THE UNIVERSITY OF CHICAGO

INTERROGATING THE 3D STRUCTURE OF PRIMATE GENOMES

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES
AND THE PRITZKER SCHOOL OF MEDICINE
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF HUMAN GENETICS

BY
ITTAI ETHAN ERES

CHICAGO, ILLINOIS
DECEMBER 2020

"If I am not for myself, who will be for me?

But if I am only for myself, who am I?

And if not now, when?"

Rabbi Hillel

Table of Contents

List of Figures

List of Tables [1]

1. Note: Due to the large size of many of the tables, the tables have been provided in a supplementary file accompanying the dissertation. In such cases, the page number provided below directs the reader to a table's title.

ACKNOWLEDGMENTS

I would like to express extreme gratitude to all the people who supported, guided, mentored, and assisted me before and throughout my PhD. None of this work would have been possible without the mental, emotional, and fiscal support of numerous other trainees and faculty.

I am very thankful to my advisor, Yoav Gilad. Our interactions and conversations throughout the course of the PhD have not only significantly shaped my scientific thinking, but have also made my grad school experience an enjoyable and educational one. I found Yoav as an advisor to be a very understanding and compassionate, two qualities that made for a very enriching training experience. I am especially grateful that he encouraged me to pursue comparative primate genomics through the lens of my own interest in 3D genomics, while also always keeping me grounded in pursuing tangible results and publication. Additionally, Yoav was flexible and versatile when it was most needed (i.e. during the COVID-19 pandemic), and helped ensure I was still able to do quality scientific work and graduate. In so many ways, I could not have asked for a better PhD advisor.

I am also thankful to my thesis committee members, Marcelo Nobrega, John Novembre, and Xin He. Their guidance and insight throughout our committee meetings helped steer me in the right direction to make interesting inferences from my data, and, ultimately, to graduate. I also greatly enjoyed our "big picture" discussions at these meetings, that helped contextualize my work in the broader space of epigenetics, evolution, and 3D genome structure. Additionally, I am thankful to Matthew Stephens for his statistical advice and explanations at different times throughout my PhD.

I would like to thank my collaborators, Kevin Luo, Lauren Blake, and Joyce Hsiao. Kevin was instrumental in helping direct my early analyses of Hi-C data, especially with respect to quality control metrics. Lauren and Joyce were both extraordinarily helpful for properly carrying out the mediation analysis assessing the effect of 3D chromatin structure on gene expression differences.

In general, I would like to thank all members of the Gilad lab. My time in the lab was an extremely enjoyable and educational experience. I would like to give special thanks to Bryan Pavlovic for mentoring me early on in my PhD studies. Numerous other lab members provided helpful scientific guidance and were great friends to me during my time in the lab; in particular I would like to thank Jonathan Burnett, Nick Banovich, Seb Pott, Genevieve Housman, Kenneth Barr, Benjamin Fair, Ben Umans, Reem Elorbany, Katie Rhodes, Wenhe Lin, Erik McIntire, Sidney Wang, Courtney Burrows, Po-Yuan Tung, John Blischak, Briana Mittleman, Natalia Gonzales, Irene Gallego Romero, Amy Mitrano, Emilie Briscoe, Stephanie Lozano, Marsha Myrthil, and Claudia Chavarria. I am also grateful to other members of the broader human genetics community during my time at University of Chicago: Abhishek Sarkar, Joe Marcus, Arjun Biddanda, Nick Knoblauch, Kevin Magnaye, Sarah Urbut, Oni Basu, Vincent Lynch, and countless other trainees, faculty, and staff.

A wide variety of different administrators provided me with unparalleled support during my PhD. I am especially thankful to Sue Levison, who has always gone above and beyond the call of duty to make me and other students feel welcome, and often helped assuage my worries. I would also like to thank Diane Hall, Melissa Lindberg, Candice Lewis, Anita Williams, and Carolyn Brown. In the first few years of my PhD, I was lucky to be funded by the Genetics and Regulation Training Grant (T32GM007197) from the National Institutes of Health. Many thanks to the NIH and to Lucia Rothman-Denes, the grant manager.

I would like to thank my various friends throughout grad school who provided many different kinds of support: Andrew Tremain, Amelia Joslin, UnJin Lee, Frances Lee, Charlie Lang, Erin Fry, Katie Mika, Aarti Venkat, Ryan Duncombe, Chris Stamper, Jess Fessler, Andrés Moya-Rodriguez, Keelan Armstrong, Ben Ward, Addison Hughes, Michael Hustedde, Drew Waford, Asher Mayerson, Max Sloan, Sam Broer, Seth Brown, Paul Finkelstein, Alex Advani, Toufic Mayassi, Sangman Kim, Jason Lui, Logan Poole, Charlie Dulberger, and countless others. I am also grateful for the different groups that I played both IM and pick-

up soccer with throughout grad school. Finally, I would like to thank my partner, Hannah Morley, for the many different ways in which she has supported me throughout this process.

Lastly and most importantly, I would like to thank my family. My parents, Avichai and Ronit, supported me in so many different ways growing up, and gave me the educational and experiential opportunities that enabled me to pursue a PhD to begin with. I could not be more grateful for all their support throughout my life. My siblings, Tomer and Merav, have also been instrumental to my development both emotionally and intellectually, and I am thankful for the different ways in which they have supported and lifted me up. None of what I have accomplished would have been possible without the amazing support my whole family has continuously provided, and I am forever indebted to them.

ABSTRACT

A primary goal in human genetics is to understand how genetic variation affects phenotypic variation observed between different individuals. Elucidating these connections is crucial to understanding the molecular mechanisms and causes that lead to differences in traits and diseases among humans, and enables the discovery of therapeutic interventions that can improve human health. Many of the genetic variants associated with trait variation thus far are in non-coding regions of the genome, emphasizing a need to characterize non-coding sequence elements through functional genomics approaches. As opposed to variation in protein-coding sequences, genetic variation at non-coding loci has not been decoded, impeding a rapid understanding of its downstream functional effects. What is clear is that many non-coding loci are likely implicated in gene regulation, and understanding precisely what they are and how they act will require application of a wide variety of functional genomics techniques. In particular, one emerging technique assesses 3D genome structure, which can help connect regulatory loci to the genes they affect, and is rapidly expanding our understanding of gene regulatory networks and the noncoding genome. Various epigenetic features, such as 3D genome structure, can now be compared genome-wide across humans and non-human primates, broadening our understanding of evolution and gene regulation. My thesis work uses these paradigms to understand the interplay between 3D chromatin organization and gene expression across evolution. In Chapter 2, I apply RNA sequencing and chromosome conformation capture sequencing to human and chimpanzee induced pluripotent stem cells, in order to understand how divergence in 3D regulatory landscapes across these species affects expression divergence. As expected, this work demonstrates that reorganization of 3D genome structure contributes to gene regulatory evolution in primates. In Chapter 3, I critically assess existing evidence from other studies that have led many to conclude there is high evolutionary conservation of topologically associating domains (TADs, a large-scale feature of 3D genome organization). A thorough examination of the available data suggests such

a conclusion may be unwarranted. Finally, in Chapter 4, I summarize the insights gained from this work, assess the state of the 3D genome field in general, and suggest next steps for future research.

CHAPTER 1

INTRODUCTION

1.1 The evolution of gene regulation

The study of human genetics seeks to understand the genetic basis behind human phenotypic variation. Specifically, human geneticists are interested in characterizing how genetics impacts differences in traits between species, as well as between individuals within our own species. Insights gleaned from this field of study are useful not only for understanding and potentially improving human health, but also for furthering our understanding of biology and evolution more broadly. One particularly compelling approach that addresses both these aims is to compare genetic and phenotypic variation between humans and closely-related primate species. Early comparisons between the human and chimpanzee genomes revealed that the two species share the vast majority (99%) of their DNA, with an especially high level of conservation in protein-coding sequences [147, 225, 302]. These observations provided amazing evolutionary insight, suggesting that differences observed between humans and chimpanzees are not driven by inherent differences in the proteins (genes), but, rather, by differences in how, when, and where these proteins are expressed–i.e. differences in gene regulation [147]. More recently, genome-wide association studies (GWAS) within the human species have connected inter-individual differences in traits and diseases with thousands of genetic variants, the vast majority of which are non-coding [78, 110]. Additionally, numerous estimates suggest the human genome is comprised of up to 98% noncoding DNA, much of which is functional, [218, 47, 139, 156, 169], further underscoring the importance of understanding gene regulation.

The hypothesis that gene regulatory differences may be major drivers of phenotypic variation was first posited more than 50 years ago [28, 29]. Although there is still some debate about the extent to which adaptation and speciation are driven by mutations in gene

regulatory loci vs. mutations in protein-coding genes [119, 36], it is evident that gene expression variance plays a crucial role in phenotypic divergence within and between species [35, 311, 107, 92, 277, 328]. Advances in molecular biology and sequencing technologies over the last several decades have enabled rigorous investigations of gene expression levels and the regulatory mechanisms affecting them. RNA-sequencing (RNA-seq) represents a major improvement over prior microarray technology for accurate genome-wide measurement of gene expression levels and other transcriptomic phenotypes [326, 182, 299, 184]. Simultaneously, an array of emerging molecular and computational techniques allow for genome-wide assessment of different regulatory mechanisms such as DNA methylation, histone modification, chromatin state, transcription factor binding, and more [43]. Widespread use of these techniques is needed, because, in contrast to protein-coding genes, there is no clear code connecting primary sequence to downstream function for the noncoding portions of the genome. The outstanding challenges are thus to characterize the different genetic and epigenetic mechanisms regulating gene expression, to understand their relative evolutionary contributions to adaptation and speciation, and to ultimately connect them back to variation in primary DNA sequence.

While my work and this introduction focuses on inter-species primate comparative genomics in order to understand the evolution of gene regulation, it should be noted that research on gene expression differences within species has yielded valuable insights, suggesting expression levels are heritable and connected to genetic variation through expression quantitative trait loci (eQTLs) [181, 108]. These findings further highlight the functional relevance of gene expression levels as a molecular phenotype affecting higher-order traits, providing a tractable avenue for understanding the evolution of a wide variety of phenotypes. In the rest of this introduction, I will review key findings from comparative primate genomics studies about the evolution of gene regulation, and explain why characterizing the 3-dimensional structure of the genome is a critical next step in understanding the evolution

of gene regulation and the relationships between genotypes and phenotypes.

1.2 Gene regulatory evolution insights from comparative primate genomics

Why focus on primates? Numerous efforts have made progress in decoding the non-coding genome and furthering our understanding of gene regulation. In particular, work from the ENCODE (Encyclopedia of DNA Elements) consortia and many others have helped to identify and characterize functional regulatory elements across a number of species, including humans [16, 48, 50, 38]. Inter-species comparative genomics studies have been especially effective in identifying functional regulatory sequences [208]. In addition to characterizing specific regulatory loci, studies in model organisms have also revealed that expression divergence between species is primarily driven by mutations in cis-regulatory elements (CREs), rather than trans elements [289]. The former operate in an allele-specific manner, typically on the same chromosome, while the latter operate more broadly and can often diffuse throughout the genome. Although these research endeavors have expanded knowledge on gene regulatory mechanisms and associated loci more broadly, their use of distantly-related species precludes the possibility of categorizing their findings as highly conserved or human-specific [123]. Grounding genetic and epigenetic observations in humans by comparison with closely-related non-human primates (NHP) is crucial for providing evolutionary context. Without such comparisons, it is impossible to obtain a comprehensive understanding of how different loci and mechanisms of gene regulation have evolved in the human lineage, and, consequently, how these features may affect human-specific phenotypes [239]. In addition to providing evolutionary context, using NHP in comparative genomics and biomedical research can yield useful insights into human diseases, which are harder to model in more distantly related species with more divergent physiologies [238]. There are thus a number of different ways in which comparative genomics studies utilizing NHP may provide unique insights into

human evolution and the lexicon of human gene regulation, that cannot be obtained by focusing exclusively on model organisms and/or more distantly related species.

However, using primate comparative genomics to understand gene regulation still entails a number of challenges. The paucity of human and chimpanzee primary tissues, as well as obvious ethical limitations on experimentation in the two species, represent major barriers in the study of gene regulation [239]. Comparison of biological samples between primate species is possible with post-mortem collection of flash-frozen tissues, but this approach is problematic for several reasons. Due to the inherently opportunistic sample collection, sample sizes are typically quite small, with some studies using only a few individuals from each species [21, 214, 228]. Additionally, post-mortem tissue samples may be subject to variance induced by technical factors such as sample collection and shipping [19, 44]. These issues may be mitigated by utilizing induced pluripotent stem cells (iPSCs), which can be reprogrammed from and differentiated into a wide variety of different cell types [285, 284, 279], allowing for controlled experimentation on larger panels of human and NHP cells [240, 183]. Regardless of the biological samples being compared, care must also be taken in the study design and analysis methods employed in a comparative genomics setting. Without careful study design, batch effects may have a strong impact on the data, leading to inferences and conclusions that are driven more by technical variables than by true biological differences. In one intriguing recent example, researchers observed gene expression data from human and mouse clustering by species, rather than by tissue (as would be expected based on prior research) [50, 319]. A reanalysis of the data found that this unexpected observation was due to flawed study design confounding sequencing batch with species, and that the data do indeed cluster by tissue after accounting for sequencing batch effects [105]. With respect to the analysis of comparative genomics data, thoughtful normalization and orthology-calling techniques must also be employed to ensure comparisons made between sequences and features lead to valid biological inferences [294, 332, 20, 31]. Intentional study

designs, sample collection methods, and analytical techniques can attenuate many of the aforementioned limitations.

Even when researchers do well to address technical limitations and minimize confounding variables, there are still challenges in inferring regulatory evolutionary dynamics from comparative primate genomics studies. Numerous models and methodologies exist to infer the action of natural selection on primary DNA sequence [297]. Such methods are particularly effective when applied to protein-coding genes, where the functional effect of a mutation can be understood readily, given our knowledge of the link between sequence codons and the amino acids they are translated into. However, the links between primary sequence and trait variation are considerably less clear in the context of epigenetic, regulatory, and other molecular phenotypes. In these cases, the action of natural selection on the trait in question can be inferred and statistically tested based on observed deviations from null models (e.g. a neutral model with no selection). These types of approaches can be extremely useful in model organisms, where one can directly measure the necessary parameters (e.g. mutation rate) to generate a reasonable null model [106]. Unfortunately, these parameters are difficult to measure even in model organisms, and can often be practically impossible to estimate in humans and NHP. For most functional genomic traits that are studied, there is not yet a robust, well-formulated null model of no selection. Thus, more ad-hoc and empirical approaches must be utilized to understand the action of natural selection on intermediate molecular phenotypes. For instance, if gene expression levels show low variation both within and between species, it may be inferred that regulation of these genes is evolving under stabilizing selection (i.e. expression extremes are selected against). Conversely, if expression variance is low within species, but mean expression is much higher or lower in one species compared to others, this may suggest directional selection is affecting regulation of the gene in that species [239]. Empirical approaches such as these can therefore elucidate evolutionary dynamics, but greater care must be taken in interpretation of their results. In the previ-

ous example, more extreme expression in one species may actually be due to environmental factors and/or reduced action of stabilizing selection on that lineage, rather than directional selection. Functional follow-up studies, consideration of interspecies environmental differences, and accounting for different possible evolutionary trajectories can help exclude alternative explanations for specific evolutionary inferences. Consequently, when executed and interpreted properly, comparative primate genomics studies have vastly expanded our understanding of the mechanisms and loci involved in the evolution of gene regulation.

What have we learned? For many genes, results from comparative studies suggest that expression levels in primates are evolving under natural selection [142, 21, 27]. One early study looked at RNA levels in liver tissues from humans, chimpanzees, orangutans, and rhesus macaques, finding a set of genes with relatively invariant expression levels across species [107]. If regulatory mutations are in general selectively neutral (as many other mutations have been speculated to be [146]), expression levels for most genes would show more substantial inter-primate variation. The observed low variance in expression across diverse primate lineages suggests stabilizing selection has acted on these genes, preventing extreme expression levels [162]. Other studies comparing RNA levels across primate species have largely corroborated this notion [21, 219, 27, 42]. It is interesting to note that interspecies conservation of expression levels is particularly high for genes thought to be critical for defining cell type identity. In turn, gene expression patterns are more similar in the same tissue across different species than across different tissues within a species [278]. Taken together, these results bolster the idea that regulatory changes may be crucial drivers of evolution and adaptation. Compared to protein-coding sequence mutations, regulatory mutations can act in a more tissue-specific manner, and are thus less susceptible to pleiotropic effects that could be deleterious across multiple organs [304, 21].

While most genes show evidence for their expression evolving under stabilizing selection in primates, there has also been great interest in finding examples of directional selection.

The impact of such examples is fairly intuitive: a strong motivation in comparative primate genomics is to identify the genetic and epigenetic facets underlying differences between humans and NHP, in an effort to understand the biology behind human-specific traits and diseases. The exact proportion of genes whose regulation appears to be evolving under directional selection in primates differs dependent upon the tissue or cell type being considered, but still represents the minority of genes [239]. As discussed above, not every gene showing lineage-specific expression differences compared to other primates is necessarily evolving under positive (directional) selection on its regulation, but many likely are. Since some of the most striking phenotypic differences between humans and NHP are cognitive, many primate comparative genomics studies have focused on the brain and specific cell types therein [210, 104]. Results from studies utilizing bulk RNA-seq [271], and from more recent work examining RNA transcripts in single cells [195, 334, 274], suggest that interspecies expression variation in tissue location (heterotopy) and timing (heterochrony) during brain development may play a role in cognitive differences observed between humans and NHP. There are also examples in other organs of directional selection acting on gene regulation. An RNA-seq study on livers from 16 species revealed expression changes in some primate lineages that could be tied to dietary adaptations [219]. A similar study examining livers, kidneys, and hearts in humans, chimpanzees, and rhesus macaques found subsets of genes in each tissue displaying lineage-specific expression in humans [21]. Yet another study found marked expression differences in blood leukocytes, livers, and brains of humans, chimpanzees, orangutans, and rhesus macaques [79]. A more recent study assayed RNA levels in heart, kidney, liver, and lung tissue samples from humans, chimpanzees, and rhesus macaques, and also found a minority of genes whose expression patterns imply some directional selection [19]. Specific and functionally validated examples of genes with expression levels evolving under directional selection are scarce, and it is difficult to confidently connect them to higher-order phenotypes [8, 301]. More examples will hopefully be discovered and

characterized as our understanding of gene-phenotype connections increases, and as single-cell and other technologies enable better sampling from different locations, time points, and, consequently, cell types within a tissue. Regardless, numerous lines of evidence from inferences about directional selection suggest that a complex network of different mechanisms regulates gene expression. Gene sets with some evidence for directional selection on their regulation are often enriched for transcription factors [21, 107], are regulated by fewer enhancers than genes with more conserved expression patterns [57, 18], and do not necessarily display signatures of directional selection in the abundance of their corresponding proteins [143]. This last observation is particularly intriguing, and suggests that protein levels may be under more selective constraint than RNA levels and other gene regulatory mechanisms [298, 4]. An emerging notion from this and other work is that regulatory mechanisms utilize such buffering and redundancy to ensure appropriate downstream functional outcomes [180]. Together, these findings highlight the importance of measuring a wide variety of epigenetic features in order to understand the evolution of gene regulation.

Indeed, more recent work has moved from merely characterizing expression differences across primate species, to attempting to understand variation in regulatory mechanisms driving these differences. Understanding the regulatory mechanisms and loci responsible for expression variation should help de-mystify the noncoding portions of the genome, identifying functional elements that may have an impact on human health [52]. One recent study examined DNA methylation and gene expression in livers, kidneys, hearts, and lungs from humans, chimpanzees, and rhesus macaques, finding that methylation differences can only explain a small proportion of expression variation between tissues and species [19]. An earlier study found similar results when comparing human and chimpanzee livers, hearts, and kidneys [214]. Differential abundance of microRNAs, which regulate mRNA transcript decay, was also observed to account for only a small proportion (<5%) of expression differences in prefrontal cotex across humans, chimpanzees, and rhesus macaques [124, 272]. Alterna-

tive splicing of genes, which could (in theory) easily introduce regulatory novelty, has also been shown to have little effect on differential expression between humans and chimpanzees [32]. Comparable observations have been made for histone marks: one early study examined H3K4me3 in lymphoblastoid cell lines (LCLs) from humans, chimpanzees, and rhesus macaques, and found that interspecies differences in this histone modification explain only 7% of interspecies expression variation [31]. These results are perhaps not surprising, given that only a small minority of loci show human-specific increase or decrease of H3K4me3 in prefrontal cortex samples from the same species [262]. A later analysis in LCLs from the same species integrated RNA polymerase II occupancy, H3K4me1, H3k4me3, H3K27ac, and H3K27me3, and found that roughly 40% of interspecies gene expression variance can be explained by these marks combined [331]. Similarly, an even more recent study integrated a wide variety of histone marks as well as chromatin accessibility and methylation status in LCLs from great apes and macaques, and found that these features combined can explain approximately 67% of interspecies expression variance [101]. It would thus appear that the effect of any single epigenetic mechanism on gene expression is relatively modest, but, in concert, these mechanisms have a large effect on expression levels.

One principle that emerges from these findings and others in model organisms is that chromatin state and its effects on cis-regulatory elements (CREs) play a major role in the evolution of gene expression [239, 59]. Many of the aforementioned epigenetic modifications and mechanisms are associated with how 'open' or 'closed' chromatin is at a given locus, making the locus and the CREs it encompasses more or less accessible to transcription factors and other regulatory machinery [33]. It is therefore important to assess not only whether a given CRE is conserved, but also if the chromatin state at that locus is comparable across species. A variety of methods exist to assay chromatin accessibility, such as DNase-seq and ATAC-seq [148, 292, 30]. Broadly, these methods have found that the vast majority of the genome is not accessible in any given cell type, and that most transcrip-

tion factor binding events occur within regions of open chromatin identified by these assays [288]. Consequently, regions of open chromatin are often likely to harbor active regulatory elements [273, 150]. Consortia such as ENCODE have used ChIP-seq and other techniques to produce copious histone mark and other epigenetic data that, through careful analysis, have helped identify and characterize different classes of CREs at these putative regulatory regions[48, 166, 81, 120]. Some of the most well-studied CREs are enhancers, DNA modules that interface with transcription factors and associated proteins to make contact with gene promoters, thereby affecting gene expression. Although the measured extent of inter-primate divergence or conservation in enhancers and other CREs is dependent on the technology used and the number of species examined [77, 280], there are intriguing enhancer differences across primate species. Enhancers are considerably less conserved amongst primate lineages than gene expression levels [296, 18] and less conserved than promoters [291], highlighting their evolutionary relevance. Surveys of enhancers across primate and mammalian evolution have found interspecies differences in their activity [149, 228, 260], and evidence for high evolutionary turnover of enhancers as compared to promoters [34, 296].

While tremendous work has been done to identify and characterize enhancers and other CREs, numerous outstanding questions remain. Identification of enhancers is still imperfect because there is no single epigenetic mark that perfectly predicts enhancer regions. Enhancers affect gene expression via chromatin looping but also through other mechanisms that are less well understood (e.g. enhancer transcription), and disease-associated genetic variation at enhancer regions is difficult to functionally characterize [218]. Although we have been able to map regulatory quantitative trait loci (QTL, genetic variants with discrete effects on expression or other intermediate molecular phenotypes), most disease-associated variants in regulatory regions do not appear as QTL [293]. This is likely due in part to sampling strategies (e.g. not assaying gene expression in the appropriate cell type or condition relevant to the disease), but progress in understanding these mutations is also broadly stymied by a lack

of knowledge about which gene(s) a given CRE regulates [293]. Determining CRE targets is of particular importance both because CREs act in a distance-independent manner (only 40% of enhancers are linked to the nearest gene [51]), and because many CREs are tissue-specific in their activity [163, 213, 3, 309]. Ultimately, CREs' tissue-specific effects on gene expression are likely to be principally determined by local chromatin state and by what gene promoter(s) CREs come into contact with [100]. Connecting regulatory elements directly with their targets thus represents a crucial step towards obtaining a complete picture of how CREs modify expression, and how mutations in CREs affect higher-order phenotypes. My thesis work addresses this issue by examining expression divergence between humans and chimpanzees, with a focus on comparative assessment of CRE-gene contacts in 3-dimensional genome space.

1.3 The growing importance of the 3D genome

Given the vast wealth of genomic information that each cell has to carry in its nucleus, eukaryotic genomes must be packaged in a highly complex and structured fashion. At the highest level, individual chromosomes preferentially occupy discrete regions of the nucleus ("chromosome territories"), and these territories are fairly conserved across primate lineages [190, 194, 286]. Early studies of 3D genome structure largely relied on imaging techniques such as FISH (fluorescence in-situ hybridization), and did well to uncover chromosome territories; the advent of molecular methodologies like chromosome conformation capture (3C) have enabled interrogation of 3D genome structure at even finer scales [60]. Since the inception of 3C, the technology has developed through a variety of iterations and improvements that have increased its genomic resolution and throughput [94]. The latest version, known as Hi-C, pairs the original method's proximity-based ligation with high throughput next-generation sequencing to find DNA-DNA contacts genome-wide [167]. These techniques have revealed that, beneath the scale of chromosome territories, individual chromosomes are

also partitioned into two types of large-scale "compartments": A compartments, representing open and transcriptionally accessible chromatin, and B compartments, representing closed chromatin [167, 205]. At an even finer scale, loci within a compartment appear to form self-interacting regions on the scale of a megabase, termed topologically associating domains (TADs) [69, 209, 122, 258]. As I discuss further in the third chapter of this thesis, the definition of a TAD is still changing as Hi-C libraries are more deeply sequenced and new TAD inference algorithms arise. Regardless, loci within a TAD make contact with one another much more frequently than they do with loci outside of the TAD, suggesting that TADs may represented neighborhoods of insulated gene regulation that constrain the possible set of gene-CRE interactions [5, 282, 257]. Lastly and perhaps most importantly, at the lowest scale, Hi-C and related techniques have uncovered individual DNA looping interactions that bring linearly distant CREs into proximity with the genes they regulate [233, 133].

Numerous lines of evidence point to the functional significance of 3D genome organization at different scales. Studies have found that a significant fraction of genomic regions switch between A and B compartments in different ways during organismal development and cellular differentiation [200, 68, 324, 172]. Such compartment switching can move individual loci between permissive and repressed states along differentiation pathways, in part explaining regulatory differences between different cell types [168, 65]. Some similar results have been observed in TADs, suggesting TAD locations and intra- and inter-TAD contact frequencies change during cellular development [39, 68, 93, 24, 40], although the extent of these alterations is less clear given variance in TAD identification [56]. Despite uncertainty about the role of TAD variance in cellular differentiation, it is clear that TADs play an important role in genome organization and function. Genes located within the same TAD can have strongly correlated expression patterns and are often co-regulated during cell differentiation [209, 323, 232]. TAD boundaries are strongly correlated with replication-timing domain boundaries [223], and are enriched for insulator elements such as CCCTC-binding

factor (CTCF) [69, 233]. Disruptions to normative TAD structure have also been implicated in a number of human pathologies [174, 127, 91]. Similarly, 3D genome organization at the lowest scale (i.e. individual gene-CRE loops) affects gene expression and regulation [113, 10, 63, 196], and characterizing these interactions helps connect genetic variation with human trait and disease variation [49, 193, 325, 199]. While GWAS have done well to identify variants associated with many human diseases, analysis of chromatin conformation capture data is often necessary to understand understand how these variants exert their effects, which genes and loci they interact with, and, consequently, what therapeutic options may be effective [83]. In one particularly compelling example, researchers long thought that an obesity-associated mutation in an intron of the *FTO* gene increased risk of obesity and type 2 diabetes by affecting the gene itself [114]. This was not an unreasonable assumption, especially since follow-up studies found *FTO* expression levels affect body mass in mice [45, 88, 99]. Only through the application of chromosome conformation capture was it finally discovered that the variant in question actually regulates the expression of a transcription factor gene several megabases away, *IRX3* [265]. There are countless other examples where integration of Hi-C data has aided in identification of novel pathways and genes involved in the progression of various human pathologies [187, 186, 188, 312]. Without a doubt, collecting and analyzing chromosome conformation capture data across species and cell types represents an exciting novel frontier in broadening our understanding of gene regulation, development, and evolution.

Overall, there is a dearth of comparative studies examining 3D genome architecture across species. In Chapter 2, I address this gap by collecting, integrating, and analyzing Hi-C and RNA-seq data from human and chimpanzee iPSCs. In Chapter 3, I challenge the prevailing notion that TADs are highly conserved across species by critically analyzing the existing data that support this claim. Finally, in Chapter 4, I discuss the evolutionary and gene regulatory implications of my findings, before providing some perspective on the state of the

3D genome field and recommendations for future research directions.

CHAPTER 2

REORGANIZATION OF 3D GENOME STRUCTURE MAY CONTRIBUTE TO GENE REGULATORY EVOLUTION IN PRIMATES

2.1 Abstract[1]

A growing body of evidence supports the notion that variation in gene regulation plays a crucial role in both speciation and adaptation. However, a comprehensive functional understanding of the mechanisms underlying regulatory evolution remains elusive. In primates, one of the crucial missing pieces of information towards a better understanding of regulatory evolution is a comparative annotation of interactions between distal regulatory elements and promoters. Chromatin conformation capture technologies have enabled genome-wide quantifications of such distal 3D interactions. However, relatively little comparative research in primates has been done using such technologies. To address this gap, we used Hi-C to characterize 3D chromatin interactions in induced pluripotent stem cells (iPSCs) from humans and chimpanzees. We also used RNA-seq to collect gene expression data from the same lines. We generally observed that lower-order, pairwise 3D genomic interactions are conserved in humans and chimpanzees, but higher order genomic structures, such as topologically associating domains (TADs), are not as conserved. Inter-species differences in 3D genomic interactions are often associated with gene expression differences between the species. To provide additional functional context to our observations, we considered previously published chromatin data from human stem cells. We found that inter-species differences in 3D genomic interactions, which are also associated with gene expression differences between

1. Citation for chapter: Eres IE, Luo K, Hsiao CJ, Blake LE, Gilad Y. 2019. Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. PLoS Genetics 15, e1008278. doi:10.1371/journal.pgen.1008278

the species, are enriched for both active and repressive marks. Overall, our data demonstrate that, as expected, an understanding of 3D genome reorganization is key to explaining regulatory evolution.

2.2 Introduction

A growing body of evidence indicates that variation in gene regulation plays a key role in phenotypic divergence between species [29, 147, 35, 107, 311, 21, 135]. Inferring the causal relationship between inter-species regulatory differences and phenotypic differences between species remains challenging, but compelling examples of regulatory adaptations have been published in a large number of species, including primates [224, 8, 301, 170, 237, 222]. The molecular mechanisms that underlie regulatory adaptation have also been the focus of much research. Studies in mice, flies, yeast, and primates have revealed that expression divergence between species is often driven by mutations or epigenetic modifications within cis-regulatory elements (CREs), rather than trans elements (e.g. transcription factors [311, 224, 8, 301, 170, 237, 218, 213]). This makes intuitive sense, because transcription factors can operate broadly across multiple functional contexts and throughout the genome (affecting many genes), whereas CREs often have more specific functional outcomes [35, 305].

The ability to measure epigenetic marks, chromatin structure, and other functional genomic data has enabled us to identify and classify CREs into different types of regulators with distinct effects on gene expression (e.g. enhancers, silencers, insulators) [3, 48, 309]. Despite significant advances in our ability to identify and predict the functional role of CREs, we still lack a comprehensive characterization of the functional relationships between CREs and the genes they regulate. In many cases, we still do not know which genes are regulated by which CREs, or when and how often these relationships change. Connecting CREs to their target genes is crucial for understanding how regulatory architecture changes in response to different spatial, temporal and organismal contexts [213, 3, 48, 309, 163, 73, 84].

Ultimately, the effects of CREs on gene expression are likely to depend on which promoter(s) they contact, which is inherently related to the 3D structure of the genome [96, 126].

The proximity and frequency of CRE-gene contacts can be measured in vivo using chromosome conformation capture techniques [60]. Chromosome conformation affects how genes are expressed within a cell [7, 102, 155, 26, 72, 121, 132, 234]. For example, 3D genome structures may bring linearly distant loci into close proximity, connecting genes with CREs [259, 236, 54, 256, 130, 191, 152]. Expressed genes have been observed to spatially localize with distant CREs in 3D FISH experiments [236, 252]. The latest chromosome conformation capture based technique, Hi-C, pairs the original method's proximity-based ligation with high-throughput sequencing to identify DNA-DNA contacts on a genome-wide scale [167]. With enough sequencing coverage, Hi-C data can ultimately yield a comprehensive map of the 3D structure of an entire genome at high resolution [295].

Divergence in 3D genome structure may lead to regulatory evolution and ultimately to adaptation of new phenotypes. Currently, however, there are only a small number of comparative Hi-C data sets that can be used to test this notion [245, 233, 69], and even fewer comparative data sets in primates [58, 160]. Most Hi-C studies to date have focused primarily on variation in chromatin contact frequencies within a single species [167, 251, 145, 165]. The few comparative Hi-C studies published to date typically draw comparisons between distantly related species (such as human and mouse [233, 69]), use cancerous or otherwise transformed cell lines [233], and rely on low resolution genome-wide Hi-C data (typically collecting 100-600M reads from most samples [245, 69, 71]). These comparative studies typically collect data in only a single individual from each species, and often compare contacts that are inferred from Hi-C libraries with large differences in read depth between species, a property that leads to differences in power to infer 3D genome structures at multiple scales [245, 233, 69].

Thus, to conduct a comparative Hi-C study in primates and address these challenges, we collected high resolution Hi-C data from iPSCs derived from four human and four chimpanzee

individuals. The human and chimpanzee genomes share a high degree of synteny [320, 321, 250, 137, 37, 161], thus allowing us to consider a comparison of both low and high order chromatin interactions. Using our data, we were able to characterize 'lower-order' locus-locus contacts and to infer 'higher-order' structural features, such as TADs and TAD boundaries. We also quantified gene expression levels using RNA-seq data from the same eight cell lines. We considered our data with existing functional annotations, including histone marks and chromatin accessibility data, and evaluated the extent to which inter-species variation in 3D genome structure and epigenetic profiles are associated with gene expression divergence between humans and chimpanzees.

2.3 Results

We performed in situ Hi-C as previously described [233] on a sex-balanced panel of four human and four chimpanzee integration-free iPSC lines that were previously generated and quality-checked by the Gilad lab [240]. Using HiCUP [306] and HOMER [117] (see Methods), we obtained genome-wide Hi-C contact maps at 10 kb resolution for all eight individuals, with each map containing approximately one billion sequencing reads. Since there is currently no gold standard for Hi-C normalization and statistical modeling, we also used an alternative method, Juicer [74], to confirm that our results are robust with respect to the choices of normalization schemes and modeling (2.8–2.13 Figs). We also demonstrated the robustness of our results by performing certain analyses using different resolutions (from 10 kb to 500 kb). In the main text we report the results obtained using the HOMER pipeline at 10 kb resolution. Results using the alternative pipelines are shown in the supplement.

We used HOMER to independently classify between 779,503–883,438 contacts ($P < 0.01$) in the Hi-C data obtained from each individual (genomic coordinates of all contacts in all individuals are provided in S1–S8 Tables). We define a 'contact' as a pair of 10 kb regions which we observed to be in physical proximity more often than expected by chance.

Throughout the paper, we refer to Hi-C contacts as 'lower-order' or 'pairwise' interactions in order to distinguish them from higher-order, chromosome-scale structures (i.e. TADs and TAD boundaries).

Our goal was to compare Hi-C contacts between humans and chimpanzees. One intuitive approach to do so might be to identify the orthologous locations of each contact in the two species and classify such contacts as shared or unshared. However, this could lead to an inflated estimate of inter-species differences due to incomplete power to identify contacts in one species or the other. Instead, we collected the coordinates of all contacts identified in at least one individual into a single database. For each contact in the database, we independently identified the pair of corresponding orthologous regions in the human and chimpanzee genomes (using reciprocal searches, to avoid bias). Using this approach, we excluded about 18% of contacts because we failed to identify clear orthologous regions in the genomes of the two species (see Methods and S9 Table). Following the orthology based filtering, we extracted the normalized contact frequencies (log2 observed:expected read count ratios) for all pairs of loci in the database, regardless of whether they were classified as contacts by HOMER. Thus, our analysis is not biased by potential differences in power to detect contacts in one species over the other. That said, we observed that the variance in contact frequency was lower for interactions that were independently identified in a greater number of samples, regardless of species (2.14 Fig). We thus filtered out interactions that were independently classified as significant in fewer than four individuals. This approach allowed us to compare contact frequencies between species for 347,206 interactions while largely sidestepping the problem of incomplete power.

2.3.1 Inter-species differences in 3D genomic interactions

We used limma [267] to perform pairwise cyclic loess normalization and minimize the effects of technical variables on our data (2.15 Fig). Following normalization, principal components

analysis (PCA) and unsupervised hierarchical clustering of the Hi-C data revealed that, as expected, samples cluster by species (Fig 2.1).

To identify inter-species differences in contact frequencies, we analyzed the data using a linear model with fixed effects for species, sex, and processing batch (see Methods). At an FDR of 5%, we classified 13,572 contacts (about 4%) as having differential normalized contact frequency between humans and chimpanzees. Analysis of the orthologous regions anchoring these contacts suggested that approximately 4,000 of these differences might be explained by large inter-species differences in distance between mates of a contact pair (because read count is correlated with distance between the mates; see Methods and 2.16 Fig). We thus conservatively excluded locus pairs whose distance varied by more than 20 kb across species. Ultimately, we classified with confidence 9,661 Hi-C contacts (of 292,070; about 3.3%) with a significant difference in normalized contact frequency between the two species. We refer to these contacts as inter-species differentially contacting (DC) regions (S10 Table). Our observations thus suggest that lower-order contacts are generally conserved between humans and chimpanzees. That said, if we assume that all of the contacts we filtered out (either due to lack of orthology or because the distance between the anchor regions differed across species) are in fact DC, divergence in contact frequency would have been observed for 16% of the Hi-C contacts (assuming similar properties to the current data set). However, we find it more likely that a large subset of the contacts we excluded are not truly DC, but, rather, not comparable between the species due to differences in genome assembly quality.

Across all DC regions, 55% exhibited a higher contact frequency in chimpanzees, while 45% showed a higher frequency in humans (Fig 2.2A, see Fig 2.3 and 2.17 Fig for examples). We observed that some chromosomes were associated with greater asymmetry in inter-species contact frequencies than others (Fig 2.2B). Greater asymmetry seems to be present more often in chromosomes with large inter-species rearrangements. Specifically, in our data, 8 of the 9 chromosomes with known large-scale pericentric inversions between the species (1, 4, 5,

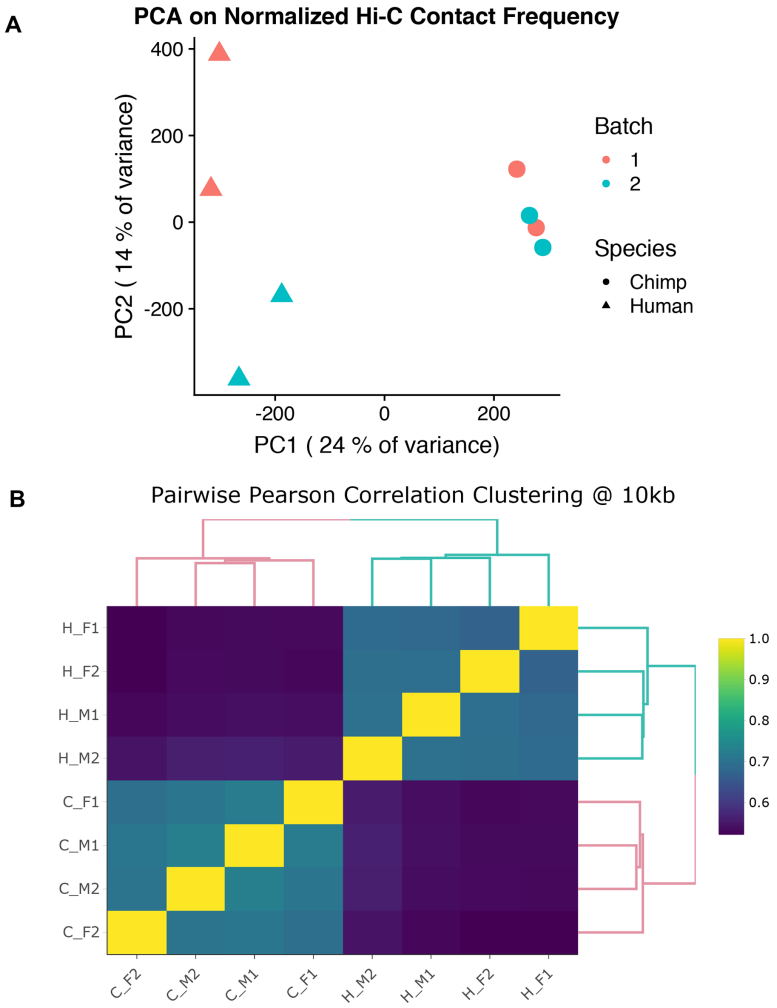

Figure 2.1: **General patterns in Hi-C data.** (A) Principal components analysis (PCA) of HOMER-normalized interaction frequencies for the union of all contacts in humans (triangles) and chimpanzees (circles). PC1 is highly correlated with species ($r = 0.98$; $P < 10^{-5}$). (B) Unsupervised hierarchical clustering of the pairwise correlations (Pearson's r^2) of HOMER-normalized interaction frequencies at 10 kb resolution. The first letter in the labels demarcates the species (H for human and C for chimpanzee), and the following symbols indicate sex (male, M or female, F) and batch (1 or 2).

