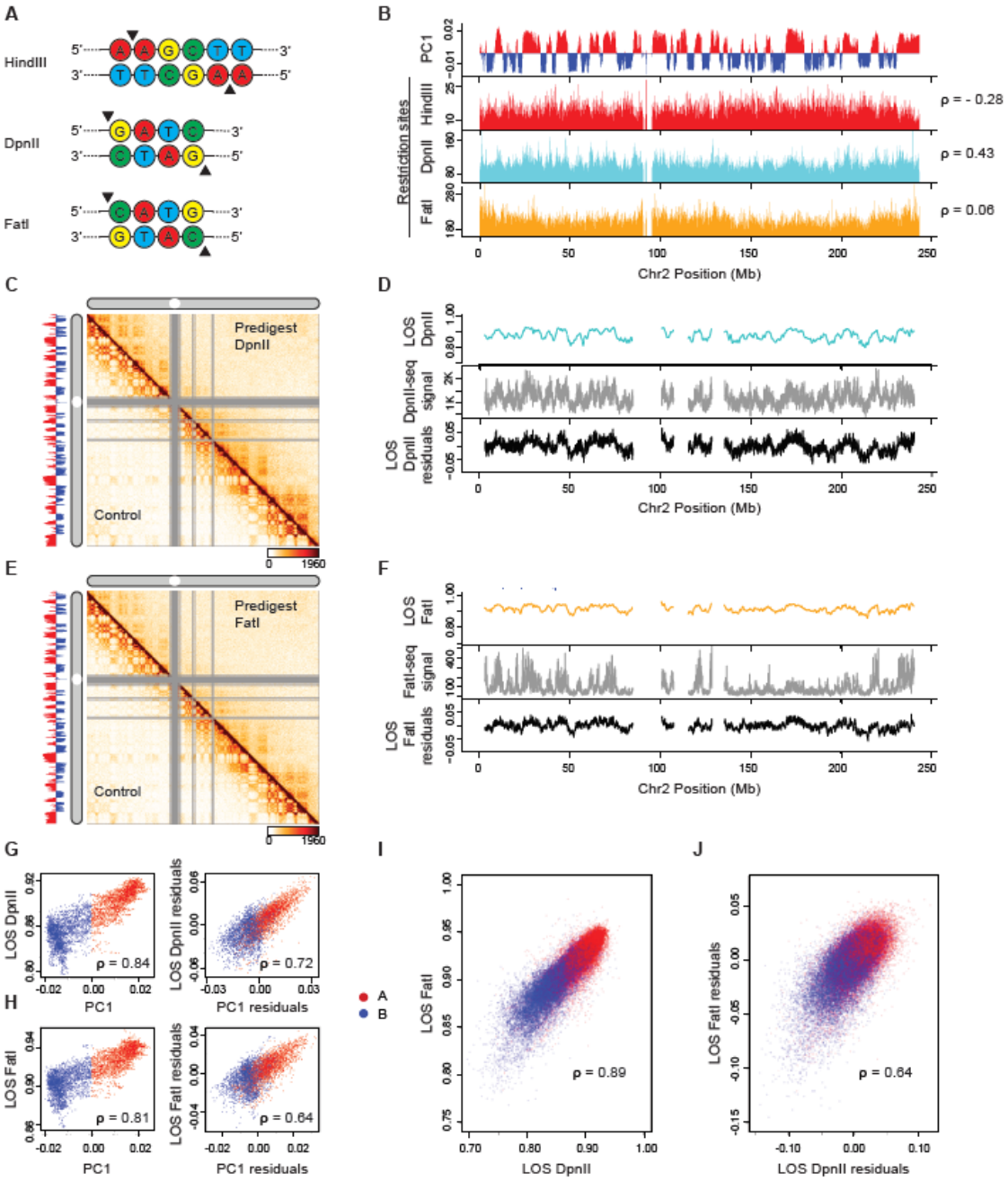


Figure 2.8 | Liquid chromatin Hi-C results are reproducible using the restriction enzyme FatI.

(**A**) Restriction sites for the selected restriction enzymes. Black triangles denote cut sites. (**B**) Top plot: Eigenvector 1 values (PC1, 40 kb resolution) across a section of chromosome 2, representing A (red) and B (blue) compartments. Bottom three plots: Coverage of restriction sites (40kb resolution). Spearman

correlation between restriction site coverage and PC1 is given for each restriction site track. (**C**) Third replicate of DpnII predigest liquid chromatin Hi-C. Hi-C interaction maps of chromosome 2 binned at 500 kb. Bottom left: control nuclei in restriction buffer for 4 hours. Top right: nuclei digested for 4 hours with DpnII prior to Hi-C. Left track: Eigenvector 1 values (PC1, 40 kb resolution) across a section of chromosome 2, representing A (red) and B (blue) compartments. (**D**) Top plot: LOS along chromosome 2 at 40 kb resolution for nuclei pre-digested with DpnII. Middle plot: DpnII-seq signal. Bottom plot: LOS-residuals for nuclei pre-digested with DpnII after correction for DpnII-seq signal. (**E**) FatI predigest liquid chromatin Hi-C. Hi-C interaction maps of chromosome 2 binned at 500 kb. Bottom left: control nuclei in restriction buffer for 4 hours. Top right: nuclei digested for 4 hours with FatI prior to Hi-C. Left track: Eigenvector 1 values (PC1, 40 kb resolution) across a section of chromosome 2, representing A (red) and B (blue) compartments. (**F**) Top plot: LOS along chromosome 2 at 40 kb resolution for nuclei pre-digested with FatI. Middle plot: FatI-seq signal. Bottom plot: LOS-residuals for nuclei pre-digested with FatI after correction for FatI-seq signal. (**G**) Left plot: Correlation between LOS for nuclei pre-digested with DpnII and PC1. Right plot: partial correlation between residuals of LOS for nuclei pre-digested with DpnII and residuals of PC1 after correcting for correlations between LOS and DpnII-seq and PC1 and DpnII-seq signal (for chromosome 2, Spearman correlation values are indicated). (**H**) Left plot: Correlation between LOS for nuclei pre-digested with FatI and PC1. Right plot: partial correlation between residuals of LOS for nuclei pre-digested with FatI and residuals of PC1 after correcting for correlations between LOS and FatI-seq and PC1 and FatI-seq signal (for chromosome 2, Spearman correlation values are indicated). (**I**) Correlation between LOS for nuclei pre-digested with FatI and LOS for nuclei pre-digested with DpnII (genome wide, Spearman correlation values are indicated). (**J**) Correlation between residuals of LOS for nuclei pre-digested with FatI and residuals of LOS for nuclei pre-digested with DpnII after correcting for correlations between FatI LOS and FatI-seq and DpnII LOS and DpnII-seq (genome wide, Spearman correlation values are indicated).

Dissociation kinetics of chromatin interactions and compartments

The loss of conformation after DpnII pre-digestion allowed us to measure the dissociation kinetics of compartments and stability of chromatin interactions. We first determined the kinetics of chromatin fragmentation (Figure 2.9, A, Figure 2.10, A). We digested nuclei with DpnII for 5 minutes up to 16 hours. After 5 minutes the size range of fragments was between 3 and 15 kb (80% of

fragments; Figure 2.11, A). After one hour 80% of DNA fragments were smaller than 7 kb and after 16 hours 85% of fragments were smaller than 3.5 kb. We sequenced DNA ends to determine the distribution of DpnII cuts across the genome (Figure 2.11, B). At all timepoints the number of DpnII cuts per 40 kb bin was correlated with PC1 (Figure 2.11, B) and the pattern did not change over time (Figure 2.11, B, correlation matrix, Figure 2.9, A).

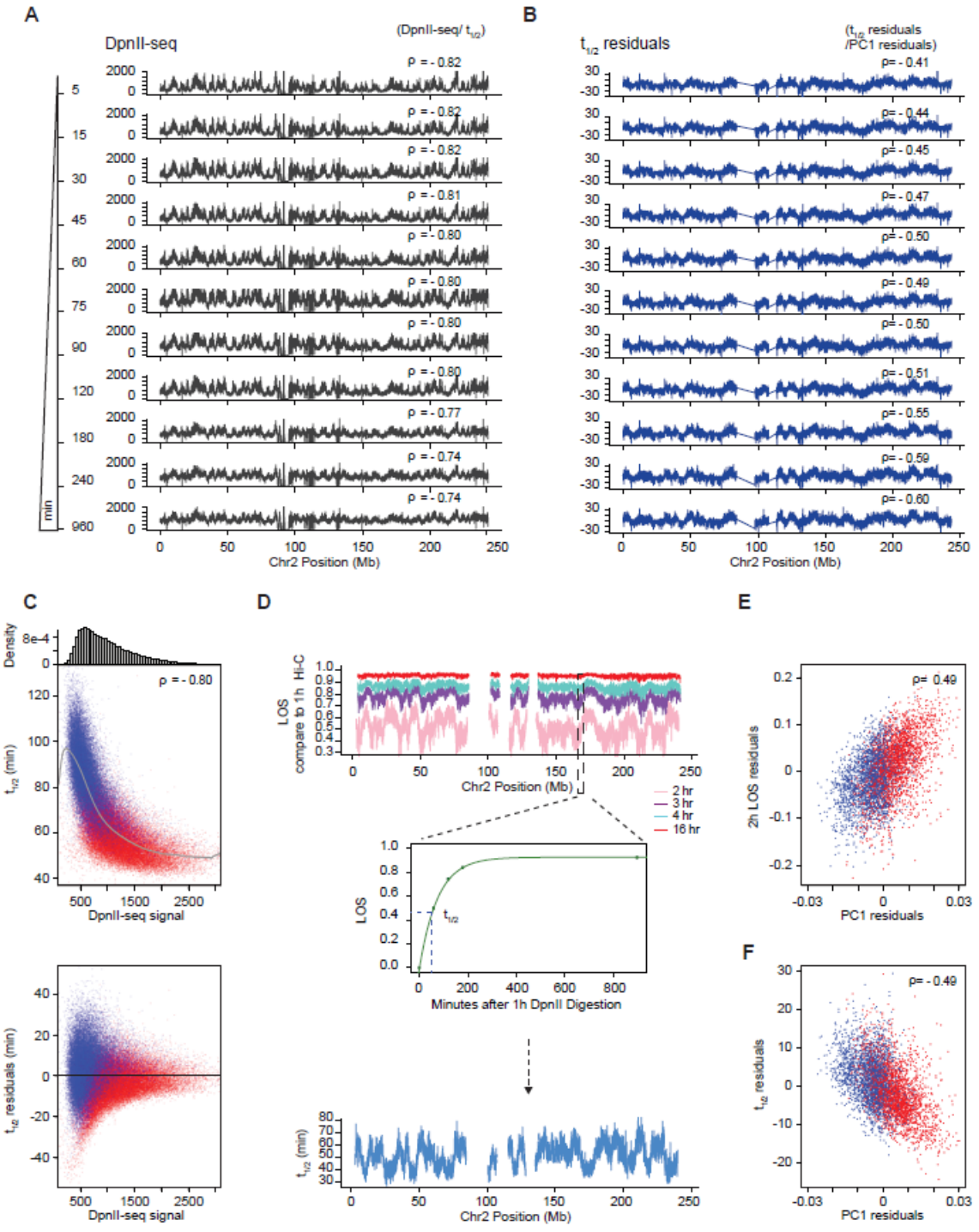


Figure 2.9 | Variations in Half-life and LOS are not explained by DpnII digestion kinetics.
(**A**) DpnII-seq signals along chromosome 2 after indicated times of digestion.

Spearman correlations between DpnII-seq and $t_{1/2}$ at each timepoint is indicated. (**B**) $t_{1/2}$ residuals along chromosome 2 after correcting $t_{1/2}$ values by the correlation between $t_{1/2}$ and DpnII-seq signals shown on the left obtained after the indicated times of digestion. Spearman correlation between $t_{1/2}$ residuals and PC1 residuals are indicated. (**C**) Top: Genome wide scatterplot of $t_{1/2}$ versus 1 hour DpnII-seq signal. Gray line: moving average. Bar plot above shows the number of loci displaying various levels of DpnII-mediated cuts. Bottom: residuals of $t_{1/2}$ calculated by subtracting $t_{1/2}$ from the corresponding average $t_{1/2}$ (gray line in top plot) plotted vs. number of DpnII cuts. Red dots: loci in the A compartment; Blue dots: loci in the B compartment. The majority of loci have 500-1100 cuts. When comparing loci with similar number of DpnII cut we observe that loci in the A compartment have shorter $t_{1/2}$ values as compared to loci in the B compartment. (**D**) Top: LOS along chromosome 2 at the indicated timepoints of digestion and calculated by comparison to Hi-C data obtained after 1 hour of digestion. Middle: calculation of $t_{1/2}$ from LOS at different timepoints. Bottom: $t_{1/2}$ along chromosome 2. This $t_{1/2}$ is calculated using the Hi-C data obtained after 1 hour of pre-digestion as starting point. (**E**) Partial correlation between LOS and PC1 after correcting for their correlations with DpnII-seq. LOS (at 2 hours) is calculated as in panel C using the Hi-C data obtained after 1 hour of pre-digestion as starting point (**F**) Partial correlation between $t_{1/2}$ and PC1 after correcting for their correlations with DpnII seq. $t_{1/2}$ is calculated as in panel D using the Hi-C data obtained after 1 hour of pre-digestion as starting point. Spearman correlations are indicated.

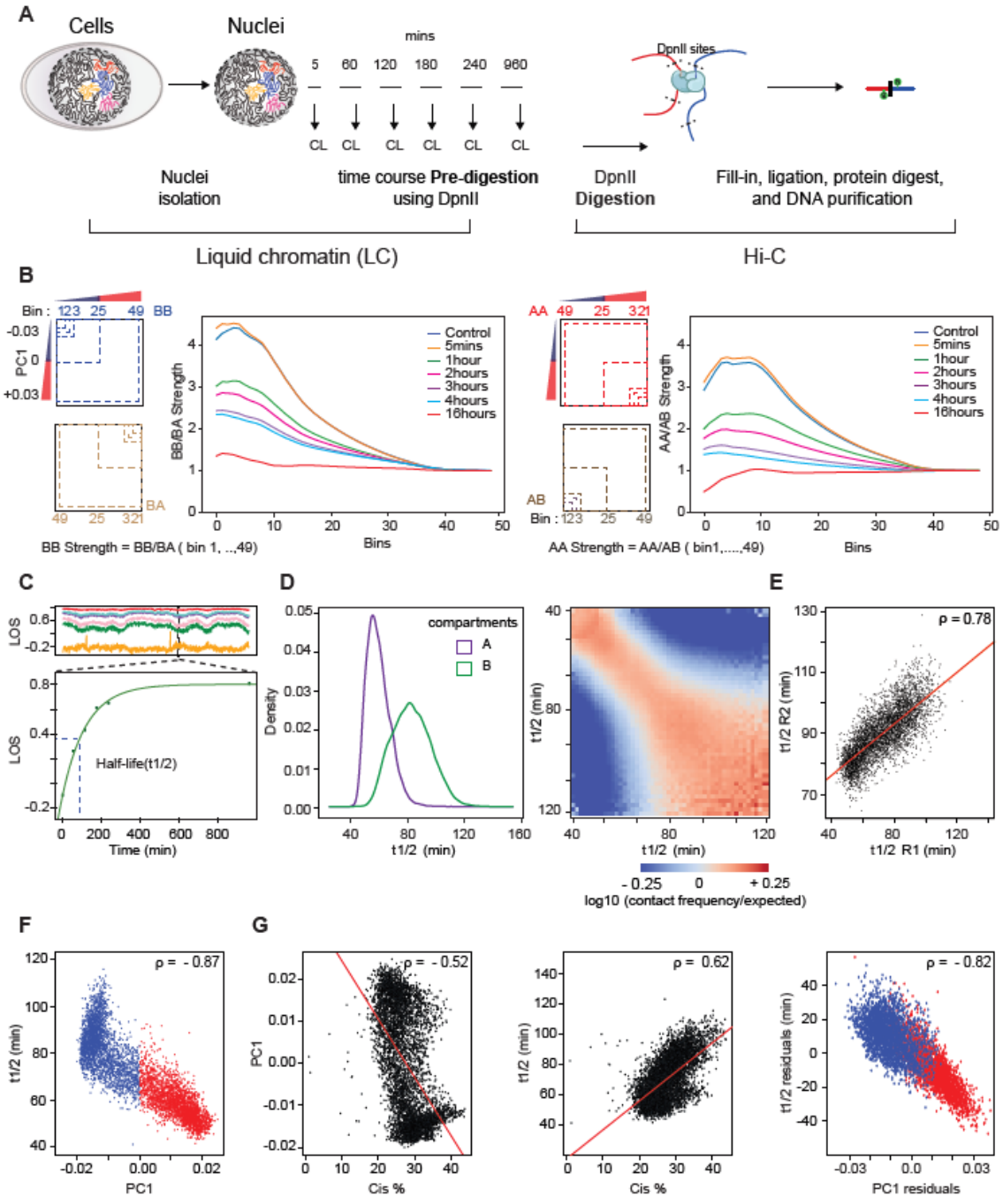


Figure 2.10 | Liquid chromatin-Hi-C protocol and quantification of loss of structure after chromatin pre-digestion.
(**A**) Workflow for Liquid chromatin Hi-C timecourse. CL = cross-linking step. (**B**) Compartment strength derived from compartment saddle plots (See Methods). Left: Diagram depicting compartment strength calculation for B-B interactions.

Plot to the right of diagram: B-B interaction strength as a function of bin number for all timepoints of the time course. Right: Diagram depicting compartment strength calculation for A-A interactions. Plot to the right of diagram: A-A interaction strength as a function of bin number for all time points of the time course. (**C**) Top: LOS signal across a 40 Mb region on chromosome 2 calculated for indicated timepoints in the digestion timecourse. Line colors as in Figure 2.11, E. Bottom: Exponential curve fit to LOS timepoints for a single 40kb bin. $t_{1/2}$ (dashed vertical blue line) representing time elapsed to reach half saturation of LOS signal. (**D**) Left: Density distributions of $t_{1/2}$ for A and B compartments. Right: $t_{1/2}$ saddle plots: average intra-chromosomal interaction frequencies between 40 kb bins, normalized by genomic distance. Bins are sorted by their $t_{1/2}$ value derived from digestion timecourse. Bins with high $t_{1/2}$ preferentially interact (bottom right of heatmap) and bins with low $t_{1/2}$ preferentially interact (top left of heatmap). (**E**) Scatterplot of $t_{1/2}$ vs $t_{1/2}$ for two timecourse replicates (R1 and R2) on chromosome 2. Regression line (red). Spearman correlation is indicated. (**F**) Scatterplot of PC1 vs $t_{1/2}$ for chromosome 2. A compartment (red); B compartment (blue). (**G**) Left: Scatterplot of percent interactions occurring in cis within a 6 Mb distance out of total genome wide interactions for each 40 kb bin in control Hi-C map (Cis %) vs PC1. Middle: Cis% vs $t_{1/2}$. Right: Scatterplot of partial correlation between PC1 and $t_{1/2}$ controlled by Cis %. A compartment (red); B compartment (blue). Solid red lines are regression lines. Spearman correlations are indicated.

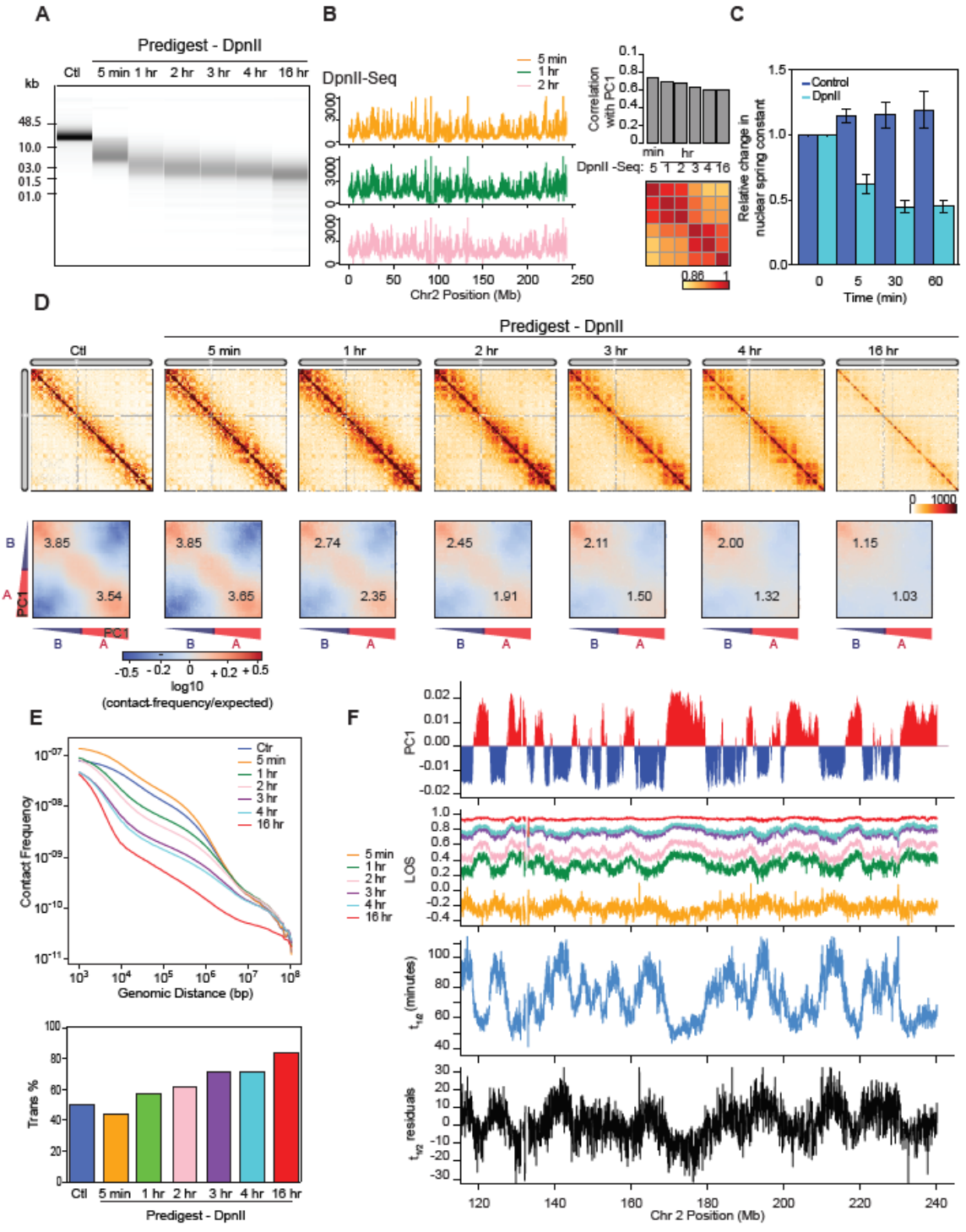


Figure 2.11 | Kinetics of chromatin fragmentation and chromatin dissolution. (**A**) DNA purified from undigested nuclei, and nuclei pre-digested with DpnII for different time points were run on a Fragment Analyzer. (**B**) Left:

DpnII-seq signals along chromosome 2 binned at 40 kb resolution after digestion for 5 minutes, 1 hour and 2 hours. Right: correlation between DpnII-seq signals and PC1 and between DpnII-seq signals at different time points. (**C**) Relative change in nuclear spring constant (nN/µm) after DpnII fragmentation at different time points. Spring constant is significantly decreased after 5 minutes and at background level by 1 hour ($p = 0.002$, two-tailed t-test). (**D**) Top row: Hi-C interaction maps of chromosome 2 binned at 500 kb. Control: nuclei in restriction buffer for 4 hours. Pre-digest DpnII: nuclei were pre-digested with DpnII for 5 minutes up to 16 hours. (Figure 2.10, A). Bottom row: compartmentalization saddle plots for the corresponding conditions. Numbers indicate strength of A-A and B-B interactions. (**E**) Top: genome-wide interaction frequency as function of genomic distance for Hi-C data shown in panel (D). Bottom: percentage of inter-chromosomal (trans) interactions genome-wide for control nuclei and for nuclei pre-digested with DpnII for up to 16 hours. (**F**) Top: PC1 along a section of 120 Mb of chromosome 2. Second plot: LOS along chromosome 2 at 40 kb resolution for all time points (Figure 2.4, B). Third plot: half-life ($t_{1/2}$) values derived from the exponential fit of the six time-points for every 40 kb bin (Figure 2.10, C). Bottom plot: residuals of $t_{1/2}$ after correcting for correlations between $t_{1/2}$ and DpnII-seq (DpnII-seq data for t = 1 hour).

Micromanipulation was again used to measure the nuclear spring constant corresponding to nuclear stiffness. Nuclei displayed a significant loss in stiffness within 5 minutes. Loss of stiffness leveled off at 30 minutes of digestion and showed a 60% decrease in nuclear rigidity, similar to previous experiments in other cell types ((Stephens *et al.*, 2017), Figure 2.11, C). Combined these analyses show that the bulk of DNA fragmentation and chromatin liquefication occurs within the first hour.

Next, we performed liquid chromatin Hi-C where nuclei were pre-digested with DpnII for 5 minutes up to 16 hours (Figure 2.10, A). Interestingly, after 5 minutes of pre-digestion chromosome conformation and compartmentalization are intact, even though chromatin was fragmented to 3-15 kb segments before

fixation and nuclear stiffness was significantly reduced (Figure 2.11, C and D). The percentage of intra-chromosomal interactions especially for loci separated by <1 Mb was increased (Figure 2.11, E).

At subsequent time points, when most chromatin fragments are <7 kb long we observe increased loss of intra-chromosomal interactions and concomitant increased inter-chromosomal interactions genome-wide (Figure 2.11, D and E). Compartmentalization, as quantified by the preference of A-A and B-B interactions over A-B interactions, is progressively lost (Figure 2.11, D lower row of heatmaps, Figure 2.10, B). A-A interactions disappear faster than B-B interactions. After 16 hours, only a low level of preferential B-B interaction remains.

Quantification of the half-life of chromosome conformation across the genome

To quantify the kinetics of loss of chromosome conformation and compartmentalization, we calculated LOS genome-wide for each time point (Figure 2.11, F). At t = 5 minutes, LOS is generally negative indicating a gain in chromatin interactions: on average ~25% gain of intra-chromosomal interactions between loci separated by <6 Mb, consistent with the initial increase in overall intra-chromosomal interactions described above (Figure 2.11, E). LOS is inversely correlated with PC1, indicating that loci located within A compartments gain more intra-chromosomal interactions than loci located within B compartments (Spearman ρ = -0.53 for chromosome 2, Spearman ρ = -0.49

genome-wide). A “block copolymer” model predicts that partial DNA digestion can lead to a strengthening of compartmentalization by removing covalent linkages between A and B blocks, as long as the fragments are still large enough so that attractive forces between them are sufficient for phase segregation (see Methods). At subsequent time points, LOS is increasingly positive as intra-chromosomal interactions are progressively lost and inter-chromosomal interactions are gained. LOS is the highest for loci located in the A compartment. At t = 16 hours, LOS is generally as high as 90%, intra-chromosomal interactions are low (<20% of total), and only preferential B-B interactions are still observed in the Hi-C interaction map (Figure 2.11, D). Similar results were obtained with an independent replicate time course experiment (see below).

Next we determined for each 40 kb locus at which time LOS has reached 50% of its maximal value at t = 16 hours. We refer to this time as the half-life, $t_{1/2}$ (minutes), of chromatin interactions at each locus (Figure 2.11, F). To identify $t_{1/2}$ we plotted LOS as a function of time for each 40 kb locus and fit the data to an exponential curve so that $t_{1/2}$ could be determined when LOS equals 50% of its maximum (Figure 2.10, C). Examining $t_{1/2}$ along chromosomes, we observe a strong inverse correlation with PC1 (Spearman $\rho = -0.87$, Figure 2.10, F): interactions in the A compartment dissolve relatively fast ($t_{1/2}$ = 40-80 minutes) while interactions in the B compartment dissolve slower ($t_{1/2}$ = 60-120 minutes, Figure 2.10, D). We also calculated $t_{1/2}$ genome-wide for the second independent time course experiment and find a strong correlation between $t_{1/2}$ calculated from

the two datasets (Spearman ρ = 0.78 for chromosome 2, Spearman ρ = 0.76 genome-wide, Figure 2.10, E). The value of $t_{1/2}$ is proportional to a dissociation rate constant and thus independent of the initial level of intra-chromosomal interactions for a given locus. $t_{1/2}$ remains highly correlated with PC1 after correcting for the initial level of intra-chromosomal (<6 Mb) interactions for each bin (Spearman ρ = -0.82, Figure 2.10, F and G).

Similar to LOS, $t_{1/2}$ is correlated with DpnII digestion frequency at all timepoints (Figure 2.9, A). We calculated the partial correlation between $t_{1/2}$ and PC1 after correcting for correlations between PC1 and $t_{1/2}$ with DpnII cutting frequency. We find that $t_{1/2}$ and PC1 remain strongly correlated (Figure 2.11, F), regardless of which DpnII fragmentation dataset (genome wide Spearman ρ ranging from -0.41 to -0.60, t = 5 min up to t = 16 hours) was used for the calculation of the partial correlation (Figure 2.9, A and B). Although loci in the A compartment are often cut more frequently than loci in the B compartment, when comparing loci cut with similar frequency, loci in the A compartments had shorter half-lives (Figure 2.9, C). Similar results were obtained when $t_{1/2}$ was corrected for average fragment size for each bin (Figure 2.7, I, Spearman ρ = -0.85 for chromosome 2, Spearman ρ = -0.76 genome-wide).

We considered whether we could have overestimated the $t_{1/2}$ for loci in the B compartment because fragmentation of these loci could be slower than for loci in the A compartment. We reasoned that because after 1 hour incubation with DpnII digestion is largely complete, calculation of LOS using the Hi-C data at t =

1 hour as starting condition would provide an estimate of dissolution kinetics starting at a timepoint when A and B compartments are both extensively fragmented. We find that LOS, and $t_{1/2}$ calculated this way are still strongly correlated with PC1, and this correlation remains strong after correcting for fragmentation level (Figure 2.9 D to F).

Dissociation kinetics of chromatin interactions at different sub-nuclear structures

The A1 sub-compartment is most enriched in active histone modifications and found near nuclear speckles (Chen *et al.*, 2018). B2 and B3 are located near the nuclear lamina (B2 and B3) and the nucleolus (B2) (Rao *et al.*, 2014; Chen *et al.*, 2018; Quinodoz *et al.*, 2018). B1 is enriched in the repressive H3K27me3 mark, which is often associated with polycomb binding. To relate sub-compartment status to chromatin dissociation rates, we compared the residuals of $t_{1/2}$ (after correcting for fragmentation level using DpnII-seq) for loci located in the 5 sub-compartments defined for K562 cells ((Xiong and Ma, 2018), Figure 2.12, A). We find that residual $t_{1/2}$ varies greatly between sub-compartments: $t_{1/2}(A1) \sim t_{1/2}(B1) < t_{1/2}(A2) < t_{1/2}(B2) \sim t_{1/2}(B3)$.

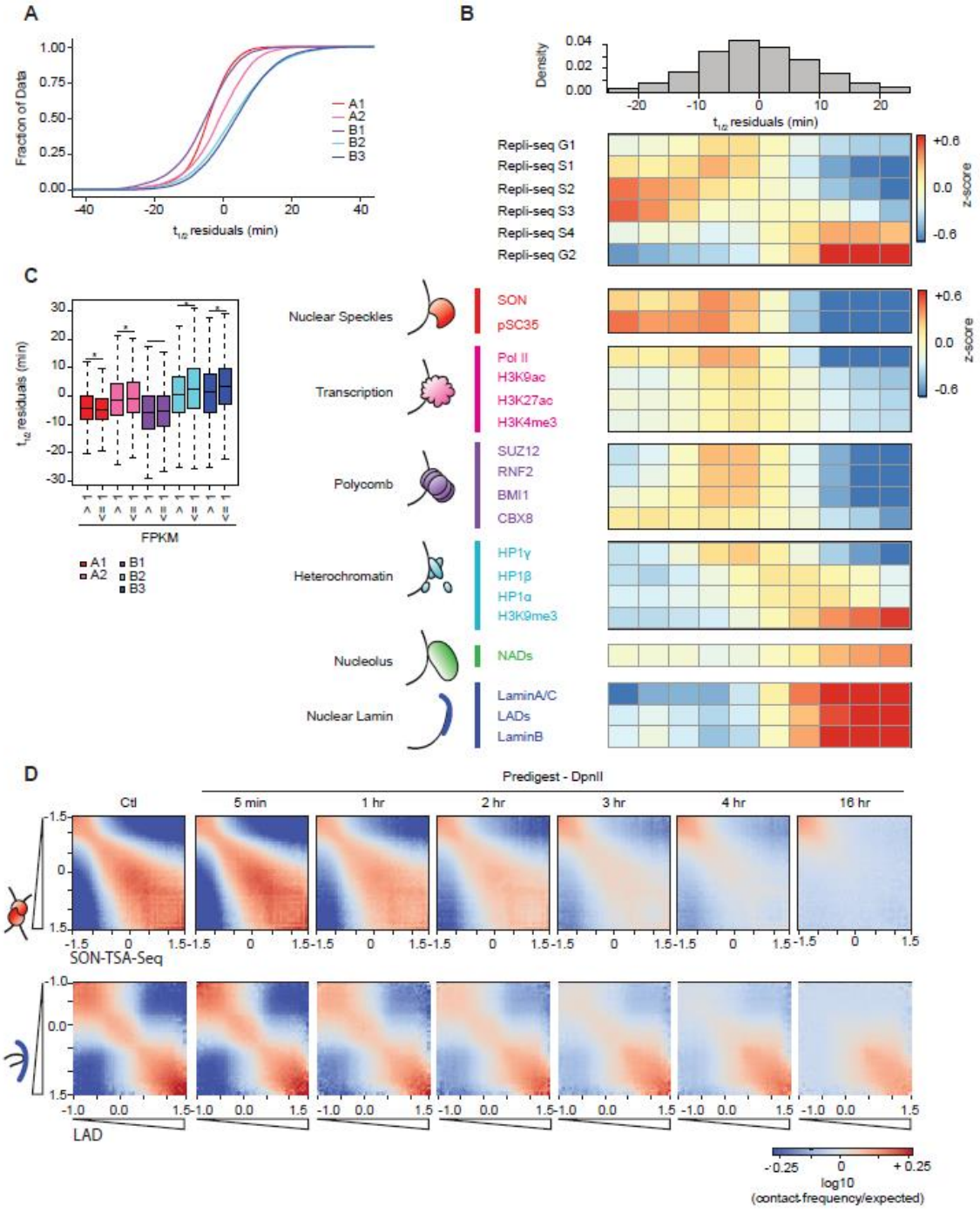


Figure 2.12 | Dissociation kinetics of chromatin interactions at different sub-nuclear structures.

(**A**) Cumulative distributions of residuals of $t_{1/2}$ (in minutes) for each of the five annotated sub-compartments. (**B**) Top: the genome was split into 10 bins, where each bin corresponds to sets of loci that share the same $t_{1/2}$ residual interval.

Middle: For each $t_{1/2}$ residual interval a heatmap of mean z-score signal of Repli-Seq data in different phases of the cell cycle G1, S1-4, G2. Bottom: For each $t_{1/2}$ residual interval a heatmap of mean z-score signal enrichment was quantified for various markers of sub-nuclear structures (See Methods). (**C**) Boxplot of $t_{1/2}$ residuals for bins with expressed genes (mean FPKM > 1) and bins with low or no expression (mean FPKM <=1) stratified by sub-compartment. Significance determined by two-sample two tailed t-test *($p < 0.003$). (**D**) Homotypic interaction saddle plots for loci ranked by their association with speckles (as detected by SON-TSA-seq, top 3) and by their association with the nuclear lamina. Preferential pair-wise interactions between loci associated with the lamina can still be observed after several hours, whereas preferential pair-wise interactions between loci associated with speckles are lost more quickly.

It is noteworthy that interaction dissociation rates for loci in the polycomb related B1 sub-compartment are as high or higher (residuals of $t_{1/2}$ as or more negative) than those for loci in the active and open A1 sub-compartment. Interactions between Lamin-associated loci in the B3 sub-compartments dissociate the slowest while interactions between loci in the B2 sub-compartment dissociate somewhat faster. These observations indicate that loci associated with different sub-nuclear structures display a range of interaction stabilities.

We coarse-grained $t_{1/2}$ residuals by splitting them into 10 intervals and then assigned these intervals genome-wide. We then explored the enrichment for varying chromatin features for each $t_{1/2}$ residual interval (Figure 2.12, B). Chromatin interactions for early replicating domains had short half-lives, while interactions for loci in later replicating domains were more stable (Figure 2.12, B). Interestingly, loci with the shortest $t_{1/2}$ residuals replicate in the middle of S-phase.

We find that loci near the speckle-associated proteins pSC35 or SON are engaged in the most unstable interactions. Similarly, transcriptionally active loci, identified by ChIP-seq for a range of histone modifications and factors associated with open chromatin such as H3K4me3 and RNA PolII, were also involved in relatively unstable chromatin interactions.

Interactions for loci bound by polycomb complexes (a subset of which are in the B1 sub-compartment) were as unstable as active speckle associated loci (Figure 2.12, B, Figure 2.13, B). Half-lives differed for loci bound by different polycomb subunits. Loci with the shortest $t_{1/2}$ residual values are enriched for binding the CBX8 subunit. An example of a large polycomb-bound domain in K562 cells is the HoxD cluster. The cluster is ~100 kb and covered by polycomb subunits Suz12, RNF2, CBX8 and BMI1 and the histone modification H3K27me3 (Figure 2.13, C). The half-life of chromatin interactions for loci in the HoxD cluster is relatively short.

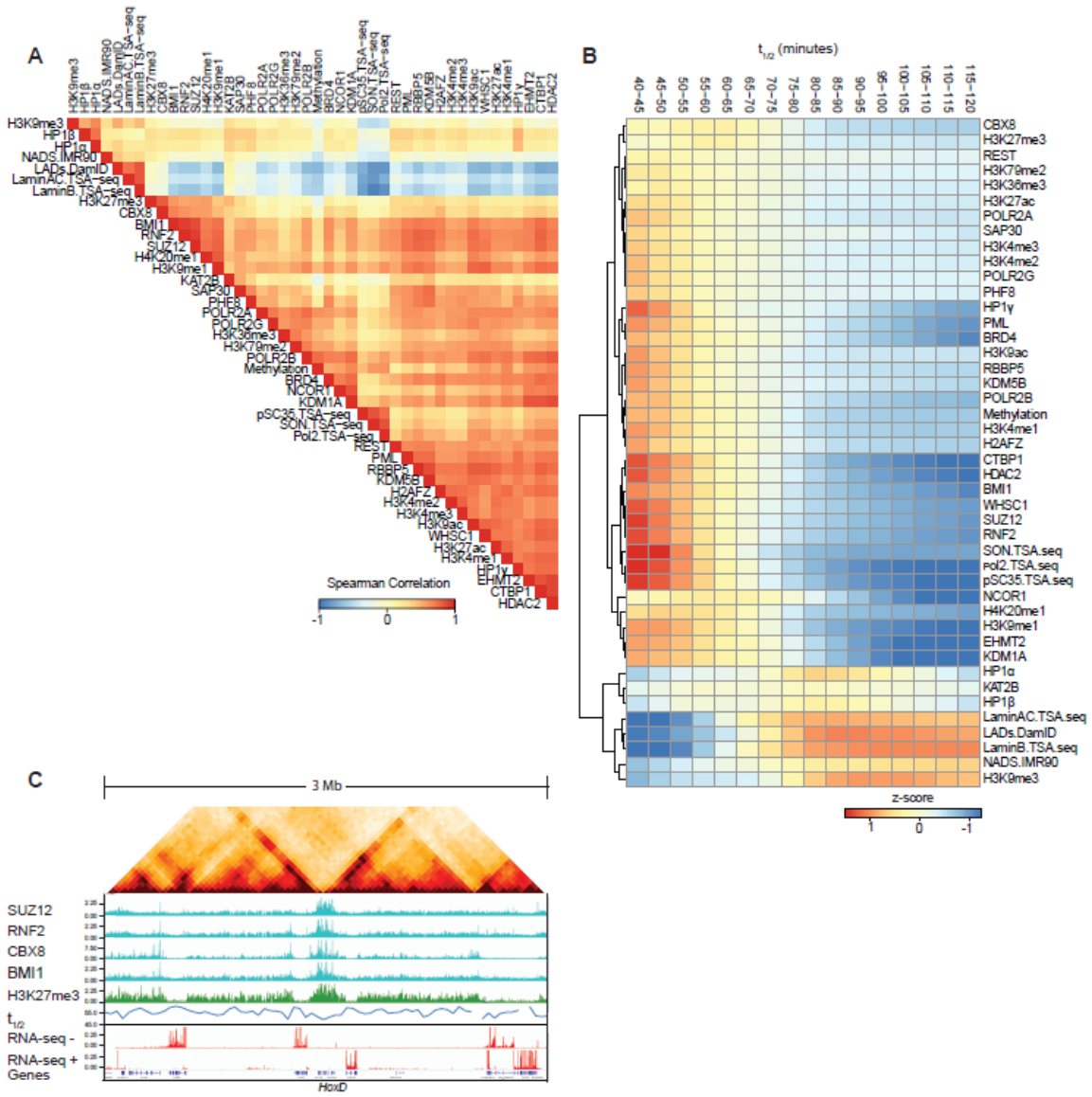


Figure 2.13 | Associations between sub-nuclear structures and chromatin interaction stability.

(**A**) Spearman correlation matrix between signals for various chromatin state markers of various sub-nuclear structures, chromatin remodellers and histone modifications with row order determined by hierarchical clustering. (**B**) The genome was split into 16 bins, where each bin corresponds to sets of loci that share the same $t_{1/2}$ interval. For each $t_{1/2}$ interval the mean z-score signal enrichment for various markers of sub-nuclear structures, chromatin remodellers and histone modifications was calculated and shown as a heatmap. Row order determined by hierarchical clustering. (**C**) 3 Mb region surrounding HoxD locus. Top: Hi-C contact map for K562 control nuclei showing the position of the HoxD

locus. Tracks: ChIP-seq tracks for polycomb subunits (cyan) and the polycomb associated histone modification H3K27me3 (green). $t_{1/2}$ (blue). Minus strand and plus-strand signal of total RNA-seq (red). Refseq Genes (blue/black). The polycomb-bound domain displays shorter half-life compared to expressed genes in flanking regions.

Silent and closed chromatin loci around the nucleolus or at the nuclear lamina were engaged in the most stable interactions (Figure 2.12, B). Chromatin interactions for loci associated with HP1γ (CBX3) were relatively unstable while interactions for loci associated with HP1β (CBX1) or HP1α (CBX5) were more stable. This variation is in agreement with the chromosomal locations and dynamics of these three HP1 proteins. HP1γ is associated with active chromatin and mobile, while HP1α and HP1β are typically found in constitutive heterochromatin near (peri) centromeres and are much less mobile (Dialynas *et al.*, 2007).

For each sub-compartment, we split loci into expressed (FPKM>=1) or not expressed (FPKM<1) categories (Figure 2.12, C). We find that sub-compartment status is the major determinant of chromatin interaction stability, irrespective of transcriptional status. However, transcriptional status modulates $t_{1/2}$ to some extent: in general, loci located in B2 and B3 sub-compartments are engaged in relatively stable chromatin interactions, but interactions that involve loci that are expressed have shorter half-lives. The expression status of loci located in the A1, A2 sub-compartments had only very minor effect on the $t_{1/2}$.

The differential stability of pair-wise chromatin interactions at different sub-nuclear structures can be quantified by plotting interaction frequencies between pairs of 40 kb loci arranged by their level of factor binding to obtain homotypic interaction saddle plots (Figure 2.12, D). In these plots, pair-wise interactions between loci enriched in factor binding are shown in the lower right corner, and pair-wise interactions between loci not bound by the factor are shown in the upper left corner. After chromatin fragmentation we observe loss of preferential interactions between speckle associated loci, while preferential interactions between non-speckle associated loci can be observed even after 16 hours. Conversely, preferential interactions between lamin-associated loci remain detectable even at late time points, while interactions between loci not at the lamina disappear relatively fast**.**

Chromatin loops dissociate upon chromatin fragmentation

Enriched point-to-point looping interactions are detected as “dots” in Hi-C interaction maps. The majority of these represent cohesin-mediated interactions between pairs of convergent CTCF sites (Rao *et al.*, 2014). We aggregated Hi-C data at pairs of sites that had previously been shown to engage in looping interactions in K562 cells (Rao *et al.*, 2014). We readily detected these loops in intact purified nuclei (Figure 2.14, A). After fragmentation with HindIII for 4 hours, loops appeared to become slightly stronger. Fragmenting chromatin with DpnII resulted in loss of loops over time.

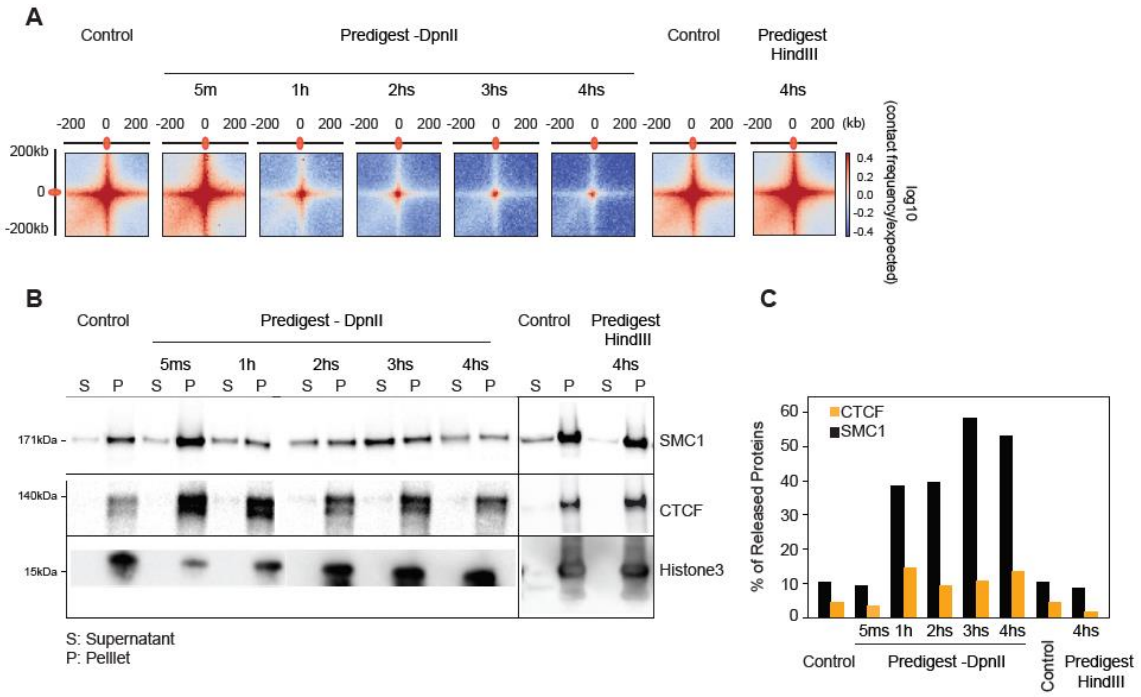


Figure 2.14 | Chromatin loop dissociation upon fragmentation.
(**A**) Aggregated distance-normalized Hi-C interactions around 6,057 loops detected in K562 cells by HiCCUPS (Rao *et al.*, 2014) at 10 kb resolution, for control nuclei and nuclei digested with DpnII up to 4 hours, and for nuclei digested with HindIII for 4 hours. (**B**) Western blot analysis of CTCF, cohesin and Histone H3 abundance in soluble and chromatin-bound fractions obtained from control nuclei and from nuclei pre-digested with DpnII up to 4 hours and HindIII for 4 hours. (**C**) Quantification of the data shown in panel B. Percentage of released protein is the ratio of protein level in the soluble fraction divided by the sum of the levels in the soluble and chromatin-bound fractions.

We assessed whether CTCF and cohesin binding to chromatin is affected by chromatin fragmentation. We fractionated proteins in chromatin-bound and soluble fractions ((Liang and Stillman, 1997), Methods). In intact nuclei, most of the CTCF and cohesin is associated with chromatin (Figure 2.14, B and C). Digesting chromatin with HindIII did not lead to dissociation of CTCF or cohesin.

However, fragmenting chromatin with DpnII led to dissociation of cohesin after 1 hour, while CTCF binding was only weakly affected. We conclude that DNA fragmentation to <6 kb fragments, but not to 10-25 kb fragments, leads to loss of cohesin binding and loss of looping interactions. These results are consistent with earlier observations that showed that in yeast stable chromatin binding by cohesin requires intact DNA (Ciosk *et al.*, 2000). These data can be interpreted in the context of the model where cohesin rings encircle DNA (pseudo-) topologically (Srinivasan *et al.*, 2018). Possibly, when DNA is fragmented, the cohesin ring can slide off nearby free ends**.**

Discussion

Using liquid chromatin Hi-C we obtained a view of the dynamics of chromatin interactions throughout the nucleus and the genome (Figure 2.15, A). Previously, live cell imaging experiments found differences in mobility dependent on sub-nuclear position and chromatin state and activity (Marshall *et al.*, 1997; Hediger *et al.*, 2002; Thakar, Gordon and Csink, 2006; Bronshtein *et al.*, 2009, 2015; Therizolsa *et al.*, 2010; Shinkai *et al.*, 2016; Nagashima *et al.*, 2019). In such experiments the movement detected is strongly constrained by the fact that loci are part of very long chromosomes. A previous study, which inspired the current work, aimed to identify factors that determine intrinsic locus-locus interactions and locus mobility by removing the polymeric constraint due to linkage (Gartenberg *et al.*, 2004). In that work a silent locus was excised from the chromosome (Gartenberg *et al.*, 2004) and its mobility and preference for

association with other silent loci and the nuclear periphery was found to depend on specific silencing complexes. In our liquid chromatin Hi-C experiments, the polymeric constraint on movement is removed for all loci simultaneously, in effect performing a genome-wide variant of the experiments performed by Gartenberg et al.

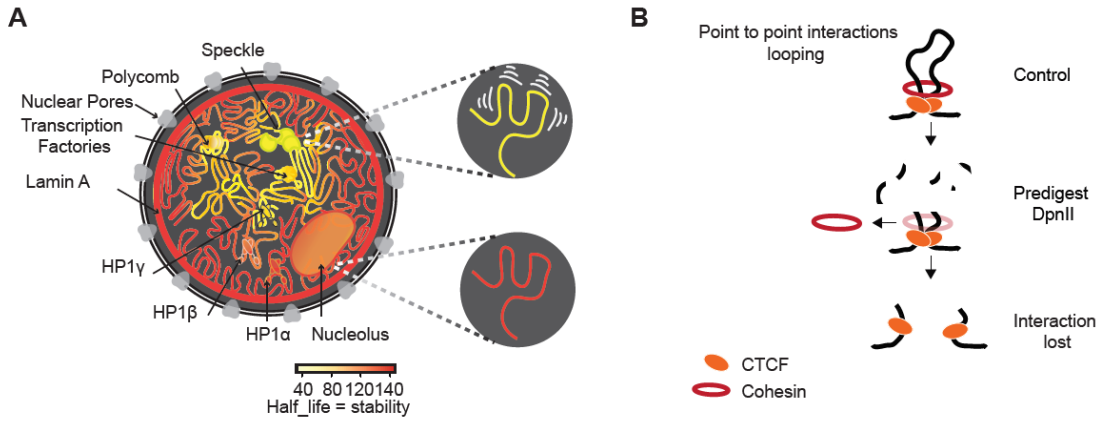


Figure 2.15 | Illustration of chromatin interaction dynamics in the nucleus and model for cohesin loss after chromatin digestion.
(**A**) Left: Schematic representation of varying chromatin interactions dynamics at different sub-nuclear domains. Shortest half-life reflects the least stable interactions (yellow), while longest half-life reflects the most stable interactions (dark orange). Nuclear subdomains differ greatly in their stability. Top right: Chromatin anchored at speckles is driven by the most dynamic interactions. Bottom right: Chromatin anchored at the nuclear lamina involves the most stable interactions. (**B**) Model for how cohesin rings stabilize CTCF-CTCF loops by encircling loop bases. Top: Cohesin ring encircles loop bases at convergent CTCF sites. Middle: Pre-digestion with DpnII cuts loop into chromatin fragments <6 kb, and the cohesin ring can slide off nearby ends. Bottom: CTCF remains bound to digested chromatin fragments but interactions between CTCF-bound sites are lost

Chromosomal compartmentalization tolerates genome-wide fragmentation with HindIII in >10-25 kb fragments. Micro-mechanical measurements also show

that chromosomes remain mechanically fully connected. We conclude that stable chromosome conformation and phase segregation can occur when blocks of a particular chromatin state are at least 10 kb. Our results obtained with DpnII digestion where the genome is fragmented in <6 kb (average ~3.1 kb) fragments show that these fragments are too short to maintain phase-segregated domains. The stability of interactions between <6 kb fragments depends on their chromatin state and association with sub-nuclear structures: interactions at the nuclear lamina are relatively stable, those near nuclear speckles and polycomb complexes are highly unstable, while interactions for loci associated with different heterochromatin proteins and the nucleolus displayed a range of intermediate stabilities. The dynamics of associations between loci are therefore determined by chromatin-associated factors, and may also be determined directly by the biochemical properties of histone tail modifications. For instance, the Rosen lab found that chromatin fragments can form droplets in vitro and that the dynamics of chromatin fragments within these droplets are dependent upon both H1 binding and histone acetylation (Gibson *et al.*, 2019).

Loci associated with the three different HP1 proteins display different dissociation kinetics. Interestingly, these differences correlate with different dynamics and sub-nuclear locations of the HP1 proteins in the nucleus. HP1γ (CBX3) binds relatively transiently to euchromatin, which may explain the dynamic nature of chromatin interactions between loci bound by this protein (Dialynas *et al.*, 2007). On the other hand, interactions between loci enriched in

HP1α (CBX5) are more stable, which likely is the result of more stable HP1α binding to heterochromatin (Strom *et al.*, 2017). Our finding that lamin-associated and HP1α-associated loci are engaged in the most stable interactions is consistent with recent modeling of inverted nuclei where heterochromatin is located in the center of the nucleus. Such inverted arrangement can only occur when heterochromatic interactions are much more stable or abundant than interactions between active chromatin loci (Falk *et al.*, 2019).

Liquid chromatin Hi-C showed differences in chromatin interaction stability between facultative heterochromatic domains marked by polycomb and constitutive heterochromatic domains marked by lamina association or binding of HP1α/HP1β proteins. While many chromatin contacts in constitutive heterochromatin were maintained even after 16 hours of digestion, the half-life for chromatin contacts at polycomb-bound regions was short, on a scale similar to active regions of the genome. The compacted states of polycomb and HP1α bound chromatin appear to form via a similar phase-separation mechanism mediated by multivalent interactions between specific CBX homologs. *In vitro* and *in vivo*, both CBX2 (polycomb subunit) and CBX5 (HP1α) are capable of forming condensates of polycomb bodies and constitutive heterochromatin, respectively (Larson *et al.*, 2017; Plys *et al.*, 2019; Tatavosian *et al.*, 2019). Our data indicate that these different condensates and associated chromatin have very different properties: the stability of interactions between loci mediated by these factors is distinct, possibly related to differences in affinity between CBX

proteins and chromatin: the binding affinity of CBX5 (HP1α) for its preferential histone modification H3K9me3 is higher than the affinity of CBX2 for its preferential mark H3K27me3 (Kaustov *et al.*, 2011). This differential could have consequences for the ability of cells to regulate genes embedded within these different types of heterochromatin.

Our results allow a crude estimate of the Flory-Huggins χ parameter for A/B segregation of chromatin in the nucleus. Given that HindIII and DpnII cut chromatin into segments of approximately 17±7 kb and 3±1 kb respectively (with FatI also generating approximately 3 kb average sized fragments), a reasonable estimate of the minimum length of fragments necessary to drive A/B segregation is N* = 10±4 kb. Given that these fragments are small compared to the A/B compartmentalization scale of a few Mb, the fragments will be essentially A- or B-type (euchromatin or heterochromatin) homopolymers. The similar ~1.2 Mb average size of the A and B compartments, as well as the remarkably similar ~3 kb fragment sizes observed for DpnII digestion of A and B regions (Figure 2.7, F) indicate that the simple Flory-Huggins model can be applied to a concentrated solution of symmetrical (equal length and concentration) A and B homopolymers.

For homopolymers, the Flory-Huggins model predicts a critical length needed for phase separation of $N^*=2/\chi$ (de Gennes, 1979), indicating $\chi = 0.20 \pm 0.07$/kb ($\chi = 0.036 \pm 0.013$ /nucleosome). Given that χ is in k_BT units (1 k_BT = 0.6 kcal/mol at physiological temperatures), its small value indicates that the effective demixing interaction between nucleosomes is weak, consistent with a liquid-like phase-

separation picture for A and B compartments, where regulation of chromatin organization and compartmentalization is possible by relatively small changes in nucleosome interactions (e.g., via histone modifications). While extremely crude (e.g., we have used the χ estimate for a dense polymer melt rather than for a concentrated solution) our data clearly indicate a weak value for χ, and that the interactions that are driving compartmentalization are a small fraction of a kcal/(mol·nucleosome). Our result is within a factor of 2 of a recent theoretical estimate (critical size for demixing of 20 kb, (MacPherson, Beltran and Spakowitz, 2018)). We note that we have been able to rather use a simple application of Flory-Huggins theory by virtue of the similar A/B fragment and compartment sizes. We emphasize that these properties do need to be examined before similar estimates can be done for other cell types or species.

We also note that our estimated χ assumes that the fragments are able to equilibrate their positions in both the DpnII and HindIII cases. It is conceivable that entanglements, cross-bridging or other effects may strongly retard or preclude demixing in the HindIII case. Such effects would be a confounding factor in estimation of χ, but the DpnII data do solidly constrain the critical size for demixing to be larger than 3 kb (i.e., $\chi < 0.66$/kb). Further experiments on the kinetics of fragment demixing would be very interesting in this regard, as they would shed further light on the physical processes underlying A/B compartment formation in vivo.

It is important to point out that during the liquid chromatin Hi-C procedure some chromatin factors and RNAs may dissociate from the purified nuclei, and this could affect the locus mixing behavior we observe. The current work analyzed the intrinsic chromatin interaction strengths and dissolution kinetics of chromosome conformation within otherwise inactive nuclei. Future work should focus on how these kinetic properties change in cells or nuclei, where active processes such as transcription, replication, chromatin compaction and condensation, and loop extrusion are also acting, and on determining the roles of RNAs, protein complexes, and histone modifications in modulating the attractive forces between loci and the dynamics of genome folding in general.

Chromatin Feature	Assay	ENCODE Accession ID	Output type	PubMed ID
H3K36me3	ChIP-seq	ENCFF223BKS	fold change over control	
H3K27ac	ChIP-seq	ENCFF006RIO	fold change over control	
H3K4me1	ChIP-seq	ENCFF463AQS	fold change over control	
H3K4me2	ChIP-seq	ENCFF778DNU	fold change over control	
H3K4me3	ChIP-seq	ENCFF330TJF	fold change over control	
H4K20me1	ChIP-seq	ENCFF242XLB	fold change over control	
H2AFZ	ChIP-seq	ENCFF430FSQ	fold change over control	
H3K9me1	ChIP-seq	ENCFF526UWC	fold change over control	
H3K9me3	ChIP-seq	ENCFF700FQH	fold change over control	
H3K9ac	ChIP-seq	ENCFF527JUP	fold change over control	
H3K27me3	ChIP-seq	ENCFF936BVT	fold change over control	
H3K79me2	ChIP-seq	ENCFF619JRY	fold change over control	
HP1β	ChIP-seq	ENCFF985MBQ	fold change over control	
HP1γ	ChIP-seq	ENCFF630YDI	fold change over control	
HP1α	ChIP-seq	ENCFF014XHT	fold change over control	
EHMT2	ChIP-seq	ENCFF982RMW	fold change over control	
CBX8	ChIP-seq	ENCFF206WVX	fold change over control	
RNF2	ChIP-seq	ENCFF071CIY	fold change over control	
BMI1	ChIP-seq	ENCFF514QBY	fold change over control	
SUZ12	ChIP-seq	ENCFF363DWX	fold change over control	
RBBP5	ChIP-seq	ENCFF058KTS	fold change over control	
CTBP1	ChIP-seq	ENCFF112LWM	fold change over control	
KAT2B	ChIP-seq	ENCFF164NLF	fold change over control	
BRD4	ChIP-seq	ENCFF260JHC	fold change over control	
NCOR1	ChIP-seq	ENCFF292YLV	fold change over control	
KDM5B	ChIP-seq	ENCFF293III	fold change over control	
HDAC2	ChIP-seq	ENCFF915GWT	fold change over control	
SAP30	ChIP-seq	ENCFF816KCQ	fold change over control	
WHSC1	ChIP-seq	ENCFF864WOR	fold change over control	
PHF8	ChIP-seq	ENCFF427QTV	fold change over control	
REST	ChIP-seq	ENCFF518QUW	fold change over control	
KDM1A	ChIP-seq	ENCFF734YKJ	fold change over control	
POLR2A	ChIP-seq	ENCFF749YKR	fold change over control	
POLR2B	ChIP-seq	ENCFF452WGO	fold change over control	
POLR2G	ChIP-seq	ENCFF015NSS	fold change over control	
PML	ChIP-seq	ENCFF157IUB	fold change over control	
Methylation	WGBS	ENCFF867JRG	methylation state at CpG	
NADS.IMR90	aCGH		NAD state	28575119
SON	TSA-seq		log2(pull-down / input)	30154186
pSC35	TSA-seq		log2(pull-down / input)	30154186
LaminA/C	TSA-seq		log2(pull-down / input)	30154186

LaminB	TSA-seq		log2(pull-down / input)	30154186
Pol2	TSA-seq		log2(pull-down / input)	30154186
LADs	DamID		log2(Dam-LaminB1/Dam)	30154186
G1	Repli-seq	ENCFF001GRX	percentage normalized signal	
S1	Repli-seq	ENCFF001GSF	percentage normalized signal	
S2	Repli-seq	ENCFF001GSJ	percentage normalized signal	
S3	Repli-seq	ENCFF001GSN	percentage normalized signal	
S4	Repli-seq	ENCFF001GSP	percentage normalized signal	
G2	Repli-seq	ENCFF001GSB	percentage normalized signal	

Table 2.1 | Public datasets used to associate liquid chromatin Hi-C measured stability with various chromatin features
Data used to associate $t_{1/2}$ with various chromatin histone modifications, transcription factors, DNA-binding proteins, sub-nuclear structures, and replication timing (Figure 2.12, B and D, Figure 2.13, A and B).

Materials and methods

Digestion, cross-linking and copolymer architecture and hetero/euchromatin phase separation

Chromatin in the G1 nucleus can be considered as a set of blocks of euchromatin and heterochromatin (the A and B compartments consisting of regions of predominantly euchromatin vs heterchromatin, respectively), which are constrained to be near each other by being part of the same linear chromosomes, i.e., effectively being long many-block copolymers. We suppose that the A and B heterochromatin/euchromatin monomers have a weak tendency to repel one another (or equivalently that A-A or B-B attract one another, for example via protein-mediated nucleosome-nucleosome interactions acting preferentially on euchromatin or heterochromatin, or even via physio-chemical effects such as relative hydrophobicity of more methylated nucleosomes).

If we suppose the A and B blocks to be on average N monomers long (roughly nucleosomes for the sake of this discussion), then under melt-like conditions the standard Flory theory of polymer phase separation predicts that if we were to cut the polymers into pure A and B blocks at the block boundaries (i.e., at a spacing of N monomers commensurate with the block sizes), they would phase separate for a segment-segment interaction strength stronger than $\chi^* = 2/N$ (de Gennes, 1979). Note that this level of interaction (given approximately in k_BT units) is proportional to 1/N where N is crudely in nucleosome units; for 200 kilobase blocks, we have approximately N=1000,

indicating that small fractions of a k_BT in effective A/B repulsion or A-A or B-B attraction is sufficient to drive strong euchromatin/heterochromatin phase separation (Marko and Siggia, 1997).

Now if we were to instead cut less frequently than this, say at every second block boundary, so as to arrive at a system of AB linear diblock copolymers each of length 2N (N monomers of A followed by N monomers of B), the constraint that the A and B blocks be connected suppresses phase separation, increasing the critical interaction (all other factors held constant) to $\chi^* = 5.3/N$ (Leibler, 1980). In this case bulk phase separation cannot occur, but instead local, or "microphase separation" occurs, with formation of micelle-like or layered phase-separated structures. Nevertheless, for $\chi >> \chi^*$, strong segregation of the A and B monomers can still occur.

If we were to not cut at all, but rather to suppose that the chromosomes are very long multiblock copolymers, with many blocks each of N monomers alternating between A and B ("ABABABAB... multiblock copolymers"), the critical interaction strength will rise with increasing number of blocks, approaching the limit $\chi^* = 7.5/N$ for many blocks (Matsen and Schick, 1994). Therefore, starting from this limit, the tendency for chromosome domains to phase separate will be enhanced by cutting the chromosomes up into successively smaller pieces: as chromatin cutting increases from no cutting, we expect to see intensification of A/B compartment contrast in the Hi-C map.

Now, if we cut too frequently, when the cuts become spaced smaller than the block size (cut spacing M < N monomers), we will have the situation that the critical interaction strength will become $\chi^* = 2/M > 2/N$, i.e., the cuts are frequent enough to suppress phase separation by decreasing the amount of interaction enthalpy per polymer "molecule". Therefore we expect that overly frequent cutting will cause a reduction in A/B compartment Hi-C map contrast, i.e., for some intermediate level of cutting similar to the sizes of the A and B blocks, one will see a maximum level of A/B compartment contrast.

There is also likely an effect of “crosslinking” ("chromatin cross-bridging"), which provides an additional level of constraint suppressing phase separation, above the linear-multiblock architecture of chromosomes. For example, taking linear diblock copolymers (N A monomers followed by N B monomers) and circularizating them raises the critical interaction for microphase separation from 5.3/N to 8.9/N, nearly a factor of 2 (Marko, 1993).

Similarly, if we start with A and B homopolymers each of length N, constraining them to have their ends at a flat surface, thus forcing them to mix at the surface, increases the critical interaction for phase separation from 2/N up to 4.5/N (Marko and Witten, 1991), with microphase separation again occurring in the constrained case. Releasing chromatin crosslinking/cross-bridging constraints (which also will occur for chromatin cutting) will in general also reduce the interaction strength needed to drive phase separation, increasing A/B compartment contrast in Hi-C maps.

In conclusion, basic polymer phase separation theory predicts that gradually increasing the cleavage of chromatin will gradually increase the intensity of A/B compartment contrast in Hi-C maps until the cuts are spaced by approximately one A or B block; further cutting will reduce the intensity of phase separation and A/B compartment contrast. Notably, the nature of the segregation can be expected to be "microphase segregation" rather than bulk phase separation, until the number of cuts is sufficient to liberate A or B "homopolymer" segments.

K562 nuclei purification

Three sucrose cushions were made before starting nuclei purification. 30 mL of 30% sucrose [10 mM PIPES pH 7.4, 10 mM KCl, 2 mM MgCl2, pH adjusted to 7.4 using 1 N KOH, 30% sucrose, 1 mM DTT (added prior to use), 1:100 protease inhibitor (Thermo Fisher 78438) (added prior to use)] was transferred to a 50 mL tube, then 5 mL of 10% sucrose [10 mM PIPES pH 7.4, 10 mM KCl, 2 mM MgCl2, 10% Sucrose, 1 mM DTT (added prior to use), 1:100 protease inhibitor (added prior to use)] was slowly loaded in top of 30% sucrose, and the tubes were incubated at 4°C until needed. K562 cell pellets (100 million cells) were lysed using the following nuclear isolation procedure. After the cells were spun, the pellets were washed twice with 10 mL HBSS, then pelleted after each wash at 300 rpm for 10 min at 4°C. Cell pellets were dissolved in 15 mL nuclear isolation buffer [10 mM PIPES PH 7.4, 10 mM KCl, 2 mM MgCl2, 1 mM DTT (added prior to use), 1:100 protease inhibitor (added prior to use)], pH

adjusted to 7.4 using 1 M KOH]. Then, cells were lysed on ice in a 15 mL Dounce homogenizer with pestle A (KIMBLE Kontes 885002-0015) by moving the pestle slowly up and down 20 times, followed by incubation on ice for 20 min and another 20 strokes. Next, each 5 mL of lysed extract was loaded slowly on top of a sucrose cushion prepared earlier. Then the tubes were spun for 15 min at 800 g at 4°C. The supernatant was removed carefully for a good recovery of the nuclei pellet in the bottom of the tube. Nuclei pellets were resuspended in 1 mL of HBSS, then spun for 5 min at 5,000 g at 4°C using a benchtop refrigerated centrifuge. Then, the nuclei pellet was resuspended in 3 mL HBSS, and 1 µL was taken to quantify the nuclei before the 3 mL was split over two microfuge tubes and spun for 5 min at 5,000 g at 4°C using a benchtop refrigerated centrifuge. Finally, the nuclei pellet was dissolved into an adequate total volume to obtain 1 million nuclei per 0.1 mL of Nuclei storage buffer (NSB) [10 mM PIPES pH 7.4, 10 mM KCl, 2 mM MgCl2, 50% glycerol, 8.5% sucrose, 1 mM DTT (added prior to use), 1:100 protease inhibitor (added prior to use)]. Each 0.5 mL of NSB containing 5 million nuclei was transferred to a microfuge tube and stored at -80ºC.

3C (Chromosome Conformation Capture)

3C was performed as described in “From cells to chromatin: Capturing snapshots of genome organization with 5C technology” (Ferraiuolo *et al.*, 2012).

Crosslinking: 1.25 mL of 37% formaldehyde was added to 40 mL of HBSS. 50 million cells or nuclei were washed twice using 20 mL of HBSS and then pelleted

at 500 g for 10 min. The pellet was resuspended in 5 mL HBSS and then added to 41.25 mL of HBSS and formaldehyde (final formaldehyde concentration was 1%). The sample was incubated at RT for 10 min on a rocking platform. Afterward, to stop cross-linking 2.5 mL of 2.5 M glycine was added and samples were incubated at RT for 5 min on a rotating platform. To pellet the crosslinked cells or nuclei the sample was centrifuged at 800 g for 10 min at 4°C. After discarding the supernatant the pellet was washed twice using HBSS. Next, the pellet was either processed immediately as described below or was stored at -80 ºC after flash freezing using liquid nitrogen.

Cell lysis: (This step was included when cells are used, but was skipped for 3C with purified nuclei).Cells were lysed by adding 2 mL of cold lysis buffer [10 mM Tris-HCl (pH=8.0), 10 mM NaCl, 0.2% Igepal CA-630 (NP40)] and 20 µL of 100x Protease inhibitors. The sample was incubated on ice for 15 min to let the cells swell. The cells were lysed on ice using the homogenizer with pestle A (KIMBLE Kontes 885300-0002) by moving the pestle slowly up and down 30 times and incubating on ice for 1 min followed by another 30 strokes. The sample was transferred to two 1.5 mL microcentrifuge tubes, spun at 5,000 g at RT for 5 min using a benchtop centrifuge.

Digestion: each pellet was washed using 1 mL cold 1X NEBuffer 2.1, then spun at 5,000 g for 5 min at RT using a benchtop centrifuge, afterward each pellet was resuspended in 250 µL of 1X NEB2.1 buffer, and the two pellets were pooled (~ 500 µL). 50 µL aliquots of the suspension were transferred to 10 new 1.5 mL

microfuge tube and 292 µL of 1x NEBuffer 2.1 was added to each tube. Next, 38 µL of 1% SDS was added per tube and mixed well, the samples were incubated at 65ºC for 10 min, then placed on ice. 44 µL of 10% Triton X-100 was added to each tube to quench SDS. Finally, 400 U of EcoRI (NEB R0101L) was added per tube and incubated at 37ºC overnight on a thermocycler (with 900 rpm for 30 sec every 4 min).

Ligation: 86 µL of 10% SDS was added to the digested samples and the samples were then incubated at 65ºC for 30 min for EcoRI inactivation after which the tubes were placed on ice. Each sample was then transferred to a 15 mL conical tube and 7.69 mL of ligation mix was added [820 µL 10% Triton X-100, 820 µL 10x ligation buffer (500 mM Tris-HCl pH7.5, 100 mM MgCl2, 100mM DTT), 82 µL 10 mg/mL BSA, 82 µL 100 mM ATP and 5.886 µL ultrapure distilled water]. Finally, 10 U of T4 ligase (Invitrogen 15224090) was added per tube before incubation at 16ºC for 2 hr on a thermocycler (with 900 rpm for 30 sec every 4 min).

Reverse Crosslinking: 50 µL of 10 mg/mL proteinase K (Fisher BP1750I-400) was added per tube, the sample was incubated at 65°C for 4 hr followed by a second addition of 50 µL 10 mg/mL Proteinase K and overnight incubation at 65°C on a thermocycler (with 900 rpm for 30 sec every 4 min).

DNA purification: Tubes were cooled at room temperature, at this stage each tube contains ~ 8.21 mL final volume. The samples from every two tubes were

combined to a 50 mL conical tube (~16,42 mL) to have five tubes in total. DNA was extracted by adding an equal volume of 17 mL of saturated phenol pH8.0: chloroform (1:1) (Fisher BP1750I-400) and vortexing for 3 min. Then the mix was transferred to a 15 mL phase-lock tube (Quiagen 129073) followed by spinning tubes at 5,000 g for 10 min. The upper phase was taken to a 50 mL tube to start the second extraction. We added an equal volume of 17 mL saturated phenol pH 8.0: chloroform (1:1), vortexing for 1 min. Then the upper phase was transferred to a 15 mL phase-lock tube, and tubes were centrifuged at 5,000 g for 10 min. We pooled all the upper phases from all 5 tubes ~ 85 mL into a single 300 mL high-speed centrifuge tube to precipitate the DNA. 8.5 mL (1/10 volume) of 3M sodium acetate pH 5.2 was added and brief vortexing was performed, then 212 mL (2.5 volumes) of ice-cold 100% ethanol was added, and the tube was inverted slowly several times and incubated at -80° C for 1 hr. Afterward, the DNA was pelleted at 16,000 g for 30 min at 4°C. The supernatant was discarded and the pellet was dissolved in 500 µL 1X TLE and transferred to a 0.5 mL AMICON Ultra Centrifuge filter (UFC5030BK EMD Millipore). The column was centrifuged for 5 min at 14,000 g and the flow-through was discarded. The column was washed 4 times using 450 µL of 1X TLE for desalting DNA. After the final wash, the library remaining in the column (~50 µL) was eluted in 30 µL of 1XTLE, the column was flipped upside down into a new tube to collect DNA by centrifugation for 3 min at 4,000 g. RNA was degraded by adding 1 µL of 10 mg/mL RNAase A and incubation for 30 min at 37°C.

Quality control assessment: to test the quality of the 3C library we used PCR to amplify a specific ligation product formed by two nearby restriction fragments, using the following primers:

GPF33: GACCTCTGCACTAGGAATGGAAGGTTAGCC

GPF23: GACTAATTCCTGACACTACTTGAGGGATAC

The amplicon was digested with EcoRI to assess the efficiency of 3C ligation.

BAC library for 3C-PCR

BAC DNA was generated as described (Dostie *et al.*, 2006). A control ligation library covering the Beta-globin locus (ENCODE region ENm009) was generated using BACs overlapping the region. Starting with a mixture of DNA of seven BACs (CTC-775N13, RP11-715G8, CTD-3048C22, CTD3055E11, CTD-2643I7, CTD-3234J1, and RP11-589G14) (Invitrogen), mixed in equimolar ratios, we used the same steps described in the 3C protocol above starting from the digestion step. BAC clones were digested with EcoRI, then randomly ligated, and the DNA was purified. The BAC ligation library reflects random ligation of EcoRI fragments throughout the beta-globin locus, so any difference in PCR signal for 3C primer pairs along the beta-globin locus due to differences in primer efficiency can be corrected by normalizing the amount of PCR product obtained with the 3C library to the amount obtained with the BAC ligation library.

Experimental design

Probes were designed as described (Dostie *et al.*, 2006). 213 5C probes were designed for a ~1 Mb region (chr11:4730996 -5729937; hg18) around the Beta-globin locus at EcoRI restriction sites using publicly available 5C primer design tools (Lajoie *et al.*, 2009). Probes were designed according to a single alternating scheme exactly as described before (Lajoie *et al.*, 2009) and the genomic uniqueness of all primers was verified with the SSAHA algorithm. For each EcoRI fragment at the 1 Mb target region a primer was designed. 104 5' forward (FOR) and 109 5' reverse (REV) primers were designed.

Generation of 5C libraries

5C libraries were generated as described before (Ferraiuolo *et al.*, 2012), with three modifications. First, we skipped the gel purification after the adaptor ligation and replaced this with a 1:1 Ampure step to remove unligated DNA and adaptors. Second, barcoded Illumina adaptors were used. Third, we performed the final PCR using TruSeq DNA LT kit Set A (REF 15041757).

Annealing: The 5C probes were pooled and combined with the 3C template each reaction contained 800,000 genome copies of 3C template and 0.2 fmol per 5C probe [800,000 genome copies of 3C template, 2 µL of 10X NEB4 (NEB B7004S), 2.75 µL of Salmon Sperm DNA (250 ng; (Invitrogen™ 15632011), 0.25 µL of 1 fmol/µL probes , up to 20 µL ultrapure distilled water]. We set up 8

annealing reactions for each library in a 96-well PCR plate. We then incubated the samples in a PCR machine and ran the following program [95°C for 9 min, Ramp 0.1°C/sec to 55 C, then keep at 55°C for 12 hr].

Ligation: We ligated 5C probe pairs, which represent a specific ligation junction in the 3C library, by adding 20 µL of ligation mix 2 µL of [10X Taq DNA ligase buffer (NEB B0208S), 0.25 µL Taq DNA ligase (NEB M0208S), 17.75 uL ultrapure distiller water] while the samples are kept in the PCR block at 55°C. We then incubated the reactions for 1 hr at 55°C followed by a 10 min incubation at 65°C; samples were then cooled to 4°C. Negative controls (no ligase, no template, no 5C oligonucleotide) were included to ensure the absence of any contamination.

PCR amplification: Universal emulsion primers were used for amplification of the ligated product by using 5C forward and reverse emulsion primers [Forward_primer: CCTCTCTATGGGCAGTCGGTGAT. Reverse_primer : CTGCCCCGGGTTCCTCATTCTCT] for 25 PCR cycles [6 µL of ligation product, 2.5 µL of 10XPCR (600 mM Tris-SO4, pH 8.9, 180 mM (NH4)2SO4), 1.8 mM MgCl2, 0.2 mM dNTP, 0.5 µL F-emulsion primer (80 µM), 0.5 µL R-emulsion primer (80 µM), 0.225 µL AmpliTaqÒ Gold DNA polymerase, ultrapure distilled water to bring volume up to 25 µL]. We then amplified DNA using this PCR program: [95° 9 min, 25 cycles (95°C 30 s, 65°C 30 s, 72° 30 s), 95°C 30 s, 65°C 30 s, 72°C 8 min, 4°C].

We pooled all the PCR reactions for the same library together and concentrated the DNA to 50 µL using 0.5 mL AMICON Ultra Centrifuge filter (UFC5030BK EMD Millipore). DNA was then loaded on a 2% agarose gel, along with a low molecular weight ladder, and the gel was run in a 4°C room at 200 volts for 90 min. The 150 bp DNA that corresponded to the ligated 5C probes was isolated from the gel using the QIAquick Gel Extraction Kit Protocol (QIAGEN 28115). DNA was finally eluted in 32 µL of 1XTLE.

A-tailing: A dATP was added to the 3' ends of the 5C library by adding 18 µL of A-tailing mix [5 µL NEB buffer 2.1, 10 µl of 1 mM dATP, 3 µL Klenow exo (NEB M0212S)] to the 32 µL of DNA sample from the previous step. The reaction was then incubated in a PCR machine [at 37°C for 30 min, then at 65°C for 20 min, and finally cooled down to 4°C]. Next, the tube was placed on ice immediately. 1:1 Ampure was used to remove unligated adaptors. The DNA was finally eluted in 40 µL 1X T4 DNA Ligase buffer (Invitrogen).

Illumina adapter ligation and paired-end PCR: For this step, we used the TruSeq DNA LT kit Set A (REF 15041757). 10 µL of ligation mix [5 µL Illumina paired-end adapters, 3 µL T4 DNA ligase Invitrogen, 2 µL 5x T4 DNA ligase buffer (Invitrogen 5X)] was added to the 40 µL sample from the previous step. The ligation sample was then incubated at RT for 2 hours on a tube rotator. Afterward, the sample was run on a 2% agarose gel in a cold room 4°C at 150 volts for 120 min along with a low molecular weight ladder. The 270 bp band that corresponds to 5C products (150 bp) ligated to the two adaptors (64 bp) was

extracted from the gel and isolated using the QIAquick Gel Extraction Kit (QIAGEN 28115). DNA was finally eluted in 30 µL 1XTLE.

Pre-digestion of nuclei (liquify chromatin)

Purified nuclei as described above (K562 nuclei purification) were placed on ice and 1 mL of HBSS was added to each 0.5 mL of 5 million frozen nuclei. After thawing, nuclei were centrifuged 5 min at 5,000 g. The nuclei pellet was washed twice with 1XNEB3.1 for nuclei that would be digested with DpnII or 1XNEB2.1 for nuclei that would be digested with HindIII. The nuclei were pelleted for 5 min at 5,000 g after each wash.

Isolated nuclei: a sample of 5 million nuclei was resuspended in 1,250 µL of 1X NEB3.1 as control, and then processed immediately for Hi-C starting at the crosslinking step (see below Hi-C 2.0 protocol).

Undigested nuclei: Each sample of two million nuclei was resuspended in 500 µL of 1X NEB3.1 on ice, as and control for the pre-digestion and then treated as described immediately below.

DpnII pre-digestion: Each sample of two million nuclei was resuspended in 500 µL of 1X NEB3.1 on ice. Next, 120 U of DpnII (NEB R0543S) was added to the sample in order to obtain 10 U DpnII/µg DNA and then treated as described immediately below.

FatI pre-digestion: Each sample of two million nuclei was resuspended in 500 µL of 1X NEB3.1 on ice. Next, 120 U of FatI (NEB R0650L) was added to the sample in order to obtain 10 U FatI/µg DNA and then treated as described immediately below.

HindIII pre-digestion: Each sample of two million nuclei was resuspended in 500 µL of 1X NEB2.1 on ice. Next, 600 U of HindIII (NEB R0104T) was added to the sample in order to obtain 50 U HindIII/µg of DNA and then treated as described immediately below.

Next, control and pre-digestion samples were incubated at 37°C on a thermocycler (900 rpm for 30 sec every 4 min) for 5 min up to 16 h. Afterward, samples were placed on ice for 10 min.For DpnII-seq and assessment of fragmentation level , a final volume of 10 mM of EDTA was added to inactivate the endonuclease, followed immediately by the DpnII-seq protocol (details of protocols below. DpnII-Seq) or DNA purification for fragment analyzer analysis. For Hi-C, we proceeded immediately to the first step of the protocol (crosslinking as described below). For microscopy, nuclei samples were cross-linked with a 4% final concentration of paraformaldehyde.

Hi-C 2.0

Hi-C was performed as described (Belaghzal, Dekker and Gibcus, 2017) with some modifications in the crosslinking and lysis step as described below. Hi-C was performed following the exact same protocol for mock treated and for pre-

digested nuclei, even though for pre-digested nuclei chromatin was already fragmented prior to fixation. There are several reasons to perform the full Hi-C procedure even for pre-digested nuclei. First, pre-digestion (e.g. with DpnII) leads to partial digestion. During the subsequent Hi-C procedure chromatin is digested with DpnII again, but this time the chromatin has been opened with SDS treatment. This allows more complete digestion. Second, by performing the full Hi-C procedure for pre-digested nuclei allows direct comparison of data obtained with mock treated nuclei, HindIII pre-digested nuclei, DpnII pre-digested nuclei and FatI pre-digested nuclei.

Crosslinking: isolated, undigested, and pre-digested (with liquified chromatin) nuclei were not pelleted after the pre-digestion step above but were crosslinked immediately as follows: for each sample 1,250 µL volume of nuclei in the digestion buffer was transferred to a 21.875 mL mix [625 µL of 37% formaldehyde + 21.25 mL of HBSS]. For intact cells: 5 million K562 cells or nuclei were washed twice with 15 mL of HBSS and pelleted at 300 g for 10 min, then resuspended in 2.5 mL of HBSS. The sample was transferred to 20.625 mL crosslinking mix [625 µL of 37% formaldehyde + 20 mL of HBSS].

All samples were incubated at RT for 10 min on a rocking platform. Next, to stop cross-linking 1.25 mL of 2.5 M glycine was added to each sample and the mix was incubated at RT for 5 min on a rocking platform. To pellet the crosslinked cells/nuclei, the sample was centrifuged at 1,000 g for 10 min at 4°C. The

supernatant was discarded and the pellet was washed twice with HBSS before going to the next step or storing samples at -80°C.

Cells lysis: This step is not needed for isolated, undigested, and pre-digested (with liquified chromatin) nuclei. For Hi-C with intact cells: the 5 million crosslinked cells were lysed by adding 1 mL cold lysis buffer [10 mM Tris-HCl (pH=8.0), 10 mM NaCl, 0.2% Igepal CA-630 (NP40)] and 10 µL of 100X Protease inhibitors. The sample was incubated on ice for 15 min to let the cells swell. The cells were lysed on ice using a dounce homogenizer with pestle A (KIMBLE Kontes 885300-0002) by moving the pestle slowly up and down 30 times and incubating on ice for 1 min followed by another 30 strokes. The sample was transferred to a new 1.5 mL microcentrifuge tube, and the sample was centrifuged at 5,000 g at RT for 5 min.

Digestion: from each sample (isolated undigested, and pre-digested (with liquified chromatin) nuclei and lysed cells) the pellet was resuspended in 500 µL of ice-cold 1X NEBuffer 3.1, and pelleted for 5 min at 4,000 g. The pellet was washed twice using 500 µL of ice-cold 1X NEBuffer 3.1. After the last wash, the pellet was resuspended in 350 µL 1X NEBuffer 3.1, and 8 µL was taken and kept at 4°C to assess the DNA integrity later. 38 µL of 1% SDS was added to 342 µL (380 µL total volume), and the mixture was resuspended and incubated for 10 min at 65°C. The tube was placed on ice immediately afterward. Next 43 µL of 10% Triton X-100 was added and the sample was mixed gently by pipetting. The

tubes were placed at room temperature and 12 µL of 10X NEBuffer 3.1 was added. Then 400 U of DpnII (R0543L) was added and mixed gently before an overnight incubation at 37°C on a thermocycler (with 900 rpm for 30 sec every 4 min).

Biotin Fill-in: After overnight digestion, the sample was incubated at 65˚C for 20 min in order to inactivate the restriction enzyme. Then, 10 µL of the digested sample was taken and kept at 4°C to assess the digestion efficiency later. DNA ends were marked with biotin-14-dATP by adding 60 µL of biotin fill-in master mix [1XNEB 3.1, 0.25 mM dCTP, 0.25 mM dGTP, 0.25 mM dTTP, 0.25 mM biotin-dATP (ThermoFisher#19524016), 50U Klenow polymerase Polymerase I (NEB M0210L)]. Next, the sample was incubated for 4 h at 23°C on a thermocycler (with 900 rpm for 30 sec every 4 min). Finally, the sample was placed on ice immediately for 15 min before proceeding to the next step.

Ligation: After fill-in, the total sample volume was ~535 µL. Ligation was performed by adding 665 µL of ligation mix [240 µL of 5x ligation buffer (1.8X) (Invitrogen), 120 µL 10% Triton X-100, 12 µL of 10 mg/mL BSA, 50 µL T4 DNA ligase (Invitrogen 15224090), and 243 µL ultrapure distilled water (Invitrogen)], to make a total volume of 1,200 µL. The reaction was then incubated at 16°C for 4 hours in a Thermomixer with interval shake.

Reverse Crosslinking: 50 μL of 10 mg/mL proteinase K (Fisher BP1750I-400) was added after ligation, the sample was incubated at 65°C for 4 hr followed by a second addition of 50 μL 10 mg/mL Proteinase K and overnight incubation 65°C

DNA purification: Reactions were cooled to room temperature and the 1.3 mL total volume was transferred to a 15 mL tube. The DNA was extracted by adding an equal volume of 1.3 mL of saturated phenol pH 8.0: chloroform (1:1) (Fisher BP1750I-400) and vortexing for 1 min. Then the total volume of 2.6 mL was transferred to a 15 ml phase-lock tube (Quiagen #129065) and tubes were centrifuged at 5,000 g for 10 min. The upper phase was transferred to a 15 mL tube to start the second extraction. An equal volume of 1.3 mL saturated phenol pH8.0: chloroform (1:1) was added and the sample was vortexed for 1 min. Then the mix was transferred to 15 ml phase-lock tube (Quiagen #129065) followed by spinning tubes at 5,000 g for 10 min. The upper phase of ~1.3 mL was transferred to a 15 mL tube (high speed) to precipitate the DNA. 1/10 volume (130 μL) of 3 M sodium acetate pH 5.2 was added and the sample was briefly vortexed. Then, 2.5 volumes of ice-cold 100% ethanol 3.25 mL was added, the tube was inverted slowly several times and then incubated at -80° C for 1hr. Next, the DNA was pelleted at 16,000 g for 30 min at 4°C. The supernatant was discarded and the pellet was dissolved in 500 μL 1X TLE and transferred to a 0.5 mL AMICON Ultra Centrifuge filter (UFC5030BK EMD Millipore). The column was spun for 5 min at 14,000 g and the flow-through was discarded. The column was washed 4 times using 450 μL of 1X TLE for desalting of DNA. After the final

wash the DNA remaining in the column (~50 µL) was eluted in 52 µL of 1XTLE. The column flipped upside down into a new tube to collect DNA and spun for 3 min at 4,000 g, the volume was adjusted to 102 µL. RNA was degraded by adding 1 µL of 10 mg/mL RNAase A and incubation for 30 min at 37°C. To quantify the DNA concentration, 2 µL of the final DNA sample along with the first 8 µL sample taken before digestion, the 10 µL sample taken after digestion, and various amounts of the 1 kb ladder (NEB#N3232s) were run on 1% Agarose gel.

Removal of Biotin from unligated ends: To remove biotinylated nucleotides at DNA ends that did not ligate, the Hi-C sample was treated with T4 DNA polymerase. For each Hi-C sample, we assembled the following reaction: [up to 5 µg of Hi-C library, 5 µL 10x NEBuffer 3.1, 0.025 mM dATP, 0.025 mM dGTP and 15 U T4 DNA polymerase (NEB # M0203L). The samples were brought up to 50 µL total volume adding ultrapure distilled water . Reactions were incubated at 20°C for 4 hours, the enzyme was then inactivated by incubation of the reaction for 20 mins at 75°C and placed at 4°C. Next, the samples were pooled and the volume was brought up to 130 µL 1XTLE in preparation for sonication.

Sonication: the DNA was sheared to a size of 100-300 bp using a Covaris instrument [Duty cycle 10%, Intensity 5, Cycles per Burst 200, set Mode Frequency sweeping, continuous degassing, process time 60 sec, Number of cycles] for 3 cycles. The volume was brought up to 500 µL using TLE for Ampure fractionation.

Size fractionation using AMpure XP: 500 µL AMpure beads (Beckman Coulter A63881) were added to a 1.5 mL tube labeled as 1.1X. Then the tube was placed on the Magnetic Particle Separator (MPS) for 5 min, and the supernatant was removed. Beads were resuspended in 150 µL AMpure mixture in order to make the 1.1X solution. 400 µL of AMpure mixture was added to 500 µL of sonicated DNA from the previous step and the tube was labeled 0.8X. The sample was vortexed and spun down briefly followed by incubation at RT for 10 min on a rotating platform. Then the tube was placed on the MPS for 5 min at RT. The supernatants were collected and added to the 1.1X tube, the tube was briefly vortexed and spun down followed by incubation at RT for 10 min on a rotating platform. Then the tube was placed on the MPS for 5 min at RT. The supernatant was discarded and the beads in 0.8X and 1.1X tubes were washed twice with 1 mL 70% ethanol. Beads were reclaimed by the MPS for 5 min. Beads were then air-dried on the MPS until ethanol had evaporated completely. Next, 51 µL of 1XTLE was added to the 0.8X and 1.1X tubes to resuspend the DNA from the beads. Tubes were incubated at RT on a rotating platform for 10 min. Then the tubes with AMpure beads from both 0.8X and 1.1X tubes were placed on the MPS for 5 min. Finally, the supernatants were transferred to 1.7 mL tubes labeled 0.8X and 1.1X. Our sample with DNA that ranges from 100-300 bp is in the 1.1X sample, the 0.8X sample was kept in case more DNA was needed. DNA from both samples 0.8X and 1.1X were quantified by running 1 µL on a 2%

agarose gel along with different amounts of low molecular weight DNA ladder (100 ng, 200 ng, 400 ng).

End Repair: 50 µL of Hi-C sample was transferred to a PCR tube, then 20 µL of the end-repair mix [3.5X NEB ligation buffer (NEB B0202S), 17.5 mM dNTP mix, 7.5 U T4 DNA polymerase (NEBM0203L), 25 U T4 polynucleotide kinase (NEB M0201S), 2.5 U Klenow polymerase Polymerase I (NEB M0210L)] was added. The 70 µL total volume reaction was then incubated at 37°C for 30 min, followed by incubation at 65°C for 20 min to inactivate Klenow polymerase, and then the sample was put at 4°C. The volume was brought up to 400 µL using 1X TLE for the next step.

Biotin pull-down: All the following steps were performed with 1.5 mL loBind tubes (Eppendorf 22431021). 15 µL of MyOne streptavidin C1 beads mix (Thermo Fisher 65001) was transferred to a 1.5 mL tube. The beads were washed twice by adding 400 µL of TWB [5 mM Tris-HCl pH8.0, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween20] followed by incubation for 3 min at RT. The tube was then placed on an MPS for 1 min and the supernatant was removed. After the washes, the beads were resuspended in 400 µL of 2X Binding Buffer (BB) [10 mM Tris-HCl pH8, 1 mM EDTA, 2 M NaCl] and mixed with the 400 µL DNA from the previous step in a new 1.5 mL tube. The mixture was incubated for 15 min at RT with rotation, the tube was then placed on the MPS for 1 min and the supernatant was removed. The DNA bound to the beads was washed by adding 400 µL of 1X BB and transferred to a new tube. The beads were reclaimed

against the MPS for 1 min, and the supernatant was discarded. The second wash used 100 µL of 1X TLE, beads were reclaimed against MPS for 1 min, and the supernatant was discarded. Finally, the DNA bound to the beads was eluted in 32 µL of 1X TLE.

A-tailing: A dATP was added to the 3’ ends by adding 18 µL of A-tailing mix [5 µL NEB buffer 3.1, 10 µL of 1 mM dATP, 3 U Klenow exo (NEB M0212S)] to the 32 µL of DNA sample from the previous step. The reaction was incubated in a PCR machine [at 37°C for 30 min, then at 65°C for 20 min, followed by cool down to 4°C]. Next, the tube was placed on ice immediately. The sample was transferred to a 1.5 mL loBind tube, the tube was placed on the MPS for 1 min and the supernatant was removed. The streptavidin beads bound to DNA were washed twice using 100 µL 1X T4 DNA Ligase Buffer (Invitrogen). Finally, streptavidin beads bound to DNA were resuspended in 40 µL 1X T4 DNA Ligase buffer (Invitrogen).

Illumina adapter ligation and paired-end PCR: For this step, the TruSeq DNA LT kit Set A (REF#15041757) was used. 10 µL of ligation mix [5 µL Illumina paired-end adapters, 3 µL T4 DNA ligase Invitrogen, 2 µL 5x T4 DNA ligase buffer (Invitrogen 5X)] was added to the 40 µL Hi-C sample from the previous step. The ligation sample was then incubated at RT for 2 hours on a rotator. The sample was transferred to a 1.5 mL loBind tube, the tube was placed on the MPS for 1 min and the supernatant was removed. The streptavidin beads bound to DNA

were washed twice with 400 µl of TWB, then twice using 100 µL 1X TLE. Finally, the sample was resuspended in 20 µL 0f 1XTLE.

Illumina Truseq Kit for PCR: We performed three trial PCR reactions as follows [2.5 µL DNA bound to beads, 2 µL of Primers mix (TruSeq DNA LT kit Set A 15041757)), 10 µL Master Mix (TruSeq DNA LT kit Set A 15041757), 10.5 µL of ultrapure distilled water (Invitrogen)]. We split the 25 µL over three PCR tubes (5 µL, 5 µL, 15 µL per tube). Each of the three samples was then amplified with different numbers of PCR cycles (6, 8, 10 respectively) to assess the Hi-C library quality: [30 sec at 98°C, n cycles of (30 sec at 98°C, 30 sec at 65°C, 30 sec at 72°C), 5 min at 72°C, hold at 10°C]. 10 µL was taken from the 15 µL sample (with 10 PCR cycles), the 10 µL sample was then digested with ClaI for 1 h by adding 10 µL of digestion mix [1.5 µL 10x NEB Cutsmart buffer, 1.5 µL ClaI (NEB R0197S), 7 µL ultrapure distilled water]. The 5 µL of each PCR cycle sample along with the 20 µL digested sample, and titration of the low molecular ladder (100 ng, 200 ng, 400 ng) (NEB) were run on a 2% Agarose gel. After digestion with ClaI, a downward shift of the amplified DNA to smaller sizes is expected, which indicates DNA ends were correctly filled in and ligated (creating a ClaI site). The number of PCR cycles to generate the final Hi-C material for deep sequencing was chosen based the minimum number of PCR cycles in the PCR titration that was needed to obtain sufficient amounts of DNA for sequencing using the remaining 17.5 µL Hi-C sample.

DpnII-Seq

For each DpnII-Seq library, 10 million nuclei were used right after the pre-digestion procedure described above (Pre-digestion of nuclei). The pre-digested nuclei were then treated as follows:

Proteinase K: 50 μL of 10 mg/mL proteinase K (ThermoFisher # 25530) was added to each 500 μL pre-digested nuclei sample (2 million nuclei) (See Methods: Pre-digestion) and the 5 tubes were incubated at 65°C for 3 hours.

DNA purification: Tubes were cooled to room temperature and all 5 samples were pooled in a single 15 mL tube (2.75 mL total volume). The DNA was extracted by adding an equal volume of 2.75 mL of saturated phenol pH8.0: chloroform (1:1) (Fisher BP1750I-400), followed by vortexing for 1 min. The sample (5.5 mL) was transferred to a 15 mL phase-lock tube (Quiagen #129065) followed by centriguation at 5,000 g for 10 min. The upper phase was transferred to a 15 mL tube to start the second extraction. An equal volume of 2.75 mL saturated phenol pH8.0: chloroform (1:1) was added, followed by vortexing for 1 min. Then the mix was transferred to a 15 mL phase-lock tube (Quiagen #129065) followed by centrifugation at 5,000 g for 10 min. The upper phase of ~ 2.75 mL was transferred to a 15 mL tube (high speed), 1/10 volume (275 μL) 3M sodium acetate pH 5.2 was added followed by brief vortexing and then 2.5 volumes of ice-cold 100% ethanol (6.875 mL) were added. The tube was inverted slowly several times, incubated at -80°C for 1 hr and then DNA was pelleted by

centrifugation at 16,000 g for 30 min at 4°C. The supernatant was discarded and the pellet was dissolved in 500 µL 1X NEB3.1 and transferred to a 0.5 mL AMICON Ultra Centrifuge filter (UFC5030BK EMD Millipore). The column was centrifuged for 5 min at 14,000 g and the flow-through was discarded. The column was washed 4 times using 450 µL of 1X NEB3.1 for desalting of DNA. After the final wash, the library remaining in the column (~50 µL) was eluted in 450 µL of 1XNEB3.1; the column was flipped upside down into a new tube to collect DNA and centrifuged for 3 min at 4,000 g. ~500 µL of DNA was recovered. RNA was degraded by adding 1 µL of 10 mg/mL RNAase A and incubation for 30 min at 37°C. The amount of DNA was estimated by running an aliquot on a 1% Agarose gel along with a 1kb ladder (NEB#N3232s).

Biotin Fill-in: 1XNEB3.1 was added the reaction to a final volume of 680 µL, and then the 680 µL was split over 2 1.5 mL tubes. DNA ends were filled in and marked with biotin-14-dATP. To each tube 60 µL of biotin fill-in master mix was added: [1xNEB2.1, 0.25 mM dCTP, 0.25 mM dGTP, 0.25 mM dTTP, 0.25 mM biotin-dATP (ThermoFisher#19524016), 50 U Klenow polymerase Polymerase I (NEB M0210L)]. Samples were incubated at 37°C in a Thermocycler for 75 mins. Next, the tubes were placed on ice immediately for 15 mins, and samples from the 2 tubes were combined to obtain a final volume ~800 µL. Amicon filters were used to reduce the volume of the final sample from 801 µL to 130 µL.

Sonication: DNA was sonicated to a size of 100 – 300 bp using a Covaris instrument (Duty Cycle 10%, Intensity 5, Cycles per Burst 200, set Mode

Frequency sweeping, continuous degassing, process time 60 sec, Number of cycles) for 4 cycles. The 130 µL of sonicated DNA was transferred to a 1.5 mL tube and 1XTLE was added to a total volume of 500 µL. DNA fragment size was determined by running 2 µL of DNA along with low molecular ladder (NEB) on a 2% agarose gel.

Size fractionation using AMpure XP: 500 µL AMpure beads (Beckman Coulter A63881) were added to a 1.5 mL tube labeled as 1.1X. The tube was placed on the MPS for 5 min, and the supernatant was removed. Beads were resuspended in 150 µL AMpure mixture in order to make 1.1X. 400 µL of AMpure mixture was added to 500 µL of sonicated DNA from the previous step and the tube was labeled 0.8X. The sample was vortexed and centrifuged briefly using a tabletop small centrifuge followed by incubation at RT for 10 min on a rocking platform. Then the tube was placed on the MPS for 5 min at RT. The 0.8X supernatants were collected and added to the 1.1X tube, the tube was briefly vortexed and centrifuged followed by incubation at RT for 10 min on a rocking platform. The tube was placed on the MPS for 5 min at RT, and the supernatant discarded. Beads in the 0.8X and 1.1X tubes were washed twice with 1 mL 70% ethanol, reclaiming beads against the MPS for 5 min. Beads on the MPS were then dried until ethanol had evaporated completely. Next, 51 µL of 1XTLE was added to the 0.8X and 1.1X tubes to elute DNA from the beads. Tubes were incubated at RT on a rocking platform for 10 min. The 0.8X and 1.1X tubes were placed on the MPS for 5 min. Finally, the supernatants were transferred to 1.7 mL tubes

labeled 0.8X and 1.1X. The 1.1X sample contains DNA that ranges in size from 100-300 bp. The DNA in the 0.8X sample was kept in case more DNA was required, in which case the DNA would be sonicated using 2 cycles followed by a similar round of size fractionation as described above. The amount of DNA from both samples 0.8X and 1.1X was quantified by running 1 µL on a 2% agarose gel along with a titration of low molecular weight DNA ladder (100 ng, 200 ng, 400 ng).

End Repair: 50 µL from the 1.1X sample was transferred to a PCR tube, and 20 µL of end repair mix was added: [3.5X NEB ligation buffer (NEB B0202S), 0.875 mM dNTP mix, 0.375 U/µL T4 DNA polymerase, 1.25 U/µL T4 polynucleotide kinase, 0.125 U/µL Klenow DNA polymerase]. The 70 µl total volume reaction was incubated for 30 min at 20°C in a PCR machine and then placed on ice. The DNA was purified by 1:2 Ampure, by adding 140 µL 2X Ampure solution to the 70 µL DNA sample followed by incubation for 5 min at RT. The tube was placed on the MPS for 4 min to reclaim the beads and the supernatant was discarded. The beads were washed twice with 1 mL of 70% ethanol while on the MPS. After beads were dried DNA was eluted in 32 µL TLE (pH 8.0) and incubation for 10 min at RT. The supernatant was transferred to a 1.5 mL tube.

A-tailing: A dATP was added to the 3’ ends by adding 18 µL of A-tailing mix [5 µL NEB buffer 3.1, 10 µL of 1 mM dATP, 3 U Klenow exo (NEB M0212S)] to the 32 µL of DNA sample from the previous step. The reaction was then incubated in a PCR machine at 37°C for 30 min followed by incubation 65°C for 20 min and

cooling down to 4°C. The tube was placed on ice. The volume was brought to 100 µL by adding 1X NEB2.1. The DNA was then purified by adding 1:2 Ampure mix (200 µL of Ampure was added to the 100 µL final DNA volume). Finally, the DNA was eluted in 40 µL of 1X T4 DNA ligase buffer (Invitrogen 5X).

Illumina adapter ligation and paired-end PCR: For this step we used the TruSeq DNA LT kit Set A (REF#15041757). 50 µL of ligation mix [25 µL Illumina paired-end adapters, 15 µL T4 DNA ligase Invitrogen, 10 µL 5X T4 DNA ligase buffer (Invitrogen 5X)] was added to the 40 µL sample from the previous step. The ligation sample was then incubated at RT for 2 hours on a rotator. Next, the DNA was purified by adding 1:1 Ampure solution (180 µL of Ampure mix was added to the 90 µL sample), the supernatant was discarded and beads were washed twice with 1 mL of 70% ethanol. After the last wash step, the beads were resuspended in 400 µL of 1X TLE and incubated at RT on a rocking platform for 10 mins. The tube was placed on the MPS for 4 mins. Finally, the 400 µL supernatant was transferred to a new tube.

Biotin pull-down: All the following steps are done using 1.5 mL loBind tube (Eppendorf 22431021). 15 µL of MyOne streptavidin C1 beads mix (Thermo Fisher 65001) was transferred to a 1.5 mL tube. The beads were washed twice with 400 µL of TWB [5 mM Tris-HCl pH8.0, 0.5 mM EDTA, 1 M NaCl, 0.05% Tween20] by incubation for 3 min at RT. After each wash, the tube was placed on the MPS for 1 min and the supernatant was removed. After the washes, the beads were resuspended in 400 µL of 2X Binding Buffer (BB) [10 mM Tris-HCl

pH8.0, 1 mM EDTA, 2 M NaCl] and mixed with the 400 µL DNA from the previous step in a new 1.5 mL. The mixture was incubated for 15 min at RT with rotation. The tube was then placed on the MPS for 1 min and the supernatant was removed. The DNA bound to the beads was washed first by adding 400 µL of 1X BB and transferring to a new tube. The beads were reclaimed against the MPS for 1 min, and the supernatant discarded. 100 µL of 1X TLE was added and the beads were reclaimed against the MPS for 1 min, then the supernatant was discarded. Finally, the DNA bound to the beads was eluted in 32.5 µL of 1X TLE.

PCR optimization: The Illumina Truseq Kit (DNA LT kit Set A (REF#15041757)) was used for PCR amplification of DNA for DpnII-Seq. The trial PCR reaction was set up as follows: [2.5 µL DNA bound to beads, 2 µL of Primers mix (Truseq kit), 10 µL Master Mix (Truseq kit), 10.5 µL of ultrapure distilled water (Invitrogen)]. The 25 µL was split over four PCR tubes (5 µL/per tube). Each of the four samples was incubated for different PCR cycles (6, 8, 10, or 12 cycles): [30 sec at 98°C, n cycles of (30 sec at 98°C, 30 sec at 65°C, 30 sec at 72°C), 7 min at 72°C, hold at 10°C]. The optimal PCR cycle number needed to get enough DNA for sequencing was determined by running the 4 PCR reactions on a 2% agarose gel along with low molecular ladder titration (100 ng, 200 ng, 400 ng). Three PCR reactions of 50 µL volume were then performed: [5 µL DNA bound to beads, 4 µL of Primers mix (Truseq kit), 20 µL Master Mix (Truseq kit), 21 µL of ultrapure distilled water (Invitrogen)]. The 3 PCR reactions were pooled together to obtain 150 µL total volume. The samples were reclaimed against the MPS for

1 min, then the PCR products (supernatant) were taken to new 1.5 mL tubes. 1:1 Ampure was performed for removal of primer dimers (150 µL of Ampure and 150 µL DNA sample). Finally, beads were resuspended in 35 µL of TLE to elute the DNA. DNA that remained bound to beads was saved after a first wash using TBW followed by two washes with 1X TLE and then resuspended in 30 µL of 1X TLE.

DpnII Pre-digestion size assessment

4 million cells were pre-digested for 4 hours using DpnII procedure described above (Pre-digestion of nuclei). The pre-digested nuclei were then treated as follows:

Proteinase K: 50 µL of 10 mg/mL proteinase K (ThermoFisher # 25530) was added to each 500 µL pre-digested nuclei sample (2 million nuclei) (See Methods: Pre-digestion) and the 2 tubes were incubated at 65°C for 3 hours.

DNA purification: Tubes were cooled to room temperature and all 2 samples were pooled in a single 15 mL tube (1.1 mL total volume). The DNA was extracted by adding an equal volume of 1.1 of saturated phenol pH8.0: chloroform (1:1) (Fisher BP1750I-400), followed by vortexing for 1 min. The sample (2.2 mL) was transferred to a 15 mL phase-lock tube (Quiagen #129065) followed by centriguation at 5,000 g for 10 min. The upper phase was transferred to a 15 mL tube to start the second extraction. An equal volume of 1.1 mL saturated phenol pH8.0: chloroform (1:1) was added, followed by vortexing for 1

min. Then the mix was transferred to a 15 mL phase-lock tube (Quiagen #129065) followed by centrifugation at 5,000 g for 10 min. The upper phase of ~ 1.1 mL was transferred to a 15 mL tube (high speed), 1/10 volume (110 µL) 3M sodium acetate pH 5.2 was added followed by brief vortexing and then 2.5 volumes of ice-cold 100% ethanol (2.75 mL) were added. The tube was inverted slowly several times, incubated at -80°C for 1 hr and then DNA was pelleted by centrifugation at 16,000 g for 30 min at 4°C. The supernatant was discarded, and the pellet was dissolved in 500 µL 1X NEB3.1 and transferred to a 0.5 mL AMICON Ultra Centrifuge filter (UFC5030BK EMD Millipore). The column was centrifuged for 5 min at 14,000 g and the flow-through was discarded. The column was washed 4 times using 450 µL of 1X NEB3.1 for desalting of DNA. After the final wash, the DNA remaining in the column (~50 µL) was eluted in 70 µL of 1XNEB3.1; the column was flipped upside down into a new tube to collect DNA and centrifuged for 3 min at 4,000 g. ~70 µL of DNA was recovered. RNA was degraded by adding 1 µL of 10 mg/mL RNAase A and incubation for 30 min at 37°C. The amount of DNA was estimated by running an aliquot on a 1% Agarose gel along with a 1kb ladder (NEB#N3232s).

Blunting overhang: DNA ends were filled in by adding 60 µL of fill-in master mix was added: [1xNEB2.1, 0.25 mM dCTP, 0.25 mM dGTP, 0.25 mM dTTP, 0.25 mM dATP (ThermoFisher#19524016), 50 U Klenow polymerase Polymerase I (NEB M0210L)]. Samples were incubated at 37°C in a Thermocycler for 75 mins. Next, the tubes were placed on ice immediately for 15 mins.

Gel Extraction for size-selection: a 1% gel was prepared. 25µL of loading dye (blue dark) was added to the 130µL DNA sample from the previous step. 25µL of 1kb DNA ladder was loaded in one well and all sample in the remaining wells. After elecrophoresis the DNA was isolated from the gel in three size intervals: less than 1kb, 1kb-3kb, and larger than 3kb.

DNA purification from agarose gel: The DNA was extracted from the agarose gel using GFX™ PCR DNA and Gel Band Purification Kit (GE28-9034-70 Millipore Sigma). After DNA extraction the sample was washed five times with TLE buffer using a 1.5 Amicon column followed by elution with 50 µl TLE. Aliquots of each DNA samples was analyzed on a fragment analyzer.

Sonication: DNA from each size fraction was sonicated to a size of 100 – 300 bp using a Covaris instrument (Duty Cycle 10%, Intensity 5, Cycles per Burst 200, set Mode Frequency sweeping, continuous degassing, process time 60 sec, Number of cycles) for 4 cycles. The 130 µL of sonicated DNA was transferred to a 1.5 mL tube and 1XTLE was added to a total volume of 500 µL. DNA fragment size was determined by running 2 µL of DNA along with low molecular ladder (NEB) on a 2% agarose gel.

Size fractionation using AMpure XP: 500 µL AMpure beads (Beckman Coulter A63881) were added to a 1.5 mL tube labeled as 1.1X. The tube was placed on the MPS for 5 min, and the supernatant was removed. Beads were resuspended in 150 µL AMpure mixture in order to make 1.1X. 400 µL of AMpure mixture was

added to 500 µL of sonicated DNA from the previous step and the tube was labeled 0.8X. The sample was vortexed and centrifuged briefly using a tabletop small centrifuge followed by incubation at RT for 10 min on a rocking platform. Then the tube was placed on the MPS for 5 min at RT. The 0.8X supernatants were collected and added to the 1.1X tube, the tube was briefly vortexed and centrifuged followed by incubation at RT for 10 min on a rocking platform. The tube was placed on the MPS for 5 min at RT, and the supernatant discarded. Beads in the 0.8X and 1.1X tubes were washed twice with 1 mL 70% ethanol, reclaiming beads against the MPS for 5 min. Beads on the MPS were then dried until ethanol had evaporated completely. Next, 51 µL of 1XTLE was added to the 0.8X and 1.1X tubes to elute DNA from the beads. Tubes were incubated at RT on a rocking platform for 10 min. The 0.8X and 1.1X tubes were placed on the MPS for 5 min. Finally, the supernatants were transferred to 1.7 mL tubes labeled 0.8X and 1.1X. The 1.1X sample contains DNA that ranges in size from 100-300 bp. The DNA in the 0.8X sample was kept in case more DNA was required, in which case the DNA would be sonicated using 2 cycles followed by a similar round of size fractionation as described above. The amount of DNA from both samples 0.8X and 1.1X was quantified by running 1 µL on a 2% agarose gel along with a titration of low molecular weight DNA ladder (100 ng, 200 ng, 400 ng).

End Repair: 50 µL from the 1.1X sample was transferred to a PCR tube, and 20 µL of end repair mix was added: [3.5X NEB ligation buffer (NEB B0202S), 0.875

mM dNTP mix, 0.375 U/µL T4 DNA polymerase, 1.25 U/µL T4 polynucleotide kinase, 0.125 U/µL Klenow DNA polymerase]. The 70 µl total volume reaction was incubated for 30 min at 20°C in a PCR machine and then placed on ice. The DNA was purified by 1:2 Ampure, by adding 140 µL 2X Ampure solution to the 70 µL DNA sample followed by incubation for 5 min at RT. The tube was placed on the MPS for 4 min to reclaim the beads and the supernatant was discarded. The beads were washed twice with 1 mL of 70% ethanol while on the MPS. After beads were dried DNA was eluted in 32 µL TLE (pH 8.0) and incubation for 10 min at RT. The supernatant was transferred to a 1.5 mL tube.

A-tailing: A dATP was added to the 3’ ends by adding 18 µL of A-tailing mix [5 µL NEB buffer 3.1, 10 µL of 1 mM dATP, 3 U Klenow exo (NEB M0212S)] to the 32 µL of DNA sample from the previous step. The reaction was then incubated in a PCR machine at 37°C for 30 min followed by incubation 65°C for 20 min and cooling down to 4°C. The tube was placed on ice. The volume was brought to 100 µL by adding 1X NEB2.1. The DNA was then purified by adding 1:2 Ampure mix (200 µL of Ampure was added to the 100 µL final DNA volume). Finally, the DNA was eluted in 40 µL of 1X T4 DNA ligase buffer (Invitrogen 5X).

Illumina adapter ligation and paired-end PCR: For this step we used the TruSeq DNA LT kit Set A (REF#15041757). 50 µL of ligation mix [25 µL Illumina paired-end adapters, 15 µL T4 DNA ligase Invitrogen, 10 µL 5X T4 DNA ligase buffer (Invitrogen 5X)] was added to the 40 µL sample from the previous step. The ligation sample was then incubated at RT for 2 hours on a rotator. Next, the DNA

was purified by adding 1:1 Ampure solution (180 µL of Ampure mix was added to the 90 µL sample), the supernatant was discarded and beads were washed twice with 1 mL of 70% ethanol. After the last wash step, the beads were resuspended in 400 µL of 1X TLE and incubated at RT on a rocking platform for 10 mins. The tube was placed on the MPS for 4 mins. Finally, the 400 µL supernatant was transferred to a new tube.

PCR optimization: The Illumina Truseq Kit (DNA LT kit Set A (REF#15041757)) was used for PCR amplification of DNA for DpnII-Seq. The trial PCR reaction was set up as follows: [2.5 µL DNA bound to beads, 2 µL of Primers mix (Truseq kit), 10 µL Master Mix (Truseq kit), 10.5 µL of ultrapure distilled water (Invitrogen)]. The 25 µL was split over four PCR tubes (5 µL/per tube). Each of the four samples was incubated for different PCR cycles (6, 8, 10, or 12 cycles): [30 sec at 98°C, n cycles of (30 sec at 98°C, 30 sec at 65°C, 30 sec at 72°C), 7 min at 72°C, hold at 10°C]. The optimal PCR cycle number needed to get enough DNA for sequencing was determined by running the 4 PCR reactions on a 2% agarose gel along with low molecular ladder titration (100 ng, 200 ng, 400 ng). Three PCR reactions of 50 µL volume were then performed: [5 µL DNA bound to beads, 4 µL of Primers mix (Truseq kit), 20 µL Master Mix (Truseq kit), 21 µL of ultrapure distilled water (Invitrogen)]. The 3 PCR reactions were pooled together to obtain 150 µL total volume. The samples were reclaimed against the MPS for 1 min, then the PCR products (supernatant) were taken to new 1.5 mL tubes. 1:1 Ampure was performed for removal of primer dimers (150 µL of Ampure and 150

µL DNA sample). Finally, beads were resuspended in 35 µL of TLE to elute the DNA. DNA that remained bound to beads was saved after a first wash using TBW followed by two washes with 1X TLE and then resuspended in 30 µL of 1X TLE.

Lamin A Immunofluorescence and DAPI

For nuclei immunofluorescence, we prepared a coverslip by adding 1 mL of 0.1% Poly-L-lysine solution (Sigma SLBQ5716V) for 10 min, then coverslips were dried using Whatman papers. Each coverslip was transferred to a single well of an eight wells plate. The coverslips were washed twice using PBS. Next 500 µL of 30% sucrose with 1 mM DTT was added on top of the coverslips to protect nuclei from an abrupt contact with coverslip during spinning. 1 million control nuclei or nuclei after chromatin digestion were crosslinked for 20 min using a 4% final concentration of Paraformaldehyde immediately after pre-digestion. Next, nuclei were added slowly on top of he sucrose solutions on the coverslips and spun for 15 mins at 2,500 g at 4°C. Next, nuclei were assumed to be attached to the coverslips which were then transferred to a new 8 well plate. The coverslips were washed five times with 1% PBS. Next, non-specific binding of the primary antibody was blocked by adding 500 µL of the blocking buffer [3% BSA, 1X PBS, 0.1% Triton X-100 (Sigma 9002-93-1)] and incubating for 60 min at RT. Afterward, lamin A antibody (ab 26300) was diluted 1:1000 in blocking buffer, and the coverslip was incubated face-down on top of a 250 µL of lamin A antibody droplet that was placed on parafilm for 120 min at RT. Then, the

coverslip was placed back in the well of a new plate face-up and washed five times with washing buffer (1X PBS, 0.1% Triton X-100). The secondary antibody Goat Anti-Rabbit (ab150077) was diluted 1:1000 in blocking buffer, and the coverslip was incubated face-down on top of a 250 µL droplet of the secondary antibody (Goat Abti-Rabbit (ab150077) that was placed on parafilm for 60 min at RT. Next, the coverslip was placed back in the well of a new plate face-up and washed five times with washing buffer (1X PBS, 0.1% Triton X-100) and twice with 1X PBS. The slide was mounted and sealed using 10 µL antifade mountant with DAPI (Invitrogen P36931).

