

Electronic Thesis and Dissertation Repository

3-5-2020 9:00 AM

Genomic Domains And Tissue Type Conservation Of Chromatin Accessibility Differences Between Human Metaphase Homologues

Seana L. Hill, *The University of Western Ontario*

Supervisor: Knoll, Joan HM, *The University of Western Ontario*

A thesis submitted in partial fulfillment of the requirements for the Master of Science degree in Pathology and Laboratory Medicine

© Seana L. Hill 2020

Follow this and additional works at: <https://ir.lib.uwo.ca/etd>



Part of the [Genetics and Genomics Commons](#)

Recommended Citation

Hill, Seana L., "Genomic Domains And Tissue Type Conservation Of Chromatin Accessibility Differences Between Human Metaphase Homologues" (2020). *Electronic Thesis and Dissertation Repository*. 6845. <https://ir.lib.uwo.ca/etd/6845>

This Dissertation/Thesis is brought to you for free and open access by Scholarship@Western. It has been accepted for inclusion in Electronic Thesis and Dissertation Repository by an authorized administrator of Scholarship@Western. For more information, please contact wlsadmin@uwo.ca.

ABSTRACT

During mitosis, interphase chromatin structures change dramatically to allow formation of discrete chromosomes. The mechanisms that follow, allowing rapid and reproducible re-establishment of functional interphase organizations, remain elusive. Our laboratory identified locus-specific condensation differences (referred to as differential accessibility [DA]) in metaphase chromosome homologues by visualizing genome sequence-defined single-copy (sc) DNA probes using fluorescence *in situ* hybridization (FISH). Original identifications of DA loci were performed with individual scFISH probes (1.5–4kb) in lymphocyte chromosomes. In this study, we computationally designed multiple adjacent scFISH probes for 6 different DA loci and determined that DA occurs in domains. Domain lengths varied from ~16kb-130kb. DA was also investigated in other cell types (bone marrow and fibroblast samples) using 5 known DA probes and found to be maintained. DA is a conserved, structural feature in metaphase chromosomes that may play a role in the maintenance of chromosome memory from parent to daughter cells.

Keywords: Mitosis, human metaphase chromosomes, homologs, chromatin accessibility, single copy DNA probes, chromosome condensation, epigenetics, topologically associated domains, lymphocytes, fibroblasts, bone marrow, fluorescence *in situ* hybridization, chromatin structure

LAY AUDIENCE SUMMARY

Human DNA is organized into 23 different pairs of chromosomes. One member of each pair is inherited from each parent. Chromosome structure changes from being dispersed and string-like during periods of cell growth to highly compacted individual units during cell division. During the compaction phase, important functional structures disappear, however, chromosomes must return to their string-like form to function properly for cell growth. These changes in structure are necessary to form healthy daughter cells identical to the parent cell. This cycle continues for each new generation. The memory that allows this cycle of accurate chromosome reorganization from one cell generation to the next is not well understood.

Highly compacted chromosomes can be seen using a microscope and individual pairs of chromosomes can be identified. Our laboratory examines different regions of compacted chromosomes using a technique called FISH (fluorescence *in situ* hybridization) that identifies short regions of DNA by colouring the targeted region with fluorescence using DNA probes. Using FISH with short, unique human DNA probes, we previously identified novel non-random structural differences within individual compacted chromosomes pairs. Within a chromosome pair, using these DNA probes, particular regions of one chromosome of the pair shone brighter than the other. We refer to this finding as differential accessibility (DA). Identified in white blood cells in short individual regions on multiple chromosomes, DA results from different amounts of compaction at the same region between members of a chromosome pair.

In this study, new DA regions were identified. Six regions were expanded to include neighbouring, unique DNA sequences to determine if DA extended past the edges of regions identified by one probe. DA was found to extend outside of these individual regions. Five

regions investigated in blood, bone and skin cells showed DA to be conserved at the same locations across the three tissues. This supports DA as a structural feature of compacted human chromosomes present during cell division that spans beyond the edges of individual unique regions at the same location in different cell types. DA may help to accurately transmit the memory of important structures between parent and daughter cells.

ACKNOWLEDGMENTS

First, I'd like to express my immense gratitude to my supervisor, Dr. Joan Knoll. I would not have been able to complete this degree without her continuous support and guidance. Thank you for all of the time taken to train and teach, and for always being willing to lend your expertise to review and check chromosome IDs. Thank you for always being available for any question, quandary, and concern and for knowing that the first step is always a cup of tea. I'm very appreciative for the opportunity to work on this project with you and Dr. Rogan. Thank you to Dr. Peter Rogan, without his help and guidance I would not have been able to complete this project. Thank you for all your input and insight for all facets of this project. The time and effort to review proposals, posters, and this thesis were incredibly appreciated. I am forever grateful for the invaluable experiences and knowledge I have gained from this opportunity.

Thank you to my advisory committee, Dr. Rogan and Dr. Martin Duennwald, for your time, support, and feedback. Thank you Dr. Duennwald for always being open and available for both revision and help when needed.

A massive thank you to the members of the Rogan/Knoll lab. Thank you to John Mucaki, Ben Shirley, and Dr. Rogan for your expertise in all major computer crises. Thank you John for your technical expertise and helping in my training including all your help with probe design, always being there to answer any question big or small, and always letting me tag along to product shows (i.e. free coffee and cookies). To Ben for always helping with technical support, asking how I'm doing and the funny conversations. Thank you to Dr. Ruipeng Li for your help and support especially with CTCF binding site prediction. Thank you to Jerry Wang, I couldn't have asked for a better person to share the bench with. Thank you for all the conversations during wait times, I learned a lot more about basketball than I ever

thought I would. And thank you Dr. Wahab Khan, for all your work in identifying and characterizing DA that was the foundation of this project and for being open to answering technical questions.

I am also grateful to have had the financial support from the Ontario Graduate Scholarship (OGS) and Western Graduate Research Scholarship (WGRS) during this degree.

Thank you to Tracey, Susan and the rest of the administrative staff in the Pathology and Laboratory Medicine department for the great support and for answering any and all questions.

Thank you to my family for all your support and knowing just when I need some quality dog-sitting time. Thank you to my Mom and Dad, I couldn't have gotten this far without you. I always find it difficult to put into words how much your constant support and love means to me through stress and abnormal schedule but thank you for everything that you've done for me and your constant encouragement. I also couldn't have gotten through this without friends both old and new. Thank you to everyone for your support through it all. Erin, Bailey, Kristin, Emily, Allison, Eleane, Celia, Dom, and Alison for being constant throughout the last 2 years. Srinitya, Janice, Jina, Maddi, Vy, Christine, and all the other wonderful students I've met in this department. You get it.

A final thank you to everyone, including those I didn't name, who have helped me through this degree. It's difficult to articulate my gratitude, I wouldn't have made it this far without the help and support of so many people. It is immensely appreciated.

TABLE OF CONTENTS

<i>ABSTRACT</i>	<i>ii</i>
<i>LAY AUDIENCE SUMMARY</i>	<i>iii</i>
<i>ACKNOWLEDGMENTS</i>	<i>v</i>
<i>TABLE OF CONTENTS</i>	<i>vii</i>
<i>LIST OF TABLES</i>	<i>xi</i>
<i>LIST OF FIGURES</i>	<i>xii</i>
<i>ACRONYMS AND ABBREVIATIONS</i>	<i>xv</i>
<i>Chapter 1: Introduction</i>	1
1.1 Introduction.....	1
1.2 Higher-Order Chromosome Structures.....	2
1.2.1 Chromatin Organization during Interphase.....	5
1.2.2 Organization of Chromatin into Metaphase Chromosomes during Mitosis.....	9
1.2.3 Mitotic Memory and Bookmarking.....	14
1.3 Chromosome Organization and Disease.....	15
1.4 Localized Regions of Differential Chromatin Condensation between Homologous Metaphase Chromosomes.....	17
1.4.1 Initial Observation of Differential Accessibility (DA) – Development of Single- copy Fluorescence <i>in situ</i> Hybridization (scFISH) Technique.....	17
1.4.2 Characterizing DA Between Metaphase Homologues.....	18
1.5 Study Objectives.....	21

Chapter 2: Materials and Methods	24
2.1 Design and Development of Single-copy DNA FISH probes	24
2.1.1 Oligonucleotide Primer Design and Amplicon Production.....	24
2.1.2 Labeling of Purified Amplicon by Nick Translation	30
2.1.3 Probe Nomenclature System	33
2.2 Probe Selection to Determine DA Domain Size	33
2.3 Probe Selection to Examine if DA Occurs in Different Tissues	35
2.4 Cell Culture and Preparation	36
2.5 Single-copy Fluorescence <i>in situ</i> Hybridization (ScFISH).....	38
2.6 Scoring of Hybridized Probe Fluorescence Intensity Differences Between Metaphase Homologues	40
2.7 Analysis of Epigenetic Characteristics Present During Interphase for Individual Probe Loci and DA Domains	42
Chapter 3: Results	45
3.1 Identification of New Regions of Differential Accessibility (DA)	45
3.1.1 New Sc Probes Developed and DA Confirmed by Qualitative ScFISH on Human Metaphase Chromosomes	45
3.1.2 Quantification of DA by Gradient Vector Flow Analysis (GVF) in ScFISH Probes	52
3.1.3 Comparing Open Chromatin Marks Between DA and Equivalent Accessibility (EA) Regions.....	54

3.2 Defining DA Domains in Human Metaphase Homologous Chromosomes	60
3.2.1 Identification of Differentially Accessible Domains	60
3.2.2 Epigenetic Characteristics of DA Domains and the ScFISH Probes Defining DA Domains – Topologically Associated Domains (TADs)	70
3.3 DA Loci Are Conserved Between Peripheral Blood, Bone Marrow, and Fibroblast Tissues.	78
3.3.1 Expression in DA Regions Across T-Lymphocytes, Bone Marrow Cells and Fibroblasts.....	78
3.3.2 Conservation of DA Loci in T-Lymphocytes, Cells from Bone Marrow, and Fibroblasts.....	82
<i>Chapter 4: Discussion</i>	89
4.1 Identification and Validation of New DA Loci.....	89
4.2 Definition of Six DA Domains.....	92
4.3 Conservation of DA Loci Across Tissue Types.....	99
4.4 Conclusion.....	102
<i>REFERENCES.....</i>	105
<i>APPENDICES</i>	I
Appendix I: Gene Names.....	I
Appendix II: Copyright Release for Use of Image	II

Appendix III: Python script reporting mean expression values of RefSeq genes in lymphocytes, bone marrow, and fibroblast cells from the Human Protein Atlas (HPA) and Genotype-Tissue Expression (GTEx) databases.....	IV
Appendix IV: Details for the production of all new ScFISH probes developed in this study	IX
Appendix V: Frequency of copy number variants (CNVs) that overlap DA intervals used for scFISH probes.....	XI
Appendix VI: Comparison between DA domain location in metaphase and interphase TAD and sub-TAD structures for the XDH, FGF6, and TPM1 domains	XIII
Appendix VII: DA and EA cell count for each probe and tissue type examined with results of significance testing between DA and EA proportions per probe and between individuals	XVI

LIST OF TABLES

Table 2.1: PCR cycling parameters for single-copy probe amplification using long PCR.....	27
Table 3.1: ScFISH probes developed and validated to evaluate the extent of DA domains established with an anchor scFISH probe and conservation of DA between tissues.....	50
Table 3.2: Integrated intensity values of open chromatin marks in each DA and EA interval from this study.....	59
Table 3.3: List of DA domains with each scFISH probe defining each domain.....	63
Table 3.4: Median CTCF binding affinity/kilobase across DA domains compared to sequences of equal length beyond the boundaries of the domain within the same TAD and in the adjacent TAD.....	77
Appendix IV: Details for the production of all new ScFISH probes developed in this study....	IX
Appendix V: Frequency of copy number variants (CNVs) that overlap DA intervals used for scFISH probes.....	XI
Appendix VII: DA and EA cell count for each probe and tissue type examined with results of significance testing between DA and EA proportions per probe and between individuals....	XVI

LIST OF FIGURES

Figure 1.1: Organization of chromatin during interphase from linear DNA through to higher-order chromatin structures and chromosome territories to fit the genome into the nucleus.....	4
Figure 1.2: The progression of chromatin compaction during mitosis to achieve discrete metaphase chromosomes with supercoiling differences between metaphase homologues.....	12
Figure 2.1: Gel image example of an optimal and poor scFISH probe amplification produced in this study.....	29
Figure 2.2: Example image of check gel to assess length of nick translation products of 3.3_1p36, ONECUT1_F2701, and XDH_C2501 probes.....	32
Figure 3.1: Differential and equivalent accessibility hybridization patterns between human metaphase homologues detected by single-copy probe fluorescence <i>in situ</i> hybridization.....	48
Figure 3.2: Frequency of cells exhibiting DA on homologous chromosome regions by sc probe FISH analysis.....	49
Figure 3.3: Quantification of probe signal fluorescence between homologues shown by box and whisker plots of normalized integrated fluorescence intensity ratios.....	53
Figure 3.4: Distribution of integrated intensity data for each open chromatin mark in new DA intervals (n=18) and previously reported EA intervals ⁷ (n=59).....	56
Figure 3.5: Genomic map demonstrating the difference in enrichment of open chromatin mark H3K27ac at the <i>SCAMP2_IVS1</i> DA probe outlier compared to 2 neighbouring DA probe loci <i>SCAMP2_IVS7-IVS4</i> and <i>COX5A_tel20100</i>	57
Figure 3.6: Open chromatin marks at DA loci have lower mean integrated intensities compared to EA loci.....	58
Figure 3.7: Genomic map of XDH domain.....	64

Figure 3.8: Genomic map of HMGB1P5 domain.....	65
Figure 3.9: Genomic map of FGF6 domain.....	66
Figure 3.10: Genomic map of TPM1 domain.....	67
Figure 3.11: Genomic map of COX5A domain.....	68
Figure 3.12: Genomic map of HMGB1P1 domain.....	69
Figure 3.13: Localization of the COX5A domain in metaphase cells relative to TAD structures present in interphase cells.....	74
Figure 3.14: Localization of the HMGB1P1 domain in metaphase cells relative to TAD structures present in interphase cells.....	75
Figure 3.15: Localization of the HMGB1P5 domain in metaphase cells relative to TAD structures present in interphase cells.....	76
Figure 3.16: Mean gene expression in transcripts per million (TPM) reported for each gene in which a confirmed DA scFISH probe localized.....	80
Figure 3.17: Mean gene expression in transcripts per million (TPM) reported for each gene in which a confirmed EA scFISH probe localized.....	81
Figure 3.18: The <i>PCK1_cen180-IVS6</i> region shows the conservation of DA within a gene in metaphase T-lymphocyte, bone marrow and fibroblast cells.....	83
Figure 3.19: The <i>TPM1_tel3200</i> region shows conservation of DA in an intergenic region in metaphase T-lymphocyte, bone marrow and fibroblast cells.....	84
Figure 3.20: The 3.3_1p36 region shows conservation of EA in metaphase T-lymphocyte, bone marrow and fibroblast cells.....	85
Figure 3.21: Proportion of cells scored as DA and EA within lymphocytes, bone marrow cells, and fibroblasts are not significant.....	86

Appendix VI: Figure 1: Localization of the XDH domain in metaphase cells relative to TAD structures present in interphase cells.....XIII

Appendix VI: Figure 2: Localization of the FGF6 domain in metaphase cells relative to TAD structures present in interphase cells.....XIV

Appendix VI: Figure 3: Localization of the TPM1 domain in metaphase cells relative to TAD structures present in interphase cells.....XV

ACRONYMS AND ABBREVIATIONS

BAC – Bacterial Artificial Chromosome

BLAST® – Basic Local Alignment Search Tool

BLAT – BLAST Like Alignment Tool

bp – Base Pair

BSA – Bovine Serum Albumin

ChIP-seq – Chromatin Immunoprecipitation Assay with Sequencing

CNV – Copy Number Variation

CO₂ – Carbon Dioxide

CTCF – CCCTC Binding Factor Protein

DA – Differential Accessibility

dATP – Deoxyadenine Triphosphate

dCTP – Deoxycytosine Triphosphate

dGTP – Deoxyguanine Triphosphate

DAPI – 4',6-diamidino-2-phenylindole

DNA – Deoxyribonucleic Acid

DNase I HS – Deoxyribonuclease I Hypersensitivity

dNTP – Deoxyribonucleotide Triphosphate

dTTP – Deoxythymidine Triphosphate

dUTP – Deoxyuridine Triphosphate

EA – Equivalent Accessibility

EBV- Epstein-Barr Virus

EDTA – Ethylenediaminetetraacetic Acid

ENCODE – Encyclopedia of DNA Elements

FAIRE – Formaldehyde-Assisted Isolation of Regulatory Elements

FBS – Fetal Bovine Serum

FISH – Fluorescence *In Situ* Hybridization

Gene names – (see Appendix I)

GTE_x – Genotype-Tissue Expression database

GVF – Gradient Vector Flow

H3K4_{me} – Histone 3 Lysine 4 Methylation

H3K4_{me2} - Histone 3 Lysine 4 Dimethylation

H3K9_{ac} – Histone 3 Lysine 9 Acetylation

H3K27_{ac} – Histone 3 Lysine 27 Acetylation

kb – Kilobase

KCl – Potassium Chloride

M – Molar (moles/liter)

Mb – Megabase

mL – Milliliter

mM – Millimolar (mole/milliliter)

mmol – Millimole

nmol – Nanomole

MgCl₂ – Magnesium Chloride

PCR – Polymerase Chain Reaction

PHA – Phytohemagglutinin

RNA – Ribonucleic Acid

sc – Single-copy

SSC – Saline Citrate

TAD – Topologically Associated Domain

TAE buffer – Tris-acetate-EDTA buffer

Tris-HCl – Tris (hydroxymethyl) Aminomethane Hydrochloride

UCSC – University of California Santa Cruz

µg – Microgram

µL – Microliter

Chapter 1: Introduction

1.1 Introduction

The blueprint of life exists within the sequence of nucleotide bases in DNA. The 3-dimensional organization of that sequence, both spatially and temporally, is required for proper execution of this blueprint¹. Structural organization, an essential epigenetic mechanism, regulates differential gene expression programs required for processes such as cell growth, division, differentiation and survival¹⁻³. Chromatin organization achieves two purposes: fitting the entire genome in a readily accessible manner within a spatially restrictive nucleus and allowing the interaction of genes and distant genetic elements like promoters and enhancers as needed⁴. A continuous cycle of condensation and decondensation causes dynamic chromatin organization throughout a cell's lifetime. A high degree of condensation is necessary to ensure high fidelity segregation during cell division, however a more relaxed chromatin organization is required for proper access of regulatory and transcriptional machinery to the genome to ensure normal cell function during interphase^{3,5}. Despite constant changes in organization through the cell cycle, new generations of cells are able to re-establish correct gene programming consistent with that of parent cells⁶. The understanding of this mechanism remains incomplete, though it is proposed as an epigenetic memory that allows correct replication of genome organization in cell progeny¹.

Our laboratory identified a novel, non-random, stable condensation difference between homologous metaphase chromosomes in human lymphocytes^{7,8}. This difference is designated differential accessibility or DA^{7,8}. It was detected by fluorescence *in situ* hybridization (FISH) using short, single-copy (sc) DNA probes^{7,9,10}. It is not understood why DA is present in metaphase chromosomes and if there is a relationship between DA in metaphase homologues

and the chromatin structures present in interphase. Further investigation is necessary to address what purpose DA may have. This study examines the genomic distribution of DA beyond single scFISH probe boundaries to define lengths of contiguous DA intervals and assesses whether DA is conserved across different tissues.

1.2 Higher-Order Chromosome Structures

A well-defined cell cycle is followed to produce new healthy, functional cell generations. This cycle splits the diploid parent cell into two identical diploid daughter cells¹¹. The cell cycle involves complete replication and equal segregation of the genome between two daughter cells, as well as the distribution of cytoplasm and organelles between the two. There are 4 stages in the typical cell cycle: G₁, S, G₂, and M phases. Interphase includes i) G₁, the phase where the diploid cell is metabolically active and preparing for DNA replication, ii) S, the phase where DNA is replicated, and iii) G₂, the phase following replication where the cell is metabolically active and preparing for cell division (M phase)¹¹. M phase or mitosis, in which the parent cell divides into 2 daughter cells, passes through a sequence of stages; prophase, metaphase, anaphase and telophase, respectively^{6,12}. The events describing each stage include defined changes in chromatin condensation¹²⁻¹⁴. Metaphase is the stage in which chromatin is at its most condensed and when discrete chromosomes can be visualized using microscopy^{12,13}. An incredible amount of condensation in metaphase chromosomes, which is >100 fold compared to that of interphase chromatin, is necessary for the accurate segregation of discrete metaphase chromosomes between daughter cells during mitotic cell division^{5,14-16}. Once M phase is complete, the two daughter cells are separated by cytokinesis, enter G₁ and the cycle begins again.

The chromatin organization present in interphase must be lost during mitosis to allow the formation of compact chromatin units^{5,14}. To understand chromatin structure during cell division and the methods of maintaining interphase organizations between cell generations, knowledge of metaphase structure must be paired with knowledge of the interphase structures lost and faithfully re-established in future cell generations^{1,13}. Chromatin at a lower compaction level is relatively well understood; with chromatin organization beginning with the association of architectural proteins with the DNA strand creating a “bead-on-string” chromatin fiber (Figure 1.1)^{6,17}. The same cannot be said for higher-order organization of both interphase and metaphase chromatin^{5,13,14,18,19}. Higher-order structures organize chromatin into defined locations, domains, and compartments that are spatially programmed in the nucleus of the interphase cell (Figure 1.1). These locations are important to facilitate and anchor interactions across and between chromosomes. They are involved in regulating transcription spatially and temporally during interphase. These include an important level of organization, topologically associated domains (TADs)^{6,20–22}.

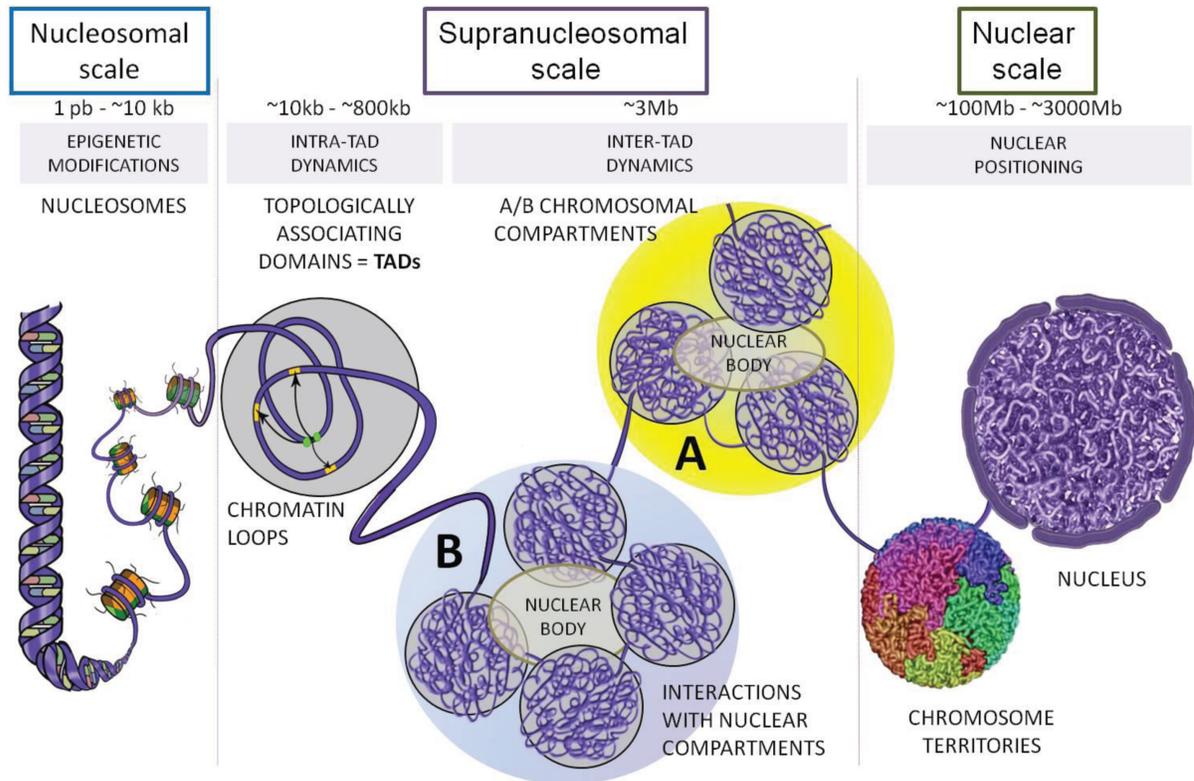


Figure 1.1: Organization of chromatin during interphase from linear DNA to higher-order chromatin structures and chromosome territories to fit the genome into the nucleus. At the nucleosomal scale, organization of the DNA strand begins by wrapping around histone octamers to form nucleosomes. This is followed by formation of chromatin loops, organizing into topologically associating domains then further into active A and inactive B compartments at the supranucleosomal scale. At the nuclear scale, chromatin is organized into chromosome territories. The increasing levels of organizations allow the genome to fit into the nucleus. Reprinted from Ref [23] copyright licensed under a Creative Commons Attribution 4.0 International License <https://creativecommons.org/licenses/by/4.0/legalcode>.

1.2.1 Chromatin Organization during Interphase

Chromatin is defined as the string of nucleic acids with the associated architectural proteins that allow organization beyond the linear DNA sequence. The basic level of chromatin organization are nucleosomes, eight histone proteins wrapped twice by DNA (Figure 1.1)^{13,17,23,24}. The prevailing belief of chromatin organization has been a uniform polymer of nucleosomes. Formed by an average 147 bp of DNA wrapped around a histone octamer, each nucleosome is linked by a linker histone, H1, to 20-60 bp of unassociated DNA^{17,24}. This forms the 10nm diameter “bead-on-string” fiber that is considered the base of chromosome organization throughout the cell cycle^{13,23}. This histone octamer core is composed of two H3 and two H4 histone subunits forming a tetramer, and 2 heterodimers made up of one H2A and one H2B histone^{13,17,23}. At this level of organization there are a number of modifications with important regulatory functions applied to both the DNA strand and histones. These include chemical modifications such as methylation and acetylation to the common histone subunits as well as different structural histone variant substitutes^{23,25}.

The next level of condensation has been referred to as the 30nm strand, in which the 10nm fiber folds to form a strand of chromatin 30nm in diameter decreasing accessibility to the underlying DNA sequence^{18,26,27}. Two different methods of folding into this strand have been proposed. The “solenoid” model, where the strand wraps around itself in a helical manner with adjacent nucleosomes in contact with each other. The other is the “zigzag” model, where nucleosomes are in contact with every other nucleosome, not with the one directly beside it, forming a zigzag pattern¹⁸. Chromatin was believed to continue to fold in a hierarchical manner into uniform strands of larger and larger diameters to allow the level of condensation observed in interphase and beyond into metaphase¹³.

New evidence contradicts this hierarchical folding pattern. It proposes that the base chromatin strand is not a uniform 10nm fiber but rather a strand varying in diameter (5 – 24 nm) with a heterogeneous composition of varying nucleosome arrangement, chromatin concentration densities and conformations¹³. This was observed in both resting and mitotic cells¹³. There is also disagreement about the organization of the 30nm fiber and whether it occurs *in vivo*^{19,28}. Data describing both the 10nm and 30nm fiber were primarily observed in controlled *in vitro* environments with low ionic concentrations and controlled histone concentrations^{13,18}. Further investigations across different eukaryotic organisms and the use of computational models suggest, rather than a distinguishable 30nm fiber, interphase chromatin is organized by the irregular folding of the base chromatin fiber into higher-order chromatin structures^{18,28}. Irregular folding of a less rigid 10nm strand is more consistent with reports of chromatin dynamics during interphase that facilitate accessibility to the DNA strand, compared to the dynamics of a less flexible uniform 30nm strand^{2,18,28,29}.

Though there is debate over defined diameters of hierarchical chromatin organization, there is no doubt that there are other gradations of chromatin organization formed by different layers of chromatin structure that have been observed and studied^{6,21,23,30}. It is clear that the way in which higher-order chromatin structures are formed is important to the proper regulation of cell functions^{6,30}. These higher-order structures include chromosome looping, topologically associated domains (TADs), chromatin compartments, and chromosome territories (Figure 1.1)⁶. TADs specifically are of interest as they have been proposed as the basic unit of chromatin folding³⁰.