9, 12, 15, 16, 17, and 18; [320, 321, 138, 206, 302, 64]) show particularly strong asymmetry in inter-species contact frequencies. We also observed asymmetry in inter-species contact frequencies in human chromosome 2, a fusion of the ancestral chromosomes giving rise to chimpanzee chromosomes 2A and 2B [302], as well as in chromosome 7, which has the highest number of un-localized sequences of any chromosome in the panTro5 genome.

Next, we turned our attention to higher-order chromosomal structures by characterizing TADs in each species. Previous studies indicate that the human and chimpanzee genomes share a high degree of synteny [320, 321, 250, 137, 37, 161], a property we confirmed by tiling each genome into various bin sizes and using a reciprocal best hits liftOver method to identify syntenic regions (see Methods and 2.18 Fig). To infer steady-state TAD structures, we pooled reads across all individuals within each species to create 'high-density consensus' Hi-C maps for humans and chimpanzees [74]. We used the Arrowhead algorithm at 10 kb resolution [74] to independently infer 11,298 TADs in humans and 10,505 TADs in chimpanzees (see Methods). We then used liftOver to identify orthologous genomic regions that corresponded to these TADs, and removed 10% of domains for which orthology could not be identified (S11 and S12 Tables list the TADs identified in each species; S13 Table lists the orthologous locations of the combined TADs). Once orthology has been established, for each TAD, we considered the domain conserved in humans and chimpanzees when 90% of the TAD interval overlapped reciprocally between species (see Methods). Using this approach, we found that only ∼43% of TADs discovered in humans and chimpanzees are shared (Fig 2.4A).

The observation that TADs are generally not as conserved as practically all other regulatory phenotypes studied in humans and chimpanzees was unexpected. We thus thoroughly tested the robustness of this inference. To do so, we performed a large number of alternative analyses. We analyzed the data at different resolutions (from 10 kb to 500 kb—each time repeating the reciprocal liftOver analysis). We analyzed the data by considering, instead of pooled data, TADs identified independently in a single and in up to 4 individuals within each

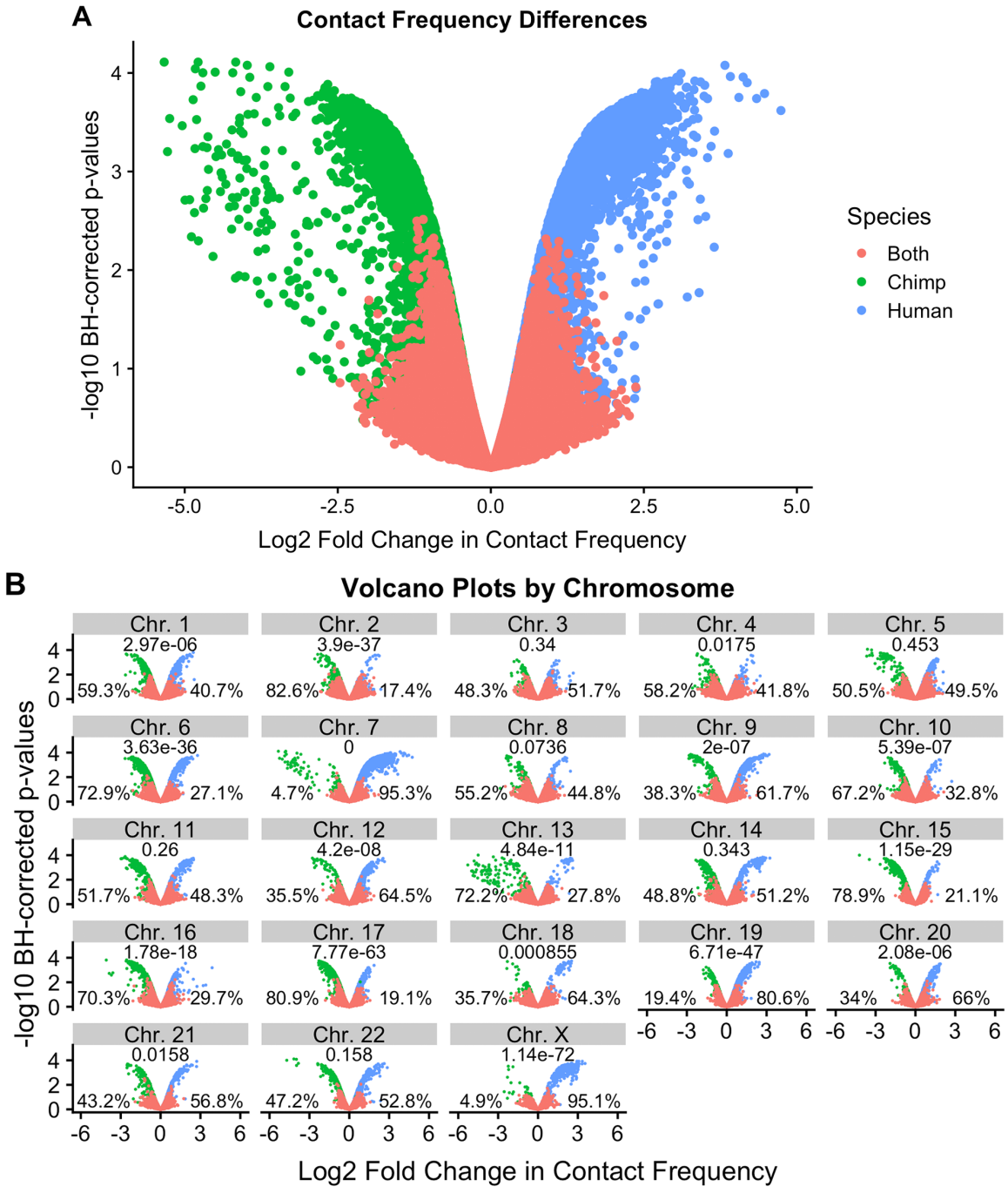

Figure 2.2: **Linear modeling reveals large-scale chromosomal differences in contact frequency.** (A) Volcano plot of $\log_2$ fold change in contact frequency between humans and chimpanzees (x-axis) against Benjamini-Hochberg FDR (y-axis), after filtering non-orthologus regions (results for unfiltered data are plotted in S9 Fig). Data are colored by the species in which the contact was originally identified as significant. (B) Per-chromosome volcano plot using the same legend as in A. P-values provided for a binomial test of the null that inter-species differences in contact frequencies are evenly distributed. The percentage of contacts with significant higher frequency in each species is noted.

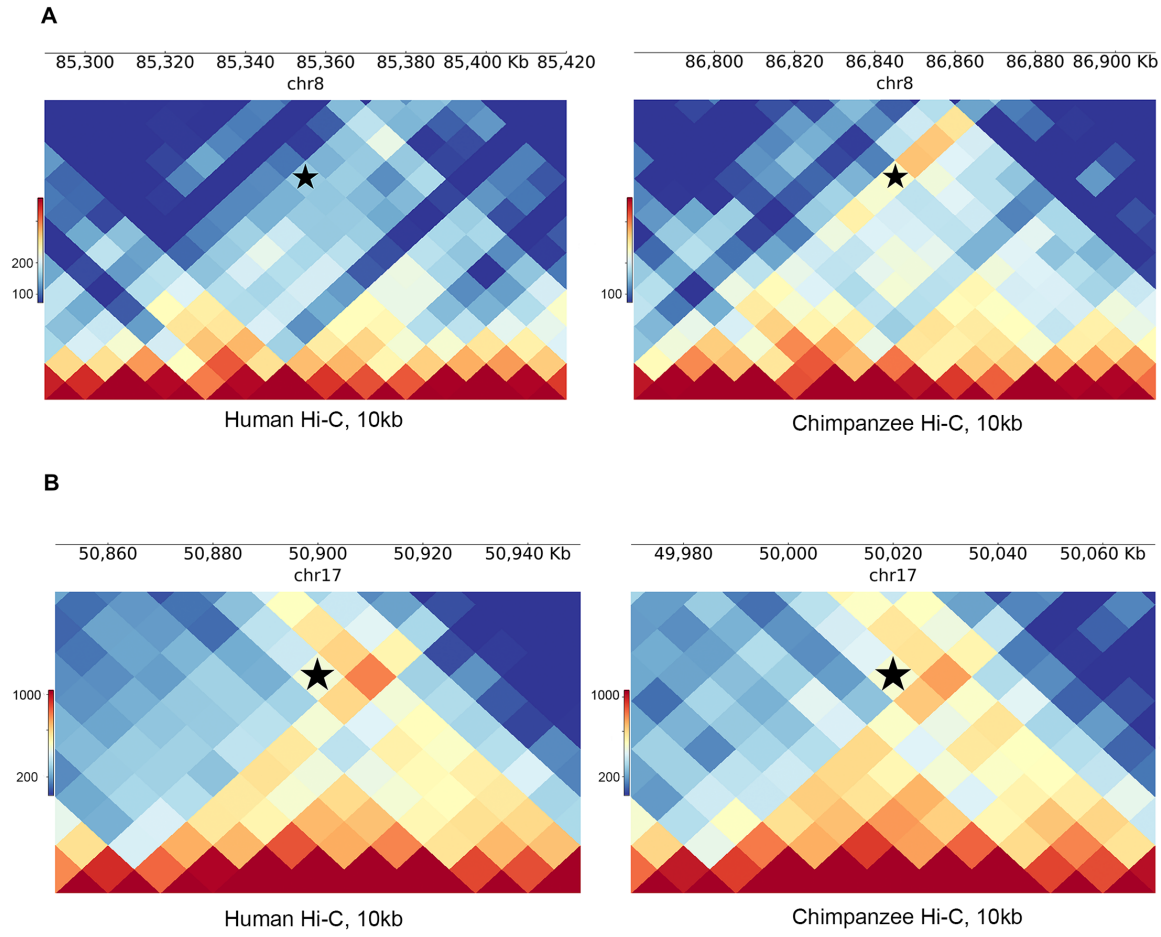

Figure 2.3: **Examples of DC and non-DC interactions.** (A) PyGenomeTracks plots [232] of a chromosome 8 interaction between bins 130kb away for human (left panel) and chimpanzee (right). The bin pair tested is indicated by a black star, and was found to be DC between species. (B) Same as A, but for a conserved (non-DC) interaction on chromosome 17 separated by 100kb.

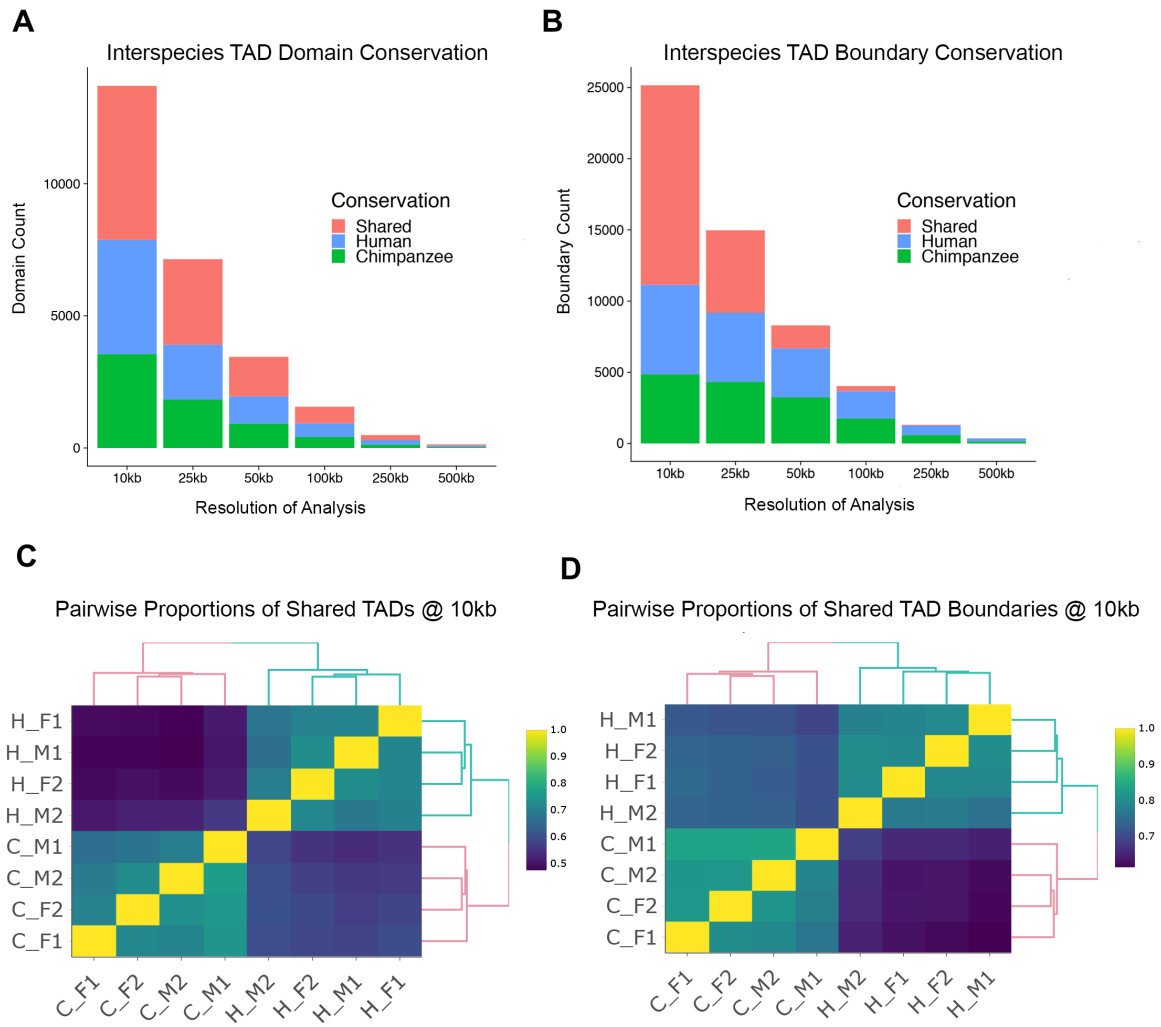

Figure 2.4: **Higher-order chromosomal structure in humans and chimpanzees.** (A) Across different resolutions (x-axis), we plotted the number of shared and species-specific domains (y-axis) identified with Arrowhead [74] using the consensus map from each species (alternative approaches plotted in 2.19–2.21 Figs). (B) Same as A, but for TAD boundaries instead of the domains themselves. Boundaries were defined as 15kb flanking regions at the edges of inferred Arrowhead domains. (C) Unsupervised hierarchical clustering of pairwise comparison of TADs across all individuals. These proportions were obtained using Arrowhead TAD inferences on each individual at 10kb resolution. Proportions indicated by color scale on right. Similar plots using analysis at different resolution are available in 2.19 and 2.21 Figs. (D) Similar to C, but for TAD boundaries instead of the domains themselves.

species (Fig 2.4C and 2.19 Fig), and we did this across the different resolutions. We analyzed the data by classifying conservation based on the approach of Rao et al. [233] instead of relying on an overlap of 90% of the domain; we analyzed the pooled data using panTro6 as a reference genome rather than the panTro5 assembly (2.20 Fig). We analyzed the data by focusing on boundaries instead of the entire domains (Fig 2.4B and 2.4D); we used multiple alternative definitions of boundaries, and repeated this analysis across all resolutions and with boundaries identified in different numbers of individuals within species (2.19 Fig). Finally, we identified TADs using an alternative algorithm, TopDom [261], and repeated all of the alternative analyses mentioned above using this algorithm (2.21 Fig).

The results of many of these alternative analyses are reported in the supplement (2.17 and 2.19–2.21 Figs). All of the alternative analyses produced consistent results and an inference that TADs and TAD boundaries are much less conserved between humans and chimpanzees than any other regulatory phenotype studied to date [214, 262, 32, 291, 228, 144]. The Arrowhead analysis of TADs that are independently identified in four individuals within either species, at 10 kb resolution, where conservation is classified based on the less stringent approach of Rao et al. [233], resulted in the highest estimate of conservation, with 78% of domains and 83% of boundaries shared between the species (2.19B and 2.19D Fig). The restriction to TADs or boundaries identified in all 4 individuals of either species results in far fewer features that can be examined (2.19 and 2.21 Figs), yet even in this analysis conservation of domains and boundaries is modest (see Fig 2.5 and 2.17 Fig for examples).

2.3.2 The relationship between inter-species differences in contacts and gene expression

We previously collected RNA sequencing data from the same human and chimpanzee iPSC lines [217]. We jointly analyzed the Hi-C and RNA-sequencing data to learn how often inter-species differences in 3D genomic contact frequencies are associated with inter-species

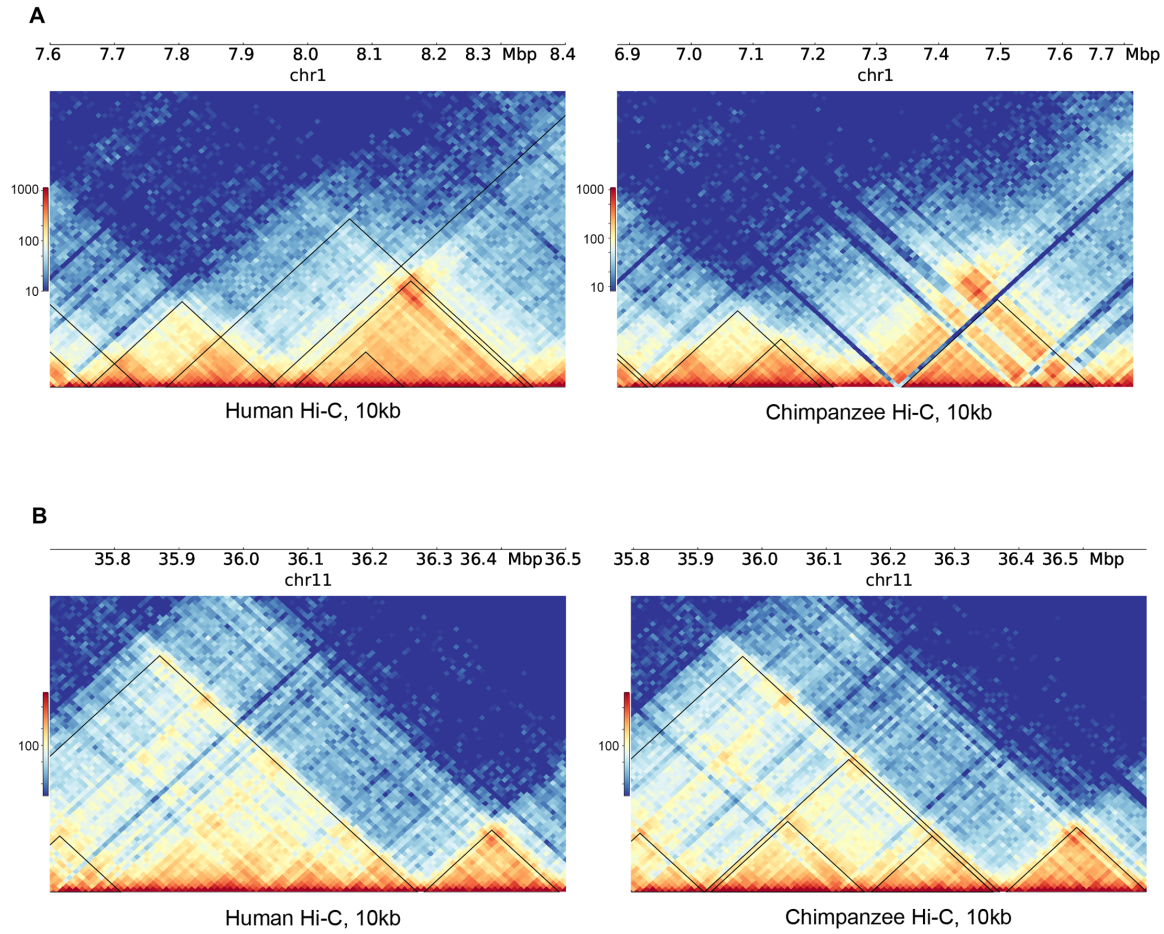

Figure 2.5: **Examples of conserved and divergent TADs.**(A) A region on chromosome 1 with examples of both conserved and divergent Arrowhead [74] TAD inferences (black lines). Both the larger TADs seen in the chimpanzee map (right) appear to be conserved in the human map (left), whereas several of the TADs inferred in the human map are noticeably absent from the chimpanzee map. (B) A region on chromosome 11, once again showing examples of conserved and divergent Arrowhead TAD inferences (black lines). All the TADs seen in the human map (left) appear conserved in the chimpanzee map (right), whereas three smaller TADs inferred in the chimpanzee map are not found in the human map, suggesting divergence.

differences in gene expression. We first identified 7,764 orthologous genes for which we have expression and Hi-C data anchored at a region that overlaps the gene's transcription start site (TSS; see Methods). A single genomic region that overlaps a TSS can have multiple contacts to other genomic regions. For the purpose of our analysis, we conservatively considered only the contact that shows the highest inter-species divergence for each gene.

We did not observe a correlation between gene expression and contact frequency when we considered data from all 7,764 genes. However, when we focused on the 1,401 genes classified as differentially expressed (DE) between humans and chimpanzees (at FDR ≤ 0.05), we observed an excess of both positive and negative correlations between inter-species differences in gene expression and inter-species differences in Hi-C contacts (2.22 Fig). Indeed, genes whose TSS is associated with inter-species DC are more likely to be DE between species (χ^2 test; P = 0.01; Fig 2.6A and 2.6B). The association between Hi-C contacts and gene expression divergence was somewhat stronger if instead of focusing on the contact with the highest divergence, we obtained a summary P-value [189] for testing the null hypothesis that there are no differences between the species in any of the contacts associated with the TSS for a given gene (P = 0.001; 2.27C and 2.27D Fig).

A combined analysis of functional genomic data does not allow us to infer a direct causal relationship between chromatin contacts and gene expression patterns. Nevertheless, independent evidence strongly suggests that changes in 3D genomic structure can affect interactions between regulatory elements and promoters [233, 66, 41, 134, 173], which may ultimately drive differences in gene expression levels [234, 66, 41, 134, 173, 263, 207]. We thus sought to quantitatively estimate the extent to which inter-species DC might explain gene expression differences between the species in our data. To do so, we estimated and compared the effect of species on expression before and after accounting for the corresponding contact frequencies (see Methods; [9]).

Specifically, we performed a mediation analysis using linear models to assess the effect

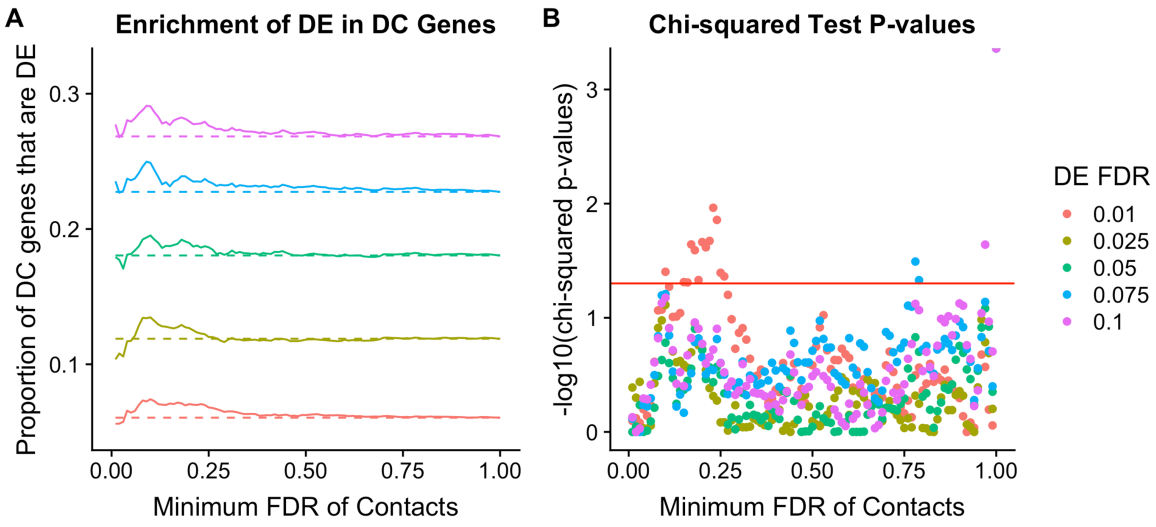

Figure 2.6: **Differentially contacting Hi-C loci show enrichment for differentially expressed genes.** A) Enrichment of inter-species differentially expressed (DE) genes with corresponding differences in Hi-C contact frequencies (DC) between the species. The proportion of DC genes that are significantly DE (y-axis) is shown across a range of DC FDRs (x-axis). Colors indicate different DE FDR thresholds, and dashed lines indicate the proportion of DE genes expected by chance alone. (B) P values of Chi-squared tests of the null that there is no difference in proportion of DE genes among DC genes (y-axis), shown for a range of DC FDRs (x-axis). In both panels, DC regions were chosen to have the minimum FDR supporting inter-species difference in contact frequency. We plotted results using the weighted p-value combination instead of the minimum FDR in 2.27 Fig.

of contact on expression divergence (95% confidence interval based on the Monte Carlo test of significance; see Methods). For approximately 8% of DE genes (116/1401) we were able to reject the null hypothesis that the indirect effect is zero (2.23 Fig). Taken together, these data suggest that a subset of inter-species differences in gene expression levels can be explained by divergence in Hi-C contacts.

2.3.3 The chromatin and epigenetic context of inter-species differences in 3D genome structure

Finally, we reasoned that species-specific contacts (i.e. significant DC regions) would be more likely to involve active, functional regulatory elements. This seems intuitive if one assumes most genomic contacts are functionally relevant, and not simply the result of pure noise. To test this hypothesis, we assessed the overlap between our Hi-C data and publicly available chromHMM annotations based on histone modification data from human embryonic stem cells [48]. We assigned each Hi-C locus to an epigenetic state based on its maximum weighted base pair overlap with 15-state chromHMM annotations (see Methods and 2.24 Fig). Our approach to classify Hi-C regions with a functional assignment based on majority sequence overlap is arbitrary, but our conclusions are robust with respect to alternative approaches to analyze the Hi-C data (2.11 and 2.24 Figs).

We found marked differences in the chromHMM annotations between genomic regions that are inferred to physically contact a promoter and those that do not contact a promoter (Fig 2.7A and 2.7B). For example, genomic regions in physical contact with a promoter are enriched with genic enhancer annotations (χ^2 test; P = 0.0002, S14 Table), as might be expected. Perhaps more novel is the observation that inter-species DC regions were also enriched with genic enhancers, in contrast to regions that did not differ in contact frequency between the two species (P = 0.04, S14 Table). We note that this latter observation is not robust with respect to different annotations of enhancers, and we do not find this association

if we simply combine all regions annotated as 'enhancers' in the data set.

We repeated the enrichment analysis of Hi-C regions using existing human iPSC histone mark data, including H3K27ac, H3K4me1, and H3K4me3, and h1-hESC histone mark data, including H3K27me3 and DNase I hypersensitivity sites (DHS [48]). As expected, Hi-C regions in contact with a promoter showed greater overlap with DHS peaks than Hi-C regions that did not contact a promoter (t-test, $P < 2.2 * 10^{-16}$; 2.25A Fig). When we focused on contacts involving a promoter, we found that inter-species DCs that are also associated with DE genes showed the largest overlap with DHS peaks, followed by DE genes that were not associated with DC regions ($P = 0.01$). Regions that were not associated with either DC or DE showed the least amount of overlap with DHS ($P = 0.0006$; 2.25B Fig).

Remarkably, apart from the heterochromatic, repressive marker H3K27me3 (where the sign of the effect was the same, but the enrichment was not significant), Hi-C regions that are DC and are also associated with DE genes are more likely to overlap all other histone marks in our data set compared with Hi-C regions that are not DC and are not associated with a DE gene (all enrichment $P < 0.03$; Fig 2.7C and 2.7D, 2.25C and 2.25D Fig). In other words, inter-species DCs associated with DE genes are more likely to occur in genomic regions that are marked by histone modification, and are thus likely to have a regulatory function.

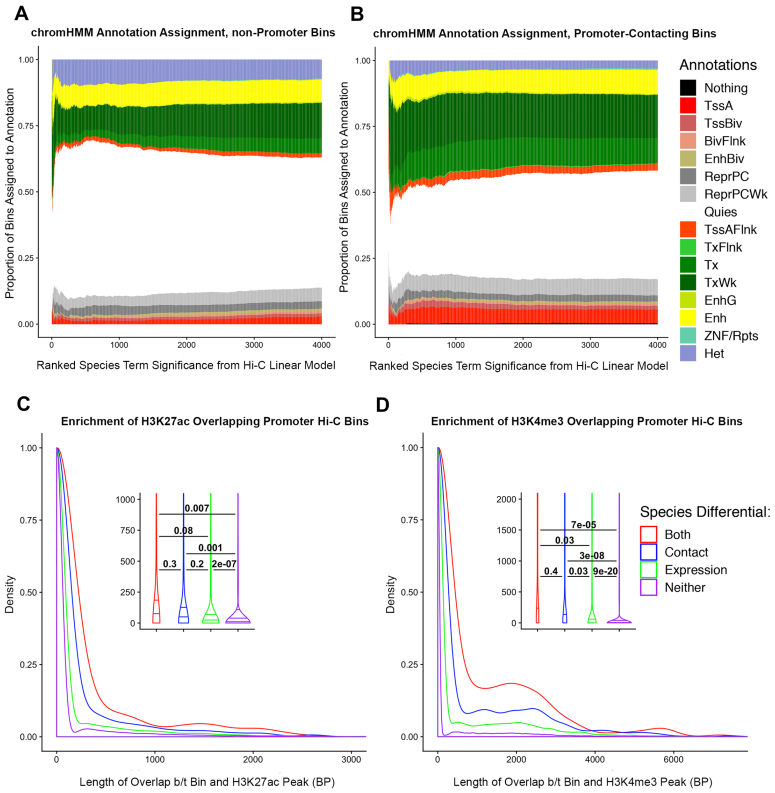

Figure 2.7: **Overlap of epigenetic signatures and Hi-C contacts.** (A) Hi-C loci that do not make contact with promoters are ranked in order of decreasing DC FDR (x-axis). The y-axis shows cumulative proportion of chromHMM annotation assignments for all Hi-C loci at the given FDR or lower. (TssA-Active TSS, TSSBiv-Bivalent/Poised TSS, BivFlnk-Flanking Bivalent TSS/Enh, EnhBiv-Bivalent Enhancer, ReprPC-Repressed PolyComb, ReprPCWk-Weak Repressed PolyComb, Quies-Quiescent/Low, TssAFlnk-Flanking Active TSS, TxFlnk-Transcription at gene 5' and 3', Tx-Strong transcription, TxWk-Weak transcription, EnhG-Genic Enhancers, Enh-Enhancers, ZNF/Rpts-ZNF genes and repeats, Het-Heterochromatin). (B) Same as A, but only considering Hi-C loci making contact with promoter bins. Results of separating promoter-contacting bins between DE and non-DE genes can be seen in 2.26 Fig. (C) Density plot of the base pair overlap between different classes of Hi-C contact loci and H3K27ac. Histone mark data were obtained from ENCODE in experiments carried out in human iPSCs. We grouped contacts into 4 classes, indicated by color: those that show differential contact between species, those that show differential expression between species, those that show both, and those that show neither. We used pairwise t-tests to compare differences in the mean overlap among the four classes of Hi-C loci. (D) Same as C, but performed on H3K4me3 data obtained from ENCODE, collected in hESCs. Results with other histone marks can be seen in 2.25 Fig.

2.4 Discussion

In general, we observed lower-order, pairwise chromatin contacts in iPSCs to be conserved between humans and chimpanzees. We believe that this observation is intuitive, though we acknowledge that with only four individuals from each species, and given the challenges in identifying orthologous regions, we are likely to somewhat underestimate the degree of divergence in pairwise chromatin contacts.

In contrast to the conservation of lower order pairwise contacts, we did not find higher-order chromatin structures, such as TADs and TAD boundaries, to be generally conserved between human and chimpanzee iPSCs. Because this observation seems to contradict previous reports suggesting that TADs are strongly conserved across species [233, 69], we performed a large number of alternative analyses to demonstrate the robustness of our inference. Even in our most lenient analysis, we observed that only 78% of domains are shared between humans and chimpanzees—a much lower conservation than observed for any other regulatory phenotypes between these two species (when similar sample sizes are considered; [214, 262, 32, 291, 228]).

While all of the alternative analyses supported our inference, these analyses also demonstrated the known difficulty of robustly inferring TADs and TAD boundaries based on Hi-C data alone [232, 56]. Indeed, the algorithms used to infer TADs and TAD boundaries themselves are not very robust, as has been discussed previously [56, 89]. Given our observations and the difficulty obtaining robust definitions of TADs and TAD boundaries, we carefully examined the previous evidence for high conservation of TADs between species.

The conclusion of our literature analysis is that the evidence for strong domain conservation is weak, and thus that our inference does not actually contradict previous data. A few of the studies typically cited as providing evidence for strong conservation of TADs across species did not actually perform a genome-wide assessment TADs, but inferred conservation based on a few examples [245, 116, 111]. Rudan et al. [245], for instance, reported functional

conservation of TAD boundaries in liver cells from rhesus macaque, dog, rabbit, and mouse, but did not report the number or proportion of conserved regions they observed. Instead, they presented correlations of 0.5 between contact frequencies across these species in subsets of contacts binned by the distance between mates, without further considering TADs or boundaries.

In contrast to studies that focused on specific examples, Dixon et al. [69], who originally described megabase-sized TADs at 40 kb resolution, reported that TAD boundaries were conserved in human and mouse embryonic stem cells. At a greater sequencing depth and finer resolution (1 kb), Rao et al. [233] observed TADs with a median size of 185 kb, and similar to Dixon et al., concluded that the domains were conserved in human and mouse B-lymphoblasts. However, the evidence for conservation in both studies is not strong. First, the actual conservation reported, though described as high, seems in fact to be modest: Dixon et al. reported that 54% of human boundaries are shared with mouse (76% if the comparison is reversed), and Rao et al. reported that 45% of mouse domains are shared with human. Second, and more importantly, in both studies, conservation estimates were made unilaterally, by considering the proportion of TADs identified in the species for which they had less data that are also identified in the species for which they had more data. This approach results in an overestimate of sharing of domains because only the very strong TADs can be identified in the species with less data, and these are more likely to be shared across species. Indeed, if we perform a similar analysis using our own data (assessing sharing of the top 10% of TADs identified in one species), we observe a much higher conservation (85%). Conversely, if we use the data from Rao et al. to estimate reciprocal TAD sharing across species, conservation is even lower than originally reported, at $\sim$30% instead of 45%.

Thus, based on our analysis of the literature, we believe that the common notion that TADs are highly conserved in their placement across species is not well supported. Indeed, recent evidence from yeast [290], different Drosophila tissues [242], and plant species [71]

suggests that TADs and TAD-like domains may not be particularly conserved, which raises questions about the stability of these higher-order structures and the significance of their role in the evolution of gene regulation across different lineages. However, the extent (or lack) of inter-species TAD conservation is difficult to falsify with existing data, partially because there is no standard method for identifying TADs, nor for comparing them across species [56, 89]. The ability to reliably identify TADs also depends on the quality of the genome assemblies used, the approach for inferring synteny, sequencing depth and coverage, and various other parameters. We acknowledge that our estimates of inter-species differences in TADs may be somewhat inflated due to incomplete power to detect TAD structures in each genome. Unfortunately, the outputs of the available algorithms do not allow us to directly address this potential caveat in the same way we addressed incomplete power in the comparative analysis of lower-order interactions.

More generally, as many studies indicated [56, 315, 247], including ours, it is difficult to reconcile the visual examination of contact maps with TADs inferred based on algorithms. In our case (Fig 2.5 and 2.17 Fig), we found many examples where visual inspection naively suggests high conservation, but the algorithms do not indicate sharing of domains or boundaries. This is not surprising; a previous comprehensive analysis of numerous TAD algorithm inferences found very little concordance when compared to manual visual annotations of TADs [56]. Obviously, comparing all TAD inferences based on manual visual assessment is not feasible. Yet, the lack of stability of TAD algorithms means that it is possible that a better computational analysis will emerge and will indicate that domains or boundaries are indeed conserved. Currently, however, neither our own nor previously published data provides support for strong conservation of these structures.

2.4.1 Contribution of variation in 3D genome structure to expression divergence

We considered our Hi-C data along with gene expression data previously collected from the same cell lines [217] and assessed the extent to which inter-species variation in 3D genome contacts could potentially explain gene expression divergence between species. Previous studies have observed that spatial co-expression of genes is associated with chromatin interaction profiles [7, 72, 121, 252, 75]. A number of studies have focused on differentially expressed genes following a treatment or perturbation and observed that such genes are often associated with corresponding differences in nearby chromatin contacts [66, 41]. Consistent with these reports, we found an enrichment of inter-species differences in pairwise chromatin contacts that involve promoters of differentially expressed genes between the species. Our observations are robust with respect to a range of data processing decisions and the statistical cutoffs we used. Under the common assumption that changes in chromatin contacts are more likely to explain differences in gene expression than vice versa, our results support the notion that species-specific 3D genomic contacts play an important role in the evolution of gene regulation.

Our observation that inter-species differences in pairwise genomic contacts are associated with regulatory evolution more than differences in large scale TAD boundaries is also consistent with previous reports. For example, Rao et al. [234] found that the degradation of cohesin, one of the proteins involved in maintaining TAD boundaries and large-scale loops, is associated with only modest effects on gene expression. In contrast, a number of other studies found strong correlations between differences in fine-scale genomic contacts and differences in the expression of nearby genes [233, 134].

Previous studies have identified a wide variety of regulatory phenotypes that contribute to inter-primate differences in gene expression levels [301, 170, 222, 214, 331, 22, 31]; 3D genome conformation is only one of the putative upstream factors in the evolution of gene regulation.

Our results argue for a model whereby inter-species differences in pairwise contact frequencies are among the main drivers of expression divergence between humans and chimpanzees. Given the low 10-kb resolution of our Hi-C data, it is likely that we have underestimated the contribution of inter-species variation in 3D genome structure to gene expression divergence between species. Future comparative Hi-C studies that sequence deeply enough to obtain higher, sub-kilobase resolutions, will allow researchers to resolve variation in contact frequency at even smaller scales, augmenting predictive power.

2.4.2 Functional annotations

Finally, we considered our data in the context of functional chromatin annotations available for the human genome. Previous studies have shown that 3D contact maps produced by Hi-C can be accurately recapitulated by epigenetic marks [221, 333]. Other reports have found enrichments for various chromatin accessibility and histone marks among interactions inferred from chromosome conformation capture data [244, 241].

Our results corroborate and expand upon these findings. The differences we observed in chromHMM state assignments in our comparisons (namely, more active and less repressive states in promoter-involved contacts and contacts overlapping differentially expressed genes), provide additional support for the functional relevance of our inferences. We acknowledge that these differences could potentially be more pronounced with higher-resolution Hi-C data and with chromHMM inferences made from ChIP-seq experiments in the same cell lines. While our study design does not allow us to directly infer causality between chromatin interactions and gene expression, the functional enrichments we observed for different epigenetic marks suggest that 3D genome conformation may be one of the upstream elements in the chain of events driving the evolution of gene expression. Although this notion is intuitive to us and is consistent with our data, it is still possible that differences in epigenetic marks are the true drivers of divergence in gene expression levels and/or chromatin contacts

between humans and chimpanzees.

Future studies integrating similar data types could explore these possibilities by examining epigenetic marks across species (only human data were available to us), which would enable researchers to polarize the regulatory differences in orthologous sequences between humans and chimpanzees. This would also allow for a sharper definition of the functional classes of inter-species differences in lower-order chromatin contacts.

2.5 Materials and methods

2.5.1 Ethics statement

We collected human fibroblasts with written informed consent obtained from all human participants under University of Chicago IRB protocol 11—0524. We obtained fibroblasts from chimpanzees from the Yerkes Primate Research Center of Emory University under protocol 006—12, in full compliance with IACUC protocols [240]. All experimental methods are in accordance with the Helsinki Declaration.

2.5.2 Induced pluripotent stem cells (iPSCs)

As described previously, the Gilad lab has derived panels of both human and chimpanzee iPSCs via episomal reprogramming [240]. To ensure their quality, we validated iPSCs from both species as pluripotent at high passages (>10). Quality control checks included an embryoid body assay confirming their ability to differentiate into all three germ layers, qPCR of endogenous transcription factors associated with pluripotency, PCR to confirm the absence of exogenous pluripotency genes (both from residual episomal plasmid or genomic integration), and PluriTest [197], a bioinformatics classifier that assesses pluripotency based on gene expression data [240]. In the current study, we grew all cell lines in the same incubator in two passage-matched batches, which were also balanced across species and sex, in order

to avoid batch effects in our data.