For image acquisition, we used a Nikon Eclipse Ti microscope. Imaging was performed using an Apo TIRF, N.A. 1.49, 60X oil immersion objective (Nikon), and a Zyla sCMOS camera (Andor). Z-series of 0.2 µm slices were acquired using Nikon Elements software (Version 4.4).

Chromatin fractionation assay

Chromatin-bound proteins were isolated and separated from free proteins. A sample of 2 million control nuclei or pre-digested nuclei (obtained as described above “Pre-digestion of nuclei”) was centrifuged at 5,000 g for 5 min at 4°C. The supernatant was transferred to an Amicon column to reduce the volume from 500 µL to 100 µL by centrifugation for 4 min at 14,000 g. This sample contains the free protein fraction. Next, 26 µL of glycerol and 1.3 µL of 100X protease inhibitor cocktail were added to the 100 µL free proteins sample. The pellet containing the nuclei was resuspended in 100 µL of nuclei purification buffer with Triton (10 mM

PIPES pH 7.4, 10 mM KCl, 2 mM MgCl2, 0.25% Triton, 1% Protease inhibitor, 1mM DTT) and incubated for 10 min on ice. Then, in order to protect protein structure during sonication, 25 µL of glycerol was added to the 100 µL pellet sample to have 20% final glycerol concentration. The sample was sonicated using a Covaris instrument at 4°C as follows: (Duty Cycle 10%, Intensity 5, Cycles per Burst 200, set Mode Frequency sweeping, continuous degassing, process time 60 sec, 4 cycles). The pellet sample contains chromatin-bound proteins, was transferred to a 1.5 mL tube. All samples were stored at -20. These samples contain the protein bound CTCF and cohesin. Note: when these samples were centrifuged after the triton solubilization, we found that no SMC3 or CTCF could be detected in the supernatant. These results indicate that non-chromatin-bound proteins exit the nuclei and were recovered in the supernatant prior to triton solubilization step.

For analysis of CTCF and SMC1 chromatin binding: 15 µL from each protein sample (supernatant or pellet) was mixed with 5 µL of 5X Lane Marker Reducing Sample Buffer (Thermo Fisher 39000), then the mix was boiled for 10 min. The samples were cooled down to RT before loading them on a 3-8% Tris-Acetate Protein Gels (Invitrogen EA0375PK2). Next, the gel was run in 1X Tris-Acetate SDS Running Buffer (Invitrogen LA0041) for 75 min at 150V. For Histone H3: 1 µL of protein sample was mixed with 14 µL of PBS containing 1% Protease inhibitor, 5 µL of 5X Lane Marker Reducing Sample Buffer was added to the mix and boiled for 10 min. The samples were cooled down to RT before loading them

in Tris-Base 4-12% (Invitrogen NP0322BOX), then the gel was run in 1X MES-SDS running buffer (Invitrogen B0002) for 60 min at 150V. The proteins were transferred from the gel to nitrocellulose membrane using 1X western blot transfer buffer (Thermo science 35040). The transfer was 120 min for SMC1 and CTCF and 75 min for H3. The nitrocellulose membranes were washed using 1X TBST [50 mM Tris-Cl, pH 7.6; 150 mM NaCl, 0.1 mL of Tween 20], then Blocked for 120 min using 5% milk (1 g milk in 20 mL 1X TBST). The membrane when then incubated overnight at 4°C with primary antibody diluted in 5% milk [1:1000 CTCF antibody cell signaling (activeMotif 61311), 1:2000 SMC1 (Bethyl Antibody, A300-055A), 1:4000 H3 Abcam (ab1791)] . Next, the membranes were washed 6 times for 10 min per wash using 1X TBST. The secondary antibody anti-rabbit IgG HRP from cell signaling was diluted using 5% milk for CTCF and SMC1 [1:1000 for CTCF, 1:2000 SMC1] and in 1% milk for H3 1:5000 dilution. Membranes were incubated for 120 min at RT. Finally, membranes were washed 6 times for 10 min using 1X TBST. Finally, the membranes were developed using luminol-based enhanced chemiluminescence(Thermo science 34076).

Micromanipulation force measurement and treatments of an isolated nuclei

Micromanipulation force measurements were conducted as described previously in Stephens et al. (Stephens *et al.*, 2017). K562 cells were grown in microscope slide wells and treated with 1 µg/mL latrunculin A (Enzo Life Sciences) for ~45 min before single nucleus isolation. The nucleus was isolated by using small amounts of detergent (0.05% Triton X-100 in PBS) locally sprayed

onto a living cell via an “isolation” micropipette. This gentle lysis allows the use of a second micropipette to retrieve the nucleus from the cell, using slight aspiration and non-specific adherence to the inside of the micropipette. A third micropipette was then attached to the opposite end of the nucleus in a similar fashion. This last “force” micropipette was pre-calibrated for its deflection spring constant, which is on the order of 2 nN/μm. A custom computer program written in LabView was then run to move the “pull” micropipette and track the position of both the “pull” and “force” pipettes. The “pull” pipette was instructed to move 5 μm at 45 nm/sec. The program then tracked the distance between the pipettes to provide a measure of nucleus extension ~3 μm. Tracking the distance that the “force” pipette moved/deflected multiplied by the pre-measured spring constant provides a calculation of force exerted. Calculations were done in Excel (Microsoft) to produce a force-extension plot from which the best-fit slope of the line would provide a spring constant of the nucleus (nN/μm). Isolated nuclei were measured twice initially to establish the native spring constant prior to treatment. After 50 uL of buffer only (control), 100 units DpnII (|GATC) with NEB buffer 3.1, or 100 units HindIII (A|AGCTT) with NEB buffer 2.1 was added to the 1.5 mL imaging well and mixed gently. Force measurements were performed 5 min, 30 min, and 60 min post-treatment.

3C-PCR

The human β-globin locus is an ideal region to examine looping interactions between enhancers and genes because of the strong looping

interactions between the LCR and HBG globin gene in the erythroleukemia cell line K562, which highly expresses the globin genes (Dostie *et al.*, 2006). 3C libraries were generated from: (1) K562 cells that have an LCR-HBG interaction, (2) GM12878 cells in which the LCR-HBG looping interaction is absent, and (3) beta-globin BAC (ENm009) control to normalize for primer bias. To investigate the interaction between the LCR and HBG gene, 3C primers from (Dostie *et al.*, 2006) were used. 16 forward primers of 28-33 bp length were designed 40-60 bp upstream of each EcoRI site throughout a 110 kb region around the Beta Globin locus (chr11: 5221788- 5337325). The EcoRI fragment overlapping with the LCR (HS3,4,5) was used as an anchor to detect the interaction frequencies between the LCR and EcoRI fragments throughout the β-globin locus. For each primer pair, triplicate PCR reactions were set up, and the mean of the three was normalized to the BAC signal for the same primer pair before plotting normalized interaction frequency in the y-axis, the distance from EcoRI fragment overlapping with LCR to neighboring EcoRI fragments is plotted in the x-axis. Error bars are the standard error of the mean (SEM).

5C data processing

The fastq files for 5C sequencing data were processed as described in https://github.com/dekkerlab/5C-CBFb-SMMHC-Inhib/blob/master/data_processing_steps.md

The Fastq files were mapped using novoalign to a reference genome built from the pool of all 277 probes. After mapping, we combined the read-pairs. The

results were then transferred to a matrix format, and interactions were filtered as previously described (Lajoie *et al.*, 2009; Sanyal *et al.*, 2012). First, interactions that belong to the same EcoRI fragment were removed. Second, outliers that are overrepresented as a result of overamplification were also removed. Outliers were defined as the interactions with a Z-score greater than 20 in all datasets. Third, probes that strongly over or underperform which leads to strongly enriched or depleted interactions in a whole row of interactions, were also removed. The four matrices were then scaled to the same number of total reads. Finally, data were binned at 20 Kb (median) with a sliding window with 2.5 Kb steps

Hi-C data processing

Hi-C read mapping, filtering, binning and matrix normalization were performed using the cMapping pipeline available at https://github.com/dekkerlab/cMapping (Lajoie, Dekker and Kaplan, 2015). In brief, Hi-C reads were mapped to reference human genome assembly hg19 using an iterative mapping strategy and Bowtie 2 (Langmead *et al.*, 2009). Successfully mapped reads were then filtered to remove reads mapping to the same restriction fragment and to remove PCR duplicates. Interaction frequency versus distance plots displayed high variance for interactions below 1 kb for all samples. Hence, after mapping of valid pair, we removed all pairs with a genomic distance less than 1 kb. The remaining valid read pairs were then binned to 500 kb, 40 kb, and 10 kb resolution matrices. Outlier bins of these matrices with low signal were assigned values of NA. Then as a bias correction step, matrices

were normalized such that the sum of interactions in each row/column are approximately equivalent via an iterative correction procedure (ICE) (Imakaev *et al.*, 2012). Lastly, for comparison between samples, matrices were scaled such that the total interactions for a genome-wide matrix equals one billion for each sample. These ICEd scaled matrices were used for subsequent analyses.

A/B compartments

All reads from Hi-C in control K562 samples were pooled to identify A (active) and B (inactive) compartments in K562 cells. A/B compartments were identified at 40 kb resolution following the procedure described in (Lieberman-Aiden *et al.*, 2009) using matrix2compartment.pl in https://github.com/dekkerlab/cworld-dekker. Briefly, each cis interaction matrix was first transformed into a z-score matrix followed by transformation into a correlation matrix. PCA was performed on the correlation matrix and the first eigenvector (PC1) of the PCA analysis was used to identify compartments for each chromosome. A/B compartments were assigned based on gene density such that the A-compartment was more gene-dense than the B-compartment. Positive PC1 values indicate gene-rich A compartments and negative PC1 values indicate gene-poor B compartments. For chromosome 9 the compartments were called for each chromosome arm separately as PC1 captured preferences for interactions within the same arm as opposed to canonical compartment preferences.

LOS and half-life calculation

To measure the 3D structure changes resulting from DpnII, HindIII, or FatI pre-digestion we quantified the amount of cis interactions lost or gained in a 6 Mb window centered at every 40 kb bin genome wide. We note that we did not observe detectable amounts of DNA in the supernatant after chromatin fractionation indicating the large majority of liquefied DNA remains within the nuclei. Even if some DNA is lost, ICE balancing of Hi-C matrices ensures any biases in sequence coverage are removed. For each 40 kb bin, the percent of interactions occurring within its 6 Mb window (corresponding to interactions less than or equal to 3 Mb in distance either upstream or downstream from 40 kb bin) out of total interactions for the 40 kb bin (cis and trans) was calculated. These 6 Mb cis percentages were calculated for control, DpnII pre-digested, HindIII-pre-digested nuclei, and FatI pre-digested nuclei. The change in 3D structure relative to control using these cis percentages was given by the following loss of structure (LOS) metric:

$$LOS = \frac{Control_{Cis\%} - Predigest_{Cis\%}}{Control_{Cis\%}}$$

Hence, LOS values in the range (0, 1) represent a loss in short range contacts after pre-digestion; LOS values < 0 represent an increase in short range contacts after pre-digest, and an LOS equal to zero would indicate no change in structure after pre-digestion. A window of 6 Mb was chosen as we sought here to quantify interactions disrupted by pre-digestion. Many longer range interactions

increased after pre-digestion, potentially due to random ligations of cut fragments that start to mix. Difference noted in A and B stability was preserved when LOS was calculated using cis percentages for entire chromosomes as opposed to a 6 Mb window, however the size of chromosomes did bias results by giving 40 kb bins in small chromosomes greater LOS. We note that any loci that may have been lost from the nuclei will not be included in the Hi-C dataset. LOS represents the relative redistribution of short-range interactions to longer-range and inter-chromosomal interactions for the set of loci that remained contained within the nucleus after pre-digestion and we assume this re-distribution would not be affected by any lost loci. FatI-pre digested libraries were of lower sequencing coverage and hence had a lower signal to noise ratio compared with DpnII and HindIII-pre digested libraries. To reduce noise, we applied a loess based smoothing with an α smoothing parameter of 0.01 to the signal track of LOS for nuclei pre-digested using FatI. Correlations between FatI LOS and PC1 were evident both before and after smoothing. Correlations also remained evident before and after smoothing between FatI LOS residuals and PC1 residuals corrected for digestion efficiency by FatI-seq.

To quantify the timing of disrupted interactions we generated a half-life track utilizing the Hi-C matrices from the DpnII timecourses. For each 40 kb bin we fit a curve to the LOS of each timepoint following an exponential decay of the form (Figure 2.10, C):

$$LOS = a - (b \times e^{-c \times minutes})$$

such that a,b and c are parameters to fit.The half-life, $t_{1/2}$, was defined as the time required to reach half saturation, saturation being the 16 hour timepoint where maximal cis interactions have been lost. Half-life values were then computed for every 40 kb bin genome wide. To remove noisy and less reliable $t_{1/2}$ data, we first removed all extreme outliers bins where the sum of squared residuals (SSR) for the exponential fit was greater than 0.1. Then all bins with an SSR greater than two standard deviations from the mean were deemed as outliers and also removed from analyses.

As LOS and $t_{1/2}$ are both dependent on digestion efficiency we also generated residual LOS and $t_{1/2}$ tracks to account for bin to bin variation in digestion efficiency. We used a moving average approach to calculate residuals for LOS as a function of DpnII-seq signal and also $t_{1/2}$ as a function of DpnII-seq signal since the relationships between these variables were non-linear (Figure 2.5, F left, Figure 2.9, C). For both stability metrics LOS and $t_{1/2}$, a sliding window of 200 DpnII-seq signal with a step size of one was used to calculate mean LOS or $t_{1/2}$ signal for each DpnII-seq signal increment (Figure 2.5, F left, Figure 2.9, C). Window and step size were selected by manual inspection of moving averages and compromising between over and underfitting. These moving averages were used to calculate residuals such that a positive LOS residual indicates more structure loss than expected by given digestion efficiency and a negative LOS residual indicates less structure loss than expected. As $t_{1/2}$ is inversely related to LOS, positive $t_{1/2}$ residuals indicate less structure loss than

expected and negative $t_{1/2}$ residuals indicate more structure loss than expected. Moving averages were also used to generate residuals for DpnII-seq as a function of PC1 and LOS as a function of PC1 (Figure 2.5, G right).

Similar to the digestion efficiency correction by DpnII-seq, we also used estimated average fragment size as an independent measure to correct LOS and $t_{1/2}$ for biases in digestion efficiency (Figure 2.7). For both stability metrics LOS and $t_{1/2}$, a sliding window of 200 bp with a step size of one was used to calculate mean LOS or $t_{1/2}$ signal for each average fragment size bp increment (Figure 2.7, G). These moving averages were then used to calculate LOS residuals or $t_{1/2}$ residuals as in the DpnII-seq correction approach described previously.

DpnII-seq data analysis

Sequenced reads were mapped to the hg19 genome using the Bowtie read aligner (Langmead *et al.*, 2009) and reads mapping to multiple sites of the genome were removed. As expected, a high percentage of reads mapped precisely to their associated restriction cut site (Figure 2.6). To remove potential artificial biases, we filtered out paired-end reads from fragments whose start or end coordinate was more than three nucleotides from an appropriate restriction cut site. Filtered reads were then binned to 500 kb or 40 kb resolutions. The K562 cell line has a primarily triploid karyotype with regions of the genome in diploid and tetraploid states. Copy number state assignments for each 500 kb or 40 kb bin were assigned using publicly available K562 copy number data from the Catalogue of Somatic Mutations In Cancer (COSMIC) database

(https://cancer.sanger.ac.uk/cell_lines/download). Copy number segments in the COSMIC dataset were identified by PICNIC analysis of Affymetrix SNP6.0 array data (PMID:19837654). Read coverage files at 500 kb and 40 kb were corrected to a genome wide diploid state using the copy number state assignments and dividing coverage by appropriate correction value (diploid = 1, triploid =1.5, tetraploid = 2, etc.) per bin. (Figure 2.6, D and E). Final copy number corrected coverage files were used for all downstream analysis. DpnII-seq computational workflow is maintained at https://github.com/tborrman/DpnII-seq

FatI-seq data analysis

Computational workflow for FatI-seq analysis was identical to previously described DpnII-seq analysis, with the exception that FatI restriction sites were used in the filtering step as opposed to DpnII restriction sites. The DpnII-seq workflow maintained at https://github.com/tborrman/DpnII-seq has options for analyzing restriction enzyme-seq experiments using the following enzymes: DpnII, HindIII, and FatI.

DpnII pre-digestion average fragment size analysis

Pre-digestion by DpnII leads to variable DNA fragment sizes across the genome. To estimate the average fragment size for a genomic bin after a 4-hour DpnII pre-digestion, we first separated 4-hour DpnII pre-digestion DNA into three slices: less than 1kb, 1kb-3kb, and larger than 3kb (See previously described above: DpnII Pre-digestion size assessment, Figure 2.7 A and B). DNA fragments purified from these slices were sequenced and sequenced read pairs

were mapped to the hg19 genome using the Bowtie read aligner (Langmead *et al.*, 2009). Mapped reads were then binned to 40 kb resolution, normalized for sequencing depth, and corrected for copy number state as in the DpnII-seq workflow (See previously described above: DpnII-seq data analysis). This resulted in a coverage track for each of the three DpnII pre-digested slices: less than 1 kb, 1 kb – 3 kb, and larger than 3kb (Figure 2.7, C).

To estimate average fragment size for a given genomic bin we used the following formula:

$$Average\ fragment\ size = \frac{p_1 q_1 s_1 + p_2 q_2 s_2 + p_3 q_3 s_3}{p_1 q_1 + p_2 q_2 + p_3 q_3}$$

such that p_x equals the percent of normalized slice x reads that mapped to bin, q_x equals quantity of slice x fragments (ng/µL), and s_x equals mean size of fragments from slice x (bp). The variable x represents one of the three slice intervals (1: less than 1 kb, 2: 1 kb - 3 kb, and 3: larger than 3 kb). Hence, the average fragment size for a given bin estimates the quantity of fragments from each slice size mapping to the bin over the total quantity of fragments mapped to the bin. The values for p_x are extracted from our coverage tracks and vary bin to bin, while the values for q_x and s_x are extracted from the Fragment Analyzer analysis and are constants (s_1 = 643 bp, s_2 = 2332 bp, s_3 = 5495 bp, q_1 = 1.6562 ng/µL, q_2 = 2.544 ng/µL, q_3 = 2.4632 ng/µL, Figure 2.7, B). The average fragment size track was then used as an independent metric for measuring 4-hour DpnII digestion efficiency as compared to the DpnII-seq signal track.

Subcompartments

Rao et al. (2014) divided the canonical A/B compartments into five primary subcompartments A1, A2, B1, B2, B3 based on each subcompartment's preferential Hi-C interactions in GM12878 cells. Subcompartments were annotated using high resolution (~1 kb) Hi-C data and were shown to display unique genomic and epigenomic profiles. K562 subcompartments were annotated in (Xiong and Ma, 2019) via the method SNIPER using lower resolution Hi-C data. In short, SNIPER infers subcompartments via a neural network approach to accurately annotate subcompartments using Hi-C datasets with moderate coverage (~500 million mapped read pairs). Xiong et al.'s K562 SNIPER subcompartments showed a substantial conservation with GM12878 annotations from Rao et al. (Rao *et al.*, 2014) and were also enriched in similar epigenetic features, hence we utilized these SNIPER annotations to compare subcompartment status with chromatin stability. K562 SNIPER subcompartments were annotated at 100 kb resolution. To compare with our 40kb resolution liquid chromatin Hi-C data, we binned the 100 kb subcompartment annotations to 40kb such that any 40 kb bin overlapping a boundary of two separate subcompartments was assigned a value of NA. Upon piling up K562 subcompartment boundaries, we also found enrichment and depletion of various chromatin features consistent with those described in both Rao et al. (Rao *et al.*, 2014; Xiong and Ma, 2019).

Sub-nuclear structures

To assess the effect of sub-nuclear structures on chromatin stability we utilized the extensive genetic and epigenetic data publicly available for K562 cells (Table 2.1).

Fold change over control ChIP-seq tracks for histone modifications, chromatin remodellers, and other various proteins were downloaded from the ENCODE Portal. To compare ChIP-seq data with $t_{1/2}$, or residuals of $t_{1/2}$ after correction for DpnII-signal, we binned the ChIP-seq signal tracks into 40 kb such that each 40 kb bin represented the mean signal found across the bin. Bins with no overlapping signal were designated a value of NA.

To examine the association between methylation state and $t_{1/2}$ or residuals of $t_{1/2}$ after correction for DpnII-signal, we downloaded methylation state at CpG Whole-Genome Bisulfite Sequencing (WGBS) tracks from ENCODE. As the methylation data was mapped to hg38, we used the UCSC LiftOver program to convert coordinates to hg19. Then percentage methylation at CpG sites was binned to 40kb resolution using the mean.

As there is currently no nucleolus associated domains (NADs) data available for K562, we analyzed a binary NADs state track for the human embryonic fibroblast IMR90 cell line (Dillinger, Straub and Nemeth, 2017). Dillinger et al. annotated NADs via a two-state hidden Markov model of aCGH data from DNA of isolated nucleoli. Using these annotated NADs, coverage of

each 40 kb bin for NADs was assessed and used for all our downstream analyses.

Mapping of nuclear speckle, nuclear lamina and PolII associated loci for K562 cells was accomplished recently via the TSA-seq protocol (Chen *et al.*, 2018). Signal tracks of log2(pull-down/input) were downloaded from GEO and binned to 40 kb as previously described for ChIP-seq files. Microarray data for LaminB1 associated domains identified through the DamID protocol was also available from that study. We used the UCSC LiftOver program to convert coordinates from hg18 to hg19. We then binned the log2(Dam-LaminB1/Dam) signal to 40 kb bins as previously described for ChIP-seq files.

To analyze cell cycle relationship with chromatin stability we downloaded Repli-seq data for K562 cells from ENCODE. Actively replicating regions are quantified as a percentage normalized signal for FACS sorted cells in G1 phase, four stages of S phase (S1-S4) and G2 phase. Signal tracks for Repli-seq data were binned to 40 kb as previously described for ChIP-seq files.

Binning of data was performed using the bedtools/v2.26.0 software. To assess the quality of the publicly downloaded data we generated the spearman correlation matrix of all binned signal tracks (Figure 2.13, A). Hierarchical clustering of rows of the correlation matrix position heterochromatic marks (H3K9me3, HP1α, HP1β, NADs, and LADs) near one another as expected. The majority active marks form a larger cluster, with the markers for polycomb

regions (H3K27me3, CBX8, BMI1, RNF2, and SUZ12) representing facultative heterochromatin clustered together segregating active from inactive marks.

Gene Expression

To assess the effect of gene expression on chromatin stability we utilized processed gene expression quantifications of total RNA-seq for K562 cells available from ENCODE (Accession ID: ENCFF782PCD). Gene locations were mapped using the hg19 ensGene table from UCSC Table Browser. To compare expression values with 40 kb resolution $t_{1/2}$ or residuals of $t_{1/2}$ after correction for DpnII-signal tracks, fragments per kilobase million (FPKM) values for each gene were binned to 40 kb such that each 40 kb bin represented the mean FPKM for all genes overlapping that bin. Bins without any genes were assigned a value of NA. Binned FPKM >=1 was determined to be a reasonable cutoff for expression by inspection of the full distribution of FPKM values.

Compartmentalization saddle plots

Saddle plot calculations were performed using tools in the cooltools repository: (https://github.com/hms-dbmi/hic-data-analysis-bootcamp/tree/master/notebooks/04_analysis_cooltools-eigenvector-saddle.ipynb).

To measure the strength of compartments, intra-chromosomal interaction frequencies were first normalized by the average interaction frequency at a given genomic distance (observed/expected Hi-C maps) at a resolution of 40 kb. Then

the distance corrected interaction frequencies were sorted based on PC1 values of a pair of bins that define a given interaction. Finally, sorted frequencies were aggregated into 50 by 50 groups according to their PC1 values and averaged to obtain a compartmentalization saddle plot. In a compartmentalization saddle plot, preferential B-B interactions are in the upper left corner, and preferential A-A interactions are in the lower right corner.

Homotypic interaction saddle plots

Intra-chromosomal interactions frequencies between 40 kb bins were normalized by the average interaction frequency at a given genomic distance (observed/expected Hi-C maps). Then, the distance corrected interaction frequencies were sorted based on signal values (TSA-seq, DamID) of a pair of bins that define a given interaction, for a given factor (SON, Lamin). Finally, sorted frequencies were aggregated into 50 by 50 groups according to their signal values and averaged, to obtain homotypic interaction saddle plots. In these plots, pair-wise interactions between loci enriched in factor binding are shown in the lower right corner, and pair-wise interactions between loci not bound by the factor are shown in the upper left corner.

Scaling plot

The script to generate scaling plots was adapted from cooltools (https://github.com/mirnylab/cooltools/tree/master/cooltools). Genome-wide curves of normalized contact frequency $P(s)$ is plotted as a function of genomic

distance for all intra-chromosomal interactions. Each library was normalized by total valid interactions

Mean z-score heatmap

Each genome wide 40kb signal vector for a sub-nuclear structure was cleaned for outliers above three standard deviations of the vector's mean. Each cleaned vector was z-score transformed and then partitioned based on the different $t_{1/2}$ residual intervals for associated bins. The mean z-score for all bins within a given $t_{1/2}$ residual interval is plotted as a square in the heatmap.

CHAPTER III: NEURON-SPECIFIC SIGNATURES IN THE CHROMOSOMAL CONNECTOME ASSOCIATED WITH SCHIZOPHRENIA RISK

Preface

This chapter comprises work published in Science by Prashanth Rajarajan, myself, Will Liao, Nadine Schrode, Erin Flaherty, Charlize Casiño, Samuel Powell, Chittampalli Yashaswini, Elizabeth A. LaMarca, Bibi Kassim, Behnam Javidfar, Sergio Espeso-Gil, Aiqun Li, Hyejung Won, Daniel H. Geschwind, Seok-Man Ho, Matthew MacDonald, Gabriel E. Hoffman, Panos Roussos, Bin Zhang, Chang-Gyu Hahn, Zhiping Weng, Kristen J. Brennand, and Schahram Akbarian. The publication reference is “Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk” *Science*. Vol. 362 Issue 6420 Dec. 2018 (Rajarajan *et al.*, 2018)

Cell culture work including Hi-C, 3C, RNA-seq, ATAC-seq, Cas9, and dCas9 (epi)genome editing was performed by Prashanth Rajarajan, Charlize Casiño, Chittampalli Yashaswini, Bibi Kassim, Behnam Javidfar, Samuel Powell, Elizabeth A. LaMarca, Bin Zhang, Seok-Man Ho, and Aiqun Li. Biocomputing, informatics, and genomic analyses was performed by myself, Will Liao, Nadine Schrode, Sergio Espeso-Gil, Erin Flaherty, Gabriel E. Hoffman., and Zhiping Weng. Specifically, I computationally processed Hi-C sequencing datasets, performed loop calling, performed loop analysis, developed differential loop

calling method, performed gene ontology analyses, performed downstream RNA-seq analyses, and performed coregulation analyses. Materials were contributed by Seok-Man Ho, Erin Flaherty, Hyejung Won, and Daniel H. Geschwind. Research was supervised by Zhiping Weng, Panos Roussos, Kristen J. Brennand and Schahram Akbarian. Experiments were conceived and designed by Prashanth Rajarajan, Kristen J. Brennand and Schahram Akbarian. The paper was written and figures were produced by Prashanth Rajarajan, myself, Will Liao, Kristen J. Brennand, and Schahram Akbarian with contributions from all coauthors.

Abstract

To explore the developmental reorganization of the three-dimensional genome of the brain in the context of neuropsychiatric disease, we monitored chromosomal conformations in differentiating neural progenitor cells. Neuronal and glial differentiation was associated with widespread developmental remodeling of the chromosomal contact map and included interactions anchored in common variant sequences that confer heritable risk for schizophrenia. We describe cell type–specific chromosomal connectomes composed of schizophrenia risk variants and their distal targets, which altogether show enrichment for genes that regulate neuronal connectivity and chromatin remodeling, and evidence for coordinated transcriptional regulation and proteomic interaction of the participating genes. Developmentally regulated chromosomal conformation changes at schizophrenia-relevant sequences

disproportionally occurred in neurons, highlighting the existence of cell type–specific disease risk vulnerabilities in spatial genome organization.

Introduction

Spatial genome organization is highly regulated and critically important for normal brain development and function (Rajarajan *et al.*, 2016). Many of the risk variants contributing to the heritability of complex genetic psychiatric disorders are located in noncoding sequences (Ripke *et al.*, 2014) , presumably embedded in “three-dimensional genome” (3DG) structures important for transcriptional regulation, such as chromosomal loop formations that bypass linear genome on a kilobase (or megabase) scale and topologically associated domains (TADs) (Dixon *et al.*, 2012; Nora *et al.*, 2012) that assemble in nested fashion across hundreds of kilobases (Bharadwaj *et al.*, 2013, 2014; Won *et al.*, 2016; Jiang *et al.*, 2017). By linking noncoding schizophrenia associated genetic variants with distal gene targets, 3DG mapping with Hi-C (Lieberman-Aiden *et al.*, 2009; Rao *et al.*, 2014) and other genome-scale approaches could inform how higher-order chromatin organization affects genetic risk for psychiatric disease. To date, only a very limited number of Hi-C datasets exist for the human brain: two generated from bulk tissue of developing forebrain structures (Won *et al.*, 2016) and adult brain (Schmitt *et al.*, 2016) and one from neural stem cells (Dixon *et al.*, 2015). Although such datasets have advanced our understanding of the genetic risk architecture of psychiatric disease (Won *et al.*, 2016; de la Torre-Ubieta *et al.*, 2018), 3DG mapping from postmortem tissue lacks cell type–specific resolution

and may not capture higher-order chromatin structures sensitive to the autolytic process (Mitchell *et al.*, 2014). We monitored developmentally regulated changes in chromosomal conformations during the course of isogenic neuronal and glial differentiation, describing large-scale pruning of chromosomal contacts during the transition from neural progenitor cells (NPCs) to neurons. Furthermore, we uncovered an expanded

3DG risk space for schizophrenia—with a functional network of disease-relevant regulators of neuronal connectivity, synaptic signaling, and chromatin remodeling—and demonstrate neural cell type–specific coordination at the level of the chromosomal connectome, transcriptome, and proteome.

Results

Neural progenitor differentiation is associated with dynamic 3DG remodeling

We applied in situ Hi-C (Rao *et al.*, 2014) to map the 3DG of two male human induced pluripotent stem cell (hiPSC)–derived neural progenitor cells (NPCs) (Topol *et al.*, 2016), together with isogenic populations of induced excitatory neurons (“neuron”) generated through viral overexpression of the transcription factor *NGN2* (Ho *et al.*, 2016) and differentiations of astrocyte-like glial cells (“glia”) (Figure 3.1, A and B, and Table 3.1) (TCW *et al.*, 2017). Transcriptome RNA sequencing (RNA-seq) comparison with published datasets (Hoffman *et al.*, 2017) confirmed that the NPCs, but not glia, from subjects S1 and S2 clustered together with NPCs from independent donors, whereas S1 and

S2 NGN2 neurons closely aligned with directed differentiation forebrain neurons (Brennand *et al.*, 2011) and prenatal brain datasets (Figure 3.2, A and B). As with our transcriptomic datasets, hierarchical clustering of our Hi-C datasets after initial processing (Figure 3.3, A) also showed clear separation by cell type (Figure 3.1 A and Figure 3.3, B). Genome-scale interaction matrices were enriched for intrachromosomal conformations (Figure 3.3, C), with the exception of the negative control ("No Ligase") NPC library, in which we omitted the ligase step (Materials and methods) and observed an interaction map with no signal due to the loss of ligated chimeric fragments (Figure 3.3, D). Given the observed correlation between technical replicates of Hi-C assays from the same donor and cell type, and the correlation between cell type–specific Hi-C from the two donors ($R_{\text{technical replicates}}$, range = 0.970 to 0.979; $R_{\text{subject1-subject 2 by cell type}}$, range = 0.962 to 0.970), we pooled by cell type for subsequent analyses (Figure 3.3, E).

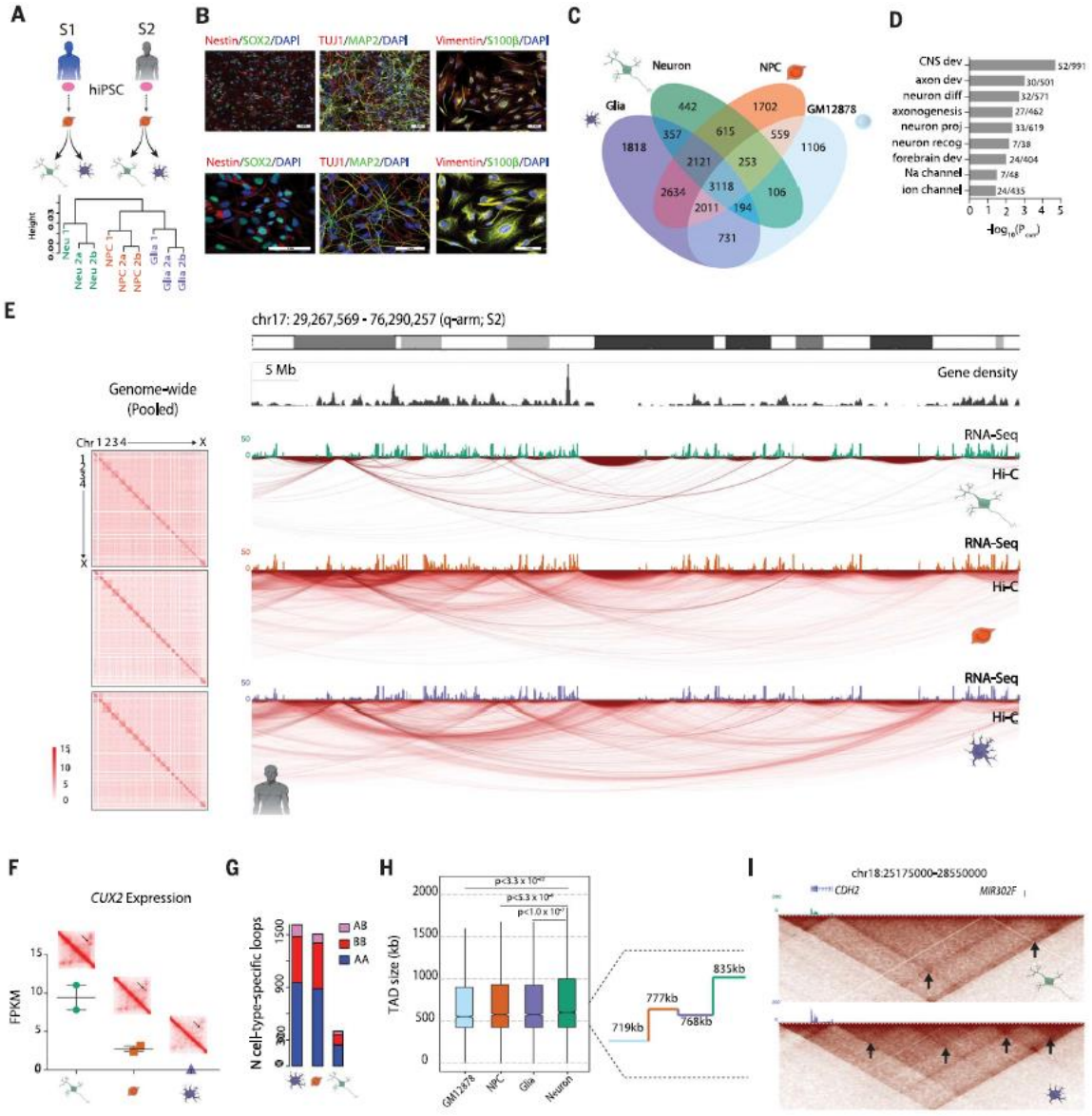
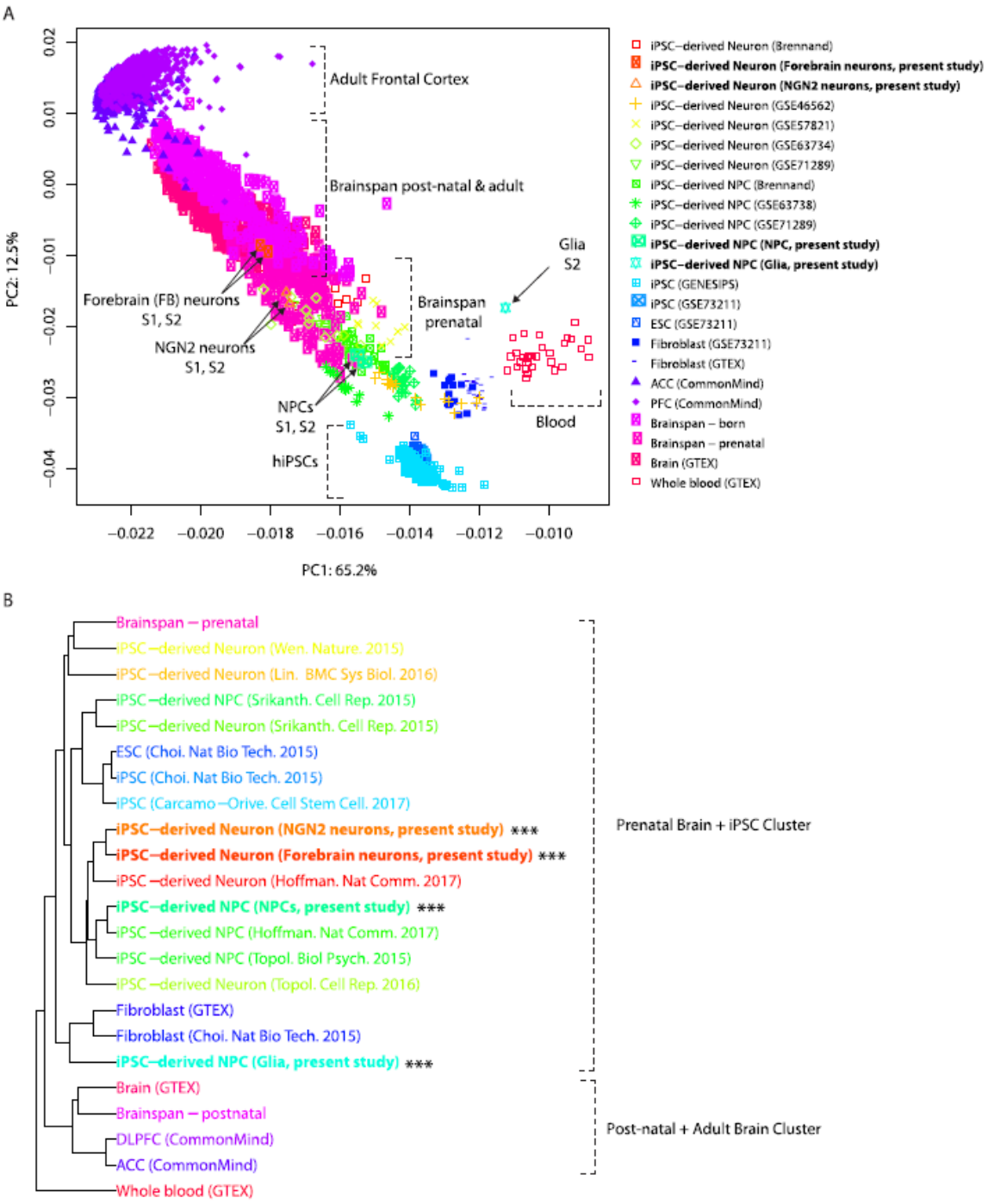


Figure 3.1 | Neural differentiation is associated with large-scale remodeling of the 3D genome.

(**A**) (Top) Derivation scheme of isogenic cell types from two male control cell lines. Pink oval, donor hiPSC; orange, NPC; green, neuron; purple, glia. (Bottom) Hierarchical clustering of intrachromosomal interactions (see Materials and methods) from six in situ Hi-C libraries. a and b are technical replicates of the same library; height corresponds to the distance between libraries (see Materials and methods) (Figure 3.3, B). (**B**) Immunofluorescent staining of characteristic cell markers for NPCs (Nestin and SOX2), neurons (TUJ1 and MAP2), and glia (Vimentin and S100β). (**C**) Venn diagram of loop calls specific to and shared by different subsets of cells, including previously published GM12878

lymphoblastoid Hi-C data. (**D**) Gene ontology (GO) enrichment (significant terms only) of genes overlapping anchors of loops shared by NPCs, neurons, and glia but absent in GM12878. (**E**) (Left) Cell-type pooled whole-genome heatmaps at 500-kb resolution (Figure 3.3, C). (Right) “Arc map” showing intrachromosomal interactions at 40-kb resolution of the q-arm of chr17 for isogenic neurons, NPCs, and glia, as indicated, from subject 2. RNA-seq tracks for each cell type shown on top of arc maps. Green, neuron; orange, NPC; purple, glia. (**F**) FPKM gene expression of *CUX2* across three cell types with heatmap zoomed in on *CUX2* loop (black arrow) (Figure 3.4). (**G**) Number of loops specific to each cell type (not shared with other cell types) with one anchor in an A compartment and another in a B compartment (pink), both in B compartments (red), or both in A compartments (blue). (**H**) (Left) Box and-whisker distribution plot of TAD size across four cell types. (Right) Mean TAD length for each of the four cell types. (**I**) Heatmaps at 40-kb resolution for a 3-Mb window at the *CDH2* locus on chr18. (Bottom) Nested TAD landscape in glia with multiple subTADs (black arrows) called, which (top) is absent from neuronal Hi-C. RNA-seq tracks: green, neuron; purple, glia (Figures 3.2 to 3.6).



3.2 | Cell- and tissue-specific gene expression profiles.
(**A**) Multidimensional scaling with samples colored as indicated in cell- and tissue-specific manner. Note that neuronal cultures and NPC from subjects S1 and S2 cluster together with fetal and perinatal brain tissue, intermingled in related cell types from other donors. (**B**) Pairwise distance matrix was computed

for all samples, and the median distance between all samples in each category were used to create a summary distance matrix in order to perform the final clustering. Note that samples from the present study (marked by ***) align with independently generated datasets for same cell types from different donors.

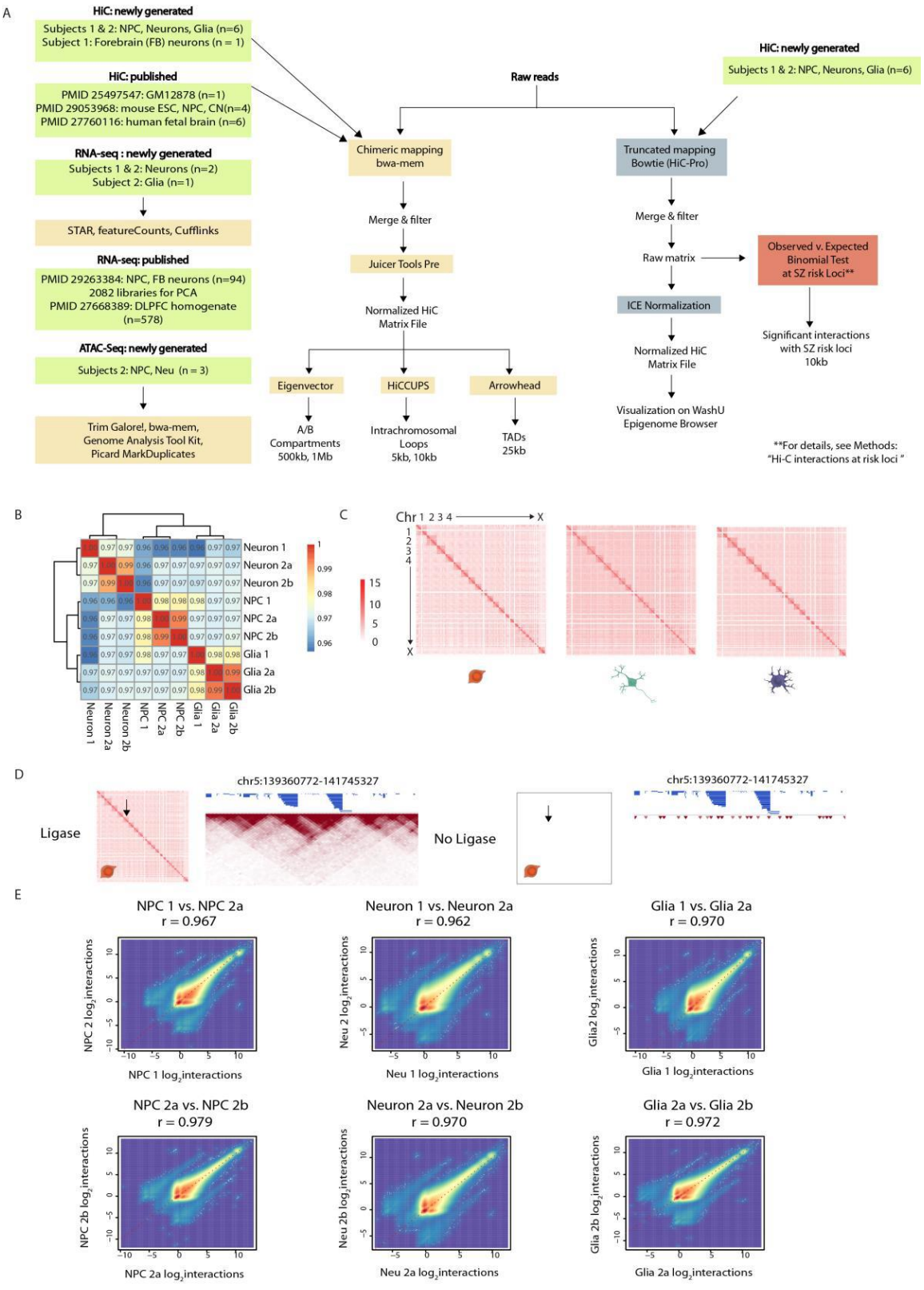

A
HiC: newly generated
Subjects 1 & 2: NPC, Neurons, Glia (n=6)
Subject 1: Forebrain (FB) neurons (n = 1)
HiC: published
PMID 25497547: GM12878 (n=1)
PMID 29053968: mouse ESC, NPC, CN(n=4)
PMID 27760116: human fetal brain (n=6)
RNA-seq : newly generated
Subjects 1 & 2: Neurons (n=2)
Subject 2: Glia (n=1)
STAR, featureCounts, Cufflinks
RNA-seq: published
PMID 29263384: NPC, FB neurons (n=94)
2082 libraries for PCA
PMID 27668389: DLPFC homogenate
(n=578)
ATAC-Seq: newly generated
Subjects 2: NPC, Neu (n = 3)
Trim Galore!, bwa-mem,
Genome Analysis Tool Kit,
Picard MarkDuplicates
Raw reads
Chimeric mapping
bwa-mem
Merge & filter
Juicer Tools Pre
Normalized HiC
Matrix File
Eigenvector
A/B
Compartments
500kb, 1Mb
HiCCUPS
Intrachromosomal
Loops
5kb, 10kb
Arrowhead
TADs
25kb
HiC: newly generated
Subjects 1 & 2: NPC, Neurons, Glia (n=6)
Truncated mapping
Bowtie (HiC-Pro)
Merge & filter
Raw matrix
ICE Normalization
Normalized HiC
Matrix File
Visualization on WashU
Epigenome Browser
Observed v. Expected
Binomial Test
at SZ risk Loci**
Significant interactions
with SZ risk loci
10kb
**For details, see Methods:
"Hi-C interactions at risk loci "
B
Neuron 1
Neuron 2a
Neuron 2b
NPC 1
NPC 2a
NPC 2b
Glia 1
Glia 2a
Glia 2b
C
Chr 1 2 3 4 X
D
Ligase
chr5:139360772-141745327
No Ligase
chr5:139360772-141745327
E
NPC 1 vs. NPC 2a
r = 0.967
NPC 2 log2interactions
NPC 1 log2interactions
Neuron 1 vs. Neuron 2a
r = 0.962
Neu 2 log2interactions
Neu 1 log2interactions
Glia 1 vs. Glia 2a
r = 0.970
Glia2 log2interactions
Glia 1 log2interactions
NPC 2a vs. NPC 2b
r = 0.979
NPC 2b log2interactions
NPC 2a log2interactions
Neuron 2a vs. Neuron 2b
r = 0.970
Neu 2b log2interactions
Neu 2a log2interactions
Glia 2a vs. Glia 2b
r = 0.972
Glia 2b log2interactions
Glia 2a log2interactions

Figure 3.3 | Bioinformatic pipeline overview and basic library characteristics.
(**A**) Bioinformatic approaches involved in processing raw reads to arrive at Hi-C contact matrices, intrachromosomal loops and interactions, TADs, and compartments. Hi-C and other datasets used, both newly generated and previously published, are listed in green boxes. (**B**) Pearson correlation of log-transformed interaction bin counts for the 10% most variable intrachromosomal interactions at 500kb resolution, showing separation of libraries by cell type and not by individual identity. (See hierarchical clustering in Figure 3.1 A.) (**C**) Genome-wide Hi-C interaction matrix for NPC, neuron, and glia from donor S2 at 500kb resolution. (**D**) Genome-wide Hi-C heatmap of pooled NPC library (left) compared to No Ligase negative control library generated from NPC 1 (right), which shows a severely contact depleted map without the crucial ligation step. Black arrow points to chr5. Zoomed in TAD landscape at the clustered PROTOCADHERIN locus in pooled NPC library (left) and No Ligase NPC 1 (right). (**E**) Pearson correlation at 500kb resolution of intra- and inter-chromosomal interaction frequency between replicates, showing high level of correlation between different individuals ("1" v. "2") and technical replicates ("a" v. "b") within each cell type.

We first focused on intrachromosomal loop formations, which are conservatively defined as distinct contacts between two loci in the absence of similar interactions in the surrounding sequences (Rao *et al.*, 2014). Our comparative analyses included published (Rao *et al.*, 2014) in situ Hi-C data from the B lymphocyte–derived cell line GM12878 (Table 3.1). When analyzed with the HiCCUPS pipeline (5- and 10-kb loop resolutions combined, subsampled to 372 million valid-intrachromosomal read pairs to reflect the library with the fewest reads after filtration) (Rao *et al.*, 2014), 17,767 distinct loops were called: n = 3118 (17.5%) were shared among all four cell types, whereas n = 5068 (28.5%) were specific to only one of the four cell types (Figure 3.1, C). Biologically relevant terms such as "central nervous system development," "forebrain development," and "neuron differentiation" were among the top gene ontology

(GO) enrichments from genes overlapping loops shared between NPCs, glia, and neurons (brain-specific) but not identified in lymphocytes (Figure 3.1, D and Table 3.2), indicating strong tissue-specific loop signatures that were also confirmed in individual cell types (Figure 3.4, A and Tables 3.3 to 3.6).

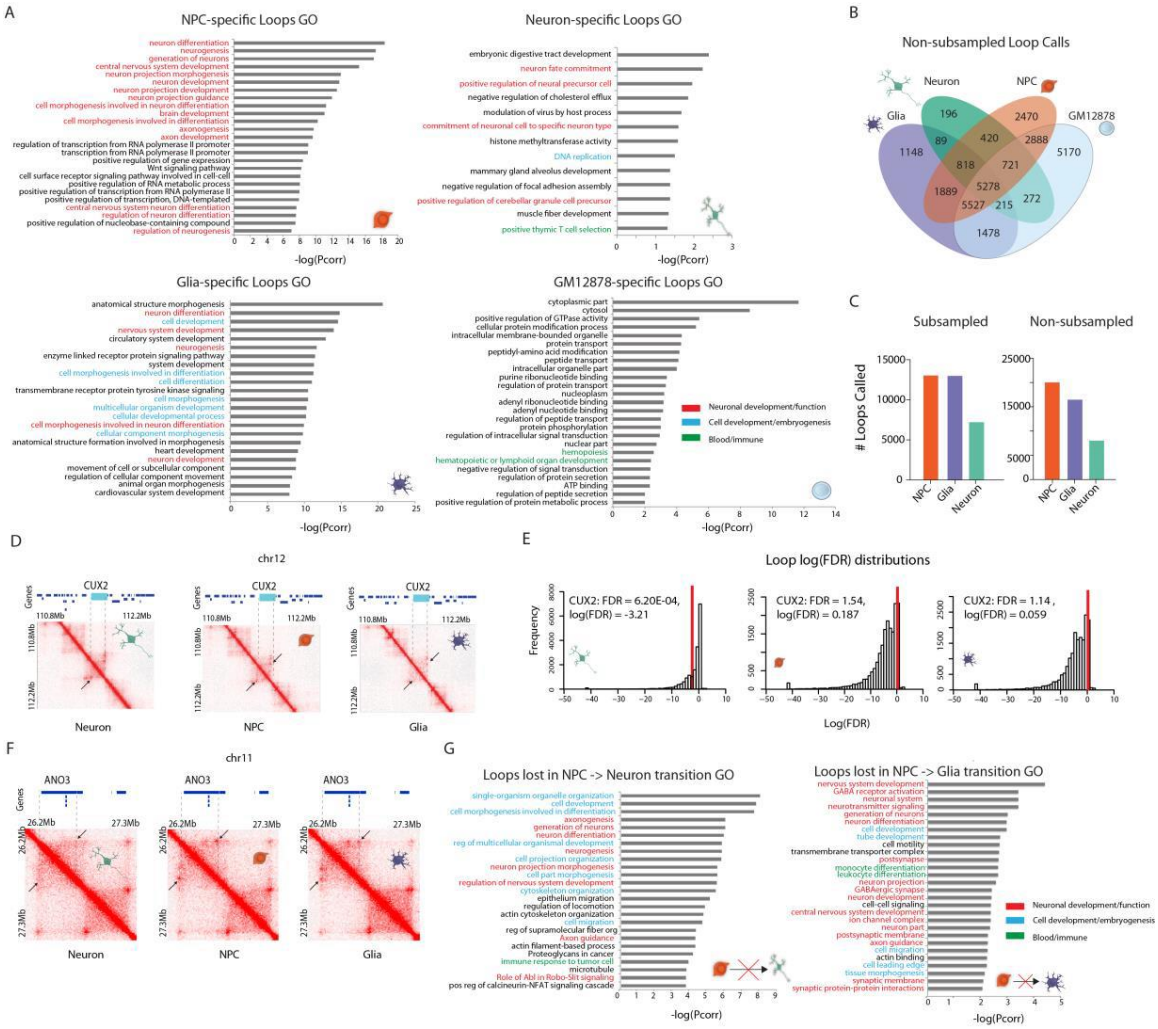


Figure 3.4 | Cell type-specific features of the 3D genome.
(**A**) GO enrichment of genes overlapping cell type-specific loops from non-subsampled datasets, with genes overlapping loops specific to neural cell types involved in neurodevelopment while no such pattern is observed in GM12878. GO terms are color coded by functional categories (see legend). (**B**) Venn diagram of loops called from the non-subsampled datasets for each cell type. (**C**)

Count of loops called in each cell type in the subsampled (to 372M cis contacts) and non-subsampled datasets, showing drastically reduced number of loops called in neurons. (**D**) Heatmaps at 10kb resolution with a neuron-specific loop (arrows; left) at CUX2 gene locus called in neurons but not in NPCs (middle) or glia (right), respectively. (**E**) Frequency distribution of log(FDR) of 17,767 loops (5kb and 10kb resolutions combined) in neurons (left), NPCs (middle), and glia (right). Vertical red line denotes the log(FDR) for the 10kb-resolution loop overlapping CUX2 in each cell type. Note that CUX2 loop is significant in neurons and not in NPCs or glia. (**F**) Representative example of a loop called in NPCs and glia but not in neurons. This loop spans 370kb, overlapping the gene ANO3. (**G**) GO enrichment of genes overlapping anchors of loops that are lost in the transition from NPC to neuron (left) and NPC to glia (right). GO terms are color coded by functional categories (see legend).

Unexpectedly, there was a reduction (~40 to 50% decrease) in the total number of chromosomal loops in neurons relative to isogenic glia and NPCs (Figure 3.4, B and C). Reduced densities of chromosomal conformations were also evident in genome browser visualization of chromosomal arms, including chr17q (Figure 3.1, E). Although both glia and NPCs harbored ~13,000 loop formations, only 7206 were identified in neurons (Figure 3.1, C; Figure 3.4, B and C; and table S1), including 442 neuron-specific loop formations. One such neuron-specific loop was at *CUX2*, a transcription factor whose expression marks a subset of cortical projection neurons (Gil-Sanz *et al.*, 2015) and that is highly expressed in our *NGN2*-induced neurons (Figure 3.1, F and Figure 3.4, D and E). Examples of loops lost in neurons include one spanning the Ca^{2+} channel and dystonia-risk gene, *ANO3* (Figure 3.4, F) (Charlesworth *et al.*, 2012). Furthermore, NPCs, neurons, and glia had similar proportions of loops anchored in solely active (A) compartments, solely inactive (B) compartments, or in both, indicating no preferential loss of either active or inactive loops in neurons (Figure

3.1, G). However, among the genes overlapping anchors of loops that underwent pruning during the course of the NPC-to-neuron transition, regulators of cell proliferation, morphogenesis, and neurogenesis ranked prominently in the top 25 GO terms with significant enrichment (Benjamini-Hochberg corrected P_{range} = 10^{-3} – 10^{-8}) (Figure 3.4, G and Table 3.4, B), which is consistent with a departure from precursor stage toward postmitotic neuronal identity (Wang *et al.*, 2017). Likewise, loops lost during NPC-to-glia transition were significantly enriched (Benjamini-Hochberg corrected P_{range} = 10^{-2} – 10^{-5}) for neuron-specific functions, including “transmission across chemical synapse,” “g-aminobutyric acid (GABA) receptor activation,” and “postsynapse” (Figure 3.4, G and Table 3.4, C), which is consistent with non-neuronal lineage commitment.

We defined “loop genes” as genes that either have gene body or transcription start site (TSS) overlap with a loop anchor (5- or 10-kb bins forming the points of contact in a chromatin loop). Genes with loop-bound gene bodies (one-tailed *Z* test, Z_{range} = 42.1 to 59.2, $P < 10^{-324}$ for all) or loop-bound TSS (one-tail *Z*-test, Z_{range} = 15.2 to 28.8, $P_{range} = 2.32 \times 10^{-52}$ to 4.40×10^{-182}) both showed significantly greater expression [mean $\log_{10}$(FPKM + 1); FPKM, fragments per kilobase of exon per million fragments mapped] than that of background (all genes for all brain cell types) (Figure 3.5, A), suggesting that looping architecture was associated with increased gene expression. Furthermore, 3% of loops shared by NPCs, neurons, and glia (brain-specific loops) interconnected a brain expression quantitative trait locus (eQTL) single

nucleotide polymorphism (SNP) with its destined target gene(s), representing significant enrichment over background as determined with 1000 random distance- and functional annotation–matched loop samplings, (random sampling, one-sided empirical $P = 0.012$) (Materials and methods) (Figure 3.5, B).

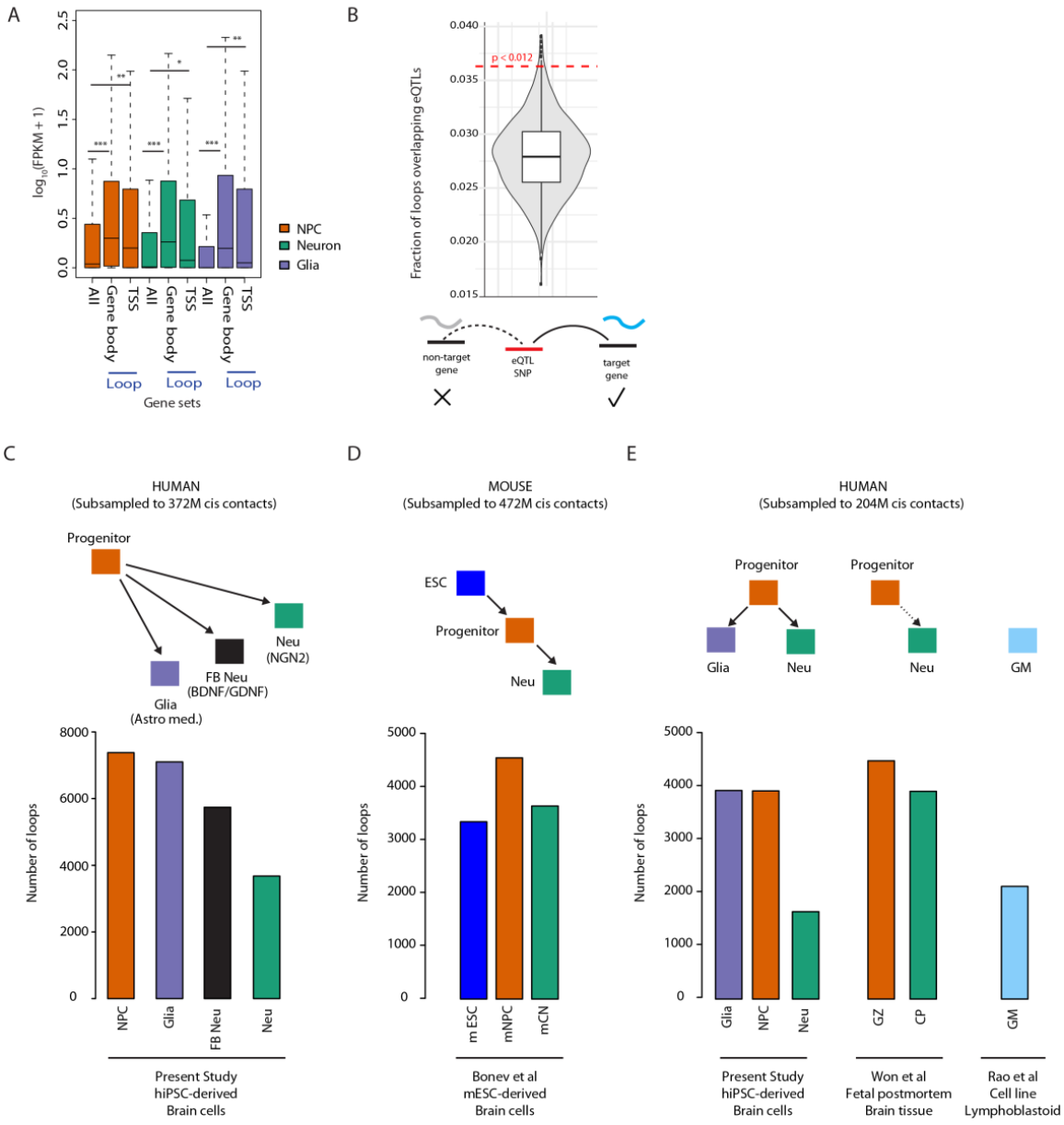


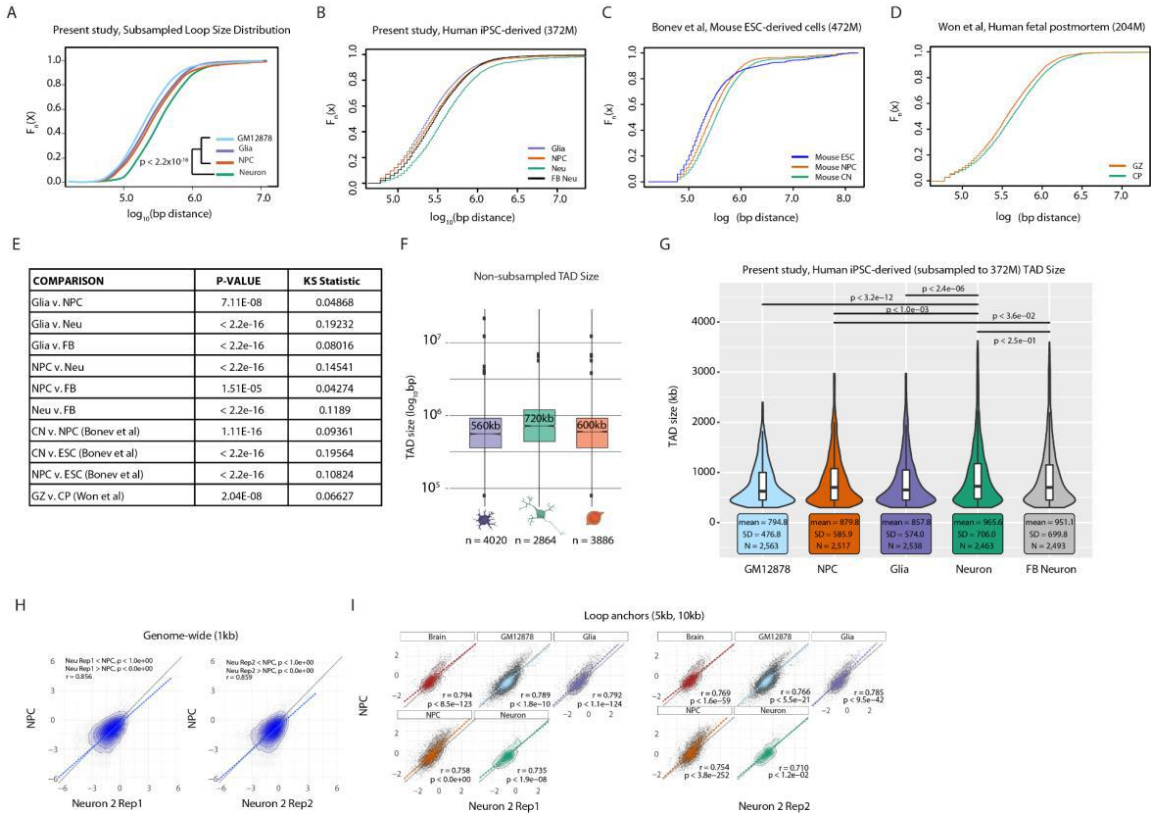
Figure 3.5 | Loop functional features and cell type-specific patterns in multiple model systems.
(**A**) Box-and-whisker plot comparing gene expression (log10(FPKM+1)) across all genes ("All"), the subset of genes located within the loop anchor ("Gene body"), and the subset of genes whose TSSs specifically are overlapping the loop anchor ("TSS"). * $P < 10^{-50}$, ** $P < 10^{-150}$, *** $P < 10^{-300}$. $P < 10^{-324}$ for All v. Gene body across all 3 cell types (***); $P = 4.40 \times 10^{-182}$ All v. TSS in NPCs; $P = 2.32 \times 10^{-52}$ All v. TSS in neurons; $P = 8.23 \times 10^{-173}$ All v. TSS in glia (non-subsampled datasets). (**B**) Violin plot of the distribution of fraction of overlaps between 1000 background loop sets (generated in silico) and brain eQTL SNP-target gene pairs. An eQTL loop was counted if one bin contained the eQTL SNP (i.e., eSNP) and the other bin contained the target gene (bottom). Red dashed line indicates the fraction of overlap observed between brain-specific loops and brain eSNP-gene pairs. (**C**) Count of 10kb resolution loops called by HiCCUPS in hiPSC-derived libraries only, including directed differentiation forebrain neurons from S1 ("FB Neuron"), all subsampled to 372M valid cis contacts. (**D**) Count of loops called in mouse ESC-derived libraries that were all subsampled to 472M valid cis contacts. (**E**) Count of loops after all human libraries, including previously published fetal brain Hi-C, were subsampled to 204M valid cis contacts.

We aimed to confirm that the observed net loss of loop formations during the NPC to neuron transition could be replicated across a variety of independent cell culture and in vivo approaches and was not specific to our methodological choice of *NGN2*-induction. We conducted an additional Hi-C experiment on cells differentiated from hiPSC-NPCs by means of a non-*NGN2* protocol that used only differentiation medium and yielded a heterogeneous population of hiPSC forebrain-neurons in addition to a small subset of glia (Brennand *et al.*, 2011). In addition, we reanalyzed Hi-C datasets generated from a mouse model of neural differentiation, consisting of mouse embryonic stem cell (mESCs), mESC derived NPCs (mNPC), and cortical neurons (mCN) differentiated from the mNPCs via inhibition of the Sonic Hedgehog (SHH) pathway (Bonev *et al.*, 2017). To

examine whether such genome-wide chromosomal loop remodeling also occurred in the developing brain in vivo, we reanalyzed Hi-C data from human fetal cortical plate (CP), mostly composed of young neurons, and forebrain germinal zone (GZ), primarily harboring dividing neural precursor cells in addition to a smaller subset of newly generated neurons (Won *et al.*, 2016). Across both the hiPSC-NPC-to-forebrain neuron and mESC-mNPC-mCN differentiation, in vitro neurons showed a 20% decrease in loops compared with their neural progenitors (Figure 3.5, C and D). Consistent with this, in vivo CP (neuron) compared with GZ (progenitor) showed a 13% decrease in loops genome-wide (Figure 3.5, E). The highly replicative cell types included here, mouse ESCs and human lymphoblastoid GM12878 cells, exhibited loop numbers very similar to their neuronal counterparts (Figure 3.5, D and E), suggesting that the changes in 3DG architecture from NPC to neurons do not simply reflect a generalized effect explained by mitotic potential.

Along with having fewer total loops, neurons exhibited a greater proportion of longer-range (>100 kb) loops than did NPCs or glia (two-sample two-tailed Kolmogorov-Smirnov test, KS_{range} = 0.1269 to 0.2317, $P < 2.2 \times 10^{-16}$ for three comparisons: Neu versus NPC/Glia/GM) (Figure 3.6, A). Likewise, in each of the alternative in vitro and in vivo analyses considered above, neurons exhibited a greater proportion of longer range (>100 kb) loops than did NPCs or glia [two-sample two-tailed Kolmogorov-Smirnov test, KS = 0.0427, $P = 1.5 \times 10^{-5}$ for hiPSC-NPC versus forebrain neuron; KS = 0.0936, $P = 1.1 \times 10^{-16}$ for mESC-

NPC versus mCN; *KS* = 0.0663, $P = 2.04 \times 10^{-8}$ for fetal CP (neuron) compared with GZ (progenitor)] (Figure 3.6, B, C, D, and E). Therefore, multiple in vitro and in vivo approaches comparing, in human and mouse, neural precursors to young neurons consistently show a reduced number of loops in neuron-enriched cultures and tissues, primarily affecting shorter-range loops.



COMPARISON	P-VALUE	KS Statistic
Glia v. NPC	7.11E-08	0.04868
Glia v. Neu	< 2.2e-16	0.19232
Glia v. FB	< 2.2e-16	0.08016
NPC v. Neu	< 2.2e-16	0.14541
NPC v. FB	1.51E-05	0.04274
Neu v. FB	< 2.2e-16	0.1189
CN v. NPC (Bonev et al)	1.11E-16	0.09361
CN v. ESC (Bonev et al)	< 2.2e-16	0.19564
NPC v. ESC (Bonev et al)	< 2.2e-16	0.10824
GZ v. CP (Won et al)	2.04E-08	0.06627

Figure 3.6 | Loop/TAD size comparisons across multiple datasets.
(**A**) Cumulative loop size distribution where $F_n(x)$ is the percent of loops whose size is less than or equal to the base pair distance (x). P-value is calculated using the Kolmogorov-Smirnov (KS) test. Neurons have a significantly larger proportion of loops that are longer in size than do other neural cells and non-neural GM12878 cells. (**B**) Cumulative loop size distribution of hiPSC-derived cell types subsampled to 372M cis contacts in the present study, including forebrain neurons. (**C**) Cumulative loop size distribution of mouse embryonic stem cells

(ESCs), ESC-derived neural progenitor cells (NPCs), and ESC-derived cortical neurons (CN). (**D**) Cumulative loop size distribution of germinal zone (GZ) and cortical plate (CP) (**E**) Table of p-values and corresponding KS test statistics for each comparison presented in B-D. (**F**) Box-and- whisker plot of the distribution of TAD sizes across cell types in non-subsampled datasets, showing larger TADs in neurons than in the other two neural cell types. (**G**) Violin plot of TAD size across hiPSC-derived neural cell types, including directed differentiation FB neurons, and non-neural GM12878 cell lines. P-value is calculated using the Wilcoxon-Mann-Whitney test. (**H**) 2D density plots with coverage-balanced, log_2-transformed ATAC signal for neurons on the x-axis and for NPCs on the y-axis. Open chromatin signatures of these two cell types were compared in 1kb bins genome-wide. P-values are reported from two-sided, paired t-tests. (**I**) 2D density plots with coverage-balanced, log_2-transformed ATAC signal for neurons on the x-axis and for NPCs on the y-axis. Open chromatin signatures of these two cell types were compared in HiCCUPS loop anchors (5 and 10kb) that were shared by NPCs, neurons, and glia ("Brain"); specific to GM12878; or specific to only glia, NPCs, or neurons. P-values are reported from two-sided, paired t-tests.

Consistent with studies in peripheral tissues reporting conservation of the overall TAD landscape across developmental stages, tissues, and species (when considering syntenic loci) (Dixon *et al.*, 2012, 2015), overall TAD landscapes (Rao *et al.*, 2014) remained similar between neurons, glia, and NPCs. Nonetheless, TADs also showed a subtle (~10%) increase in average size in neurons compared with isogenic NPCs, independent of the differentiation protocol applied (Wilcoxon-Mann-Whitney test, $P < 3.6 \times 10^{-2}$) (Figure 3.1, H and Figure 3.6, F and G), as highlighted here at a 3.4-Mb TAD at the *CDH2* cell adhesion gene locus (Figure 3.1, I). TAD remodeling may therefore reflect restructuring of nested subdomains within larger neuronal TADs (Tables 3.7 and 3.8). To examine whether such developmental reorganization of the brain's spatial genomes was associated with a generalized shift in chromatin structure, we applied the assay for transposase accessible chromatin with high-throughput

sequencing (ATAC-seq) to map open chromatin sequences before and after *NGN2*-neuronal induction (Table 3.1). Genome-wide distribution profiles for transposase-accessible chromatin were only minimally different between NPCs and neurons (Figure 3.6, H) and further revealed that both NPCs and neurons showed low to moderate chromatin accessibility [$-2.5 < \log_2(\textit{ATAC signal}) < 1$] for ≥ 89% of the anchor sequences comprising cell type–specific and shared "brain" loops in our cell culture system (Figure 3.6, I). These findings, taken together, point to widespread 3DG changes during the NPC-to-neuron transition and NPC-to-glia transition in human and mouse brain that are unlikely attributable to global chromatin accessibility differences. This includes highly cell type–specific signatures in gene ontologies of differentiation-induced loop prunings, reflecting neuronal and glial (non-neuronal) lineage commitment (Figure 3.4, A and G, and Table 3.4, B and C), and a subtle widening of average loop and TAD length in young neurons (Figure 3.1, H and Figure 3.6, A to G).

Chromosomal contacts associated with schizophrenia risk sequences

Because many schizophrenia risk variants lie in noncoding regions in proximity to several genes, we predicted that chromosomal contact mapping could resolve putative regulatory elements capable of conferring schizophrenia risk via their physical proximity (bypassing linear genome) to the target gene, as has been demonstrated in tissue in vivo (Won *et al.*, 2016; de la Torre-Ubieta *et al.*, 2018). We overlaid our cell type–specific interactions onto the 145 risk loci associated with schizophrenia risk (Ripke *et al.*, 2014; Pardiñas *et al.*, 2018).

Because only very few loops (defined as distinct pixels with greater contact frequency than neighboring pixels on a contact map) (Rao *et al.*, 2014) were associated with schizophrenia risk loci (*n* = 212, 81, and 17 loops in NPC, glia, and neurons, respectively) (Table 3.9), we applied an established alternative approach to more comprehensively explore the 3DG in context of disease-relevant sequences (Won *et al.*, 2016). This approach defines interactions as those filtered contacts that stand out over the global background and applies binomial statistics to identify chromosomal contacts anchored at disease-relevant loci (Won *et al.*, 2016). To begin, we examined the 40 loci with strongest statistical evidence for colocalization of an adult postmortem brain eQTL and schizophrenia genome-wide association study (GWAS) signal (Dobbyn *et al.*, 2018). Chromosomal contacts were called for 29 of the 46 eQTLs present in the 40 loci, with 8 of 29 (28%) of the loci showing significant interactions (binomial test, –log *q* value range = 1.33 to 11.0) between the eQTL-SNPs (eSNPs) in the one contact anchor and the transcription start site of the associated gene(s) in the other anchor (Table 3.10). We conclude that ~30% of risk locus–associated eQTLs with strong evidence for colocalization with GWAS signal bypass the linear genome and are in physical proximity to the proximal promoter and transcription start site of the target gene, resonating with previous findings in fetal brain tissue that used a similar contact mapping strategy (Won *et al.*, 2016).

Cell type–specific contact maps with 10-kb-wide bins, queried for the schizophrenia-associated loci, frequently revealed differential chromosomal

conformations in NPCs, glia, and neurons. For example, the risk locus upstream of the *PROTOCADHERIN* cell adhesion molecule gene clusters (chromosome 5), which is critically relevant for neuronal connectivity in developing and adult brain (Figure 3.7, A) (Yagi, 2012; Chen and Maniatis, 2013), showed through both observed/expected interaction matrix (Durand, Robinson, *et al.*, 2016) and global background-filtered contact mapping (Won *et al.*, 2016) a bifurcated bundle of interactions in NPCs, with one bundle emanating to sequences 5′ and the other bundle to sequences 3′ from the risk locus. In neurons, the 3′ bundle was maintained, but the 5′ bundle was "pruned," whereas glia showed the opposite pattern; these differences between the three cell types were highly significant (observed/expected Wilcoxon rank sum $P < 10^{-9}$ to 10^{-15}) (Figure 3.8, A to C). Dosage of the noncoding schizophrenia risk-SNP (rs111896713) at the *PCDH* locus significantly increased the expression of multiple *PROTOCADHERIN* genes (*PCDHA2*, *PCDHA4*, *PCDHA7*, *PCDHA8*, *PCDHA9*, *PCDHA10*, and *PCDHA13*) in adult frontal cortex of a large cohort of 579 individuals, including cases with schizophrenia and controls (Figure 3.7, B and Table 3.11) (Fromer *et al.*, 2016). The affected genes were interconnected to the disease relevant noncoding sequence in neurons and NPCs but not in glia (Figure 3.7, C). Therefore, cell-type-specific Hi-C identified chromosomal contacts anchored in schizophrenia-associated risk sequences that affected expression of the target gene(s). On the basis of earlier chromosome conformation capture assays at the site of candidate genes, the underlying

mechanisms may include alterations in transcription factor and other nucleoprotein binding at loop-bound cis-regulatory elements (Bharadwaj *et al.*, 2014) or even local disruption of chromosomal conformations (Bharadwaj *et al.*, 2013).

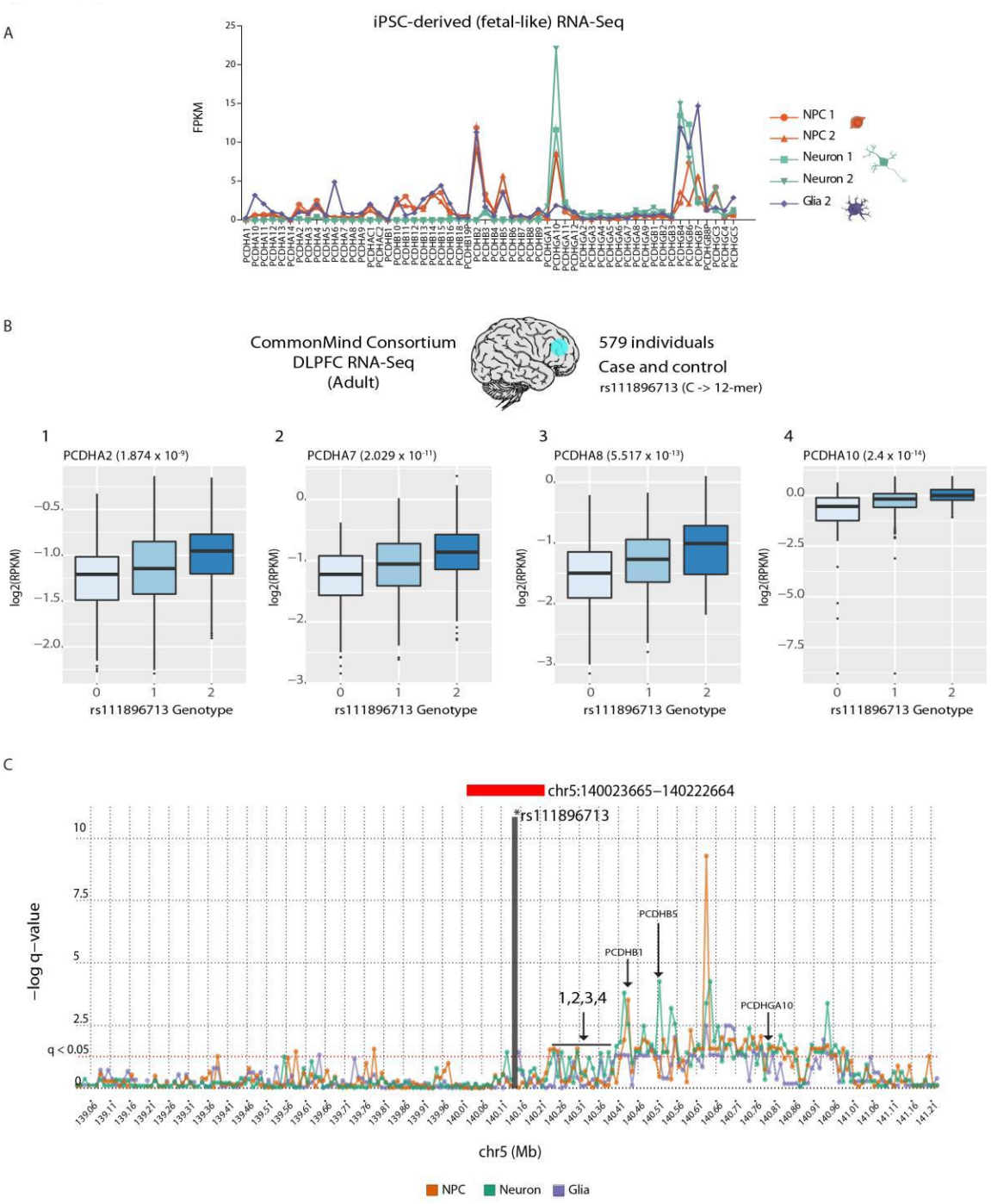

A
iPSC-derived (fetal-like) RNA-Seq
FPKM
NPC 1
NPC 2
Neuron 1
Neuron 2
Glia 2
B
CommonMind Consortium
DLPFC RNA-Seq
(Adult)
579 individuals
Case and control
rs111896713 (C -> 12-mer)
1
PCDHA2 (1.874 x 10-9)
2
PCDHA7 (2.029 x 10-11)
3
PCDHA8 (5.517 x 10-13)
4
PCDHA10 (2.4 x 10-14)
log2(RPKM)
rs111896713 Genotype
C
chr5:140023665–140222664
*rs111896713
−log q−value
q < 0.05
PCDHB5
PCDHB1
1,2,3,4
PCDHGA10
chr5 (Mb)
NPC
Neuron
Glia

Figure 3.7 | Risk-associated chromosomal contact mapping.
(**A**) RNA-seq FPKM expression for all clustered *PCDH* genes in NPCs (orange), neurons (green), and glia (purple). (**B**) The association of expression of *PCDHA2* (1; $P = 1.874 \times 10^{-9}$), *PCDHA7* (2; $P = 2.029 \times 10^{-11}$), *PCDHA8* (3; $P = 5.517 \times 10^{-13}$), *PCDHA10* (4; $P = 2.4 \times 10^{-14}$) with schizophrenia risk allele at the GWAS index SNP rs111896713, upstream of the clustered *PCDH* genes, using the CommonMind Consortium (CMC) postmortem prefrontal cortex RNA-seq dataset consisting of 258 schizophrenia subjects and 279 controls. (**C**) Binomial interaction landscape with 10kb anchor bin (gray vertical box) containing rs111896713 within the larger schizophrenia risk locus (solid red horizontal box), showing significant (above horizontal red dotted line) interactions with the four PROTOCADHERIN genes no. 1 to no. 4 significantly affected by risk polymorphisms (shown in B) in neurons (green) and NPCs (orange), but not in glia (purple).

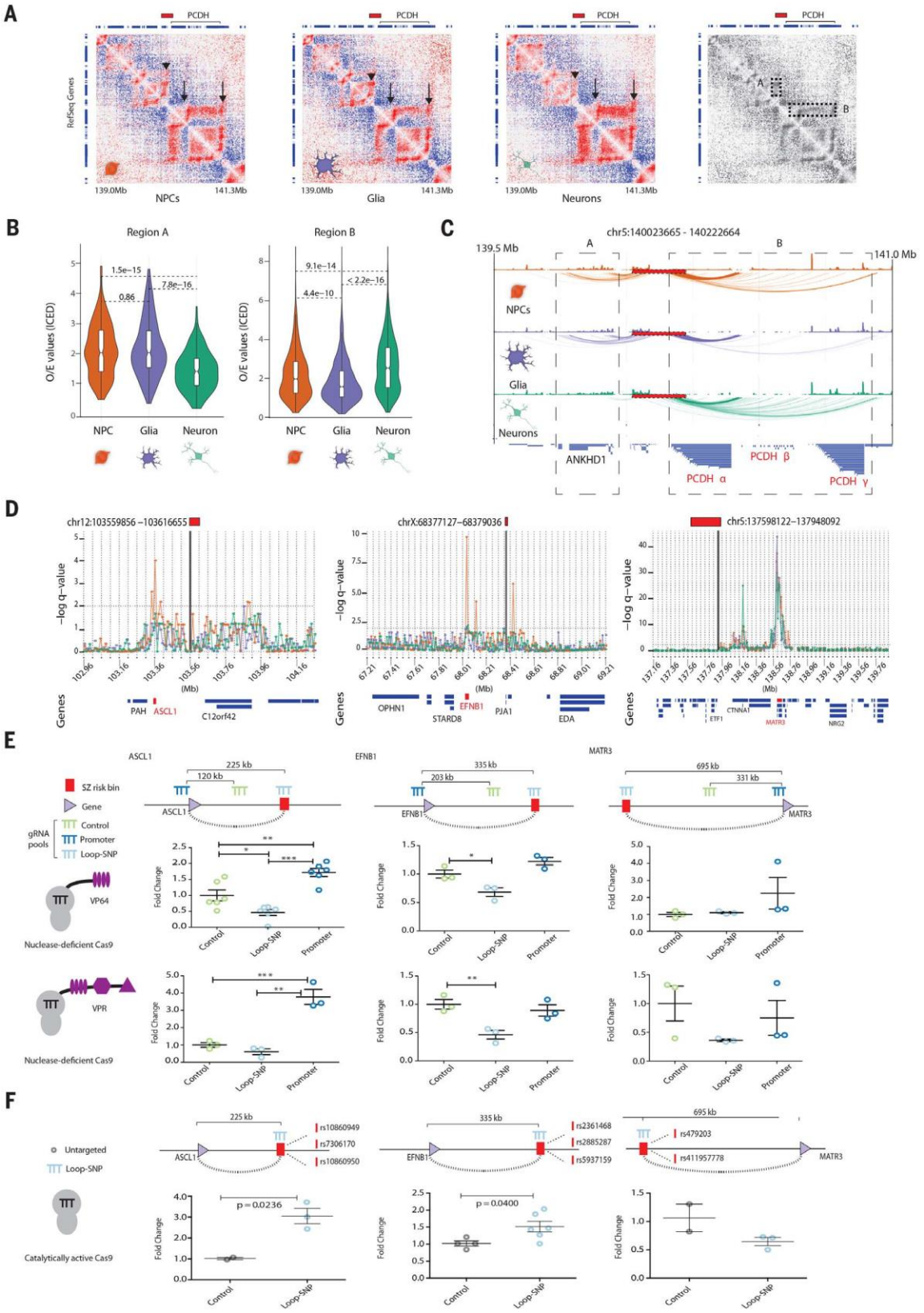

A
PCDH
RefSeq Genes
139.0Mb
141.3Mb
NPCs
Glia
Neurons
B
Region A
Region B
O/E values (ICED)
1.5e−15
0.86
7.8e−16
9.1e−14
4.4e−10
< 2.2e−16
NPC
Glia
Neuron
C
chr5:140023665 - 140222664
139.5 Mb
141.0 Mb
NPCs
Glia
Neurons
ANKHD1
PCDH α
PCDH β
PCDH γ
D
chr12:103559856 −103616655
chrX:68377127−68379036
chr5:137598122−137948092
−log q−value
(Mb)
Genes
PAH
ASCL1
C12orf42
OPHN1
STARD8
EFNB1
PJA1
EDA
ETF1
CTNNA1
MATR3
NRG2
E
ASCL1
EFNB1
MATR3
225 kb
120 kb
335 kb
203 kb
695 kb
331 kb
SZ risk bin
Gene
gRNA pools
Control
Promoter
Loop-SNP
VP64
Nuclease-deficient Cas9
VPR
Fold Change
F
Untargeted
Loop-SNP
Catalytically active Cas9
rs10860949
rs7306170
rs10860950
rs2361468
rs2885287
rs5937159
rs479203
rs411957778
p = 0.0236
p = 0.0400

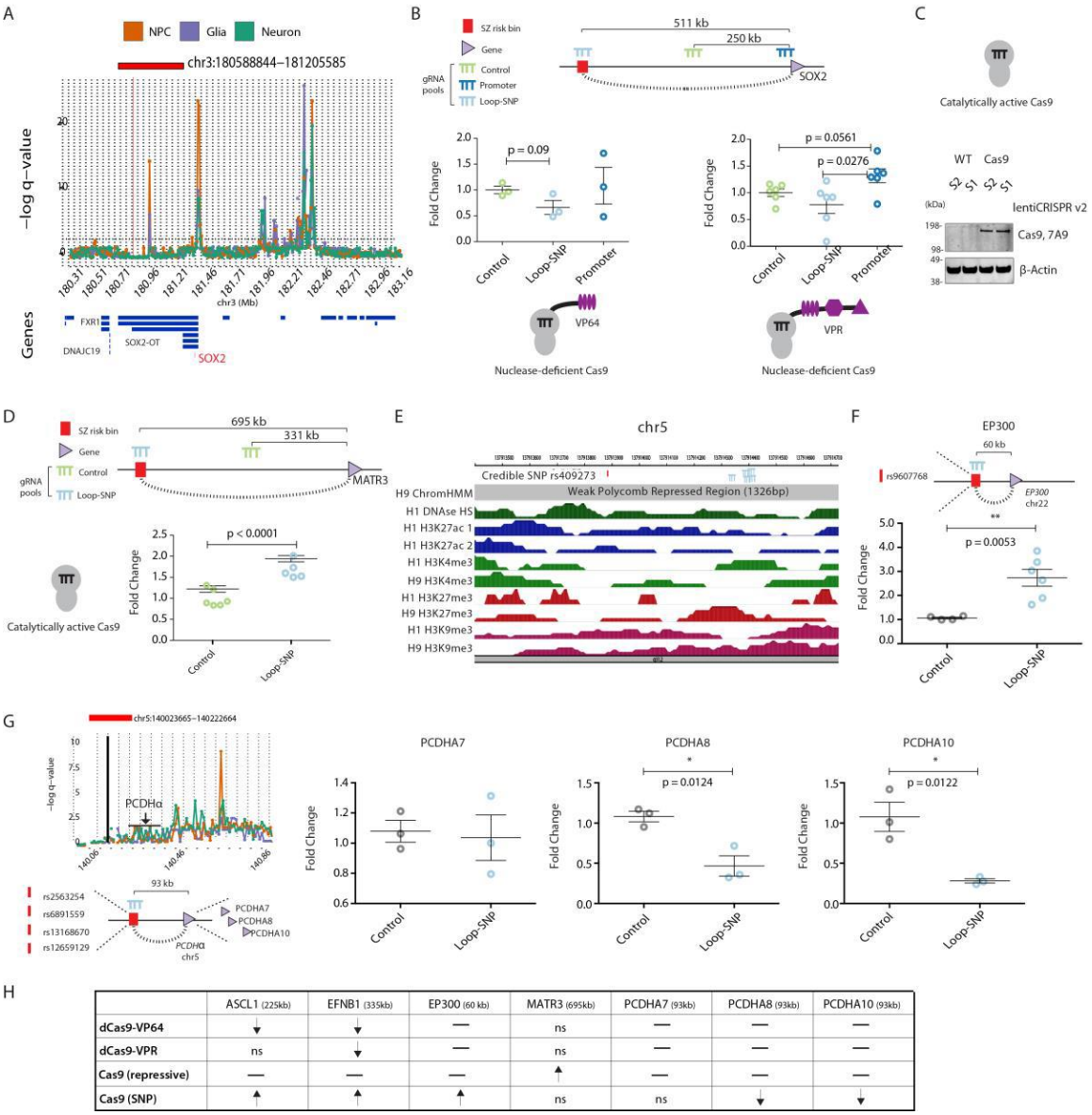
Figure 3.8 | Cell type–specific chromosomal contact maps at schizophrenia risk loci.
(**A**) Juicebox observed/expected interaction heatmaps at 10-kb resolution for the risk associated clustered *PCDH* locus chr5:140023665−140222664 for NPC, glia, and neurons as indicated. (Far right) Grayscale heatmap depicts areas of highly cell-specific contact enrichments: upstream genes including *ANKHD1* (dotted rectangle “A” and arrowhead) and downstream *PCDH* gene clusters (dotted rectangle “B” and arrows). Clustered *PCDH* gene expression patterns are available in Figure 3.7, A. (**B**) Violin plots of observed/expected interaction values in the regions A and B highlighted in (A). (**C**) Map of contacts identified by binomial statistics. Red box with dashed black line represents the schizophrenia risk locus, dotted boxes regions “A” and “B” in heatmaps. (**D**) Cell-type resolved contact map of 10-kb bins (bold, black vertical lines) within risk sequences on chr12 (left), chrX (middle), and chr5 (right); NPC (orange), neuron (green), glia (purple); –log *q* value, significance of contact between schizophrenia risk locus and each 10-kb bin; gene models (“Genes”) below with SNP-loop target gene highlighted in red. (**E**) Epigenomic editing (CRISPRa with nuclease-deficient dCas9 in NPCs) for three risk SNP-target gene pairs and their respective control sequences (top), measured with quantitative reverse transcription polymerase chain reaction (RT-PCR) (fold change from baseline) for VP64 (middle) and VPR (bottom) transcriptional activators. (**F**) Quantitative PCR gene expression changes upon directing catalytically active Cas9 to schizophrenia risk-associated credible SNPs (vertical red dashes with rsIDs) interacting via chromosomal contacts with promoters of *ASCL1*, *EFNB1*, and *MATR3* in NPCs. Targeting strategy and contact distances depicted above; *P < 0.05, **P < 0.01, ***P < 0.0001 (Figures 3.7 and 3.9).

Transcriptional profiles of hiPSC derived NPCs and neurons most closely resemble those of the human fetus in the first trimester (Brennand *et al.*, 2015); moreover, a portion of the genetic risk architecture of schizophrenia matches to regulatory elements that are highly active during prenatal development (Gulsuner *et al.*, 2013). We surveyed in our Hi-C datasets seven loci encompassing 36 “credible” (potentially causal) schizophrenia-risk SNPs with known chromosomal interactions in fetal brain to genes important for neuron development and function (Won *et al.*, 2016). We found that risk-associated chromosomal contacts were

conserved between our hiPSC NPCs and the published human fetal CP and germinal zone Hi-C datasets (Won *et al.*, 2016) for five of the seven loci (71%) tested (*CHRNA2*, *EFNB1*, *MATR3*, *PCDH*, and *SOX2*, but not *ASCL1* or *DRD2*) (Table 3.12). To test the regulatory function of these conserved risk sequence bound conformations, we performed single-guide RNA (sgRNA)-based epigenomic editing experiments on isogenic antibiotic-selected NPCs that stably express nuclease-deficient dCas9-VP64 (Maeder *et al.*, 2013; Perez-Pinera *et al.*, 2013) or dCas9-VPR (Chavez *et al.*, 2016; Ho *et al.*, 2017) transactivators (Table 3.13). Previous studies in peripheral cell lines succeeded in inducing gene expression changes by placing dCas9-repressor fusion proteins at the site of chromosomal contacts separated by up to 2 Mb of linear genome from the promoter target (Fulco *et al.*, 2016). We tested *ASCL1*-, *EFNB1*-, *MATR3*-, and *SOX2*- bound chromosomal contacts separated by 200- to 700-kb interspersed sequences (Figure 3.8, D and E; Figure 3.9, A; and Table 3.14). Pools of five individual sgRNAs directed against a risk-associated noncoding sequence bypassing 225 and 355 kb of genome consistently resulted in significantly decreased expression of *ASCL1* [one-way analysis of variance (ANOVA), $F_{VP64}(2, 15) = 22.20$, $P < 0.0001$; Dunnett's $P_{VP64} = 0.023$] and *EFNB1* target genes [one-way ANOVA, $F_{VP64}(2, 6) = 14.47$, $P = 0.0051$, Dunnett's $P_{VP64} = 0.0356$; $F_{VPR}(2, 6) = 1.46$, $P = 0.0111$, Dunnett's $P_{VPR} = 0.0088$], in comparison with positive (promoter bound) and negative (linear genome) control sgRNAs. Epigenomic editing of risk sequence 500 to 600 kb distant from the *SOX2* and

MATR3 loci did not alter target gene expression (Figure 3.8, D and E, and Figure 3.9, A and B), which could reflect practical limitations in nonintegrative transfection based (as opposed to viral) methods, impact of epigenetic landscape, or suboptimal guide RNA positioning (Ho *et al.*, 2017), further limited by the 10-kb contact map resolution. Because portions of the *MATR3*-bound risk sequences are embedded in repressive chromatin, we directed five sgRNAs for Cas9 nuclease mutagenesis toward a 138–base pair (bp) sequence within a *MATR3* long-range contact that was enriched with trimethyl-histone H3K27me3, commonly associated with *Polycomb* repressive chromatin remodeling, in order to disrupt it (Figure 3.9, C to E). This strategy produced a significant increase in *MATR3* expression upon ablation of the putative repressor sequence, whereas targeting *MATR3* (linear genome) control sequence remained ineffective (Figure 3.9, D and E). We conducted additional genomic mutagenesis assays, with sgRNAs directly overlapping credible SNPs participating in chromatin contacts with *ASCL1*, *EFNB1*, *EP300*, *MATR3*, *PCDHA7*, *PCDHA8*, and *PCDHA10* (Table 3.13). Cas9 nuclease deletion of interacting credible SNPs significantly increased gene expression of *ASCL1*, *EFNB1*, and *EP300* (P_{range} = 0.0053 to 0.04, t_{range} = 2.449 to 4.265) (Figure 3.8, F and Figure 3.9, F). Similar targeting of four credible SNPs upstream of the clustered *PCDH* locus significantly decreased levels, by ~50 to 60%, of *PCDHA8* and *PCDHA10* (P_{range} = 0.0122 to 0.0124, t_{range} = 4.326 to 4.343), two of the genes whose expression increased with dosage of the risk SNP rs111896713 in adult postmortem brain (Figure 3.7, B and Figure 3.9, G).

Taken together, our (epi)genomic editing assays (Figure 3.9, H) demonstrate that chromosomal contacts anchored in schizophrenia risk loci potentially affect target gene expression across hundreds of kilobases, which is consistent with predictions from chromosomal conformation maps from hiPSC-derived brain cells described here, and from developing (Won *et al.*, 2016; de la Torre-Ubieta *et al.*, 2018) and adult (Bharadwaj *et al.*, 2014) human brain tissue.



	ASCL1 (225kb)	EFNB1 (335kb)	EP300 (60 kb)	MATR3 (695kb)	PCDHA7 (93kb)	PCDHA8 (93kb)	PCDHA10 (93kb)
dCas9-VP64	↓	↓	—	ns	—	—	—
dCas9-VPR	ns	↓	—	ns	—	—	—
Cas9 (repressive)	—	—	—	↑	—	—	—
Cas9 (SNP)	↑	↑	↑	ns	ns	↓	↓

Figure 3.9 | Epigenomic and genomic editing at schizophrenia risk-associated chromosomal conformations.
(**A**) Interaction landscape at the *SOX2* locus on chr3, where highlighted (red) 10kb bin in *SOX2-OT* has a strong interaction with *SOX2*, a key transcription factor important in maintaining pluripotency in neural precursors. (**B**) *SOX2* epigenomic editing strategy (top) with results from VP64 (middle) and VPR (bottom) platforms, demonstrating that promoter-targeted gRNAs show increased gene expression and schizophrenia (SZ) risk SNP-targeted gRNAs potentially

decrease expression. (**C**) Western blot for Cas9 protein levels in wild type (left) and NPCs transduced with lentiviral vector containing Cas9 (right). (**D**) CRISPR deletion strategy (top) for contact-associated schizophrenia risk region (top) and results (bottom), which shows increased MATR3 after noncoding deletion ~700kb from promoter. Note that no SNPs were targeted by sgRNAs in this experiment, and the closest risk-associated SNP was 400bp from the nearest sgRNA. Positions of sgRNA and of nearest risk SNP are shown in Panel E browser window. (**E**) ENCODE epigenomic landscape of H9/H1 NPCs at the targeted region in panel D, suggesting weak polycomb repressed region. (**F**) qPCR gene expression upon directing catalytically active Cas9 to schizophrenia risk-associated credible SNPs (vertical red dashes with rsIDs) interacting with the promoters of *EP300*. Targeting strategy and interaction distance depicted above. (**G**) PCDH interaction landscape of 4 credible SNPs (vertical red dashes) with members of the PCDHα cluster, spanning roughly a 93kb distance (left). qPCR expression results for *PCDHA7*, *PCDHA8*, and *PCDHA10* after Cas9 mutagenesis of PCDH credible SNPs. (**H**) Summary table of CRISPR epigenomic and genomic editing experiments.

Cell type–specific schizophrenia-related chromosomal connectomes are associated with gene co-regulation and protein-protein association networks

Having shown that the chromosomal contact maps anchored in sequences associated with schizophrenia heritability undergo cell type–specific regulation (Figure 3.8, A to C), are reproducible in neural cell culture and fetal brain (Table 3.12), frequently harbor risk-associated eQTLs (Table 3.10), and bypass extensive stretches of linear genome to affect target gene expression in genomic and epigenomic editing assays (Figure 3.8, D to F, and Figure 3.9), we investigated chromosomal contacts for all 145 GWAS-defined schizophrenia risk loci together (Pardiñas *et al.*, 2018).We refer to the resulting “network” of risk loci and their 3D proximal genes as the “schizophrenia-related chromosomal connectome.”

Earlier studies in adult brain had shown that open-chromatin-associated histone modification and other “linear epigenome” mappings strongly link the genetic risk architecture of schizophrenia specifically with neuronal, as opposed to non-neuronal, chromatin (Girdhar *et al.*, 2018), which would suggest that similar cell-specific signatures may emerge in the risk-associated 3DG. Neurons and NPCs, but not the isogenic glia, showed a high preponderance of chromosomal contacts with schizophrenia-associated risk loci (Figure 3.10, A). There were 1203 contacts involving schizophrenia risk sequences that were highly specific to neurons (median distance between risk and target bins = 510 kb), 1100 highly specific for NPCs (median distance between risk and target bins = 520 kb), whereas only 425 highly specific for glia (median distance between risk and target bins = 580 kb) (Figure 3.10, A; Figure 3.11, A and B). There were also unexpectedly robust cell type- and gene-ontology-specific signatures, including genes associated with neuronal connectivity and synaptic signaling (Figure 3.10, B and Tables 3.15 and 3.16). Separate analysis of the Psychiatric Genomics Consortium “PGC2” 108 risk loci (Ripke *et al.*, 2014) yielded similar results (Figure 3.12, A and B).

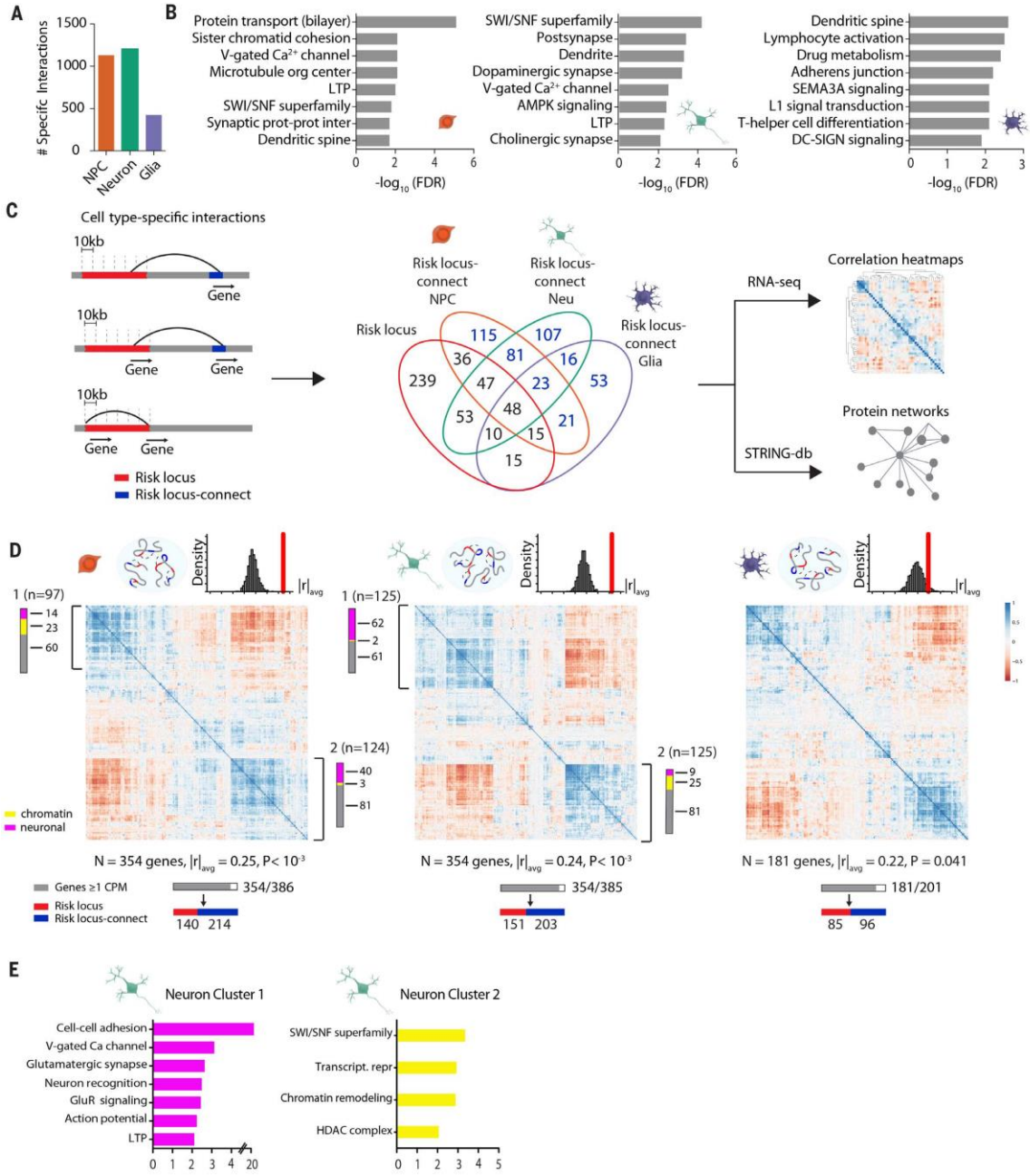


Figure 3.10 | Expanded GWAS risk connectome is associated with gene coregulation.

(**A**) Counts of highly cell type–specific contacts associated with schizophrenia risk in each of the three hiPSC-derived cell types. (**B**) GO enrichment of genes located in schizophrenia risk contacts in NPCs (left), neurons (middle), and glia (right). (**C**) (Left) Schematic workflow of analyses performed with cell type specific contact genes, distinguished as “risk locus” and “risk locus–connect” genes. (Middle) Venn diagram of genes located in the 145 loci and those found in

cell type–specific contacts, with numbers in blue indicating “risk locus–connect” genes. (Right) Schematic workflow of analyses performed with combined set of “risk locus” and “risk locus–connect” genes. (**D**) RNA Pearson transcriptomic correlation heatmaps consisting of risk locus and risk locus–connect genes derived from the cell type–specific contacts of NPCs (left), neurons (middle), and glia (right). Organization scores ($|r|_{avg}$) for each heatmap are reported with *P* values from sampling analysis. Schematics above heatmaps are representations of each cell type’s particular connectome (blue oval) and frequency distribution of organization scores from permutation analyses of randomly generated heatmaps (red, observed organization score of heatmap being tested). The gray bar corresponds to *n* genes that have at least 1 count per million in RNA-seq dataset out of the total number of genes and are used to construct the heatmap; red and blue bars indicate how many of the genes in the heatmap are in a risk locus (red) and are risk locus–connect (blue). Fuchsia, neuron connectivity/synaptic function genes; yellow, chromatin remodeling genes as determined from gene ontology analysis in (**E**). Additional information on coexpression clusters is provided in tables S22 and S23 (Figure 3.11 and 3.12).

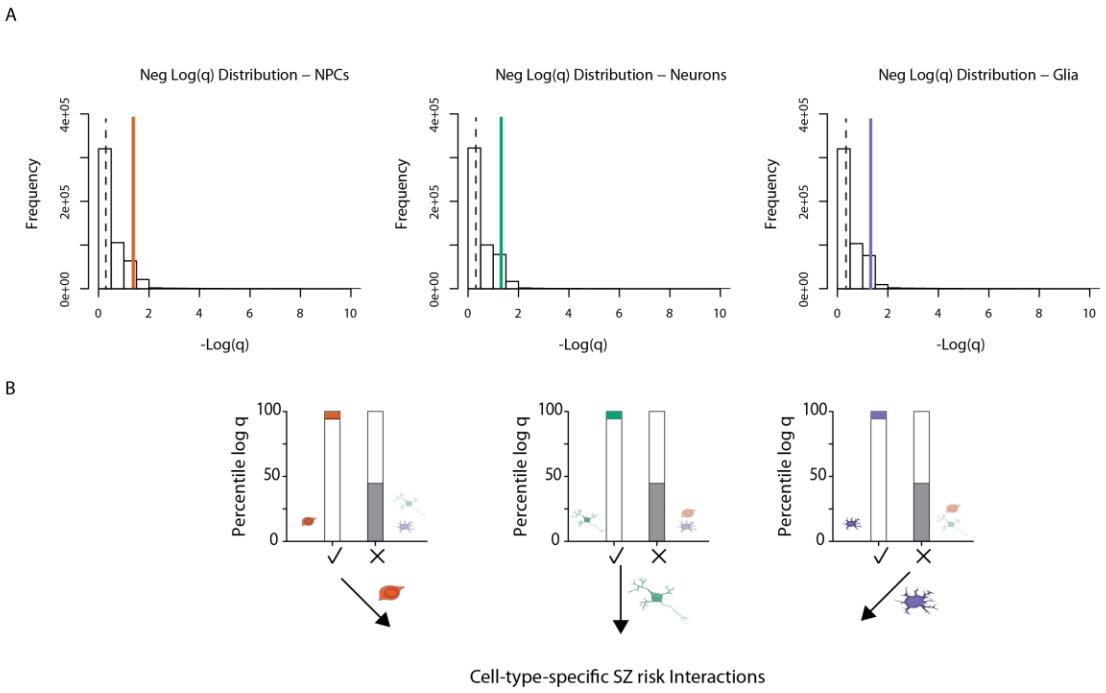


Figure 3.11 | Determining cell type-specific PGC interactions.
(**A**) Frequency histograms of the distribution of –log(*q*) values of binomial interactions for each cell type. Colored lines = 95th percentile of the union set of

all –log(*q*) (1.47); grey dashed lines = 50th percentile of the union set of all –log(*q*) (0.33). (**B**) Work flow for determining cell type-specific PGC interactions. Interactions that had -log(*q*) values ≥ 95th percentile in one cell type (color highlight) and -log(*q*) values < 50th percentile in the remaining two cell types (grey highlight) were defined as cell type-specific interactions.

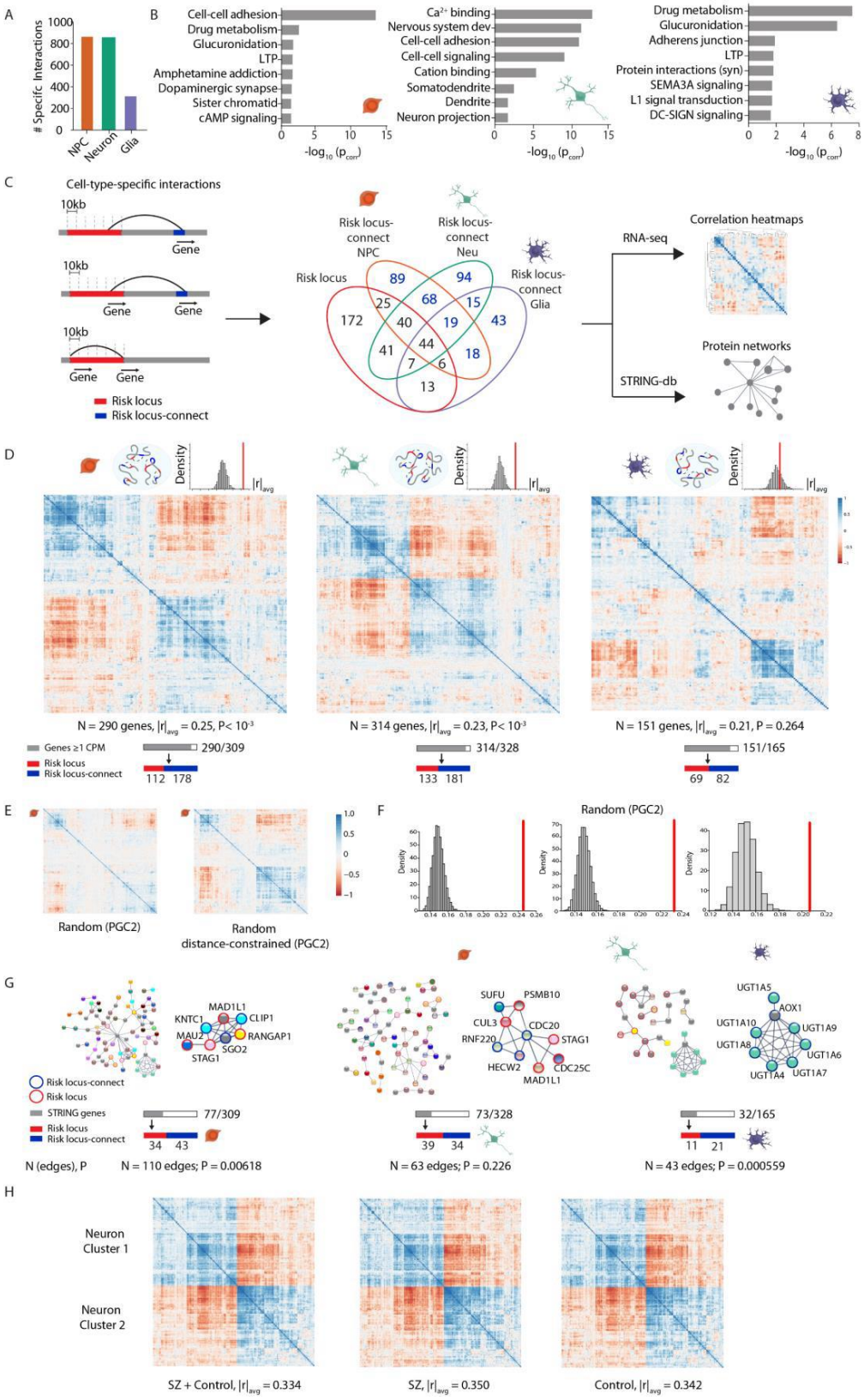

A
Specific Interactions
NPC
Neuron
Glia
B
Cell-cell adhesion
Drug metabolism
Glucuronidation
LTP
Amphetamine addiction
Dopaminergic synapse
Sister chromatid
cAMP signaling
Ca2+ binding
Nervous system dev
Cell-cell adhesion
Cell-cell signaling
Cation binding
Somatodendrite
Dendrite
Neuron projection
Drug metabolism
Glucuronidation
Adherens junction
LTP
Protein interactions (syn)
SEMA3A signaling
L1 signal transduction
DC-SIGN signaling
C
Cell-type-specific interactions
10kb
Gene
Risk locus
Risk locus-connect
Risk locus-connect NPC
Risk locus-connect Neu
Risk locus-connect Glia
Risk locus
RNA-seq
Correlation heatmaps
Protein networks
STRING-db
D
Density
N = 290 genes, |r|avg = 0.25, P< 10-3
N = 314 genes, |r|avg = 0.23, P< 10-3
N = 151 genes, |r|avg = 0.21, P = 0.264
Genes ≥1 CPM
290/309
314/328
151/165
E
Random (PGC2)
Random distance-constrained (PGC2)
F
Random (PGC2)
G
MAD1L1
KNTC1
CLIP1
MAU2
RANGAP1
SGO2
STAG1
SUFU
PSMB10
CUL3
CDC20
RNF220
STAG1
HECW2
CDC25C
MAD1L1
UGT1A5
AOX1
UGT1A10
UGT1A9
UGT1A8
UGT1A6
UGTTA4
UGT1A7
STRING genes
77/309
73/328
32/165
N (edges), P
N = 110 edges; P = 0.00618
N = 63 edges; P = 0.226
N = 43 edges; P = 0.000559
H
Neuron Cluster 1
Neuron Cluster 2
SZ + Control, |r|avg = 0.334
SZ, |r|avg = 0.350
Control, |r|avg = 0.342

Figure 3.12 | Expanded GWAS risk connectome is associated with significant gene coregulation and protein-protein association networks. (**A**) Counts of highly cell type-specific significant interactions in each of the 3 hiPSC-derived cell types. (**B**) GO enrichment of genes located in significant schizophrenia risk- interactions in NPCs (left), neurons (middle), and glia (right). (**C**) Schematic workflow of analyses performed with cell type-specific interaction genes, distinguished as “risk locus” and “risk locus-connect” genes (left). Venn diagram (middle) of genes located in the 108 loci and those found in cell type-specific interactions, with numbers in blue indicating “risk locus-connect” genes. (**D**) RNA Pearson correlation heatmaps consisting of risk locus and risk locus-connect genes derived from the cell type-specific interactions of NPCs (left), neurons (middle), and glia (right). Organization scores ($|r|_{avg}$) for each heatmap are reported with *P*-values from permutation analysis. Schematics above heatmaps are representations of each cell type’s unique connectome (blue oval) and frequency distribution of organization scores from permutation analyses of randomly generated heatmaps (red = observed organization score of heatmap being tested);grey bar corresponds to *n* genes that have at least 1 count per million in RNA-seq dataset out of the total number of genes and are used to construct the heatmap; red and blue bars indicate how many of the genes in the heatmap are in a risk locus (red) and are risk locus-connect (blue). (**E**) Representative example of a heatmap generated by randomly sampling an identical number of genes as in the heatmap being queried without (left) and with (right) distance constraints (see Materials and methods for inclusion criteria and further details). (**F**) Frequency distribution of permuted organization scores for NPCs (left), neurons (middle), and glia (right) for random permutation analyses without distance-constraint in the PGC2-anchored interactions. (**G**) Overview and representative examples (zoomed in) of protein-protein association networks in NPCs (left), neurons (middle), and glia (right). Numbers of edges connecting the proteins in each network and STRING-computed *P*-values are reported below. Grey bar indicates the subset of these genes whose proteins are involved in the network out of the total number of genes from cell type-specific interactions; red and blue bars indicate how many of the genes in the network are in a risk locus (red) and are risk locus-connect (blue). Risk locus (red circle outline) and risk locus-connect (blue circle outline) are marked in the representative examples. (**H**) Subset heatmaps of only those genes in the neuronal coexpression clusters 1 and 2 (refer to Figure 3.10 D) from all 94 (hiPSC-derived NPC and neurons; N = 47 libraries from 14 schizophrenia cases and N = 47 libraries from 12 controls) samples (left), schizophrenia cases only (middle), and control only (right).