The organization of chromatin into discrete domains was initially observed during investigations into DNA replication timing^{30–32}. Replication sites, on average a megabase in

length, were identified by 2 independent groups in 2 different cells types. In both studies, these domains were preserved through multiple cell divisions³⁰⁻³². Similar domain structures were subsequently identified, using new chromatin confirmation capture technologies developed to examine chromatin structure^{22,23,33,34}. These methods, including 3C, 4C, 5C, and Hi-C, hold chromatin interactions in time by chemically-crosslinking interacting segments of genome and are used in both genome-wide and targeted analyses²³. Independent studies identified patterns of chromatin domains where contact was largely restricted within areas of 220kb-1Mb in length with an average length of 880kb using Hi-C and 5C respectively^{22,33}. These domains were also conserved through multiple cell divisions and between different cell types and were found to be consistent with the replication timing domains^{22,33,35}. These domains have been discussed using multiple names: topologically associated domains, topological domains, and contact domains³⁰. We will refer to the domains as TADs (topologically associated domains) for further description in this thesis.

TAD boundaries are conserved across cell types and species, and are stable throughout cell division^{21,22,33}. They have been identified in both human and mouse embryonic stem cells, as well as a number of pluripotent and differentiated cell types^{22,33}. Due to this broad level of conservation, TADs are considered a basic organizational unit of interphase chromatin structure³⁰. TADs are broadly defined as genomic regions where there is a high frequency of interaction within the boundaries of a given TAD, but little interaction with adjacent compartments and other chromatin beyond the boundaries of an individual TAD. This can be simplified as self-association and insulation properties^{22,30}. There is an inner TAD framework of varying chromatin organizations. These inner structures compartmentalize interactions

within the greater TAD structure, they include chromatin loops and sub-TAD interactions^{21,22}. It is these internal organizations that vary between different cell lineages^{21,30}.

Beyond being a basic organizational unit of chromatin, TADs have been implicated in regulation of transcription and DNA replication timing. TADs serve as both structural and functional compartments of chromatin that bring together promoters and enhancers with both single and multiple gene targets^{21,22,33}. This allows co-regulation of different genetic elements by the same regulatory elements within a single TAD. This action is predominantly isolated to regulatory elements and genes within the TAD itself^{21,30}. The TAD boundaries act to contain the action of these regulatory elements within the individual structural compartment^{21,22}. The boundaries are enriched with insulator proteins. One key protein is CCCTC binding factor (CTCF), which is a zinc-finger containing DNA binding protein. CTCF is one factor with a role in facilitating chromatin loop formation including the chromatin loops and sub-TAD interactions on the interior of a TAD^{21,22}. Beyond TADs, interphase chromatin is arranged into A/B compartments^{6,36}. These large compartments are often multiple megabases in length. They are independent of TAD formation and are more associated with active and repressed chromatin marks rather than structural organization of chromatin interactions^{6,23}. The A compartments are correlated with active transcription and open chromatin; and the B compartments are characterized by decreased transcriptional activity and corresponding repressing epigenetic marks^{20,23}. These have been compared to conventional descriptions of chromatin accessibility with respect to euchromatin and heterochromatin. Euchromatin is associated with loosely organized chromatin with increased accessibility allowing active transcription and is observed in metaphase chromatin as light Giemsa staining bands. Heterochromatin is considered more highly condensed, is associated with low transcriptional

activity, and is visualized by dark Giemsa staining bands of metaphase chromatin^{37,38}. However, it is not known to what extent this metaphase conformation and interphase A/B compartment correspond^{5,14}. The organization into the A/B compartments are cell type specific^{21,36}. All of these higher levels of chromatin organization and function, including TADs and A/B compartments, are lost as the chromatin enters the mitotic phase of the cell cycle and condenses to form discrete chromosomes for cell division^{5,14}.

1.2.2 Organization of Chromatin into Metaphase Chromosomes during Mitosis

Discrete chromosomes are required for faithful and complete genetic propagation in future cell generations produced during cell division. This requires chromatin to undergo a high degree of condensation with extensive reorganization of interphase structural arrangements^{3,13}. The predominate theory had been that this condensation occurred in a hierarchical manner. The 10 nm and 30 nm diameter strands, discussed in Section 1.2.1, continue to fold over itself into strands of increasing diameters: 120 nm [i.e chromonema], 300-700 nm [i.e chromatid], and 1400nm [i.e metaphase chromosomes]¹³. This model, however, much like the organization into uniform 10 and 30 nm strands primarily observed in purified chromatin *in vitro*, has been challenged by recent findings of chromatin organization *in vivo*^{5,13,14}. The manner in which mitotic chromosomes organize and their internal conformation have largely remained a mystery. A number of models, including different loops-on-scaffolds and hierarchical models, have been proposed using a variety of methods including conventional light and electron microscopy^{5,13}. Recent work examining metaphase condensation and chromatin structures with chromatin capture technologies (Hi-C, 5C)⁵ as well as electron microscopy tomography with a fluorescent labeling technique *in situ*¹³ have improved understanding of mitotic chromosome structure.

During mitosis, similar to proposed irregular organizations of interphase, chromatin appears to have a heterogeneous structural conformation, rather than a fiber of a consistent diameter as the hierarchical model would suggest. The diameter of the chromatin strand was found to cover the same range as the strands in interphase, 5 nm to 24 nm in diameter^{5,13,14}. The method of compaction greatly increases density of chromatin by increasing the number of chromatin interactions, with a collapse into arrays of small compact loops around a central axis. It is proposed that it is this density of compaction that differs between interphase and metaphase organizations not the base organization of the strand itself^{13,14}. Recent evidence supports a loop extrusion model as the general mechanism of mitotic chromosome formation, specifically a 2-state loop extrusion model has been proposed (Figure 1.2)^{5,14,39}.

This model begins upon entrance into prophase with the formation of consecutive chromatin loops extending from a central axis allowing linear compaction, followed by axial compression. The reorganization of higher-order chromatin organization continues in metaphase with 80 - 120 kb linear chromatin loops emitting from the central axis primarily composed of topoisomerase II α and condensins (Figure 1.2)^{5,14}. A nested loop structure was later determined, with the 80-120 kb loops forming within larger 200 - 400 kb loops during prometaphase, evidence of which was observed *in vivo* and supported by polymer modeling (Figure 1.2)¹⁴. Axial compression is achieved by the twisting of the central axis into a helix, compressing the length of looped chromatin formation from the longer strand observed in prophase into the short rods of metaphase¹⁴. The formation of this structure is largely mediated by condensin I and II, condensin I is required for the inner loop structures and condensin II for both the larger loops and twisting of the central axis into a helix (Figure 1.2)^{14,39}. A member of the structural maintenance of chromosomes (SMC) family of protein complexes, condensin

is formed by 5 subunits^{39,40}. It has been determined that one of its functions is creating chromatin loops in an ATP dependent manner. This mitotic organization appears to be consistent between cell types^{5,14}.

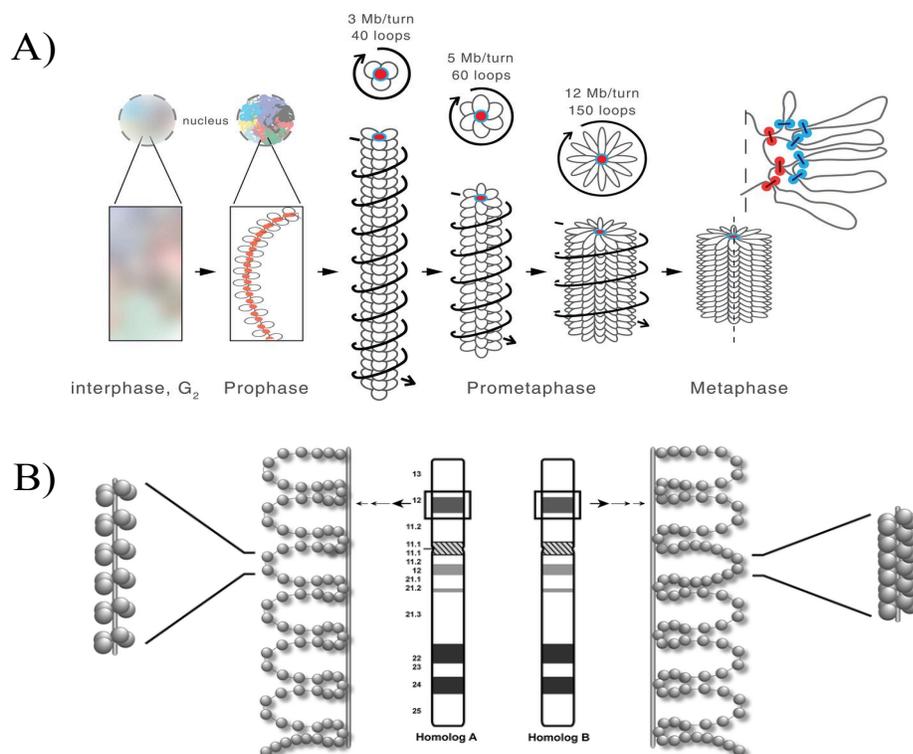


Figure 1.2: The progression of chromatin compaction during mitosis to achieve discrete metaphase chromosomes with supercoiling differences occurring between metaphase homologues. A) Pathway of chromosome condensation from interphase into mitosis proposed by Gibcus *et al.*¹⁴. Condensin II (red) begins compaction of chromatin by forming consecutive loops around a central axis of condensin and topoisomerase II. In prometaphase, condensin II mediated-loops continue to increase in size. A nested loop structure is produced as condensin I (blue) begins to form smaller more peripheral loops of the elongating condensin II mediated loops. The central axis twists into a helical structure that allows axial compression from prometaphase to mature metaphase chromosomes. Image reprinted from Reference [14] with permission from AAAS (Appendix II).

B) Working model of localized differences in solenoidal supercoiling producing differential accessibility (DA) within homologue pairs based on topoisomerase II studies (see Section 1.4) as proposed by Khan, Rogan, and Knoll, 2015⁸. Homologue A shows less compaction compared to homologue B at the same loci. Compaction differences occur within individual chromatin loops without altering loop size and frequency between homologues. Reprinted from Reference [8], copyright licensed under a Creative Commons Attribution 4.0 International License. <https://creativecommons.org/licenses/by/4.0/legalcode>.

A consequence of this reorganization is a loss of higher-order chromatin structures observed during interphase^{2,5}. The organization of chromatin into the compartments and domains outlined in Section 1.2.1 are crucial during cell stages of growth and development⁶. They all serve important roles in regulating accessibility of transcription factors to underlying functional sequences allowing active transcription^{6,30}. Distinct functional compartments and TADs though present in interphase are not preserved during mitosis and is also accompanied by the disassociation of a number of architectural proteins and transcription factors^{12,14}. These structures are no longer present by late prophase, with condensins helping facilitate this loss of architecture while assembling chromatin in their mitotic organization, stopping transcription for the duration of mitosis. The mechanism controlling this loss is still unknown¹⁴. However following mitosis, during entry into G1 phase, there is a quick restoration of functional interphase chromatin structures, re-association of lost epigenetic modifications as well as correct transcription and regulatory activity^{41,42}. It is known that there exists some form of memory of previous epigenetic configurations including organizational domains, histone modifications and protein binding sites that allows a quick restoration of these features following cell division to ensure correct function within the new cell generation^{42,43}. The main focus of epigenetic memory has been the rapid restoration of transcription following exit from mitosis and into the G1 phase of interphase. Preservation of certain histone modifications and transcription factors have been shown in a process called genomic or mitotic bookmarking⁴¹⁻⁴⁴. The variety of factors and mechanisms involved in the bookmarking process still remain unclear including memory of interphase chromatin structures needed when re-establishing functional interphase organization in the next cell generation^{42,43}.

1.2.3 Mitotic Memory and Bookmarking

Overall, mitosis corresponds with a global decrease in transcription^{42,45,46}. This is due to dissociation of transcription factors and their targeted inactivation and degradation accompanied by a global change in histone modifications^{45,47}. A significant reorganization of overall chromatin structure with the loss of functional structures including TADs^{5,14} also occurs. This reorganization also includes modifications to nucleosome structure and position to block interactions with transcription factors^{14,46}. However, in order re-establish the functional chromatin structures and modifications required for transcription in daughter cells there must be some system in place to act as a memory of the parental cell's genomic organization of transcriptional and regulatory signals. The identification and characterization of mechanisms of mitotic memory and bookmarking is an ongoing investigation, of which most details remain unknown.

Currently, the conservation of chromatin organization between generations has been proposed to occur in both an active and passive manner⁴⁸. Investigations have begun to propose and identify mechanisms of active transmission, however passive methods remain undefined⁴⁸. Active mechanisms concern regulatory factors. More specifically described as mitotic bookmarking, it is the conservation of regulatory factors or their specific position through mitosis. This includes transcription factors and epigenetic modifications on mitotic chromatin that allow the faithful inheritance of regulatory information from parent to daughter cells^{42,44,48}. Histone phosphorylation, for example, a modification to histone structure that occurs in mitosis that is reversed following re-entry into interphase, has been linked to aiding in proper chromatin condensation as well as speculated to mark gene regulatory elements^{25,49}. Another histone modification, H3K27ac, well characterized in interphase, is also maintained during

mitosis in a locus-specific manner^{44,50}. Some sites are maintained between different cell types, reported at the promoters of housekeeping genes, and some differ between cell types, observed at enhancers of more cell-type specific genes⁴⁴. A number of transcription factors have been identified to remain bound to mitotic chromosomes, however the nature of this retained association, if they are occurring in a binding site specific or non-specific manner, is not clear⁴². These include factors such as FoxA1, involved in liver differentiation, CTCF, an important factor in chromatin architecture, and GATA1, an important factor in haematopoiesis^{42,43}. Mitotic bookmarking is thought to not only transmit structural and regulatory information to daughter cells, but also to be involved in the rapid reactivation of transcription which occurs upon re-entry into interphase. Thereby, maintaining positions at specific enhancers and promoters to jump start housekeeping and cell-type specific functioning as soon as mitosis is complete^{42,44}. It is important to note that if a particular factor is to be considered a mitotic memory or bookmark, evidence of the disruption in daughter cell function, structure, or viability following interference with that factor during mitosis will need to be established⁴⁸.

1.3 Chromosome Organization and Disease

In addition to aberrant coding regions in the genome that are involved in various mechanisms of disease, there are also disruptions in non-coding regions implicated in abnormalities as severe as those in coding regions^{51,52}. These non-coding regions are necessary for normal cellular function. When linking abnormal phenotypes to their underlying genetic causes, the first area inspected is structural rearrangements that disrupt the genetic sequence within genes. This includes point mutations, deletions, duplications, inversions, and unbalanced rearrangements^{51,52}. There are, however, aberrant phenotypes that cannot be linked entirely to disruptions in coding regions of genes. Recently attention has been directed towards

mutations and variants in non-coding regions that disrupt the functional chromatin structures needed for normal cell function⁵¹⁻⁵³. Disruption of these chromatin organizations necessary for proper regulation of coding regions have been linked to aberrant phenotypes⁵¹⁻⁵³.

Chromatin organization is a key epigenetic regulator of genes. Rearrangements that interfere with the specific architecture of these higher-order chromatin structures, including changing or eliminating TAD boundaries, are implicated in aberrant expression and abnormal phenotypes^{51,52}. As previously mentioned, TADs are considered a functional unit of chromatin, allowing the interaction of regulatory elements with their gene targets and insulating those genes from action of other regulators acting outside the boundaries of a given TAD. Elimination of CTCF binding sites at TAD boundaries was found to cause the merging of neighbouring units, and result in nontypical interactions and aberrant transcription in both human cells and mouse models^{33,54}. The clearest examples of TAD disruption causing abnormal phenotypes are those rearrangements modifying the boundaries of the TAD containing *EPHA4* and its two neighbouring TADs. Alterations of these boundaries resulting in ectopic gene regulation have been implicated in aberrant development⁵¹. Large deletions of multiple CTCF binding sites as well as other mutations (including deletions, inversions and duplications) occurring at TAD boundaries resulted in abnormalities of limb development in mouse models⁵¹. The abnormal phenotypes resulting from each structural rearrangement in mice corresponded with those observed in human samples derived from individuals with limb malformations including brachydactyly, syndactyly, and polydactyly⁵¹. Similar disruptions of TAD boundaries causing ectopic gene expression from abnormal interactions between genes and distant enhancers have been associated with other limb development abnormalities as well as neuropathies⁵⁵⁻⁵⁷.

Beyond topologically associated domains, there are also other changes in chromatin structure related to other abnormal phenotypes. For example, association of chromatin with the tau protein, one of the proteins accumulated in the brain of Alzheimer patients, is connected with widespread chromatin remodeling at the level of higher-order chromatin structure⁵⁸. This study was conducted focussing on H3K9ac, a marker of open chromatin, with tau showing a broad effect on histone acetylation in the brain⁵⁸. Connections have also been made between the aberrant regulation associated with cancer and the reorganization of TAD boundaries^{3,52}. Changes in the epigenome are not restricted to single cell generations. Abnormal TAD boundary locations and other epigenetic changes affecting the action of enhancers and promoters will be maintained through mitosis.

1.4 Localized Regions of Differential Chromatin Condensation between

Homologous Metaphase Chromosomes

A novel, stable, and heritable difference in chromatin condensation between homologous metaphase chromosomes in humans was first identified in our laboratory using single-copy probe technology combined with fluorescence *in situ* hybridization⁷⁻¹⁰. Initially observed as a difference in fluorescence probe intensity between metaphase homologues, it was designated as differential accessibility (DA), after the observed difference in probe intensity was determined to be a localized difference in chromatin supercoiling between metaphase homologues⁷⁻¹⁰.

1.4.1 Initial Observation of Differential Accessibility (DA) – Development of Single-copy Fluorescence *in situ* Hybridization (scFISH) Technique

Single-copy (sc) probe technology, first described in 2001, is a method of using the human genome sequence to design and develop nucleic acid probes for any unique sequence region

present⁹. This method excluded repetitive sequences during the design and development of the probes⁹. Sc probe technology can be used with FISH or other genomic technologies to examine the genome with high definition. ScFISH probes range in length from ~1.5 to 5 kb. Their precise genomic coordinates are known and they can be designed to target single-copy regions within genes, including introns, exons, and promoters, as well as in intergenic regions⁹. In contrast, traditional FISH probes span much larger genomic targets (~100 kb to several megabases), contain repetitive sequences, and are not precisely defined by genomic coordinates^{9,59,60}. ScFISH probes can visualize much smaller genomic regions with greater granularity^{7,9}. Initial scFISH studies focused on targeting clinically relevant areas of the genome and concentrated on genes disrupted in cytogenetic abnormalities of interest^{9,10}.

During scFISH probe validation, a difference in probe fluorescence hybridization intensities between homologues was first observed in a minority of probes. At the time, no clear factor, including chromosome morphology and probe characteristics, could be attributed to causing this difference in intensity¹⁰.

1.4.2 Characterizing DA Between Metaphase Homologues

The first investigations determined that the characteristic difference in hybridized probe fluorescence intensity between metaphase homologues was due to a distinct difference between the chromatin condensation at the same loci^{7,8}. Differential accessibility (DA), was determined to be a stable, non-random, localized, and parentally-derived difference in metaphase chromatin supercoiling between homologous pairs^{7,8}. DA was observed in the majority of cells ($\geq 2/3$) in samples from multiple individuals and at multiple loci⁷. Our laboratory has identified these localized condensation differences between homologues in ~10% of >300 scDNA probes mapped in the human genome, many of which had been

intentionally designed to interrogate copy number and contextual changes in cytogenetically and clinically relevant regions of the human genome^{7,9,10}. Probes that show DA were found to cover exonic, intronic and intergenic genomic regions⁷.

DA is stable, and was observed at the same loci across multiple individuals on multiple regions across the genome⁷. Common copy number variants (CNVs) on homologues were excluded as the source of the fluorescent intensity differences by comparing the genomic coordinates of DA probes to the genomic locations of CNVs observed in two normal control populations⁷. The qualitative analyses of hybridized cells identified with DA or equivalent accessible (EA) homologues at a given locus were supported by quantitative analyses using gradient vector flow analysis (GVF)^{7,61}. DA loci had significantly higher average differences in fluorescence signal between homologues than EA loci⁷.

Super resolution 3D structured illumination microscopy (3D-SIM) provided clear evidence that DA resulted from a significant difference between homologues in the volume and depth occupied by the probe. EA loci had no significant accessibility differences between homologues⁷. In addition, examination of open chromatin marks present in EA and DA loci identified a measurable reduction in open chromatin marks in DA regions compared to EA regions⁷. The open chromatin marks measured were: deoxyribonuclease I hypersensitivity (DNase I HS), formaldehyde-assisted isolation of regulatory elements (FAIRE) analysis, histone 3 lysine 4 methylation (H3K4me), histone 3 lysine 9 acetylation (H3K9ac), histone 3 lysine 27 acetylation (H3K27ac), and histone 3 lysine 4 dimethylation (H3K4me2)⁷. Overall, the levels of all open chromatin marks in DA intervals were significantly lower compared to EA intervals⁷.

The non-random and heritable localization preference of the more accessible (brighter) hybridization to one homologue over the other was determined using samples from individuals with homologues that could be distinguished from each other⁷. The observation of DA was determined to be independent of the presence or absence of the chromosomal heteromorphism or abnormality differentiating the homologues⁷. When all probe targets were considered, the more accessible allele tracked with one parental-derived homologue within an individual, independent of the presence or absence of the abnormality, though between individuals it could track with either the maternal or paternal homologue. There also appeared to be a preference to parental origin of the more accessible hybridization between generations in the same pedigree⁷. This is unlike genomic imprinted loci, specific loci in which the same single parent allele is preferentially transcribed while the other is silenced across individuals^{7,62}.

Another strong source of evidence for a biochemical basis for DA came from treating lymphoblast cell lines with different reagents that targeted chromosome condensation through different epigenetic factors⁸. Only inhibition of topoisomerase II α with ICRF-193, an inhibitor of the ATPase activity that stabilizes the enzyme in an ineffective closed conformation, affected the presence of DA at a given probe locus^{8,63}. The inhibition of topoisomerase II α implicated differences in chromatin catenation levels between homologues as the basis for DA, with a decreased amount of chromatin supercoiling of both homologues when inhibited, equalizing the accessibility between homologues. The effects of topoisomerase II α were confirmed by 3D-SIM and GVF⁸.

All observation and characterization of DA was completed in human metaphase chromosomes derived from lymphocyte and lymphoblast samples^{7,8}. However, other preliminary investigation in our laboratory suggested DA also occurred in fibroblasts.

Intriguingly, there has also been potential evidence of DA reported from a FISH gene mapping study recently conducted on chromosome 2 of metaphase onion root spreads⁶⁴. There were no repetitive elements within the probe⁶⁴. This suggests that DA may also be present in metaphase chromosomes of other organisms.

All the studies completed up to the present have shown DA to be a stable, non-random difference in chromatin catenation between homologues that is heritable between cell generations^{7,8}. This results in a difference in chromatin supercoiling between homologues creating the difference in accessibility observed. Current evidence supports a potential role for DA to distinguish between homologous regions during metaphase by accessibility of chromatin. We suggest that DA could be a manifestation of structural memory that contributes to maintaining information of interphase chromatin structure in these regions. In this way, a direct link can be made between the chromatin organization of parent and daughter cells during cell division^{7,8}. The ability to re-establish functional structures present in interphase but lost during metaphase is crucial to maintain correct gene regulation and cell functioning^{3,42}. The presence of DA has only been examined in peripheral lymphocytes and lymphoblasts with some suggestion it may be present in other tissues as well as the genomes of other organisms^{7,8,64}. The specific role of DA in chromosome condensation, mitotic memory or other possible nuclear functions, remains to be determined.

1.5 Study Objectives

The observation and characterization of DA implicates an exciting new avenue in the investigation into chromatin structure, leading to questions involving both the observed metaphase structure and the possible corresponding structures in interphase. The initial DA studies showed the source of the difference in hybridized DNA probe fluorescence intensity

between homologous metaphase chromosomes was due to differences in chromatin superhelicity at the same loci between homologues. This present study's objectives set out to address both expanding knowledge about how DA is distributed throughout the genome and if it is present in cell and tissue types besides differentiated lymphocytic cells. The previous studies were achieved using scFISH probes designed from genomic regions of clinical cytogenetic interest in which various pathological rearrangements were already known. These scFISH probes were widely dispersed throughout the genome, with most only confirming DA within the genomic coordinates of individual sc probe boundaries. This led to the question: Does DA expand beyond the boundaries of these scFISH probes occupying larger genomic areas i.e. what is the extent of each domain of DA chromatin?

In addition, all published data concerning DA were collected by analyzing chromosomes of peripheral lymphocytes and lymphoblasts from multiple individuals. This prompted the second research question: does DA occur in other tissue types or is it specifically present only in T-lymphocytes and Epstein-Barr virus (EBV) transformed lymphoblast cells? My hypotheses were:

- 1) Differential accessibility, DA, is a structural feature of metaphase chromatin that extends beyond the original scFISH intervals that were previously investigated in our laboratory. In analyzing intervals adjacent to confirmed DA regions the size of DA domains can be determined.
- 2) DA is unrelated to normal tissue specific chromatin structural programming and therefore will be conserved across the same loci in different cell and tissue types (lymphocytes, bone marrow, fibroblasts).

To address my hypotheses, 3 specific aims were identified:

Aim 1a: Examine genomic length and frequency of DA domains by developing scFISH probes between and adjacent to confirmed DA regions to map size and frequency of DA domains.

Aim 1b: Examine potential interphase features related to DA in metaphase including epigenetic marks present in intervals exhibiting DA compared to regions demonstrating EA using publicly available epigenetic data.

Aim 2: Determine if DA is present and maintained at the same loci across different cell and tissue types using scFISH techniques. Cell types examined will include bone marrow samples to examine lymphocytes at various stages of differentiation and skin derived fibroblasts to examine cells from a different germ layer.

It is becoming more and more apparent that gathering knowledge of chromatin structure and its organization through time and space is essential to properly understanding gene regulation and normal cell function¹. This includes how it is maintained through multiple cell generations. DA is a new structural feature of metaphase chromatin of which little is known, including what its role is during cell division. Further characterizing the way in which DA is distributed in the genome as well as the different cell types in which DA is found, will advance knowledge of *in situ* metaphase chromosome structure at high resolution. The approach proposed here may also indicate possible connections to chromatin organization in interphase. Each of these are key elements required to address the broad question regarding the purpose of a difference in accessibility between metaphase homologues.

Chapter 2: Materials and Methods

2.1 Design and Development of Single-copy DNA FISH probes

Single-copy (sc) DNA sequence intervals from the human genome with either no repetitive elements or highly divergent repetitive sequences (>20%), were selected to develop scFISH probes as previously described^{9,65}. Each sc interval was precisely defined by specific human genome coordinates and ranged in length from ~1.4 – 4 kilobases (kb) in length. The area to which an scFISH probe hybridizes is interchangeably referred to as a locus, region or interval. To develop a probe, the sc interval needed to be amplified from human genomic DNA with polymerase chain reactions (PCR) optimized for long products, the amplicon purified, and labelled with a modified nucleotide prior to performing scFISH on metaphase chromosomes. The labelled probe was detected with a fluorescent antibody during scFISH.

2.1.1 Oligonucleotide Primer Design and Amplicon Production

Primer pairs for each selected sc interval were designed using Primer-BLAST⁶⁶. Sc intervals were identified using both RepeatMasker (UCSC Genome Browser) and CytoVA (CytoGnomix Inc, London ON). The DNA sequence (hg19/GRCh37 genome assembly) for the full sc interval was the PCR template used to generate all primer pair options. The sequence was obtained from the UCSC Genome Browser⁶⁷. Generally, 15-20 primer pairs were designed for each sc interval. The maximum length of the PCR product was the length of the sc interval in base pairs (bp) and a minimum length of 200-500 bp less than the maximum. The selected primer melting temperature (T_m) range was 58.0°C - 65.0°C, with an optimal T_m of 62.0°C. The maximum T_m difference between a pair of primers was limited to 2°C. Primer pair specificity was verified using the “RefSeq representation genome” database for alignment with the human genome by BLAST® (Basic Local Alignment Search Tool)⁶⁸. The coordinates of

the primer pairs were reported from the hg38 genome assembly, therefore the “view in other genomes” option in the UCSC genome browser was used to convert coordinates to the hg19 assembly. The primer pair that was selected minimized the self-complementarity of individual primers in the pair as well as the T_m difference between the pair. Primers in which the PCR product had unintended targets were avoided as well as those beyond the 40 – 60% GC content range. Longer primers were preferred (>25 bp).

Common copy number variants were also excluded from the intervals to rule out any downstream fluorescence intensity differences between homologous regions produced by common chromosomal polymorphisms ($\geq 1\%$ of the population), such as deletions and gains, following the protocol from Khan *et al.* 2014⁷. Independent microarray datasets, Ontario Population Genomics Platforms (n= 873 individuals of European ancestry; minimum 25 probes per CNV, Database of Genomic Variants), and Healthy sample set (n = ~400 individuals; minimum 35 probes per CNV, Affymetrix, Inc), were used to identify common CNVs using ChAS (Chromosome Analysis Suite) software on ThermoFisher (formally Affymetrix, Inc) CytoScan HD array. From each single-copy interval lacking common CNVs, DNA probes were developed and produced to both define DA domains and examine presence of DA between tissues.