2.5.3 In-situ Hi-C library preparation and sequencing

We performed *in situ* Hi-C with the restriction enzyme MboI, as previously described [233] on the iPSCs from both species. We grew cells in feeder-free conditions [204] to approximately 80% confluence before adding formaldehyde to crosslink the proteins mediating DNA-DNA contacts. We flash-froze pellets of 5 million cells each before beginning the in situ Hi-C protocol [233]. We used MboI to cut the DNA at each of its 4-bp recognition sites (GATC) throughout the genome. Ligation of proximal fragments with T4 DNA ligase yielded chimeric DNA molecules representing two distinct loci. Libraries were created in two balanced batches identical to the cell growth batches and sequenced (100bp paired-end) on an Illumina Hi-Seq 4000 at the University of Chicago Genomics Core Facility. To avoid batch effects resulting from differences in flow cells, libraries were sequenced across three lanes, each on separate flow cells balanced for species.

2.5.4 Hi-C read mapping, filtering, and normalization

We preprocessed, mapped, and filtered the resulting FastQ sequence files using HiCUP version 0.5.9 [306]. We also used HiCUP to truncate the reads at ligation junctions. Thereafter, we used bowtie2 version 2.2.9 [157] to independently map the two mates of paired-end sequences to either the hg38 or panTro5 genomes, and removed reads with low quality scores (MAPQ < 30). We carried out further HiCUP filtering as previously described based on an *in silico* genome digest in order to remove experimental artifacts [306]. We then used HOMER version 4.9.1, a foundational statistical analysis suite for Hi-C data [117], to tile the genome into a matrix of 10 kb bins and assign reads to their corresponding intersecting bins. We subsequently used HOMER to normalize Hi-C contact bins as previously described [117], accounting for known technical biases in Hi-C data. Finally, we called statistically significant

interactions independently in each individual using HOMER, based on a null expectation of read counts falling into bins in a cumulative binomial distribution [117]. We retained interactions with an unadjusted P value ≤ 0.01, the default recommendation by HOMER. As other studies have noted [74], a traditional multiple testing correction paradigm is overly conservative for Hi-C data due to the high number of tests, and because the spatial nature of the data means that individual tests are highly correlated (and thus not independent).

2.5.5 Creation of a union list of orthologous Hi-C contacts across species

In order to ensure that the contact frequencies we compared across species were from representative orthologous sequences in humans and chimpanzees, we used liftOver with a reciprocal best hits method [300, 140] to transfer interaction bin coordinates across both genomes. For each called contact, we used liftOver to independently map the coordinates of the two anchor bins in the other species' genome, obtaining coordinates in both genomes for all contacts. We then rounded the coordinates to the nearest 10 kb bin, in order to align properly with a Hi-C bin. We required both anchor bins to have orthologous bins in the other species in order to retain a contact for comparison; statistics on the number of called contacts and the number retained after our liftOver procedure are available in S9 Table. In order to assess the extent of contacts lost due to lack of orthology, we also compared the retention of genome-wide 10 kb bins in both genomes with the retention of unique 10 kb bins found within each of our individuals. We found that our Hi-C bins tended to have a higher rate of orthologous mappability across species (S9 Table). For all contacts in this union list, we then extracted the HOMER-normalized interaction frequencies from each individual's 10 kb Hi-C matrix. Including interactions discovered in fewer than 4 individuals increased the variance in our data (2.14 Fig). Therefore, we retained only the Hi-C contacts that were independently discovered by HOMER in at least 4 individuals, for a total of 347,206 interactions. As we describe in the next section, we also later filtered out contacts where the distance

between bins showed a difference of > 20 kb across species, retaining 292,070 interactions.

2.5.6 Linear modeling of Hi-C interaction frequencies

In an effort to quantify inter-species differences in the Hi-C interaction frequency values, we used the following linear model:

$$Y_{ij} = \beta_0 + \beta_{sp} s_i + \beta_{sx} x_j + \beta_{btc} b_i + \epsilon_{ij} \tag{2.1}$$

Y_{ij} represents the observed Hi-C interaction frequency of a contact from individual j in species i. β_0 is the intercept. β_{sp}, β_{sx}, and β_{btc} are effect sizes for species, sex, and batch, respectively, with their corresponding variables s_i, x_j, and b_i, and an error term ϵ_{ij}. We used the R/Bioconductor package limma [267, 159] to test for inter-species differences in Hi-C interaction frequency. We applied Benjamini-Hochberg multiple testing correction and found 13,572 interaction pairs where the species term is significant at a 5% false discovery rate (FDR).

Initial visualization of the linear modeling results for the species term revealed a stark asymmetry (2.16A Fig) suggesting that on a global level, the contacts identified as significant in chimpanzees were much stronger than those identified in humans. This was surprising to us; we reasoned that this asymmetry could be due to a technical factor. For example, liftOver conversion of genome coordinates between species to identify orthologous bins can create differences in both the Hi-C locus size and in the genomic distance between mates of a contact pair (mate-pair distance). We investigated the impact of these two factors on the proportion of contacts classified as differential across species in our data. We discovered that while changes in Hi-C locus size had little effect on the proportion of interspecies DCs, differences in mate-pair distances > 20 kb across species created a noticeable inflation in this proportion at an FDR of 5% (2.16B Fig). We believe this makes intuitive sense, as bins that are farther apart will have fewer read counts due to the proximity-based ligation in Hi-C.

Thus, a mate-pair distance difference across the genomes could induce what appears to be a differential contact, because the contact inherently has more read support in the species where the mates are closer. However, we note that it is impossible to ascertain the relative biological and/or technical relevance of the differences seen in these contacts. We thus took a conservative approach to minimize false positives and removed contacts with a >20 kb mate-pair distance difference between species from our downstream analyses (2.16 Fig), accepting that the number of inter-species differences we observe may be underestimated.

2.5.7 Identification of orthologous topologically associating domains (TADs) and boundaries

We chose to perform TAD analyses on both individual-level data and on representative species consensus data. For our analysis comparing TAD boundaries on species consensus Hi-C maps, we combined all the preprocessed Juicer files from all our individuals within a species and used the *juicer_mega.sh* script [74] to create higher density contact maps for each species. We then ran the Arrowhead algorithm across resolutions to infer TADs, and then we extended the edges of TADs 7.5 kb in each direction to create 15 kb boundaries (accounting for imprecision in boundary inference). We used a reciprocal best hits liftOver strategy [300, 140] to obtain orthologously mappable TADs and boundaries. To confirm high synteny of large-scale linear genomic intervals between the species, we employed this same orthology analysis on genome-wide tilings of the hg38 and panTro5 genome assemblies, with varying window sizes created with bedtools [229] *makewindows* (2.18 Fig). In the case of TADs, we then assessed number of domains found in one species that were also found in the other species (conserved domains) with reciprocal bedtools [229] *intersect –c –f 0.9 -r* calls. These parameters will only define a domain as overlapping if there is a domain in the other species such that each domain shares 90% of their interval with the other. We used the larger of the two estimates of shared TADs across the species as the conserved domain count (to

be conservative), and divided this by the sum of the conserved and species-specific domains identified in order to assess conservation. As an alternative analysis, we also employed the method previously described by Rao et al. [233]; namely, we called a TAD conserved in one species if it and a TAD from the other species displayed a Euclidean distance less than the smaller of 50 kb or half the given TAD's size. We analyzed boundary conservation using bedtools *intersect –c*, considering any overlap as indication of conservation (i.e. even a single base pair overlap of boundaries meant a boundary was classified as conserved).

To examine individual-level data and to ensure robustness of our results, we separately used both Arrowhead [74] and TopDom [261] (with *window = 20*) across resolutions to call TADs independently in each individual sample. Though we performed essentially the same analyses on both outputs, it should be noted that Arrowhead provides nested TADs only, from which we inferred boundaries as described above, whereas TopDom provides separate domain and boundary inferences. We used a reciprocal best hits liftOver method [300, 140] to obtain a set of orthologous domains and boundaries. We assessed interspecies conservation by performing left outer joins (bedtools *intersect –loj*) of each individual's domains against all the others, once again requiring 90% reciprocal overlap. We then took the average species-specific and shared domain counts across these individual comparisons to produce a single estimate of conservation (2.19 and 2.21 Figs). The individuals' pairwise percentages of shared domains were used in hierarchical clustering analysis (Fig 2.4C, 2.19 and 2.21 Figs). We also once again checked the robustness of our results using the conservation calling method from Rao et al. [233] described above. In the case of boundaries, we reasoned that, given the nested nature of the TADs, as well as variance between individuals in their exact placement, it would make sense to merge the boundaries (using bedtools *merge*) in order to obtain a list of unique boundary elements. We added a column of individual identifiers to each set of boundaries and then merged all together, thereafter assessing conservation by examining what percentage of boundaries were independently found in both species out of the total set

of unique boundaries. We also applied hierarchical clustering analysis to individual pairwise percentages of shared boundaries in this union merged file (Fig 2.4D, 2.19 and 2.21 Figs). Further descriptions of these analyses can be found on our GitHub repository (/data/TADs folder), and 10 kb individual Arrowhead inferences are available in S17–S24 Tables.

2.5.8 Differential expression analysis

Previously, the Gilad lab generated RNA-seq expression data on the same iPSC lines from this study (GEO accession GSE110471 [217]). We computed reads per kilobase per million mapped reads (RPKM) for every gene, as the orthologous genes are not constrained to be the same length across species. We retained 11,074 genes that had at least half of the individuals (2 observations) in each species with $\log_2$RPKM ≥ 0.4. We then used the limma-voom pipeline with precision weights [267, 159] to test for differential expression across species, using a linear model including a species effect and a sex effect. Using this approach, we found 2,086 differentially expressed genes (at 5% FDR).

2.5.9 Broad integration of Hi-C and gene expression data

We obtained the overlap between our gene expression data and our Hi-C data by applying bedtools *overlap* [229] to the Hi-C loci and the first exon of each gene. Using a curated file of orthologous gene coordinates between humans and chimpanzees [217], we extracted a one-base-pair interval at the beginning of each first exon to use as a proxy for transcription start sites (TSSs).

As described in the main text, the difference in dimensionality between the two datasets presented a challenge. While every gene has only one expression value per individual, a given Hi-C locus can and frequently does make contact with many other loci. When a given gene overlapped a Hi-C locus making multiple contacts, we chose the contact with the smallest species term FDR (i.e. the most species-specific contact) in our DC analysis to represent the

interaction frequency for that gene. Accordingly, we interpreted the FDR-adjusted P value for the chosen contact as the gene's differential contact significance. To examine correlations between normalized Hi-C contact frequency and $\log_2$RPKM gene expression, we considered the correlation between gene expression values across all 8 individuals with the corresponding interaction frequency values across the same 8 individuals.

2.5.10 Enrichment of differential expression in differential contacts

We examined the enrichment of differential expression in genes with differential contact (Fig 2.6A and 2.27C Fig) across a continuous range of DC FDRs and a discrete range of DE FDRs (1%, 2.5%, 5%, 7.5%, and 10%). We used Pearson's chi-squared test to quantify significance of the enrichment at each FDR (Fig 2.6B and 2.27D Fig). We also examined the reciprocal enrichment; that is, DC enrichment amongst DE genes (2.27A and 2.27B Fig).

2.5.11 Assessing the quantitative contribution of Hi-C contact frequencies to gene expression levels

We assessed the hypothesis that expression divergence may be mediated by contact frequency using linear models [9]. The intuition behind this approach is that the effect of species (X) on expression (Y) can be partitioned into its indirect effect on expression mediated through contact frequency (M) and its direct effect on expression. Therefore, a significant indirect effect would suggest that expression divergence is causally mediated by contact frequency. To test our mediation hypothesis, we computed the indirect effect of species on expression ($X \rightarrow M \rightarrow Y$: causal effect of X on Y through M) by taking the product of the effect of species on contact frequency (α: $X \rightarrow M$) and the effect of contact frequency on expression after controlling for species (β: $M \rightarrow Y$). The indirect effect ($\alpha^*\beta$) is conceptually equivalent to the difference between the effect of species on expression and the effect of species on expression after controlling for contact frequency, but is more mathematically tractable and

commonly used in mediation analyses [269, 268, 178]. We obtained α as the species effect size in a simple linear model attempting to predict Hi-C interaction frequency based solely on a species term. We estimated β as the contact frequency effect size in a linear model predicting expression based on both species and contact frequency per gene. To determine statistical significance of the indirect effect, we applied the Monte Carlo test of significance to construct the 95% confidence interval. The primary benefits of the Monte Carlo method are that it requires no distributional assumptions of the data and is robust against type I error in small samples [179, 227, 226]. Thus, we choose the Monte Carlo test over Sobel test, the conventional approach to significance testing of mediation, which relies on the data following normal distribution [269, 268].

2.5.12 Integration with epigenetic annotations

We obtained chromHMM 15-state model peak calls in human iPSC cells from ENCODE [48] (S15 Table). We subsequently found the overlap between the human coordinates of our orthologous Hi-C contact loci and the chromHMM peak calls and quantified the extent of base pair overlap between each locus and all overlapping chromHMM peaks. We assigned each individual locus a single chromHMM annotation based on the peak with the highest base pair overlap with that locus. However, the distribution of overlaps of different chromHMM annotation peaks with our Hi-C bins were quite variable in size. To account for this, we normalized each annotation's overlap length in each locus by multiplying it by the reciprocal of its mean base pair overlap across all our bins (2.24 Fig). After removing duplicate Hi-C loci, we then assigned individual loci to chromHMM annotations based on these normalized base pair overlaps. We started with a small set of the top ten most differentially contacting loci (i.e. the ten lowest FDR loci from our Hi-C linear modeling), and tabulated proportions of which annotations were represented amongst them. We then iteratively added the next-lowest FDR contact (i.e. two Hi-C loci at a time) to this tabulation, re-calculating

proportions on the new set of contacts. We ran this same cumulative proportions analysis separately on contacts not overlapping promoters, contacts overlapping promoters, contacts overlapping promoters of DE genes, and contacts overlapping promoters of genes that were not DE (Fig 2.7A and 2.7B, 2.26 Fig).

We also obtained data on H3K4me1, H3K4me3, and H3K27ac collected in human iPS-18A cells, and data on H3K27me3 and DNase hypersensitivity sites collected in H1-hESCs, all from ENCODE [48] (S15 Table). We used bedtools *intersect* [229] to find the base pair overlap between each of these different marks and our Hi-C contact loci. We then removed duplicate Hi-C loci from the dataset and used a pairwise t-test to identify significant differences in the overlapping distributions for different sets of Hi-C classes (based on differential contact and differential expression, Fig 2.7C and 2.7D).

2.6 Acknowledgments

We thank the members of the Gilad, Nobrega, and Stephens labs for helpful discussions, particularly Matthew Stephens, Bryan J. Pavlovic, Débora R. Sobreira, Abhishek Sarkar, and Lindsey E. Montefiori. We thank Natalia Gonzales for help editing the paper.

2.6.1 Author contributions

IEE and YG conceived of the study and designed the experiments. IEE performed the experiments. IEE analyzed the results (with input from CJH, KL, and LEB). YG supervised the project. IEE and YG wrote the paper. All authors read and approved the final manuscript.

2.7 Supplementary Figures

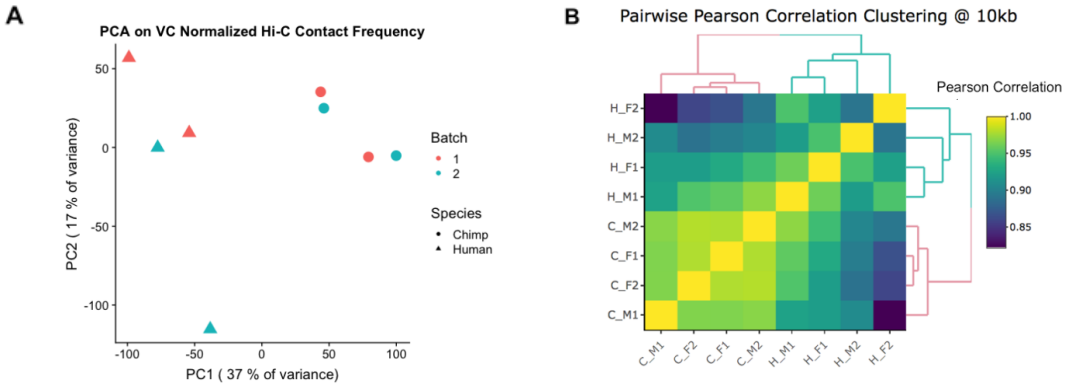

Figure 2.8: **S1. Regulatory landscapes cluster by species, Juicer.** (A) Principal components analysis (PCA) of Juicer vanilla coverage (VC)-normalized interaction frequencies for the union of all contacts in humans (triangles) and chimpanzees (circles). PC1 is highly correlated with species (r = 0.89; P = 0.0004). (B) Unsupervised hierarchical clustering of the pairwise correlations (Pearson's r^2) of Juicer VC-normalized interaction frequencies at 10 kb resolution. The first letter in the labels demarcates the species (H for human and C for chimpanzee), and the following symbols indicate sex (male, M or female, F) and batch (1 or 2).

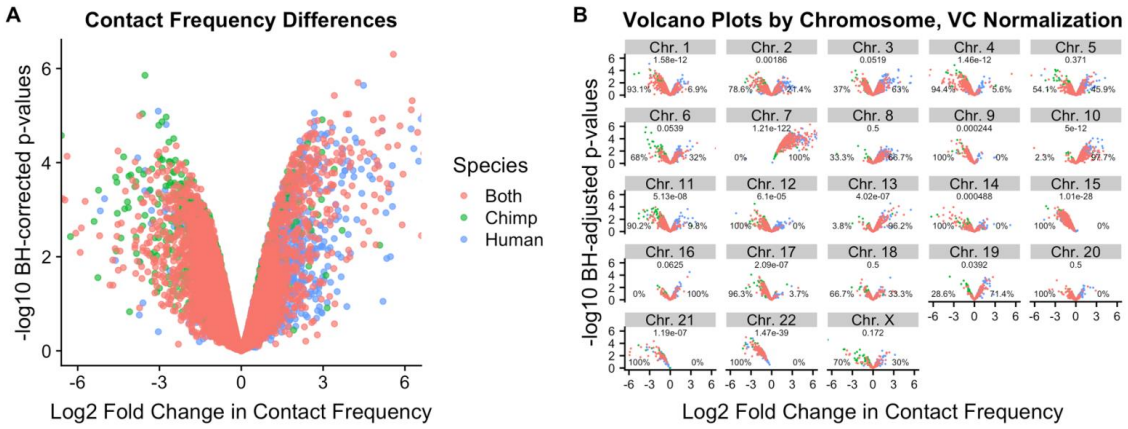

Figure 2.9: **S2. Linear modeling reveals large-scale chromosomal differences in contact frequency, Juicer.** (A) Volcano plot of $\log_2$ fold change in contact frequency between humans and chimpanzees (x-axis) against Benjamini-Hochberg FDR (y-axis), after filtering non-orthologous regions. Data are colored by the species in which the contact was originally identified as significant. (B) Per-chromosome volcano plot using the same legend as in A. P-values provided for a binomial test of the null that inter-species differences in contact frequencies are evenly distributed. The percentage of contacts with significant higher frequency in each species is noted. Of note is that many of the same chromosomal asymmetries in contact strength observed here are in the same chromosomes as those observed in the HOMER-normalized data (Fig 2.2).

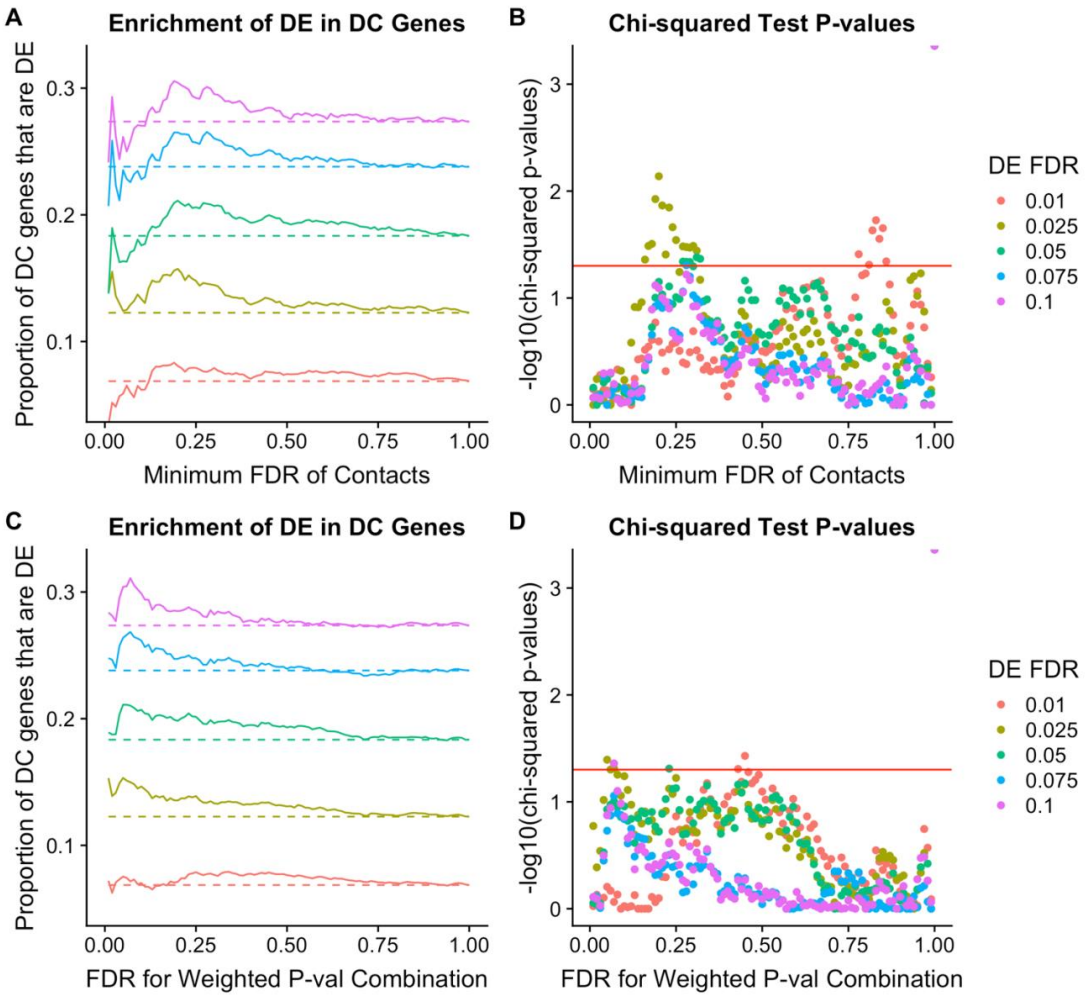

Figure 2.10: **S3. Differentially expressed genes show enrichment for differential Hi-C contacts, Juicer.** (A) Enrichment of inter-species differentially expressed (DE) genes with corresponding differences in Hi-C contact frequencies (DC) between the species. The proportion of DC genes that are significantly DE (y-axis) is shown across a range of DC FDRs (x-axis). Colors indicate different DE FDR thresholds, and dashed lines indicate the proportion of DE genes expected by chance alone. (B) P values of Chi-squared tests of the null that there is no difference in proportion of DE genes among DC genes (y-axis), shown for a range of DC FDRs (x-axis). In both panels, DC regions were chosen to have the minimum FDR supporting inter-species difference in contact frequency. (C) Same as A, but this time, a weighted p-value combination technique [189] was used to integrate each Hi-C bin's DC FDR across all of its contacts. (D) Same as B, but for the weighted p-value combination instead of the minimum FDR contact.

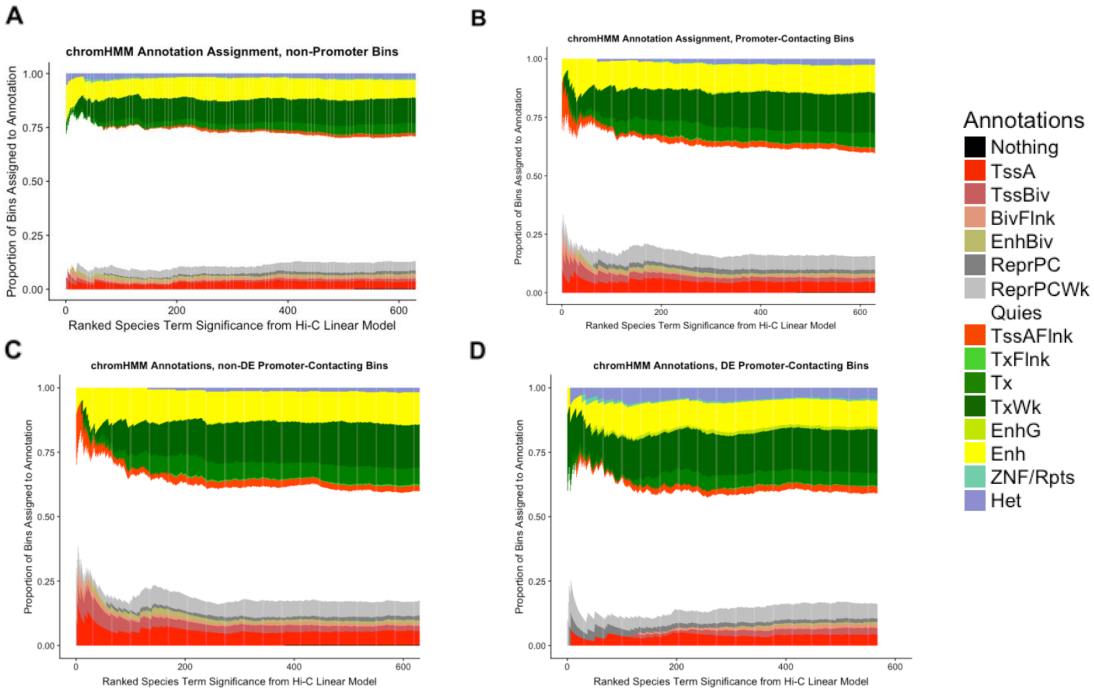

Figure 2.11: **S4. Dynamics of chromHMM state among significant Hi-C contacts, Juicer.** (A) Hi-C loci that do not make contact with promoters are ranked in order of decreasing DC FDR (x-axis). The y-axis shows cumulative proportion of chromHMM annotation assignments for all Hi-C loci at the given FDR or lower. (TssA-Active TSS, TSSBiv-Bivalent/Poised TSS, BivFlnk-Flanking Bivalent TSS/Enh, EnhBiv-Bivalent Enhancer, ReprPC-Repressed PolyComb, ReprPCWk-Weak Repressed PolyComb, Quies-Quiescent/Low, TssAFlnk-Flanking Active TSS, TxFlnk-Transcription at gene 5' and 3', Tx-Strong transcription, TxWk-Weak transcription, EnhG-Genic Enhancers, Enh-Enhancers, ZNF/Rpts-ZNF genes and repeats, Het-Heterochromatin). (B) Same as A, but only considering Hi-C loci making contact with promoter bins. (C) Same as B, but only considering Hi-C loci making contact with promoters of genes that are not differentially expressed (DE). (D) Same as C, but only considering Hi-C loci making contact with promoters of genes that are differentially expressed (DE).

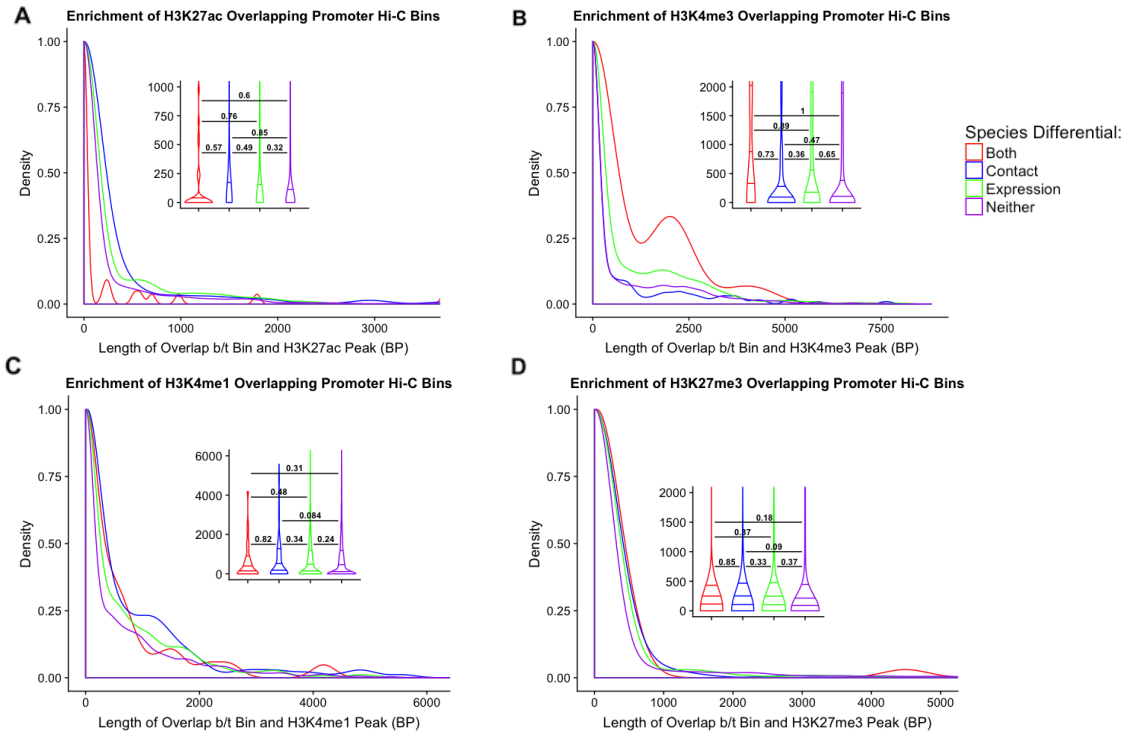

Figure 2.12: **S5. Overlap of activating and repressive histone marks among Hi-C contacts, Juicer.** (A) Density plot of the base pair overlap between different classes of Hi-C contact loci and H3K27ac. Histone mark data were obtained from ENCODE in experiments carried out in human iPSCs. We grouped contacts into 4 classes, indicated by color: those that show differential contact between species, those that show differential expression between species, those that show both, and those that show neither. We used pairwise t-tests to compare differences in the mean overlap among the four classes of Hi-C loci. Unlike in the HOMER-normalized data, we do not observe statistically significant differences in overlaps with H3K27ac between different locus classes. This may reflect the previous observation that the hiccups algorithm for assigning statistical significance of loops in Hi-C data is much more conservative than HOMER's significance calling method [89]. (B) Same as A, but performed on H3K4me3 data obtained from ENCODE, collected in hESCs. (C) Same as A and B, but performed on H3K4me1 data obtained from ENCODE, collected in human iPSCs. (D) Same as A, B, and C, but performed on H3K27me3 data obtained from ENCODE, collected in human iPSCs.

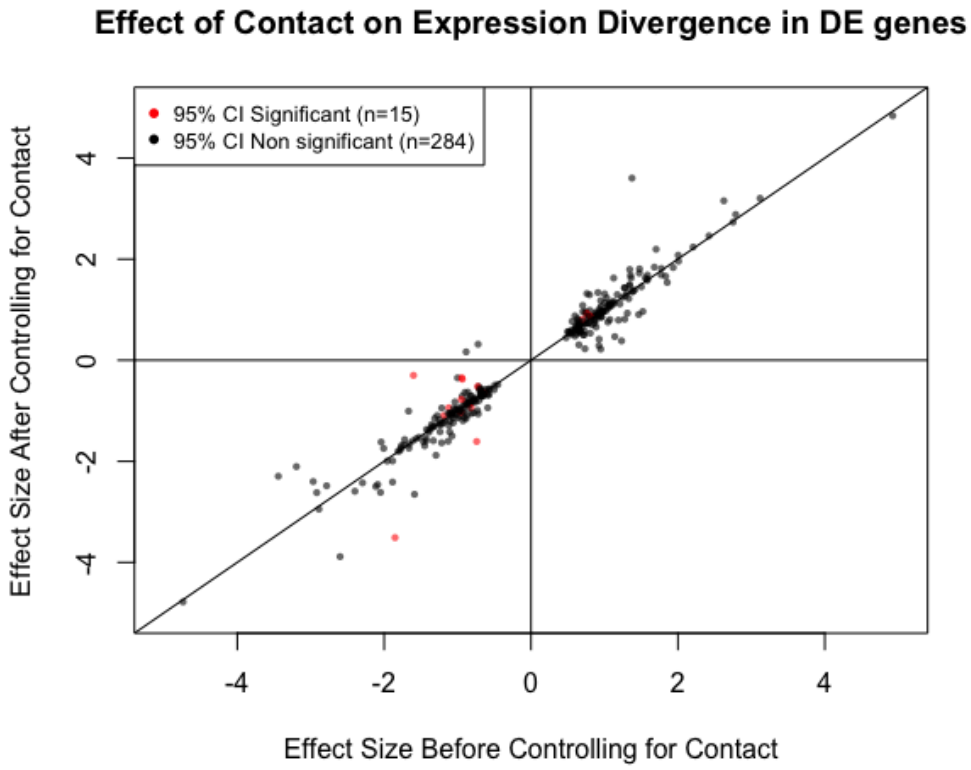

Figure 2.13: **S6. Gene expression variance is explained by chromatin contacts for 5% of DE genes, Juicer.** Plot of the species effect size in DE genes between models before (x-axis) and after (y-axis) conditioning on contact frequency. The Monte Carlo test of significance was used to construct the 95% confidence interval and evaluate the significance of the indirect effect (species' effect on expression mediated through contact). Amongst DE genes, 5% (15/299) of genes showed a statistically significant effect of Hi-C contacts on expression levels (i.e. their 95% confidence interval does not include zero).

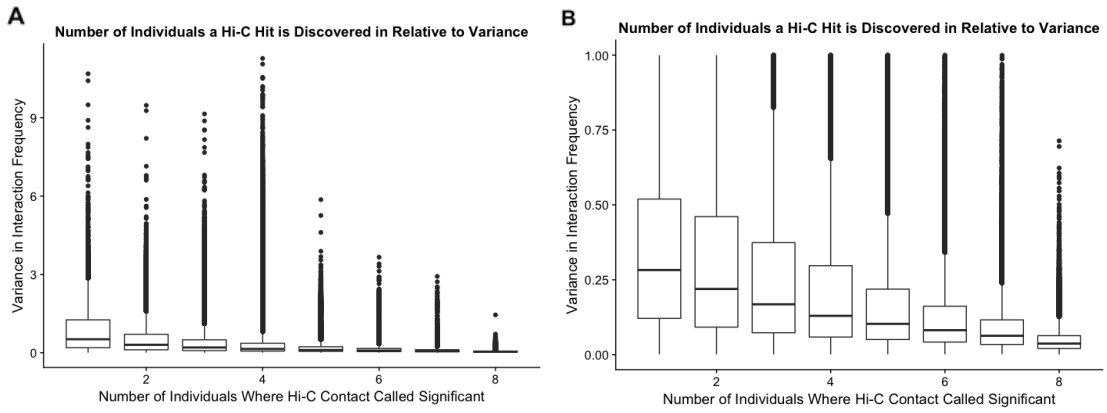

Figure 2.14: **S7. Variance in interaction frequency as a function of the number of individuals in which a significant interaction is independently discovered.** (A) Boxplots of variance in contact frequency across all 8 individuals on the y-axis, binned by the number of individuals in which an interaction is independently called significant on the x-axis. (B) Same as A, but zoomed in on the y-axis to visualize finer-scale variation.

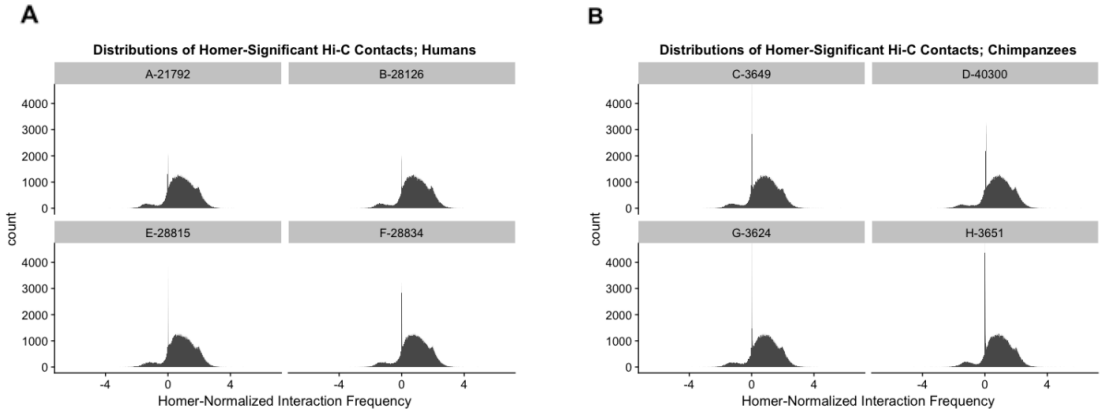

Figure 2.15: **S8. Distributions of HOMER-normalized interaction frequencies are remarkably similar across species.** (A) Histogram of $\log_2$(observed/expected) HOMER-normalized interaction frequencies in all four human samples used in this study, after applying pairwise cyclic loess normalization with limma [267]. (B) Same as A, but in chimpanzees.

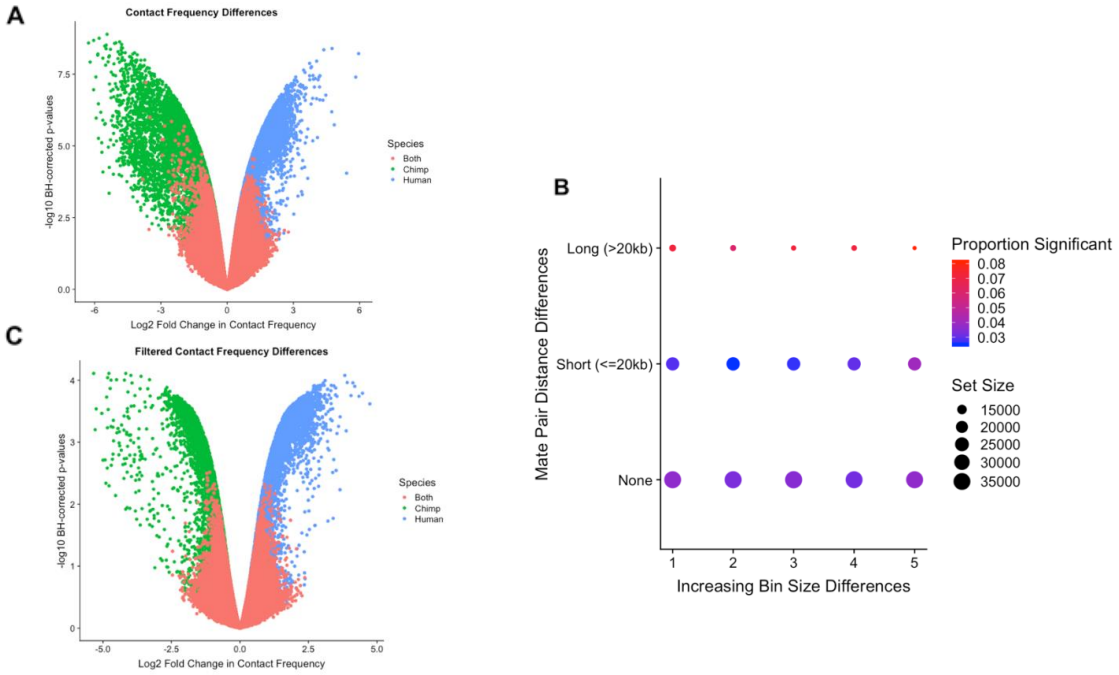

Figure 2.16: **S9. Volcano plot asymmetry quality control.** (A) Volcano plot of $\log_2$ fold change in contact frequency between humans and chimpanzees (x-axis) against Benjamini-Hochberg FDR (y-axis). This plot shows data only filtered for independent discovery in at least 4 individuals. Data are colored by the species in which the contact was originally identified as significant. (B) Scatter plot of sets of Hi-C contacts, with proportion of contacts significant in our linear modeling of interaction frequency shown based on color. Contacts are binned by mate-pair distance differences (y-axis) and bin size differences (x-axis). Circle size is proportional to the size of the set of Hi-C contacts falling into each criteria. Red indicates that the data were filtered out after this step, and blue/purple indicates that the data were retained for further analysis. (C) Volcano plot as in A, but after removing contacts with large mate-pair distance differences across the species.