Because spatial 3DG proximity of genes is an indicator for potential coregulation (Kustatscher, Grabowski and Rappsilber, 2017), we tested whether

the neural cell type–specific schizophrenia-related chromosomal connectome showed evidence of coordinated transcriptional regulation and proteomic interaction of the participating genes. To this end, we generated lists of genes anchored in the most highly cell type–specific schizophrenia risk associated contacts (Materials and methods) (Figure 3.10, C, Figure 3.11, B, and Table 3.15). Thus, for the NPC-specific contacts, we counted 386 genes, including 146 within the risk loci and another 240 genes positioned elsewhere in the linear genome but connected via an intrachromosomal contact to within-risk-locus sequences. Similarly, for the neuron-specific contacts, we identified 385 genes, including 158 within risk loci and 227 outside of risk loci (Figure 3.10, C). Last, for glia-specific contacts, we identified 201 genes, including 88 within and 113 outside of risk loci. We labeled the intrachromosomal contact genes located outside of schizophrenia risk loci as “risk locus-connect,” which we define as a collection of genes identified only through Hi-C interaction data, expanding—depending on cell type—by 50 to 150% the current network of known genes overlapping risk sequences that is informed only by GWAS (Figure 3.10, C).

To examine whether such types of disease associated, cell-type-specific chromosomal connectomes were linked to a coordinated program of gene expression, we analyzed a merged transcriptome dataset (comprised of 47 hiPSC-NPC and 47 hiPSC-forebrain neuron RNA-seq libraries from 22 schizophrenia and control donors not related to those of our Hi-C datasets) (Hoffman *et al.*, 2017). We examined pair-wise correlations of the collective sets

of the 386 NPC, 385 neuron, and 201 glia genes representing “risk locus” and “risk locus-connect” genes (cell-type-specific “risk connectomes”). The risk connectome for each cell type showed extremely strong pair-wise correlations, with two of the largest clusters visualized on the neuron and NPC correlation matrices involving an admixture of 354 “risk locus” and “risk locus connect” genes each, and similarly 181 genes from the glia matrix (Figure 3.10, D and Table 3.17). The averaged gene-by-gene transcript correlation index for each matrix overall, defined here as “organization score” ($|r|_{avg}$), for the NPCs, neurons, and glia were 0.22 to 0.25. Such levels of organized gene expression were robustly significant for NPC and neurons, after controlling for linear genomic distance (1000 random samplings, $|r|_{avg}$, $P < 0.001$ for NPC and for neuron; $P = 0.041$ for glia) (Figure 3.10, D, Figure 3.12, E, and Table 3.18). There were four large clusters in the correlation matrices of the neuronal and NPC risk connectome: neuronal connectivity and synaptic signaling proteins (neuron cluster 1 and NPC cluster 2) and epigenetic regulators (neuron cluster 2 and NPC cluster 1). For example, within neuron cluster 1 (Figure 3.10, D, middle), 62 of 125 genes encoded neural cell adhesion and synaptic molecules, voltage-gated ion channels, and other neuron specific genes (Figure 3.10, E and Tables 3.19 and 3.20). We thus conclude that the chromosomal connectomes associated with schizophrenia risk are cell type specific, with the neuronal risk connectome particularly enriched for genes pertaining to neuronal connectivity, synaptic signaling, and chromatin remodeling (Figure 3.10, D and E). Analyses of

the subset of PGC2 risk loci (108 and 145) provided similar results (Figure 3.12, C to F). Additionally, organization scores for neuron cluster 1 and cluster 2 genes were similar between hiPSC-derived NPCs and forebrain neurons from schizophrenia cases (n = 47) and control (n=47), suggesting that many risk locus–connect and risk locus genes are coregulated across individuals (Figure 3.12, H).

Numerous proteins encoded by risk locus and risk locus–connect genes were associated with synaptic signaling (Table 3.21). The cell type–specific risk locus–connect and risk locus genes show significant protein-protein interaction network effects for NPCs (P = 0.0004) and neurons (P = 0.009) but not glia (Figure 3.13, A, Figures 3.14 to 3.16, and Table 3.21) when examined by using the STRING database v10.5 (Szklarczyk *et al.*, 2015, 2017). We observed many proteomic clusters, including large groups of epigenomic regulators associated with the SWI/SNF (SWItch/Sucrose Non-Fermentable) chromatin remodeling complex and histone lysine methyltransferases and demethylases (Figure 3.13, A and Figures 3.14 and 3.15), many of which were the genes identified in NPC cluster 1 and neuron cluster 2 of the transcriptome analysis (Figure 3.10, D and E). The transcriptomic correlation heatmaps for these protein networks (“STRING” genes), when compared with randomly generated subset heatmaps from the overall (“Full”) schizophrenia-related chromosomal connectome (Figure 3.10, D), had higher organization scores in NPCs and neurons (NPC $|r|_{avg}$ = 0.2963, P = 0.007; neuron $|r|_{avg}$ = 0.2877, P = 0.008, glia $|r|_{avg}$ = 0.2225, P =

0.595, STRING versus full permutation test) (Materials and methods) (Figure 3.13, B, Figures 3.17 to 3.19, and table S21). Because the transcriptomic correlation heatmap for the schizophrenia-related chromosomal connectome was significantly decreased by the removal specifically of the NPC STRING protein network genes ($P < 10^{-3}$) (Table 3.21), this subset of STRING-interacting proteins may drive the observed orchestrated coregulation. Within these transcriptome- and proteome-based regulatory networks were numerous occasions of coregulated (RNA) and interacting (protein) risk locus and risk locus–connect genes that share the same TAD, including *CDC20*, which regulates dendrite development (Puram *et al.*, 2011; Watanabe, Khodosevich and Monyer, 2014) and is associated at the protein level with *RNF220*, an E3 ubiquitin-ligase and β-catenin stabilizer (Figure 3.13, C) (Ma *et al.*, 2014).

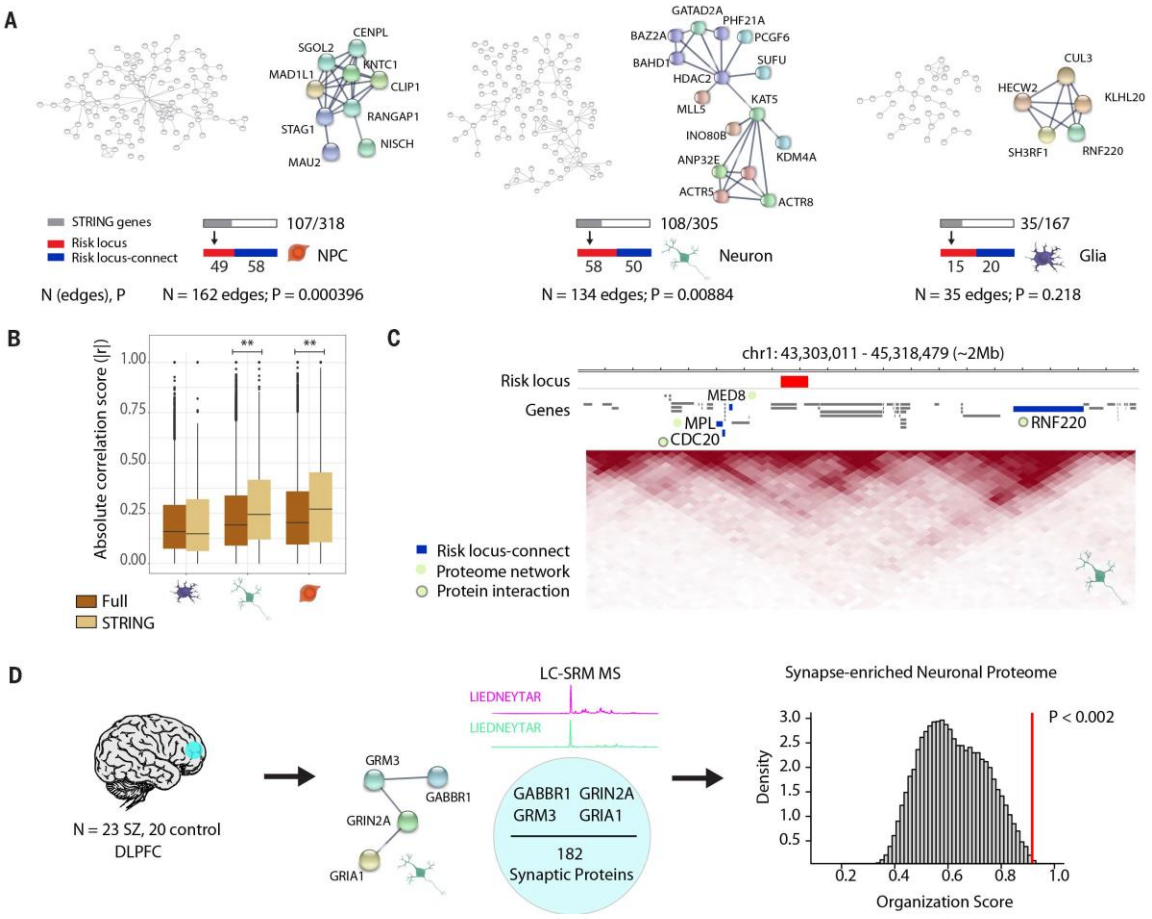


Figure 3.13 | Expanded GWAS risk connectome is linked to protein-protein association networks.

(**A**) Overview and representative examples (zoomed in) of protein-protein association networks in NPCs (left), neurons (middle), and glia (right). Numbers of edges connecting the proteins in each network and STRING-computed *P* values are reported below. Gray bar indicates the subset of these genes whose proteins are involved in the network out of the total number of genes from cell type–specific interactions; red and blue bars indicate how many of the genes in the network are in a risk locus (red) and are risk locus–connect (blue).

(**B**) Comparison of organization scores between the full RNA transcriptomic correlation heatmaps (brown) (Figure 3.10, D) and the "STRING" heatmaps (tan) (Figures 3.17 to 3.19), consisting of only those genes in protein networks for each cell type. Permutation test, **P<0.01. (**C**) Representative neuronal TAD landscape (chr1, ~2 Mb) depicting a schizophrenia risk–associated locus (red) with its risk locus–connect genes (blue), *MED8*, *MPL*, *CDC20*, and *RNF220*, which are members of the neuronal schizophrenia protein network (green circle). *CDC20* and *RNF220* interact at the protein level (green circle with gray border).

(**D**) (Left) Liquid chromotography–selected reaction monitoring (LC-SRM) mass spectrometry (MS) was performed on dorsolateral prefrontal cortex (DLPFC) tissue from 43 adult postmortem brains (23 schizophrenia, 20 control). (Middle) 182 neuronal proteins were reliably quantified, and four of them were observed to have associations in the neuron protein network in (A). (Right) GABBR1, GRM3, GRIN2A, and GRIA1 proteins were found to have significantly more correlated expression than expected by random permutation analysis. Additional information on protein-protein interactions is provided in Figures 3.12 to 3.19.

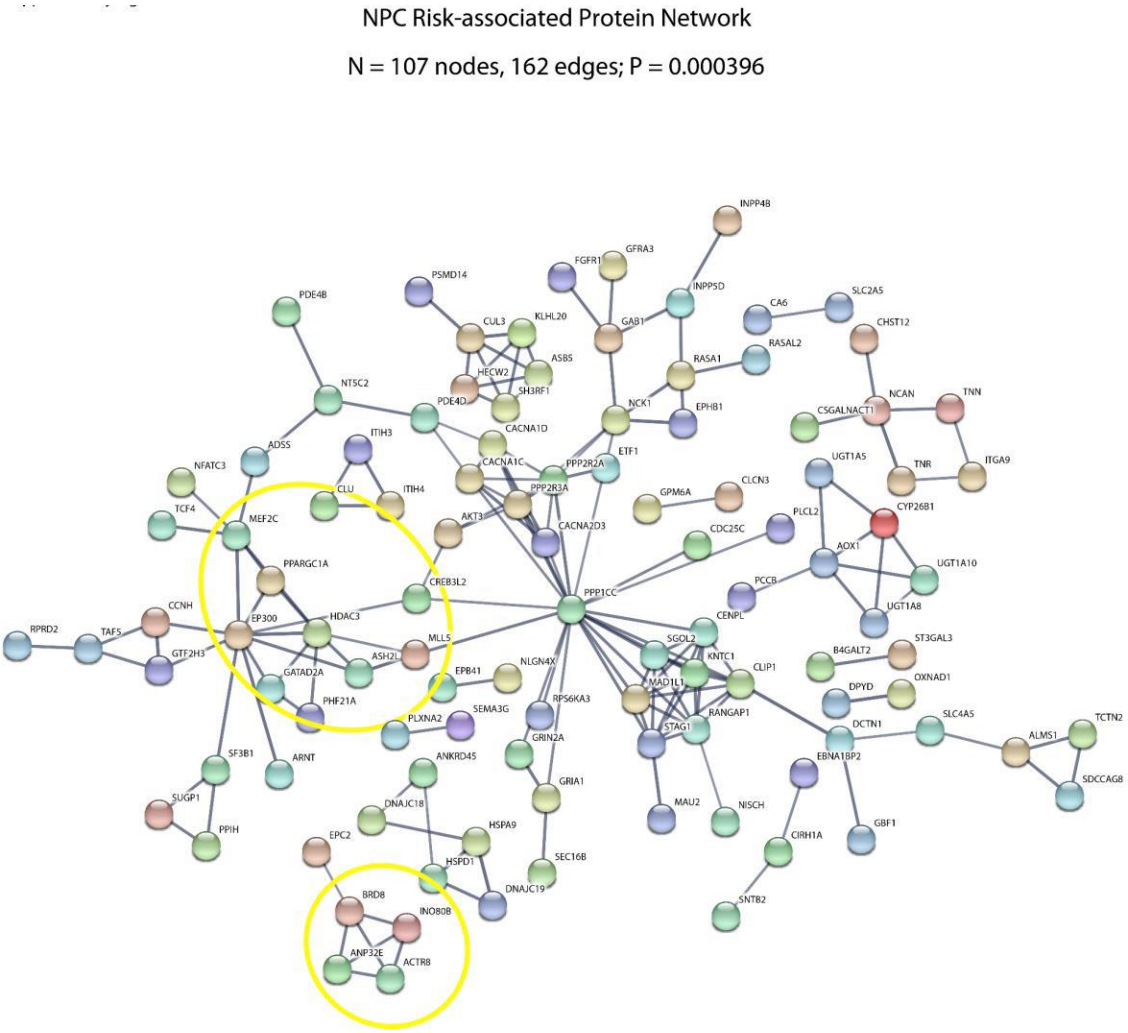


Figure 3.14 | NPC schizophrenia risk-associated protein-protein association network.
Protein-protein network (STRING) derived from genes located in NPC-specific schizophrenia risk interactions. Yellow circles roughly outline groups of genes associated with chromatin remodeling, as highlighted in Figure 3.10, D and E.

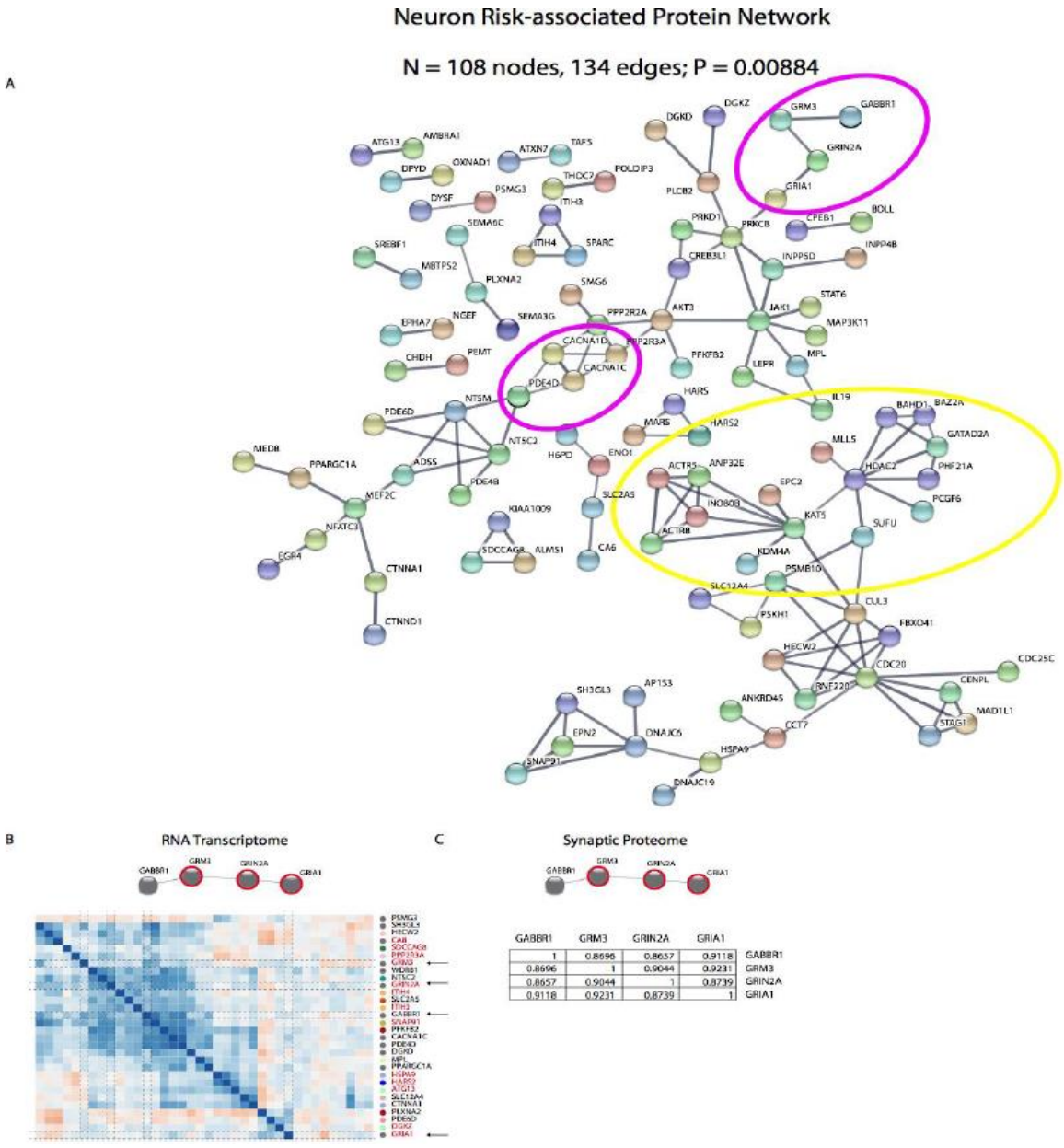


Figure 3.15 | Neuron schizophrenia risk-associated protein-protein association network.
(**A**) Protein-protein network (STRING) derived from genes located in neuron-specific schizophrenia risk interactions. Yellow and fuchsia circles roughly outline groups of genes associated with chromatin remodeling and neuronal connectivity, respectively, as highlighted in Fig 3.10, D and E. (**B**) RNA correlation heatmap (PGC2) zoomed in to a cluster of coexpression that contains 4 genes of interest (*GABBR1*, *GRM3*, *GRIN2A*, *GRIA1*) that were identified in the

protein network in panel A. Red = risk locus genes are in red; arrows indicate 4 synaptic genes of interest; each color circle represents a different TAD whereby genes co-localized with one or more genes in same TAD or adjacent TADs share the same color. (**C**) Pearson correlation matrix of protein measures for the 4 proteins of interest from LC-SRM MS of adult postmortem DLPFC.

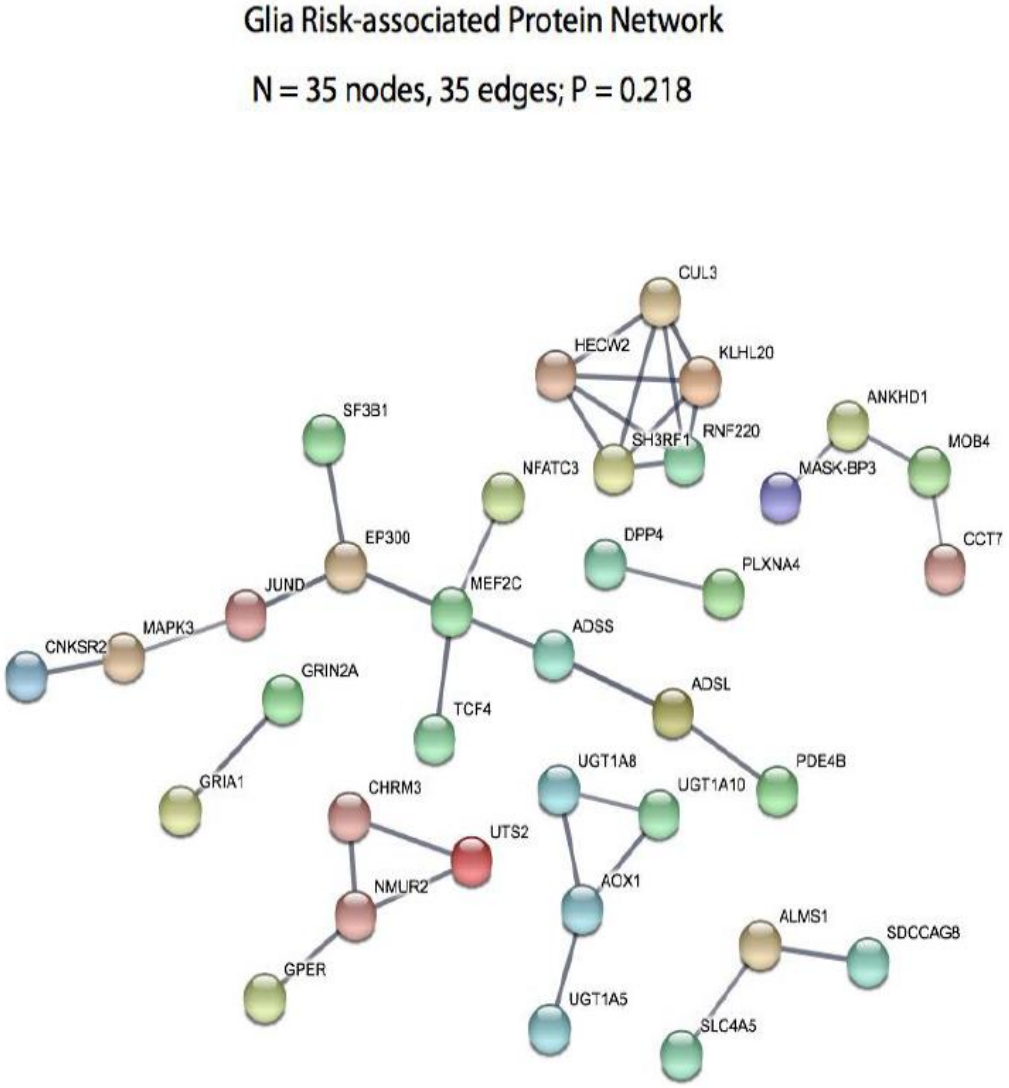


Figure 3.16 | Glia schizophrenia risk-associated protein-protein association network.
Protein-protein network (STRING) derived from genes located in glia-specific schizophrenia risk interactions.

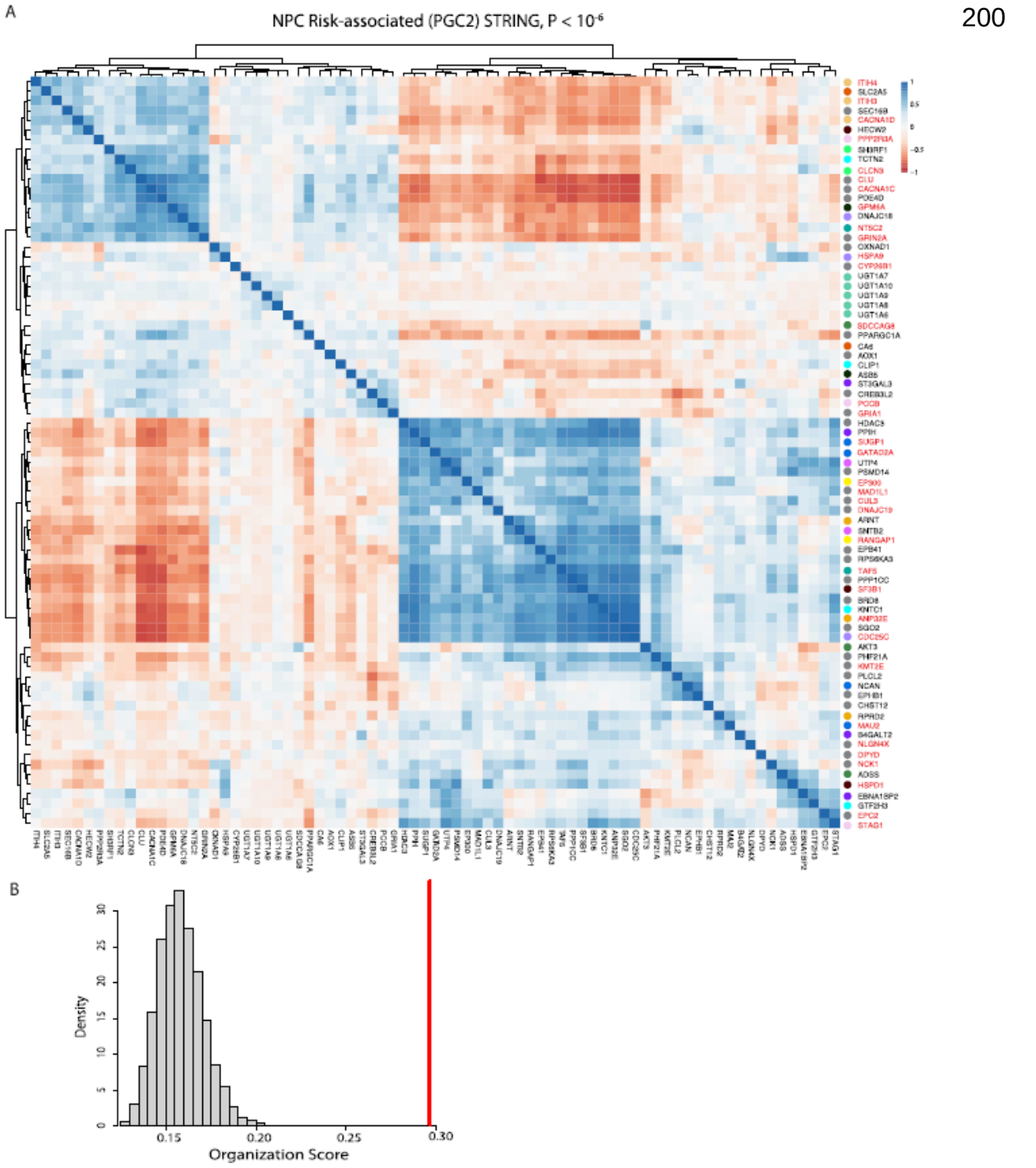

A
NPC Risk-associated (PGC2) STRING, P < 10⁻⁶
B
Density
Organization Score
0.15
0.20
0.25
0.30

Figure 3.17 | NPC schizophrenia risk-associated STRING subset genes show greater transcriptional organization than full risk connectome gene list (PGC2).

(**A**) Pearson correlation heatmap of genes participating in NPC-specific STRING protein-protein network. Risk locus gene (red), gene present in TAD with no other genes in the network (grey circle), gene co-localized with one or more genes in same TAD or adjacent TADs (colored circles; each color represents a different TAD; see Table S24). (**B**) Frequency distribution of permuted organization scores, randomly sampling from all genes in NPC risk connectome. Red line = organization score for risk-associated Pearson correlation in panel A.

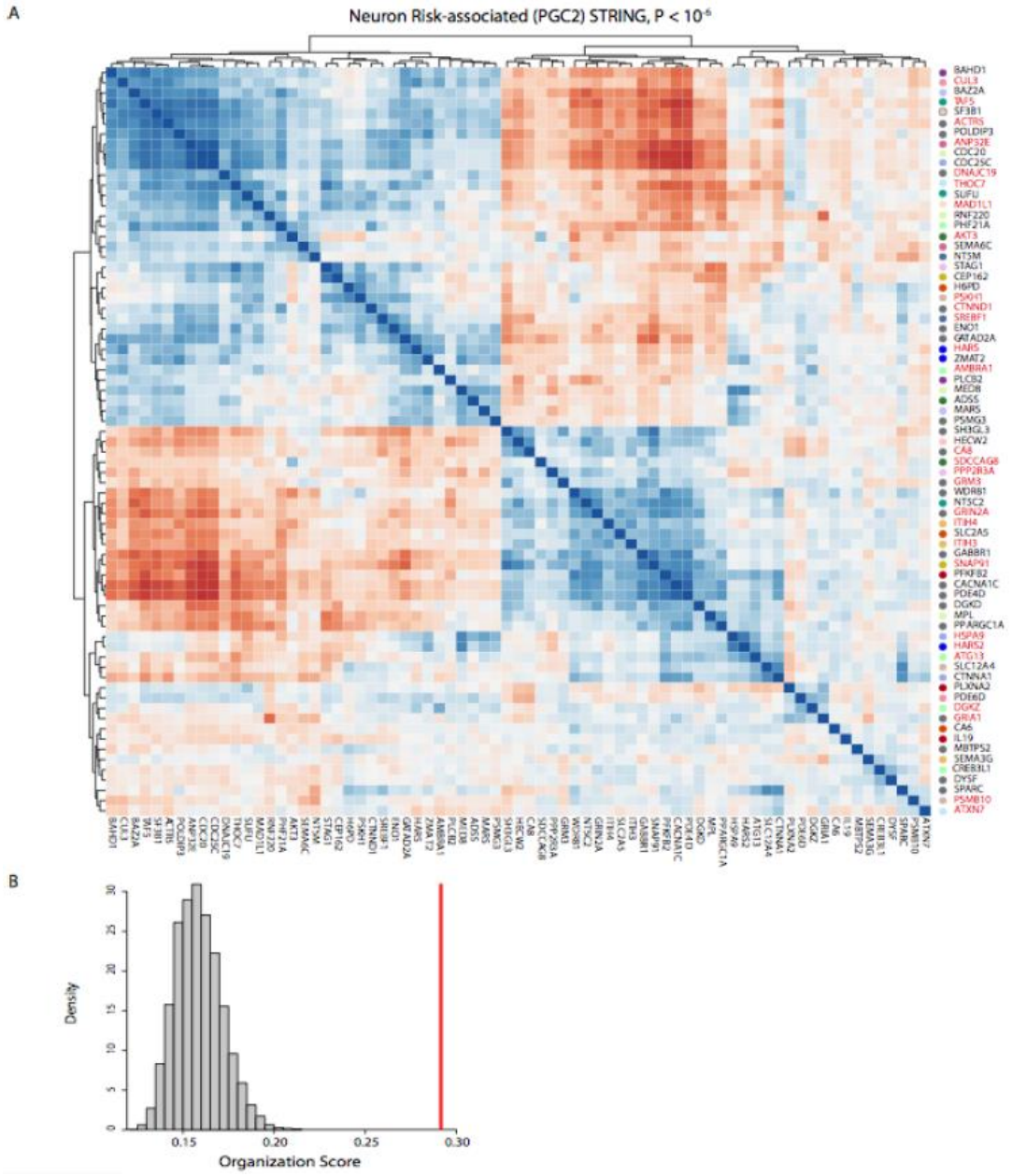

A
Neuron Risk-associated (PGC2) STRING, P < 10⁻⁶
B
Density
Organization Score
0.15
0.20
0.25
0.30

Figure 3.18 | Neuron schizophrenia risk-associated STRING subset genes show greater transcriptional organization than full risk connectome gene list (PGC2).
(**A**) Pearson correlation heatmap of genes participating in neuron-specific STRING protein-protein network. Risk locus gene (red), gene present in TAD with no other genes in the network (grey circle), gene co-localized with one or more genes in same TAD or adjacent TADs (colored circles; each color represents a different TAD; see Table S24). (**B**) Frequency distribution of permuted organization scores, randomly sampling from all genes in neuron risk connectome. Red line = organization score for risk-associated Pearson correlation in panel A.

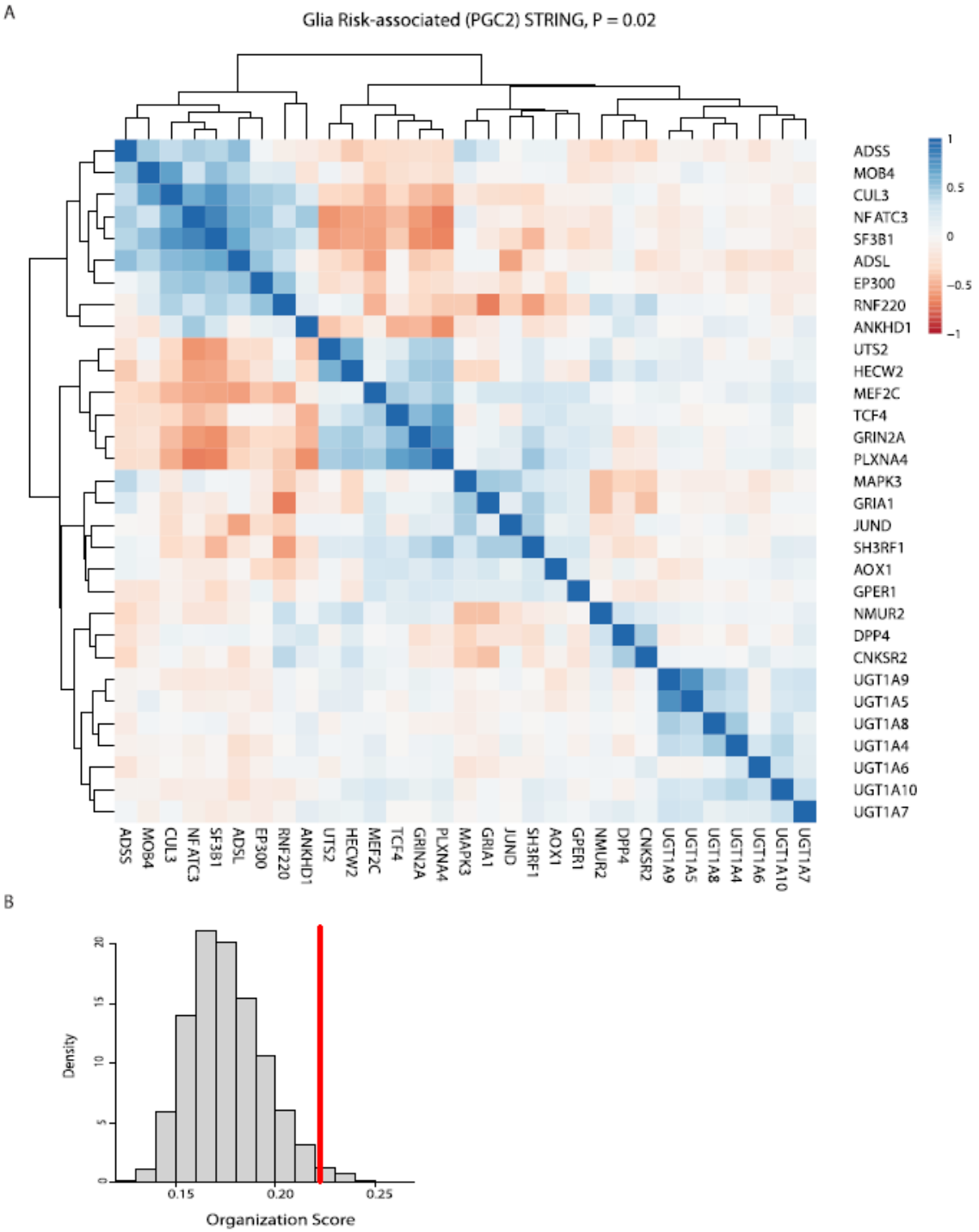

A
Glia Risk-associated (PGC2) STRING, P = 0.02
ADSS
MOB4
CUL3
NFATC3
SF3B1
ADSL
EP300
RNF220
ANKHD1
UTS2
HECW2
MEF2C
TCF4
GRIN2A
PLXNA4
MAPK3
GRIA1
JUND
SH3RF1
AOX1
GPER1
NMUR2
DPP4
CNKSR2
UGT1A9
UGT1A5
UGT1A8
UGT1A4
UGT1A6
UGT1A10
UGT1A7
1
0.5
0
−0.5
−1
B
Density
0
5
10
15
20
0.15
0.20
0.25
Organization Score

Figure 3.19 | Glia schizophrenia risk-associated STRING subset genes do not show greater transcriptional organization than full risk connectome gene list (PGC2).
(**A**) Pearson correlation heatmap of genes participating in glia-specific STRING protein-protein network. (**B**) Frequency distribution of permuted organization scores, randomly sampling from all genes in glia risk connectome. Red line = organization score for risk-associated Pearson correlation in panel A.

To examine whether such coregulation could be representative of the prefrontal cortex proteome of the adult brain, we screened a newly generated mass spectrometry–based dataset of 182 neuronal proteins, the majority of which were synaptic, quantified from prefrontal cortex of n = 23 adult schizophrenia and n = 20 control subjects (Table 3.22) (Hahn *et al.*, 2009). Among the 182 proteins, there were four from the risk-associated neuronal protein network (Figure 3.13, D): $GABA_B$ receptor subunit GABBR1 and ionotropic (GRIA1 and GRIN2A) and metabotropic glutamate receptor subunits (GRM3). Protein-protein correlation scores were significantly higher for these four risk-associated proteins than expected from random permutation analysis from the pool of 182 proteins ($P < 0.002$) across patients and controls. We conclude that the schizophrenia-related chromosomal connectome, tethering other portions of the genome to the sequences associated with schizophrenia heritability, provides a structural foundation for a functional connectome that reflects coordinated regulation of gene expression and interactions within the proteome.

Discussion

Neural progenitor differentiation into neurons and glia is associated with dynamic remodeling of chromosomal conformations, including loss of many NPC-specific chromosomal contacts, with differentiation-induced loop pruning primarily affecting a subset of genes important for neurogenesis (NPC-to-neuron loss) and neuronal function (NPC-to-glia loss). These findings broadly resonate with a recent report linking neural differentiation to multiple scales of 3DG folding, governed by multiple mechanisms, including CTCF-dependent loop alterations, repressive chromatin remodeling, and cell- and lineage-specific transcription factor networks (Bonev *et al.*, 2017). Our results suggest that developmental 3DG remodeling affects a substantial portion of sequences that confer liability for schizophrenia; furthermore, these genes in 3D physical proximity with schizophrenia-risk variants show a surprisingly strong correlation at the level of the transcriptome and proteome. How might the disease-relevant reorganization of the spatial genome (the “chromosomal connectome”) provide a structural foundation for coordinated regulation of expression? Recent Hi-C studies in mouse brain showed that chromosomal contacts preferentially occurred between loci targeted by the same transcription factors (Bonev *et al.*, 2017), and likewise, multiple schizophrenia risk loci could converge on intra- and interchromosomal hubs sharing a similar regulatory architecture including specific enhancers as well as transcription and splicing factors (Lomvardas *et al.*, 2006; Khanna, Hu and Belmont, 2014; Quinodoz *et al.*, 2018). Intriguingly, the three major

functional categories associated with the genetic risk architecture of schizophrenia—neuronal connectivity, synaptic signaling, and chromatin remodeling (Gilman *et al.*, 2012; O'dushlaine *et al.*, 2015)—were heavily represented within the cell type–specific chromosomal connectomes of neurons and NPCs described here (Fig. 3.10, B and E) and in whole tissue in vivo (Won *et al.*, 2016; de la Torre-Ubieta *et al.*, 2018). Cell type–specific 3DG reorganization during the course of neural progenitor differentiation, as shown here, could therefore have profound implications for our understanding of the genetic underpinnings of psychiatric disease. For example, inclusion of the cell type–specific risk (sequence) associated chromosomal connectome may lead to refinements of cumulative schizophrenia risk allele burden estimates, including "polygenic risk score" (PRS) or "biologically informed multilocus profile scores" (BIMPS), which currently only explain a small portion of disease risk (Bogdan, Baranger and Agrawal, 2018). Cell type–specific intersection of 3DG and genetic risk maps are of clinical interest beyond psychiatric disorders; for example, risk variants that confer susceptibility to autoimmune disease were embedded in physically interacting chromosomal loci in lymphoblastoid cells (Grubert *et al.*, 2015). Our 3DG maps from neural progenitors and their isogenic neurons and glia are accessible through the PsychENCODE Knowledge Portal (https://synapse.org) and more than double the number of currently available Hi-C datasets from human brain (Dixon *et al.*, 2015; Schmitt *et al.*, 2016; Won *et al.*, 2016), providing investigators with a resource to chart the expanded genome

space associated with cognitive and neuropsychiatric disease in context of cell type–specific remodeling of chromosomal conformations during early development.

Tables

Hyperlink to Excel format of all tables (ChapterIII_Tables)

Hi-C

LIBRARY	TOTAL READS	ALIGNED READS	CHIMERIC READS	Duplicates	FR PAIRS	Valid cis interactions	Subsampled contacts	Loop Calling	5kb (Non-Sub) Loops	10kb (Non-Sub) Loops	5+10kb (Non-Sub) Loops	5kb (Sub) Loops	10kb (Sub) Loops	5+10kb (Sub) Loops	Accession ID
Neuron 1	171,049,419	87%	21%	2%	26%	372,787,143	372,787,143	Reevaluation	2661	5348	8009	1704	5502	7206	
Neuron 1	171,049,419	87%	21%	2%	26%	372,787,143	372,787,143	Standard HiCCUPS					3676		
Neuron 2 (Rep1)	191,616,387	89%	22%	3%	26%										
Neuron 2 (Rep2)	277,560,748	88%	22%	3%	26%										
FB Neuron 1	1,148,145,724	90%	38%	5%	24%	491,798,369	372,787,143	Standard HiCCUPS					5741		
Glia 1	166,079,544	87%	20%	6%	27%	440,950,529	372,787,143	Reevaluation	7289	9153	16442	4531	8453	12984	
Glia 1	166,079,544	87%	20%	6%	27%	440,950,529	372,787,143	Standard HiCCUPS					7103		
Glia 2 (Rep1)	211,708,831	88%	21%	4%	27%										
Glia 2 (Rep2)	312,440,552	86%	21%	3%	27%										
NPC 1	239,221,938	87%	19%	4%	31%	616,826,359	372,787,143	Reevaluation	9070	10941	20011	4584	8429	13013	
NPC 1	239,221,938	87%	19%	4%	31%	616,826,359	372,787,143	Standard HiCCUPS					7378		
NPC 2 (Rep1)	205,102,780	88%	20%	3%	28%										
NPC 2 (Rep 2)	251,568,531	84%	20%	2%	28%										
No Ligase NPC 1	77,814,521	87%	0%	14%	41%										
GM12878	3,587,190,419	See Rao et al. (Ref. 15)				1,989,154,220	372,787,143	Reevaluation	10281	11268	21549	2769	5309	8078	GSE63525_GM12878_insitu_primary.hic
Fetal Germinal Zone	2,256,596,541	See Won et al. (2016) (Ref. 6)				545,496,617	204,476,188	Standard HiCCUPS		9309			4498		GZ(GSM2054567,GSM2054568,GSM2054569)
Fetal Cortical Plate	2,202,056,563	"				540,079,810	204,476,188	Standard HiCCUPS		9159			3921		CP (GSM2054564,GSM2054565,GSM2054566)
Mouse ESC	1,035,319,145	See Bonev et al. (2017) (Ref. 17)				471,503,539	471,503,539	Standard HiCCUPS		3496			3496		SRR5339748.,SRR5339753
Mouse NPC	1,284,813,522	"				619,932,857	471,503,539	Standard HiCCUPS		5896			4756		SRR5339786.,SRR5339792
Mouse Cortical Neuron	1,298,855,792	"				615,843,081	471,503,539	Standard HiCCUPS		4774			3807		SRR5339829.,SRR5339835

Hi-C

POOLED	
CELL TYPE	TOTAL READS
Neurons	640,226,554
Glia	690,228,927
NPCs	695,893,249

ATAC-seq

Sample	Total # Reads	% Uniquely Mapped	non chrM/Duplicates	% Duplicates
NPC 2	134,380,774	99.02	74,493,904	36.33
Neuron 2 Rep 1	96,205,900	99.07	54,130,694	33.5
Neuron 2 Rep 2	81,385,524	99.2	46,418,801	32.46

Table 3.1 | Hi-C and ATAC-Seq sequencing summary and quality controls.
Sequencing and quality control metrics for Hi-C libraries and loops called by cell type. Total reads, number of unfiltered reads; FR pairs, forward-reverse pairs;
*Loops are only called from contacts pooled by cell type
**Reevaluation= loops called from union set of initial loop calls (see Methods); Standard HiCCUPS= loop calls with no reevaluation
*** Non-Sub = all cis contacts passing filters; Sub= cis contacts randomly sampled from all cell types

GOID	GOTerm	Term PValue	Term PValue Corrected with Bonferroni step down	Neg log(P_corr)	% Associated	Nr. Genes	Total genes in Term	Associated Genes Found
GO:0007417	central nervous system development	78.0E-9	21.0E-6	4.677780705	5.247225025	52	991	[ARSB, BASP1, CDH11, CENPF, CEP120, CEP290, CLU, DNER, EN1, EPHA4, FUT10, GARS, GNG12, GRIK1, HOXA2, HOXB2, IGF2BP1, INHBA, IRS2, KDM4A, KIRREL3, LHX8, MACROD2, MAP2, MKKS, MYO1D, NAV2, NCOA3, NR2F2, NRG1, NRP2, PBX3, PDHX, PITX2, POU3F2, PTCH1, PTN, PTPRZ1, RCAN1, ROBO1, RTN4RL1, SALL1, SEMA3A, SLITRK5, SOX2, SOX4, SPOCK1, SUDS3, TOP2B, TRA2B, ZBTB16, ZEB1]
GO:0035108	limb morphogenesis	750.0E-9	200.0E-6	3.698970004	10.45751634	16	153	[ALX1, B9D1, BMPR1B, CYP26B1, EN1, FMN1, HOXD12, MAP3K20, PITX2, PTCH1, SALL1, SALL4, SOX4, SP9, TGFB2, ZBTB16]
GO:0072001	renal system development	1.0E-6	270.0E-6	3.568636236	7.590759076	23	303	[ADAMTS1, ADAMTS16, AGTR1, BASP1, CENPF, CEP290, COL4A4, DCN, EFNB2, EPHA4, FMN1, GARS, ITGA6, KIRREL3, NFIA, PTCH1, SALL1, SMAD2, SOX4, SULF1, SULF2, TGFB2, ZBTB16]
GO:0000904	cell morphogenesis involved in differentiation	1.5E-6	410.0E-6	3.387216143	5.52407932	39	706	[ADGRB3, ALCAM, ANK3, ART4, BMPR1B, BVES, CDH11, EPHA4, FBLN1, FEZ1, FLRT3, FMN1, GAP43, GFRA1, HEG1, HOXA2, IRS2, KLF7, LATS1, MAP1B, MAP2, NCOA3, NRG1, NRP2, PARD3, PIK3R1, PLXNB1, POU3F2, PTPRZ1, RCC2, RGMA, ROBO1, S100A10, SDC2, SEMA3A, SLITRK5, STRC, TGFB2, TOP2B]
GO:0061564	axon development	3.7E-6	970.0E-6	3.013228266	5.988023952	30	501	[ALCAM, ANK3, ART4, BMPR1B, CDH11, EPHA4, FEZ1, FLRT3, GAP43, GFRA1, HOXA2, IRS2, KLF7, MAP1B, NCAM2, NRG1, NRP2, PARD3, PCDHA4, PIK3R1, PLXNB1, POU3F2, PTPRZ1, RGMA, ROBO1, RTN4RL1, SEMA3A, SLITRK5, TGFB2, TOP2B]
GO:0001822	kidney development	4.9E-6	1.2E-3	2.920818754	7.368421053	21	285	[ADAMTS1, ADAMTS16, AGTR1, BASP1, CENPF, CEP290, COL4A4, DCN, EFNB2, EPHA4, FMN1, GARS, KIRREL3, PTCH1, SALL1, SMAD2, SOX4, SULF1, SULF2, TGFB2, ZBTB16]
GO:0048667	cell morphogenesis involved in neuron differentiation	7.8E-6	2.0E-3	2.698970004	5.604203152	32	571	[ADGRB3, ALCAM, ANK3, ART4, BMPR1B, CDH11, EPHA4, FEZ1, FLRT3, FMN1, GAP43, GFRA1, HOXA2, IRS2, KLF7, MAP1B, MAP2, NRG1, NRP2, PARD3, PIK3R1, PLXNB1, POU3F2, PTPRZ1, RGMA, ROBO1, SDC2, SEMA3A, SLITRK5, STRC, TGFB2, TOP2B]
GO:0030326	embryonic limb morphogenesis	15.0E-6	4.1E-3	2.387216143	9.848484848	13	132	[ALX1, B9D1, CYP26B1, EN1, HOXD12, MAP3K20, PITX2, PTCH1, SALL1, SALL4, SP9, TGFB2, ZBTB16]
GO:0007409	axonogenesis	17.0E-6	4.5E-3	2.346787486	5.844155844	27	462	[ALCAM, ANK3, ART4, BMPR1B, CDH11, EPHA4, FEZ1, FLRT3, GAP43, GFRA1, HOXA2, IRS2, KLF7, MAP1B, NRG1, NRP2, PARD3, PIK3R1, PLXNB1, POU3F2, PTPRZ1, RGMA, ROBO1, SEMA3A, SLITRK5, TGFB2, TOP2B]
GO:0048812	neuron projection morphogenesis	18.0E-6	4.7E-3	2.327902142	5.331179321	33	619	[ADGRB3, ALCAM, ANK3, ART4, BMPR1B, CDH11, EPHA4, FEZ1, FLRT3, FMN1, GAP43, GFRA1, HOXA2, IRS2, KIRREL3, KLF7, MAP1B, MAP2, NRG1, NRP2, PARD3, PIK3R1, PLXNB1, POU3F2, PTPRZ1, RGMA, RIMS2, ROBO1, SDC2, SEMA3A, SLITRK5, TGFB2, TOP2B]
GO:0008038	neuron recognition	27.0E-6	6.9E-3	2.161150909	18.42105263	7	38	[EPHA4, GAP43, NCAM2, OPCML, PCDHA4, ROBO1, SEMA3A]
GO:0030900	forebrain development	39.0E-6	10.0E-3	2	5.940594059	24	404	[CEP120, FUT10, GNG12, IGF2BP1, INHBA, KIRREL3, LHX8, MKKS, MYO1D, NCOA3, NR2F2, NRG1, NRP2, PITX2, POU3F2, PTN, ROBO1, RTN4RL1, SALL1, SEMA3A, SLITRK5, SOX2, TOP2B, TRA2B]
GO:0031053	primary miRNA processing	58.0E-6	14.0E-3	1.853871964	40	4	10	[GARS, HNRNPA2B1, NCBP1, SMAD2]
GO:0007507	heart development	80.0E-6	20.0E-3	1.698970004	5.263157895	29	551	[ADAMTS1, BASP1, BVES, CALCRL, CASP7, DAND5, EFNB2, FLRT3, GARS, HDAC9, HEG1, KCNK2, KDM6A, MKKS, NEBL, NRG1, NRP2, PITX2, PRICKLE1, PTCH1, PTN, ROBO1, SALL1, SALL4, SMAD2, SOX4, SUMO1, TGFB2, ZMIZ1]
GO:0005261	cation channel activity	94.0E-6	23.0E-3	1.638272164	6.211180124	20	322	[CACNB2, CLU, FAM155A, GRIK1, GRIN2B, HCN4, ITPR2, KCND2, KCNE2, KCNH1, KCNK16, KCNK2, KCNMB1, KCNS1, NALCN, SCN1A, SCN9A, TRPM3, TRPM4, TRPM8]
GO:0005272	sodium channel activity	120.0E-6	32.0E-3	1.494850022	14.58333333	7	48	[CLU, GRIK1, HCN4, NALCN, SCN1A, SCN9A, TRPM4]
GO:0005216	ion channel activity	150.0E-6	37.0E-3	1.431798276	5.517241379	24	435	[CACNB2, CLCN4, CLU, FAM155A, GABRP, GRIK1, GRIN2B, HCN4, ITPR2, KCND2, KCNE2, KCNH1, KCNK16, KCNK2, KCNMB1, KCNS1, NALCN, SCN1A, SCN9A, SLC26A2, SLC26A7, TRPM3, TRPM4, TRPM8]
GO:0048706	embryonic skeletal system development	190.0E-6	48.0E-3	1.318758763	8.8	11	125	[ACVR2A, ALX1, HOXA2, HOXA5, HOXB2, HOXB6, IRX5, SMAD2, SULF1, SULF2, ZEB1]

Table 3.2 | Gene ontology (GO) of Brain-specific loops.
Gene ontology enrichment terms and details for genes in brain-specific loops. GOID, ClueGO term ID; Term PValue, uncorrected p-value; Term PValue Corrected with Bonferroni step down, p-value corrected for multiple comparisons; Neg log(P_corr), negative log-transformed corrected p-values; % Associated, percentage of genes in GO category found in input list of genes; Nr. Genes, number of genes from input list that match the GO category; Total genes in Term, number of genes present in the term according to ClueGo.

GOID	GOTerm	Neg log (P)	Ontology Source	Term PValue	Term PValue Corrected with Bonferroni step down	negative log(P)	% Associated Genes	Nr. Genes	Total genes in term	Associated Genes Found
GO:0048566	embryonic digestive tract development	2.397940009	GO_BiologicalProcess-GOA_23.02.2017_10h01	140.0E-6	4.0E-3	2.4	11.43	4	35	[PCSK5, PDGFRA, SHH, SHOX2]
GO:0048663	neuron fate commitment	2.236572006	GO_BiologicalProcess-GOA_23.02.2017_10h01	220.0E-6	5.8E-3	2.2	7.04	5	71	[DMRTA2, GATA2, GBX1, SATB2, SHH]
GO:2000179	positive regulation of neural precursor cell proliferation	1.958607315	GO_BiologicalProcess-GOA_23.02.2017_10h01	470.0E-6	11.0E-3	2.0	8.51	4	47	[DMRTA2, EGF, INSM1, SHH]
GO:0090370	negative regulation of cholesterol efflux	1.853871964	GO_BiologicalProcess-GOA_23.02.2017_10h01	580.0E-6	14.0E-3	1.9	40.00	2	5	[EGF, SHH]
GO:0019054	modulation by virus of host process	1.677780705	GO_BiologicalProcess-GOA_23.02.2017_10h01	920.0E-6	21.0E-3	1.7	12.00	3	25	[ATG7, KPNA7, VAPA]
GO:0021902	commitment of neuronal cell to specific neuron type in forebrain	1.602059991	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.2E-3	25.0E-3	1.6	28.57	2	7	[GATA2, SATB2]
GO:0042054	histone methyltransferase activity	1.585026652	GO_MolecularFunction-GOA_23.02.2017_10h01	1.2E-3	26.0E-3	1.6	6.67	4	60	[PRMT8, SETD3, SMYD2, SMYD3]
GO:0003030	DNA replication	1.508638306	KEGG_01.03.2017	31.0E-3	31.0E-3	1.5	5.56	2	36	[MCM6, POLA2]
GO:0021940	positive regulation of cerebellar granule cell precursor proliferation	1.387216143	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.0E-3	41.0E-3	1.4	22.22	2	9	[EGF, SHH]
GO:0051895	negative regulation of focal adhesion assembly	1.387216143	GO_BiologicalProcess-GOA_23.02.2017_10h01	10.0E-3	41.0E-3	1.4	10.00	2	20	[BCAS3, RCC2]
GO:0060749	mammary gland alveolus development	1.387216143	GO_BiologicalProcess-GOA_23.02.2017_10h01	10.0E-3	41.0E-3	1.4	10.00	2	20	[EGF, ERBB4]
GO:0048747	muscle fiber development	1.337242168	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.3E-3	46.0E-3	1.3	5.36	3	56	[HDAC9, MYO18B, SHOX2]
GO:0045059	positive thymic T cell selection	1.318758763	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.5E-3	48.0E-3	1.3	20.00	2	10	[DOCK2, SHH]

Table 3.3 | Neuron-specific loops GO.
Gene ontology enrichment terms and details for genes in neuron-specific loops (See Table 3.2 for more information)

A

GOID	GOTerm	Neg log(P)	Ontology Source	Term PValue	Term PValue Corrected with Bonferroni step down	% Associated Genes	Nr. Genes	Total genes in Term
GO:0030182	neuron differentiation	18.39794	GO_BiologicalProcess-GOA_23.02.2017_10h01	260.0E-24	400.0E-21	15.75	210.00	1333.00
GO:0022008	neurogenesis	17.26706	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.5E-21	5.4E-18	14.85	232.00	1562.00
GO:0048699	generation of neurons	17.04095	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.0E-21	9.1E-18	15.10	221.00	1464.00
GO:0007417	central nervous system development	15.22914	GO_BiologicalProcess-GOA_23.02.2017_10h01	390.0E-21	590.0E-18	16.45	163.00	991.00
GO:0048812	neuron projection morphogenesis	13	GO_BiologicalProcess-GOA_23.02.2017_10h01	71.0E-18	100.0E-15	18.26	113.00	619.00
GO:0048666	neuron development	12.85372	GO_BiologicalProcess-GOA_23.02.2017_10h01	93.0E-18	140.0E-15	15.51	165.00	1064.00
GO:0031175	neuron projection development	12.53760	GO_BiologicalProcess-GOA_23.02.2017_10h01	190.0E-18	290.0E-15	16.11	146.00	906.00
GO:0097485	neuron projection guidance	11.95807	GO_BiologicalProcess-GOA_23.02.2017_10h01	730.0E-18	1.1E-12	24.90	61.00	245.00
GO:0048667	cell morphogenesis involved in neuron differentiation	11.22184	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.0E-15	6.0E-12	18.04	103.00	571.00
GO:0007420	brain development	11.05517	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.8E-15	8.8E-12	16.56	125.00	755.00
GO:0000904	cell morphogenesis involved in differentiation	10.22184	GO_BiologicalProcess-GOA_23.02.2017_10h01	39.0E-15	60.0E-12	16.57	117.00	706.00
GO:0007409	axonogenesis	9.721246	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-15	190.0E-12	18.61	86.00	462.00
GO:0061564	axon development	9.602060	GO_BiologicalProcess-GOA_23.02.2017_10h01	160.0E-15	250.0E-12	18.16	91.00	501.00
GO:0006357	regulation of transcription from RNA polymerase II promoter	9.075720	GO_BiologicalProcess-GOA_23.02.2017_10h01	550.0E-15	840.0E-12	12.46	249.00	1999.00
GO:0006366	transcription from RNA polymerase II promoter	9.026872	GO_BiologicalProcess-GOA_23.02.2017_10h01	620.0E-15	940.0E-12	12.20	269.00	2205.00
GO:0010628	positive regulation of gene expression	8.387216	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.7E-12	4.1E-9	12.53	230.00	1836.00
GO:0016055	Wnt signaling pathway	8.221848	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.0E-12	6.0E-9	17.36	88.00	507.00
GO:1905114	cell surface receptor signaling pathway involved in cell-cell signaling	8.086186	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.4E-12	8.2E-9	16.32	101.00	619.00
GO:0051254	positive regulation of RNA metabolic process	8.060480	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.8E-12	8.7E-9	12.93	199.00	1539.00
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	8	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.9E-12	10.0E-9	13.85	158.00	1141.00
GO:0045893	positive regulation of transcription, DNA-templated	7.920818	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.2E-12	12.0E-9	13.04	192.00	1472.00
GO:0021953	central nervous system neuron differentiation	7.619788	GO_BiologicalProcess-GOA_23.02.2017_10h01	16.0E-12	24.0E-9	23.71	46.00	194.00
GO:0045664	regulation of neuron differentiation	7.568636	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-12	27.0E-9	16.30	97.00	595.00
GO:0045935	positive regulation of nucleobase-containing compound metabolic process	7.508638	GO_BiologicalProcess-GOA_23.02.2017_10h01	20.0E-12	31.0E-9	12.36	224.00	1813.00
GO:0050767	regulation of neurogenesis	7.004365	GO_BiologicalProcess-GOA_23.02.2017_10h01	66.0E-12	99.0E-9	15.12	111.00	734.00
GO:0051960	regulation of nervous system development	6.920818	GO_BiologicalProcess-GOA_23.02.2017_10h01	84.0E-12	120.0E-9	14.65	121.00	826.00
GO:0031328	positive regulation of cellular biosynthetic process	6.568636	GO_BiologicalProcess-GOA_23.02.2017_10h01	180.0E-12	270.0E-9	12.02	229.00	1905.00
GO:0060070	canonical Wnt signaling pathway	6.221848	GO_BiologicalProcess-GOA_23.02.2017_10h01	400.0E-12	600.0E-9	19.22	59.00	307.00
GO:0010557	positive regulation of macromolecule biosynthetic process	6.091515	GO_BiologicalProcess-GOA_23.02.2017_10h01	540.0E-12	810.0E-9	12.11	212.00	1751.00
GO:0030900	forebrain development	5.958607	GO_BiologicalProcess-GOA_23.02.2017_10h01	760.0E-12	1.1E-6	17.33	70.00	404.00
GO:0060284	regulation of cell development	5.886056	GO_BiologicalProcess-GOA_23.02.2017_10h01	920.0E-12	1.3E-6	14.10	120.00	851.00
GO:0044212	transcription regulatory region DNA binding	5.795788	GO_MolecularFunction-GOA_23.02.2017_10h01	1.1E-9	1.6E-6	14.13	119.00	842.00
GO:0009887	animal organ morphogenesis	5.721246	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.2E-9	1.9E-6	13.38	140.00	1046.00
GO:0009887	animal organ morphogenesis	5.721246	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.2E-9	1.9E-6	13.38	140.00	1046.00
GO:0030111	regulation of Wnt signaling pathway	4.853872	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.6E-9	14.0E-6	17.66	59.00	334.00
GO:0000122	negative regulation of transcription from RNA polymerase II promoter	4.823987	GO_BiologicalProcess-GOA_23.02.2017_10h01	10.0E-9	15.0E-6	13.83	112.00	810.00
GO:0021954	central nervous system neuron development	4.431798	GO_BiologicalProcess-GOA_23.02.2017_10h01	25.0E-9	37.0E-6	29.49	23.00	78.00
GO:2000113	negative regulation of cellular macromolecule biosynthetic process	4.292430	GO_BiologicalProcess-GOA_23.02.2017_10h01	35.0E-9	51.0E-6	12.15	168.00	1383.00
GO:0010558	negative regulation of macromolecule biosynthetic process	4.167491	GO_BiologicalProcess-GOA_23.02.2017_10h01	46.0E-9	68.0E-6	11.98	176.00	1469.00
GO:0000976	transcription regulatory region sequence-specific DNA binding	4.148741	GO_MolecularFunction-GOA_23.02.2017_10h01	48.0E-9	71.0E-6	14.18	96.00	677.00
GO:1990837	sequence-specific double-stranded DNA binding	4.119186	GO_MolecularFunction-GOA_23.02.2017_10h01	51.0E-9	76.0E-6	13.95	100.00	717.00
GO:0090287	regulation of cellular response to growth factor stimulus	3.958607	GO_BiologicalProcess-GOA_23.02.2017_10h01	80.0E-9	110.0E-6	18.36	47.00	256.00
GO:0045596	negative regulation of cell differentiation	3.886056	GO_BiologicalProcess-GOA_23.02.2017_10h01	88.0E-9	130.0E-6	14.09	94.00	667.00
GO:0021955	central nervous system neuron axonogenesis	3.795788	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-9	160.0E-6	44.83	13.00	29.00

GO:0045892	negative regulation of transcription, DNA-templated	3.7695511	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-9	170.0E-6	12.39	145.00	1170.00
GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	3.7447275	GO_MolecularFunction-GOA_23.02.2017_10h01	120.0E-9	180.0E-6	14.41	85.00	590.00
GO:0004360	Axon guidance	3.7212464	KEGG_01.03.2017	130.0E-9	190.0E-6	20.57	36.00	175.00
GO:0035108	limb morphogenesis	3.7212464	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	190.0E-6	21.57	33.00	153.00
GO:0010721	negative regulation of cell development	3.69897	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	200.0E-6	17.33	52.00	300.00
GO:1903507	negative regulation of nucleic acid-templated transcription	3.69897	GO_BiologicalProcess-GOA_23.02.2017_10h01	140.0E-9	200.0E-6	12.27	149.00	1214.00
GO:0051253	negative regulation of RNA metabolic process	3.6197888	GO_BiologicalProcess-GOA_23.02.2017_10h01	160.0E-9	240.0E-6	12.10	155.00	1281.00
GO:0051961	negative regulation of nervous system development	3.552842	GO_BiologicalProcess-GOA_23.02.2017_10h01	190.0E-9	280.0E-6	17.50	49.00	280.00
GO:0031327	negative regulation of cellular biosynthetic process	3.3665315	GO_BiologicalProcess-GOA_23.02.2017_10h01	290.0E-9	430.0E-6	11.60	180.00	1552.00
GO:0045597	positive regulation of cell differentiation	3.2839967	GO_BiologicalProcess-GOA_23.02.2017_10h01	350.0E-9	520.0E-6	12.84	117.00	911.00
GO:0060828	regulation of canonical Wnt signaling pathway	3.1191864	GO_BiologicalProcess-GOA_23.02.2017_10h01	520.0E-9	760.0E-6	17.72	45.00	254.00
GO:0033267	axon part	2.9586073	GO_CellularComponent-GOA_23.02.2017_10h01	780.0E-9	1.1E-3	18.72	38.00	203.00
GO:0010975	regulation of neuron projection development	2.9586073	GO_BiologicalProcess-GOA_23.02.2017_10h01	800.0E-9	1.1E-3	14.97	66.00	441.00
GO:0045934	negative regulation of nucleobase-containing compound metabolic process	2.8860566	GO_BiologicalProcess-GOA_23.02.2017_10h01	950.0E-9	1.3E-3	11.59	167.00	1441.00
GO:0001764	neuron migration	2.853872	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.0E-6	1.4E-3	20.53	31.00	151.00
GO:0030178	negative regulation of Wnt signaling pathway	2.8239087	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.0E-6	1.5E-3	18.57	39.00	210.00
GO:0010629	negative regulation of gene expression	2.8239087	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.0E-6	1.5E-3	11.23	194.00	1727.00
GO:0044304	main axon	2.7212464	GO_CellularComponent-GOA_23.02.2017_10h01	1.3E-6	1.9E-3	27.54	19.00	69.00
GO:0061035	regulation of cartilage development	2.6197888	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.6E-6	2.4E-3	27.14	19.00	70.00
GO:0048588	developmental cell growth	2.6197888	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.6E-6	2.4E-3	18.45	38.00	206.00
GO:2000112	regulation of cellular macromolecule biosynthetic process	2.6197888	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.6E-6	2.4E-3	9.87	411.00	4166.00
GO:0015630	microtubule cytoskeleton	2.5086383	GO_CellularComponent-GOA_23.02.2017_10h01	2.1E-6	3.1E-3	11.94	136.00	1139.00
GO:0044428	nuclear part	2.4814861	GO_CellularComponent-GOA_23.02.2017_10h01	2.3E-6	3.3E-3	9.77	431.00	4411.00
GO:0045665	negative regulation of neuron differentiation	2.4089354	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.7E-6	3.9E-3	18.23	37.00	203.00
GO:0000987	core promoter proximal region sequence-specific DNA binding	2.3665315	GO_MolecularFunction-GOA_23.02.2017_10h01	2.9E-6	4.3E-3	15.20	57.00	375.00
GO:0051252	regulation of RNA metabolic process	2.3565473	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.0E-6	4.4E-3	9.89	389.00	3933.00
GO:0090090	negative regulation of canonical Wnt signaling pathway	2.2757241	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.7E-6	5.3E-3	18.86	33.00	175.00
GO:0031981	nuclear lumen	2.251812	GO_CellularComponent-GOA_23.02.2017_10h01	3.9E-6	5.6E-3	9.84	398.00	4045.00
GO:0000978	RNA polymerase II core promoter proximal region sequence-specific DNA binding	2.1191864	GO_MolecularFunction-GOA_23.02.2017_10h01	5.3E-6	7.6E-3	15.15	55.00	363.00
GO:0030901	midbrain development	2.1079054	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.4E-6	7.8E-3	21.82	24.00	110.00
GO:0006355	regulation of transcription, DNA-templated	2.09691	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.5E-6	8.0E-3	9.89	372.00	3763.00
GO:0016070	RNA metabolic process	2.0555173	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.0E-6	8.8E-3	9.61	462.00	4809.00
GO:0016070	RNA metabolic process	2.0555173	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.0E-6	8.8E-3	9.61	462.00	4809.00
GO:1990138	neuron projection extension	2.0222764	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.5E-6	9.5E-3	19.61	30.00	153.00
GO:1903506	regulation of nucleic acid-templated transcription	2.0043648	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.9E-6	9.9E-3	9.87	374.00	3791.00
GO:0032330	regulation of chondrocyte differentiation	2	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.9E-6	10.0E-3	29.41	15.00	51.00
GO:0051962	positive regulation of nervous system development	1.9208188	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.5E-6	12.0E-3	13.80	69.00	500.00
GO:0005829	cytosol	1.8860566	GO_CellularComponent-GOA_23.02.2017_10h01	9.0E-6	13.0E-3	9.53	475.00	4982.00
GO:0030323	respiratory tube development	1.8860566	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.1E-6	13.0E-3	17.99	34.00	189.00
GO:0005634	nucleus	1.853872	GO_CellularComponent-GOA_23.02.2017_10h01	10.0E-6	14.0E-3	9.16	665.00	7258.00
GO:0005634	nucleus	1.853872	GO_CellularComponent-GOA_23.02.2017_10h01	10.0E-6	14.0E-3	9.16	665.00	7258.00
GO:0006351	transcription, DNA-templated	1.7695511	GO_BiologicalProcess-GOA_23.02.2017_10h01	12.0E-6	17.0E-3	9.77	384.00	3930.00
GO:0021872	forebrain generation of neurons	1.7212464	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-6	19.0E-3	24.66	18.00	73.00
GO:0070013	intracellular organelle lumen	1.69897	GO_CellularComponent-GOA_23.02.2017_10h01	14.0E-6	20.0E-3	9.48	485.00	5118.00
GO:0048568	embryonic organ development	1.6777807	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-6	21.0E-3	13.97	63.00	451.00
GO:0021537	telencephalon development	1.6777807	GO_BiologicalProcess-GOA_23.02.2017_10h01	15.0E-6	21.0E-3	16.09	42.00	261.00
GO:0045666	positive regulation of neuron differentiation	1.6777807	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-6	21.0E-3	14.90	52.00	349.00
GO:0005856	cytoskeleton	1.6382722	GO_CellularComponent-GOA_23.02.2017_10h01	16.0E-6	23.0E-3	10.51	224.00	2131.00
GO:0051216	cartilage development	1.6382722	GO_BiologicalProcess-GOA_23.02.2017_10h01	16.0E-6	23.0E-3	17.35	34.00	196.00
GO:0043009	chordate embryonic development	1.60206	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-6	25.0E-3	13.06	79.00	605.00

GO:0061036	positive regulation of cartilage development	1.4948 5	GO_BiologicalProcess-GOA_23.02.2017_10h01	23.0E-6	32.0E-3	34.38	11.00	32.00		
GO:0007423	sensory organ development	1.4814 861	GO_BiologicalProcess-GOA_23.02.2017_10h01	23.0E-6	33.0E-3	13.20	73.00	553.00		
GO:0050769	positive regulation of neurogenesis	1.4814 861	GO_BiologicalProcess-GOA_23.02.2017_10h01	23.0E-6	33.0E-3	13.96	61.00	437.00		
GO:0010769	regulation of cell morphogenesis involved in differentiation	1.4559 32	GO_BiologicalProcess-GOA_23.02.2017_10h01	24.0E-6	35.0E-3	15.69	43.00	274.00		
GO:0030334	regulation of cell migration	1.3979 4	GO_BiologicalProcess-GOA_23.02.2017_10h01	28.0E-6	40.0E-3	12.45	91.00	731.00		
GO:2000027	regulation of organ morphogenesis	1.3979 4	GO_BiologicalProcess-GOA_23.02.2017_10h01	28.0E-6	40.0E-3	15.52	43.00	277.00		
GO:2000027	regulation of organ morphogenesis	1.3979 4	GO_BiologicalProcess-GOA_23.02.2017_10h01	28.0E-6	40.0E-3	15.52	43.00	277.00		
GO:0004310	Wnt signaling pathway	1.3665 315	KEGG_01.03.2017	30.0E-6	43.0E-3	18.88	27.00	143.00		
GO:0032774	RNA biosynthetic process	1.3279 021	GO_BiologicalProcess-GOA_23.02.2017_10h01	33.0E-6	47.0E-3	9.64	393.00	4077.00		

B

GOID	GOTerm	Ontology Source	Term PValue	Term PValue Corrected with Benjamini-Hochberg	negative log(FDR)	Group PValue	Group PValue Corrected with Benjamini-Hochberg	GOLevels	GOGroups	% Associated Genes
GO:1902589	single-organism organelle organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	26.0E-12	12.0E-9	7.9E+0	1.7E-9	12.0E-9	[3, 4]	Group11	4.57
GO:0048468	cell development	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-12	21.0E-9	7.7E+0	57.0E-12	790.0E-12	[3, 4, 5]	Group13	4.24
GO:0000904	cell morphogenesis involved in differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-12	27.0E-9	7.6E+0	2.5E-9	12.0E-9	[4, 5, 6]	Group10	6.23
GO:0000904	cell morphogenesis involved in differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-12	27.0E-9	7.6E+0	57.0E-12	790.0E-12	[4, 5, 6]	Group13	6.23
GO:0007409	axonogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-9	1.2E-6	5.9E+0	2.5E-9	12.0E-9	[6, 7, 8, 9, 10, 11, 12]	Group10	6.71
GO:0048699	generation of neurons	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-9	1.2E-6	5.9E+0	57.0E-12	790.0E-12	[6, 7]	Group13	4.37
GO:0007409	axonogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-9	1.2E-6	5.9E+0	57.0E-12	790.0E-12	[6, 7, 8, 9, 10, 11, 12]	Group13	6.71
GO:0050770	regulation of axonogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-9	1.3E-6	5.9E+0	57.0E-12	790.0E-12	[6, 7, 8, 9, 10, 11, 12, 13]	Group13	10.98
GO:0032989	cellular component morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-9	1.4E-6	5.9E+0	57.0E-12	790.0E-12	[3, 4]	Group13	4.83
GO:0045595	regulation of cell differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	30.0E-9	1.4E-6	5.9E+0	57.0E-12	790.0E-12	[3, 4, 5]	Group13	4.20
GO:0030182	neuron differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	23.0E-9	1.4E-6	5.9E+0	57.0E-12	790.0E-12	[5, 7, 8]	Group13	4.50
GO:0000902	cell morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	29.0E-9	1.5E-6	5.8E+0	57.0E-12	790.0E-12	[4, 5]	Group13	4.97
GO:2000026	regulation of multicellular organismal development	GO_BiologicalProcess-GOA_23.02.2017_10h01	58.0E-9	1.9E-6	5.7E+0	57.0E-12	790.0E-12	[3, 4]	Group13	4.02
GO:0022008	neurogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	52.0E-9	1.9E-6	5.7E+0	57.0E-12	790.0E-12	[5, 6]	Group13	4.23
GO:0048667	cell morphogenesis involved in neuron differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	50.0E-9	2.0E-6	5.7E+0	2.5E-9	12.0E-9	[5, 6, 7, 8, 9, 10]	Group10	5.95
GO:0030030	cell projection organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	47.0E-9	2.0E-6	5.7E+0	57.0E-12	790.0E-12	[3]	Group13	4.32
GO:0048667	cell morphogenesis involved in neuron differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	50.0E-9	2.0E-6	5.7E+0	57.0E-12	790.0E-12	[5, 6, 7, 8, 9, 10]	Group13	5.95
GO:0048812	neuron projection morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-9	3.3E-6	5.5E+0	2.5E-9	12.0E-9	[5, 6, 7, 8, 10, 11]	Group10	5.65
GO:0050767	regulation of neurogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	100.0E-9	3.3E-6	5.5E+0	57.0E-12	790.0E-12	[5, 6, 7, 8]	Group13	5.31
GO:0048812	neuron projection morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-9	3.3E-6	5.5E+0	57.0E-12	790.0E-12	[5, 6, 7, 8, 10, 11]	Group13	5.65
GO:0032990	cell part morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	3.4E-6	5.5E+0	2.5E-9	12.0E-9	[4, 5]	Group10	5.51
GO:0032990	cell part morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	3.4E-6	5.5E+0	57.0E-12	790.0E-12	[4, 5]	Group13	5.51
GO:0045664	regulation of neuron differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	3.5E-6	5.5E+0	57.0E-12	790.0E-12	[6, 7, 8, 9]	Group13	5.71
GO:0051960	regulation of nervous system development	GO_BiologicalProcess-GOA_23.02.2017_10h01	120.0E-9	3.6E-6	5.4E+0	57.0E-12	790.0E-12	[4, 5, 6]	Group13	5.08
GO:0007010	cytoskeleton organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	170.0E-9	4.0E-6	5.4E+0	1.7E-9	12.0E-9	[4]	Group11	4.42
GO:0048666	neuron development	GO_BiologicalProcess-GOA_23.02.2017_10h01	210.0E-9	5.0E-6	5.3E+0	57.0E-12	790.0E-12	[4, 5, 6, 8, 9]	Group13	4.61
GO:0060284	regulation of cell development	GO_BiologicalProcess-GOA_23.02.2017_10h01	260.0E-9	5.6E-6	5.3E+0	57.0E-12	790.0E-12	[4, 5, 6]	Group13	4.94
GO:0010769	regulation of cell morphogenesis involved in differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	360.0E-9	7.5E-6	5.1E+0	57.0E-12	790.0E-12	[5, 6, 7]	Group13	7.66
GO:0090132	epithelium migration	GO_BiologicalProcess-GOA_23.02.2017_10h01	440.0E-9	8.9E-6	5.1E+0	750.0E-9	2.1E-6	[4]	Group12	7.87
GO:0031175	neuron projection development	GO_BiologicalProcess-GOA_23.02.2017_10h01	570.0E-9	11.0E-6	5.0E+0	2.8E-9	9.8E-9	[4, 5, 6, 7, 9, 10]	Group09	4.75
GO:0031175	neuron projection development	GO_BiologicalProcess-GOA_23.02.2017_10h01	570.0E-9	11.0E-6	5.0E+0	57.0E-12	790.0E-12	[4, 5, 6, 7, 9, 10]	Group13	4.75
GO:0010975	regulation of neuron projection development	GO_BiologicalProcess-GOA_23.02.2017_10h01	730.0E-9	13.0E-6	4.9E+0	57.0E-12	790.0E-12	[5, 6, 7, 8, 9, 10, 11]	Group13	6.12
GO:0040012	regulation of locomotion	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.0E-6	17.0E-6	4.8E+0	750.0E-9	2.1E-6	[2, 3]	Group12	4.77
GO:0030036	actin cytoskeleton organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.2E-6	21.0E-6	4.7E+0	1.7E-9	12.0E-9	[4, 5]	Group11	5.27
GO:0022604	regulation of cell morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.4E-6	23.0E-6	4.6E+0	57.0E-12	790.0E-12	[4, 5, 6]	Group13	5.91

GO:0016477	cell migration	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.5E-6	24.0E-6	4.6E+0	2.5E-9	12.0E-9	[3, 4, 5]	Group10	4.05
GO:0016477	cell migration	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.5E-6	24.0E-6	4.6E+0	750.0E-9	2.1E-6	[3, 4, 5]	Group12	4.05
GO:1902903	regulation of supramolecular fiber organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.5E-6	55.0E-6	4.3E+0	1.7E-9	12.0E-9	[4, 5]	Group11	6.65
GO:0022603	regulation of anatomical structure morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.8E-6	57.0E-6	4.2E+0	57.0E-12	790.0E-12	[3, 4]	Group13	4.38
GO:0004360	Axon guidance	KEGG_01.03.2017	4.3E-6	61.0E-6	4.2E+0	2.5E-9	12.0E-9	[-1]	Group10	8.57
GO:0030029	actin filament-based process	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.2E-6	61.0E-6	4.2E+0	1.7E-9	12.0E-9	[3]	Group11	4.84
GO:0004360	Axon guidance	KEGG_01.03.2017	4.3E-6	61.0E-6	4.2E+0	57.0E-12	790.0E-12	[-1]	Group13	8.57
GO:2000145	regulation of cell motility	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.7E-6	64.0E-6	4.2E+0	750.0E-9	2.1E-6	[3, 4, 5]	Group12	4.71
GO:0005205	Proteoglycans in cancer	KEGG_01.03.2017	6.2E-6	83.0E-6	4.1E+0	6.2E-6	14.0E-6	[-1]	Group03	7.88
GO:0051270	regulation of cellular component movement	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.8E-6	88.0E-6	4.1E+0	750.0E-9	2.1E-6	[3, 4]	Group12	4.55
GO:0008092	cytoskeletal protein binding	GO_MolecularFunction-GOA_23.02.2017_10h01	7.0E-6	89.0E-6	4.1E+0	1.7E-9	12.0E-9	[3]	Group11	4.45
GO:0043005	neuron projection	GO_CellularComponent-GOA_23.02.2017_10h01	10.0E-6	120.0E-6	3.9E+0	2.8E-9	9.8E-9	[3, 4]	Group09	4.22
GO:0010632	regulation of epithelial cell migration	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.8E-6	120.0E-6	3.9E+0	750.0E-9	2.1E-6	[3, 4, 5, 6, 7, 8]	Group12	8.02
GO:0043005	neuron projection	GO_CellularComponent-GOA_23.02.2017_10h01	10.0E-6	120.0E-6	3.9E+0	57.0E-12	790.0E-12	[3, 4]	Group13	4.22
GO:0003779	actin binding	GO_MolecularFunction-GOA_23.02.2017_10h01	11.0E-6	130.0E-6	3.9E+0	1.7E-9	12.0E-9	[4]	Group11	5.65
GO:0031344	regulation of cell projection organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	12.0E-6	140.0E-6	3.9E+0	57.0E-12	790.0E-12	[4, 5]	Group13	5.02
GO:0045596	negative regulation of cell differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-6	140.0E-6	3.9E+0	57.0E-12	790.0E-12	[3, 4, 5, 6]	Group13	4.80
GO:0002418	immune response to tumor cell	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-6	150.0E-6	3.8E+0	14.0E-6	28.0E-6	[3, 4]	Group06	31.25
GO:0097435	supramolecular fiber organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-6	150.0E-6	3.8E+0	1.7E-9	12.0E-9	[3]	Group11	4.97
GO:0044463	cell projection part	GO_CellularComponent-GOA_23.02.2017_10h01	16.0E-6	160.0E-6	3.8E+0	2.8E-9	9.8E-9	[2, 3, 4]	Group09	4.11
GO:0030334	regulation of cell migration	GO_BiologicalProcess-GOA_23.02.2017_10h01	15.0E-6	160.0E-6	3.8E+0	750.0E-9	2.1E-6	[4, 5, 6]	Group12	4.65
GO:0044463	cell projection part	GO_CellularComponent-GOA_23.02.2017_10h01	16.0E-6	160.0E-6	3.8E+0	57.0E-12	790.0E-12	[2, 3, 4]	Group13	4.11
GO:0051093	negative regulation of developmental process	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-6	180.0E-6	3.7E+0	57.0E-12	790.0E-12	[2, 3, 4]	Group13	4.36
GO:0005874	microtubule	GO_CellularComponent-GOA_23.02.2017_10h01	20.0E-6	200.0E-6	3.7E+0	1.7E-9	12.0E-9	[5, 6, 7, 8, 9, 10]	Group11	5.60
GO:0001052	Role of Abl in Robo-Slit signaling	REACTOME_Pathways_01.03.2017	21.0E-6	210.0E-6	3.7E+0	21.0E-6	38.0E-6	[-1]	Group04	44.44
GO:0070886	positive regulation of calcineurin-NFAT signaling cascade	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-6	210.0E-6	3.7E+0	35.0E-6	55.0E-6	[6, 7, 8, 9, 10, 11]	Group07	44.44
GO:0051961	negative regulation of nervous system development	GO_BiologicalProcess-GOA_23.02.2017_10h01	27.0E-6	260.0E-6	3.6E+0	57.0E-12	790.0E-12	[3, 4, 5, 6, 7]	Group13	6.43
GO:0001525	angiogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	32.0E-6	290.0E-6	3.5E+0	750.0E-9	2.1E-6	[3, 4, 5, 7, 8, 9, 10]	Group12	5.31
GO:0001667	ameboidal-type cell migration	GO_BiologicalProcess-GOA_23.02.2017_10h01	38.0E-6	350.0E-6	3.5E+0	750.0E-9	2.1E-6	[4, 5, 6]	Group12	5.85
GO:0048646	anatomical structure formation involved in morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	46.0E-6	400.0E-6	3.4E+0	750.0E-9	2.1E-6	[2, 3, 4]	Group12	4.05
GO:0051493	regulation of cytoskeleton organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	50.0E-6	440.0E-6	3.4E+0	1.7E-9	12.0E-9	[5, 6]	Group11	5.16
GO:1903729	regulation of plasma membrane organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	59.0E-6	490.0E-6	3.3E+0	59.0E-6	83.0E-6	[4, 5, 6]	Group02	10.84
GO:0030424	axon	GO_CellularComponent-GOA_23.02.2017_10h01	57.0E-6	490.0E-6	3.3E+0	2.8E-9	9.8E-9	[4, 5]	Group09	5.24
GO:0030424	axon	GO_CellularComponent-GOA_23.02.2017_10h01	57.0E-6	490.0E-6	3.3E+0	57.0E-12	790.0E-12	[4, 5]	Group13	5.24
GO:0007026	negative regulation of microtubule depolymerization	GO_BiologicalProcess-GOA_23.02.2017_10h01	60.0E-6	500.0E-6	3.3E+0	60.0E-6	77.0E-6	[5, 6, 7, 8, 9, 10]	Group00	23.81
GO:0010721	negative regulation of cell development	GO_BiologicalProcess-GOA_23.02.2017_10h01	68.0E-6	550.0E-6	3.3E+0	57.0E-12	790.0E-12	[4, 5, 6, 7]	Group13	6.00
GO:0060548	negative regulation of cell death	GO_BiologicalProcess-GOA_23.02.2017_10h01	78.0E-6	620.0E-6	3.2E+0	70.0E-6	82.0E-6	[3, 4, 5]	Group08	4.07
GO:0031589	cell-substrate adhesion	GO_BiologicalProcess-GOA_23.02.2017_10h01	87.0E-6	660.0E-6	3.2E+0	35.0E-6	55.0E-6	[3]	Group07	5.69
GO:1901215	negative regulation of neuron death	GO_BiologicalProcess-GOA_23.02.2017_10h01	86.0E-6	670.0E-6	3.2E+0	70.0E-6	82.0E-6	[4, 5, 6]	Group08	7.00
GO:0048514	blood vessel morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	85.0E-6	670.0E-6	3.2E+0	750.0E-9	2.1E-6	[3, 4, 6, 7, 8, 9]	Group12	4.82
GO:0070997	neuron death	GO_BiologicalProcess-GOA_23.02.2017_10h01	94.0E-6	700.0E-6	3.2E+0	70.0E-6	82.0E-6	[4]	Group08	5.65
GO:0001952	regulation of cell-matrix adhesion	GO_BiologicalProcess-GOA_23.02.2017_10h01	98.0E-6	720.0E-6	3.1E+0	35.0E-6	55.0E-6	[4, 5]	Group07	9.17
GO:0036477	somatodendritic compartment	GO_CellularComponent-GOA_23.02.2017_10h01	100.0E-6	740.0E-6	3.1E+0	2.8E-9	9.8E-9	[3, 4]	Group09	4.35
GO:0015631	tubulin binding	GO_MolecularFunction-GOA_23.02.2017_10h01	100.0E-6	760.0E-6	3.1E+0	1.7E-9	12.0E-9	[4]	Group11	5.79
GO:0006935	chemotaxis	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-6	780.0E-6	3.1E+0	2.5E-9	12.0E-9	[3, 4]	Group10	4.64
GO:0006935	chemotaxis	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-6	780.0E-6	3.1E+0	750.0E-9	2.1E-6	[3, 4]	Group12	4.64
GO:0006935	chemotaxis	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-6	780.0E-6	3.1E+0	57.0E-12	790.0E-12	[3, 4]	Group13	4.64
GO:0050919	negative chemotaxis	GO_BiologicalProcess-GOA_23.02.2017_10h01	120.0E-6	830.0E-6	3.1E+0	120.0E-6	120.0E-6	[4, 5]	Group05	20.83
GO:0008017	microtubule binding	GO_MolecularFunction-GOA_23.02.2017_10h01	120.0E-6	850.0E-6	3.1E+0	1.7E-9	12.0E-9	[5]	Group11	6.44
GO:0007548	sex differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-6	920.0E-6	3.0E+0	130.0E-6	130.0E-6	[3, 4]	Group01	5.88

GO:0002040	sprouting angiogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	140.0E-6	940.0E-6	3.0E+0	750.0E-9	2.1E-6	[4, 5, 6, 8, 9, 10, 11]	Group12	10.96	

C

GOID	GOTerm	Ontology Source	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Neg log (FDR)	Group PValue	Group PValue Corrected with Benjamini-Hochberg	GOLevels	GOGroups	% Associated Genes	Nr. Genes
GO:0007399	nervous system development	GO_BiologicalProcess-GOA_23.02.2017_10h01	78.0E-9	45.0E-6	4.3E+0	13.0E-9	880.0E-9	[4, 5]	Group69	4.61	109.00
GO:0009653	anatomical structure morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	200.0E-9	58.0E-6	4.2E+0	200.0E-9	4.4E-6	[2, 3]	Group39	4.47	114.00
GO:0098590	plasma membrane region	GO_CellularComponent-GOA_23.02.2017_10h01	900.0E-9	170.0E-6	3.8E+0	1.4E-6	18.0E-6	[2, 3, 4, 5, 6]	Group52	5.50	57.00
GO:0006928	movement of cell or subcellular component	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.6E-6	230.0E-6	3.6E+0	2.9E-6	32.0E-6	[3]	Group62	4.56	93.00
GO:0000053	Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell	REACTOME_Pathways_01.03.2017	3.8E-6	440.0E-6	3.4E+0	6.1E-6	58.0E-6	[-1]	Group63	7.62	26.00
GO:0000054	Transmission across Chemical Synapses	REACTOME_Pathways_01.03.2017	3.8E-6	440.0E-6	3.4E+0	6.1E-6	58.0E-6	[-1]	Group63	7.62	26.00
GO:0000055	Neuronal System	REACTOME_Pathways_01.03.2017	3.8E-6	440.0E-6	3.4E+0	6.1E-6	58.0E-6	[-1]	Group63	7.62	26.00
GO:0001844	GABA A (rho) receptor activation	REACTOME_Pathways_01.03.2017	3.8E-6	440.0E-6	3.4E+0	6.1E-6	58.0E-6	[-1]	Group63	7.62	26.00
GO:0001845	GABA receptor activation	REACTOME_Pathways_01.03.2017	3.8E-6	440.0E-6	3.4E+0	6.1E-6	58.0E-6	[-1]	Group63	7.62	26.00
GO:0048699	generation of neurons	GO_BiologicalProcess-GOA_23.02.2017_10h01	15.0E-6	1.1E-3	3.0E+0	13.0E-9	880.0E-9	[6, 7]	Group69	4.71	69.00
GO:0035239	tube morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	20.0E-6	1.2E-3	2.9E+0	260.0E-6	1.2E-3	[3, 4, 5]	Group51	6.93	26.00
GO:0042995	cell projection	GO_CellularComponent-GOA_23.02.2017_10h01	14.0E-6	1.2E-3	2.9E+0	32.0E-6	260.0E-6	[2, 3]	Group58	4.41	88.00
GO:0048468	cell development	GO_BiologicalProcess-GOA_23.02.2017_10h01	12.0E-6	1.2E-3	2.9E+0	13.0E-9	880.0E-9	[3, 4, 5]	Group69	4.39	89.00
GO:0030182	neuron differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	20.0E-6	1.2E-3	2.9E+0	13.0E-9	880.0E-9	[5, 7, 8]	Group69	4.80	64.00
GO:0022008	neurogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	26.0E-6	1.3E-3	2.9E+0	13.0E-9	880.0E-9	[5, 6]	Group69	4.61	72.00
GO:0035295	tube development	GO_BiologicalProcess-GOA_23.02.2017_10h01	55.0E-6	2.1E-3	2.7E+0	260.0E-6	1.2E-3	[3, 4]	Group51	5.75	36.00
GO:0048870	cell motility	GO_BiologicalProcess-GOA_23.02.2017_10h01	53.0E-6	2.2E-3	2.7E+0	2.9E-6	32.0E-6	[2, 3, 4]	Group62	4.52	70.00
GO:1902495	transmembrane transporter complex	GO_CellularComponent-GOA_23.02.2017_10h01	51.0E-6	2.3E-3	2.6E+0	1.1E-3	3.5E-3	[2, 3, 4, 5]	Group68	7.01	23.00
GO:0098794	postsynapse	GO_CellularComponent-GOA_23.02.2017_10h01	50.0E-6	2.4E-3	2.6E+0	1.4E-6	18.0E-6	[2, 3]	Group52	6.36	28.00
GO:0002521	leukocyte differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	71.0E-6	2.5E-3	2.6E+0	1.2E-3	3.7E-3	[5, 6, 7, 8]	Group56	6.07	30.00
GO:0030224	monocyte differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	74.0E-6	2.5E-3	2.6E+0	650.0E-6	2.3E-3	[7, 8, 9, 10]	Group67	18.42	7.00
GO:0043005	neuron projection	GO_CellularComponent-GOA_23.02.2017_10h01	93.0E-6	3.0E-3	2.5E+0	32.0E-6	260.0E-6	[3, 4]	Group58	4.89	51.00
GO:0030030	cell projection organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-6	3.2E-3	2.5E+0	13.0E-9	880.0E-9	[3]	Group69	4.53	65.00
GO:0005033	Nicotine addiction	KEGG_01.03.2017	100.0E-6	3.2E-3	2.5E+0	150.0E-9	5.0E-6	[-1]	Group70	17.50	7.00
GO:0004727	GABAergic synapse	KEGG_01.03.2017	160.0E-6	4.3E-3	2.4E+0	150.0E-9	5.0E-6	[-1]	Group70	11.36	10.00
GO:0048666	neuron development	GO_BiologicalProcess-GOA_23.02.2017_10h01	160.0E-6	4.4E-3	2.4E+0	13.0E-9	880.0E-9	[4, 5, 6, 8, 9]	Group69	4.79	51.00
GO:0044463	cell projection part	GO_CellularComponent-GOA_23.02.2017_10h01	190.0E-6	4.5E-3	2.3E+0	32.0E-6	260.0E-6	[2, 3, 4]	Group58	4.75	52.00
GO:0002573	myeloid leukocyte differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	180.0E-6	4.5E-3	2.3E+0	650.0E-6	2.3E-3	[6, 7, 8, 9]	Group67	7.88	16.00
GO:0005032	Morphine addiction	KEGG_01.03.2017	220.0E-6	4.5E-3	2.3E+0	150.0E-9	5.0E-6	[-1]	Group70	10.99	10.00
GO:0007267	cell-cell signaling	GO_BiologicalProcess-GOA_23.02.2017_10h01	240.0E-6	4.6E-3	2.3E+0	240.0E-6	1.3E-3	[3]	Group42	4.30	72.00
GO:0044224	juxtaparanode region of axon	GO_CellularComponent-GOA_23.02.2017_10h01	240.0E-6	4.6E-3	2.3E+0	600.0E-6	2.3E-3	[4, 5, 6, 7, 8]	Group57	33.33	4.00
GO:0007163	establishment or maintenance of cell polarity	GO_BiologicalProcess-GOA_23.02.2017_10h01	210.0E-6	4.7E-3	2.3E+0	210.0E-6	1.3E-3	[3]	Group19	8.11	15.00
GO:1990778	protein localization to cell periphery	GO_BiologicalProcess-GOA_23.02.2017_10h01	230.0E-6	4.7E-3	2.3E+0	25.0E-3	25.0E-3	[5]	Group53	7.20	18.00
GO:0090002	establishment of protein localization to plasma membrane	GO_BiologicalProcess-GOA_23.02.2017_10h01	260.0E-6	4.7E-3	2.3E+0	25.0E-3	25.0E-3	[5, 6, 7]	Group53	8.78	13.00
GO:0034702	ion channel complex	GO_CellularComponent-GOA_23.02.2017_10h01	180.0E-6	4.7E-3	2.3E+0	1.1E-3	3.5E-3	[3, 4, 5, 6]	Group68	6.92	20.00
GO:0007417	central nervous system development	GO_BiologicalProcess-GOA_23.02.2017_10h01	210.0E-6	4.7E-3	2.3E+0	13.0E-9	880.0E-9	[4, 5, 6]	Group69	4.84	48.00
GO:0060562	epithelial tube morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	280.0E-6	4.9E-3	2.3E+0	110.0E-6	820.0E-6	[4, 5, 6]	Group66	6.51	22.00
GO:0097458	neuron part	GO_CellularComponent-GOA_23.02.2017_10h01	290.0E-6	5.0E-3	2.3E+0	32.0E-6	260.0E-6	[2, 3]	Group58	4.42	63.00
GO:0045211	postsynaptic membrane	GO_CellularComponent-GOA_23.02.2017_10h01	400.0E-6	6.0E-3	2.2E+0	1.4E-6	18.0E-6	[3, 4, 5, 6, 7, 8]	Group52	7.11	17.00
GO:0004360	Axon guidance	KEGG_01.03.2017	400.0E-6	6.1E-3	2.2E+0	400.0E-6	1.7E-3	[-1]	Group13	8.00	14.00
GO:0016477	cell migration	GO_BiologicalProcess-GOA_23.02.2017_10h01	370.0E-6	6.2E-3	2.2E+0	2.9E-6	32.0E-6	[3, 4, 5]	Group62	4.41	62.00
GO:0072073	kidney epithelium development	GO_BiologicalProcess-GOA_23.02.2017_10h01	380.0E-6	6.2E-3	2.2E+0	110.0E-6	820.0E-6	[4, 5, 6, 7, 8]	Group66	8.44	13.00
GO:0007044	cell-substrate junction assembly	GO_BiologicalProcess-GOA_23.02.2017_10h01	400.0E-6	6.3E-3	2.2E+0	140.0E-6	980.0E-6	[5]	Group60	10.20	10.00
GO:0003779	actin binding	GO_MolecularFunction-GOA_23.02.2017_10h01	460.0E-6	6.5E-3	2.2E+0	2.7E-3	6.6E-3	[4]	Group59	5.88	25.00

GO:0031252	cell leading edge	GO_CellularComponent-GOA_23.02.2017_10h01	450.0E-6	6.6E-3	2.2E+0	140.0E-6	980.0E-6	[2, 3]	Group60	6.12	24.00
GO:0010464	regulation of mesenchymal cell proliferation	GO_BiologicalProcess-GOA_23.02.2017_10h01	500.0E-6	7.0E-3	2.2E+0	500.0E-6	2.0E-3	[4, 5]	Group25	16.22	6.00
GO:2000026	regulation of multicellular organismal development	GO_BiologicalProcess-GOA_23.02.2017_10h01	530.0E-6	7.2E-3	2.1E+0	13.0E-9	880.0E-9	[3, 4]	Group69	4.13	75.00
GO:0048729	tissue morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	590.0E-6	7.9E-3	2.1E+0	450.0E-9	7.4E-6	[3, 4]	Group55	5.19	34.00
GO:0097060	synaptic membrane	GO_CellularComponent-GOA_23.02.2017_10h01	690.0E-6	8.9E-3	2.1E+0	1.4E-6	18.0E-6	[2, 3, 4, 5, 6, 7]	Group52	6.35	20.00
GO:0001464	Interactions of neurexins and neuroligins at synapses	REACTOME_Pathways_01.03.2017	780.0E-6	9.6E-3	2.0E+0	150.0E-9	5.0E-6	[-1]	Group70	11.27	8.00
GO:0001465	Protein-protein interactions at synapses	REACTOME_Pathways_01.03.2017	780.0E-6	9.6E-3	2.0E+0	150.0E-9	5.0E-6	[-1]	Group70	11.27	8.00
GO:0032989	cellular component morphogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	770.0E-6	9.7E-3	2.0E+0	13.0E-9	880.0E-9	[3, 4]	Group69	4.55	49.00

Table 3.4 | NPC-specific loops GO.
(**A**) Gene ontology enrichment terms and details for genes in NPC-specific loops. (**B**) Gene ontology enrichment terms and details for genes in significant NPC loops that are lost in neurons. Negative log(FDR) from Benjamini-Hochberg correction. (**C**) Gene ontology enrichment terms and details for genes in significant NPC loops that are lost in glia. Negative log(FDR) from Benjamini-Hochberg correction. (See Table 3.2 for more information).

GOID	GOTerm	Neg log(P)	Ontology Source	Term PValue	Term PValue Corrected with Bonferroni step down	negative log (P)	% Associated Genes	Nr. Genes	Total genes in term
GO:0009653	anatomical structure morphogenesis	20.60206	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.4E-24	2.5E-21	20.6	8.36	213	2549
GO:0030182	neuron differentiation	14.76955	GO_BiologicalProcess-GOA_23.02.2017_10h01	980.0E-21	1.7E-15	14.8	9.53	127	1333
GO:0048468	cell development	14.55242	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.5E-18	2.8E-15	14.6	8.29	168	2026
GO:0007399	nervous system development	13.95860	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.3E-18	11.0E-15	14.0	7.83	185	2363
GO:0072359	circulatory system development	12.88605	GO_BiologicalProcess-GOA_23.02.2017_10h01	76.0E-18	130.0E-15	12.9	10.09	102	1011
GO:0022008	neurogenesis	11.63827	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.3E-15	2.3E-12	11.6	8.51	133	1562
GO:0007167	enzyme linked receptor protein signaling pathway	11.45593	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.9E-15	3.5E-12	11.5	9.67	101	1044
GO:0048731	system development	11.34687	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.5E-15	4.5E-12	11.3	6.28	296	4711
GO:0000904	cell morphogenesis involved in differentiation	11.21467	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.5E-15	6.1E-12	11.2	11.05	78	706
GO:0030154	cell differentiation	11	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.0E-15	10.0E-12	11.0	6.50	261	4013
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	10.52287	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-15	30.0E-12	10.5	10.73	78	727
GO:0000902	cell morphogenesis	10.48148	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-15	33.0E-12	10.5	9.64	95	985
GO:0007275	multicellular organism development	10.29243	GO_BiologicalProcess-GOA_23.02.2017_10h01	29.0E-15	51.0E-12	10.3	6.03	319	5288
GO:0048869	cellular developmental process	10.20760	GO_BiologicalProcess-GOA_23.02.2017_10h01	35.0E-15	62.0E-12	10.2	6.36	267	4199
GO:0048667	cell morphogenesis involved in neuron differentiation	9.958607	GO_BiologicalProcess-GOA_23.02.2017_10h01	63.0E-15	110.0E-12	10.0	11.56	66	571
GO:0032989	cellular component morphogenesis	9.823987	GO_BiologicalProcess-GOA_23.02.2017_10h01	89.0E-15	150.0E-12	9.8	9.20	99	1076
GO:0048646	anatomical structure formation involved in morphogenesis	9.508683	GO_BiologicalProcess-GOA_23.02.2017_10h01	170.0E-15	310.0E-12	9.5	9.26	96	1037
GO:0007507	heart development	9.161150	GO_BiologicalProcess-GOA_23.02.2017_10h01	390.0E-15	690.0E-12	9.2	11.43	63	551
GO:0048666	neuron development	8.886056	GO_BiologicalProcess-GOA_23.02.2017_10h01	750.0E-15	1.3E-9	8.9	9.02	96	1064
GO:0006928	movement of cell or subcellular component	8.853872	GO_BiologicalProcess-GOA_23.02.2017_10h01	840.0E-15	1.4E-9	8.9	7.45	152	2041
GO:0051270	regulation of cellular component movement	8.366531	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.4E-12	4.3E-9	8.4	9.56	82	858
GO:0009887	animal organ morphogenesis	8.050611	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.1E-12	8.9E-9	8.1	8.89	93	1046
GO:0009888	tissue development	8	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.0E-12	10.0E-9	8.0	7.36	146	1983
GO:0072358	cardiovascular system development	8	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.9E-12	10.0E-9	8.0	10.32	68	659
GO:0048812	neuron projection morphogenesis	7.769551	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.9E-12	17.0E-9	7.8	10.50	65	619
GO:0001568	blood vessel development	7.638272	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-12	23.0E-9	7.6	10.40	65	625
GO:0070848	response to growth factor	7.585026	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-12	26.0E-9	7.6	9.97	70	702
GO:0009790	embryo development	7.494855	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-12	32.0E-9	7.5	8.80	90	1023
GO:0061061	muscle structure development	7.481486	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-12	33.0E-9	7.5	10.41	64	615
GO:0061061	muscle structure development	7.481486	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-12	33.0E-9	7.5	10.41	64	615
GO:0048513	animal organ development	7.387216	GO_BiologicalProcess-GOA_23.02.2017_10h01	23.0E-12	41.0E-9	7.4	6.35	220	3464
GO:0048513	animal organ development	7.387216	GO_BiologicalProcess-GOA_23.02.2017_10h01	23.0E-12	41.0E-9	7.4	6.35	220	3464
GO:0009966	regulation of signal transduction	7.327902	GO_BiologicalProcess-GOA_23.02.2017_10h01	27.0E-12	47.0E-9	7.3	6.60	193	2924
GO:0009966	regulation of signal transduction	7.327902	GO_BiologicalProcess-GOA_23.02.2017_10h01	27.0E-12	47.0E-9	7.3	6.60	193	2924
GO:0009966	regulation of signal transduction	7.327902	GO_BiologicalProcess-GOA_23.02.2017_10h01	27.0E-12	47.0E-9	7.3	6.60	193	2924
GO:0031175	neuron projection development	7.180456	GO_BiologicalProcess-GOA_23.02.2017_10h01	38.0E-12	66.0E-9	7.2	9.05	82	906
GO:0040012	regulation of locomotion	7.167491	GO_BiologicalProcess-GOA_23.02.2017_10h01	39.0E-12	68.0E-9	7.2	9.20	79	859
GO:2000145	regulation of cell motility	7	GO_BiologicalProcess-GOA_23.02.2017_10h01	58.0E-12	100.0E-9	7.0	9.43	74	785
GO:0071363	cellular response to growth factor stimulus	6.744727	GO_BiologicalProcess-GOA_23.02.2017_10h01	100.0E-12	180.0E-9	6.7	9.82	66	672
GO:0048518	positive regulation of biological process	6.481486	GO_BiologicalProcess-GOA_23.02.2017_10h01	190.0E-12	330.0E-9	6.5	5.65	319	5645
GO:0048518	positive regulation of biological process	6.481486	GO_BiologicalProcess-GOA_23.02.2017_10h01	190.0E-12	330.0E-9	6.5	5.65	319	5645
GO:0048518	positive regulation of biological process	6.481486	GO_BiologicalProcess-GOA_23.02.2017_10h01	190.0E-12	330.0E-9	6.5	5.65	319	5645
GO:0048514	blood vessel morphogenesis	6.481486	GO_BiologicalProcess-GOA_23.02.2017_10h01	190.0E-12	330.0E-9	6.5	10.58	57	539
GO:0048522	positive regulation of cellular process	6.443675	GO_BiologicalProcess-GOA_23.02.2017_10h01	200.0E-12	360.0E-9	6.4	5.78	291	5037
GO:0048522	positive regulation of cellular process	6.443675	GO_BiologicalProcess-GOA_23.02.2017_10h01	200.0E-12	360.0E-9	6.4	5.78	291	5037

GO:0048522	positive regulation of cellular process	6.4436975	GO_BiologicalProcess-GOA_23.02.2017_10h01	200.0E-12	360.0E-9	6.4	5.78	291	5037
GO:0007409	axonogenesis	6.3467875	GO_BiologicalProcess-GOA_23.02.2017_10h01	260.0E-12	450.0E-9	6.3	11.04	51	462
GO:0030334	regulation of cell migration	6.2676062	GO_BiologicalProcess-GOA_23.02.2017_10h01	310.0E-12	540.0E-9	6.3	9.44	69	731
GO:0007166	cell surface receptor signaling pathway	6.2006595	GO_BiologicalProcess-GOA_23.02.2017_10h01	360.0E-12	630.0E-9	6.2	6.46	188	2909
GO:0007166	cell surface receptor signaling pathway	6.2006595	GO_BiologicalProcess-GOA_23.02.2017_10h01	360.0E-12	630.0E-9	6.2	6.46	188	2909
GO:0010646	regulation of cell communication	6.0809219	GO_BiologicalProcess-GOA_23.02.2017_10h01	480.0E-12	830.0E-9	6.1	6.28	204	3247
GO:0010646	regulation of cell communication	6.0809219	GO_BiologicalProcess-GOA_23.02.2017_10h01	480.0E-12	830.0E-9	6.1	6.28	204	3247
GO:0010646	regulation of cell communication	6.0809219	GO_BiologicalProcess-GOA_23.02.2017_10h01	480.0E-12	830.0E-9	6.1	6.28	204	3247
GO:0051094	positive regulation of developmental process	6.0222764	GO_BiologicalProcess-GOA_23.02.2017_10h01	550.0E-12	950.0E-9	6.0	7.94	100	1260
GO:0051094	positive regulation of developmental process	6.0222764	GO_BiologicalProcess-GOA_23.02.2017_10h01	550.0E-12	950.0E-9	6.0	7.94	100	1260
GO:0003007	heart morphogenesis	6.0043648	GO_BiologicalProcess-GOA_23.02.2017_10h01	570.0E-12	990.0E-9	6.0	14.05	34	242
GO:0023051	regulation of signaling	5.8538072	GO_BiologicalProcess-GOA_23.02.2017_10h01	820.0E-12	1.4E-6	5.9	6.24	206	3300
GO:0023051	regulation of signaling	5.8538072	GO_BiologicalProcess-GOA_23.02.2017_10h01	820.0E-12	1.4E-6	5.9	6.24	206	3300
GO:0023051	regulation of signaling	5.8538072	GO_BiologicalProcess-GOA_23.02.2017_10h01	820.0E-12	1.4E-6	5.9	6.24	206	3300
GO:0030029	actin filament-based process	5.7695511	GO_BiologicalProcess-GOA_23.02.2017_10h01	980.0E-12	1.7E-6	5.8	9.27	67	723
GO:0016477	cell migration	5.7212464	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.1E-9	1.9E-6	5.7	7.61	107	1406
GO:0005737	cytoplasm	5.6989	GO_CellularComponent-GOA_23.02.2017_10h01	1.1E-9	2.0E-6	5.7	4.92	558	11349
GO:0048598	embryonic morphogenesis	5.6777807	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.2E-9	2.1E-6	5.7	9.71	60	618
GO:0035556	intracellular signal transduction	5.6382722	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.3E-9	2.3E-6	5.6	6.41	183	2857
GO:0010604	positive regulation of macromolecule metabolic process	5.4089354	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.2E-9	3.9E-6	5.4	6.27	192	3061
GO:0048870	cell motility	5.3372422	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.6E-9	4.6E-6	5.3	7.35	114	1550
GO:0051240	positive regulation of multicellular organismal process	5.3279021	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.7E-9	4.7E-6	5.3	7.35	114	1552
GO:0051240	positive regulation of multicellular organismal process	5.3279021	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.7E-9	4.7E-6	5.3	7.35	114	1552
GO:0009893	positive regulation of metabolic process	5.30103	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.9E-9	5.0E-6	5.3	6.16	204	3311
GO:0050793	regulation of developmental process	5.2757241	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.0E-9	5.3E-6	5.3	6.60	158	2393
GO:0050793	regulation of developmental process	5.2757241	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.0E-9	5.3E-6	5.3	6.60	158	2393
GO:0050793	regulation of developmental process	5.2757241	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.0E-9	5.3E-6	5.3	6.60	158	2393
GO:0001525	angiogenesis	5.2596373	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.2E-9	5.5E-6	5.3	10.62	48	452
GO:0007423	sensory organ development	5.2291048	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.4E-9	5.9E-6	5.2	9.95	55	553
GO:0045595	regulation of cell differentiation	5.1870866	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.8E-9	6.5E-6	5.2	7.23	117	1619
GO:0045595	regulation of cell differentiation	5.1870866	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.8E-9	6.5E-6	5.2	7.23	117	1619
GO:0010557	positive regulation of macromolecule biosynthetic process	5.1366771	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.2E-9	7.3E-6	5.1	7.08	124	1751
GO:0001667	ameboidal-type cell migration	5.1191864	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.4E-9	7.6E-6	5.1	11.70	40	342
GO:0005622	intracellular	5.0087739	GO_CellularComponent-GOA_23.02.2017_10h01	5.7E-9	9.8E-6	5.0	4.66	676	14515
GO:0048729	tissue morphogenesis	5	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.2E-9	10.0E-6	5.0	9.31	61	655
GO:0048729	tissue morphogenesis	5	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.2E-9	10.0E-6	5.0	9.31	61	655
GO:0007417	central nervous system development	4.9586073	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.9E-9	11.0E-6	5.0	8.17	81	991
GO:0007417	central nervous system development	4.9586073	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.9E-9	11.0E-6	5.0	8.17	81	991
GO:0051960	regulation of nervous system development	4.9208188	GO_BiologicalProcess-GOA_23.02.2017_10h01	7.4E-9	12.0E-6	4.9	8.60	71	826
GO:0051960	regulation of nervous system development	4.9208188	GO_BiologicalProcess-GOA_23.02.2017_10h01	7.4E-9	12.0E-6	4.9	8.60	71	826
GO:1902531	regulation of intracellular signal transduction	4.7447275	GO_BiologicalProcess-GOA_23.02.2017_10h01	10.0E-9	18.0E-6	4.7	6.91	128	1853
GO:0007165	signal transduction	4.6989	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-9	20.0E-6	4.7	5.41	334	6170
GO:0007165	signal transduction	4.6989	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-9	20.0E-6	4.7	5.41	334	6170
GO:0022603	regulation of anatomical structure morphogenesis	4.6382722	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-9	23.0E-6	4.6	8.07	81	1004
GO:0060429	epithelium development	4.6382722	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-9	23.0E-6	4.6	7.53	97	1289
GO:0009891	positive regulation of biosynthetic process	4.6197888	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-9	24.0E-6	4.6	6.82	132	1935
GO:0051173	positive regulation of nitrogen compound metabolic process	4.6197888	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-9	24.0E-6	4.6	6.83	132	1933
GO:0030030	cell projection organization	4.6197888	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-9	24.0E-6	4.6	7.31	105	1436
GO:0098590	plasma membrane region	4.6020	GO_CellularComponent-GOA_23.02.2017_10h01	14.0E-9	25.0E-6	4.6	8.00	83	1037
GO:0030036	actin cytoskeleton organization	4.5086383	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-9	31.0E-6	4.5	9.27	58	626

GO:0035295	tube development	4.5086383	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-9	31.0E-6	4.5	9.27	58	626
GO:0007010	cytoskeleton organization	4.49485	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-9	32.0E-6	4.5	7.55	94	1245
GO:0048583	regulation of response to stimulus	4.49485	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-9	32.0E-6	4.5	5.84	231	3955
GO:0048583	regulation of response to stimulus	4.49485	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-9	32.0E-6	4.5	5.84	231	3955
GO:0048583	regulation of response to stimulus	4.49485	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-9	32.0E-6	4.5	5.84	231	3955
GO:0010628	positive regulation of gene expression	4.4814861	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-9	33.0E-6	4.5	6.86	126	1836
GO:0010631	epithelial cell migration	4.4685211	GO_BiologicalProcess-GOA_23.02.2017_10h01	20.0E-9	34.0E-6	4.5	12.75	32	251
GO:0045597	positive regulation of cell differentiation	4.455932	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-9	35.0E-6	4.5	8.23	75	911
GO:0045597	positive regulation of cell differentiation	4.455932	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-9	35.0E-6	4.5	8.23	75	911
GO:0034330	cell junction organization	4.4436975	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-9	36.0E-6	4.4	12.23	34	278
GO:0071495	cellular response to endogenous stimulus	4.2839967	GO_BiologicalProcess-GOA_23.02.2017_10h01	30.0E-9	52.0E-6	4.3	7.40	98	1324
GO:0045664	regulation of neuron differentiation	4.2441251	GO_BiologicalProcess-GOA_23.02.2017_10h01	34.0E-9	57.0E-6	4.2	9.24	55	595
GO:0045216	cell-cell junction organization	4.0861861	GO_BiologicalProcess-GOA_23.02.2017_10h01	48.0E-9	82.0E-6	4.1	12.55	31	247
GO:0031328	positive regulation of cellular biosynthetic process	4.0809219	GO_BiologicalProcess-GOA_23.02.2017_10h01	49.0E-9	83.0E-6	4.1	6.72	128	1905
GO:0031325	positive regulation of cellular metabolic process	4.0757207	GO_BiologicalProcess-GOA_23.02.2017_10h01	50.0E-9	84.0E-6	4.1	6.06	188	3100
GO:0044424	intracellular part	4.0315171	GO_CellularComponent-GOA_23.02.2017_10h01	54.0E-9	93.0E-6	4.0	4.65	660	14198
GO:0051128	regulation of cellular component organization	4.0315171	GO_BiologicalProcess-GOA_23.02.2017_10h01	55.0E-9	93.0E-6	4.0	6.36	155	2437
GO:0060284	regulation of cell development	4.0043648	GO_BiologicalProcess-GOA_23.02.2017_10h01	59.0E-9	99.0E-6	4.0	8.23	70	851
GO:0060284	regulation of cell development	4.0043648	GO_BiologicalProcess-GOA_23.02.2017_10h01	59.0E-9	99.0E-6	4.0	8.23	70	851
GO:0031589	cell-substrate adhesion	3.9586073	GO_BiologicalProcess-GOA_23.02.2017_10h01	70.0E-9	110.0E-6	4.0	11.08	37	334
GO:0048568	embryonic organ development	3.9586073	GO_BiologicalProcess-GOA_23.02.2017_10h01	68.0E-9	110.0E-6	4.0	9.98	45	451
GO:0003205	cardiac chamber development	3.9586073	GO_BiologicalProcess-GOA_23.02.2017_10h01	69.0E-9	110.0E-6	4.0	14.81	24	162
GO:0043009	chordate embryonic development	3.9586073	GO_BiologicalProcess-GOA_23.02.2017_10h01	70.0E-9	110.0E-6	4.0	9.09	55	605
GO:0010632	regulation of epithelial cell migration	3.9208188	GO_BiologicalProcess-GOA_23.02.2017_10h01	75.0E-9	120.0E-6	3.9	13.90	26	187
GO:0060537	muscle tissue development	3.853872	GO_BiologicalProcess-GOA_23.02.2017_10h01	84.0E-9	140.0E-6	3.9	10.53	40	380
GO:0032879	regulation of localization	3.853872	GO_BiologicalProcess-GOA_23.02.2017_10h01	86.0E-9	140.0E-6	3.9	6.23	165	2649
GO:0060537	muscle tissue development	3.853872	GO_BiologicalProcess-GOA_23.02.2017_10h01	84.0E-9	140.0E-6	3.9	10.53	40	380
GO:0051962	positive regulation of nervous system development	3.853872	GO_BiologicalProcess-GOA_23.02.2017_10h01	87.0E-9	140.0E-6	3.9	9.60	48	500
GO:0034329	cell junction assembly	3.7447275	GO_BiologicalProcess-GOA_23.02.2017_10h01	100.0E-9	180.0E-6	3.7	12.96	28	216
GO:0090287	regulation of cellular response to growth factor stimulus	3.7447275	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-9	180.0E-6	3.7	12.11	31	256
GO:0051093	negative regulation of developmental process	3.7212464	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-9	190.0E-6	3.7	8.05	72	894
GO:0051093	negative regulation of developmental process	3.7212464	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-9	190.0E-6	3.7	8.05	72	894
GO:0051093	negative regulation of developmental process	3.7212464	GO_BiologicalProcess-GOA_23.02.2017_10h01	110.0E-9	190.0E-6	3.7	8.05	72	894
GO:0003279	cardiac septum development	3.6777807	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	210.0E-6	3.7	18.18	18	99
GO:0051239	regulation of multicellular organismal process	3.6575773	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	220.0E-6	3.7	6.13	172	2805
GO:0051239	regulation of multicellular organismal process	3.6575773	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	220.0E-6	3.7	6.13	172	2805
GO:0051239	regulation of multicellular organismal process	3.6575773	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	220.0E-6	3.7	6.13	172	2805
GO:0051239	regulation of multicellular organismal process	3.6575773	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-9	220.0E-6	3.7	6.13	172	2805
GO:0019222	regulation of metabolic process	3.5228787	GO_BiologicalProcess-GOA_23.02.2017_10h01	180.0E-9	300.0E-6	3.5	5.24	348	6639
GO:0060255	regulation of macromolecule metabolic process	3.49485	GO_BiologicalProcess-GOA_23.02.2017_10h01	190.0E-9	320.0E-6	3.5	5.31	326	6143
GO:0014031	mesenchymal cell development	3.4814861	GO_BiologicalProcess-GOA_23.02.2017_10h01	200.0E-9	330.0E-6	3.5	22.58	14	62
GO:0007420	brain development	3.4685211	GO_BiologicalProcess-GOA_23.02.2017_10h01	200.0E-9	340.0E-6	3.5	8.34	63	755
GO:0007420	brain development	3.4685211	GO_BiologicalProcess-GOA_23.02.2017_10h01	200.0E-9	340.0E-6	3.5	8.34	63	755
GO:0003206	cardiac chamber morphogenesis	3.4685211	GO_BiologicalProcess-GOA_23.02.2017_10h01	200.0E-9	340.0E-6	3.5	16.13	20	124
GO:0038084	vascular endothelial growth factor signaling pathway	3.4436975	GO_BiologicalProcess-GOA_23.02.2017_10h01	210.0E-9	360.0E-6	3.4	33.33	10	30
GO:0060322	head development	3.4089354	GO_BiologicalProcess-GOA_23.02.2017_10h01	230.0E-9	390.0E-6	3.4	8.23	65	790
GO:0060322	head development	3.4089354	GO_BiologicalProcess-GOA_23.02.2017_10h01	230.0E-9	390.0E-6	3.4	8.23	65	790
GO:0099568	cytoplasmic region	3.3767507	GO_CellularComponent-GOA_23.02.2017_10h01	250.0E-9	420.0E-6	3.4	10.87	35	322
GO:0060562	epithelial tube morphogenesis	3.3372422	GO_BiologicalProcess-GOA_23.02.2017_10h01	280.0E-9	460.0E-6	3.3	10.65	36	338
GO:0060485	mesenchyme development	3.3098039	GO_BiologicalProcess-GOA_23.02.2017_10h01	290.0E-9	490.0E-6	3.3	12.08	29	240

GO:0048519	negative regulation of biological process	3.30103	GO_BiologicalProcess-GOA_23.02.2017_10h01	300.0E-9	500.0E-6	3.3	5.47	275	5031
GO:0048519	negative regulation of biological process	3.30103	GO_BiologicalProcess-GOA_23.02.2017_10h01	300.0E-9	500.0E-6	3.3	5.47	275	5031
GO:0048519	negative regulation of biological process	3.30103	GO_BiologicalProcess-GOA_23.02.2017_10h01	300.0E-9	500.0E-6	3.3	5.47	275	5031
GO:0042692	muscle cell differentiation	3.22914	GO_BiologicalProcess-GOA_23.02.2017_10h01	350.0E-9	590.0E-6	3.2	10.18	39	383
GO:0071310	cellular response to organic substance	3.22187	GO_BiologicalProcess-GOA_23.02.2017_10h01	360.0E-9	600.0E-6	3.2	6.24	150	2403
GO:2000026	regulation of multicellular organismal development	3.21467	GO_BiologicalProcess-GOA_23.02.2017_10h01	370.0E-9	610.0E-6	3.2	6.62	120	1814
GO:2000026	regulation of multicellular organismal development	3.21467	GO_BiologicalProcess-GOA_23.02.2017_10h01	370.0E-9	610.0E-6	3.2	6.62	120	1814
GO:2000026	regulation of multicellular organismal development	3.21467	GO_BiologicalProcess-GOA_23.02.2017_10h01	370.0E-9	610.0E-6	3.2	6.62	120	1814
GO:0045944	positive regulation of transcription from RNA polymerase II promoter	3.20760	GO_BiologicalProcess-GOA_23.02.2017_10h01	370.0E-9	620.0E-6	3.2	7.36	84	1141
GO:0050767	regulation of neurogenesis	3.18045	GO_BiologicalProcess-GOA_23.02.2017_10h01	390.0E-9	660.0E-6	3.2	8.31	61	734
GO:0050767	regulation of neurogenesis	3.18045	GO_BiologicalProcess-GOA_23.02.2017_10h01	390.0E-9	660.0E-6	3.2	8.31	61	734
GO:0014069	postsynaptic density	3.15490	GO_CellularComponent-GOA_23.02.2017_10h01	420.0E-9	700.0E-6	3.2	12.44	27	217
GO:0000981	RNA polymerase II transcription factor activity, sequence-specific DNA binding	3.15490	GO_MolecularFunction-GOA_23.02.2017_10h01	420.0E-9	700.0E-6	3.2	8.28	61	737
GO:0030900	forebrain development	3.12493	GO_BiologicalProcess-GOA_23.02.2017_10h01	450.0E-9	750.0E-6	3.1	9.90	40	404
GO:0048523	negative regulation of cellular process	3.12493	GO_BiologicalProcess-GOA_23.02.2017_10h01	450.0E-9	750.0E-6	3.1	5.54	252	4552
GO:0048523	negative regulation of cellular process	3.12493	GO_BiologicalProcess-GOA_23.02.2017_10h01	450.0E-9	750.0E-6	3.1	5.54	252	4552
GO:0001501	skeletal system development	3.11918	GO_BiologicalProcess-GOA_23.02.2017_10h01	460.0E-9	760.0E-6	3.1	9.18	47	512
GO:0048705	skeletal system morphogenesis	3.11918	GO_BiologicalProcess-GOA_23.02.2017_10h01	460.0E-9	760.0E-6	3.1	12.39	27	218
GO:0097485	neuron projection guidance	3.11918	GO_BiologicalProcess-GOA_23.02.2017_10h01	460.0E-9	760.0E-6	3.1	11.84	29	245
GO:0043168	anion binding	3.08618	GO_MolecularFunction-GOA_23.02.2017_10h01	490.0E-9	820.0E-6	3.1	6.01	172	2860
GO:0035239	tube morphogenesis	3.03151	GO_BiologicalProcess-GOA_23.02.2017_10h01	560.0E-9	930.0E-6	3.0	10.13	38	375
GO:0048562	embryonic organ morphogenesis	3.02687	GO_BiologicalProcess-GOA_23.02.2017_10h01	560.0E-9	940.0E-6	3.0	10.86	33	304

Table 3.5 | Glia-specific loops GO.
Gene ontology enrichment terms and details for genes in glia-specific loops (See Table 3.2 for more information).