Primer pairs were then synthesized by Integrated DNA Technologies, Inc (Toronto, ON). Amplicons of unlabeled sc probe were produced by scaling up optimized PCR reactions for long products by 26x (fold). A 1x reaction includes the following: 3.25 μL each of the forward and reverse primers (2 μM), 1 μL of 30 ng/ μL genomic DNA (Promega Corporation, Madison WI) with 5 μL of hot start DNA polymerase (Kappa HiFi; Roche) and bringing the volume to 25 μL with autoclaved nano water. A long PCR program was run on a gradient PCR

thermocycler (Eppendorf vapo.protectTM Hamburg, Germany). The extension temperature was optimized for each probe, temperatures ranged from 63°C - 72°C. The time added to each extension for cycles 15-33 was also optimized. The general PCR parameters used to produce the amplicon of all sc intervals are given in Table 2.1.

Table 2.1: PCR cycling parameters for single-copy probe amplification using long PCR.

Amplification Stage	Temperature	Time
Initial Denaturation	94 °C	4 min
Cycle 1-14		
Denaturation	94°C	20 sec
Extension	Optimized for each probe (62 °C -72 °C)	5 min
Cycle 15-33		
Denaturation	98°C	20 sec
Extension	Optimized for each probe (62°C -72°C)	5 min with a +15 sec or +30 sec extension added to each cycle. Optimized for each probe
Annealing	72°C	10 min

A 0.85% (w/v) agarose gel in 1x TAE buffer (Tris-acetate-EDTA buffer) was used to check the result of the amplification to ensure that the amplicon produced was uniform and of the correct size. For those amplifications that synthesized the expected product, amplification was followed by ethanol precipitation of the amplicons in a volume of 3 M sodium acetate equal to 0.1 x PCR reaction volume and a volume of cold anhydrous ethanol equal to 2.5 x PCR reaction volume. Amplicon was precipitated for a minimum of 12 hours at -20°C. The GeneAid Gel/ DNA PCR extraction kit (New Taipei, Taiwan) was used to purify the amplicon after separation of products by length using gel electrophoresis with 0.85% low melting point agarose gel (Avantor; A426-05) submerged in 1x TAE buffer. The final concentration of purified amplicon was measured using a NanoDrop 2000 spectrophotometer (ThermoFisher Scientific, Waltham MA). Figure 2.1 shows an example of an optimal (left) and poor (right) probe amplicon purification gel. The two lanes for TPM1_F2294 show the desired high concentration of amplicon (identified by size) with minimal banding above and below the band. This is compared to an amplification of 3.3_1p36, a probe designed for a known EA region, with smearing visible below the band of amplicon. Extraneous DNA overlaying the desired amplicon can be a potential source of background in FISH and complicate or prevent analysis. This probe was later re-optimized to eliminate the undesired amplification products causing the smear.

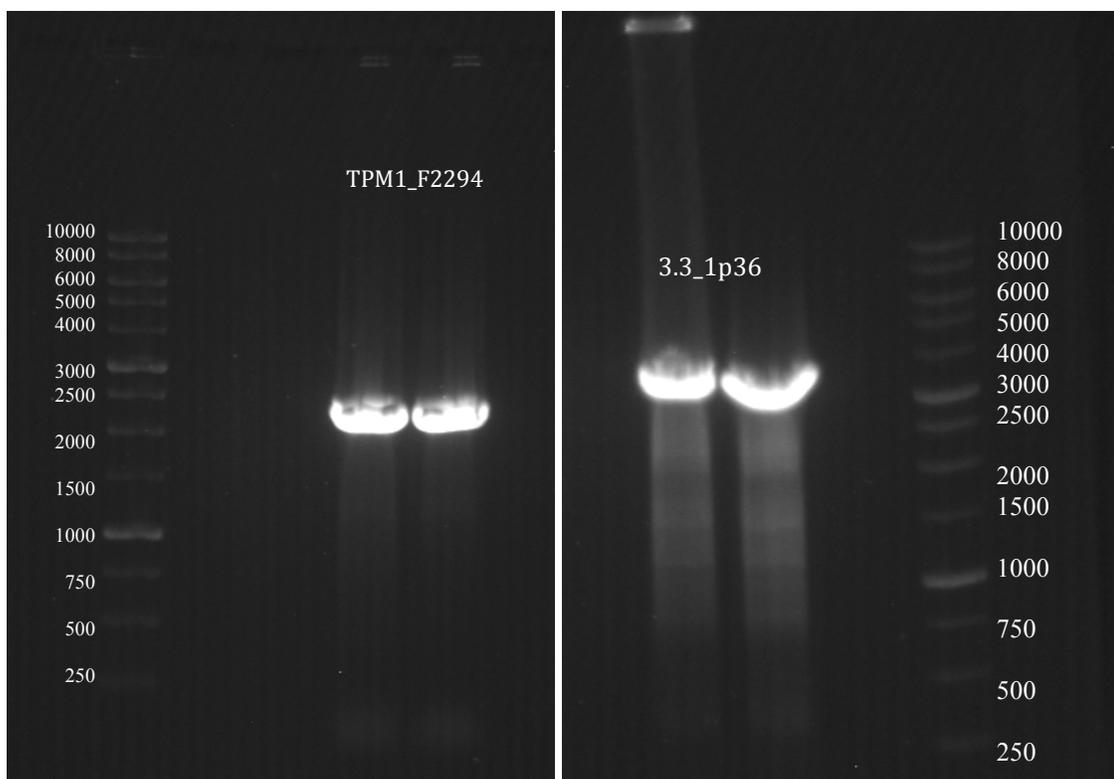


Figure 2.1: Gel image example of an optimal and poor scFISH probe amplification produced in this study. TPM1_F2294 (left) provides an example of an optimal amplification, 2 lanes with bands of high probe concentration, with little smearing above and below band of probe. 3.3_1p36 (right), is an example of a probe from this study that was re-optimized due to smearing below the desired amplicon. Extraneous amplifications pose a risk for non-target hybridizations causing high background and inhibiting analysis of metaphase chromosomes. Amplicon was run on a 0.85% TAE agarose gel in 1x TAE buffer. Gel is stained using Invitrogen™ SYBR™ safe (ThermoFisher) and visualized using UV light. Ladder is 1Kb RTU from Froggabio (Toronto, ON) and size units are in bp.

2.1.2 Labeling of Purified Amplicon by Nick Translation

All amplicons were labeled with digoxigenin-11-dUTP (Roche Diagnostics, Indianapolis, IN) by nick translation as described by Knoll and Lichter in Current Protocols in Molecular Biology⁶⁹. Nick translations were performed at 15°C. The nick translation reaction to label 1 µg of purified amplicon (5-20 µL) in a final volume of 100 µL was: 10 µL of a nick translation solution [working solution of 0.2 mM of dATP, dCTP, and dGTP, 0.1 mM of dTTP (Roche Diagnostics), 0.5 M Tris-HCl pH 7.8, 2.86 M 2-mercaptoethanol (Millipore-Sigma), 0.1 µg/mL bovine serum albumin (Roche Diagnostics) and varied concentrations of MgCl₂ dependent on the DNA polymerase I manufacturer], 4 µL of DNA polymerase I (5 units/µL), 5 µL of 100 mM digoxigenin -dUTP (Roche Diagnostics) with 5 µL of a 5:1000 5 µg/µL DNase I dilution (Worthington Biochemical, Lakewood, NJ), and sterile high quality water to the final volume of 100 µL⁶⁹. The MgCl₂ concentration within the nick translation solution varied depending on the DNA polymerase I storage buffer composition of the manufacturer. DNA polymerase I from the Promega Corporation required 0.2 M MgCl₂ in the final 100 µL nick translation reaction volume whereas DNA polymerase I from Roche Diagnostics required 0.05 M MgCl₂ for optimal activity. The DNase I dilution was optimized for each new batch received from the manufacturer.

The enzymes for the nick translation reactions were calibrated so that the amplicon was digested to 250-750 bp in length over 150 minutes (Figure 2.2). After digestion, a 15 µL aliquot of the reaction was removed to check the size of the nicked products by gel electrophoresis. Before loading the reaction onto the gel, the reaction was stopped by adding 1.5 µL of 0.5 M EDTA, pH 8 and incubated at 70°C for 5 minutes. The remainder of the reaction was stored at -80°C until the size was confirmed. If the size was optimal, the reaction was stopped with 2

μL of 0.5M EDTA, pH 8 and incubated at 70 °C for 5 minutes as done with the test aliquot. If the probe was too large, the reaction was placed at 15 °C for additional time and/or additional enzymes were added. After labelling, the probe was ethanol precipitated (same method described for the precipitation of amplicon), lyophilized to dryness and reconstituted in 10 μL of sterile nanopure water for a final concentration of 0.085 $\mu\text{g}/\mu\text{L}$. Labelled amplicon was stored at -20 °C until use for fluorescence *in situ* hybridization (FISH).

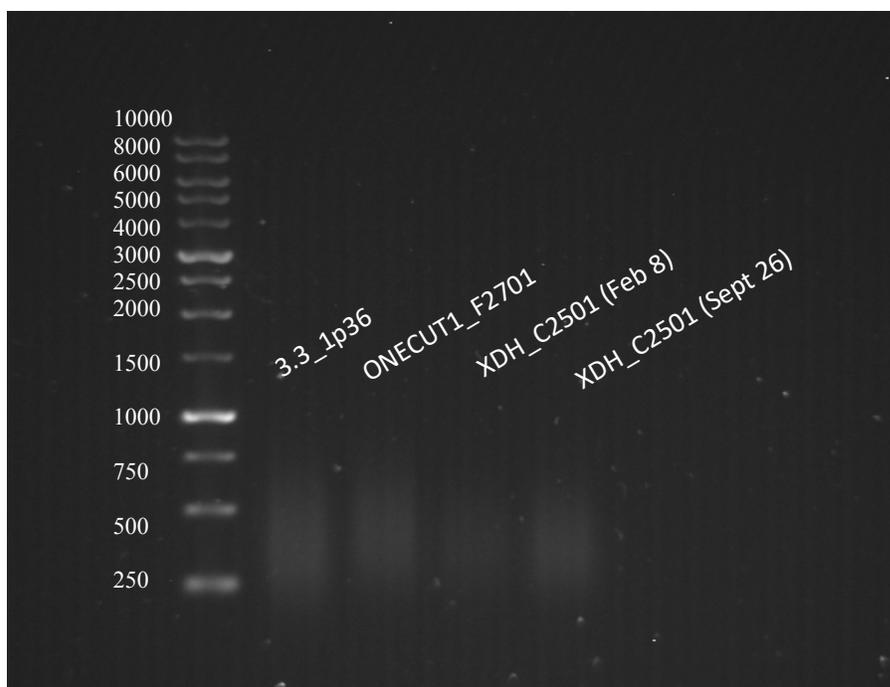


Figure 2.2: Example image of check gel to assess length of nick translation products of 3.3_1p36, ONECUT1_F2701, and XDH_C2501 probes. All lanes with labeled probe show the desired length of probe between 250 – 750 bp. The labeled amplicon was run on a 0.85% TAE agarose gel in 1x TAE buffer. Gel is stained using Invitrogen™ SYBR™ safe (ThermoFisher) and visualized using UV light. The ladder, on the left, is 1Kb RTU from Froggabio (Toronto, ON), and the size units are in bp.

2.1.3 Probe Nomenclature System

Nomenclature for probes under validation followed a different naming convention than fully validated probes. Multiple single-copy intervals were identified within large suspected DA regions detected by a literature review of historical FISH gene mapping studies predating completion of the human genome project. Therefore, new probes were named by the gene target of the mapping study and the letter corresponding to the individual single-copy interval within the suspected DA region with the length in base pairs of the probe separated by an underscore (ex. XDH_C2501). Gel images of probes maintain this naming convention as probes were still undergoing optimization and validation protocols.

Validated probes were given names more descriptive of their individual position within the genome. They were named as follows:

Intergenic: *Coding gene closest to probe interval* (longest RefSeq isoform)_ position of probe relative to gene (centromere[cen] or telomere[tel]) plus distance in base pairs between nearest boundary of the gene and that of the probe (ex. *RBM38_tel25076*).

Within gene: *Gene* _ interval of exons and introns spanned by the probe following Human Genome Variation Society (HGVS) conventions⁷⁰ (ex. *TPMI_IVS5-IVS8*).

2.2 Probe Selection to Determine DA Domain Size

The DA regions selected to address the genomic length and frequency of DA intervals and to classify DA domains were identified during development of a systematic method for identifying new regions of DA across the genome rather than focussing on regions relevant to clinical syndromes^{7,9,10,71}. This was achieved by a PubMed search and review of images in

publications with human genes or genomic sequences mapped to chromosomal bands on metaphase chromosomes using FISH⁷¹. Publications were not selected based on areas of clinical significance but rather by gene mapping studies with published FISH images. These historical gene mapping studies contributed to the assembly and completion of the human genome sequence.

The published FISH images were examined for visual differences in fluorescence intensity of hybridized probes between homologues. Those probes that appeared different were categorized as potential DA regions for further characterization in the laboratory⁷¹. These probes used in the historical gene mapping studies⁷²⁻⁷⁷ were large human genomic sequences that ranged in size from 50kb to several hundred kb, cloned into vectors such as cosmids or BACs. The precise genomic sequences of these large probes were not known. In contrast, scFISH probes have much shorter targets, generally ranging in length from 1.2- 4.0 kb, their precise genomic sequence and coordinates within the human genome are known and are not generally cloned. To confirm if the suspected DA regions were true DA or EA, the candidate region was established using the gene target of the historical probe as a genomic anchor and extending both upstream and downstream of the anchor by the reported length of the historical probe. This produced the overall genomic region from which scFISH probes were developed using the method outlined in Section 2.1.

Multiple single-copy intervals were identified across each candidate genomic region and probes were developed as described above. Candidate genomic regions with a successful hybridization of 2 different single-copy intervals, both of which demonstrated DA, were selected to be investigated further. Interpreted as contiguous DA domains, these are areas of the genome where multiple DA intervals are identified in close proximity within the same

region. Single-copy FISH probes were designed adjacent to the confirmed DA loci. The identified domains were XDH (2p23), HMGB1P5 (3p24), FGF6 (12p13), TPM1 (15q22), COX5A (15q25), and HMGB1P1 (20q13). Domains were named using the target of the original gene mapping study that identified the candidate region.

2.3 Probe Selection to Examine if DA Occurs in Different Tissues

A set of probes that demonstrated DA in lymphocyte or lymphoblast cells were selected to determine if DA was conserved across bone marrow and fibroblasts. All probes representing both DA and EA intervals were selected from our archive of previously validated probes. DA intervals present within genes (intronic and exonic) as well as in intergenic intervals, were selected from to establish DA across different tissues in both gene coding and noncoding intervals. To avoid confounding factors such as differential tissue expression, sc probes within genes were selected with no to low expression (0.0 – 5.0 transcripts per million [TPM]) across all tissues of interest (lymphocytes/blasts, bone marrow, fibroblast).

Due to the large file sizes generated from whole transcriptome sequencing, the identification of expression values from only specific genes with known DA and EA intervals and only within selected tissues of interest was achieved by a Python script (Appendix III) written to accelerate the search. The script searches publicly available RNA expression data generated across different tissues and reports mean expression values for selected genes. The script was run separately for DA and EA intervals within genes. Expression data were downloaded from the Genotype-Tissue Expression (GTEx)⁷⁸ and Human Protein Atlas^{79,80} databases. Expression data were reported in transcripts per million (TPM). Data available from GTEx provided expression values for EBV transformed lymphocytes and fibroblasts from multiple samples representing each tissue; the script calculated the mean and standard

deviation across these samples⁷⁸. Results for bone marrow were obtained using data from Human Protein Atlas of which data were reported as mean expression values for each gene in each tissue^{79,80}.

DA probes were selected according to the following criteria:

- a) localized within a gene
- b) gene showed no/low expression (0.0–5.0 TPM) in the different cell types analyzed
- c) probe hybridized well in lymphocytes (low background, good hybridization intensity/efficiency)

Two DA probes in intergenic regions were chosen to determine if DA is also conserved in regions where there is no gene coding. One intergenic EA probe from Khan *et al.* 2014⁷ was selected to control for a locus where DA was not observed in lymphocytes. Following the above analysis, the probes selected were *XDH_IVS30-IVS27*, *PCK1_cen180-IVS6*, and *DUOX1_IVS1-IVS3*. These probes are within the genes *XDH*, *PCK1*, and *DUOX1* respectively. Intergenic DA regions selected were *TPMI_tel3200* and *CTCFL_cen34302*. Intergenic EA probe selected is 3.3_1p36.

2.4 Cell Culture and Preparation

Metaphase cells were identified by dropping cytogenetic methanol-acetic acid fixed cell pellets onto microscope slides. Fixed cell pellets were obtained from blood, bone marrow, and fibroblast samples. Fibroblast samples, specifically, were obtained from cultured cell lines. The cytogenetic samples were derived from de-identified residual cell pellets that remained after routine cytogenetic diagnostic procedures were completed at the London Health Sciences Center Clinical Cytogenetics Laboratory (University of Western Ontario Office of Research Ethics: Study #15345E; 5453). Cytogenetically normal cell pellets were used for bone marrow

samples. Fixed cell pellets were produced following routine cytogenetic protocols for cell culture and harvest^{81,82}. Specific preparation for metaphase cells included arresting cells in metaphase by inhibiting microtubule formation with colcemid followed by inducing cell swelling using a hypotonic solution of 0.075 M potassium chloride (KCl). Both lymphocytes and bone marrow cells were treated with colcemid for 30 minutes. Lymphocytes were treated with 150 μ L - 200 μ L and bone marrow treated with 200 μ L of colcemid (stock: 10 μ g/mL) when in 10mL of medium^{81,82}. Cells were then fixed in this state with a 3:1 methanol: glacial acetic acid solution (Carnoy's fixative) for preservation and storage. Lymphocytes were stimulated with phytohemagglutinin (PHA) 48-72 hours prior to arresting in metaphase⁸¹.

Cytogenetic preparations were also prepared from control fibroblast samples derived from epidermal biopsy⁸³. These samples were obtained by Dr. Knoll prior to arriving at the University of Western Ontario and stored in liquid nitrogen. Fibroblasts were grown in 8mL of DMEM – Dulbecco's Modified Eagle Medium (Gibco; 11960-044) supplemented with 15% heat-inactivated fetal bovine serum (FBS) (Hyclone; SH30396.03) and 1% penicillin/streptomycin (Hyclone; SV30010) in a T25 flask at 37°C/5% CO₂ until they reached ~70% confluence for metaphase cell harvest. Medium was replenished every 3-4 days and changed 24 hours prior to harvest.

Harvesting protocol followed The AGT Cytogenetics Laboratory Manual. Fibroblast cells were arrested in metaphase with Karyomax® colcemid (Gibco; 15212-012) for 4 hours at a ratio of 10 μ L of 10 μ g/mL colcemid per 1 mL of medium. To maximize the number of mitotic cells, two cell fractions were collected: i) the mitotic fraction of rounded cells following colcemid treatment in the culture medium ii) the fraction of cells (metaphase and interphase) that remain attached to the tissue culture flask. The first fraction was collected by transferring

the culture medium into 15mL conical polypropylene tubes. The second fraction was collected after treating the remaining attached cells in the T25 flask with 4mL of trypsin-EDTA (0.05%), Phenol-red (Gibco; 25300-062) incubated at 37°C until cells were detached from the flask wall (~3 minutes). The trypsin was inactivated by adding an equal volume of DMEM medium supplemented with FBS. Each fraction was spun at 400 x g in a bench-top clinical centrifuge. The supernatant was removed, and the pellets combined in 15mL conical polypropylene tubes. Cells were resuspended in 5 mL of 0.075M hypotonic pre-warmed KCl solution followed by incubation at 37°C, for a total of 25 minutes (including resuspension and incubation times). Following incubation, a partial fixation of the cells was performed. Two mL of 3:1 methanol: glacial acetic acid solution was slowly added (6-12 drops a time with gentle mixing) to the 5mL of KCl solution with cells. This was followed by centrifugation, removal of the supernatant and two complete fixations of the cell pellet using the same fixative.

2.5 Single-copy Fluorescence *in situ* Hybridization (ScFISH)

The single-copy probes selected and prepared as described in Section 2.1 were used in fluorescence *in situ* hybridization where sc probes were hybridized to metaphase cells and detected with a fluorescently labeled antibody. The FISH protocols from Knoll and Lichter, and Rogan *et al.* 2001 were followed with some modification to washing steps^{9,69}. Cell preparations with spread metaphase cells on glass microscope slides were denatured at 70°C in 70% deionized formamide/2X saline-sodium citrate (SSC) solution for 2 minutes. Slides with denatured cells were then immediately dehydrated for 2 minutes each in 70% ethanol (on ice), 80%, 95% and 100% ethanol. Slides were then left to air dry at room temperature until sc probe hybridization solution was prepared.

For one slide, 150-200 ng of dig-11-dUTP labeled scDNA probe, 11 μ L of deionized formamide (BioBasics; FB0211) and 1 μ L of C₀t-1 DNA (1 μ g/ μ L; Roche Diagnostics) were combined in a 0.5 mL sterile microcentrifuge tube and denatured at 70°C for 5 minutes. The probe mixture was placed at 39°C for 30 minutes then combined with an equal volume of hybridization buffer solution (working solution: 10% w/v sterile dextran sulfate, 2mg/mL nuclease free bovine serum albumin (BSA) (Roche Diagnostics) and 4x SSC). The complete volume (~22 μ L) was applied to a 22mm² region of cells on each slide and covered with a 22mm x 22mm plastic coverslip. The slides were sealed in a parafilm pillow and placed in a dry incubator at 38°C overnight to allow hybridization of the probe with the chromosomal DNA. The following day, the slides were washed in 3 solutions sequentially for 30 minutes each: in 50% formamide/ 2X SSC (39°C), 1X SSC (39°C) and 0.5X SSC (room temperature) solutions. The first two washes were shaken manually at 10-minute intervals and the third at room temperature was agitated on an orbital shaker for 30 minutes at 100-200 rotations per minute. The post-hybridization washes removed non-specific hybridizations and residual hybridization solution.

The hybridized probe was then detected using Cy3-conjugated antidigoxin, an IgG monoclonal mouse antibody (Jackson ImmunoResearch, West Grove, PA), diluted 1:200 (v/v) in detection buffer (4xSSC with 1% BSA [filtered Cohn V. Fraction]). The hybridized region (22mm²) on the slides were incubated in the dark with 50 μ L of Cy3-conjugated antibody for 45-60 minutes at room temperature. Three post-detection washes followed, consisting of 0.5X SSC, 0.5X SSC/0.3% Triton-X 100 (Millipore-Sigma), and 0.5X SSC for 20-30 minutes each on an orbital shaker. The cells were stained with 50 μ L of 4',6-diamidino-2-phenylindole (DAPI, 0.1 μ g/mL phosphate buffered saline), in the dark for 20-40 minutes, followed by a

brief rinse in McIlvaine buffer (0.1M citric acid, 0.2M disodium phosphate; pH 7.2). Slide preparations were then mounted in 7 μ L of antifade⁸⁴ (1mg 1,4 phenylenediamine per 1mL of sterile glycerol [FisherBioreagents; BP229-1]), covered with a 22mm x 22mm No.0 glass coverslip and then sealed with nail polish. Cells were imaged using a Zeiss epifluorescence microscope system and analyzed as described in the next section.

2.6 Scoring of Hybridized Probe Fluorescence Intensity Differences Between Metaphase Homologues

An AxioImager Z.2 epifluorescence microscope system (ZEISS; Thornwood, NY) with 3 optical filters (blue [DAPI], green [FITC], red [customized for Cy3]) operating with Metasystems software, Metafer4 (V3.8.12), was used to capture metaphase images of the labelled probe and stained chromosomes at different colour wavelengths. The filters were used to capture DAPI blue fluorescence to visualize the staining of chromosomes and their specific banding patterns that allow chromosome identification and Cy3 red fluorescence to capture the Cy3 antibody labelled probes. For cell imaging, DAPI fluorescence was set for capture with automatic exposure control while Cy3 fluorescence was captured with a fixed exposure time. A low power scan of each slide was completed using a 10X objective (Zeiss Plan Apochromat 10X/0.45), and a subset of metaphase cells were selected by the analyst and imaged at higher magnification (Zeiss Plan-Apochromat 63X/1.4 Oil). Metasystems Metafer software was used to collect metaphase images. Metasystems Isis (V 5.3) package was used for manual image review and analysis.

Evaluation of differences in the fluorescence intensity between homologues was completed following previously reported methods of identifying DA^{7,8}. After identification of the chromosome and chromosome band to which a specific probe was designed to bind, the

fluorescence intensity of the signal on each homolog was scored on an empirical scale of nil/0, dim, medium, bright, and very bright. Chromosome identification and scoring of fluorescence signals was completed independently by a minimum of 2 analysts. A metaphase cell was considered to show differential accessibility (DA) if homologues were scored with different intensities (ex. Bright/ medium, bright/dim medium/dim). A cell was scored as equivalently accessible (EA) when homologues were scored with equivalent intensities (ex. Bright/bright, medium/ medium). Any scores of dim/dim, nil/nil, or dim/nil were excluded to avoid inefficient probe hybridization bias. Hybridized chromosomes involved in overlap at or near the location of probe hybridization were also excluded as physical interference in hybridization could also result in a difference in fluorescence intensity not derived from accessibility of chromosome structures. For most samples, 25-75 cells were scored, and a minimum of 2 samples were evaluated per scFISH probe. For a given DA interval, a two-tailed binomial test with normal approximation was used to determine if there was a significant difference between the proportion of cells identified with DA compared to the proportion of identified EA cells. As 2 different samples were hybridized per probe, a two proportion Z-test was also used to test if the proportion of DA or EA cells of a given probe differed between samples. A cell was confirmed to be DA or EA if there was a significant difference between the proportion of cells identified as DA compared to EA. Both statistical tests were performed at $\alpha = 0.05$.

The difference in fluorescence intensity was quantified using integrated gradient vector flow analysis (GVF) for a subset of probes^{7,61}. DA probes *XDH_IVS30-IVS27*, *ZNF385D_tel678016*, *DUOX1_IVS1-IVS3*, *TPMI_tel3200*, *PCK1_cen180-IVS6* and EA probe *3.3_1p36* were selected. GVF analysis used an algorithm previously developed in the laboratory to establish boundaries of probe signals and quantify the intensity of the signal

within that boundary⁶¹. Using the grey-scale image of the fluorescence signal, the integrated intensity value (pixels) was calculated within the active binary contour generated around the signal⁶¹. To determine if there was a difference between the signals on each homolog of a cell, a normalized intensity ratio was calculated by the following:

$$\text{Intensity Ratio} = \frac{|(\text{integrated intensity homolog 1} - \text{integrated intensity homolog 2})|}{(\text{integrated intensity homolog 1} + \text{integrated intensity homolog 2})}$$

Values approaching 0 indicate homologues with fluorescence signals of similar intensities and therefore equivalent accessibility whereas values approaching 1 indicate a difference in signal intensity indicative of DA⁷. A Mann-Whitney U test was used to determine if the intensity ratios differed between the DA probes developed for this study and EA probes.

2.7 Analysis of Epigenetic Characteristics Present During Interphase for Individual Probe Loci and DA Domains

The precise genomic coordinates of each DA and EA locus were known from the probe design process and genomic location confirmed by scFISH in metaphase chromosomes from T-lymphocytes. This allowed the comparison of specific genomic intervals with confirmed DA or EA classifications during metaphase with multiple interphase epigenetic characteristics including open chromatin marks and functional higher-order chromatin structures at the same loci. Epigenetic features characteristic of open chromatin were analyzed following the same methods as for previously reported DA and EA probes⁷. Open chromatin data were obtained from the Encyclopedia of DNA Elements (ENCODE)⁸⁵ for DNase I hypersensitivity (Duke DNase1 HS), Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) (University of North Carolina FAIRE seq) and histone marks H3K4me1, H3K9ac, H3K27ac, and H3K4me2 (Broad Institute histone modifications custom tracks). All open chromatin marks reported were derived from data collected from the EBV transformed lymphoblastoid cell line,

GM12878, in which DA had previously been characterized⁷. All histone modification data were derived from ChIP-seq (chromatin immunoprecipitation assay with sequencing) signal intensities for all probe intervals. The sum for each open chromatin mark was calculated for each interval for all probes, and a mean integrated intensity was calculated for DA and EA groups individually. Box and whisker plots (with the whisker limits determined by the Tukey method) of each mark for both EA and DA were plotted. Unpaired t tests with Welch correction for unequal variances were used to test for significant differences between the mean integrated intensity of each open chromatin mark between DA and EA intervals ($\alpha = 0.05$).