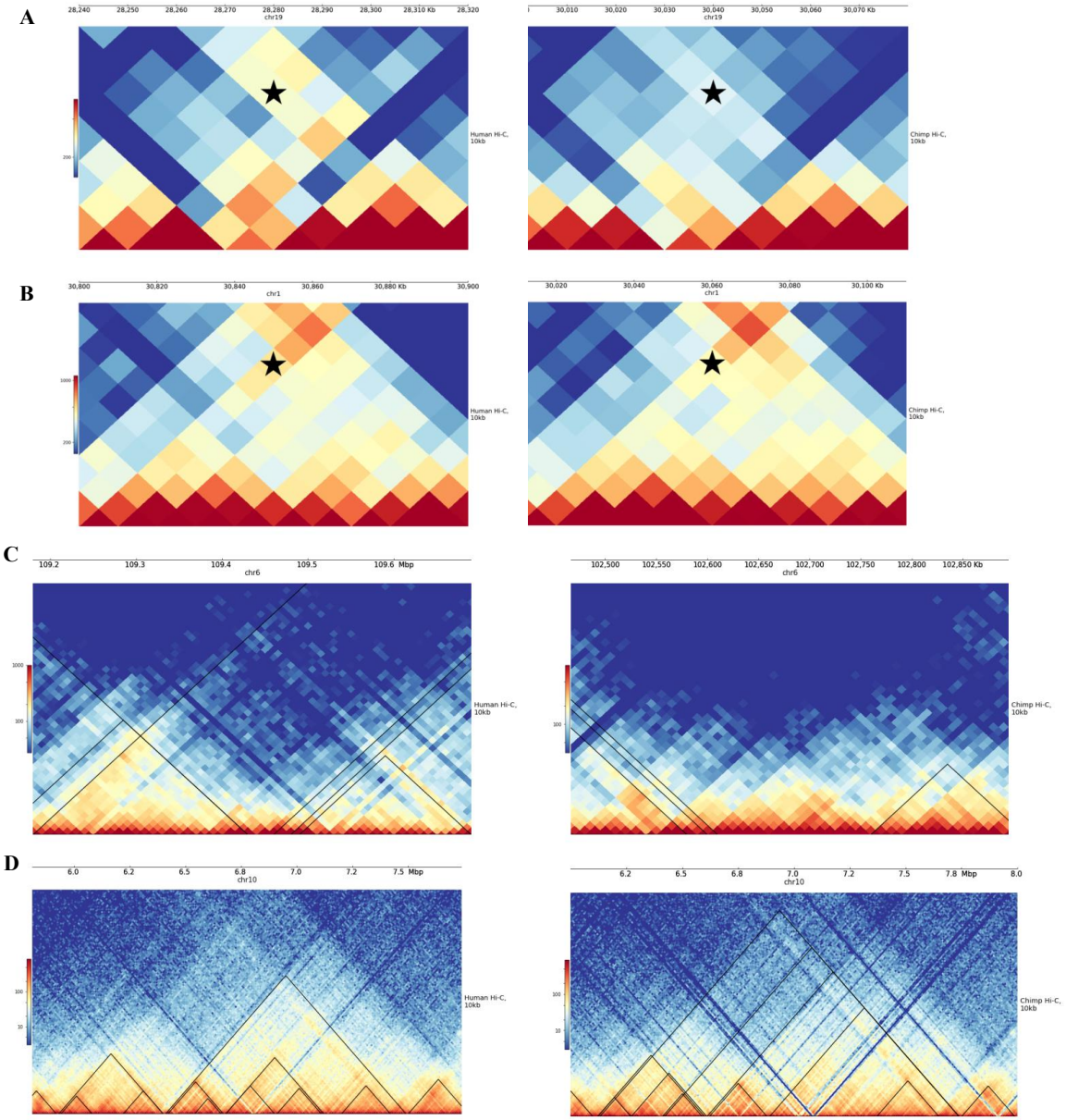

Figure 2.17: **S10. Further visual examples of DC and non-DC interactions; conserved and divergent TADs.** (A) PyGenomeTracks plots [232] of a chromosome 19 interaction between bins 80 kb away for human (left panel) and chimpanzee (right panel). The bin pair tested is indicated by a black star, and was found to be DC between species. (B) Same as A, but for a conserved (non-DC) interaction on chromosome 1 separated by 100kb...(continued on next page)

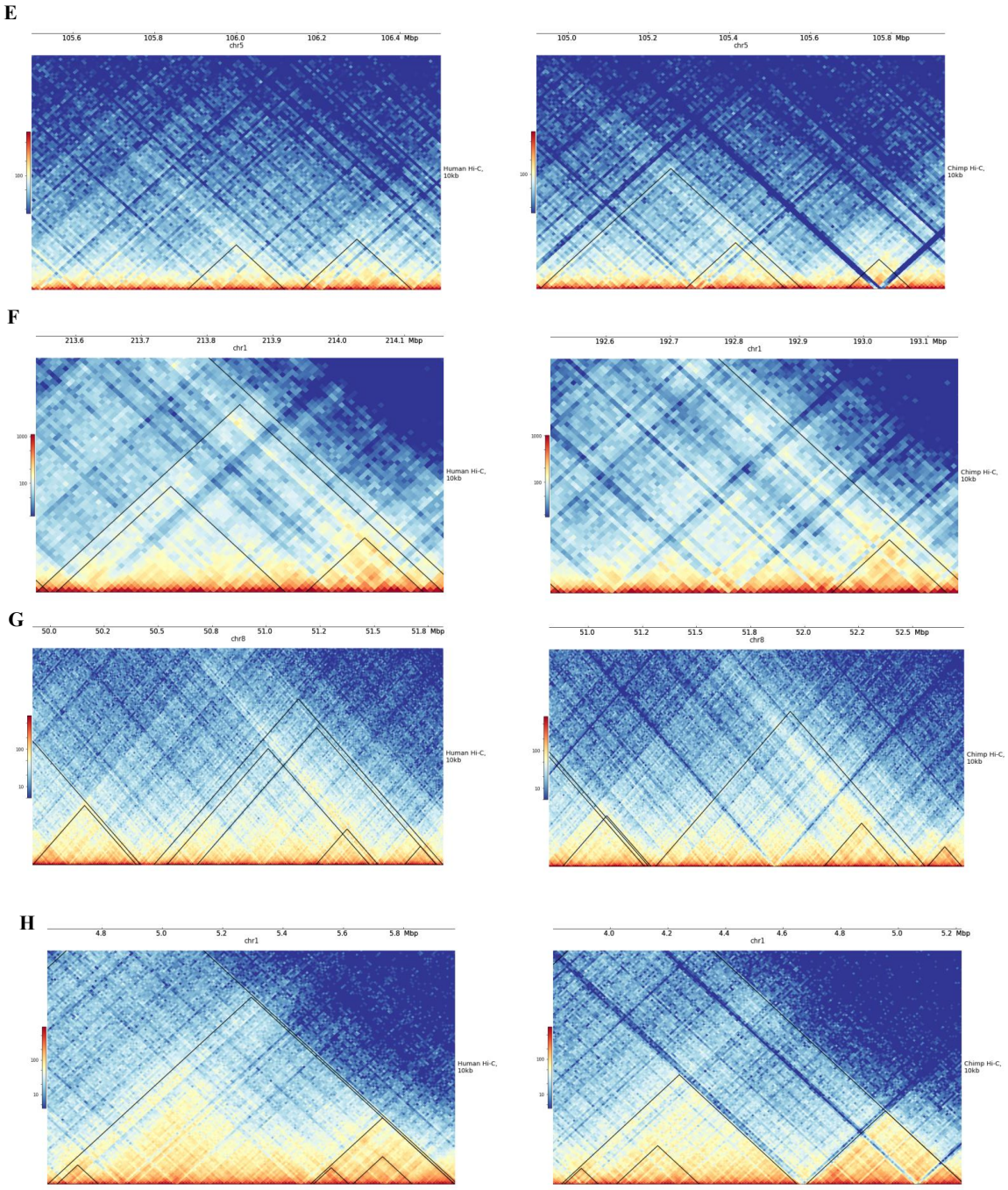

Figure 2.17: (continued from previous page) (C-H) Examples of contact maps (created with PyGenomeTracks [232]) and Arrowhead-inferred TAD structures (black lines) in humans (left) and chimpanzees (right), across a number of different chromosomes. In most examples, inference based on the algorithm indicates shared and species-specific domains, yet these are difficult to ascertain based on visual inspection, as discussed.

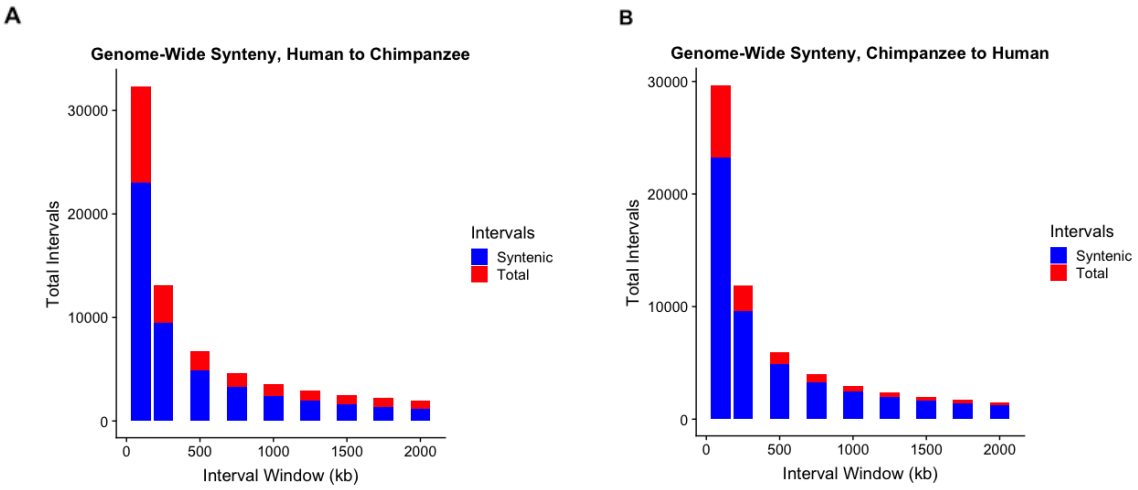

Figure 2.18: **S11. Synteny of large scale linear genomic intervals between human and chimpanzee.** (A) Across different window sizes (x-axis) for a genome-wide tiling of hg38, we plotted the number of total and syntenic linear intervals (y-axis), identified using the reciprocal best hits liftOver method [300, 140] we employed throughout the paper. (B) Same as A, but for a genome-wide tiling of panTro5.

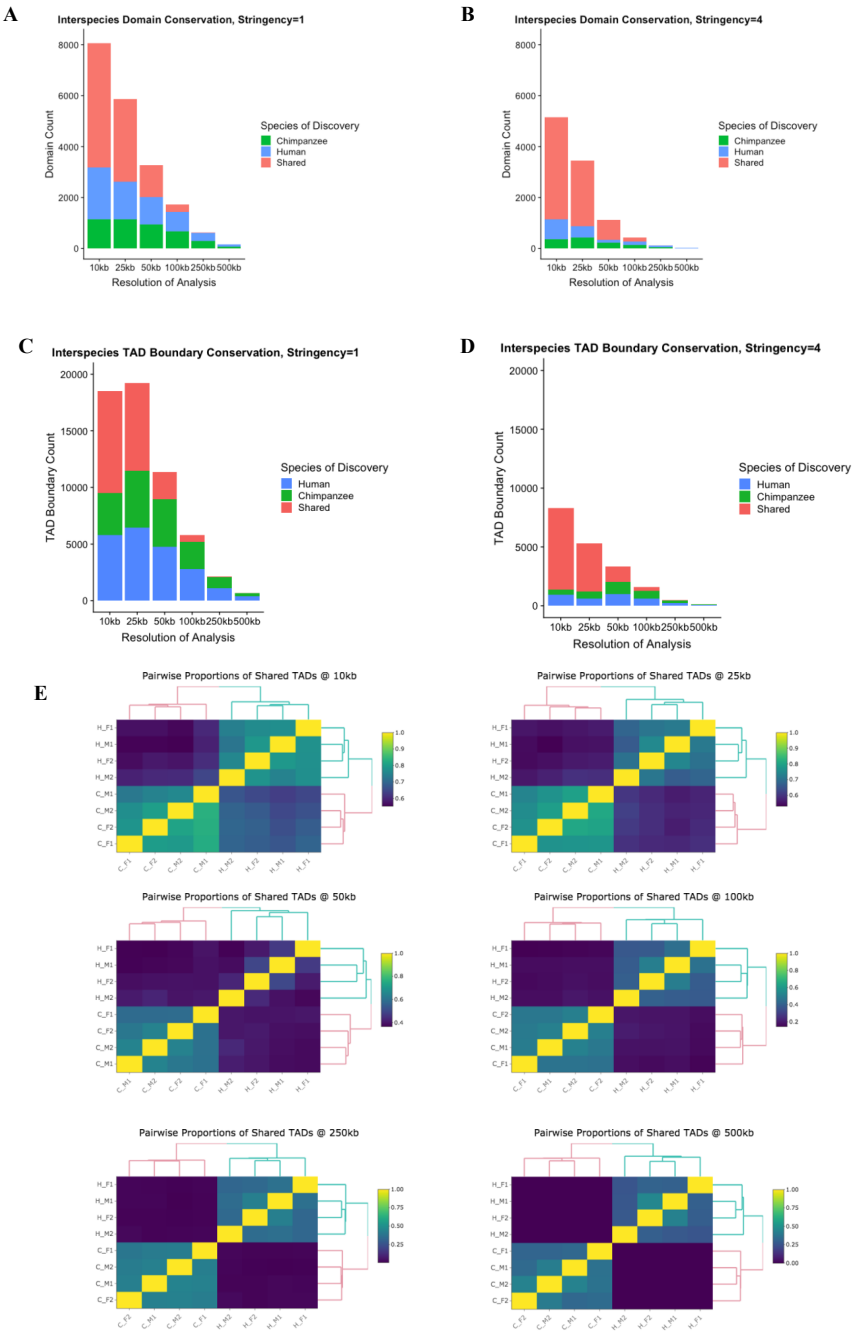

Figure 2.19: **S12. Higher-order chromosomal structure in humans and chimpanzees with alternative analysis choices.** (A) Across different resolutions (x-axis), we plotted the number of shared and species-specific domains (y-axis) identified with Arrowhead [74] on Juicer VC-normalized Hi-C maps from each individual. We called domain conservation here based on the method of Rao et al. [233] (highly similar results were observed with our 90% reciprocal overlap method, described in the text and available in the github repository associated with the paper)...(continued on next page)

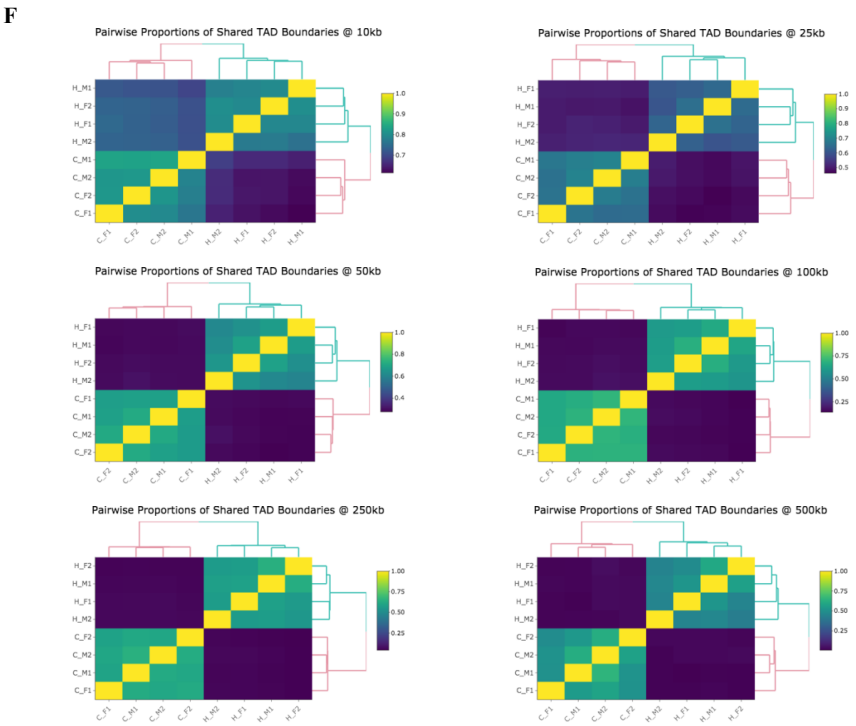

Figure 2.19: (continued from previous page) Domain count values represent the average interspecies sharing across all individuals, with no filtering for domain robustness (that is, assessing all domains discovered and orthologously mappable). Under this analysis paradigm we observe relatively low sharing across species (∼60% at 10kb). (B) Same as A, but this time, only considering TADs that were found across all 4 individuals within either one of the species (fixed TADs). Restricting to this subset increases the percentage of conservation to 78%, although the set of TADs being examined is much smaller. (C) Same as A, but for boundaries instead of domains. Boundaries were defined as 15kb flanking regions at the edges of inferred Arrowhead domains. Because the TADs called by Arrowhead are nested, we merged boundaries here to obtain unique genomic intervals, rather than counting boundaries repeatedly. We then considered boundaries shared between individuals if they had any overlap. (D) Same as B, but for boundaries instead of domains (i.e. considering only boundaries fixed within species). Here, the highest estimate of conservation we obtain is 83% of boundaries conserved across species at 10kb resolution. (E) Unsupervised hierarchical clustering of the pairwise proportions of shared TADs between all individuals in our study at a variety of resolutions, using the Rao et al. [233] methodology for calling conservation. The first letter in the labels demarcates the species (H for human and C for chimpanzee), and the following symbols indicate sex (male, M or female, F) and batch (1 or 2). Heatmaps are not necessarily symmetric because different numbers of TADs were discovered in different individuals; rows represent an individual's shared proportion of TADs (individual total) with each other individual. Highly similar clustering results were observed when using our domain conservation calling paradigm (shown in github repository associated with paper). (F) Same as E, but for boundaries instead of domains.

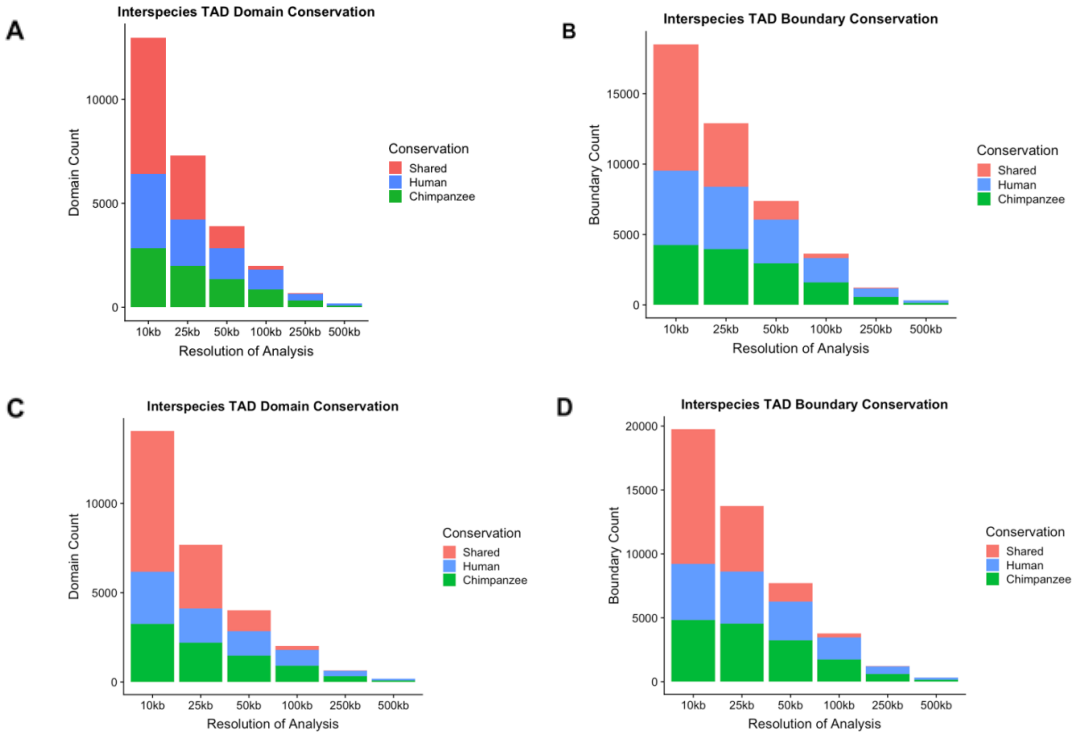

Figure 2.20: **S13. Higher-order chromosomal structure in humans and chimpanzees with alternative analysis choices and genome builds.** (A) Across different resolutions (x-axis), we plotted the number of shared and species-specific domains (y-axis) identified with Arrowhead [74] using the consensus map from each species. To call a domain as conserved here, we required that the Euclidean distance between the domain across species be less than the minimum of 50kb or 50% the length of the TAD, based on the conservation calling method employed by Rao et al [233]. Results are highly similar to those seen in Fig 2.4A. (B) Same as A, but for TAD boundaries instead of the domains themselves. Boundaries were defined as 15 kb flanking regions at the edges of inferred Arrowhead domains. In this case, conservation was called if there was any base pair overlap between boundaries. Unlike in Fig 2.4B, boundaries were merged before calling conservation, in order to find unique boundary elements. This difference in analysis paradigms could have important consequences with a nested TAD caller such as Arrowhead [74], but results are highly similar to those seen in Fig 2.4B. (C) Same as A, but this time, performed on 'high-density consensus' Hi-C maps that have been mapped to the hg38 and panTro6 genomes (rather than panTro5). Results are highly similar despite the improvement in genome quality build. (D) Same as B, but this time, on the hg38 and panTro6 genome assemblies.

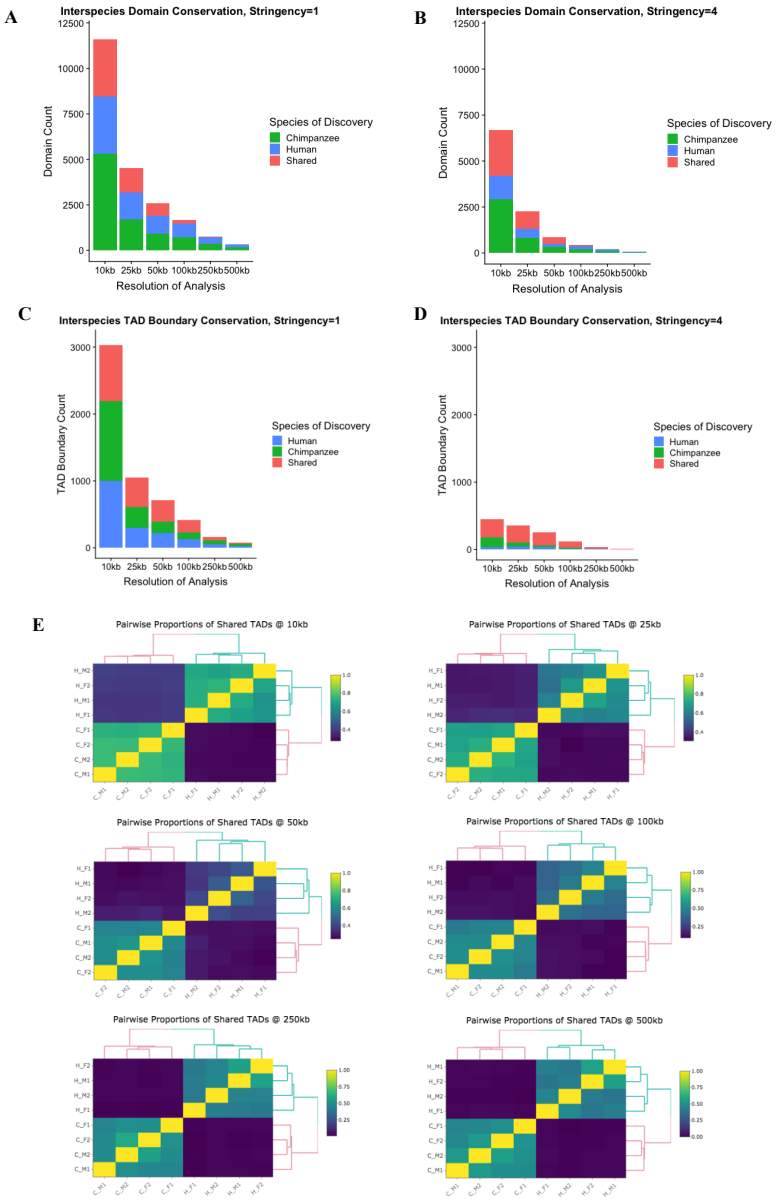

Figure 2.21: **S14. Higher-order chromosomal structure in humans and chimpanzees with alternative algorithms (TopDom).** (A) Across different resolutions (x-axis), we plotted the number of shared and species-specific domains (y-axis) identified with TopDom [261] on HOMER-normalized Hi-C maps from each individual. We called domain conservation here based on the method of Rao et al. [233] (highly similar results were observed with our 90% reciprocal overlap method, described in the text and available in the github repository associated with the paper)...(continued on next page)

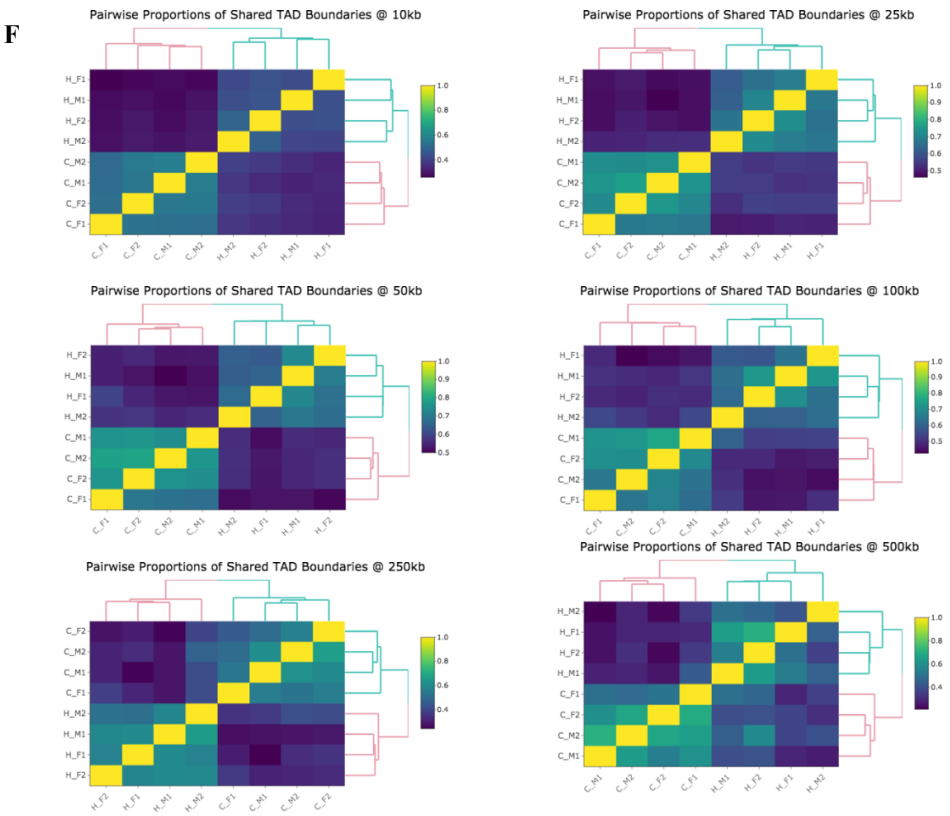

Figure 2.21: (continued from previous page) Domain count values represent the average interspecies sharing across all individuals, with no filtering for domain robustness (that is, assessing all domains discovered and orthologously mappable). Under this analysis paradigm we observe relatively low sharing across species (maximum of 30% at 25 kb). (B) Same as A, but this time, only considering TADs that were found across all 4 individuals within either one of the species (fixed TADs). Restricting to this subset increases the maximum percentage of conservation to 42% at 25 kb resolution, although the set of TADs being examined is much smaller. (C) Same as A, but for TopDom [261] boundary inferences instead of domains. We considered boundaries shared between individuals if they had any overlap. (D) Same as B, but for boundaries instead of domains (i.e. considering only boundaries fixed within species). Here, the highest estimate of conservation we obtain is 76% of boundaries conserved across species at 50 kb resolution. (E) Unsupervised hierarchical clustering of the pairwise proportions of shared TADs between all individuals in our study at a variety of resolutions, using the Rao et al. [233] methodology for calling conservation. The first letter in the labels demarcates the species (H for human and C for chimpanzee), and the following symbols indicate sex (male, M or female, F) and batch (1 or 2). Heatmaps are not necessarily symmetric because different numbers of TADs were discovered in different individuals; rows represent an individual's shared proportion of TADs (individual total) with each other individual. Highly similar clustering results were observed when using our domain conservation calling paradigm (shown in github repository associated with paper). (F) Same as E, but for boundaries instead of domains.

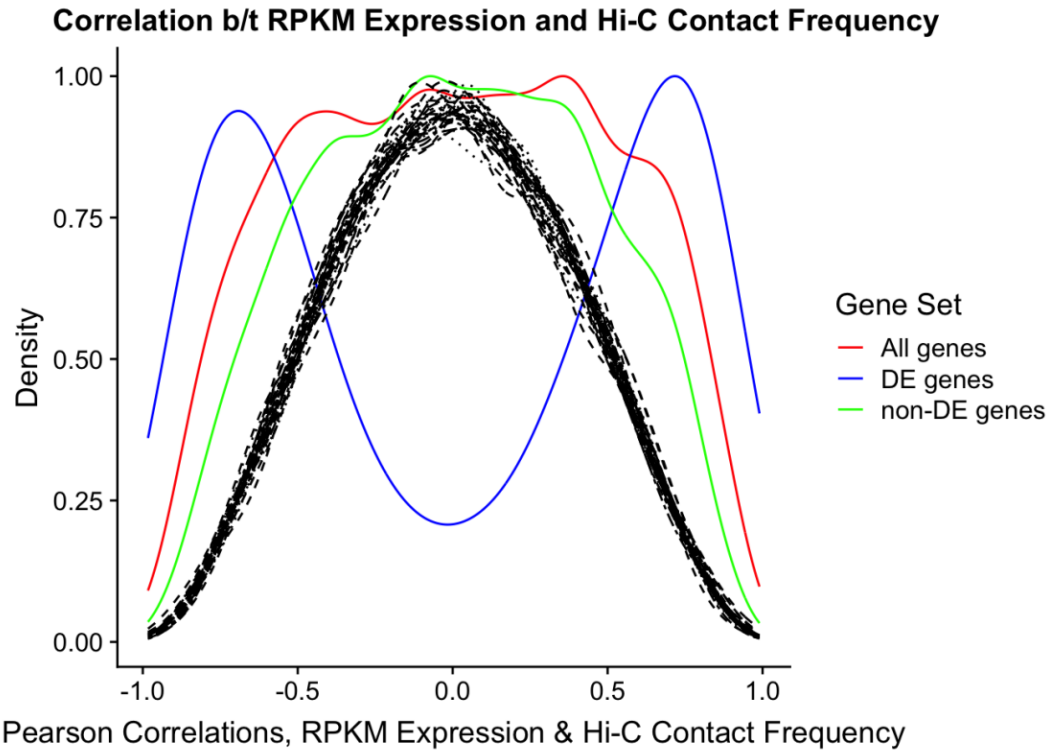

Figure 2.22: **S15. Correlations between Hi-C and expression.** Density of Pearson correlations between RPKM expression values and $\log_2$ HOMER-normalized contact frequencies across all 8 individuals. Solid lines indicate different sets of the observed data and dotted lines represent 10 permutations of the data. The Hi-C contact frequency chosen is that with the minimum FDR from linear modeling of contact frequency on species (see main text). The strong bimodal distribution of correlations between expression and contact suggests many instances where a contact difference between the species can lead to an increase (enhancer) or decrease (suppressor) of expression in the species where the contact is stronger.

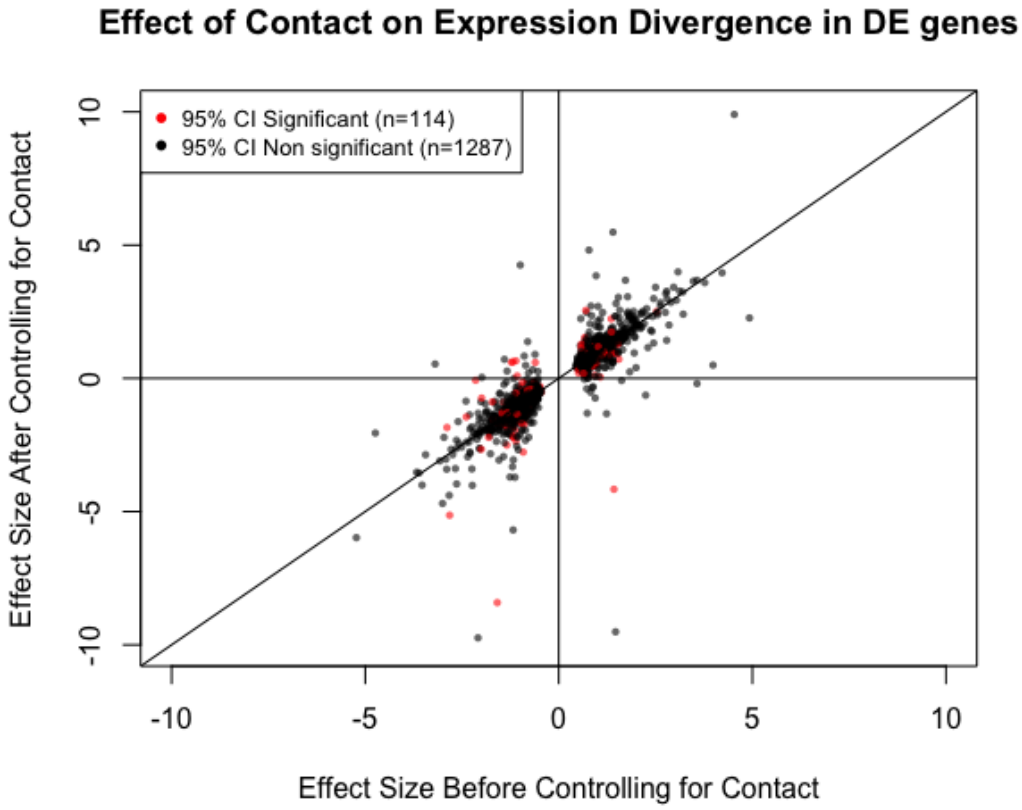

Figure 2.23: **S16. Gene expression variance is explained by chromatin contacts for 8% of DE genes.** Plot of the species effect size in DE genes between models before (x-axis) and after (y-axis) conditioning on contact frequency. The Monte Carlo test of significance was used to construct the 95% confidence interval and evaluate the significance of the indirect effect (species' effect on expression mediated through contact). Amongst DE genes, 8% (116/1401) of genes showed a statistically significant effect of Hi-C contacts on expression levels (i.e. their 95% confidence interval does not include zero).

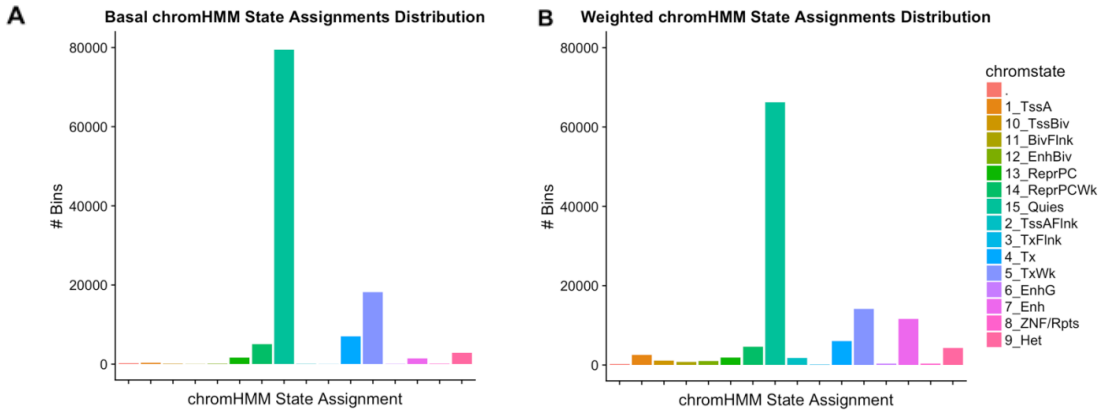

Figure 2.24: **S17. Using a weighting scheme for chromHMM annotations increases the proportion of transcriptional and enhancer-like annotations.** (A) Histogram showing the number of Hi-C loci (y-axis) assigned to each chromHMM annotation (x-axis) using maximum base pair overlap to assign each locus to a state. In the legend, '.' denotes that no annotation was found for a given bin. (TssA-Active TSS, TSSBiv-Bivalent/Poised TSS, BivFlnk-Flanking Bivalent TSS/Enh, EnhBiv-Bivalent Enhancer, ReprPC-Repressed PolyComb, ReprPCWk-Weak Repressed PolyComb, Quies-Quiescent/Low, TssAFlnk-Flanking Active TSS, TxFlnk-Transcription at gene 5' and 3', Tx-Strong transcription, TxWk-Weak transcription, EnhG-Genic Enhancers, Enh-Enhancers, ZNF/Rpts-ZNF genes and repeats, Het-Heterochromatin). (B) Same as A, only here, we assigned annotations after weighting chromHMM elements' overlaps with Hi-C loci by the reciprocal of their mean overlap in all our loci. This approach increases the number of 10kb Hi-C bins that are assigned to chromHMM annotations associated with transcriptional and enhancer activity (i.e. TssA, TssBiv, TssAFlnk, EnhG, Enh).

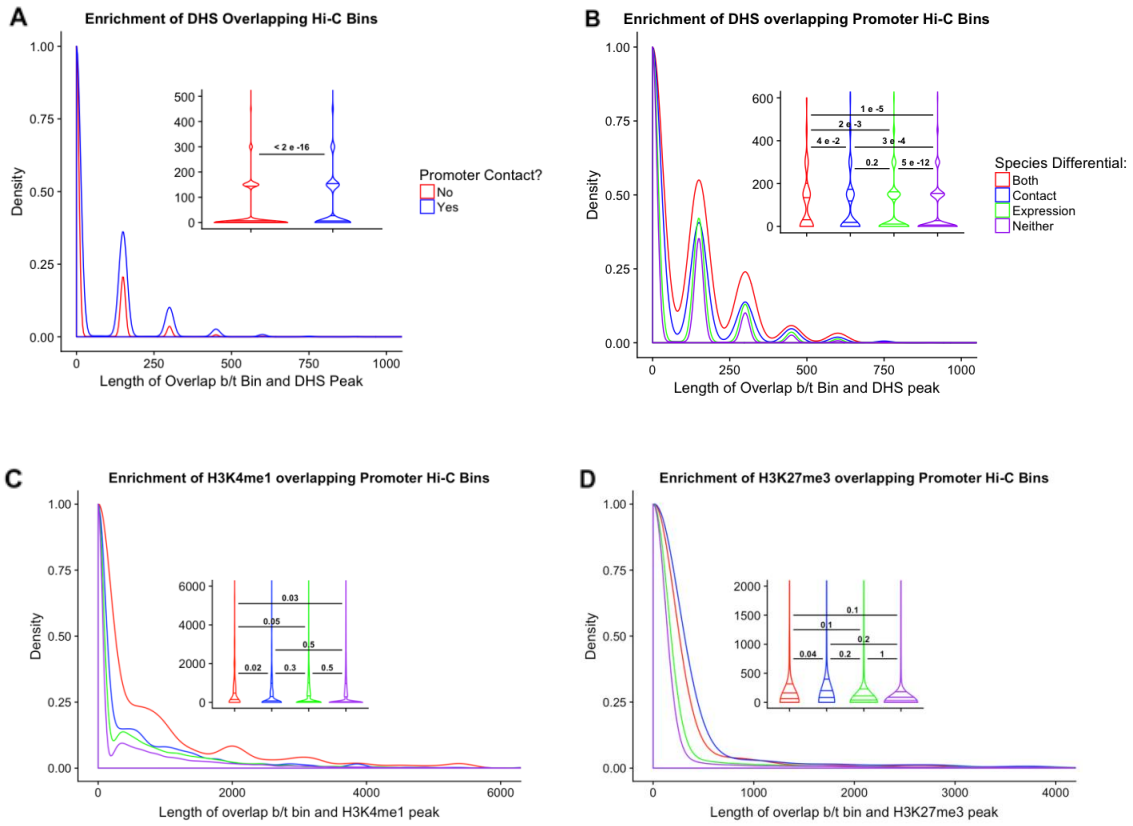

Figure 2.25: **S18. Overlap of epigenetic signatures and Hi-C contacts.** (A) Density distribution of the base pair overlap between DHS peaks downloaded from ENCODE and our Hi-C loci. Plot is split between Hi-C loci that contact a promoter and those that do not. Inlay is a violin plot of the same distributions, with lines and numbers indicating pairwise t-tests of the mean, and their corresponding significance levels. (B) Density plot similar to A, but only considering Hi-C loci involved in contact with a promoter, and separating contacts into 4 classes, indicated by color: those that show differential contact between species, those that show differential expression between species, those that show both, and those that show neither. We used pairwise t-tests to compare differences in the mean overlap among the four classes of Hi-C loci. (C) Same as in B, but for the active histone mark H3K4me1. (D) Same as in B and C, but for the repressive histone mark H3K27me3.

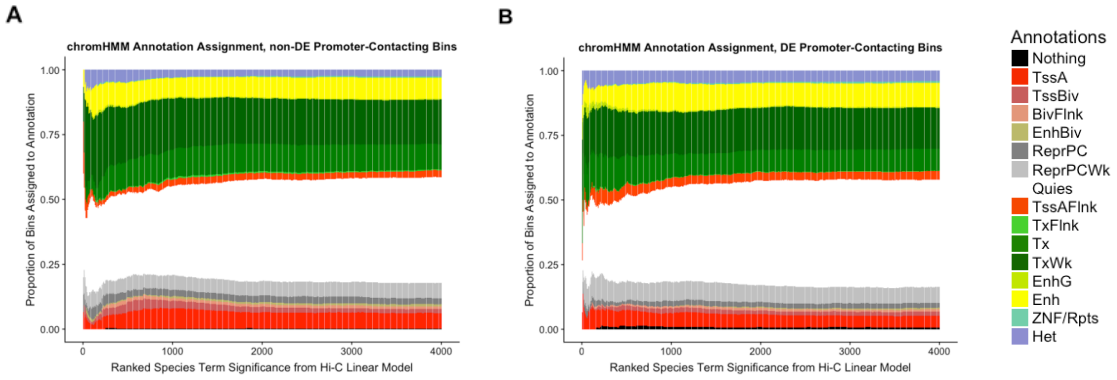

Figure 2.26: **S19. Dynamics of chromHMM state among significant Hi-C contacts overlapping DE or non-DE genes.** (A) Hi-C loci that make contact with promoters of genes that are not differentially expressed (DE) across species are ranked in order of decreasing DC FDR (x-axis). The y-axis shows cumulative proportion of chromHMM annotation assignments for all Hi-C loci at the given FDR or lower. (TssA-Active TSS, TSSBiv-Bivalent/Poised TSS, BivFlnk-Flanking Bivalent TSS/Enh, EnhBiv-Bivalent Enhancer, ReprPC-Repressed PolyComb, ReprPCWk-Weak Repressed PolyComb, Quies-Quiescent/Low, TssAFlnk-Flanking Active TSS, TxFlnk-Transcription at gene 5' and 3', Tx-Strong transcription, TxWk-Weak transcription, EnhG-Genic Enhancers, Enh-Enhancers, ZNF/Rpts-ZNF genes and repeats, Het-Heterochromatin). (B) Same as A, but only considering Hi-C loci making contact with promoters of genes that are differentially expressed (DE).