GOID	GOTerm	Neg log(P)	Ontology Source	Term PValue	Term PValue Corrected with Bonferroni step down	% Associated Genes	Nr. Genes	Total genes in term
GO:0044444	cytoplasmic part	11.7	GO_CellularComponent-GOA_23.02.2017_10h01	1.1E-15	2.1E-12	17.24	1624	9421
GO:0005829	cytosol	8.6	GO_CellularComponent-GOA_23.02.2017_10h01	1.3E-12	2.5E-9	18.27	910	4982
GO:0043547	positive regulation of GTPase activity	5.4	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.0E-9	3.8E-6	23.62	163	690
GO:0006464	cellular protein modification process	5.2	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.1E-9	5.9E-6	18.12	740	4085
GO:0043231	intracellular membrane-bounded organelle	4.3	GO_CellularComponent-GOA_23.02.2017_10h01	25.0E-9	48.0E-6	16.39	1762	10749
GO:0015031	protein transport	4.3	GO_BiologicalProcess-GOA_23.02.2017_10h01	26.0E-9	50.0E-6	19.33	405	2095
GO:0018193	peptidyl-amino acid modification	4.2	GO_BiologicalProcess-GOA_23.02.2017_10h01	35.0E-9	66.0E-6	20.66	264	1278
GO:0015833	peptide transport	4.1	GO_BiologicalProcess-GOA_23.02.2017_10h01	40.0E-9	75.0E-6	19.25	410	2130
GO:0044446	intracellular organelle part	4.0	GO_CellularComponent-GOA_23.02.2017_10h01	51.0E-9	95.0E-6	16.64	1481	8901
GO:0032555	purine ribonucleotide binding	3.4	GO_MolecularFunction-GOA_23.02.2017_10h01	220.0E-9	410.0E-6	19.19	379	1975
GO:0051223	regulation of protein transport	3.3	GO_BiologicalProcess-GOA_23.02.2017_10h01	250.0E-9	460.0E-6	21.73	178	819
GO:0005654	nucleoplasm	3.2	GO_CellularComponent-GOA_23.02.2017_10h01	310.0E-9	580.0E-6	17.99	619	3441
GO:0032559	adenyl ribonucleotide binding	3.2	GO_MolecularFunction-GOA_23.02.2017_10h01	360.0E-9	670.0E-6	19.58	316	1614
GO:0030554	adenyl nucleotide binding	3.1	GO_MolecularFunction-GOA_23.02.2017_10h01	390.0E-9	730.0E-6	19.56	318	1626
GO:0090087	regulation of peptide transport	3.0	GO_BiologicalProcess-GOA_23.02.2017_10h01	510.0E-9	970.0E-6	21.44	182	849
GO:0006468	protein phosphorylation	3.0	GO_BiologicalProcess-GOA_23.02.2017_10h01	540.0E-9	1.0E-3	19.04	376	1975
GO:1902531	regulation of intracellular signal transduction	3.0	GO_BiologicalProcess-GOA_23.02.2017_10h01	620.0E-9	1.1E-3	19.16	355	1853
GO:0044428	nuclear part	2.7	GO_CellularComponent-GOA_23.02.2017_10h01	970.0E-9	1.8E-3	17.46	770	4411
GO:0030097	hemopoiesis	2.6	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.4E-6	2.7E-3	21.52	164	762
GO:0048534	hematopoietic or lymphoid organ development	2.4	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.1E-6	3.9E-3	21.24	171	805
GO:0009968	negative regulation of signal transduction	2.4	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.3E-6	4.3E-3	20.02	239	1194
GO:0005524	ATP binding	2.3	GO_MolecularFunction-GOA_23.02.2017_10h01	2.5E-6	4.7E-3	19.29	304	1576
GO:0050708	regulation of protein secretion	2.3	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.5E-6	4.7E-3	23.62	103	436
GO:0002791	regulation of peptide secretion	2.0	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.1E-6	9.6E-3	23.06	107	464
GO:0051247	positive regulation of protein metabolic process	2.0	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.3E-6	9.9E-3	19.07	312	1636
GO:0004015	Rap1 signaling pathway	2.0	KEGG_01.03.2017	6.2E-6	11.0E-3	27.14	57	210
GO:0031981	nuclear lumen	2.0	GO_CellularComponent-GOA_23.02.2017_10h01	6.4E-6	11.0E-3	17.40	704	4045
GO:0019932	second-messenger-mediated signaling	1.9	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.9E-6	12.0E-3	25.77	67	260
GO:0007169	transmembrane receptor protein tyrosine kinase signaling pathway	1.8	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.5E-6	15.0E-3	21.18	154	727
GO:0004115	3',5'-cyclic-AMP phosphodiesterase activity	1.8	GO_MolecularFunction-GOA_23.02.2017_10h01	8.7E-6	16.0E-3	66.67	10	15
GO:0032270	positive regulation of cellular protein metabolic process	1.7	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-6	20.0E-3	19.08	292	1530
GO:0031399	regulation of protein modification process	1.7	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-6	21.0E-3	18.75	333	1776
GO:0009306	protein secretion	1.7	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-6	21.0E-3	22.06	120	544
GO:0048666	neuron development	1.5	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-6	34.0E-3	19.83	211	1064
GO:0002790	peptide secretion	1.5	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-6	35.0E-3	21.70	125	576
GO:0007167	enzyme linked receptor protein signaling pathway	1.4	GO_BiologicalProcess-GOA_23.02.2017_10h01	23.0E-6	43.0E-3	19.83	207	1044

Table 3.6 | GM12878 lymphoblastoid-specific loops GO.
Gene ontology enrichment terms and details for genes in GM12878-specific loops (See Table 3.2 for more).

chr(Neu TAD)	x1(Neu TAD)	x2(Neu TAD)	No. Overlap Glia TADs	chr (Glia Gene)	y1 (Glia Gene)	y2 (Glia Gene)	Gene ID (ENSG)	Gene Symbol	Strand	Gene length (bp)
chr2	139600000	143750000	11	chr2	139654893	139656744	ENSG00000229104	YY1P2	+	1851
chr2	139600000	143750000	11	chr2	141750111	141750228	ENSG00000201892	Y_RNA	+	117
chr2	139600000	143750000	11	chr2	141924825	141924926	ENSG00000252015	U6	-	101
chr2	139600000	143750000	11	chr2	143635194	143799885	ENSG00000115919	KYNU	+	114806
chr2	139600000	143750000	11	chr2	140988995	142889270	ENSG00000168702	LRP1B	-	1900275
chr2	140425000	143750000	11	chr2	141750111	141750228	ENSG00000201892	Y_RNA	+	117
chr2	140425000	143750000	11	chr2	141924825	141924926	ENSG00000252015	U6	-	101
chr2	140425000	143750000	11	chr2	143635194	143799885	ENSG00000115919	KYNU	+	114806
chr2	140425000	143750000	11	chr2	140988995	142889270	ENSG00000168702	LRP1B	-	1900275
chr2	139600000	142950000	10	chr2	139654893	139656744	ENSG00000229104	YY1P2	+	1851
chr2	139600000	142950000	10	chr2	141750111	141750228	ENSG00000201892	Y_RNA	+	117
chr2	139600000	142950000	10	chr2	141924825	141924926	ENSG00000252015	U6	-	101
chr2	139600000	142950000	10	chr2	140988995	142889270	ENSG00000168702	LRP1B	-	1900275
chr18	25175000	28550000	9	chr18	25373549	25373648	n/a	Mir_384	n/a	99
chr18	25175000	28550000	9	chr18	25613650	25613747	n/a	Mir_340	n/a	97
chr18	25175000	28550000	9	chr18	27878875	27878926	n/a	MIR302F	n/a	51
chr18	25175000	28550000	9	chr18	25530929	25757445	ENSG00000170558	CDH2	-	226516
chr18	25175000	28550000	9	chr18	24916342	25175128	ENSG00000264151	AK127888	-	128
chr9	80925000	85725000	9	chr9	82478897	82478939	n/a	DQ575560	n/a	42
chr9	80925000	85725000	9	chr9	82479022	82479053	n/a	DQ583878	n/a	31
chr9	80925000	85725000	9	chr9	82479147	82479179	n/a	DQ589820	n/a	32
chr9	80925000	85725000	9	chr9	82479196	82479300	n/a	DQ580124	n/a	104
chr9	80925000	85725000	9	chr9	82479444	82479479	n/a	DQ595182	n/a	35
chr9	80925000	85725000	9	chr9	82479496	82479529	n/a	DQ596925	n/a	33
chr9	80925000	85725000	9	chr9	82479564	82479595	n/a	DQ597713	n/a	31
chr9	80925000	85725000	9	chr9	82479638	82479675	n/a	DQ586158	n/a	37
chr9	80925000	85725000	9	chr9	82479836	82479872	n/a	DQ588659	n/a	36
chr9	80925000	85725000	9	chr9	82479911	82479966	n/a	DQ599976	n/a	55
chr9	80925000	85725000	9	chr9	82480085	82480149	n/a	DQ574305	n/a	64
chr9	80925000	85725000	9	chr9	82480265	82480311	n/a	DQ600106	n/a	46
chr9	80925000	85725000	9	chr9	82480444	82480513	n/a	DQ574306	n/a	69
chr9	80925000	85725000	9	chr9	82480639	82480689	n/a	DQ599155	n/a	50
chr9	80925000	85725000	9	chr9	82480842	82480885	n/a	DQ574330	n/a	43
chr9	80925000	85725000	9	chr9	82481004	82481059	n/a	DQ579956	n/a	55
chr9	80925000	85725000	9	chr9	82481263	82481303	n/a	DQ594965	n/a	40
chr9	80925000	85725000	9	chr9	82481394	82481432	n/a	DQ596802	n/a	38
chr9	80925000	85725000	9	chr9	82481509	82481544	n/a	DQ574826	n/a	35
chr9	80925000	85725000	9	chr9	82481703	82481771	n/a	DQ592136	n/a	68
chr9	80925000	85725000	9	chr9	82482216	82482251	n/a	DQ577348	n/a	35
chr9	80925000	85725000	9	chr9	82482361	82482391	n/a	DQ593172	n/a	30
chr9	80925000	85725000	9	chr9	82482630	82482673	n/a	DQ592375	n/a	43
chr9	80925000	85725000	9	chr9	82483105	82483135	n/a	DQ570523	n/a	30
chr9	80925000	85725000	9	chr9	82483941	82483980	n/a	DQ573195	n/a	39
chr9	80925000	85725000	9	chr9	82973117	82973151	n/a	DQ591664	n/a	34
chr9	80925000	85725000	9	chr9	84528351	84534842	ENSG00000240632	SPATA31D5P	+	6491
chr9	80925000	85725000	9	chr9	84531745	84531773	n/a	DQ592725	n/a	28
chr9	80925000	85725000	9	chr9	84532087	84532118	n/a	DQ577940	n/a	31
chr9	80925000	85725000	9	chr9	84533136	84533203	n/a	DQ578305	n/a	67
chr9	80925000	85725000	9	chr9	84534683	84534718	n/a	DQ582032	n/a	35
chr9	80925000	85725000	9	chr9	84535666	84535704	n/a	DQ584769	n/a	38
chr9	80925000	85725000	9	chr9	84543342	84549913	ENSG00000189357	SPATA31D4	+	6571
chr9	80925000	85725000	9	chr9	84547133	84547164	n/a	DQ577940	n/a	31
chr9	80925000	85725000	9	chr9	84548208	84548275	n/a	DQ578305	n/a	67
chr9	80925000	85725000	9	chr9	84549755	84549790	n/a	DQ582032	n/a	35
chr9	80925000	85725000	9	chr9	84550738	84550776	n/a	DQ584769	n/a	38
chr9	80925000	85725000	9	chr9	84558414	84565009	ENSG00000186788	SPATA31D3	+	6595
chr9	80925000	85725000	9	chr9	84562225	84562256	n/a	DQ577940	n/a	31
chr9	80925000	85725000	9	chr9	84563300	84563367	n/a	DQ578305	n/a	67
chr9	80925000	85725000	9	chr9	84891466	84891571	n/a	Mir_544	n/a	105
chr9	80925000	85725000	9	chr9	85045810	85045838	n/a	DQ588544	n/a	28
chr9	80925000	85725000	9	chr9	85202301	85202353	n/a	Mir_1302	n/a	52
chr9	80925000	85725000	9	chr9	80912058	80945009	ENSG00000135069	PSAT1	+	20009
chr9	80925000	85725000	9	chr9	84304627	84391814	ENSG00000233926	BC036431	+	87187
chr9	80925000	85725000	9	chr9	84545140	84592004	ENSG00000267559	AK097447	-	46864
chr9	80925000	85725000	9	chr9	84603686	84610171	ENSG00000214929	SPATA31D1	+	6485
chr9	80925000	85725000	9	chr9	85594499	85678043	ENSG00000165105	RASEF	-	83544
chr9	80925000	85725000	9	chr9	82186877	82341656	ENSG00000106829	TLE4	+	154779
chr9	80925000	85725000	9	chr9	84198597	84303596	ENSG00000196781	TLE1	-	104999
chr10	60100000	62425000	8	chr10	60474774	60477293	ENSG00000183055	FAM133CP	+	2519
chr10	60100000	62425000	8	chr10	61019597	61019717	n/a	Mir_584	n/a	120
chr10	60100000	62425000	8	chr10	61821742	61822047	n/a	Metazoa_SRP	n/a	305
chr10	60100000	62425000	8	chr10	61841498	61841612	ENSG00000202190	Y_RNA	+	114
chr10	60100000	62425000	8	chr10	60094738	60130513	ENSG00000072401	UBE2D1	+	30513
chr10	60100000	62425000	8	chr10	60144902	60158990	ENSG00000108064	TFAM	+	14088
chr10	60100000	62425000	8	chr10	61410521	61469649	ENSG00000165449	SLC16A9	-	59128
chr10	60100000	62425000	8	chr10	61496747	61513203	ENSG00000227877	LINC00948	-	16456
chr10	60100000	62425000	8	chr10	61715187	61720671	ENSG00000235931	C10orf40	-	5484
chr10	60100000	62425000	8	chr10	60272903	60588845	ENSG00000122870	BICC1	+	315942
chr10	60100000	62425000	8	chr10	60936347	61007534	ENSG00000165443	PHYHIPL	+	71187
chr10	60100000	62425000	8	chr10	61005888	61122352	ENSG00000148541	FAM13C	-	116464
chr10	60100000	62425000	8	chr10	61548505	61666414	ENSG00000108091	CCDC6	-	117909
chr10	60100000	62425000	8	chr10	61786055	62149634	ENSG00000151150	ANK3	-	363579
chr3	58525000	62275000	8	chr3	59956575	59958982	n/a	NPCR	n/a	2407
chr3	58525000	62275000	8	chr3	60842060	60842277	ENSG00000212211	U3	-	217
chr3	58525000	62275000	8	chr3	61068514	61068637	n/a	5S_rRNA	n/a	123

chr3	58525000	62275000	8	chr3	58549844	58563491	ENSG00000168309	FAM107A	-	13647
chr3	58525000	62275000	8	chr3	58619669	58652561	ENSG00000198643	FAM3D	-	32892
chr3	58525000	62275000	8	chr3	58727736	59035715	ENSG00000163689	C3orf67	-	307979
chr3	58525000	62275000	8	chr3	58810196	59004819	ENSG00000242428	AK090895	+	194623
chr3	58525000	62275000	8	chr3	62247493	62304622	ENSG00000241472	PTPRG-AS1	-	27507
chr3	58525000	62275000	8	chr3	59735035	61237133	ENSG00000189283	FHIT	-	1502098
chr3	58525000	62275000	8	chr3	61547242	62280573	ENSG00000144724	PTPRG	+	727758
chr4	175250000	177250000	8	chr4	175344945	175345015	ENSG00000265846	MIR4276	+	70
chr4	175250000	177250000	8	chr4	175474212	175474248	n/a	HH834010	n/a	36
chr4	175250000	177250000	8	chr4	175848510	175848545	n/a	HH834010	n/a	35
chr4	175250000	177250000	8	chr4	177019313	177019436	ENSG00000201516	SNORA51	+	123
chr4	175250000	177250000	8	chr4	175411327	175444049	ENSG00000164120	HPGD	-	32722
chr4	175250000	177250000	8	chr4	176711457	176733340	ENSG00000249106	BC038536	+	21883
chr4	175250000	177250000	8	chr4	177105724	177116822	ENSG00000150628	SPATA4	-	11098
chr4	175250000	177250000	8	chr4	177134825	177190373	ENSG00000164122	ASB5	-	55548
chr4	175250000	177250000	8	chr4	177241089	177253396	ENSG00000129128	SPCS3	+	8911
chr4	175250000	177250000	8	chr4	175205054	175254531	ENSG00000164118	CEP44	+	4531
chr4	175250000	177250000	8	chr4	175563197	175750465	ENSG00000145451	GLRA3	-	187268
chr4	175250000	177250000	8	chr4	175750818	175791938	ENSG00000168594	AK093264	+	41120
chr4	175250000	177250000	8	chr4	175839508	175899331	ENSG00000168594	ADAM29	+	59823
chr4	175250000	177250000	8	chr4	176554087	176923842	ENSG00000150625	GPM6A	-	369755
chr4	175250000	177250000	8	chr4	176986984	177103979	ENSG00000150627	WDR17	+	116995
chr5	162950000	167650000	8	chr5	165036445	165036684	ENSG00000252794	7SK	+	239
chr5	162950000	167650000	8	chr5	167460095	167460210	ENSG00000253065	SNORA40	-	115
chr5	162950000	167650000	8	chr5	163781145	164029423	ENSG00000241956	BC011998	+	248278
chr5	162950000	167650000	8	chr5	166711842	167691162	ENSG00000145934	TENM2	+	938158
chr1	215800000	218375000	7	chr1	216825011	216825068	n/a	Mir_598	n/a	57
chr1	215800000	218375000	7	chr1	218066241	218094146	ENSG00000231814	LINC00210	+	27905
chr1	215800000	218375000	7	chr1	216676587	216896814	ENSG00000196482	ESRRG	-	220227
chr1	215800000	218375000	7	chr1	217603833	217804409	ENSG00000092978	GPATCH2	-	200576
chr1	215800000	218375000	7	chr1	217804694	218040484	ENSG00000162814	SPATA17	+	235790
chr1	215800000	218375000	7	chr1	215796235	216596738	ENSG00000042781	USH2A	-	796738
chr1	71550000	75175000	7	chr1	72259914	72302695	ENSG00000228853	NEGR1-IT1	-	42781
chr1	71550000	75175000	7	chr1	75043113	75091782	ENSG00000234497	CR627203	+	48669
chr1	71550000	75175000	7	chr1	75171171	75199092	ENSG00000116791	CRYZ	-	3829
chr1	71550000	75175000	7	chr1	71547006	71703406	ENSG00000229956	ZRANB2-AS2	+	153406
chr1	71550000	75175000	7	chr1	73771852	73804560	ENSG00000233973	BC041341	+	32708
chr1	71550000	75175000	7	chr1	74491701	74663871	ENSG00000162620	LRRIQ3	-	172170
chr1	71550000	75175000	7	chr1	74663895	75010116	ENSG00000116783	FPGT-TNNI3K	+	346221
chr1	71550000	75175000	7	chr1	75033794	75139422	ENSG00000178965	C1orf173	-	105628
chr1	71550000	75175000	7	chr1	71868624	72748405	ENSG00000172260	NEGR1	-	879781
chr10	65600000	69550000	7	chr10	66585284	66586634	ENSG00000216740	ANXA2P3	+	1350
chr10	65600000	69550000	7	chr10	66585343	66585369	n/a	DJ439558	n/a	26
chr10	65600000	69550000	7	chr10	66586101	66586126	n/a	DJ439576	n/a	25
chr10	65600000	69550000	7	chr10	66586203	66586227	n/a	DJ439561	n/a	24
chr10	65600000	69550000	7	chr10	68254456	68254557	ENSG00000252203	SnoU40	+	101
chr10	65600000	69550000	7	chr10	69524260	69524407	n/a	HI650153	n/a	147
chr10	65600000	69550000	7	chr10	68685791	68860867	ENSG00000198739	LRRTM3	+	175076
chr10	65600000	69550000	7	chr10	67679724	69455949	ENSG00000183230	CTNNA3	-	1776225
chr12	96925000	98800000	7	chr12	97490690	97490764	n/a	TRNA_Pseudo	n/a	74
chr12	96925000	98800000	7	chr12	97506929	97507042	n/a	Mir_584	n/a	113
chr12	96925000	98800000	7	chr12	97866326	97868402	n/a	AK129935	n/a	2076
chr12	96925000	98800000	7	chr12	97885686	97885756	ENSG00000221479	MIR1251	+	70
chr12	96925000	98800000	7	chr12	97945914	97946006	ENSG00000251844	SnoMe28S_Am2634	-	92
chr12	96925000	98800000	7	chr12	97957589	97957689	ENSG00000207586	MIR135A2	+	100
chr12	96925000	98800000	7	chr12	98389160	98389226	ENSG00000263890	MIR4303	-	66
chr12	96925000	98800000	7	chr12	98752079	98752163	n/a	Mir_548	n/a	84
chr12	96925000	98800000	7	chr12	96883352	96934765	ENSG00000188596	C12orf55	+	9765
chr12	96925000	98800000	7	chr12	97301000	97347469	ENSG00000139350	NEDD1	+	46469
chr12	96925000	98800000	7	chr12	98123825	98150295	ENSG00000257501	LOC643711	-	26470
chr12	96925000	98800000	7	chr12	96933914	97024429	ENSG00000188596	AX747187	+	90515
chr12	96925000	98800000	7	chr12	97041752	97159032	ENSG00000188596	C12orf63	+	117280
chr12	96925000	98800000	7	chr12	97858798	97927544	ENSG00000255794	RMST	+	68746
chr16	63375000	66400000	7	chr16	65420321	65420381	n/a	JB153694	n/a	60
chr16	63375000	66400000	7	chr16	66335711	66335843	ENSG00000201999	5S_rRNA	+	132
chr16	63375000	66400000	7	chr16	65318401	65610203	ENSG00000261742	LINC00922	-	291802
chr16	63375000	66400000	7	chr16	64980682	65155919	ENSG00000140937	CDH11	-	175237
chr2	140500000	142900000	7	chr2	141750111	141750228	ENSG00000201892	Y_RNA	+	117
chr2	140500000	142900000	7	chr2	141924825	141924926	ENSG00000252015	U6	-	101
chr2	140500000	142900000	7	chr2	140988995	142889270	ENSG00000168702	LRP1B	-	1900275
chr2	33875000	37175000	7	chr2	33951127	33953284	ENSG00000239649	MYADML	-	2157
chr2	33875000	37175000	7	chr2	34628738	34628822	ENSG00000212025	Mir_548	-	84
chr2	33875000	37175000	7	chr2	36581891	36582713	n/a	LOC100288911	n/a	822
chr2	33875000	37175000	7	chr2	36779403	36825332	ENSG00000171055	FEZ2	-	45929
chr2	33875000	37175000	7	chr2	36923832	37041937	ENSG00000205221	VIT	+	118105
chr2	33875000	37175000	7	chr2	37064840	37193615	ENSG00000115808	STRN	-	110160
chr2	33875000	37175000	7	chr2	36583369	36778278	ENSG00000150938	CRIM1	+	194909
chr2	62125000	64150000	7	chr2	62432960	62433052	ENSG00000266097	MIR5192	+	92
chr2	62125000	64150000	7	chr2	62489521	62489806	ENSG00000239958	Metazoa_SRP	-	285
chr2	62125000	64150000	7	chr2	62727355	62733604	ENSG00000186889	TMEM17	-	6249
chr2	62125000	64150000	7	chr2	62953770	62953875	ENSG00000252436	Y_RNA	-	105
chr2	62125000	64150000	7	chr2	63271099	63274846	ENSG00000231609	LOC100132215	-	3747
chr2	62125000	64150000	7	chr2	63277936	63284966	ENSG00000115507	OTX1	+	7030
chr2	62125000	64150000	7	chr2	63344985	63346677	ENSG00000242412	DBIL5P2	-	1692
chr2	62125000	64150000	7	chr2	62296587	62374016	ENSG00000229839	BC071802	-	77429
chr2	62125000	64150000	7	chr2	62423261	62451866	ENSG00000170340	B3GNT2	+	28605
chr2	62125000	64150000	7	chr2	62839006	62889763	ENSG00000226622	BC038779	-	50757
chr2	62125000	64150000	7	chr2	62132802	62363205	ENSG00000173163	COMMD1	+	230403
chr2	62125000	64150000	7	chr2	62933000	63273621	ENSG00000115504	EHBP1	+	340621
chr2	62125000	64150000	7	chr2	63348534	63815867	ENSG00000143951	WDPCP	-	467333
chr2	62125000	64150000	7	chr2	63816284	63834330	ENSG00000014641	MDH1	+	18046

chr2	62125000	64150000	7	chr2	64069013	64118696	ENSG00000169764	UGP2	+	49683
chr2	62125000	64150000	7	chr2	64119666	64246214	ENSG00000143952	VPS54	-	30334
chr4	181975000	184250000	7	chr4	182756017	182756261	ENSG00000251742	7SK	-	244
chr4	181975000	184250000	7	chr4	183060158	183065668	ENSG00000177822	MGC45800	-	5510
chr4	181975000	184250000	7	chr4	183090445	183090531	ENSG00000221227	MIR1305	+	86
chr4	181975000	184250000	7	chr4	183469811	183469932	ENSG00000252343	U2	+	121
chr4	181975000	184250000	7	chr4	184018173	184020352	ENSG00000251359	WWC2-AS2	-	2179
chr4	181975000	184250000	7	chr4	184090500	184090570	ENSG00000252702	U7	-	70
chr4	181975000	184250000	7	chr4	184239219	184241927	ENSG00000177300	CLDN22	-	2708
chr4	181975000	184250000	7	chr4	184242916	184243579	ENSG00000185758	CLDN24	-	663
chr4	181975000	184250000	7	chr4	183811243	183838630	ENSG00000129187	DCTD	-	27387
chr4	181975000	184250000	7	chr4	183958817	183961272	ENSG00000230219	FAM92A1P2	+	2455
chr4	181975000	184250000	7	chr4	181985242	182080302	ENSG00000248197	LINC00290	-	95060
chr4	181975000	184250000	7	chr4	182896308	183006289	ENSG00000177822	AK056196	-	109981
chr4	181975000	184250000	7	chr4	184020462	184241929	ENSG00000151718	WWC2	+	221467
chr4	181975000	184250000	7	chr4	184154780	184161787	ENSG00000251128	BX537950	-	7007
chr4	181975000	184250000	7	chr4	183245098	183724177	ENSG00000218336	TENM3	+	479079
chr9	6775000	9925000	7	chr9	6834701	6835011	ENSG00000230581	DQ580140	+	310
chr9	6775000	9925000	7	chr9	7167235	7167357	n/a	Mir_584	n/a	122
chr9	6775000	9925000	7	chr9	8858133	8861724	ENSG00000225706	AK094342	+	3591
chr9	6775000	9925000	7	chr9	8888135	8888207	ENSG00000251699	SNORD27	+	72
chr9	6775000	9925000	7	chr9	9442059	9442347	ENSG00000265735	Metazoa_SRP	+	288
chr9	6775000	9925000	7	chr9	7796490	7799799	ENSG00000137038	C9orf123	-	3309
chr9	6775000	9925000	7	chr9	6757640	7175648	ENSG00000107077	KDM4C	+	400648
chr9	6775000	9925000	7	chr9	8314245	10612723	ENSG00000153707	PTPRD	-	1610755
chr1	95650000	98950000	6	chr1	95699710	95712781	ENSG00000122481	RWDD3	+	13071
chr1	95650000	98950000	6	chr1	95970531	95970615	ENSG00000216037	Mir_548	-	84
chr1	95650000	98950000	6	chr1	97161411	97161723	ENSG00000223229	7SK	-	312
chr1	95650000	98950000	6	chr1	98510798	98510907	ENSG00000225206	MIR2682	-	109
chr1	95650000	98950000	6	chr1	95582893	95663161	ENSG00000152078	TMEM56	+	13161
chr1	95650000	98950000	6	chr1	95975895	95981020	ENSG00000228971	BC067883	+	5125
chr1	95650000	98950000	6	chr1	98453555	98515249	ENSG00000225206	MIR137HG	-	61694
chr1	95650000	98950000	6	chr1	95628774	95699538	ENSG00000226026	AK090700	-	49538
chr1	95650000	98950000	6	chr1	95940292	95944912	ENSG00000233907	FLJ31662	+	4620
chr1	95650000	98950000	6	chr1	97187174	97280605	ENSG00000117569	PTBP2	+	93431
chr1	95650000	98950000	6	chr1	97543299	98386615	ENSG00000188641	DPYD	-	843316
chr1	95650000	98950000	6	chr1	97561478	97788511	ENSG00000232878	DPYD-AS1	+	227033
chr1	95650000	98950000	6	chr1	98676266	98738214	ENSG00000226053	LOC729987	+	61948
chr10	60100000	61800000	6	chr10	60474774	60477293	ENSG00000183055	FAM133CP	+	2519
chr10	60100000	61800000	6	chr10	61019597	61019717	n/a	Mir_584	n/a	120
chr10	60100000	61800000	6	chr10	60094738	60130513	ENSG00000072401	UBE2D1	+	30513
chr10	60100000	61800000	6	chr10	60144902	60158990	ENSG00000108064	TFAM	+	14088
chr10	60100000	61800000	6	chr10	61410521	61469649	ENSG00000165449	SLC16A9	-	59128
chr10	60100000	61800000	6	chr10	61496747	61513203	ENSG00000227877	LINC00948	-	16456
chr10	60100000	61800000	6	chr10	61715187	61720671	ENSG00000235931	C10orf40	-	5484
chr10	60100000	61800000	6	chr10	60272903	60588845	ENSG00000122870	BICC1	+	315942
chr10	60100000	61800000	6	chr10	60936347	61007534	ENSG00000165443	PHYHIPL	+	71187
chr10	60100000	61800000	6	chr10	61005888	61122352	ENSG00000148541	FAM13C	-	116464
chr10	60100000	61800000	6	chr10	61548505	61666414	ENSG00000108091	CCDC6	-	117909
chr10	60100000	61800000	6	chr10	61786055	62149634	ENSG00000151150	ANK3	-	13945
chr11	34475000	36525000	6	chr11	35639734	35642421	ENSG00000179431	FJX1	+	2687
chr11	34475000	36525000	6	chr11	35835881	35839556	ENSG00000261355	DQ573949	+	3675
chr11	34475000	36525000	6	chr11	35887569	35887666	n/a	Mir_652	n/a	97
chr11	34475000	36525000	6	chr11	36031647	36031754	ENSG00000263389	MIR3973	+	107
chr11	34475000	36525000	6	chr11	34500341	34535347	ENSG00000135374	ELF5	-	35006
chr11	34475000	36525000	6	chr11	34654010	34684834	ENSG00000135373	EHF	+	30824
chr11	34475000	36525000	6	chr11	34903842	34937939	ENSG00000149089	APIP	-	34097
chr11	34475000	36525000	6	chr11	34963383	34963467	ENSG00000251862	MIR1343	+	84
chr11	34475000	36525000	6	chr11	35160416	35253949	ENSG00000026508	CD44	+	93533
chr11	34475000	36525000	6	chr11	36084833	36099448	n/a	BC036209	n/a	14615
chr11	34475000	36525000	6	chr11	36505316	36531863	ENSG00000175104	TRAF6	-	19684
chr11	34475000	36525000	6	chr11	34460471	34493607	ENSG00000121691	CAT	+	18607
chr11	34475000	36525000	6	chr11	34937676	35017675	ENSG00000110435	PDHX	+	79999
chr11	34475000	36525000	6	chr11	35272751	35441105	ENSG00000110436	SLC1A2	-	168354
chr11	34475000	36525000	6	chr11	35453375	35547176	ENSG00000149090	PAMR1	-	93801
chr11	34475000	36525000	6	chr11	35684352	35830930	ENSG00000166326	TRIM44	+	146578
chr11	34475000	36525000	6	chr11	35965611	36252841	ENSG00000179241	LDLRAD3	+	287230
chr11	34475000	36525000	6	chr11	36293841	36310999	ENSG00000110442	COMMD9	-	17158
chr11	34475000	36525000	6	chr11	36397534	36486754	ENSG00000135362	PRR5L	+	89220
chr12	31475000	32850000	6	chr12	31477249	31478879	ENSG00000177340	FLJ13224	+	1630
chr12	31475000	32850000	6	chr12	31516414	31522235	ENSG00000256232	BC039477	-	5821
chr12	31475000	32850000	6	chr12	31747026	31747140	ENSG00000252390	U5	-	114
chr12	31475000	32850000	6	chr12	31944118	31945175	ENSG00000188375	H3F3C	-	1057
chr12	31475000	32850000	6	chr12	32146772	32146802	n/a	RNU6-78P	n/a	30
chr12	31475000	32850000	6	chr12	31742856	31768285	ENSG00000255867	DENND5B-AS1	+	25429
chr12	31475000	32850000	6	chr12	31812608	31822016	ENSG00000139160	METTL20	+	9408
chr12	31475000	32850000	6	chr12	32832136	32898584	ENSG00000087470	DNM1L	+	17864
chr12	31475000	32850000	6	chr12	31535156	31743952	ENSG00000170456	DENND5B	-	208796
chr12	31475000	32850000	6	chr12	31824070	31882108	ENSG00000151743	AMN1	-	58038
chr12	31475000	32850000	6	chr12	32112352	32146043	ENSG00000174718	KIAA1551	+	33691
chr12	31475000	32850000	6	chr12	32655040	32798984	ENSG00000139132	FGD4	+	143944
chr12	31475000	32850000	6	chr12	31433519	31479159	ENSG00000139146	FAM60A	-	4159
chr12	31475000	32850000	6	chr12	32260184	32531141	ENSG00000151746	BICD1	+	270957
chr12	58300000	62625000	6	chr12	59762831	59762932	ENSG00000212599	U6	-	101
chr12	58300000	62625000	6	chr12	58335444	58351052	ENSG00000166896	XRCC6BP1	+	15608
chr12	58300000	62625000	6	chr12	59265936	59314319	ENSG00000139263	LRIG3	-	48383
chr12	58300000	62625000	6	chr12	58325231	58329947	ENSG00000257698	LOC100506844	-	4716
chr12	58300000	62625000	6	chr12	58959742	59175498	ENSG00000257259	AK093124	-	215756
chr12	58300000	62625000	6	chr12	59989820	60183635	ENSG00000118596	SLC16A7	+	193815
chr12	58300000	62625000	6	chr12	62102028	62586620	ENSG00000198673	FAM19A2	-	484592
chr13	30025000	31700000	6	chr13	30050651	30061887	ENSG00000179141	MTUS2-AS1	-	11236
chr13	30025000	31700000	6	chr13	31032878	31040081	ENSG00000189403	HMGB1	-	7203
chr13	30025000	31700000	6	chr13	31248100	31248174	n/a	TRNA_Asn	n/a	74
chr13	30025000	31700000	6	chr13	31377342	31384782	ENSG00000237879	LINC00398	+	7440
chr13	30025000	31700000	6	chr13	30446749	30462542	ENSG00000224329	LINC00297	-	15793

chr13	30025000	31700000	6	chr13	30510667	30524625	ENSG00000122043	LINC00544	+	13958
chr13	30025000	31700000	6	chr13	31480311	31499709	ENSG00000102802	MEDAG	+	19398
chr13	30025000	31700000	6	chr13	31506833	31549153	ENSG00000175664	TEX26	+	42320
chr13	30025000	31700000	6	chr13	29598747	30080084	ENSG00000132938	MTUS2	+	55084
chr13	30025000	31700000	6	chr13	30083550	30169825	ENSG00000139514	SLC7A1	-	86275
chr13	30025000	31700000	6	chr13	30776766	30881191	ENSG00000102781	KATNAL1	-	104425
chr13	30025000	31700000	6	chr13	30914406	30948036	ENSG00000238121	LINC00426	-	33630
chr13	30025000	31700000	6	chr13	31191829	31233686	ENSG00000132952	USPL1	+	41857
chr13	30025000	31700000	6	chr13	31287614	31338565	ENSG00000132965	ALOX5AP	+	50951
chr13	30025000	31700000	6	chr13	30338544	30424820	ENSG00000122042	UBL3	-	86276
chr13	30025000	31700000	6	chr13	31456971	31506745	ENSG00000224743	TEX26-AS1	-	49774

Table 3.7 | TADs expanded in neurons compared to glia.
TAD calls in neurons that encompass multiple TADs in glia (Table 3.7) and NPCs (Table 3.8). Chr(Neu TAD), x1 (Neu TAD), x2 (Neu TAD) – coordinates for TADs in neurons; No. Overlap Glia TADs, number of TADs in glia that overlap with the TAD called in neurons (columns A-C). Chr(Glia/NPC gene), x1 (Glia/NPC gene), x2 (Glia/NPC gene) – coordinates for genes located in TADs in column D in glia (Table 3.7) and NPCs (Table 3.8), respectively. Gene ID (ENSG), ENSEMBL gene IDs.

chr(Neu TAD)	x1(Neu TAD)	x2(Neu TAD)	No. Overlap NPC TADs	chr (NPC gene)	y1 (NPC Gene)	y2 (NPC Gene)	Gene ID (ENSG)	Gene Symbol	Strand	Gene length (bp)
chr10	60100000	62425000	9	chr10	60474774	60477293	ENSG00000183055	FAM133CP	+	2519
chr10	60100000	62425000	9	chr10	61019597	61019717	n/a	Mir_584	n/a	120
chr10	60100000	62425000	9	chr10	61821742	61822047	n/a	Metazoa_SRP	n/a	305
chr10	60100000	62425000	9	chr10	61841498	61841612	ENSG00000202190	Y_RNA	+	114
chr10	60100000	62425000	9	chr10	60094738	60130513	ENSG00000072401	UBE2D1	+	30513
chr10	60100000	62425000	9	chr10	60144902	60158990	ENSG00000108064	TFAM	+	14088
chr10	60100000	62425000	9	chr10	61410521	61469649	ENSG00000165449	SLC16A9	-	59128
chr10	60100000	62425000	9	chr10	61496747	61513203	ENSG00000227877	LINC00948	-	16456
chr10	60100000	62425000	9	chr10	61715187	61720671	ENSG00000235931	C10orf40	-	5484
chr10	60100000	62425000	9	chr10	60272903	60588845	ENSG00000122870	BICC1	+	315942
chr10	60100000	62425000	9	chr10	60936347	61007534	ENSG00000165443	PHYHIPL	+	71187
chr10	60100000	62425000	9	chr10	61005888	61122352	ENSG00000148541	FAM13C	-	116464
chr10	60100000	62425000	9	chr10	61548505	61666414	ENSG00000108091	CCDC6	-	117909
chr10	60100000	62425000	9	chr10	61786055	62149634	ENSG00000151150	ANK3	-	363579
chr6	20525000	24375000	8	chr6	20567407	20567491	n/a	Mir_548	n/a	84
chr6	20525000	24375000	8	chr6	22134830	22147422	ENSG00000260455	LOC729177	-	12592
chr6	20525000	24375000	8	chr6	22569677	22570750	ENSG00000112273	HDGFL1	+	1073
chr6	20525000	24375000	8	chr6	24357130	24358512	ENSG00000146049	KAAG1	+	1382
chr6	20525000	24375000	8	chr6	21593971	21598849	ENSG00000124766	SOX4	+	4878
chr6	20525000	24375000	8	chr6	22287472	22303082	ENSG00000172179	PRL	-	15610
chr6	20525000	24375000	8	chr6	24126413	24147757	ENSG00000152954	NRSN1	+	21344
chr6	20525000	24375000	8	chr6	22146882	22194616	ENSG00000272168	LINC00340	+	47734
chr6	20525000	24375000	8	chr6	24171982	24358280	ENSG00000146038	DCDC2	-	186298
chr6	20525000	24375000	8	chr6	20534687	21232634	ENSG00000145996	CDKAL1	+	697947
chr6	20525000	24375000	8	chr6	21666674	22194616	ENSG00000272168	LINC00340	+	527942
chr9	6775000	9925000	8	chr9	6834701	6835011	ENSG00000230581	DQ580140	+	310
chr9	6775000	9925000	8	chr9	7167235	7167357	n/a	Mir_584	n/a	122
chr9	6775000	9925000	8	chr9	8858133	8861724	ENSG00000225706	AK094342	+	3591
chr9	6775000	9925000	8	chr9	8888135	8888207	ENSG00000251699	SNORD27	+	72
chr9	6775000	9925000	8	chr9	9442059	9442347	ENSG00000265735	Metazoa_SRP	+	288
chr9	6775000	9925000	8	chr9	7796490	7799799	ENSG00000137038	C9orf123	-	3309
chr9	6775000	9925000	8	chr9	6757640	7175648	ENSG00000107077	KDM4C	+	400648
chr9	6775000	9925000	8	chr9	8314245	10612723	ENSG00000153707	PTPRD	-	1610755
chr10	60100000	61800000	7	chr10	60474774	60477293	ENSG00000183055	FAM133CP	+	2519
chr10	60100000	61800000	7	chr10	61019597	61019717	n/a	Mir_584	n/a	120
chr10	60100000	61800000	7	chr10	60094738	60130513	ENSG00000072401	UBE2D1	+	30513
chr10	60100000	61800000	7	chr10	60144902	60158990	ENSG00000108064	TFAM	+	14088
chr10	60100000	61800000	7	chr10	61410521	61469649	ENSG00000165449	SLC16A9	-	59128
chr10	60100000	61800000	7	chr10	61496747	61513203	ENSG00000227877	LINC00948	-	16456
chr10	60100000	61800000	7	chr10	61715187	61720671	ENSG00000235931	C10orf40	-	5484
chr10	60100000	61800000	7	chr10	60272903	60588845	ENSG00000122870	BICC1	+	315942
chr10	60100000	61800000	7	chr10	60936347	61007534	ENSG00000165443	PHYHIPL	+	71187
chr10	60100000	61800000	7	chr10	61005888	61122352	ENSG00000148541	FAM13C	-	116464
chr10	60100000	61800000	7	chr10	61548505	61666414	ENSG00000108091	CCDC6	-	117909
chr10	60100000	61800000	7	chr10	61786055	62149634	ENSG00000151150	ANK3	-	13945
chr4	52675000	55125000	7	chr4	53578620	53580305	ENSG00000226950	DANCR	+	1685
chr4	52675000	55125000	7	chr4	53578848	53578914	ENSG00000264585	MIR4449	+	66
chr4	52675000	55125000	7	chr4	53579415	53579537	ENSG00000212588	SNORA26	+	122
chr4	52675000	55125000	7	chr4	53609683	53617807	ENSG00000226887	ERVMER34-1	-	8124
chr4	52675000	55125000	7	chr4	53728494	53733002	ENSG00000128045	RASL11B	+	4508
chr4	52675000	55125000	7	chr4	54851665	54853449	ENSG00000229585	RPL21P44	-	1784
chr4	52675000	55125000	7	chr4	54966247	54968122	ENSG00000180613	GSX2	+	1875
chr4	52675000	55125000	7	chr4	52709275	52783003	ENSG00000109184	DCUN1D4	+	73728
chr4	52675000	55125000	7	chr4	52859865	52883786	ENSG00000188993	LRRC66	-	23921
chr4	52675000	55125000	7	chr4	52886860	52904485	ENSG00000163069	SGCB	-	17625
chr4	52675000	55125000	7	chr4	53656160	53681631	ENSG00000250302	LOC152578	+	25471
chr4	52675000	55125000	7	chr4	53811820	53824254	ENSG00000248115	AK055055	+	12434
chr4	52675000	55125000	7	chr4	54294193	54326103	ENSG00000145216	FIP1L1	+	31910
chr4	52675000	55125000	7	chr4	54366243	54389443	ENSG00000250930	LNX1-AS1	+	23200
chr4	52675000	55125000	7	chr4	54459122	54471548	ENSG00000248494	LNX1-AS2	+	12426
chr4	52675000	55125000	7	chr4	55095263	55164412	ENSG00000134853	PDGFRA	+	29737
chr4	52675000	55125000	7	chr4	52917592	52963458	ENSG00000163071	SPATA18	+	45866
chr4	52675000	55125000	7	chr4	53739150	54232242	ENSG00000184178	SCFD2	-	493092
chr4	52675000	55125000	7	chr4	54326436	54457753	ENSG00000072201	LNX1	-	131317
chr4	52675000	55125000	7	chr4	54875957	54930815	ENSG00000109220	CHIC2	-	54858
chr4	52675000	55125000	7	chr4	53457126	53525502	ENSG00000109189	USP46	-	68376
chr5	162950000	167650000	7	chr5	165036445	165036684	ENSG00000252794	7SK	+	239
chr5	162950000	167650000	7	chr5	167460095	167460210	ENSG00000253065	SNORA40	-	115
chr5	162950000	167650000	7	chr5	163781145	164029423	ENSG00000241956	BC011998	+	248278
chr5	162950000	167650000	7	chr5	166711842	167691162	ENSG00000145934	TENM2	+	938158
chr9	97300000	99100000	7	chr9	97572243	97572339	ENSG00000252153	MIR2278	+	96
chr9	97300000	99100000	7	chr9	97781172	97784562	ENSG00000148120	AL137535	+	3390
chr9	97300000	99100000	7	chr9	97847489	97848370	ENSG00000207563	MIR23B	+	881
chr9	97300000	99100000	7	chr9	97848295	97848376	ENSG00000207617	MIR3074	-	81
chr9	97300000	99100000	7	chr9	97884155	97886342	n/a	BC041030	n/a	2187
chr9	97300000	99100000	7	chr9	98225890	98232301	n/a	LOC100507346	n/a	6411
chr9	97300000	99100000	7	chr9	98680445	98680471	ENSG00000237631	DQ588075	+	26
chr9	97300000	99100000	7	chr9	98782013	98784037	ENSG00000225194	LINC00092	-	2024
chr9	97300000	99100000	7	chr9	97317422	97330409	ENSG00000231806	BC080653	+	12987
chr9	97300000	99100000	7	chr9	97321002	97356075	ENSG00000130957	FBP2	-	35073
chr9	97300000	99100000	7	chr9	98205263	98270831	ENSG00000185920	PTCH1	-	65568
chr9	97300000	99100000	7	chr9	98568369	98638259	ENSG00000175611	LINC00476	-	69890
chr9	97300000	99100000	7	chr9	98868942	98878693	n/a	LOC158434	n/a	9751
chr9	97300000	99100000	7	chr9	98997588	99064434	ENSG00000130948	HSD17B3	-	66846
chr9	97300000	99100000	7	chr9	97365420	97401923	ENSG00000165140	FBP1	-	36503

chr9	97300000	99100000	7	chr9	97521930	97849500	ENSG00000148120	C9orf3	+	327570
chr9	97300000	99100000	7	chr9	97861335	98079991	ENSG00000158169	FANCC	-	218656
chr9	97300000	99100000	7	chr9	98637899	98731122	ENSG00000182150	ERCC6L2	+	93223
chr9	97300000	99100000	7	chr9	98828120	98864194	ENSG00000237212	LOC158435	+	36074
chr9	97300000	99100000	7	chr9	99082987	99145992	ENSG00000130958	SLC35D2	-	17013
chr1	71550000	75175000	6	chr1	72259914	72302695	ENSG00000228853	NEGR1-IT1	-	42781
chr1	71550000	75175000	6	chr1	75043113	75091782	ENSG00000234497	CR627203	+	48669
chr1	71550000	75175000	6	chr1	75171171	75199092	ENSG00000116791	CRYZ	-	3829
chr1	71550000	75175000	6	chr1	71547006	71703406	ENSG00000229956	ZRANB2-AS2	+	153406
chr1	71550000	75175000	6	chr1	73771852	73804560	ENSG00000233973	BC041341	+	32708
chr1	71550000	75175000	6	chr1	74491701	74663871	ENSG00000162620	LRRIQ3	-	172170
chr1	71550000	75175000	6	chr1	74663895	75010116	ENSG00000116783	FPGT-TNNI3K	+	346221
chr1	71550000	75175000	6	chr1	75033794	75139422	ENSG00000178965	C1orf173	-	105628
chr1	71550000	75175000	6	chr1	71868624	72748405	ENSG00000172260	NEGR1	-	879781
chr1	95650000	98950000	6	chr1	95699710	95712781	ENSG00000122481	RWDD3	+	13071
chr1	95650000	98950000	6	chr1	95970531	95970615	ENSG00000216037	Mir_548	-	84
chr1	95650000	98950000	6	chr1	97161411	97161723	ENSG00000223229	7SK	-	312
chr1	95650000	98950000	6	chr1	98510798	98510907	ENSG00000225206	MIR2682	-	109
chr1	95650000	98950000	6	chr1	95582893	95663161	ENSG00000152078	TMEM56	+	13161
chr1	95650000	98950000	6	chr1	95975895	95981020	ENSG00000228971	BC067883	+	5125
chr1	95650000	98950000	6	chr1	98453555	98515249	ENSG00000225206	MIR137HG	-	61694
chr1	95650000	98950000	6	chr1	95628774	95699538	ENSG00000226026	AK090700	-	49538
chr1	95650000	98950000	6	chr1	95940292	95944912	ENSG00000233907	FLJ31662	+	4620
chr1	95650000	98950000	6	chr1	97187174	97280605	ENSG00000117569	PTBP2	+	93431
chr1	95650000	98950000	6	chr1	97543299	98386615	ENSG00000188641	DPYD	-	843316
chr1	95650000	98950000	6	chr1	97561478	97788511	ENSG00000232878	DPYD-AS1	+	227033
chr1	95650000	98950000	6	chr1	98676266	98738214	ENSG00000226053	LOC729987	+	61948
chr10	60225000	61375000	6	chr10	60474774	60477293	ENSG00000183055	FAM133CP	+	2519
chr10	60225000	61375000	6	chr10	61019597	61019717	n/a	Mir_584	n/a	120
chr10	60225000	61375000	6	chr10	60272903	60588845	ENSG00000122870	BICC1	+	315942
chr10	60225000	61375000	6	chr10	60936347	61007534	ENSG00000165443	PHYHIPL	+	71187
chr10	60225000	61375000	6	chr10	61005888	61122352	ENSG00000148541	FAM13C	-	116464
chr10	60475000	61375000	6	chr10	60474774	60477293	ENSG00000183055	FAM133CP	+	2293
chr10	60475000	61375000	6	chr10	61019597	61019717	n/a	Mir_584	n/a	120
chr10	60475000	61375000	6	chr10	60272903	60588845	ENSG00000122870	BICC1	+	113845
chr10	60475000	61375000	6	chr10	60936347	61007534	ENSG00000165443	PHYHIPL	+	71187
chr10	60475000	61375000	6	chr10	61005888	61122352	ENSG00000148541	FAM13C	-	116464
chr12	69725000	72125000	6	chr12	69978501	69978603	ENSG00000264405	MIR3913-1	-	102
chr12	69725000	72125000	6	chr12	69978502	69978602	n/a	MIR3913-2	n/a	100
chr12	69725000	72125000	6	chr12	70002344	70004942	ENSG00000198812	LRRC10	-	2598
chr12	69725000	72125000	6	chr12	70195707	70195791	n/a	Mir_548	n/a	84
chr12	69725000	72125000	6	chr12	71300696	71300798	ENSG00000207387	Y_RNA	+	102
chr12	69725000	72125000	6	chr12	71869656	71869910	n/a	7SK	n/a	254
chr12	69725000	72125000	6	chr12	69742133	69748013	ENSG00000090382	LYZ	+	5880
chr12	69725000	72125000	6	chr12	69753531	69784576	ENSG00000127337	YEATS4	+	31045
chr12	69725000	72125000	6	chr12	69864128	69973562	ENSG00000166225	FRS2	+	109434
chr12	69725000	72125000	6	chr12	70047388	70093196	ENSG00000127325	BEST3	-	45808
chr12	69725000	72125000	6	chr12	70132630	70216984	ENSG00000127328	RAB3IP	+	84354
chr12	69725000	72125000	6	chr12	70320436	70352503	ENSG00000166268	C12orf28	+	32067
chr12	69725000	72125000	6	chr12	70910631	71031220	ENSG00000127329	PTPRB	-	120589
chr12	69725000	72125000	6	chr12	71518876	71551779	ENSG00000127324	TSPAN8	-	32903
chr12	69725000	72125000	6	chr12	72003378	72057749	ENSG00000133858	ZFC3H1	-	54371
chr12	69725000	72125000	6	chr12	72057676	72074428	ENSG00000173451	THAP2	+	16752
chr12	69725000	72125000	6	chr12	69979207	69995357	ENSG00000166226	CCT2	+	16150
chr12	69725000	72125000	6	chr12	70107414	70132348	ENSG00000247131	BC042465	-	24934
chr12	69725000	72125000	6	chr12	70636773	70748773	ENSG00000111596	CNOT2	+	112000
chr12	69725000	72125000	6	chr12	70760061	70828072	ENSG00000135643	KCNMB4	+	68011
chr12	69725000	72125000	6	chr12	70861864	70932592	ENSG00000258168	BC031864	+	70728
chr12	69725000	72125000	6	chr12	71833549	71980088	ENSG00000139292	LGR5	+	146539
chr12	69725000	72125000	6	chr12	72079877	72097839	ENSG00000139291	TMEM19	+	17962
chr12	69725000	72125000	6	chr12	71031852	71314584	ENSG00000153233	PTPRR	-	282732
chr12	96925000	98800000	6	chr12	97490690	97490764	n/a	TRNA_Pseudo	n/a	74
chr12	96925000	98800000	6	chr12	97506929	97507042	n/a	Mir_584	n/a	113
chr12	96925000	98800000	6	chr12	97866326	97868402	n/a	AK129935	n/a	2076
chr12	96925000	98800000	6	chr12	97885686	97885756	ENSG00000221479	MIR1251	+	70
chr12	96925000	98800000	6	chr12	97945914	97946006	ENSG00000251844	SnoMe28S_Am2634	-	92
chr12	96925000	98800000	6	chr12	97957589	97957689	ENSG00000207586	MIR135A2	+	100
chr12	96925000	98800000	6	chr12	98389160	98389226	ENSG00000263890	MIR4303	-	66
chr12	96925000	98800000	6	chr12	98752079	98752163	n/a	Mir_548	n/a	84
chr12	96925000	98800000	6	chr12	96883352	96934765	ENSG00000188596	C12orf55	+	9765
chr12	96925000	98800000	6	chr12	97301000	97347469	ENSG00000139350	NEDD1	+	46469
chr12	96925000	98800000	6	chr12	98123825	98150295	ENSG00000257501	LOC643711	-	26470
chr12	96925000	98800000	6	chr12	96933914	97024429	ENSG00000188596	AX747187	+	90515
chr12	96925000	98800000	6	chr12	97041752	97159032	ENSG00000188596	C12orf63	+	117280
chr12	96925000	98800000	6	chr12	97858798	97927544	ENSG00000255794	RMST	+	68746
chr14	25850000	29100000	6	chr14	26641372	26641456	ENSG00000212037	Mir_548	-	84
chr14	25850000	29100000	6	chr14	27377847	27377931	ENSG00000265165	MIR4307	+	84
chr14	25850000	29100000	6	chr14	29063026	29063048	ENSG00000257662	DD413682	+	22
chr14	25850000	29100000	6	chr14	28081793	28108842	ENSG00000258548	LINC00645	+	27049
chr14	25850000	29100000	6	chr14	26915088	27066960	ENSG00000139910	NOVA1	-	151872
chr14	25850000	29100000	6	chr14	27791205	28142425	ENSG00000258932	BC148262	-	351220
chr16	78125000	80550000	6	chr16	79627744	79634622	ENSG00000178573	MAF	-	6878
chr16	78125000	80550000	6	chr16	78133326	79246564	ENSG00000186153	WWOX	+	1113238
chr18	49100000	51725000	6	chr18	51105992	51107426	ENSG00000260433	BC034434	-	1434
chr18	49100000	51725000	6	chr18	51677970	51751158	ENSG00000134046	MBD2	-	47030
chr18	49100000	51725000	6	chr18	49866541	51062273	ENSG00000187323	DCC	+	1195732
chr18	49350000	51725000	6	chr18	51105992	51107426	ENSG00000260433	BC034434	-	1434
chr18	49350000	51725000	6	chr18	51677970	51751158	ENSG00000134046	MBD2	-	47030
chr18	49350000	51725000	6	chr18	49866541	51062273	ENSG00000187323	DCC	+	1195732
chr2	103425000	105550000	6	chr2	103486031	103486051	n/a	DD413674	n/a	20
chr2	103425000	105550000	6	chr2	105471968	105473471	ENSG00000198914	POU3F3	+	1503
chr2	103425000	105550000	6	chr2	105481954	105488840	ENSG00000229743	AK095498	-	6886
chr2	103425000	105550000	6	chr2	104995307	105024790	n/a	LOC100287010	n/a	29483
chr2	103425000	105550000	6	chr2	105363094	105374177	ENSG00000234177	LOC284998	-	11083
chr2	103425000	105550000	6	chr2	105421882	105467934	ENSG00000233639	LOC100506421	-	46052
chr2	103425000	105550000	6	chr2	103378489	103434138	ENSG00000170417	TMEM182	+	9138

chr2	103425000	105550000	6	chr2	105050804	105129215	ENSG00000235597	LOC150568	+	78411
chr2	188350000	190325000	6	chr2	189162218	189162315	ENSG00000207951	MIR561	+	97
chr2	188350000	190325000	6	chr2	189842817	189842886	ENSG00000221502	MIR1245A	+	69
chr2	188350000	190325000	6	chr2	189842819	189842887	n/a	MIR1245B	n/a	68
chr2	188350000	190325000	6	chr2	189860355	189860418	n/a	MIR3606	n/a	63
chr2	188350000	190325000	6	chr2	189997761	189997837	ENSG00000264725	MIR3129	-	76
chr2	188350000	190325000	6	chr2	189598464	189654831	ENSG00000174325	DIRC1	-	56367
chr2	188350000	190325000	6	chr2	189839098	189877472	ENSG00000168542	COL3A1	+	38374
chr2	188350000	190325000	6	chr2	188328957	188419219	ENSG00000003436	TFPI	-	69219
chr2	188350000	190325000	6	chr2	189156395	189460652	ENSG00000144366	GULP1	+	304257
chr2	188350000	190325000	6	chr2	189896640	190044605	ENSG00000204262	COL5A2	-	147965
chr2	188350000	190325000	6	chr2	190306158	190340264	ENSG00000115368	WDR75	+	18842
chr2	33875000	37175000	6	chr2	33951127	33953284	ENSG00000239649	MYADML	-	2157
chr2	33875000	37175000	6	chr2	34628738	34628822	ENSG00000212025	Mir_548	-	84
chr2	33875000	37175000	6	chr2	36581891	36582713	n/a	LOC100288911	n/a	822
chr2	33875000	37175000	6	chr2	36779403	36825332	ENSG00000171055	FEZ2	-	45929
chr2	33875000	37175000	6	chr2	36923832	37041937	ENSG00000205221	VIT	+	118105
chr2	33875000	37175000	6	chr2	37064840	37193615	ENSG00000115808	STRN	-	110160
chr2	33875000	37175000	6	chr2	36583369	36778278	ENSG00000150938	CRIM1	+	194909
chr2	39550000	42025000	6	chr2	39620662	39620759	ENSG00000252239	U6	-	97
chr2	39550000	42025000	6	chr2	41559792	41559915	n/a	Mir_584	n/a	123
chr2	39550000	42025000	6	chr2	39741286	39828484	ENSG00000231312	LOC728730	+	87198
chr2	39550000	42025000	6	chr2	39893034	39945104	ENSG00000152154	TMEM178A	+	52070
chr2	39550000	42025000	6	chr2	39476406	39664453	ENSG00000011566	MAP4K3	-	114453
chr2	39550000	42025000	6	chr2	39963199	40006416	ENSG00000138050	THUMPD2	-	43217
chr2	39550000	42025000	6	chr2	40144773	40482349	ENSG00000227028	SLC8A1-AS1	+	337576
chr2	39550000	42025000	6	chr2	40339285	40657444	ENSG00000183023	SLC8A1	-	318159
chr2	62125000	64150000	6	chr2	62432960	62433052	ENSG00000266097	MIR5192	+	92
chr2	62125000	64150000	6	chr2	62489521	62489806	ENSG00000239958	Metazoa_SRP	-	285
chr2	62125000	64150000	6	chr2	62727355	62733604	ENSG00000186889	TMEM17	-	6249
chr2	62125000	64150000	6	chr2	62953770	62953875	ENSG00000252436	Y_RNA	-	105
chr2	62125000	64150000	6	chr2	63271099	63274846	ENSG00000231609	LOC100132215	-	3747
chr2	62125000	64150000	6	chr2	63277936	63284966	ENSG00000115507	OTX1	+	7030
chr2	62125000	64150000	6	chr2	63344985	63346677	ENSG00000242412	DBIL5P2	-	1692
chr2	62125000	64150000	6	chr2	62296587	62374016	ENSG00000229839	BC071802	-	77429
chr2	62125000	64150000	6	chr2	62423261	62451866	ENSG00000170340	B3GNT2	+	28605
chr2	62125000	64150000	6	chr2	62839006	62889763	ENSG00000226622	BC038779	-	50757
chr2	62125000	64150000	6	chr2	62132802	62363205	ENSG00000173163	COMMD1	+	230403
chr2	62125000	64150000	6	chr2	62933000	63273621	ENSG00000115504	EHBP1	+	340621
chr2	62125000	64150000	6	chr2	63348534	63815867	ENSG00000143951	WDPCP	-	467333
chr2	62125000	64150000	6	chr2	63816284	63834330	ENSG00000014641	MDH1	+	18046
chr2	62125000	64150000	6	chr2	64069013	64118696	ENSG00000169764	UGP2	+	49683
chr2	62125000	64150000	6	chr2	64119666	64246214	ENSG00000143952	VPS54	-	30334
chr21	30400000	32450000	6	chr21	30552745	30552960	ENSG00000212479	U3	-	215
chr21	30400000	32450000	6	chr21	31538240	31538971	ENSG00000156282	CLDN17	-	731
chr21	30400000	32450000	6	chr21	31581468	31584101	ENSG00000227342	LINC00307	-	2633
chr21	30400000	32450000	6	chr21	31661462	31661832	ENSG00000232263	KRTAP25-1	-	370
chr21	30400000	32450000	6	chr21	31691449	31692607	ENSG00000197683	KRTAP26-1	-	1158
chr21	30400000	32450000	6	chr21	31709330	31710012	ENSG00000206107	KRTAP27-1	-	682
chr21	30400000	32450000	6	chr21	31720716	31720924	ENSG00000186980	KRTAP23-1	-	208
chr21	30400000	32450000	6	chr21	31743708	31744557	ENSG00000182816	KRTAP13-2	-	849
chr21	30400000	32450000	6	chr21	31747611	31747696	ENSG00000265007	MIR4327	-	85
chr21	30400000	32450000	6	chr21	31797710	31798230	ENSG00000240432	KRTAP13-3	-	520
chr21	30400000	32450000	6	chr21	31802593	31803076	ENSG00000186971	KRTAP13-4	+	483
chr21	30400000	32450000	6	chr21	31812645	31813098	ENSG00000186970	KRTAP15-1	+	453
chr21	30400000	32450000	6	chr21	31852363	31852636	ENSG00000184351	KRTAP19-1	-	273
chr21	30400000	32450000	6	chr21	31859508	31859667	ENSG00000186965	KRTAP19-2	-	159
chr21	30400000	32450000	6	chr21	31863781	31864275	ENSG00000244025	KRTAP19-3	-	494
chr21	30400000	32450000	6	chr21	31869173	31869428	ENSG00000186967	KRTAP19-4	-	255
chr21	30400000	32450000	6	chr21	31874189	31874408	ENSG00000186977	KRTAP19-5	-	219
chr21	30400000	32450000	6	chr21	31913853	31914181	ENSG00000186925	KRTAP19-6	-	328
chr21	30400000	32450000	6	chr21	31933416	31933608	ENSG00000244362	KRTAP19-7	-	192
chr21	30400000	32450000	6	chr21	31962423	31962716	ENSG00000206106	KRTAP22-2	-	293
chr21	30400000	32450000	6	chr21	31971004	31971193	ENSG00000186930	KRTAP6-2	-	189
chr21	30400000	32450000	6	chr21	31973439	31973586	ENSG00000186924	KRTAP22-1	+	147
chr21	30400000	32450000	6	chr21	31986004	31986223	ENSG00000184724	KRTAP6-1	-	219
chr21	30400000	32450000	6	chr21	31988773	31988944	ENSG00000244624	KRTAP20-1	+	171
chr21	30400000	32450000	6	chr21	31992945	31993169	ENSG00000206105	KRTAP20-4	+	224
chr21	30400000	32450000	6	chr21	32007582	32007780	ENSG00000184032	KRTAP20-2	+	198
chr21	30400000	32450000	6	chr21	32015182	32015455	ENSG00000206104	KRTAP20-3	+	273
chr21	30400000	32450000	6	chr21	32090842	32091095	ENSG00000231068	KRTAP21-3	-	253
chr21	30400000	32450000	6	chr21	32119268	32119520	ENSG00000187026	KRTAP21-2	-	252
chr21	30400000	32450000	6	chr21	32127456	32127696	ENSG00000187005	KRTAP21-1	-	240
chr21	30400000	32450000	6	chr21	32185014	32185570	ENSG00000183640	KRTAP8-1	-	556
chr21	30400000	32450000	6	chr21	32201357	32202051	ENSG00000256386	KRTAP7-1	-	694
chr21	30400000	32450000	6	chr21	32252963	32253874	ENSG00000182591	KRTAP11-1	-	911
chr21	30400000	32450000	6	chr21	32410477	32410795	ENSG00000206102	KRTAP19-8	-	318
chr21	30400000	32450000	6	chr21	30428647	30446010	ENSG00000156261	CCT8	-	17363
chr21	30400000	32450000	6	chr21	30565814	30660526	ENSG00000215533	LINC00189	+	94712
chr21	30400000	32450000	6	chr21	30671219	30718469	ENSG00000156273	BACH1	+	47250
chr21	30400000	32450000	6	chr21	30968359	31003067	ENSG00000183653	GRIK1-AS2	+	34708
chr21	30400000	32450000	6	chr21	31120493	31136325	ENSG00000174680	GRIK1-AS1	+	15832
chr21	30400000	32450000	6	chr21	31653626	31655276	ENSG00000188694	KRTAP24-1	-	1650
chr21	30400000	32450000	6	chr21	31768391	31769138	ENSG00000198390	KRTAP13-1	+	747
chr21	30400000	32450000	6	chr21	31964758	31965374	ENSG00000212938	KRTAP6-3	+	616
chr21	30400000	32450000	6	chr21	30452872	30548202	ENSG00000156265	MAP3K7CL	+	95330
chr21	30400000	32450000	6	chr21	30925865	31312282	ENSG00000171189	GRIK1	-	386417
chr21	30400000	32450000	6	chr21	31586323	31588469	ENSG00000156284	CLDN8	-	2146
chr21	30400000	32450000	6	chr21	30396937	30426807	ENSG00000156256	USP16	+	26807
chr3	180425000	182400000	6	chr3	180586548	180588578	ENSG00000114416	BC034416	+	2030
chr3	180425000	182400000	6	chr3	180701497	180707562	ENSG00000205981	DNAJC19	-	6065
chr3	180425000	182400000	6	chr3	180949524	180949631	ENSG00000206932	U6	+	107
chr3	180425000	182400000	6	chr3	181413636	181413690	n/a	JA611300	n/a	54
chr3	180425000	182400000	6	chr3	181429711	181432223	ENSG00000181449	SOX2	+	2512
chr3	180425000	182400000	6	chr3	181686246	181686469	ENSG00000252257	7SK	+	223
chr3	180425000	182400000	6	chr3	181723101	181725535	n/a	DQ571917	n/a	2434
chr3	180425000	182400000	6	chr3	180630233	180700539	ENSG00000114416	FXR1	+	70306

chr3	180425000	182400000	6	chr3	181670166	181721828	ENSG00000242512	BC036236	+	51662
chr3	180425000	182400000	6	chr3	180425376	180587966	ENSG00000145075	DKFZp434A128	-	162590
chr3	180425000	182400000	6	chr3	182164757	182204150	ENSG00000241098	FLJ46066	-	39393
chr3	180425000	182400000	6	chr3	180774467	181460013	ENSG00000242808	SOX2-OT	+	685546
chr3	58525000	62275000	6	chr3	59956575	59958982	n/a	NPCR	n/a	2407
chr3	58525000	62275000	6	chr3	60842060	60842277	ENSG00000212211	U3	-	217
chr3	58525000	62275000	6	chr3	61068514	61068637	n/a	5S_rRNA	n/a	123
chr3	58525000	62275000	6	chr3	58549844	58563491	ENSG00000168309	FAM107A	-	13647
chr3	58525000	62275000	6	chr3	58619669	58652561	ENSG00000198643	FAM3D	-	32892
chr3	58525000	62275000	6	chr3	58727736	59035715	ENSG00000163689	C3orf67	-	307979
chr3	58525000	62275000	6	chr3	58810196	59004819	ENSG00000242428	AK090895	+	194623
chr3	58525000	62275000	6	chr3	62247493	62304622	ENSG00000241472	PTPRG-AS1	-	27507
chr3	58525000	62275000	6	chr3	59735035	61237133	ENSG00000189283	FHIT	-	1502098
chr3	58525000	62275000	6	chr3	61547242	62280573	ENSG00000144724	PTPRG	+	727758

Table 3.8 | TADs expanded in neurons compared to NPC.
TAD calls in neurons that encompass multiple TADs in glia (Table 3.7) and NPCs (Table 3.8). Chr(Neu TAD), x1 (Neu TAD), x2 (Neu TAD) – coordinates for TADs in neurons; No. Overlap NPC TADs, number of TADs in NPC that overlap with the TAD called in neurons (columns A-C). Chr(Glia/NPC gene), x1 (Glia/NPC gene), x2 (Glia/NPC gene) – coordinates for genes located in TADs in column D in glia (Table 3.7) and NPCs (Table 3.8), respectively. Gene ID (ENSG), ENSEMBL gene IDs.

NEURONS												
Riskchr	Risk start	Risk end	Risk overlap	chr1	x1	x2	chr2	y1	y2	GliafdrDonut	NeufdrDonut	NPCfdrDonut
chr10	104570118	105059896	5000	chr10	104590000	104595000	chr10	104660000	104665000	0.4254983	0.06027781	0.14398974
chr10	104570118	105059896	5000	chr10	104590000	104595000	chr10	104660000	104665000	0.4254983	0.06027781	0.14398974
chr18	52747689	53804156	5000	chr18	53075000	53080000	chr18	54240000	54245000	1.2518088	0.05274656	0.44168428
chr2	200547937	201309547	10000	chr2	199180000	199190000	chr2	200720000	200730000	0.51119745	2.03E-08	0.30081078
chr2	200547937	201309547	10000	chr2	200320000	200330000	chr2	200720000	200730000	1.5162903	0.00013409	1.5750872
chr22	39840130	40091818	10000	chr22	39720000	39730000	chr22	39910000	39920000	0.7976765	6.04E-05	0.8887403
chr4	176717618	176904037	10000	chr4	176480000	176490000	chr4	176730000	176740000	1.0397903	6.22E-05	1.0504069
chr5	87676693	88195380	5000	chr5	88025000	88030000	chr5	88595000	88600000	1.1398252	0	1.3084083
chr5	88580998	88748452	5000	chr5	88025000	88030000	chr5	88595000	88600000	1.1398252	0	1.3084083
chr5	137838122	137948140	10000	chr5	137930000	137940000	chr5	138200000	138210000	1.2840025	1.70E-05	1.7241881
chr6	24988105	33842877	10000	chr6	25260000	25270000	chr6	25590000	25600000	0.20888025	0.05431648	0.40973878
chr6	24988105	33842877	10000	chr6	25260000	25270000	chr6	25590000	25600000	0.20888025	0.05431648	0.40973878
chr6	24988105	33842877	5000	chr6	26120000	26125000	chr6	27110000	27115000	1.0039818	0	1.2458384
chr6	24988105	33842877	5000	chr6	26120000	26125000	chr6	27110000	27115000	1.0039818	0	1.2458384
chr6	24988105	33842877	10000	chr6	28460000	28470000	chr6	28630000	28640000	1.5588554	5.11E-05	0.98528725
chr6	24988105	33842877	10000	chr6	28460000	28470000	chr6	28630000	28640000	1.5588554	5.11E-05	0.98528725
chr7	110850439	111180544	10000	chr7	108260000	108270000	chr7	111120000	111130000	1.4305197	0.01244852	1.6058803
GLIA												
Riskchr	Risk start	Risk end	Risk overlap	chr1	x1	x2	chr2	y1	y2	GliafdrDonut	NeufdrDonut	NPCfdrDonut
chr1	8392592	8701288	10000	chr1	8430000	8440000	chr1	8770000	8780000	3.85E-06	0.44497925	0.7925938
chr1	98341152	98559093	9093	chr1	98550000	98560000	chr1	98660000	98670000	0.00097779	1.5657233	1.7241881
chr1	149998923	150214166	5000	chr1	149935000	149940000	chr1	150020000	150025000	2.52E-05	1.060373	0.8423837
chr1	200253612	200269903	9903	chr1	200260000	200270000	chr1	200450000	200460000	0.00100757	1.5745742	1.0504069
chr10	104570118	105059896	10000	chr10	104480000	104490000	chr10	104810000	104820000	0.00022482	0.90376765	0.3040932
chr10	104570118	105059896	5000	chr10	104485000	104490000	chr10	104665000	104670000	0.01798354	1.2801908	0.9664832
chr10	104570118	105059896	5000	chr10	104485000	104490000	chr10	104660000	104665000	6.75E-05	0.5831852	0.1140143
chr10	104570118	105059896	5000	chr10	104590000	104595000	chr10	104675000	104680000	0.00142198	1.3248246	1.3478131
chr10	104570118	105059896	5000	chr10	104590000	104595000	chr10	104675000	104680000	0.00142198	1.3248246	1.3478131
chr10	104570118	105059896	10000	chr10	104850000	104860000	chr10	105000000	105010000	0.00171292	1.5044254	0.85600626
chr10	104570118	105059896	10000	chr10	104850000	104860000	chr10	105000000	105010000	0.00171292	1.5044254	0.85600626
chr11	113317745	113451229	10000	chr11	113420000	113430000	chr11	113480000	113490000	0.00300015	1.1429281	1.0907023
chr12	123447928	123902361	10000	chr12	123520000	123530000	chr12	123710000	123720000	2.77E-08	1.508456	0.630625
chr12	123447928	123902361	10000	chr12	123520000	123530000	chr12	123710000	123720000	2.77E-08	1.508456	0.630625
chr13	79855297	80162555	10000	chr13	80130000	80140000	chr13	80930000	80940000	3.23E-05	0.90376765	0.17745346
chr14	30000405	30190316	5000	chr14	30015000	30020000	chr14	30680000	30685000	1.14E-06	1.363864	0.44168428
chr15	82827938	83391537	10000	chr15	83250000	83260000	chr15	83650000	83660000	0.00023944	1.5335083	1.2896612
chr15	84703470	85392298	5000	chr15	85130000	85135000	chr15	85585000	85590000	0.07324936	1.363864	0.24052724
chr16	29924422	30117253	5000	chr16	29710000	29715000	chr16	30030000	30035000	0.08700617	1.3652908	1.2216492
chr16	29924422	30117253	10000	chr16	30070000	30080000	chr16	30400000	30410000	0.00022482	1.5745742	1.688994
chr16	67708897	68305708	5000	chr16	67550000	67555000	chr16	67960000	67965000	0	1	1.3505517
chr16	67708897	68305708	5000	chr16	67850000	67855000	chr16	67960000	67965000	0.01798354	1.3262548	0.5672295
chr16	67708897	68305708	5000	chr16	67850000	67855000	chr16	67960000	67965000	0.01798354	1.3262548	0.5672295
chr16	67708897	68305708	10000	chr16	68120000	68130000	chr16	68360000	68370000	0.00817463	1.5335083	1.6341001
chr16	67708897	68305708	5000	chr16	68290000	68295000	chr16	68360000	68365000	0.02715562	1.0222301	0.21247724
chr17	17649172	17967397	10000	chr17	17810000	17820000	chr17	17930000	17940000	8.20E-07	1.1149824	0.52612877
chr17	17649172	17967397	10000	chr17	17810000	17820000	chr17	17930000	17940000	8.20E-07	1.1149824	0.52612877
chr17	17649172	17967397	5000	chr17	17825000	17830000	chr17	17870000	17875000	0.00082981	1.521011	0.96269846
chr17	17649172	17967397	5000	chr17	17825000	17830000	chr17	17870000	17875000	0.00082981	1.521011	0.96269846
chr18	52747689	53804156	5000	chr18	53170000	53175000	chr18	53205000	53210000	0.00299272	1.6471775	1.3091516
chr18	52747689	53804156	5000	chr18	53170000	53175000	chr18	53205000	53210000	0.00299272	1.6471775	1.3091516
chr19	19358332	19657632	5000	chr19	19255000	19260000	chr19	19600000	19605000	0	1.061395	1.3407764
chr2	57950104	58484172	10000	chr2	58130000	58140000	chr2	58460000	58470000	0.05253198	1.5455453	1.6341001
chr2	57950104	58484172	10000	chr2	58130000	58140000	chr2	58460000	58470000	0.05253198	1.5455453	1.6341001
chr2	57950104	58484172	10000	chr2	58350000	58360000	chr2	58670000	58680000	2.43E-08	1.5479501	0.5771854
chr2	198146381	198940251	5000	chr2	197975000	197980000	chr2	198245000	198250000	9.33E-05	1.3912318	0.1050422
chr2	198146381	198940251	5000	chr2	198055000	198060000	chr2	198175000	198180000	0.06144834	0.9483586	0.7659539
chr2	198146381	198940251	5000	chr2	198055000	198060000	chr2	198175000	198180000	0.06144834	0.9483586	0.7659539
chr2	200547937	201309547	5000	chr2	200835000	200840000	chr2	201560000	201565000	2.32E-05	0.110981	0.6510733
chr22	39840130	40091818	10000	chr22	39800000	39810000	chr22	39890000	39900000	0.0029721	1.1993746	1.4838825
chr22	42315790	42689370	10000	chr22	42360000	42370000	chr22	42480000	42490000	0.02962909	1.5869178	0.1819355
chr22	42315790	42689370	10000	chr22	42360000	42370000	chr22	42480000	42490000	0.02962909	1.5869178	0.1819355
chr3	17221017	17888256	5000	chr3	17300000	17305000	chr3	17455000	17460000	0.00032593	1.3415203	0.19510272
chr3	17221017	17888256	5000	chr3	17300000	17305000	chr3	17455000	17460000	0.00032593	1.3415203	0.19510272
chr3	36843149	36945794	5000	chr3	36770000	36775000	chr3	36895000	36900000	0.00071711	1.3308439	1.3718083
chr3	52965713	53175017	10000	chr3	52880000	52890000	chr3	53090000	53100000	8.95E-05	0.84550905	0.504883
chr3	135807609	136615268	5000	chr3	135785000	135790000	chr3	136470000	136475000	0.00010323	0.27722827	0.44168428
chr3	135807609	136615268	5000	chr3	135790000	135795000	chr3	136475000	136480000	0.00967255	1.312808	0.18504989
chr3	180571624	181207851	5000	chr3	180510000	180515000	chr3	180890000	180895000	0.00659433	1.1958251	1.3511878
chr3	180571624	181207851	5000	chr3	181040000	181045000	chr3	181405000	181410000	0.00071711	1.1443537	1.1120648
chr4	170357792	170646003	10000	chr4	169930000	169940000	chr4	170480000	170490000	2.37E-09	1.3162088	0.7925938
chr5	87676693	88195380	10000	chr5	87070000	87080000	chr5	87840000	87850000	3.70E-05	0.22795595	0.40951368
chr5	87676693	88195380	10000	chr5	87680000	87690000	chr5	87780000	87790000	0.00987681	1.4089538	0.95696616
chr5	87676693	88195380	10000	chr5	87680000	87690000	chr5	87780000	87790000	0.00987681	1.4089538	0.95696616
chr5	137838122	137948140	5000	chr5	137910000	137915000	chr5	138600000	138605000	0.00010323	0.5248362	0.11458015
chr6	24988105	33842877	5000	chr6	26610000	26615000	chr6	27050000	27055000	0.09176145	1.1728417	0.3633115
chr6	24988105	33842877	5000	chr6	26610000	26615000	chr6	27050000	27055000	0.09176145	1.1728417	0.3633115
chr6	24988105	33842877	10000	chr6	27140000	27150000	chr6	27430000	27440000	9.69E-05	1.5393742	0.3642763
chr6	24988105	33842877	10000	chr6	27140000	27150000	chr6	27430000	27440000	9.69E-05	1.5393742	0.3642763
chr6	24988105	33842877	5000	chr6	28235000	28240000	chr6	28300000	28305000	0.04845583	1.2806816	1.1394541
chr6	24988105	33842877	5000	chr6	28235000	28240000	chr6	28300000	28305000	0.04845583	1.2806816	1.1394541
chr6	24988105	33842877	5000	chr6	30710000	30715000	chr6	30845000	30850000	3.10E-05	0.768764	0.21247724
chr6	24988105	33842877	5000	chr6	30710000	30715000	chr6	30845000	30850000	3.10E-05	0.768764	0.21247724
chr6	24988105	33842877	5000	chr6	31645000	31650000	chr6	31800000	31805000	0.00290042	0.14871757	0.6141024
chr6	24988105	33842877	5000	chr6	31645000	31650000	chr6	31800000	31805000	0.00290042	0.14871757	0.6141024
chr6	24988105	33842877	5000	chr6	32085000	32090000	chr6	32120000	32125000	4.14E-05	1.1549484	0.14398974
chr6	24988105	33842877	5000	chr6	32085000	32090000	chr6	32120000	32125000	4.14E-05	1.1549484	0.14398974

chr6	24988105	33842877	10000	chr6	32850000	32860000	chr6	32900000	32910000	6.56E-06	1.5971392	0.20099184
chr6	24988105	33842877	10000	chr6	32850000	32860000	chr6	32900000	32910000	6.56E-06	1.5971392	0.20099184
chr6	24988105	33842877	10000	chr6	32850000	32860000	chr6	32900000	32910000	6.56E-06	1.5971392	0.20099184
chr6	24988105	33842877	10000	chr6	32850000	32860000	chr6	32900000	32910000	6.56E-06	1.5971392	0.20099184
chr6	24988105	33842877	10000	chr6	32980000	32990000	chr6	33040000	33050000	2.64E-05	1.601542	0.26288363
chr6	24988105	33842877	10000	chr6	32980000	32990000	chr6	33040000	33050000	2.64E-05	1.601542	0.26288363
chr6	84264202	84409255	10000	chr6	84320000	84330000	chr6	84920000	84930000	8.82E-06	1.5061508	0.8328693
chr7	86403263	86948073	5000	chr7	86560000	86565000	chr7	86770000	86775000	0.00199059	1.3308439	1.3042823
chr7	86403263	86948073	5000	chr7	86560000	86565000	chr7	86770000	86775000	0.00199059	1.3308439	1.3042823
chr7	104597669	105063372	5000	chr7	104620000	104625000	chr7	104655000	104660000	0.00326091	1.0854295	1.2470313
chr7	104597669	105063372	5000	chr7	104620000	104625000	chr7	104655000	104660000	0.00326091	1.0854295	1.2470313
chr8	26190836	26279173	5000	chr8	26230000	26235000	chr8	26305000	26310000	0.00317194	1.3397491	0.4989114
chr8	38014429	38310910	10000	chr8	37960000	37970000	chr8	38240000	38250000	0.00277824	1.5559133	0.5771854
chr8	89221915	89462854	10000	chr8	88700000	88710000	chr8	89330000	89340000	0.00277824	0.55052704	0.17356487
NPCs												
Riskchr	**Risk start**	**Risk end**	**Risk overlap**	**chr1**	**x1**	**x2**	**chr2**	**y1**	**y2**	**GliafdrDonut**	**NeufdrDonut**	**NPCfdrDonut**
chr1	8392592	8701288	10000	chr1	8070000	8080000	chr1	8400000	8410000	0.19700903	1.5615494	0.01706708
chr1	8392592	8701288	10000	chr1	8470000	8480000	chr1	8780000	8790000	1.5955924	1.3734437	0.00030532
chr1	8392592	8701288	10000	chr1	8520000	8530000	chr1	8570000	8580000	0.47351536	1.8601022	0.00419528
chr1	8392592	8701288	10000	chr1	8520000	8530000	chr1	8570000	8580000	0.47351536	1.8601022	0.00419528
chr1	98341152	98559093	5000	chr1	97710000	97715000	chr1	98515000	98520000	1.2857124	1.1945096	7.96E-09
chr1	98341152	98559093	10000	chr1	98290000	98300000	chr1	98510000	98520000	0.3418393	0.22795595	3.03E-06
chr1	149998923	150214166	5000	chr1	149930000	149935000	chr1	150155000	150160000	1.046633	1.3828198	0.00012025
chr1	200253612	200269903	6388	chr1	200250000	200260000	chr1	200450000	200460000	0.14504352	0.84550905	5.19E-05
chr1	243503764	243685760	5000	chr1	243505000	243510000	chr1	244010000	244015000	1.0040472	1.3334736	6.39E-07
chr1	243503764	243685760	10000	chr1	243630000	243640000	chr1	244210000	244220000	0.2424752	1.5495831	0.00119237
chr10	104570118	105059896	5000	chr10	104590000	104595000	chr10	104665000	104670000	0.11233018	0.6618322	3.07E-05
chr10	104570118	105059896	5000	chr10	104590000	104595000	chr10	104665000	104670000	0.11233018	0.6618322	3.07E-05
chr11	24367339	24412992	5000	chr11	23560000	23565000	chr11	24405000	24410000	0.8606538	1.3652908	6.39E-07
chr11	30240470	30395895	5000	chr11	30365000	30370000	chr11	30940000	30945000	0.61549026	0.12672402	7.96E-09
chr11	46342942	46751495	10000	chr11	46260000	46270000	chr11	46550000	46560000	0.51119745	0.6351586	0.00517559
chr11	65378028	65485218	10000	chr11	65470000	65480000	chr11	65760000	65770000	0.96181405	1.4031175	0.0005021
chr11	65378028	65485218	5218	chr11	65480000	65490000	chr11	65750000	65760000	1.5454209	0.18674426	0.01933173
chr11	113317745	113451229	10000	chr11	112080000	112090000	chr11	113400000	113410000	1.4454288	1.5345914	0.00173531
chr12	2321868	2523772	5000	chr12	2165000	2170000	chr12	2390000	2395000	0.9105591	1.3828198	1.94E-05
chr12	39448519	39533484	10000	chr12	38530000	38540000	chr12	39490000	39500000	1.4776708	1.3464317	0.00176357
chr12	39448519	39533484	5000	chr12	39300000	39305000	chr12	39495000	39500000	0.18078998	0.5831852	5.75E-08
chr12	39448519	39533484	5000	chr12	39315000	39320000	chr12	39495000	39500000	0.10347593	0.82978237	0.00040699
chr12	57483524	57682956	5000	chr12	57400000	57405000	chr12	57500000	57505000	1.2518088	1.3637018	0.05757648
chr12	57483524	57682956	5000	chr12	57405000	57410000	chr12	57505000	57510000	1.3117417	0.768764	6.82E-07
chr12	57483524	57682956	5000	chr12	57460000	57465000	chr12	57500000	57505000	0.6582037	0.29175246	2.68E-09
chr12	57483524	57682956	5000	chr12	57515000	57520000	chr12	57570000	57575000	0.31741977	1.6471775	4.71E-07
chr12	57483524	57682956	5000	chr12	57515000	57520000	chr12	57570000	57575000	0.31741977	1.6471775	4.71E-07
chr12	57483524	57682956	5000	chr12	57515000	57520000	chr12	57570000	57575000	0.31741977	1.6471775	4.71E-07
chr12	57483524	57682956	5000	chr12	57515000	57520000	chr12	57570000	57575000	0.31741977	1.6471775	4.71E-07
chr12	57483524	57682956	5000	chr12	57605000	57610000	chr12	57640000	57645000	1.2080998	0.21955356	0.00049333
chr12	57483524	57682956	5000	chr12	57605000	57610000	chr12	57640000	57645000	1.2080998	0.21955356	0.00049333
chr12	103361112	103382325	3888	chr12	103360000	103365000	chr12	103405000	103410000	1.0532553	1.406275	0.00465611
chr12	123447928	123902361	10000	chr12	123450000	123460000	chr12	123630000	123640000	0.48641446	0.14620744	0.00015493
chr12	123447928	123902361	10000	chr12	123450000	123460000	chr12	123630000	123640000	0.48641446	0.14620744	0.00015493
chr12	123447928	123902361	5000	chr12	123755000	123760000	chr12	124065000	124070000	0.5114943	1.0074358	2.63E-06
chr14	30000405	30190316	5000	chr14	29220000	29225000	chr14	30005000	30010000	0.4999548	1.1687775	0.00664212
chr14	30000405	30190316	5000	chr14	30010000	30015000	chr14	30680000	30685000	0.61155725	1.312808	6.39E-07
chr14	30000405	30190316	10000	chr14	30020000	30030000	chr14	30910000	30920000	1.0302607	1.5480927	0.02061128
chr14	59864362	60033892	10000	chr14	59950000	59960000	chr14	60380000	60390000	1.5195316	0.6351586	2.58E-09
chr14	59864362	60033892	10000	chr14	59970000	59980000	chr14	60640000	60650000	1.217849	1.4720036	7.04E-05
chr14	104188920	104537680	5000	chr14	104270000	104275000	chr14	104445000	104450000	0.38935196	1.214352	0.05757648
chr14	104188920	104537680	5000	chr14	104270000	104275000	chr14	104375000	104380000	0.75055546	1.3912318	5.85E-14
chr14	104188920	104537680	5000	chr14	104270000	104275000	chr14	104375000	104380000	0.75055546	1.3912318	5.85E-14
chr14	104188920	104537680	5000	chr14	104270000	104275000	chr14	104375000	104380000	0.75055546	1.3912318	5.85E-14
chr14	104188920	104537680	5000	chr14	104270000	104275000	chr14	104375000	104380000	0.75055546	1.3912318	5.85E-14
chr14	104188920	104537680	5000	chr14	104270000	104275000	chr14	104445000	104450000	0.38935196	1.214352	0.05757648
chr15	40566759	40602256	5000	chr15	40570000	40575000	chr15	40610000	40615000	1.265262	0.5248362	0.08441468
chr15	84703470	85392298	10000	chr15	82530000	82540000	chr15	84760000	84770000	0.5006912	0.7435584	0.00023499
chr15	84703470	85392298	10000	chr15	85120000	85130000	chr15	85470000	85480000	0.5458053	0.84550905	0.0005021
chr16	58538662	58684945	5000	chr16	58565000	58570000	chr16	58610000	58615000	0.15176348	0.78110456	2.92E-06
chr16	58538662	58684945	5000	chr16	58565000	58570000	chr16	58610000	58615000	0.15176348	0.78110456	2.92E-06
chr16	67708897	68305708	10000	chr16	67840000	67850000	chr16	67960000	67970000	0.14010376	1.5178461	0.00205718
chr16	67708897	68305708	10000	chr16	67840000	67850000	chr16	68070000	68080000	0.5458053	0.2988238	2.33E-05
chr16	67708897	68305708	10000	chr16	67840000	67850000	chr16	67960000	67970000	0.14010376	1.5178461	0.00205718
chr16	67708897	68305708	10000	chr16	67840000	67850000	chr16	68070000	68080000	0.5458053	0.2988238	2.33E-05
chr16	67708897	68305708	10000	chr16	68120000	68130000	chr16	68250000	68260000	0.10775846	0.77173877	0.0001776
chr16	67708897	68305708	10000	chr16	68120000	68130000	chr16	68250000	68260000	0.10775846	0.77173877	0.0001776
chr17	17649172	17967397	5000	chr17	17575000	17580000	chr17	17760000	17765000	1.0688708	1.3500707	2.69E-05
chr17	17649172	17967397	5000	chr17	17580000	17585000	chr17	17760000	17765000	0.10347593	0.5248362	0.00024168
chr17	17649172	17967397	10000	chr17	17710000	17720000	chr17	17760000	17770000	1.2335014	1.7054206	5.20E-10
chr17	17649172	17967397	10000	chr17	17710000	17720000	chr17	17760000	17770000	1.2335014	1.7054206	5.20E-10
chr17	17649172	17967397	5000	chr17	17715000	17720000	chr17	17760000	17765000	1.2496098	1.2806816	1.86E-08
chr17	17649172	17967397	5000	chr17	17715000	17720000	chr17	17760000	17765000	1.2496098	1.2806816	1.86E-08
chr17	17649172	17967397	5000	chr17	17795000	17800000	chr17	17935000	17940000	1.0495049	0.5248362	5.57E-05
chr17	17649172	17967397	5000	chr17	17795000	17800000	chr17	17935000	17940000	1.0495049	0.5248362	5.57E-05
chr17	17649172	17967397	5000	chr17	17795000	17800000	chr17	17935000	17940000	1.0495049	0.5248362	5.57E-05
chr17	17649172	17967397	5000	chr17	17795000	17800000	chr17	17935000	17940000	1.0495049	0.5248362	5.57E-05
chr18	52747689	53804156	5000	chr18	53080000	53085000	chr18	54240000	54245000	0.5891176	1.3511385	4.20E-06
chr18	52747689	53804156	10000	chr18	53250000	53260000	chr18	54240000	54250000	0.1354248	0.2903848	1.89E-05
chr19	19358332	19657632	10000	chr19	19640000	19650000	chr19	20060000	20070000	0.9354744	0.35116497	6.07E-07
chr19	50106208	50182697	10000	chr19	50010000	50020000	chr19	50170000	50180000	0.14845291	0.45033494	3.03E-06
chr2	57950104	58484172	10000	chr2	57980000	57990000	chr2	58460000	58470000	1.173447	1.5335083	0.00251959
chr2	57950104	58484172	10000	chr2	57980000	57990000	chr2	58460000	58470000	1.173447	1.5335083	0.00251959
chr2	57950104	58484172	10000	chr2	58130000	58140000	chr2	58780000	58790000	1.5449245	1.2797885	0.00020904
chr2	57950104	58484172	5000	chr2	58135000	58140000	chr2	58670000	58675000	1.2857124	1.1327559	0.00071491
chr2	198146381	198940251	10000	chr2	197160000	197170000	chr2	198240000	198250000	1.4342196	1.5084682	8.24E-05
chr2	198146381	198940251	5000	chr2	197710000	197715000	chr2	198270000	198275000	1.2287014	1.34677	2.69E-05
chr2	198146381	198940251	10000	chr2	197760000	197770000	chr2	198260000	198270000	1.173447	1.3162088	7.00E-09
chr2	198146381	198940251	10000	chr2	197790000	197800000	chr2	198170000	198180000	1.0634867	0.14082912	0.00205718
chr2	198146381	198940251	5000	chr2	198055000	198060000	chr2	198245000	198250000	0.11233018	1.0938406	0.00040699
chr2	198146381	198940251	5000	chr2	198055000	198060000	chr2	198245000	198250000	0.11233018	1.0938406	0.00040699

chr2	198146381	198940251	5000	chr2	198275000	198280000	chr2	198350000	198355000	0.33233196	1.1542732	0.00040699
chr2	198146381	198940251	5000	chr2	198275000	198280000	chr2	198350000	198355000	0.33233196	1.1542732	0.00040699
chr2	198146381	198940251	5000	chr2	198365000	198370000	chr2	198570000	198575000	0.5114943	1.3262548	8.71E-06
chr2	198146381	198940251	5000	chr2	198365000	198370000	chr2	198570000	198575000	0.5114943	1.3262548	8.71E-06
chr2	200547937	201309547	5000	chr2	199175000	199180000	chr2	200605000	200610000	1.2815038	0.25484702	0.00180604
chr2	200547937	201309547	5000	chr2	200565000	200570000	chr2	200605000	200610000	0.884727	1.3637018	0.00103107
chr2	200547937	201309547	5000	chr2	200565000	200570000	chr2	200605000	200610000	0.884727	1.3637018	0.00103107
chr22	41408754	41684093	5000	chr22	41345000	41350000	chr22	41415000	41420000	1.0219431	1.3248246	0.00267979
chr22	41408754	41684093	10000	chr22	41550000	41560000	chr22	42030000	42040000	1.4649647	1.3464317	2.30E-05
chr22	42315790	42689370	10000	chr22	42350000	42360000	chr22	42480000	42490000	1.5955924	1.5869178	3.77E-05
chr22	42315790	42689370	10000	chr22	42350000	42360000	chr22	42480000	42490000	1.5955924	1.5869178	3.77E-05
chr3	10800703	10807667	4297	chr3	10800000	10805000	chr3	11100000	11105000	1.2339293	0.5064022	7.70E-10
chr3	10800703	10807667	6964	chr3	10800000	10810000	chr3	10920000	10930000	1.4067187	1.5559133	0.00251959
chr3	17221017	17888256	5000	chr3	16640000	16645000	chr3	17455000	17460000	0.7089102	1.1532058	4.27E-06
chr3	17221017	17888256	10000	chr3	17290000	17300000	chr3	17880000	17890000	0.4119865	0.86441755	0.01132488
chr3	17221017	17888256	8256	chr3	17290000	17300000	chr3	17880000	17890000	0.4119865	0.86441755	0.01132488
chr3	17221017	17888256	10000	chr3	17300000	17310000	chr3	17880000	17890000	1.069242	1.3684161	0.00194236
chr3	17221017	17888256	8256	chr3	17300000	17310000	chr3	17880000	17890000	1.069242	1.3684161	0.00194236
chr3	17221017	17888256	5000	chr3	17470000	17475000	chr3	17875000	17880000	0.34376985	1.0329467	1.63E-05
chr3	17221017	17888256	5000	chr3	17470000	17475000	chr3	17875000	17880000	0.34376985	1.0329467	1.63E-05
chr3	52965713	53175017	10000	chr3	52880000	52890000	chr3	53100000	53110000	0.31834856	1.5495831	4.66E-05
chr3	63792668	64003983	10000	chr3	62600000	62610000	chr3	63800000	63810000	1.2053969	0.6132447	0.00451189
chr3	135807609	136615268	5000	chr3	135920000	135925000	chr3	135965000	135970000	0.31253472	1.3427322	0.03763996
chr3	135807609	136615268	5000	chr3	135920000	135925000	chr3	135965000	135970000	0.31253472	1.3427322	0.03763996
chr3	135807609	136615268	10000	chr3	135970000	135980000	chr3	136460000	136470000	1.579057	1.3394774	0.00162069
chr3	135807609	136615268	10000	chr3	135970000	135980000	chr3	136460000	136470000	1.579057	1.3394774	0.00162069
chr3	135807609	136615268	10000	chr3	136110000	136120000	chr3	136720000	136730000	1.5838375	1.5335083	0.06340116
chr3	135807609	136615268	5000	chr3	136265000	136270000	chr3	136470000	136475000	1.2750337	1.269171	7.88E-06
chr3	135807609	136615268	5000	chr3	136265000	136270000	chr3	136470000	136475000	1.2750337	1.269171	7.88E-06
chr3	180571624	181207851	5000	chr3	180705000	180710000	chr3	180890000	180895000	0.884727	1.1728417	5.75E-08
chr3	180571624	181207851	5000	chr3	180705000	180710000	chr3	180890000	180895000	0.884727	1.1728417	5.75E-08
chr3	180571624	181207851	10000	chr3	180770000	180780000	chr3	181430000	181440000	1.4305197	1.5375824	2.30E-05
chr3	180571624	181207851	10000	chr3	181170000	181180000	chr3	181320000	181330000	1.5225048	1.5741836	3.97E-05
chr4	103001649	103198082	10000	chr4	102080000	102090000	chr4	103180000	103190000	0.31834856	1.5084682	6.39E-11
chr4	143742567	143883013	10000	chr4	143200000	143210000	chr4	143760000	143770000	0.2424752	1.466538	0.00010735
chr4	170357792	170646003	10000	chr4	170180000	170190000	chr4	170400000	170410000	0.26364303	1.4433368	0.0005892
chr4	170357792	170646003	10000	chr4	170530000	170540000	chr4	170970000	170980000	0.51119745	0.4597334	4.66E-05
chr5	49441779	49884022	10000	chr5	46320000	46330000	chr5	49500000	49510000	74.00363	196.37476	9.41E-08
chr5	49441779	49884022	10000	chr5	46320000	46330000	chr5	49500000	49510000	74.00363	196.37476	9.41E-08
chr5	87676693	88195380	5000	chr5	86705000	86710000	chr5	87685000	87690000	1.2059175	1.0492826	0.00019324
chr5	87676693	88195380	5000	chr5	87240000	87245000	chr5	87690000	87695000	1.3303624	1.2508715	2.52E-08
chr5	87676693	88195380	5000	chr5	87240000	87245000	chr5	87960000	87965000	1.0688708	1.2623683	1.84E-09
chr5	87676693	88195380	5000	chr5	87440000	87445000	chr5	87840000	87845000	0.5114943	1.1590292	0.00024168
chr5	87676693	88195380	10000	chr5	87500000	87510000	chr5	87960000	87970000	0.10775846	0.84550905	6.22E-13
chr5	87676693	88195380	5000	chr5	87690000	87695000	chr5	87840000	87845000	0.38935196	0.5064022	8.71E-06
chr5	87676693	88195380	5000	chr5	87690000	87695000	chr5	87840000	87845000	0.38935196	0.5064022	8.71E-06
chr5	87676693	88195380	5000	chr5	87695000	87700000	chr5	87780000	87785000	0.11233018	1.34677	2.15E-06
chr5	87676693	88195380	5000	chr5	87695000	87700000	chr5	87780000	87785000	0.11233018	1.34677	2.15E-06
chr5	87676693	88195380	5000	chr5	87710000	87715000	chr5	87765000	87770000	0.6582037	1.3248246	0.01746155
chr5	87676693	88195380	5000	chr5	87710000	87715000	chr5	87765000	87770000	0.6582037	1.3248246	0.01746155
chr5	87676693	88195380	5000	chr5	87780000	87785000	chr5	87960000	87965000	0.500902	0.12672402	8.71E-06
chr5	87676693	88195380	5000	chr5	87780000	87785000	chr5	87960000	87965000	0.500902	0.12672402	8.71E-06
chr5	87676693	88195380	5000	chr5	87850000	87855000	chr5	87950000	87955000	1.2184821	1.269171	2.63E-06
chr5	87676693	88195380	5000	chr5	87850000	87855000	chr5	87950000	87955000	1.2184821	1.269171	2.63E-06
chr5	137838122	137948140	1878	chr5	137785000	137790000	chr5	137835000	137840000	0.4254983	1.2680601	7.56E-12
chr6	24988105	33842877	5000	chr6	24855000	24860000	chr6	25060000	25065000	1.0038705	0.9483586	0.02084518
chr6	24988105	33842877	5000	chr6	24860000	24865000	chr6	25015000	25020000	0.8402052	1.4453112	0.00296772
chr6	24988105	33842877	5000	chr6	25220000	25225000	chr6	25600000	25605000	0.6925214	1.3397491	6.44E-12
chr6	24988105	33842877	10000	chr6	25220000	25230000	chr6	25880000	25890000	1.4305197	1.5657233	0.00011976
chr6	24988105	33842877	5000	chr6	25220000	25225000	chr6	25600000	25605000	0.6925214	1.3397491	6.44E-12
chr6	24988105	33842877	10000	chr6	25220000	25230000	chr6	25880000	25890000	1.4305197	1.5657233	0.00011976
chr6	24988105	33842877	5000	chr6	25270000	25275000	chr6	25490000	25495000	0.11233018	1.3367205	0.00588869
chr6	24988105	33842877	5000	chr6	25270000	25275000	chr6	25490000	25495000	0.11233018	1.3367205	0.00588869
chr6	24988105	33842877	5000	chr6	25275000	25280000	chr6	25485000	25490000	0.6925214	1.1728417	0.00117906
chr6	24988105	33842877	5000	chr6	25275000	25280000	chr6	25485000	25490000	0.6925214	1.1728417	0.00117906
chr6	24988105	33842877	10000	chr6	26170000	26180000	chr6	26570000	26580000	0.74615896	1.5682914	0.01017091
chr6	24988105	33842877	10000	chr6	26170000	26180000	chr6	26570000	26580000	0.74615896	1.5682914	0.01017091
chr6	24988105	33842877	10000	chr6	26330000	26340000	chr6	27040000	27050000	1.5449245	0.8654066	1.48E-06
chr6	24988105	33842877	10000	chr6	26330000	26340000	chr6	27040000	27050000	1.5449245	0.8654066	1.48E-06
chr6	24988105	33842877	10000	chr6	26810000	26820000	chr6	58210000	58220000	3.5411448	22.384825	7.29E-16
chr6	24988105	33842877	10000	chr6	26810000	26820000	chr6	58210000	58220000	3.5411448	22.384825	7.29E-16
chr6	24988105	33842877	5000	chr6	27120000	27125000	chr6	27860000	27865000	0.32161906	0.5831852	0.00068245
chr6	24988105	33842877	5000	chr6	27120000	27125000	chr6	27860000	27865000	0.32161906	0.5831852	0.00068245
chr6	24988105	33842877	5000	chr6	27155000	27160000	chr6	27755000	27760000	1.2981611	1.3262548	8.30E-05
chr6	24988105	33842877	5000	chr6	27155000	27160000	chr6	27755000	27760000	1.2981611	1.3262548	8.30E-05
chr6	24988105	33842877	10000	chr6	27480000	27490000	chr6	27550000	27560000	1.3893585	1.6099682	3.71E-05
chr6	24988105	33842877	10000	chr6	27480000	27490000	chr6	27550000	27560000	1.3893585	1.6099682	3.71E-05
chr6	24988105	33842877	10000	chr6	27650000	27660000	chr6	27860000	27870000	1.3705888	1.3260291	1.89E-05
chr6	24988105	33842877	10000	chr6	27650000	27660000	chr6	27860000	27870000	1.3705888	1.3260291	1.89E-05
chr6	24988105	33842877	5000	chr6	28105000	28110000	chr6	28190000	28195000	0.75055546	1.3637018	0.00267979
chr6	24988105	33842877	5000	chr6	28105000	28110000	chr6	28190000	28195000	0.75055546	1.3637018	0.00267979
chr6	24988105	33842877	10000	chr6	28460000	28470000	chr6	28650000	28660000	0.1354248	0.45033494	1.03E-05
chr6	24988105	33842877	10000	chr6	28460000	28470000	chr6	28650000	28660000	0.1354248	0.45033494	1.03E-05
chr6	24988105	33842877	10000	chr6	28650000	28660000	chr6	29550000	29560000	0.13810997	0.15189412	6.27E-06
chr6	24988105	33842877	10000	chr6	28650000	28660000	chr6	29550000	29560000	0.13810997	0.15189412	6.27E-06
chr6	24988105	33842877	5000	chr6	29810000	29815000	chr6	29975000	29980000	0.38935196	1.3514338	1.04E-06
chr6	24988105	33842877	5000	chr6	29810000	29815000	chr6	29975000	29980000	0.38935196	1.3514338	1.04E-06
chr6	24988105	33842877	5000	chr6	29815000	29820000	chr6	29975000	29980000	0.16963711	0.49460578	3.02E-06
chr6	24988105	33842877	5000	chr6	29815000	29820000	chr6	29975000	29980000	0.16963711	0.49460578	3.02E-06
chr6	24988105	33842877	5000	chr6	30450000	30455000	chr6	30565000	30570000	0.61549026	1.2254188	0.00023257
chr6	24988105	33842877	5000	chr6	30450000	30455000	chr6	30565000	30570000	0.61549026	1.2254188	0.00023257
chr6	24988105	33842877	10000	chr6	30500000	30510000	chr6	31050000	31060000	0.10125773	1.2104888	1.76E-05
chr6	24988105	33842877	10000	chr6	30500000	30510000	chr6	31050000	31060000	0.10125773	1.2104888	1.76E-05
chr6	24988105	33842877	5000	chr6	31510000	31515000	chr6	31630000	31635000	1.2857124	0.76937705	0.00682281
chr6	24988105	33842877	5000	chr6	31510000	31515000	chr6	31630000	31635000	1.2857124	0.76937705	0.00682281
chr6	24988105	33842877	5000	chr6	31760000	31765000	chr6	31800000	31805000	0.9526036	1.4984149	0.00296772
chr6	24988105	33842877	5000	chr6	31760000	31765000	chr6	31800000	31805000	0.9526036	1.4984149	0.00296772

chr6	24988105	33842877	10000	chr6	32080000	32090000	chr6	32140000	32150000	0.44272268	1.6237167	4.96E-05
chr6	24988105	33842877	10000	chr6	32080000	32090000	chr6	32140000	32150000	0.44272268	1.6237167	4.96E-05
chr6	24988105	33842877	5000	chr6	32380000	32385000	chr6	32665000	32670000	0.38935196	0.5867754	4.44E-10
chr6	24988105	33842877	10000	chr6	32380000	32390000	chr6	32830000	32840000	0.8012636	1.1520559	1.76E-05
chr6	24988105	33842877	10000	chr6	32380000	32390000	chr6	32830000	32840000	0.8012636	1.1520559	1.76E-05
chr6	24988105	33842877	5000	chr6	32380000	32385000	chr6	32665000	32670000	0.38935196	0.5867754	4.44E-10
chr6	24988105	33842877	10000	chr6	32380000	32390000	chr6	32830000	32840000	0.8012636	1.1520559	1.76E-05
chr6	24988105	33842877	10000	chr6	32380000	32390000	chr6	32830000	32840000	0.8012636	1.1520559	1.76E-05
chr6	24988105	33842877	5000	chr6	32850000	32855000	chr6	32940000	32945000	0.4254983	1.3397491	0.01301809
chr6	24988105	33842877	5000	chr6	32850000	32855000	chr6	32925000	32930000	0.58261454	1.3912318	5.14E-07
chr6	24988105	33842877	5000	chr6	32850000	32855000	chr6	32925000	32930000	0.58261454	1.3912318	5.14E-07
chr6	24988105	33842877	5000	chr6	32850000	32855000	chr6	32940000	32945000	0.4254983	1.3397491	0.01301809
chr6	24988105	33842877	5000	chr6	33550000	33555000	chr6	33675000	33680000	0.21398759	0.18502843	1.39E-07
chr6	24988105	33842877	5000	chr6	33550000	33555000	chr6	33675000	33680000	0.21398759	0.18502843	1.39E-07
chr6	24988105	33842877	5000	chr6	33735000	33740000	chr6	33995000	34000000	0.16174258	1.214352	0.00071491
chr6	24988105	33842877	5000	chr6	33735000	33740000	chr6	34120000	34125000	1.2104913	0.17672323	6.03E-05
chr6	24988105	33842877	5000	chr6	33735000	33740000	chr6	34120000	34125000	1.2104913	0.17672323	6.03E-05
chr6	84264202	84409255	10000	chr6	84320000	84330000	chr6	84930000	84940000	0.21959442	0.31036413	0.0002823
chr7	86403263	86948073	5000	chr7	86565000	86570000	chr7	86780000	86785000	0.32161906	1.0722134	0.00021275
chr7	86403263	86948073	5000	chr7	86565000	86570000	chr7	86780000	86785000	0.32161906	1.0722134	0.00021275
chr7	86403263	86948073	10000	chr7	86690000	86700000	chr7	87580000	87590000	74.00363	1.4095246	0.04428426
chr8	18396405	18429406	9406	chr8	18050000	18060000	chr8	18420000	18430000	1.4860443	1.1758273	1.23E-07
chr8	26190836	26279173	5000	chr8	26230000	26235000	chr8	26510000	26515000	0.34376985	1.2175577	0.00071491
chr8	27327841	27453762	3762	chr8	27170000	27175000	chr8	27450000	27455000	0.7089102	1.3237675	5.86E-05
chr8	27327841	27453762	3762	chr8	27170000	27175000	chr8	27450000	27455000	0.7089102	1.3237675	5.86E-05
chr8	38014429	38310910	10000	chr8	37630000	37640000	chr8	38020000	38030000	1.3000284	0.2903848	4.66E-05
chr8	38014429	38310910	5000	chr8	37975000	37980000	chr8	38030000	38035000	1.2364298	1.3860595	3.32E-06
chr8	38014429	38310910	5000	chr8	37975000	37980000	chr8	38030000	38035000	1.2364298	1.3860595	3.32E-06
chr8	38014429	38310910	5000	chr8	38035000	38040000	chr8	38575000	38580000	0.34376985	1.0722134	0.00021275
chr8	60475926	60954059	5000	chr8	59630000	59635000	chr8	60835000	60840000	0.10347593	1.2623683	2.74E-05
chr8	60475926	60954059	5000	chr8	59720000	59725000	chr8	60830000	60835000	0.8606538	0.8023862	9.22E-05
chr8	89221915	89462854	10000	chr8	87710000	87720000	chr8	89330000	89340000	1.3222934	0.45391688	3.55E-07
chr9	101065115	101076627	1627	chr9	101075000	101080000	chr9	101535000	101540000	0.17700288	1.2175577	5.86E-05

Table 3.9 | Loops anchored in schizophrenia (SZ) risk sequence.
Cell-type-specific HiCCUPS loops (i.e., FDR < 0.1 in one cell type and FDR > 0.1 in the remaining two) overlapping SZ risk-associated GWAS loci in at least one anchor. Riskchr, Risk start, Risk end = coordinates of SZ risk locus overlapping loop anchor; Risk overlap = # bp overlap between loop anchor and risk locus; chr1, x1, x2 = anchor 1 coordinates; chr2, y1, y2 = anchor 2 coordinates; Glia/Neu/NPC-fdrDonut = HiCCUPS FDR value from the "donut" local neighborhood.

eSNP	Chr	GWAS Locus Start	GWAS Locus End	eSNP position	Gene	ROI	anchor.bin.coord	anchor.bin.genes	target.bin.coord	target.bin.genes	Glia	NPC	Neuron	Tissue	Cell Type	Developmental Period	Primary/Conditional
rs12037821	1	2372401	2402501	2387715	SLC35E2	chr1:2372402-2402501	chr1:2380001-2390000		chr1:1670001-1680000	SLC35E2B|0-1|-,SLC35E2|0-4|-	NA	NA	NA	–	–	early mid-prenatal	conditional
rs138050288	1	8355697	8638984	8460247	RERE	chr1:8411185-8638984	chr1:8460001-8470000	RERE|1|-	chr1:8880001-8890000	RERE|0|-	1.57E+00	2.25E+00	1.78E+00	–	–	–	primary
rs201524	1	30412551	30443951	30428943	PTPRU	chr1:30412552-30437271	chr1:30420001-30430000		chr1:29560001-29570000	PTPRU|0-1|+,MECR|0|-	1.679003	1.130746	1.760513	–	neurons	early mid-prenatal	primary
rs2621129	2	198148577	198835577	198265350	SF3B1	chr2:198148578-198835577	chr2:198260001-198270000	SF3B1|13-26|-	chr2:198290001-198300000	SF3B1|0-1|-	6.28E+00	7.76E+00	1.82E+00	–	–	–	primary
rs35220450	2	200715237	201247789	200780737	FTCDNL1, AC073043.2	chr2:200715238-200848037	chr2:200780001-200790000	C2orf69|2|+	chr2:200710001-200720000	FTCDNL1|0-3|-	5.62E-01	0.038190 44	0.15	–	–	adult	primary
rs186546506	2	200715237	201247789	200779303	LINC01792, AC007163.3	chr2:200715238-200848037	chr2:200770001-200780000	C2orf69|0|-1|+	chr2:201600001-201610000	LOC100507140|0|-,AOX2P|7-11|+	1.33E+00	2.21E-01	4.18E-01	putamen (basal ganglia)	–	adult	conditional
rs9834970	3	36843183	36945783	36856030	DCLK3	chr3:36843184-36945783	chr3:36850001-36860000		chr3:36780001-36790000	DCLK3|0-2|-	1.572272 52	5.13E-01	0.960857 03	nerve - tibial	neurons	infant	primary
rs6801235	3	52281078	53539269	53009595	PPM1M	chr3:52965713-53175017	chr3:53000001-53010000	SFMBT1|2|-	chr3:52270001-52280000	TWF2|0-1|-,PPM1M|0-1|+	0.777369 68	0.285430 16	0.223966 09	–	neurons	late prenatal	conditional
rs113386200	3	63792650	64004050	63832212	THOC7	chr3:63792651-64004050	chr3:63800001-63810000	C3orf49|0|-2|+	chr3:63850001-63860000	ATXN7|0-2|+,THOC7|0|-	7.46E-03	7.35E-02	2.19E-01	–	–	–	primary
rs109355184	3	135807405	136615405	136153468	PCCB	chr3:135807406-136615405	chr3:136150001-136160000	STAG1|20|-	chr3:135960001-135970000	PCCB|0-1|+	0.038329 76	0.101138 13	0.015655 63	–	–	–	primary
rs7438	4	170357552	170646052	170642246	CLCN3	chr4:170357553-170646052	chr4:170640001-170650000	CLCN3|16|+	chr4:170540001-170550000	CLCN3|0-1|+	4.61E+00	1.10E+01	2.016706 55	–	–	–	primary
rs9292918	5	45291475	46404116	45301035	BRCAT54, RP11-53O19.1	chr5:45291476-45393775	chr5:45300001-45310000	HCN1|6|-	chr5:44740001-44750000	(BRCAT54)	0.659013 99	2.21E-01	0.712464 12	–	–	adult	primary
rs2016358	6	83779798	84407274	84283733	SNAP91	chr6:84279923-84407274	chr6:84280001-84290000	SNAP91|32-33|-	chr6:84420001-84430000	SNAP91|0|-	0.083956 96	0.058077 64	0.550498 17	cerebellar hemisphere	–	–	primary
rs17055186	8	26181524	26279124	26260910 1	SDAD1P	chr8:26190836-26279173	chr8:26260001-26270000	BNIP3L|6-8|+	chr8:26230001-26240000	PPP2R2A|18|+,BNIP3L|0|+,SDAD1P1	0.207700 33	0.675092 72	3.28E-01	testis	–	adult	conditional
rs201999919	8	38020424	38310924	382918441	WHSC1L1	chr8:38014429-38310910	chr8:38290001-38300000	FGFR1|5|-	chr8:38240001-38250000	LETM2|0-3|+,WHSC1L1|0|-	1.059031 73	6.29E-01	0.221717 08	–	–	early prenatal	primary
rs169385506	11	46340213	46751213	46454313	MDK	chr11:46342944-46751213	chr11:46450001-46460000	AMBRA1|17-18|-	chr11:46390001-46400000	MIR4688|0-1|+,MDK|0|+,DGKZ|8,12-34|+	9.32E-02	0.000108 69	0.094945 4	–	–	early mid-prenatal	primary
rs4559	12	57428314	57497814	57489648	STAT6	chr12:57428315-57682971	chr12:57480001-57490000	NAB2|0-7|+,STAT6|25|-	chr12:57520001-57530000	LRP1|0-1|+,STAT6|0-1|-	4.14E-02	3.87E-02	2.74E-02	–	microglia	adolescent	conditional
rs7171869	15	78803032	78926732	78900909	IREB2	chr15:78803033-78926732	chr15:78900001-78910000	CHRNA3|4|-	chr15:78720001-78730000	IREB2|0|+	0.282781 52	1.042098 59	0.226348 13	–	–	early prenatal	conditional
rs35677834	15	84661161	85153461	84933868	LOC101929479, RP11-561C5.3	chr15:84661162-85153461	chr15:84930001-84940000	LOC642423||-	chr15:85740001-85750000	LOC642423|0-1|-	NA	NA	NA	ovary	–	early mid-prenatal	primary
rs4702	15	91416560	91436560	91426560	FURIN	chr15:91416561-91429040	chr15:91420001-91430000	FES|0-4|+,FURIN|5-16|+	chr15:91400001-91410000	FURIN|0|+	3.41E-02	3.43E-01	2.53E-01	–	endothelial cells	adolescent	primary
rs4788203	16	29924377	30144877	29978827 9	TMEM21	chr16:29924378-30144877	chr16:29970001-29980000	SLX1B-SULT1A4||+,TMEM219|2-6|+,SMG1P2||-,SLX1A||+,BOLA2||-,LOC606724||+,LOC613038||-	chr16:29940001-29950000	SLX1B-SULT1A4||+,**TMEM219|0|+**,KCTD13|0|-,SMG1P2||-,SLX1A||+,BOLA2||-,LOC606724||+,LOC613038||-	3.52E+00	5.98E+00	4.65E+00	–	–	–	primary
rs3935873	16	29924377	30144877	30018500	INO80E	chr16:29924378-30144877	chr16:30010001-30020000	SLX1B-SULT1A4||+,INO80E|4-11|+,SMG1P2||-,SLX1A||+,BOLA2||-,LOC606724||+,LOC613038||-	chr16:30000001-30010000	SLX1B-SULT1A4||+,INO80E|0-3|+,SMG1P2||-,SLX1A||+,BOLA2||-,LOC606724||+,LOC613038||-,HIRIP3|0-	0.141626 01	0.073098 34	0.490891 01	–	neurons	–	primary

								8||-,DOC2A|11-17|-,HIRIP3|0|-		9|-,TAOK2|22-24|+						
rs4787491	16	29924377	30144877	30015337	DOC2A	chr16:29924378-30144877	chr16:30010001-30020000	SLX1B-SULT1A4||+,INO80E|4-11|+,SMG1P2||-,SLX1A||+,BOLA2||-,LOC606724||+,LOC613038||-,DOC2A|11-17|-,HIRIP3|0|-	chr16:30000001-30010000	SLX1B-SULT1A4||+,INO80E|0-3|+,SMG1P2||-,SLX1A||+,BOLA2||-,LOC606724||+,LOC613038||-,HIRIP3|0-9|-,TAOK2|22-24|+	0.14162601	0.07309834	0.49089101	brain - cortex	neurons	adolescent	conditional
rs1164797	16	58669293	58691393	58680936	CNOT1	chr16:58669294-58682833	chr16:58680001-58690000		chr16:58670001-58670000	CNOT1|0-1|-	8.66E-01	9.55E-01	6.07E-01	–	–	–	primary
rs4072739	17	17722402	18030202	17884660	DRG2	chr17:17722403-18030202	chr17:17880001-17890000	DRC3|4-5|+	chr17:17980001-17990000	DRG2|0|+	0.3760446	0.19674481	0.37681556	–	–	–	primary
rs72986630	19	11839736	11859736	11849736	ZNF823	chr19:11849736-11849736	chr19:11840001-11850000	ZNF823|0-1|-	chr19:11850001-11860000	ZNF823|0|-	0.00738382	0.04647734	7.22E-08	–	endothelial cells	early prenatal	primary
rs2965199	19	19374022	19658022	19475088	GATAD2A	chr19:19374023-19658022	chr19:19470001-19480000		chr19:19490001-19500000	GATAD2A|0-1|+	2.81E+00	5.43E+00	1.44233463	–	–	–	primary
rs5023763	19	50067499	50135399	50072067	SNRNP70	chr19:50067500-50135399	chr19:50070001-50080000	NOSIP||-	chr19:49580001-49590000	SNRNP70|0-2|+	0.75323226	9.12E-01	5.46E-01	–	–	–	primary
rs200447424	22	41408556	42689414	41587548	RANGAP1	chr22:41408557-41675156	chr22:41580001-41590000		chr22:41680001-41690000	RANGAP1|0-3|-	7.81E-01	0.39254406	0.79789577	–	–	–	primary

Table 3.10 | Chromosomal contacts anchored in SZ GWAS co-localized eQTLs.