The 3-D genome browser⁸⁶ was used to analyze higher-order chromatin structures, topologically associated domains (TADs) as well as sub-TADs, present within DA intervals to compare the presence of defined DA domains in metaphase with functional chromatin organization in interphase. Chromatin capture data (Hi-C) collected from a genome wide study published in 2014²¹ were accessed using this database. Chromatin interaction frequency heatmaps were also generated using the database and additional genomic information was determined by integrating a UCSC (University of Santa Cruz) genome browser window⁶⁷. All data accessed were collected from experiments on the lymphoblast cell line, GM12878, with the GRCh37/Hg19 genome assembly. The data that were generated used the Lieberman-raw format at a resolution of 25kb. Correspondence of domains with TADs and other sub-TAD interactions was completed visually using scaled heat-map and genome browser outputs from the 3-D genome browser and UCSC Genome Browser respectively^{67,86}.

TAD boundaries are characterized by the clustering of insulator proteins. CTCF is one of the important insulators involved in defining the edges of each TAD, as well as aiding the formation of other higher-order chromatin structures in its role producing chromatin loops. To

investigate the clustering of CTCF protein, within and at the boundaries of TADs, an information theory-based approach developed by Lu *et al.* 2017⁸⁷ was used to predict CTCF binding sites within the boundaries and beyond the boundaries of each DA domain⁸⁷. Prediction of CTCF binding site loci and binding affinity were generated using information theory-based transcription factor binding site motifs within the defined genomic coordinates of each domain \pm 1Mb. As average TAD size is \sim 880kb, data generated for each domain reached into a minimum of the adjacent TADs²². Sites generated with binding affinities (R_i) equal to or greater than half the overall mean binding affinity (R_{sequence}) were reported.

Within each domain the sum of binding affinities of each CTCF binding site was calculated and normalized by dividing by the number of kilobases in each domain. Two different classifications of areas outside of DA domains were investigated to determine if there was a difference in strength of CTCF binding within domains compared to areas beyond their boundaries. These were areas of equal size within the same chromosome, one area selected beyond the boundaries of the DA domain but within the same TAD and one area selected in the adjacent TAD to the DA domain. A Kruskal-Wallis test was used to determine if there was a difference between the 3 groups.

Chapter 3: Results

3.1 Identification of New Regions of Differential Accessibility (DA)

3.1.1 New Sc Probes Developed and DA Confirmed by Qualitative ScFISH on Human Metaphase Chromosomes

All 18 sc probes (1459-3553bp) developed for this study were localized by FISH to the expected regions in metaphase chromosomes of PHA stimulated lymphocytes of peripheral blood (Primer details; Appendix IV). The probes ranged in size from 1459 to 3553 bp in length and mapped to chromosomes 2, 3, 12, 15, and 20. Three of the probes, *TPMI_IVS5-IVS8*, *SCAMP2_IVS7-IVS4*, and *COX5A_tel20100* showed cross-hybridization on non-targeted chromosomes. *SCAMP2_IVS7-IVS4* had consistent hybridization on the p arm/satellite of acrocentric chromosomes in addition to the clear hybridization of the target. *TPMI_IVS5-IVS8* had a cross hybridization at approximately 1q21 whereas *COX5A_tel20100* had nontarget hybridizations at the terminus of 1p and 6q in addition to a centromere hybridization on chromosome 12. It was somewhat surprising that the BLAST® did not find any areas of the human genome with highly similar sequences, and similarly that BLAT (BLAST-like alignment tool) from the UCSC Genome Browser also did not predict these areas of non-targeted hybridization. The cross-hybridizations of acrocentric short arms, centromeric, and heterochromatic regions may be related to the fact that the DNA sequence of these regions of the genome are known to be incomplete, however the autosomal regions of cross-hybridization were unexpected. Nevertheless, these cross-hybridizations did not interfere with the analysis of DA.

Eighteen of these previously uncharacterized sequence intervals exhibited DA in metaphase chromosomes. All probes developed for this study were shown to hybridize

according to the previously established criteria for either DA or EA patterns⁷. Probes demonstrating DA showed the characteristic difference in probe fluorescence intensities between homologous chromosomes, in contrast to one previously reported probe⁷ showing EA on both homologues. Figure 3.1 shows examples of metaphase cells with differences in fluorescence intensity between homologues for 3 DA probes (*SCAMP2_IVS7-IVS4*, *ZNF385D_tel678016*, *FGF6_IVS2*) and an example of hybridization of one cell with similar fluorescence intensity of EA probe 3.3_1p36. DA and EA regions differ significantly from each other by the proportion of cells with observed DA. When the majority of cells ($\geq 2/3$) have a DA hybridization pattern, the probe is designated as a DA probe and similarly when the majority of cells have an EA pattern, the probe is designated as an EA probe. A minority of cells in each sample exhibit the opposite hybridization pattern than the probe designation i.e. a DA probe has predominately DA hybridization patterns on metaphase cells with a small subset with EA patterns. We evaluated the consistency of this definition of DA or EA for all probes using a two-tailed binomial test with normal approximation that excluded the null hypothesis i.e. the equivalent hypothesis (Table 3.1). DA probes demonstrated a significantly higher proportion of DA cells compared to EA cells (p values $\leq 1.5E-04$) across 2 samples (Figure 3.2). EA probe, 3.3_1p36, was confirmed to have a significantly higher proportion of cells with EA hybridizations than DA hybridizations (p = $1.4E-04$, Table 3.1). This indicates that by qualitative scoring of probe fluorescence intensity differences between homologues, each probe developed for this study hybridized to a DA region in human T-lymphocytes.

Prior to developing the sc probes, the sc genomic intervals within the regions of interest were determined computationally to have no or rare copy number variants (CNVs) within the regions (Section 2.1.1). Intervals with common CNV gains or losses are quite common in

normal and abnormal genomes of humans as well as other species. These intervals were excluded from probe development so CNVs would not be the source of probe fluorescence intensity differences between homologues. The frequencies of CNV gains or losses within the sc probe genome intervals from this study are shown in Appendix V.

For each probe, metaphase cells from T-lymphocyte samples of 2 different individuals were hybridized. Cells were scored as DA or EA as outlined in Section 2.6 and the number of cells scored as DA and EA were totaled. Cell numbers examined for each probe hybridization varied between samples based on mitotic index. These data are presented in Table 3.1. A two proportion Z-test ($\alpha = 0.05$) demonstrated that there was no evidence of a significant statistical difference between the fraction of cells scored as DA between the different patient samples (n=23) used to identify the accessibility pattern of 17 of 18 probes (Table 3.1). One DA probe, *SCAMP2_IVS7-IVS4* showed a difference (p=0.02) between the fraction of cells with DA between samples, however both samples clearly showed DA (>2/3 DA). This may represent a stochastic difference or be individual related. For EA probe, 3.3_1p36, there was also no significant statistical difference found between patient samples in the proportion of EA cells scored for each individual (Table 3.1).

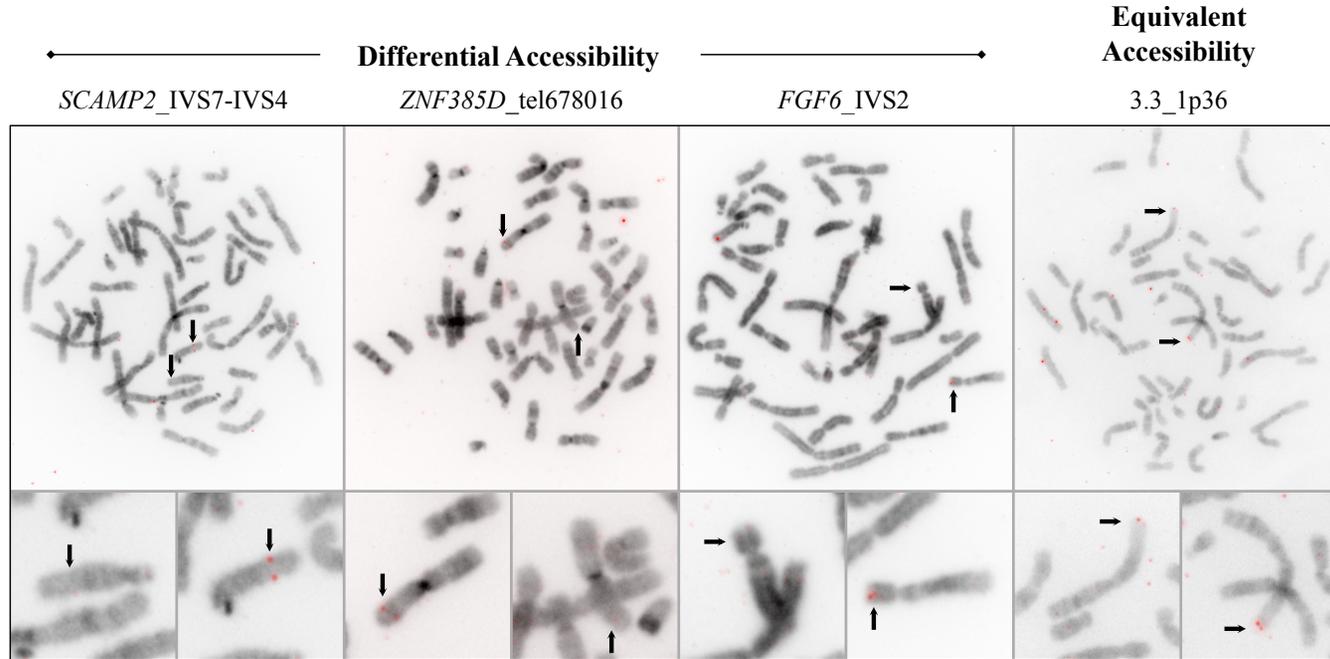


Figure 3.1: Differential and equivalent accessibility hybridization patterns between human metaphase homologues detected by single-copy probe fluorescence *in situ* hybridization. Chromosomes from single metaphase cells hybridized with single-copy FISH probes developed for the intervals *SCAMP2_IVS7-IVS4* (3002 bp; 15q24.1), *ZNF385D_tel678016* (2251 bp; 3p24.3) and *FGF6_IVS2* (3430 bp; 12p12.3) [left to right] show differential probe fluorescence hybridization between homologues. Differential probe fluorescence from left to right: 1 dim compared to 2 bright hybridizations, 1 bright and 1 dim compared to no hybridization, and 0 compared to 1 medium hybridization. Arrows indicate the expected location of probe hybridizations on each homologue in full metaphase and magnified images. ScFISH probe 3.3_1p36 (3354 bp; 1p36) shows similar fluorescence intensity (or equivalent accessibility [EA]) between homologues. Chromosomes were stained with DAPI and probes were labelled with digoxigenin-11-dUTP and detected with Cy3-digoxin antibody.

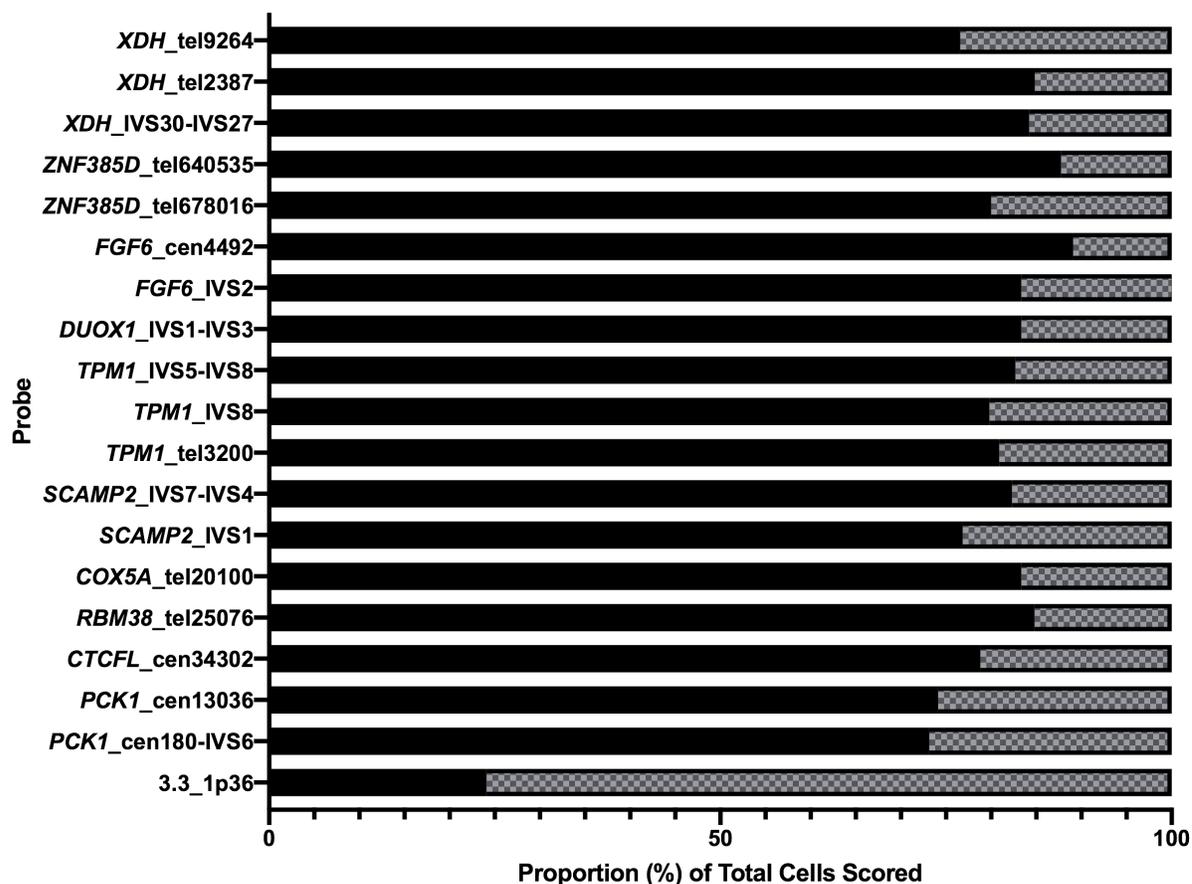


Figure 3.2: Frequency of cells exhibiting DA on homologous chromosome regions by sc probe FISH analysis. The black and grey shading represent the relative proportion of cells with DA and EA, respectively. Each row represents the results for each sc probe from samples of 2 individuals. All probes show DA with the exception of the EA probe, 3.3_1p36 (last row). All DA probes (n=18) had a statistically significant larger proportion (73-89%) of cells demonstrating DA compared to one EA probe 3.3_1p36, bottom row, (24%) which had a significantly larger proportion of EA. Significance was demonstrated using a two-tailed binomial test with normal approximation ($\alpha = 0.05$).

Table 3.1: ScFISH probes developed and validated to evaluate the extent of DA domains established with an anchor scFISH probe and conservation of DA between tissues. Location of each probe is indicated by chromosome band and genomic coordinates. The number of cells scored as DA and EA per sample is indicated along with the total number of DA and EA cells scored per interval for both samples. The p-value results are given from two proportion Z-test testing if there is a statistical difference between individuals and a two-tailed binomial test with normal approximation testing statistical difference between proportion of DA and EA cells scored overall for each probe.

Probe Name	Chromosome Band	Genomic Coordinates [GRCh37/hg19]	Length (bp)	Genomic Position	FISH Pattern	Sample*			p-value ($\alpha = 0.05$)	
						1	2	Total	Two proportion Z-test between samples	Two-tailed binomial test between DA and EA
Differential Accessibility										
<i>XDH_tel9264</i>	2p23.1	chr2:31,545,815-31,547,924	2110	Intergenic	DA	50	35	85	0.25	2.1E-08
					EA	12	14	26		
<i>XDH_tel2387</i>	2p23.1	chr2:31,551,816-31,554,801	2986	Intergenic	DA	24	23	56	0.26	1.5E-08
					EA	3	4	10		
<i>XDH_IVS30-IVS27</i> †	2p23.1	chr2:31,568,769-31,571,269	2501	<i>XDH</i>	DA	30	18	48	0.90	2.4E-07
					EA	7	2	9		
<i>ZNF385D_tel640535</i>	3p24.3	chr3:22,433,351-22,436,333	3968	Intergenic	DA	65	28	93	0.27	7.8E-15
					EA	11	2	13		
<i>ZNF385D_tel678016</i>	3p24.3	chr3:22,470,832-22,473,082	2251	Intergenic	DA	51	17	68	0.093	3.2E-08
					EA	13	4	17		
<i>FGF6_cen4492</i>	12p12.3	chr12:4,537,157-4,538,816	1660	Intergenic	DA	42	33	75	0.48	6.7E-09
					EA	6	9	15		
<i>FGF6_IVS2</i>	12p12.3	chr12:4,549776-4,553,205	3430	<i>FGF6</i>	DA	30	19	49	0.38	2.5E-10
					EA	3	3	6		
<i>DUOX1_IVS1-IVS3</i> †	15q21.1	chr15:45,422,890-45,424,597	1708	<i>DUOX1</i>	DA	45	50	95	0.67	1.1E-12
					EA	8	11	19		
<i>TPM1_IVS5-IVS8</i> ^			2408	<i>TPM1</i>	DA	30	13	43	0.49	2.4E-06

	15q22.2	chr15:63,353,573-63,355,980			EA	7	2	9		
<i>TPMI_IVS8</i>	15q22.2	chr15:63,357,346-63,360,645	3300	<i>TPMI</i>	DA	31	52	83	0.60	1.2E-09
					EA	6	15	21		
<i>TPMI_tel3200</i> †	15q22.2	chr15:63,367,314-63,369,607	2294	Intergenic	DA	19	36	55	0.80	3.5E-07
					EA	6	7	13		
<i>SCAMP2_IVS7-IVS4</i> ^	15q24.1	chr15:75,142,349-75,145,350	3002	<i>SCAMP2</i>	DA	49	44	93	0.025	6.5E-12
					EA	16	4	20		
<i>SCAMP2_IVS1</i>	15q24.1	chr15:75,161,783-75,163,308	1526	<i>SCAMP2</i>	DA	27	40	67	0.89	4.7E-07
					EA	10	10	20		
<i>COX5A_tel20100</i> ^	15q24.1	chr15:75,250,595-75,252,319	1725	Intergenic	DA	25	35	60	0.63	1.5E-08
					EA	3	9	12		
<i>RBM38_tel25076</i>	20q13.3	chr20:56,009,465-56,011,485	2021	Intergenic	DA	36	27	63	0.55	8.7E-06
					EA	8	14	22		
<i>CTCFL_cen34302</i> †	20q13.3	chr20:56,033,167-56,036,719	3553	Intergenic	DA	32	20	52	0.59	2.9E-06
					EA	6	8	14		
<i>PCK1_cen13036</i>	20q13.3	chr20:56,119,569-56,123,101	3533	Intergenic	DA	18	49	67	0.21	6.1E-10
					EA	3	9	12		
<i>PCK1_cen180-IVS6</i> †	20q13.3	chr20:56,135,957-56,139,048	3092	<i>PCK1</i>	DA	31	18	49	0.28	1.5E-04
					EA	12	6	18		
Equivalent Accessibility										
<i>3.3_1p36</i> †	1p36.3	chr1:1,171,789-1,175,143	3354	Intergenic	DA	3	10	13	0.55	1.4E-04
					EA	13	28	41		
* Samples 1 and 2 are from different individuals. The same samples were not used across all probes. Samples from 23 different individuals were used for hybridizations. A minimum of one sample is always from a male.										
† Probe selected to investigate conservation of DA between tissues										
^ Probe showed cross-hybridization with non-target chromosomes										

3.1.2 Quantification of DA by Gradient Vector Flow Analysis (GVF) in ScFISH Probes

Gradient vector flow (GVF) analysis of integrated probe intensities was used to validate the qualitative scoring of homologous loci^{7,61}. This method analyzed differences between the previously identified fluorescence intensity comparisons of the same sequences on different homologues (none/dim, medium, bright, very bright) in metaphase images. GVF quantified the fluorescence intensity of each probe hybridization over all pixels in each metaphase image. From these values, an integrated intensity ratio was calculated between homologues for a sample of cells for each probe within a subset to classify them as either DA or EA. Intensity ratio scores approaching or equal to 0.0 indicate equivalent (or similar) hybridization intensities between homologues whereas intensity ratio scores approaching or equal to 1.0 indicate a greater difference in probe hybridization intensities between homologues. Integrated intensity ratio formula is provided in section 2.6.

Hybridizations of 5 DA probes (*XDH_IVS30-IVS27*, *ZNF385D_tel678016*, *DUOX1_IVS1-IVS3*, *TPMI_tel3200*, *PCK1_cen180-IVS6*) and one EA probe (3.3_1p36) were analyzed by GVF using one pair of chromosome homologues per cell for 25 cells for each sc probe (DA n=125 diploid cells, EA n=26 diploid cells). Distributions of the normalized integrated intensity ratios of DA and EA regions were compared in a box and whisker plot (Figure 3.3). A significant difference ($p < 0.0001$) was determined between the median intensity ratio of DA (0.82) and EA regions (0.23) using a Mann-Whitney non-parametric test, after a non-normal distribution was determined with the Shapiro-Wilk test ($p < 0.0001$). This trend was consistent with previous characterization of published DA and EA regions⁷.

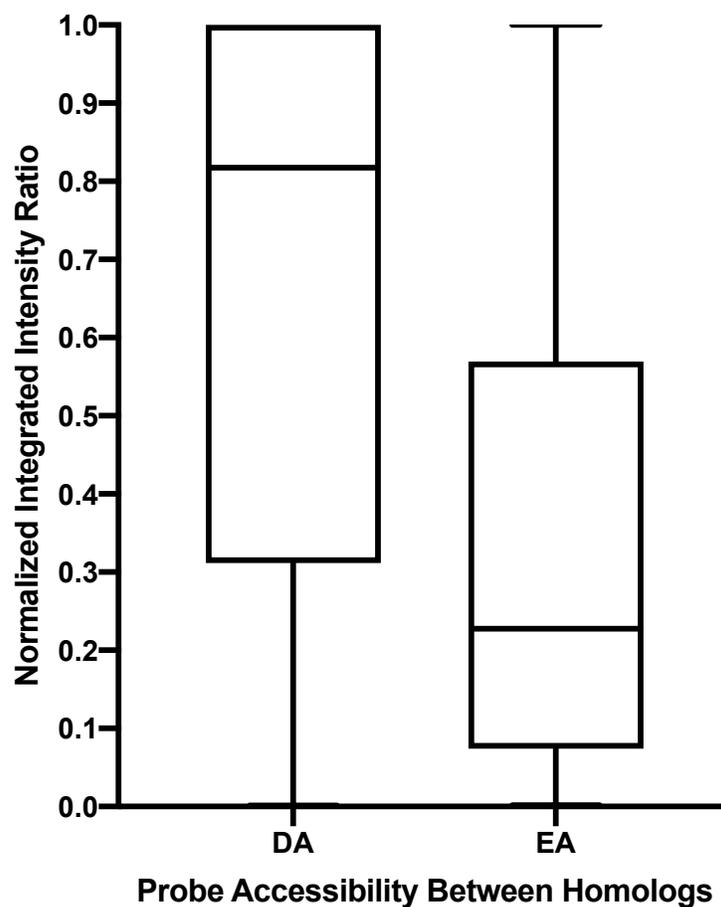


Figure 3.3: Quantification of probe signal fluorescence between homologues shown by box and whisker plots of normalized integrated fluorescence intensity ratios. The limits of the whiskers were determined using the Tukey method for box plots. Single-copy probes (*XDH_IVS30-IVS27*, *ZNF385D_tel678016*, *DUOX1_IVS1-IVS3*, *TPMI_tel3200*, *PCK1_cen180-IVS6*; n=125) detecting DA showed a large difference in hybridization intensities between homologues relative to the single-copy probe detecting EA (3.3_1p36; n=26). A significant difference was determined between the median integrated intensity ratio of DA (median = 0.82) and EA regions (median = 0.23) using a Mann-Whitney U test. The interquartile range for DA regions is 0.31 – 1.00 whereas that of the EA region is 0.07 – 0.57.

3.1.3 Comparing Open Chromatin Marks Between DA and Equivalent Accessibility (EA)

Regions

Previous investigation into connections between previously reported DA sequences and sequence specific chromatin accessibility marks present in interphase found that overall there was lower enrichment of 6 open chromatin marks in DA sequences compared to EA sequences. The same trend was observed between 17 of 18 new DA sequences from this study compared to the 59 EA sequences previously reported⁷. The profiles of the same 6 open chromatin marks (DNase 1 hypersensitivity (DNase 1 HS), Formaldehyde Assisted Isolation of Regulatory Elements (FAIRE), and histone modifications H3K4me, H3K9ac, H3K27ac, and H3K4me2) were compared between the new DA loci identified in this study (n=18) and the previously characterized EA loci⁷ (n=59). All DA probes were investigated using integrated intensity ENCODE data⁸⁵ from interphase cells of EBV transformed lymphoblastoid cell line GM12878. Using a box and whisker plot, a single outlier (red dot) can be observed in 5 of the 6 open chromatin marks. These points were all derived from a single DA interval, *SCAMP2_IVS1* (Figure 3.4). Only enrichment of H3K4me at this locus was within the boundaries of the whiskers. There is a pronounced enrichment of open chromatin marks at *SCAMP2_IVS1* compared to other DA loci, both identified in this study and those previously published⁷. Figure 3.5 shows an example of open chromatin mark enrichment of H3K27ac at *SCAMP2_IVS1* relative to the 2 neighbouring DA regions, *SCAMP2_IVS7-IVS4* and *COX5A_tel20100*. Due to this clear single point deviation from other DA loci, *SCAMP2_IVS1* open chromatin mark data were removed from the other DA interval data sets prior to statistical testing between DA and EA loci.

Overall lower mean integrated intensities of all 6 open chromatin marks (DNase 1 HS, FAIRE, H3K4me, H3K9ac, H3K27ac, and H3K4me2), were observed in 17 DA regions relative to the same open chromatin marks in 59 EA regions (Figure 3.6). This is consistent with the trend observed in previous work⁷.

An unpaired t-test with Welch's correction for unequal variances identified a significantly lower integrated intensity of DA intervals compared to EA intervals for open chromatin marks DNase I HS ($p = 0.0006$), H3K9ac ($p=0.0042$), H3K27ac ($p=0.0025$), and H3K4me2 ($p=0.0070$). No significant difference was identified between the FAIRE ($p=0.3232$) and H3K4me ($p=0.1727$) marks present at DA and EA intervals (Figure 3.6). Individual values for each open chromatin mark across all DA and EA regions investigated in this study are found in Table 3.2.

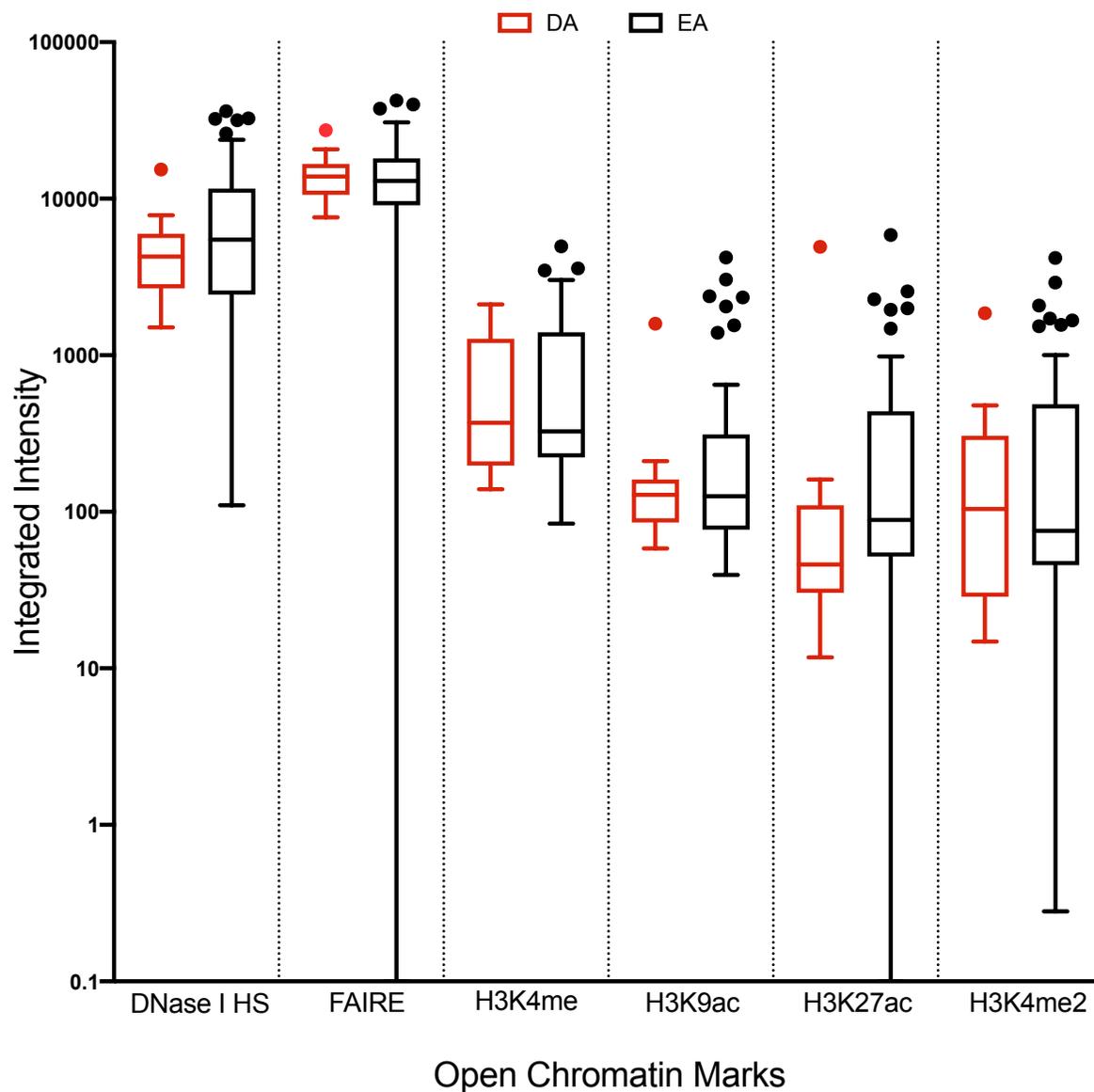


Figure 3.4: Distribution of integrated intensity data for each open chromatin mark in new DA intervals (n=18) and previously reported EA intervals⁷ (n=59). Data are presented in a box and whisker plot with the limits of each whisker determined by Tukey for each open chromatin mark (x-axis). Center line of each box represents the median. Outliers are represented by dots beyond the limits of the whiskers of each box plot. A single outlier from the DA group identified in 5 of 6 open chromatin marks was derived from the same interval, *SCAMP2_IVS1*.