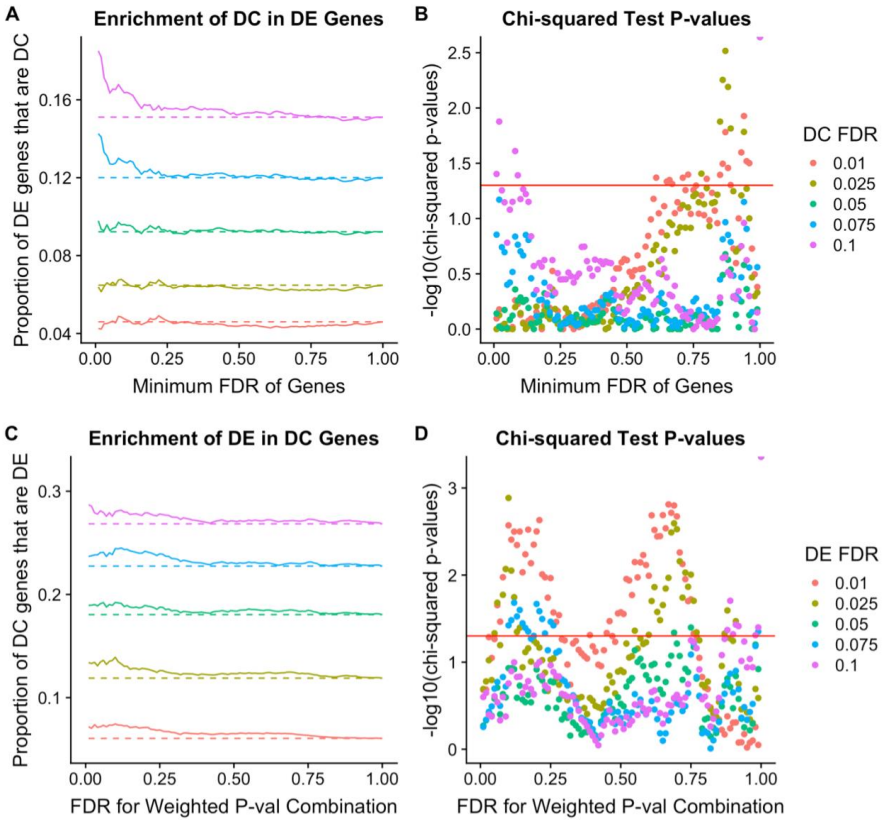

Figure 2.27: **S20. Reciprocal enrichments of differential expression and differential contact.** (A) Enrichment of inter-species differentially contacting (DC) loci in genes with corresponding differences in expression (DE) between the species. The proportion of DE genes that are significantly DC (y-axis) is shown across a range of DE FDRs (x-axis). Colors indicate different DC FDR thresholds, and dashed lines indicate the proportion of DC loci expected by chance alone. (B) P values of Chi-squared tests of the null that there is no difference in proportion of DC loci among DE genes (y-axis), shown for a range of DE FDRs (x-axis). In both panels, the DE genes overlapping Hi-C loci were chosen to have the minimum FDR supporting inter-species difference in expression. (C) Similar to Fig 2.6A, but using a weighted p-value combination technique [189] to integrate DC FDR across regions, instead of using the minimum FDR DC region. Once again, we observe enrichment of inter-species differentially expressed (DE) genes with corresponding differences in Hi-C contact frequencies (DC) between the species. The proportion of DC genes that are significantly DE (y-axis) is shown across a range of DC FDRs (x-axis). Colors indicate different DE FDR thresholds, and dashed lines indicate the proportion of DE genes expected by chance alone. (D) P values of Chi-squared tests of the null that there is no difference in proportion of DE genes among DC genes (y-axis), shown for a range of DC FDRs (x-axis).

2.8 Supplemental Tables

Table 2.1: **S1. HOMER-called contacts, H21792.**

Table 2.2: **S2. HOMER-called contacts, H28126.**

Table 2.3: **S3. HOMER-called contacts, H28815.**

Table 2.4: **S4. HOMER-called contacts, H28834.**

Table 2.5: **S5. HOMER-called contacts, C3649.**

Table 2.6: **S6. HOMER-called contacts, C40300.**

Table 2.7: **S7. HOMER-called contacts, C3624.**

Table 2.8: **S8. HOMER-called contacts, C3651.**

Table 2.9: **S9. Orthology calling statistics.**

Table 2.10: **S10. Differentially contacting (DC) regions.**

Table 2.11: **S11. Human consensus Arrowhead-inferred TADs.**

Table 2.12: **S12. Chimpanzee consensus Arrowhead-inferred TADs.**

Table 2.13: **S13. Human-Chimpanzee orthologous TAD coordinates.**

Table 2.14: **S14. ChromHMM genic enhancer annotation enrichments.**

Table 2.15: **S15. ENCODE chromatin mark sources.**

Table 2.16: **S16. Sample metadata.**

Table 2.17: **S17. 10kb Arrowhead TAD inferences, H21792.**

Table 2.18: **S18. 10kb Arrowhead TAD inferences, H28126.**

Table 2.19: **S19. 10kb Arrowhead TAD inferences, H28815.**

Table 2.20: **S20. 10kb Arrowhead TAD inferences, H28834.**

Table 2.21: **S21. 10kb Arrowhead TAD inferences, C3649.**

Table 2.22: **S22. 10kb Arrowhead TAD inferences, C40300.**

Table 2.23: **S23. 10kb Arrowhead TAD inferences, C3624.**

Table 2.24: **S24. 10kb Arrowhead TAD inferences, C3651.**

CHAPTER 3

A TAD SKEPTIC: IS 3D GENOME TOPOLOGY EVOLUTIONARILY CONSERVED?

3.1 Abstract[1]

The notion that topologically associating domains (TADs) are highly conserved across species has practically become an axiom in the field of 3D genome research. But what exactly do we mean by 'highly conserved', and what are the actual comparative data that support this notion? To address these questions, we performed a historical review of the relevant literature, and retraced numerous citation chains to reveal the primary data that were used as the basis for the widely accepted conclusion that TADs are highly conserved across evolution. A thorough review of the available evidence suggests the answer may be more complex than what is commonly presented.

3.2 What are TADs?

Some of the most fascinating features to have emerged from research into 3D genome conformation are topologically associating domains (TADs). Originally discovered through the analysis of Hi-C data [69, 209, 122, 258], TADs appear on a chromatin contact map as large squares of enhanced contact frequency rising off the diagonal. The original method for algorithmic TAD inference used a 'directionality index' [69] to define TADs as segments of the genome that are more connected to each other than to other regions. In turn, TAD boundaries were defined as segments of the genome that are characterized by a sharp transition between upstream and downstream highly connected regions. Originally applying directionality index to Hi-C data at the relative low resolution of 40 kb, early studies reported

1. Citation for abbreviated version of this chapter: Eres, Ittai E, and Gilad, Y. A TAD Skeptic: Is 3D Genome Topology Conserved? *Manuscript in review at Trends in Genetics.*

that TADs are non-overlapping, highly self-interacting megabase-scale structures. More recent studies, using higher-resolution contact maps, as well as different inference algorithms, have revealed TAD structures at much smaller scales, and often nested within each other [220, 15, 233]. The precise nature of these features is still a matter of debate, with various definitions of TADs shifting as new algorithms arise and new discoveries are made about the mechanisms behind TAD formation (e.g. loop extrusion and compartmentalization) [220, 233, 86, 67, 323, 211, 95, 246]. Previous studies have found relatively low concordance of TADs defined by different algorithms and across various resolutions and parameters, further impeding a robust definition of these structures [56, 89, 335]. While efforts have been made to functionally delineate between TADs at different scales [67, 11, 264, 283], most studies, especially those who rely solely on Hi-C data, do not make these distinctions.

Regardless of the challenge of defining TADs, it is clear that these 3D structures play an important role in genome organization and function [15, 67, 11, 5, 90, 313, 68, 62, 266]. Studies assessing the direct transcriptional effects of TADs have found mixed results, with some locus-specific work suggesting a strong impact of TAD disruption on gene expression [173, 118, 115, 158], while other genome-wide results imply only mild effects on expression [103, 336, 82, 234]. Despite some uncertainty about the magnitude of regulatory changes induced by TAD disruptions, multiple independent lines of evidence suggest TADs are functionally relevant. Genes located within the same TAD can have strongly correlated expression patterns and are often coregulated during cell differentiation [209, 323, 232]. TAD boundaries are strongly correlated with replication-timing domain boundaries [223], and are enriched for insulator elements such as CCCTC-binding factor (CTCF) [69, 233]. Disruptions in normative TAD structures have also been implicated in a number of human pathologies [174, 127, 91]. While precise and robust TAD and boundaries definitions are still elusive, a general feature that is robust to the specific definition is that loci within a TAD make contact more frequently with other loci in the same TAD than with loci outside

it. The common paradigm is that TADs represent insulated neighborhoods, constraining the possible set of interactions between cis regulatory elements (CREs) and target genes [5, 90, 68, 62, 266, 282, 287]. It is believed that TADs are critical features of the genome, serving to sustain specific sets of regulatory interactions while preventing ectopic interactions between regulatory elements and the wrong target genes [5, 282, 257].

Numerous other papers and reviews have delved into the history and functionality of TADs. Previous papers and reviews discuss the variance in algorithmic identification of TADs, the inconsistency of TAD calls at different resolutions, and the lack of robust approach to identify TAD boundaries based on Hi- C data [67, 11, 56, 89, 335, 264, 5, 90, 253, 243, 94, 6]. The notion that TADs are highly conserved across species and cell types is prevalent and often considered a foregone conclusion. Determining TAD variability across cell types is important for understanding the extent to which 3D genome structure affects differential gene regulation during development, enabling the regulatory and functional novelty observed in different cell lineages. In turn, assessing TAD conservation across evolution could help reveal the regulatory loci and mechanisms responsible for speciation and adaptation. We do not discuss further the issues related to similarities and differences in TADs across cell types and tissues, which have been previously discussed [5, 247, 248, 251, 327, 23, 55]. In this review, we wish to specifically focus on the notion that TADs and their boundaries are highly conserved across species.

Many studies are cited as reporting this conclusion, but it is difficult to trace the origin of this claim. If TADs and their boundaries are indeed highly conserved across spices, the origin of regulatory novelty must be elsewhere. However, if genome organization is not highly conserved, it is possible that changes in TADs and insulation boundaries may play an important role in underlying adaptation and speciation through changes in gene regulation. In other words, the answer to the question of TAD conservation has important implications for evolutionary research. We thus set out to thoroughly review the evidence

for TAD conservation and we found that, in fact, only a few studies collected relevant data and provide direct evidence to support this notion.

3.3 Conservation in Context

In order to evaluate the evidence for conservation of TADs, it is important to consider what we actually mean when we refer to conservation of genetic and epigenetic features across species. At the level of a single feature—a single locus for example—conservation is easily defined when the state of the feature is identical across species. More generally, however, when studies refer to genome-wide properties, they typically do not use a specific standard for the definition of conservation. When studies refer to a general property (for example, chromatin accessibility) as conserved, it implies that this property evolves under natural selection to maintain similarity across species. However, very few studies formally test this hypothesis. In fact, for most functional genomic traits that are comparatively studied, we have not yet formulated as null model of 'no selection.' Without a formal test, what do we typically mean when we conclude that a molecular trait is conserved? In most studies, conservation simply means 'highly similar' across species. While this is typically not a formal process, the degree of similarity, or variance, is evaluated and benchmarked based on other relevant comparisons. For example, if the level of observed variation in a trait is similar within and between species, the trait is typically deemed highly similar across species and therefore conserved. If variation between species is consistently low regardless of the time to the most recent common ancestors of the species, the trait is likely to be conserved. If different molecular features show a range of inter-species variability, the features with the lowest variance across species are assumed to be conserved. All of these examples point to ad-hoc definition of conservation, but this does not mean that they are wrong.

Let us consider specific examples. Comparative studies have reported that genome-wide, the overlap of histone modification H3K4me3 locations in humans and chimpanzees is

around 70% [31]. Remarkably, the genome-wide overlap of H3K4me3 locations in humans and mouse is also around 70% [310]. With these figures, not much could be said about the conservation of H3K4me3 locations in primates, but we can probably conclude with confidence that H3K4me3 locations are quite conserved between human and mouse. That said, the best genomic context for evaluating the degree of conservation in these comparisons may be other histone modifications, but even the minimal context provided here illustrates the importance of benchmarking similarity values in order to understand what they imply about conservation across species.

To date, comparative studies of chromatin conformations and TADs did not put forward a formal null model with which to evaluate levels of conservation. Instead, statements regarding the conservation of TAD and boundaries were made by using the ad-hoc rationale we discussed above. With this in mind, we now turn to critically examine the existing evidence for evolutionary conservation of TADs.

3.4 Indirect evidence for conservation

The notion that TADs and are highly conserved appears to be supported by a number of studies. One class of such studies, however, does not perform direct comparative assessment of TADs and boundaries across species. Instead, the indirect inference of TAD conservation is based on comparative functional genomic data that are independently associated with TADs. The most common approach in this class of studies is to directly map TADs in one species, then infer the locations of TADs in other species based on genomic features that are associated with TADs, such as CTCF binding sites, high gene density, or regions of active transcription [69, 209, 245, 141, 61]. To date, however, no single genomic feature can be used to effectively predict TAD locations and boundaries. Thus, the inference of TAD conservation based on other functional genomic features is indirect and might not be accurate.

One of the most commonly cited studies supporting TAD conservation, Rudan et al. 2015 [245], used such an indirect inference approach. Rudan et al. 2015 [245] collected comparative Hi-C data in liver cells from mouse, macaque, rabbit, and dog, but most of their comparative inference was based on the placement and orientation of CTCF binding sites. The authors conclusion that there is "extensive genome-wide interspecies conservation of chromosome structure" was based on comparisons of a broader set of contacts, not specifically of TADs. In fact, Rudan et al. did not report the total number of TADs identified in each species (only in mouse and dog), nor did they directly estimate the proportion of TADs that are found to be conserved across species. Instead, they used an indirect measure, estimating the interspecies correlations of inferred insulator activity [270] at different distances from orthologous genes. Rudan et al only reported species pairwise comparisons that involved the mouse data, resulting in Spearman correlations that ranged from 0.34 to 0.61. These correlations values may indicate some degree of conservation of 3D genome structure, but it is difficult to conclude from these analyses that TADs are indeed highly conserved across species. Moreover, Rudan et al. data collection was uneven across species, with ∼275 million reads sequenced from mouse, ∼150 million from rabbit, ∼100 million from macaque, and ∼550 million for dog. The large differences in read count result in a difference in the power to identify 3D genome structures across species and hence complicate the interpretation of the reported results.

There are other widely cited studies that concluded that TADs are highly conserved based on indirect evidence. Harmston et al. 2017 [116], for instance, identified genomic regulatory blocks (GRBs, regions with a high density of conserved noncoding elements) in human, opossum, chicken, and spotted gar. They reported that GRBs are often quite conserved across species. Using previously collected Hi-C data from human and Drosophila, Harmston et al. have shown that GRBs often fall within TADs and/or have edges proximal to TAD boundaries in these two species. Based on these data, the authors concluded that TADs

are generally conserved ancient features of the genome and that TAD boundaries are largely invariant between all the species in their study. However, the data reported by Harmston et al. shows that only about a third of the TADs were associated with GRBs; thus, even if one accepts the indirect inference based on GRBs as correct, up two thirds of TADs may still not be conserved in these species, as no direct evidence for TAD conservation was presented in this study. Indeed, in their concluding statement, Harmston et al. are careful to note that only the subset of GRB-associated TADs appears to be ancient conserved structures. However, this paper is often cited as providing strong evidence for general TAD conservation across species.

Similarly, Krefting et al. 2018 [153] considered TADs that were previously directly identified in humans [69, 233] along with genomic rearrangement breakpoints they identified in 13 species. Based on observed enrichment of these breakpoints at TAD boundaries and depletion within TAD bodies, the authors made relatively strong claims about TAD stability across evolutionary timescales. However, no direct comparison of TADs was made across species; the conclusions are essentially only based on the inter-species comparison of the rearrangement breakpoints. Given this, that the enrichments observed were only in comparison to TADs found in humans, and a fair amount of inconsistency in the degree of enrichment observed between the two different TAD sets used, we do not believe strong claims of conservation are warranted. Motivated by the findings of Harmston et al. 2017 [116], the same study also examined correlation of gene expression for orthologous genes in humans and mice within TADs associated with GRBs vs. orthologous genes in non-GRB associated TADs. While they found that gene expression within GRB-associated TADs is significantly slightly more correlated than the expression of non-GRB TAD genes, they also noted that more than 60% of hESC TADs don't overlap GRBs, perhaps suggesting that only a small subset of TADs are actually conserved. The authors should be commended for making note of this, as well as including the caveat that the enrichment they observe may be

due to chromatin accessibility differences between TAD boundaries and TAD bodies. Still, the majority of the paper makes claims of TAD conservation across evolutionary timescales, which simply are not supported due to inconsistent enrichments across different TAD sets and a lack of TAD inferences in any species aside from humans.

In another comparable approach, Lazar et al. 2018 [160] chose to compare human and gibbon genomes in LCLs, given the large number of chromosomal rearrangements and high DNA sequence identity (96%) between the two species. The authors cite previous studies to substantiate the claim that TADs are largely conserved across species [69, 245], and ultimately conclude that most TADs have been maintained as 'intact modules' during genomic divergence between humans and gibbons. Such claims may be overstated, given that this work did not undertake a direct comparison of TAD locations between the species, instead largely focusing on the overlap of multiple species' TAD boundaries with 67 rearrangement breakpoints identified in the gibbon genome (in comparison to human). Though their results indicate a high overlap of TAD boundaries across multiple species with gibbon breakpoints above what would be expected at random, they do not, to us, indicate strong conservation of the TADs themselves across species. Our interpretation is further supported by the study's finding that only 19 of the 67 breakpoints ($\sim$28%) overlapped TAD boundaries in all other species compared (human, rhesus, mouse, dog, and rabbit). Conversely, the authors found almost no evidence for new TADs being created between humans and gibbons based on rearrangement breakpoints, but we note again that this analysis is focused on the breakpoints rather than the TADs, and that the absence of evidence does not indicate evidence of absence.

Finally, we wish to highlight one exemplary study that also did not perform a direct assessment of TADs across species, but was measured in its conclusions and made a significant methodological contribution to comparative analysis of 3D genome topology. Yang et al. 2019 [316] represents one of the only papers with an explicit aim of comparing 3D

genome organization across a number of primate species, and sequenced a similar number of Hi-C reads from LCLs in chimpanzees, bonobos, and gorillas, in addition to examining previously-published human Hi-C data in the same cell type [233]. The analysis framework they present for interspecies comparisonphylo-HMRF (hidden markov random field)is sorely needed for interspecies comparative Hi-C. The robustness of the method is also underscored by the proximity between its identified boundaries of Hi-C evolutionary pattern blocks and previously-inferred TAD boundaries [233]. We speculate that the authors focused on this overlap, rather than directly inferring TADs in the novel primate Hi-C datasets collected, because direct TAD inference and comparison would not be robust, just as general TAD inference is not robust. Despite similar sequencing depth across species, TAD detection would likely be confounded by differences in reference genome quality, a consideration the authors apparently took quite seriously, given that all Hi-C reads in the method are ultimately mapped back to the human reference genome. To our knowledge, this is one of the only papers presenting an analytical framework for interspecies Hi-C comparison, with the other being more focused on similarity of genomic contacts within orthologous TADs between species, rather than the locations of the TADs [212]. As both of these methods are relatively recent, they have not yet been more broadly applied to a wide range of Hi-C datasets across different species, but we look forward to such analyses being used in the future in our own work and that of others.

In summary, all of the studies we discussed in this section did not directly identify TADs in multiple species, could not perform a direct inter-species comparison of TADs, and thus these studies do not provide direct evidence for the notion that TADs are highly conserved. In order to truly understand the extent of conservation of these structures, we must infer and examine them across a number of species, and assess if a given TAD in one species has a corresponding counterpart in others. Otherwise, claims of conservation speak to conservation of features of TADs and 3D genome topology more broadly, rather than conservation of the

individual structures themselves.

3.5 Direct but anecdotal evidence for conservation

The second class of studies that are widely cited as providing evidence for general TAD conservation provide only anecdotal evidence. These are studies that provide direct and strong evidence for conservation of TADs, but only in a small number of well-studied cases. It is thus difficult to generalize from these studies and conclude with confidence that TADs are generally highly conserved across species.

Woltering et al. 2014 [308], for example, found that Hox loci across zebrafish and mouse tend to have similar TAD structure, and Gomez-Marin et al. 2015 found comparable TAD structures across a number of species at the Six loci [111]. Both of these studies, as well as a number of others [266, 173, 97, 98], focus on loci that are highly conserved and/or thought to be critical for normal organismal development. Though these findings underscore the functional importance of TADs, they do not provide evidence for broad and general TAD conservation. In particular, the focus on a subset of candidate loci that are more likely to contain conserved features makes it difficult to generalize these observations to a genome-wide scale. Thus, though these studies infer TAD conservation based on direct functional data, they do not provide strong support for the widely accepted notion that TADs are highly conserved.

3.6 Direct evidence for the conservation of TADs

We now turn to the relatively small body of research that studied TAD conservation by directly identifying TADs and boundaries in multiple species. This direct approach seems the most obvious. In fact, it would be challenging to find another example in the genomics field where a widely accepted conclusion was mostly supported by indirect evidence and

inference. In the case of comparative studies of TADs, only a few studies collected direct comparative data.

Dixon et al. 2012 [69] collected Hi-C data and inferred TADs in human and mouse. This study was groundbreaking as it was one of the first to discover TADs and propose an algorithm to infer them from Hi-C contact maps (directionality index). This study is often cited as providing the first evidence that TADs are highly conserved between humans and mice. The authors collected 475 million sequencing reads from mouse Hi-C libraries but only 330 million reads from human. TAD boundaries were considered conserved if they had any overlap in the other species, with 76% of mouse TAD boundaries found in humans but only 54% of human boundaries found in mice. If one considers the entire dataset of TAD boundaries identified in both human and mouse (rather than the reported unilateral overlaps), ∼31% of boundaries are shared across the two species.

These results were and still are interpreted as evidence for strong TAD conservation. To provide some context, we considered other functional annotations in human and mouse. There is about 60-75% overlap of loci marked by histone modification in humans and mice [310], and between half to two-thirds of candidate regulatory regions are conserved in the two species [319]. Considering the observed proportion of overlapping TAD boundaries in human and mouse in this context, we believe that there is evidence for some level of conservation, but arguably, this cannot be considered strong evidence that TADs are generally highly conserved across species.

Another study that performed a direct comparative assessment of TADs is Rao et al. 2014 [233]. The authors collected ∼6.5 billion Hi-C sequencing reads from human but only ∼1.4 billion reads from mouse. The difference in read depth resulted in a striking difference in the power to infer TADs in the two species, with more than 9000 domains identified in human but only ∼3000 domains found in mouse. The authors considered entire domains conserved if the center of a domain in one species was within 50 kb of an annotated domain in the

other species (or within half the domain size, for domains smaller than 100 kb). Ultimately, Rao et al. 2014 reported that 45% of mouse domains (where they had considerably less power to identify TADs) were also present in human. Again, there may be some evidence for conservation, but it is difficult to conclude based on these data that TADs are highly conserved.

As far as we know, Dixon et al. 2012 and Rao et al. 2014 are two of the only studies that concluded that TADs are highly conserved based on a direct analysis of TADs and boundaries in more than one species. Both works used human and mouse, and utilized an unbalanced sequencing study design across species, which makes the interpretation of the results somewhat challenging. Regardless, even if we accept the observation of Dixon et al. 2012 and Rao et al. 2014 at face value, the reported overlap of TADs and boundaries in human and mouse arguably does not indicate that these features are highly conserved.

3.7 On the other hand...

There are a few studies that suggest that TADs may not be particularly conserved across species. Berthelot et al. 2015 [17] considered the order of orthologous genes to identify genomic rearrangement breakpoints in the genomes of human, mouse, dog, cow, horse, and a genomic reconstruction of the Boreoeutherian last common ancestor. In an attempt to understand the non-random genomic distribution of these breakpoints, the authors considered the overlap of rearrangement breakpoints with TADs that were previously identified in human [69]. Because the basal set used for comparisons was breakpoints rather than TAD boundaries, this is another example of a study that relied on indirect inference. In contrast to the results described from Lazar et al. [160], the authors did not find evidence for strong overlap of TAD boundaries and breakpoints, reporting that only 8% of the identified breakpoints overlap with TAD boundaries. This would suggest that TADs do not generally contribute to the locations of genomic rearrangements.

Berthelot et al. 2015 [17] is also notable for its interpretation of the results of Dixon et al. 2012 [69]. While the vast majority of papers cite Dixon et al.'s study as providing strong evidence that TADs are highly conserved, Berthelot et al. cite the same study to provide evidence for some TAD divergence between humans and mice. That the results of Dixon et al. 2012 can be interpreted by different groups both as supporting conservation of TADs or lack thereof highlights our notion that the foundation for the claim that TADs are highly conserved is not strong.

The notion that TADs may not be particularly conserved is also supported by our own study, in which we directly inferred TADs in humans and chimpanzees [80]. Our initial analysis found only ∼43% of TADs conserved between these species, but across many different parameters (e.g. resolution, window size, genome assembly), and different downstream analysis decisions, we found that no more than 78% of domains and 83% of TAD boundaries were shared between humans and chimpanzees—a much lower percentage than what has been seen across these species for a number of other functional regulatory phenotypes.

The notion that TADs may not be particularly conserved is also supported by recent results from Hi-C data across three distantly related Drosophila species. Renschler et al. 2019 [235] inferred TADs and genomic rearrangements across three Drosophila species, and found significant overlap above what would be expected by chance alone. However, the percentages of TAD boundaries overlapping a rearrangement breakpoint were relatively low (ranging from 13-21% of boundaries depending on the comparison). The proportion of overlapping TADs was even lower, at 10%.

Other findings, particularly in plants, also suggest that TAD positions may not be conserved across species. Dong et al. 2017 [71], for instance, collected Hi-C data from maize, tomato, sorghum, foxtail millet, and rice, and found relatively little conservation of TADs across these species. Although plants lack a homolog for CTCF, a transcription factor strongly implicated in the maintenance of TAD boundaries [115, 245, 141, 111], the authors

observed TAD-like domains in contact maps across all species, and found that they share many epigenetic features with TADs inferred in mammals. Xie et al. 2019 [313] used a similar method to assess TAD conservation in two different mustard plants, *Brassica rapa* and *Brassica oleracea*, and reported that about 25% of all TADs are found in both species.

It should be noted that the existence of TADs in plants, worms, yeast, and other non-mammalian species is a matter of active debate [23]. While chromatin conformation capture experiments have revealed self-interacting TAD-like structures in many of these species, their characteristics and mechanisms of formation often differ substantially from those of mammalian TADs [283, 1]. In many cases, these species lack homologs for insulator proteins thought to be essential to the formation of mammalian TADs (e.g. CTCF) [283]. More samples and more deeply sequenced Hi-C libraries from these species, as well as a deeper understanding of possible mechanisms of TAD-like feature formation, will be necessary to thoroughly assess conservation of TAD structures across all of evolution.

3.8 Concluding remarks and future perspectives

It is important to note that we are not taking a strong position 'for' or 'against' the notion of TAD conservation. Based on the available evidence, we conclude that there is currently no satisfying answer to the question of just how conserved TADs are across evolution. While the results from certain studies suggest some degree of conservation, others often lead to much lower estimates, and flawed study designs and variable analytical choices further obscure the issue. Although many studies state that TADs are conserved across species, there are only sparse data supporting or refuting this claim. In our mind, there is no strong basis for the common and often unchallenged notion that TADs are highly conserved.

One of the largest factors affecting our ability to assess evolutionary TAD conservation is the lack of any 'gold standard,' either for inferring TADs or for comparing them across species. As others have noted, TADs are variously and poorly defined, and it seems likely

that stable TADs observed in Hi-C data represent statistical features that emerge from averaging more dynamic interactions across millions of cells [307]. The few studies that did directly compare TADs across species made somewhat arbitrary choices about how to call these features conserved.

We struggled with this and many other aforementioned issues in our own work examining 3D genome structure across humans and chimpanzees [80]. Despite our own results suggesting a fair degree of TAD divergence between the species, we were unable to find many clear visual examples where divergent TAD inferences were obvious, based on the contact map. This once again emphasizes the need for specific, robust analytical methods to compare 3D genome topology and infer TADs across species . Unfortunately, evolutionary TAD conservation may remain an open and evolving question until we arrive at a more precise definition of TADs and converge on a set of truly robust methods for TAD inference and comparison.

To be clear, we do not disagree with the notion that a subset of TADs, particularly those involved in the regulation of key developmental loci or found near genomic rearrangement breakpoints, are likely to be highly conserved across species. We simply disagree with the conclusion often made, based on TAD subsets and existing interspecies comparative data, that TADs are highly conserved across species. Certainly, the existing evidence suggests that TADs as functional units of 3D genome organization exist and have similar epigenetic features across many different species. In mammals, a copious amount of chromatin contact data suggests some degree of conservation of TAD structure. However, the existing direct comparative data and analyses do not, in our opinion, provide enough evidence to claim strong conservation of TAD positioning across evolution.

Future studies hoping to assess 3D genome conservation across species should attempt to use a wide variety of TAD algorithms and parameters, as well as new interspecies Hi-C analytical methods to assess 3D genome conservation [316, 212]. Research addressing this question should also take great care to sequence a similar number of reads across species,

and check the robustness of their results across different analytical decisions for calling TADs and their boundaries conserved. TADs represent one intriguing feature of 3D genome architecture, and evolutionary conservation of other features (e.g. regulatory loops) is even less clear. In order to understand regulatory dynamics overall, we must refine our understanding of TADs, and agree on how to infer and compare them across species and cell types.

3.9 Acknowledgments

We apologize to the authors of relevant studies whose work was not addressed due to space limitations. We thank Natalia Gonzales, Jasmin Zohren, Sergey Kolchenko, Daniel Ibrahim, and Carlos Bustamante, for useful discussions, comments, and/or edits to the manuscript. YG is supported by NIH grant R35GM131726.

CHAPTER 4

CONCLUSION

4.1 Evolutionary and gene regulatory implications of this work

The field of human genetics has already accomplished much in connecting genetic variation to complex trait and disease variation between individuals. Outstanding challenges remain in characterizing the mechanisms of action for all these variants, and understanding the relative contributions of these different mechanisms to speciation, adaptation, and inter-individual trait variation. Although our understanding is rapidly expanding, we are still far from obtaining a comprehensive picture of gene regulation. The findings detailed herein corroborate the idea that assessing 3D genome structure is a crucial next piece of the puzzle. In Chapter 2, collaborators and I measured 3D genome structure and gene expression across human and chimpanzee induced pluripotent stem cells (iPSCs), revealing that differences in 3D genome structure may contribute to differential gene expression across these species. We found that, at the lowest scale (i.e. individual gene regulatory DNA loops), human and chimpanzee chromatin contacts are fairly conserved. This would imply that, at least in iPSCs, individual gene-cis-regulatory element (CRE) interactions do not vary significantly between humans and chimpanzees. However, this does not necessarily mean that the origins of regulatory novelty must lie elsewhere. Chromatin state in iPSCs is generally more open and 'permissive' than in differentiated cell types [275], and thus we may expect lower divergence in gene-CRE loops across species in iPSCs than in other cell types. Interestingly, we also observed evidence for large sets of species-biased differences in loop strength on individual chromosomes that have experienced large-scale rearrangements between humans and chimpanzees. This suggests that such rearrangements may help drive interspecies regulatory novelty in 3D chromatin interactions, although more functional follow-up will be required to confirm this notion, given conflicting results from some other comparative studies [160, 153].

In theory, changes to gene-CRE interactions could be a major driver of phenotypic divergence amongst primates, but studies comparing more individuals across a variety of cell types will be required to confirm this notion.

Regardless of the precise level of conservation/divergence in DNA looping, we found ample evidence for pairs of loci exhibiting differential contact (DC) across species. When we overlapped these data with RNA-seq data assessing gene expression levels, we observed strong correlations between interspecies contact and expression differences for differentially expressed (DE) genes overlapping our Hi-C loci. The fact that we did not observe similarly strong correlations for non-DE genes implies that variation in chromatin contacts plays a role in DE. Further corroborating this notion, we observed a significant enrichment for DE amongst genes we classified as DC across species–and vice versa. As previously stated, the observational nature of the study meant we could not directly infer a causal relationship between DC and DE. However, our mediation analysis found that up to 8% of DE genes may have a significant portion of their expression variation explained by variation in chromatin contacts. Placed in the context of other studies that observed perturbations in chromatin contact affecting gene expression [173, 263], our findings suggest that species-specific differences in 3D genomic contacts are indeed a driver of species-specific expression. This conclusion is also supported by our observation that, compared to contacts not involving a promoter, promoter-associated contacts are enriched for more active chromHMM state assignments. The intuitive conclusion from this result is that loci making contact with a promoter are likely involved in active regulation of the corresponding gene. Similarly, we observed that joint DE/DC loci identified in our study are enriched for a wide variety of functional epigenetic marks as compared to non-DE/DC loci. Under the common paradigm that most chromatin is not accessible and thus not active in any given cell type [288], these functional annotation enrichments suggest that the identified joint DE/DC loci represent functionally relevant stretches of DNA between the species. Based on all these

results, it may be tempting to speculate that 3D genome conformation is one of the most basal elements laying the groundwork for a broad cascade of events dictating gene regulation. Unfortunately, this conclusion seems somewhat premature absent a bevy of mechanistic perturbation studies, and given more recent conflicting results about the order and nature of events with respect to genome conformation and observed differences in gene expression [131, 103, 82, 129, 2, 14]. Although cause and consequence are still difficult to disentangle in this framework, there is no doubt that 3D genome conformation is an important facet affecting the evolution of gene regulation.

Beyond exploring regulatory loop conservation and its effects on expression, the data collected in Chapter 2 also enabled us to examine higher-order chromatin structure, such as topologically associating domains (TADs), across the species. We found relatively weak conservation of TAD structures as compared to regulatory loops and other epigenetic phenotypes previously compared between humans and chimpanzees. While this might point to TAD variation as a significant source of regulatory novelty, we were unable to find concrete examples of interspecies TAD differences affecting differential expression. This does not, however, preclude a significant role for TAD variation in speciation and interspecies expression divergence; as I discuss further below, this lack of signal could be due to TADs being poorly defined and difficult to robustly infer [56]. Regardless, the observed low interspecies TAD conservation was surprising, given the prevailing notion in the field that TADs are highly conserved across species [69, 233]. In large part, this incongruence motivated our critical assessment of the evidence for evolutionary TAD conservation, detailed in Chapter 3. A thorough review of the available data suggest that, while there is certainly some evidence for TAD conservation across mammalian species, it is not compelling enough to claim TADs are highly conserved. The validity of this notion is important to consider because, if true, it implies that TAD variation does not play a significant role in speciation. As addressed further in the final section of this chapter, analytical and definitional issues stymie a robust

assessment of interspecies TAD variation and preclude a thorough understanding of TADs' impact on gene expression. This much is evident from the fact that, although the TAD inference algorithms we employed found numerous differences between the species, visualizations of the corresponding contact maps did not appear significantly different between humans and chimpanzees. Thus, the results of our own TAD comparisons do not necessarily conflict with previous findings. It is possible that differences in TADs play a significant role in differences observed between primate species, but it is difficult to support or refute this notion with any confidence, given the current state of the field. Despite this, my thesis work has broadly confirmed the idea that 3D genome organization is an integral feature affecting the evolution of gene regulation. At the same time, it is important to understand the limitations of this work, and consequently, avenues for future research.

4.2 Limitations and next steps

There are a number of limitations to this research program that should be considered to inform avenues for future research. For one, the true extent of interspecies divergence in gene-CRE interactions may be higher than our estimate, given the underpowered nature of the Hi-C assay [13] and our limited number of individuals from each species. Still, our analysis was carried out in a robust quantitative fashion that is likely to give more accurate estimates of inter-species conservation than simplistic approaches other studies have used (i.e. assessing conservation via a venn diagram of overlap in significant chromatin contacts per-species) [69, 233]. Directly testing each contact identified as significant in any individual for inter-species differences allowed us to largely sidestep the issue of incomplete power, avoiding an overinflated estimate of divergence. While numerous methods have been proposed to quantitatively compare regulatory loop strength across different biological conditions [171, 216, 70, 85, 233], sparse novel techniques have only very recently emerged for running similar comparisons across species [316, 212]. When we began analyzing the data

collected in Chapter 2, these techniques had not yet been published, but applying them to these and similar data in the future would be of great interest for robustly assessing primate divergence in regulatory chromatin looping. In a similar vein, the sequencing depth in our own study allowed for assessment of chromatin contacts at a 10 kb resolution, but a comprehensive picture of inter-primate differences in chromatin loops will require deeper sequencing to enable sub-kilobase resolution and analysis of finer-scale loops. Lastly and as noted above, our comparisons were only performed in iPSCs, which tend to have more permissive regulatory landscapes than differentiated cell types [275]. Comparison of 3D chromatin structure across species in other cell types would thus be highly desirable, and may reveal greater divergence in gene-CRE loops than that observed in our own work. Ideally, this would be performed on isogenic samples to reduce the confounding effects of genetic variation, which could be accomplished by differentiating the same iPSC lines into a variety of terminal cell types.

Another important limitation to consider is that, when examining functional enrichments, we only overlapped our DE/DC loci with publicly-available epigenetic data from humans. The precise functional significance and evolutionary impact of these loci could be more thoroughly assessed and polarized in the future by adding chimpanzee epigenetic mark data. Such an analysis would be particularly interesting for assessing which DC loci have undergone differential CRE evolution between species, and thus are more likely to have species-specific effects on gene regulation. The mediation analysis we utilized to assess the impact of DC on DE could be expanded to include epigenetic mark data across species, improving power to predict gene expression differences [136]. More broadly, our analyses integrating Hi-C data and RNA-seq data could be improved in a number of ways that may find more expression variation explained by chromatin contact variation. As is also the case for assessment of conservation, deeper sequencing could provide better resolution of individual gene-CRE interactions, making it easier to tease out the effects of chromatin contact on expression. At

the 10 kb resolution we used, there were instances of multiple genes being assigned to the same Hi-C bin, likely obscuring interesting signals that could be observed in finer-scale data. Similarly, greater signals of association between chromatin contact and expression might have been observed if the RNA-seq and Hi-C data were collected concomitantly. Although our RNA-seq data came from the same cell lines, they were collected previously by different researchers culturing the cells in slightly different conditions. Concomitant collection would be more likely to maintain (and thus detect) weaker links between chromatin conformation and gene expression that may have been concealed in our own data. In some sense, our observational collection of these data across species represents an experimental paradigm for 'natural perturbation' of chromatin structure and gene expression. At the same time, a thorough understanding of DC affecting DE would require more precise targeted perturbations, altering regulatory loop strength in one species and expecting to see corresponding expression 'rescue' to comparable levels observed in the other species.

Our inferences regarding conservation of TAD structure could also be improved upon with functional and perturbational follow-up studies. As mentioned in the previous section, our algorithmic inference of TAD divergence often appeared fairly conserved upon visual inspection. This is probably largely due to issues with TAD identification (discussed further below), but could be mitigated with further functional characterizations. In particular, divergence in TAD structure could be more confidently characterized by also collecting ChIP-seq for CCCTC-binding factor (CTCF), a protein centrally involved in anchoring chromatin loops and demarcating TAD boundaries [233, 336, 220, 69, 209]. Overlaying these data with Hi-C data could help differentiate between instances of TAD divergence that are reflective of true biology (i.e. their boundaries show differences in CTCF binding across species), and divergence that is driven by technical issues (i.e. minimal difference in CTCF binding, but the inference algorithm fails to detect a TAD in one species where it nonetheless appears present). Situations falling into the former category could be further validated

and characterized via CRISPR-based perturbations to the differential CTCF binding site in both species. Specifically, one could abrogate the binding site in the species where CTCF is strongly bound, verify reduced binding via ChIP-seq, and observe subsequent effects on TAD structure and corresponding gene expression. Carrying out the reciprocal experiment (i.e. creating a binding site in the species where CTCF is not bound) would also be useful. In both instances, these perturbations should drive TAD structure and local gene expression to appear more similar to the species where CTCF binding was not altered. If this expectation ends up being incorrect, that is in and of itself an extremely interesting result, and could spur further research into the mechanisms driving TAD formation and divergence across species. Furthermore, integrating the other data types collected with CTCF occupancy differences across species would add another intriguing layer of regulation that could help explain individual instances of DC and DE, and their effect on one another. Lastly, although it has been examined in some previous studies within a single species [248, 247, 251, 68], estimating TAD sharing between tissues and cell types across primate species would be of great interest for elucidating the role these structures may play in differentiation and development, and how this may differ across evolution. Unfortunately, as I discuss in Chapter 3 and elaborate upon below, robust inferences regarding TAD conservation, function, and tissue-specificity are still severely hampered by imprecise and variable TAD definitions.