HiC interactions overlapping GWAS co-localized eQTL SNP-gene pairs (Dobbyn *et al.*, 2018). Columns A-E are from Dobbyn et al (2018) Table 2. eSNP, eQTL SNP; Chr, GWAS Locus Start and GWAS Locus End together indicate the coordinates for the corresponding risk locus within which the eSNP resides; Gene, the gene corresponding to the eSNP; eSNP position, coordinate of the eSNP in hg19; ROI, region of interest corresponding to the SZ risk loci; anchor.bin.coord, coordinate of the 10kb used as anchor in the binomial test for significant interactions; anchor.bin.genes, genes overlapping the 10kb anchor bin; target.bin.coord, coordinate of the 10kb queried as a target of the anchor in the binomial test for significant interactions; target.bin.genes, genes overlapping the 10kb target bin; Glia, NPC, Neuron columns refer to the log (q-value) for each interaction determined by binomial statistics.

						Genotype 0 (n=150)		Genotype 1 (n=277)		Genotype 2 (n=152)	
Ensembl ID	Gene ID	Gene Start (bp)	Gene End (bp)	p-value	p-adjusted	Mean	SE	Mean	SE	Mean	SE
ENSG00000250120	PCDHA10	140235595	140391929	2.40E-14	1.61E-12	-0.9450845	0.14005154	-0.3220529	0.0469093	-0.018637	0.03260897
ENSG00000204962	PCDHA8	140220907	140391929	5.52E-13	3.70E-11	-1.5623850	0.04569762	-1.3153710	0.03233461	-1.085131	0.04282544
ENSG00000204963	PCDHA7	140213969	140391929	2.03E-11	1.36E-09	-1.2849008	0.04151695	-1.0832615	0.02988216	-0.9140139	0.04085353
ENSG00000204969	PCDHA2	140174444	140391929	1.87E-09	1.25E-07	-1.2637450	0.03214275	-1.1634137	0.02443782	-0.9951632	0.0306389
ENSG00000239389	PCDHA13	140261793	140391929	2.65E-08	1.78E-06	-0.3379225	0.03420221	-0.4584249	0.02614051	-0.5841679	0.03465996
ENSG00000204967	PCDHA4	140186659	140391929	1.28E-05	0.0008576	-0.0940208	0.03133687	0.07152288	0.02254759	0.08552427	0.03129946
ENSG00000204961	PCDHA9	140227048	140391929	0.0003829	0.0256543	-0.8885906	0.13530321	-0.531843	0.04152144	-0.4422395	0.04931668
ENSG00000249158	PCDHA11	140248689	140391929	0.0008857	0.0593419	0.4383577	0.03023422	0.5256985	0.02074671	0.5673561	0.02727161
ENSG00000204970	PCDHA1	140165876	140391929	0.2997	1	-1.1683680	0.03560745	-1.2121490	0.02525289	-1.238437	0.03251574
ENSG00000255408	PCDHA3	140180783	140391929	0.09267	1	-0.3266051	0.02698158	-0.2652614	0.02019158	-0.2768451	0.03101834
ENSG00000204965	PCDHA5	140201222	140391929	0.3493	1	0.013753010	0.03100332	0.064261130	0.02347947	0.044756	0.03029636
ENSG00000081842	PCDHA6	140207563	140391929	0.4351	1	-0.0261231	0.0399647	-0.0368595	0.02697582	-0.0722259	0.03390551
ENSG00000249034	AC005609.1	140240341	140243224	0.8416	1	-1.6517620	0.06461126	-1.734715	0.0462817	-1.657964	0.06411573
ENSG00000251664	PCDHA12	140255058	140391929	0.01968	1	-0.5597011	0.03440314	-0.4958071	0.02650548	-0.4345344	0.03334305
ENSG00000248383	PCDHAC1	140306302	140391929	0.04019	1	0.1123253	0.04139802	0.1829496	0.02673045	0.2158291	0.03283328
ENSG00000243232	PCDHAC2	140345820	140391936	0.589	1	2.3616290	0.03826713	2.3767940	0.03004392	2.3263270	0.03968978
ENSG00000112852	PCDHB2	140474227	140476962	0.3131	1	0.6907204	0.03089572	0.6421203	0.02787192	0.6447874	0.0344836
ENSG00000272154	AC005754.7	140479829	140481794	0.881	1	-0.0156119	0.04615792	0.055654370	0.03715139	-0.0173929	0.06464357
ENSG00000113205	PCDHB3	140480234	140483406	0.02615	1	0.4022594	0.03103061	0.3829122	0.0218286	0.3135084	0.02927212
ENSG00000272108	AC005754.8	140498262	140500347	0.04383	1	0.6543306	0.04894567	0.5441159	0.03794151	0.5393526	0.05292806
ENSG00000081818	PCDHB4	140501581	140505201	0.8963	1	1.8412280	0.02971057	1.8153920	0.02339597	1.8104160	0.03254467
ENSG00000113209	PCDHB5	140514800	140517703	0.224	1	1.5671020	0.02299553	1.5496670	0.02068541	1.51964	0.02580371
ENSG00000113211	PCDHB6	140529683	140532868	0.3184	1	0.2656840	0.03583055	0.2689495	0.02605475	0.3133934	0.03659252
ENSG00000255622	PCDHB17	140535577	140538639	0.7226	1	-0.3756098	0.04920646	-0.4890143	0.03970313	-0.4072474	0.05548462
ENSG00000113212	PCDHB7	140552243	140555957	0.3499	1	0.4885163	0.03826745	0.4943283	0.03128195	0.5439796	0.04520774
ENSG00000120322	PCDHB8	140557371	140560081	0.9269	1	0.010015230	0.04438598	-0.013271500	0.03983904	0.025983590	0.05385032
ENSG00000196963	PCDHB16	140560980	140565793	0.4212	1	0.025974490	0.04194141	-0.027519400	0.03339609	0.090181980	0.04660624
ENSG00000177839	PCDHB9	140566893	140571111	0.03744	1	0.674292	0.0381622	0.7036802	0.02482808	0.7873397	0.03631906
ENSG00000120324	PCDHB10	140571942	140575215	0.0332	1	1.5509950	0.02669294	1.5585070	0.01814735	1.626481	0.02711403
ENSG00000197479	PCDHB11	140579183	140582618	0.7363	1	1.24371	0.02766519	1.1981840	0.02279798	1.224654	0.02874177
ENSG00000120328	PCDHB12	140588269	140591696	0.5195	1	1.5782920	0.02579869	1.5687640	0.01984231	1.613458	0.02583847
ENSG00000187372	PCDHB13	140593509	140596993	0.5773	1	1.1445940	0.03861705	1.1173810	0.02907133	1.186299	0.04333161

ENSG00000120327	PCDHB14	140602931	140605858	0.6245	1	2.372460	0.01787361	2.345718	0.01396584	2.359431	0.01816899
ENSG00000146001	PCDHB18	140613938	140617101	0.7843	1	0.5287241	0.04483872	0.4868465	0.03350162	0.556907	0.04718459
ENSG00000262096	PCDHB19P	140619518	140621864	0.6678	1	0.5610128	0.05018575	0.5222347	0.036843	0.6020646	0.04864062
ENSG00000113248	PCDHB15	140625147	140627799	0.2016	1	1.2757230	0.02478772	1.2598740	0.02024715	1.326387	0.0259284
ENSG00000178913	TAF7	140698057	140700330	0.6472	1	5.2875930	0.02872701	5.3071010	0.02267996	5.242052	0.03666379
ENSG00000272070	AC005618.6	140705777	140708924	0.3824	1	0.2658314	0.02889967	0.2744418	0.02361841	0.2928478	0.03017335
ENSG00000204956	PCDHGA1	140710252	140892546	0.08634	1	-0.9367046	0.04188522	-0.8589053	0.0321514	-0.8368188	0.03893686
ENSG00000081853	PCDHGA2	140718539	140892546	0.3214	1	-0.4044892	0.04268331	-0.4066491	0.03242702	-0.3469972	0.03727468
ENSG00000254245	PCDHGA3	140723601	140892546	0.699	1	0.1035317	0.04510637	0.0829532	0.03484245	0.1544447	0.0494062
ENSG00000254221	PCDHGB1	140729828	140892546	0.05279	1	0.372262	0.03450643	0.3869631	0.02540657	0.4593994	0.03457968
ENSG00000262576	PCDHGA4	140734768	140892546	0.6408	1	-0.0668598	0.03192729	-0.0684422	0.02606557	-0.0257894	0.03763937
ENSG00000253910	PCDHGB2	140739703	140892546	0.01694	1	0.3736417	0.037329	0.4069007	0.03197805	0.5231219	0.04426136
ENSG00000253485	PCDHGA5	140743898	140892546	0.6391	1	-0.401642	0.03938976	-0.373267	0.03426557	-0.3313155	0.0488054
ENSG00000262209	PCDHGB3	140749831	140892546	0.1391	1	-0.9051775	0.04431799	-0.8966564	0.03587637	-0.7654312	0.05014739
ENSG00000253731	PCDHGA6	140753651	140892546	0.6019	1	-0.3094883	0.0401174	-0.2804153	0.03370213	-0.3146994	0.05037824
ENSG00000253537	PCDHGA7	140762467	140892546	0.485	1	-0.4691286	0.03963365	-0.4766255	0.03066884	-0.4103881	0.04403638
ENSG00000253953	PCDHGB4	140767452	140892546	0.2756	1	-0.4760418	0.03919649	-0.4802712	0.02919548	-0.412954	0.04021008
ENSG00000253767	PCDHGA8	140772381	140892546	0.363	1	-0.9928499	0.05395769	-0.9891268	0.0443002	-0.8983144	0.06304307
ENSG00000261934	PCDHGA9	140782520	140892546	0.419	1	0.7850766	0.03107475	0.7785777	0.0252662	0.8391602	0.03686301
ENSG00000253305	PCDHGB6	140787770	140892546	0.438	1	0.0813423	0.04342428	0.0939199	0.03328246	0.1441245	0.04685704
ENSG00000253846	PCDHGA10	140792743	140892546	0.9656	1	0.7094627	0.03317664	0.7157928	0.02714398	0.7139836	0.03741461
ENSG00000254122	PCDHGB7	140797427	140892546	0.2273	1	-0.0062796	0.04391279	0.0008358	0.03474804	0.1018263	0.05136511
ENSG00000253873	PCDHGA11	140800762	140891835	0.7571	1	0.1503583	0.03974685	0.1630275	0.03141137	0.1841329	0.04277974
ENSG00000248449	PCDHGB8P	140805853	140808219	0.4083	1	0.1226163	0.05903111	0.0846888	0.0463076	0.2108079	0.06345998
ENSG00000253159	PCDHGA12	140810185	140892546	0.5601	1	0.4870029	0.04397322	0.4898922	0.03426719	0.521935	0.04629193
ENSG00000240184	PCDHGC3	140855580	140892542	0.246	1	3.3354960	0.04493861	3.3904390	0.03367271	3.4003170	0.04292565
ENSG00000242419	PCDHGC4	140864741	140892546	0.6337	1	1.5117940	0.04101957	1.5379310	0.02512287	1.5224940	0.03315068
ENSG00000240764	PCDHGC5	140868808	140892546	0.8556	1	2.8724480	0.04664224	2.8885230	0.03236965	2.84832	0.04243448
ENSG00000131504	DIAPH1	140894583	140998622	0.8943	1	2.4751230	0.02367114	2.4466330	0.01934177	2.4866940	0.02561167
ENSG00000248106	AC005609.2	140143695	140144406	NA	NA	NA	NA	NA	NA	NA	NA
ENSG00000249504	PCDHA14	140240860	140243104	NA	NA	NA	NA	NA	NA	NA	NA
ENSG00000171815	PCDHB1	140430979	140433512	NA	NA	NA	NA	NA	NA	NA	NA
ENSG00000120329	SLC25A2	140682196	140683612	NA	NA	NA	NA	NA	NA	NA	NA
ENSG00000255729	AC005618.1	140699661	140700339	NA	NA	NA	NA	NA	NA	NA	NA
ENSG00000242020	RN7SL68P	140858883	140859190	NA	NA	NA	NA	NA	NA	NA	NA

Table 3.11 | Gene-level single-SNP eQTLs for clustered PCDH.

Gene-level single-SNP eQTL analysis testing for association of clustered PCDH gene expression with SZ risk SNP rs111896713. Significant genes are highlighted in green.

bp	credsnp	indexsnp	R_to_index	hgnc_CP	hgnc_GZ	hgnc_ES	hgnc_IMR90	NPC	Neuron	Glia
103573575	rs10860949	rs10860964	0.976583					ASCL1, C12orf42	NA	NA
103574202	rs7306170	rs10860964	0.910154					ASCL1, C12orf42	NA	NA
103575583	rs10778221	rs10860964	0.973776	C12orf42	NA	NA	NA	ASCL1, C12orf42	NA	NA
103575787	rs10860950	rs10860964	0.936245	C12orf42	NA	NA	NA	ASCL1, C12orf42	NA	NA
68377126	rs5937157	rs5937157	1	EFNB1	NA	NA	NA	EFNB1	NA	EFNB1
68377204	rs62606711	rs5937157	0.912256	EFNB1	NA	NA	NA	EFNB1	NA	EFNB1
68377205	rs62606712	rs5937157	0.90829	EFNB1	NA	NA	NA	EFNB1	NA	EFNB1
68377485	rs2361468	rs5937157	0.993235	EFNB1	NA	NA	NA	EFNB1	NA	EFNB1
68377499	rs2885287	rs5937157	0.993235	EFNB1	NA	NA	NA	EFNB1	NA	EFNB1
68379039	rs5937159	rs5937157	0.92888	EFNB1	NA	NA	NA	EFNB1	NA	EFNB1
137913882	rs409273	rs3849046	0.887143	CTNNA1,LRRTM2,SIL1,MATR3,PAIP2	SIL1,MATR3,PAIP2,SLC23A1	CTNNA1,SIL1,MATR3	CTNNA1,SIL1	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2, SLC23A1, SPATA24, PROB1, DNAJC18	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2, SLC23A1, SPATA24, DNAJC18
137917639	rs7378744	rs3849046	0.88466	CTNNA1,LRRTM2,SIL1,MATR3,PAIP2	SIL1,MATR3,PAIP2,SLC23A1	CTNNA1,SIL1,MATR3	CTNNA1,SIL1	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2, SLC23A1, SPATA24, PROB1, DNAJC18	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2, SLC23A1, SPATA24, DNAJC18
137917826	rs11957778	rs3849046	0.880865	CTNNA1,LRRTM2,SIL1,MATR3,PAIP2	SIL1,MATR3,PAIP2,SLC23A1	CTNNA1,SIL1,MATR3	CTNNA1,SIL1	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2, SLC23A1, SPATA24, PROB1, DNAJC18	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2	CTNNA1, LRRTM2, SIL1, SNHG4, MATR3, SNORA74A, PAIP2, SLC23A1, SPATA24, DNAJC18
180928466	rs10804885	rs33972009	0.769784	CCDC39,SOX2	CCDC39,SOX2	NA	NA	SOX2, SOX-OT	SOX2, SOX-OT	SOX2, SOX-OT, FLJ46066
113364647	rs4245150	rs2514218	0.952883	DRD2	DRD2	NA	NA	TTC12	TTC12	TTC12
113364691	rs17602038	rs2514218	0.952883	DRD2	DRD2	NA	NA	TTC12	TTC12	TTC12
113364803	rs4938021	rs2514218	0.952883	DRD2	DRD2	NA	NA	TTC12	TTC12	TTC12
113365084	rs4936275	rs2514218	0.952883	DRD2	DRD2	NA	NA	TTC12	TTC12	TTC12
113365141	rs4936276	rs2514218	0.952883	DRD2	DRD2	NA	NA	TTC12	TTC12	TTC12
27442127	rs73229090	rs73229090	1	DPYSL2,PTK2B,TRIM35,CHRNA2,EPHX2	PTK2B,TRIM35,CHRNA2	CHRNA2,EPHX2	EPHX2,SCARA5	CHRNA2, PTK2B,TRIM35	CHRNA2, PTK2B,TRIM35, EPHX2	CHRNA2, PTK2B,TRIM35, EPHX2
27442329	rs73229093	rs73229090	0.849192	DPYSL2,PTK2B,TRIM35,CHRNA2,EPHX2	PTK2B,TRIM35,CHRNA2	CHRNA2,EPHX2	EPHX2,SCARA5	CHRNA2, PTK2B,TRIM35	CHRNA2, PTK2B,TRIM35, EPHX2	CHRNA2, PTK2B,TRIM35, EPHX2
27452241	rs35598594	rs73229090	0.81184	PTK2B,TRIM35,CHRNA2,EPHX2	PTK2B,TRIM35,CHRNA2,EPHX2	PTK2B,CHRNA2,EPHX2	NA	CHRNA2, PTK2B,TRIM35	CHRNA2, PTK2B,TRIM35, EPHX2	CHRNA2, PTK2B,TRIM35, EPHX2
27453579	rs35236974	rs73229090	0.815871	PTK2B,TRIM35,CHRNA2,EPHX2	PTK2B,TRIM35,CHRNA2,EPHX2	PTK2B,CHRNA2,EPHX2	NA	CHRNA2, PTK2B,TRIM35	CHRNA2, PTK2B,TRIM35, EPHX2	CHRNA2, PTK2B,TRIM35, EPHX2
140140239	rs2563258	rs111896713	0.963304	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB7	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505646	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA11
140140529	rs2563257	rs111896713	0.987169	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB6	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505647	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA10
140140853	rs2337516	rs111896713	0.965827	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB5	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505648	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA9
140141779	rs2563256	rs111896713	0.987169					PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB4	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505649	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA8

140142547	rs2563255	rs111896713	0.987169					PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB3	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505650	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA7
140142701	rs2563254	rs111896713	0.982092					PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB2	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505651	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA6
140143664	rs111896713	rs111896713	1	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB1	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505652	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA5
140145206	rs6891559	rs111896713	0.965827	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB0	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505653	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA4
140145949	rs11741994	rs111896713	0.965827	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB1	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505654	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA3
140146345	rs7737424	rs111896713	0.984453	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB2	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505655	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA2
140146821	rs13168514	rs111896713	0.924772	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB3	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505656	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA1
140146899	rs13168670	rs111896713	0.955721	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB4	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505657	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA0
140147248	rs12659129	rs111896713	0.987169	PCDHA2,PCDHA3,PCDHA1,PCDHB1,PCDHB4	PCDHA1,PCDHA2,PCDHA3,PCDHA4,PCDHA5,PCDHA6,PCDHA7,PCDHA8,PCDHA9,PCDHA10,PCDHA11,PCDHA12,PCDHA13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGB1,PCDHGA4,PCDHGB2		NA	PCDHB1,PCDHB12,PCDHB13,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA8,PCDHGA9,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGB4,PCDHGB5	PCDHB1,PCDHB2,PCDHB3,PCDHB5,PCDHB7,PCDHB8,PCDHB16,PCDHGB1,PCDHGB2,PCDHGB3,PCDHGA1,PCDHGA2,PCDHGA3,PCDHGA4,PCDHGA5,PCDHGA6,PCDHGA7,PCDHGA9,PCDHGA10,PCDHGA11,PCDHGB5,PCDHGB6,PCDHGB7,PCDHGA12,DIAPH1,LOC100505658	PCDHB7,PCDHB8,PCDHB16,SLC25A2,TAF7,PCDHGA1

Table 3.12 | Locus-specific chromosomal contacts in fetal brain compared to hiPSC-derived cells.

Comparison of genes identified as interacting with 36 credible SNPs of interest (Won *et al.*, 2016) between fetal brain and hiPSC Hi-C interactions. Credsnp, credible SNP called by CAVIAR; indexsnp, index SNP of the haplotype in which

credible SNP lies; R_to_index, correlation between credible SNP and its index SNP; hgnc_CP, gene interacting with credible SNP in cortical plate samples; hgnc_GZ, gene interacting with credible SNP in germinal zone samples; hgnc_ES, gene interacting with credible SNP in embryonic stem cell samples; hgnc_IMR90, gene interacting with credible SNP in IMR90 samples; NPC, Neuron, Glia – genes interacting with credible SNPs in cell types from the present study.

CRISPRa

ID	TYPE	COORDINATES (Region, gRNA)	gRNA SEQUENCE	REVERSE COMPLEMENT	IVT F1 PRIMER	IVT R1 PRIMER	GENE
AS1	Loop-SNP1	(chr12:103160000-104160000) 103573681-103573700 FWD	TTCTAGTTTAGACAACAGGA	TCCTGTTGTCTAAACTAGAA	TAATACGACTCACTATAGCTAGTTTAGACAACAGGA	TTCTAGCTCTAAAACTCCTGTTGTCTAAACTAG	ASCL1
AS2	Loop-SNP2	(chr12:103160000-104160000) 103573733-103573752 REV	CTTCAGATCATTCTTCCTGG	CCAGGAAGAATGATCTGAAG	TAATACGACTCACTATAGTCAGATCATTCTTCCTGG	TTCTAGCTCTAAAACCCAGGAAGAATGATCTGA	ASCL1
AS3	Loop-SNP3	(chr12:103160000-104160000) 103573810-103573829 FWD	ACAGTGGGACAGAGTTTAGG	CCTAAACTCTGTCCCACTGT	TAATACGACTCACTATAGAGTGGGACAGAGTTTAGG	TTCTAGCTCTAAAACCCTAAACTCTGTCCCACT	ASCL1
AS4	Loop-SNP4	(chr12:103160000-104160000) 103574251-103574270 REV	AACTGTACTAAGCTATCCAG	CTGGATAGCTTAGTACAGTT	TAATACGACTCACTATAGCTGTACTAAGCTATCCAG	TTCTAGCTCTAAAACCTGGATAGCTTAGTACAG	ASCL1
AS5	Loop-SNP5	(chr12:103160000-104160000) 103574252-103574271 REV	GAACTGTACTAAGCTATCCA	TGGATAGCTTAGTACAGTTC	TAATACGACTCACTATAGACTGTACTAAGCTATCCA	TTCTAGCTCTAAAACTGGATAGCTTAGTACAGT	ASCL1
AS6	Control1	(chr12:103160000-104160000) 103469557-103469576 REV	CTCTCCCATAAACACTACCC	GGGTAGTGTTTATGGGAGAG	TAATACGACTCACTATAGCTCCCATAAACACTACCC	TTCTAGCTCTAAAACGGGTAGTGTTTATGGGAG	ASCL1
AS7	Control2	(chr12:103160000-104160000) 103469613-103469632 FWD	TAAGACTCTATACATGACCA	TGGTCATGTATAGAGTCTTA	TAATACGACTCACTATAGAGACTCTATACATGACCA	TTCTAGCTCTAAAACTGGTCATGTATAGAGTCT	ASCL1
AS8	Control3	(chr12:103160000-104160000) 103469622-103469641 FWD	ATACATGACCATGGGTGATG	CATCACCCATGGTCATGTAT	TAATACGACTCACTATAGACATGACCATGGGTGATG	TTCTAGCTCTAAAACCATCACCCATGGTCATGT	ASCL1
AS9	Control4	(chr12:103160000-104160000) 103470279-103470298 REV	CCAGTTCTAGTATGAATCTG	CAGATTCATACTAGAACTGG	TAATACGACTCACTATAGAGTTCTAGTATGAATCTG	TTCTAGCTCTAAAACCAGATTCATACTAGAACT	ASCL1
AS10	Control5	(chr12:103160000-104160000) 103470309-103470328 FWD	AATTTTACAAAGTCGCAAGG	CCTTGCGACTTTGTAAAATT	TAATACGACTCACTATAGTTTTACAAAGTCGCAAGG	TTCTAGCTCTAAAACCCTTGCGACTTTGTAAAA	ASCL1
AS11	Promoter1	(chr12:103160000-104160000) 103350547-103350566 REV	TATGGCAGGGACGTCCCCCT	AGGGGGACGTCCCTGCCATA	TAATACGACTCACTATAGTGGCAGGGACGTCCCCCT	TTCTAGCTCTAAAACAGGGGGACGTCCCTGCCA	ASCL1
AS12	Promoter2	(chr12:103160000-104160000) 103350598-103350617 FWD	TCCTCGGTGTCGCTTCCCCG	CGGGGAAGCGACACCGAGGA	TAATACGACTCACTATAGCTCGGTGTCGCTTCCCCG	TTCTAGCTCTAAAACCGGGGAAGCGACACCGAG	ASCL1
AS13	Promoter3	(chr12:103160000-104160000) 103350602-103350621 REV	GCCGCGGGGAAGCGACACCG	CGGTGTCGCTTCCCCGCGGC	TAATACGACTCACTATAGCGCGGGGAAGCGACACCG	TTCTAGCTCTAAAACCGGTGTCGCTTCCCCGCG	ASCL1
AS14	Promoter4	(chr12:103160000-104160000) 103350875-103350894 FWD	TCCAATTTCTAGGGTCACCG	CGGTGACCCTAGAAATTGGA	TAATACGACTCACTATAGCAATTTCTAGGGTCACCG	TTCTAGCTCTAAAACCGGTGACCCTAGAAATTG	ASCL1
AS15	Promoter5	(chr12:103160000-104160000) 103351166-103351185 REV	CTCCCGGCTGAATAAACAGG	CCTGTTTATTCAGCCGGGAG	TAATACGACTCACTATAGCCCGGCTGAATAAACAGG	TTCTAGCTCTAAAACCCTGTTTATTCAGCCGGG	ASCL1
EF1	Loop-SNP1	(chrX:67810000-68810000) 68378608-68378627 REV	CCGACGGGCACCACCTAAGG	CCTTAGGTGGTGCCCGTCGG	TAATACGACTCACTATAGGACGGGCACCACCTAAGG	TTCTAGCTCTAAAACCCTTAGGTGGTGCCCGTC	EFNB1
EF2	Loop-SNP2	(chrX:67810000-68810000) 68378823-68378842 REV	CACTGCAAGTACAGAAACGT	ACGTTTCTGTACTTGCAGTG	TAATACGACTCACTATAGCTGCAAGTACAGAAACGT	TTCTAGCTCTAAAACACGTTTCTGTACTTGCAG	EFNB1
EF3	Loop-SNP3	(chrX:67810000-68810000) 68379023-68379042 REV	GGCCGGGGTTACTTCTAGGG	CCCTAGAAGTAACCCCGGCC	TAATACGACTCACTATAGCCGGGGTTACTTCTAGGG	TTCTAGCTCTAAAACCCCTAGAAGTAACCCCGG	EFNB1
EF4	Loop-SNP4	(chrX:67810000-68810000) 68379038-68379057 REV	ACATGCCTAACTGATGGCCG	CGGCCATCAGTTAGGCATGT	TAATACGACTCACTATAGATGCCTAACTGATGGCCG	TTCTAGCTCTAAAACCGGCCATCAGTTAGGCAT	EFNB1
EF5	Loop-SNP5	(chrX:67810000-68810000) 68379044-68379063 REV	AAGGCGACATGCCTAACTGA	TCAGTTAGGCATGTCGCCTT	TAATACGACTCACTATAGGGCGACATGCCTAACTGA	TTCTAGCTCTAAAACTCAGTTAGGCATGTCGCC	EFNB1
EF6	Control1	(chrX:67810000-68810000) 68249998-68250017 REV	GGCCTCCCCCAAAATATGTG	CACATATTTTGGGGGAGGCC	TAATACGACTCACTATAGCCTCCCCCAAAATATGTG	TTCTAGCTCTAAAACCACATATTTTGGGGGAGG	EFNB1
EF7	Control2	(chrX:67810000-68810000) 68250062-68250081 REV	GCCCTAGTGCTAGGGGAATG	CATTCCCCTAGCACTAGGGC	TAATACGACTCACTATAGCCTAGTGCTAGGGGAATG	TTCTAGCTCTAAAACCATTCCCCTAGCACTAGG	EFNB1
EF8	Control3	(chrX:67810000-68810000) 68250069-68250088 REV	AAAAAAAGCCCTAGTGCTAG	CTAGCACTAGGGCTTTTTT	TAATACGACTCACTATAGAAAAAGCCCTAGTGCTAG	TTCTAGCTCTAAAACCTAGCACTAGGGCTTTTT	EFNB1
EF9	Control4	(chrX:67810000-68810000) 68250158-68250177 FWD	CTGGGCCACTCTAACCAACA	TGTTGGTTAGAGTGGCCCAG	TAATACGACTCACTATAGGGGCCACTCTAACCAACA	TTCTAGCTCTAAAACTGTTGGTTAGAGTGGCCC	EFNB1
EF10	Control5	(chrX:67810000-68810000) 68250187-68250206 FWD	CCACTTAACAGTCCTTGTCA	TGACAAGGACTGTTAAGTGG	TAATACGACTCACTATAGACTTAACAGTCCTTGTCA	TTCTAGCTCTAAAACTGACAAGGACTGTTAAGT	EFNB1
EF11	Promoter1	(chrX:67810000-68810000) 68047822-68047841 FWD	GCACGTGGGTACGTCCTCTG	CAGAGGACGTACCCACGTGC	TAATACGACTCACTATAGACGTGGGTACGTCCTCTG	TTCTAGCTCTAAAACCAGAGGACGTACCCACGT	EFNB1
EF12	Promoter2	(chrX:67810000-68810000) 68048066-68048085 REV	GATAAAGAAAGACACCTCGA	TCGAGGTGTCTTTCTTTATC	TAATACGACTCACTATAGTAAAGAAAGACACCTCGA	TTCTAGCTCTAAAACTCGAGGTGTCTTTCTTTA	EFNB1
EF13	Promoter3	(chrX:67810000-68810000) 68048398-68048417 FWD	AGCCCACTAAAGCCTTACGT	ACGTAAGGCTTTAGTGGGCT	TAATACGACTCACTATAGCCCACTAAAGCCTTACGT	TTCTAGCTCTAAAACACGTAAGGCTTTAGTGGG	EFNB1
EF14	Promoter4	(chrX:67810000-68810000) 68048437-68048456 FWD	ACAGGCTGTCCTTGACACGG	CCGTGTCAAGGACAGCCTGT	TAATACGACTCACTATAGAGGCTGTCCTTGACACGG	TTCTAGCTCTAAAACCCGTGTCAAGGACAGCCT	EFNB1
EF15	Promoter5	(chrX:67810000-68810000) 68048547-68048566 FWD	TGGGGTGTTGACCAGACGCG	CGCGTCTGGTCAACACCCCA	TAATACGACTCACTATAGGGGTGTTGACCAGACGCG	TTCTAGCTCTAAAACCGCGTCTGGTCAACACCC	EFNB1

MA1	Loop-SNP1	**(chr5:137860000-138860000) 137914322-137914341 FWD**	TACCGATGCAAAGAAGACAG	CTGTCTTCTTTGCATCGGTA	TAATACGACTCACTATAGCCGATGCAAAGAAGACAG	TTCTAGCTCTAAAACCTGTCTTCTTTGCATCGG	MATR3
MA2	Loop-SNP2	**(chr5:137860000-138860000) 137914389-137914408 FWD**	TGTTCTTACGTTGTACATGA	TCATGTACAACGTAAGAACA	TAATACGACTCACTATAGTTCTTACGTTGTACATGA	TTCTAGCTCTAAAACTCATGTACAACGTAAGAA	MATR3
MA3	Loop-SNP3	**(chr5:137860000-138860000) 137914391-137914410 FWD**	TTCTTACGTTGTACATGACG	CGTCATGTACAACGTAAGAA	TAATACGACTCACTATAGCTTACGTTGTACATGACG	TTCTAGCTCTAAAACCGTCATGTACAACGTAAG	MATR3
MA4	Loop-SNP4	**(chr5:137860000-138860000) 137914394-137914413 FWD**	TTACGTTGTACATGACGGGG	CCCCGTCATGTACAACGTAA	TAATACGACTCACTATAGACGTTGTACATGACGGGG	TTCTAGCTCTAAAACCCCCGTCATGTACAACGT	MATR3
MA5	Loop-SNP5	**(chr5:137860000-138860000) 137914401-137914420 FWD**	GTACATGACGGGGCGGAGAA	TTCTCCGCCCCGTCATGTAC	TAATACGACTCACTATAGACATGACGGGGCGGAGAA	TTCTAGCTCTAAAACTTCTCCGCCCCGTCATGT	MATR3
MA6	Control1	**(chr5:137860000-138860000) 138278537-138278556 REV**	CCTCACACAAGGGAGCCGCA	TGCGGCTCCCTTGTGTGAGG	TAATACGACTCACTATAGTCACACAAGGGAGCCGCA	TTCTAGCTCTAAAACTGCGGCTCCCTTGTGTGA	MATR3
MA7	Control2	**(chr5:137860000-138860000) 138278547-138278566 REV**	GGCCTGGGAACCTCACACAA	TTGTGTGAGGTTCCCAGGCC	TAATACGACTCACTATAGCCTGGGAACCTCACACAA	TTCTAGCTCTAAAACTTGTGTGAGGTTCCCAGG	MATR3
MA8	Control3	**(chr5:137860000-138860000) 138278824-138278843 REV**	AAGTGTGCAGACTCCAGATG	CATCTGGAGTCTGCACACTT	TAATACGACTCACTATAGGTGTGCAGACTCCAGATG	TTCTAGCTCTAAAACCATCTGGAGTCTGCACAC	MATR3
MA9	Control4	**(chr5:137860000-138860000) 138278931-138278950 REV**	GGTGAGAATAGATGACACAG	CTGTGTCATCTATTCTCACC	TAATACGACTCACTATAGTGAGAATAGATGACACAG	TTCTAGCTCTAAAACCTGTGTCATCTATTCTCA	MATR3
MA10	Control5	**(chr5:137860000-138860000) 138278933-138278952 REV**	GTGGTGAGAATAGATGACAC	GTGTCATCTATTCTCACCAC	TAATACGACTCACTATAGGGTGAGAATAGATGACAC	TTCTAGCTCTAAAACGTGTCATCTATTCTCACC	MATR3
MA11	Promoter1	**(chr5:137860000-138860000) 138608459-138608478 REV**	TAAAACCAGGCAGATTGGGT	ACCCAATCTGCCTGGTTTTA	TAATACGACTCACTATAGAAACCAGGCAGATTGGGT	TTCTAGCTCTAAAACACCCAATCTGCCTGGTTT	MATR3
MA12	Promoter2	**(chr5:137860000-138860000) 138608898-138608917 REV**	CCATTGCCAGGTGAACCCCT	AGGGGTTCACCTGGCAATGG	TAATACGACTCACTATAGATTGCCAGGTGAACCCCT	TTCTAGCTCTAAAACAGGGGTTCACCTGGCAAT	MATR3
MA13	Promoter3	**(chr5:137860000-138860000) 138609219-138609238 FWD**	CGGCGAAGAATCCCACTGCA	TGCAGTGGGATTCTTCGCCG	TAATACGACTCACTATAGGCGAAGAATCCCACTGCA	TTCTAGCTCTAAAACTGCAGTGGGATTCTTCGC	MATR3
MA14	Promoter4	**(chr5:137860000-138860000) 138609310-138609329 FWD**	TGGAAAGATCCCGAAGACCG	CGGTCTTCGGGATCTTTCCA	TAATACGACTCACTATAGGAAAGATCCCGAAGACCG	TTCTAGCTCTAAAACCGGTCTTCGGGATCTTTC	MATR3
MA15	Promoter5	**(chr5:137860000-138860000) 138609317-138609336 FWD**	ATCCCGAAGACCGTGGACCA	TGGTCCACGGTCTTCGGGAT	TAATACGACTCACTATAGCCCGAAGACCGTGGACCA	TTCTAGCTCTAAAACTGGTCCACGGTCTTCGGG	MATR3
SO1	Loop-SNP1	**chr3:180560000-181560000 180928342-180928361 REV**	TAGGAACAGAAAATTATGCT	AGCATAATTTTCTGTTCCTA	TAATACGACTCACTATAGGGAACAGAAAATTATGCT	TTCTAGCTCTAAAACAGCATAATTTTCTGTTCC	SOX2
SO2	Loop-SNP2	**chr3:180560000-181560000 180928424-180928443 FWD**	ATGAAGTTCCACCCAGTTTG	CAAACTGGGTGGAACTTCAT	TAATACGACTCACTATAGGAAGTTCCACCCAGTTTG	TTCTAGCTCTAAAACCAAACTGGGTGGAACTTC	SOX2
SO3	Loop-SNP3	**chr3:180560000-181560000 180928425-180928444 FWD**	TGAAGTTCCACCCAGTTTGG	CCAAACTGGGTGGAACTTCA	TAATACGACTCACTATAGAAGTTCCACCCAGTTTGG	TTCTAGCTCTAAAACCCAAACTGGGTGGAACTT	SOX2
SO4	Loop-SNP4	**chr3:180560000-181560000 180928530-180928549 REV**	TATTAGGAAAAATCCGCCAG	CTGGCGGATTTTTCCTAATA	TAATACGACTCACTATAGTTAGGAAAAATCCGCCAG	TTCTAGCTCTAAAACCTGGCGGATTTTTCCTAA	SOX2
SO5	Loop-SNP5	**chr3:180560000-181560000 180928608-180928627 FWD**	ACTCTAAAGTTTCATCAGGA	TCCTGATGAAACTTTAGAGT	TAATACGACTCACTATAGTCTAAAGTTTCATCAGGA	TTCTAGCTCTAAAACTCCTGATGAAACTTTAGA	SOX2
SO6	Control1	**(chr3:180560000-181560000) 181180191-181180210 FWD**	TGGCTTAATAGTGAGTTACG	CGTAACTCACTATTAAGGCCA	TAATACGACTCACTATAGGCTTAATAGTGAGTTACG	TTCTAGCTCTAAAACCGTAACTCACTATTAAGC	SOX2
SO7	Control2	**(chr3:180560000-181560000) 181180624-181180643 FWD**	TGAGCCATAAAATAGTCACT	AGTGACTATTTTATGGCTCA	TAATACGACTCACTATAGAGCCATAAAATAGTCACT	TTCTAGCTCTAAAACAGTGACTATTTTATGGCT	SOX2
SO8	Control3	**(chr3:180560000-181560000) 181180998-181181017 FWD**	TCTGTTGTGTATCAGGACCA	TGGTCCTGATACACAACAGA	TAATACGACTCACTATAGTGTTGTGTATCAGGACCA	TTCTAGCTCTAAAACTGGTCCTGATACACAACA	SOX2
SO9	Control4	**(chr3:180560000-181560000) 181181058-181181077 FWD**	ACTTTAAAAAATCCCACACG	CGTGTGGGATTTTTTAAAGT	TAATACGACTCACTATAGTTTAAAAAATCCCACACG	TTCTAGCTCTAAAACCGTGTGGGATTTTTTAAA	SOX2
SO10	Control5	**(chr3:180560000-181560000) 181181105-181181124 FWD**	AATAACTAATAGCTACAAGt	aCTTGTAGCTATTAGTTATT	TAATACGACTCACTATAGTAACTAATAGCTACAAGt	TTCTAGCTCTAAAACaCTTGTAGCTATTAGTTA	SOX2
SO11	Promoter1	**(chr3:180560000-181560000) 181428583-181428602 FWD**	CACCACAATGGAAATCTACG	CGTAGATTTCCATTGTGGTG	TAATACGACTCACTATAGCCACAATGGAAATCTACG	TTCTAGCTCTAAAACCGTAGATTTCCATTGTGG	SOX2
SO12	Promoter2	**(chr3:180560000-181560000) 181428799-181428818 FWD**	CCGAAACCCTTCTTACGGGG	CCCCGTAAGAAGGGTTTCGG	TAATACGACTCACTATAGGAAACCCTTCTTACGGGG	TTCTAGCTCTAAAACCCCCGTAAGAAGGGTTTC	SOX2
SO13	Promoter3	**(chr3:180560000-181560000) 181429039-181429058 FWD**	TAAGAACAGAGCAAGTTACG	CGTAACTTGCTCTGTTCTTA	TAATACGACTCACTATAGAGAACAGAGCAAGTTACG	TTCTAGCTCTAAAACCGTAACTTGCTCTGTTCT	SOX2
SO14	Promoter4	**(chr3:180560000-181560000) 181429420-181429439 REV**	AAACAGCACTAAGACTACGT	ACGTAGTCTTAGTGCTGTTT	TAATACGACTCACTATAGACAGCACTAAGACTACGT	TTCTAGCTCTAAAACACGTAGTCTTAGTGCTGT	SOX2
SO15	Promoter5	**(chr3:180560000-181560000) 181429421-181429440 REV**	TAAACAGCACTAAGACTACG	CGTAGTCTTAGTGCTGTTTA	TAATACGACTCACTATAGAACAGCACTAAGACTACG	TTCTAGCTCTAAAACCGTAGTCTTAGTGCTGTT	SOX2

Cas9

ID	TYPE	COORDINATES (Region, gRNA)	gRNA SEQUENCE	REVERSE COMPLEMENT	IVT F1 PRIMER	IVT R1 PRIMER	GENE	Targeted credible SNP rsID
PC1	Loop-SNP-del1	**(chr5:140109000-141100000) 140142674-140142693 REV**	agagttatccttctgccaaa	tttggcagaaggataactct	TAATACGACTCACTATAGagttatccttctgccaaa	TTCTAGCTCTAAAACtttggcagaaggataact	PCDH	rs2563254
PC2	Loop-SNP-del2	**(chr5:140109000-141100000) 140142705-140142724 REV**	gaggccatatgtccaagatc	gatcttggacatatggcctc	TAATACGACTCACTATAGggccatatgtccaagatc	TTCTAGCTCTAAAACgatcttggacatatggcc	PCDH	rs2563254
PC3	Loop-SNP-del3	**(chr5:140109000-141100000) 140145209-140145228 FWD**	taaaatgtgtctcctgtaca	tgtacaggagacacattttag	TAATACGACTCACTATAGaaatgtgtctcctgtaca	TTCTAGCTCTAAAACtgtacaggagacacattt	PCDH	rs6891559
PC4	Loop-SNP-del4	**(chr5:140109000-141100000) 140146895-140146914 REV**	ttcccagatagggccaggcg	cgcctggccctatctgggaa	TAATACGACTCACTATAGcccagatagggccaggcg	TTCTAGCTCTAAAACcgcctggccctatctggg	PCDH	rs13168670
PC5	Loop-SNP-del5	**(chr5:140109000-141100000) 140147218-140147237 FWD**	ttgatgtctgtcattaaatg	catttaatgacagacatcaa	TAATACGACTCACTATAGgatgtctgtcattaaatg	TTCTAGCTCTAAAACcatttaatgacagacatc	PCDH	rs12659129

AS16	Loop-SNP-del1	**(chr12:103350000-103580000) 103573566-103573585 FWD**	ATTCAGAATGACG GATTTGG	CCAAATCCGTCAT TCTGAAT	TAATACGACTCACTATAGTCAG AATGACGGATTTGG	TTCTAGCTCTAAAACCCAAA TCCGTCATTCTGA	ASCL1	rs10860949
AS17	Loop-SNP-del2	**(chr12:103350000-103580000) 103573577-103573596 FWD**	CGGATTTGGTGGT GCAGTTG	CAACTGCACCACC AAATCCG	TAATACGACTCACTATAGGATT TGGTGGTGCAGTTG	TTCTAGCTCTAAAACCAACT GCACCACCAAATC	ASCL1	rs10860949
AS18	Loop-SNP-del3	**(chr12:103350000-103580000) 103574184-103574203 FWD**	TCAAAGAGAGTTG ACTTCGT	ACGAAGTCAACTC TCTTTGA	TAATACGACTCACTATAGAAAG AGAGTTGACTTCGT	TTCTAGCTCTAAAACACGAA GTCAACTCTCTTT	ASCL1	rs7306170
AS19	Loop-SNP-del4	**(chr12:103350000-103580000) 103575773-103575792 REV**	GAGTGGTCATTCT AATTACT	AGTAATTAGAATG ACCACTC	TAATACGACTCACTATAGGTG GTCATTCTAATTACT	TTCTAGCTCTAAAACAGTAA TTAGAATGACCAC	ASCL1	rs10860950
AS20	Loop-SNP-del5	**(chr12:103350000-103580000) 103575779-103575798 FWD**	TAGAATGACCACT CTTCTGG	CCAGAAGAGTGGT CATTCTA	TAATACGACTCACTATAGGAAT GACCACTCTTCTGG	TTCTAGCTCTAAAACCCAGA AGAGTGGTCATTC	ASCL1	rs10860950
MA16	Loop-SNP-del1	**chr5:137911382-137916382 137913870-137913889 FWD**	ggagaaaaggaaataag gaa	tccttctttccctccttccc	TAATACGACTCACTATAGagaaa aggaaataaggaa	TTCTAGCTCTAAAACtccttctttc cctccttc	MATR3	rs409273
MA17	Loop-SNP-del2	**chr5:137915326-137920326 137917793-137917812 FWD**	ggaattagtgcctttacaag	cttgtaaaggcactaattcc	TAATACGACTCACTATAGaattag tgcctttacaag	TTCTAGCTCTAAAACcttgtaaa ggcactaatt	MATR3	rs11957778
EF16	Loop-SNP-del1	**chrX:68374626-68379626 68377097-68377116 FWD**	TTTTCAGGAGCTG GCAAATT	AATTTGCCAGCTC CTGAAAA	TAATACGACTCACTATAGTTCA GGAGCTGGCAAATT	TTCTAGCTCTAAAACAATTT GCCAGCTCCTGAA	EFNB1	rs5937157
EF17	Loop-SNP-del2	**chrX:68374704-68379704 68377479-68377498 FWD**	AGAGCTGGAAAGG ACATTTC	GAAATGTCCTTTC CAGCTCT	TAATACGACTCACTATAGAGCT GGAAAGGACATTTC	TTCTAGCTCTAAAACGAAAT GTCCTTTCCAGCT	EFNB1	rs2361468
EF18	Loop-SNP-del3	**chrX:68374704-68379704 68377486-68377505 FWD**	GAAAGGACATTTC GGGTCAG	CTGACCCGAAATG TCCTTTC	TAATACGACTCACTATAGAAG GACATTTCGGGTCAG	TTCTAGCTCTAAAACCTGAC CCGAAATGTCCTT	EFNB1	rs2885287
EP16	Loop-SNP-del1	**(chr22:41400000-41600000) 41427455-41427474 FWD**	tcactgcgaactccgcccc c	gggggcggagttcgcagt ga	TAATACGACTCACTATAGactgc gaactccgccccc	TTCTAGCTCTAAAACgggggc ggagttcgcagt	EP300	rs9607768

Table 3.13 | Oligonucleotide sequences for gRNA in vitro transcription. Oligos used to generate gRNAs through in vitro transcription (IVT). ID, unique oligo name; Type, condition of the experiment; Name, contains coordinate and strand information (from Benchling); Sequence, gRNA sequence; F1 and R1 Primer, blue corresponds to T7 Promoter sequence (F1) and tracrRNA (R1) as per manual, p.9-11 (https://assets.thermofisher.com/TFS-Assets/LSG/manuals/geneart_precision_gRNA_synthesis_kit_man.pdf)

NAME	SEQUENCE
ASCL1-1-F	GGAGCTTCTCGACTTCACCA
ASCL1-1-R	AACGCCACTGACAAGAAAGC
EFNB1-1-F	CTCCCTGGGTGACTCTGATG
EFNB1-1-R	TGATGAGCAGGAAGATGACG
MATR3-3-F	AAGCAAGAGCTTGGACGTGT
MATR3-3-R	AACCATTGCCATTGCATCTT
SOX2-F	GCTAGTCTCCAAGCGACGAA
SOX2-R	GCAAGAAGCCTCTCCTTGAA
EP300-F	ATGACACACTGCCAGTCAGG
EP300-R	TTTTTGAGGGGGAGACACAC
PCDHA7-F	ACCCAAGACAGACCTCATGG
PCDHA7-R	AATGCCAGCCTCCTCTAGGT
PCDHA8-F	AGAGGGTGTGCTCTGGTGAG
PCDHA8-R	GAGGCAGAGTAACGCCAGTC
PCDHA10-F	GTAGATGTGGACGGGGAAGA
PCDHA10-R	CTCAGGGAGGCAGAGTAACG

Table 3.14 | qPCR primer sequences for RNA quantification in CRISPR experiments.
Primer sequences to measure gene expression of target genes from CRISPR experiments.

Genes located in 10kb bins corresponding to specific interactions anchored in PGC2 loci			Genes located in 10kb bins corresponding to specific interactions anchored in PGC2+ CLOZUK loci			Genes located in 10kb bins corresponding to specific interactions anchored in PGC2+ CLOZUK loci; with clusters of co-localized genes (e.g., PCDH) randomly removed such that only one gene is present		
NPC (309)	**Neurons (328)**	**Glia (165)**	**NPC (386)**	**Neurons (385)**	**Glia (201)**	**NPC (clusters removed)**	**Neurons (clusters removed)**	**Glia (clusters removed)**
AADAT	ABCB9	ABCB9	AADAT	ABCB9	ABCB9	AADAT	ABCB9	ABCB9
ADSS	ACTR5	ADSL	ABCD3	ACTG1P17	ADSL	ABCD3	ACTG1P17	ADSL
AKT3	ADAMTS9-AS2	ADSS	ACTR8	ACTR5	ADSS	ACTR8	ACTR5	ADSS
AMBRA1	ADAMTSL3	AKT3	ADGRG6	ACTR8	AKT3	ADGRG6	ACTR8	AKT3
ANKRD44	ADGRA3	ANKHD1	ADSS	ADAMTS9-AS2	ALMS1	ADSS	ADAMTS9-AS2	ALMS1
ANP32E	ADIG	ANKHD1-EIF4EBP3	AIG1	ADAMTSL3	ALMS1P1	AIG1	ADAMTSL3	ALMS1P1
AOX1	ADSS	ANKRD44	AKT3	ADGRA3	ANKHD1	AKT3	ADGRA3	ANKHD1
AOX2P	AKAP13	AOX1	ALAS1	ADIG	ANKHD1-EIF4EBP3	ALAS1	ADIG	ANKHD1-EIF4EBP3
AP1S3	AKT3	AOX2P	ALMS1	ADSS	ANKRD44	ALMS1	ADSS	ANKRD44
ARNT	AMBRA1	ARL6IP4	ALMS1P1	AIG1	AOX1	ALMS1P1	AIG1	AOX1
ASB5	ANKRD44	ARTN	AMBRA1	AKAP13	AOX2P	AMBRA1	AKAP13	AOX2P
ASTN1	ANP32E	ATF7IP2	ANKRD44	AKT3	AP3B2	ANKRD44	AKT3	AP3B2
ATP2A2	AP1S3	BOLA2	ANKRD45	ALAS1	ARL6IP4	ANKRD45	ALAS1	ARL6IP4
ATP6V0A2	ARHGAP40	BOLL	ANP32E	ALKBH5	ARTN	ANP32E	ALKBH5	ARTN
B4GALT2	ARL6IP4	C12orf65	AOX1	ALMS1	ATF7IP2	AOX1	ALMS1	ATF7IP2
BANK1	ASTN1	C1orf100	AOX2P	ALMS1P1	AUP1	AOX2P	ALMS1P1	AUP1
BNIPL	ATG13	C1orf101	AP1S3	AMBRA1	BOLA2	AP1S3	AMBRA1	BOLA2
BOLL	ATPAF2	C2orf69	ARL14	ANKRD44	BOLL	ARL14	ANKRD44	BOLL
BRD8	ATXN7	C7orf50	ARNT	ANKRD45	BRINP3	ARNT	ANKRD45	BRINP3
BRINP2	BAHD1	CCDC150	ASB5	ANP32E	C12orf65	ASB5	ANP32E	C12orf65
C11orf49	BAZ2A	CCDC39	ASH2L	AP1S3	C1orf100	ASH2L	AP1S3	C1orf100
C1orf101	BOLA2	CD46	ASTN1	AP3B2	C1orf101	ASTN1	AP3B2	C1orf101
C2orf69	BOLL	CNBD1	ATP2A2	ARHGAP15	C2orf69	ATP2A2	ARHGAP15	C2orf69
C4BPA	BRINP2	CNKSR2	ATP6V0A2	ARHGAP40	C7orf50	ATP6V0A2	ARHGAP40	C7orf50
C7orf31	C1orf100	CNTN4	B4GALT2	ARL6IP4	CCDC150	B4GALT2	ARL6IP4	CCDC150
CA6	C1orf101	COQ10B	BANK1	ASTN1	CCDC39	BANK1	ASTN1	CCDC39
CACNA1C	C1orf116	CSMD1	BNIPL	ATG13	CCT7	BNIPL	ATG13	CCT7
CACNA1D	C2orf69	CTNND1	BOLA3	ATPAF2	CD46	BOLA3	ATPAF2	CD46
CBR4	C2orf82	CUL3	BOLA3-AS1	ATXN7	CHRM3	BOLA3-AS1	ATXN7	CHRM3
CCDC150	C3orf49	DHX35	BOLL	BAHD1	CNBD1	BOLL	BAHD1	CNBD1
CCDC24	C7orf50	DNAH10	BRD8	BAZ2A	CNKSR2	BRD8	BAZ2A	CNKSR2
CCDC30	CA6	DOCK10	BRINP2	BNIP3L	CNTN4	BRINP2	BNIP3L	CNTN4
CCDC39	CA8	DPP4	BRINP3	BOLA2	COQ10B	BRINP3	BOLA2	COQ10B
CD46	CACNA1C	DPYD	C11orf49	BOLL	CPEB1-AS1	C11orf49	BOLL	CPEB1-AS1
CDC25C	CACNA1C-AS4	EP300	C1orf101	BRINP2	CSMD1	C1orf101	BRINP2	CSMD1
CEP170	CACNA1C-IT3	ERMAP	C2orf69	BRINP3	CTNND1	C2orf69	BRINP3	CTNND1
CETN3	CBR4	FAM57B	C4BPA	C1orf100	CUL3	C4BPA	C1orf100	CUL3
CFAP57	CCDC150	FAM83D	C7orf31	C1orf101	DAAM1	C7orf31	C1orf101	DAAM1
CHST12	CCDC39	FANCL	CA6	C1orf116	DHX35	CA6	C1orf116	DHX35
CLCN3	CCDC62	FAT2	CACNA1C	C2orf69	DNAH10	CACNA1C	C2orf69	DNAH10
CLIP1	CD46	FCAMR	CACNA1D	C2orf78	DOCK10	CACNA1D	C2orf78	DOCK10
CLU	CDC20	FLJ40288	CACNA2D3	C2orf82	DPP4	CACNA2D3	C2orf82	DPP4
CNKSR2	CDC25C	FLJ46066	CBR4	C3orf49	DPYD	CBR4	C3orf49	DPYD
COQ10B	CDHR3	FPGT-TNNI3K	CCDC150	C7orf50	DQX1	CCDC150	C7orf50	DQX1
CREB3L2	CEP162	GALNT10	CCDC24	CA6	EMX1	CCDC24	CA6	EMX1
CSMD1	CHRNA3	GBA3	CCDC30	CA8	EP300	CCDC30	CA8	EP300
CUL3	CNKSR2	GIGYF2	CCDC39	CACNA1C	ERMAP	CCDC39	CACNA1C	ERMAP
CYP26B1	CNNM2	GPER1	CCNH	CACNA1C-AS4	EXOC6B	CCNH	CACNA1C-AS4	EXOC6B
CYSTM1	COQ10B	GPM6A	CD46	CACNA1C-IT3	FAM57B	CD46	CACNA1C-IT3	FAM57B
DAZL	COX20	GPX5	CDC25C	CACNA1D	FAM83D	CDC25C	CACNA1D	FAM83D
DDX60L	CR1L	GRAMD1B	CENPL	CACYBP	FANCL	CENPL	CACYBP	FANCL
DENR	CREB3L1	GRIA1	CEP170	CBR4	FAT2	CEP170	CBR4	FAT2
DESI2	CSMD1	GRIN2A	CETN3	CCDC150	FCAMR	CETN3	CCDC150	FCAMR
DFNA5	CTNNA1	HCN1	CFAP57	CCDC39	FLJ40288	CFAP57	CCDC39	FLJ40288
DGKD	CTNND1	HECW2	CHDH	CCDC62	FLJ46066	CHDH	CCDC62	FLJ46066
DGKZ	CTRL	HIRIP3	CHST12	CCT7	FOXP1	CHST12	CCT7	FOXP1
DNAJC18	CTSS	HSPE1-MOB4	CLCN3	CD46	FPGT-TNNI3K	CLCN3	CD46	FPGT-TNNI3K
DNAJC19	CUL3	IMMP2L	CLIP1	CDC20	GALNT10	CLIP1	CDC20	GALNT10
DOCK10	CYP26B1	INO80E	CLU	CDC25C	GALNT15	CLU	CDC25C	GALNT15
DPH3	DDX60L	ITGA9	CMTR2	CDHR3	GBA3	CMTR2	CDHR3	GBA3
DPP4	DESI2	JUND	CNKSR2	CENPL	GIGYF2	CNKSR2	CENPL	GIGYF2
DPYD	DGKD	KCNJ13	COQ10B	CEP162	GPER1	COQ10B	CEP162	GPER1
DPYD-AS1	DGKZ	KDM4A	CREB3L2	CHDH	GPM6A	CREB3L2	CHDH	GPM6A
DUS2	DIS3L2	KMT5A	CSGALNACT1	CHRNA3	GPX5	CSGALNACT1	CHRNA3	GPX5
EBNA1BP2	DNAJC19	LINC01122	CSMD1	CNKSR2	GRAMD1B	CSMD1	CNKSR2	GRAMD1B
EFHD1	DOCK10	LOC100507091	CUL3	CNNM2	GRIA1	CUL3	CNNM2	GRIA1
EIF2B1	DPEP3	LOC606724	CYP26B1	COQ10B	GRIN2A	CYP26B1	COQ10B	GRIN2A
EIF3B	DPH3	LOC613038	CYSTM1	COX20	HCN1	CYSTM1	COX20	HCN1
EP300	DPYD	LOC729987	DARS2	CPEB1	HECW2	DARS2	CPEB1	HECW2
EPB41	DPYD-AS1	LRRIQ3	DAZL	CPEB1-AS1	HIRIP3	DAZL	CPEB1-AS1	HIRIP3
EPC2	DUS2	LSMEM1	DCP1A	CR1L	HSPE1-MOB4	DCP1A	CR1L	HSPE1-MOB4
EPHB1	DYSF	MAD1L1	DCTN1	CREB3L1	HTRA2	DCTN1	CREB3L1	HTRA2
ERI3	EFHD1	MAN2A1	DCTN1-AS1	CSMD1	HYDIN	DCTN1-AS1	CSMD1	HYDIN
ESAM	EIF4E2	MAPK3	DDX60L	CTNNA1	IBTK	DDX60L	CTNNA1	IBTK
ETF1	ELOVL1	MBTPS2	DENR	CTNND1	IMMP2L	DENR	CTNND1	IMMP2L
FAM114A2	ENO1	MEF2C	DESI2	CTRL	INO80E	DESI2	CTRL	INO80E
FAM13B	EPC2	MIR137	DFNA5	CTSS	ITGA9	DFNA5	CTSS	ITGA9
FAM53C	ERI3	MIR137HG	DGKD	CUL3	JUND	DGKD	CUL3	JUND
FANCL	FAM124B	MIR2682	DGKZ	CYP26B1	KCNJ13	DGKZ	CYP26B1	KCNJ13
FAT2	FAM53C	MIR29C	DNAJC18	DARS2	KDM4A	DNAJC18	DARS2	KDM4A

FCHSD1	FAM57B	MIR3188	DNAJC19	DCDC1	KLHL20	DNAJC19	DCDC1	KLHL20
FER	FANCL	MIR3714	DOCK10	DCDC5	KMT5A	DOCK10	DCDC5	KMT5A
FLJ46066	FLJ46066	MIR4304	DPH3	DCP1A	LINC00461	DPH3	DCP1A	LINC00461
FPGT-TNNI3K	FPGT-TNNI3K	MKL1	DPP4	DCTN1	LINC01122	DPP4	DCTN1	LINC01122
FTCDNL1	FSIP2	MOB4	DPYD	DDHD2	LOC100506023	DPYD	DDHD2	LOC100506023
FXR1	FTCDNL1	MPHOSPH9	DPYD-AS1	DDX60L	LOC100507091	DPYD-AS1	DDX60L	LOC100507091
GALNT10	FXR1	MPP6	DUS2	DESI2	LOC153910	DUS2	DESI2	LOC153910
GALNT15	G3BP1	MYLPF	EBNA1BP2	DGKD	LOC440704	EBNA1BP2	DGKD	LOC440704
GATAD2A	GABBR1	MYO1A	EFHD1	DGKZ	LOC606724	EFHD1	DGKZ	LOC606724
GBF1	GALNT10	NDUFA13	EIF2B1	DIS3L2	LOC613038	EIF2B1	DIS3L2	LOC613038
GFRA3	GATAD2A	NEK1	EIF3B	DNAJC19	LOC729987	EIF3B	DNAJC19	LOC729987
GIGYF2	GFRA3	NFATC3	EMB	DNAJC6	LOXL3	EMB	DNAJC6	LOXL3
GLRA1	GID4	NLGN4X	EMX1	DOCK10	LRRIQ3	EMX1	DOCK10	LRRIQ3
GPM6A	GIGYF2	NMUR2	EP300	DPEP3	LSMEM1	EP300	DPEP3	LSMEM1
GPX5	GLRA1	NRN1L	EPB41	DPH3	MAD1L1	EPB41	DPH3	MAD1L1
GPX6	GPM6A	NUGGC	EPC2	DPYD	MAN2A1	EPC2	DPYD	MAN2A1
GRIA1	GPX5	OGFOD2	EPHB1	DPYD-AS1	MAPK3	EPHB1	DPYD-AS1	MAPK3
GRIN2A	GRIA1	OR1S1	ERI3	DUS2	MBTPS2	ERI3	DUS2	MBTPS2
GTF2H3	GRIN2A	OR1S2	ESAM	DUSP11	MEF2C	ESAM	DUSP11	MEF2C
GTF3C3	GRM3	OR8B3	ETF1	DYSF	MIR137	ETF1	DYSF	MIR137
HCN1	H6PD	OR9Q1	EXOC6B	EFHD1	MIR137HG	EXOC6B	EFHD1	MIR137HG
HDAC3	HARS	PAPPA2	FAM114A2	EGR4	MIR2682	FAM114A2	EGR4	MIR2682
HECW2	HARS2	PDE6D	FAM13B	EIF4E2	MIR29C	FAM13B	EIF4E2	MIR29C
HPF1	HCN1	PITPNM2	FAM53C	ELOVL1	MIR3188	FAM53C	ELOVL1	MIR3188
HSPA9	HECW2	PLCL1	FANCL	EMX1	MIR3714	FANCL	EMX1	MIR3714
HSPD1	HSPA9	PLCL2	FAT2	ENO1	MIR4304	FAT2	ENO1	MIR4304
HSPE1-MOB4	IL19	PLD5	FCHSD1	EPC2	MIR9-2	FCHSD1	EPC2	MIR9-2
IK	INPP5D	PLEKHO1	FER	EPHA7	MKL1	FER	EPHA7	MKL1
IL19	IPO11	PLXNA4	FGFR1	EPN2	MOB1A	FGFR1	EPN2	MOB1A
IMMP2L	ITIH1	PPP1R16B	FLJ46066	EPN2-IT1	MOB4	FLJ46066	EPN2-IT1	MOB4
INPP5D	ITIH3	PSKH1	FPGT-TNNI3K	ERI3	MPHOSPH9	FPGT-TNNI3K	ERI3	MPHOSPH9
IQCE	ITIH4	PTCHD1-AS	FTCDNL1	EXOC6B	MPP6	FTCDNL1	EXOC6B	MPP6
IQCF1	KCNH7	PTPRF	FXR1	FAM124B	MYLPF	FXR1	FAM124B	MYLPF
ITGA9	KCNJ13	RERE	GAB1	FAM53C	MYO1A	GAB1	FAM53C	MYO1A
ITIH3	KDM4A	RFTN2	GALNT10	FAM57B	NAT8	GALNT10	FAM57B	NAT8
ITIH4	KMT2E	RILPL1	GALNT15	FANCL	NDUFA13	GALNT15	FANCL	NDUFA13
KCNH7	KMT2E-AS1	RNF220	GATAD2A	FBXO41	NEK1	GATAD2A	FBXO41	NEK1
KCNJ13	LAPTM5	SAP30L-AS1	GBF1	FLJ46066	NFATC3	GBF1	FLJ46066	NFATC3
KDM3B	LINC00634	SATB2	GFRA3	FOXP1	NLGN4X	GFRA3	FOXP1	NLGN4X
KDM4A	LINC00698	SBNO1	GIGYF2	FPGT-TNNI3K	NMUR2	GIGYF2	FPGT-TNNI3K	NMUR2
KMT2E	LINC01004	SDCCAG8	GLRA1	FRMD8	NRN1L	GLRA1	FRMD8	NRN1L
KNTC1	LINC01122	SF3B1	GPM6A	FSIP2	NUGGC	GPM6A	FSIP2	NUGGC
LINC01122	LOC100129620	SH3RF1	GPR52	FTCDNL1	OGFOD2	GPR52	FTCDNL1	OGFOD2
LOC100129620	LOC100652758	SLC4A10	GPX5	FXR1	OPCML	GPX5	FXR1	OPCML
LOC100130452	LOC339529	SLC6A9	GPX6	G3BP1	OR1S1	GPX6	G3BP1	OR1S1
LOC100130880	LOC339862	SLX1A	GRIA1	GABBR1	OR1S2	GRIA1	GABBR1	OR1S2
LOC100506085	LOC606724	SLX1B-SULT1A4	GRIN2A	GALNT10	OR8B3	GRIN2A	GALNT10	OR8B3
LOC100507091	LOC613038	SMG1P2	GTF2H3	GALNT15	OR9Q1	GTF2H3	GALNT15	OR9Q1
LOC100507140	LOC642423	SMS	GTF3C3	GAS5	PAPPA2	GTF3C3	GAS5	PAPPA2
LOC148696	LRP1	SNAP91	GUSBP5	GAS5-AS1	PDE4B	GUSBP5	GAS5-AS1	PDE4B
LOC339529	LRRIQ3	SOX2-OT	HCN1	GATAD2A	PDE6D	HCN1	GATAD2A	PDE6D
LOC339862	LYPD6B	ST3GAL3	HDAC3	GFRA3	PGM3	HDAC3	GFRA3	PGM3
LOC729987	MAD1L1	STK31	HECW2	GID4	PITPNM2	HECW2	GID4	PITPNM2
LRRC43	MAN2A1	TAOK2	HPF1	GIGYF2	PLCL1	HPF1	GIGYF2	PLCL1
LRRIQ3	MARS	TBC1D10B	HSPA9	GLRA1	PLCL2	HSPA9	GLRA1	PLCL2
LYPD6	MBTPS2	TBC1D5	HSPD1	GLT8D1	PLD5	HSPD1	GLT8D1	PLD5
MAD1L1	MDK	TCF4	HSPE1-MOB4	GNL3	PLEKHO1	HSPE1-MOB4	GNL3	PLEKHO1
MAN2A1	MED8	TMX2-CTNND1	HYDIN	GPM6A	PLXNA4	HYDIN	GPM6A	PLXNA4
MARS2	MIR137HG	TNFRSF9	IK	GPR52	PPP1R16B	IK	GPR52	PPP1R16B
MAU2	MIR22	TNNI3K	IL17RB	GPX5	PSKH1	IL17RB	GPX5	PSKH1
MBTPS2	MIR22HG	TNRC6B	IL19	GRIA1	PTCHD1-AS	IL19	GRIA1	PTCHD1-AS
MDK	MIR29C	TOX	IMMP2L	GRIN2A	PTPRF	IMMP2L	GRIN2A	PTPRF
MFAP3	MIR339	TRANK1	INO80B	GRM3	RABGAP1L	INO80B	GRM3	RABGAP1L
MIR137HG	MIR33B	TSSK6	INO80B-WBP1	H6PD	RC3H1	INO80B-WBP1	H6PD	RC3H1
MIR2682	MIR4677	TTYH3	INPP4B	HARS	RERE	INPP4B	HARS	RERE
MIR3160-1	MIR4688	UGT1A10	INPP5D	HARS2	RFTN2	INPP5D	HARS2	RFTN2
MIR3160-2	MIR548A2	UGT1A4	IQCE	HCN1	RILPL1	IQCE	HCN1	RILPL1
MIR4688	MOG	UGT1A5	IQCF1	HDAC2	RNF220	IQCF1	HDAC2	RNF220
MOB4	MPHOSPH9	UGT1A6	ITGA9	HECW2	SAP30L-AS1	ITGA9	HECW2	SAP30L-AS1
MPHOSPH9	MPL	UGT1A7	ITIH3	HIVEP2	SATB2	ITIH3	HIVEP2	SATB2
MPP6	MRAP2	UGT1A8	ITIH4	HSPA9	SBNO1	ITIH4	HSPA9	SBNO1
MUSTN1	MSL2	UGT1A9	KCNH7	IL17RB	SDCCAG8	KCNH7	IL17RB	SDCCAG8
MYO1A	MUSTN1	UNCX	KCNJ13	IL19	SF3B1	KCNJ13	IL19	SF3B1
MYOT	MYO1A	USP40	KDM3B	IL34	SFXN5	KDM3B	IL34	SFXN5
NCAN	NDUFA4L2	UTS2	KDM4A	INO80B	SH3RF1	KDM4A	INO80B	SH3RF1
NCK1	NEK1	VPS45	KLHL20	INO80B-WBP1	SLC4A10	KLHL20	INO80B-WBP1	SLC4A10
NEK1	NEK4	VRK2	KLHL29	INPP4B	SLC4A5	KLHL29	INPP4B	SLC4A5
NFATC3	NEMP1	WDFY1	KMT2E	INPP5D	SLC6A9	KMT2E	INPP5D	SLC6A9
NLGN4X	NFATC3	YJEFN3	KNTC1	IPO11	SLC9C2	KNTC1	IPO11	SLC9C2
NME5	NGEF	ZNF165	LINC00461	ITIH1	SLX1A	LINC00461	ITIH1	SLX1A
NMUR2	NLGN4X	ZNF48	LINC01122	ITIH3	SLX1B-SULT1A4	LINC01122	ITIH3	SLX1B-SULT1A4
NOSIP	NT5C2	ZNF691	LOC100129620	ITIH4	SMG1P2	LOC100129620	ITIH4	SMG1P2
NPY6R	NT5M	ZNF804A	LOC100130452	JAK1	SMS	LOC100130452	JAK1	SMS
NT5C2	NYAP2	ZSCAN16-AS1	LOC100130880	KAT5	SNAP91	LOC100130880	KAT5	SNAP91
NUBP1	OGFOD2	ZSWIM6	LOC100506023	KCNH7	SOX2-OT	LOC100506023	KCNH7	SOX2-OT
NYAP2	P2RX3		LOC100506085	KCNJ13	ST3GAL3	LOC100506085	KCNJ13	ST3GAL3
OR2B2	PALLD		LOC100507091	KDM4A	STAB2	LOC100507091	KDM4A	STAB2
ORC5	PARK7		LOC100507140	KLHL29	STK31	LOC100507140	KLHL29	STK31
OTUD7B	PBRM1		LOC148696	KMT2E	TAOK2	LOC148696	KMT2E	TAOK2
OXNAD1	PCCB		LOC339529	KMT2E-AS1	TBC1D10B	LOC339529	KMT2E-AS1	TBC1D10B
PAIP2B	PCDHA1		LOC339862	LAPTM5	TBC1D5	LOC339862	LAPTM5	TBC1D5
PALLD	PCDHA10		LOC440704	LEPR	TCF4	LOC440704	LEPR	TCF4
PCCB	PCDHA11		LOC729987	LINC00634	TMEM161B-AS1	LOC729987	LINC00634	TMEM161B-AS1
PCDHA1	PCDHA12		LOC730159	LINC00698	TMX2-CTNND1	LOC730159	LINC00698	TMX2-CTNND1
PCDHA10	PCDHA13		LRRC43	LINC01004	TNFRSF9	LRRC43	LINC01004	TNFRSF9
PCDHA11	PCDHA2		LRRIQ3	LINC01122	TNNI3K	LRRIQ3	LINC01122	TNNI3K
PCDHA12	PCDHA3		LYPD6	LOC100129620	TNRC6B	LYPD6	LOC100129620	TNRC6B

PCDHA13	PCDHA4		M1AP	LOC100506023	TOX	M1AP	LOC100506023	TOX
PCDHA2	PCDHA5		MAD1L1	LOC100652758	TRANK1	MAD1L1	LOC100652758	TRANK1
PCDHA3	PCDHA6		MAN2A1	LOC338963	TSSK6	MAN2A1	LOC338963	TSSK6
PCDHA4	PCDHA7		MARS2	LOC339529	TTYH3	MARS2	LOC339529	TTYH3
PCDHA5	PCDHA8		MAU2	LOC339862	UGT1A10	MAU2	LOC339862	UGT1A10
PCDHA6	PCDHA9		MBTPS2	LOC440704	UGT1A4	MBTPS2	LOC440704	UGT1A5
PCDHA7	PCDHAC1		MDK	LOC606724	UGT1A5	MDK	LOC606724	UGT1A8
PCDHA8	PCDHAC2		MEF2C	LOC613038	UGT1A6	MEF2C	LOC613038	UNCX
PCDHA9	PCDHGA1		MFAP3	LOC642423	UGT1A7	MFAP3	LOC642423	USP40
PCDHAC1	PCDHGA10		MIR137HG	LRP1	UGT1A8	MIR137HG	LRP1	UTS2
PCDHAC2	PCDHGA11		MIR2682	LRRIQ3	UGT1A9	MIR2682	LRRIQ3	VPS45
PCDHGA1	PCDHGA12		MIR3160-1	LYPD6B	UNCX	MIR3160-1	LYPD6B	VRK2
PCDHGA10	PCDHGA2		MIR3160-2	MAD1L1	USP40	MIR3160-2	MAD1L1	WDFY1
PCDHGA11	PCDHGA3		MIR4688	MAN2A1	UTS2	MIR4688	MAN2A1	YJEFN3
PCDHGA12	PCDHGA4		MOB4	MAP3K11	VPS45	MOB4	MAP3K11	ZNF165
PCDHGA2	PCDHGA5		MPHOSPH9	MARS	VRK2	MPHOSPH9	MARS	ZNF48
PCDHGA3	PCDHGA6		MPP6	MBTPS2	WDFY1	MPP6	MBTPS2	ZNF691
PCDHGA4	PCDHGA7		MSRA	MDK	YJEFN3	MSRA	MDK	ZNF804A
PCDHGA5	PCDHGA8		MUSTN1	MED8	ZNF165	MUSTN1	MED8	ZSCAN16-AS1
PCDHGA6	PCDHGA9		MYO1A	MEF2C	ZNF48	MYO1A	MEF2C	ZSWIM6
PCDHGA7	PCDHGB1		MYOT	MIR137HG	ZNF691	MYOT	MIR137HG	
PCDHGA8	PCDHGB2		NAT8	MIR22	ZNF804A	NAT8	MIR22	
PCDHGA9	PCDHGB3		NCAN	MIR22HG	ZSCAN16-AS1	NCAN	MIR22HG	
PCDHGB1	PCDHGB4		NCK1	MIR29C	ZSWIM6	NCK1	MIR29C	
PCDHGB2	PCDHGB5		NEK1	MIR339		NEK1	MIR339	
PCDHGB3	PCDHGB6		NFATC3	MIR33B		NFATC3	MIR33B	
PCDHGB4	PCDHGB7		NISCH	MIR4677		NISCH	MIR4677	
PCDHGB5	PCDHGC3		NLGN4X	MIR4688		NLGN4X	MIR4688	
PCDHGB6	PCGF6		NME5	MIR548A2		NME5	MIR548A2	
PCDHGB7	PDE4D		NMUR2	MIR548AI		NMUR2	MIR548AI	
PCGF6	PDE6D		NOSIP	MOG		NOSIP	MOG	
PDE4D	PEMT		NPY6R	MOGS		NPY6R	MOGS	
PER3	PEX5L		NT5C2	MPHOSPH9		NT5C2	MPHOSPH9	
PEX5L	PFKFB2		NTM	MPL		NTM	MPL	
PGAP1	PGAP1		NUBP1	MRAP2		NUBP1	MRAP2	
PHEX	PGM3		NYAP2	MRPS14		NYAP2	MRPS14	
PHF21A	PHEX		OPCML	MSL2		OPCML	MSL2	
PIH1D1	PHF21A		OR2B2	MSRA		OR2B2	MSRA	
PITPNM2	PITPNM2		ORC5	MUSTN1		ORC5	MUSTN1	
PKD2L2	PLCB2		OTUD7B	MYO15A		OTUD7B	MYO15A	
PLCL1	PLCL1		OXNAD1	MYO1A		OXNAD1	MYO1A	
PLCL2	PLCL2		PAIP2B	NAT8		PAIP2B	NAT8	
PLD5	PLD5		PALLD	NDUFA4L2		PALLD	NDUFA4L2	
PLEKHO1	PLEKHO1		PARP8	NEK1		PARP8	NEK1	
PLPPR5	PLPPR5		PCCB	NEK4		PCCB	NEK4	
PLXNA2	PLXNA2		PCDHA1	NEMP1		PCDHA1	NEMP1	
PODXL	POLDIP3		PCDHA10	NFATC3		PCDHAC1	NFATC3	
PPARGC1A	PPARGC1A		PCDHA11	NGEF		PCDHGA2	NGEF	
PPIH	PPP2R3A		PCDHA12	NLGN4X		PCGF6	NLGN4X	
PPP1CC	PRUNE1		PCDHA13	NT5C2		PDE4B	NT5C2	
PPP2R3A	PSKH1		PCDHA2	NT5M		PDE4D	NT5M	
PSMD14	PSMB10		PCDHA3	NTM		PDGFRL	NTM	
PTCHD1-AS	PSMG3		PCDHA4	NYAP2		PER3	NYAP2	
PTPRF	PSMG3-AS1		PCDHA5	OGFOD2		PEX5L	OGFOD2	
RAD54L2	PTBP2		PCDHA6	OPCML		PGAP1	OPCML	
RANBP10	PTCHD1-AS		PCDHA7	OXNAD1		PHEX	OXNAD1	
RANGAP1	PTPRF		PCDHA8	P2RX3		PHF21A	P2RX3	
RASAL2	R3HDM2		PCDHA9	PALLD		PIH1D1	PALLD	
REEP2	RAI1		PCDHAC1	PARK7		PITPNM2	PARK7	
RELL2	RASAL2		PCDHAC2	PBRM1		PKD2L2	PBRM1	
RERE	RERE		PCDHGA1	PCCB		PLCL1	PCCB	
RFTN1	RFTN1		PCDHGA10	PCDHA1		PLCL2	PCDHA1	
RFTN2	RFTN2		PCDHGA11	PCDHA10		PLD5	PCDHAC1	
RILPL2	RNF220		PCDHGA12	PCDHA11		PLEKHO1	PCDHGA10	
RPRD2	RPS6KA3		PCDHGA2	PCDHA12		PLPPR5	PCGF6	
RPS6KA3	RRP7BP		PCDHGA3	PCDHA13		PLXNA2	PCNX3	
RSRC2	RTN4RL1		PCDHGA4	PCDHA2		POC1A	PDE4B	
SAP30L-AS1	RTN4RL2		PCDHGA5	PCDHA3		PODXL	PDE4D	
SATB1	RWDD2A		PCDHGA6	PCDHA4		PPARGC1A	PDE6D	
SATB2	SAP30L-AS1		PCDHGA7	PCDHA5		PPIH	PEMT	
SBNO1	SATB2		PCDHGA8	PCDHA6		PPM1L	PEX5L	
SCG2	SBNO1		PCDHGA9	PCDHA7		PPP1CC	PFKFB2	
SDCCAG8	SDCCAG8		PCDHGB1	PCDHA8		PPP2R2A	PGAP1	
SEC16B	SEMA3G		PCDHGB2	PCDHA9		PPP2R3A	PGM3	
SF3B1	SEMA6C		PCDHGB3	PCDHAC1		PRDX6	PHEX	
SGO2	SERHL2		PCDHGB4	PCDHAC2		PSD3	PHF21A	
SH3RF1	SERPINE2		PCDHGB5	PCDHGA1		PSMD14	PITPNM2	
SLC17A7	SERPINF1		PCDHGB6	PCDHGA10		PTCHD1-AS	PLCB2	
SLC25A42	SF3B1		PCDHGB7	PCDHGA11		PTPRF	PLCL1	
SLC2A5	SFMBT1		PCGF6	PCDHGA12		RABGAP1L	PLCL2	
SLC35G2	SFXN2		PDE4B	PCDHGA2		RAD54L2	PLD5	
SLC39A8	SH3GL3		PDE4D	PCDHGA3		RANBP10	PLEKHO1	
SLC4A10	SH3PXD2A		PDGFRL	PCDHGA4		RANGAP1	PLPPR5	
SLC6A9	SLC12A4		PER3	PCDHGA5		RASA1	PLXNA2	
SMPX	SLC25A33		PEX5L	PCDHGA6		RASAL2	POC1A	
SMS	SLC2A5		PGAP1	PCDHGA7		RC3H1	POLDIP3	
SNTB2	SLC2A7		PHEX	PCDHGA8		REEP2	PPARGC1A	
SNX7	SLC35G2		PHF21A	PCDHGA9		RELL2	PPP2R2A	
SNX8	SLC4A10		PIH1D1	PCDHGB1		RERE	PPP2R3A	
SOX2	SLX1A		PITPNM2	PCDHGB2		RFT1	PRADC1	
SOX2-OT	SLX1B-SULT1A4		PKD2L2	PCDHGB3		RFTN1	PRKCB	
SPATA24	SMG1P2		PLCL1	PCDHGB4		RFTN2	PRKD1	
SRPK2	SMG6		PLCL2	PCDHGB5		RILPL2	PRUNE1	
ST3GAL3	SMPX		PLD5	PCDHGB6		RPRD2	PSD3	
STAG1	SMS		PLEKHO1	PCDHGB7		RPS6KA3	PSKH1	
STK31	SNAP91		PLPPR5	PCDHGC3		RSRC2	PSMB10	
SUGP1	SNORD32B		PLXNA2	PCGF6		RTKN	PSMG3	

SUGP2	SNORD63		POC1A	PCNX3		SAP30L-AS1	PSMG3-AS1	
TAF5	SNX7		PODXL	PDE4B		SATB1	PTBP2	
TBC1D5	SNX8		PPARGC1A	PDE4D		SATB2	PTCHD1-AS	
TCF4	SOX2		PPIH	PDE6D		SBNO1	PTPRF	
TCTN2	SOX2-OT		PPM1L	PEMT		SCG2	R3HDM2	
TFAMP1	SPARC		PPP1CC	PEX5L		SDCCAG8	RAI1	
TKT	SPATS2L		PPP2R2A	PFKFB2		SEC16B	RASAL2	
TMEFF2	SREBF1		PPP2R3A	PGAP1		SEMA3G	RERE	
TMEM110	SRPK2		PRDX6	PGM3		SEMA4F	RFTN1	
TMEM110-MUSTN1	SRR		PSD3	PHEX		SERPINC1	RFTN2	
TMTC1	SSRP1		PSMD14	PHF21A		SF3B1	RNF220	
TNNI3K	ST3GAL3		PTCHD1-AS	PITPNM2		SFMBT1	RPS6KA3	
TRANK1	STAC3		PTPRF	PLCB2		SFXN5	RRP7BP	
TSPAN9	STAG1		RABGAP1L	PLCL1		SGO2	RTN4RL1	
TVP23A	STAT6		RAD54L2	PLCL2		SH3RF1	RTN4RL2	
TYW5	SUFU		RANBP10	PLD5		SLC17A7	RWDD2A	
UGT1A10	SYNPR		RANGAP1	PLEKHO1		SLC25A42	SAP30L-AS1	
UGT1A6	TAF5		RASA1	PLPPR5		SLC2A5	SATB2	
UGT1A7	TBC1D5		RASAL2	PLXNA2		SLC35G2	SBNO1	
UGT1A8	THOC7		RC3H1	POC1A		SLC39A8	SDCCAG8	
UGT1A9	TLCD2		REEP2	POLDIP3		SLC4A10	SEMA3G	
UTP4	TLR9		RELL2	PPARGC1A		SLC4A5	SEMA6C	
VPS33A	TMEFF2		RERE	PPP2R2A		SLC6A9	SERHL2	
VPS45	TMEM110		RFT1	PPP2R3A		SLC9C2	SERPINE2	
VRK2	TMEM110-MUSTN1		RFTN1	PRADC1		SMPX	SERPINF1	
VSIG2	TMEM161A		RFTN2	PRKCB		SMS	SF3B1	
WBP1L	TMX2-CTNND1		RILPL2	PRKD1		SNTB2	SFMBT1	
WDR55	TNNI3K		RPRD2	PRUNE1		SNX7	SFXN2	
ZBED9	TOM1L2		RPS6KA3	PSD3		SNX8	SH3GL3	
ZBTB18	TOX		RSRC2	PSKH1		SOX2	SH3PXD2A	
ZKSCAN4	TTC14		RTKN	PSMB10		SOX2-OT	SLC12A4	
ZNF391	TWF2		SAP30L-AS1	PSMG3		SPATA24	SLC25A33	
ZNF804A	TYW5		SATB1	PSMG3-AS1		SRPK2	SLC2A5	
ZSCAN23	USP40		SATB2	PTBP2		ST3GAL3	SLC2A7	
ZSWIM6	VPS37B		SBNO1	PTCHD1-AS		STAG1	SLC35G2	
	VPS45		SCG2	PTPRF		STK31	SLC4A10	
	VRK2		SDCCAG8	R3HDM2		SUGP1	SLX1A	
	VTRNA1-1		SEC16B	RAI1		SUGP2	SLX1B-SULT1A4	
	WBP1L		SEMA3G	RASAL2		TAF5	SMG1P2	
	WDFY1		SEMA4F	RERE		TBC1D5	SMG6	
	WDR81		SERPINC1	RFTN1		TCF4	SMPX	
	YY2		SF3B1	RFTN2		TCTN2	SMS	
	ZBED9		SFMBT1	RNF220		TET3	SNAP91	
	ZBTB18		SFXN5	RPS6KA3		TFAMP1	SNORD32B	
	ZDHHC5		SGO2	RRP7BP		TKT	SNORD63	
	ZFP57		SH3RF1	RTN4RL1		TMEFF2	SNX7	
	ZKSCAN3		SLC17A7	RTN4RL2		TMEM110	SNX8	
	ZMAT2		SLC25A42	RWDD2A		TMEM110-MUSTN1	SOX2	
	ZNF592		SLC2A5	SAP30L-AS1		TMTC1	SOX2-OT	
	ZNF804A		SLC35G2	SATB2		TNFSF4	SPARC	
	ZSCAN12		SLC39A8	SBNO1		TNN	SPATS2L	
	ZSCAN23		SLC4A10	SDCCAG8		TNNC1	SREBF1	
	ZSCAN31		SLC4A5	SEMA3G		TNNI3K	SRPK2	
	ZSWIM6		SLC6A9	SEMA6C		TNR	SRR	
			SLC9C2	SERHL2		TRANK1	SSRP1	
			SMPX	SERPINE2		TSPAN9	ST3GAL3	
			SMS	SERPINF1		TVP23A	STAC3	
			SNTB2	SF3B1		TYW5	STAG1	
			SNX7	SFMBT1		UGT1A5	STAT6	
			SNX8	SFXN2		UGT1A8	SUFU	
			SOX2	SH3GL3		UGT1A10	SYNPR	
			SOX2-OT	SH3PXD2A		USP38	TAF5	
			SPATA24	SLC12A4		UTP4	TBC1D5	
			SRPK2	SLC25A33		VPS33A	THOC7	
			ST3GAL3	SLC2A5		VPS45	TLCD2	
			STAG1	SLC2A7		VRK2	TLR9	
			STK31	SLC35G2		VSIG2	TMEFF2	
			SUGP1	SLC4A10		WBP1L	TMEM110	
			SUGP2	SLX1A		WDR55	TMEM110-MUSTN1	
			TAF5	SLX1B-SULT1A4		WHSC1L1	TMEM161A	
			TBC1D5	SMG1P2		ZBED9	TMX2-CTNND1	
			TCF4	SMG6		ZBTB18	TNNI3K	
			TCTN2	SMPX		ZBTB37	TOM1L2	
			TET3	SMS		ZEB2	TOX	
			TFAMP1	SNAP91		ZKSCAN4	TTC14	
			TKT	SNORD32B		ZNF391	TWF2	
			TMEFF2	SNORD63		ZNF804A	TYW5	
			TMEM110	SNX7		ZSCAN23	USP40	
			TMEM110-MUSTN1	SNX8		ZSWIM6	VPS37B	
			TMTC1	SOX2				
			TNFSF4	SOX2-OT				
			TNN	SPARC				
			TNNC1	SPATS2L				
			TNNI3K	SREBF1				
			TNR	SRPK2				
			TRANK1	SRR				
			TSPAN9	SSRP1				
			TVP23A	ST3GAL3				
			TYW5	STAC3				
			UGT1A10	STAG1				
			UGT1A6	STAT6				
			UGT1A7	SUFU				
			UGT1A8	SYNPR				
			UGT1A9	TAF5				
			USP38	TBC1D5				

			UTP4	THOC7				
			VPS33A	TLCD2				
			VPS45	TLR9				
			VRK2	TMEFF2				
			VSIG2	TMEM110				
			WBP1L	TMEM110-MUSTN1				
			WDR55	TMEM161A				
			WHSC1L1	TMX2-CTNND1				
			ZBED9	TNNI3K				
			ZBTB18	TOM1L2				
			ZBTB37	TOX				
			ZEB2	TTC14				
			ZKSCAN4	TWF2				
			ZNF391	TYW5				
			ZNF804A	USP40				
			ZSCAN23	VPS37B				
			ZSWIM6					

Table 3.15 | Genes located in cell-type specific SZ risk associated chromosomal contacts.

NPC

GOID	GOTerm	Ontology Source	Term PValue	Term PValue Corrected with Bonferroni step down	Neg Log (Pcorr)	% Associated Genes	Nr. Genes
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-33	1.6E-30	29.79588	22.09	36.00
GO:0098609	cell-cell adhesion	GO_BiologicalProcess-GOA_23.02.2017_10h01	750.0E-18	68.0E-15	13.1674911	5.64	47.00
GO:0051552	flavone metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.5E-9	310.0E-9	6.50863831	83.33	5.00
GO:0032594	protein transport within lipid bilayer	GO_BiologicalProcess-GOA_23.02.2017_10h01	32.0E-9	2.8E-6	5.55284197	62.50	5.00
GO:1904224	negative regulation of glucuronosyltransferase activity	GO_BiologicalProcess-GOA_23.02.2017_10h01	32.0E-9	2.8E-6	5.55284197	62.50	5.00
GO:0052697	xenobiotic glucuronidation	GO_BiologicalProcess-GOA_23.02.2017_10h01	71.0E-9	6.2E-6	5.20760831	55.56	5.00
GO:0052696	flavonoid glucuronidation	GO_BiologicalProcess-GOA_23.02.2017_10h01	28.0E-6	2.5E-3	2.60205999	60.00	3.00
GO:0000053	Ascorbate and aldarate metabolism	KEGG_01.03.2017	37.0E-6	3.1E-3	2.50863831	18.52	5.00
GO:0000830	Retinol metabolism	KEGG_01.03.2017	40.0E-6	3.4E-3	2.46852108	10.77	7.00
GO:0000983	Drug metabolism	KEGG_01.03.2017	48.0E-6	4.0E-3	2.39794001	13.04	6.00
GO:0000040	Pentose and glucuronate interconversions	KEGG_01.03.2017	110.0E-6	9.7E-3	2.01322827	14.71	5.00
GO:0015020	glucuronosyltransferase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	110.0E-6	9.7E-3	2.01322827	14.71	5.00
GO:0010677	negative regulation of cellular carbohydrate metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	130.0E-6	11.0E-3	1.95860731	10.91	6.00
GO:0000860	Porphyrin and chlorophyll metabolism	KEGG_01.03.2017	320.0E-6	26.0E-3	1.58502665	11.90	5.00
GO:0000210	Glucuronidation	REACTOME_Pathways_01.03.2017	350.0E-6	28.0E-3	1.55284197	16.67	4.00
GO:0004720	Long-term potentiation	KEGG_01.03.2017	400.0E-6	31.0E-3	1.50863831	8.96	6.00
GO:0005031	Amphetamine addiction	KEGG_01.03.2017	430.0E-6	34.0E-3	1.46852108	8.82	6.00
GO:0000982	Drug metabolism	KEGG_01.03.2017	500.0E-6	39.0E-3	1.40893539	8.57	6.00
GO:0051001	negative regulation of nitric-oxide synthase activity	GO_BiologicalProcess-GOA_23.02.2017_10h01	580.0E-6	44.0E-3	1.35654732	25.00	3.00
GO:0004728	Dopaminergic synapse	KEGG_01.03.2017	590.0E-6	44.0E-3	1.35654732	6.15	8.00
GO:0007062	sister chromatid cohesion	GO_BiologicalProcess-GOA_23.02.2017_10h01	620.0E-6	45.0E-3	1.34678749	6.11	8.00
GO:0004024	cAMP signaling pathway	KEGG_01.03.2017	610.0E-6	45.0E-3	1.34678749	5.05	10.00

NEURON

GOID	GOTerm	Ontology Source	Term PValue	Term PValue Corrected with Bonferroni step down	Neg log (Pcorr)	% Associated Genes	Nr. Genes
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	GO_BiologicalProcess-GOA_23.02.2017_10h01	57.0E-33	13.0E-30	28.9E+0	22.09	36.00
GO:0005509	calcium ion binding	GO_MolecularFunction-GOA_23.02.2017_10h01	970.0E-18	230.0E-15	12.6E+0	6.02	45.00
GO:0007399	nervous system development	GO_BiologicalProcess-GOA_23.02.2017_10h01	23.0E-15	5.7E-12	11.2E+0	3.51	83.00
GO:0098609	cell-cell adhesion	GO_BiologicalProcess-GOA_23.02.2017_10h01	52.0E-15	12.0E-12	10.9E+0	5.40	45.00
GO:0007267	cell-cell signaling	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.9E-12	950.0E-12	9.0E+0	3.76	63.00
GO:0031226	intrinsic component of plasma membrane	GO_CellularComponent-GOA_23.02.2017_10h01	72.0E-12	17.0E-9	7.8E+0	3.58	61.00
GO:0005887	integral component of plasma membrane	GO_CellularComponent-GOA_23.02.2017_10h01	290.0E-12	69.0E-9	7.2E+0	3.56	58.00
GO:0044459	plasma membrane part	GO_CellularComponent-GOA_23.02.2017_10h01	21.0E-9	5.1E-6	5.3E+0	2.79	77.00
GO:0043169	cation binding	GO_MolecularFunction-GOA_23.02.2017_10h01	22.0E-9	5.3E-6	5.3E+0	2.44	107.00
GO:0048731	system development	GO_BiologicalProcess-GOA_23.02.2017_10h01	35.0E-9	8.2E-6	5.1E+0	2.38	112.00
GO:0007275	multicellular organism development	GO_BiologicalProcess-GOA_23.02.2017_10h01	990.0E-9	230.0E-6	3.6E+0	2.21	117.00
GO:0036477	somatodendritic compartment	GO_CellularComponent-GOA_23.02.2017_10h01	15.0E-6	3.5E-3	2.5E+0	3.67	27.00
GO:0030425	dendrite	GO_CellularComponent-GOA_23.02.2017_10h01	87.0E-6	20.0E-3	1.7E+0	3.91	20.00
GO:0043005	neuron projection	GO_CellularComponent-GOA_23.02.2017_10h01	99.0E-6	22.0E-3	1.7E+0	3.07	32.00
GO:0044297	cell body	GO_CellularComponent-GOA_23.02.2017_10h01	190.0E-6	44.0E-3	1.4E+0	3.68	20.00

GLIA

GOID	GOTerm	Ontology Source	Term PValue	Term PValue Corrected with Bonferroni step down	Neg log (Pcorr)	% Associated Genes	Nr. Genes
GO:0052697	xenobiotic glucuronidation	GO_BiologicalProcess-GOA_23.02.2017_10h01	31.0E-15	930.0E-15	12.0E+0	77.78	7.00
GO:1904224	negative regulation of glucuronosyltransferase activity	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.5E-12	100.0E-12	10.0E+0	75.00	6.00
GO:0000053	Ascorbate and aldarate metabolism	KEGG_01.03.2017	690.0E-12	19.0E-9	7.7E+0	25.93	7.00
GO:0000983	Drug metabolism	KEGG_01.03.2017	1.2E-9	33.0E-9	7.5E+0	17.39	8.00
GO:0000040	Pentose and glucuronate interconversions	KEGG_01.03.2017	4.0E-9	100.0E-9	7.0E+0	20.59	7.00
GO:0000210	Glucuronidation	REACTOME_Pathways_01.03.2017	15.0E-9	380.0E-9	6.4E+0	25.00	6.00
GO:0000860	Porphyrin and chlorophyll metabolism	KEGG_01.03.2017	19.0E-9	460.0E-9	6.3E+0	16.67	7.00
GO:0000830	Retinol metabolism	KEGG_01.03.2017	21.0E-9	490.0E-9	6.3E+0	12.31	8.00
GO:0000982	Drug metabolism	KEGG_01.03.2017	39.0E-9	860.0E-9	6.1E+0	11.43	8.00
GO:0000140	Steroid hormone biosynthesis	KEGG_01.03.2017	190.0E-9	4.0E-6	5.4E+0	12.07	7.00
GO:0000980	Metabolism of xenobiotics by cytochrome P450	KEGG_01.03.2017	1.0E-6	21.0E-6	4.7E+0	9.46	7.00
GO:0005204	Chemical carcinogenesis	KEGG_01.03.2017	2.1E-6	40.0E-6	4.4E+0	8.54	7.00
GO:0045833	negative regulation of lipid metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.7E-6	66.0E-6	4.2E+0	7.87	7.00
GO:0008194	UDP-glycosyltransferase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	11.0E-6	190.0E-6	3.7E+0	5.44	8.00
GO:0010675	regulation of cellular carbohydrate metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	92.0E-6	1.4E-3	2.9E+0	4.79	7.00
GO:0000205	Phase II conjugation	REACTOME_Pathways_01.03.2017	100.0E-6	1.5E-3	2.8E+0	5.83	6.00
GO:0000208	Cytosolic sulfonation of small molecules	REACTOME_Pathways_01.03.2017	100.0E-6	1.5E-3	2.8E+0	5.83	6.00
GO:0042440	pigment metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	190.0E-6	2.7E-3	2.6E+0	6.76	5.00
GO:0001972	retinoic acid binding	GO_MolecularFunction-GOA_23.02.2017_10h01	270.0E-6	3.6E-3	2.4E+0	16.67	3.00
GO:0043197	dendritic spine	GO_CellularComponent-GOA_23.02.2017_10h01	460.0E-6	5.5E-3	2.3E+0	4.41	6.00
GO:0004520	Adherens junction	KEGG_01.03.2017	1.8E-3	14.0E-3	1.9E+0	5.56	4.00
GO:0004662	B cell receptor signaling pathway	KEGG_01.03.2017	14.0E-3	14.0E-3	1.9E+0	4.23	3.00
GO:0004720	Long-term potentiation	KEGG_01.03.2017	1.4E-3	15.0E-3	1.8E+0	5.97	4.00
GO:0001464	Interactions of neurexins and neuroligins at synapses	REACTOME_Pathways_01.03.2017	1.7E-3	15.0E-3	1.8E+0	5.63	4.00
GO:0001465	Protein-protein interactions at synapses	REACTOME_Pathways_01.03.2017	1.7E-3	15.0E-3	1.8E+0	5.63	4.00
GO:0001073	Unblocking of NMDA receptor, glutamate binding and activation	REACTOME_Pathways_01.03.2017	1.7E-3	17.0E-3	1.8E+0	9.09	3.00
GO:0001082	CREB phosphorylation through the activation of CaMKII	REACTOME_Pathways_01.03.2017	1.7E-3	17.0E-3	1.8E+0	9.09	3.00
GO:0001733	SALM protein interactions at the synapses	REACTOME_Pathways_01.03.2017	1.7E-3	17.0E-3	1.8E+0	9.09	3.00
GO:0051310	metaphase plate congression	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.2E-3	18.0E-3	1.7E+0	5.77	3.00
GO:0000521	Neurophilin interactions with VEGF and VEGFR	REACTOME_Pathways_01.03.2017	3.2E-3	22.0E-3	1.7E+0	7.32	3.00
GO:0000973	Sema3A PAK dependent Axon repulsion	REACTOME_Pathways_01.03.2017	3.2E-3	22.0E-3	1.7E+0	7.32	3.00

GO:0000974	SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion	REACTOME_Pathways_01.03.2017	3.2E-3	22.0E-3	1.7E+0	7.32	3.00
GO:0001093	Signal transduction by L1	REACTOME_Pathways_01.03.2017	3.2E-3	22.0E-3	1.7E+0	7.32	3.00
GO:0021879	forebrain neuron differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-3	23.0E-3	1.6E+0	4.62	3.00
GO:0005211	Renal cell carcinoma	KEGG_01.03.2017	11.0E-3	23.0E-3	1.6E+0	4.62	3.00
GO:0000556	Nuclear Events (kinase and transcription factor activation)	REACTOME_Pathways_01.03.2017	5.9E-3	23.0E-3	1.6E+0	5.88	3.00
GO:0000558	ERK/MAPK targets	REACTOME_Pathways_01.03.2017	5.9E-3	23.0E-3	1.6E+0	5.88	3.00
GO:0000564	CREB phosphorylation	REACTOME_Pathways_01.03.2017	5.9E-3	23.0E-3	1.6E+0	5.88	3.00
GO:0001116	MAPK targets/ Nuclear events mediated by MAP kinases	REACTOME_Pathways_01.03.2017	5.9E-3	23.0E-3	1.6E+0	5.88	3.00
GO:0001291	CD209 (DC-SIGN) signaling	REACTOME_Pathways_01.03.2017	5.9E-3	23.0E-3	1.6E+0	5.88	3.00
GO:0000493	Pre-NOTCH Processing in Golgi	REACTOME_Pathways_01.03.2017	4.1E-3	25.0E-3	1.6E+0	6.67	3.00
GO:0000494	Pre-NOTCH Expression and Processing	REACTOME_Pathways_01.03.2017	4.1E-3	25.0E-3	1.6E+0	6.67	3.00
GO:0030071	regulation of mitotic metaphase/anaphase transition	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.3E-3	26.0E-3	1.6E+0	6.12	3.00

NPC (PGC2 + CLOZUK)

GOID	GOTerm	Ontology Source	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Neg log (FDR)	% Associated Genes	Nr. Genes
GO:0032594	protein transport within lipid bilayer	GO_BiologicalProcess-GOA_23.02.2017_10h01	59.0E-9	8.6E-6	5.1E+0	62.50	5.00
GO:0051552	flavone metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	82.0E-6	6.0E-3	2.2E+0	50.00	3.00
GO:0004261	Adrenergic signaling in cardiomyocytes	KEGG_01.03.2017	120.0E-6	6.1E-3	2.2E+0	6.94	10.00
GO:0004261	Adrenergic signaling in cardiomyocytes	KEGG_01.03.2017	120.0E-6	6.1E-3	2.2E+0	6.94	10.00
GO:0004261	Adrenergic signaling in cardiomyocytes	KEGG_01.03.2017	120.0E-6	6.1E-3	2.2E+0	6.94	10.00
GO:0052697	xenobiotic glucuronidation	GO_BiologicalProcess-GOA_23.02.2017_10h01	330.0E-6	6.9E-3	2.2E+0	33.33	3.00
GO:0007062	sister chromatid cohesion	GO_BiologicalProcess-GOA_23.02.2017_10h01	290.0E-6	7.2E-3	2.1E+0	6.87	9.00
GO:0005891	voltage-gated calcium channel complex	GO_CellularComponent-GOA_23.02.2017_10h01	510.0E-6	7.4E-3	2.1E+0	12.20	5.00
GO:0004024	cAMP signaling pathway	KEGG_01.03.2017	410.0E-6	7.5E-3	2.1E+0	5.56	11.00
GO:0004024	cAMP signaling pathway	KEGG_01.03.2017	410.0E-6	7.5E-3	2.1E+0	5.56	11.00
GO:0014888	striated muscle adaptation	GO_BiologicalProcess-GOA_23.02.2017_10h01	570.0E-6	7.6E-3	2.1E+0	11.90	5.00
GO:0014888	striated muscle adaptation	GO_BiologicalProcess-GOA_23.02.2017_10h01	570.0E-6	7.6E-3	2.1E+0	11.90	5.00
GO:0044450	microtubule organizing center part	GO_CellularComponent-GOA_23.02.2017_10h01	480.0E-6	7.8E-3	2.1E+0	5.88	10.00
GO:0004728	Dopaminergic synapse	KEGG_01.03.2017	280.0E-6	8.1E-3	2.1E+0	6.92	9.00
GO:0004728	Dopaminergic synapse	KEGG_01.03.2017	280.0E-6	8.1E-3	2.1E+0	6.92	9.00
GO:0004728	Dopaminergic synapse	KEGG_01.03.2017	280.0E-6	8.1E-3	2.1E+0	6.92	9.00
GO:0004720	Long-term potentiation	KEGG_01.03.2017	760.0E-6	9.3E-3	2.0E+0	8.96	6.00
GO:0005031	Amphetamine addiction	KEGG_01.03.2017	820.0E-6	9.3E-3	2.0E+0	8.82	6.00
GO:0004720	Long-term potentiation	KEGG_01.03.2017	760.0E-6	9.3E-3	2.0E+0	8.96	6.00
GO:0005031	Amphetamine addiction	KEGG_01.03.2017	820.0E-6	9.3E-3	2.0E+0	8.82	6.00
GO:0005814	centriole	GO_CellularComponent-GOA_23.02.2017_10h01	960.0E-6	10.0E-3	2.0E+0	6.45	8.00
GO:0051393	alpha-actinin binding	GO_MolecularFunction-GOA_23.02.2017_10h01	270.0E-6	10.0E-3	2.0E+0	13.89	5.00
GO:0035641	locomotory exploration behavior	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.6E-3	14.0E-3	1.9E+0	20.00	3.00
GO:0031011	Ino80 complex	GO_CellularComponent-GOA_23.02.2017_10h01	1.6E-3	14.0E-3	1.9E+0	20.00	3.00
GO:0000819	sister chromatid segregation	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.5E-3	14.0E-3	1.9E+0	4.74	11.00
GO:0000775	chromosome, centromeric region	GO_CellularComponent-GOA_23.02.2017_10h01	1.6E-3	14.0E-3	1.9E+0	5.03	10.00
GO:0050651	dermatan sulfate proteoglycan biosynthetic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.6E-3	14.0E-3	1.9E+0	20.00	3.00
GO:0070603	SWI/SNF superfamily-type complex	GO_CellularComponent-GOA_23.02.2017_10h01	2.0E-3	15.0E-3	1.8E+0	7.41	6.00
GO:0006446	regulation of translational initiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.1E-3	15.0E-3	1.8E+0	7.32	6.00
GO:0005410	Hypertrophic cardiomyopathy (HCM)	KEGG_01.03.2017	2.3E-3	15.0E-3	1.8E+0	7.23	6.00
GO:0015020	glucuronosyltransferase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	2.1E-3	15.0E-3	1.8E+0	11.76	4.00
GO:0005410	Hypertrophic cardiomyopathy (HCM)	KEGG_01.03.2017	2.3E-3	15.0E-3	1.8E+0	7.23	6.00
GO:0072523	purine-containing compound catabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.5E-3	15.0E-3	1.8E+0	8.62	5.00
GO:0002190	cap-independent translational initiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.0E-3	16.0E-3	1.8E+0	18.75	3.00
GO:1903510	mucopolysaccharide metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.5E-3	17.0E-3	1.8E+0	5.83	7.00
GO:0030220	platelet formation	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.9E-3	17.0E-3	1.8E+0	16.67	3.00
GO:0019932	second-messenger-mediated signaling	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.6E-3	17.0E-3	1.8E+0	4.23	11.00
GO:0009066	aspartate family amino acid metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.3E-3	18.0E-3	1.7E+0	8.06	5.00
GO:0005414	Dilated cardiomyopathy	KEGG_01.03.2017	3.5E-3	18.0E-3	1.7E+0	6.67	6.00
GO:0014898	cardiac muscle hypertrophy in response to stress	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.9E-3	18.0E-3	1.7E+0	15.00	3.00
GO:0000830	Retinol metabolism	KEGG_01.03.2017	4.1E-3	18.0E-3	1.7E+0	7.69	5.00
GO:0072509	divalent inorganic cation transmembrane transporter activity	GO_MolecularFunction-GOA_23.02.2017_10h01	3.1E-3	18.0E-3	1.7E+0	4.92	9.00
GO:0005414	Dilated cardiomyopathy	KEGG_01.03.2017	3.5E-3	18.0E-3	1.7E+0	6.67	6.00
GO:0055119	relaxation of cardiac muscle	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.4E-3	18.0E-3	1.7E+0	15.79	3.00
GO:0071871	response to epinephrine	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.4E-3	18.0E-3	1.7E+0	15.79	3.00
GO:0072509	divalent inorganic cation transmembrane transporter activity	GO_MolecularFunction-GOA_23.02.2017_10h01	3.1E-3	18.0E-3	1.7E+0	4.92	9.00
GO:0004602	glutathione peroxidase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	4.5E-3	19.0E-3	1.7E+0	14.29	3.00
GO:0048872	homeostasis of number of cells	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.8E-3	19.0E-3	1.7E+0	4.07	11.00
GO:0030641	regulation of cellular pH	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.5E-3	19.0E-3	1.7E+0	6.32	6.00
GO:0035966	response to topologically incorrect protein	GO_BiologicalProcess-GOA_23.02.2017_10h01	4.3E-3	19.0E-3	1.7E+0	4.69	9.00
GO:0000140	RORA activates gene expression	REACTOME_Pathways_01.03.2017	5.3E-3	19.0E-3	1.7E+0	7.25	5.00
GO:0000141	BMAL1:CLOCK,NPAS2 activates circadian gene expression	REACTOME_Pathways_01.03.2017	5.3E-3	19.0E-3	1.7E+0	7.25	5.00
GO:0000979	Circadian Clock	REACTOME_Pathways_01.03.2017	5.3E-3	19.0E-3	1.7E+0	7.25	5.00
GO:0000860	Porphyrin and chlorophyll metabolism	KEGG_01.03.2017	4.7E-3	19.0E-3	1.7E+0	9.52	4.00
GO:0004022	cGMP-PKG signaling pathway	KEGG_01.03.2017	5.3E-3	19.0E-3	1.7E+0	4.91	8.00
GO:0004022	cGMP-PKG signaling pathway	KEGG_01.03.2017	5.3E-3	19.0E-3	1.7E+0	4.91	8.00
GO:0006555	methionine metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.2E-3	20.0E-3	1.7E+0	13.64	3.00
GO:0043501	skeletal muscle adaptation	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.7E-3	20.0E-3	1.7E+0	12.50	3.00
GO:0043501	skeletal muscle adaptation	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.7E-3	20.0E-3	1.7E+0	12.50	3.00
GO:0001464	Interactions of neurexins and neuroligins at synapses	REACTOME_Pathways_01.03.2017	5.9E-3	20.0E-3	1.7E+0	7.04	5.00
GO:0001465	Protein-protein interactions at synapses	REACTOME_Pathways_01.03.2017	5.9E-3	20.0E-3	1.7E+0	7.04	5.00
GO:0035637	multicellular organismal signaling	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.5E-3	20.0E-3	1.7E+0	4.39	9.00
GO:0043501	skeletal muscle adaptation	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.7E-3	20.0E-3	1.7E+0	12.50	3.00
GO:0030552	cAMP binding	GO_MolecularFunction-GOA_23.02.2017_10h01	6.7E-3	20.0E-3	1.7E+0	12.50	3.00
GO:0043197	dendritic spine	GO_CellularComponent-GOA_23.02.2017_10h01	6.9E-3	21.0E-3	1.7E+0	5.15	7.00
GO:0002062	chondrocyte differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	7.7E-3	21.0E-3	1.7E+0	5.66	6.00
GO:0045948	positive regulation of translational initiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.9E-3	21.0E-3	1.7E+0	13.04	3.00
GO:0000776	kinetochore	GO_CellularComponent-GOA_23.02.2017_10h01	7.4E-3	21.0E-3	1.7E+0	5.07	7.00
GO:0051788	response to misfolded protein	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.9E-3	21.0E-3	1.7E+0	13.04	3.00
GO:0030206	chondroitin sulfate biosynthetic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	7.5E-3	21.0E-3	1.7E+0	12.00	3.00
GO:0042440	pigment metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	7.1E-3	21.0E-3	1.7E+0	6.76	5.00
GO:0071300	cellular response to retinoic acid	GO_BiologicalProcess-GOA_23.02.2017_10h01	7.1E-3	21.0E-3	1.7E+0	6.76	5.00
GO:0000983	Drug metabolism	KEGG_01.03.2017	6.5E-3	21.0E-3	1.7E+0	8.70	4.00
GO:0042440	pigment metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	7.1E-3	21.0E-3	1.7E+0	6.76	5.00
GO:0005412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	KEGG_01.03.2017	6.3E-3	21.0E-3	1.7E+0	6.94	5.00
GO:0019722	calcium-mediated signaling	GO_BiologicalProcess-GOA_23.02.2017_10h01	6.3E-3	21.0E-3	1.7E+0	4.76	8.00
GO:0004931	Insulin resistance	KEGG_01.03.2017	8.1E-3	22.0E-3	1.7E+0	5.61	6.00
GO:0006941	striated muscle contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.3E-3	22.0E-3	1.7E+0	4.55	8.00
GO:0022406	membrane docking	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.4E-3	23.0E-3	1.6E+0	4.44	8.00