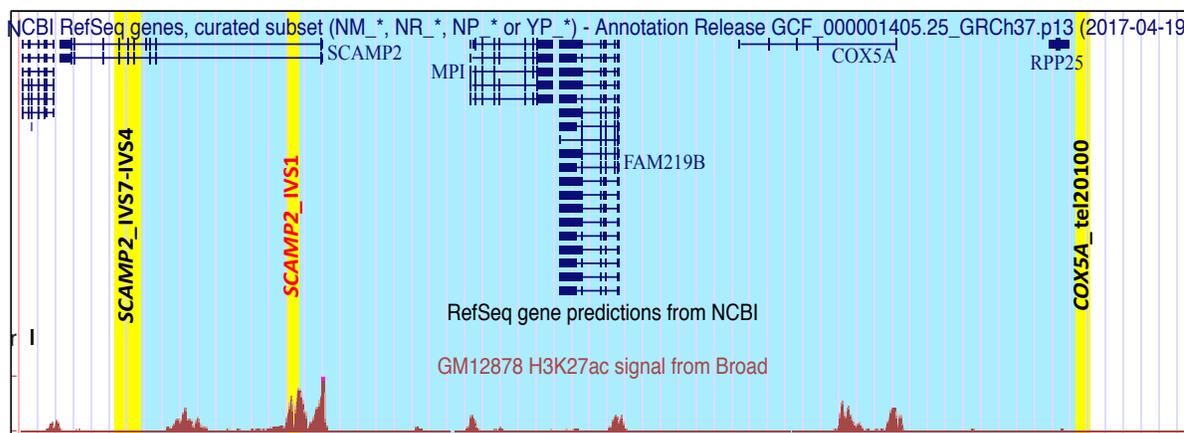


Figure 3.5: Genomic map demonstrating the difference in enrichment of open chromatin mark H3K27ac at the *SCAMP2_IVS1* DA probe outlier, compared to 2 neighbouring DA probe loci *SCAMP2_IVS7-IVS4* and *COX5A_tel20100*. H3K27ac (burgundy) is enriched at the *SCAMP2_IVS1* locus compared to all other DA loci identified in this study and those previously reported. The image was produced using the UCSC genome browser with the hg19 genome assembly. RefSeq genes with isoforms are in dark blue. H3K27ac signal (burgundy) is ChIP-seq data from the the GM12878 lymphoblastoid cell line (Broad Institute). Yellow bars show the locations of 3 sc probes with DA loci within chromosomes region 15q24.1. Light blue indicates the domain defined by these probes (to be discussed in section 3.2).

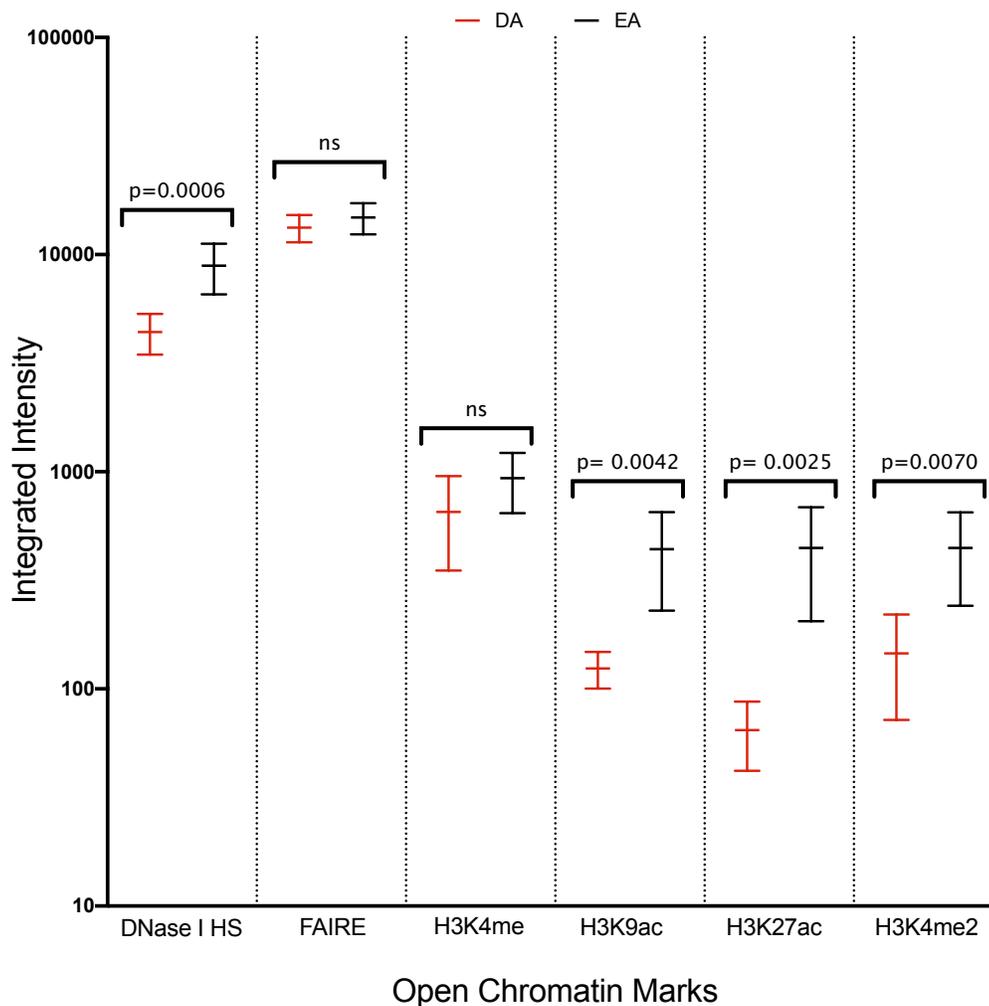


Figure 3.6: Open chromatin marks at DA loci have lower mean integrated intensities compared to EA loci. Integrated intensity values of DA regions were significantly lower than EA regions of DNase I HS, H3K9ac, H3K27ac, and H3K4me2 using an unpaired t-test with Welch's correction for unequal variances. No significant difference was found between the mean integrated intensity values of DA and EA regions for FAIRE and H3K4me ($p > 0.05$). The 95% confidence intervals for the DA ($n = 17$, excluding *SCAMP2_IVS1*) and previously reported EA intervals ($n=59$)⁷ are shown.

Table 3.2: Integrated intensity values of open chromatin marks in each DA and EA interval from this study. Present in interphase, integrated intensity values were retrieved from the ENCODE project⁸⁵ as averaged replications from the GM12878 cell line over each DA and EA interval identified in metaphase.

Probe Name	Total Integrated Intensity					
	DNase1 HS ¹	FAIRE ²	H3K4me	H3K9ac	H3K27ac	H3K4me2
<i>XDH_tel</i> 9264	5797	16484	1397.64	121.64	96.2	403.24
<i>XDH_tel</i> 2387	4172	14177	200.12	181.32	53.16	51.8
<i>XDH_IVS</i> 30-IVS27	2685	7588	139.4	66.6	30.64	22.2
<i>ZNF385D_tel</i> 640535	7757	16562	181.24	68.88	29.6	22.2
<i>ZNF385D_tel</i> 678016	1508	12601	214.96	88.8	14.6	37.0
<i>FGF6_cen</i> 4492	2201	8667	189.92	58.32	29.6	143.12
<i>FGF6_IVS</i> 2	5924	13874	712.68	132.8	29.6	143.12
<i>DUOX1_IVS</i> 1-IVS3	3906	9097	229.48	75	63.44	25.96
<i>TPMI_IVS</i> 5-IVS8	2592	12419	168.68	94.12	42.72	29.6
<i>TPMI_IVS</i> 8	4115	18354	373.2	137.96	47.56	115.88
<i>TPMI_tel</i> 3200	4316	13824	1602.12	154.16	160.56	478.4
<i>SCAMP2_IVS</i> 7-IVS4	2652	11393	918.52	210.28	150.28	111.2
<i>SCAMP2_IVS</i> 1 [†]	15355	27321	2108.4	1589.36	4924.44	1856.52
<i>COX5A_tel</i> 20100	4638	8371	1234.48	148.36	44.4	292.68
<i>RBM38_tel</i> 25076	6082	14144	1949.36	202.28	110.44	343.36
<i>CTCF</i> cen34302	7848	20677	996.36	112	110	111
<i>PCK1_cen</i> 13036	4225	16928	366.56	127.04	44.4	51.8
<i>PCK1_cen</i> 180-IVS6	4323	11077	227.12	129.96	42.12	97.2
3.3_1p36	11592	8710	544.0	168.4	115.6	120.6

¹DNase I Hypersensitivity
² Formaldehyde Assisted Isolation of Regulatory Elements
[†]Data excluded from ANOVA analysis (Section 3.1.3 justification)

3.2 Defining DA Domains in Human Metaphase Homologous Chromosomes

3.2.1 Identification of Differentially Accessible Domains

DA domains were defined as areas of the genome over which multiple single-copy regions demonstrating DA were confirmed to be neighbouring each other using scFISH probes. Six DA domains were identified by developing adjacent sc probes within a genomic area in which there was a confirmed DA probe. These domains were: *XDH* (2p23), *HMGB1P5* (3p24), *FGF6* (12p13), *TPM1* (15q22), *COX5A*(15q25) and *HMGB1P1* (20q13). Domains were named for the target of the original gene mapping study examined for which scFISH probes identifying DA intervals were designed. The boundaries of each domain were defined by the smallest and largest genome coordinates of the two DA probes that were the greatest distance apart from each other. The area of a given domain is inferred to be continuous between scFISH probes. The DA probes defining each domain are listed in Table 3.3.

The DA domains range in size from 16 kb to 129.6 kb with 2 to 4 DA regions identified within each of these domains (Table 3.3). The *XDH* domain is defined by 3 DA regions and spans 25.5 kb of chromosome 2p23.1 (Figure 3.7). Two DA regions within the *XDH* domain are intergenic, *XDH_tel9264* [2110 bp] and *XDH_tel2387* [2986 bp], while the third region covers intron 30 to intron 27 of the *XDH* gene, *XDH_IVS30-IVS27* [2501 bp]. The *HMGB1P5* domain is an entirely intergenic domain, found within chromosome 3p24.3. It is defined by 2 DA probes that span 39.7 kb (Figure 3.8), *ZNF385D_tel640535* [3968 bp] and *ZNF385D_tel678016* [2251 bp]. The two smallest domains, *FGF6* (Figure 3.9) and *TPM1* (Figure 3.10), were identified on chromosomes 12p12.3 and 15q22.2, respectively. *FGF6* spans 16.0 kb and is defined by 2 DA regions, one that is intergenic, *FGF6_cen4492* [1660 bp], and the other that is found within intron 2 of *FGF6*, *FGF6_IVS2* [3430 bp]. The *TPM1* domain

also spans ~16.0 kb. It is defined by 3 DA probes. Two probes span intron 5 to intron 8 of the *TPM1* gene, *TPM1_IVS5-IVS8* [2408 bp] and *TPM1_IVS8* [3300 bp] and one is within an intergenic region, *TPM1_tel3200* [2294 bp]. The two largest domains are COX5A (Figure 3.11) located on 15q24.1 and HMGB1P1 (Figure 3.12) located on 20q13.3. Spanning 110 kb and 129.6 kb respectively, these domains cover large areas of genome. COX5A is defined by 2 DA regions found in the *SCAMP2* gene, one spanning intron 7 to 4, *SCAMP2_IVS7-IVS4* [3002 bp] and the other in intron 1, *SCAMP2_IVS1* [1526 bp] and 1 intergenic region, *COX5A_tel20100* [1725 bp]. HMGB1P1 is defined by 4 DA regions. Three are in intergenic regions, *RBM38_tel25076* [2021 bp], *CTCFI_cen34302* [3553 bp], *PCK1_cen13036* [3533 bp] and 1 that covers just beyond the centromeric edge of the *PCK1* gene to intron 6, *PCK1_cen180-IVS6* [3092 bp]. In both the COX5A and TPM1 domains, there are large areas not covered by scFISH probes. Probe development was attempted in these areas, however a number of constraints on scFISH probe design restricted hybridization within these prospective regions. These included the restriction of probe design by the presence of repetitive elements and optimization failures of prospective probes designed in these areas.

To investigate DA domains, probes from 5 different chromosomes were designed in intervals within genes (coding regions) and in intervals within intergenic regions (non-gene coding regions). The genomic contents as well as precise genomic coordinates of each probe are given in Table 3.3. Within the TPM1 domain (Figure 3.8), 3 regions of DA spanned a short distance of 16.0 kb. Within this area, 8.0 kb was confirmed as DA by sc probe hybridization. The XDH and FGF6 domains spanned similarly small regions, 25.5 kb and 16.0 kb, with 2 and 3 DA regions defining the domains respectively. ScFISH probe hybridizations identified DA in 7.6 kb and 5.1 kb of the XDH and FGF6 domains. The proximity of these DA regions within

such a short span of genome suggests that differential accessibility between homologues extends beyond the defined coordinates of a single probe and are therefore found in areas larger than those covered by single scFISH probes (1.5-5 kb).

Demonstration of up to 4 DA intervals within 129.6 kb (HMGB1P1 domain; Figure 3.12) and 3 DA intervals within 110.0 kb (COX5A domain; Figure 3.11) suggest even larger domain sizes. However, these larger domains may not actually cover a continuous region of DA. In the HMGB1P1 domain, this can be observed with a large gap of 82.9 kb between the *CTCF* and *PCK1* loci, and smaller gaps of <25 kb between the other probes defining this domain. This is also evident in the COX5A domain (Figure 3.11), where 3 probes span 110kb with 87.3 kb not covered by scFISH probes between DA loci *SCAMP2_IVS1* and *COX5A_tel20100*. However, the proximity of these regions, as well as other domains demonstrated in this study suggest DA is a feature of metaphase chromatin present in neighbouring single-copy regions. The location of the sc regions hybridized to define each domain provides evidence that DA occurs more frequently within neighbouring sc genomic intervals than was demonstrated by previously published evidence of individual DA probes that were distributed across larger genomic intervals.

Table 3.3: List of DA domains with each scFISH probe defining each domain.

Domain Name [length]	Chromosome band	Probe Name	Probe Coordinates [GRCh37/hg19]	Genomic Position*
XDH [25 454 bp]	2p23.1	<i>XDH_tel9264</i> <i>XDH_tel2387</i> <i>XDH_IVS30-IVS27</i>	chr2:31,545,815-31,547,924 chr2:31,551,816-31,554,801 chr2:31,568,769-31,571,269	Intergenic Intergenic <i>XDH</i>
HMGB1P5 [39 731 bp]	3p24.3	<i>ZNF385D_tel640535</i> <i>ZNF385D_tel678016</i>	chr3:22,433,351-22,436,333 chr3:22,470,832-22,473,082	Intergenic Intergenic
FGF6 [16 048 bp]	12p12.3	<i>FGF6_cen4492</i> <i>FGF6_IVS2</i>	chr12:4,537,157-4,538,816 chr12:4,549776-4,553,205	Intergenic <i>FGF6</i>
TPM1 [16 034 bp]	15q22.2	<i>TPM1_IVS5-IVS8</i> <i>TPM1_IVS8</i> <i>TPM1_tel3200</i>	chr15:63,353,573-63,355,980 chr15:63,357,346-63,360,645 chr15:63,367,314-63,369,607	<i>TPM1</i> <i>TPM1</i> Intergenic
COX5A [109 970 bp]	15q24.1	<i>SCAMP2_IVS7-IVS4</i> <i>SCAMP2_IVS1</i> <i>COX5A_tel20100</i>	chr15:75,142,349-75,145,350 chr15:75,161,783-75,163,308 chr15:75,250,595-75,252,319	<i>SCAMP2</i> <i>SCAMP2</i> Intergenic
HMGB1P1 [129 583 bp]	20q13.3	<i>RBM38_tel25076</i> <i>CTCFI_cen34302</i> <i>PCK1_cen13036</i> <i>PCK1_cen180-IVS6</i>	chr20:56,009,465-56,011,485 chr20:56,033,167-56,036,719 chr20:56,119,569-56,123,101 chr20:56,135,957-56,139,048	Intergenic Intergenic Intergenic <i>PCK1</i>

Bp = base pair
*Genomic position refers to the genetic characteristics of the sequence hybridized by each individual sc probe. Intergenic refers to a sc probe target outside of gene coding regions. A gene name indicates that the sc probe target is within that given gene.

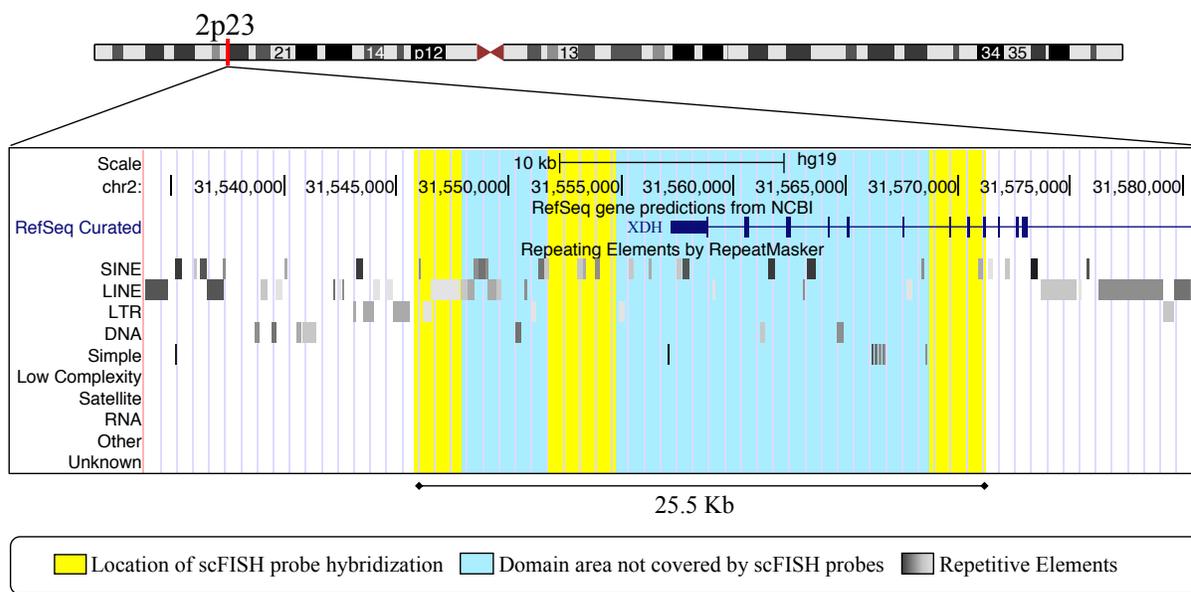


Figure 3.7: Genomic map of XDH domain. A human chromosome 2 ideogram with a magnified view within chromosome band 2p23 (red) that contains the XDH DA domain. The image was created within the UCSC browser⁶⁷ using the GRCh37/hg19 human genome assembly. The domain is represented by turquoise and yellow bars with adjacent chromatin outside the domain in white. The domain size spans ~25.5 kilobases (kb) [chr2:31,545,815-31,571,269]. Yellow bars indicate specific locations of the 3 hybridized scFISH probes in this domain; left to right *XDH_tel9264*, *XDH_tel2387*, *XDH_IVS30-IVS27*. The left margin indicates what is presented in each track. Genomic coordinates are provided in 5 kb intervals followed by curated genes (RefSeq, dark blue) within the region and then multiple tracks of different repetitive sequences. These repeating elements (RepeatMasker) are represented using a greyscale with the lighter the grey, the higher the divergence between DNA sequences within the same family. ScFISH probes are located within regions that either have no repeating elements or repeating elements that have a sequence of divergence of >20%. The different repetitive elements shown include: short interspersed nuclear elements (SINE), long interspersed nuclear elements (LINE), long terminal repeat elements (LTR), DNA repeat elements, simple repeats, low complexity repeats and others.

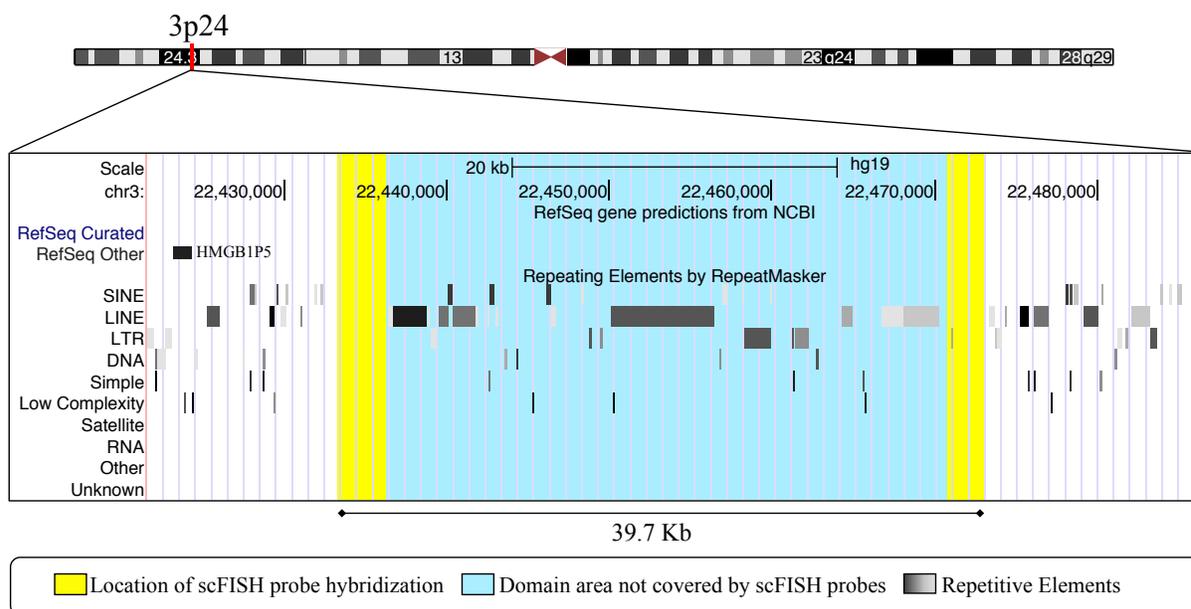


Figure 3.8: Genomic map of HMGB1P5 domain. A human chromosome 3 ideogram with a magnified view within chromosome band 3p24 (red) that contains the HMGB1P5 DA domain. The image was created within the UCSC browser⁶⁷ using the GRCh37/hg19 human genome assembly. The domain is represented by turquoise and yellow bars with adjacent chromatin outside the domain in white. The domain size spans ~39.7 kb [chr3:22,433,351-22,473,082]. Yellow bars indicate specific locations of the 2 hybridized scFISH probes in this domain; left to right *ZNF385D_tel640535*, *ZNF385D_tel678016*. The left margin indicates what is presented in each track. Genomic coordinates are provided in 10 kb intervals followed by curated genes (RefSeq, dark blue) within the region and then multiple tracks of different repetitive sequences. These repeating elements (RepeatMasker) are represented using a greyscale with the lighter the grey, the higher the divergence between DNA sequences within the same family. ScFISH probes are located within regions that either have no repeating elements or repeating elements that have a sequence of divergence of >20%. The different repetitive elements shown include: short interspersed nuclear elements (SINE), long interspersed nuclear elements (LINE), long terminal repeat elements (LTR), DNA repeat elements, simple repeats, low complexity repeats and others.

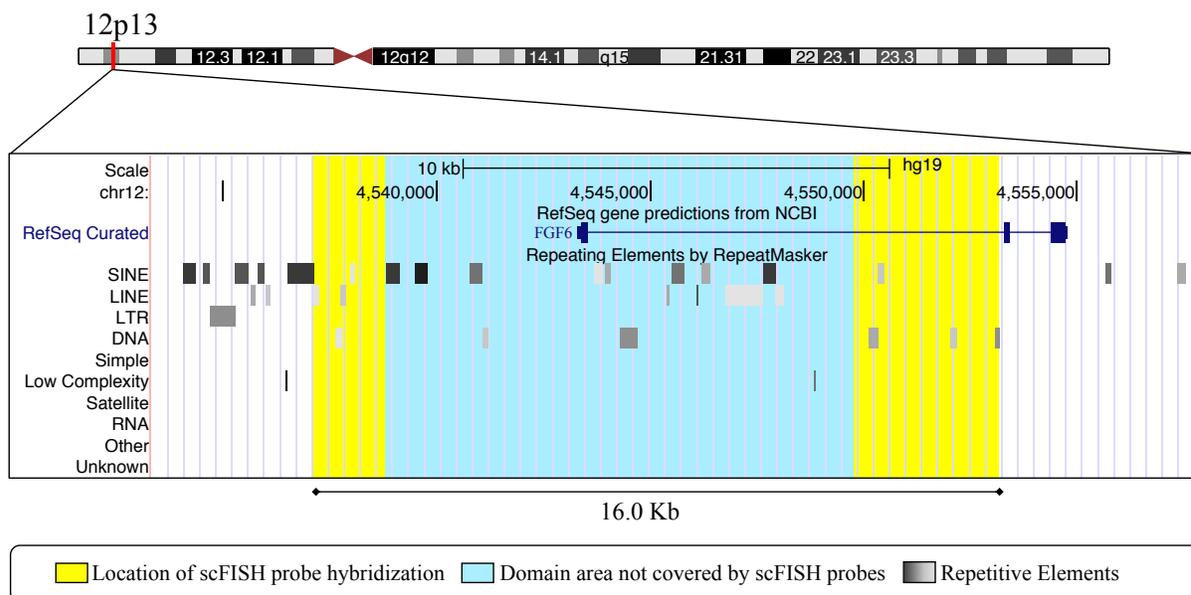


Figure 3.9: Genomic map of FGF6 domain. A human chromosome 12 ideogram with a magnified view within chromosome band 12p13 (red) that contains the FGF6 DA domain. The image was created within the UCSC browser⁶⁷ using the GRCh37/hg19 human genome assembly. The domain is represented by turquoise and yellow bars with adjacent chromatin outside the domain in white. The domain size spans ~16.0 kb [chr12:4,537,157- 4,553,205]. Yellow bars indicate specific locations of the 2 hybridized scFISH probes in this domain; left to right *FGF6_cen4492*, *FGF6_IVS2*. The left margin indicates what is presented in each track. Genomic coordinates are provided in 5 kb intervals followed by curated genes (RefSeq, dark blue) within the region and then multiple tracks of different repetitive sequences. These repeating elements (RepeatMasker) are represented using a greyscale with the lighter the grey, the higher the divergence between DNA sequences within the same family. ScFISH probes are located within regions that either have no repeating elements or repeating elements that have a sequence of divergence of >20%. The different repetitive elements shown include: short interspersed nuclear elements (SINE), long interspersed nuclear elements (LINE), long terminal repeat elements (LTR), DNA repeat elements, simple repeats, low complexity repeats and others.