4.3 The state of the 3D genome field, challenges, and future perspectives

3D genomics is an exciting and relatively young field of epigenetic research. While the novelty of the field allows for high-impact discoveries and inferences, it also presents unique challenges. Hi-C is regarded as a revolutionary technology enabling genome-wide assessment of 3D genomic contacts, but it is also known to have a poor signal:noise ratio [154, 317]. This has improved somewhat as researchers have transitioned away from dilution Hi-C to

in-nucleus Hi-C, reducing the number of spurious interchromosomal trans reads [203], but still remains a problem. Several factors contribute to this issue. The method creates chimeric molecules by ligating proximal restriction fragments together, and the number of possible pairwise fragment interactions is very high, regardless of the restriction enzyme used. This means that, in order to achieve statistical power to detect significant contacts, reads from Hi-C data must typically be binned into fixed-size intervals tiling the genome [215]. Even when sequencing depth is great enough to achieve individual restriction fragment resolution, these fragments are often not the same size, resulting in differences in power to infer contact at different loci [314]. Differences in restriction fragment length, chromatin accessibility, and GC content also affect the efficiency of ligation, restriction enzyme cutting, and sequence amplification, respectively, exacerbating power differences between restriction fragments [314]. In addition, proximity ligation can introduce spurious ligation products (e.g. self-circularized ligations) that add more noise, as they do not actually represent chimeric molecules connecting two linearly distant loci in physical proximity [12]. That Hi-C data have many sources of systematic bias is evident from the plethora of studies proposing models to analyze Hi-C and address these biases, both explicitly (normalizing for specific sources of bias) and implicitly (normalizing the data to achieve equal visibility across loci) [314, 125, 128, 233, 53, 151, 168, 276, 306, 255, 249]. A very recent paper proposed a different method with enhanced crosslinking to assess chromatin organization that seems less noisy than Hi-C data, but its full utility is difficult to ascertain before it sees more widespread use [318]. Future research should endeavor to use this and other methods that may arise to comprehensively assay 3D genome structure with fewer sources of bias, but, for the time being, Hi-C remains the dominant technique. Newer techniques could also be useful for characterizing cell-to-cell variability in 3D genome structure; single-cell methods exist for Hi-C, but they are plagued by issues with data sparsity, genome coverage, and incomplete power, even moreso than bulk Hi-C [230, 231, 201, 202, 330]. Studies using clustering and deconvolution

of single-cell and bulk Hi-C data, respectively, may help improve the utility of single-cell Hi-C, better connect these data with bulk data, and define specific cell subpopulations with similar 3D regulatory interactions [329, 254, 322, 177].

In much the same way that there is no single agreed upon method to address Hi-C biases and normalize the data, the field also lacks a 'gold standard' method for assessing significant chromatin contacts. A wide variety of statistical paradigms have been proposed, but no single method stands out. Studies comparing significance calling algorithms typically recommend researchers choose an algorithm that will work well for their downstream analyses, given differences in the quantity and characteristics of significant loops identified by each option [89, 6, 175]. There is thus a pressing need to converge upon a 'gold standard' method for identification of significant interactions. Sadly, such a convergence does not appear likely any time soon, since the field cannot even agree upon a standard format for storing Hi-C data, let alone analyzing it [215, 185]. In the meantime, much as was done in Chapter 2, studies seeking to examine significant chromatin contacts should utilize a number of different algorithms, in order to ensure their resulting inferences are robust. It is important to note that the field of 3D genome research has at least arrived at a (relative) consensus definition of what constitutes a significant interaction: a pair of loci with Hi-C reads connecting them more often than would be expected by chance, given their linear genomic distance. The lack of agreement seems centered primarily on how to statistically assess the significance of these interactions, and, relatedly, what is an appropriate null model for 'no significant contact.' Similarly, although finer demarcations of A/B compartments have been observed with higher-resolution Hi-C data [233], the field appears to largely agree upon the broad nature of A/B compartments (active/inactive chromatin), and the class of methods used to identify them (i.e. principal components analysis and clustering) [192]. Thus, although sometimes disparate methods exist to quantify the 3D genome at these two scales (chromatin loops and A/B compartments), there is at least some consensus about the foundational

definitions of these features. As alluded to in Chapter 3, no such consensus exists with respect to TADs.

Indeed, many of the aforementioned issues with TADs stem from the lack of a clear definition. When TADs were first discovered, they were defined in an analytical (rather than biological) fashion: as large squares of enhanced contact frequency arising off the diagonal in Hi-C maps [69, 209, 122, 258]. At the relatively low resolution of these original studies (40 kb), TADs emerged as megabase-scale, non-overlapping, highly self-interacting regions of the genome. Importantly, loci within a TAD not only make contact with other loci in the same TAD more often, but appear somewhat insulated from making contacts with loci outside of the TAD. TADs were systematically inferred with a hidden Markov model based on a 'directionality index' that quantified the degree of upstream or downstream contact bias at each Hi-C bin, under the intuition that TAD boundaries should display sharp transitions in this bias state [69]. While subsequent studies achieved improved Hi-C resolution with greater sequencing depth and identified nested and overlapping TADs at much smaller scales, they were still defined based on technical features of the data [233]. Despite the discovery that TADs exist at different scales, many novel TAD inference algorithms proposed in the years since have not made these distinctions, and are generally billed as TAD predictors, irrespective of scale [56]. Some TAD inference algorithms have been built for analysis of specific TAD sizes and hierarchies [233, 303], but most have not. This is problematic because different resolutions and algorithmic parameters are necessary for robust detection of different hierarchies of TADs: a low-resolution Hi-C experiment might easily be able to detect megabase-scale TADs (sometimes termed "meta-TADs" [93]), but will not have sufficient power to detect smaller TADs at the scale of several hundred kilobases (sometimes termed "sub-TADs" [220]). Overall, identification of TADs remains an outstanding issue, as highlighted by a number of studies that have found low concordance between different TAD algorithms [56, 335, 89]. It may be tempting to blame these issues on the algorithms

themselves, but the core of the problem truly resides in the definition of a TAD.

In the years since their discovery, TADs have been functionally characterized in many ways. TAD boundaries are enriched for CTCF and cohesin binding sites [69, 233, 25], active histone marks, and transcription start sites (TSS) of housekeeping genes [283, 122, 232]. The functional significance of TADs is highlighted by these findings, as well as the observations that genes within the same TAD exhibit strongly correlated expression patterns [209, 232, 282], and that enhancer-promoter contacts largely occur within the same TAD [24, 281, 266, 62]. There has also been great interest in elucidating the mechanisms behind TAD formation. Outstanding questions remain, but accumulating evidence suggests these structures are largely formed via two mechanisms: loop extrusion, and compartmentalization [211, 95, 246, 242, 76]. Despite all these functional and mechanistic characterizations, the definition of a TAD has not changed much. Some recent reviews have proposed refining TAD definitions based on scale, overlap with other 3D chromatin features, and putative mechanisms of formation [67, 11], but the field has not yet widely adopted these distinctions. Such delineations will be crucial to furthering our understanding of 3D genome structure and its regulatory effects moving forward, particularly in light of recent observations from imaging and single-cell studies that suggest TADs are much more dynamic than originally thought [109, 87]. These and other findings imply that TADs identified from bulk Hi-C data may be statistical artifacts that emerge from averaging chromatin conformation in millions of cells [307], further underscoring the need for updated TAD definitions that reflect differences in specific factors and/or mechanisms involved in their formation. In the future, I suggest researchers combine mechanistic, functional, and single-cell techniques to thoroughly characterize different classes of TADs, ideally giving them different names. This would be a tremendous boon in helping define a 'ground truth' for TADs, against which to test the output of various inference algorithms. It may be infeasible to functionally characterize TADs genome-wide, but perhaps higher-resolution research and more widespread

use of the distinctions already proposed could reveal facets of the data that are sufficient for distinguishing different classes of TADs from one another *in silico*. These efforts could be greatly aided by employing techniques such as HiChIP, that enable simultaneous assessment of chromatin contact and protein binding genome-wide [198]. While a few studies have utilized HiChIP to examine 3D genome architecture as mediated through TAD-associated proteins [198, 164, 176], more similar research will be needed to lay the groundwork for a biology-based definition of TADs. Until then, future studies assessing TAD structure should consider utilizing a wide array of different algorithms, in order to establish a confident set of TADs. Simultaneously, I hope more work emerges using CRISPR and other techniques to carry out perturbational follow-up experiments assessing the impact of specific TADs on gene regulation across a variety of cell types, conditions, and species. When it comes to TADs, a vast frontier of exciting discoveries clearly remains.

3D genomics has done much to expand our understanding of the evolutionary and developmental aspects of gene regulation, but more precise definitions and robust methods will be necessary to ensure continued impact moving forward. As the cost of sequencing continues to decrease, more studies will hopefully be able to assess chromatin conformation in larger panels with more individuals. One recent seminal work using 20 individuals found quantitative trait loci (QTL) that affect several facets of higher-order 3D genome structure, dependent upon the genotype at the QTL [112]. The significance of such studies will only grow as we continue to appreciate the full extent of genomic structural variation between human individuals [46]. In conclusion, I am proud to have been a part of this research community during my PhD, and know that, as the 3D genome field evolves and technology continues to improve, it will help fulfill a central promise of human genetics: to understand the connection between genetic variation and phenotypic variation.

References

[1] Rafael D. Acemel, Ignacio Maeso, and José Luis Gómez-Skarmeta. Topologically associated domains: a successful scaffold for the evolution of gene regulation in animals. *Wiley Interdisciplinary Reviews: Developmental Biology*, 6(3):e265, 2017.

[2] Jeffrey M Alexander, Juan Guan, Bingkun Li, Lenka Maliskova, Michael Song, Yin Shen, Bo Huang, Stavros Lomvardas, and Orion D Weiner. Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *eLife*, 8:e41769, 2019.

[3] C D Allis and Thomas Jenuwein. The molecular hallmarks of epigenetic control. *Nature Reviews Genetics*, 17(8):487–500, 2016.

[4] Jordan A Anderson, Tauras P Vilgalys, and Jenny Tung. Broadening primate genomics: new insights into the ecology and evolution of primate gene regulation. *Current Opinion in Genetics & Development*, 62:16–22, 2020.

[5] Guillaume Andrey and Stefan Mundlos. The three-dimensional genome: regulating gene expression during pluripotency and development. *Development*, 144(20):3646–3658, 2017.

[6] Ferhat Ay and William S Noble. Analysis methods for studying the 3D architecture of the genome. *Genome biology*, 16:183, 2015.

[7] Sepideh Babaei, Ahmed Mahfouz, Marc Hulsman, Boudewijn P Lelieveldt, Jeroen de Ridder, and Marcel Reinders. Hi-C Chromatin Interaction Networks Predict Co-expression in the Mouse Cortex. *PLoS computational biology*, 11(5):e1004221, 2015.

[8] Courtney C Babbitt, Jesse S Silverman, Ralph Haygood, Jennifer M Reininga, Matthew V Rockman, and Gregory A Wray. Multiple Functional Variants in cis Modulate PDYN Expression. *Molecular biology and evolution*, 27(2):465–79, 2010.

[9] Reuben M. Baron and David A. Kenny. The Moderator-Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations. *Journal of Personality and Social Psychology*, 51(6):1173–1182, 1986.

[10] Caroline R. Bartman, Sarah C. Hsu, Chris C-S. Hsiung, Arjun Raj, and Gerd A. Blobel. Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Molecular Cell*, 62(2):237–247, 2016.

[11] Jonathan A Beagan and Jennifer E Phillips-Cremins. On the existence and functionality of topologically associating domains. *Nature genetics*, 52(1):8–16, 2020.

[12] Houda Belaghzal, Job Dekker, and Johan H. Gibcus. Hi-C 2.0: An optimized Hi-C procedure for high-resolution genome-wide mapping of chromosome conformation. *Methods*, 123:56–65, 2017.

[13] Jon-Matthew Belton, Rachel Patton McCord, Johan Harmen Gibcus, Natalia Naumova, Ye Zhan, and Job Dekker. Hi–C: A comprehensive technique to capture the conformation of genomes. *Methods*, 58(3):268–276, 2012.

[14] Nezha S. Benabdallah, Iain Williamson, Robert S. Illingworth, Lauren Kane, Shelagh Boyle, Dipta Sengupta, Graeme R. Grimes, Pierre Therizols, and Wendy A. Bickmore. Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation. *Molecular Cell*, 76(3):473–484.e7, 2019.

[15] Soizik Berlivet, Denis Paquette, Annie Dumouchel, David Langlais, Josée Dostie, and Marie Kmita. Clustering of tissue-specific sub-TADs accompanies the regulation of HoxA genes in developing limbs. *PLoS genetics*, 9(12):e1004018, 2013.

[16] Bradley E Bernstein, John A Stamatoyannopoulos, Joseph F Costello, Bing Ren, Aleksandar Milosavljevic, Alexander Meissner, Manolis Kellis, Marco A Marra, Arthur L Beaudet, Joseph R Ecker, Peggy J Farnham, Martin Hirst, Eric S Lander, Tarjei S Mikkelsen, and James A Thomson. The NIH Roadmap Epigenomics Mapping Consortium. *Nature Biotechnology*, 28(10):1045–1048, 2010.

[17] Camille Berthelot, Matthieu Muffato, Judith Abecassis, and Hugues Roest Crollius. The 3D Organization of Chromatin Explains Evolutionary Fragile Genomic Regions. *Cell Reports*, 10(11):1913–1924, 2015.

[18] Camille Berthelot, Diego Villar, Julie E. Horvath, Duncan T. Odom, and Paul Flicek. Complexity and conservation of regulatory landscapes underlie evolutionary resilience of mammalian gene expression. *Nature Ecology & Evolution*, 2(1):152–163, 2018.

[19] Lauren E. Blake, Julien Roux, Irene Hernando-Herraez, Nicholas E. Banovich, Raquel Garcia Perez, Chiaowen Joyce Hsiao, Ittai Eres, Claudia Cuevas, Tomas Marques-Bonet, and Yoav Gilad. A comparison of gene expression and DNA methylation patterns across tissues and species. *Genome Research*, 30(2):250–262, 2020.

[20] Ran Blekhman, John C Marioni, Paul Zumbo, Matthew Stephens, and Yoav Gilad. Sex-specific and lineage-specific alternative splicing in primates. *Genome research*, 20(2):180–9, 2010.

[21] Ran Blekhman, Alicia Oshlack, Adrien E. Chabot, Gordon K. Smyth, and Yoav Gilad. Gene Regulation in Primates Evolves under Tissue-Specific Selection Pressures. *PLoS Genetics*, 4(11), 2008.

[22] Ran Blekhman, Alicia Oshlack, and Yoav Gilad. Segmental duplications contribute to gene expression differences between humans and chimpanzees. *Genetics*, 182(2):627–30, 2009.

[23] B Bonev and Cavalli G. Organization and function of the 3D genome. *Nature Reviews Genetics*, 2016.

[24] Boyan Bonev, Netta Mendelson Cohen, Quentin Szabo, Lauriane Fritsch, Giorgio L. Papadopoulos, Yaniv Lubling, Xiaole Xu, Xiaodan Lv, Jean-Philippe Hugnot, Amos Tanay, and Giacomo Cavalli. Multiscale 3D Genome Rewiring during Mouse Neural Development. *Cell*, 171(3):557–572.e24, 2017.

[25] Kevin Van Bortle, Michael H Nichols, Li Li, Chin-Tong Ong, Naomi Takenaka, Zhaohui S Qin, and Victor G Corces. Insulator function and topological domain border strength scale with architectural protein occupancy. *Genome Biology*, 15(5):R82, 2014.

[26] Marco Botta, Syed Haider, Ian X Y Leung, Pietro Lio, and Julien Mozziconacci. Intra- and inter-chromosomal interactions correlate with CTCF binding genome wide. *Molecular Systems Biology*, 6(1):426, 2010.

[27] David Brawand, Magali Soumillon, Anamaria Necsulea, Philippe Julien, Gábor Csárdi, Patrick Harrigan, Manuela Weier, Angélica Liechti, Ayinuer Aximu-Petri, Martin Kircher, Frank W. Albert, Ulrich Zeller, Philipp Khaitovich, Frank Grützner, Sven Bergmann, Rasmus Nielsen, Svante Pääbo, and Henrik Kaessmann. The evolution of gene expression levels in mammalian organs. *Nature*, 478(7369):343–348, 2011.

[28] Roy J. Britten and Eric H. Davidson. Gene Regulation for Higher Cells: A Theory. *Science*, 165(3891):349–357, 1969.

[29] Roy J Britten and Eric H Davidson. Repetitive and Non-Repetitive DNA Sequences and a Speculation on the Origins of Evolutionary Novelty. *The Quarterly Review of Biology*, 46(2):111–138, 1971.

[30] Jason D. Buenrostro, Beijing Wu, Howard Y. Chang, and William J. Greenleaf. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Current Protocols in Molecular Biology*, 109(1):21.29.1–21.29.9, 2015.

[31] Carolyn E Cain, Ran Blekhman, John C Marioni, and Yoav Gilad. Gene expression differences among primates are associated with changes in a histone epigenetic modification. *Genetics*, 187(4):1225–34, 2011.

[32] John A. Calarco, Yi Xing, Mario Cáceres, Joseph P. Calarco, Xinshu Xiao, Qun Pan, Christopher Lee, Todd M. Preuss, and Benjamin J. Blencowe. Global analysis of alternative splicing differences between humans and chimpanzees. *Genes & Development*, 21(22):2963–2975, 2007.

[33] Eliezer Calo and Joanna Wysocka. Modification of Enhancer Chromatin: What, How, and Why? *Molecular Cell*, 49(5):825–837, 2013.

[34] Francesco N. Carelli, Angélica Liechti, Jean Halbert, Maria Warnefors, and Henrik Kaessmann. Repurposing of promoters and enhancers during mammalian evolution. *Nature Communications*, 9(1):4066, 2018.

[35] Sean B Carroll. Evolution at two levels: on genes and form. *PLoS biology*, 3(7):e245, 2005.

[36] Sean B Carroll. Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell*, 134(1):25–36, 2008.

[37] Claudia Rita Catacchio, Flavia Angela Maria Maggiolini, Pietro D'Addabbo, Miriana Bitonto, Oronzo Capozzi, Martina Lepore Signorile, Mattia Miroballo, Nicoletta Archidiacono, Evan E Eichler, Mario Ventura, and Francesca Antonacci. Inversion variants in human and primate genomes. *Genome Research*, 28(6):910–920, 2018.

[38] Susan E Celniker, Laura A L Dillon, Mark B Gerstein, Kristin C Gunsalus, Steven Henikoff, Gary H Karpen, Manolis Kellis, Eric C Lai, Jason D Lieb, David M MacAlpine, Gos Micklem, Fabio Piano, Michael Snyder, Lincoln Stein, Kevin P White, Robert H Waterston, and modENCODE Consortium. Unlocking the secrets of the genome. *Nature*, 459(7249):927–930, 2009.

[39] Keerthi T. Chathoth and Nicolae Radu Zabet. Chromatin architecture reorganization during neuronal cell differentiation in Drosophila genome. *Genome Research*, 29(4):613–625, 2019.

[40] Changya Chen, Wenbao Yu, Joanna Tober, Peng Gao, Bing He, Kiwon Lee, Tuan Trieu, Gerd A. Blobel, Nancy A. Speck, and Kai Tan. Spatial Genome Re-organization between Fetal and Adult Hematopoietic Stem Cells. *Cell Reports*, 29(12):4200–4211.e7, 2019.

[41] H Chen, L Seaman, S Liu, T Ried, and Rajapakse I Nucleus. Chromosome conformation and gene expression patterns differ profoundly in human fibroblasts grown in spheroids versus monolayers. *Nucleus*, 2017.

[42] Jenny Chen, Ross Swofford, Jeremy Johnson, Beryl B Cummings, Noga Rogel, Kerstin Lindblad-Toh, Wilfried Haerty, Federica di Palma, and Aviv Regev. A quantitative framework for characterizing the evolutionary history of mammalian gene expression. *Genome Research*, 29(1):53–63, 2018.

[43] Zhen Chen, Shuai Li, Shankar Subramaniam, John Y.-J. Shyy, and Shu Chien. Epigenetic Regulation: A New Frontier for Biomedical Engineers. *Annual Review of Biomedical Engineering*, 19(1):1–25, 2016.

[44] Ia Chevyreva, Richard L.M. Faull, Colin R. Green, and Louise F.B. Nicholson. Assessing RNA quality in postmortem human brain tissue. *Experimental and Molecular Pathology*, 84(1):71–77, 2008.

[45] Chris Church, Lee Moir, Fiona McMurray, Christophe Girard, Gareth T Banks, Lydia Teboul, Sara Wells, Jens C Brüning, Patrick M Nolan, Frances M Ashcroft, and Roger D Cox. Overexpression of Fto leads to increased food intake and results in obesity. *Nature Genetics*, 42(12):1086–1092, 2010.

[46] Ryan L. Collins, Harrison Brand, Konrad J. Karczewski, Xuefang Zhao, Jessica Alföldi, Laurent C. Francioli, Amit V. Khera, Chelsea Lowther, Laura D. Gauthier, Harold

Wang, Nicholas A. Watts, Matthew Solomonson, Anne O'Donnell-Luria, Alexander Baumann, Ruchi Munshi, Mark Walker, Christopher W. Whelan, Yongqing Huang, Ted Brookings, Ted Sharpe, Matthew R. Stone, Elise Valkanas, Jack Fu, Grace Tiao, Kristen M. Laricchia, Valentin Ruano-Rubio, Christine Stevens, Namrata Gupta, Caroline Cusick, Lauren Margolin, Jessica Alföldi, Irina M. Armean, Eric Banks, Louis Bergelson, Kristian Cibulskis, Ryan L. Collins, Kristen M. Connolly, Miguel Covarrubias, Beryl Cummings, Mark J. Daly, Stacey Donnelly, Yossi Farjoun, Steven Ferriera, Laurent Francioli, Stacey Gabriel, Laura D. Gauthier, Jeff Gentry, Namrata Gupta, Thibault Jeandet, Diane Kaplan, Konrad J. Karczewski, Kristen M. Laricchia, Christopher Llanwarne, Eric V. Minikel, Ruchi Munshi, Benjamin M. Neale, Sam Novod, Anne H. O'Donnell-Luria, Nikelle Petrillo, Timothy Poterba, David Roazen, Valentin Ruano-Rubio, Andrea Saltzman, Kaitlin E. Samocha, Molly Schleicher, Cotton Seed, Matthew Solomonson, Jose Soto, Grace Tiao, Kathleen Tibbetts, Charlotte Tolonen, Christopher Vittal, Gordon Wade, Arcturus Wang, Qingbo Wang, James S. Ware, Nicholas A. Watts, Ben Weisburd, Nicola Whiffin, Carlos A. Aguilar Salinas, Tariq Ahmad, Christine M. Albert, Diego Ardissino, Gil Atzmon, John Barnard, Laurent Beaugerie, Emelia J. Benjamin, Michael Boehnke, Lori L. Bonnycastle, Erwin P. Bottinger, Donald W. Bowden, Matthew J. Bown, John C. Chambers, Juliana C. Chan, Daniel Chasman, Judy Cho, Mina K. Chung, Bruce Cohen, Adolfo Correa, Dana Dabelea, Mark J. Daly, Dawood Darbar, Ravindranath Duggirala, Josée Dupuis, Patrick T. Ellinor, Roberto Elosua, Jeanette Erdmann, Tõnu Esko, Martti Färkkilä, Jose Florez, Andre Franke, Gad Getz, Benjamin Glaser, Stephen J. Glatt, David Goldstein, Clicerio Gonzalez, Leif Groop, Christopher Haiman, Craig Hanis, Matthew Harms, Mikko Hiltunen, Matti M. Holi, Christina M. Hultman, Mikko Kallela, Jaakko Kaprio, Sekar Kathiresan, Bong-Jo Kim, Young Jin Kim, George Kirov, Jaspal Kooner, Seppo Koskinen, Harlan M. Krumholz, Subra Kugathasan, Soo Heon Kwak, Markku Laakso, Terho Lehtimäki, Ruth J. F. Loos, Steven A. Lubitz, Ronald C. W. Ma, Daniel G. MacArthur, Jaume Marrugat, Kari M. Mattila, Steven McCarroll, Mark I. McCarthy, Dermot McGovern, Ruth McPherson, James B. Meigs, Olle Melander, Andres Metspalu, Benjamin M. Neale, Peter M. Nilsson, Michael C. O'Donovan, Dost Ongur, Lorena Orozco, Michael J. Owen, Colin N. A. Palmer, Aarno Palotie, Kyong Soo Park, Carlos Pato, Ann E. Pulver, Nazneen Rahman, Anne M. Remes, John D. Rioux, Samuli Ripatti, Dan M. Roden, Danish Saleheen, Veikko Salomaa, Nilesh J. Samani, Jeremiah Scharf, Heribert Schunkert, Moore B. Shoemaker, Pamela Sklar, Hilkka Soininen, Harry Sokol, Tim Spector, Patrick F. Sullivan, Jaana Suvisaari, E. Shyong Tai, Yik Ying Teo, Tuomi Tiinamaija, Ming Tsuang, Dan Turner, Teresa Tusie-Luna, Erkki Vartiainen, James S. Ware, Hugh Watkins, Rinse K. Weersma, Maija Wessman, James G. Wilson, Ramnik J. Xavier, Kent D. Taylor, Henry J. Lin, Stephen S. Rich, Wendy S. Post, Yii-Der Ida Chen, Jerome I. Rotter, Chad Nusbaum, Anthony Philippakis, Eric Lander, Stacey Gabriel, Benjamin M. Neale, Sekar Kathiresan, Mark J. Daly, Eric Banks, Daniel G. MacArthur, and Michael E. Talkowski. A structural variation reference for medical and population genetics. *Nature*, 581(7809):444–451, 2020.

[47] Nathaniel Comfort. Genetics: We are the 98%. *Nature*, 520(7549):615–616, 2015.

[48] ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489(7414):57–74, 2012.

[49] GTEx Consortium, Eric R Gamazon, Ayellet V Segrè, Martijn van de Bunt, Xiaoquan Wen, Hualin S Xi, Farhad Hormozdiari, Halit Ongen, Anuar Konkashbaev, Eske M Derks, François Aguet, Jie Quan, Dan L Nicolae, Eleazar Eskin, Manolis Kellis, Gad Getz, Mark I McCarthy, Emmanouil T Dermitzakis, Nancy J Cox, and Kristin G Ardlie. Using an atlas of gene regulation across 44 human tissues to inform complex disease- and trait-associated variation. *Nature Genetics*, 50(7):956–967, 2018.

[50] Mouse ENCODE Consortium, John A Stamatoyannopoulos, Michael Snyder, Ross Hardison, Bing Ren, Thomas Gingeras, David M Gilbert, Mark Groudine, Michael Bender, Rajinder Kaul, Theresa Canfield, Erica Giste, Audra Johnson, Mia Zhang, Gayathri Balasundaram, Rachel Byron, Vaughan Roach, Peter J Sabo, Richard Sandstrom, A Sandra Stehling, Robert E Thurman, Sherman M Weissman, Philip Cayting, Manoj Hariharan, Jin Lian, Yong Cheng, Stephen G Landt, Zhihai Ma, Barbara J Wold, Job Dekker, Gregory E Crawford, Cheryl A Keller, Weisheng Wu, Christopher Morrissey, Swathi A Kumar, Tejaswini Mishra, Deepti Jain, Marta Byrska-Bishop, Daniel Blankenberg, Bryan R Lajoie, Gaurav Jain, Amartya Sanyal, Kaun-Bei Chen, Olgert Denas, James Taylor, Gerd A Blobel, Mitchell J Weiss, Max Pimkin, Wulan Deng, Georgi K Marinov, Brian A Williams, Katherine I Fisher-Aylor, Gilberto Desalvo, Anthony Kiralusha, Diane Trout, Henry Amrhein, Ali Mortazavi, Lee Edsall, David McCleary, Samantha Kuan, Yin Shen, Feng Yue, Zhen Ye, Carrie A Davis, Chris Zaleski, Sonali Jha, Chenghai Xue, Alex Dobin, Wei Lin, Meagan Fastuca, Huaien Wang, Roderic Guigo, Sarah Djebali, Julien Lagarde, Tyrone Ryba, Takayo Sasaki, Venkat S Malladi, Melissa S Cline, Vanessa M Kirkup, Katrina Learned, Kate R Rosenbloom, W James Kent, Elise A Feingold, Peter J Good, Michael Pazin, Rebecca F Lowdon, and Leslie B Adams. An encyclopedia of mouse DNA elements (Mouse ENCODE). *Genome Biology*, 13(8):418, 2012.

[51] The FANTOM Consortium, Robin Andersson, Claudia Gebhard, Irene Miguel-Escalada, Ilka Hoof, Jette Bornholdt, Mette Boyd, Yun Chen, Xiaobei Zhao, Christian Schmidl, Takahiro Suzuki, Evgenia Ntini, Erik Arner, Eivind Valen, Kang Li, Lucia Schwarzfischer, Dagmar Glatz, Johanna Raithel, Berit Lilje, Nicolas Rapin, Frederik Otzen Bagger, Mette Jørgensen, Peter Refsing Andersen, Nicolas Bertin, Owen Rackham, A Maxwell Burroughs, J Kenneth Baillie, Yuri Ishizu, Yuri Shimizu, Erina Furuhata, Shiori Maeda, Yutaka Negishi, Christopher J Mungall, Terrence F Meehan, Timo Lassmann, Masayoshi Itoh, Hideya Kawaji, Naoto Kondo, Jun Kawai, Andreas Lennartsson, Carsten O Daub, Peter Heutink, David A Hume, Torben Heick Jensen, Harukazu Suzuki, Yoshihide Hayashizaki, Ferenc Müller, Alistair R R Forrest, Piero Carninci, Michael Rehli, and Albin Sandelin. An atlas of active enhancers across human cell types and tissues. *Nature*, 507(7493):455–461, 2014.

[52] Gregory M. Cooper and Jay Shendure. Needles in stacks of needles: finding disease-

causal variants in a wealth of genomic data. *Nature Reviews Genetics*, 12(9):628–640, 2011.

[53] Axel Cournac, Hervé Marie-Nelly, Martial Marbouty, Romain Koszul, and Julien Mozziconacci. Normalization of a chromosomal contact map. *BMC Genomics*, 13(1):436, 2012.

[54] T Cremer and Cremer C. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature Reviews Genetics*, 2001.

[55] Caelin Cubeñas-Potts and Victor G Corces. Topologically Associating Domains: An invariant framework or a dynamic scaffold? *Nucleus (Austin, Tex.)*, 6(6):430–4, 2015.

[56] Rola Dali and Mathieu Blanchette. A critical assessment of topologically associating domain prediction tools. *Nucleic acids research*, 45(6):2994–3005, 2017.

[57] Charles G. Danko, Lauren A. Choate, Brooke A. Marks, Edward J. Rice, Zhong Wang, Tinyi Chu, Andre L. Martins, Noah Dukler, Scott A. Coonrod, Elia D. Tait Wojno, John T. Lis, W. Lee Kraus, and Adam Siepel. Dynamic evolution of regulatory element ensembles in primate CD4+ T cells. *Nature Ecology & Evolution*, 2(3):537–548, 2018.

[58] Emily M. Darrow, Miriam H. Huntley, Olga Dudchenko, Elena K. Stamenova, Neva C. Durand, Zhuo Sun, Su-Chen Huang, Adrian L. Sanborn, Ido Machol, Muhammad Shamim, Andrew P. Seberg, Eric S. Lander, Brian P. Chadwick, and Erez Lieberman Aiden. Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proceedings of the National Academy of Sciences*, 113(31):E4504–E4512, 2016.

[59] Jacob F. Degner, Athma A. Pai, Roger Pique-Regi, Jean-Baptiste Veyrieras, Daniel J. Gaffney, Joseph K. Pickrell, Sherryl De Leon, Katelyn Michelini, Noah Lewellen, Gregory E. Crawford, Matthew Stephens, Yoav Gilad, and Jonathan K. Pritchard. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature*, 482(7385):390–394, 2012.

[60] J Dekker, K Rippe, M Dekker, and Kleckner N. Capturing chromosome conformation. *Science*, 2002.

[61] Job Dekker and Edith Heard. Structural and functional diversity of Topologically Associating Domains. *FEBS letters*, 589(20 Pt A):2877–84, 2015.

[62] O Delaneau, M Zazhytska, C Borel, G Giannuzzi, G Rey, C Howald, S Kumar, H Ongen, K Popadin, D Marbach, G Ambrosini, D Bielser, D Hacker, L Romano, P Ribaux, M Wiederkehr, E Falconnet, P Bucher, S Bergmann, S E Antonarakis, A Reymond, and E T Dermitzakis. Chromatin three-dimensional interactions mediate genetic effects on gene expression. *Science (New York, N.Y.)*, 364(6439), 2019.

[63] Wulan Deng, Jeremy W. Rupon, Ivan Krivega, Laura Breda, Irene Motta, Kristen S. Jahn, Andreas Reik, Philip D. Gregory, Stefano Rivella, Ann Dean, and Gerd A. Blobel. Reactivation of Developmentally Silenced Globin Genes by Forced Chromatin Looping. *Cell*, 158(4):849–860, 2014.

[64] Briana K Dennehey, Diane G Gutches, Edwin H McConkey, and Kenneth S Krauter. Inversion, duplication, and changes in gene context are associated with human chromosome 18 evolution. *Genomics*, 83(3):493–501, 2004.

[65] Vishnu Dileep, Korey A. Wilson, Claire Marchal, Xiaowen Lyu, Peiyao A. Zhao, Ben Li, Axel Poulet, Daniel A. Bartlett, Juan Carlos Rivera-Mulia, Zhaohui S. Qin, Allan J. Robins, Thomas C. Schulz, Michael J. Kulik, Rachel Patton McCord, Job Dekker, Stephen Dalton, Victor G. Corces, and David M. Gilbert. Rapid Irreversible Transcriptional Reprogramming in Human Stem Cells Accompanied by Discordance between Replication Timing and Chromatin Compartment. *Stem Cell Reports*, 13(1):193–206, 2019.

[66] Le F Dily, D Baù, A Pohl, GP Vicent, Serra F, D Soronellas, G Castellano, R Wright, C Ballare, G Filion, and M Marti-Renom. Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation. *Genes Dev*, 2014.

[67] Jesse R Dixon, David U Gorkin, and Bing Ren. Chromatin Domains: The Unit of Chromosome Organization. *Molecular Cell*, 62(5):668–680, 2016.

[68] Jesse R Dixon, Inkyung Jung, Siddarth Selvaraj, Yin Shen, Jessica E Antosiewicz-Bourget, Ah Y Lee, Zhen Ye, Audrey Kim, Nisha Rajagopal, Wei Xie, Yarui Diao, Jing Liang, Huimin Zhao, Victor V Lobanenkov, Joseph R Ecker, James A Thomson, and Bing Ren. Chromatin architecture reorganization during stem cell differentiation. *Nature*, 518(7539):331–6, 2015.

[69] Jesse R Dixon, Siddarth Selvaraj, Feng Yue, Audrey Kim, Yan Li, Yin Shen, Ming Hu, Jun S Liu, and Bing Ren. Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature*, 485(7398):376–380, 2012.

[70] Mohamed Nadhir Djekidel, Yang Chen, and Michael Q. Zhang. FIND: difFerential chromatin INteractions Detection using a spatial Poisson process. *Genome Research*, 28(3):412–422, 2018.

[71] Pengfei Dong, Xiaoyu Tu, Po-Yu Chu, Peitao Lü, Ning Zhu, Donald Grierson, Baijuan Du, Pinghua Li, and Silin Zhong. 3D Chromatin Architecture of Large Plant Genomes Determined by Local A/B Compartments. *Molecular plant*, 10(12):1497–1509, 2017.

[72] Xiao Dong, Chao Li, Yunqin Chen, Guohui Ding, and Yixue Li. Human transcriptional interactome of chromatin contribute to gene co-expression. *BMC genomics*, 11:704, 2010.

[73] Robin D Dowell. The similarity of gene expression between human and mouse tissues. *Genome biology*, 12(1):101, 2011.

[74] Neva C Durand, Muhammad S Shamim, Ido Machol, Suhas S Rao, Miriam H Huntley, Eric S Lander, and Erez L Aiden. Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. *Cell systems*, 3(1):95–8, 2016.

[75] Z Duren, X Chen, R Jiang, Y Wang, and WH Wong. Modeling gene regulation from paired expression and chromatin accessibility data. *Proceedings of the National Academy of Sciences*, 2017.

[76] Kyle P. Eagen, Erez Lieberman Aiden, and Roger D. Kornberg. Polycomb-mediated chromatin loops revealed by a subkilobase-resolution chromatin interaction map. *Proceedings of the National Academy of Sciences*, 114(33):8764–8769, 2017.

[77] Lee E Edsall, Alejandro Berrio, William H Majoros, Devjanee Swain-Lenz, Shauna Morrow, Yoichiro Shibata, Alexias Safi, Gregory A Wray, Gregory E Crawford, and Andrew S Allen. Evaluating chromatin accessibility differences across multiple primate species using a joint modelling approach. *Genome Biology and Evolution*, 11(10):3035–3053, 2019.

[78] Stacey L. Edwards, Jonathan Beesley, Juliet D. French, and Alison M. Dunning. Beyond GWASs: Illuminating the Dark Road from Association to Function. *The American Journal of Human Genetics*, 93(5):779–797, 2013.

[79] Wolfgang Enard, Philipp Khaitovich, Joachim Klose, Sebastian Zöllner, Florian Heissig, Patrick Giavalisco, Kay Nieselt-Struwe, Elaine Muchmore, Ajit Varki, Rivka Ravid, Gaby M Doxiadis, Ronald E Bontrop, and Svante Pääbo. Intra- and interspecific variation in primate gene expression patterns. *Science (New York, N.Y.)*, 296(5566):340–3, 2002.

[80] Ittai E Eres, Kaixuan Luo, Chiaowen Joyce Hsiao, Lauren E Blake, and Yoav Gilad. Reorganization of 3D genome structure may contribute to gene regulatory evolution in primates. *PLOS Genetics*, 15(7):e1008278, 2019.

[81] Jason Ernst and Manolis Kellis. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods*, 9(3):215–216, 2012.

[82] Sergio Martin Espinola, Markus Götz, Jean-Bernard Fiche, Maelle Bellec, Christophe Houbron, Andrés M. Cardozo Gizzi, Mounia Lagha, and Marcelo Nollmann. Cis-regulatory chromatin loops arise before TADs and gene activation, and are independent of cell fate during development. *bioRxiv*, page 2020.07.07.191015, 2020.

[83] Tayaza Fadason, William Schierding, Thomas Lumley, and Justin M. O'Sullivan. Chromatin interactions and expression quantitative trait loci reveal genetic drivers of multimorbidities. *Nature Communications*, 9(1):5198, 2018.

[84] Linn Fagerberg, Björn M M Hallström, Per Oksvold, Caroline Kampf, Dijana Djureinovic, Jacob Odeberg, Masato Habuka, Simin Tahmasebpoor, Angelika Danielsson, Karolina Edlund, Anna Asplund, Evelina Sjöstedt, Emma Lundberg, Cristina A Szigyarto, Marie Skogs, Jenny O Takanen, Holger Berling, Hanna Tegel, Jan Mulder, Peter Nilsson, Jochen M Schwenk, Cecilia Lindskog, Frida Danielsson, Adil Mardinoglu, Asa Sivertsson, Kalle von Feilitzen, Mattias Forsberg, Martin Zwahlen, IngMarie Olsson, Sanjay Navani, Mikael Huss, Jens Nielsen, Fredrik Ponten, and Mathias Uhlén. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Molecular & cellular proteomics : MCP*, 13(2):397–406, 2014.

[85] Lindsey R. Fernandez, Thomas G. Gilgenast, and Jennifer E. Phillips-Cremins. 3DeFDR: statistical methods for identifying cell type-specific looping interactions in 5C and Hi-C data. *Genome Biology*, 21(1):219, 2020.

[86] Darya Filippova, Rob Patro, Geet Duggal, and Carl Kingsford. Identification of alternative topological domains in chromatin. *Algorithms for Molecular Biology*, 9(1):14, 2014.

[87] Elizabeth H Finn, Gianluca Pegoraro, Hugo B Brandão, Anne-Laure Valton, Marlies E Oomen, Job Dekker, Leonid Mirny, and Tom Misteli. Extensive Heterogeneity and Intrinsic Variation in Spatial Genome Organization. *Cell*, 176(6):1502–1515.e10, 2019.