GO:0004260	Cardiac muscle contraction	KEGG_01.03.2017	8.8E-3	23.0E-3	1.6E+0	6.41	5.00
GO:0008287	protein serine/threonine phosphatase complex	GO_CellularComponent-GOA_23.02.2017_10h01	9.4E-3	23.0E-3	1.6E+0	7.84	4.00
GO:0000053	Ascorbate and aldarate metabolism	KEGG_01.03.2017	9.3E-3	23.0E-3	1.6E+0	11.11	3.00
GO:0060078	regulation of postsynaptic membrane potential	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.0E-3	23.0E-3	1.6E+0	4.90	7.00
GO:0004260	Cardiac muscle contraction	KEGG_01.03.2017	8.8E-3	23.0E-3	1.6E+0	6.41	5.00
GO:0008287	protein serine/threonine phosphatase complex	GO_CellularComponent-GOA_23.02.2017_10h01	9.4E-3	23.0E-3	1.6E+0	7.84	4.00
GO:0061337	cardiac conduction	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.0E-3	23.0E-3	1.6E+0	4.90	7.00
GO:0055117	regulation of cardiac muscle contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.8E-3	23.0E-3	1.6E+0	6.41	5.00
GO:0005231	excitatory extracellular ligand-gated ion channel activity	GO_MolecularFunction-GOA_23.02.2017_10h01	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0000223	Mitochondrial biogenesis	REACTOME_Pathways_01.03.2017	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0000260	Mitochondrial transcription initiation	REACTOME_Pathways_01.03.2017	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0000673	Transcriptional activation of mitochondrial biogenesis	REACTOME_Pathways_01.03.2017	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0001680	Transcription from mitochondrial promoters	REACTOME_Pathways_01.03.2017	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0000223	Mitochondrial biogenesis	REACTOME_Pathways_01.03.2017	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0000260	Mitochondrial transcription initiation	REACTOME_Pathways_01.03.2017	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0000673	Transcriptional activation of mitochondrial biogenesis	REACTOME_Pathways_01.03.2017	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0001680	Transcription from mitochondrial promoters	REACTOME_Pathways_01.03.2017	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0008194	UDP-glycosyltransferase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	10.0E-3	24.0E-3	1.6E+0	4.76	7.00
GO:0015276	ligand-gated ion channel activity	GO_MolecularFunction-GOA_23.02.2017_10h01	10.0E-3	24.0E-3	1.6E+0	4.79	7.00
GO:0005231	excitatory extracellular ligand-gated ion channel activity	GO_MolecularFunction-GOA_23.02.2017_10h01	10.0E-3	24.0E-3	1.6E+0	7.69	4.00
GO:0000480	Glutathione metabolism	KEGG_01.03.2017	11.0E-3	26.0E-3	1.6E+0	7.41	4.00
GO:0070252	actin-mediated cell contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-3	26.0E-3	1.6E+0	5.17	6.00
GO:0070252	actin-mediated cell contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	11.0E-3	26.0E-3	1.6E+0	5.17	6.00
GO:0006479	protein methylation	GO_BiologicalProcess-GOA_23.02.2017_10h01	12.0E-3	27.0E-3	1.6E+0	4.23	8.00
GO:0001491	Regulation of TP53 Activity through Acetylation	REACTOME_Pathways_01.03.2017	12.0E-3	27.0E-3	1.6E+0	10.00	3.00
GO:0035967	cellular response to topologically incorrect protein	GO_BiologicalProcess-GOA_23.02.2017_10h01	12.0E-3	27.0E-3	1.6E+0	4.58	7.00
GO:0000579	Chondroitin sulfate biosynthesis	REACTOME_Pathways_01.03.2017	12.0E-3	27.0E-3	1.6E+0	10.00	3.00
GO:0000580	Dermatan sulfate biosynthesis	REACTOME_Pathways_01.03.2017	12.0E-3	27.0E-3	1.6E+0	10.00	3.00
GO:0000585	CS/DS degradation	REACTOME_Pathways_01.03.2017	12.0E-3	27.0E-3	1.6E+0	10.00	3.00
GO:0000852	Defective CHST3 causes SEDCJD	REACTOME_Pathways_01.03.2017	12.0E-3	27.0E-3	1.6E+0	10.00	3.00
GO:0000853	Defective CHST14 causes EDS, musculocontractural type	REACTOME_Pathways_01.03.2017	12.0E-3	27.0E-3	1.6E+0	10.00	3.00
GO:0000854	Defective CHSY1 causes TPBS	REACTOME_Pathways_01.03.2017	12.0E-3	27.0E-3	1.6E+0	10.00	3.00
GO:0001964	startle response	GO_BiologicalProcess-GOA_23.02.2017_10h01	12.0E-3	27.0E-3	1.6E+0	10.00	3.00
GO:0086065	cell communication involved in cardiac conduction	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-3	27.0E-3	1.6E+0	7.14	4.00
GO:0086065	cell communication involved in cardiac conduction	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-3	27.0E-3	1.6E+0	7.14	4.00
GO:0032733	positive regulation of interleukin-10 production	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-3	28.0E-3	1.6E+0	9.68	3.00
GO:0032733	positive regulation of interleukin-10 production	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-3	28.0E-3	1.6E+0	9.68	3.00
GO:0032733	positive regulation of interleukin-10 production	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-3	28.0E-3	1.6E+0	9.68	3.00
GO:0006937	regulation of muscle contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	13.0E-3	28.0E-3	1.6E+0	4.52	7.00
GO:0051216	cartilage development	GO_BiologicalProcess-GOA_23.02.2017_10h01	15.0E-3	30.0E-3	1.5E+0	4.08	8.00
GO:0001782	B cell homeostasis	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-3	30.0E-3	1.5E+0	9.38	3.00
GO:0001065	Golgi Associated Vesicle Biogenesis	REACTOME_Pathways_01.03.2017	16.0E-3	30.0E-3	1.5E+0	4.80	6.00
GO:0001474	Signaling by BRAF and RAF fusions	REACTOME_Pathways_01.03.2017	16.0E-3	30.0E-3	1.5E+0	4.80	6.00
GO:0001477	Oncogenic MAPK signaling	REACTOME_Pathways_01.03.2017	16.0E-3	30.0E-3	1.5E+0	4.80	6.00
GO:0030173	Integral component of Golgi membrane	GO_CellularComponent-GOA_23.02.2017_10h01	14.0E-3	30.0E-3	1.5E+0	6.90	4.00
GO:0032743	positive regulation of interleukin-2 production	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-3	30.0E-3	1.5E+0	9.38	3.00
GO:0004110	Cell cycle	KEGG_01.03.2017	15.0E-3	31.0E-3	1.5E+0	4.84	6.00
GO:0006370	7-methylguanosine mRNA capping	GO_BiologicalProcess-GOA_23.02.2017_10h01	16.0E-3	31.0E-3	1.5E+0	9.09	3.00
GO:0000287	Rev-mediated nuclear export of HIV RNA	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	5.49	5.00
GO:0000350	Export of Viral Ribonucleoproteins from Nucleus	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	5.49	5.00
GO:0000358	NEP/NS2 Interacts with the Cellular Export Machinery	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	5.49	5.00
GO:0000417	Interactions of Rev with host cellular proteins	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	5.49	5.00
GO:0000685	Downregulation of TGF-beta receptor signaling	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	5.49	5.00
GO:0000686	TGF-beta receptor signaling activates SMADs	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	5.49	5.00
GO:0001124	HuR (ELAVL1) binds and stabilizes mRNA	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	5.49	5.00
GO:0001549	Cyclin A/B1 associated events during G2/M transition	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	5.49	5.00
GO:0001073	Unblocking of NMDA receptor, glutamate binding and activation	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	9.09	3.00
GO:0001082	CREB phosphorylation through the activation of CaMKII	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	9.09	3.00
GO:0001733	SALM protein interactions at the synapses	REACTOME_Pathways_01.03.2017	16.0E-3	31.0E-3	1.5E+0	9.09	3.00
GO:0005244	voltage-gated ion channel activity	GO_MolecularFunction-GOA_23.02.2017_10h01	16.0E-3	31.0E-3	1.5E+0	4.02	8.00
GO:0005244	voltage-gated ion channel activity	GO_MolecularFunction-GOA_23.02.2017_10h01	16.0E-3	31.0E-3	1.5E+0	4.02	8.00
GO:0071354	cellular response to interleukin-6	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	32.0E-3	1.5E+0	8.82	3.00
GO:0098739	Import across plasma membrane	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	32.0E-3	1.5E+0	8.82	3.00
GO:0001662	behavioral fear response	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	32.0E-3	1.5E+0	8.82	3.00
GO:0042100	B cell proliferation	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	32.0E-3	1.5E+0	5.38	5.00
GO:0000040	Pentose and glucuronate interconversions	KEGG_01.03.2017	17.0E-3	32.0E-3	1.5E+0	8.82	3.00
GO:0043620	regulation of DNA-templated transcription in response to stress	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-3	33.0E-3	1.5E+0	4.69	6.00
GO:0002823	negative regulation of adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-3	33.0E-3	1.5E+0	8.57	3.00
GO:0008376	acetylgalactosaminyltransferase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	19.0E-3	33.0E-3	1.5E+0	8.57	3.00
GO:0050777	negative regulation of immune response	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-3	34.0E-3	1.5E+0	4.62	6.00
GO:0050777	negative regulation of immune response	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-3	34.0E-3	1.5E+0	4.62	6.00
GO:0004713	Circadian entrainment	KEGG_01.03.2017	20.0E-3	35.0E-3	1.5E+0	5.21	5.00
GO:0004713	Circadian entrainment	KEGG_01.03.2017	20.0E-3	35.0E-3	1.5E+0	5.21	5.00
GO:0006901	vesicle coating	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-3	36.0E-3	1.4E+0	6.15	4.00
GO:0005211	Renal cell carcinoma	KEGG_01.03.2017	21.0E-3	36.0E-3	1.4E+0	6.15	4.00
GO:0032729	positive regulation of interferon-gamma production	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-3	36.0E-3	1.4E+0	6.15	4.00
GO:0015698	Inorganic anion transport	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-3	36.0E-3	1.4E+0	4.12	7.00
GO:0030968	endoplasmic reticulum unfolded protein response	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-3	36.0E-3	1.4E+0	4.55	6.00
GO:0032729	positive regulation of interferon-gamma production	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-3	36.0E-3	1.4E+0	6.15	4.00
GO:0001756	somitogenesis	GO_BiologicalProcess-GOA_23.02.2017_10h01	24.0E-3	39.0E-3	1.4E+0	5.88	4.00
GO:0048199	vesicle targeting, to, from or within Golgi	GO_BiologicalProcess-GOA_23.02.2017_10h01	24.0E-3	39.0E-3	1.4E+0	5.88	4.00
GO:0006986	response to unfolded protein	GO_BiologicalProcess-GOA_23.02.2017_10h01	24.0E-3	39.0E-3	1.4E+0	4.02	7.00
GO:1901863	positive regulation of muscle tissue development	GO_BiologicalProcess-GOA_23.02.2017_10h01	24.0E-3	39.0E-3	1.4E+0	5.88	4.00
GO:0004360	Axon guidance	KEGG_01.03.2017	24.0E-3	40.0E-3	1.4E+0	4.00	7.00
GO:0015103	inorganic anion transmembrane transporter activity	GO_MolecularFunction-GOA_23.02.2017_10h01	24.0E-3	40.0E-3	1.4E+0	4.38	6.00
GO:0006821	chloride transport	GO_BiologicalProcess-GOA_23.02.2017_10h01	27.0E-3	42.0E-3	1.4E+0	4.81	5.00
GO:0000982	Drug metabolism	KEGG_01.03.2017	27.0E-3	42.0E-3	1.4E+0	5.71	4.00
GO:0005033	Nicotine addiction	KEGG_01.03.2017	27.0E-3	42.0E-3	1.4E+0	7.50	3.00
GO:0006821	chloride transport	GO_BiologicalProcess-GOA_23.02.2017_10h01	27.0E-3	42.0E-3	1.4E+0	4.81	5.00
GO:0018023	peptidyl-lysine trimethylation	GO_BiologicalProcess-GOA_23.02.2017_10h01	28.0E-3	44.0E-3	1.4E+0	7.32	3.00
GO:0001072	Post NMDA receptor activation events	REACTOME_Pathways_01.03.2017	28.0E-3	44.0E-3	1.4E+0	7.32	3.00
GO:0001083	CREB phosphorylation through the activation of Ras	REACTOME_Pathways_01.03.2017	28.0E-3	44.0E-3	1.4E+0	7.32	3.00
GO:0001085	Activation of NMDA receptor upon glutamate binding and postsynaptic events	REACTOME_Pathways_01.03.2017	28.0E-3	44.0E-3	1.4E+0	7.32	3.00
GO:0001086	Ras activation uopn Ca2+ infux through NMDA receptor	REACTOME_Pathways_01.03.2017	28.0E-3	44.0E-3	1.4E+0	7.32	3.00
GO:0004520	Adherens junction	KEGG_01.03.2017	29.0E-3	45.0E-3	1.3E+0	5.56	4.00

GO:0000777	condensed chromosome kinetochore	GO_CellularComponent-GOA_23.02.2017_10h01	30.0E-3	46.0E-3	1.3E+0	4.67	5.00
GO:0000218	Signaling by NOTCH	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000549	Signaling by NOTCH1	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000550	Signaling by NOTCH2	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000654	Activated NOTCH1 Transmits Signal to the Nucleus	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000743	Signaling by NOTCH1 PEST Domain Mutants in Cancer	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000744	Signaling by NOTCH1 in Cancer	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000746	Constitutive Signaling by NOTCH1 PEST Domain Mutants	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000756	Signaling by NOTCH1 HD Domain Mutants in Cancer	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000757	Constitutive Signaling by NOTCH1 HD Domain Mutants	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000765	Signaling by NOTCH1 HD+PEST Domain Mutants in Cancer	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000766	Constitutive Signaling by NOTCH1 HD+PEST Domain Mutants	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:0000768	NOTCH2 Activation and Transmission of Signal to the Nucleus	REACTOME_Pathways_01.03.2017	31.0E-3	46.0E-3	1.3E+0	4.63	5.00
GO:1903779	regulation of cardiac conduction	GO_BiologicalProcess-GOA_23.02.2017_10h01	31.0E-3	46.0E-3	1.3E+0	5.48	4.00
GO:0051489	regulation of filopodium assembly	GO_BiologicalProcess-GOA_23.02.2017_10h01	32.0E-3	47.0E-3	1.3E+0	6.98	3.00
GO:0006904	vesicle docking involved in exocytosis	GO_BiologicalProcess-GOA_23.02.2017_10h01	32.0E-3	47.0E-3	1.3E+0	6.98	3.00
GO:0008135	translation factor activity, RNA binding	GO_MolecularFunction-GOA_23.02.2017_10h01	32.0E-3	47.0E-3	1.3E+0	4.59	5.00
GO:0006661	phosphatidylinositol biosynthetic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	34.0E-3	48.0E-3	1.3E+0	4.05	6.00
GO:0021549	cerebellum development	GO_BiologicalProcess-GOA_23.02.2017_10h01	35.0E-3	48.0E-3	1.3E+0	4.50	5.00
GO:0031571	mitotic G1 DNA damage checkpoint	GO_BiologicalProcess-GOA_23.02.2017_10h01	35.0E-3	48.0E-3	1.3E+0	5.26	4.00
GO:0048678	response to axon injury	GO_BiologicalProcess-GOA_23.02.2017_10h01	35.0E-3	48.0E-3	1.3E+0	5.26	4.00
GO:0051304	chromosome separation	GO_BiologicalProcess-GOA_23.02.2017_10h01	33.0E-3	48.0E-3	1.3E+0	5.33	4.00
GO:0004973	Carbohydrate digestion and absorption	KEGG_01.03.2017	34.0E-3	48.0E-3	1.3E+0	6.82	3.00
GO:0000309	Complement cascade	REACTOME_Pathways_01.03.2017	34.0E-3	48.0E-3	1.3E+0	6.82	3.00
GO:0000310	Lectin pathway of complement activation	REACTOME_Pathways_01.03.2017	34.0E-3	48.0E-3	1.3E+0	6.82	3.00
GO:0000311	Initial triggering of complement	REACTOME_Pathways_01.03.2017	34.0E-3	48.0E-3	1.3E+0	6.82	3.00
GO:0000313	Creation of C4 and C2 activators	REACTOME_Pathways_01.03.2017	34.0E-3	48.0E-3	1.3E+0	6.82	3.00
GO:0000759	Ficolins bind to repetitive carbohydrate structures on the target cell surface	REACTOME_Pathways_01.03.2017	34.0E-3	48.0E-3	1.3E+0	6.82	3.00
GO:0003022	Basal transcription factors	KEGG_01.03.2017	36.0E-3	49.0E-3	1.3E+0	6.67	3.00
GO:0033555	multicellular organismal response to stress	GO_BiologicalProcess-GOA_23.02.2017_10h01	36.0E-3	49.0E-3	1.3E+0	5.19	4.00
GO:0032580	Golgi cisterna membrane	GO_CellularComponent-GOA_23.02.2017_10h01	36.0E-3	49.0E-3	1.3E+0	5.19	4.00
GO:0000493	Pre-NOTCH Processing in Golgi	REACTOME_Pathways_01.03.2017	36.0E-3	49.0E-3	1.3E+0	6.67	3.00
GO:0000494	Pre-NOTCH Expression and Processing	REACTOME_Pathways_01.03.2017	36.0E-3	49.0E-3	1.3E+0	6.67	3.00

NEURON (PGC2 + CLOZUK)

GOID	GOTerm	Ontology Source	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Neg log (FDR)	% Associated Genes	Nr. Genes
GO:0097346	INO80-type complex	GO_CellularComponent-GOA_23.02.2017_10h01	830.0E-9	52.0E-6	4.3E+0	27.27	6.00
GO:0070603	SWI/SNF superfamily-type complex	GO_CellularComponent-GOA_23.02.2017_10h01	520.0E-9	66.0E-6	4.2E+0	12.35	10.00
GO:0044297	cell body	GO_CellularComponent-GOA_23.02.2017_10h01	5.2E-6	220.0E-6	3.7E+0	4.41	24.00
GO:0006338	chromatin remodeling	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.2E-6	260.0E-6	3.6E+0	7.50	12.00
GO:0098794	postsynapse	GO_CellularComponent-GOA_23.02.2017_10h01	21.0E-6	380.0E-6	3.4E+0	4.55	20.00
GO:0043025	neuronal cell body	GO_CellularComponent-GOA_23.02.2017_10h01	21.0E-6	450.0E-6	3.3E+0	4.40	21.00
GO:0030425	dendrite	GO_CellularComponent-GOA_23.02.2017_10h01	19.0E-6	490.0E-6	3.3E+0	4.30	22.00
GO:0004728	Dopaminergic synapse	KEGG_01.03.2017	37.0E-6	600.0E-6	3.2E+0	7.69	10.00
GO:0004728	Dopaminergic synapse	KEGG_01.03.2017	37.0E-6	600.0E-6	3.2E+0	7.69	10.00
GO:0042578	phosphoric ester hydrolase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	45.0E-6	630.0E-6	3.2E+0	4.62	18.00
GO:0031011	Ino80 complex	GO_CellularComponent-GOA_23.02.2017_10h01	69.0E-6	880.0E-6	3.1E+0	26.67	4.00
GO:0032526	response to retinoic acid	GO_BiologicalProcess-GOA_23.02.2017_10h01	92.0E-6	1.0E-3	3.0E+0	7.69	9.00
GO:0071300	cellular response to retinoic acid	GO_BiologicalProcess-GOA_23.02.2017_10h01	150.0E-6	1.6E-3	2.8E+0	9.46	7.00
GO:0090568	nuclear transcriptional repressor complex	GO_CellularComponent-GOA_23.02.2017_10h01	290.0E-6	2.8E-3	2.6E+0	13.16	5.00
GO:0090568	nuclear transcriptional repressor complex	GO_CellularComponent-GOA_23.02.2017_10h01	290.0E-6	2.8E-3	2.6E+0	13.16	5.00
GO:0071229	cellular response to acid chemical	GO_BiologicalProcess-GOA_23.02.2017_10h01	370.0E-6	3.4E-3	2.5E+0	5.39	11.00
GO:0005891	voltage-gated calcium channel complex	GO_CellularComponent-GOA_23.02.2017_10h01	420.0E-6	3.5E-3	2.5E+0	12.20	5.00
GO:0005891	voltage-gated calcium channel complex	GO_CellularComponent-GOA_23.02.2017_10h01	420.0E-6	3.5E-3	2.5E+0	12.20	5.00
GO:0005031	Amphetamine addiction	KEGG_01.03.2017	660.0E-6	4.2E-3	2.4E+0	8.82	6.00
GO:0014069	postsynaptic density	GO_CellularComponent-GOA_23.02.2017_10h01	630.0E-6	4.2E-3	2.4E+0	5.07	11.00
GO:0005031	Amphetamine addiction	KEGG_01.03.2017	660.0E-6	4.2E-3	2.4E+0	8.82	6.00
GO:0014069	postsynaptic density	GO_CellularComponent-GOA_23.02.2017_10h01	630.0E-6	4.2E-3	2.4E+0	5.07	11.00
GO:0004152	AMPK signaling pathway	KEGG_01.03.2017	620.0E-6	4.4E-3	2.4E+0	6.61	8.00
GO:0004152	AMPK signaling pathway	KEGG_01.03.2017	620.0E-6	4.4E-3	2.4E+0	6.61	8.00
GO:0004973	Carbohydrate digestion and absorption	KEGG_01.03.2017	590.0E-6	4.6E-3	2.3E+0	11.36	5.00
GO:0004720	Long-term potentiation	KEGG_01.03.2017	610.0E-6	4.6E-3	2.3E+0	8.96	6.00
GO:0004720	Long-term potentiation	KEGG_01.03.2017	610.0E-6	4.6E-3	2.3E+0	8.96	6.00
GO:0004973	Carbohydrate digestion and absorption	KEGG_01.03.2017	590.0E-6	4.6E-3	2.3E+0	11.36	5.00
GO:0000118	histone deacetylase complex	GO_CellularComponent-GOA_23.02.2017_10h01	1.0E-3	6.3E-3	2.2E+0	8.11	6.00
GO:0000118	histone deacetylase complex	GO_CellularComponent-GOA_23.02.2017_10h01	1.0E-3	6.3E-3	2.2E+0	8.11	6.00
GO:0004931	Insulin resistance	KEGG_01.03.2017	1.4E-3	7.0E-3	2.2E+0	6.54	7.00
GO:0045211	postsynaptic membrane	GO_CellularComponent-GOA_23.02.2017_10h01	1.3E-3	7.0E-3	2.2E+0	4.60	11.00
GO:0005231	excitatory extracellular ligand-gated ion channel activity	GO_MolecularFunction-GOA_23.02.2017_10h01	1.2E-3	7.0E-3	2.2E+0	9.62	5.00
GO:0004931	Insulin resistance	KEGG_01.03.2017	1.4E-3	7.0E-3	2.2E+0	6.54	7.00
GO:0001101	response to acid chemical	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.2E-3	7.1E-3	2.1E+0	4.01	14.00
GO:0043197	dendritic spine	GO_CellularComponent-GOA_23.02.2017_10h01	1.3E-3	7.1E-3	2.1E+0	5.88	8.00
GO:0004725	Cholinergic synapse	KEGG_01.03.2017	1.8E-3	8.2E-3	2.1E+0	6.25	7.00
GO:0004725	Cholinergic synapse	KEGG_01.03.2017	1.8E-3	8.2E-3	2.1E+0	6.25	7.00
GO:0004925	Aldosterone synthesis and secretion	KEGG_01.03.2017	1.7E-3	8.3E-3	2.1E+0	7.32	6.00
GO:0060078	regulation of postsynaptic membrane potential	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.8E-3	8.4E-3	2.1E+0	5.59	8.00
GO:0000422	mitophagy	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.1E-3	8.7E-3	2.1E+0	7.06	6.00
GO:0004724	Glutamatergic synapse	KEGG_01.03.2017	2.0E-3	8.8E-3	2.1E+0	6.14	7.00
GO:0004724	Glutamatergic synapse	KEGG_01.03.2017	2.0E-3	8.8E-3	2.1E+0	6.14	7.00
GO:0051393	alpha-actinin binding	GO_MolecularFunction-GOA_23.02.2017_10h01	2.3E-3	9.1E-3	2.0E+0	11.11	4.00
GO:0051393	alpha-actinin binding	GO_MolecularFunction-GOA_23.02.2017_10h01	2.3E-3	9.1E-3	2.0E+0	11.11	4.00
GO:0030224	monocyte differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.8E-3	10.0E-3	2.0E+0	10.53	4.00
GO:0000257	Post-translational modification: synthesis of GPI-anchored proteins	REACTOME_Pathways_01.03.2017	2.8E-3	10.0E-3	2.0E+0	6.67	6.00
GO:0016575	histone deacetylation	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.7E-3	10.0E-3	2.0E+0	6.74	6.00
GO:0004070	Phosphatidylinositol signaling system	KEGG_01.03.2017	4.1E-3	12.0E-3	1.9E+0	6.19	6.00
GO:0000706	SeMet incorporation into proteins	REACTOME_Pathways_01.03.2017	4.0E-3	12.0E-3	1.9E+0	9.52	4.00
GO:0000889	Cytosolic tRNA aminoacylation	REACTOME_Pathways_01.03.2017	4.0E-3	12.0E-3	1.9E+0	9.52	4.00
GO:0000890	tRNA Aminoacylation	REACTOME_Pathways_01.03.2017	4.0E-3	12.0E-3	1.9E+0	9.52	4.00
GO:0071869	response to catecholamine	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.7E-3	12.0E-3	1.9E+0	9.76	4.00
GO:0035258	steroid hormone receptor binding	GO_MolecularFunction-GOA_23.02.2017_10h01	3.5E-3	12.0E-3	1.9E+0	6.38	6.00
GO:0071869	response to catecholamine	GO_BiologicalProcess-GOA_23.02.2017_10h01	3.7E-3	12.0E-3	1.9E+0	9.76	4.00
GO:0004070	Phosphatidylinositol signaling system	KEGG_01.03.2017	4.1E-3	12.0E-3	1.9E+0	6.19	6.00
GO:0008081	phosphoric diester hydrolase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	4.3E-3	13.0E-3	1.9E+0	6.12	6.00

GO:0004024	cAMP signaling pathway	KEGG_01.03.2017	4.0E-3	13.0E-3	1.9E+0	4.55	9.00
GO:0008081	phosphoric diester hydrolase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	4.3E-3	13.0E-3	1.9E+0	6.12	6.00
GO:0004713	Circadian entrainment	KEGG_01.03.2017	3.9E-3	13.0E-3	1.9E+0	6.25	6.00
GO:0004024	cAMP signaling pathway	KEGG_01.03.2017	4.0E-3	13.0E-3	1.9E+0	4.55	9.00
GO:0004713	Circadian entrainment	KEGG_01.03.2017	3.9E-3	13.0E-3	1.9E+0	6.25	6.00
GO:0044450	microtubule organizing center part	GO_CellularComponent-GOA_23.02.2017_10h01	5.3E-3	15.0E-3	1.8E+0	4.71	8.00
GO:0016581	NuRD complex	GO_CellularComponent-GOA_23.02.2017_10h01	5.2E-3	15.0E-3	1.8E+0	13.04	3.00
GO:0016581	NuRD complex	GO_CellularComponent-GOA_23.02.2017_10h01	5.2E-3	15.0E-3	1.8E+0	13.04	3.00
GO:0031103	axon regeneration	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.6E-3	16.0E-3	1.8E+0	8.70	4.00
GO:0043523	regulation of neuron apoptotic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	5.8E-3	16.0E-3	1.8E+0	4.29	9.00
GO:0030552	cAMP binding	GO_MolecularFunction-GOA_23.02.2017_10h01	5.9E-3	16.0E-3	1.8E+0	12.50	3.00
GO:0030552	cAMP binding	GO_MolecularFunction-GOA_23.02.2017_10h01	5.9E-3	16.0E-3	1.8E+0	12.50	3.00
GO:0004812	aminoacyl-tRNA ligase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	7.6E-3	19.0E-3	1.7E+0	8.00	4.00
GO:0004114	3',5'-cyclic-nucleotide phosphodiesterase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	8.3E-3	19.0E-3	1.7E+0	11.11	3.00
GO:0004261	Adrenergic signaling in cardiomyocytes	KEGG_01.03.2017	7.5E-3	19.0E-3	1.7E+0	4.86	7.00
GO:0004114	3',5'-cyclic-nucleotide phosphodiesterase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	8.3E-3	19.0E-3	1.7E+0	11.11	3.00
GO:0004261	Adrenergic signaling in cardiomyocytes	KEGG_01.03.2017	7.5E-3	19.0E-3	1.7E+0	4.86	7.00
GO:0021549	cerebellum development	GO_BiologicalProcess-GOA_23.02.2017_10h01	7.9E-3	20.0E-3	1.7E+0	5.41	6.00
GO:0010508	positive regulation of autophagy	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.2E-3	20.0E-3	1.7E+0	5.36	6.00
GO:0015276	ligand-gated ion channel activity	GO_MolecularFunction-GOA_23.02.2017_10h01	8.0E-3	20.0E-3	1.7E+0	4.79	7.00
GO:0051306	mitotic sister chromatid separation	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.3E-3	21.0E-3	1.7E+0	7.55	4.00
GO:0004435	phosphatidylinositol phospholipase C activity	GO_MolecularFunction-GOA_23.02.2017_10h01	9.2E-3	21.0E-3	1.7E+0	10.71	3.00
GO:1903599	positive regulation of mitophagy	GO_BiologicalProcess-GOA_23.02.2017_10h01	10.0E-3	22.0E-3	1.7E+0	10.34	3.00
GO:0006418	tRNA aminoacylation for protein translation	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.9E-3	22.0E-3	1.7E+0	7.41	4.00
GO:0000289	Activation of AKT2	REACTOME_Pathways_01.03.2017	10.0E-3	23.0E-3	1.6E+0	5.88	5.00
GO:0000847	Integrin alphaIIb beta3 signaling	REACTOME_Pathways_01.03.2017	10.0E-3	23.0E-3	1.6E+0	5.88	5.00
GO:0000929	CTLA4 inhibitory signaling	REACTOME_Pathways_01.03.2017	10.0E-3	23.0E-3	1.6E+0	5.88	5.00
GO:0001088	RSK activation	REACTOME_Pathways_01.03.2017	10.0E-3	23.0E-3	1.6E+0	5.88	5.00
GO:0001177	VEGFR2 mediated vascular permeability	REACTOME_Pathways_01.03.2017	10.0E-3	23.0E-3	1.6E+0	5.88	5.00
GO:0001685	Platelet Aggregation (Plug Formation)	REACTOME_Pathways_01.03.2017	10.0E-3	23.0E-3	1.6E+0	5.88	5.00
GO:0004911	Insulin secretion	KEGG_01.03.2017	10.0E-3	23.0E-3	1.6E+0	5.88	5.00
GO:0004911	Insulin secretion	KEGG_01.03.2017	10.0E-3	23.0E-3	1.6E+0	5.88	5.00
GO:0001491	Regulation of TP53 Activity through Acetylation	REACTOME_Pathways_01.03.2017	11.0E-3	24.0E-3	1.6E+0	10.00	3.00
GO:0001491	Regulation of TP53 Activity through Acetylation	REACTOME_Pathways_01.03.2017	11.0E-3	24.0E-3	1.6E+0	10.00	3.00
GO:0001491	Regulation of TP53 Activity through Acetylation	REACTOME_Pathways_01.03.2017	11.0E-3	24.0E-3	1.6E+0	10.00	3.00
GO:0001491	Regulation of TP53 Activity through Acetylation	REACTOME_Pathways_01.03.2017	11.0E-3	24.0E-3	1.6E+0	10.00	3.00
GO:0006473	protein acetylation	GO_BiologicalProcess-GOA_23.02.2017_10h01	12.0E-3	25.0E-3	1.6E+0	4.08	8.00
GO:0004727	GABAergic synapse	KEGG_01.03.2017	12.0E-3	25.0E-3	1.6E+0	5.68	5.00
GO:0004727	GABAergic synapse	KEGG_01.03.2017	12.0E-3	25.0E-3	1.6E+0	5.68	5.00
GO:0030426	growth cone	GO_CellularComponent-GOA_23.02.2017_10h01	13.0E-3	26.0E-3	1.6E+0	4.32	7.00
GO:0005814	centriole	GO_CellularComponent-GOA_23.02.2017_10h01	13.0E-3	26.0E-3	1.6E+0	4.84	6.00
GO:0016836	hydro-lyase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	12.0E-3	26.0E-3	1.6E+0	6.90	4.00
GO:0004211	Longevity regulating pathway	KEGG_01.03.2017	14.0E-3	26.0E-3	1.6E+0	5.49	5.00
GO:0072523	purine-containing compound catabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	12.0E-3	26.0E-3	1.6E+0	6.90	4.00
GO:1901657	glycosyl compound metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	14.0E-3	27.0E-3	1.6E+0	4.29	7.00
GO:1903578	regulation of ATP metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	15.0E-3	27.0E-3	1.6E+0	6.56	4.00
GO:0001065	Golgi Associated Vesicle Biogenesis	REACTOME_Pathways_01.03.2017	13.0E-3	27.0E-3	1.6E+0	4.80	6.00
GO:0001474	Signaling by BRAF and RAF fusions	REACTOME_Pathways_01.03.2017	13.0E-3	27.0E-3	1.6E+0	4.80	6.00
GO:0001477	Oncogenic MAPK signaling	REACTOME_Pathways_01.03.2017	13.0E-3	27.0E-3	1.6E+0	4.80	6.00
GO:1903578	regulation of ATP metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	15.0E-3	27.0E-3	1.6E+0	6.56	4.00
GO:0004022	cGMP-PKG signaling pathway	KEGG_01.03.2017	14.0E-3	27.0E-3	1.6E+0	4.29	7.00
GO:0001073	Unblocking of NMDA receptor, glutamate binding and activation	REACTOME_Pathways_01.03.2017	14.0E-3	27.0E-3	1.6E+0	9.09	3.00
GO:0001082	CREB phosphorylation through the activation of CaMKII	REACTOME_Pathways_01.03.2017	14.0E-3	27.0E-3	1.6E+0	9.09	3.00
GO:0001733	SALM protein interactions at the synapses	REACTOME_Pathways_01.03.2017	14.0E-3	27.0E-3	1.6E+0	9.09	3.00
GO:0004022	cGMP-PKG signaling pathway	KEGG_01.03.2017	14.0E-3	27.0E-3	1.6E+0	4.29	7.00
GO:0001073	Unblocking of NMDA receptor, glutamate binding and activation	REACTOME_Pathways_01.03.2017	14.0E-3	27.0E-3	1.6E+0	9.09	3.00
GO:0001082	CREB phosphorylation through the activation of CaMKII	REACTOME_Pathways_01.03.2017	14.0E-3	27.0E-3	1.6E+0	9.09	3.00
GO:0001733	SALM protein interactions at the synapses	REACTOME_Pathways_01.03.2017	14.0E-3	27.0E-3	1.6E+0	9.09	3.00
GO:0009066	aspartate family amino acid metabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	15.0E-3	28.0E-3	1.6E+0	6.45	4.00
GO:0004914	Progesterone-mediated oocyte maturation	KEGG_01.03.2017	17.0E-3	30.0E-3	1.5E+0	5.21	5.00
GO:0030902	hindbrain development	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	30.0E-3	1.5E+0	4.14	7.00
GO:2001252	positive regulation of chromosome organization	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-3	30.0E-3	1.5E+0	4.09	7.00
GO:1901983	regulation of protein acetylation	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	30.0E-3	1.5E+0	6.25	4.00
GO:0005143	African trypanosomiasis	KEGG_01.03.2017	17.0E-3	30.0E-3	1.5E+0	8.57	3.00
GO:1901983	regulation of protein acetylation	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	30.0E-3	1.5E+0	6.25	4.00
GO:0005143	African trypanosomiasis	KEGG_01.03.2017	17.0E-3	30.0E-3	1.5E+0	8.57	3.00
GO:1901983	regulation of protein acetylation	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	30.0E-3	1.5E+0	6.25	4.00
GO:1903169	regulation of calcium ion transmembrane transport	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	30.0E-3	1.5E+0	4.55	6.00
GO:0005143	African trypanosomiasis	KEGG_01.03.2017	17.0E-3	30.0E-3	1.5E+0	8.57	3.00
GO:0051427	hormone receptor binding	GO_MolecularFunction-GOA_23.02.2017_10h01	19.0E-3	31.0E-3	1.5E+0	4.05	7.00
GO:0000970	Aminoacyl-tRNA biosynthesis	KEGG_01.03.2017	19.0E-3	32.0E-3	1.5E+0	6.06	4.00
GO:0000230	Purine metabolism	KEGG_01.03.2017	20.0E-3	32.0E-3	1.5E+0	4.00	7.00
GO:0001425	DNA Damage Recognition in GG-NER	REACTOME_Pathways_01.03.2017	21.0E-3	32.0E-3	1.5E+0	7.89	3.00
GO:0006306	DNA methylation	GO_BiologicalProcess-GOA_23.02.2017_10h01	19.0E-3	32.0E-3	1.5E+0	6.06	4.00
GO:0000230	Purine metabolism	KEGG_01.03.2017	20.0E-3	32.0E-3	1.5E+0	4.00	7.00
GO:0030594	neurotransmitter receptor activity	GO_MolecularFunction-GOA_23.02.2017_10h01	20.0E-3	32.0E-3	1.5E+0	5.00	5.00
GO:0004723	Retrograde endocannabinoid signaling	KEGG_01.03.2017	21.0E-3	33.0E-3	1.5E+0	4.95	5.00
GO:0086003	cardiac muscle cell contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-3	33.0E-3	1.5E+0	5.88	4.00
GO:0004723	Retrograde endocannabinoid signaling	KEGG_01.03.2017	21.0E-3	33.0E-3	1.5E+0	4.95	5.00
GO:0086003	cardiac muscle cell contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	21.0E-3	33.0E-3	1.5E+0	5.88	4.00
GO:0000140	RORA activates gene expression	REACTOME_Pathways_01.03.2017	22.0E-3	34.0E-3	1.5E+0	5.80	4.00
GO:0000141	BMAL1:CLOCK,NPAS2 activates circadian gene expression	REACTOME_Pathways_01.03.2017	22.0E-3	34.0E-3	1.5E+0	5.80	4.00
GO:0000979	Circadian Clock	REACTOME_Pathways_01.03.2017	22.0E-3	34.0E-3	1.5E+0	5.80	4.00
GO:0006304	DNA modification	GO_BiologicalProcess-GOA_23.02.2017_10h01	22.0E-3	34.0E-3	1.5E+0	4.85	5.00
GO:0043044	ATP-dependent chromatin remodeling	GO_BiologicalProcess-GOA_23.02.2017_10h01	22.0E-3	34.0E-3	1.5E+0	5.80	4.00
GO:0000260	Glycine, serine and threonine metabolism	KEGG_01.03.2017	24.0E-3	35.0E-3	1.5E+0	7.50	3.00
GO:1901658	glycosyl compound catabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	24.0E-3	35.0E-3	1.5E+0	7.50	3.00
GO:0003300	cardiac muscle hypertrophy	GO_BiologicalProcess-GOA_23.02.2017_10h01	26.0E-3	36.0E-3	1.4E+0	5.56	4.00
GO:0004662	B cell receptor signaling pathway	KEGG_01.03.2017	24.0E-3	36.0E-3	1.4E+0	5.63	4.00
GO:0009166	nucleotide catabolic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	26.0E-3	36.0E-3	1.4E+0	5.56	4.00
GO:0001464	Interactions of neurexins and neuroligins at synapses	REACTOME_Pathways_01.03.2017	24.0E-3	36.0E-3	1.4E+0	5.63	4.00
GO:0001465	Protein-protein interactions at synapses	REACTOME_Pathways_01.03.2017	24.0E-3	36.0E-3	1.4E+0	5.63	4.00
GO:0004662	B cell receptor signaling pathway	KEGG_01.03.2017	24.0E-3	36.0E-3	1.4E+0	5.63	4.00
GO:0005161	Hepatitis B	KEGG_01.03.2017	25.0E-3	37.0E-3	1.4E+0	4.17	6.00
GO:0001072	Post NMDA receptor activation events	REACTOME_Pathways_01.03.2017	25.0E-3	37.0E-3	1.4E+0	7.32	3.00
GO:0001083	CREB phosphorylation through the activation of Ras	REACTOME_Pathways_01.03.2017	25.0E-3	37.0E-3	1.4E+0	7.32	3.00

GO:0001085	Activation of NMDA receptor upon glutamate binding and postsynaptic events	REACTOME_Pathways_01.03.2017	25.0E-3	37.0E-3	1.4E+0	7.32	3.00
GO:0001086	Ras activation uopn Ca2+ infux through NMDA receptor	REACTOME_Pathways_01.03.2017	25.0E-3	37.0E-3	1.4E+0	7.32	3.00
GO:0005161	Hepatitis B	KEGG_01.03.2017	25.0E-3	37.0E-3	1.4E+0	4.17	6.00
GO:0001072	Post NMDA receptor activation events	REACTOME_Pathways_01.03.2017	25.0E-3	37.0E-3	1.4E+0	7.32	3.00
GO:0001083	CREB phosphorylation through the activation of Ras	REACTOME_Pathways_01.03.2017	25.0E-3	37.0E-3	1.4E+0	7.32	3.00
GO:0001085	Activation of NMDA receptor upon glutamate binding and postsynaptic events	REACTOME_Pathways_01.03.2017	25.0E-3	37.0E-3	1.4E+0	7.32	3.00
GO:0001086	Ras activation uopn Ca2+ infux through NMDA receptor	REACTOME_Pathways_01.03.2017	25.0E-3	37.0E-3	1.4E+0	7.32	3.00
GO:0004407	histone deacetylase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	27.0E-3	38.0E-3	1.4E+0	7.14	3.00
GO:0043524	negative regulation of neuron apoptotic process	GO_BiologicalProcess-GOA_23.02.2017_10h01	28.0E-3	39.0E-3	1.4E+0	4.05	6.00
GO:0004918	Thyroid hormone synthesis	KEGG_01.03.2017	28.0E-3	39.0E-3	1.4E+0	5.41	4.00
GO:0004918	Thyroid hormone synthesis	KEGG_01.03.2017	28.0E-3	39.0E-3	1.4E+0	5.41	4.00
GO:0006497	protein lipidation	GO_BiologicalProcess-GOA_23.02.2017_10h01	30.0E-3	41.0E-3	1.4E+0	4.00	6.00
GO:0055117	regulation of cardiac muscle contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	33.0E-3	45.0E-3	1.3E+0	5.13	4.00
GO:0055117	regulation of cardiac muscle contraction	GO_BiologicalProcess-GOA_23.02.2017_10h01	33.0E-3	45.0E-3	1.3E+0	5.13	4.00
GO:0032204	regulation of telomere maintenance	GO_BiologicalProcess-GOA_23.02.2017_10h01	35.0E-3	46.0E-3	1.3E+0	5.06	4.00
GO:0004919	Thyroid hormone signaling pathway	KEGG_01.03.2017	35.0E-3	46.0E-3	1.3E+0	4.31	5.00
GO:0004919	Thyroid hormone signaling pathway	KEGG_01.03.2017	35.0E-3	46.0E-3	1.3E+0	4.31	5.00
GO:0004919	Thyroid hormone signaling pathway	KEGG_01.03.2017	35.0E-3	46.0E-3	1.3E+0	4.31	5.00
GO:0000197	Inositol phosphate metabolism	REACTOME_Pathways_01.03.2017	36.0E-3	47.0E-3	1.3E+0	6.38	3.00
GO:0000453	Synthesis of IP2, IP, and Ins in the cytosol	REACTOME_Pathways_01.03.2017	36.0E-3	47.0E-3	1.3E+0	6.38	3.00
GO:0004071	Sphingolipid signaling pathway	KEGG_01.03.2017	37.0E-3	47.0E-3	1.3E+0	4.24	5.00
GO:1901019	regulation of calcium ion transmembrane transporter activity	GO_BiologicalProcess-GOA_23.02.2017_10h01	36.0E-3	47.0E-3	1.3E+0	5.00	4.00
GO:0004071	Sphingolipid signaling pathway	KEGG_01.03.2017	37.0E-3	47.0E-3	1.3E+0	4.24	5.00
GO:0001754	eye photoreceptor cell differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	38.0E-3	48.0E-3	1.3E+0	6.25	3.00
GO:0043204	perikaryon	GO_CellularComponent-GOA_23.02.2017_10h01	38.0E-3	48.0E-3	1.3E+0	4.20	5.00
GLIA (PGC2 + CLOZUK)							
GOID	GOTerm	Ontology Source	Term PValue	Term PValue Corrected with Benjamini-Hochberg	Neg log (FDR)	% Associated Genes	Nr. Genes
GO:0052697	xenobiotic glucuronidation	GO_BiologicalProcess-GOA_23.02.2017_10h01	48.0E-6	1.4E-3	2.9E+0	33.33	3.00
GO:0008081	phosphoric diester hydrolase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	180.0E-6	1.8E-3	2.7E+0	6.12	6.00
GO:0043197	dendritic spine	GO_CellularComponent-GOA_23.02.2017_10h01	160.0E-6	2.4E-3	2.6E+0	5.15	7.00
GO:0002285	lymphocyte activation involved in immune response	GO_BiologicalProcess-GOA_23.02.2017_10h01	400.0E-6	3.0E-3	2.5E+0	4.43	7.00
GO:0000983	Drug metabolism	KEGG_01.03.2017	610.0E-6	3.7E-3	2.4E+0	8.70	4.00
GO:0004114	3',5'-cyclic-nucleotide phosphodiesterase activity	GO_MolecularFunction-GOA_23.02.2017_10h01	1.5E-3	5.6E-3	2.3E+0	11.11	3.00
GO:0004435	phosphatidylinositol phospholipase C activity	GO_MolecularFunction-GOA_23.02.2017_10h01	1.6E-3	5.6E-3	2.3E+0	10.71	3.00
GO:0000053	Ascorbate and aldarate metabolism	KEGG_01.03.2017	1.5E-3	5.6E-3	2.3E+0	11.11	3.00
GO:0004520	Adherens junction	KEGG_01.03.2017	3.2E-3	5.8E-3	2.2E+0	5.56	4.00
GO:0001464	Interactions of neurexins and neuroligins at synapses	REACTOME_Pathways_01.03.2017	3.1E-3	5.8E-3	2.2E+0	5.63	4.00
GO:0001465	Protein-protein interactions at synapses	REACTOME_Pathways_01.03.2017	3.1E-3	5.8E-3	2.2E+0	5.63	4.00
GO:0000982	Drug metabolism	KEGG_01.03.2017	2.9E-3	5.9E-3	2.2E+0	5.71	4.00
GO:0001073	Unblocking of NMDA receptor, glutamate binding and activation	REACTOME_Pathways_01.03.2017	2.7E-3	6.2E-3	2.2E+0	9.09	3.00
GO:0001082	CREB phosphorylation through the activation of CaMKII	REACTOME_Pathways_01.03.2017	2.7E-3	6.2E-3	2.2E+0	9.09	3.00
GO:0001733	SALM protein interactions at the synapses	REACTOME_Pathways_01.03.2017	2.7E-3	6.2E-3	2.2E+0	9.09	3.00
GO:0002292	T cell differentiation involved in immune response	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.4E-3	6.3E-3	2.2E+0	6.90	4.00
GO:0000040	Pentose and glucuronate interconversions	KEGG_01.03.2017	2.9E-3	6.3E-3	2.2E+0	8.82	3.00
GO:0004720	Long-term potentiation	KEGG_01.03.2017	2.5E-3	6.3E-3	2.2E+0	5.97	4.00
GO:0032206	positive regulation of telomere maintenance	GO_BiologicalProcess-GOA_23.02.2017_10h01	1.3E-3	6.5E-3	2.2E+0	7.14	4.00
GO:0001782	B cell homeostasis	GO_BiologicalProcess-GOA_23.02.2017_10h01	2.4E-3	6.7E-3	2.2E+0	9.38	3.00
GO:0000830	Retinol metabolism	KEGG_01.03.2017	2.2E-3	6.8E-3	2.2E+0	6.15	4.00
GO:0000521	Neurophilin interactions with VEGF and VEGFR	REACTOME_Pathways_01.03.2017	5.0E-3	8.4E-3	2.1E+0	7.32	3.00
GO:0000973	Sema3A PAK dependent Axon repulsion	REACTOME_Pathways_01.03.2017	5.0E-3	8.4E-3	2.1E+0	7.32	3.00
GO:0000974	SEMA3A-Plexin repulsion signaling by inhibiting Integrin adhesion	REACTOME_Pathways_01.03.2017	5.0E-3	8.4E-3	2.1E+0	7.32	3.00
GO:0001093	Signal transduction by L1	REACTOME_Pathways_01.03.2017	5.0E-3	8.4E-3	2.1E+0	7.32	3.00
GO:0000860	Porphyrin and chlorophyll metabolism	KEGG_01.03.2017	5.4E-3	8.5E-3	2.1E+0	7.14	3.00
GO:0000493	Pre-NOTCH Processing in Golgi	REACTOME_Pathways_01.03.2017	6.5E-3	9.8E-3	2.0E+0	6.67	3.00
GO:0000494	Pre-NOTCH Expression and Processing	REACTOME_Pathways_01.03.2017	6.5E-3	9.8E-3	2.0E+0	6.67	3.00
GO:0030071	regulation of mitotic metaphase/anaphase transition	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.3E-3	11.0E-3	2.0E+0	6.12	3.00
GO:0042093	T-helper cell differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	8.3E-3	11.0E-3	2.0E+0	6.12	3.00
GO:0001776	leukocyte homeostasis	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.0E-3	12.0E-3	1.9E+0	4.17	4.00
GO:0051310	metaphase plate congression	GO_BiologicalProcess-GOA_23.02.2017_10h01	9.7E-3	12.0E-3	1.9E+0	5.77	3.00
GO:0000556	Nuclear Events (kinase and transcription factor activation)	REACTOME_Pathways_01.03.2017	9.2E-3	12.0E-3	1.9E+0	5.88	3.00
GO:0000558	ERK/MAPK targets	REACTOME_Pathways_01.03.2017	9.2E-3	12.0E-3	1.9E+0	5.88	3.00
GO:0000564	CREB phosphorylation	REACTOME_Pathways_01.03.2017	9.2E-3	12.0E-3	1.9E+0	5.88	3.00
GO:0001116	MAPK targets/ Nuclear events mediated by MAP kinases	REACTOME_Pathways_01.03.2017	9.2E-3	12.0E-3	1.9E+0	5.88	3.00
GO:0001291	CD209 (DC-SIGN) signaling	REACTOME_Pathways_01.03.2017	9.2E-3	12.0E-3	1.9E+0	5.88	3.00
GO:0000480	Glutathione metabolism	KEGG_01.03.2017	10.0E-3	13.0E-3	1.9E+0	5.56	3.00
GO:0000140	Steroid hormone biosynthesis	KEGG_01.03.2017	13.0E-3	15.0E-3	1.8E+0	5.17	3.00
GO:0021879	forebrain neuron differentiation	GO_BiologicalProcess-GOA_23.02.2017_10h01	17.0E-3	19.0E-3	1.7E+0	4.62	3.00
GO:0002312	B cell activation involved in immune response	GO_BiologicalProcess-GOA_23.02.2017_10h01	18.0E-3	19.0E-3	1.7E+0	4.55	3.00
GO:0005211	Renal cell carcinoma	KEGG_01.03.2017	17.0E-3	19.0E-3	1.7E+0	4.62	3.00
GO:0004662	B cell receptor signaling pathway	KEGG_01.03.2017	22.0E-3	23.0E-3	1.6E+0	4.23	3.00
GO:0000980	Metabolism of xenobiotics by cytochrome P450	KEGG_01.03.2017	25.0E-3	25.0E-3	1.6E+0	4.05	3.00

Table 3.16 | GO for genes in SZ risk-associated chromosomal contacts. GO enrichment of genes in Table 3.15 See description of Table 3.2-3.6 for legend; Benjamini Hochberg was used as correction for multiple comparisons.

	NPC (PGC+CLOZUK)		NEURON (PGC+CLOZUK)		GLIA (PGC+CLOZUK)		NPC (PGC)		NEURON (PGC)		GLIA (PGC)		NPC-String (PGC)		NEURON-String (PGC)		GLIA-String (PGC)
1	EIF2B1	1	PCDHA1	1	ANKRD44	1	PHF21A	1	EPC2	1	SNAP91	1	ITIH4	1	BAHD1	1	ADSS
2	CUL3	2	FSIP2	2	SNAP91	2	ZBTB18	2	DIS3L2	2	ANKRD44	2	SLC2A5	2	CUL3	2	MOB4
3	MARS2	3	PCDHA2	3	BRINP3	3	FANCL	3	SFMBT1	3	GRAMD1B	3	ITIH3	3	BAZ2A	3	CUL3
4	DNAJC19	4	CSMD1	4	ZNF804A	4	MAD1L1	4	NLGN4X	4	ZNF804A	4	SEC16B	4	TAF5	4	NFATC3
5	ORC5	5	PCDHA11	5	RABGAP1L	5	KDM4A	5	KMT2E	5	SLC4A10	5	CACNA1D	5	SF3B1	5	SF3B1
6	ZEB2	6	CHRNA3	6	LINC01122	6	KDM3B	6	SEMA6C	6	LINC01122	6	HECW2	6	ACTR5	6	ADSL
7	PHF21A	7	ATPAF2	7	SLC4A10	7	SOX2-OT	7	NT5M	7	TCF4	7	PPP2R3A	7	POLDIP3	7	EP300
8	ZBTB18	8	PCDHGA10	8	MIR137HG	8	SAP30L-AS1	8	PLCL2	8	ANKHD1-EIF4EBP3	8	SH3RF1	8	ANP32E	8	RNF220
9	ITGA9	9	IL17RB	9	GRAMD1B	9	ARNT	9	PLPPR5	9	C1orf100	9	TCTN2	9	CDC20	9	ANKHD1
10	ALMS1	10	PSMG3-AS1	10	OPCML	10	FXR1	10	PBRM1	10	CSMD1	10	CLCN3	10	CDC25C	10	UTS2
11	TET3	11	C3orf49	11	CSMD1	11	CD46	11	RAI1	11	PLD5	11	CLU	11	DNAJC19	11	HECW2
12	CCDC150	12	PPP2R3A	12	HECW2	12	EPB41	12	RNF220	12	HECW2	12	CACNA1C	12	THOC7	12	MEF2C
13	MPHOSPH9	13	BRINP3	13	PLD5	13	RAD54L2	13	PLEKHO1	13	UTS2	13	PDE4D	13	SUFU	13	TCF4
14	KNTC1	14	FBXO41	14	DOCK10	14	RPS6KA3	14	RERE	14	PTCHD1-AS	14	GPM6A	14	MAD1L1	14	GRIN2A
15	POC1A	15	LINC01122	15	CHRM3	15	RSRC2	15	INPP5D	15	CCDC39	15	DNAJC18	15	RNF220	15	PLXNA4
16	ANP32E	16	GRM3	16	GPM6A	16	GTF3C3	16	GIGYF2	16	ZNF165	16	NT5C2	16	PHF21A	16	MAPK3
17	CDC25C	17	PCDHGA1	17	TRANK1	17	SBNO1	17	SOX2	17	PITPNM2	17	GRIN2A	17	AKT3	17	GRIA1
18	SGO2	18	PCDHA6	18	GALNT15	18	TAF5	18	FAM53C	18	RILPL1	18	OXNAD1	18	SEMA6C	18	JUND
19	CENPL	19	PCDHA8	19	PDE4B	19	PPP1CC	19	PTPRF	19	COQ10B	19	HSPA9	19	NT5M	19	SH3RF1
20	PPIH	20	PCDHA7	20	PTCHD1-AS	20	SF3B1	20	FPGT-TNNI3K	20	JUND	20	CYP26B1	20	STAG1	20	AOX1
21	BRD8	21	PCDHA9	21	UTS2	21	DESI2	21	AKT3	21	ARTN	21	UGT1A7	21	CEP162	21	GPER1
22	SUGP1	22	PCDHAC1	22	C1orf100	22	NFATC3	22	RRP7BP	22	SH3RF1	22	UGT1A10	22	H6PD	22	NMUR2
23	CD46	23	PLD5	23	EMX1	23	PCGF6	23	SERHL2	23	NEK1	23	UGT1A9	23	PSKH1	23	DPP4
24	EPB41	24	KCNH7	24	ALMS1P1	24	NOSIP	24	SAP30L-AS1	24	GRIN2A	24	UGT1A8	24	CTNND1	24	CNKSR2
25	RSRC2	25	SYNPR	25	NAT8	25	PPIH	25	SATB2	25	PLXNA4	25	UGT1A6	25	SREBF1	25	UGT1A9
26	GTF3C3	26	CA8	26	HYDIN	26	CETN3	26	ZBTB18	26	MEF2C	26	SDCCAG8	26	ENO1	26	UGT1A5
27	SBNO1	27	FTCDNL1	27	SLC9C2	27	SMS	27	PHF21A	27	TRANK1	27	PPARGC1A	27	GATAD2A	27	UGT1A8
28	PPP1CC	28	NTM	28	NEK1	28	KNTC1	28	CBR4	28	DOCK10	28	CA6	28	HARS	28	UGT1A4
29	SF3B1	29	OPCML	29	GRIN2A	29	ANP32E	29	COX20	29	GPM6A	29	AOX1	29	ZMAT2	29	UGT1A6
30	TAF5	30	ARHGAP40	30	PLXNA4	30	CDC25C	30	ZSCAN23	30	CNKSR2	30	CLIP1	30	AMBRA1	30	UGT1A10
31	ZSWIM6	31	MIR137HG	31	MEF2C	31	SGO2	31	SMG1P2	31	DPP4	31	ASB5	31	PLCB2	31	UGT1A7
32	RAD54L2	32	EFHD1	32	SDCCAG8	32	CCDC150	32	PLCB2	32	PLCL2	32	ST3GAL3	32	MED8		
33	RPS6KA3	33	SERPINE2	33	ARTN	33	MPHOSPH9	33	SLC35G2	33	ABCB9	33	CREB3L2	33	ADSS		
34	RANGAP1	34	PCDHGA2	34	SH3RF1	34	BRD8	34	ARL6IP4	34	FAM57B	34	PCCB	34	MARS		
35	SUGP2	35	PCDHGA8	35	PITPNM2	35	SUGP1	35	PLXNA2	35	PLCL1	35	GRIA1	35	PSMG3		
36	ARNT	36	PCDHGA9	36	STAB2	36	SNTB2	36	TOX	36	FAT2	36	HDAC3	36	SH3GL3		
37	FXR1	37	NDUFA4L2	37	KCNJ13	37	ZSWIM6	37	RASAL2	37	PPP1R16B	37	PPIH	37	HECW2		
38	MDK	38	STAC3	38	NDUFA13	38	RANGAP1	38	PCDHGA6	38	TMX2-CTNND1	38	SUGP1	38	CA8		
39	NCK1	39	IL19	39	SLC4A5	39	SUGP2	39	PCDHGB3	39	LSMEM1	39	GATAD2A	39	SDCCAG8		
40	TKT	40	ITIH1	40	YJEFN3	40	EIF2B1	40	PCDHGA4	40	YJEFN3	40	UTP4	40	PPP2R3A		
41	NUBP1	41	BOLL	41	JUND	41	CUL3	41	PCDHGA5	41	NDUFA13	41	PSMD14	41	GRM3		
42	HDAC3	42	EGR4	42	COQ10B	42	MARS2	42	SERPINF1	42	KCNJ13	42	EP300	42	WDR81		
43	PIH1D1	43	SDCCAG8	43	RILPL1	43	MOB4	43	TMEM110	43	MIR137HG	43	MAD1L1	43	NT5C2		
44	PSMD14	44	PCDHA10	44	PGM3	44	AMBRA1	44	DPYD	44	BOLL	44	CUL3	44	GRIN2A		
45	EIF3B	45	PCDHA5	45	LSMEM1	45	ZKSCAN4	45	P2RX3	45	NUGGC	45	DNAJC19	45	ITIH4		
46	GATAD2A	46	MIR339	46	TMX2-CTNND1	46	TYW5	46	PTBP2	46	GBA3	46	ARNT	46	SLC2A5		
47	SNX7	47	SMPX	47	KLHL20	47	GIGYF2	47	RTN4RL2	47	MIR4304	47	SNTB2	47	ITIH3		
48	CETN3	48	CYP26B1	48	ERMAP	48	INPP5D	48	PGAP1	48	GPX5	48	RANGAP1	48	GABBR1		
49	FGFR1	49	ADIG	49	SFXN5	49	RASAL2	49	TYW5	49	UGT1A9	49	EPB41	49	SNAP91		
50	PPP2R2A	50	MIR548AI	50	TBC1D10B	50	VSIG2	50	C2orf69	50	UGT1A5	50	RPS6KA3	50	PFKFB2		
51	DESI2	51	NT5C2	51	USP40	51	ITGA9	51	ZSCAN12	51	UGT1A8	51	TAF5	51	CACNA1C		
52	NFATC3	52	ALMS1P1	52	VRK2	52	C2orf69	52	CD46	52	UGT1A4	52	PPP1CC	52	PDE4D		
53	SMS	53	NAT8	53	EXOC6B	53	PGAP1	53	MSL2	53	OR9Q1	53	SF3B1	53	DGKD		
54	NOSIP	54	PHEX	54	MBTPS2	54	ADSS	54	SBNO1	54	FCAMR	54	BRD8	54	MPL		
55	PCGF6	55	TOM1L2	55	MAPK3	55	ETF1	55	BAZ2A	55	UGT1A7	55	KNTC1	55	PPARGC1A		
56	FAM53C	56	MSRA	56	GRIA1	56	HSPD1	56	ZSWIM6	56	UGT1A6	56	ANP32E	56	HSPA9		
57	DUS2	57	NEK1	57	MAN2A1	57	MDK	57	TAF5	57	UGT1A10	57	SGO2	57	HARS2		
58	RFT1	58	CCDC39	58	IMMP2L	58	NCK1	58	ACTR5	58	OR1S2	58	CDC25C	58	ATG13		
59	PRDX6	59	ANKRD45	59	ST3GAL3	59	TKT	59	RPS6KA3	59	FLJ40288	59	AKT3	59	SLC12A4		
60	ADSS	60	DCDC1	60	AUP1	60	DUS2	60	KMT2E-AS1	60	STK31	60	PHF21A	60	CTNNA1		
61	ETF1	61	PTCHD1-AS	61	GALNT10	61	SNX7	61	GID4	61	SDCCAG8	61	KMT2E	61	PLXNA2		
62	HSPD1	62	FAM124B	62	GPER1	62	ORC5	62	FXR1	62	CNBD1	62	PLCL2	62	PDE6D		
63	FANCL	63	ITIH4	63	MKL1	63	DNAJC19	63	WDFY1	63	TNNI3K	63	NCAN	63	DGKZ		
64	MAD1L1	64	C1orf100	64	SLC6A9	64	HPF1	64	BAHD1	64	MPP6	64	EPHB1	64	GRIA1		
65	KDM4A	65	CR1L	65	TNRC6B	65	ATP6V0A2	65	CUL3	65	DNAH10	65	CHST12	65	CA6		
66	KDM3B	66	CTSS	66	TCF4	66	EP300	66	DESI2	66	C12orf65	66	RPRD2	66	IL19		
67	SOX2-OT	67	DOCK10	67	CCDC39	67	RANBP10	67	POLDIP3	67	C7orf50	67	MAU2	67	MBTPS2		
68	STAG1	68	LRP1	68	ZNF165	68	PSMD14	68	NFATC3	68	CTNND1	68	B4GALT2	68	SEMA3G		
69	HPF1	69	PFKFB2	69	GBA3	69	UTP4	69	SF3B1	69	TTYH3	69	NLGN4X	69	CREB3L1		
70	ASH2L	70	CPEB1	70	NUGGC	70	EIF3B	70	MPHOSPH9	70	PSKH1	70	DPYD	70	DYSF		
71	DARS2	71	JAK1	71	BOLL	71	GATAD2A	71	CCDC150	71	VPS45	71	NCK1	71	SPARC		
72	VPS45	72	AIG1	72	MIR4304	72	NUBP1	72	ANP32E	72	TBC1D5	72	ADSS	72	PSMB10		
73	ABCD3	73	PDE4B	73	NMUR2	73	HDAC3	73	NEMP1	73	PLEKHO1	73	HSPD1	73	ATXN7		
74	TBC1D5	74	PDE4D	74	ANKHD1-EIF4EBP3	74	PIH1D1	74	G3BP1	74	RERE	74	EBNA1BP2				
75	B4GALT2	75	HIVEP2	75	HCN1	75	MAU2	75	CDC25C	75	NLGN4X	75	GTF2H3				

76	MAU2	76	CACNA1C	76	PAPPA2	76	B4GALT2	76	CDC20	76	TAOK2	76	EPC2
77	ADGRG6	77	CACNA1C-AS4	77	OR9Q1	77	TBC1D5	77	PCGF6	77	GALNT10	77	STAG1
78	SOX2	78	CACNA1C-IT3	78	CNBD1	78	VPS45	78	SSRP1	78	MKL1		
79	CMTR2	79	GALNT15	79	STK31	79	AADAT	79	IPO11	79	GPER1		
80	RANBP10	80	PRUNE1	80	FCAMR	80	EPC2	80	SOX2-OT	80	MAPK3		
81	ATP6V0A2	81	ADAMTS9-AS2	81	UGT1A7	81	STAG1	81	MAD1L1	81	GRIA1		
82	EP300	82	EPN2	82	UGT1A6	82	DENR	82	VPS37B	82	MAN2A1		
83	FPGT-TNNI3K	83	GPM6A	83	UGT1A4	83	VPS33A	83	DNAJC19	83	IMMP2L		
84	PTPRF	84	TMEFF2	84	OR1S2	84	DPYD	84	KDM4A	84	ST3GAL3		
85	OXNAD1	85	EMX1	85	UGT1A10	85	TMEM110	85	SUFU	85	UNCX		
86	PODXL	86	PCDHGA12	86	UGT1A8	86	SATB2	86	FANCL	86	TNFRSF9		
87	AADAT	87	MEF2C	87	UGT1A5	87	TMEM110-MUSTN1	87	SFXN2	87	AOX1		
88	EPC2	88	PCDHGA11	88	UGT1A9	88	PLXNA2	88	THOC7	88	NRN1L		
89	SLC39A8	89	DGKD	89	CNTN4	89	PCDHGB1	89	AMBRA1	89	MBTPS2		
90	CHST12	90	SLC2A5	90	SLX1A	90	PCDHGB2	90	HARS	90	VRK2		
91	GFRA3	91	ITIH3	91	MYO1A	91	PCDHGA6	91	EIF4E2	91	USP40		
92	LYPD6	92	PRKCB	92	AOX2P	92	PCDHGB3	92	PSMG3	92	LRRIQ3		
93	NLGN4X	93	CDHR3	93	MYLPF	93	PCDHGA4	93	MARS	93	BOLA2		
94	SATB1	94	LINC00634	94	UNCX	94	PCDHGA5	94	MED8	94	OGFOD2		
95	PLEKHO1	95	PCDHAC2	95	TNFRSF9	95	SLC39A8	95	ADSS	95	ERMAP		
96	RERE	96	BRINP2	96	AOX1	96	CHST12	96	SLC25A33	96	TBC1D10B		
97	RPRD2	97	CACNA1D	97	NRN1L	97	GFRA3	97	DPH3	97	SLC6A9		
98	SAP30L-AS1	98	KLHL29	98	RFTN2	98	LYPD6	98	ZMAT2	98	TNRC6B		
99	DPYD	99	PCDHA4	99	LINC00461	99	NLGN4X	99	PARK7	99	NMUR2		
100	SATB2	100	SH3PXD2A	100	TMEM161B-AS1	100	SATB1	100	TWF2	100	HCN1		
101	M1AP	101	GRIN2A	101	PLEKHO1	101	PLEKHO1	101	C7orf50	101	PAPPA2		
102	SLC35G2	102	MRAP2	102	RERE	102	RERE	102	PDE6D	102	C2orf69		
103	EMB	103	INO80B	103	NLGN4X	103	RPRD2	103	H6PD	103	CUL3		
104	RASAL2	104	SLC2A7	104	TAOK2	104	DPP4	104	PSKH1	104	SBNO1		
105	VSIG2	105	TLR9	105	TNNI3K	105	PLCL2	105	CTNND1	105	WDFY1		
106	TYW5	106	PPARGC1A	106	CNKSR2	106	CEP170	106	SREBF1	106	SMS		
107	C2orf69	107	CCDC62	107	MPP6	107	KMT2E	107	VPS45	107	CD46		
108	PGAP1	108	MPL	108	DNAH10	108	CBR4	108	CEP162	108	ZSCAN16-AS1		
109	GIGYF2	109	ADAMTSL3	109	FAT2	109	ZSCAN23	109	TBC1D5	109	CCDC150		
110	INPP5D	110	PCDHA13	110	ABCB9	110	AKT3	110	ADGRA3	110	FAM83D		
111	AMBRA1	111	HECW2	111	DQX1	111	PAIP2B	111	STAG1	111	MPHOSPH9		
112	ZKSCAN4	112	PEX5L	112	DPP4	112	FPGT-TNNI3K	112	ELOVL1	112	DHX35		
113	LRRIQ3	113	ASTN1	113	PLCL2	113	PTPRF	113	ZDHHC5	113	ZSWIM6		
114	TMEM110-MUSTN1	114	PCDHA12	114	PPP1R16B	114	FAM53C	114	GATAD2A	114	SF3B1		
115	MYO1A	115	ANKRD44	115	CPEB1-AS1	115	SOX2	115	MDK	115	NFATC3		
116	OR2B2	116	NYAP2	116	AP3B2	116	IMMP2L	116	SMS	116	KMT5A		
117	PLXNA2	117	SLC4A10	117	FAM57B	117	PODXL	117	TMEM161A	117	ITGA9		
118	PCDHGA6	118	GPR52	118	PLCL1	118	OXNAD1	118	SNX7	118	RNF220		
119	PCDHGB3	119	SNAP91	119	ADSS	119	SLC25A42	119	DUS2	119	ZNF691		
120	PCDHGA4	120	LYPD6B	120	DPYD	120	AOX2P	120	ENO1	120	ZNF48		
121	PCDHGA5	121	DNAJC6	121	HTRA2	121	UGT1A6	121	TMEFF2	121	FANCL		
122	CBR4	122	SH3GL3	122	MOB4	122	MIR3160-1	122	GPM6A	122	HIRIP3		
123	ZSCAN23	123	R3HDM2	123	IBTK	123	UGT1A8	123	PCDHGA12	123	INO80E		
124	KMT2E	124	IL34	124	PDE6D	124	LRRIQ3	124	GRIN2A	124	ADSL		
125	RC3H1	125	SRPK2	125	SATB2	125	MYO1A	125	SLC2A5	125	EP300		
126	AKT3	126	ERI3	126	TSSK6	126	OR2B2	126	NT5C2	126	KDM4A		
127	PAIP2B	127	NEK4	127	LRRIQ3	127	EBNA1BP2	127	NEK1	127	SOX2-OT		
128	BOLA3	128	PEMT	128	BOLA2	128	DPH3	128	TOM1L2	128	MAD1L1		
129	MOB4	129	SRR	129	OGFOD2	129	IK	129	CACNA1C-AS4	129	RFTN2		
130	DCTN1	130	USP40	130	CTNND1	130	FAM114A2	130	CACNA1C-IT3	130	SLX1A		
131	DENR	131	ACTR8	131	TTYH3	131	GTF2H3	131	PDE4D	131	CNTN4		
132	VPS33A	132	MRPS14	132	C12orf65	132	SLC35G2	132	CACNA1C	132	AKT3		
133	EBNA1BP2	133	FRMD8	133	C7orf50	133	HSPA9	133	ADAMTS9-AS2	133	PTPRF		
134	DPH3	134	GALNT10	134	PSKH1	134	WDR55	134	PRUNE1	134	FPGT-TNNI3K		
135	IK	135	PCNX3	135	TBC1D5	135	C7orf31	135	CDHR3	135	ARL6IP4		
136	RTKN	136	AKAP13	136	VPS45	136	EFHD1	136	PCDHAC2	136	ANKHD1		
137	FAM114A2	137	SMG6	137	ITGA9	137	TMTC1	137	GABBR1	137	SMG1P2		
138	GTF2H3	138	AP1S3	138	RNF220	138	PTCHD1-AS	138	LINC00634	138	SATB2		
139	SFMBT1	139	NGEF	139	ATF7IP2	139	TSPAN9	139	SPATS2L	139	ATF7IP2		
140	ERI3	140	DGKZ	140	GIGYF2	140	PCDHA12	140	WDR81	140	SAP30L-AS1		
141	SLC6A9	141	EXOC6B	141	CCT7	141	PCDHA4	141	SNX8	141	MIR3188		
142	ACTR8	142	DPYD-AS1	142	ADSL	142	CCDC24	142	CTSS	142	PDE6D		
143	EXOC6B	143	PCDHGA3	143	INO80E	143	DAZL	143	DOCK10	143	TOX		
144	C11orf49	144	CREB3L1	144	ALMS1	144	CCDC39	144	LRP1	144	ADSS		
145	FER	145	DYSF	145	ZNF48	145	CFAP57	145	MIR22HG	145	DPYD		
146	SFXN5	146	MBTPS2	146	FANCL	146	NME5	146	PFKFB2	146	MOB4		
147	WBP1L	147	HCN1	147	HIRIP3	147	MIR137HG	147	PHEX	147	GIGYF2		
148	PCDHGB1	148	INPP4B	148	ZNF691	148	PCDHGA1	148	TLCD2	148	AOX2P		
149	PCDHGB2	149	ST3GAL3	149	EP300	149	NYAP2	149	SNORD32B	149	MYLPF		
150	DPYD-AS1	150	DDX60L	150	KDM4A	150	SLC4A10	150	ITIH3	150	MYO1A		
151	PCDHGA2	151	PSMB10	151	MAD1L1	151	ANKRD44	151	WBP1L	151	TSSK6		
152	PCDHGA7	152	RWDD2A	152	SOX2-OT	152	CACNA1D	152	PCDHGA10				
153	AOX2P	153	CHDH	153	SMS	153	LINC01122	153	ATPAF2				
154	ARL14	154	PCDHGB4	154	SF3B1	154	ZNF804A	154	PCDHGB6				
155	UGT1A7	155	PCDHGB7	155	KMT5A	155	PCDHA6	155	RWDD2A				
156	DGKZ	156	STAT6	156	NFATC3	156	PCDHA8	156	CHRNA3				
157	AOX1	157	GRIA1	157	MOB1A	157	PCDHA7	157	CCDC39				
158	ESAM	158	MAN2A1	158	WDFY1	158	PCDHA9	158	ZSCAN31				
159	GUSBP5	159	CA6	159	CCDC150	159	PCDHAC1	159	ITIH4				
160	PCDHGA3	160	C2orf78	160	FAM83D	160	PLD5	160	FAM124B				
161	GALNT10	161	SNORD63	161	DHX35	161	HECW2	161	CR1L				
162	VRK2	162	HSPA9	162	MPHOSPH9	162	PCDHA13	162	C1orf100				

163	CA6	163	ALAS1	163	LOXL3	163	ZBED9	163	PSMG3-AS1										
164	KLHL20	164	HARS2	164	ZSWIM6	164	ZNF391	164	PPP2R3A										
165	HCN1	165	OXNAD1	165	C2orf69	165	PCDHA11	165	C3orf49										
166	INPP4B	166	PGM3	166	CUL3	166	CSMD1	166	PCDHA1										
167	MBTPS2	167	PCDHGC3	167	SBNO1	167	FTCDNL1	167	PCDHA2										
168	PDGFRL	168	PCDHGB6	168	MIR3188	168	NEK1	168	FSIP2										
169	ATP2A2	169	PITPNM2	169	RC3H1	169	NT5C2	169	PCDHA11										
170	USP38	170	ATG13	170	DAAM1	170	GRIN2A	170	SNORD63										
171	AP1S3	171	COQ10B	171	CD46	171	CCDC30	171	CSMD1										
172	SEMA4F	172	SNX8	172	ZSCAN16-AS1	172	IQCE	172	SDCCAG8										
173	NMUR2	173	SPATS2L	173	SAP30L-AS1	173	SPATA24	173	ADIG										
174	MYOT	174	MIR22HG	174	FOXP1	174	PER3	174	PCDHA10										
175	NPY6R	175	TLCD2	175	TOX	175	TFAMP1	175	PCDHA5										
176	GLRA1	176	PALLD	176	ARL6IP4	176	BRINP2	176	CYP26B1										
177	PLCL1	177	PCCB	177	ANKHD1	177	ASTN1	177	MIR339										
178	KCNH7	178	SPARC	178	SMG1P2	178	PEX5L	178	ZFP57										
179	MPP6	179	SLC12A4	179	AKT3	179	SRPK2	179	SMPX										
180	FAT2	180	CTNNA1	180	FPGT-TNNI3K	180	DNAJC18	180	NGEF										
181	STK31	181	RFTN1	181	PTPRF	181	PCDHAC2	181	SERPINE2										
182	C7orf31	182	PCDHGA6			182	BANK1	182	EFHD1										
183	EFHD1	183	PCDHGB3			183	ITIH3	183	PCDHA6										
184	PCDHA6	184	PCDHGA4			184	ITIH4	184	PCDHA8										
185	PCDHA8	185	PCDHGA5			185	ASB5	185	PCDHA7										
186	PCDHA7	186	SERPINF1			186	SEC16B	186	PCDHA9										
187	PCDHA9	187	PCDHGB1			187	TMEFF2	187	HECW2										
188	NTM	188	PCDHGB2			188	FAM13B	188	CA8										
189	OPCML	189	ACTG1P17			189	PHEX	189	FTCDNL1										
190	PCDHAC1	190	LINC01004			190	SLC2A5	190	SYNPR										
191	PLD5	191	TMEM110-MUSTN1			191	GALNT15	191	PLD5										
192	FTCDNL1	192	C1orf116			192	LRRC43	192	PCDHAC1										
193	ZNF391	193	MYO1A			193	CLU	193	ZBED9										
194	PCDHA11	194	OGFOD2			194	REEP2	194	MIR137HG										
195	CSMD1	195	BOLA2			195	CACNA1C	195	ARHGAP40										
196	SERPINC1	196	CTRL			196	PDE4D	196	LINC01122										
197	PCDHA1	197	LRRIQ3			197	DOCK10	197	LYPD6B										
198	PCDHA2	198	MIR33B			198	GPM6A	198	SH3GL3										
199	PARP8	199	LAPTM5			199	PCDHGA11	199	ZNF804A										
200	PPP2R3A	200	ARHGAP15			200	PCDHGA12	200	NYAP2										
201	PCDHA10	201	SLX1A			201	DGKD	201	SLC4A10										
202	PCDHA5	202	LINC00698			202	SCG2	202	SNAP91										
203	C4BPA	203	PCDHGA7			203	TRANK1	203	ANKRD44										
204	SMPX	204	GFRA3			204	COQ10B	204	DGKD										
205	PCDHGA1	205	SEMA3G			205	CYSTM1	205	SRPK2										
206	DCTN1-AS1	206	KMT2E			206	AP1S3	206	MRAP2										
207	CACNA2D3	207	SEMA6C			207	PITPNM2	207	C2orf82										
208	MIR137HG	208	NT5M			208	CLIP1	208	SH3PXD2A										
209	EPHB1	209	PLCL2			209	MFAP3	209	PTCHD1-AS										
210	NCAN	210	PLPPR5			210	SH3RF1	210	PCDHA12										
211	PCDHA3	211	RNF220			211	PCDHGB6	211	PCDHA4										
212	PSD3	212	RAI1			212	TVP23A	212	PCDHA13										
213	PPM1L	213	PLEKHO1			213	CLCN3	213	ADAMTSL3										
214	CNKSR2	214	RERE			214	TCTN2	214	PEX5L										
215	PLPPR5	215	DIS3L2			215	RILPL2	215	ASTN1										
216	CEP170	216	DCTN1			216	SNX8	216	BRINP2										
217	DPP4	217	EPC2			217	RFTN1	217	R3HDM2										
218	PLCL2	218	KCNJ13			218	CREB3L2	218	CCDC62										
219	RFTN2	219	MYO15A			219	PALLD	219	PPARGC1A										
220	CCNH	220	TNNI3K			220	GRIA1	220	MOG										
221	LINC00461	221	DPEP3			221	MAN2A1	221	MPL										
222	RASA1	222	MIR4677			222	PCCB	222	GPX5										
223	TNNI3K	223	CNNM2			223	GBF1	223	SLC2A7										
224	SLC25A42	224	TMX2-CTNND1			224	ST3GAL3	224	TLR9										
225	UGT1A10	225	RFTN2			225	CYP26B1	225	PCDHGA11										
226	UGT1A9	226	TTC14			226	PCDHGA2	226	PCDHGB7										
227	MIR3160-1	227	ABCB9			227	PCDHGA7	227	PCDHGB4										
228	UGT1A8	228	RTN4RL1			228	DGKZ	228	PCDHGA1										
229	PKD2L2	229	CPEB1-AS1			229	MBTPS2	229	GRM3										
230	SEMA3G	230	AP3B2			230	VRK2	230	BOLL										
231	UGT1A6	231	GLRA1			231	DDX60L	231	STAC3										
232	PEX5L	232	FAM57B			232	GSDME	232	NDUFA4L2										
233	BRINP2	233	PLCL1			233	AOX1	233	TMX2-CTNND1										
234	ASTN1	234	PCDHA3			234	ESAM	234	CNNM2										
235	TNR	235	CNKSR2			235	PCDHGB4	235	RFTN2										
236	ALMS1P1	236	PSD3			236	PCDHGB7	236	TTC14										
237	NAT8	237	NLGN4X			237	GALNT10	237	GFRA3										
238	PCDHA4	238	PRKD1			238	CA6	238	VTRNA1-1										
239	DAZL	239	EPHA7			239	HCN1	239	LINC00698										
240	CSGALNACT1	240	SFMBT1			240	FCHSD1	240	SEMA3G										
241	PCDHA12	241	G3BP1			241	RELL2	241	LAPTM5										
242	HECW2	242	CDC20			242	KCNJ13	242	MIR33B										
243	PCDHA13	243	CDC25C			243	PKD2L2	243	SNORD63										
244	BRINP3	244	CCDC150			244	PPARGC1A	244	SLX1A										
245	ZNF804A	245	MPHOSPH9			245	BNIPL	245	LRRIQ3										
246	LINC01122	246	POC1A			246	IQCF1	246	TMEM110-MUSTN1										
247	RABGAP1L	247	SSRP1			247	UGT1A10	247	MYO1A										
248	GPR52	248	CENPL			248	UGT1A9	248	C1orf116										
249	NYAP2	249	ANP32E			249	IL19	249	OGFOD2										
250	SLC4A10	250	NEMP1			250	UGT1A7	250	BOLA2										
251	ANKRD44	251	IPO11			251	PCDHGA8	251	CTRL										
252	CACNA1D	252	PCGF6			252	PCDHGA9	252	LINC01004										
253	KLHL29	253	CCT7			253	DPYD-AS1	253	DPEP3										

254	MEF2C	254	SF3B1			254	PCDHGA3	254	MIR4677										
255	EMX1	255	NFATC3			255	PCDHA10	255	KCNJ13										
256	PCDHGA12	256	DESI2			256	PCDHA5	256	TNNI3K										
257	DOCK10	257	POLDIP3			257	C4BPA	257	PCDHA3										
258	MSRA	258	RPS6KA3			258	SMPX	258	CNKSR2										
259	GPM6A	259	HDAC2			259	BOLL	259	ZKSCAN3										
260	SEC16B	260	ACTR5			260	SLC17A7	260	ABCB9										
261	GRIN2A	261	TAF5			261	GPX5	261	RTN4RL1										
262	TNFSF4	262	BAZ2A			262	GPX6	262	KCNH7										
263	GALNT15	263	MSL2			263	RFTN2	263	GLRA1										
264	LRRC43	264	CD46			264	SDCCAG8	264	FAM57B										
265	CLU	265	KMT2E-AS1			265	TNNI3K	265	PLCL1										
266	REEP2	266	GNL3			266	PCDHA3	266	PCDHGB1										
267	AIG1	267	PPP2R2A			267	EPHB1	267	PCDHGB2										
268	CACNA1C	268	MAP3K11			268	NCAN	268	PCDHGA2										
269	PDE4B	269	SMS			269	CNKSR2	269	PCDHGA7										
270	PDE4D	270	MDK			270	PLPPR5	270	DPYD-AS1										
271	FAM13B	271	BNIP3L			271	NMUR2	271	PCDHGA3										
272	PHEX	272	MOGS			272	MYOT	272	CREB3L1										
273	SLC2A5	273	TMEM161A			273	NPY6R	273	DYSF										
274	TMEFF2	274	FXR1			274	FAT2	274	MBTPS2										
275	SCG2	275	GID4			275	STK31	275	VRK2										
276	TRANK1	276	TYW5			276	GLRA1	276	USP40										
277	DGKD	277	C2orf69			277	PLCL1	277	ERI3										
278	BOLA3-AS1	278	CUL3			278	KCNH7	278	NEK4										
279	SRPK2	279	BAHD1			279	MPP6	279	SRR										
280	INO80B	280	SBNO1			280	ATP2A2	280	PEMT										
281	RILPL2	281	STAG1			281	ERI3	281	AKAP13										
282	SPATA24	282	ADGRA3			282	SLC6A9	282	SMG6										
283	PER3	283	KDM4A			283	PPP2R3A	283	ZNF592										
284	TFAMP1	284	SUFU			284	TCF4	284	DDX60L										
285	CCDC39	285	PBRM1			285	PCDHA1	285	PSMB10										
286	CFAP57	286	ALMS1			286	PCDHA2	286	ST3GAL3										
287	NME5	287	FANCL			287	C11orf49	287	ATXN7										
288	NISCH	288	SOX2-OT			288	FER	288	PGM3										
289	PCDHGB6	289	MAD1L1			289	PCDHGA10	289	HSPA9										
290	GAB1	290	VPS37B			290	WBP1L	290	HARS2										
291	ANKRD45	291	DNAJC19					291	STAT6										
292	TCTN2	292	DARS2					292	GRIA1										
293	NEK1	293	SFXN2					293	MAN2A1										
294	NT5C2	294	THOC7					294	PALLD										
295	DNAJC18	295	PGAP1					295	SPARC										
296	PCDHAC2	296	RTN4RL2					296	PCCB										
297	CCDC24	297	RASAL2					297	SLC12A4										
298	HYDIN	298	DDHD2					298	CTNNA1										
299	IQCE	299	TOX					299	RFTN1										
300	CCDC30	300	PHF21A					300	ATG13										
301	SLC9C2	301	P2RX3					301	COQ10B										
302	MFAP3	302	PTBP2					302	PITPNM2										
303	CLCN3	303	AMBRA1					303	AP1S3										
304	SNX8	304	C7orf50					304	CA6										
305	IL17RB	305	HARS					305	HCN1										
306	TCF4	306	PRADC1					306	PCDHGA9										
307	CHDH	307	EIF4E2					307	PCDHGA8										
308	PCDHGA10	308	KAT5					308	SNORD63										
309	TMTC1	309	PDE6D					309	IL19										
310	PTCHD1-AS	310	PARK7					310	ITIH1										
311	TSPAN9	311	TWF2					311	DGKZ										
312	CLIP1	312	DPH3					312	PCDHGC3										
313	ITIH3	313	SLC25A33					313	GALNT10										
314	ITIH4	314	MARS					314	YY2										
315	TVP23A	315	ADSS																
316	ASB5	316	CACYBP																
317	BANK1	317	PSMG3																
318	FCHSD1	318	SNX7																
319	RELL2	319	DUS2																
320	KCNJ13	320	ENO1																
321	SLC17A7	321	ELOVL1																
322	SLC4A5	322	GATAD2A																
323	ZBTB37	323	GLT8D1																
324	BNIPL	324	CTNND1																
325	IQCF1	325	SREBF1																
326	PPARGC1A	326	CEP162																
327	TNN	327	TBC1D5																
328	IMMP2L	328	H6PD																
329	PCCB	329	PSKH1																
330	GBF1	330	ALKBH5																
331	ST3GAL3	331	MED8																
332	GRIA1	332	DUSP11																
333	MAN2A1	333	SLC35G2																
334	CREB3L2	334	ARL6IP4																
335	PALLD	335	GAS5																
336	PCDHGA8	336	LEPR																
337	PCDHGA9	337	SMG1P2																
338	CYP26B1	338	GIGYF2																
339	BOLL	339	INPP5D																
340	IL19	340	FAM53C																
341	DDX60L	341	SOX2																
342	SDCCAG8	342	FPGT-TNNI3K																
343	TNNC1	343	PTPRF																
344	PCDHGA11	344	AKT3																
345	PCDHGB4	345	RRP7BP																
346	PCDHGB7	346	SERHL2																
347	HSPA9	347	CBR4																
348	ALAS1	348	COX20																
349	WDR55	349	SAP30L-AS1																
350	RFTN1	350	SATB2																

351	SH3RF1	351	FOXP1														
352	PITPNM2	352	PLXNA2														
353	COQ10B	353	DPYD														
354	CYSTM1	354	PLCB2														

Table 3.17 | List of genes shown in RNA correlation heatmaps.
List of genes (by row) in the order they appear in each cell type's full connectome and STRING RNA Pearson correlation heatmaps (See Figure 3.10 and Figure 3.12, 3.17-19).

list name	length	length after deletions	absolute correlation mean	# of samplings/permutations	P-value (random)	P-value (distance constraint)	P-value (full v. string)	P-value (full v. string; string removed from random)
NPC vs COS	308	290	0.2455	1000000, 1000	< 0.000001	< 0.001		
NEU vs COS	358	314	0.2329	1000000, 1000	< 0.000001	< 0.001		
GLIA vs COS	165	151	0.2066	1000000, 1000	0.00001	0.264		
string NPC	77	77	0.2963	1000000	< 0.000001		0.007	< 0.001
string NEU	73	73	0.2877	1000000	< 0.000001		0.008	
string GLIA	31	31	0.2225	10000	0.02		0.595	
string NPC medium	139	138	0.2725747	1000000	< 0.000001		0.012	
string NEU medium	138	138	0.2654931	1000000	< 0.000001		0.003	
string GLIA medium	58	58	0.2183428	10000	0.001		0.434	

Table 3.18 | Summary of results from RNA-seq sampling/permutation anslyses.

Summary of results from RNA-seq sampling/permutation analyses (see Materials and methods, "RNA transcriptomic correlation heatmaps"). List name, list being compared against a background (COS = Childhood Onset Schizophrenia cohort); length, number of genes in the input list; length after deletions, number of genes remaining after filtering out genes with CPM < 1 across 30% of individual RNA-seq experiments; absolute correlation mean, the organization score for each heatmap; # of samplings/permutations, (random analysis, distance-constrained analysis); P-value (random), P-value from randomized sampling analysis without distance constraint; P-value (distance constraint), P-value calculated from distance-constrained randomized sampling (x1,000) analysis (featured in Fig 3.10, D); P-value (full v. string), P-value from the comparison between STRING heatmaps against those from the full connectome background (featured in Fig 3.13, B); P-value (full v. string; string removed from random), P-value from a version of the comparison that removed the subset STRING genes from the full connectome gene list before randomly sampling 1000 times

NPC 1	NPC 2	NEURON 1	NEURON 2	GLIA 1	GLIA 2
EIF2B1	UGT1A6	PCDHA1	SFMBT1	ANKRD44	OGFOD2
CUL3	PEX5L	FSIP2	G3BP1	SNAP91	CTNND1
MARS2	BRINP2	PCDHA2	CDC20	BRINP3	TTYH3
DNAJC19	ASTN1	CSMD1	CDC25C	ZNF804A	C12orf65
ORC5	TNR	PCDHA11	CCDC150	RABGAP1L	C7orf50
ZEB2	ALMS1P1	CHRNA3	MPHOSPH9	LINC01122	PSKH1
PHF21A	NAT8	ATPAF2	POC1A	SLC4A10	TBC1D5
ZBTB18	PCDHA4	PCDHGA10	SSRP1	MIR137HG	VPS45
ITGA9	DAZL	IL17RB	CENPL	GRAMD1B	ITGA9
ALMS1	CSGALNACT1	PSMG3-AS1	ANP32E	OPCML	RNF220
TET3	PCDHA12	C3orf49	NEMP1	CSMD1	ATF7IP2
CCDC150	HECW2	PPP2R3A	IPO11	HECW2	GIGYF2
MPHOSPH9	PCDHA13	BRINP3	PCGF6	PLD5	CCT7
KNTC1	BRINP3	FBXO41	CCT7	DOCK10	ADSL
POC1A	ZNF804A	LINC01122	SF3B1	CHRM3	INO80E
ANP32E	LINC01122	GRM3	NFATC3	GPM6A	ALMS1
CDC25C	RABGAP1L	PCDHGA1	DESI2	TRANK1	ZNF48
SGO2	GPR52	PCDHA6	POLDIP3	GALNT15	FANCL
CENPL	NYAP2	PCDHA8	RPS6KA3	PDE4B	HIRIP3
PPIH	SLC4A10	PCDHA7	HDAC2	PTCHD1-AS	ZNF691
BRD8	ANKRD44	PCDHA9	ACTR5	UTS2	EP300
SUGP1	CACNA1D	PCDHAC1	TAF5	C1orf100	KDM4A
CD46	KLHL29	PLD5	BAZ2A	EMX1	MAD1L1
EPB41	MEF2C	KCNH7	MSL2	ALMS1P1	SOX2-OT
RSRC2	EMX1	SYNPR	CD46	NAT8	SMS
GTF3C3	PCDHGA12	CA8	KMT2E-AS1	HYDIN	SF3B1
SBNO1	DOCK10	FTCDNL1	GNL3	SLC9C2	KMT5A
PPP1CC	MSRA	NTM	PPP2R2A	NEK1	NFATC3
SF3B1	GPM6A	OPCML	MAP3K11	GRIN2A	MOB1A
TAF5	SEC16B	ARHGAP40	SMS	PLXNA4	WDFY1
ZSWIM6	GRIN2A	MIR137HG	MDK		CCDC150
RAD54L2	TNFSF4	EFHD1	BNIP3L		FAM83D
RPS6KA3	GALNT15	SERPINE2	MOGS		DHX35
RANGAP1	LRRC43	PCDHGA2	TMEM161A		MPHOSPH9
SUGP2	CLU	PCDHGA8	FXR1		LOXL3
ARNT	REEP2	PCDHGA9	GID4		ZSWIM6
FXR1	AIG1	NDUFA4L2	TYW5		C2orf69
MDK	CACNA1C	STAC3	C2orf69		CUL3
NCK1	PDE4B	IL19	CUL3		SBNO1
TKT	PDE4D	ITIH1	BAHD1		MIR3188
NUBP1	FAM13B	BOLL	SBNO1		RC3H1
HDAC3	PHEX	EGR4	STAG1		DAAM1
PIH1D1	SLC2A5	SDCCAG8	ADGRA3		CD46
PSMD14	TMEFF2	PCDHA10	KDM4A		ZSCAN16-AS1
EIF3B	SCG2	PCDHA5	SUFU		SAP30L-AS1
GATAD2A	TRANK1	MIR339	PBRM1		FOXP1
SNX7	DGKD	SMPX	ALMS1		TOX
CETN3	BOLA3-AS1	CYP26B1	FANCL		ARL6IP4
FGFR1	SRPK2	ADIG	SOX2-OT		ANKHD1
PPP2R2A	INO80B	MIR548AI	MAD1L1		SMG1P2
DESI2	RILPL2	NT5C2	VPS37B		AKT3
NFATC3	SPATA24	ALMS1P1	DNAJC19		FPGT-TNNI3K
SMS	PER3	NAT8	DARS2		PTPRF
NOSIP	TFAMP1	PHEX	SFXN2		
PCGF6	CCDC39	TOM1L2	THOC7		
FAM53C	CFAP57	MSRA	PGAP1		
DUS2	NME5	NEK1	RTN4RL2		
RFT1	NISCH	CCDC39	RASAL2		
PRDX6	PCDHGB6	ANKRD45	DDHD2		
ADSS	GAB1	DCDC1	TOX		
ETF1	ANKRD45	PTCHD1-AS	PHF21A		
HSPD1	TCTN2	FAM124B	P2RX3		
FANCL	NEK1	ITIH4	PTBP2		
MAD1L1	NT5C2	C1orf100	AMBRA1		
KDM4A	DNAJC18	CR1L	C7orf50		
KDM3B	PCDHAC2	CTSS	HARS		
SOX2-OT	CCDC24	DOCK10	PRADC1		
STAG1	HYDIN	LRP1	EIF4E2		
HPF1	IQCE	PFKFB2	KAT5		
ASH2L	CCDC30	CPEB1	PDE6D		
DARS2	SLC9C2	JAK1	PARK7		
VPS45	MFAP3	AIG1	TWF2		
ABCD3	CLCN3	PDE4B	DPH3		
TBC1D5	SNX8	PDE4D	SLC25A33		
B4GALT2	IL17RB	HIVEP2	MARS		
MAU2	TCF4	CACNA1C	ADSS		
ADGRG6	CHDH	CACNA1C-AS4	CACYBP		
SOX2	PCDHGA10	CACNA1C-IT3	PSMG3		
CMTR2	TMTC1	GALNT15	SNX7		
RANBP10	PTCHD1-AS	PRUNE1	DUS2		
ATP6V0A2	TSPAN9	ADAMTS9-AS2	ENO1		
EP300	CLIP1	EPN2	ELOVL1		
FPGT-TNNI3K	ITIH3	GPM6A	GATAD2A		
PTPRF	ITIH4	TMEFF2	GLT8D1		
OXNAD1	TVP23A	EMX1	CTNND1		
PODXL	ASB5	PCDHGA12	SREBF1		
AADAT	BANK1	MEF2C	CEP162		

EPC2	FCHSD1	PCDHGA11	TBC1D5		
SLC39A8	RELL2	DGKD	H6PD		
CHST12	KCNJ13	SLC2A5	PSKH1		
GFRA3	SLC17A7	ITIH3	ALKBH5		
LYPD6	SLC4A5	PRKCB	MED8		
NLGN4X	ZBTB37	CDHR3	DUSP11		
SATB1	BNIPL	LINC00634	SLC35G2		
PLEKHO1	IQCF1	PCDHAC2	ARL6IP4		
RERE	PPARGC1A	BRINP2	GAS5		
RPRD2	TNN	CACNA1D	LEPR		
	MMP2L	KLHL29	SMG1P2		
	PCCB	PCDHA4	GIGYF2		
	GBF1	SH3PXD2A	INPP5D		
	ST3GAL3	GRIN2A	FAM53C		
	GRIA1	MRAP2	SOX2		
	MAN2A1	INO80B	FPGT-TNNI3K		
	CREB3L2	SLC2A7	PTPRF		
	PALLD	TLR9	AKT3		
	PCDHGA8	PPARGC1A	RRP7BP		
	PCDHGA9	CCDC62	SERHL2		
	CYP26B1	MPL	CBR4		
	BOLL	ADAMTSL3	COX20		
	IL19	PCDHA13	SAP30L-AS1		
	DDX60L	HECW2	SATB2		
	SDCCAG8	PEX5L	FOXP1		
	TNNC1	ASTN1	PLXNA2		
	PCDHGA11	PCDHA12	DPYD		
	PCDHGB4	ANKRD44	PLCB2		
	PCDHGB7	NYAP2			
	HSPA9	SLC4A10			
	ALAS1	GPR52			
	WDR55	SNAP91			
	RFTN1	LYPD6B			
	SH3RF1	DNAJC6			
	PITPNM2	SH3GL3			
	COQ10B	R3HDM2			
	CYSTM1	IL34			
		SRPK2			

Table 3.19 | Neuronal signaling and chromatin regulatory genes clustered in RNA heatmap.

Lis of genes in co-expression clusters from Figure 3.10, D

NPC CLUSTER 1

GOID	GOTerm	Term PValue Corrected with Benjamini-Hochberg	% Associated Genes	Nr. Genes	Total # Genes
R-HSA:2555396	Mitotic Metaphase and Anaphase	6.64E-06	4.43	9.00	203.00
R-HSA:68882	Mitotic Anaphase	9.56E-06	4.46	9.00	202.00
GO:0007062	sister chromatid cohesion	1.20E-05	5.76	8.00	139.00
R-HSA:2500257	Resolution of Sister Chromatid Cohesion	2.20E-05	5.51	7.00	127.00
R-HSA:2467813	Separation of Sister Chromatids	2.62E-05	4.19	8.00	191.00
R-HSA:68877	Mitotic Prometaphase	2.63E-05	4.00	8.00	200.00
R-HSA:141424	Amplification of signal from the kinetochores	2.82E-05	6.25	6.00	96.00
R-HSA:141444	Amplification of signal from unattached kinetochores via a MAD2 inhibitory signal	2.82E-05	6.25	6.00	96.00
GO:0000775	chromosome, centromeric region	2.95E-05	4.02	8.00	199.00
GO:0000922	spindle pole	4.13E-05	4.43	7.00	158.00
R-HSA:69618	Mitotic Spindle Checkpoint	5.47E-05	5.36	6.00	112.00
R-HSA:5663220	RHO GTPases Activate Formins	1.82E-04	4.26	6.00	141.00
GO:0000777	condensed chromosome kinetochore	4.68E-04	4.63	5.00	108.00
GO:0000779	condensed chromosome, centromeric region	7.32E-04	4.13	5.00	121.00
KEGG:04110	Cell cycle	7.61E-04	4.03	5.00	124.00
GO:0000118	histone deacetylase complex	8.51E-04	5.56	4.00	72.00
GO:0000123	histone acetyltransferase complex	1.50E-03	4.71	4.00	85.00
GO:0018023	peptidyl-lysine trimethylation	1.60E-03	7.69	3.00	39.00
GO:0031248	protein acetyltransferase complex	2.10E-03	4.17	4.00	96.00
GO:1902493	acetyltransferase complex	2.10E-03	4.17	4.00	96.00
GO:0051568	histone H3-K4 methylation	3.13E-03	5.88	3.00	51.00
GO:0008287	protein serine/threonine phosphatase complex	3.14E-03	5.77	3.00	52.00
GO:1903293	phosphatase complex	3.14E-03	5.77	3.00	52.00
GO:0021695	cerebellar cortex development	3.34E-03	5.56	3.00	54.00
GO:0005876	spindle microtubule	3.36E-03	5.45	3.00	55.00
GO:0061733	peptide-lysine-N-acetyltransferase activity	3.92E-03	5.08	3.00	59.00
GO:0009066	aspartate family amino acid metabolic process	4.13E-03	4.92	3.00	61.00
GO:0043967	histone H4 acetylation	4.96E-03	4.55	3.00	66.00
KEGG:04720	Long-term potentiation	4.97E-03	4.48	3.00	67.00
GO:0034212	peptide N-acetyltransferase activity	4.97E-03	4.48	3.00	67.00
GO:0003725	double-stranded RNA binding	5.20E-03	4.35	3.00	69.00
R-HSA:400253	Circadian Clock	5.22E-03	4.29	3.00	70.00
KEGG:04520	Adherens junction	5.45E-03	4.17	3.00	72.00
R-HSA:1234174	Regulation of Hypoxia-inducible Factor (HIF) by oxygen	5.91E-03	4.00	3.00	75.00
R-HSA:2262749	Cellular response to hypoxia	5.91E-03	4.00	3.00	75.00

NPC CLUSTER 2

GOID	GOTerm	Term PValue Corrected with Benjamini-Hochberg	% Associated Genes	Nr. Genes	Total # Genes
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	8.88E-09	7.55	12.00	159.00
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	6.32E-07	4.88	12.00	246.00
GO:0032596	protein transport into membrane raft	6.25E-05	50.00	3.00	6.00
GO:0032594	protein transport within lipid bilayer	1.34E-04	16.67	4.00	24.00
GO:1903044	protein localization to membrane raft	1.55E-04	33.33	3.00	9.00
KEGG:05031	Amphetamine addiction	4.09E-04	7.35	5.00	68.00
GO:0097440	apical dendrite	4.71E-04	21.43	3.00	14.00
GO:1903115	regulation of actin filament-based movement	7.56E-04	8.89	4.00	45.00
GO:1903115	regulation of actin filament-based movement	7.56E-04	8.89	4.00	45.00
GO:0005891	voltage-gated calcium channel complex	7.79E-04	9.09	4.00	44.00
GO:0005891	voltage-gated calcium channel complex	7.79E-04	9.09	4.00	44.00
GO:0071871	response to epinephrine	8.58E-04	15.79	3.00	19.00
GO:0005234	extracellularly glutamate-gated ion channel activity	9.14E-04	15.00	3.00	20.00
KEGG:04713	Circadian entrainment	9.70E-04	5.21	5.00	96.00
GO:0051453	regulation of intracellular pH	1.00E-03	5.26	5.00	95.00
GO:0030641	regulation of cellular pH	1.04E-03	5.05	5.00	99.00
GO:0006885	regulation of pH	1.33E-03	4.72	5.00	106.00
GO:0070252	actin-mediated cell contraction	1.70E-03	4.35	5.00	115.00
GO:0070252	actin-mediated cell contraction	1.70E-03	4.35	5.00	115.00
KEGG:04724	Glutamatergic synapse	1.74E-03	4.39	5.00	114.00
GO:0030004	cellular monovalent inorganic cation homeostasis	1.74E-03	4.27	5.00	117.00
GO:0086003	cardiac muscle cell contraction	1.76E-03	5.88	4.00	68.00
GO:0034704	calcium channel complex	1.76E-03	5.88	4.00	68.00
GO:0086003	cardiac muscle cell contraction	1.76E-03	5.88	4.00	68.00
GO:0034704	calcium channel complex	1.76E-03	5.88	4.00	68.00
GO:0099094	ligand-gated cation channel activity	1.76E-03	5.88	4.00	68.00
GO:0071300	cellular response to retinoic acid	1.78E-03	5.71	4.00	70.00
GO:0005262	calcium channel activity	1.82E-03	4.13	5.00	121.00
GO:0032743	positive regulation of interleukin-2 production	1.89E-03	9.38	3.00	32.00
GO:0051393	alpha-actinin binding	2.55E-03	8.33	3.00	36.00
GO:0051393	alpha-actinin binding	2.55E-03	8.33	3.00	36.00
GO:0086004	regulation of cardiac muscle cell contraction	2.97E-03	7.69	3.00	39.00
GO:0086004	regulation of cardiac muscle cell contraction	2.97E-03	7.69	3.00	39.00
GO:0015297	antiporter activity	2.98E-03	4.65	4.00	86.00
GO:0071867	response to monoamine	3.07E-03	7.32	3.00	41.00
GO:0071869	response to catecholamine	3.07E-03	7.32	3.00	41.00
GO:0035249	synaptic transmission, glutamatergic	3.07E-03	4.76	4.00	84.00
GO:0007215	glutamate receptor signaling pathway	3.07E-03	4.76	4.00	84.00
KEGG:05033	Nicotine addiction	3.07E-03	7.50	3.00	40.00
GO:0001659	temperature homeostasis	3.40E-03	6.98	3.00	43.00
GO:0042805	actinin binding	3.51E-03	6.82	3.00	44.00
GO:0042805	actinin binding	3.51E-03	6.82	3.00	44.00
R-HSA:1592230	Mitochondrial biogenesis	3.62E-03	4.21	4.00	95.00
GO:0014888	striated muscle adaptation	3.62E-03	6.67	3.00	45.00
GO:0006195	purine nucleotide catabolic process	4.09E-03	6.25	3.00	48.00
GO:0005231	excitatory extracellular ligand-gated ion channel activity	4.47E-03	6.00	3.00	50.00
GO:0032663	regulation of interleukin-2 production	5.68E-03	5.45	3.00	55.00

GO:0072523	purine-containing compound catabolic process	5.81E-03	5.36	3.00	56.00
R-HSA:2151201	Transcriptional activation of mitochondrial biogenesis	5.81E-03	5.36	3.00	56.00
GO:0015296	anion:cation symporter activity	5.94E-03	5.26	3.00	57.00
GO:0086065	cell communication involved in cardiac conduction	6.08E-03	5.17	3.00	58.00
GO:0071398	cellular response to fatty acid	6.08E-03	5.17	3.00	58.00
GO:0086065	cell communication involved in cardiac conduction	6.08E-03	5.17	3.00	58.00
GO:0098739	import across plasma membrane	7.46E-03	4.76	3.00	63.00
GO:0099516	ion antiporter activity	7.46E-03	4.76	3.00	63.00
GO:0032623	interleukin-2 production	7.46E-03	4.76	3.00	63.00
GO:0051966	regulation of synaptic transmission, glutamatergic	7.93E-03	4.62	3.00	65.00
KEGG:04720	Long-term potentiation	8.41E-03	4.48	3.00	67.00
GO:0003341	cilium movement	8.55E-03	4.41	3.00	68.00
GO:0006901	vesicle coating	8.55E-03	4.41	3.00	68.00
GO:0032729	positive regulation of interferon-gamma production	8.55E-03	4.41	3.00	68.00
GO:0009166	nucleotide catabolic process	8.55E-03	4.41	3.00	68.00
GO:1901863	positive regulation of muscle tissue development	8.70E-03	4.35	3.00	69.00
GO:0048199	vesicle targeting, to, from or within Golgi	9.55E-03	4.17	3.00	72.00
GO:0050795	regulation of behavior	9.70E-03	4.11	3.00	73.00
GO:1901292	nucleoside phosphate catabolic process	9.85E-03	4.05	3.00	74.00

NEURON CLUSTER 1

GOID	GOTerm	Term PValue Corrected with Benjamini-Hochberg	% Associated Genes	Nr. Genes	Total # Genes
GO:0007156	homophilic cell adhesion via plasma membrane adhesion molecules	5.18E-23	13.84	22.00	159.00
GO:0098742	cell-cell adhesion via plasma-membrane adhesion molecules	4.52E-19	8.94	22.00	246.00
GO:0051001	negative regulation of nitric-oxide synthase activity	4.06E-04	25.00	3.00	12.00
GO:0005891	voltage-gated calcium channel complex	7.33E-04	9.09	4.00	44.00
GO:0032769	negative regulation of monooxygenase activity	7.63E-04	18.75	3.00	16.00
GO:0071871	response to epinephrine	8.69E-04	15.79	3.00	19.00
GO:0051966	regulation of synaptic transmission, glutamatergic	2.08E-03	6.15	4.00	65.00
KEGG:05031	Amphetamine addiction	2.20E-03	5.88	4.00	68.00
GO:0086003	cardiac muscle cell contraction	2.20E-03	5.88	4.00	68.00
GO:0034704	calcium channel complex	2.20E-03	5.88	4.00	68.00
KEGG:04724	Glutamatergic synapse	2.28E-03	4.39	5.00	114.00
GO:0051354	negative regulation of oxidoreductase activity	2.77E-03	9.09	3.00	33.00
GO:0008038	neuron recognition	3.25E-03	8.33	3.00	36.00
GO:0086004	regulation of cardiac muscle cell contraction	3.48E-03	7.69	3.00	39.00
GO:0007215	glutamate receptor signaling pathway	3.64E-03	4.76	4.00	84.00
GO:0035249	synaptic transmission, glutamatergic	3.64E-03	4.76	4.00	84.00
GO:0071867	response to monoamine	3.74E-03	7.32	3.00	41.00
GO:0071869	response to catecholamine	3.74E-03	7.32	3.00	41.00
KEGG:04713	Circadian entrainment	4.20E-03	4.17	4.00	96.00
GO:0004867	serine-type endopeptidase inhibitor activity	4.27E-03	4.08	4.00	98.00
GO:0030212	hyaluronan metabolic process	4.28E-03	6.82	3.00	44.00
KEGG:04973	Carbohydrate digestion and absorption	4.28E-03	6.82	3.00	44.00
GO:0099601	regulation of neurotransmitter receptor activity	4.28E-03	6.67	3.00	45.00
GO:1903115	regulation of actin filament-based movement	4.28E-03	6.67	3.00	45.00
GO:0006195	purine nucleotide catabolic process	4.34E-03	6.25	3.00	48.00
GO:0050999	regulation of nitric-oxide synthase activity	5.47E-03	5.66	3.00	53.00
GO:0098900	regulation of action potential	5.78E-03	5.45	3.00	55.00
GO:0072523	purine-containing compound catabolic process	5.81E-03	5.36	3.00	56.00
GO:0086065	cell communication involved in cardiac conduction	6.14E-03	5.17	3.00	58.00
GO:0030374	ligand-dependent nuclear receptor transcription coactivator activity	6.17E-03	5.08	3.00	59.00
GO:0030276	clathrin binding	7.43E-03	4.69	3.00	64.00
GO:0032768	regulation of monooxygenase activity	7.79E-03	4.55	3.00	66.00
KEGG:04720	Long-term potentiation	7.82E-03	4.48	3.00	67.00
GO:0032729	positive regulation of interferon-gamma production	7.85E-03	4.41	3.00	68.00
GO:0009166	nucleotide catabolic process	7.85E-03	4.41	3.00	68.00
GO:1901863	positive regulation of muscle tissue development	7.89E-03	4.35	3.00	69.00
GO:0071300	cellular response to retinoic acid	7.94E-03	4.29	3.00	70.00
GO:1901292	nucleoside phosphate catabolic process	8.95E-03	4.05	3.00	74.00

NEURON CLUSTER 2

GOID	GOTerm	Term PValue Corrected with Benjamini-Hochberg	% Associated Genes	Nr. Genes	Total # Genes
GO:0070603	SWI/SNF superfamily-type complex	1.11E-04	7.79	6.00	77.00
GO:1904949	ATPase complex	2.81E-04	5.88	6.00	102.00
GO:0090568	nuclear transcriptional repressor complex	2.87E-04	11.11	4.00	36.00
GO:0090568	nuclear transcriptional repressor complex	2.87E-04	11.11	4.00	36.00
GO:0006338	chromatin remodeling	3.20E-04	4.24	7.00	165.00
GO:0006338	chromatin remodeling	3.20E-04	4.24	7.00	165.00
GO:0097346	INO80-type complex	1.17E-03	13.64	3.00	22.00
GO:0000118	histone deacetylase complex	2.15E-03	5.56	4.00	72.00
KEGG:04110	Cell cycle	2.29E-03	4.03	5.00	124.00
R-HSA:6804758	Regulation of TP53 Activity through Acetylation	2.49E-03	10.00	3.00	30.00
KEGG:04914	Progesterone-mediated oocyte maturation	4.00E-03	4.04	4.00	99.00
GO:0004812	aminoacyl-tRNA ligase activity	4.28E-03	6.38	3.00	47.00
GO:0016575	histone deacetylation	4.28E-03	4.35	4.00	92.00
GO:0016706	oxidoreductase activity, acting on paired donors, with incorporation or reduction of molecular oxygen, 2-oxoglutarate as one donor, and incorporation of one atom each of oxygen into both donors	4.36E-03	6.52	3.00	46.00
GO:0050681	androgen receptor binding	4.36E-03	6.52	3.00	46.00
GO:0000422	autophagy of mitochondrion	4.38E-03	4.21	4.00	95.00
GO:0061726	mitochondrion disassembly	4.38E-03	4.21	4.00	95.00
R-HSA:379724	tRNA Aminoacylation	4.47E-03	7.14	3.00	42.00
GO:0006418	tRNA aminoacylation for protein translation	4.69E-03	5.88	3.00	51.00
GO:0043038	amino acid activation	5.13E-03	5.45	3.00	55.00
GO:0000049	tRNA binding	5.17E-03	5.56	3.00	54.00
GO:0043039	tRNA aminoacylation	5.17E-03	5.56	3.00	54.00
GO:0050661	NADP binding	5.36E-03	5.26	3.00	57.00
GO:1903146	regulation of autophagy of mitochondrion	5.36E-03	5.26	3.00	57.00
GO:0070491	repressing transcription factor binding	6.72E-03	4.76	3.00	63.00
GO:1901983	regulation of protein acetylation	7.22E-03	4.48	3.00	67.00
GO:0051881	regulation of mitochondrial membrane potential	7.22E-03	4.48	3.00	67.00
KEGG:00970	Aminoacyl-tRNA biosynthesis	7.27E-03	4.55	3.00	66.00
KEGG:04662	B cell receptor signaling pathway	8.09E-03	4.23	3.00	71.00
GO:0030521	androgen receptor signaling pathway	8.31E-03	4.05	3.00	74.00
GO:0043044	ATP-dependent chromatin remodeling	8.35E-03	4.11	3.00	73.00

GLIA CLUSTER 2

GOID	GOTerm	Term PValue Corrected with Benjamini-Hochberg	% Associated Genes	Nr. Genes	Total # Genes
GO:0002292	T cell differentiation involved in immune response	1.05E-04	6.45	4.00	62.00
GO:0002286	T cell activation involved in immune response	2.95E-04	4.17	4.00	96.00
GO:0051310	metaphase plate congression	4.94E-04	5.26	3.00	57.00
GO:0002287	alpha-beta T cell activation involved in immune response	5.86E-04	5.36	3.00	56.00
GO:0002293	alpha-beta T cell differentiation involved in immune response	5.86E-04	5.36	3.00	56.00
GO:0043367	CD4-positive, alpha-beta T cell differentiation	6.34E-04	4.55	3.00	66.00
KEGG:04520	Adherens junction	7.01E-04	4.17	3.00	72.00
GO:0002294	CD4-positive, alpha-beta T cell differentiation involved in immune response	7.40E-04	5.45	3.00	55.00

Table 3.20 | GOs of gene clusters with high correlation scores from Table 3.19.