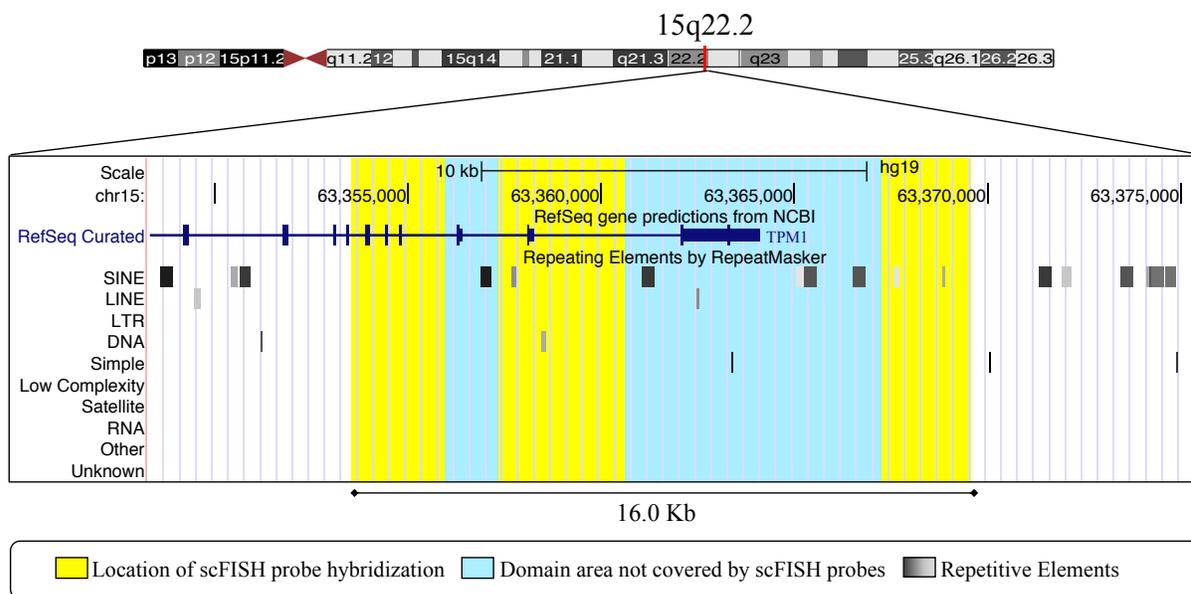


Figure 3.10: Genomic map of TPM1 domain. A human chromosome 15 ideogram with a magnified view within chromosome band 15q22.2 (red) that contains the TPM1 DA domain. The image was created within the UCSC browser⁶⁷ using the GRCh37/hg19 human genome assembly. The domain is represented by turquoise and yellow bars with adjacent chromatin outside the domain in white. The domain size spans ~16.0 kb [chr15:63,353,573 - 63,369,607]. Yellow bars indicate specific locations of the 3 hybridized scFISH probes in this domain; left to right *TPM1_IVS5-IVS8*, *TPM1_IVS8*, *TPM1_tel3200*. The left margin indicates what is presented in each track. Genomic coordinates are provided in 5 kb intervals followed by curated genes (RefSeq, dark blue) within the region and then the multiple tracks of different repetitive sequences. These repeating elements (RepeatMasker) are represented using a greyscale with the lighter the grey, the higher the divergence between DNA sequences within the same family. ScFISH probes are located within regions that either have no repeating elements or repeating elements that have a sequence of divergence of >20%. The different repetitive elements shown include: short interspersed nuclear elements (SINE), long interspersed nuclear elements (LINE), long terminal repeat elements (LTR), DNA repeat elements, simple repeats, low complexity repeats and others.

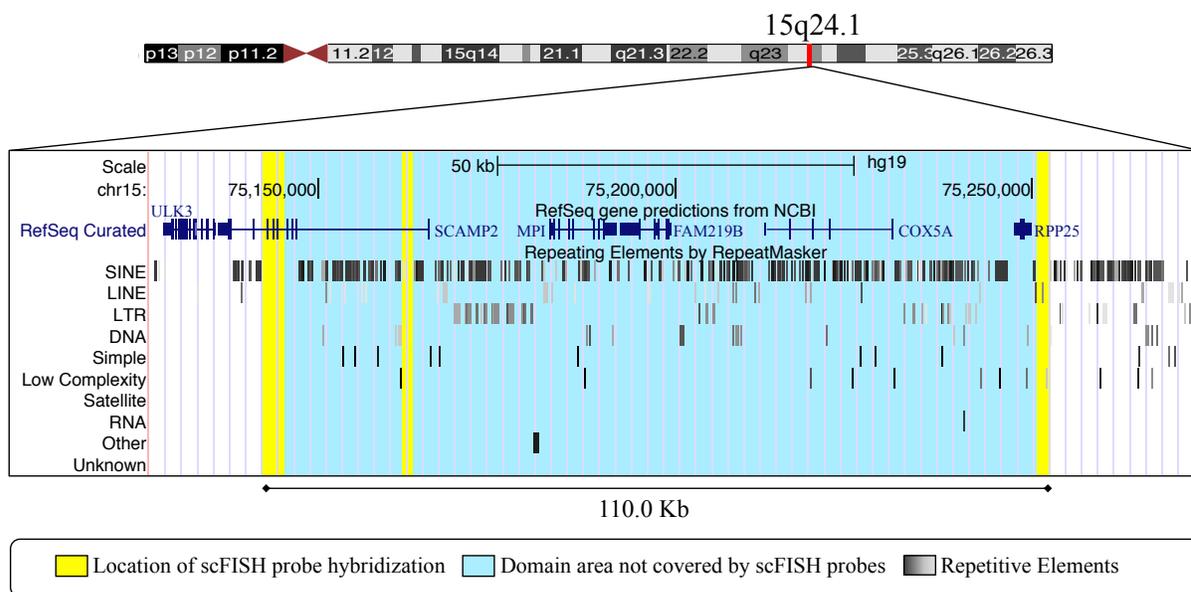


Figure 3.11: Genomic map of COX5A domain. A human chromosome 15 ideogram with a magnified view within chromosome band 15q24.1 (red) that contains the COX5A DA domain. The image was created within the UCSC browser⁶⁷ using the GRCh37/hg19 human genome assembly. The domain is represented by turquoise and yellow bars with adjacent chromatin outside the domain in white. The domain size spans ~110.0 kb [chr15:75,142,349-75,252,319]. Yellow bars indicate specific locations of the 3 hybridized scFISH probes in this domain; left to right *SCAMP2_IVS7-IVS4*, *SCAMP2_IVS1*, *COX5A_tel20100*. The left margin indicates what is presented in each track. Genomic coordinates are provided in 50 kb intervals followed by curated genes (RefSeq, dark blue) within the region and then multiple tracks of different repetitive sequences. These repeating elements (RepeatMasker) are represented using a greyscale with the lighter the grey, the higher the divergence between DNA sequences within the same family. ScFISH probes are located within regions that either have no repeating elements or repeating elements that have a sequence of divergence of >20%. The different repetitive elements shown include: short interspersed nuclear elements (SINE), long interspersed nuclear elements (LINE), long terminal repeat elements (LTR), DNA repeat elements, simple repeats, low complexity repeats and others.

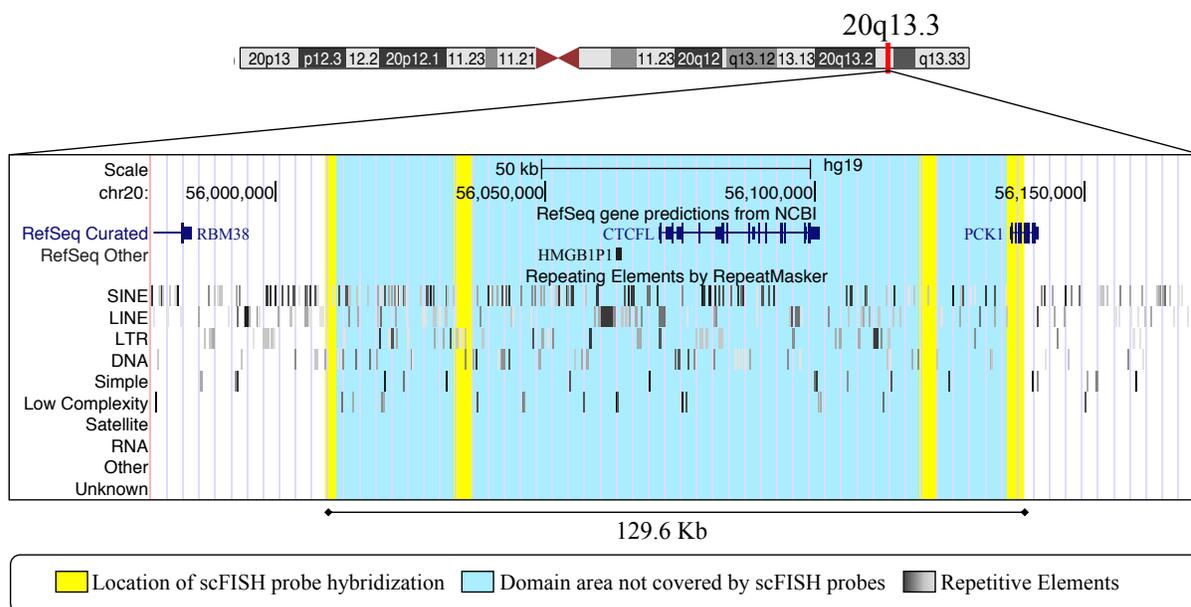


Figure 3.12: Genomic map of HMGB1P1 domain. A human chromosome 20 ideogram with a magnified view within chromosome band 20q13.3 (red) that contains the HMGB1P1 DA domain. The image was created within the UCSC browser⁶⁷ using the GRCh37/hg19 human genome assembly. The domain is represented by turquoise and yellow bars with adjacent chromatin outside the domain in white. The domain size spans ~129.6 kb [chr20:56,009,465-56,139,048]. Yellow bars indicate specific locations of the 4 hybridized scFISH probes in this domain; left to right *RBM38_tel25076*, *CTCF_L_cen34302*, *PCK1_cen13036*, *PCK1_cen180-IVS6*. The left margin indicates what is presented in each track. Genomic coordinates are provided in 50 kb intervals followed by curated genes (RefSeq, dark blue) within the region and then multiple tracks of different repetitive sequences. These repeating elements (RepeatMasker) are represented using a greyscale with the lighter the grey, the higher the divergence between DNA sequences within the same family. ScFISH probes are located within regions that either have no repeating elements or repeating elements that have a sequence of divergence of >20%. The different repetitive elements shown include: short interspersed nuclear elements (SINE), long interspersed nuclear elements (LINE), long terminal repeat elements (LTR), DNA repeat elements, simple repeats, low complexity repeats and others.

3.2.2 Epigenetic Characteristics of DA Domains and the ScFISH Probes Defining DA Domains – Topologically Associated Domains (TADs)

Characterisation of epigenetic structures in genomic regions corresponding to DA probe intervals was expanded to include higher-order levels of chromatin folding related to gene expression and regulation during interphase^{6,30,42}. Topologically associated domains (TADs) have been suggested to form structural units in interphase chromatin^{22,30}. TADs facilitate interactions with regulatory elements and their gene targets within the defined boundaries of chromatin scaffolds. Adjacent TADs are separated by insulator sequences that isolate interactions of regulatory elements within an individual TAD and protect from interaction of regulatory elements outside the defined boundaries that are intended for other gene targets^{22,30}. Publicly available data of chromatin confirmation capture information from Hi-C analysis in lymphoblast cell line GM12878²¹ was accessed and visualized using the 3D Genome Browser⁸⁶. Hi-C arrests the 3D chromatin structure of a population of cells by holding points of contact in place across the genome, then sequencing is used to determine the frequency at which two distinct points in the genome are within contact within that population. The Hi-C data were used to assess the correspondence between loci of DA in metaphase and interphase TAD structures. The 3D genome browser⁸⁶ was used to visualize TADs and sub-TAD interactions over the DA domains defined in metaphase chromosomes and the surrounding chromatin (>100kb up and downstream) within the same TAD as well as adjacent chromatin folding structures.

Five of the six DA domains (XDH, FGF6, COX5A, TPM1, HMGB1P1) in metaphase chromosomes that we defined in Section 3.2.1 were each contained within a single TAD. The other domain, HMGB1P5, was located between 2 TADs. The COX5A (Figure 3.13),

HMGB1P1 (Figure 3.14), and HMGB1P5 (Figure 3.15) domains are presented as examples in figures 3.13 through 3.15. The other 3 domains, XDH, FGF6, and TPM1, are presented in figures 1 through 3 in Appendix VI.

The frequency of chromatin interactions within the TAD and in the area surrounding each DA domain are observed in the contact frequency heat maps (red). To interpret the frequency of contact between two genomic points, a triangle connecting the two points is drawn and the intensity of red at the point of the peak (where the lines from the two genomic points intersect) indicates how frequent contact is between these points. The frequency of contact data presented were the average frequencies within the entire cell population examined. Four of the 5 domains (XDH, FGF6, COX5A, HMGB1P1) contained within a separate TAD show clear, high frequency, intra-TAD interactions (evidence of sub-TAD structures) occurring between the chromatin within the DA domain and areas beyond the domain within the same TAD. This is evident in both the COX5A domain (Figure 3.13) and HMGB1P1 domain (Figure 3.14) where there are clear triangles of high frequency sequence interactions (bright red) above the domain with areas in close proximity within the same TAD, as well as larger contact points connecting more distant loci within the TAD. All four domains show the bright red associated with the highest frequency of interaction between points. This shows domains in contact with multiple areas within the TAD during interphase. Similar observations of sub-TAD interactions in close and distant loci of the TAD can be observed within the XDH and FGF6 domains (Appendix VI. Fig 1 & 2).

Weaker points of contact were observable between the TPM1 domain (Appendix VI Fig. 3) and surrounding chromatin within the same TAD. Overall the entire TAD in which the TPM1 domain is located has few observable high frequency contacts suggesting overall weak

contact levels. Weak intra-TAD interactions indicate that few sub-TAD structures are present across this particular TAD in the GM12878 cell line. The lower level of contact within the TPM1 domain is similar to the HMGB1P5 domain (Fig 3.15) which is found between the boundaries of 2 different TADs in very close proximity to the boundary of the blue-grey TAD. However, the observation of 5 of 6 metaphase DA domains, each within an individual TAD, also with sub-TAD interactions is consistent with a possible link between TAD structures that are present during interphase and regions of DA present in metaphase homologous chromatin at 2p23, 12p13, 15q22.2, 15q24.1, and 20q13.3.

A number of architectural proteins have been implicated in the formation of TAD structures, the most well-established connection being with CTCF insulator proteins. The boundaries of topologically associated domains can be characterized by clustering of CTCF binding sites^{21,22}. Maintaining knowledge of the location of these clusters is important to re-establish functional TAD structures at the same location after their loss during metaphase. Though all TAD boundaries are clusters of CTCF, not all CTCF binding sites are TAD boundaries as they also facilitate chromatin looping within TADs involved in producing sub-TAD structures. As 4 of 6 DA domains occurred within a single TAD, with observable high frequencies of intra-TAD interaction indicative of sub-TAD looping structures, binding of CTCF would be required to help facilitate the formation of these structures as well as delimit the overall boundaries of each TAD. Prediction of CTCF binding site loci and their binding affinity were generated using information theory-based transcription factor binding site motifs using an algorithm written and validated by Lu *et al.* 2017⁸⁷. Sites generated with binding affinities (R_i) equal to or greater than half the overall mean binding affinity (R_{seq}) were reported across the genomic areas of each domain, plus and minus 1Mb.

Within each domain the sum of binding affinities of each CTCF binding site was calculated and normalized by dividing by the number of kilobases in each domain. Genomic areas of equal size in kilobases within the same chromosome, one area beyond the boundaries of the domain but within the same TAD and one area in the adjacent TAD to the domain, were selected to determine if there was a difference in strength of CTCF binding sites within domains compared to other areas of the genome. No significant difference ($p = 0.5367$) was found between the median binding affinity per kilobase in each group using a Kruskal Wallis test with multiple comparisons (Table 3.2). This indicates that the domains within the boundaries defined in this study did not differ in the level of CTCF binding site affinity per kilobase compared to areas outside the domains. This analysis does not support the hypothesis that DA domains present in metaphase differ in CTCF binding site affinity compared to areas outside the domains during interphase. However, this is a small sample size of domains of which the boundaries have not been defined by adjacent equivalently accessible region; therefore, further investigation into DA domains as a potential bookmark of CTCF binding clusters in metaphase should not be ruled out.

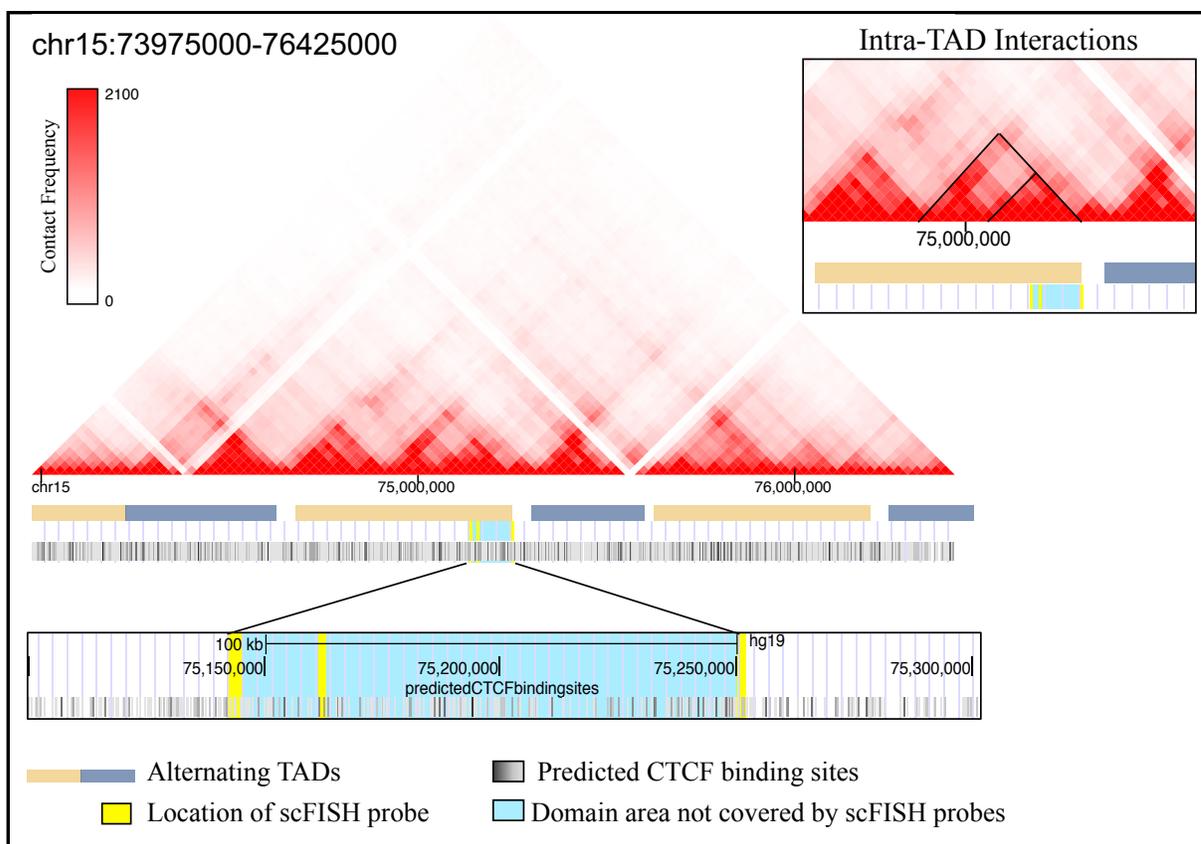


Figure 3.13: Localization of the COX5A domain in metaphase cells relative to TAD structures present in interphase cells. The heat map (3D Genome Browser)⁸⁶ shows interaction frequencies between chromatin within the COX5A domain (indicated by yellow-light blue area in UCSC genome browser image⁶⁷) and surrounding chromatin. Alternating blue-grey and light tan bars represent alternating TADs. The COX5A domain is clear at one end of a single light tan TAD. Intensity of red increases with increased frequency of interaction. The scale in the top left measures the intensity of contact frequency as the normalized number of contacts between 2 points. There is high frequency of intra-TAD interactions (bright red triangles) within and around this domain. Two areas of intra-TAD interaction over the domain are outlined in a magnified view in the top right corner. Predicted CTCF binding sites⁸⁷ are presented in grey-scale, with the colour darkening with increased site binding affinity. Magnified image of DA domain (UCSC Genome Browser; GRCh37/hg19 assembly) shows full domain in light blue with yellow indicating location of hybridized scFISH probes.

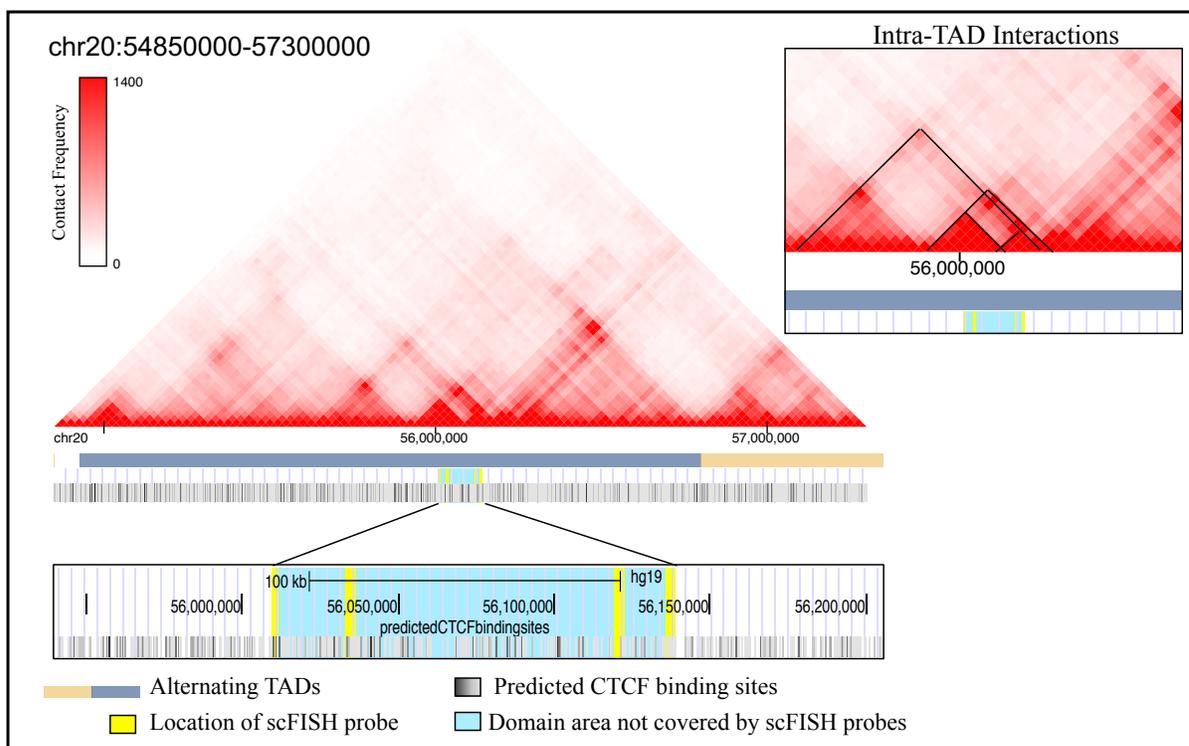


Figure 3.14: Localization of the HMGB1P1 domain in metaphase cells relative to TAD structures present in interphase cells. The heat map (3D Genome Browser)⁸⁶ shows interaction frequencies between chromatin within the HMGB1P1 domain (indicated by yellow-light blue area in UCSC genome browser image⁶⁷) and surrounding chromatin. Alternating blue-grey and light tan bars represent alternating TADs. The HMGB1P1 domain is in the middle of a single blue-grey TAD. Intensity of red increases with increased frequency of interaction. The scale in the top left measures the intensity of contact frequency as the normalized number of contacts between 2 points. There is a high frequency of intra-TAD interactions within and around this domain, with a clear triangle of frequent contact occurring above the domain. Four of these intra-TAD interactions over the HMGB1P1 domain are indicated in the magnified image in the top right corner. Predicted CTCF binding sites⁸⁷ are presented in grey-scale, with the colour darkening with increased site binding affinity. Magnified image of DA domain (UCSC Genome Browser; GRCh37/hg19 assembly) shows full domain in light blue with yellow indicating location of hybridized scFISH probes.

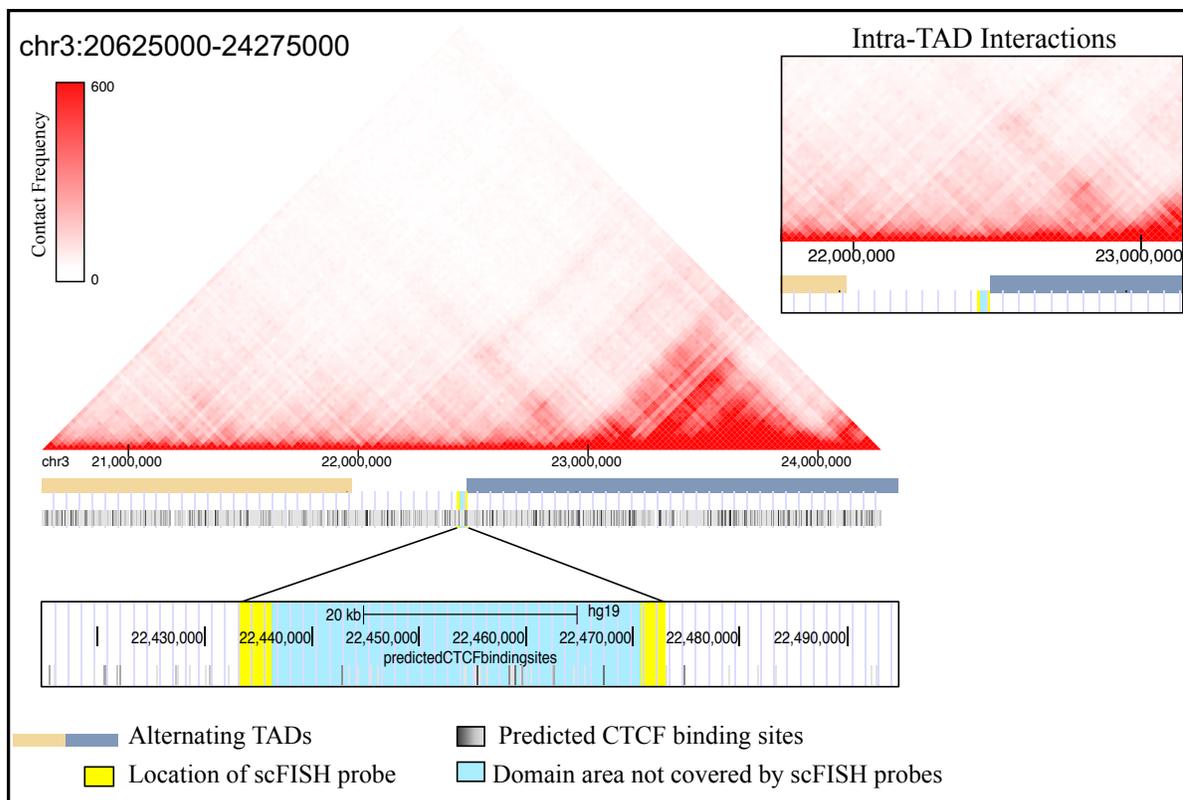


Figure 3.15: Localization of the HMGB1P5 domain in metaphase cells relative to TAD structures present in interphase cells. HMGB1P5 is the only domain identified that is not found within a TAD, but rather in the space between two TADs. The heat map (3D Genome Browser)⁸⁶ shows interaction frequencies between chromatin within the HMGB1P5 domain (indicated by yellow-light blue area in UCSC genome browser image⁶⁷) and surrounding chromatin. Alternating blue-grey and light tan bars represent alternating TADs. Intensity of red increases with increased frequency of interaction. The scale in the top left measures the intensity of contact frequency as the normalized number of contacts between 2 points. Short intra-TAD interactions can be observed above the location of the domain. A magnified view of the area over the HMGB1P5 domain is shown in the top right corner. Predicted CTCF binding sites⁸⁷ are presented in grey-scale, with the colour darkening with increased site binding affinity. Magnified image of DA domain (UCSC Genome Browser; GRCh37/hg19 assembly) shows full domain in light blue with yellow indicating location of hybridized scFISH probes.

Table 3.4: Median CTCF binding affinity/kilobase across DA domains compared to sequences of equal length beyond the boundaries of the domain within the same TAD and in the adjacent TAD.