[88] Julia Fischer, Linda Koch, Christian Emmerling, Jeanette Vierkotten, Thomas Peters, Jens C. Brüning, and Ulrich Rüther. Inactivation of the Fto gene protects from obesity. *Nature*, 458(7240):894–898, 2009.

[89] Mattia Forcato, Chiara Nicoletti, Koustav Pal, Carmen M Livi, Francesco Ferrari, and Silvio Bicciato. Comparison of computational methods for Hi-C data analysis. *Nature methods*, 2017.

[90] Martin Franke and José Luis Gómez-Skarmeta. An evolutionary perspective of regulatory landscape dynamics in development and disease. *Current opinion in cell biology*, 55:24–29, 2018.

[91] Martin Franke, Daniel M. Ibrahim, Guillaume Andrey, Wibke Schwarzer, Verena Heinrich, Robert Schöpflin, Katerina Kraft, Rieke Kempfer, Ivana Jerković, Wing-Lee Chan, Malte Spielmann, Bernd Timmermann, Lars Wittler, Ingo Kurth, Paola Cambiaso, Orsetta Zuffardi, Gunnar Houge, Lindsay Lambie, Francesco Brancati, Ana Pombo, Martin Vingron, Francois Spitz, and Stefan Mundlos. Formation of new chromatin domains determines pathogenicity of genomic duplications. *Nature*, 538(7624):265–269, 2016.

[92] Nicolás Frankel, Shu Wang, and David L Stern. Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 109(51):20975–9, 2012.

[93] James Fraser, Carmelo Ferrai, Andrea M Chiariello, Markus Schueler, Tiago Rito, Giovanni Laudanno, Mariano Barbieri, Benjamin L Moore, Dorothee CA Kraemer, Stuart Aitken, Sheila Q Xie, Kelly J Morris, Masayoshi Itoh, Hideya Kawaji, Ines Jaeger, Yoshihide Hayashizaki, Piero Carninci, Alistair RR Forrest, The FANTOM Consortium, Colin A Semple, Josée Dostie, Ana Pombo, and Mario Nicodemi. Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation. *Molecular Systems Biology*, 11(12):852, 2015.

[94] James Fraser, Iain Williamson, Wendy A. Bickmore, and Josée Dostie. An Overview of Genome Organization and How We Got There: from FISH to Hi-C. *Microbiology and Molecular Biology Reviews*, 79(3):347–372, 2015.

[95] Geoffrey Fudenberg, Maxim Imakaev, Carolyn Lu, Anton Goloborodko, Nezar Abdennur, and Leonid A Mirny. Formation of Chromosomal Domains by Loop Extrusion. *Cell reports*, 15(9):2038–49, 2016.

[96] Takashi Fukaya, Bomyi Lim, and Michael Levine. Enhancer Control of Transcriptional Bursting. *Cell*, 166(2):358–368, 2016.

[97] Rafael Galupa and Edith Heard. X-Chromosome Inactivation: A Crossroads Between Chromosome Architecture and Gene Regulation. *Annual Review of Genetics*, 52(1):535–566, 2018.

[98] Rafael Galupa, Elphège Pierre Nora, Rebecca Worsley-Hunt, Christel Picard, Chris Gard, Joke Gerarda van Bemmel, Nicolas Servant, Yinxiu Zhan, Fatima El Marjou, Colin Johanneau, Patricia Diabangouaya, Agnès Le Saux, Sonia Lameiras, Juliana Pipoli da Fonseca, Friedemann Loos, Joost Gribnau, Sylvain Baulande, Uwe Ohler, Luca Giorgetti, and Edith Heard. A Conserved Noncoding Locus Regulates Random Monoallelic Xist Expression across a Topological Boundary. *Molecular Cell*, 77(2):352–367.e8, 2020.

[99] Xue Gao, Yong-Hyun Shin, Min Li, Fei Wang, Qiang Tong, and Pumin Zhang. The Fat Mass and Obesity Associated Gene FTO Functions in the Brain to Regulate Postnatal Growth in Mice. *PLoS ONE*, 5(11):e14005, 2010.

[100] Estela García-González, Martín Escamilla-Del-Arenal, Rodrigo Arzate-Mejía, and Félix Recillas-Targa. Chromatin remodeling effects on enhancer activity. *Cellular and molecular life sciences : CMLS*, 73(15):2897–910, 2016.

[101] Raquel García-Pérez, Paula Esteller-Cucala, Glòria Mas, Irene Lobón, Valerio Di Carlo, Meritxell Riera, Martin Kuhlwilm, Arcadi Navarro, Antoine Blancher, Luciano Di Croce, José Luis Gómez-Skarmeta, David Juan, and Tomàs Marquès-Bonet. Epigenomic profiling of primate LCLs reveals the coordinated evolution of gene expression and epigenetic signals in regulatory architectures. *bioRxiv*, page 2019.12.18.872531, 2020.

[102] Mark B Gerstein, Anshul Kundaje, Manoj Hariharan, Stephen G Landt, Koon-Kiu K Yan, Chao Cheng, Xinmeng J Mu, Ekta Khurana, Joel Rozowsky, Roger Alexander, Renqiang Min, Pedro Alves, Alexej Abyzov, Nick Addleman, Nitin Bhardwaj, Alan P Boyle, Philip Cayting, Alexandra Charos, David Z Chen, Yong Cheng, Declan Clarke, Catharine Eastman, Ghia Euskirchen, Seth Frietze, Yao Fu, Jason Gertz, Fabian Grubert, Arif Harmanci, Preti Jain, Maya Kasowski, Phil Lacroute, Jing J Leng, Jin Lian, Hannah Monahan, Henriette O'Geen, Zhengqing Ouyang, E C Partridge, Dorrelyn Patacsil, Florencia Pauli, Debasish Raha, Lucia Ramirez, Timothy E Reddy, Brian Reed, Minyi Shi, Teri Slifer, Jing Wang, Linfeng Wu, Xinqiong Yang, Kevin Y Yip, Gili Zilberman-Schapira, Serafim Batzoglou, Arend Sidow, Peggy J Farnham, Richard M Myers, Sherman M Weissman, and Michael Snyder. Architecture of the human regulatory network derived from ENCODE data. *Nature*, 489(7414):91–100, 2012.

[103] Yad Ghavi-Helm, Aleksander Jankowski, Sascha Meiers, Rebecca R. Viales, Jan O. Korbel, and Eileen E. M. Furlong. Highly rearranged chromosomes reveal uncoupling between genome topology and gene expression. *Nature Genetics*, pages 1–11, 2019.

[104] Thomas Giger, Philipp Khaitovich, Mehmet Somel, Anna Lorenc, Esther Lizano, Laura W. Harris, Margaret M. Ryan, Martin Lan, Matthew T. Wayland, Sabine Bahn, and Svante Pääbo. Evolution of Neuronal and Endothelial Transcriptomes in Primates. *Genome Biology and Evolution*, 2:284–292, 2010.

[105] Yoav Gilad and Orna Mizrahi-Man. A reanalysis of mouse ENCODE comparative gene expression data. *F1000Research*, 4:121, 2015.

[106] Yoav Gilad, Alicia Oshlack, and Scott A. Rifkin. Natural selection on gene expression. *Trends in Genetics*, 22(8):456–461, 2006.

[107] Yoav Gilad, Alicia Oshlack, Gordon K. Smyth, Terence P. Speed, and Kevin P. White. Expression profiling in primates reveals a rapid evolution of human transcription factors. *Nature*, 440(7081), 2006.

[108] Yoav Gilad, Scott A. Rifkin, and Jonathan K. Pritchard. Revealing the architecture of gene regulation: the promise of eQTL studies. *Trends in Genetics*, 24(8):408–415, 2008.

[109] Andrés M Cardozo Gizzi, Diego I Cattoni, and Marcelo Nollmann. TADs or no TADS: Lessons From Single-cell Imaging of Chromosome Architecture. *Journal of Molecular Biology*, 432(3):682–693, 2020.

[110] Brian S. Gloss and Marcel E. Dinger. Realizing the significance of noncoding functionality in clinical genomics. *Experimental & Molecular Medicine*, 50(8):97, 2018.

[111] Carlos Gómez-Marín, Juan J Tena, Rafael D Acemel, Macarena López-Mayorga, Silvia Naranjo, Elisa de la Calle-Mustienes, Ignacio Maeso, Leonardo Beccari, Ivy Aneas,

Erika Vielmas, Paola Bovolenta, Marcelo A Nobrega, Jaime Carvajal, and José Gómez-Skarmeta. Evolutionary comparison reveals that diverging CTCF sites are signatures of ancestral topological associating domains borders. *Proceedings of the National Academy of Sciences*, 112(24):7542–7547, 2015.

[112] David U Gorkin, Yunjiang Qiu, Ming Hu, Kipper Fletez-Brant, Tristin Liu, Anthony D Schmitt, Amina Noor, Joshua Chiou, Kyle J Gaulton, Jonathan Sebat, Yun Li, Kasper D Hansen, and Bing Ren. Common DNA sequence variation influences 3-dimensional conformation of the human genome. *Genome Biology*, 20(1):255, 2019.

[113] William W. Greenwald, He Li, Paola Benaglio, David Jakubosky, Hiroko Matsui, Anthony Schmitt, Siddarth Selvaraj, Matteo D'Antonio, Agnieszka D'Antonio-Chronowska, Erin N. Smith, and Kelly A. Frazer. Subtle changes in chromatin loop contact propensity are associated with differential gene regulation and expression. *Nature Communications*, 10(1):1054, 2019.

[114] Louise G. Grunnet, Emma Nilsson, Charlotte Ling, Torben Hansen, Oluf Pedersen, Leif Groop, Allan Vaag, and Pernille Poulsen. Regulation and Function of FTO mRNA Expression in Human Skeletal Muscle and Subcutaneous Adipose Tissue. *Diabetes*, 58(10):2402–2408, 2009.

[115] Ya Guo, Quan Xu, Daniele Canzio, Jia Shou, Jinhuan Li, David U. Gorkin, Inkyung Jung, Haiyang Wu, Yanan Zhai, Yuanxiao Tang, Yichao Lu, Yonghu Wu, Zhilian Jia, Wei Li, Michael Q. Zhang, Bing Ren, Adrian R. Krainer, Tom Maniatis, and Qiang Wu. CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell*, 162(4):900–10, 2015.

[116] Nathan Harmston, Elizabeth Ing-Simmons, Ge Tan, Malcolm Perry, Matthias Merkenschlager, and Boris Lenhard. Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nature communications*, 8(1):441, 2017.

[117] Sven Heinz, Christopher Benner, Nathanael Spann, Eric Bertolino, Yin C. Lin, Peter Laslo, Jason X. Cheng, Cornelis Murre, Harinder Singh, and Christopher K. Glass. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, 38(4):576–589, 2010.

[118] Denes Hnisz, Abraham S. Weintraub, Daniel S. Day, Anne-Laure Valton, Rasmus O. Bak, Charles H. Li, Johanna Goldmann, Bryan R. Lajoie, Zi Peng Fan, Alla A. Sigova, Jessica Reddy, Diego Borges-Rivera, Tong Ihn Lee, Rudolf Jaenisch, Matthew H. Porteus, Job Dekker, and Richard A. Young. Activation of proto-oncogenes by disruption of chromosome neighborhoods. *Science*, 351(6280):1454–1458, 2016.

[119] Hopi E. Hoekstra and Jerry A. Coyne. The Locus Of Evolution: Evo Devo And The Genetics Of Adaptation. *Evolution*, 61(5):995–1016, 2007.

[120] Michael M Hoffman, Orion J Buske, Jie Wang, Zhiping Weng, Jeff A Bilmes, and William Stafford Noble. Unsupervised pattern discovery in human chromatin structure through genomic segmentation. *Nature Methods*, 9(5):473–476, 2012.

[121] D Homouz and Kudlicki AS. The 3D organization of the yeast genome correlates with co-expression and reflects functional relations between genes. *PLoS One*, 2013.

[122] Chunhui Hou, Li Li, Zhaohui S Qin, and Victor G Corces. Gene density, transcription, and insulators contribute to the partition of the Drosophila genome into physical domains. *Molecular cell*, 48(3):471–84, 2012.

[123] Genevieve Housman and Yoav Gilad. Prime time for primate functional genomics. *Current opinion in genetics & development*, 62:1–7, 2020.

[124] Hai Yang Hu, Song Guo, Jiang Xi, Zheng Yan, Ning Fu, Xiaoyu Zhang, Corinna Menzel, Hongyu Liang, Hongyi Yang, Min Zhao, Rong Zeng, Wei Chen, Svante Pääbo, and Philipp Khaitovich. MicroRNA Expression and Regulation in Human, Chimpanzee, and Macaque Brains. *PLoS Genetics*, 7(10):e1002327, 2011.

[125] Ming Hu, Ke Deng, Siddarth Selvaraj, Zhaohui Qin, Bing Ren, and Jun S Liu. HiCNorm: removing biases in Hi-C data via Poisson regression. *Bioinformatics (Oxford, England)*, 28(23):3131–3, 2012.

[126] Peng Huang, Cheryl A Keller, Belinda Giardine, Jeremy D Grevet, James OJ Davies, Jim R Hughes, Ryo Kurita, Yukio Nakamura, Ross C Hardison, and Gerd A Blobel. Comparative analysis of three-dimensional chromosomal architecture identifies a novel fetal hemoglobin regulatory element. *Genes & development*, 31(16):1704–1713, 2017.

[127] Jonas Ibn-Salem, Sebastian Köhler, Michael I Love, Ho-Ryun Chung, Ni Huang, Matthew E Hurles, Melissa Haendel, Nicole L Washington, Damian Smedley, Christopher J Mungall, Suzanna E Lewis, Claus-Eric Ott, Sebastian Bauer, Paul N Schofield, Stefan Mundlos, Malte Spielmann, and Peter N Robinson. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biology*, 15(9):423, 2014.

[128] Maxim Imakaev, Geoffrey Fudenberg, Rachel P McCord, Natalia Naumova, Anton Goloborodko, Bryan R Lajoie, Job Dekker, and Leonid A Mirny. Iterative correction of Hi-C data reveals hallmarks of chromosome organization. *Nature methods*, 9(10):999–1003, 2012.

[129] Elizabeth Ing-Simmons, Roshan Vaid, Mattias Mannervik, and Juan M. Vaquerizas. Independence of 3D chromatin conformation and gene regulation during Drosophila dorsoventral patterning. *bioRxiv*, page 2020.07.07.186791, 2020.

[130] Dekker J. Gene regulation in the third dimension. *Science*, 2008.

[131] Yongpeng Jiang, Jie Huang, Kehuan Lun, Boyuan Li, Haonan Zheng, Yuanjun Li, Rong Zhou, Wenjia Duan, Chenlu Wang, Yuanqing Feng, Hong Yao, Cheng Li, and Xiong Ji. Genome-wide analyses of chromatin interactions after the loss of Pol I, Pol II, and Pol III. *Genome Biology*, 21(1):158, 2020.

[132] Fulai Jin, Yan Li, Jesse R Dixon, Siddarth Selvaraj, Zhen Ye, Ah Y Lee, Chia-An A Yen, Anthony D Schmitt, Celso A Espinoza, and Bing Ren. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature*, 503(7475):290–4, 2013.

[133] Stephan Kadauke and Gerd A. Blobel. Chromatin loops in gene regulation. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, 1789(1):17–25, 2009.

[134] MH Kagey, JJ Newman, S Bilodeau, Y Zhan, DA Orlando, NL Van Berkum, CC Ebmeier, J Goossens, PB Rahl, SS Levine, DJ Taatjes, J Dekker, and Young RA. Mediator and cohesin connect gene expression and chromatin architecture. *Nature*, 2010.

[135] Mazen W Karaman, Marlys L Houck, Leona G Chemnick, Shailender Nagpal, Daniel Chawannakul, Dominick Sudano, Brian L Pike, Vincent V Ho, Oliver A Ryder, and Joseph G Hacia. Comparative analysis of gene-expression patterns in human and African great ape cultured fibroblasts. *Genome research*, 13(7):1619–30, 2003.

[136] Rosa Karlić, Ho-Ryun Chung, Julia Lasserre, Kristian Vlahoviček, and Martin Vingron. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences*, 107(7):2926–2931, 2010.

[137] Hildegard Kehrer-Sawatzki and David N. Cooper. Understanding the recent evolution of the human genome: insights from human–chimpanzee genome comparisons. *Human Mutation*, 28(2):99–130, 2007.

[138] Hildegard Kehrer-Sawatzki, Justyna M. Szamalek, Simone Tänzer, Matthias Platzer, and Horst Hameister. Molecular characterization of the pericentric inversion of chimpanzee chromosome 11 homologous to human chromosome 9. *Genomics*, 85(5):542–550, 2005.

[139] Manolis Kellis, Barbara Wold, Michael P. Snyder, Bradley E. Bernstein, Anshul Kundaje, Georgi K. Marinov, Lucas D. Ward, Ewan Birney, Gregory E. Crawford, Job Dekker, Ian Dunham, Laura L. Elnitski, Peggy J. Farnham, Elise A. Feingold, Mark Gerstein, Morgan C. Giddings, David M. Gilbert, Thomas R. Gingeras, Eric D. Green, Roderic Guigo, Tim Hubbard, Jim Kent, Jason D. Lieb, Richard M. Myers, Michael J. Pazin, Bing Ren, John A. Stamatoyannopoulos, Zhiping Weng, Kevin P. White, and Ross C. Hardison. Defining functional DNA elements in the human genome. *Proceedings of the National Academy of Sciences*, 111(17):6131–6138, 2014.

[140] WJ Kent, CW Sugnet, TS Furey, KM Roskin, TH Pringle, AM Zahler, and Haussler D. The human genome browser at UCSC. *Genome Research*, 2002.

[141] Elissavet Kentepozidou, Sarah J Aitken, Christine Feig, Klara Stefflova, Ximena Ibarra-Soria, Duncan T Odom, Maša Roller, and Paul Flicek. Clustered CTCF binding is an evolutionary mechanism to maintain topologically associating domains. *Genome biology*, 21(1):5, 2020.

[142] Philipp Khaitovich, Wolfgang Enard, Michael Lachmann, and Svante Pääbo. Evolution of primate gene expression. *Nature Reviews Genetics*, 7(9):693–702, 2006.

[143] Zia Khan, Michael J. Ford, Darren A. Cusanovich, Amy Mitrano, Jonathan K. Pritchard, and Yoav Gilad. Primate Transcript and Protein Expression Levels Evolve Under Compensatory Selection Pressures. *Science*, 342(6162):1100–1104, 2013.

[144] Dong Seon Kim and Yoonsoo Hahn. Identification of novel phosphorylation modification sites in human proteins that originated after the human–chimpanzee divergence. *Bioinformatics*, 27(18):2494–2501, 2011.

[145] Seungsoo Kim, Ivan Liachko, Donna G Brickner, Kate Cook, William S Noble, Jason H Brickner, Jay Shendure, and Maitreya J Dunham. The dynamic three-dimensional organization of the diploid yeast genome. *eLife*, 6:e23623, 2017.

[146] Motoo Kimura. Evolutionary Rate at the Molecular Level. *Nature*, 217(5129):624–626, 1968.

[147] MC King and Wilson AC. Evolution at two levels in humans and chimpanzees. *Science*, 1975.

[148] David C. Klein and Sarah J. Hainer. Genomic methods in profiling DNA accessibility and factor localization. *Chromosome Research*, 28(1):69–85, 2020.

[149] Jason C. Klein, Aidan Keith, Vikram Agarwal, Timothy Durham, and Jay Shendure. Functional characterization of enhancer evolution in the primate lineage. *Genome Biology*, 19(1):99, 2018.

[150] Sandy L. Klemm, Zohar Shipony, and William J. Greenleaf. Chromatin accessibility and the regulatory epigenome. *Nature Reviews Genetics*, 20(4):207–220, 2019.

[151] P A Knight and D Ruiz. A fast algorithm for matrix balancing. *IMA Journal of Numerical Analysis*, 33(3):1029–1047, 2012.

[152] ST Kosak and Groudine M. Form follows function: the genomic organization of cellular differentiation. *Genes Dev*, 2004.

[153] Jan Krefting, Miguel A Andrade-Navarro, and Jonas Ibn-Salem. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. *BMC biology*, 16(1):87, 2018.

[154] Bryan R. Lajoie, Job Dekker, and Noam Kaplan. The Hitchhiker's guide to Hi-C analysis: Practical guidelines. *Methods*, 72:65–75, 2015.

[155] Xun Lan, Heather Witt, Koichi Katsumura, Zhenqing Ye, Qianben Wang, Emery H Bresnick, Peggy J Farnham, and Victor X Jin. Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages. *Nucleic acids research*, 40(16):7690–7704, 2012.

[156] E S Lander, L M Linton, B Birren, C Nusbaum, M C Zody, J Baldwin, K Devon, K Dewar, M Doyle, W FitzHugh, R Funke, D Gage, K Harris, A Heaford, J Howland, L Kann, J Lehoczky, R LeVine, P McEwan, K McKernan, J Meldrim, J P Mesirov, C Miranda, W Morris, J Naylor, C Raymond, M Rosetti, R Santos, A Sheridan, C Sougnez, Y Stange-Thomann, N Stojanovic, A Subramanian, D Wyman, J Rogers, J Sulston, R Ainscough, S Beck, D Bentley, J Burton, C Clee, N Carter, A Coulson, R Deadman, P Deloukas, A Dunham, I Dunham, R Durbin, L French, D Grafham, S Gregory, T Hubbard, S Humphray, A Hunt, M Jones, C Lloyd, A McMurray, L Matthews, S Mercer, S Milne, J C Mullikin, A Mungall, R Plumb, M Ross, R Shownkeen, S Sims, R H Waterston, R K Wilson, L W Hillier, J D McPherson, M A Marra, E R Mardis, L A Fulton, A T Chinwalla, K H Pepin, W R Gish, S L Chissoe, M C Wendl, K D Delehaunty, T L Miner, A Delehaunty, J B Kramer, L L Cook, R S Fulton, D L Johnson, P J Minx, S W Clifton, T Hawkins, E Branscomb, P Predki, P Richardson, S Wenning, T Slezak, N Doggett, J F Cheng, A Olsen, S Lucas, C Elkin, E Uberbacher, M Frazier, R A Gibbs, D M Muzny, S E Scherer, J B Bouck, E J Sodergren, K C Worley, C M Rives, J H Gorrell, M L Metzker, S L Naylor, R S Kucherlapati, D L Nelson, G M Weinstock, Y Sakaki, A Fujiyama, M Hattori, T Yada, A Toyoda, T Itoh, C Kawagoe, H Watanabe, Y Totoki, T Taylor, J Weissenbach, R Heilig, W Saurin, F Artiguenave, P Brottier, T Bruls, E Pelletier, C Robert, P Wincker, D R Smith, L Doucette-Stamm, M Rubenfield, K Weinstock, H M Lee, J Dubois, A Rosenthal, M Platzer, G Nyakatura, S Taudien, A Rump, H Yang, J Yu, J Wang, G Huang, J Gu, L Hood, L Rowen, A Madan, S Qin, R W Davis, N A Federspiel, A P Abola, M J Proctor, R M Myers, J Schmutz, M Dickson, J Grimwood, D R Cox, M V Olson, R Kaul, C Raymond, N Shimizu, K Kawasaki, S Minoshima, G A Evans, M Athanasiou, R Schultz, B A Roe, F Chen, H Pan, J Ramser, H Lehrach, R Reinhardt, W R McCombie, M de la Bastide, N Dedhia, H Blöcker, K Hornischer, G Nordsiek, R Agarwala, L Aravind, J A Bailey, A Bateman, S Batzoglou, E Birney, P Bork, D G Brown, C B Burge, L Cerutti, H C Chen, D Church, M Clamp, R R Copley, T Doerks, S R Eddy, E E Eichler, T S Furey, J Galagan, J G Gilbert, C Harmon, Y Hayashizaki, D Haussler, H Hermjakob, K Hokamp, W Jang, L S Johnson, T A Jones, S Kasif, A Kaspryzk, S Kennedy, W J Kent, P Kitts, E V Koonin, I Korf, D Kulp, D Lancet, T M Lowe, A McLysaght, T Mikkelsen, J V Moran, N Mulder, V J Pollara, C P Ponting, G Schuler, J Schultz, G Slater, A F Smit, E Stupka, J Szustakowki, D Thierry-Mieg, J Thierry-Mieg, L Wagner, J Wallis, R Wheeler, A Williams, Y I Wolf, K H Wolfe, S P Yang, R F Yeh, F Collins, M S Guyer, J Peterson, A Felsenfeld, K A Wetterstrand, A Patrinos, M J Morgan, P de Jong, J J Catanese, K Osoegawa, H Shizuya, S Choi, Y J Chen, J Szustakowki, and International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature*, 409(6822):860–921, 2001.

[157] Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357, 2012.

[158] Magdalena Laugsch, Michaela Bartusel, Rizwan Rehimi, Hafiza Alirzayeva, Agathi Karaolidou, Giuliano Crispatzu, Peter Zentis, Milos Nikolic, Tore Bleckwehl, Petros Kolovos, Wilfred F.J. van Ijcken, Tomo Šarić, Katrin Koehler, Peter Frommolt, Katherine Lachlan, Julia Baptista, and Alvaro Rada-Iglesias. Modeling the Pathological Long-Range Regulatory Effects of Human Structural Variation with Patient-Specific hiPSCs. *Cell Stem Cell*, 24(5):736–752.e12, 2019.

[159] Charity W Law, Yunshun Chen, Wei Shi, and Gordon K Smyth. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, 15(2):R29, 2014.

[160] Nathan H Lazar, Kimberly A Nevonen, Brendan O'Connell, Christine McCann, Rachel J O'Neill, Richard E Green, Thomas J Meyer, Mariam Okhovat, and Lucia Carbone. Epigenetic maintenance of topological domains in the highly rearranged gibbon genome. *Genome research*, 28(7):983–997, 2018.

[161] Jongin Lee, Woon-young Hong, Minah Cho, Mikang Sim, Daehwan Lee, Younhee Ko, and Jaebum Kim. Synteny Portal: a web-based application portal for synteny block analysis. *Nucleic Acids Research*, 44(W1):W35–W40, 2016.

[162] Bernardo Lemos, Colin D. Meiklejohn, Mario Cceres, and Daniel L. Hartl. Rates Of Divergence In Gene Expression Profiles Of Primates, Mice, And Flies: Stabilizing Selection And Variability Among Functional Categories. *Evolution*, 59(1):126–137, 2005.

[163] Mike Levine. Transcriptional enhancers in animal development and evolution. *Current biology : CB*, 20(17):R754–63, 2010.

[164] Jiao Li, Kaimeng Huang, Gongcheng Hu, Isaac A. Babarinde, Yaoyi Li, Xiaotao Dong, Yu-Sheng Chen, Liping Shang, Wenjing Guo, Junwei Wang, Zhaoming Chen, Andrew P. Hutchins, Yun-Gui Yang, and Hongjie Yao. An alternative CTCF isoform antagonizes canonical CTCF occupancy and changes chromatin architecture to promote apoptosis. *Nature Communications*, 10(1):1535, 2019.

[165] Li Li, Xiaowen Lyu, Chunhui Hou, Naomi Takenaka, Huy Q. Nguyen, Chin-Tong Ong, Caelin Cubeñas-Potts, Ming Hu, Elissa P. Lei, Giovanni Bosco, Zhaohui S. Qin, and Victor G. Corces. Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing. *Molecular Cell*, 58(2):216–231, 2015.

[166] Yifeng Li, Wenqiang Shi, and Wyeth W. Wasserman. Genome-wide prediction of cis-regulatory regions using supervised deep learning methods. *BMC Bioinformatics*, 19(1):202, 2018.

[167] Erez Lieberman-Aiden, Nynke L van Berkum, Louise Williams, Maxim Imakaev, Tobias Ragoczy, Agnes Telling, Ido Amit, Bryan R Lajoie, Peter J Sabo, Michael O Dorschner, Richard Sandstrom, Bradley Bernstein, M A Bender, Mark Groudine, Andreas Gnirke, John Stamatoyannopoulos, Leonid A Mirny, Eric S Lander, and Job Dekker. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science (New York, N.Y.)*, 326(5950):289–93, 2009.

[168] Yin C Lin, Christopher Benner, Robert Mansson, Sven Heinz, Kazuko Miyazaki, Masaki Miyazaki, Vivek Chandra, Claudia Bossen, Christopher K Glass, and Cornelis Murre. Global changes in the nuclear positioning of genes and intra- and interdomain genomic interactions that orchestrate B cell fate. *Nature Immunology*, 13(12):1196–1204, 2012.

[169] Peter F.R. Little. Structure and function of the human genome. *Genome Research*, 15(12):1759–1766, 2005.

[170] Dagan A Loisel, Matthew V Rockman, Gregory A Wray, Jeanne Altmann, and Susan C Alberts. Ancient polymorphism and functional variation in the primate MHC-DQA1 5' cis-regulatory region. *Proceedings of the National Academy of Sciences of the United States of America*, 103(44):16331–6, 2006.

[171] Aaron T Lun and Gordon K Smyth. diffHic: a Bioconductor package to detect differential genomic interactions in Hi-C data. *BMC bioinformatics*, 16:258, 2015.

[172] Zhengyu Luo, Xiaorong Wang, Hong Jiang, Ruoyu Wang, Jian Chen, Yusheng Chen, Qianlan Xu, Jun Cao, Xiaowen Gong, Ji Wu, Yungui Yang, Wenbo Li, Chunsheng Han, C. Yan Cheng, Michael G. Rosenfeld, Fei Sun, and Xiaoyuan Song. Reorganized 3D genome structures support transcriptional regulation in mouse spermatogenesis. *iScience*, 23(4):101034, 2020.

[173] Darío G Lupiáñez, Katerina Kraft, Verena Heinrich, Peter Krawitz, Francesco Brancati, Eva Klopocki, Denise Horn, Hülya Kayserili, John M Opitz, Renata Laxova, Fernando Santos-Simarro, Brigitte Gilbert-Dussardier, Lars Wittler, Marina Borschiwer, Stefan A Haas, Marco Osterwalder, Martin Franke, Bernd Timmermann, Jochen Hecht, Malte Spielmann, Axel Visel, and Stefan Mundlos. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell*, 161(5):1012–25, 2015.

[174] Darío G. Lupiáñez, Malte Spielmann, and Stefan Mundlos. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends in Genetics*, 32(4):225–237, 2016.

[175] Hongqiang Lyu, Erhu Liu, and Zhifang Wu. Comparison of normalization methods for Hi-C data. *BioTechniques*, 68(2):56–64, 2020.

[176] Xiaowen Lyu, M. Jordan Rowley, and Victor G. Corces. Architectural Proteins and Pluripotency Factors Cooperate to Orchestrate the Transcriptional Response of hESCs to Temperature Stress. *Molecular Cell*, 71(6):940–955.e7, 2018.

[177] Xiaoyan Ma, Daphne Ezer, Boris Adryan, and Tim J. Stevens. Canonical and single-cell Hi-C reveal distinct chromatin interaction sub-networks of mammalian transcription factors. *Genome Biology*, 19(1):174, 2018.

[178] David P. MacKinnon, Chondra M. Lockwood, Jeanne M. Hoffman, Stephen G. West, and Virgil Sheets. A Comparison of Methods to Test Mediation and Other Intervening Variable Effects. *Psychological Methods*, 7(1):83–104, 2002.

[179] David P. MacKinnon, Chondra M. Lockwood, and Jason Williams. Confidence Limits for the Indirect Effect: Distribution of the Product and Resampling Methods. *Multivariate Behavioral Research*, 39(1):99–128, 2004.

[180] Lesley T. MacNeil and Albertha J.M. Walhout. Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research*, 21(5):645–657, 2011.

[181] Jacek Majewski and Tomi Pastinen. The study of eQTL variations by RNA-seq: from SNPs to phenotypes. *Trends in Genetics*, 27(2):72–79, 2011.

[182] Kirk J Mantione, Richard M. Kream, Hana Kuzelova, Radek Ptacek, Jiri Raboch, Joshua M. Samuel, and George B. Stefano. Comparing Bioinformatic Gene Expression Profiling Methods: Microarray and RNA-Seq. *Medical Science Monitor Basic Research*, 20:138–141, 2014.

[183] Maria C. N. Marchetto, Iñigo Narvaiza, Ahmet M. Denli, Christopher Benner, Thomas A. Lazzarini, Jason L. Nathanson, Apuã C. M. Paquola, Keval N. Desai, Roberto H. Herai, Matthew D. Weitzman, Gene W. Yeo, Alysson R. Muotri, and Fred H. Gage. Differential L1 regulation in pluripotent stem cells of humans and apes. *Nature*, 503(7477):525–529, 2013.

[184] John C. Marioni, Christopher E. Mason, Shrikant M. Mane, Matthew Stephens, and Yoav Gilad. RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research*, 18(9):1509–1517, 2008.

[185] Marc A. Marti-Renom, Genevieve Almouzni, Wendy A. Bickmore, Kerstin Bystricky, Giacomo Cavalli, Peter Fraser, Susan M. Gasser, Luca Giorgetti, Edith Heard, Mario Nicodemi, Marcelo Nollmann, Modesto Orozco, Ana Pombo, and Maria-Elena Torres-Padilla. Challenges and guidelines toward 4D nucleome data and model standards. *Nature Genetics*, 50(10):1352–1358, 2018.

[186] Paul Martin, Amanda McGovern, Jonathan Massey, Stefan Schoenfelder, Kate Duffus, Annie Yarwood, Anne Barton, Jane Worthington, Peter Fraser, Stephen Eyre, and Gisela Orozco. Identifying Causal Genes at the Multiple Sclerosis Associated Region 6q23 Using Capture Hi-C. *PLOS ONE*, 11(11):e0166923, 2016.

[187] Paul Martin, Amanda McGovern, Gisela Orozco, Kate Duffus, Annie Yarwood, Stefan Schoenfelder, Nicholas J. Cooper, Anne Barton, Chris Wallace, Peter Fraser, Jane Worthington, and Steve Eyre. Capture Hi-C reveals novel candidate genes and complex long-range interactions with related autoimmune risk loci. *Nature Communications*, 6(1):10069, 2015.

[188] Nana Matoba, Ivana Y Quiroga, Douglas H Phanstiel, and Hyejung Won. Mapping Alzheimer's Disease Variants to Their Target Genes Using Computational Analysis of Chromatin Configuration. *Journal of Visualized Experiments*, (155), 2020.

[189] Whitlock MC. Combining probability from independent tests: the weighted Z-method is superior to Fisher's approach. *J Evol Biol*, 2005.

[190] Karen J. Meaburn and Tom Misteli. Chromosome territories. *Nature*, 445(7126):379–381, 2007.

[191] Tom Misteli. Beyond the Sequence: Cellular Organization of Genome Function. *Cell*, 128(4):787–800, 2007.

[192] Hisashi Miura, Rawin Poonperm, Saori Takahashi, and Ichiro Hiratani. X-Chromosome Inactivation, Methods and Protocols. *Methods in molecular biology (Clifton, N.J.)*, 1861:221–245, 2018.

[193] Lindsey E Montefiori, Debora R Sobreira, Noboru J Sakabe, Ivy Aneas, Amelia C Joslin, Grace T Hansen, Grazyna Bozek, Ivan P Moskowitz, Elizabeth M McNally, and Marcelo A Nóbrega. A promoter interaction map for cardiovascular disease genetics. *eLife*, 7:e35788, 2018.

[194] Laia Mora, Inma Sánchez, Montserrat Garcia, and Montserrat Ponsà. Chromosome territory positioning of conserved homologous chromosomes in different primate species. *Chromosoma*, 115(5):367–375, 2006.

[195] Felipe Mora-Bermúdez, Farhath Badsha, Sabina Kanton, J Gray Camp, Benjamin Vernot, Kathrin Köhler, Birger Voigt, Keisuke Okita, Tomislav Maricic, Zhisong He, Robert Lachmann, Svante Pääbo, Barbara Treutlein, and Wieland B Huttner. Differences and similarities between human and chimpanzee neural progenitors during cerebral cortex development. *eLife*, 5:e18683, 2016.

[196] Stefanie L. Morgan, Natasha C. Mariano, Abel Bermudez, Nicole L. Arruda, Fangting Wu, Yunhai Luo, Gautam Shankar, Lin Jia, Huiling Chen, Ji-Fan Hu, Andrew R. Hoffman, Chiao-Chain Huang, Sharon J. Pitteri, and Kevin C. Wang. Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nature Communications*, 8(1):15993, 2017.

[197] FJ Müller, BM Schuldt, R Williams, D Mason, G Altun, E Papapetrou, S Danner, J Goldmann, A Herbst, N Schmidt, J Aldenhoff, L Laurent, and Loring J. A bioinformatic assay for pluripotency in human cells. *Nature Methods*, 2011.

[198] Maxwell R Mumbach, Adam J Rubin, Ryan A Flynn, Chao Dai, Paul A Khavari, William J Greenleaf, and Howard Y Chang. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nature Methods*, 13(11):919–922, 2016.

[199] Maxwell R Mumbach, Ansuman T Satpathy, Evan A Boyle, Chao Dai, Benjamin G Gowen, Seung Woo Cho, Michelle L Nguyen, Adam J Rubin, Jeffrey M Granja, Katelynn R Kazane, Yuning Wei, Trieu Nguyen, Peyton G Greenside, M Ryan Corces, Josh Tycko, Dimitre R Simeonov, Nabeela Suliman, Rui Li, Jin Xu, Ryan A Flynn, Anshul Kundaje, Paul A Khavari, Alexander Marson, Jacob E Corn, Thomas Quertermous, William J Greenleaf, and Howard Y Chang. Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nature Genetics*, 49(11):1602–1612, 2017.

[200] Luis Augusto Eijy Nagai, Sung-Joon Park, and Kenta Nakai. Analyzing the 3D chromatin organization coordinating with gene expression regulation in B-cell lymphoma. *BMC Medical Genomics*, 11(Suppl 7):127, 2019.

[201] Takashi Nagano, Yaniv Lubling, Tim J. Stevens, Stefan Schoenfelder, Eitan Yaffe, Wendy Dean, Ernest D. Laue, Amos Tanay, and Peter Fraser. Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature*, 502(7469):59–64, 2013.

[202] Takashi Nagano, Yaniv Lubling, Eitan Yaffe, Steven W Wingett, Wendy Dean, Amos Tanay, and Peter Fraser. Single-cell Hi-C for genome-wide detection of chromatin interactions that occur simultaneously in a single cell. *Nature Protocols*, 10(12):1986–2003, 2015.

[203] Takashi Nagano, Csilla Várnai, Stefan Schoenfelder, Biola-Maria Javierre, Steven W. Wingett, and Peter Fraser. Comparison of Hi-C results using in-solution versus in-nucleus ligation. *Genome Biology*, 16(1):175, 2015.

[204] M Nakagawa, Y Taniguchi, S Senda, N Takizawa, T Ichisaka, K Asano, A Morizane, D Doi, J Takahashi, M Nishizawa, Y Yoshida, T Toyoda, K Osafune, K Sekiguchi, and Yamanaka S. A novel efficient feeder-free culture system for the derivation of human induced pluripotent stem cells. *Scientific Reports*, 2014.

[205] Natalia Naumova and Job Dekker. Integrating one-dimensional and three-dimensional maps of genomes. *Journal of Cell Science*, 123(12):1979–1988, 2010.

[206] Elizabeth Nickerson and David L. Nelson. Molecular Definition of Pericentric Inversion Breakpoints Occurring during the Evolution of Humans and Chimpanzees. *Genomics*, 50(3):368–372, 1998.

[207] Henri Niskanen, Irina Tuszynska, Rafal Zaborowski, Merja Heinäniemi, Seppo Ylä-Herttuala, Bartek Wilczynski, and Minna U Kaikkonen. Endothelial cell differentiation is encompassed by changes in long range interactions between inactive chromatin regions. *Nucleic acids research*, 46(4):1724–1740, 2018.

[208] Marcelo A. Nobrega and Len A. Pennacchio. Comparative genomic analysis as a tool for biological discovery. *The Journal of Physiology*, 554(1):31–39, 2004.

[209] Elphège P P Nora, Bryan R Lajoie, Edda G Schulz, Luca Giorgetti, Ikuhiro Okamoto, Nicolas Servant, Tristan Piolot, Nynke L van Berkum, Johannes Meisig, John Sedat, Joost Gribnau, Emmanuel Barillot, Nils Blüthgen, Job Dekker, and Edith Heard. Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature*, 485(7398):381–5, 2012.

[210] Katja Nowick, Tim Gernat, Eivind Almaas, and Lisa Stubbs. Differences in human and chimpanzee gene expression patterns define an evolving network of transcription factors in brain. *Proceedings of the National Academy of Sciences*, 106(52):22358–22363, 2009.

[211] Johannes Nuebler, Geoffrey Fudenberg, Maxim Imakaev, Nezar Abdennur, and Leonid A. Mirny. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proceedings of the National Academy of Sciences*, 115(29):201717730, 2018.

[212] M Nuriddinov and V Fishman. C-InterSecture—a computational tool for interspecies comparison of genome architecture. *Bioinformatics*, 35(23):4912–4921, 2019.

[213] Chin-Tong T Ong and Victor G Corces. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nature Reviews Genetics*, 12(4):283–93, 2011.

[214] Athma A Pai, Jordana T Bell, John C Marioni, Jonathan K Pritchard, and Yoav Gilad. A genome-wide study of DNA methylation patterns and gene expression levels in multiple human and chimpanzee tissues. *PLoS genetics*, 7(2):e1001316, 2011.