Significant GO terms from gene lists for each coexpression cluster from Figure 3.10, D.

NPC (PGC + CLOZUK)	NEURON (PGC + CLOZUK)	GLIA (PGC + CLOZUK)	NPC (PGC)	NEURON (PGC)	GLIA (PGC)
ACTR8	ACTR5	ADSL	CA6	ENO1	UTS2
ADSS	ACTR8	ADSS	SLC2A5	CA6	RNF220
AKT3	ADSS	ALMS1	EPB41	SLC2A5	ADSS
ALMS1	AKT3	ANKHD1	PPIH	H6PD	GRIN2A
ANKRD45	ALMS1	AOX1	EBNA1BP2	MPL	MAPK3
ANP32E	AMBRA1	CCT7	ST3GAL3	CDC20	NFATC3
AOX1	ANKRD45	CHRM3	B4GALT2	MED8	TCF4
ARNT	ANP32E	CNKSR2	DPYD	RNF220	JUND
ASB5	AP1S3	CUL3	ANP32E	ANP32E	DPP4
ASH2L	ATG13	DPP4	RPRD2	SEMA6C	HECW2
B4GALT2	ATXN7	EP300	ARNT	IL19	SF3B1
BRD8	BAHD1	GPER	SEC16B	PFKFB2	MOB4
CA6	BAZ2A	GRIA1	SDCCAG8	PLXNA2	AOX1
CACNA1C	BOLL	GRIN2A	AKT3	SDCCAG8	CUL3
CACNA1D	CA6	HECW2	ADSS	AKT3	UGT1A8
CACNA2D3	CACNA1C	JUND	NT5C2	ADSS	UGT1A10
CCNH	CACNA1D	KLHL20	TAF5	SUFU	UGT1A9
CDC25C	CCT7	MAPK3	PHF21A	NT5C2	UGT1A7
CENPL	CDC20	MASK-BP3	CACNA1C	TAF5	UGT1A6
CHST12	CDC25C	MEF2C	PPP1CC	PHF21A	UGT1A5
CIRH1A	CENPL	MOB4	CLIP1	CREB3L1	UGT1A4
CLCN3	CHDH	NFATC3	KNTC1	DGKZ	ADSL
CLIP1	CPEB1	NMUR2	GTF2H3	AMBRA1	EP300
CLU	CREB3L1	PDE4B	TCTN2	ATG13	SH3RF1
CREB3L2	CTNNA1	PLXNA4	GRIN2A	CTNND1	MEF2C
CSGALNACT1	CTNND1	RNF220	UTP4	CACNA1C	ANKHD1
CUL3	CUL3	SDCCAG8	SNTB2	BAZ2A	NMUR2
CYP26B1	DGKD	SF3B1	NCAN	MARS	GRIA1
DCTN1	DGKZ	SH3RF1	SUGP1	PLCB2	GPER1
DNAJC18	DNAJC19	SLC4A5	MAU2	BAHD1	PLXNA4
DNAJC19	DNAJC6	TCF4	GATAD2A	SH3GL3	CNKSR2
DPYD	DPYD	UGT1A10	CYP26B1	GRIN2A	
EBNA1BP2	DYSF	UGT1A5	EPC2	PSKH1	
EP300	EGR4	UGT1A8	PSMD14	PSMB10	
EPB41	ENO1	UTS2	HECW2	SLC12A4	
EPC2	EPC2		SF3B1	WDR81	
EPHB1	EPHA7		HSPD1	NT5M	
ETF1	EPN2		SGO2	SREBF1	
FGFR1	FBXO41		AOX1	GATAD2A	
GAB1	GABBR1		CUL3	DYSF	
GATAD2A	GATAD2A		UGT1A8	HECW2	
GBF1	GRIA1		UGT1A10	SF3B1	
GFRA3	GRIN2A		UGT1A9	CUL3	
GPM6A	GRM3		UGT1A7	PDE6D	
GRIA1	H6PD		UGT1A6	DGKD	
GRIN2A	HARS		EP300	ACTR5	
GTF2H3	HARS2		RANGAP1	POLDIP3	
HDAC3	HDAC2		OXNAD1	SEMA3G	
HECW2	HECW2		PLCL2	ITIH3	
HSPA9	HSPA9		ITIH3	ITIH4	
HSPD1	IL19		ITIH4	THOC7	
INO80B	INO80B		CACNA1D	ATXN7	
INPP4B	INPP4B		EPHB1	PPP2R3A	
INPP5D	INPP5D		PPP2R3A	STAG1	
ITGA9	ITIH3		PCCB	DNAJC19	
ITIH3	ITIH4		STAG1	PPARGC1A	
ITIH4	JAK1		NCK1	PDE4D	
KLHL20	KAT5		DNAJC19	CDC25C	
KNTC1	KDM4A		PPARGC1A	HSPA9	
MAD1L1	KIAA1009		SH3RF1	CTNNA1	
MAU2	LEPR		CLCN3	HARS	
MEF2C	MAD1L1		GPM6A	HARS2	
MLL5	MAP3K11		ASB5	ZMAT2	
NCAN	MARS		PDE4D	SPARC	
NCK1	MBTPS2		BRD8	GRIA1	
NFATC3	MED8		CDC25C	GABBR1	
NISCH	MEF2C		HSPA9	SNAP91	
NLGN4X	MLL5		DNAJC18	CEP162	
NT5C2	MPL		HDAC3	PSMG3	
OXNAD1	NFATC3		GRIA1	MAD1L1	
PCCB	NGEF		MAD1L1	GRM3	
PDE4B	NT5C2		CHST12	CA8	
PDE4D	NT5M		KMT2E	MBTPS2	
PHF21A	OXNAD1		CREB3L2		
PLCL2	PCGF6		CLU		
PLXNA2	PDE4B		NLGN4X		
PPARGC1A	PDE4D		RPS6KA3		
PPIH	PDE6D				
PPP1CC	PEMT				
PPP2R2A	PFKFB2				
PPP2R3A	PHF21A				
PSMD14	PLCB2				
RANGAP1	PLXNA2				
RASA1	POLDIP3				
RASAL2	PPARGC1A				

RPRD2	PPP2R2A				
RPS6KA3	PPP2R3A				
SDCCAG8	PRKCB				
SEC16B	PRKD1				
SEMA3G	PSKH1				
SF3B1	PSMB10				
SGOL2	PSMG3				
SH3RF1	RNF220				
SLC2A5	SDCCAG8				
SLC4A5	SEMA3G				
SNTB2	SEMA6C				
ST3GAL3	SH3GL3				
STAG1	SLC12A4				
SUGP1	SLC2A5				
TAF5	SMG6				
TCF4	SNAP91				
TCTN2	SPARC				
TNN	SREBF1				
TNR	STAG1				
UGT1A10	STAT6				
UGT1A5	SUFU				
UGT1A8	TAF5				
	THOC7				

Table 3.21 | Protein networks (string-db) in SZ risk connections.
Proteins recognized by string-db as part of an association network from the list of genes in Table 3.15 for NPCs, neurons, and glia. Highlighted groups of rows (in PGC2 set) for indicated genes that are located in the same or adjacent TADs (i.e., chromatin colocalization). Different colors are merely for ease of visualization.

	Schizophrenia (n=23)	Healthy/Control (n=20)
Age		
Mean	82.2	83.3
SD	8.6	10
Sex, n (%)		
Male	8 (35)	7 (35)
Female	15 (65)	13(65)
PMI		
Mean	11.4	8
SD	3.6	4.2

Table 3.22 | Demographics of brain cohort used for proteomic analysis. Demographic information of postmortem adult brains that were assayed for neuronal protein quantification with LC-SRM MS. PMI, post-mortem interval

Materials and methods

In situ Hi-C from hiPSC-derived cells

In situ Hi-C libraries were generated from 2 million to 5 million cultured hiPSC derived NPCs, glia, and neurons as described in (Rao *et al.*, 2014) without modifications in the protocol. Briefly, in situ Hi-C consists of 7 steps: (i) crosslinking cells with formaldehyde, (ii) digesting the DNA using a 4-cutter restriction enzyme (e.g., MboI) within intact permeabilized nuclei, (iii) filling in and biotinylating the resulting 5′-overhangs, (iv) ligating the blunt ends, (v) shearing the DNA, (vi) pulling down the biotinylated ligation junctions with streptavidin beads, and (vii) analyzing these fragments using paired end sequencing. As quality control (QC) steps, we checked for efficient restriction with an agarose DNA gel and for appropriate size selection using the Agilent Bioanalyzer after steps (v) and (vi). For the final QC, we performed superficial sequencing on the Illumina MiSeq (~2-3Mreads/sample) to assess quality of the libraries using metrics such as percent of reads passing filter, percent of chimeric reads, and percent of forward-reverse pairs (Table 3.1). For the forebrain directed differentiation neuronal library from subject S1, the Arima Hi-C kit (Arima Genomics, San Diego) was used according to the manufacturer's instructions.

Hi-C read mapping and matrix generation

The Hi-C libraries were sequenced on the Illumina HiSeq1000 platform (125bp paired-end) (New York Genome Center). Technical replicates of subject

S2 NPCs, neurons, and glia were also sequenced to enhance resolution. Initial processing of the raw 2 ×125 bp read pair FASTQ files was performed using the HiC-Pro analysis pipeline (Servant *et al.*, 2015). In brief, HiC-Pro performs four major tasks: aligning short reads, filtering for valid pairs, binning, and normalizing contact matrices. HiCPro implements the truncation-based alignment strategy using Bowtie v2.2.3 (Langmead *et al.*, 2009), mapping full reads end-to-end or the 5′ portion of reads preceding a GATCGATC ligation site that results from restriction enzyme digestion with MboI followed by end ligation. Invalid interactions such as same-strand, dangling-end, self-cycle, and single-end pairs are not retained. Binning was performed in 10kb, 40 kb and 100 kb nonoverlapping, adjacent windows across the genome and resulting contact matrices were normalized using iterative correction and eigenvector decomposition (ICE) as previously described (Imakaev *et al.*, 2012), using HiC-Pro's default settings of 100 maximum iterations, filtering of the sparse bins (lowest 2%), and a relative result increment of 0.1 before declaring convergence (http://nservant.github.io/HiC-Pro/MANUAL.html). Data are reported in browser-extensible-data-like (BED) format and visualized in the Washington University Epigenome Browser (http://epigenomegateway.wustl.edu). Hierarchical clustering was performed on the ICE-corrected intrachromosomal contact matrices after the bins with the 1% most extreme interaction values were excluded as largely artifactual. Clustering was performed using Ward's method on the 1, 5, 10, 25, 50, and 100% most variable remaining bins using (1-

correlation) as a distance metric. The results using the 10% most variable interaction bins, shown here in a cluster dendrogram and a Pearson correlation matrix, are representative of these results.

Hi-C loop calls using Juicer

Loop calling was performed using the software HiCCUPS (Rao *et al.*, 2014). To format data for HiCCUPS input, we remapped reads from Hi-C libraries using the Juicer pipeline (Durand, Shamim, *et al.*, 2016). Similar to HiC-Pro, the Juicer pipeline performs read alignment, filtering, binning, and matrix normalization. Samples were pooled for each cell type (S1 and 2 technical replicates from S2) to generate the maximum amount of coverage required for accurate loop calling. The resulting .hic matrix files (MAPQ > 0) were then used as input to HiCCUPS. The following parameters were set for HiCCUPS following the analysis in (Rao *et al.*, 2014): FDR threshold (*f*) = 0.10, 0.10; peak width (*p*) = 4, 2; window width (*i*) = 7, 5; merge distance (*d*) = 20 kb, 20 kb. Values for parameters correspond to calls made at 5kb and 10kb, respectively. Representative neuronal and non-neuronal loops are presented in Figure 3.4. As the number of loops called is dependent upon the number of Hi-C contacts in the matrix (Forcato *et al.*, 2017), we also generated matrices with equivalent total Hi-C contacts via subsampling. hiPSC-derived Hi-C interaction matrices were randomly subsampled to 372,787,143 cis only contacts (the lowest number of cis contacts across all cell types) and HiCCUPS was rerun on the subsampled matrices. After loops were called for each cell type, we performed a reevaluation

on this union set of loop loci. HiCCUPS was rerun using the union set of loop loci as input to produce *q*-values for each loop in the union set for every cell type. By default, HiCCUPS does not output a *q*-value for every pixel. Hence, this reevaluation produced q-values for pixels in cells that did not pass the significance threshold. We then defined any pixel from the union set with a *q*-value < 0.10 with respect to the donut neighborhood surrounding the pixel to be a loop and defined the loop to be shared with any cell types having a *q*-value < 0.10 for the same pixel.

These loop calls were used for comparing loop calls between cell types. Loops were also called and subsampled as above for the GM12878 cell line using the processed data from (Rao *et al.*, 2014) found here: www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE63525. Loop calls were overlapped with compartment calls (Materials and methods), such that AA, BB, and AB refer to loops with both anchors in A, both anchors in B, and one anchor in A and other anchor in B, respectively. Loops in chromosomes 4, 18, 19, and X were removed from this compartment analysis since the first principle component most likely corresponded to p versus q arm distinctions and not A versus B compartments.

Hi-C interactions at risk loci

To approach 3DG conformation in context of the disease-relevant sequences, we adapted the binomial statistics based mapping strategy previously described by Won et al (Won *et al.*, 2016). The set of schizophrenia

risk loci used in this study included the original (PGC2, Psychiatric Genomics Consortium) (Ripke *et al.*, 2014) risk sequences, or 108 physically distinct association loci defined by 128 index SNPs (corrected P 10^{-8}) and an additional 37 loci from the CLOZUK (a series of UK cases registered for clozapine treatment with a clinical diagnosis of schizophrenia) study for a total of 145 loci defined by 179 independent genome-wide significant SNPs (corrected $P < 5 \times 10^{-8}$), determined by GWAS in 40,675 cases and 64,643 controls (Pardiñas *et al.*, 2018). A risk locus is defined as a collection of (SNPs) existing in linkage disequilibrium, ranging from 1bp to 8.9Mb (average 256.2 kb) in length and in total equivalent to approximately 0.012% of human genomic sequence.

To identify significantly enriched interactions involving a bin of interest with another bin, our principal approach was to first estimate the expected interaction counts for each interaction distance by calculating the mean of all intrachromosomal bin-bin interactions of the same separation distance throughout the raw intrachromosomal contact matrix. We used the R package, HiTC (Servant *et al.*, 2012), to facilitate manipulation of our HiC-Pro-produced raw contact matrices and estimation of the expected counts at various interaction distances. The probability of observing an interaction between a bin-of-interest and another bin was then defined as the expected interaction between those two bins divided by the sum of all expected interactions between the bin-of-interest and all other intrachromosomal bins. A *P* value was then calculated as binomial probability of observing the number of interaction counts or more between the

bin-of interest and some other bin where the number of successes was defined as the observed interaction count, the number of tries as the total number of observed interactions between the bin-of-interest and all other intrachromosomal bins, and the success probability as the probability of observing the bin-bin interaction estimated from the expected mean interaction counts. The Benjamini-Hochberg method was used to control false discovery rate (FDR) for *P* values determined for all interactions with a bin-of interest (includes all bins 1Mb up and downstream in our tests).

Generation of stable selected dCas9-VP64/VPR and Cas9 NPCs

All CRISPR-based epigenomic editing assays were performed on antibiotic selected dCas9-VP64 (VP64 as the tetrameric VP16 activator domain) and dCas9-VPR (VPR as the tripartite activator, VP64-p65-Rta) NPCs derived as described in (Ho *et al.*, 2017). For generation of Cas9 stable, selected NPCs, we used a plasmid of lentiCRISPR v2 gifted by Feng Zhang (Addgene plasmid # 52961). DNA sequencing with a U6 primer confirmed the identity. Lentiviral production and titration were performed as described previously (Ho *et al.*, 2016). Control S1 and S2 NPCs were spinfected with lentiCRISPR v2 virus as described (Ho *et al.*, 2017). 48 hours post-transduction, cells were selected by exposure to puromycin at 0.3 µg/mL. Without transduction, all control cells died within around 5 days after the antibiotic addition. The puromycin-selected NPCs were subject to Western blot analysis of Cas9 expression. 30 µg of proteins were electrophoresed in NuPAGE 4-12% Bis-Tris Protein Gels (NP0323PK2, Life

Technologies) in 1× MES running buffer, 200 V constant, 35 min. Proteins were transferred onto nitrocellulose membrane (IB23002, Life Technologies) on the iBlot® 2 Dry Blotting System (program P3, 7:00 min). The membranes were incubated with primary antibodies against Cas9 (1:250, monoclonal, clone 7A9, Millipore) and b-Actin (1:10,000, mouse, 1406030, Ambion) overnight at 4°C. Then, membranes were incubated with the IRDye-labeled secondary antibodies for 45 min at RT in the dark on the rocker. Fluorescence was visualized using a Li-CorOdyssey Imaging System.

In vitro transcription and transfection of gRNAs

Guide RNAs (gRNAs) were designed on Benchling (www.benchling.com) using the CRISPR tool. gRNAs were generated via in vitro transcription (IVT) with the GeneArt Precision gRNA Synthesis Kit (Thermo Fisher Scientific, A29377) as per manufacturer instructions. Five gRNAs were designed per condition (i.e., “loop SNP”, negative control, and positive control) and pooled for transfection. The genomic ranges within which loop-SNP gRNAs were designed (i.e., region spanning the SNP of interest and all gRNAs in the condition) were roughly 600 bp for *ASCL1*, 550 bp for *MATR3*, 460 bp for *EFNB1* (with 2/5 gRNAs directly overlapping the SNP), 300 bp for *SOX2*. Puromycin-selected (1µg/mL in NPC media; Sigma, #P7255) dCas9-VP64 and dCas9- VPR NPCs (Ho *et al.*, 2017) were seeded at a density of ~400,000 per well on Matrigel-coated (BD Biosciences) 24-well plates. Pooled IVT gRNAs (500 ng total RNA/well) and 2 µL EditPro Stem lipofectamine (MTI-GlobalStem, #GST-2174;

now, ThermoFisher, STEM00003) were diluted in 50 µL Opti-MEM (Thermo Fisher Scientific, #31985062) and added dropwise to each well. Cells were harvested with TRIzol for total RNA extraction 48 hours later. All experiments were conducted with 3 to 6 biological replicates from 1 donor (subject S1), generated in parallel, with the donor contributing isogenic dCas9-VP64 and dCas9-VPR effector cells. Each data point in Figure 3.8, D to F, represents one biological replicate within each condition. For each target gene promoter and candidate loop, control gRNAs were strategically placed into the middle third of the (linear) genome portion bypassed by the candidate loop. CRISPRa results were analyzed on PRISM with a one-way ANOVA across 3 conditions with a Dunnett's test for multiple comparisons. Cas9 mutagenesis was also performed as described above with the exception of the negative control, which in these experiments consisted of an empty transfection (i.e., lipofectamine + Opti-MEM without any gRNA). Cas9 results were analyzed with an unpaired *t* test comparing the loop-SNP and negative control conditions.

RNA transcriptomic correlation heatmaps

Pearson correlation coefficient matrices were calculated for gene expression in the childhood onset schizophrenia data set (Hoffman *et al.*, 2017) using R from lists of genes that are located in cell-type-specific loops anchored at schizophrenia risk loci and, as a subset of this list, sets of genes whose proteins participate in an association network for each of the three cell types (see below). Significance was computed calculating the absolute mean correlation coefficient

of each correlation matrix (“organization score”) as a test statistic against a null distribution generated by random gene sampling. Randomized gene lists were drawn only from the pool of genes with over 1 count per million (CPM) in at least 30% of the experiments described in (Hoffman *et al.*, 2017). To generate a null distribution of organization scores for a given cell type that accounted for genomic distance and neighborhood effects, we began by randomly selecting a significant PGC interaction for that cell type. Using the bp genomic distance of this interaction we randomly selected two 10kb bins from the genome separated by the same distance. All genes overlapping these bins were then added to the list of genes with which to calculate the organization score. This process was iterated until enough genes were added to the list to match the number of genes used in the original cell-type-specific organization score. Finally, this protocol was repeated 1000 times to generate the null distribution of random organization scores. This distribution was then used to calculate significance of co-regulation (i.e., P = number of times $|r|_{avg}$ of the null exceeded that of the test heatmap / 1000). Note that STRING gene network transcriptomic analyses (Figure 3.13, B) were performed with 1000 random permutations of genes sampled from the full schizophrenia risk connectome (i.e., risk locus + risk locus connect genes) for each cell type.

Generation of hiPSC-derived cell types

NPCs were derived from human iPSCs, as described (Brennand *et al.*, 2011). Briefly, hiPSCs reprogrammed from two independent controls (Subjects

S1 and S2) fibroblasts were cultured as embryoid bodies and maintained in suspension in N2 medium (Invitrogen) for 7 days. They were then plated onto polyornithine/laminin-coated plates and visible rosettes were manually dissected after 1 week and cultured in NPC medium (Brennand *et al.*, 2011). NPCs were maintained at high density on Matrigel (BD Biosciences) and split approximately 1:4 once a week with Accutase (Millipore). In our hands, doubling time of hiPSC-derived forebrain NPCs is 3.69±0.05 days, with the percentage of cells in G1 cell cycle phase is 62.3 ± 0.2%, in S phase is 24.6±0.1% and G2 phase is 6.5 ± 0.03% (Brennand *et al.*, 2015).

NPCs were further differentiated along the glial lineage as previously described (TCW *et al.*, 2017). Briefly, NPCs described above were seeded at a density of 15,000 cells/cm^2 on Matrigel-coated plates in astrocyte medium (ScienCell). Cells were split at 95% confluency, approximately every 6-7 days, to the initial seeding density with Accutase (Millipore). After 30 days of differentiation, glia were used in assays or were maintained in astrocyte medium (with 2% FBS), splitting 1:3 every week with Accutase.

To arrive at hiPSC-neurons via Neurogenin2 induction as previously described (Ho *et al.*, 2016), NPCs were seeded at low density on Matrigel-coated plates and transduced with lentiviral constructs to overexpress *NGN2*, driven by a TetO promoter, in NPC-medium. Doxycycline (Sigma, #D9891) was added in NPC medium to activate the system and puromycin (Sigma, #P7255) was introduced for selection. 2 days after doxycycline addition, puromycin is

withdrawn and NPC medium was switched to neuron medium (Ho *et al.*, 2016). Cells were maintained with regular half-medium changes until day 20 after first doxycycline addition. At this point, neurons were ready to be used for assays. In our hands, as described, combining puromycin induction and Ara-C treatment to block proliferation of dividing cells results in pure or near-pure (approaching 99-100% homogeneity by cell type) populations of postmitotic excitatory (glutamatergic) neurons, with other cell types undetectable in the culture (Ho *et al.*, 2016).

For forebrain neuronal differentiation, as previously described (Brennand *et al.*, 2011), for S1, NPCs were seeded at low density and cultured in neural differentiation medium (DMEM/F12, 1xN2, 1xB27-RA, 20 ng ml−1 BDNF (Peprotech), 20 ng ml−1 GDNF (Peprotech), 1mM dibutyryl-cyclic AMP (Sigma), 200nM ascorbic acid (Sigma) and 1 μgml−1 laminin (ThermoFisher Scientific) 1–2 days later. Cells were maintained in differentiation medium for 7.5 weeks before harvesting. Note that forebrain differentiated cultures reflect the dual lineage potential of the NPC, and therefore contain a neural population of βIII-TUBULIN-positive neurons, which comprise 70-80% of all cells in the culture. Another 20-30% of cells in culture are glial fibrillary acidic protein (GFAP)-positive astrocytes. In addition, while the majority of forebrain hiPSC neurons express the glutamatergic marker VGLUT1, another (approximately 30%) subset of neurons expresses the GAD67 GABAergic marker gene (Topol, Tran and Brennand, 2015).

Immunofluorescent staining

NPCs: NPCs on coverslips (24-well plate) were fixed in ice cold 4% PFA (Electron Microscopy Sciences, #15714) for 10 minutes and rinsed with PBS (containing Ca^{2+} and Mg^{2+}; Thermo Fisher Scientific 21300-058). Blocking was performed with a buffer consisting 0.5% BSA/0.1% Triton-X (Sigma, #T8787)/PBS for 30 minutes at 4°C. Cells were incubated overnight at 4 degrees with primary antibody that was diluted in the blocking buffer. Primary antibodies used were Nestin (Alexa 647-conjugated, BD. #560341; 1:100) and SOX2 (Goat, Santa Cruz, #sc-17320; 1:500). Cells were then washed twice with PBS for 5 minutes. Then, secondary Alexa 488 donkey anti-goat (Jackson Immuno, #705-545-147; 1:400) was added in 300 µL PBS for 1-2 hours at room temperature. Cells were then washed again twice with PBS. Nuclei were counterstained with DAPI (Sigma, #D9542) for 10-15 minutes in 300 µL PBS. Coverslips containing the fluorescent stained NPCs were mounted onto slides with AquaPolymount mounting solution (Polysciences Inc., #18606-20) that was equilibrated to room temperature. Mounted coverslips were then left to air-dry overnight in the dark.

Neurons: Followed the staining protocol as per (Ho *et al.*, 2016). Primary antibodies used were TUJ1 (Rabbit, Covance, #PRB-435P; 1:1000) and MAP2 (Chicken, ABcam, #ab5392; 1:500). Secondary antibodies used were Alexa 488 donkey anti-chicken (Jackson Immuno, #703-545-155) and Cy3 donkeyanti-rabbit (Jackson Immuno, #711-165-152), both at 1:400.

Glia: Followed the staining protocol as per (TCW *et al.*, 2017). Primary antibodies used were S100β (Mouse, Sigma-Aldrich, #S2532; 1:1000) and Vimentin (Rabbit, Cell Signaling, #3932; 1:500). Secondary antibodies used were Alexa 488 donkey anti-mouse (Jackson Immuno, #715-545-151) and Alexa 647 donkey-a-rabbit (Jackson Immuno, 711-605-152), both at 1:400.

All imaging was performed with a Zeiss LSM 780 confocal microscope.

Hi-C A/B compartment calling

To identify boundaries of low-resolution A/B compartments associated with open and closed chromatin, respectively, the first eigenvector from PCA on the Pearson's correlation matrix of Knight-Ruiz balanced observed/expected interaction frequencies (Knight and Ruiz, 2013) was calculated using `pearson` and `eigenvector` functions within the pre module of the Juicer tools suite (https://github.com/theaidenlab/juicer/wiki/Pre) at 500kb and 1Mb resolution (Durand, Shamim, *et al.*, 2016). Positively and negatively signed PC1 regions were used to establish compartments. For each library, the PC1 region that had higher gene density was defined as open-chromatin-associated compartment A while the other region was defined as closed-chromatin-associated compartment B as per previously described methods (Lieberman-Aiden *et al.*, 2009).

Hi-C topologically associating domain calling

Topologically associated domains (TADs) were called using the `arrowhead` algorithm in the Juicer tools suite with a 10kb resolution and 2kb sliding windows, `-r 10000 –m 2000 –ignore -sparsity`, on the ICE-corrected

contact matrices. The distributions of all called TADs were compared using contact matrices sampled to have similar cis interactions (see below). Statistically significant differences in distributions were determined using the Wilcoxon-Mann-Whitney test.

Cumulative distribution of loop size

Cumulative plots were generated using loop calls from subsampled Hi-C data. Two-sample two-tail Kolmogorov-Smirnov test was used to determine the *p*-value, testing the hypothesis that two samples come from the same distribution. Refer to the script “loop_distance.R” at the following link: https://github.com/tborrman/Schahram-project.

Gene expression v. loop analyses

To test whether looping architecture was associated with increased gene expression, we overlapped genes with our non-subsampled HiCCUPS loop calls. We defined Gene Body Loop genes to be genes that overlap with loop anchor loci and defined TSS Loop genes to be genes whose TSS lies within a loop anchor loci. We then calculated the mean $\log_{10}$(FPKM + 1) for Gene Body loop genes and TSS Loop genes, and performed one-tail *Z*-tests to determine if expressions in gene sets were significantly high. Refer to the scripts “get_overlapping_genes.py” and “FPKM_cell_type_specific_boxplots.R” at the following link: https://github.com/tborrman/Schahram-project.

Gene ontology

To examine the functional role of significant interactions, genes present in bins participating in interactions of interest (e.g., brain-specific loops, celltype-specific schizophrenia risk interactions, etc) were tested for enrichment of gene ontology categories using ClueGO (adjusted *p*-value threshold of < 0.05 with Bonferroni step-down or Benjamini-Hochberg correction, at “Medium” or “Medium-high” network specificity, and with “GO Term Fusion” enabled to avoid redundant terms (Bindea *et al.*, 2009). For the brain-specific loops GO, we took the 818 loops shared by all NPCs, glia, and astrocytes from the non-subsampled datasets. For the GO of loops lost in the neurons or glia upon transitioning from NPCs, the union set of all loops called with HiCCUPS was filtered to extract only those that were *q*-value < 0.1 (“present”) in NPCs and *q*-value > 0.1 (“lost”) in neurons or glia. Genes were annotated from anchors of the top 1000 most significant loops lost in neurons or glia.

ATAC-seq and accessibility processing

To assess chromatin accessibility, ATAC-seq was performed as described in (Buenrostro *et al.*, 2015). Briefly, 50,000 cells from each sample were snap frozen, resuspended in cold lysis buffer (10 mM Tris-HCl pH 7, 10 mM NaCl, 3 mM $MgCl_2$, 0.01% Igepal CA-630) and spun down at 500 x g for 10 min at 4°C. Cells were incubated with 25 µL TD buffer, 2.5 µL Tn5 transposase (Illumina; Cat No: #FC-121-1030) and 22.5 µL nuclease-free H_20 at 37°C for 30 min. After purification (Qiagen; Cat No: #28004), transposed DNA fragments were amplified

with custom Nextera PCR primers (Buenrostro *et al.*, 2013) for a total of 9 cycles. To minimize GC and size bias, the final PCR cycle number needed to minimally amplify libraries was determined by a separate qPCR reaction as described in Buenrostro et al., 2015. Purified libraries were sequenced on an Illumina HiSeq 2500.

To evaluate accessibility, ATAC-seq raw FASTQs were quality (minimum Q20) and adapter trimmed using Trim Galore! v0.4.1 (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/). Trimmed reads were aligned to hg19 using `bwa mem` v0.7.17. Aligned reads were then Insertion-Deletion realigned and Base Quality Score Realigned using the Genome Analysis Tool Kit v3.1.1 (McKenna *et al.*, 2010). Duplicates were marked using Picard MarkDuplicates v1.137 (https://broadinstitute.github.io/picard/). Genome-wide accessibility signal was computed from R1 alignments with 5' ends shifted 4 bases and 3' ends shifted 5 bases towards the center of the purported transposition event. Shifted R1 alignments were then aggregated at each base pair. *NGN2*-neuron Rep1 and NPC libraries were subsampled to approximately 227 million reads, comparable to the lowest coverage library, *NGN2*-neuron Rep2. Aggregated coverage in the subsampled libraries was log2-converted. To evaluate genome-wide accessibility, the mean accessibility signal across 1kb-tiled windows was calculated. The mean ATAC-seq signal within HiCCUPS loop anchor bin coordinates was used to assess the accessibility of loop calls.

Estimating significance of rate of loop/eQTL overlap

In order to assess whether the observed rate of overlap of our brain-specific (i.e., loops shared by NPCs, neurons, and glia) Hi-C loops with the Common Minds Consortium brain cis-eQTLs (with FDR < 10^{-5}) was significant, we first constructed a set of background loops. To do so, we annotated the anchors of the observed loops with functional classes using ChIPseeker (Yu, Wang and He, 2015). We used separate annotations to represent the 0-1kb, 1-2kb, 2-3kb, 3-4kb, and 4-5kb upstream promoter. To retain any hidden nuance in exon/intron annotations, we also preserved intron and exon numbers 1 through 5 and a separate merged class of intron and exon numbers 6 or greater. Downstream intergenic and unassociated intergenic served as additional sets of annotations. We similarly annotated the set of all 10kb genomic bins used in creating the Hi-C contact matrix. To create the null sets, we randomly selected a genomic bin that matched the annotation of the first anchor of our observed loops, then scanned up and downstream for a second bin a similar distance away (within ±25% of the observed loop distance), with a matching functional annotation to the second loop anchor. For each loop, we tried up to 100 random bin selection iterations to find a suitable bin pair match (with a failure rate of approximately 0.1%). Thus, each null set was similar or identical in size to the observed loop set. All bins associated with our observed loops and the bins ±20kb flanking its anchors were excluded from the null sets. We overlapped the background loops with the Common Minds cis-eQTLs with FDR<10^{-5} to estimate

the null distribution of the rate of loop/eQTL overlap. To do so, we randomly sampled 1000 null sets as described and calculated the fraction of each that fully overlapped an eQTL—requiring both SNP and gene-body to overlap each loop anchor. The one-sided empirical *p*-value was defined as the fraction of null-loop/eQTL overlap rates that were greater than or equal to our observed brain-specific-loop/eQTL overlap rate.

Comparing interaction intensity within interactions with PCDH locus

Observed/Expected (O/E) interaction scores were extracted from the subsampled Glia, NPC, and Neuron *.hic files using juicer-tools (https://github.com/aidenlab/juicer/wiki/Data-Extraction) (Durand, Shamim, *et al.*, 2016). Interactions falling within the regions of interest were selected:

1. chr5:140120000:140222664 || chr5:140110382:140960850
2. chr5:139730111-139952633 || chr5:140023665-140119999

Violin plots for all bin-bin interactions falling within the regions of interest were plotted and significant differences in O/E scores were tested for using a Wilcoxon-Rank Sum test.

Single SNP-level eQTL analysis

For the 61 genes within the PCDH locus with above-threshold expression, gene-level gene expression quantitative trait loci (eQTL) were obtained for the rs111896713 genotype across the N = 579 genetically inferred samples within the published CommonMind Consortium dataset (Fromer *et al.*, 2016) A linear

model on the imputed genotype dosages was used followed by an *f*-test with Bonferroni correction for multiple testing.

Real-time qPCR

Quantitative expression analysis was performed with a QuantStudio 7 Flex Real-Time qPCR System using the Power SYBR Green RNA-to-Ct Real Time qPCR kit for primers (Thermo Fisher Scientific). 50 ng of RNA template was added to PCR master mix containing primers. qPCR conditions were 48°C for 15 minutes, 95°C for 10 minutes, and 40 cycles of 95°C for 15 seconds followed by 60°C for 60 seconds.

Determining cell-type-specific interactions

The distribution of all $-\log_{10} q$-values from Hi-C interaction mapping of PGC risk loci indicated the 95th percentile log *q*-value = 1.47. For an interaction to be cell type-specific, it has to exceed this threshold in one cell type while that same interaction must have log *q*-value below the 50th percentile ($\log_{10} q = 0.33$) in the other two cell types. Similar thresholds were applied for contacts anchored in the newer CLOZUK risk loci (95th percentile $-\log_{10} q = 1.42$, 50th percentile $-\log_{10} q = 0.30$).

RNA-seq and transcriptome processing

A reference transcriptome, built from recently published RNA-seq datasets filtered through a stringent set of quality controls, included 47 NPC and 47 neuronal cultures from 22 independent donors including 11 subjects diagnosed with childhood onset schizophrenia (5 females, 6 males) (using unmodified DSM

criteria, with age of onset ranging from 4-12 years of age) and 11 controls (5 females 6 males) (Hoffman *et al.*, 2017). The donors range in age at biopsy was 8-30 years, as described (see Table 1 and Supplemental Data 1, 2 from reference (Hoffman *et al.*, 2017)). Using a cut-off of > 1 counts per million (CPM) in at least 30% of the experiments, the reference set included altogether 56,632 expressed sequences, including annotated genes, non-coding RNA, etc. Random distance-matched samplings (1,000x) were run to assess the significance of pair-wise gene correlation strength for each of the risk-associated chromosomal connectomes of the 3 cell types separately.

In addition, we generated RNA-seq libraries of glia from donor #2 and NPCs and excitatory neurons from both donors #1 and #2 that were used for the Hi-C studies (who are independent from the donors for the reference transcriptome described above). Libraries were prepared with the Kapa Total RNA library prep kit with ribo-depletion and strand-specific cDNA library construction (Kapa Biosystems). The libraries were then sequenced (125bp paired-end) on the Illumina HiSeq2000 platform (New York Genome Center). RNA-seq reads were aligned with STAR v2.4.0g1 (Dobin *et al.*, 2013) to GRCh37, after which uniquely mapped reads overlapping genes were counted with featureCounts v1.4.4 (Liao, Smyth and Shi, 2014) (with ENSEMBL v70 annotations). FPKM values were generated using Cufflinks (Trapnell *et al.*, 2012) and then compared by hierarchical clustering, multi-dimensional scaling and principal component analyses to 2082 RNA-seq data sets from stem cells and

their differentiated counterparts, and adult and fetal brain tissue, and blood as described in (Hoffman *et al.*, 2017). Genes with > 1 count per million in 10% of the samples in each data set were retained. This left 12,670 genes in common across all data sets. All expression values were converted to $\log_2$ RPKM, resulting in 12,670 genes common to all 2082 datasets. Of note, RNA-seq of subject S1 and S2 hiPSC-derived NPC clustered together with these previous NPC RNAseq and prenatal BRAINSPAN datasets from different individuals. However, S1 and S2 neuronal cultures (differentiated by NGN2 into glutamatergic neurons and also by neural differentiation medium and BDNF and GDNF growth factors into a more mixed population primarily comprised of forebrain-like neurons) clustered together with the previously generated RNA-seq from neuronal cultures (from a different set of individual) and in close proximity to pre- and postnatal BRAINSPAN brain tissue samples. Finally, our RNA-seq from the astrocyte-like glial cells (differentiated from NPC of Subject S1) did not co-cluster with either NPC or neurons, as expected. Thus, transcriptome mapping further confirms the specificity of our various cell culture systems.

Protein-protein interaction network analysis

To assess whether the proteins of the genes associated with schizophrenia risk loops are linked by physical functions or interactions, we input lists of genes from cell type-specific schizophrenia risk chromatin contacts for each cell type into the STRING v10.5 database (http://string-db.org) (Szklarczyk *et al.*, 2015, 2017). Default settings were used except for 1) “confidence” as the

meaning of network edges and 2) “high confidence (0.700)” for the minimum required interaction score. We noticed that in the NPC and glia high-confidence networks the significance could have been driven by numerous connections (“edges”) among the UGT gene cluster, all located within one 10kb bin on the genome. While continuing our analyses with the remaining highest confidence protein-protein associations, we re-ran the same gene lists, with all but one UGT gene (*UGTA10*) removed, at medium confidence to confirm significant networks (P_{NPC} = 0.0438; P_{Neuron} = 0.0876 ; P_{Glia} = 0.0134).

Human postmortem prefrontal cortex tissue

Postmortem brain tissues were derived from the University of Pennsylvania Brain bank, for which subjects with schizophrenia were prospectively accrued. Subjects with psychiatric illnesses were evaluated for Diagnostic and Statistical Manual of Mental Disorders Fourth Edition (DSM-IV) diagnostic criteria for SCZ. Diagnosis was determined by consensus of at least two board-certified research psychiatrists after comprehensive review of medical records, direct clinical assessments and interviews with caregivers. Autopsy consent was obtained from the next of kin or a legal guardian in all cases, based on a protocol approved by the Institutional Review Board at the University of Pennsylvania. Control and psychiatric subjects were matched for sex, age, smoking history and postmortem interval Table 3.22. All subjects with schizophrenia were treated with antipsychotics except 6 (24%) at the time of death.

Biochemical fractionation

Whole tissue homogenates were prepared from gray matter of the dorsolateral prefrontal cortex (DLPFC), BA 9 or 46, of SCZ and matched control subjects (Table 3.22) using a variation of the protocol we have developed and validated in human postmortem brain tissue. 50 mg grey matter was homogenized in .5 ml solution A (0.32 M sucrose, 1mM $MgCl^2$ and 0.1mM $CaCl^2$) with a Teflon pestle. ~45 µl of the homogenate (H) was saved, solubilized with 1% SDS and clarified by centrifugation.

Sample preparation and LC-SRM/MS

20 µg whole tissue homogenate preparations were mixed with the [$^{13}C_6$] brain ISTD at a ratio of 1:1 (µg/µg) and processed for LC-SRM/MS (liquid chromatography selected reaction monitoring mass spectrometry) analysis by on-gel trypsin digestion as described (MacDonald *et al.*, 2012). LC-SRM/MS analyses were conducted on a TSQ Vantage triple stage quadrupole mass spectrometer (ThermoFisher Scientific) with an Eksigent 2Dnano LC (Eksigent) and a CaptiveSpray source (Michrom). 5 µl (~2.5µg protein) sample was loaded on to a Magic C18 column (Michrom) at 1 µl/min for 12 min, and eluted at 750nl/min over a 25 min gradient from 3-35% mobile phase B (ACN containing .1% formic acid). SRM transitions were timed using 1 – 1.5 min retention windows, depending on the number of SRMs to be assayed. Transitions were monitored, allowing for a cycle time of 1 sec, resulting in a dynamic dwell time, never falling below 10 msec. The MS instrument parameters were as follows:

capillary temperature 275 oC, spray voltage 1100 V, and a collision gas of 1.4 mTorr (argon). The resolving power of the instrument was set to 0.7 Da (FWHM) for Q1 and Q3. Data were acquired using a chrom filter peak width of 4.0 sec.

Mass spectrometry data processing, informatics and statistics

Peak areas and area ratios were calculated within Pinpoint (Thermo Scientific). Raw files generated by LC-SRM/MS analysis were loaded into Pinpoint files containing target proteins/peptides/transitions. All individual SRM transitions and integration areas were manually inspected. Transitions for which the signal-to-noise ratio was below 3 were excluded from analysis. The ratios of the integrated areas for “light” endogenous peptides and “heavy” [$^{13}C_6$]brain ISTD peptides were calculated to obtain peptide measures using multiple transitions per peptide. Peptide measures from the same protein were averaged to obtain the measure for that protein. Prior to comparison of SCZ and control groups, protein measures were normalized in each fraction. Outlier subjects were identified by manual inspection for the number of individual protein or enrichment value outliers and principle component analysis of the entire dataset, run as a plug-in in Excel. Permutation analysis to determine significant coexpression of synaptic proteins of interest was performed by randomly sampling 4 proteins from the pool of 182 synaptic proteins and calculating the organization score for each set of 4.

CHAPTER IV: ATLAS: A DATABASE LINKING BINDING AFFINITIES WITH STRUCTURES FOR WILD-TYPE AND MUTANT TCR-PMHC COMPLEXES

Preface

This chapter comprises work published in PROTEINS: Structure, Function, and Bioinformatics by myself, Jennifer Cimons, Michael Cosiano, Michael Purcaro, Brian G. Pierce, Brian M. Baker, and Zhiping Weng. The publication reference is “ATLAS: A database linking binding affinities with structures for wild-type and mutant TCR-pMHC complexes” *PROTEINS: Structure, Function, and Bioinformatics*. Vol. 85 Issue 5 May 2017 (Borrman *et al.*, 2017)

Data curation was performed by myself, Jennifer Cimons and Michael Cosiano. Frontend and backend of the ATLAS website was developed by myself and Michael Purcaro. All data analysis and regression modeling was performed by myself. Research was supervised by Zhiping Weng, Brian G. Pierce and Brian M. Baker. The paper was written and figures were produced by myself, Brian G. Pierce, Brian M. Baker, and Zhiping Weng with contributions from all coauthors.

Abstract

The ATLAS (Altered TCR Ligand Affinities and Structures) database (https://zlab.umassmed.edu/atlas/web/) is a manually curated repository containing the binding affinities for wild-type and mutant T cell receptors (TCRs)

and their antigens, peptides presented by the major histocompatibility complex (pMHC). The database links experimentally measured binding affinities with the corresponding three dimensional (3D) structures for TCR-pMHC complexes. The user can browse and search affinities, structures, and experimental details for TCRs, peptides, and MHCs of interest. We expect this database to facilitate the development of next-generation protein design algorithms targeting TCR-pMHC interactions. ATLAS can be easily parsed using modeling software that builds protein structures for training and testing. As an example, we provide structural models for all mutant TCRs in ATLAS, built using the Rosetta program. Utilizing these structures, we report a correlation of 0.63 between experimentally measured changes in binding energies and our predicted changes.

Introduction

The binding of a T cell receptor (TCR) to an antigenic peptide presented by a major histocompatibility complex (pMHC) is a fundamental step in cell-mediated immunity. To eliminate pathogens and diseased cells, TCRs recognize foreign antigens displayed by MHC molecules on the surface of antigen presenting cells (Zinkernagel and Doherty, 1974; Babbitt *et al.*, 1985) (Figure 4.1, a). This recognition triggers T cell activation and an ensuing signaling cascade that leads to an antigen-directed cellular immune response.

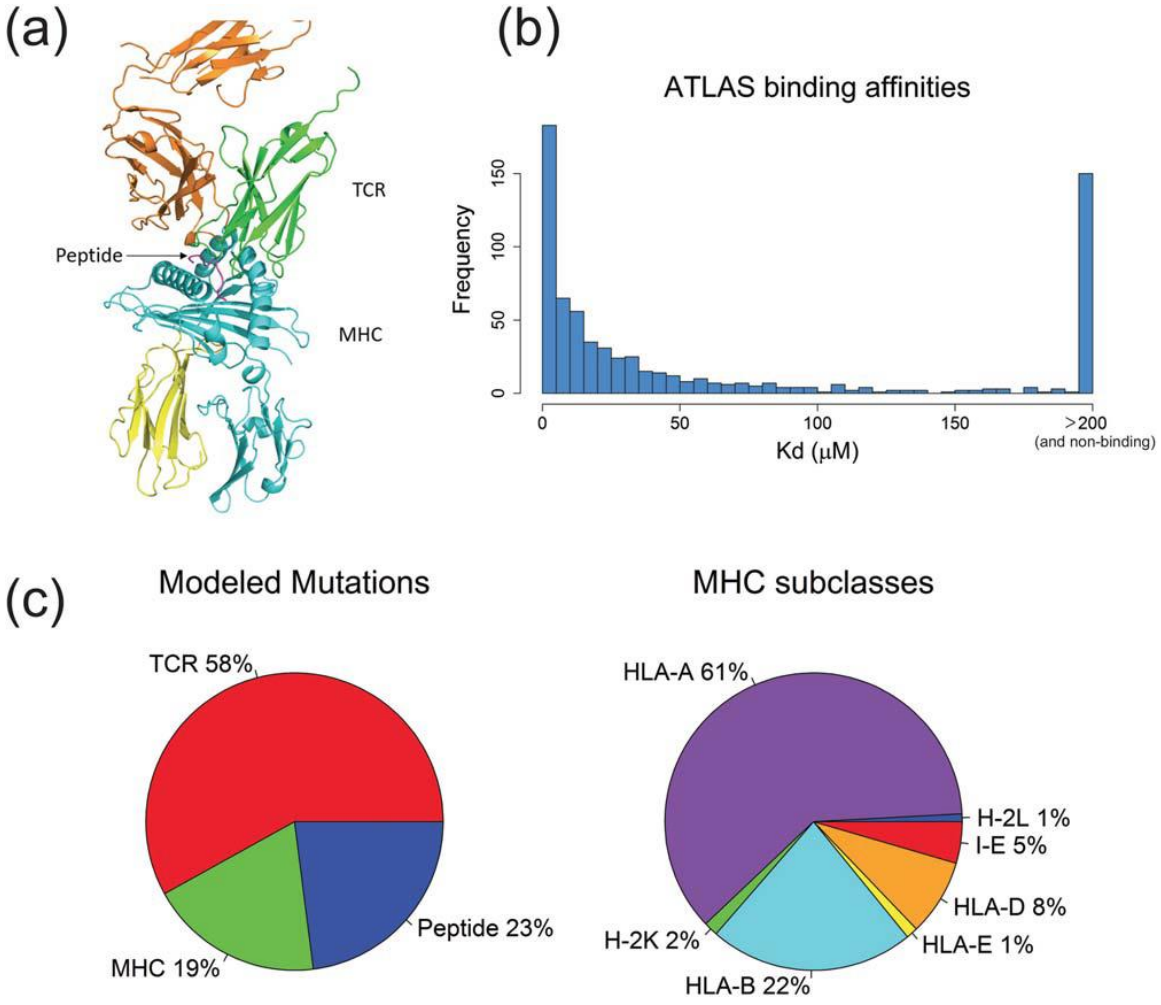


Figure 4.1 | ATLAS data statistics.
(**a**) A6/Tax/HLA-A*02:01 TCR-pMHC complex (PDB: 1AO7). TCR α and β chains (green, orange), MHC molecule (cyan), peptide (magenta), and β_2 microglobulin (yellow) are shown in cartoon style. (**b**) Histogram of the binding affinities of TCR-pMHC complexes in ATLAS. (**c**) Pie charts for percentage of entries with modeled mutations made to the TCR, the MHC or the peptide (left), percentage of MHC subclasses (right).

The ability to manipulate TCR-pMHC recognition has broad applications in a variety of biomedical arenas. One example is adoptive T cell transfer, which

uses tumor recognizing T cells to eradicate cancer cells (Restifo, Dudley and Rosenberg, 2012). The TCRs of these T cells can be genetically engineered to enhance their affinities toward specific tumor antigens (Varela-Rohena *et al.*, 2008; Linette *et al.*, 2013; Morgan *et al.*, 2013). In a related approach, high affinity TCRs can be used as soluble biologic therapeutics to target tumor-associated antigens (Oates and Jakobsen, 2013). Another example is peptide-based vaccination, which uses peptides to selectively stimulate T cells capable of battling infections or cancers (Purcell, McCluskey and Rossjohn, 2007). Many tumor associated peptides are derived from self proteins and are only weakly immunogenic because TCRs that strongly recognize self-antigens have been eliminated during negative selection in the thymus. Thus attempts have been made to develop modified peptides that can selectively enhance T cell activation (Chen *et al.*, 2005; Cole *et al.*, 2010).

TCR-pMHC binding strength is an important parameter in determining the quality of the ensuing immune response. TCR affinity has been shown to correlate positively with T cell activation (Holler and Kranz, 2003; Aleksic *et al.*, 2010); however, robust immune responses appear to result from TCR affinities in an optimal range, which is not necessarily high (Stone and Kranz, 2013). Past an apparent affinity threshold, the strength of the T cell response may plateau or attenuate.

Another consideration while striving for a desirable immune response is to avoid the cross-recognition of TCRs between foreign and self peptides, which

can lead to autoimmune diseases, such as multiple sclerosis, type 1 diabetes, and paraneoplastic syndromes (Wooldridge *et al.*, 2012; Sethi *et al.*, 2013). Before any potential therapeutic use of T cells, it is vital to identify off-target binding (Obenaus *et al.*, 2015); this consideration is particularly important for engineered TCRs, as demonstrated by adverse events in clinical trials (Linette *et al.*, 2013; Morgan *et al.*, 2013).

Precise prediction and manipulation of both TCR affinity and specificity is therefore essential for designing effective T-cell-based therapeutics. A number of methods have been developed for altering TCR-pMHC interactions, including in vitro molecular evolution and structure-guided protein design (Holler *et al.*, 2000; Li *et al.*, 2005; Varela-Rohena *et al.*, 2008; Borbulevych *et al.*, 2009; Zoete *et al.*, 2013; Malecek *et al.*, 2014; Pierce *et al.*, 2014). Structure guided design algorithms can alter affinity and specificity directly and efficiently, but are limited by the accuracy of their scoring functions.

Prediction of protein-protein binding affinity from protein-complex structure is a challenging problem. When nine unique scoring functions developed for docking programs or web servers were tested on a benchmark of 81 protein complexes, correlations between scores and binding affinities were low or nonexistent (r ranging from -0.18 to 0.32) (Kastritis and Bonvin, 2010). More recent studies utilizing supervised learning methods have increased correlations between predicted and experimental affinities, and there is still room for improvement (Vreven *et al.*, 2012; Xue *et al.*, 2016). Prediction of changes in

binding energy due to point mutations has seen greater success. Correlations between predicted and experimental $\Delta\Delta G$ in a study analyzing >1,500 point mutations ranged from 0.28 to 0.61 depending on the prediction method used (Geng, Vangone and Bonvin, 2016). Progress in $\Delta\Delta G$ prediction is critical to the field of TCR design where point mutants may be made to increase a TCR's affinity toward an antigen to trigger a robust immune response.

The improvement of TCR design algorithms requires access to both structural and binding data. We have built the ATLAS (Altered TCR Ligand Affinities and Structures) database (https://zlab.umassmed.edu/atlas/web/) to meet this demand. ATLAS links measurements of TCR affinity with structural information, and allows a user to query for a TCR, MHC, or peptide of interest. Results from such queries include details on affinity, mutation information, and structures of associated TCR-pMHC complexes that exist in the Protein Data Bank (Berman *et al.*, 2000). ATLAS includes structural and binding data for point-mutant TCRs that have been studied. If PDB structures for the relevant mutant complexes are not available, the database provides computationally modeled TCR-pMHC structures.

The immune epitope database (IEDB) (Vita *et al.*, 2015) and the Anti-Jen database (Toseland *et al.*, 2005) both provide binding affinities for TCR-pMHC complexes; however, these databases are peptide-epitope-centric and do not allow the user to query specific TCRs. Furthermore, there is no direct link between affinity and structural data in these databases. The IEDB does allow the

user to filter queries based on the availability of X-ray crystallography and surface plasmon resonance (SPR) experiments; however, in many cases a query using one peptide epitope will return multiple TCR-pMHC complexes that contain the peptide. Hence, to correctly match a TCR-pMHC complex with its reported binding affinity, the user needs to manually inspect the literature.

In comparison with IEDB and AntiJen, ATLAS allows the user to search specific TCRs, MHCs, and peptides. Full datasets in ATLAS can also be downloaded as flat files. With the goal of providing a repository to train and test next generation TCR design strategies and scoring functions, ATLAS also provides experimental details such as the resolutions of the structures and references for each of its entries. As low resolution structural data may skew scoring results, this information will be critical for the selection of a subset of the data to optimize prediction algorithms. As of this writing, the database includes data only for $\alpha\beta$ TCRs, but can be readily extended as more experimental data for the $\gamma\delta$ TCR family becomes available.

Results

The ATLAS database currently contains affinity and structural data on human and mouse TCR-pMHC complexes, with a total of 694 measured affinities ranging from high nanomolar affinities to low affinities with $K_D > 200$ μM (Figure 4.1, b, c). As more affinity data exist than structural data, ATLAS also contains a set of models for all TCR-pMHC affinity entries lacking crystal structures (see Materials and methods). Below, we first illustrate a usage case for the data in

ATLAS, and then we describe how to query the ATLAS database via the web interface.

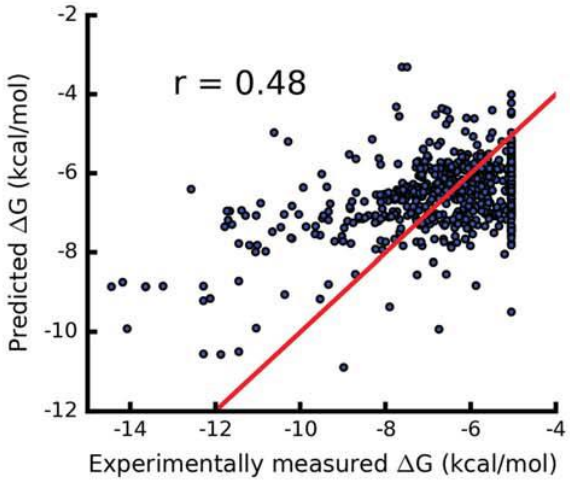
Using the data in ATLAS to develop energy functions

As a proof of concept, we performed multiple linear regressions using ATLAS data to develop scoring functions capable of affinity prediction. We examined two cases of energy prediction: prediction of TCR-pMHC binding energy, ΔG, and prediction of change in binding energy upon mutation, $\Delta\Delta G$ (see Materials and methods). Prior to regression analysis, no correlation was found between experimental and predicted ΔGs (or $\Delta\Delta G$s) using the Rosetta modeling software's standard scoring function (Leaver-Fay *et al.*, 2011). Eight energy terms from Rosetta—solvation, hydrogen bonding expressed in four terms, attractive and repulsive van der Waals, and a statistical pair potential—were used in the regression model (Figure 4.2, a).

(a)

Energy term	Coefficient	P-value
Lennard Jones van der Waals energy; attractive	2.16	10^{-41}
Lennard Jones van der Waals energy; repulsive	0.03	0.67
Lazaridis-Karplus solvation energy	1.67	10^{-23}
statistics-based pairwise energy	0.16	0.04
backbone-backbone hydrogen bonds; distant in primary sequence	-0.60	10^{-14}
backbone-backbone hydrogen bonds; close in primary sequence	-0.24	10^{-5}
sidechain-backbone hydrogen bonds	-0.31	10^{-5}
sidechain-sidechain hydrogen bonds	0.49	10^{-12}

(b)



(c)

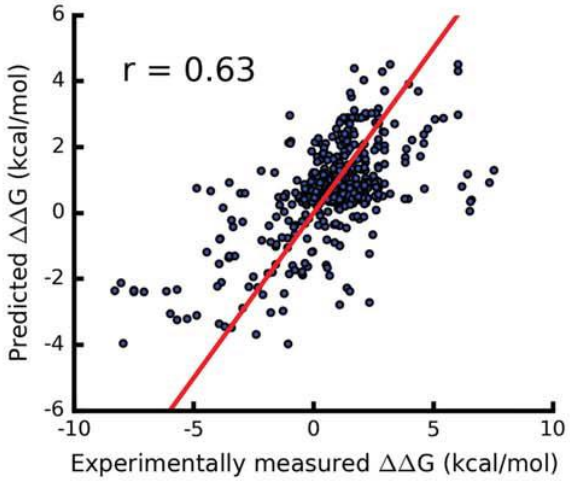


Figure 4.2 | Results of predicting binding free energies in ATLAS. (**a**) Table of coefficients and p-values for all energy features of the regression analysis. One insignificant feature (repulsive van der Waals) is highlighted in red. (**b**) Scatterplot of predicted ΔG versus ΔG determined by SPR for all ATLAS entries. Linear regression analysis was performed to predict ΔG using the following best set of features: attractive van der Waals energy, Lazaridis–Karplus solvation energy and all four hydrogen bond energy terms. r = 0.48 and RMSE = 1.48 kcal mol^{-1}. (**c**) Scatterplot of predicted $\Delta\Delta G$ versus $\Delta\Delta G$ determined by SPR for 575 mutant ATLAS entries determined by regression analysis using the following best set of features: attractive van der Waals energy, Lazaridis–Karplus solvation energy and side-chain-side-chain hydrogen bond energy. r = 0.63 and RMSE = 1.58 kcal mol^{-1}. Red line represents perfect prediction.

Performing leave-one-out cross-validation (LOO-CV) on all 694 ATLAS entries, we report a correlation of 0.65 between predicted ΔG and experimentally measured ΔG with a root mean square error (RMSE) of ~1.09 kcal mol^{-1}. However, many of the ATLAS entries differed by only a few residues and used the same PDB structure as the template for structure modeling. To accurately assess scoring function performance, we ran a cross-validation scheme where the training set for each prediction excluded any entries that used the same PDB structure as the template to model the structures. For example, many ATLAS entries were mutants of the A6-Tax/HLA-A2 TCR-pMHC complex and used the 1AO7 PDB structure as the template for modeling the mutant structures. For prediction of any of these mutants, we exclude from the training set all other mutants that were also modeled using 1AO7 as the template. Following this leave-one-complex-out cross-validation (LOCO-CV) scheme, we report a correlation r of 0.45 between experimental and predicted ΔGs and an RMSE of 1.52 kcal mol^{-1}.

The attractive van der Waals energy along with the solvation energy were the most important features for prediction, judged by the *P* values for these features (10^{-41} and 10^{-23}). We asked whether steric clashes in the modeled structures might have caused the poor performance of the repulsive van der Waals term. However, even after removing outlier entries with unfavorable repulsive van der Waals terms from the regression model, the coefficient for the repulsive van der Waals term remained insignificant. To extract the best predicting combination of features, we implemented LOCO-CV on all 255 combinations of the eight energy features. A slight increase in performance was seen when the van der Waals repulsive and pair potential terms were removed from the regression model, $r = 0.48$ and RMSE = 1.48 kcal mol^{-1} (Figure 4.2, b).

We then proceeded to build multiple linear regression models for prediction of $\Delta\Delta G$s. We included all multiple residue mutation cases in this study; however, 73% of the $\Delta\Delta G$ mutations were single residue mutations. Following the LOCO-CV scheme and analyzing all feature combinations, we report the maximal correlation of 0.63 between experimental and predicted $\Delta\Delta G$ with an RMSE of 1.58 kcal mol^{-1} (Figure 4.2, c). This best performing model used only the attractive van der Waals energy, solvation energy, and side-chain–side-chain hydrogen bond energy to predict $\Delta\Delta G$.

Given that the majority of structural data used for training in the regression model was designed via Rosetta, we assessed the accuracy of these modeled structures. Twenty-one out of the 694 ATLAS entries had both wildtype and

mutant crystal structures available (Table 4.1). Seven of these 21 entries were point mutations to residues with at least one side-chain dihedral angle (χ). The other 14 entries were either point mutations to Ala or were complexes designed with multiple mutations. We used these seven entries to evaluate the accuracy of the modeled mutant structures in ATLAS. The Dunbrack rotamer library was used to assess whether Rosetta designed mutant side chains had the same rotamers as those of the mutant crystal structures (Shapovalov and Dunbrack, 2011). Here, a side-chain conformation is defined as one of the combinations of multiple χ angle conformations for a side-chain. We found that four of the seven mutant side chains were of identical conformation to the side chain in the crystal structure. For two of the three side chains with incorrect conformations, only the terminal χ angle was inaccurate (χ^2 in Asp and χ^3 in Glu). Hence, only one side-chain, a mutation from Val to Arg, was completely mismodeled (Table 4.2). Although this is only a small sample of the entire dataset, it provides some evidence that the majority of the designed ATLAS mutations have correctly modeled side-chain conformations.

The modeled TCR-pMHC complexes from the regression analysis were generated using a fixed backbone approach. However, the complementarity determining region (CDR) loops of TCRs can change conformations upon binding with pMHC (Gagnon et al., 2006; J. B. Reiser et al., 2003; J. B. Reiser et al., 2002). To assess whether modeling flexibility of CDR loops could improve affinity prediction performance, we also generated another set of modeled

complexes via the Rosetta backrub application which accounted for flexibility of the CDR loops. These structures were then employed in our LOCO-CV scheme. Analyzing all features combinations, we saw a modest reduction in ΔG prediction performance, r = 0.45 and RMSE = 1.50 kcal mol^{-1} for the set of best performing features. A larger reduction in performance was found in prediction of $\Delta\Delta G$, r = 0.47 and RMSE = 1.80 kcal mol^{-1} (compared with r = 0.63 and RMSE = 1.58 kcal mol^{-1} for the fixed backbone approach) (Figure 4.3). This result was not entirely surprising as previous studies have reported poorer correlations when modeling backbone flexibility compared with using fixed backbone calculations (Mandell and Kortemme, 2009).

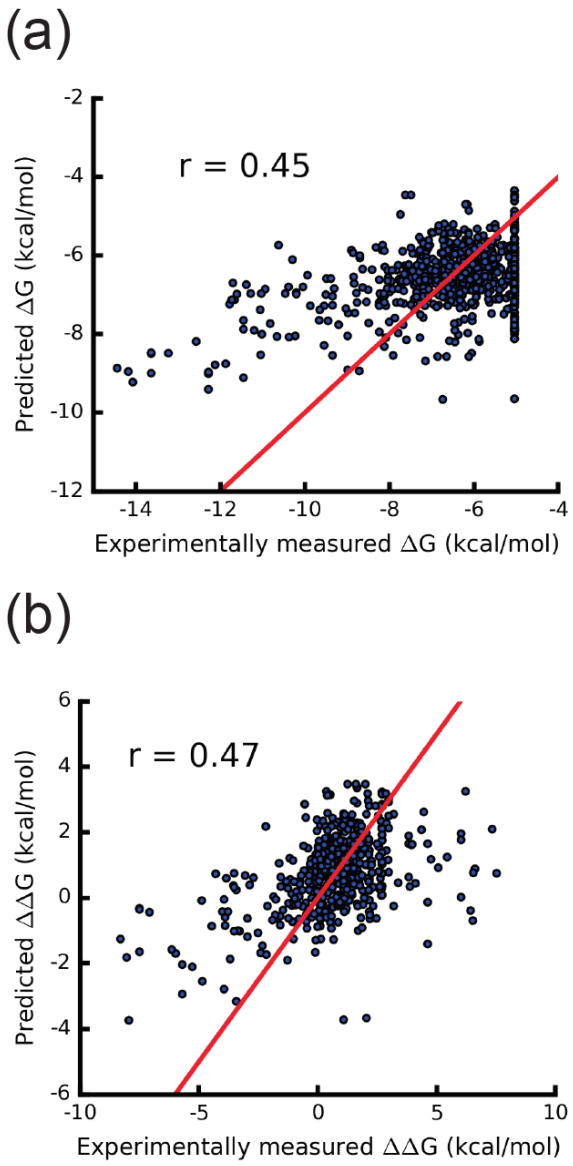


Figure 4.3 | Results of predicting binding free energies in ATLAS after modeling flexibility of CDR loops.
(**a**) Scatterplot of predicted Δ*G* versus Δ*G* determined by SPR for all ATLAS entries. Linear regression analysis was performed to predict Δ*G* using the following best set of features: attractive van der Waals energy, Lazaridis–Karplus solvation energy and backbone-backbone hydrogen bond energy close in primary sequence. r = 0.45 and RMSE = 1.50 kcal mol^{-1}. (**b**) Scatterplot of predicted ΔΔ*G* versus ΔΔ*G* determined by SPR for 575 mutant ATLAS entries determined by regression analysis using the following best set of features: attractive van der Waals energy, Lazaridis–Karplus solvation energy, statistics-

based pairwise energy, and repulsive van der Waals energy. r = 0.47 and RMSE = 1.80 kcal mol^{-1}. Red line represents perfect prediction.

Although there is room for improvement, TCR-pMHC binding affinity prediction is feasible through the use of the structural information in ATLAS. As a starting point for engineering TCRs, peptides, or MHCs to enhance binding affinity, the ATLAS database is a useful resource to guide the design process.

The web-based user interface of ATLAS

As many users may be only interested in a specific TCR or peptide, ATLAS provides a searchable interface so that the user can extract the relevant data of interest. To browse the entire ATLAS dataset, the user may simply submit a search leaving all fields with their default parameters.

As a demo, we queried for all entries that contain the human A6 TCR, the MHC allele HLA-A*02:01, and a peptide whose amino sequence contained LFGYPVY, with binding free energies lower than -6 kcal mol^{-1} (Figure 4.4, a). Note that the user may also search ATLAS by specifying *TRAV* or *TRBV* genes, as well as MHC allele or class. Submission of the search form brings the user to the results page (Figure 4.4, b). Each ATLAS entry (row) of the search results corresponds to a unique TCR-pMHC complex with an experimentally determined binding affinity and a 3D structure which can be used as a template for design. The binding affinity is reported in both K_D (μM) and ΔG (kcal mol^{-1}). The *PDB* column provides the PDB ID for a structure matching the TCR-pMHC complex with the reported experimental binding affinity. For many entries an exact

structure corresponding to the recorded binding affinity does not exist. To make use of such binding data, the *Template PDB* column refers to the PDB ID for a template structure for which a TCR-pMHC complex matching the reported binding affinity may be generated by modeling the mutations described in the *TCR mutation*, *MHC mutation*, and *Peptide mutation* columns of the entry. For further information on each entry, the results page provides a link to the abstract of the publication in which the binding affinity was determined in the *PMID* column. Lastly, the query results can be downloaded as individual files directly from the results page.

(a) Search ATLAS

Search through the ATLAS database by selecting a specific TCR, MHC allele, energy upper bound, or peptide sequence motif. The ATLAS search returns entries satisfying **all** selected search fields. To browse the entire dataset simply click 'Search' leaving all fields with their default parameters. For a more extensive description of search categories and criteria please see the help page.

1 TCR

Name: A6

TRAV: all

TRBV: all

2 MHC

Class: all

Allele: HLA-A*02:01

3 Energy

ΔG (kcal mol^{-1}) < -6.00

4 Peptide

Peptide Sequence: LFGYPVY

Search

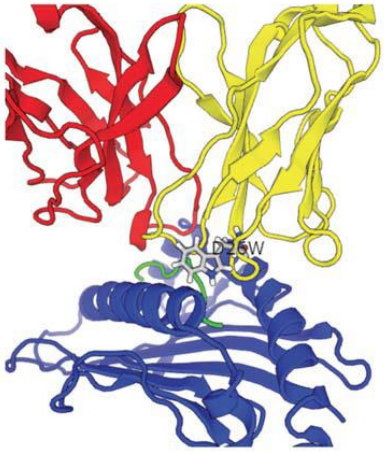
(b) Search Results

Show 10 entries

TCR name	TCR mutation	TCR mutation chain	MHC allele	MHC mutation	MHC mutation chain	Peptide	Peptide mutation	K_D (μM)	ΔG (kcal mol^{-1})	1 PDB	2 Template PDB	PMID
A6	WT	nan	HLA-A*02:01	WT	nan	LLFGYPVYV	WT	2.113502935	-7.74	1AO7		18767161
A6	D26M	A	HLA-A*02:01	WT	nan	LLFGYPVYV	WT	3.014285714	-7.53		1AO7	18767161
A6	D26V	A	HLA-A*02:01	WT	nan	LLFGYPVYV	WT	21.1	-6.38		1AO7	18767161
A6	D26W	A	HLA-A*02:01	WT	nan	LLFGYPVYV	WT	0.327868853	-8.84		1AO7	18767161
A6	R27F	A	HLA-A*02:01	WT	nan	LLFGYPVYV	WT	1.469489415	-7.95		1AO7	18767161

(c)

PV viewer



Download

Template PDB 1AO7:

Template

Designed model of 1AO7 with the following mutation(s):

TCR mutations: D26W

TCR mutation chains: A

MHC mutations: None

MHC mutation chains: None

Peptide mutations: None

Designed

Figure 4.4 | ATLAS web interface and data accession.
(**a**) The search page for querying the ATLAS database. ATLAS is searchable by TCR (1) and MHC (2) features, binding energies ∆G below a user specified kcal/mol (3) and by case-insensitive peptide sequence motifs (4). (**b**) The search results page linking binding energies to complex structures. The PDB structure specified in (1) refers to a structure identical to the TCR-pMHC used in the binding assay. The template PDB structure in (2) can be designed to replicate the TCR-pMHC used in the binding assay by modeling the mutations listed in the TCR mutation, MHC mutation, and Peptide mutation columns. (**c**) Protein Viewer and individual PDB downloads. An example shows the selection of PDB ID 1AO7 from the Template PDB column for the D26W TCR mutant. Template structures and Rosetta designed structures are both available for download.

Selecting a PDB ID from the results page brings the user to the PV Javascript Protein Viewer (Marco Biasini, 2015) and downloadable PDB content. Continuing the demo, we selected the template PDB 1AO7 for the D26W TCR mutant from our previous results page (Figure 4.4, c). The Rosetta modeled mutant structure is displayed in the PV viewer. The modeled mutant tryptophan side-chain of the TCR is highlighted in ball-and-stick style in the protein complex. From this page the user can download the individual template PDB complex along with the Rosetta-designed mutant PDB complex, both structures adjusted for consistency as described in the Materials and methods section.

Downloading data tables of ATLAS

The Downloads page provides the four tables used to build ATLAS in Microsoft Excel format. The *TCR gene table* contains the *TRAV* and *TRBV* genes for all TCRs in ATLAS. The *MHC class table* contains the classes for all MHC alleles. The *ATLAS table* provides all of the structural and affinity data for each ATLAS entry and is the extended version of tables found by browsing or

searching the web interface. Lastly, we also provide the set of consistency adjusted TCR-pMHC structures described earlier, *TCR-pMHC structures*, which contain the template PDB structures and the mutant structures. We provide two sets of mutant structures: (1) structures predicted using the fixed backbone approach and (2) structures predicted allowing flexibility in CDR loops via the Rosetta backrub application. All mutant structures are generated using the mutation information recorded in ATLAS.

Discussion

The multiple search parameters of ATLAS make it particularly useful for studying specific subsets of TCR or MHC. One recent application involved identifying TCRs that recognized the human Class I MHC allele HLA-A* 02:01 and TCRs that recognized human Class I MHC alleles that were not HLA-A*02:01. The links to PDB structures were used to make structural comparisons between the two groups of TCRs (Blevins *et al.*, 2016). Additionally, the option to search entries by *TRAV/TRBV* genes, MHC allele, and peptide sequence allow for comparisons to be made involving these parameters. For example, searching by a variable chain segment can allow the user to compare the effects that mutations within the shared chain have on binding affinity. The accumulation of affinity values for all published binding studies also allows for the identification of potentially important residues for pMHC recognition. Similarly, searching by peptide sequence can identify all TCRs known to recognize a particular peptide

(or a substring of residues in a peptide), as well as how mutations on the peptide impact TCR binding.

As a further example to demonstrate the utility of a large TCR database, ATLAS was recently used in a separate study to identify single point mutations when training a generalized approach for engineering TCRs (Riley *et al.*, 2016). In this study, nearly 200 point mutations in multiple HLA-A2 restricted TCRs (A6, B7, DMF5, and DMF4) and one HLA-B8 restricted TCR (LC13) were modeled using Rosetta and utilized in a multiple linear regression model. Using Rosetta energy terms and molecular dynamics derived flexibility terms as predictor variables and the experimental binding energies as the response variable, a score function was parameterized which emphasized van der Waals forces, solvation effects, and flexibility. This score function was rigorously cross-validated and found to estimate the effects of any given mutation relative to wild type with an average error of <1.5 kcal mol^{-1} and was used to identify additional affinity enhancing mutations in the B7, DMF5, and DMF4 TCRs.

We have developed the ATLAS database as a centralized resource to link structural and binding data for TCR-pMHC complexes, with an emphasis on the impacts of mutations within TCR-pMHC interfaces. The database can be queried multiple ways, and when structures do not exist, ATLAS provides modeled structures, as well as the means to visualize experimental or modeled structures. We anticipate that ATLAS will be useful in the design and optimization of TCRs, including the development of next-generation design algorithms for TCR-pMHC

interactions. It can also be used in combination with other large datasets of structural and affinity data, such as the AB-Bind database of antibody affinities (Sirin *et al.*, 2015), which would be useful in structure-based immune receptor design. Beyond this, the database may serve as a resource for studies aiming to correlate structural and biophysical binding data with immunological outcomes.

Wildtype PDB	Mutant PDB	TCR mutation	TCR mutation chain	MHC mutation	MHC mutation chain	Peptide mutation
2VLJ	2VLR	S99A	B	WT	nan	WT
3MV7	3MV9	Q55A	B	WT	nan	WT
3MV7	3MV8	Q55H	B	WT	nan	WT
1AO7	1QSE	WT	nan	WT	nan	V7R
1AO7	1QRN	WT	nan	WT	nan	P6A
1AO7	1QSF	WT	nan	WT	nan	Y8A
2AK4	3KXF	WT	nan	Q65A | T69A | Q155A	A | A | A	WT
1AO7	3QFJ	WT	nan	WT	nan	Y5F
3HG1	4JFF	D27F | R28L | I50T | S52R | N53E | Y71H | V93D | A94G | K96R | S97L | L44P | S50W | V51G | G52P | I53F | T100M | E102G | L103W | F104Q	A | A | A | A | A | A | A | A | A | B | B | B | B | B | B | B | B | B	WT	nan	WT
4JFF	4JFD	WT	nan	WT	nan	G4A
4JFF	4JFE	WT	nan	WT	nan	I7A
1AO7	4FTV	A99M | G100S | G101A | R102Q	B |B | B | B	WT	nan	WT
3VXR	3VXS	WT	nan	WT	nan	F6L
4PRI	4PRH	WT	nan	WT	nan	E5D
4G8G	4G9F	WT	nan	WT	nan	L6M
2F54	2F53	Q50P | S51F | S52W | G49S | A50V | I52M	A | A | A | B | B | B	WT	nan	WT
2BNR	2P5E	Q51T | S52P | S53W | T95L | S96L | G97D | S99T | G50A | A51I | G52Q | I53T | V95L	A | A | A | A | A | A | A | B | B | B | B | B	WT	nan	WT
2BNR	2P5W	Q51T | S52P | S53W | G50S | A51V | I53M | T95L | S96L | G97D | S99T | V95L	A | A | A | B | B | B | A | A | A | A | B	WT	nan	WT
3QDG	4L3E	D26Y | L98W	A | B	WT	nan	WT
3QIU	3QIW	WT	nan	WT	nan	K9E
4P23	4P46	Y31A	A	WT	nan	WT

Table 4.1 | ATLAS entries with both wildtype and mutant crystal structures available.

Designed PDB	Experimental PDB	Designed Rotamer	Experimental Rotamer	Designed χ Angles	Experimental χ Angles
3MV7_WT_nan_Q55H_B_WT.pdb	3MV8.pdb	2,6	2,6	175.53, -150.03	167.17, -143.99
1AO7_WT_nan_WT_nan_V7R.pdb	1QSE.pdb	2,2,2,3	3,1,2,2	-163.59, -173.07, -170.48, -86.61	-110.06, 46.25, -175.26, 162.23
1AO7_WT_nan_WT_nan_Y5F.pdb	3QFJ.pdb	3,1	3,1	-61.62, 97.64	-47.74, -77.15
3VXR_WT_nan_WT_nan_F6L.pdb	3VXS.pdb	3,2	3,2	-65.24, 166.16	-69.84, 160.56
4PRI_WT_nan_WT_nan_E5D.pdb	4PRH.pdb	3,1	3,5	-72.25, -15.16	-95.35, 115.36
4G8G_WT_nan_WT_nan_L6M.pdb	4G9F.pdb	3,2,1	3,2,1	-59.85, 169.33, 72.42	-63.51, 177.08, 68.47
3QIU_WT_nan_WT_nan_K9E.pdb	3QIW.pdb	3,2,5	3,2,3	-64.02, 165.74, -55.63	-42.88, -165.85, 60.11

Table 4.2 | Rotamer analysis for designed mutations.

Materials and methods

Data collection

To collect data suitable for training and testing TCR-pMHC scoring functions, we required all ATLAS entries to meet the following two criteria: (1) The affinity of the TCR-pMHC must be measured experimentally with purified proteins (most frequently) using SPR or isothermal titration calorimetry (ITC); and (2) The 3D structure of the complex has been determined experimentally, or for mutants, a template wild-type structure exists in the PDB. To provide the most comprehensive list of TCR-pMHC complexes, we did not make any quality restrictions pertaining to the affinity or structure data; instead, we recorded the resolution of crystallographic structures in the full dataset flat files available in the Downloads page.

To identify TCR-pMHC complexes for inclusion in ATLAS, we first found all crystallographic structures of TCR-pMHC complexes in the IMGT database (Ehrenmann, Kaas and Lefranc, 2010) verified by a careful inspection of the corresponding PDB entries. We next manually searched the literature for experiments measuring the affinity of each TCR-pMHC complex, including measurements with TCR or MHC mutants and/or peptide variants. If we could identify quantitative data on binding affinity, we then proceeded to include the TCR-pMHC complex in ATLAS, along with experimental details as metadata for the entry. We describe these metadata fields in the following sections.

Metadata fields

As many of the binding experiments recorded in ATLAS tested the effects of mutations on TCR-pMHC affinity, it follows that the majority of ATLAS entries pair mutant affinity data with wild-type template structures as opposed to the actual mutant TCR-pMHC complexes to which the binding affinities refer. Thus we recorded detailed information on the mutations such that one could build a 3D structure model of the mutant complex that corresponded to the affinity data, given the template structure. With this application in mind, the data tables of ATLAS are designed not only for information, but also for easy parsing by protein design software. The following five fields represent the information required for modeling a mutant TCR-pMHC complex structure from a template structure.

- *TCR mutation*: <wild-type residue><residue number><mutant residue>
- *TCR mutation chain*: <chain>; A for α or B for β chain
- *MHC mutation*: <wild-type residue><residue number><mutant residue>
- *MHC mutation chain*: <chain>; A for α chain B for β chain
- *Peptide mutation*: <wild-type residue><residue number><mutant residue>

Some fields may be left empty if one or more molecules in the template structure are the same as the molecules used to measure affinity. In the case of a complete match between the complex structure and the complex tested for binding, all mutation fields may be left empty. As mentioned previously, all entries are required to have at least a template structure and in some case both experimental and template structures for mutants exist. These cases could be

particularly helpful in assessing the accuracy of the structural modeling.

Inconsistent chain naming, structure boundaries, and prevalence of water molecules complicate the design and scoring process. To overcome these challenges, we also supply a set of files for all experimentally determined TCR pMHC structures in ATLAS with the following consistency adjustments: renaming of chains, truncation of chains to the binding domains, and removal of water molecules. When there were multiple complexes in the asymmetric unit of a crystal structure, the first complex was selected.

Protein modeling

As a proof of principle, we wrote a script to parse ATLAS and build models using Rosetta for all listed mutations upon the adjusted template structures. These models were built using the fixed backbone design option of Rosetta, fixbb, and are available for download. The example parser script, "build_models.py," is also available at the Github site of ATLAS (https://github.com/weng-lab/ATLAS). Although this example is specific for design via the Rosetta protein modeling software, the ATLAS database can be easily integrated with any other design software.

Regression analysis

Many ATLAS entries have low affinities with unreliable ΔGs or undetectable binding. Hence, we assigned all entries with a ΔG > -5.05 kcal mol^{-1} (K_D > 200 μM) and all non-binding entries to have ΔG = -5.05 for the regression analysis.

The following equation was used to calculate ΔG for each independent

variable in the regression model:

$$\Delta G = \Delta G_{COMPLEX} - (\Delta G_{TCR} + \Delta G_{pMHC}), \qquad (1)$$

where $\Delta G_{COMPLEX}$ is the energy of a feature in the regression model for the TCR-pMHC complex and ΔG_{TCR} and ΔG_{pMHC} are the energies of a feature for the isolated TCR structure and isolated pMHC structure, respectively.

Similarly, we use the following equation to define $\Delta\Delta G$ for changes in energy upon mutation:

$$\Delta\Delta G = \Delta G_{MUT} - \Delta G_{WT}, \qquad (2)$$

where ΔG_{WT} is the ΔG for a wild type TCR-pMHC entry and ΔG_{MUT} is the ΔG for a mutated version of the wild type (mutations may involve multiple residues).

Coefficients for the regression models were estimated by the ordinary least squares method. P-values were calculated from the *t* statistics of the coefficients using a two tailed *t* test. All regression calculations were made using the python statistics module statsmodels.

Architecture

The backend of ATLAS was built using the archetypal web service solution stack, LAMP, consisting of Linux (Ubuntu version 14.04), Apache (version 2.4.7), MySQL (version 5.5.41) and PHP (version 5.5.9). The front end was designed using the Bootstrap framework (version 3.3.5). All programs related to ATLAS are available at github (https://github.com/weng-lab/ATLAS).

CHAPTER V: HIGH-THROUGHPUT MODELING AND SCORING OF TCR-PMHC COMPLEXES TO PREDICT CROSS-REACTIVE PEPTIDES

Preface

This chapter is adapted from a manuscript currently under review at Bioinformatics authored by myself, Brian G. Pierce, Thom Vreven, Brian Baker, and Zhiping Weng titled: High-throughput modeling and scoring of TCR-pMHC complexes to predict cross-reactive peptides.

Peptide sequence extraction and filtering, modeling of TCR-pMHC complexes, computational scoring, and prediction performance analysis was performed by myself. Original project hypothesis was conceived by Brian G. Pierce. Research was supervised by Thom Vreven, Zhiping Weng and Brian M. Baker. The paper was written and figures were produced by myself, Brian G. Pierce, Thom Vreven, Brian M. Baker, and Zhiping Weng.

Abstract

The binding of T cell receptors (TCRs) to their target peptide MHC (pMHC) ligands initializes the cell-mediated immune response. In autoimmune diseases such as multiple sclerosis, the TCR erroneously recognizes self-peptides as foreign and activates an immune response against healthy cells. Such responses can be triggered by cross-recognition of the autoreactive TCR with foreign peptides. Hence, it would be desirable to identify such foreign-

antigen triggers to provide a mechanistic understanding of autoimmune diseases. However, the large sequence space of foreign antigens presents an obstacle in the identification of cross-reactive peptides.

Here, we present an *in silico* modeling and scoring method which exploits the structural properties of TCR-pMHC complexes to predict the binding of cross-reactive peptides. We analyzed three mouse TCRs and one human TCR isolated from a patient with multiple sclerosis. Cross-reactive peptides for these TCRs were previously identified via yeast display coupled with deep sequencing, providing a robust dataset for evaluating our method. Our method accurately selected the top binding peptides from sets containing more than a hundred thousand unique peptides.

Introduction

As a surveillance mechanism against pathogens and cancer, T cells of the host immune system use their αβ T cell receptors (TCRs) to inspect other cells. Targets recognized by TCRs are peptides bound and presented by the host major histocompatibility complex (MHC) proteins on the outer surface of the cellular membrane, and the peptide epitope may derive, for instance, from a viral protein. TCR recognition triggers complex signaling pathways that lead to a variety of outcomes, such as the destruction of infected or diseased cells, T cell proliferation, and release of pro-immune cytokines.

Determining peptide epitopes that can be recognized by TCRs is of considerable interest, impacting fields ranging from virology to cancer immunotherapy. Peptide immunogenicity involves three steps, each of which have been addressed via various predictive algorithms: peptide processing (Bhasin and Raghava, 2004; Nielsen *et al.*, 2005), peptide binding to an MHC (Andreatta and Nielsen, 2015; Jurtz *et al.*, 2017; O'Donnell *et al.*, 2018), and TCR recognition of the peptide-MHC (pMHC) complex (Tung *et al.*, 2011; Pierce and Weng, 2013; Schneidman-Duhovny *et al.*, 2018; Lanzarotti, Marcatili and Nielsen, 2019; Ogishi and Yotsuyanagi, 2019; Riley *et al.*, 2019). While progress has been made in predicting the outcome of each step, the fixed size of the TCR repertoire relative to the much larger number of possible peptide epitopes that T cells may encounter presents a particularly significant challenge.

Even with a TCR repertoire estimated to lie in the tens of millions, estimates are that any particular TCR would need to recognize at least one million different pMHC complexes in order to provide sufficient immune coverage (Mason, 1998; Sewell, 2012). This high level of cross-reactivity has been verified using combinatorial peptide libraries (Maynard *et al.*, 2005; Wooldridge *et al.*, 2012). Thus, although specificity is considered a hallmark of immunity, TCRs display significant cross-reactivity. Even if such cross-reactivity can be rationalized at a high level from structural and biophysical principles (Singh *et al.*, 2017), determining the range of peptides recognized by a specific TCR remains a major goal in immunology. Demonstrating the biological significance of the

problem, TCR cross-recognition of self-peptides is believed to underlie various autoimmune disorders (Gravano and Hoyer, 2013), and patient deaths have occurred due to unanticipated “off-target” recognition of TCRs used in clinical trials for cancer immunotherapy (Linette *et al.*, 2013; Morgan *et al.*, 2013).

Given the availability of TCR-pMHC structural information, together with advances in protein design and prediction methodologies, in principle, the peptide specificity profile of a TCR should be predictable using *in silico* methods. One challenge, however, is the availability of detailed experimental datasets against which such prediction methods could be benchmarked. In addition to combinatorial peptide libraries, Garcia and colleagues have used yeast display of pMHC libraries coupled with TCR staining and deep sequencing to assess the specificity profiles of TCRs (Birnbaum *et al.*, 2014; Adams *et al.*, 2016; Gee *et al.*, 2018). With each yeast cell expressing a unique random peptide, these libraries allow for affinity-based interrogation of over one hundred million peptides against a query TCR. Affinity based selection proceeds through multiple rounds where yeast libraries are enriched for yeast that bound bead-multimerized TCR. Subsequent deep sequencing of yeast DNA from final selection rounds produces enrichment counts for peptides selected by the query TCR. Such experiments provide rich datasets for developing and benchmarking *in silico* approaches to evaluating TCR specificity.

Here, we used structure-based *in silico* methods to predict the specificity profiles for four TCRs assessed using yeast display and deep sequencing: 2B4,

226, 5cc7, and Ob.1A12 (Birnbaum *et al.*, 2014). Three of these TCRs recognize a peptide derived from moth cytochrome C presented by the murine class II MHC protein I-E^k (Newell *et al.*, 2011). The fourth (Ob.1A12) was isolated from a patient with relapsing-remitting multiple sclerosis and recognizes a peptide derived from the myelin basic protein presented by the human class II MHC protein HLA-DR2) (Wucherpfennig *et al.*, 1994). The deep sequencing data provided more than 100,000 peptides for each TCR, including both binders and non-binders, ideal for benchmarking structure-based *in silico* methods.

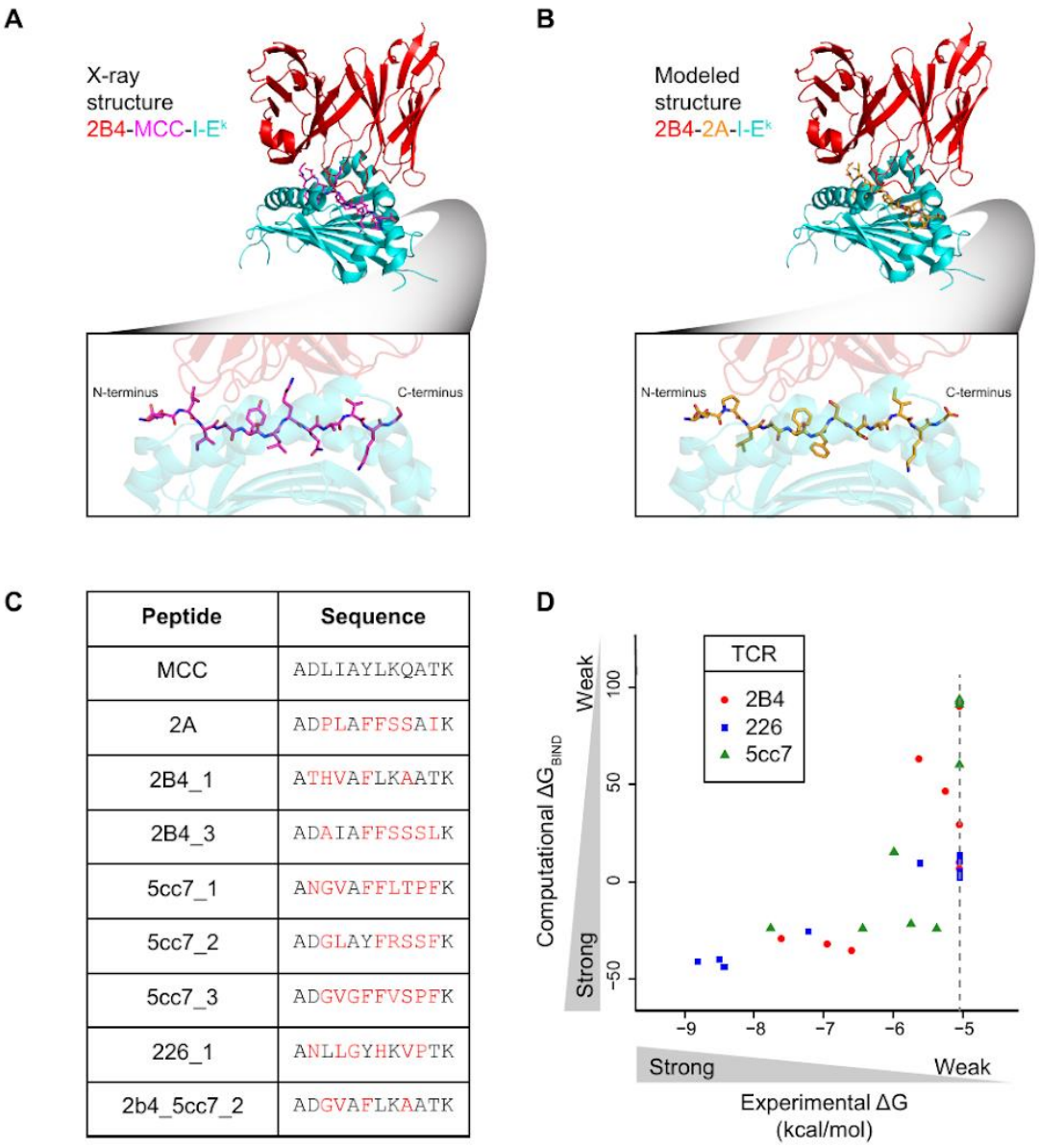
Using the crystal structures for the four TCR-pMHC complexes (Hahn *et al.*, 2005; Newell *et al.*, 2011; Birnbaum *et al.*, 2014), we modeled all of the query peptides within the TCR-pMHC complexes and scored the structural models to predict cross-reactive peptides for each of the four TCRs. Our modeling and scoring approach was capable of recovering cross-reactive peptides from large pools of primarily non-binding peptides for each TCR tested. We further show that our method outperforms the strategy of selecting peptides closest in sequence to each TCR's cognate peptide epitope (i.e., the target peptide found in crystallographic structure), underscoring the value of including structural information in epitope prediction.

Results

High-throughput modeling reproduces experimentally observed enrichment of binder peptides

Previously described experimental yeast display and deep sequencing generated libraries that were enriched for peptides specifically recognized by four TCRs (Birnbaum *et al.*, 2014). Beginning with the crystallographic structures of the 2B4, 226, 5cc7, and Ob.1A12 TCRs in complex with their cognate pMHC complexes, we computationally modeled and scored the peptides with sequences in the preselection libraries and four sequential selection rounds—347,210 peptides for 2B4, 811,481 peptides for 226, 809,156 peptides for 5cc7, and 514,906 peptides for Ob.1A12, and 2,482,753 peptides in total (Birnbaum *et al.*, 2014).

To increase the computational throughput in modeling the structures of these approximately 2.5 million peptides, we performed a restricted structural modeling procedure using Rosetta's fixed backbone design application, fixbb, which optimizes side-chain conformations on a fixed backbone using the Rosetta energy function (Leaver-Fay, Snoeyink and Kuhlman, 2008). We retained TCR and MHC side chains in the conformations adopted in the crystallographic structures with their cognate pMHCs (Figure 5.1, A and B).



Peptide	Sequence
MCC	ADLIAYLKQATK
2A	ADPLAFFSSAIK
2B4_1	ATHVAFLKAATK
2B4_3	ADAIAFFSSSLK
5cc7_1	ANGVAFFLTPFK
5cc7_2	ADGLAYFRSSFK
5cc7_3	ADGVGFFVSPFK
226_1	ANLLGYHKVPTK
2b4_5cc7_2	ADGVAFLKAATK

Figure 5.1 | Prediction of TCR-pMHC binding free energies.
(**A**) Crystal structure of TCR-pMHC interface for the 2B4 TCR (red) interacting with the MCC peptide (magenta) displayed by the I-E^k MHC (cyan). Bottom box: profile of the MCC peptide. (**B**) Fixed backbone model structure of TCR-pMHC interface for the 2B4 TCR (red) interacting with the 2A peptide (yellow) displayed by the I-E^k MHC (cyan). 2A peptide was modeled onto the backbone of the MCC peptide in (A), TCR and MHC protein structures remain identical to (A). Bottom box: profile of the 2A peptide. (**C**) Table of peptide names and amino acid sequences. Amino acids in red differ from the MCC peptide at the corresponding

position. (**D**) Scatter plot of ΔG_{BIND} from computational modeling and scoring versus ΔG from experimental binding energies determined by SPR from Birnbaum et al. for peptides in Table (C). Pearson correlations are 0.73, 0.98 and 0.70 for TCRs 2B4, 226, and 5cc7, respectively. Correlation for the entire set of peptides = 0.70. Peptides with unreliable K_Ds due to weak or non-binding interactions were assigned a K_D of 200 μM (ΔG = -5.05 kcal/mol, gray dotted line). Individual TCR correlations were similar if we assigned a ΔG of 0 kcal/mol for these weak/non-binding peptides and the correlation for the entire set remained the same (r = 0.70).

Once each TCR-pMHC model was generated, we scored the full complex ($G_{COMPLEX}$) and isolated components ($G_{TCR/MHC}$, $G_{PEPTIDE}$) using Rosetta's score application (see Materials and methods). These scores were combined to produce a binding score, ΔG_{BIND}, which accounted for the peptide's interaction energy with both the MHC and the TCR.

To quantitatively assess our peptide modeling and scoring approach, we examined sets of peptides for which experimental binding free energies were available for the 2B4, 226, and 5cc7 TCRs (Figure 5.1, C) (Birnbaum *et al.*, 2014). Correlations between ΔG_{BIND} and experimentally measured binding free energies were greater than 0.69 for all TCRs, and 0.70 for the entire set together (Figure 5.1, D).

We next examined the distributions of ΔG_{BIND} across the four experimental selection rounds of the yeast display library where each successive round was further enriched in cross-reactive peptides via TCR selection. Indeed, ΔG_{BIND} scoring of modeled complexes revealed an increasing enrichment of favorable energy scores for peptides in each subsequent selection round, in congruence with the subsequent enrichment of cross-reactive peptides for each round (Figure

5.2). Thus, relying on a relatively simple structural modeling method to enable computational throughput permits the recovery of experimentally determined peptides bound by TCR.

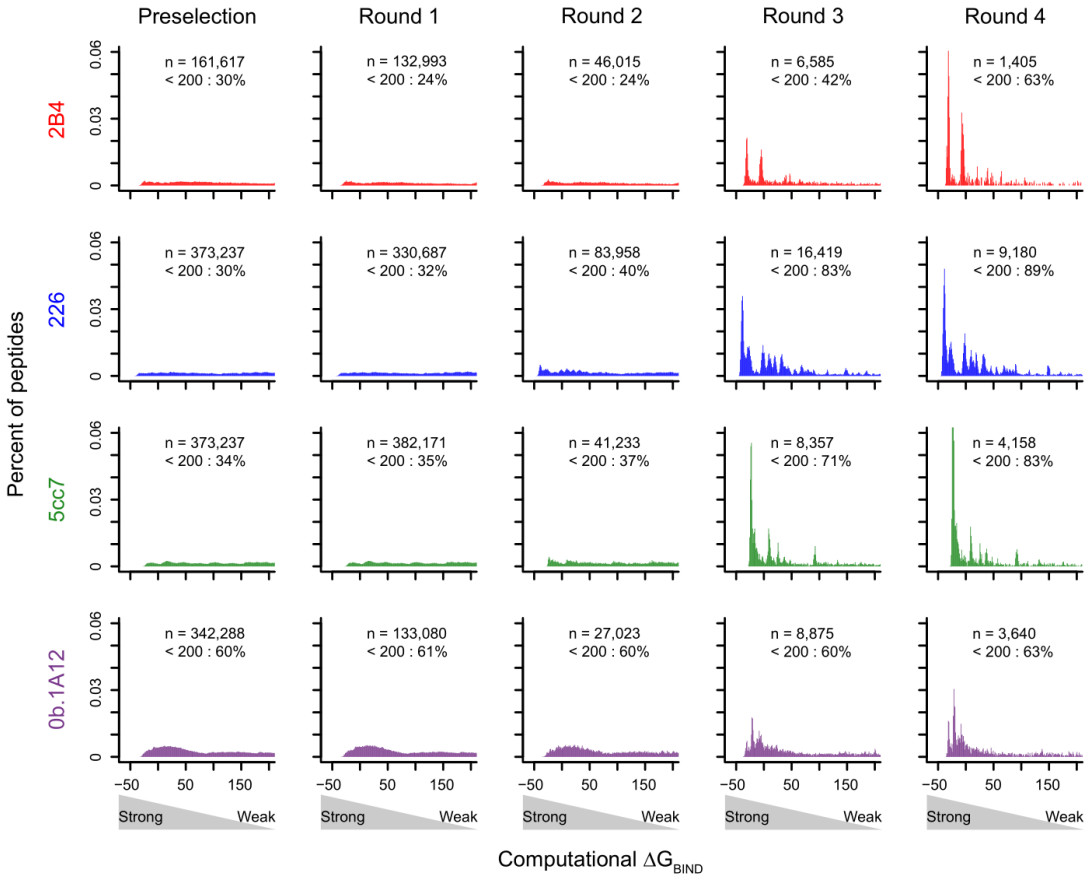


Figure 5.2 | Distributions of ΔG_{BIND} for peptides recovered from different selection rounds.
We generated structural models of TCR-pMHC complexes using peptide sequences from all experimental selection libraries. For each unique peptide recovered in each selection round, we modeled its structure bound to MHC and TCR and computed ΔG_{BIND} for the TCR-pMHC complex. The probability densities for ΔG_{BIND} are plotted for each round of selection for the four TCRs analyzed in this study, 2B4, 226, 5cc7 and Ob.1A12. The probability density is defined such that the histogram has a total area of one. (n = total number of unique peptides in

the given round; < 200: percent of peptides in the round with ΔG_{BIND} less than 200).

Footnote: *The deep sequencing results present only a sample of the total unique peptides in each selection round. It is thus important to note, the Round 4 peptides while a subset of the full experimental pre-selection library with >10^8 peptides, is not a strict subset of the >10^5 unique peptides recovered from sequencing the pre-selection library.*

Correlation between top computationally selected peptides and top experimentally selected peptides

To examine the extent to which our modeling and scoring approach selected the peptides recognized by the TCRs with the strongest affinities, we compared the 50 peptides with the most favorable ΔG_{BIND} to the 50 peptides with the most reads recovered by deep sequencing after the fourth round of selection for the 2B4, 226, 5cc7, and Ob.1A12 TCRs (Birnbaum et al., 2014). Rather than simply selecting from the round-four peptides, which are highly enriched in binders, we asked the computational method to identify top-scoring peptides from the pool of unique peptides in the union of the preselection and the round four libraries, providing for a more unbiased test. We note, while the full preselection pool contains all the round four peptides, the deep sequencing data present only a sample of the preselection pool peptides. Hence, we use the union of both the preselection and round four sequencing to ensure the true binders are accounted for. For a successful computational method, we would expect peptides with the most favorable ΔG_{BIND} to be members of the smaller, round-four peptide sets and to share amino acid preferences with the peptides that have the most abundant

reads in the fourth round. The amino acid preference generated using the 50 top experimentally selected peptides (50 peptides with the most abundant reads counts in round four) illustrated binding motifs distinct for each TCR (Figure 5.3, A). Among the 50 peptides that had the most favorable ΔG_{BIND} according to our modeling and scoring method, 41, 45, 24, and 38 were among the peptides in the round-four library for the 2B4, 226, 5cc7, and Ob.1A12 TCRs, respectively. Therefore, our scoring method was capable of identifying true binders within a large pool consisting primarily of non-binder peptides.

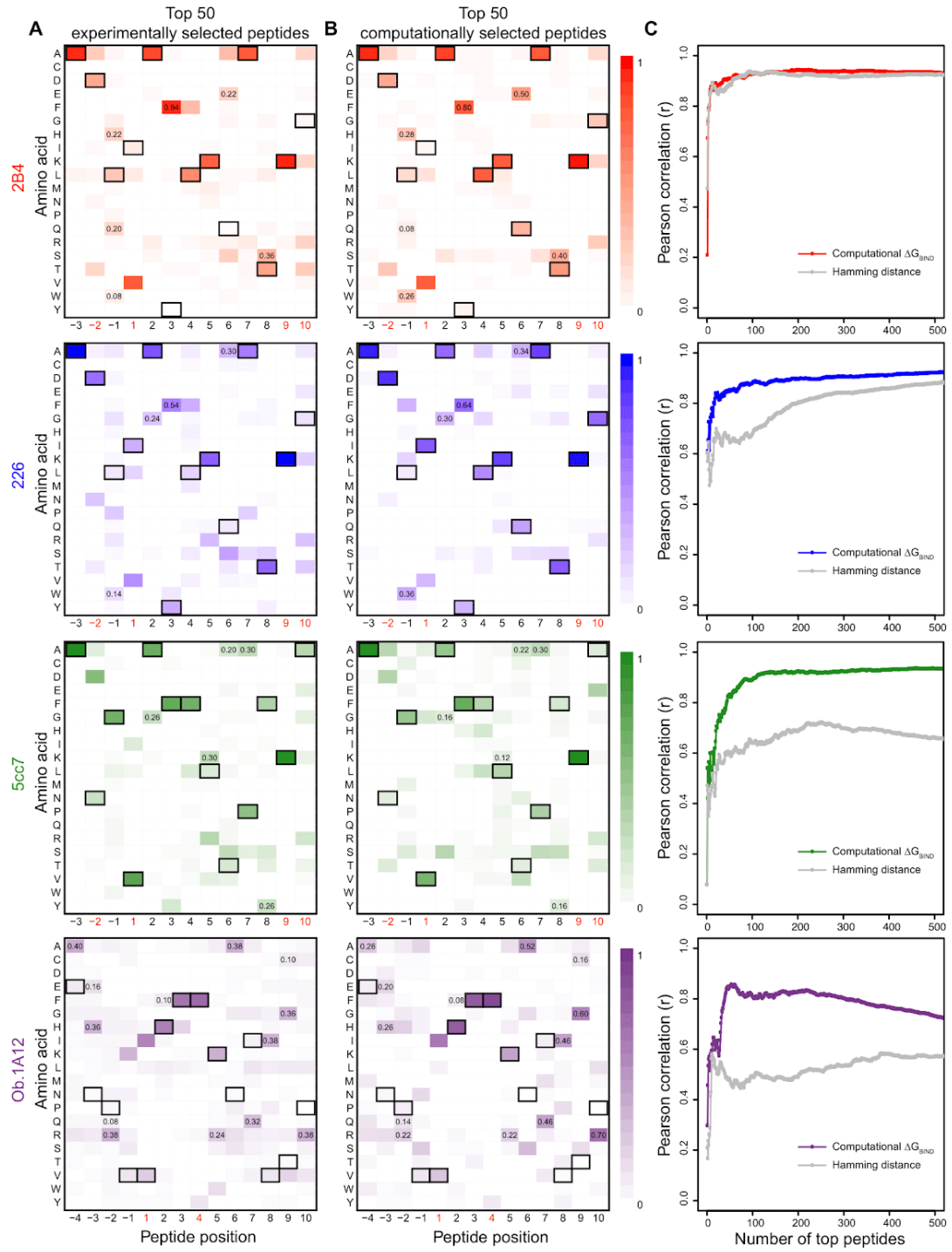

A
Top 50
experimentally selected peptides
B
Top 50
computationally selected peptides
C
2B4
226
5cc7
Ob.1A12
Amino acid
Peptide position
Pearson correlation (r)
Number of top peptides
Computational ΔG_BIND
Hamming distance

Figure 5.3 | Amino acid frequencies for top peptides selected by yeast display or by computation for mouse and human TCRs.
(**A**) Heatmaps represent the amino acid frequencies at peptide positions for the 50 peptides with the most abundant reads in the fourth round of selection for four TCRs (2B4, 226, 5cc7, and Ob.1A12). (**B**) Amino acid frequencies at peptide positions for the 50 peptides with the most favorable ΔG_{BIND}. Peptide pool for ΔG_{BIND} computation was the union of the preselection library and round 4 library ($>10^5$ peptides). The peptide residues from the template TCR-pMHC structures used for modeling, (2B4, 226) MCC, (5cc7) 5c1, and (Ob.1A12) MBP are outlined in black. Peptide positions restricted in the yeast display libraries to maintain MHC binding are marked in red beneath heatmap. Correlations of frequencies between the experimental and computational heatmap for 2B4, 226, 5cc7, and Ob.1A12 TCRs are 0.91, 0.85, 0.83, and 0.85 respectively (excluding restricted positions). (**C**) Pearson correlations between the experimental and computational ΔG_{BIND} heatmaps as a function of the number of top selected peptides used to generate the heatmaps for each TCR (red, blue, green, purple). Pearson correlations between the experimental heatmap and heatmap generated from peptides with lowest hamming distance to wildtype peptide for each TCR (2B4-MCC, 226-MCC, 5cc7-MCC, Ob.1A12-MBP) as a function of the number of top selected peptides used to generate the heatmaps (gray).

To further examine how the best scoring peptides compared to those identified experimentally after TCR selection, we compared heatmaps of amino acid preferences for the 50 top experimentally selected peptides and the top 50 computationally selected peptides (50 peptides with the most favorable ΔG_{BIND} from the union of preselection and round 4) for each TCR (Figure 5.3, A and B). Many sequence features were shared between the top-scoring peptides and the peptides with the most abundant read counts. To quantify similarity between heatmaps, we flattened the heatmap matrices into vectors and calculated the Pearson correlation between them. Excluding those anchor positions restricted in the libraries for MHC binding, correlations between the heatmaps representing

strongly selected and top-scoring peptides were 0.91, 0.85, 0.83, and 0.85 for TCRs 2B4, 226, 5cc7, and Ob.1A12, respectively.

We also assessed a different binding score that only incorporated interaction energies between TCR and pMHC. A comparison of the different binding scores revealed the inclusion of interaction energies between peptide and MHC as in ΔG_{BIND} was critical to the success of cross-reactivity predictions (Figure 5.4). A possible reason is that some of the peptides in the preselection library may not bind stably to the MHC. Because a stable peptide-MHC interaction is a prerequisite for TCR binding, the incorporation of interaction energies between peptide and MHC improved the correlation between prediction and experiment. We also tested the $G_{COMPLEX}$ score by itself. While ΔG_{BIND} represents interactions between the peptide and TCR and also the peptide and MHC, $G_{COMPLEX}$ additionally accounts for deformations of the peptide itself. However, predictive performance of $G_{COMPLEX}$ was worse than that of ΔG_{BIND} with correlations between the heatmaps representing strongly selected and top-scoring peptides being 0.81, 0.85, 0.80, and 0.26 for TCRs 2B4, 226, 5cc7, and Ob.1A12, respectively.

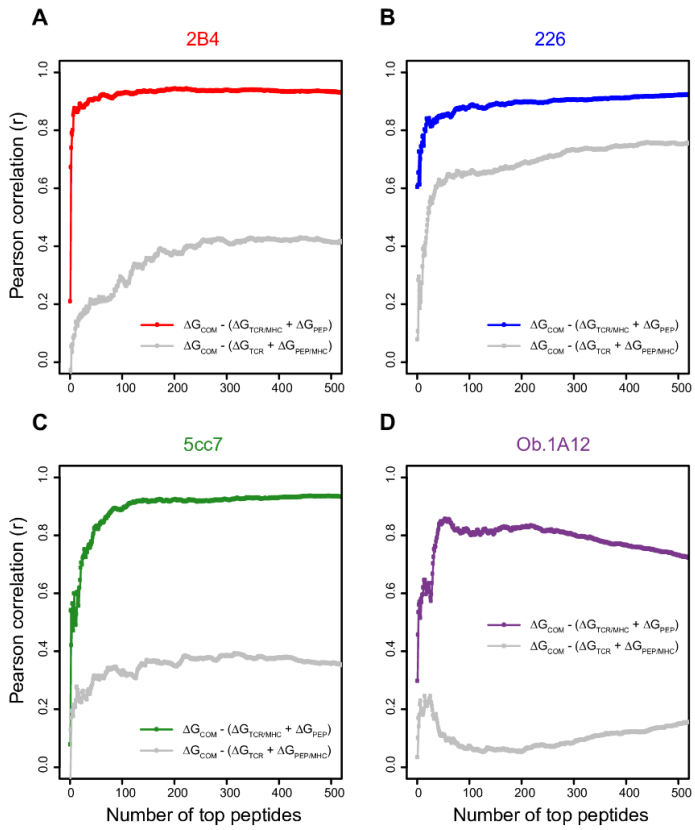


Figure 5.4 | Prediction comparison of scoring approaches.
Pearson correlations between the experimental and computational ΔG_{BIND} heatmaps of cross-reactivity profiles (See Figure 5.3) as a function of the number of top selected peptides used to generate the heatmaps for each TCR (A, B, C, D). Correlations for scoring approach such that

$\Delta G_{BIND} = G_{COMPLEX} - (G_{TCR/MHC} + G_{PEPTIDE})$

is shown in red, blue, green and purple lines. Correlations for scoring approach such that

$\Delta G_{BIND} = G_{COMPLEX} - (G_{TCR} + G_{pMHC})$

is shown in gray.

We note that in the case of Ob.1A12, our scoring method was successful in assigning favorable scores to peptides carrying the ‘HF’ motif as was found experimentally (Birnbaum *et al.*, 2014). It is evident from the experimental heatmap that Ob.1A12 is tolerant of amino acid substitutions outside the anchor residues and the central HF motif. This feature of Ob.1A12 is also captured by our modeling and scoring method, with the exception of a strong preference for Lys at position -1 in the experimental heatmap and a few other less frequent substitutions our method could not reproduce.

The 2B4, 226 and 5cc7 TCRs all recognize the MCC peptide (ADLIAYLKQATKG), which is presented in the TCR-pMHC crystal structures of 2B4 and 226, but the crystal structure for 5cc7 had a different peptide (5c1, ANGVAFFLTPFKA). Both the experimentally selected peptides and the top-scoring peptides by our modeling method revealed peptide motifs similar to these cognate peptides (their residues are in black boxes in Figure 5.3).

Because our modeling method started with the ternary complex structure containing cognate peptides, our method may simply favor peptides with similar sequences. Nevertheless, our scoring method does reproduce many amino acid substitutions seen in the top experimentally selected peptides (marked with amino acid frequency in Figure 5.3). We defined a substitution to be shared between the experimental and computed peptide sets if the frequency of the mutant amino acid at its peptide position was 2 fold higher than what would be expected by chance in both heatmaps (based on the NNK codon library used to

design the yeast-display libraries). The following substitutions are shared between the two heatmaps for 2B4: L-1H, L-1Q, L-1W, Y3F, Q6E, and T8S. For 226, the shared substitutions are L-1W, A2G, Y3F, and Q6A. For 5cc7, the shared substitutions are A2G, 5LK, T6A, P7A, and F8Y. For Ob.1A12, the shared substitutions are E-4A, N-3H, N-3E, P-2Q, P-2R, H2F, K5R, N6A, I7Q, V8I, T9G, T9C, P10R. Hence, although our modeling method may be biased toward the cognate peptide, our modeling and scoring method is still capable of identifying target peptides with beneficial or permissible mutations.

To provide a quantitative assessment of whether our method outperformed a baseline method simply based on sequence similarity to the cognate peptide, we computed the Hamming distance between the cognate peptide sequence and the sequence of each peptide in the union of the preselection and the round-four library for each TCR. The top peptides with the highest sequence similarity to the cognate peptide (lowest Hamming distance) were used to generate heatmaps of amino acid preference. The correlation between these heatmaps and heatmaps of the top experimentally selected peptides was compared to the equivalent correlation derived from the top ΔG_{BIND} peptides as described above (Figure 5.3, C). For one TCR (2B4) similar correlations were found when peptides were selected based on the ΔG_{BIND} score or Hamming distance. However, a notable improvement in correlation was detected for the other three TCRs (226, 5cc7, and Ob.1A12) when peptides were selected based on ΔG_{BIND} as opposed to mere sequence similarity with cognate

peptide. The strong improvements in correlation for three of the four TCRs highlight the value of incorporating structural information in next-generation peptide prediction algorithms**.**

Discussion

Numerous methods exist for the prediction of peptide binding to either class I or class II MHC molecules and have achieved high accuracy dependent upon the training and testing data utilized (Zhao and Sher, 2018). However, far fewer tools are available for prediction of TCR binding to pMHC and the accuracy of existing tools show room for improvement (Tung *et al.*, 2011; Pierce and Weng, 2013; Lanzarotti, Marcatili and Nielsen, 2018; Schneidman-Duhovny *et al.*, 2018; Ogishi and Yotsuyanagi, 2019). Utilizing structural information from four TCR-pMHC complexes, we present a high throughput modeling and scoring approach capable of successfully selecting cross-reactive peptides from large pools of primarily non-binding peptides. Our method outperforms the approach based on sequence similarity to the cognate peptides.

Several other groups incorporated structural information of the TCR-pMHC interface to aid in binding prediction. In a recent study, optimized FoldX and Rosetta energy terms were used to predict peptide binding given the sequences of MHC, TCR and a query peptide (Lanzarotti, Marcatili and Nielsen, 2018). The authors noted that the availability of a high sequence identity structural TCR template and successful prediction of peptide binding to MHC were vital to the success of their TCR-pMHC binding prediction. Similarly, the

method ITCell utilizes atomic statistical potentials to predict a TCR's peptide epitope from all possible peptides in the full-length parent protein when given sequences of class II MHC, the TCR variable region, and the parent protein antigen as input (Schneidman-Duhovny *et al.*, 2018). In the majority of test cases, ITCell ranked the correct peptide epitope among the top 20 peptides among all peptides that could result from the parent antigen.

Benchmarking sets for the aforementioned methods were generated by using overlapping peptides from the parent protein sequence of the cognate peptide as negatives (excluding the cognate), based on the assumption that parent protein sequence would harbor only peptide epitopes for a single TCR, which resulted in ~10^2-10^3 query peptides per TCR-pMHC test case (Lanzarotti, Marcatili and Nielsen, 2018; Schneidman-Duhovny *et al.*, 2018). A more exact set of non-binding peptides would require experimental evidence for failed binding. Here, we present deep-sequencing results from yeast display as a robust and larger benchmarking tool for TCR epitope prediction. In particular, each preselection library provided >10^5 peptides, which were not selected by the TCR of interest and are likely negative non-binding peptides. Although 10^5 peptides is still a small subset of the theoretical diversity for the 13-mer (~8.1 x 10^{16}) and 14-mer (~1.6 x 10^{18}) peptides, they represent a larger challenge than previous benchmarks for predicting TCR epitopes.

Like our study, the success of both of the aforementioned methods relied on accurate template-based modeling of the TCR-pMHC complex (Lanzarotti,

Marcatili and Nielsen, 2018; Schneidman-Duhovny *et al.*, 2018). In our work, modeling of the TCR-pMHC was simplified because crystal structures of TCR-pMHC complexes existed for all four TCRs investigated and only structural changes resulting from the different peptide sequences needed to be accounted for. Previous studies showed the TCR's complementarity determining region (CDR) loops can be flexible and change their conformations upon ligand binding (Reiser *et al.*, 2002, 2003; Gagnon *et al.*, 2006; Scott *et al.*, 2011; Pierce and Weng, 2013). Furthermore, it has been shown CDR flexibility can contribute to cross-reactivity (Reiser *et al.*, 2003; Hawse *et al.*, 2014). It may be surprising how well our modeling and scoring method performed without making any structural adjustments to the TCR molecules. It is unlikely our modeling method could predict antigens that require large backbone movements, or altered binding orientation, of the TCR for recognition. However, while our modeling method is conservative in terms of modeling any structural changes of the TCRs CDR3 loops, it appears to perform well in providing poor scores for unfavorable peptides.