	Location of Sequence		
	DA Domain	Outside Domain within the same TAD	Adjacent TAD
Median CTCF binding affinity (Rseq)/kilobase (bp)	12.680	8.262	12.720
	No significant difference between groups (p = 0.5367)		

3.3 DA Loci Are Conserved Between Peripheral Blood, Bone Marrow, and Fibroblast Tissues.

Three different tissue types (peripheral blood, bone marrow, and fibroblasts) were selected to investigate the presence and conservation of DA loci across different cell and tissue types. The cell type in which DA loci were initially identified and characterized was peripheral lymphocytes (T-lymphocytes and B-lymphoblasts)^{7,8}. All new DA probes designed for this study were validated and identified in T-lymphocytes. In bone marrow, the dividing cells represent a heterogeneous mixture of cell types including both B and T lymphocytes as well as hematopoietic stem cells and differentiating lymphoid and myeloid cells. This allowed investigation of DA beyond the terminally differentiated and activated lymphocytes in which DA was characterized. Fibroblast cell samples (originally derived from skin punch biopsies) represent different germ layers with fibroblasts derived from the ectoderm while lymphocytes and bone marrow are derived from the mesoderm⁸⁸. Preliminary unpublished work in our laboratory suggested that DA occurred in fibroblasts.

3.3.1 Expression in DA Regions Across T-Lymphocytes, Bone Marrow Cells and Fibroblasts

Five DA probes were selected as described in Methods section 2.3 to compare presence of DA between the tissue types. Four of the probes were used in the domain studies previously described. Three of the probes were within genes: *XDH_IVS30-IVS27*, *PCK1_cen180-IVS6*, and *DUOX1_IVS1-IVS3*. The two other probes were from intergenic DA regions: *TPMI_tel3200* and *CTCFL_cen34302*. One equivalent accessible (EA) probe from an intergenic region, *3.3_1p36*, was selected as a control.

DA regions were selected, in part, based on having no or low expression (0.0 to 5.0 TPM) across all tissues. This was done to control for differences in expression across one or more tissues complicating interpretation of conservation of DA between tissues. Expression data in transcripts per million (TPM) were reported for each gene in which a confirmed DA or EA scFISH probe localized using data from the GTEx⁷⁸ and Human Protein Atlas (HPA)⁷⁹ databases (Figure 3.16, Figure 3.17). The list of genes for which expression data was collected was assembled from DA and EA regions identified from the domain studies, previously published regions identified as DA or EA⁷, and studies from our laboratory identifying DA regions from examination of historical gene mapping studies⁷¹. This maximized the number of regions selected from (DA n= 41, EA n= 26) when determining the probes to be used for comparison between tissues. Data presented include EA regions within genes, however, time did not allow this study to include a probe hybridizing to an EA region within a gene.

From the data collected, the genes investigated showed variation in the mean level of expression from 0 TPM to values >100 TPM with comparable variation in expression observed across genes in both DA regions (Figure 3.16) and EA regions (Figure 3.17). There were also variable differences in expression when comparing between the three tissues of interest, lymphocytes, bone marrow, and fibroblasts, at the same locus both within DA and EA loci. Comparable variations in expression between DA and EA loci within genes, as well as between tissues at the same locus along with the observation of intergenic DA and EA regions does not support a role of DA in regulating tissue-specific expression. Of note is that data from the HPA were already presented as mean values and therefore standard deviation was unable to be calculated nor represented for bone marrow data as individual values for each tissue were not available.

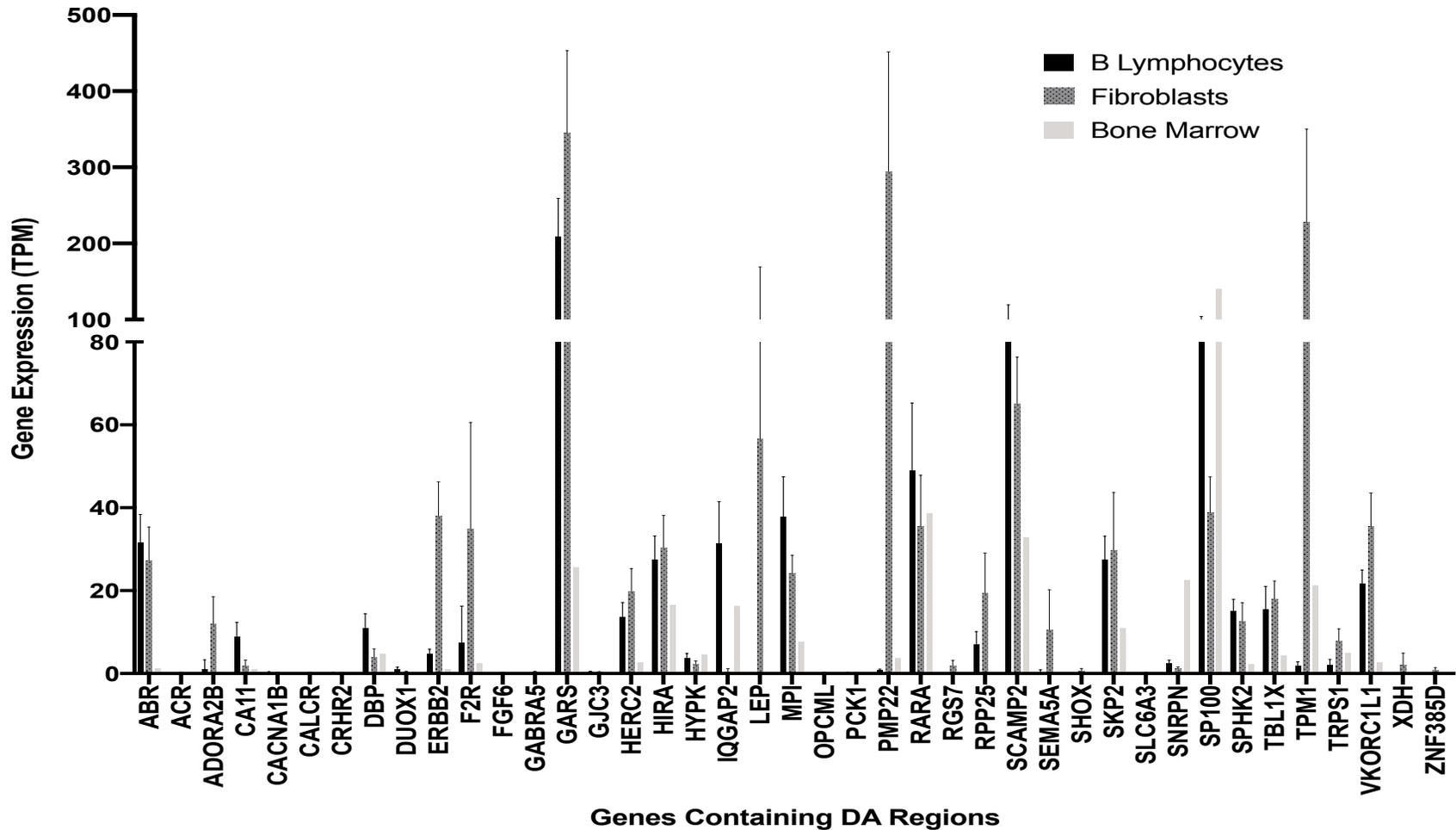


Figure 3.16: Mean gene expression in transcripts per million (TPM) reported for each gene in which a confirmed DA scFISH probe localized. Data reported includes both previously published⁷ and recently identified DA regions within genes. B lymphocyte and fibroblast data were obtained from GTEx⁷⁸. Bone marrow data were obtained from the Human Protein Atlas⁷⁹. Error bars representing standard deviations were calculated for data collected from the GTEx database⁷⁸ but not for HPA data (see text).

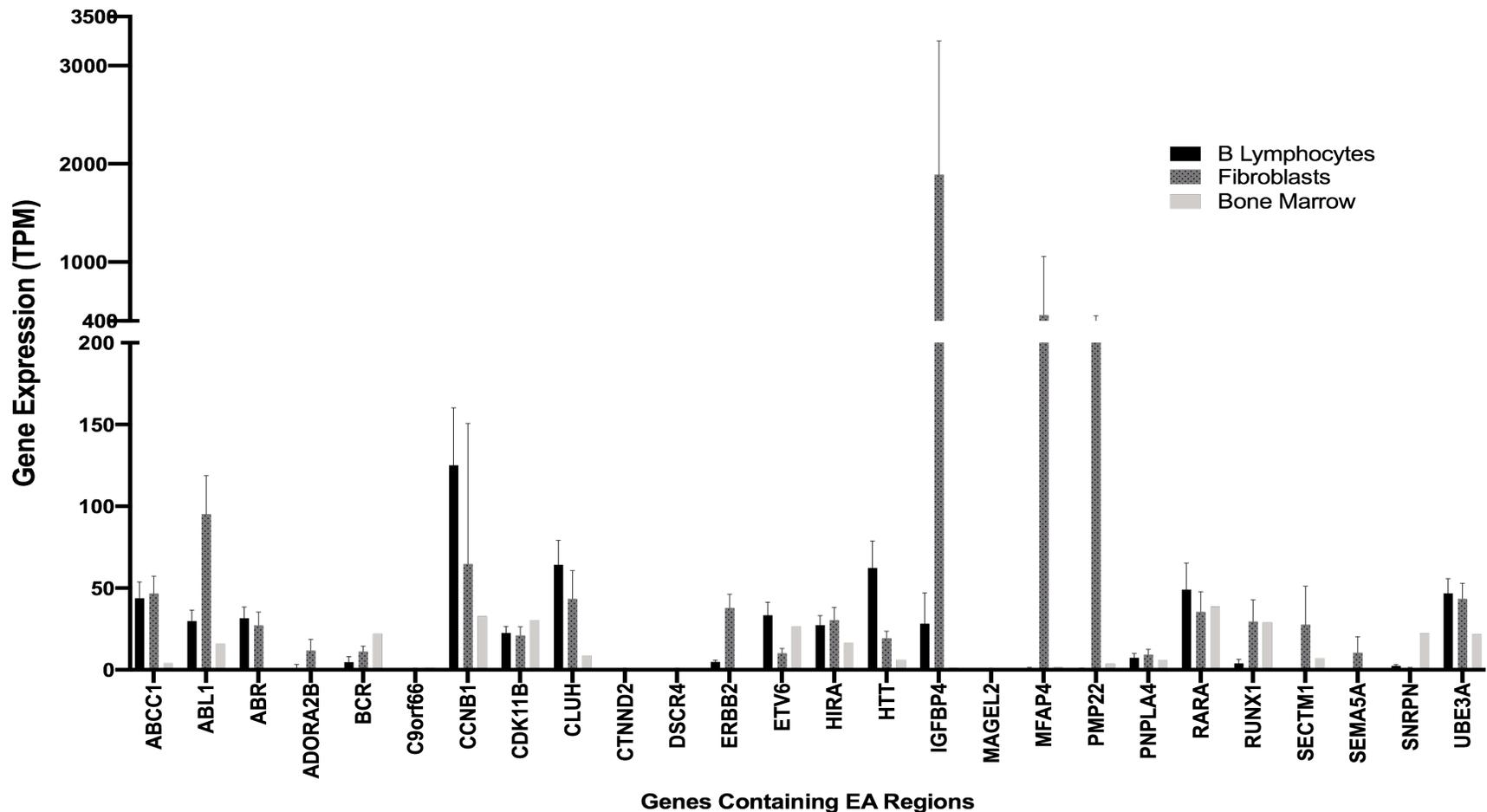


Figure 3.17: Mean gene expression in transcripts per million (TPM) reported for each gene in which a confirmed EA scFISH probe localized. Data collected was from all previously reported EA regions within genes⁷. B lymphocyte and fibroblast data were obtained from GTEx⁷⁸. Bone marrow data were obtained from Human Protein Atlas⁷⁹. Error bars representing standard deviations were calculated for data collected from the GTEx database⁷⁸ but not for HPA data (see text).

3.3.2 Conservation of DA Loci in T-Lymphocytes, Cells from Bone Marrow, and Fibroblasts

All DA loci were conserved across all three tissues. The characteristic probe fluorescence difference between homologues was observed in all loci investigated within genes *XDH_IVS30-IVS27*, *PCK1_cen180-IVS6* (Figure 3.18), and *DUOX1_IVS1-IVS3* and in intergenic regions, *TPMI_tel3200* (Figure 3.19) and *CTCFL_cen34302*. Examples of the fluorescence difference between homologues both in a gene and in an intergenic region across tissues are given in Figure 3.18 and 3.19. A significant number of cells were scored as DA at all DA loci investigated in all tissues using a two-tailed binomial test with normal approximation (Appendix VII). An intergenic EA probe, 3.3_1p36, was also included to investigate the conservation of a locus known not to show DA in T-lymphocytes between the same tissues. The presence of EA was also determined using a two-tailed binomial test with normal approximation (Appendix VII). Equivalent accessibility between homologues was conserved between lymphocytes, bone marrow and fibroblasts at the EA locus (Figure 3.20) suggesting that both EA and DA loci are maintained between tissues types. The conservation of DA in bone marrow was not surprising as both peripheral lymphocytes and bone marrow are derived from stem cells located in bone marrow.

A non-parametric Kruskal-Wallis test determined that the presence of DA was statistically indistinguishable ($p > 0.99$ at each locus) between T-lymphocytes, bone marrow, and fibroblasts (Figure 3.21). The same test found the presence of EA at 3.3_1p36 to be indistinguishable between tissues ($p > 0.99$) as well. Therefore, we assume that the accessibility at these loci are indistinguishable between these 3 tissues and define these as a conserved property at these loci in mitotic chromosomes.

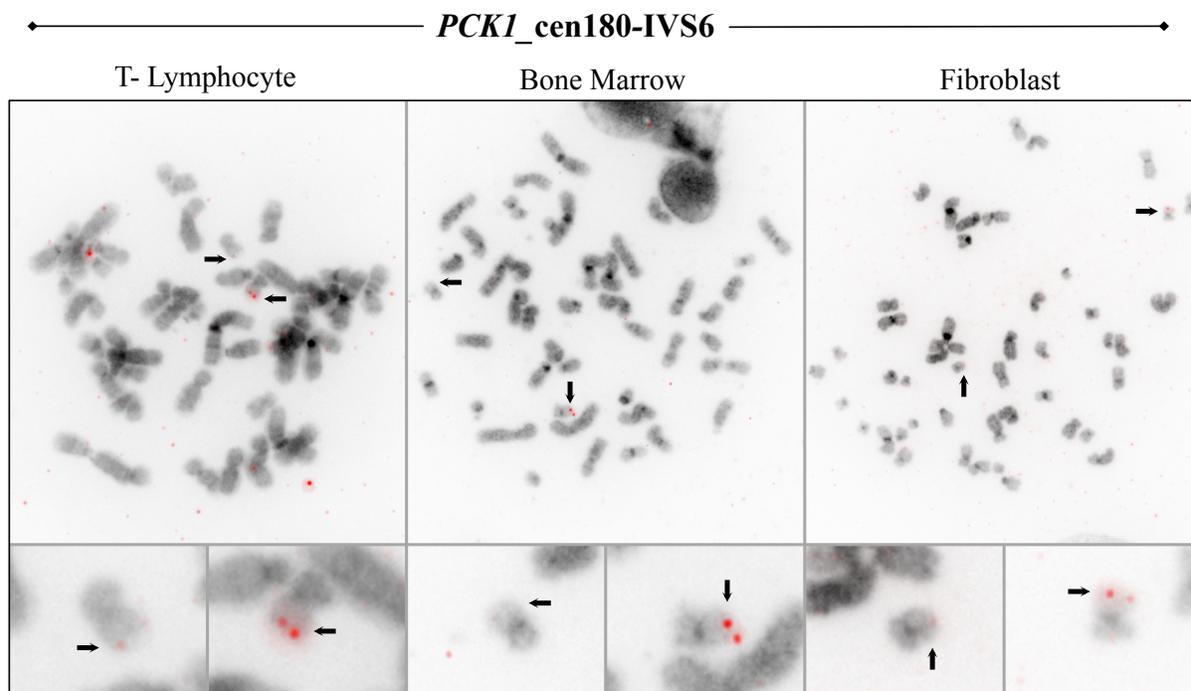


Figure 3.18: The *PCK1_cen180-IVS6* region shows the conservation of DA within a gene in metaphase T-lymphocyte, bone marrow, and fibroblast cells. Human metaphase homologous chromosomes hybridized with *PCK1_cen180-IVS6* (chr20q13.3: 56,135,957-56,139,048), a single-copy FISH probe, to T lymphocyte (left), bone marrow (center) and fibroblast (right) cells. Both homologues hybridized by the scFISH probe are indicated with arrows on the full metaphase and below in the magnified images of each homologue. The T-lymphocyte cell has 1 dim hybridization (left homologue) compared to 2 bright (right homologue). The bone marrow cell has no hybridization (left homologue) compared to 2 bright (right homologue). The fibroblast cell has a dim (left homologue) compared to 1 bright and 1 medium (right homologue). The differential hybridization intensity observed across all tissues is characteristic of DA.

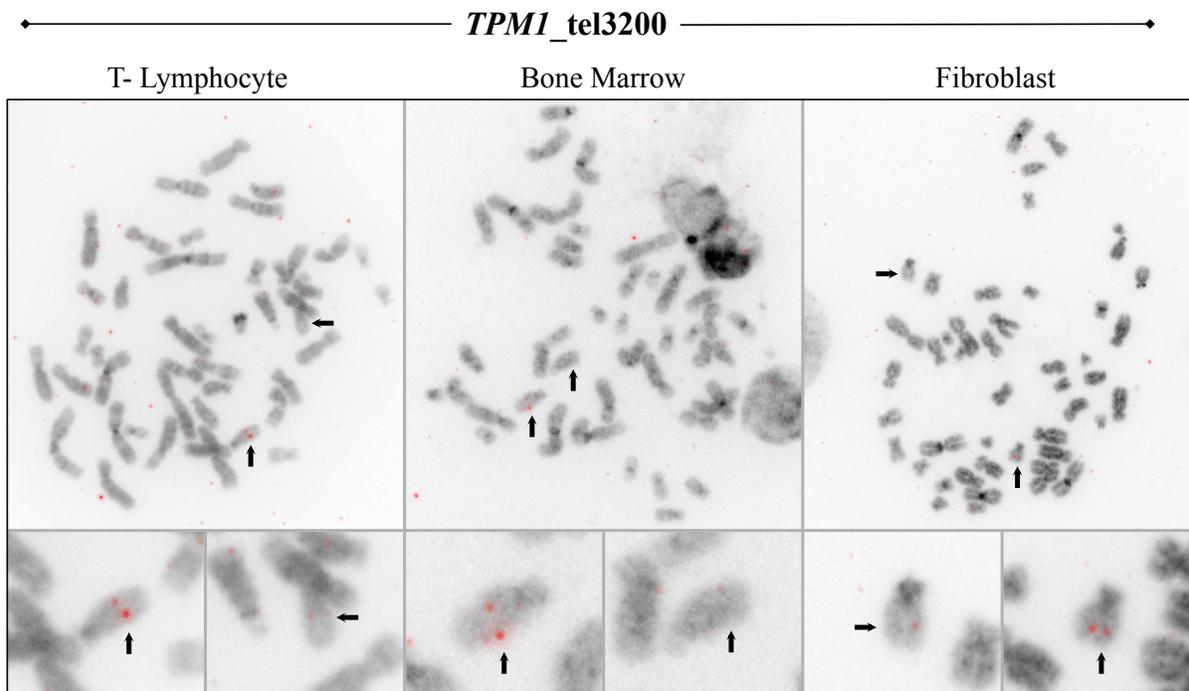


Figure 3.19: The *TPM1_tel3200* region shows conservation of DA in an intergenic region in metaphase T-lymphocyte, bone marrow, and fibroblast cells. Human metaphase homologous chromosomes hybridized with *TPM1_tel3200* (chr15q22.2: 63,367,314-63,369,607), a single-copy FISH probe, to T lymphocyte (left), bone marrow (center) and fibroblast (right) cells. Both homologues hybridized by the scFISH probe are indicated with arrows on the full metaphase and below in the magnified images of each homologue. The T-lymphocyte cell has 1 bright and 1 medium hybridization (left homologue) compared to 1 dim (right homologue). The bone marrow cell has 2 bright hybridizations (left homologue) compared to 0 (right homologue). The fibroblast cell has a medium (left homologue) compared to 2 bright (right homologue). The differential hybridization intensity observed across all tissues is characteristic of DA.

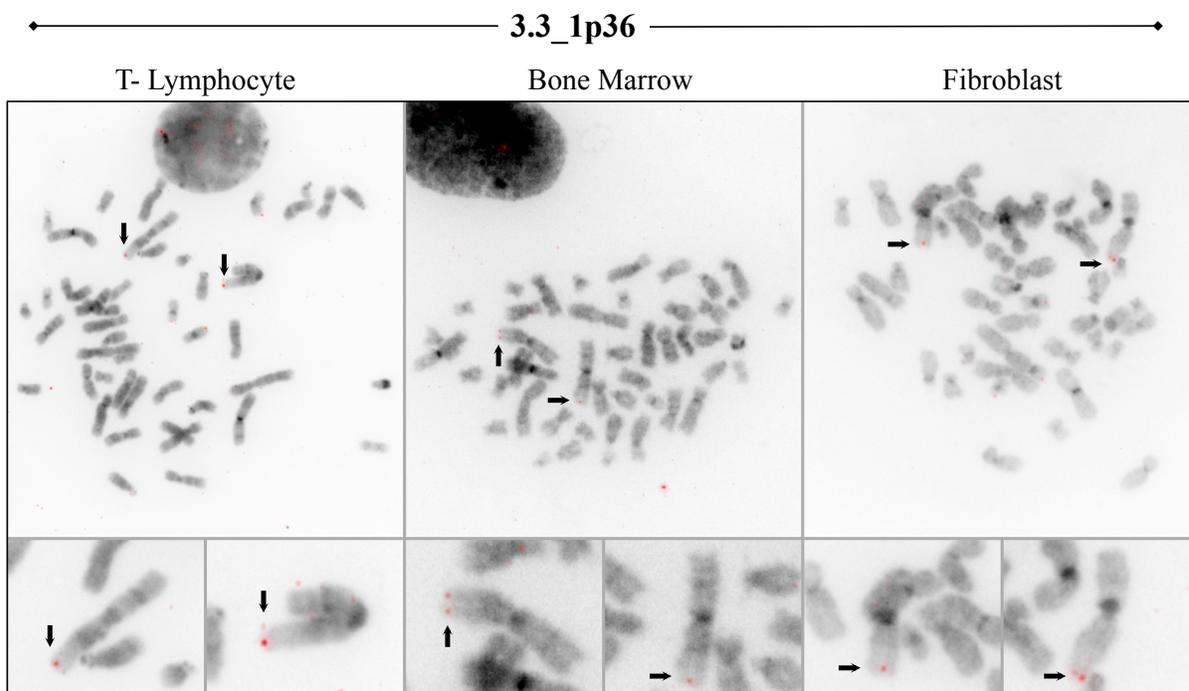


Figure 3.20: The 3.3_1p36 region shows conservation of EA in metaphase T-lymphocyte, bone marrow, and fibroblast cells. Human metaphase homologous chromosomes hybridized with 3.3_1p36 (chr1p36:1,171,789-1,175,143), a single-copy FISH probe, to T lymphocyte (left), bone marrow (center) and fibroblast (right) cells. Both homologues hybridized by the scFISH probe are indicated with arrows on the full metaphase and below in the magnified images of each homologue. In the T-lymphocyte cell both homologues have 1 bright and 1 dim hybridization. The bone marrow cell has 2 medium hybridizations (left homologue) compared to 1 medium (right homologue). The fibroblast cell has 1 bright hybridization (left homologue) compared to 1 bright and 1 medium (right homologue). The equivalent probe hybridization intensity observed across all tissues is characteristic of EA.

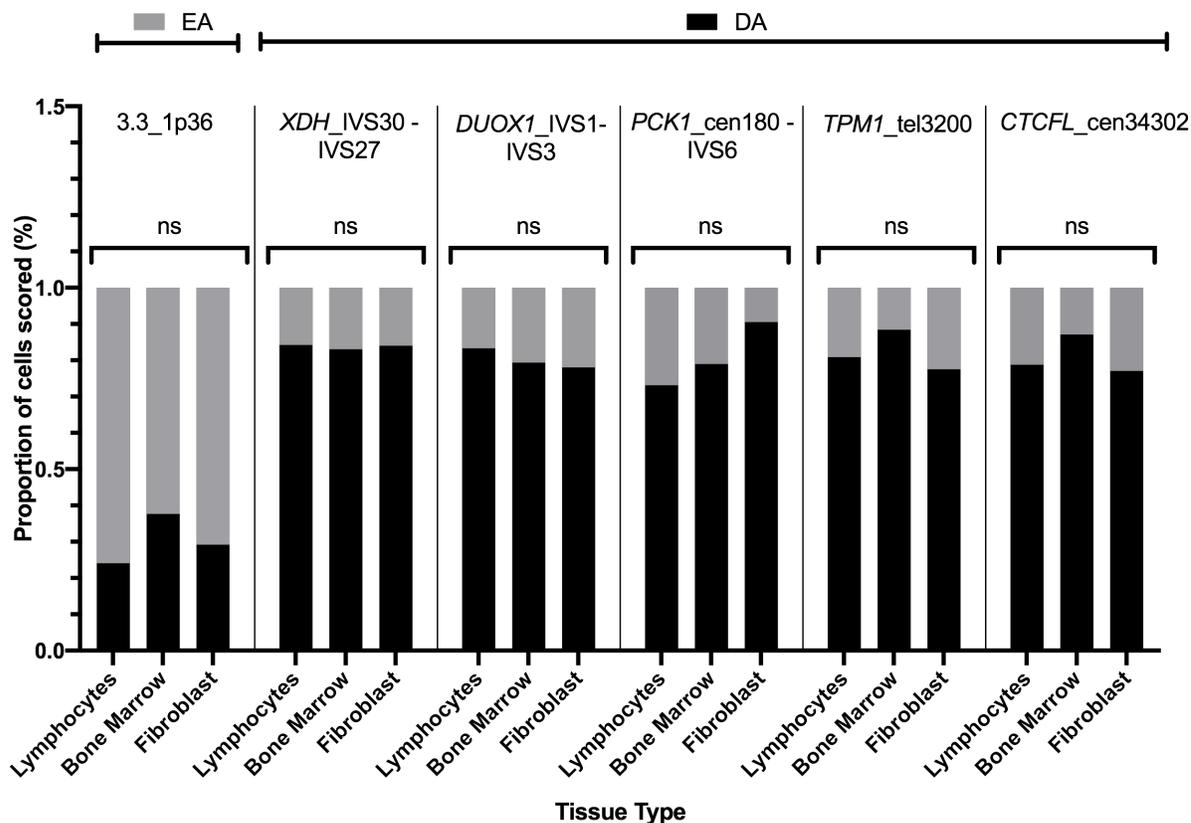


Figure 3.21: Proportion of cells scored as DA (black) and EA (grey) within lymphocytes, bone marrow cells, and fibroblasts are not significant. Regions include 3 DA regions within genes, 2 intergenic DA regions, and 1 EA region. Across the tissues examined for each DA or EA region, the accessibility between metaphase homologues remained the same, DA loci remained DA and EA loci remained EA, with a significant difference calculated between cells scored as DA and EA. No significant difference was found between the proportion of cells identified as DA between the 3 tissues investigated. Sample size differs between each tissue and each probe (Appendix VII). Significant differences were calculated using a Kruskal Wallis test ($\alpha=0.05$) when comparing between tissues and proportion of cells scored as DA and EA. (ns= not significant)

In all cases where two different samples were analyzed for each tissue type a two proportion Z-test was used to test the difference between individuals (Appendix VII). There was no significant difference found between individuals (n=8) at all DA loci and the single EA locus across all tissues, with the exception of the 2 bone marrow samples hybridized with *CTCFI_cen34302* ($p = 0.004$). Both samples showed DA, it was the proportion of DA cells scored in each sample that differed. This difference, as previously mentioned with *SCAMP2_IVS7-IVS4* in lymphocytes, could be stochastic or individual related.

DA conservation at the same loci in T-lymphocytes and bone marrow suggests DA is present and maintained at the same locus in other peripheral blood cells, T- and B-lymphocytes, as well as progenitor lymphocyte cells at various stages of differentiation. Repeating scFISH experiments in fibroblasts yielded similar results providing evidence of DA in tissues derived from both ectoderm and mesoderm germ layers. These observations of conservation of both DA and EA demonstrated that accessibility patterns between homologues during metaphase are not a feature unique to lymphocytes. Of note is that 4 of the DA loci conserved in the 3 tissues were found within 3 of the domains described in this study. *XDH_IVS30-IVS27* within the XDH domain, *TPMI_tel320* part of the TPM1 domain and *PCKI_cen180-IVS6* and *CTCFI_cen34302* both within the HMGB1P1 domain. This is evidence that the DA domains described and characterized within peripheral lymphocytes may also be maintained between bone marrow and fibroblasts.