[215] Koustav Pal, Mattia Forcato, and Francesco Ferrari. Hi-C analysis: from data generation to integration. *Biophysical Reviews*, 11(1):67–78, 2019.

[216] Jonas Paulsen, Geir Kjetil Sandve, Sveinung Gundersen, Tonje G. Lien, Kai Trengereid, and Eivind Hovig. HiBrowse: multi-purpose statistical analysis of genome-wide chromatin 3D organization. *Bioinformatics*, 30(11):1620–1622, 2014.

[217] Bryan J. Pavlovic, Lauren E. Blake, Julien Roux, Claudia Chavarria, and Yoav Gilad. A Comparative Assessment of Human and Chimpanzee iPSC-derived Cardiomyocytes with Primary Heart Tissues. *Scientific Reports*, 8(1):15312, 2018.

[218] Len A Pennacchio, Wendy Bickmore, Ann Dean, Marcelo A Nobrega, and Gill Bejerano. Enhancers: five essential questions. *Nature Reviews Genetics*, 14(4):288–295, 2013.

[219] George H. Perry, Páll Melsted, John C. Marioni, Ying Wang, Russell Bainer, Joseph K. Pickrell, Katelyn Michelini, Sarah Zehr, Anne D. Yoder, Matthew Stephens,

Jonathan K. Pritchard, and Yoav Gilad. Comparative RNA sequencing reveals substantial genetic variation in endangered primates. *Genome Research*, 22(4):602–610, 2012.

[220] Jennifer E Phillips-Cremins, Michael E Sauria, Amartya Sanyal, Tatiana I Gerasimova, Bryan R Lajoie, Joshua S Bell, Chin-Tong T Ong, Tracy A Hookway, Changying Guo, Yuhua Sun, Michael J Bland, William Wagstaff, Stephen Dalton, Todd C McDevitt, Ranjan Sen, Job Dekker, James Taylor, and Victor G Corces. Architectural protein subclasses shape 3D organization of genomes during lineage commitment. *Cell*, 153(6):1281–95, 2013.

[221] Di M Pierro, RR Cheng, EL Aiden, PG Wolynes, and Onuchic JN. De novo prediction of human chromosome structures: Epigenetic marking patterns encode genome architecture. *Proceedings of the National Academy of Sciences*, 2017.

[222] Katherine S Pollard, Sofie R Salama, Nelle Lambert, Marie-Alexandra A Lambot, Sandra Coppens, Jakob S Pedersen, Sol Katzman, Bryan King, Courtney Onodera, Adam Siepel, Andrew D Kern, Colette Dehay, Haller Igel, Manuel Ares, Pierre Vanderhaeghen, and David Haussler. An RNA gene expressed during cortical development evolved rapidly in humans. *Nature*, 443(7108):167–72, 2006.

[223] Benjamin D Pope, Tyrone Ryba, Vishnu Dileep, Feng Yue, Weisheng Wu, Olgert Denas, Daniel L Vera, Yanli Wang, R S Hansen, Theresa K Canfield, Robert E Thurman, Yong Cheng, Günhan Gülsoy, Jonathan H Dennis, Michael P Snyder, John A Stamatoyannopoulos, James Taylor, Ross C Hardison, Tamer Kahveci, Bing Ren, and David M Gilbert. Topologically associating domains are stable units of replication-timing regulation. *Nature*, 515(7527):402–5, 2014.

[224] S Prabhakar, A Visel, JA Akiyama, M Shoukry, K Lewis, A Holt, I Plazjer-Frick, H Morrison, D FitzPatrick, V Afzal, L Pennacchio, E Rubin, and J Noonan. Human-specific gain of function in a developmental enhancer. *Science*, 2008.

[225] Javier Prado-Martinez, Peter H. Sudmant, Jeffrey M. Kidd, Heng Li, Joanna L. Kelley, Belen Lorente-Galdos, Krishna R. Veeramah, August E. Woerner, Timothy D. O'Connor, Gabriel Santpere, Alexander Cagan, Christoph Theunert, Ferran Casals, Hafid Laayouni, Kasper Munch, Asger Hobolth, Anders E. Halager, Maika Malig, Jessica Hernandez-Rodriguez, Irene Hernando-Herraez, Kay Prüfer, Marc Pybus, Laurel Johnstone, Michael Lachmann, Can Alkan, Dorina Twigg, Natalia Petit, Carl Baker, Fereydoun Hormozdiari, Marcos Fernandez-Callejo, Marc Dabad, Michael L. Wilson, Laurie Stevison, Cristina Camprubí, Tiago Carvalho, Aurora Ruiz-Herrera, Laura Vives, Marta Mele, Teresa Abello, Ivanela Kondova, Ronald E. Bontrop, Anne Pusey, Felix Lankester, John A. Kiyang, Richard A. Bergl, Elizabeth Lonsdorf, Simon Myers, Mario Ventura, Pascal Gagneux, David Comas, Hans Siegismund, Julie Blanc, Lidia Agueda-Calpena, Marta Gut, Lucinda Fulton, Sarah A. Tishkoff, James C. Mullikin, Richard K. Wilson, Ivo G. Gut, Mary Katherine Gonder, Oliver A. Ryder, Beatrice H.

Hahn, Arcadi Navarro, Joshua M. Akey, Jaume Bertranpetit, David Reich, Thomas Mailund, Mikkel H. Schierup, Christina Hvilsom, Aida M. Andrés, Jeffrey D. Wall, Carlos D. Bustamante, Michael F. Hammer, Evan E. Eichler, and Tomas Marques-Bonet. Great ape genetic diversity and population history. *Nature*, 499(7459):471–475, 2013.

[226] Kristopher J. Preacher and Andrew F. Hayes. SPSS and SAS procedures for estimating indirect effects in simple mediation models. *Behavior Research Methods, Instruments, & Computers*, 36(4):717–731, 2004.

[227] Kristopher J. Preacher and James P. Selig. Advantages of Monte Carlo Confidence Intervals for Indirect Effects. *Communication Methods and Measures*, 6(2):77–98, 2012.

[228] Sara L. Prescott, Rajini Srinivasan, Maria Carolina Marchetto, Irina Grishina, Iñigo Narvaiza, Licia Selleri, Fred H. Gage, Tomek Swigut, and Joanna Wysocka. Enhancer Divergence and cis-Regulatory Evolution in the Human and Chimp Neural Crest. *Cell*, 163(1):68–83, 2015.

[229] AR Quinlan and Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*, 2010.

[230] Vijay Ramani, Xinxian Deng, Ruolan Qiu, Kevin L Gunderson, Frank J Steemers, Christine M Disteche, William S Noble, Zhijun Duan, and Jay Shendure. Massively multiplex single-cell Hi-C. *Nature methods*, 14(3):263–266, 2017.

[231] Vijay Ramani, Xinxian Deng, Ruolan Qiu, Choli Lee, Christine M Disteche, William S Noble, Jay Shendure, and Zhijun Duan. Sci-Hi-C: a single-cell Hi-C method for mapping 3D genome organization in large number of single cells. *Methods*, 170:61–68, 2019.

[232] Fidel Ramírez, Vivek Bhardwaj, Laura Arrigoni, Kin Lam, Björn A Grüning, José Villaveces, Bianca Habermann, Asifa Akhtar, and Thomas Manke. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nature communications*, 9(1):189, 2018.

[233] Suhas Rao, Miriam H Huntley, Neva C Durand, Elena K Stamenova, Ivan D Bochkov, James T Robinson, Adrian L Sanborn, Ido Machol, Arina D Omer, Eric S Lander, and Erez Aiden. A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell*, 159(7):1665–1680, 2014.

[234] Suhas S.P. Rao, Su-Chen Huang, Brian Glenn St Hilaire, Jesse M. Engreitz, Elizabeth M. Perez, Kyong-Rim Kieffer-Kwon, Adrian L. Sanborn, Sarah E. Johnstone, Gavin D. Bascom, Ivan D. Bochkov, Xingfan Huang, Muhammad S. Shamim, Jaeweon Shin, Douglass Turner, Ziyi Ye, Arina D. Omer, James T. Robinson, Tamar Schlick, Bradley E. Bernstein, Rafael Casellas, Eric S. Lander, and Erez Lieberman Aiden. Cohesin Loss Eliminates All Loop Domains. *Cell*, 171(2):305–320.e24, 2017.

[235] Gina Renschler, Gautier Richard, Claudia Isabelle Keller Valsecchi, Sarah Toscano, Laura Arrigoni, Fidel Ramírez, and Asifa Akhtar. Hi-C guided assemblies reveal conserved regulatory topologies on X and autosomes despite extensive genome shuffling. *Genes & development*, 33(21-22):1591–1612, 2019.

[236] Dietmar Rieder, Zlatko Trajanoski, and James G. McNally. Transcription factories. *Frontiers in Genetics*, 3:221, 2012.

[237] Matthew V Rockman, Matthew W Hahn, Nicole Soranzo, Fritz Zimprich, David B Goldstein, and Gregory A Wray. Ancient and recent positive selection transformed opioid cis-regulation in humans. *PLoS biology*, 3(12):e387, 2005.

[238] Jeffrey Rogers and Richard A. Gibbs. Comparative primate genomics: emerging patterns of genome content and dynamics. *Nature Reviews Genetics*, 15(5):347–359, 2014.

[239] Irene G Romero, Ilya Ruvinsky, and Yoav Gilad. Comparative studies of gene expression and the evolution of gene regulation. *Nature Reviews Genetics*, 13(7):505–16, 2012.

[240] Irene Gallego Romero, Bryan J Pavlovic, Irene Hernando-Herraez, Xiang Zhou, Michelle C Ward, Nicholas E Banovich, Courtney L Kagan, Jonathan E Burnett, Constance H Huang, Amy Mitrano, Claudia I Chavarria, Inbar Friedrich Ben-Nun, Yingchun Li, Karen Sabatini, Trevor R Leonardo, Mana Parast, Tomas Marques-Bonet, Louise C Laurent, Jeanne F Loring, and Yoav Gilad. A panel of induced pluripotent stem cells from chimpanzees: a resource for comparative functional genomics. *eLife*, 4:e07103, 2015.

[241] G Ron, Y Globerson, D Moran, and Kaplan T. Promoter-enhancer interactions identified from Hi-C data using probabilistic models and hierarchical topological domains. *Nature Communications*, 2017.

[242] M J Rowley, Michael H Nichols, Xiaowen Lyu, Masami Ando-Kuri, I Sarahi M SM Rivera, Karen Hermetz, Ping Wang, Yijun Ruan, and Victor G Corces. Evolutionarily Conserved Principles Predict 3D Chromatin Organization. *Molecular cell*, 67(5):837–852.e7, 2017.

[243] M. Jordan Rowley and Victor G. Corces. Organizational principles of 3D genome architecture. *Nature Reviews Genetics*, page 1, 2018.

[244] Sushmita Roy, Alireza Fotuhi Siahpirani, Deborah Chasman, Sara Knaack, Ferhat Ay, Ron Stewart, Michael Wilson, and Rupa Sridharan. A predictive modeling approach for cell line-specific long-range regulatory interactions. *Nucleic Acids Research*, 43(18):8694–8712, 2015.

[245] Matteo Vietri Rudan, Christopher Barrington, Stephen Henderson, Christina Ernst, Duncan T Odom, Amos Tanay, and Suzana Hadjur. Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture. *Cell reports*, 10(8):1297–309, 2015.

[246] Adrian L. Sanborn, Suhas S. P. Rao, Su-Chen Huang, Neva C. Durand, Miriam H. Huntley, Andrew I. Jewett, Ivan D. Bochkov, Dharmaraj Chinnappan, Ashok Cutkosky, Jian Li, Kristopher P. Geeting, Andreas Gnirke, Alexandre Melnikov, Doug McKenna, Elena K. Stamenova, Eric S. Lander, and Erez Lieberman Aiden. 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):E6456–E6465, 2015.

[247] Natalie Sauerwald and Carl Kingsford. Quantifying the similarity of topological domains across normal and cancer human cell types. *Bioinformatics*, 34(13):i475–i483, 2018.

[248] Natalie Sauerwald, Akshat Singhal, and Carl Kingsford. Analysis of the structural variability of topologically associated domains as revealed by Hi-C. *NAR Genomics and Bioinformatics*, 2(1), 2019.

[249] Michael EG Sauria, Jennifer E. Phillips-Cremins, Victor G. Corces, and James Taylor. HiFive: a tool suite for easy and efficient HiC and 5C data analysis. *Genome Biology*, 16(1):237, 2015.

[250] A Scally, JY Dutheil, LDW Hillier, and Jordan GE et al. Insights into hominid evolution from the gorilla genome sequence. *Nature*, 2012.

[251] Anthony D. Schmitt, Ming Hu, Inkyung Jung, Zheng Xu, Yunjiang Qiu, Catherine L. Tan, Yun Li, Shin Lin, Yiing Lin, Cathy L. Barr, and Bing Ren. A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome. *Cell Reports*, 17(8):2042–2059, 2016.

[252] S Schoenfelder, T Sexton, L Chakalova, NF Cope, A Horton, S Andrews, S Kurukuti, J Mitchell, D Umlauf, D Dimitrova, C Eskiw, Y Luo, C Wei, Y Ruan, J Biker, and Fraser P. Preferential associations between co-regulated genes reveal a transcriptional interactome in erythroid cells. *Nature Genetics*, 2010.

[253] Stefan Schoenfelder and Peter Fraser. Long-range enhancer-promoter contacts in gene expression control. *Nature Reviews Genetics*, 2019.

[254] Emre Sefer, Geet Duggal, and Carl Kingsford. Deconvolution of Ensemble Chromatin Interaction Data Reveals the Latent Mixing Structures in Cell Subpopulations. *Journal of Computational Biology*, 23(6):425–438, 2016.

[255] Nicolas Servant, Nelle Varoquaux, Bryan R Lajoie, Eric Viara, Chong-Jian J Chen, Jean-Philippe P Vert, Edith Heard, Job Dekker, and Emmanuel Barillot. HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome biology*, 16:259, 2015.

[256] T Sexton, H Schober, P Fraser, and Gasser SM. Gene regulation through nuclear organization. *Nature Structural and Molecular Biology*, 2007.

[257] Tom Sexton and Giacomo Cavalli. The Role of Chromosome Domains in Shaping the Functional Genome. *Cell*, 160(6):1049–1059, 2015.

[258] Tom Sexton, Eitan Yaffe, Ephraim Kenigsberg, Frédéric Bantignies, Benjamin Leblanc, Michael Hoichman, Hugues Parrinello, Amos Tanay, and Giacomo Cavalli. Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell*, 148(3):458–472, 2012.

[259] Y Shen, F Yue, DF McCleary, Z Ye, L Edsall, S Kuan, U Wagner, J Dixon, L Lee, VV Lobanenkov, and Ren B. A map of the cis-regulatory sequences in the mouse genome. *Nature*, 2012.

[260] Yoichiro Shibata, Nathan C. Sheffield, Olivier Fedrigo, Courtney C. Babbitt, Matthew Wortham, Alok K. Tewari, Darin London, Lingyun Song, Bum-Kyu Lee, Vishwanath R. Iyer, Stephen C. J. Parker, Elliott H. Margulies, Gregory A. Wray, Terrence S. Furey, and Gregory E. Crawford. Extensive Evolutionary Changes in Regulatory Element Activity during Human Origins Are Associated with Altered Gene Expression and Positive Selection. *PLoS Genetics*, 8(6):e1002789, 2012.

[261] Hanjun Shin, Yi Shi, Chao Dai, Harianto Tjong, Ke Gong, Frank Alber, and Xianghong J Zhou. TopDom: an efficient and deterministic method for identifying topological domains in genomes. *Nucleic acids research*, 44(7):e70, 2016.

[262] Hennady P Shulha, Jessica L Crisci, Denis Reshetov, Jogender S Tushir, Iris Cheung, Rahul Bharadwaj, Hsin-Jung J Chou, Isaac B Houston, Cyril J Peter, Amanda C Mitchell, Wei-Dong D Yao, Richard H Myers, Jiang-Fan F Chen, Todd M Preuss, Evgeny I Rogaev, Jeffrey D Jensen, Zhiping Weng, and Schahram Akbarian. Human-specific histone methylation signatures at transcription start sites in prefrontal neurons. *PLoS biology*, 10(11):e1001427, 2012.

[263] Rasmus Siersbæk, Jesper Grud Skat GS Madsen, Biola M Javierre, Ronni Nielsen, Emilie K Bagge, Jonathan Cairns, Steven W Wingett, Sofie Traynor, Mikhail Spivakov, Peter Fraser, and Susanne Mandrup. Dynamic Rewiring of Promoter-Anchored Chromatin Loops during Adipocyte Differentiation. *Molecular cell*, 66(3):420–435.e5, 2017.

[264] Natalia Sikorska and Tom Sexton. Defining functionally relevant spatial chromatin domains: it's a TAD complicated. *Journal of molecular biology*, 432(3):653–664, 2019.

[265] Scott Smemo, Juan J. Tena, Kyoung-Han Kim, Eric R. Gamazon, Noboru J. Sakabe, Carlos Gómez-Marín, Ivy Aneas, Flavia L. Credidio, Débora R. Sobreira, Nora F. Wasserman, Ju Hee Lee, Vijitha Puviindran, Davis Tam, Michael Shen, Joe Eun Son, Niki Alizadeh Vakili, Hoon-Ki Sung, Silvia Naranjo, Rafael D. Acemel, Miguel Manzanares, Andras Nagy, Nancy J. Cox, Chi-Chung Hui, Jose Luis Gomez-Skarmeta, and Marcelo A. Nóbrega. Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature*, 507(7492):371–375, 2014.

[266] Emily M Smith, Bryan R Lajoie, Gaurav Jain, and Job Dekker. Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus. *American journal of human genetics*, 98(1):185–201, 2016.

[267] Gordon K Smyth. Linear Models and Empirical Bayes Methods for Assessing Differential Expression in Microarray Experiments. *Statistical Applications in Genetics and Molecular Biology*, 3(1):1–25, 2004.

[268] Michael E Sobel. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, 13:290, 1982.

[269] Michael E Sobel. Some New Results on Indirect Effects and Their Standard Errors in Covariance Structure Models. *Sociological Methodology*, 16:159, 1986.

[270] Sevil Sofueva, Eitan Yaffe, Wen-Ching Chan, Dimitra Georgopoulou, Matteo Vietri Rudan, Hegias Mira-Bontenbal, Steven M Pollard, Gary P Schroth, Amos Tanay, and Suzana Hadjur. Cohesin-mediated interactions organize chromosomal domain architecture. *The EMBO journal*, 32(24):3119–29, 2013.

[271] Mehmet Somel, Henriette Franz, Zheng Yan, Anna Lorenc, Song Guo, Thomas Giger, Janet Kelso, Birgit Nickel, Michael Dannemann, Sabine Bahn, Maree J. Webster, Cynthia S. Weickert, Michael Lachmann, Svante Pääbo, and Philipp Khaitovich. Transcriptional neoteny in the human brain. *Proceedings of the National Academy of Sciences*, 106(14):5743–5748, 2009.

[272] Mehmet Somel, Xiling Liu, Lin Tang, Zheng Yan, Haiyang Hu, Song Guo, Xi Jiang, Xiaoyu Zhang, Guohua Xu, Gangcai Xie, Na Li, Yuhui Hu, Wei Chen, Svante Pääbo, and Philipp Khaitovich. MicroRNA-Driven Developmental Remodeling in the Brain Distinguishes Humans from Other Primates. *PLoS Biology*, 9(12):e1001214, 2011.

[273] Lingyun Song, Zhancheng Zhang, Linda L. Grasfeder, Alan P. Boyle, Paul G. Giresi, Bum-Kyu Lee, Nathan C. Sheffield, Stefan Gräf, Mikael Huss, Damian Keefe, Zheng Liu, Darin London, Ryan M. McDaniell, Yoichiro Shibata, Kimberly A. Showers, Jeremy M. Simon, Teresa Vales, Tianyuan Wang, Deborah Winter, Zhuzhu Zhang, Neil D. Clarke, Ewan Birney, Vishwanath R. Iyer, Gregory E. Crawford, Jason D. Lieb, and Terrence S. Furey. Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Research*, 21(10):1757–1767, 2011.

[274] André M. M. Sousa, Ying Zhu, Mary Ann Raghanti, Robert R. Kitchen, Marco Onorati, Andrew T. N. Tebbenkamp, Bernardo Stutz, Kyle A. Meyer, Mingfeng Li, Yuka Imamura Kawasawa, Fuchen Liu, Raquel Garcia Perez, Marta Mele, Tiago Carvalho, Mario Skarica, Forrest O. Gulden, Mihovil Pletikos, Akemi Shibata, Alexa R. Stephenson, Melissa K. Edler, John J. Ely, John D. Elsworth, Tamas L. Horvath, Patrick R. Hof, Thomas M. Hyde, Joel E. Kleinman, Daniel R. Weinberger, Mark

Reimers, Richard P. Lifton, Shrikant M. Mane, James P. Noonan, Matthew W. State, Ed S. Lein, James A. Knowles, Tomas Marques-Bonet, Chet C. Sherwood, Mark B. Gerstein, and Nenad Sestan. Molecular and cellular reorganization of neural circuits in the human lineage. *Science*, 358(6366):1027–1032, 2017.

[275] Mikhail Spivakov and Amanda G. Fisher. Epigenetic signatures of stem-cell identity. *Nature Reviews Genetics*, 8(4):263–271, 2007.

[276] John C. Stansfield, Kellen G. Cresswell, Vladimir I. Vladimirov, and Mikhail G. Dozmorov. HiCcompare: an R-package for joint normalization and comparison of HI-C datasets. *BMC Bioinformatics*, 19(1):279, 2018.

[277] David L. Stern and Virginie Orgogozo. The Loci of Evolution: How Predictable is Genetic Evolution. *Evolution*, 62(9):2155–2177, 2008.

[278] Peter H. Sudmant, Maria S. Alexis, and Christopher B. Burge. Meta-analysis of RNA-seq expression data across species, tissues and studies. *Genome Biology*, 16(1):287, 2015.

[279] Ning Sun, Nicholas J. Panetta, Deepak M. Gupta, Kitchener D. Wilson, Andrew Lee, Fangjun Jia, Shijun Hu, Athena M. Cherry, Robert C. Robbins, Michael T. Longaker, and Joseph C. Wu. Feeder-free derivation of induced pluripotent stem cells from adult human adipose stem cells. *Proceedings of the National Academy of Sciences*, 106(37):15720–15725, 2009.

[280] Devjanee Swain-Lenz, Alejandro Berrio, Alexias Safi, Gregory E Crawford, and Gregory A Wray. Comparative analyses of chromatin landscape in white adipose tissue suggest humans may have less beigeing potential than other primates. *Genome Biology and Evolution*, 11(7):1997–2008, 2019.

[281] Orsolya Symmons, Leslie Pan, Silvia Remeseiro, Tugce Aktas, Felix Klein, Wolfgang Huber, and François Spitz. The Shh Topological Domain Facilitates the Action of Remote Enhancers by Reducing the Effects of Genomic Distances. *Developmental cell*, 39(5):529–543, 2016.

[282] Orsolya Symmons, Veli Vural Uslu, Taro Tsujimura, Sandra Ruf, Sonya Nassari, Wibke Schwarzer, Laurence Ettwiller, and François Spitz. Functional and topological characteristics of mammalian regulatory domains. *Genome research*, 24(3):390–400, 2014.

[283] Quentin Szabo, Frédéric Bantignies, and Giacomo Cavalli. Principles of genome folding into topologically associating domains. *Science Advances*, 5(4):eaaw1668, 2019.

[284] Kazutoshi Takahashi, Koji Tanabe, Mari Ohnuki, Megumi Narita, Tomoko Ichisaka, Kiichiro Tomoda, and Shinya Yamanaka. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell*, 131(5):861–72, 2007.

[285] Kazutoshi Takahashi and Shinya Yamanaka. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell*, 126(4):663–76, 2006.

[286] Hideyuki Tanabe, Stefan Müller, Michaela Neusser, Johann von Hase, Enzo Calcagno, Marion Cremer, Irina Solovei, Christoph Cremer, and Thomas Cremer. Evolutionary conservation of chromosome territory arrangements in cell nuclei from higher primates. *Proceedings of the National Academy of Sciences*, 99(7):4424–4429, 2002.

[287] Amos Tanay and Giacomo Cavalli. Chromosomal domains: epigenetic contexts and functional implications of genomic compartmentalization. *Current opinion in genetics & development*, 23(2):197–203, 2013.

[288] Robert E. Thurman, Eric Rynes, Richard Humbert, Jeff Vierstra, Matthew T. Maurano, Eric Haugen, Nathan C. Sheffield, Andrew B. Stergachis, Hao Wang, Benjamin Vernot, Kavita Garg, Sam John, Richard Sandstrom, Daniel Bates, Lisa Boatman, Theresa K. Canfield, Morgan Diegel, Douglas Dunn, Abigail K. Ebersol, Tristan Frum, Erika Giste, Audra K. Johnson, Ericka M. Johnson, Tanya Kutyavin, Bryan Lajoie, Bum-Kyu Lee, Kristen Lee, Darin London, Dimitra Lotakis, Shane Neph, Fidencio Neri, Eric D. Nguyen, Hongzhu Qu, Alex P. Reynolds, Vaughn Roach, Alexias Safi, Minerva E. Sanchez, Amartya Sanyal, Anthony Shafer, Jeremy M. Simon, Lingyun Song, Shinny Vong, Molly Weaver, Yongqi Yan, Zhancheng Zhang, Zhuzhu Zhang, Boris Lenhard, Muneesh Tewari, Michael O. Dorschner, R. Scott Hansen, Patrick A. Navas, George Stamatoyannopoulos, Vishwanath R. Iyer, Jason D. Lieb, Shamil R. Sunyaev, Joshua M. Akey, Peter J. Sabo, Rajinder Kaul, Terrence S. Furey, Job Dekker, Gregory E. Crawford, and John A. Stamatoyannopoulos. The accessible chromatin landscape of the human genome. *Nature*, 489(7414):75–82, 2012.

[289] Itay Tirosh and Naama Barkai. Inferring regulatory mechanisms from patterns of evolutionary divergence. *Molecular systems biology*, 7:530, 2011.

[290] Harianto Tjong, Ke Gong, Lin Chen, and Frank Alber. Physical tethering and volume exclusion determine higher-order genome organization in budding yeast. *Genome Research*, 22(7):1295–1305, 2012.

[291] Marco Trizzino, YoSon Park, Marcia Holsbach-Beltrame, Katherine Aracena, Katelyn Mika, Minal Caliskan, George H. Perry, Vincent J. Lynch, and Christopher D. Brown. Transposable elements are the primary source of novelty in primate gene regulation. *Genome Research*, 27(10):1623–1633, 2017.

[292] Maria Tsompana and Michael J Buck. Chromatin accessibility: a window into the genome. *Epigenetics & Chromatin*, 7(1):33, 2014.

[293] Benjamin D. Umans, Alexis Battle, and Yoav Gilad. Where Are the Disease-Associated eQTLs? *Trends in Genetics*, 2020.

[294] Eric J. Vallender. Bioinformatic approaches to identifying orthologs and assessing evolutionary relationships. *Methods*, 49(1):50–55, 2009.

[295] N Varoquaux, F Ay, WS Noble, and Vert JP. A statistical approach for inferring the 3D structure of the genome. *Bioinformatics*, 2014.

[296] Diego Villar, Camille Berthelot, Sarah Aldridge, Tim F. Rayner, Margus Lukk, Miguel Pignatelli, Thomas J. Park, Robert Deaville, Jonathan T. Erichsen, Anna J. Jasinska, James M.A. Turner, Mads F. Bertelsen, Elizabeth P. Murchison, Paul Flicek, and Duncan T. Odom. Enhancer Evolution across 20 Mammalian Species. *Cell*, 160(3):554–566, 2015.

[297] Joseph J. Vitti, Sharon R. Grossman, and Pardis C. Sabeti. Detecting Natural Selection in Genomic Data. *Annual Review of Genetics*, 47(1):97–120, 2013.

[298] Sidney H. Wang, Chiaowen Joyce Hsiao, Zia Khan, and Jonathan K. Pritchard. Post-translational buffering leads to convergent protein expression levels between primates. *Genome Biology*, 19(1):83, 2018.

[299] Zhong Wang, Mark Gerstein, and Michael Snyder. RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1):57–63, 2009.

[300] N Ward and Moreno-Hagelsieb G. Quickly finding orthologs as reciprocal best hits with BLAT, LAST, and UBLAST: how much do we miss? *PLoS One*, 2014.

[301] Lisa R Warner, Courtney C Babbitt, Alex E Primus, Tonya F Severson, Ralph Haygood, and Gregory A Wray. Functional consequences of genetic variation in primates on tyrosine hydroxylase (TH) expression in vitro. *Brain research*, 1288:1–8, 2009.

[302] RH Waterson, ES Lander, RK Wilson, Chimpanzee Sequencing, and Analysis Consortium. Initial sequence of the chimpanzee genome and comparison with the human genome. *Nature*, 2005.

[303] Caleb Weinreb and Benjamin J. Raphael. Identification of hierarchical chromatin domains. *Bioinformatics*, 32(11):1601–1609, 2016.

[304] A. C. Wilson, L. R. Maxson, and V. M. Sarich. Two Types of Molecular Evolution. Evidence from Studies of Interspecific Hybridization. *Proceedings of the National Academy of Sciences*, 71(7):2843–2847, 1974.

[305] Michael D Wilson and Duncan T Odom. Evolution of transcriptional control in mammals. *Current opinion in genetics & development*, 19(6):579–85, 2009.

[306] Steven Wingett, Philip Ewels, Mayra Furlan-Magaril, Takashi Nagano, Stefan Schoenfelder, Peter Fraser, and Simon Andrews. HiCUP: pipeline for mapping and processing Hi-C data. *F1000Research*, 4:1310, 2015.

[307] Elzo de Wit. TADs as the caller calls them. *Journal of Molecular Biology*, 432(3):638–642, 2019.

[308] Joost M Woltering, Daan Noordermeer, Marion Leleu, and Denis Duboule. Conservation and divergence of regulatory strategies at Hox Loci and the origin of tetrapod digits. *PLoS biology*, 12(1):e1001773, 2014.

[309] Kyoung-Jae J Won, Iouri Chepelev, Bing Ren, and Wei Wang. Prediction of regulatory elements in mammalian genomes using chromatin signatures. *BMC bioinformatics*, 9:547, 2008.

[310] Yong H Woo and Wen-Hsiung Li. Evolutionary conservation of histone modifications in mammals. *Molecular biology and evolution*, 29(7):1757–67, 2012.

[311] Gregory A Wray. The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*, 8(3):206–16, 2007.

[312] Chong Wu and Wei Pan. Integration of Enhancer-Promoter Interactions with GWAS Summary Results Identifies Novel Schizophrenia-Associated Genes and Pathways. *Genetics*, 209(3):699–709, 2018.

[313] Ting Xie, Fu-Gui Zhang, Hong-Yu Zhang, Xiao-Tao Wang, Ji-Hong Hu, and Xiao-Ming Wu. Biased gene retention during diploidization in Brassica linked to three-dimensional genome organization. *Nature plants*, 5(8):822–832, 2019.

[314] Eitan Yaffe and Amos Tanay. Probabilistic modeling of Hi-C contact maps eliminates systematic biases to characterize global chromosomal architecture. *Nature Genetics*, 43(11):1059–65, 2011.

[315] Koon-Kiu Yan, Shaoke Lou, and Mark Gerstein. MrTADFinder: A network modularity based approach to identify topologically associating domains in multiple resolutions. *PLOS Computational Biology*, 13(7):e1005647, 2017.

[316] Yang Yang, Yang Zhang, Bing Ren, Jesse R Dixon, and Jian Ma. Comparing 3D Genome Organization in Multiple Species Using Phylo-HMRF. *Cell systems*, 8(6):494–505.e14, 2019.

[317] Galip Gürkan Yardımcı, Hakan Ozadam, Michael E. G. Sauria, Oana Ursu, Koon-Kiu Yan, Tao Yang, Abhijit Chakraborty, Arya Kaul, Bryan R. Lajoie, Fan Song, Ye Zhan, Ferhat Ay, Mark Gerstein, Anshul Kundaje, Qunhua Li, James Taylor, Feng Yue, Job Dekker, and William S. Noble. Measuring the reproducibility and quality of Hi-C data. *Genome Biology*, 20(1):57, 2019.

[318] Qiancheng You, Anthony Youzhi Cheng, Xi Gu, Bryan T. Harada, Miao Yu, Tong Wu, Bing Ren, Zhengqing Ouyang, and Chuan He. Direct DNA crosslinking with CAP-C uncovers transcription-dependent chromatin organization at high resolution. *Nature Biotechnology*, pages 1–11, 2020.

[319] Feng Yue, Yong Cheng, Alessandra Breschi, Jeff Vierstra, Weisheng Wu, Tyrone Ryba, Richard Sandstrom, Zhihai Ma, Carrie Davis, Benjamin D Pope, Yin Shen, Dmitri D Pervouchine, Sarah Djebali, Robert E Thurman, Rajinder Kaul, Eric Rynes, Anthony Kirilusha, Georgi K Marinov, Brian A Williams, Diane Trout, Henry Amrhein, Katherine Fisher-Aylor, Igor Antoshechkin, Gilberto DeSalvo, Lei-Hoon See, Meagan Fastuca, Jorg Drenkow, Chris Zaleski, Alex Dobin, Pablo Prieto, Julien Lagarde, Giovanni Bussotti, Andrea Tanzer, Olgert Denas, Kanwei Li, M A Bender, Miaohua Zhang, Rachel Byron, Mark T Groudine, David McCleary, Long Pham, Zhen Ye, Samantha Kuan, Lee Edsall, Yi-Chieh Wu, Matthew D Rasmussen, Mukul S Bansal, Manolis Kellis, Cheryl A Keller, Christapher S Morrissey, Tejaswini Mishra, Deepti Jain, Nergiz Dogan, Robert S Harris, Philip Cayting, Trupti Kawli, Alan P Boyle, Ghia Euskirchen, Anshul Kundaje, Shin Lin, Yiing Lin, Camden Jansen, Venkat S Malladi, Melissa S Cline, Drew T Erickson, Vanessa M Kirkup, Katrina Learned, Cricket A Sloan, Kate R Rosenbloom, Beatriz Lacerda de Sousa, Kathryn Beal, Miguel Pignatelli, Paul Flicek, Jin Lian, Tamer Kahveci, Dongwon Lee, W James Kent, Miguel Ramalho Santos, Javier Herrero, Cedric Notredame, Audra Johnson, Shinny Vong, Kristen Lee, Daniel Bates, Fidencio Neri, Morgan Diegel, Theresa Canfield, Peter J Sabo, Matthew S Wilken, Thomas A Reh, Erika Giste, Anthony Shafer, Tanya Kutyavin, Eric Haugen, Douglas Dunn, Alex P Reynolds, Shane Neph, Richard Humbert, R Scott Hansen, Marella De Bruijn, Licia Selleri, Alexander Rudensky, Steven Josefowicz, Robert Samstein, Evan E Eichler, Stuart H Orkin, Dana Levasseur, Thalia Papayannopoulou, Kai-Hsin Chang, Arthur Skoultchi, Srikanta Gosh, Christine Disteche, Piper Treuting, Yanli Wang, Mitchell J Weiss, Gerd A Blobel, Xiaoyi Cao, Sheng Zhong, Ting Wang, Peter J Good, Rebecca F Lowdon, Leslie B Adams, Xiao-Qiao Zhou, Michael J Pazin, Elise A Feingold, Barbara Wold, James Taylor, Ali Mortazavi, Sherman M Weissman, John A Stamatoyannopoulos, Michael P Snyder, Roderic Guigo, Thomas R Gingeras, David M Gilbert, Ross C Hardison, Michael A Beer, Bing Ren, and Mouse ENCODE Consortium. A comparative encyclopedia of DNA elements in the mouse genome. *Nature*, 515(7527):355–64, 2014.

[320] JJ Yunis and O Prakash. The origin of man: a chromosomal pictorial legacy. *Science*, 215(4539):1525–1530, 1982.

[321] JJ Yunis, Sawyer, and K Dunham. The striking resemblance of high-resolution G-banded chromosomes of man and chimpanzee. *Science*, 208(4448):1145–1148, 1980.

[322] Wanwen Zeng, Xi Chen, Zhana Duren, Yong Wang, Rui Jiang, and Wing Hung Wong. DC3 is a method for deconvolution and coupled clustering from bulk and single-cell genomics data. *Nature Communications*, 10(1):4613, 2019.

[323] Yinxiu Zhan, Luca Mariani, Iros Barozzi, Edda G. Schulz, Nils Blüthgen, Michael Stadler, Guido Tiana, and Luca Giorgetti. Reciprocal insulation analysis of Hi-C data shows that TADs represent a functionally but not structurally privileged scale in the hierarchical folding of chromosomes. *Genome Research*, 27(3):479–490, 2017.

[324] Yanxiao Zhang, Ting Li, Sebastian Preissl, Maria L Amaral, Jonathan D Grinstein, Elie N Farah, Eugin Destici, Yunjiang Qiu, Rong Hu, Ah Y Lee, Sora Chee, Kaiyue Ma, Zhen Ye, Quan Zhu, Hui Huang, Rongxin Fang, Leqian Yu, Juan C Izpisua Belmonte, Jun Wu, Sylvia M Evans, Neil C Chi, and Bing Ren. Transcriptionally active HERV-H retrotransposons demarcate topologically associating domains in human pluripotent stem cells. *Nature genetics*, 51(9):1380–1388, 2019.

[325] Quanyi Zhao, Michael Dacre, Trieu Nguyen, Milos Pjanic, Boxiang Liu, Dharini Iyer, Paul Cheng, Robert Wirka, Juyong Brian Kim, Hunter B. Fraser, and Thomas Quertermous. Molecular mechanisms of coronary disease revealed using quantitative trait loci for TCF21 binding, chromatin accessibility, and chromosomal looping. *Genome Biology*, 21(1):135, 2020.

[326] Shanrong Zhao, Wai-Ping Fung-Leung, Anton Bittner, Karen Ngo, and Xuejun Liu. Comparison of RNA-Seq and Microarray in Transcriptome Profiling of Activated T Cells. *PLoS ONE*, 9(1):e78644, 2014.

[327] Hui Zheng and Wei Xie. The role of 3D genome organization in development and cell differentiation. *Nature Reviews Molecular Cell Biology*, 20(9):535–550, 2019.

[328] Wei Zheng, Tara A. Gianoulis, Konrad J. Karczewski, Hongyu Zhao, and Michael Snyder. Regulatory Variation Within and Between Species. *Annual Review of Genomics and Human Genetics*, 12(1):327–346, 2011.

[329] Jingtian Zhou, Jianzhu Ma, Yusi Chen, Chuankai Cheng, Bokan Bao, Jian Peng, Terrence J. Sejnowski, Jesse R. Dixon, and Joseph R. Ecker. HiCluster: A Robust Single-Cell Hi-C Clustering Method Based on Convolution and Random Walk. *bioRxiv*, page 506717, 2018.

[330] Jingtian Zhou, Jianzhu Ma, Yusi Chen, Chuankai Cheng, Bokan Bao, Jian Peng, Terrence J. Sejnowski, Jesse R. Dixon, and Joseph R. Ecker. Robust single-cell Hi-C clustering by convolution- and random-walk–based imputation. *Proceedings of the National Academy of Sciences*, 116(28):14011–14018, 2019.

[331] Xiang Zhou, Carolyn E Cain, Marsha Myrthil, Noah Lewellen, Katelyn Michelini, Emily R Davenport, Matthew Stephens, Jonathan K Pritchard, and Yoav Gilad. Epigenetic modifications are associated with inter-species gene expression variation in primates. *Genome biology*, 15(12):547, 2014.

[332] Yan Zhou, Jiadi Zhu, Tiejun Tong, Junhui Wang, Bingqing Lin, and Jun Zhang. A statistical normalization method and differential expression analysis for RNA-seq data between different species. *BMC Bioinformatics*, 20(1):163, 2019.

[333] Y Zhu, Z Chen, K Zhang, M Wang, D Medovoy, JW Whitaker, B Ding, N Li, L Zheng, and Wang W. Constructing 3D interaction maps from 1D epigenomes. *Nature Communications*, 2016.

[334] Ying Zhu, André M. M. Sousa, Tianliuyun Gao, Mario Skarica, Mingfeng Li, Gabriel Santpere, Paula Esteller-Cucala, David Juan, Luis Ferrández-Peral, Forrest O. Gulden, Mo Yang, Daniel J. Miller, Tomas Marques-Bonet, Yuka Imamura Kawasawa, Hongyu Zhao, and Nenad Sestan. Spatiotemporal transcriptomic divergence across human and macaque brain development. *Science*, 362(6420):eaat8077, 2018.

[335] Marie Zufferey, Daniele Tavernari, Elisa Oricchio, and Giovanni Ciriello. Comparison of computational methods for the identification of topologically associating domains. *Genome biology*, 19(1):217, 2018.

[336] Jessica Zuin, Jesse R. Dixon, Michael I. J. A. van der Reijden, Zhen Ye, Petros Kolovos, Rutger W. W. Brouwer, Mariëtte P. C. van de Corput, Harmen J. G. van de Werken, Tobias A. Knoch, Wilfred F. J. van IJcken, Frank G. Grosveld, Bing Ren, and Kerstin S. Wendt. Cohesin and CTCF differentially affect chromatin architecture and gene expression in human cells. *Proceedings of the National Academy of Sciences*, 111(3):996–1001, 2014.