Large conformational changes of the peptide can also occur upon TCR binding. For example, the DMF5 TCR that recognizes the MART-1 melanoma antigen presented by the class I MHC protein HLA-A2 was shown to cross-react with the DRG class of peptides that are chemically distinct from MART-1 (Gee *et al.*, 2018). DMF5 TCR binding to an HLA-A2-presented DRG-class peptide led to a 'register shift' in the peptide, causing a C-terminal peptide extension from the

MHC binding groove (Riley *et al.*, 2018). Identification of cross-reactive peptides with such large structural adjustments relative to cognate peptide would be missed by our fixed-backbone peptide-modeling approach, as would instances in which MHC deformations are required (Borbulevych *et al.*, 2009; Borbulevych, Piepenbrink and Baker, 2011). However, peptides of class II pMHC complexes (i.e., those studied here) typically do not bulge from the groove and class II pMHC complexes are thus less prone to backbone rearrangements (Tynan *et al.*, 2005; Ayres, Corcelli and Baker, 2017). Hence, the success seen here with class II complexes may not fully translate when predicting cross-reactivity in class I systems, although we should anticipate success with conformationally simpler modes of cross-reactivity that involve more commonly observed molecular mimicry mechanisms (Macdonald *et al.*, 2009; Borbulevych, Piepenbrink and Baker, 2011).

Even when accounting for simple molecular mimicry mechanisms in cross-reactivity, peptide side-chain must also be precise because a single erroneous side-chain conformation could lead to false positive or false negative predictions. While we do not have an estimate for the accuracy of side-chain modeling for our modeled peptides here, our previous work showed Rosetta's side-chain optimization methods performed well, albeit on a limited set of TCR-pMHC point mutations (Borrman *et al.*, 2017). As advancements in technology allow for faster and more accurate modeling of larger conformational changes, future studies may focus on allowing for flexibility in CDR loops, MHC, and peptide backbone to

potentially identify cross-reactive peptides with distinct structural and chemical signatures.

To score the modeled TCR-pMHC structures, we accounted for the interactions made by the peptide with the TCR and the MHC using Rosetta's scoring application with default weights for energy terms. Future work could potentially improve upon our results by optimizing energy term weights using machine learning approaches and taking advantage of structural and chemical trends in known TCR-pMHC complexes. For example, there is evidence that immunogenic peptides are enriched in hydrophobic amino acids at peptide centers (Calis *et al.*, 2013), and structural modeling combined with neural-network optimized scoring has been used to predict neoantigen immunogenicity (Riley *et al.*, 2019). Here, we employed timesaving modeling and scoring methods to efficiently interrogate large pools of peptides for binding. Future studies optimizing score functions and weights for predicting TCR cross-reactivity might take into account the consequences of the weak affinities TCRs have for their ligands, which can stem from 'imperfect' interfaces that are traditionally difficult to discriminate between with default functions.

One potential application of our method is in cancer immunotherapy. Accurate identification and targeting of neoantigens (peptides derived from mutated tumor proteins) could lead to successful development of immuno-therapeutics. Recent work highlighted the importance of incorporating MHC binding strength, self-similarity to reference antigen, and peptide-centric features

to accurately predict neoantigen immunogenicity (Bjerregaard *et al.*, 2017; Smith *et al.*, 2019). Building on this and other work, structural modeling and scoring of peptide neoantigens in the context of the full TCR-pMHC complex rather than the MHC alone may provide additional insights beneficial to immunogenic prediction (Riley *et al.*, 2019).

Many efforts have been made to enhance TCR affinity for tumor and viral antigens (Holler *et al.*, 2000; Li *et al.*, 2005; Chervin *et al.*, 2008). However, enhanced affinity may lead to increased cross-reactivity (Linette *et al.*, 2013; Riley and Baker, 2018; Hellman *et al.*, 2019). To check for unwanted cross-reactivity of engineered TCRs, one may perform alanine scanning of the antigen to identify motifs essential for binding and then searching for possible self-antigens in a protein sequence database (Obenaus *et al.*, 2015). The alanine scanning can be expedited using DNA barcode-labeled MHC multimers (Bentzen *et al.*, 2016, 2018). A more direct approach is to interrogate all human peptides for cross-reactivity, like the recent T-Scan method which utilized the lentiviral delivery of an antigen library spanning the entire human proteome into antigen-presenting cells. Selected peptides confirmed the cognate MAGE-A3 epitope along with several novel cross-reactive endogenous self-peptides (Kula *et al.*, 2019). Our modeling and scoring method could represent an *in silico* approach with a similarly broad coverage. We could scan all peptides of the entire human proteome computationally for possible binding to an engineered TCR granted a template TCR-pMHC crystal structure is available. The thus identified cross-

reactive antigens could be further tested experimentally using binding assays or assays measuring immunogenic response.

Materials and methods

Sequence extraction

Sequencing data were downloaded from the Sequence Read Archive (SRA) under project code SRP040021 and converted to FASTQ files using the SRA Toolkit. Sequencing reads were split by barcode into their individual selection rounds. Nucleotide sequences were translated into amino acid sequences and peptide sequences containing stop codons or unknown amino acids were discarded. The resulting counts for each unique peptide were recorded for each round of selection for each TCR.

Peptide structure modeling

Template TCR-pMHC complex structures were downloaded from the protein data bank (PDB) with the following PDB IDs: 3QIB (2B4-MCC-I-E^k), 3QIU (226-MCC-I-E^k), 4P2R (5cc7-5c1-I-E^k), and 1YMM (Ob.1A12-MBP-HLA-DR2). To reduce computation time the structures were truncated to contain only the binding interface (up to residue 83 for the class II MHC α chain and residue 93 for the β chain). Each TCR was truncated to just contain the variable domains, excluding the constant domains that are distal from the binding interface. Water molecules were also removed to simplify scoring and for consistency across TCR-pMHC structures with different resolutions. To model peptides onto the template TCR-pMHC structures, we utilized the fixed backbone application, fixbb,

of the Rosetta suite of programs (Version 3.5) (Leaver-Fay, Snoeyink and Kuhlman, 2008; Leaver-Fay *et al.*, 2011), with parameters “extrachi_cutoff 1 –ex1 –ex2 –ex3” to increase χ angle rotamer sampling for side-chain placement of peptide residues. All side chains of the TCR-pMHC aside from those modeled on the peptide were left in their original poses. The following is an example Rosetta command used for peptide structural modeling:

```
rosetta_source/bin/fixbb.linuxgccrelease –database rosetta_database/ -s pdb -resfile my_resfile -suffix my_label -extrachi_cutoff 1 -ex1 -ex2 -ex3
```

Prediction of peptide–MHC/TCR binding free energy

Each modeled TCR-pMHC complex was scored using the following formula:

$$\Delta G_{BIND} = G_{COMPLEX} - (G_{TCR/MHC} + G_{PEPTIDE}) \quad (1)$$

such that $G_{COMPLEX}$ is the Rosetta score for the entire TCR-pMHC complex, $G_{TCR/MHC}$ is the Rosetta score for the TCR and MHC chains bound without the peptide, and $G_{PEPTIDE}$ is the score for the isolated peptide in its bound conformation. To score each component we used Rosetta’s scoring application, score (Leaver-Fay *et al.*, 2011). This scoring function is a linear combination of 19 energy terms, including van der Waals, solvation, electrostatics, and hydrogen bonding interactions along with other statistical potentials. The weights for the energy terms were left in their default settings (those specified in the standard.wts file along with the score12.wts_patch file found in Rosetta’s weights

directory). The following is an example Rosetta command used for scoring modeled TCR-pMHC structures (i.e. scoring commands for $G_{COMPLEX}$, $G_{TCR/MHC}$, and $G_{PEPTIDE}$):

```
rosetta_source/bin/score.linuxgccrelease –database rosetta_database/ -s
my_pdb.pdb -out:file:scorefile outputfile.sc
```

CHAPTER VI: DISCUSSION

Defining compartment principles by liquid chromatin Hi-C

Knowledge pertaining to mechanisms driving compartmentalization of chromatin states in the nucleus is lacking. In Chapter II, I sought to answer fundamental questions regarding compartment formation and maintenance. This goal was accomplished by applying a new experimental technique, liquid chromatin Hi-C, to quantitatively assess chromatin interaction stability and dissolution kinetics of 3D chromatin conformations.

Through liquid chromatin Hi-C, we fragmented the human genome to varying sizes and assessed whether compartment formation was perturbed. We estimate that compartmentalization requires chromatin fragments to be at least 11±4 kb in length. Digesting the genome into fragments less than 6 kb disrupts the canonical compartment checkerboard pattern and increases mixing of A and B loci.

Furthermore, digesting chromatin into less than 6 kb fragments leads to a loss of short range intra-chromosomal interactions and an increase in long range intra-chromosomal interactions as well as inter-chromosomal interactions. The loss of short range interactions is not uniform across the genome, but instead dependent on chromatin state. Specifically, we find loci in A compartments lose more short range interactions in comparison with loci in B compartments.

Tracking these changes in interactions throughout time we assign a half-life of interaction stability, $t_{1/2}$, to every 40 kb locus, representing the time it takes

to reach half of the maximal level of interaction loss. Correlating $t_{1/2}$ with various sub-nuclear structures, we find interactions with lamin associated domains to be the most stable, and interactions near nuclear speckles to be the least stable.

Our results are in accordance with polymer models describing strong heterochromatic B-B interactions driving compartmentalization in nuclei (Falk *et al.*, 2019). We find homotypic A-A interactions to be only slightly stronger than A-B interactions after a 4-hour digest of chromatin into fragments smaller than 6 kb. Furthermore, after 16 hours of digestion, where the majority of fragments are smaller than 3.5 kb, the A-A interaction preference is completely abolished. In contrast, the B-B interaction preference remains distinct in some loci even after 16 hours of digestion.

The unstable A-A interactions reported here, contradict models where stable interactions between active regions drive compartmental segregation. Nevertheless, transcription can successfully predict compartment patterning in various genomes and blocking of transcription leads to a reduction in compartment strength (Rowley *et al.*, 2017). When analyzing the effect of gene expression on chromatin stability we discovered expressed regions in B compartments had shorter half-lives than unexpressed B regions. Such a result is in line with active regions being more or less dispensable in maintaining chromosome conformations. However, when analyzing expressed versus non-expressed regions in the highly active A1 subcompartment, we find a small but significant increase in stability (longer $t_{1/2}$) for expressed regions. This longer $t_{1/2}$

would imply that regions of expression in open chromatin domains, harbor some mechanism of chromatin interaction stabilization. However, it is unclear whether such small increases in $t_{1/2}$ are a consequence of transcription, Pol II occupancy, epigenetic modifications, or other factors. In reference to this conundrum, super enhancer regions, rich in the active histone mark H3K27ac, are known to cluster together upon cohesin depletion, providing evidence that some active regions can form stable contacts, but are ordinarily obstructed by other 3D chromatin structures (Rao *et al.*, 2017).

In the context of transcription's role in compartmentalization, our results indicate transcription and active regions, in general, have a minor effect on compartment maintenance and the formation of stable homotypic chromatin interactions. This position is supported by both polymer simulations (Falk *et al.*, 2019) and by the presence of compartments seen in mouse sperm which is transcriptionally inert (Jung *et al.*, 2017). However, we cannot definitively rule out a role for transcription in compartmentalization. Liquid chromatin Hi-C investigates the architecture of purified nuclei, with a potentially altered transcriptional state compared to intact cells. Modifications to liquid chromatin Hi-C, for instance, performing the cross-linking step on cells as opposed to extracted nuclei (extracted nuclei were used for easier digestion by restriction enzymes), could help to establish the role of transcription in compartmentalization.

Liquid chromatin Hi-C also provides insight pertaining to the mobility of chromatin positioned within active or inactive compartments. Chromatin diffusion and mobility results largely from incessant Brownian motion which can be modeled as a random walk due to random bombardment of molecules (Marshall *et al.*, 1997; Carmo-Fonseca, Platani and Swedlow, 2002). We note the most extensive mixing between A and B loci only occurs after chromatin is segmented into fragments smaller than 6 kb. This upper limit in length may be informative in terms of chromatin diffusion. Assuming the diffusion of chromatin fragments through the nucleoplasm conforms to a random walk, the conformation of these fragments may affect their own mobility. The persistence length of chromatin (length below which chromatin behaves as a rigid rod) is estimated to be 2.4 kb (Dekker *et al.*, 2002; Dekker, 2008). Hence, the persistence length is on the same order of magnitude as the upper limit fragmentation length (6 kb) required for extensive mixing and compartment disruption. It is possible longer fragments, greater than 6 kb, can form higher order 3D conformations which obstruct and limit fragment mobility.

Our estimations of $t_{1/2}$ suggest heterochromatic loci exhibit slower dissociation kinetics relative to euchromatic loci, with the slowest dissociation kinetics found near lamin associated domains at the nuclear periphery. Modeling chromatin mobility as a random walk, its diffusional motion may be modulated or restricted by sub-nuclear structures such as the nucleolus or nuclear envelope (Carmo-Fonseca, Platani and Swedlow, 2002). We find proximity and

interactions with such sub-nuclear structures also modulate $t_{1/2}$. Such differences in mobility have been documented in yeast with telomeres displaying low mobility, due in part to their tethering with the nuclear periphery (Heun *et al.*, 2001). Furthermore, such telomere tethering restricts mobility of genes linearly proximal to telomeres, which is in agreement with predictions from random walk polymer models (Avşaroğlu *et al.*, 2014). Most relevant to our study, an excised repressed locus in yeast also shows constrained mobility localizing to the nuclear periphery. When the locus is derepressed, the excised locus shows enhanced mobility capable of sampling space throughout the entire nucleoplasm (Gartenberg *et al.*, 2004). Of important note in Gartenberg et al.'s results, excision of loci from the polymeric constraint of full chromosomes led to enhanced mobility in both repressed and derepressed loci. Our work builds on this study, by interrogating the spatial positioning of excised fragments with one another across the entire genome and throughout time.

One of the most difficult challenges faced in analysis of liquid chromatin Hi-C results was decoupling effects from both fragmentation and homotypic compartment attraction on chromatin interaction stability. In an ideal experiment to accurately compare interaction stabilities, one would cut the genome into equal segments at exactly the same time and then measure the dissolution kinetics. However, restriction enzymes have biases for where and when they begin cutting their restriction sites which themselves are not equally distributed across the genome. Such biases convolute comparisons of interaction stability

between loci which may be fragmented to different extents at different times. We accounted for such biases via measuring a timecourse of digestion efficiency. However, while we measured residuals of $t_{1/2}$ derived from deviations of expected $t_{1/2}$ based on digestion efficiency, small perturbations in $t_{1/2}$ may be undetected due to over or under correction.

Correlating $t_{1/2}$ with chromatin marks, we show polycomb bound chromatin have short half-lives compared to other condensed heterochromatic regions. In many cases, polycomb bound regions are located in larger A compartment regions. The neighboring chromatin environment of a loci potentially has an effect on its dissolution kinetics. For instance, the shorter $t_{1/2}$ of polycomb bound regions may be attributed, in part, to excision of a condensed piece of chromatin within an open chromatin environment. In such situations, more complex models accounting for neighboring sites may need to be investigated to accurately assign interaction stabilities to distinct loci.

The development of liquid chromatin Hi-C sets a foundation for investigating chromatin stability under a variety of experimental conditions. To test effects of nuclear positioning, liquid chromatin Hi-C could be applied to the inverted nuclei (euchromatin localized to nuclear periphery and heterochromatin localized to nuclear interior) of rod photoreceptors in nocturnal mammals. The static Hi-C maps between inverted and conventional nuclei look similar in terms of compartment formation (Falk *et al.*, 2019). However, dynamic differences related to compartment stability may exist. For example, it would be interesting to

test whether A compartments in inverted nuclei, which are relocated to the nuclear periphery, display the same levels of instability seen in conventional nuclei.

In terms of revealing principles of compartment formation, adaptations of liquid chromatin Hi-C could be applied to nuclei or cells inhibited for transcription, or depleted for factors contributing to compartment formation such as HP1α, lamins, or polycomb-group proteins.

Furthermore, while we assume our results presented here in human lymphoblastoid cells are representative of general chromatin characteristics, it would also be insightful to test whether our results are conserved across varying cell types, tissues, and species. Such experiments could provide insights to developmental processes. For instance, a switch from strong A-A interactions to strong B-B interactions occurs upon differentiation of embryonic stem cells to cortical neurons (Bonev *et al.*, 2017). Comparing $t_{1/2}$ values for A and B loci between embryonic stem cells and cortical neurons could help unravel the mechanisms driving these conformation changes.

Liquid chromatin Hi-C would also be applicable in cell cycle studies. For example, chromatin mobility is shown to be high in G1 cells but markedly slower in S phase (Heun *et al.*, 2001). Also, compartments are shown to disappear upon transitioning from G2 phase to prometaphase (Gibcus *et al.*, 2018). Examining changes in $t_{1/2}$ between these cell cycle stages could expand these results and

provide valuable insights into maintenance of chromosome structures throughout the cell cycle.

Chromatin loops and liquid chromatin Hi-C

We show fragmentation of the genome to segments smaller than 6 kb results in a loss of loop structures. Following the loop extrusion model, we propose this may be due to cohesin rings sliding off fragmented chromatin ends near loop anchors destabilizing loop contacts (Sanborn *et al.*, 2015; Fudenberg *et al.*, 2016). We support this hypothesis by showing that after one hour of digestion, loop contacts disappear in Hi-C maps and cohesin is released from DNA in chromatin fractionation assays.

At 10 kb resolution, the mean size of identified K562 loops is 670 kb. Hence, it may be surprising that when the genome is cut by HindIII into 10-25 kb sized fragments, loop contacts are still present. With average loop size being greater than half a megabase, HindIII fragmentation would lead to a fragmentation of internal loop structure in the majority of loops. This implies the anchors of loop loci do not require a continuous chromatin loop body to sustain contacts. Such a conjecture is still in accordance with loop extrusion models, whereby cohesin could be in contact with CTCF loop anchors prior to fragmentation and thus aid in the preservation of anchor contacts.

When fragments are smaller than 6 kb, we propose cohesin may slip off nearby fragmented ends destabilizing loop anchor contacts leading to a loss in cohesin chromatin binding and a disappearance of 3D looping interactions. 6 kb

of linear DNA is estimated to be ~ 2,040 nm in length (van Holde, 1989). Considering the relative size of the cohesin complex, with an estimated ring diameter of ~ 40 nm, traversing the necessary fragment length for release is not definitively reconciled (Haering *et al.*, 2002). However, as cohesin is shown to form loops at a rate up to 2.1 kb per second *in vitro*, and the conformation and DNA packing density of fragments at loop anchors is unknown, dissociation of cohesin from fragment ends may be a reasonable hypothesis (Davidson *et al.*, 2019). It should be noted, cohesin mediated loop formation is dependent on cohesin's ATPase activity (Davidson *et al.*, 2019; Kim *et al.*, 2019). As there may be insufficient ATP levels to support active loop extrusion in isolated nuclei, future modifications to liquid chromatin Hi-C, examining the effects of varying ATP concentrations, may support or contradict current models of looping mechanisms.

Visually and computationally, we identify loops in Hi-C maps as dots of significantly high interactions standing out above background (Rao *et al.*, 2014). Considering the increase in A-B interactions in our DpnII digested Hi-C maps, random mixing of fragments less than 6 kb in size may inadvertently wash out looping signal. Of particular relevance, in our aggregation plots of loop loci, looping signal was highly reduced, but not completely abolished after 4 hours of DpnII digestion. As Hi-C maps represent a population average, it is unclear whether remaining loop contacts are driven by a subset of nuclei in a particular

state. Future advancements in single cell studies, could help to resolve the effect digestion has on loop structure per cell.

Further obfuscating conclusions, the manner in which cohesin pseudo-topologically or non-topologically interacts with chromatin to form loops is still under investigation, along with the number of cohesin complexes required for loop formation (Davidson *et al.*, 2019; Kim *et al.*, 2019). Future studies defining how cohesin, CTCF, and other factors interact with chromatin to form loops will be critical to understanding how loops are maintained and how these conformations regulate nuclear processes. We propose liquid chromatin Hi-C experiments will be impactful in such studies by providing information pertaining to the dynamics and stability of loop contacts.

3D genome architecture and schizophrenia

In Chapter III, we sought to expand potential schizophrenia risk genes by investigating the 3D interactions occurring at coding and nocoding risk loci. As schizophrenia risk variants are shown to be elevated in neuronally expressed genes (Genovese *et al.*, 2016; Skene *et al.*, 2018), we applied hiPSC technology to generate near pure populations of NPCs, glial cells, and excitatory neurons (Takahashi and Yamanaka, 2006; Brennand *et al.*, 2011; Ho *et al.*, 2016; TCW *et al.*, 2017). To map the 3D architecture of the different brain cell types, we performed Hi-C experiments and investigated resultant 3D contacts in the context of schizophrenia risk (Lieberman-Aiden *et al.*, 2009; Rao *et al.*, 2014).

To identify candidate schizophrenia risk genes, we first set out to map all chromatin loop structures in our brain cell types, presuming some of these loops could link risk variants to candidate risk genes. Loop calling was accomplished through two separate approaches based on either global or local enrichments of interactions (Ay, Bailey and Noble, 2014; Rao *et al.*, 2014; Mifsud *et al.*, 2017). A large percentage of loops were conserved across our cell types and genes overlapping brain-specific loops were enriched for brain specific gene ontology terms. Notably, neurons displayed decreased loop numbers and harbored larger sized loops relative to NPCs and glial cells, suggesting differentiation rewires the 3D genome in a cell type specific manner.

Upon analyzing loops in schizophrenia risk loci, we found an increase in neuron specific and NPC specific loops overlapping risk loci in comparison to glial specific loops. We then tested whether NPC loops, linking risk loci to candidate genes, had regulatory effects via CRISR/Cas9 based editing assays. Both CRISPR activation and deletion assays in NPCs confirmed regulatory roles for some risk loci. In several cases, editing of a risk locus changed the expression of a loop linked gene located hundreds of kilobases away.

We annotated genes overlapping our cell type specific loops as either within a risk locus or connected to a risk locus via a cell type specific looping interaction. This resulted in a set of risk locus and risk locus connect-genes (potentially containing novel risk genes) for each cell type. Integrating these gene sets with independent gene expression and protein expression datasets revealed

patterns of significant coregulation across expression samples. We propose such coregulation patterns exemplify regulatory pathways involving risk loci and 3D architectures underlying the genetic etiology of schizophrenia.

Our work in Chapter III was largely inspired by a similar study linking 3D genome architecture to schizophrenia in the human brain (Won *et al.*, 2016). Won and colleagues performed Hi-C on human brain tissue from the germinal zone (composed primarily of NPCs) and the cortical plate (composed primarily of post-mitotic neurons). Significant Hi-C interactions between schizophrenia risk loci and candidate risk genes were identified and validated for regulatory potential via CRISPR/Cas9 editing assays. We employed a similar methodology, but took advantage of hiPSC technology to uncover cell-type specific 3D genome architecture changes as opposed to changes seen in bulk tissue.

Recent work of the PsychENCODE Consortium has revealed the importance of cell type in characterization of variance among transcriptional datasets (Wang *et al.*, 2018). As an example, the gene for dopamine receptor, *DRD3*, displayed expression levels that varied greater in cell type than in bulk tissue measurements across individuals in a population. Such examples highlight that cell-type proportions in individuals largely account for cross-population variation in bulk tissue expression. Further illustrating the importance of cell type, schizophrenia risk genes identified by Hi-C and eQTL data were shown to be highly expressed in single cell profiles of excitatory neurons (Wang *et al.*, 2018). Beyond cell type specific changes in transcription, we show cell type specific

changes in 3D genome architecture also take place in the brain and we provide evidence supporting functional roles for these changes.

Before our work, cell type specific 3D genome architecture changes were also presented in mouse NPCs and cortical neurons either derived from *in vitro* differentiation or from primary tissue (Bonev *et al.*, 2017). In mouse, cell type specific loop formations were associated with increased gene expression. We present similar findings here, whereby genes overlapping loop anchors have significantly higher expression compared to random gene sets. We also show changes in loop formations may be accompanied with changes in gene expression. For example, expression of the loop anchored *CUX2* gene is high in neurons possessing the loop, but lower in NPCs where the loop interaction is reduced, and not expressed in glia where the loop interaction is abolished. Our results indicate a large portion of genome architecture is conserved between brain cell types. However, cell type specific changes exist with regulatory and potentially clinical relevance.

Provided sufficient sequencing depth, loops can be visualized on Hi-C maps at 10 kb resolution. To identify loops in our cell types we thus pooled replicate samples to obtain sufficient coverage. Hence, we do not have statistically significant evidence that neurons contain less loop formations than NPCs or glia. However, processing of independent datasets in mouse and human reveal a similar loss of loops upon differentiation from NPCs to neurons (Won *et al.*, 2016; Bonev *et al.*, 2017). In our processing of mouse ES cells,

NPCs, and neurons, we note an increase in loops upon transitioning from ES cells to NPCs (Bonev *et al.*, 2017). This increase in loops is also supported by an independent study where loss of pluripotency is accompanied by a widespread gain of chromatin loops (Pękowska *et al.*, 2018). Such a gain occurs upon transitioning from mouse ES cells to neural stem cells and is proposed to effect the regulation of developmental genes (Pękowska *et al.*, 2018). However, this model of developmental reorganization could be challenged by considering the technical variance in Hi-C quality metrics for these different cell types. Our work supports a hypothesis that loops are gained upon differentiation from ES to NPCs and then lost upon further differentiation from NPCs to neurons.

Along with a loss in neuronal loops we report an increase in neuron loop size relative to NPC and glia. Once again, we find this increase in neuron loop size is also consistent with an increase in loop size upon the NPC to neuron transition when processing independent Hi-C datasets of mouse and human tissue (Won *et al.*, 2016; Bonev *et al.*, 2017). Previous studies have reported knockout of the cohesin release factor WAPL leads to an increase in larger loop structures as WAPL acts to restrict loop extension (Haarhuis *et al.*, 2017). Intriguingly, gene expression of *Wapl* in mouse is decreased in neurons relative to NPCs, potentially contributing to the larger loop structures in neurons (Bonev *et al.*, 2017). While the larger neuronal loops agree with the larger loop structures seen in WAPL knockout cells, the decreased number of loops in neurons contradicts the expected larger number of loops found in WAPL knockout cells

(Haarhuis *et al.*, 2017). As population Hi-C studies are incapable of distinguishing whether multi-dot clusters in Hi-C maps are representative of multiple loops in a single cell or variable sized loops across different cells, further studies refining annotation of loop size and number will be critical toward understanding developmental changes in 3D chromatin organization.

We note in CRISPR/Cas9 editing assays, deletion of loop-connected risk loci leads to an increase in expression in both *ASCL1* and *EFNB1* genes. Such an increase is suggestive that risk loci carry a repressive as opposed to enhancer function (Traxler *et al.*, 2016). Furthermore, CRISPR/Cas9 activation assays show a reduction in gene expression as opposed to enrichment when activators are targeted to risk loci. It has been reported that there is an inverse correlation between basal level expression and CRISPR-based activated expression across genes (Konermann *et al.*, 2015), and in some cases weak activators can cause repression rather than activation in highly expressed genes (Li *et al.*, 2017). Our RNA-seq analysis estimated *EFNB1* expression to be above the 95th percentile in NPC samples. Additionally, targeting of an activator to the promoter of *EFNB1* as opposed to risk loci led to a minimal increase in expression level or no increase dependent on activation system. Conversely, targeting of the *ASCL1* promoter caused up to 4-fold changes in expression of *ASCL1*. Hence, while we can speculate on the mechanisms driving such regulation patterns, further tests would be required to unravel precisely how these risk loci regulate their target genes. It would be particularly insightful to

investigate whether 3D architecture changes in NPCs occur upon CRISPR/Cas9 editing, as such changes may also have regulatory effects.

We show significant coregulation occurs between genes annotated as risk locus and risk locus-connect when compared to gene sets with similar genomic distance distributions. However, it remains unclear whether coregulation of gene sets is driven by participation in significant 3D interactions, or whether coregulation is a result of genes residing in and near schizophrenia risk loci. Deeper analyses could examine coregulation patterns of genes in schizophrenia risk loci which are not involved in looping contacts, and conversely examine coregulation of genes that make looping contacts in other regions of the genome. Such analyses could help further delineate the role of 3D genome architecture in the context of schizophrenia.

Our results support a hypothesis where schizophrenia pathology can be explained, in part, by variant loci at enhancer elements misregulating their spatially proximal target genes. We show CRISPR/Cas9 based editing of risk loci can lead to target gene expression changes. However, these genome edits did not identically mimic the schizophrenia mutations. Furthermore, these tests were conducted on genomes derived originally from non-schizophrenic human samples. Future studies performing Hi-C experiments on samples derived from patients with schizophrenia could examine directly whether disease specific 3D architectures exist. For instance, it would be interesting to discover the percentage of our risk locus and risk-locus connect loops that are actually

perturbed in schizophrenia genomes harboring the risk variants. Applying hiPSC technology to such studies could also further parse out the role of cell type in schizophrenia etiology.

There is currently a limited number of chromosome conformation datasets for cell types of the human nervous system. Our work represents a primary resource to the scientific community highlighting functional differences in 3D architectures between brain cell types. Similarly, a recent study described cell-type-specific 3D contacts associated with cell-type-specific expression within three different iPSC derived human neuronal cell types: excitatory neurons, hippocampal dentate gyrus-like neurons and lower motor neurons (Song *et al.*, 2019). As expression and 3D architecture changes occur even within neuronal subtypes, mapping out cell type specific architectures and regulatory pathways may prove essential for characterizing the pathology of schizophrenia and other psychiatric diseases.

Our study expanded the set of candidate genes conferring schizophrenia risk by integrating 3D interactions with previously annotated risk loci. Beyond GWAS and Hi-C assays, different groups have worked to further expand and refine the map of genetic schizophrenia risk by incorporating data from ChIP-seq, RNA-seq, NOMe-seq, and QTLs (Rhie *et al.*, 2018; Wang *et al.*, 2018). Such studies aid in the construction of enhancer profiles, and highlight their variability across cell types, individuals, and in disease. Future studies incorporating single cell protocols may help to further elucidate the mechanistics behind cell type

specific roles in disease. For instance, single cell Hi-C assays could be used to interpret the role of the cell cycle in brain specific 3D architecture relevant to schizophrenia risk (Nagano *et al.*, 2017).

Prediction of TCR-pMHC binding energies and cross-reactive peptides

Chapters IV and V focus on prediction of TCR-pMHC binding energies and prediction of cross-reactive peptides, respectively. In both cases, prediction is based on structural modeling and scoring of TCR-pMHC complex interactions.

In Chapter IV, we developed a publicly accessible and queryable database, ATLAS, to link 3D structures of TCR-pMHC complexes to associated binding energies. Utilizing ATLAS data, we performed multilinear regression analyses to assess performance in prediction of TCR-pMHC binding energy by combinations of energy features calculated from crystal structures. We report a correlation of 0.48 between experimental and predicted binding energies, along with a correlation of 0.63 between experimental changes in binding energies upon mutation and our predicted changes.

In Chapter V, we repurposed deep sequencing data from yeast display of pMHC libraries (Birnbaum *et al.*, 2014) to create a functional benchmark for *in silico* prediction of cross-reactive peptides for specific TCRs. Structural modeling and scoring of peptides in TCR-pMHC complexes resulted in a high concordance between experimental and predicted TCR cross-reactivity profiles. In addition, we show incorporation of structural information to predict cross-reactive peptides improves upon prediction of cross-reactivity based solely on sequence

similarities. Hence, our work highlights the value of including structural information in cross-reactive antigen prediction.

ATLAS expands on other immunological databases, such as IEDB and Antijen, by directly linking TCR-pMHC affinity measurements to complex structure (Toseland *et al.*, 2005; Vita *et al.*, 2015). Incorporating all affinity and structural information currently available, our work sets a baseline for TCR-pMHC binding energy prediction within a machine learning framework. We hypothesize predictions will improve as more affinity and structural data becomes available, along with advances and optimizations to machine learning methods.

Our study in Chapter V expands upon structure based peptide prediction methods such as ITCell, and Lanzarotti and colleague's TCR:p:MHC pipeline (Lanzarotti, Marcatili and Nielsen, 2018; Schneidman-Duhovny *et al.*, 2018). In particular, we integrated deep sequencing data with structural modeling of TCR-pMHC complexes. Compared to previous methods, our integrative approach increased the peptide test space by several orders of magnitude. As T cells are required to interrogate a vast space of peptide antigens for efficient immunological protection, our work pushes toward the development of larger scale clinically applicable prediction methods.

A caveat to the work presented in both Chapters IV and V is the unknown level of accuracy in our modeled mutations and peptides, respectively. We were able to determine that the majority of modeled side chain conformations were correct in a small subset of ATLAS entries containing structures of both wildtype

and mutant complexes. However, it is not clear what percentage of mutations or peptides are mismodeled among our entire datasets. It also remains to be understood the effect mismodeled side chain conformations have on binding energy or cross-reactive antigen prediction. Mismodeled residues involved in contacts between TCR and pMHC may disrupt essential interactions; however, the degree to which a single mismodeled χ angle affects prediction is unclear. Deeper analyses investigating patterns in side chain conformations at specific regions in the TCR-pMHC interface may aid in the validation of modeled structures and predictive performance.

In regard to cross-reactive peptide prediction, we show structural modeling and scoring is successful in prediction of binding peptides from large sets of non-binding peptides. This method was successful for the four TCR structures investigated in Chapter V; yet, it remains to be seen how such methods would perform on any given TCR. The large majority of TCRs engage pMHC in a canonical diagonal docking orientation (Rossjohn *et al.*, 2015). However, conformational changes in CDR loops and the antigen recognition surface have been reported across a wide selection of TCR-pMHC complexes (Baker *et al.*, 2012). The studies of Birnbaum et al. suggest some TCRs may also be more cross-reactive than others, capable of tolerating a higher number of amino acid substitutions at specific peptide positions (Birnbaum *et al.*, 2014). Applying our scoring and modeling methods to a variety of TCRs would help to answer how

different structure conformations affect cross-reactivity profiles and prediction performance.

We focused our studies on prediction of TCR-pMHC binding. While binding is a prerequisite to TCR signaling, the relationship between affinity and signaling is not clearly defined. TCR-pMHC affinity and subsequent signaling are correlated, but high affinity does not necessarily ensure a strong signaling response (Holler and Kranz, 2003; Stone and Kranz, 2013). Our work, concentrating on the first step in a T cell mediated immune response, inspires further development of models relating TCR-pMHC affinity with downstream signaling. For instance, the kinetic proofreading model hypothesizes TCRs must engage with pMHC with a sufficient dwell time to produce an activation signal, while also disengaging quickly enough to allow the pMHC molecule to interact with other TCRs in the contact zone (Mckeithan, 1995; Bridgeman *et al.*, 2012). Another important note, we only consider the 1:1 interaction between a TCR molecule and pMHC molecule; however, T cell activation results from the clustering of multiple TCRs at the cell surface into supramolecular activation clusters (Monks *et al.*, 1998). It follows that more complex models taking into account energies and kinetics on the macromolecular and supramolecular scale will be essential to accurate prediction of T cell mediated immune responses.

As more TCR-pMHC complexes are solved with matched affinity data, it will be interesting to see how incorporation of new data optimizes weight estimates in scoring functions and affects overall prediction performance. When

we applied scoring weights optimized by ATLAS to cross-reactive antigen prediction, results were poorer than using default scoring weights from Rosetta. We assume this is largely due to the fact that ATLAS is trained on structures with peptide bound to MHC. Hence, weights are not trained to account for peptides which are incapable of MHC binding. It follows, future optimization of scoring functions may need to incorporate MHC binding for successful cross-reactive prediction from peptide sequence alone.

In terms of clinical use we propose our methods could be used to advance the fields of TCR design in T cell based therapeutics and TCR testing for cross-reactivity. While our prediction performance lacks the accuracy required for cross-reactivity testing in a clinical setting, we provide evidence that computational approaches can be used to discriminate binding peptides from large pools of non-binders. We hypothesize further integration of deep sequencing and structural data will lead to improvements in antigen prediction and pave the way for interrogation of large peptide sets derived from the human proteome and foreign pathogens.

Conclusion

Much is yet to be discovered in terms of how the 3D architecture of the genome regulates cell function. This thesis presented principles governing genome compartmentalization and the formation of chromatin loops. These two structural phenomenons represent essential genomic features across a wide variety of organisms. As such, a deep understanding of the mechanisms driving

formation of compartments and loops is critical to our understanding of cellular processes. The results of this thesis contribute to characterizing these structural phenomenons and may provide further insights into disease pathology where such phenomenons are perturbed.

By identifying chromatin loop formations in hiPSC derived NPCs, neurons, and glial cells, this thesis expanded the set of schizophrenia risk genes by linking risk loci with loop linked genes. Our results define cell type specific schizophrenia related “chromosomal connectomes”, consisting of schizophrenia risk loci and their 3D proximal genes. Such connectomes harbor genes associated with 3D architecture changes during development and genes critical to neuronal and chromatin remodeling processes. We propose our reported chromosomal connectomes provide a foundation for further studies aimed at characterizing the etiology of schizophrenia.

Lastly, this thesis presents methods for prediction of TCR-pMHC binding energies and cross reactive peptides. We developed a public database, ATLAS, to facilitate modeling and scoring of TCR-pMHC interactions by the scientific community. We used ATLAS to train scoring functions for successful prediction of TCR-pMHC binding energies and changes in binding energy upon mutation. By integrating deep sequencing data with structural methods, we also present successful predictions of TCR cross-reactivity upon interrogation of large peptide sets. We hope our database, methods, and results support future *in silico* approaches to TCR design and cross-reactivity prediction.

You want to know how I did it? This is how I did it, Anton. I never saved anything for the swim back.

-Andrew Niccol (1997), *GATTACA*

BIBLIOGRAPHY

Adams, J. J. *et al.* (2016) 'Structural interplay between germline interactions and adaptive recognition determines the bandwidth of TCR-peptide-MHC cross-reactivity', *Nature Immunology*, 17(1), pp. 87–94. doi: 10.1038/ni.3310.

Aleksic, M. *et al.* (2010) 'Dependence of T Cell Antigen Recognition on T Cell Receptor-Peptide MHC Confinement Time', *Immunity*. Elsevier Ltd, 32(2), pp. 163–174. doi: 10.1016/j.immuni.2009.11.013.

Alexander, R. P. *et al.* (2010) 'Annotating non-coding regions of the genome', *Nature Reviews Genetics*. Nature Publishing Group, 11(8), pp. 559–571. doi: 10.1038/nrg2814.

Alipour, E. and Marko, J. F. (2012) 'Self-organization of domain structures by DNA-loop-extruding enzymes', *Nucleic Acids Research*, 40(22), pp. 11202–11212. doi: 10.1093/nar/gks925.

Allis, C. D. and Jenuwein, T. (2016) 'The molecular hallmarks of epigenetic control', *Nature Reviews Genetics*. Nature Publishing Group, 17(8), pp. 487–500. doi: 10.1038/nrg.2016.59.

Allshire, R. C. and Madhani, H. D. (2018) 'Ten principles of heterochromatin formation and function', *Nature Reviews Molecular Cell Biology*. Nature Publishing Group, 19(4), pp. 229–244. doi: 10.1038/nrm.2017.119.

Andersson, R. and Sandelin, A. (2020) 'Determinants of enhancer and promoter activities of regulatory elements', *Nature Reviews Genetics*. Springer US, 21(2),

pp. 71–87. doi: 10.1038/s41576-019-0173-8.

Andreatta, M. and Nielsen, M. (2015) ‘Gapped sequence alignment using artificial neural networks: Application to the MHC class i system’, *Bioinformatics*, 32(4), pp. 511–517. doi: 10.1093/bioinformatics/btv639.

Aranda, S., Mas, G. and Croce, L. Di (2015) ‘Regulation of gene transcription by Polycomb proteins’, *Science Advances*, 2(December), pp. 1–16.

Arstila, T. P. *et al.* (1999) ‘A direct estimate of the human alphabeta T cell receptor diversity.’, *Science (New York, N.Y.)*, 286(October), pp. 958–961. doi: 10.1126/science.286.5441.958.

Avşaroğlu, B. *et al.* (2014) ‘Effect of chromosome tethering on nuclear organization in yeast’, *PLoS ONE*, 9(7). doi: 10.1371/journal.pone.0102474.

Ay, F., Bailey, T. L. and Noble, W. S. (2014) ‘Statistical confidence estimation for Hi-C data reveals regulatory chromatin contacts’, *Genome Research*, 24(6), pp. 999–1011. doi: 10.1101/gr.160374.113.

Ayres, C. M., Corcelli, S. A. and Baker, B. M. (2017) ‘Peptide and peptide-dependent motions in MHC proteins: Immunological implications and biophysical underpinnings’, *Frontiers in Immunology*, 8(AUG), pp. 1–9. doi: 10.3389/fimmu.2017.00935.

Babbitt, B. P. *et al.* (1985) ‘Binding of immunogenic peptides to Ia histocompatibility molecules’, *Nature*, 317, pp. 359–361.

Baker, B. M. *et al.* (2012) ‘Structural and dynamic control of T-cell receptor specificity, cross-reactivity, and binding mechanism’, *Immunological Reviews*,

250(1), pp. 10–31. doi: 10.1111/j.1600-065X.2012.01165.x.

Banerji, J., Rusconi, S. and Schaffner, W. (1981) ‘Expression of a β-globin gene is enhanced by remote SV40 DNA sequences’, *Cell*, 27(2 PART 1), pp. 299–308. doi: 10.1016/0092-8674(81)90413-X.

Banigan, E. J., Stephens, A. D. and Marko, J. F. (2017) ‘Mechanics and Buckling of Biopolymeric Shells and Cell Nuclei’, *Biophysical Journal*. Biophysical Society, 113(8), pp. 1654–1663. doi: 10.1016/j.bpj.2017.08.034.

Bannister, A. J. *et al.* (2005) ‘Spatial distribution of di- and tri-methyl lysine 36 of histone H3 at active genes’, *Journal of Biological Chemistry*, 280(18), pp. 17732–17736. doi: 10.1074/jbc.M500796200.

Bannister, A. J. and Kouzarides, T. (2011) ‘Regulation of chromatin by histone modifications’, *Cell Research*. Nature Publishing Group, 21(3), pp. 381–395. doi: 10.1038/cr.2011.22.

Barry, M. and Bleackley, R. C. (2002) ‘Cytotoxic T lymphocytes: all roads lead to death.’, *Nature reviews. Immunology*, 2(6), pp. 401–9. doi: 10.1038/nri819.

Barski, A. *et al.* (2007) ‘High-Resolution Profiling of Histone Methylations in the Human Genome’, *Cell*, 129(4), pp. 823–837. doi: 10.1016/j.cell.2007.05.009.

Becker, J. S., Nicetto, D. and Zaret, K. S. (2016) ‘H3K9me3-Dependent Heterochromatin: Barrier to Cell Fate Changes’, *Trends in Genetics*. Elsevier Ltd, 32(1), pp. 29–41. doi: 10.1016/j.tig.2015.11.001.

Belaghzal, H., Dekker, J. and Gibcus, J. H. (2017) ‘Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome

conformation', *Methods*. Elsevier Inc. doi: 10.1101/090001.

Bentzen, A. K. *et al.* (2016) 'Large-scale detection of antigen-specific T cells using peptide-MHC-I multimers labeled with DNA barcodes', *Nature Biotechnology*, 34(10), pp. 1037–1045. doi: 10.1038/nbt.3662.

Bentzen, A. K. *et al.* (2018) 'T cell receptor fingerprinting enables in-depth characterization of the interactions governing recognition of peptide–MHC complexes', *Nature Biotechnology*, 36(12), pp. 1191–1196. doi: 10.1038/nbt.4303.

Berman, H. M. *et al.* (2000) 'The Protein Data Bank.', *Nucleic acids research*, 28(1), pp. 235–242. doi: 10.1093/nar/28.1.235.

Bharadwaj, R. *et al.* (2013) 'Conserved chromosome 2q31 conformations are associated with transcriptional regulation of GAD1 GABA synthesis enzyme and altered in prefrontal cortex of subjects with schizophrenia', *Journal of Neuroscience*, 33(29), pp. 11839–11851. doi: 10.1523/JNEUROSCI.1252-13.2013.

Bharadwaj, R. *et al.* (2014) 'Conserved higher-order chromatin regulates NMDA receptor gene expression and cognition', *Neuron*. Elsevier Inc., 84(5), pp. 997–1008. doi: 10.1016/j.neuron.2014.10.032.

Bhasin, M. and Raghava, G. P. . (2004) 'Analysis and prediction of affinity of TAP binding peptides using cascade SVM', *Protein Science*, 13(3), pp. 596–607. doi: 10.1110/ps.03373104.

Bickmore, W. A. and Van Steensel, B. (2013) 'Genome architecture: Domain

organization of interphase chromosomes', *Cell*. Elsevier Inc., 152(6), pp. 1270–1284. doi: 10.1016/j.cell.2013.02.001.

Bindea, G. *et al.* (2009) 'ClueGO: A Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks', *Bioinformatics*, 25(8), pp. 1091–1093. doi: 10.1093/bioinformatics/btp101.

Birnbaum, M. E. *et al.* (2014) 'Deconstructing the peptide-MHC specificity of t cell recognition', *Cell*, 157(5), pp. 1073–1087. doi: 10.1016/j.cell.2014.03.047.

Bjerregaard, A. M. *et al.* (2017) 'An analysis of natural T cell responses to predicted tumor neoepitopes', *Frontiers in Immunology*, 8(NOV), pp. 1–9. doi: 10.3389/fimmu.2017.01566.

Bjorkman, P. J. (2015) 'Not Second Class: The First Class II MHC Crystal Structure', *The Journal of Immunology*, 194(1), pp. 3–4. doi: 10.4049/jimmunol.1402828.

Blevins, S. J. *et al.* (2016) 'How structural adaptability exists alongside HLA-A2 bias in the human αβ TCR repertoire'. doi: 10.1073/pnas.1522069113.

Boehmer, H. von and Kisielow, P. (1990) 'Self-Nonself Discrimination by T Cells', *Science*, 248(4961), pp. 1369–1373.

Bogdan, R., Baranger, D. A. and Agrawal, A. (2018) 'Polygenic risk scores in clinical psychology: Bridging genomic risk to individual differences', *Annual Review of Clinical Psychology*, 14, pp. 119–157.

Boisvert, F. M. *et al.* (2007) 'The multifunctional nucleolus', *Nature Reviews Molecular Cell Biology*, 8(7), pp. 574–585. doi: 10.1038/nrm2184.

Bolzer, A. *et al.* (2005) ‘Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes’, *PLoS Biology*, 3(5), pp. 0826–0842. doi: 10.1371/journal.pbio.0030157.

Bonev, B. *et al.* (2017) ‘Multiscale 3D Genome Rewiring during Mouse Neural Development’, *Cell*. Elsevier Inc., 171(3), pp. 557.e1-557.e24. doi: 10.1016/j.cell.2017.09.043.

Bonev, B. and Cavalli, G. (2016) ‘Organization and function of the 3D genome’, *Nature Reviews Genetics*, 17(11), pp. 661–678. doi: 10.1038/nrg.2016.112.

Borbulevych, O. Y. *et al.* (2009) ‘T cell receptor cross-reactivity directed by antigen-dependent tuning of peptide-MHC molecular flexibility.’, *Immunity*. Elsevier Ltd, 31(6), pp. 885–96. doi: 10.1016/j.immuni.2009.11.003.

Borbulevych, O. Y., Piepenbrink, K. H. and Baker, B. M. (2011) ‘Conformational melding permits a conserved binding geometry in TCR recognition of foreign and self molecular mimics.’, *Journal of immunology (Baltimore, Md. : 1950)*, 186(5), pp. 2950–8. doi: 10.4049/jimmunol.1003150.

Borrman, T. *et al.* (2017) ‘ATLAS: A database linking binding affinities with structures for wild-type and mutant TCR-pMHC complexes’, *Proteins: Structure, Function and Bioinformatics*, (85), pp. 908–916. doi: 10.1002/prot.25260.

Boyle, S. (2001) ‘The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells’, *Human Molecular Genetics*, 10(3), pp. 211–219. doi: 10.1093/hmg/10.3.211.

Brennand, K. *et al.* (2015) ‘Phenotypic differences in hiPSC NPCs derived from

patients with schizophrenia', *Molecular Psychiatry*. Nature Publishing Group, 20(3), pp. 361–368. doi: 10.1038/mp.2014.22.

Brennand, K. J. *et al.* (2011) 'Modelling schizophrenia using human induced pluripotent stem cells', *Nature*. Nature Publishing Group, 473(7346), pp. 221–225. doi: 10.1038/nature09915.

Bridgeman, J. S. *et al.* (2012) 'Structural and biophysical determinants of αβ T-cell antigen recognition', *Immunology*, 135(1), pp. 9–18. doi: 10.1111/j.1365-2567.2011.03515.x.

Bronshtein, I. *et al.* (2009) 'Transient anomalous diffusion of telomeres in the nucleus of mammalian cells', *Physical Review Letters*. doi: 10.1103/PhysRevLett.103.018102.

Bronshtein, I. *et al.* (2015) 'Loss of lamin A function increases chromatin dynamics in the nuclear interior', *Nature Communications*. Nature Publishing Group, 6, pp. 1–9. doi: 10.1038/ncomms9044.

Brown, J. H. *et al.* (1993) 'Three-dimensional structure of the human class II histocompatibility antigen HLA-DR1', *Nature*, 364, pp. 33–39.

Buenrostro, J. D. *et al.* (2013) 'Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position', *Nature Methods*, 10(12), pp. 1213–1218. doi: 10.1038/nmeth.2688.

Buenrostro, J. D. *et al.* (2015) 'Single-cell chromatin accessibility reveals principles of regulatory variation', *Nature*, 523(7561), pp. 486–490. doi:

10.1038/nature14590.

Calis, J. J. a. *et al.* (2013) ‘Properties of MHC Class I Presented Peptides That Enhance Immunogenicity’, *PLoS Computational Biology*, 9(10), p. e1003266. doi: 10.1371/journal.pcbi.1003266.

Carmo-Fonseca, M., Platani, M. and Swedlow, J. R. (2002) ‘Macromolecular mobility inside the cell nucleus’, *Trends in Cell Biology*, 12(11), pp. 491–495. doi: 10.1016/S0962-8924(02)02387-5.

Carter, D. *et al.* (2002) ‘Long-range chromatin regulatory interactions in vivo’, *Nature Genetics*, 32(4), pp. 623–626. doi: 10.1038/ng1051.

Charlesworth, G. *et al.* (2012) ‘Mutations in ANO3 cause dominant craniocervical dystonia: Ion channel implicated in pathogenesis’, *American Journal of Human Genetics*, 91(6), pp. 1041–1050. doi: 10.1016/j.ajhg.2012.10.024.

Chavez, A. *et al.* (2016) ‘Comparison of Cas9 activators in multiple species’, *Nature Methods*, 13(7), pp. 563–567. doi: 10.1038/nmeth.3871.

Chen, J.-L. *et al.* (2005) ‘Structural and kinetic basis for heightened immunogenicity of T cell vaccines.’, *The Journal of experimental medicine*, 201(8), pp. 1243–1255. doi: 10.1084/jem.20042323.

Chen, W. V. and Maniatis, T. (2013) ‘Clustered protocadherins’, *Development (Cambridge)*, 140(16), pp. 3297–3302. doi: 10.1242/dev.090621.

Chen, Y. *et al.* (2018) ‘Mapping 3D genome organization relative to nuclear compartments using TSA-Seq as a cytological ruler’, *J Cell Biol*, p. jcb.201807108. doi: 10.1083/JCB.201807108.

Chervin, A. S. *et al.* (2008) 'Engineering higher affinity T cell receptors using a T cell display system', *Journal of Immunological Methods*. Elsevier B.V., 339(2), pp. 175–184. doi: 10.1016/j.jim.2008.09.016.

Chien, R. *et al.* (2011) 'Cohesin mediates chromatin interactions that regulate mammalian β-globin expression', *Journal of Biological Chemistry*, 286(20), pp. 17870–17878. doi: 10.1074/jbc.M110.207365.

Ciosk, R. *et al.* (2000) 'Cohesin's binding to chromosomes depends on a separate complex consisting of Scc2 and Scc4 proteins', *Molecular Cell*, 5(2), pp. 243–254. doi: 10.1016/S1097-2765(00)80420-7.

Cisse, I. I. *et al.* (2013) 'Real-Time Dynamics of RNA Polymerase II Clustering in Live Human Cells', *Science*, 245(August), pp. 664–667.

Cole, D. K. *et al.* (2007) 'Human TCR-binding affinity is governed by MHC class restriction.', *Journal of immunology (Baltimore, Md. : 1950)*, 178(9), pp. 5727–5734. doi: 10.4049/jimmunol.178.9.5727.

Cole, D. K. *et al.* (2010) 'Modification of MHC anchor residues generates heteroclitic peptides that alter TCR binding and T cell recognition.', *Journal of immunology (Baltimore, Md. : 1950)*, 185(4), pp. 2600–2610. doi: 10.4049/jimmunol.1000629.

Cole, D. K. *et al.* (2013) 'Increased peptide contacts govern high affinity binding of a modified TCR whilst maintaining a native pMHC docking mode', *Frontiers in Immunology*, 4(JUN), pp. 1–8. doi: 10.3389/fimmu.2013.00168.

Cong, L. *et al.* (2013) 'Multiplex Genome Engineering Using CRISPR/Cas

Systems', *Science*, 339(February), pp. 819–824.

Craig Venter, J. *et al.* (2001) 'The sequence of the human genome', *Science*, 291(5507), pp. 1304–1351. doi: 10.1126/science.1058040.

Cremer, T. *et al.* (1982) 'Rabl's model of the interphase chromosome arrangement tested in Chinise hamster cells by premature chromosome condensation and laser-UV-microbeam experiments', *Human Genetics*, 60(1), pp. 46–56. doi: 10.1007/BF00281263.

Crotty, S. (2015) 'A brief history of T cell help to B cells', *Nature Reviews Immunology*. Nature Publishing Group, 15(3), pp. 185–189. doi: 10.1038/nri3803.

Davidson, I. F. *et al.* (2019) 'DNA loop extrusion by human cohesin', *Science*, 366(6471), pp. 1338–1345. doi: DOI: 10.1126/science.aaz3418.

Dayhoff, M. O. and Ledley, R. S. (1962) 'Comprotein: A computer program to aid primary protein structure determination', *AFIPS Conference Proceedings - 1962 Fall Joint Computer Conference, AFIPS 1962*, pp. 262–274. doi: 10.1145/1461518.1461546.

Dekker, J. *et al.* (2002) 'Capturing chromosome conformation', *Science*, 295(5558), pp. 1306–1311. doi: 10.1126/science.1067799.

Dekker, J. (2008) 'Mapping in vivo chromatin interactions in yeast suggests an extended chromatin fiber with regional variation in compaction', *Journal of Biological Chemistry*, 283(50), pp. 34532–34540. doi: 10.1074/jbc.M806479200.

Dekker, J. *et al.* (2017) 'The 4D nucleome project', *Nature*. Nature Publishing Group, 549(7671), pp. 219–226. doi: 10.1038/nature23884.

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

Deng, W. *et al.* (2012) 'Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor', *Cell*. Elsevier Inc., 149(6), pp. 1233–1244. doi: 10.1016/j.cell.2012.03.051.

Dernburg, A. F. *et al.* (1996) 'Perturbation of nuclear architecture by long-distance chromosome interactions', *Cell*, 85(5), pp. 745–759. doi: 10.1016/S0092-8674(00)81240-4.

Dialynas, G. K. *et al.* (2007) 'Plasticity of HP1 proteins in mammalian cells', *Journal of Cell Science*, 120(19), pp. 3415–3424. doi: 10.1242/jcs.012914.

Dillinger, S., Straub, T. and Nemeth, A. (2017) 'Nucleolus association of chromosomal domains is largely maintained in cellular senescence despite massive nuclear reorganisation', *PLoS ONE*, 12(6), pp. 1–28. doi: 10.1371/journal.pone.0178821.

Dixon, J. R. *et al.* (2012) 'Topological domains in mammalian genomes identified by analysis of chromatin interactions.', *Nature*, 485(7398), pp. 376–80. doi: 10.1038/nature11082.

Dixon, J. R. *et al.* (2015) 'Chromatin architecture reorganization during stem cell differentiation', *Nature*. Nature Publishing Group, 518(7539), pp. 331–336. doi: 10.1038/nature14222.

Dobbyn, A. *et al.* (2018) 'Landscape of Conditional eQTL in Dorsolateral

Prefrontal Cortex and Co-localization with Schizophrenia GWAS', *American Journal of Human Genetics*, 102(6), pp. 1169–1184. doi: 10.1016/j.ajhg.2018.04.011.

Dobin, A. *et al.* (2013) 'STAR: Ultrafast universal RNA-seq aligner', *Bioinformatics*, 29(1), pp. 15–21. doi: 10.1093/bioinformatics/bts635.

Dong, P. *et al.* (2017) '3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments', *Molecular Plant*. Elsevier Ltd, 10(12), pp. 1497–1509. doi: 10.1016/j.molp.2017.11.005.

Dostie, J. *et al.* (2006) 'Chromosome Conformation Capture Carbon Copy (5C): A massively parallel solution for mapping interactions between genomic elements', *Genome Research*, 16(10), pp. 1299–1309. doi: 10.1101/gr.5571506.

Duan, Z. *et al.* (2010) 'A three-dimensional model of the yeast genome', *Nature*. Nature Publishing Group, 465(7296), pp. 363–367. doi: 10.1038/nature08973.

Dunham, I. *et al.* (2012) 'An integrated encyclopedia of DNA elements in the human genome', *Nature*, 489(7414), pp. 57–74. doi: 10.1038/nature11247.

Durand, N. C., Robinson, J. T., *et al.* (2016) 'Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom', *Cell Systems*. The Authors, 3(1), pp. 99–101. doi: 10.1016/j.cels.2015.07.012.

Durand, N. C., Shamim, M. S., *et al.* (2016) 'Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments', *Cell Systems*. The Authors, 3(1), pp. 95–98. doi: 10.1016/j.cels.2016.07.002.

Ehrenmann, F., Kaas, Q. and Lefranc, M.-P. (2010) 'IMGT/3Dstructure-DB and

IMGT/DomainGapAlign: a database and a tool for immunoglobulins or antibodies, T cell receptors, MHC, IgSF and MhcSF', *Nucleic Acids Research*, 38(Database), pp. D301–D307. doi: 10.1093/nar/gkp946.

Erdel, F. and Rippe, K. (2018) 'Formation of Chromatin Subcompartments by Phase Separation', *Biophysical Journal*. Biophysical Society, pp. 1–9. doi: 10.1016/j.bpj.2018.03.011.

Ernst, J. *et al.* (2011) 'Mapping and analysis of chromatin state dynamics in nine human cell types', *Nature*. Nature Publishing Group, 473(7345), pp. 43–49. doi: 10.1038/nature09906.

F. H. Crick (1958) 'On protein synthesis', *Symposia of the Society for Experimental Biology*, 12, pp. 138–163. doi: 10.1038/227561a0.

Falk, M. *et al.* (2019) 'Heterochromatin drives compartmentalization of inverted and conventional nuclei', *Nature*. Springer US, 570, pp. 395–399. doi: 10.1038/s41586-019-1275-3.

Ferraiuolo, M. A. *et al.* (2012) 'From cells to chromatin: Capturing snapshots of genome organization with 5C technology', *Methods*. Elsevier Inc., 58(3), pp. 255–267. doi: 10.1016/j.ymeth.2012.10.011.

Forcato, M. *et al.* (2017) 'Comparison of computational methods for Hi-C data analysis', *Nature Methods*. Nature Publishing Group, (June), pp. 14–19. doi: 10.1038/nmeth.4325.

Fraser, P. and Bickmore, W. (2007) 'Nuclear organization of the genome and the potential for gene regulation', *Nature*, 447(7143), pp. 413–417. doi:

10.1038/nature05916.

Fromer, M. *et al.* (2016) ‘Gene expression elucidates functional impact of polygenic risk for schizophrenia’, *Nature Neuroscience*, 19(11), pp. 1442–1453. doi: 10.1038/nn.4399.

Fudenberg, G. *et al.* (2016) ‘Formation of Chromosomal Domains by Loop Extrusion’, *CellReports*. The Author(s), 15(9), pp. 2038–2049. doi: 10.1016/j.celrep.2016.04.085.

Fudenberg, G. *et al.* (2018) ‘Emerging Evidence of Chromosome Folding by Loop Extrusion.’, *Cold Spring Harbor symposia on quantitative biology*, LXXXII, p. 034710. doi: 10.1101/sqb.2017.82.034710.

Fulco, C. P. *et al.* (2016) ‘Systematic mapping of functional enhancer–promoter connections with CRISPR interference’, *Science*, 354(6313), pp. 769–773. doi: 10.1126/science.aag2445.

Fullwood, M. J. *et al.* (2009) ‘Next-generation DNA sequencing of paired-end tags (PET) for transcriptome and genome analyses’, *Genome Research*, 19(4), pp. 521–532. doi: 10.1101/gr.074906.107.

Furlong, E. E. M. and Levine, M. (2018) ‘Developmental enhancers and chromosome topology’, *Science*, 361(6409), pp. 1341–1345. doi: 10.1126/science.aau0320.

Gagnon, S. J. *et al.* (2006) ‘T Cell Receptor Recognition via Cooperative Conformational Plasticity’, *Journal of Molecular Biology*, 363(1), pp. 228–243. doi: 10.1016/j.jmb.2006.08.045.

Garboczi, D. N. *et al.* (1996) ‘Structure of the complex between human T-cell receptor, viral peptide and HLA-A2’, *Nature*, 384(6605), pp. 134–141. doi: 10.1038/384134a0.

Garcia, K. C. *et al.* (1996) ‘An alphabeta T cell receptor structure at 2.5 A and its orientation in the TCR-MHC complex.’, *Science (New York, N.Y.)*, 274(5285), pp. 209–19. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21084669.

Gartenberg, M. R. *et al.* (2004) ‘Sir-mediated repression can occur independently of chromosomal and subnuclear contexts’, *Cell*, 119(7), pp. 955–967. doi: 10.1016/j.cell.2004.11.008.

Gauthier, J. *et al.* (2019) ‘A brief history of bioinformatics’, *Briefings in Bioinformatics*, 20(6), pp. 1981–1996. doi: 10.1093/bib/bby063.

Gee, M. H. *et al.* (2018) ‘Antigen Identification for Orphan T Cell Receptors Expressed on Tumor-Infiltrating Lymphocytes’, *Cell*. Elsevier Inc., 172(3), pp. 549-563.e16. doi: 10.1016/j.cell.2017.11.043.

Geng, C., Vangone, A. and Bonvin, A. M. J. J. (2016) ‘Exploring the interplay between experimental methods and the performance of predictors of binding affinity change upon mutations in protein complexes’, *Protein Engineering, Design and Selection*, 29(8), pp. 291–299. doi: 10.1093/protein/gzw020.

de Gennes, P. G. (1979) *Scaling Concepts in Polymer Physics*, *Cornell University Press*. doi: 10.1063/1.2914118.

Genovese, G. *et al.* (2016) ‘Increased burden of ultra-rare protein-altering variants among 4,877 individuals with schizophrenia’, *Nature Neuroscience*,

19(11), pp. 1433–1441. doi: 10.1038/nn.4402.

Germain, R. N. (2002) ‘T-cell development and the CD4-CD8 lineage decision’, *Nature Reviews Immunology*, 2(5), pp. 309–322. doi: 10.1038/nri798.

Gibcus, J. H. *et al.* (2018) ‘A Pathway for Mitotic Chromosome Formation’, *Science*, 6135(January). doi: 10.1126/science.aao6135.

Gibson, B. A. *et al.* (2019) ‘Organization of Chromatin by Intrinsic and Regulated Phase Separation’, *Cell*. Elsevier Inc., 179(2), pp. 470-484.e21. doi: 10.1016/j.cell.2019.08.037.

Gielis, S. *et al.* (2019) ‘Detection of Enriched T Cell Epitope Specificity in Full T Cell Receptor Sequence Repertoires’, *Frontiers in Immunology*, 10(November). doi: 10.3389/fimmu.2019.02820.

Gil-Sanz, C. *et al.* (2015) ‘Lineage Tracing Using Cux2-Cre and Cux2-CreERT2 Mice’, *Neuron*, 86(4), pp. 1091–1099. doi: 10.1016/j.neuron.2015.04.019.

Gilbert, N. *et al.* (2004) ‘Chromatin architecture of the human genome: Gene-rich domains are enriched in open chromatin fibers’, *Cell*, 118(5), pp. 555–566. doi: 10.1016/j.cell.2004.08.011.

Gilman, S. R. *et al.* (2012) ‘Diverse types of genetic variation converge on functional gene networks involved in schizophrenia’, *Nature Neuroscience*. Nature Publishing Group, 15(12), pp. 1723–1728. doi: 10.1038/nn.3261.

Girdhar, K. *et al.* (2018) ‘Cell-specific histone modification maps in the human frontal lobe link schizophrenia risk to the neuronal epigenome’, *Nature Neuroscience*. Springer US, 21(8), pp. 1126–1136. doi: 10.1038/s41593-018-

0187-0.

Gravano, D. M. and Hoyer, K. K. (2013) ‘Promotion and prevention of autoimmune disease by CD8+ T cells’, *Journal of Autoimmunity*. Elsevier Ltd, 45, pp. 68–79. doi: 10.1016/j.jaut.2013.06.004.

Grubert, F. *et al.* (2015) ‘Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions’, *Cell*. Elsevier Inc., 162(5), pp. 1051–1065. doi: 10.1016/j.cell.2015.07.048.

Gulsuner, S. *et al.* (2013) ‘Spatial and temporal mapping of de novo mutations in schizophrenia to a fetal prefrontal cortical network’, *Cell*, 154(3). doi: 10.1016/j.cell.2013.06.049.

Guo, Y. *et al.* (2015) ‘CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function’, *Cell*. Elsevier Inc., 162(4), pp. 900–910. doi: 10.1016/j.cell.2015.07.038.

Haarhuis, J. H. I. *et al.* (2017) ‘The Cohesin Release Factor WAPL Restricts Chromatin Loop Extension’, *Cell*, (169), pp. 693–707. doi: 10.1016/j.cell.2017.04.013.

Haddad, N., Jost, D. and Vaillant, C. (2017) ‘Perspectives: using polymer modeling to understand the formation and function of nuclear compartments’, *Chromosome Research*. Chromosome Research, 25(1), pp. 35–50. doi: 10.1007/s10577-016-9548-2.

Haering, C. H. *et al.* (2002) ‘Molecular Architecture of SMC Proteins and the Yeast Cohesin Complex’, *Molecular Cell*, 9, pp. 773–788.

Hahn, C. G. *et al.* (2009) 'The post-synaptic density of human postmortem brain tissues: An experimental study paradigm for neuropsychiatric illnesses', *PLoS ONE*, 4(4). doi: 10.1371/journal.pone.0005251.

Hahn, M. *et al.* (2005) 'Unconventional topology of self peptide-major histocompatibility complex binding by a human autoimmune T cell receptor.', *Nature immunology*, 6(5), pp. 490–6. doi: 10.1038/ni1187.

Haidar, J. N. *et al.* (2009) 'Structure-based design of a T-cell receptor leads to nearly 100-fold improvement in binding affinity for pepMHC.', *Proteins*, 74(4), pp. 948–60. doi: 10.1002/prot.22203.

Han, M. and Grunstein, M. (1988) 'Nucleosome loss activates yeast downstream promoters in vivo', *Cell*, 55(6), pp. 1137–1145. doi: 10.1016/0092-8674(88)90258-9.

Hawse, W. F. *et al.* (2014) 'TCR Scanning of Peptide/MHC through Complementary Matching of Receptor and Ligand Molecular Flexibility', *The Journal of Immunology*, 192(6), pp. 2885–2891. doi: 10.4049/jimmunol.1302953.

Hediger, F. *et al.* (2002) 'Live Imaging of Telomeres: yKu and Sir Proteins Define Redundant Telomere-Anchoring Pathways in Yeast', *Current Biology*, 12(24), pp. 2076–2089. doi: 10.1016/s0960-9822(02)01338-6.

Hellman, L. M. *et al.* (2019) 'Improving T Cell Receptor On-Target Specificity via Structure-Guided Design', *Molecular Therapy*. Elsevier Ltd., 27(2), pp. 300–313. doi: 10.1016/j.ymthe.2018.12.010.

Henikoff, S. (1997) 'Nuclear organization and gene expression: Homologous

pairing and long-range interactions', *Current Opinion in Cell Biology*, 9(3), pp. 388–395. doi: 10.1016/S0955-0674(97)80012-9.

Heun, P. *et al.* (2001) 'Chromosome dynamics in the yeast interphase nucleus', *Science*, 294(5549), pp. 2181–2186. doi: 10.1126/science.1065366.

Hilker, R. *et al.* (2018) 'Heritability of Schizophrenia and Schizophrenia Spectrum Based on the Nationwide Danish Twin Register', *Biological Psychiatry*. Elsevier Inc, 83(6), pp. 492–498. doi: 10.1016/j.biopsych.2017.08.017.

Hirano, T. and Mitchison, T. J. (1994) 'A heterodimeric coiled-coil protein required for mitotic chromosome condensation in vitro', *Cell*, 79(3), pp. 449–458. doi: 10.1016/0092-8674(94)90254-2.

Hnisz, D. *et al.* (2017) 'A Phase Separation Model for Transcriptional Control', *Cell*. Elsevier Inc., 169(1), pp. 13–23. doi: 10.1016/j.cell.2017.02.007.

Ho, S. M. *et al.* (2016) 'Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells', *Methods*. Elsevier Inc., 101, pp. 113–124. doi: 10.1016/j.ymeth.2015.11.019.

Ho, S. M. *et al.* (2017) 'Evaluating Synthetic Activation and Repression of Neuropsychiatric-Related Genes in hiPSC-Derived NPCs, Neurons, and Astrocytes', *Stem Cell Reports*. ElsevierCompany., 9(2), pp. 615–628. doi: 10.1016/j.stemcr.2017.06.012.

Hoffman, G. E. *et al.* (2017) 'Transcriptional signatures of schizophrenia in hiPSC-derived NPCs and neurons are concordant with post-mortem adult brains', *Nature Communications*. Springer US, 8(1). doi: 10.1038/s41467-017-

02330-5.

Hoffman, M. M. *et al.* (2012) ‘Unsupervised pattern discovery in human chromatin structure through genomic segmentation’, *Nature Methods*, 9(5), pp. 473–476. doi: 10.1038/nmeth.1937.

van Holde, K. E. (1989) *Chromatin*. Springer-Verlag New York Inc.

Holler, P. D. *et al.* (2000) ‘In vitro evolution of a T cell receptor with high affinity for peptide/MHC.’, *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), pp. 5387–5392. doi: 10.1073/pnas.080078297.

Holler, P. D. and Kranz, D. M. (2003) ‘Quantitative analysis of the contribution of TCR/pepMHC affinity and CD8 to T cell activation’, *Immunity*, 18(2), pp. 255–264. doi: 10.1016/S1074-7613(03)00019-0.

Hozumi, N. and Tonegawa, S. (1976) ‘Evidence for somatic rearrangement of immunoglobulin genes coding for variable and constant regions’, *Proceedings of the National Academy of Sciences of the United States of America*, 73(10), pp. 3628–3632. doi: 10.1073/pnas.73.10.3628.

Hsieh, C. L. *et al.* (2014) ‘Enhancer RNAs participate in androgen receptor-driven looping that selectively enhances gene activation’, *Proceedings of the National Academy of Sciences of the United States of America*, 111(20), pp. 7319–7324. doi: 10.1073/pnas.1324151111.

Hug, C. B. *et al.* (2017) ‘Chromatin Architecture Emerges during Zygotic Genome Activation Independent of Transcription’, *Cell*. Elsevier, 169(2), pp. 216-228.e19. doi: 10.1016/j.cell.2017.03.024.

Hughes, M. M. *et al.* (2003) ‘T cell receptor CDR3 loop length repertoire is determined primarily by features of the V(D)J recombination reaction’, *European Journal of Immunology*, 33(6), pp. 1568–1575. doi: 10.1002/eji.200323961.

Imakaev, M. *et al.* (2012) ‘Iterative correction of Hi-C data reveals hallmarks of chromosome organization’, *Nature methods*, 9(10), pp. 999–1003. doi: 10.1038/nmeth.2148.

Jiang, Y. *et al.* (2017) ‘The methyltransferase SETDB1 regulates a large neuron-specific topological chromatin domain’, *Nature Genetics*. Nature Publishing Group, 49(8), pp. 1239–1250. doi: 10.1038/ng.3906.

Jost, D. *et al.* (2014) ‘Modeling epigenome folding: Formation and dynamics of topologically associated chromatin domains’, *Nucleic Acids Research*, 42(15), pp. 9553–9561. doi: 10.1093/nar/gku698.

Jost, K. L., Bertulat, B. and Cardoso, M. C. (2012) ‘Heterochromatin and gene positioning: Inside, outside, any side?’, *Chromosoma*, 121(6), pp. 555–563. doi: 10.1007/s00412-012-0389-2.

Jung, Y. H. *et al.* (2017) ‘Chromatin States in Mouse Sperm Correlate with Embryonic and Adult Regulatory Landscapes’, *Cell Reports*. ElsevierCompany., 18(6), pp. 1366–1382. doi: 10.1016/j.celrep.2017.01.034.

Jurtz, V. *et al.* (2017) ‘NetMHCpan-4.0: Improved Peptide–MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data’, *The Journal of Immunology*, 199(9), pp. 3360–3368. doi: 10.4049/jimmunol.1700893.

Kaaij, L. J. T. *et al.* (2018) 'Systemic Loss and Gain of Chromatin Architecture throughout Zebrafish Development', *Cell Reports*, 24(1), pp. 1-10.e4. doi: 10.1016/j.celrep.2018.06.003.

Kagey, M. H. *et al.* (2010) 'Mediator and cohesin connect gene expression and chromatin architecture', *Nature*. Nature Publishing Group, 467(7314), pp. 430–435. doi: 10.1038/nature09380.

Kahn, R. S. *et al.* (2015) 'Schizophrenia', *Nature Reviews Disease Primers*, 1(November). doi: 10.1038/nrdp.2015.67.

Kalhor, R. *et al.* (2012) 'Genome architectures revealed by tethered chromosome conformation capture and population-based modeling.', *Nature biotechnology*. Nature Publishing Group, 30(1), pp. 90–8. doi: 10.1038/nbt.2057.

Kane, J. M. and Correll, C. U. (2010) 'Past and present progress in the pharmacologic treatment of schizophrenia', *Journal of Clinical Psychiatry*, 71(9), pp. 1115–1124. doi: 10.4088/JCP.10r06264yel.

Kang, Y. *et al.* (2017) 'Erythroid specific activator GATA-1-dependent interactions between CTCF sites around the β-globin locus', *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*. Elsevier B.V., 1860(4), pp. 416–426. doi: 10.1016/j.bbagrm.2017.01.013.

Kastritis, P. L. and Bonvin, A. M. J. J. (2010) 'Are Scoring Functions in Protein - Protein Docking Ready To Predict Interactomes ? Clues from a Novel Binding Affinity Benchmark research articles', *Journal of Proteome Research*, pp. 2216–2225.

Kaustov, L. *et al.* (2011) 'Recognition and Specificity Determinants of the Human Cbx Chromodomains', *Journal of Biological Chemistry*, 286(1), pp. 521–529. doi: 10.1074/jbc.m110.191411.

Khanna, N., Hu, Y. and Belmont, A. S. (2014) 'HSP70 transgene directed motion to nuclear speckles facilitates heat shock activation', *Current Biology*. Elsevier Ltd, 24(10), pp. 1138–1144. doi: 10.1016/j.cub.2014.03.053.

Kim, Y. *et al.* (2019) 'Human cohesin compacts DNA by loop extrusion', *Science*, 366(6471), pp. 1345–1349. doi: 10.1126/science.aaz4475.

Knight, P. A. and Ruiz, D. (2013) 'A fast algorithm for matrix balancing', *IMA Journal of Numerical Analysis*, 33(3), pp. 1029–1047. doi: 10.1093/imanum/drs019.

Koch, U. and Radtke, F. (2011) 'Mechanisms of T Cell Development and Transformation', *Annual Review of Cell and Developmental Biology*, 27(1), pp. 539–562. doi: 10.1146/annurev-cellbio-092910-154008.

Konermann, S. *et al.* (2015) 'Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex', *Nature*, 517(7536), pp. 583–8. doi: 10.1038/nature14136.Genome-scale.

Kornberg, R. (1974) 'Chromatin Structure : A Repeating Unit of Histones and DNA Chromatin structure is based on a repeating unit of eight', *Science*, 184, pp. 868–871.

Kula, T. *et al.* (2019) 'T-Scan: A Genome-wide Method for the Systematic Discovery of T Cell Epitopes', *Cell*. Elsevier Inc., 178(4), pp. 1016-1028.e13. doi:

10.1016/j.cell.2019.07.009.

Kundaje, A. *et al.* (2015) ‘Integrative analysis of 111 reference human epigenomes’, *Nature*, 518(7539), pp. 317–330. doi: 10.1038/nature14248.

Kustatscher, G., Grabowski, P. and Rappsilber, J. (2017) ‘Pervasive coexpression of spatially proximal genes is buffered at the protein level’, *Molecular Systems Biology*, 13(8), p. 937. doi: 10.15252/msb.20177548.

de la Torre-Ubieta, L. *et al.* (2018) ‘The Dynamic Landscape of Open Chromatin during Human Cortical Neurogenesis’, *Cell*. Elsevier Inc., 172(1–2), pp. 289-304.e18. doi: 10.1016/j.cell.2017.12.014.

Lai, F. *et al.* (2013) ‘Activating RNAs associate with Mediator to enhance chromatin architecture and transcription’, *Nature*. Nature Publishing Group, 494(7438), pp. 497–501. doi: 10.1038/nature11884.

Lajoie, B. R. *et al.* (2009) ‘My5C: Web tools for chromosome conformation capture studies’, *Nature Methods*. Nature Publishing Group, 6(10), pp. 690–691. doi: 10.1038/nmeth1009-690.

Lajoie, B. R., Dekker, J. and Kaplan, N. (2015) ‘The Hitchhiker’s guide to Hi-C analysis: Practical guidelines’, *Methods*, 72, pp. 65–75. doi: 10.1016/j.ymeth.2014.10.031.

Lally, J. and MacCabe, J. H. (2015) ‘Antipsychotic medication in schizophrenia: A review’, *British Medical Bulletin*, 114(1), pp. 169–179. doi: 10.1093/bmb/ldv017.

Lander, E. S. *et al.* (2001) ‘Initial sequencing and analysis of the human genome’, *Nature*, 412(6846), pp. 565–566. doi: 10.1038/35087627.

Langmead, B. *et al.* (2009) ‘Ultrafast and memory-efficient alignment of short DNA sequences to the human genome’, *Genome Biology*, 10(3). doi: 10.1186/gb-2009-10-3-r25.

Lanzarotti, E., Marcatili, P. and Nielsen, M. (2018) ‘Identification of the cognate peptide-MHC target of T cell receptors using molecular modeling and force field scoring’, *Molecular Immunology*. Elsevier, 94(December 2017), pp. 91–97. doi: 10.1016/j.molimm.2017.12.019.

Lanzarotti, E., Marcatili, P. and Nielsen, M. (2019) ‘T-Cell Receptor Cognate Target Prediction Based on Paired α and β Chain Sequence and Structural CDR Loop Similarities’, *Frontiers in Immunology*, 10(August), pp. 1–10. doi: 10.3389/fimmu.2019.02080.

Larson, A. G. *et al.* (2017) ‘Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin’, *Nature*. Nature Publishing Group, 547(7662), pp. 236–240. doi: 10.1038/nature22822.

Leaver-Fay, A. *et al.* (2011) ‘ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules.’, *Methods in enzymology*, 487(11), pp. 545–74. doi: 10.1016/B978-0-12-381270-4.00019-6.

Leaver-Fay, A., Snoeyink, J. and Kuhlman, B. (2008) ‘On-the-fly rotamer pair energy evaluation in protein design’, *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 4983 LNBI, pp. 343–354. doi: 10.1007/978-3-540-79450-9_32.

Lefranc, M. (2011) ‘TRB (T cell receptor beta)’, *Atlas of Genetics and*

Cytogenetics in Oncology and Haematology, 4(3), pp. 119–120. doi: 10.4267/2042/37641.

Leibler, L. (1980) ‘Theory of Microphase Separation in Block Copolymers’, *Macromolecules*, pp. 1602–1617. doi: 10.1021/ma60078a047.

Lettice, L. A. *et al.* (2003) ‘A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly’, *Human Molecular Genetics*, 12(14), pp. 1725–1735. doi: 10.1093/hmg/ddg180.

Li, G. *et al.* (2012) ‘Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation’, *Cell*. Elsevier Inc., 148(1–2), pp. 84–98. doi: 10.1016/j.cell.2011.12.014.

Li, Y. *et al.* (2005) ‘Directed evolution of human T-cell receptors with picomolar affinities by phage display.’, *Nature biotechnology*, 23(3), pp. 349–354. doi: 10.1038/nbt1070.

Li, Z. *et al.* (2017) ‘A potent Cas9-derived gene activator for plant and mammalian cells’, *Nature Plants*. Springer US, 3(12), pp. 930–936. doi: 10.1038/s41477-017-0046-0.

Liang, C. and Stillman, B. (1997) ‘Persistent initiation of DNA replication and chromatin-bound MCM proteins during the cell cycle in cdc6 mutants’, *Genes and Development*, 11(24), pp. 3375–3386. doi: 10.1101/gad.11.24.3375.

Liao, Y., Smyth, G. K. and Shi, W. (2014) ‘FeatureCounts: An efficient general purpose program for assigning sequence reads to genomic features’, *Bioinformatics*, 30(7), pp. 923–930. doi: 10.1093/bioinformatics/btt656.

Lieberman-Aiden, E. *et al.* (2009) 'Comprehensive mapping of long-range interactions reveals folding principles of the human genome.', *Science (New York, N.Y.)*, 326(5950), pp. 289–93. doi: 10.1126/science.1181369.

Linette, G. P. *et al.* (2013) 'Cardiovascular toxicity and titin cross-reactivity of affinity-enhanced T cells in myeloma and melanoma.', *Blood*, 122(6), pp. 863–71. doi: 10.1182/blood-2013-03-490565.

Lomvardas, S. *et al.* (2006) 'Interchromosomal Interactions and Olfactory Receptor Choice', *Cell*, 126(2), pp. 403–413. doi: 10.1016/j.cell.2006.06.035.

Lorch, Y., LaPointe, J. W. and Kornberg, R. D. (1987) 'Nucleosomes inhibit the initiation of transcription but allow chain elongation with the displacement of histones', *Cell*, 49(2), pp. 203–210. doi: 10.1016/0092-8674(87)90561-7.

Lupiáñez, D. G. *et al.* (2015) 'Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions', *Cell*, 161(5), pp. 1012–1025. doi: 10.1016/j.cell.2015.04.004.

Ma, P. *et al.* (2014) 'The Ubiquitin Ligase RNF220 Enhances Canonical Wnt Signaling through USP7-Mediated Deubiquitination of β-Catenin', *Molecular and Cellular Biology*, 34(23), pp. 4355–4366. doi: 10.1128/mcb.00731-14.

MacDonald, M. L. *et al.* (2012) 'Biochemical fractionation and stable isotope dilution liquid chromatography-mass spectrometry for targeted and microdomain-specific protein quantification in human postmortem brain tissue', *Molecular and Cellular Proteomics*, 11(12), pp. 1670–1681. doi: 10.1074/mcp.M112.021766.

Macdonald, W. a. *et al.* (2009) 'T Cell Allorecognition via Molecular Mimicry',

Immunity. Elsevier Ltd, 31(6), pp. 897–908. doi: 10.1016/j.immuni.2009.09.025.

MacPherson, Q., Beltran, B. and Spakowitz, A. J. (2018) 'Bottom–up modeling of chromatin segregation due to epigenetic modifications', *Proceedings of the National Academy of Sciences of the United States of America*, 115(50), pp. 12739–12744. doi: 10.1073/pnas.1812268115.

Maeder, M. L. *et al.* (2013) 'CRISPR RNA-guided activation of endogenous human genes', *Nature Methods*. Nature Publishing Group, 10(10), pp. 977–979. doi: 10.1038/nmeth.2598.

Makrantoni, V. and Marston, A. L. (2018) 'Cohesin and chromosome segregation', *Current Biology*. Elsevier, 28(12), pp. R688–R693. doi: 10.1016/j.cub.2018.05.019.

Malecek, K. *et al.* (2014) 'Specific Increase in Potency via Structure-Based Design of a TCR', *The Journal of Immunology*, 193, pp. 2587–2599. doi: 10.4049/jimmunol.1302344.

Mandell, D. J. and Kortemme, T. (2009) 'Backbone flexibility in computational protein design', *Current Opinion in Biotechnology*, 20(4), pp. 420–428. doi: 10.1016/j.copbio.2009.07.006.

Marco Biasini (2015) 'pv: v1.81'. doi: 10.5281/zenodo.20980.

Marko, J. F. (1993) 'Microphase Separation of Block Copolymer Rings', *Macromolecules*, pp. 1442–1444. doi: 10.1021/ma00058a038.

Marko, J. F. and Siggia, E. D. (1997) 'Polymer models of meiotic and mitotic chromosomes', *Molecular Biology of the Cell*, 8(11), pp. 2217–2231. doi:

10.1091/mbc.8.11.2217.

Marko, J. F. and Witten, T. A. (1991) ‘Phase separation in a grafted polymer layer’, *Physical Review Letters*, pp. 1541–1544. doi: 10.1103/PhysRevLett.66.1541.

Marshall, W. F. *et al.* (1997) ‘Interphase chromosomes undergo constrained diffusional motion in living cells’, *Current Biology*, 7(12), pp. 930–939. doi: 10.1016/S0960-9822(06)00412-X.

Mason, D. (1998) ‘A very high level of crossreactivity is an essential feature of the T-cell receptor’, *Immunology Today*, 19(9), pp. 395–404. doi: 10.1016/S0167-5699(98)01299-7.

Matsen, M. W. and Schick, M. (1994) ‘Stable and Unstable Phases of a Linear Multiblock Copolymer Melt’, *Macromolecules*, 27(24), pp. 7157–7163. doi: 10.1021/ma00102a025.

Maynard, J. *et al.* (2005) ‘Structure of an autoimmune T cell receptor complexed with class II peptide-MHC: Insights into MHC bias and antigen specificity’, *Immunity*, 22, pp. 81–92. doi: 10.1016/j.immuni.2004.11.015.

McClellan, J. M., Susser, E. and King, M. C. (2007) ‘Schizophrenia: A common disease caused by multiple rare alleles’, *British Journal of Psychiatry*, 190(MAR.), pp. 194–199. doi: 10.1192/bjp.bp.106.025585.

McGrath, J. *et al.* (2008) ‘Schizophrenia: A concise overview of incidence, prevalence, and mortality’, *Epidemiologic Reviews*, 30(1), pp. 67–76. doi: 10.1093/epirev/mxn001.

Mckeithan, T. W. (1995) ‘Kinetic proofreading in T-cell receptor signal transduction’, *Proceedings of the National Academy of Sciences of the United States of America*, 92(11), pp. 5042–5046. doi: 10.1073/pnas.92.11.5042.

McKenna, A. *et al.* (2010) ‘The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data’, *Genome Research*, 20, pp. 1297–1303. doi: 10.1101/gr.107524.110.20.

Merwe, P. A. Van Der (2000) ‘Surface plasmon resonance’, in *Protein-Ligand Interactions: Hydrodynamics and Calorimetry*, pp. 137–170. doi: 10.1007/978-1-60761-670-2.

Meselson, M. and Stahl, F. W. (1958) ‘The Replication of DNA in Escherichia Coli’, *Proceedings of the National Academy of Sciences*, 44, pp. 671–682.

Michieletto, D., Orlandini, E. and Marenduzzo, D. (2016) ‘Polymer model with epigenetic recoloring reveals a pathway for the de novo establishment and 3D organization of chromatin domains’, *Physical Review X*, 6(4), pp. 1–15. doi: 10.1103/PhysRevX.6.041047.

Mifsud, B. *et al.* (2015) ‘Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C’, *Nature Genetics*. Nature Publishing Group, 47(6), pp. 598–606. doi: 10.1038/ng.3286.

Mifsud, B. *et al.* (2017) ‘GOTHiC, a probabilistic model to resolve complex biases and to identify real interactions in Hi-C data’, *PLoS ONE*, 12(4), pp. 1–15. doi: 10.1371/journal.pone.0174744.

Miller, P. J. *et al.* (2007) ‘Single MHC Mutation Eliminates Enthalpy Associated

with T Cell Receptor Binding', *Journal of Molecular Biology*, 373(2), pp. 315–327. doi: 10.1016/j.jmb.2007.07.028.

Misteli, T., Cáceres, J. F. and Spector, D. L. (1997) 'The dynamics of pre-mRNA splicing in living cells', *Nature*, 387(May), pp. 523–527. Available at: http://www.nature.com/nature/journal/v387/n6632/pdf/387523a0.pdf.

Mitchell, A. C. *et al.* (2014) 'The genome in three dimensions: A new frontier in human brain research', *Biological Psychiatry*. Elsevier, 75(12), pp. 961–969. doi: 10.1016/j.biopsych.2013.07.015.

Miyamoto, S. *et al.* (2012) 'Pharmacological treatment of schizophrenia: A critical review of the pharmacology and clinical effects of current and future therapeutic agents', *Molecular Psychiatry*. Nature Publishing Group, 17(12), pp. 1206–1227. doi: 10.1038/mp.2012.47.

Monks, C. R. F. *et al.* (1998) 'Three-dimensional segregation of supramolecular activation clusters in T cells', *Nature*, 340(1986), pp. 764–766.

Morgan, R. a *et al.* (2013) 'Cancer regression and neurological toxicity following anti-MAGE-A3 TCR gene therapy.', *Journal of immunotherapy (Hagerstown, Md. : 1997)*, 36(2), pp. 133–51. doi: 10.1097/CJI.0b013e3182829903.

Morse, R. H. and Simpson, R. T. (1988) 'DNA in the nucleosome', *Cell*, 54(3), pp. 285–287. doi: 10.1016/0092-8674(88)90190-0.

Moult, J. *et al.* (1995) 'A large-scale experiment to assess protein structure prediction methods', *Proteins: Structure, Function, and Bioinformatics*, 23(3), pp. ii–iv. doi: 10.1002/prot.340230303.

Nagano, T. *et al.* (2013) ‘Single-cell Hi-C reveals cell-to-cell variability in chromosome structure’, *Nature*. Nature Publishing Group, 502(7469), pp. 59–64. doi: 10.1038/nature12593.

Nagano, T. *et al.* (2017) ‘Cell-cycle dynamics of chromosomal organization at single-cell resolution’, *Nature*. Nature Publishing Group, 547(7661), pp. 61–67. doi: 10.1038/nature23001.

Nagashima, R. *et al.* (2019) ‘Single nucleosome imaging reveals loose genome chromatin networks via active RNA polymerase II’, *Journal of Cell Biology*, 218(5), pp. 1511–1530. doi: 10.1083/jcb.201811090.

Nasmyth, K. (2001) ‘Disseminating the Genome: Joining, Resolving, and Separating Sister Chromatids During Mitosis and Meiosis’, *Annual Review of Genetics*, 35(1), pp. 673–745. doi: 10.1146/annurev.genet.35.102401.091334.

Naumova, N. *et al.* (2013) ‘Organization of the mitotic chromosome’, *Science*, 342(6161), pp. 948–953. doi: 10.1126/science.1236083.

Newell, E. W. *et al.* (2011) ‘Structural basis of specificity and cross-reactivity in T cell receptors specific for cytochrome c-I-E(k).’, *Journal of immunology (Baltimore, Md. : 1950)*, 186(10), pp. 5823–32. doi: 10.4049/jimmunol.1100197.

Nica, A. C. and Dermitzakis, E. T. (2013) ‘Expression quantitative trait loci: Present and future’, *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1620). doi: 10.1098/rstb.2012.0362.

Nichols, M. H. and Corces, V. G. (2015) ‘A CTCF Code for 3D Genome Architecture’, *Cell*. Elsevier Inc., 162(4), pp. 703–705. doi:

10.1016/j.cell.2015.07.053.

Nicolae, D. L. *et al.* (2010) 'Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS', *PLoS Genetics*, 6(4). doi: 10.1371/journal.pgen.1000888.

Nielsen, M. *et al.* (2005) 'The role of the proteasome in generating cytotoxic T-cell epitopes: Insights obtained from improved predictions of proteasomal cleavage', *Immunogenetics*, 57(1–2), pp. 33–41. doi: 10.1007/s00251-005-0781-7.

Nir, G. *et al.* (2018) *Walking along chromosomes with super-resolution imaging, contact maps, and integrative modeling*, *PLoS Genetics*. doi: 10.1371/journal.pgen.1007872.

Nora, E. P. *et al.* (2012) 'Spatial partitioning of the regulatory landscape of the X-inactivation centre', *Nature*, 485(7398), pp. 381–385. doi: 10.1038/nature11049.

Nora, E. P. *et al.* (2017) 'Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization', *Cell*, 169(5), pp. 930-944.e22. doi: 10.1016/j.cell.2017.05.004.

Nuebler, J. *et al.* (2018) 'Chromatin organization by an interplay of loop extrusion and compartmental segregation', *Proceedings of the National Academy of Sciences of the United States of America*, 115(29), pp. E6697–E6706. doi: 10.1073/pnas.1717730115.

O'Donnell, T. J. *et al.* (2018) 'MHCflurry: Open-Source Class I MHC Binding Affinity Prediction', *Cell Systems*. Elsevier Inc., 7(1), pp. 129-132.e4. doi:

10.1016/j.cels.2018.05.014.

O'dushlaine, C. *et al.* (2015) 'Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways', *Nature Neuroscience*, 18(2), pp. 199–209. doi: 10.1038/nn.3922.

Oates, J. and Jakobsen, B. K. (2013) 'ImmTACs', *OncoImmunology*, 2(2), p. e22891. doi: 10.4161/onci.22891.

Obenaus, M. *et al.* (2015) 'Identification of human T-cell receptors with optimal affinity to cancer antigens using antigen-negative humanized mice', *Nature Biotechnology*, (August 2014). doi: 10.1038/nbt.3147.

Ogishi, M. and Yotsuyanagi, H. (2019) 'Quantitative prediction of the landscape of T cell epitope immunogenicity in sequence space', *Frontiers in Immunology*, 10(MAR). doi: 10.3389/fimmu.2019.00827.

Osborne, C. S. *et al.* (2004) 'Active genes dynamically colocalize to shared sites of ongoing transcription', *Nature Genetics*, 36(10), pp. 1065–1071. doi: 10.1038/ng1423.

Owen, J. *et al.* (2013) *Kuby Immunology Seventh Edition*.

Pardiñas, A. F. *et al.* (2018) 'Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection', *Nature Genetics*, 50(3), pp. 381–389. doi: 10.1038/s41588-018-0059-2.

Pękowska, A. *et al.* (2018) 'Gain of CTCF-Anchored Chromatin Loops Marks the Exit from Naive Pluripotency', *Cell Systems*, 7(5), pp. 482-495.e10. doi: 10.1016/j.cels.2018.09.003.

Perez-Pinera, P. *et al.* (2013) 'RNA-guided gene activation by CRISPR-Cas9-based transcription factors', *Nature Methods*, 10(10), pp. 973–976. doi: 10.1038/nmeth.2600.

Pierce, B. G. *et al.* (2014) 'Computational design of the affinity and specificity of a therapeutic T cell receptor.', *PLoS computational biology*, 10(2), p. e1003478. doi: 10.1371/journal.pcbi.1003478.

Pierce, B. G. and Weng, Z. (2013) 'A flexible docking approach for prediction of T cell receptor-peptide-MHC complexes', *Protein Science*, 22, pp. 35–46. doi: 10.1002/pro.2181.

Di Pierro, M. *et al.* (2016) 'Transferable model for chromosome architecture', *Proceedings of the National Academy of Sciences of the United States of America*, 113(43), pp. 12168–12173. doi: 10.1073/pnas.1613607113.

Plys, A. J. *et al.* (2019) 'Phase separation of polycomb-repressive complex 1 is governed by a charged disordered region of CBX2', *Genes and Development*, 33(13–14), pp. 799–813. doi: 10.1101/gad.326488.119.

Puram, S. V. *et al.* (2011) 'A CaMKIIβ 2 signaling pathway at the centrosome regulates dendrite patterning in the brain', *Nature Neuroscience*. Nature Publishing Group, 14(8), pp. 973–983. doi: 10.1038/nn.2857.

Purcell, A. W., McCluskey, J. and Rossjohn, J. (2007) 'More than one reason to rethink the use of peptides in vaccine design', *Nature Reviews Drug Discovery*, 6(5), pp. 404–414. doi: 10.1038/nrd2224.

Purcell, S. M. *et al.* (2009) 'Common polygenic variation contributes to risk of

schizophrenia and bipolar disorder', *Nature*, 460(7256), pp. 748–752. doi: 10.1038/nature08185.

Quinodoz, S. A. *et al.* (2018) 'Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus', *Cell*. Elsevier Inc., 174(3), pp. 744-757.e24. doi: 10.1016/j.cell.2018.05.024.

Radwan, J. *et al.* (2020) 'Advances in the Evolutionary Understanding of MHC Polymorphism', *Trends in Genetics*, 36(4). doi: 10.1016/j.tig.2020.01.008.

Rajagopal, N. *et al.* (2013) 'RFECS: A Random-Forest Based Algorithm for Enhancer Identification from Chromatin State', *PLoS Computational Biology*, 9(3). doi: 10.1371/journal.pcbi.1002968.

Rajarajan, P. *et al.* (2016) 'Spatial genome organization and cognition', *Nature Reviews Neuroscience*. Nature Publishing Group, 17(11), pp. 681–691. doi: 10.1038/nrn.2016.124.

Rajarajan, P. *et al.* (2018) 'Neuron-specific signatures in the chromosomal connectome associated with schizophrenia risk', *Science*, 362(6420). doi: 10.1126/science.aat4311.

Ramani, V. *et al.* (2017) 'Massively multiplex single-cell Hi-C', *Nature Methods*, 14(3), pp. 263–266. doi: 10.1038/nmeth.4155.

Rao, S. S. P. *et al.* (2017) 'Cohesin Loss Eliminates All Loop Domains', *Cell*, 171, pp. 305–320. doi: 10.1016/j.cell.2017.09.026.

Rao, S. S. P. S. P. *et al.* (2014) 'A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping', *Cell*. Elsevier Inc., 159(7),

pp. 1665–1680. doi: 10.1016/j.cell.2014.11.021.

Reiser, J.-B. *et al.* (2003) ‘CDR3 loop flexibility contributes to the degeneracy of TCR recognition.’, *Nature immunology*, 4(3), pp. 241–7. doi: 10.1038/ni891.

Reiser, J. B. *et al.* (2002) ‘A T cell receptor CDR3β loop undergoes conformational changes of unprecedented magnitude upon binding to a peptide/MHC class I complex’, *Immunity*, 16(3), pp. 345–354. doi: 10.1016/S1074-7613(02)00288-1.

Restifo, N. P., Dudley, M. E. and Rosenberg, S. a. (2012) ‘Adoptive immunotherapy for cancer: harnessing the T cell response’, *Nature Reviews Immunology*. Nature Publishing Group, 12(4), pp. 269–281. doi: 10.1038/nri3191.

Rhie, S. K. *et al.* (2018) ‘Using 3D epigenomic maps of primary olfactory neuronal cells from living individuals to understand gene regulation’, *Science Advances*, 4(12). doi: 10.1126/sciadv.aav8550.

Rieder, C. L. and Khodjakov, A. (2003) ‘Mitosis through the microscope: Advances in seeing inside live dividing cells’, *Science*, 300(5616), pp. 91–96. doi: 10.1126/science.1082177.

Riggs, A. D. (1990) ‘DNA methylation and late replication probably aid cell memory, and type I DNA reeling could aid chromosome folding and enhancer function.’, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 326(1235), pp. 285–297. doi: 10.1098/rstb.1990.0012.

Riley, T. P. *et al.* (2016) ‘A generalized framework for computational design and mutational scanning of T-cell receptor binding interfaces’, *Protein Engineering*

Design and Selection, 29(12), pp. 1–12. doi: 10.1093/protein/gzw050.

Riley, T. P. *et al.* (2018) 'T cell receptor cross-reactivity expanded by dramatic peptide–MHC adaptability', *Nature Chemical Biology*. Springer US, 14(10), pp. 934–942. doi: 10.1038/s41589-018-0130-4.

Riley, T. P. *et al.* (2019) 'Structure based prediction of neoantigen immunogenicity', *Frontiers in Immunology*, 10(AUG). doi: 10.3389/fimmu.2019.02047.

Riley, T. P. and Baker, B. M. (2018) 'The intersection of affinity and specificity in the development and optimization of T cell receptor based therapeutics', *Seminars in Cell and Developmental Biology*. Elsevier Ltd, 84, pp. 30–41. doi: 10.1016/j.semcdb.2017.10.017.

Ripke, S. *et al.* (2014) 'Biological insights from 108 schizophrenia-associated genetic loci', *Nature*, 511(7510), pp. 421–427. doi: 10.1038/nature13595.

Robinson, J. *et al.* (2020) 'IPD-IMGT/HLA Database', *Nucleic acids research*. Oxford University Press, 48(D1), pp. D948–D955. doi: 10.1093/nar/gkz950.

Rossjohn, J. *et al.* (2015) 'T Cell Antigen Receptor Recognition of Antigen-Presenting Molecules', *Annual Review of Immunology*, 33(1), pp. 169–200. doi: 10.1146/annurev-immunol-032414-112334.

Roussos, P. *et al.* (2014) 'A Role for Noncoding Variation in Schizophrenia', *Cell Reports*. The Authors, 9(4), pp. 1417–1429. doi: 10.1016/j.celrep.2014.10.015.

Rowley, M. J. *et al.* (2017) 'Evolutionarily Conserved Principles Predict 3D Chromatin Organization', *Molecular Cell*. Elsevier Inc., 67(5), pp. 837-852.e7.

doi: 10.1016/j.molcel.2017.07.022.

Rowley, M. J. and Corces, V. G. (2018) ‘Organizational principles of 3D genome architecture’, *Nature Reviews Genetics*. Springer US, 19(12), pp. 789–800. doi: 10.1038/s41576-018-0060-8.

Rubinstein, M. and Colby, R. H. (2003) *Polymer Physics*. doi: 10.1145/2024288.2024304.

Sanborn, A. L. *et al.* (2015) ‘Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes’, *Proceedings of the National Academy of Sciences*, 112(47), pp. E6456–E6465. doi: 10.1073/pnas.1518552112.

Sanyal, A. *et al.* (2012) ‘The long-range interaction landscape of gene promoters.’, *Nature*. Nature Publishing Group, 489(7414), pp. 109–13. doi: 10.1038/nature11279.

Saurin, A. J. *et al.* (1998) ‘The Human Polycomb Group Complex Associates with Pericentromeric Heterochromatin to Form a Novel Nuclear Domain’, *The Journal of Cell Biology*, 142(4), pp. 887–898. doi: 10.1083/jcb.142.4.887.

Schmitt, A. D. *et al.* (2016) ‘A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome’, *Cell Rep*. The Authors, 17(8), pp. 2042–2059. doi: 10.1016/j.celrep.2016.10.061.

Schneidman-Duhovny, D. *et al.* (2018) ‘Predicting CD4 T-cell epitopes based on antigen cleavage, MHCII presentation, and TCR recognition’, *PLoS ONE*, 13(11), pp. 1–22. doi: 10.1371/journal.pone.0206654.

Scott, D. R. *et al.* (2011) ‘Disparate degrees of hypervariable loop flexibility control T-cell receptor cross-reactivity, specificity, and binding mechanism.’, *Journal of molecular biology*. Elsevier Ltd, 414(3), pp. 385–400. doi: 10.1016/j.jmb.2011.10.006.

Servant, N. *et al.* (2012) ‘HiTC: Exploration of high-throughput “C” experiments’, *Bioinformatics*, 28(21), pp. 2843–2844. doi: 10.1093/bioinformatics/bts521.

Servant, N. *et al.* (2015) ‘HiC-Pro: An optimized and flexible pipeline for Hi-C data processing’, *Genome Biology*, 16(1), pp. 1–11. doi: 10.1186/s13059-015-0831-x.

Sethi, D. K. *et al.* (2013) ‘Crossreactivity of a human autoimmune TCR is dominated by a single TCR loop.’, *Nature communications*. Nature Publishing Group, 4, p. 2623. doi: 10.1038/ncomms3623.

Sewell, A. K. (2012) ‘Why must T cells be cross-reactive?’, *Nature Publishing Group*. Nature Publishing Group, 12(9), pp. 669–677. doi: 10.1038/nri3279.

Sexton, T. *et al.* (2012) ‘Three-dimensional folding and functional organization principles of the Drosophila genome’, *Cell*, 148(3), pp. 458–472. doi: 10.1016/j.cell.2012.01.010.

Shapovalov, M. V. and Dunbrack, R. L. (2011) ‘A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions’, *Structure*. Elsevier Ltd, 19(6), pp. 844–858. doi: 10.1016/j.str.2011.03.019.

Shi, G. *et al.* (2018) ‘Interphase human chromosome exhibits out of equilibrium

glassy dynamics', *Nature Communications*. Springer US, 9(1). doi: 10.1038/s41467-018-05606-6.

Shinkai, S. *et al.* (2016) 'Dynamic Nucleosome Movement Provides Structural Information of Topological Chromatin Domains in Living Human Cells', *PLoS Computational Biology*, 12(10), pp. 1–16. doi: 10.1371/journal.pcbi.1005136.

Shopland, L. S. *et al.* (2006) 'Folding and organization of a contiguous chromosome region according to the gene distribution pattern in primary genomic sequence', *Journal of Cell Biology*, 174(1), pp. 27–38. doi: 10.1083/jcb.200603083.

Simonis, M. *et al.* (2006) 'Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C)', *Nature Genetics*, 38(11), pp. 1348–1354. doi: 10.1038/ng1896.

Singh, N. K. *et al.* (2017) 'Emerging Concepts in TCR Specificity: Rationalizing and (Maybe) Predicting Outcomes', *The Journal of Immunology*, 199(7), pp. 2203–2213. doi: 10.4049/jimmunol.1700744.

Sirin, S. *et al.* (2015) 'AB-Bind: Antibody binding mutational database for computational affinity predictions', *Protein Science*, 22(6), p. n/a-n/a. doi: 10.1002/pro.2829.

Skene, N. G. *et al.* (2018) 'Genetic identification of brain cell types underlying schizophrenia', *Nature Genetics*. Springer US, 50(6), pp. 825–833. doi: 10.1038/s41588-018-0129-5.

Slotkin, R. K. and Martienssen, R. (2007) 'Transposable elements and the

epigenetic regulation of the genome', *Nature Reviews Genetics*, 8(4), pp. 272–285. doi: 10.1038/nrg2072.

Smith, C. C. *et al.* (2019) 'Machine-learning prediction of tumor antigen immunogenicity in the selection of therapeutic epitopes', *Cancer Immunology Research*, 7(10), pp. 1591–1604. doi: 10.1158/2326-6066.CIR-19-0155.

Song, M. *et al.* (2019) 'Mapping cis-regulatory chromatin contacts in neural cells links neuropsychiatric disorder risk variants to target genes', *Nature Genetics*. Springer US, 51(8), pp. 1252–1262. doi: 10.1038/s41588-019-0472-1.

Splinter, E. *et al.* (2006) 'CTCF mediates long-range chromatin looping and local histone modification in the β-globin locus', *Genes and Development*, 20(17), pp. 2349–2354. doi: 10.1101/gad.399506.

Squire, L. R. *et al.* (2008) *Fundamentals of Neuroscience Third Edition*, *Academic Press*. doi: 10.1201/9781315373034.

Srinivasan, M. *et al.* (2018) 'The Cohesin Ring Uses Its Hinge to Organize DNA Using Non-topological as well as Topological Mechanisms', *Cell*. Elsevier Inc., 173(6), pp. 1508-1519.e18. doi: 10.1016/j.cell.2018.04.015.

Starr, T. K., Jameson, S. C. and Hogquist, K. A. (2003) 'Positive and Negative Selection of T Cells', *Annual Review of Immunology*, 21(1), pp. 139–176. doi: 10.1146/annurev.immunol.21.120601.141107.

Stephens, A. D. *et al.* (2017) 'Chromatin and lamin A determine two different mechanical response regimes of the cell nucleus', *Molecular Biology of the Cell*, 28(14), pp. 1984–1996. doi: 10.1091/mbc.E16-09-0653.

Stephens, A. D. *et al.* (2018) ‘Chromatin histone modifications and rigidity affect nuclear morphology independent of lamins’, *Molecular Biology of the Cell*, 29(2), pp. 220–233. doi: 10.1091/mbc.E17-06-0410.

Stone, J. D. and Kranz, D. M. (2013) ‘Role of T cell receptor affinity in the efficacy and specificity of adoptive T cell therapies.’, *Frontiers in immunology*, 4(August), p. 244. doi: 10.3389/fimmu.2013.00244.

Strom, A. R. *et al.* (2017) ‘Phase separation drives heterochromatin domain formation’, *Nature*. Nature Publishing Group, 547(7662), pp. 241–245. doi: 10.1038/nature22989.

Strominger, L. *et al.* (1987) ‘Structure of the human class I histocompatibility antigen, HLA-A2’, *Nature*, 329(8), pp. 506–512. Available at: http://www.jimmunol.org/content/174/1/6.citation%5Cnhttp://www.jimmunol.org/content/174/1/6.citation.

Sullivan, P. F., Kendler, K. S. and Neale, M. C. (2003) ‘Schizophrenia as a Complex Trait’, *Archives of General Psychiatry*, 60(12), p. 1187. doi: 10.1001/archpsyc.60.12.1187.

Sun, H. Bin, Shen, J. and Yokota, H. (2000) ‘Size-dependent positioning of human chromosomes in interphase nuclei’, *Biophysical Journal*. Elsevier, 79(1), pp. 184–190. doi: 10.1016/S0006-3495(00)76282-5.

Symmons, O. *et al.* (2016) ‘The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances’, *Developmental Cell*, 39(5), pp. 529–543. doi: 10.1016/j.devcel.2016.10.015.

Szklarczyk, D. *et al.* (2015) ‘STRING v10: Protein-protein interaction networks, integrated over the tree of life’, *Nucleic Acids Research*, 43(D1), pp. D447–D452. doi: 10.1093/nar/gku1003.

Szklarczyk, D. *et al.* (2017) ‘The STRING database in 2017: Quality-controlled protein-protein association networks, made broadly accessible’, *Nucleic Acids Research*, 45(D1), pp. D362–D368. doi: 10.1093/nar/gkw937.

Takahashi, K. and Yamanaka, S. (2006) ‘Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors’, *Cell*, 126(4), pp. 663–676. doi: 10.1016/j.cell.2006.07.024.

Tang, Z. *et al.* (2015) ‘CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription’, *Cell*. Elsevier Inc., 163(7), pp. 1611–1627. doi: 10.1016/j.cell.2015.11.024.

Tatavosian, R. *et al.* (2019) ‘Nuclear condensates of the Polycomb protein chromobox 2 (CBX2) assemble through phase separation’, *Journal of Biological Chemistry*, 294(5), pp. 1451–1463. doi: 10.1074/jbc.RA118.006620.

TCW, J. *et al.* (2017) ‘An Efficient Platform for Astrocyte Differentiation from Human Induced Pluripotent Stem Cells’, *Stem Cell Reports*. ElsevierCompany., 9(2), pp. 600–614. doi: 10.1016/j.stemcr.2017.06.018.

Thakar, R., Gordon, G. and Csink, A. K. (2006) ‘Dynamics and anchoring of heterochromatic loci during development’, *Journal of Cell Science*, 119(20), pp. 4165–4175. doi: 10.1242/jcs.03183.

Therizolsa, P. *et al.* (2010) ‘Chromosome arm length and nuclear constraints

determine the dynamic relationship of yeast subtelomeres', *Proceedings of the National Academy of Sciences of the United States of America*, 107(5), pp. 2025–2030. doi: 10.1073/pnas.0914187107.

Thurman, R. E. *et al.* (2012) 'The accessible chromatin landscape of the human genome', *Nature*. Nature Publishing Group, 489(7414), pp. 75–82. doi: 10.1038/nature11232.

Tolhuis, B. *et al.* (2002) 'Looping and interaction between hypersensitive sites in the active β-globin locus', *Molecular Cell*, 10(6), pp. 1453–1465. doi: 10.1016/S1097-2765(02)00781-5.

Topol, A. *et al.* (2016) 'Dysregulation of miRNA-9 in a Subset of Schizophrenia Patient-Derived Neural Progenitor Cells', *Cell Reports*. ElsevierCompany., 15(5), pp. 1024–1036. doi: 10.1016/j.celrep.2016.03.090.

Topol, A., Tran, N. N. and Brennand, K. J. (2015) 'A guide to generating and using hiPSC derived NPCs for the study of neurological diseases', *Journal of Visualized Experiments*, (96), pp. 1–9. doi: 10.3791/52495.

Toseland, C. P. *et al.* (2005) 'AntiJen: a quantitative immunology database integrating functional, thermodynamic, kinetic, biophysical, and cellular data.', *Immunome research*, 1(1), p. 4. doi: 10.1186/1745-7580-1-4.

Trapnell, C. *et al.* (2012) 'Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks', *Nature Protocols*, 7(3), pp. 562–578. doi: 10.1038/nprot.2012.016.

Traxler, E. A. *et al.* (2016) 'A genome-editing strategy to treat β-

hemoglobinopathies that recapitulates a mutation associated with a benign genetic condition', *Nature Medicine*, 22(9), pp. 987–990. doi: 10.1038/nm.4170.

Tung, C.-W. *et al.* (2011) 'POPISK: T-cell reactivity prediction using support vector machines and string kernels', *BMC Bioinformatics*. BioMed Central Ltd, 12(1), p. 446. doi: 10.1186/1471-2105-12-446.

Tynan, F. E. *et al.* (2005) 'T cell receptor recognition of a "super-bulged" major histocompatibility complex class I-bound peptide', *Nature Immunology*, 6(11), pp. 1114–1122. doi: 10.1038/ni1257.

Varela-Rohena, A. *et al.* (2008) 'Control of HIV-1 immune escape by CD8 T cells expressing enhanced T-cell receptor.', *Nature medicine*, 14(12), pp. 1390–1395. doi: 10.1038/nm.1779.

Vietri Rudan, M. *et al.* (2015) 'Comparative Hi-C Reveals that CTCF Underlies Evolution of Chromosomal Domain Architecture', *Cell Reports*. The Authors, 10(8), pp. 1297–1309. doi: 10.1016/j.celrep.2015.02.004.

Vieux-Rochas, M. *et al.* (2015) 'Clustering of mammalian Hox genes with other H3K27me3 targets within an active nuclear domain', *Proceedings of the National Academy of Sciences of the United States of America*, 112(15), pp. 4672–4677. doi: 10.1073/pnas.1504783112.

Visel, A. *et al.* (2007) 'VISTA Enhancer Browser - A database of tissue-specific human enhancers', *Nucleic Acids Research*, 35(SUPPL. 1), pp. 88–92. doi: 10.1093/nar/gkl822.

Vita, R. *et al.* (2015) 'The immune epitope database (IEDB) 3.0', *Nucleic Acids*

Research, 43(D1), pp. D405–D412. doi: 10.1093/nar/gku938.

Vreven, T. *et al.* (2012) 'Prediction of protein-protein binding free energies.', *Protein science : a publication of the Protein Society*, 21(3), pp. 396–404. doi: 10.1002/pro.2027.

Wang, D. *et al.* (2018) 'Comprehensive functional genomic resource and integrative model for the human brain', *Science*, 362(6420). doi: 10.1126/science.aat8464.

Wang, J. *et al.* (2017) 'Single-cell gene expression analysis reveals regulators of distinct cell subpopulations among developing human neurons', *Genome Research*, 27(11), pp. 1783–1794. doi: 10.1101/gr.223313.117.

Wang, S. *et al.* (2016) 'Spatial organization of chromatin domains and compartments in single chromosomes', *Science*, 353(6299).

Watanabe, Y., Khodosevich, K. and Monyer, H. (2014) 'Dendrite Development Regulated by the Schizophrenia-Associated Gene FEZ1 Involves the Ubiquitin Proteasome System', *Cell Reports*. The Authors, 7(2), pp. 552–564. doi: 10.1016/j.celrep.2014.03.022.

Watson, James and Crick, F. (1953) 'Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid', *Nature*, 171, pp. 737–738.

Watson, J and Crick, F. H. . (1953) 'Genetic Implications of the Structure of Deoxyribonucleic Acid', *Nature*, 171, pp. 964–967. doi: 10.1007/BF00434693.

Weintraub, A. S. *et al.* (2017) 'YY1 Is a Structural Regulator of Enhancer-Promoter Loops', *Cell*. Elsevier Inc., 171(7), pp. 1573-1588.e28. doi:

10.1016/j.cell.2017.11.008.

Wiles, E. T. and Selker, E. U. (2017) ‘H3K27 methylation: a promiscuous repressive chromatin mark’, *Current Opinion in Genetics and Development*. Elsevier Ltd, 43, pp. 31–37. doi: 10.1016/j.gde.2016.11.001.

de Wit, E. *et al.* (2015) ‘CTCF Binding Polarity Determines Chromatin Looping’, *Molecular Cell*. Elsevier Inc., 60(4), pp. 676–684. doi: 10.1016/j.molcel.2015.09.023.

Won, H. *et al.* (2016) ‘Chromosome conformation elucidates regulatory relationships in developing human brain’, *Nature*. Nature Publishing Group, pp. 1–20. doi: 10.1038/nature19847.

Wooldridge, L. *et al.* (2012) ‘A single autoimmune T cell receptor recognizes more than a million different peptides.’, *The Journal of biological chemistry*, 287(2), pp. 1168–77. doi: 10.1074/jbc.M111.289488.

Wucherpfennig, K. W. *et al.* (1994) ‘Clonal Expansion and Persistence of Human T Cells Specific for an Immunodominant Myelin Basic Protein Peptide’, *Journal of immunology (Baltimore, Md. : 1950)*, 152, pp. 5581–5592.

Xiong, K. and Ma, J. (2018) ‘Revealing Hi-C subcompartments by imputing high-resolution inter-chromosomal chromatin interactions’.

Xiong, K. and Ma, J. (2019) ‘Revealing Hi-C subcompartments by imputing inter-chromosomal chromatin interactions’, *Nature Communications*. Springer US, 10(1), pp. 1–12. doi: 10.1038/s41467-019-12954-4.

Xue, L. C. *et al.* (2016) ‘PRODIGY: a web server for predicting the binding affinity

of protein–protein complexes', *Bioinformatics*, 32(August), p. btw514. doi: 10.1093/bioinformatics/btw514.

Yagi, T. (2012) 'Molecular codes for neuronal individuality and cell assembly in the brain', *Frontiers in Molecular Neuroscience*, 5(MARCH), pp. 1–31. doi: 10.3389/fnmol.2012.00045.

Yao, L. *et al.* (2014) 'Functional annotation of colon cancer risk SNPs', *Nature Communications*. Nature Publishing Group, 5. doi: 10.1038/ncomms6114.

Yu, G., Wang, L. G. and He, Q. Y. (2015) 'ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization', *Bioinformatics*, 31(14), pp. 2382–2383. doi: 10.1093/bioinformatics/btv145.

Zhang, Y. *et al.* (2012) 'Spatial organization of the mouse genome and its role in recurrent chromosomal translocations', *Cell*. Elsevier Inc., 148(5), pp. 908–921. doi: 10.1016/j.cell.2012.02.002.

Zhao, W. and Sher, X. (2018) 'Systematically benchmarking peptide-MHC binding predictors: From synthetic to naturally processed epitopes', *PLoS Computational Biology*, 14(11), pp. 1–28. doi: 10.1371/journal.pcbi.1006457.

Zidovska, A., Weitz, D. A. and Mitchison, T. J. (2013) 'Micron-scale coherence in interphase chromatin dynamics', *Proceedings of the National Academy of Sciences of the United States of America*, 110(39), pp. 15555–15560. doi: 10.1073/pnas.1220313110.

Zinkernagel, R. M. and Doherty, P. C. (1974) 'Restriction of in vitro T cell-mediated cytotoxicity in lymphocytic choriomeningitis within a syngeneic or

semiallogeneic system.', *Nature*, 248, pp. 701–710.

Zoete, V. *et al.* (2013) 'Structure-based, rational design of T cell receptors', *Frontiers in Immunology*, 4(September), pp. 1–19. doi: 10.3389/fimmu.2013.00268.