It should be noted that not all loci could be investigated with 2 patient samples for each tissue type. In bone marrow data from only 1 individual hybridized with *TPMI_tel3200* was reported; in fibroblasts hybridization data from *DUOX1_IVS1-IVS3*, *TPMI_tel3200*, and *XDH_IVS30-IVS27* was only reported from a single individual. This was due to the quality of

chromosomes from individual sample preparations restricting the number of slides that met the quality needed for analysis. Use of a sample from a second individual for these probes would be ideal, however each probe had already been mapped and validated in 2 individuals in lymphocytes, and therefore a second individual in each tissue was not a requirement.

Chapter 4: Discussion

4.1 Identification and Validation of New DA Loci

Eighteen previously unknown DA loci were identified in this study. These loci and one previously published EA locus were validated to confirm their respective metaphase homologue accessibility. Trends observed during the initial characterization of DA were compared with those of the newly identified DA regions. These loci were not identified randomly, which based on previous studies from our laboratory would have resulted in a higher prevalence of EA loci. The experimental design of defining domains based on anchor scFISH loci known to exhibit DA resulted in enrichment of neighbouring DA sequences. The DA regions identified for this study were located on chromosomes 2, 3, 12, 15, and 20. The DA loci on chromosomes 2 (*XDH_tel9264*, *XDH_tel2387*, *XDH_IVS30-IVS27*) and 3 (*ZNF385D_tel640535*, *ZNF385D_tel678016*) were the first differentially accessible loci to be identified on those chromosomes. The DA loci reported in this study further expand the presence of DA across the genome from previously verified regions. There now are confirmed DA regions, new and previously reported⁷, identified on 16 of the 24 possible human chromosomes.

A significantly higher proportion of cells were identified as DA at all new DA loci compared to the EA locus that had a significantly higher proportion of EA cells. This is consistent with previously described characterization of differential accessibility between metaphase homologous chromosomes^{7,8}. The proportions of cells showing DA in different individuals analyzed at the same locus were not significantly different in T-lymphocytes with the exception of *SCAMP2_IVS7-IVS4*. At the *SCAMP2_IVS7-IVS4* locus, DA was evident in both samples, it was only the proportion of cells exhibiting DA that appeared to differ. This

difference could be a stochastic occurrence. This was the only occurrence in the different combinations of 2 individuals investigated per probe locus from a total of 23 individuals included in the T-lymphocyte studies. The difference could also be due to differences in the quality of the metaphase chromosomes between the samples, based on their respective chromosome morphologies and lengths (ie. resolution). Regardless of the observed variability in the proportion of DA cells, according to our definition, *SCAMP2_IVS7-IVS4* was scored the same between samples.

Observed differences in probe fluorescence intensity between metaphase homologues were confirmed by quantification of integrated probe signal intensities using GVF analysis. Quantification of the signals of 5 DA probe regions confirmed that the ratio of fluorescence intensities between homologues was different than that of the EA region. The median difference of integrated intensity ratios between DA regions (0.82) was significantly different compared to the median of the EA region (0.23) analyzed. This was a trend that was consistent with previous comparisons of reported DA and EA regions⁷. The range of both DA and EA datasets were 0.002 – 1.0, indicating DA and EA cells were represented in each dataset. This is due to the heterogeneous presentation of DA within a sample. The difference between classification as either DA or EA, the higher proportion of one accessibility pattern over the other, can be observed in the distribution of the values within the interquartile ranges. All EA data for this analysis were collected from a single EA locus. Expanding the number of EA loci included in this analysis would not be expected to change the significant difference between the fluorescence intensity ratios between DA and EA regions. However, the overall variation of the EA dataset would be expected to decrease with an increased number of EA loci than that from a single locus. This is due to more variation in the possible differences in probe intensities

at DA loci (e.g. dim/bright, bright/medium) compared to EA loci where a significantly larger number of cells have equivalent fluorescent intensities. The median values and the significant difference between them, support the classifications made qualitatively of the new DA loci.

The final point of comparison between the new DA regions identified in this study and the trends of previously characterized EA and DA regions was looking at epigenetic marks of open chromatin present during interphase. This analysis served to provide further support for the characterization of new regions as DA as well as continued investigation into possible connections between chromatin accessibility during metaphase and structural features present during interphase. Previously identified DA loci were reported to have significantly lower enrichment of marks of open chromatin (DNase 1 HS, FAIRE, and histone modifications H3K4me, H3K9ac, H3K27ac, and H3K4me2) compared to loci with equivalent accessibility⁷. The same general trend was observed for the DA loci identified in this study, with significantly lower mean integrated intensity values of DNase I, H3K9ac, H3K27ac, and H3K4me2 at DA loci compared to previously published EA regions. FAIRE and H3K4me integrated intensity values of DA regions had a lower mean value than EA regions however this difference was not statistically significant. The consistency of the trend between open chromatin mark levels of previously identified DA regions and those identified in this study further supports the classification of each previously unidentified region as differentially accessible.

The loci, *SCAMP2_IVS1*, was removed from the DA data set before analysis, as it differed substantially in the enrichment of open chromatin marks, DNase I HS, FAIRE, H3K9ac, H3K27ac, and H3K4me2, compared to other DA loci. Enrichment of these marks were indicative of open chromatin whereas other DA loci, identified in this study and previously published⁷, did not show the same level of enrichment. It is not known at this time

why this particular DA locus is so different in open chromatin mark enrichment compared to others characterized. However, previous work in our laboratory has shown that inhibiting histone dephosphorylation, deacetylation, and demethylation in interphase did not affect the presence of DA during metaphase, and that the difference between homologues is due to a difference in chromatin supercoiling facilitated by topoisomerase II α ⁸. These data show that individual histone modifications did not have an effect on the formation of DA loci during metaphase, and a difference in enrichment of a given histone modification did not dictate the formation of a given DA locus. An important consideration is that the data assembled for all open chromatin marks were from an EBV immortalized lymphoblastoid B-cell line, while DA data were reported from non-immortalized PHA-activated T lymphocytes. We assume that these are the same, given that they are from the same tissue (peripheral blood), however there could be differences in open chromatin mark enrichment profiles between T-cells and B-cell derived lymphoblasts. This may be the case at *SCAMP2_IVS1*. In terms of an identifying factor of DA and EA loci, a DA region with clear enrichment of open chromatin marks but with scoring data that clearly indicates a DA locus may suggest a more variable enrichment of open chromatin marks across DA and EA regions than currently represented and therefore may not alone be specific enough to DA regions in metaphase to provide a distinct profile to identify new DA loci. Another unrelated sample should be investigated to confirm the characterization of *SCAMP2_IVS1* as DA.

4.2 Definition of Six DA Domains

Using the new DA loci identified and validated as described in the results, 6 DA domains were defined in this study. DA domains were defined in this study as sections of the genome over which multiple DA loci have been identified in neighbouring single-copy regions with the

boundary of each domain demarcated by the genome coordinates of the two DA probes that are the greatest distance apart. An important caveat of the definition of the boundaries of these domains, is that the ends of each domain were not defined by identification of flanking EA regions but rather by the outer DA loci for each domain. This was partly due to the method of scFISH being a time intensive process, taking multiple weeks to go from probe design and development to analysis of scFISH images. This was also a constraint in incrementally increasing the borders of each domain until EA regions were identified. The domains identified in this study, however, provide evidence of DA regions that extend beyond the borders of individual scFISH probes and may occupy much larger genomic lengths than the evidence identifying DA regions with short scFISH probes implied.

The domains extend over genes and intergenic regions with DA loci confirmed in intronic, exonic and intergenic regions. These along with previously identified DA loci, demonstrate DA in gene coding and non-coding areas of the genome. Domains range in size from ~16.0 to ~129.6 kb, defined by 2 to 4 DA scFISH probes. The shortest of these domains, FGF6, TPM1, and XDH covering ~16.0, ~16.0 and ~25.5 kb, respectively, are evidence of multiple neighbouring DA loci in close proximity. The TPM1 domain contains 3 DA regions all located within 16kb, with less than 2kb separating *TPM1_IVS5-IVS8* and *TPM1_IVS8*. This along with the close proximity of the 2 DA regions of the FGF6 domain and the 3 DA regions of the XDH domain suggests that differential accessibility is a structural characteristic of metaphase chromatin that extends beyond the boundaries of the scFISH probes used to identify it.

Presently scFISH probes are the only way to detect DA, and they detect genomic lengths from 1.5 to 5 kb in length. Domains provide evidence that the differential accessibility

detected by these single-copy probes is a larger feature than the boundaries of individual scFISH probes. Initial characterization of DA listed some pairs of DA loci in close proximity, the initial design of these scFISH probes being used to investigate areas of pathologically relevant loci to identify cytogenetic abnormalities^{7,9,10}. These areas were not addressed further during the initial characterization of DA loci. Through the domains identified in the present study, this suggests that the DA probes in close proximity listed in the previous study may also be part of a larger DA domain, rather than individual short lengths of differences in accessibility. In the other three domains, HMGB1P5, COX5A, and HMGB1P1, there are larger gaps between the DA regions identified using scFISH. These areas may not be characterized as continuously differentially accessible across the defined domain. However their proximity, as either a large DA domain or multiple smaller DA domains close together, suggest that DA occurs more often within neighbouring single-copy regions than the placement of previously reported DA regions described.

As mentioned above none of the DA loci defining each domain are directly adjacent to each other and there is varying lengths of genome sequence separating them. The shortest distance between DA loci is ~ 1.4kb, and the largest ~87.3 kb. Due to constraints in designing single-copy FISH probes, regions directly in tandem were rarely able to be produced. This was mostly due to repetitive elements present within these regions. ScFISH probes are designed to hybridize to unique sequence within the genome, only occurring once. Therefore, sequences containing repetitive elements with divergent sequences < 20% are excluded to avoid nonspecific cross-hybridization across the genome. The possibility of cross-hybridization has also been reduced by including a 300 bp cushion from surrounding repetitive elements adjacent to each sc interval⁶⁵. This is known to reduce the probability of probe segments being extended

during amplification to include those repetitive elements⁶⁵. There was also a restriction on probe size, as scFISH probes cover short genomic segments (1.5 to 5 kb). Minimum probe length was kept close to ~ 1.5kb in length in order to readily visualize the hybridization with epifluorescence microscopy¹⁰.

After establishing the presence of DA domains in mitotic metaphase, the potential correspondence between DA domains and higher-order chromatin structures present during interphase was examined. Larger chromatin organizations that are involved in the regulation of cell function, including topologically associated domains (TADs), were considered as DA loci were determined to occupy larger areas than that of scFISH probes. Besides modifications at the nucleotide and histone levels, the increasing hierarchy of interphase chromatin organization including chromatin loops, TADs, and A/B compartments are also important for coordinated expression of genes in cells³⁰. TADs in particular with their compartmentalization of expression programming, allow contact within the domain but insulate gene targets from influence beyond TAD boundaries^{22,30}. These important structures are lost during mitosis (when cells are largely transcriptionally inactive), to allow condensation of chromatin^{41,42}. The mechanisms responsible for re-establishing these structures in daughter cells are still unknown. The location of DA domains and interphase chromatin organizations were examined to investigate whether the extent of DA domains and TAD boundaries were related to one another.

Of the 6 domains, 5 were found to occupy areas each within their own respective TAD. The HMGB1P5 domain was found in the interval between two TADs. None of the DA loci were found to cross any of the given TAD boundaries in the lymphoblastoid cell line analyzed. The XDH, FGF6, COX5A, and HMGB1P1 DA domains had a high degree of interaction in

interphase with other segments within the same TAD, suggesting sub-TAD structures with high frequency intra-TAD interactions occurring within a single TAD. The TPM1 domain interacted with some areas within the TAD. However, these interactions were not as strong compared to the other domains found within TADs. The HMGB1P5 domain, found between TADs, lacked high frequency interactions observed over the domains in metaphase. These findings are all qualitative associations consistent with a connection between the location of TADs in interphase and DA loci in metaphase.

A more in-depth investigation should be completed to begin to answer if there is a connection between these interphase and metaphase structures. Determination of a change in the accessibility between homologues in the adjacent TAD would provide stronger evidence of a possible connection to DA domains. New scFISH probes would be designed for the COX5A and TPM1 domains as both are close to the edge of their respective TAD and into the closest adjacent TAD. If results using these new probes suggest alternating TADs during interphase, consistent with alternating accessibility between metaphase homologues, that would be more compelling evidence of DA as a potential structural memory of TAD structures through metaphase. No change between adjacent TADs, would be inconsistent with the connection just described. In addition, the HMGB1P1 domain, found in the center of a TAD, could be expanded beyond the boundaries of the sub-TAD directly above the domain into more distant areas of intra-TAD interactions. Changes in metaphase accessibility within a single TAD, could suggest DA as a mitotic memory for the more minute sub-TADs within each TAD rather than the larger TAD itself. Observations in this study are based on 6 domains. Therefore, addressing if there is a true connection between DA domains and TADs would require a more expansive analysis of metaphase chromatin accessibility genome-wide and Hi-C chromatin

conformation data describing TAD structures. A full genome investigation into the accessibility between metaphase homologues would even further increase the ability to compare the occurrence of TADs and that of DA.

Potential relationships between the DA domains, TADs, and higher-order chromatin structures were investigated further by comparing the distributions of CTCF binding site clusters within DA domains and the location beyond the boundaries of those DA domains. CTCF binding to chromatin facilitates the formation of TADs and other chromosomal loops^{22,30,89}. TAD boundaries have been characterized by enrichment of CTCF acting to insulate the interior of the TAD from external genomic regulators, separating independently regulated transcriptional regions from one another²². The removal of CTCF binding sites during interphase leads to the loss of TAD structures⁸⁹. However, the relationship between CTCF binding and loop formation in mitotic chromosomes is controversial due to contradictory evidence of CTCF binding during mitosis⁹⁰. A recent study, however, presents evidence that CTCF-site specific binding is lost during mitosis⁹⁰. The ability to re-establish CTCF binding at specific sites is crucial to restore correct TAD and chromatin looping after cell division. The results do not support a significant difference between the distributions of predicted CTCF binding sites within a domain, in the adjacent TAD and within the same TAD beyond the borders of the domain. Notably, DA probes defining the COX5A and TPM1 domains located near the edge of their respective TADs, in addition to the HMGB1P5 domain located close to an adjacent TAD, are all near boundaries expected to be enriched with CTCF.

Despite the lack of correspondence between predicted CTCF binding site distributions and TAD boundaries, future investigations into the connection between CTCF binding sites and DA domains still deserve consideration. The genomic segments selected for comparison

were selected based off their relative location to DA domains and interphase TAD structures. They had unknown homologous chromosome metaphase accessibility, so comparisons could have been made between different differentially accessible domains. Also, data were only analyzed within the domain and to 1 Mb on either side of the domain. Repeating this analysis by expansion into known EA loci and previously identified DA loci, or even expanding genome wide, would allow for a larger sample size and comparison between binding site affinities in DA and EA loci. There is also the consideration that CTCF binding sites are not only found at TAD boundaries but rather throughout the genome promoting localized chromatin looping and accessibility for transcription factors. Another limitation of this study was that the DA domains defined were not all located near the boundaries of the TADs in which they were observed. When more extensive data of the accessibility between metaphase homologues is generated, a more robust comparison can be conducted.

An alternative domain structure that DA domains could be compared with is a newly reported chromatin domain described by *Nozaki et al.* 2017². These are dynamic domains of ~160nm formed by compaction of nucleosomes whose organization is influenced by a number of factors during interphase including cohesion and nucleosome-nucleosome interactions². They were identified using live cell imaging of nucleosomes by tracking a modified H2B histone². The main point of interest is nucleosome domains similar in structure to those observed in interphase were also observed in mitotic cells, both in fixed and live cells². This group proposes that these domains remain throughout the cell cycle acting as “building blocks” of chromosomes². Early establishment of these nucleosome domains was proposed, as the domains were not clear structures in mouse embryonic stem cells, but upon differentiation to embryonic bodies (aggregates of pluripotent cells resembling early embryos) the domains

became more defined². Early establishment of DA would also be anticipated, if DA is a stable mark, with the evidence of DA in both mesoderm and ectoderm derived cells. A similarity was drawn between these domains and TADs in the involvement of cohesion in their formation, though the preservation of these domains into metaphase separates the two. This group speculates that the role of domain formation in mitosis is taken up by condensin as cohesion largely dissociates during mitosis to allow chromosome condensation^{2,91}. Mutations in cohesion have been shown previously to not affect the presence of DA in metaphase with DA resulting from a difference in supercoiling facilitated by topoisomerase II α ⁸. The other major protein involved in chromatin condensation during mitosis, condensin, was unable to be studied due to mutations in this protein resulting in a large disruption in chromatin structure and mislocalization of topoisomerase II α ⁸. The observed conservation of these domains into mitosis through the loss of cohesion, and the possible role of condensin, a key protein involved in condensing chromatin during mitosis, offers a potential chromatin structure to investigate further and compare with DA domains.

4.3 Conservation of DA Loci Across Tissue Types

The presence of DA loci had previously only been reported in peripheral T-lymphocyte cells. This study found DA was conserved at the same loci across 3 different tissues types: lymphocytes, mitotic cells from bone marrow, and dermal fibroblasts. Confirmation that DA is not a feature specific to lymphocytes is supported as it occurs in cells at varying stages of differentiation as well as in cells originating from different germ layers during development.

The expression of each gene in which a DA or EA loci had been identified was examined using data from two different public databases. Data from both databases, GTEx and Human Protein Atlas, was included to compare expression between tissues at each locus. These

data were used to select DA loci with no to low levels of expression across all three tissue types to avoid the complication of potential result interpretation by differences in expression between tissues. Direct expression levels of the tissues examined in this study were not measured.

Three of these DA loci which occurred within genes as well as two DA loci found in intergenic regions were found to be present in all tissues. The conservation of DA loci between peripheral lymphocytes and bone marrow suggests the presence of DA at multiple levels of hematopoiesis. The mitotic cells present in bone marrow would have not only included B lymphocytes but also various lymphocyte progenitor cells at different stages of differentiation as well as those of cells of myeloid origin. Of note is that bone marrow as a sample type is a heterogenous mixture of cells dividing at various stages of differentiation and this study did not seek to isolate or identify the identities of the mitotic cell subtypes comprising the mixture of each sample. If the stem cell composition was biased towards a predominantly pre-lymphocytic progenitor population, the observed conservation of DA loci would have been anticipated.

In addition to cells from the both PHA stimulated T-lymphocytes and bone marrow cells, there was also conservation of DA loci in dermal fibroblasts. This provides evidence of DA in cells derived from both the mesoderm (lymphocytes) and ectoderm (fibroblasts). All of the selected DA loci that were shown to be conserved either had 0 to low transcription in all 3 tissues or were in intergenic regions. Also conserved was the EA region located within a non-coding region. As stated earlier, all the regions investigated in this study were selected because they showed either no or low gene expression levels in order to avoid complicating data interpretation with presence of differential expression between tissues. Further analysis that includes an additional EA control within a gene would be ideal in order to determine, if similar

to DA, EA loci are maintained in both genes and intergenic regions. Expanding the regions investigated to also include DA loci in regions with expression greater than 5 TPM would strengthen characterization of DA across different tissues. Conservation has not been examined in genes with expression >5 TPM, though the presence of DA in intergenic regions does not support a role for DA in tissue-specific programming. Regardless, these data suggest that DA is a feature present between homologous chromosomes during metaphase across a variety of cell types and is not concentrated only in lymphocytic tissues. The conservation of DA between tissue types opens more avenues of investigation into the origin and relevance of DA.

Of the 5 DA loci compared between these tissues, 4 were involved in defining 3 different DA domains. This included one locus each, from the XDH domain and the TPM1 domain, and 2 loci from the HMGB1P1 domain. The presence of these loci across different tissues types suggests that not only are loci with DA maintained but also larger DA domains defined in this study are also conserved between these tissue types. The location of TAD boundaries is well-maintained between tissue types whereas the location of sub-TAD interactions are cell-type specific^{21,22,30,92}. The establishment of TADS has been identified early in human embryogenesis, with well-defined TADs occurring by the 8-cell stage⁹². If metaphase chromatin accessibility between homologues is stable related to TAD location through multiple cell divisions and differentiation, then DA could be maintained between different tissue types originating from different germ layers of the early embryo. Although a number of high frequency intra-TAD interactions, indicating sub-TAD structures, are observed over the location of 4 of 5 DA domains during interphase, the evidence of DA in lymphocytes and fibroblasts suggest a possible relationship between the larger TAD structures rather than the internal cell-type specific sub-TADs. The identification of conserved DA loci has been

shown across genes with no or low expression and intergenic regions. DA conservation between tissues in genes showing differential expression between tissues was not studied.

An area of interest is the examination of DA loci across different stages of differentiation of lymphocytes, as well as between activated and inactivated mature lymphocytes. DA loci conserved in lymphocytes and bone marrow suggests that DA is present in progenitor B lymphocytes as well as the mature cell type, identified in both inactivated B and activated T lymphocytes. Changes in chromatin accessibility are not limited to stages of hematopoietic stem cell differentiation, large changes have also been observed and characterized in both B cells and T cells following activation with an antigen^{93,94}. This presents a specification pathway with multiple stages of chromatin rearrangements over which the conservation of DA can be investigated. The presence of DA in tissues derived from two different germ layers also brings up the question of when during development is DA established. If DA is a stable mark through development then observation of DA in both mesoderm and ectoderm derived cells would indicate early establishment in embryogenesis. Establishing the stages of development and differentiation that DA is present at the same loci would indicate when DA is established and where it is conserved. Conserved across the selection of tissues in this study, as well as potential broader conservation across other cell types at the same loci would support a deliberate establishment of specific areas of equivalent and differential accessibilities between metaphase homologues during cell division.

4.4 Conclusion

Differential accessibility (DA) between metaphase homologous chromosomes is a unique feature of metaphase chromatin. Initially observed in peripheral lymphocytes, DA is a stable locus-specific and heritable feature that has been observed in ~10% of >300 single-copy loci

investigated. The results of this study were able to show that DA is not confined to the boundaries of short scFISH probes but rather extend to form larger, contiguous domains with the same differential accessibility epigenotype between homologues. Large regions of DA were identified and defined, however clear boundaries that separate these domains from adjacent areas of equivalent accessibility were not determined. The location of 5 of these domains were found within their own respective TADs. Previously published work from our laboratory has demonstrated that 90% of single-copy probes derived from regions containing clinically relevant cytogenetic genes/regions exhibit EA⁷. It is therefore highly likely that further expansion from current DA domains will reveal these boundaries for the domains described here.

This study also established that these DA loci are conserved between blood, bone marrow, and fibroblast tissues. This suggests that differentially accessible loci are found at different stages of hematopoiesis and lymphocyte differentiation as well as in tissues derived from multiple germ layers present during early development. The identification of DA regions that are preserved between different cell types, and across the genome within genes and intergenic regions does not support a tissue-specific role of DA. The findings are consistent with the hypothesis that DA might contain or harbour structural marks of chromatin memory. The expansion of DA loci into domains and their conservation between tissue types further characterizes this metaphase structure.

From the data reported, multiple avenues could be pursued to further investigate the role of equivalent and differential accessibility between metaphase homologues. DA and EA could be the result of mitotic bookmarking, with specific loci maintained through mitosis in multiple tissue types to allow rapid restoration of necessary transcription programming

immediately following mitosis. Or it could serve in interphase as mitotic memory of functional chromatin domains, such as TADs, to preserve their exact location so following their loss in metaphase these structures can be faithfully restored in daughter cells. Disruption of any of these processes could be the cause of abnormal functions at multiple levels. Examination of aberrant phenotypes have already begun in which a dysregulated gene is not the result of a mutation within the gene itself but rather a disruption of distant regulatory elements or chromatin organization facilitating abnormal contact of regulatory elements with unintended genes⁵². Further understanding of the mechanisms and structures that allow rapid and precise re-establishment of chromatin organizations and modifications, necessary for correct regulation and function of cells, is essential.

REFERENCES

1. The 4D Nucleome Network *et al.* The 4D nucleome project. *Nature* **549**, 219–226 (2017).
2. Nozaki, T. *et al.* Dynamic Organization of Chromatin Domains Revealed by Super-Resolution Live-Cell Imaging. *Mol. Cell* **67**, 282–293.e7 (2017).
3. Yu, S., Yang, F. & Shen, W. H. Genome maintenance in the context of 4D chromatin condensation. *Cell. Mol. Life Sci.* **73**, 3137–3150 (2016).
4. Babu, D. & Fullwood, M. J. 3D genome organization in health and disease: emerging opportunities in cancer translational medicine. *Nucleus* **6**, 382–393 (2015).
5. Naumova, N. *et al.* Organization of the Mitotic Chromosome. *Science* **342**, 948–953 (2013).
6. Golloshi, R., Sanders, J. T. & McCord, R. P. Genome organization during the cell cycle: unity in division. *Wiley Interdiscip. Rev. Syst. Biol. Med.* **9**, e1389 (2017).
7. Khan, W. A., Rogan, P. K. & Knoll, J. H. Localized, non-random differences in chromatin accessibility between homologous metaphase chromosomes. *Mol. Cytogenet.* **7**, (2014).
8. Khan, W. A., Rogan, P. K. & Knoll, J. H. M. Reversing chromatin accessibility differences that distinguish homologous mitotic metaphase chromosomes. *Mol. Cytogenet.* **8**, (2015).
9. Rogan, P. K., Cazcarro, P. M. & Knoll, J. H. M. Sequence-Based Design of Single-Copy Genomic DNA Probes for Fluorescence In Situ Hybridization. *Genome Res.* **11**, 1086–1094 (2001).
10. Knoll, J. H. M. & Rogan, P. K. Sequence-Based, in situ detection of chromosomal abnormalities at high resolution. *Am. J. Med. Genet. A.* **121A**, 245–257 (2003).

11. Vermeulen, K., Bockstaele, D. R. V. & Berneman, Z. N. The cell cycle: a review of regulation, deregulation and therapeutic targets in cancer. *Cell Prolif.* **36**, 131–149 (2003).
12. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61 (2017).
13. Ou, H. D. *et al.* ChromEMT: Visualizing 3D chromatin structure and compaction in interphase and mitotic cells. *Science* **357**, eaag0025 (2017).
14. Gibcus, J. H. *et al.* A pathway for mitotic chromosome formation. *Science* **359**, eaao6135 (2018).
15. Goloborodko, A., Marko, J. F. & Mirny, L. A. Chromosome Compaction by Active Loop Extrusion. *Biophys. J.* **110**, 2162–2168 (2016).
16. Li, G., Sudlow, G. & Belmont, A. S. Interphase Cell Cycle Dynamics of a Late-Replicating, Heterochromatic Homogeneously Staining Region: Precise Choreography of Condensation/Decondensation and Nuclear Positioning. *J. Cell Biol.* **140**, 975–989 (1998).
17. Richmond, T. J. & Davey, C. A. The structure of DNA in the nucleosome core. *Nature* **423**, 145–150 (2003).
18. Maeshima, K., Ide, S., Hibino, K. & Sasai, M. Liquid-like behavior of chromatin. *Curr. Opin. Genet. Dev.* **37**, 36–45 (2016).
19. Nozaki, T. *et al.* Flexible and dynamic nucleosome fiber in living mammalian cells. *Nucleus* **4**, 349–356 (2013).
20. Nuebler, J., Fudenberg, G., Imakaev, M., Abdennur, N. & Mirny, L. A. Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci.* **115**, E6697–E6706 (2018).

21. Rao, S. S. P. *et al.* A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping. *Cell* **159**, 1665–1680 (2014).
22. Dixon, J. R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376–380 (2012).
23. Ramani, V., Shendure, J. & Duan, Z. Understanding Spatial Genome Organization: Methods and Insights. *Genomics Proteomics Bioinformatics* **14**, 7–20 (2016).
24. Kornberg, R. D. Chromatin Structure: A Repeating Unit of Histones and DNA. *Science* **184**, 868–871 (1974).
25. McManus, K. J. & Hendzel, M. J. The relationship between histone H3 phosphorylation and acetylation throughout the mammalian cell cycle This paper is one of a selection of papers published in this Special Issue, entitled 27th International West Coast Chromatin and Chromosome Conference, and has undergone the Journal's usual peer review process. *Biochem. Cell Biol.* **84**, 640–657 (2006).
26. Finch, J. T. & Klug, A. Solenoidal model for superstructure in chromatin. *Proc. Natl. Acad. Sci.* **73**, 1897–1901 (1976).
27. Song, F. *et al.* Cryo-EM Study of the Chromatin Fiber Reveals a Double Helix Twisted by Tetranucleosomal Units. *Science* **344**, 376 (2014).
28. Grigoryev, S. A. *et al.* Hierarchical looping of zigzag nucleosome chains in metaphase chromosomes. *Proc. Natl. Acad. Sci.* **113**, 1238–1243 (2016).
29. Audugé, N., Padilla-Parra, S., Tramier, M., Borghi, N. & Coppey-Moisan, M. Chromatin condensation fluctuations rather than steady-state predict chromatin accessibility. *Nucleic Acids Res.* **47**, 6184–6194 (2019).

30. Dixon, J. R., Gorkin, D. U. & Ren, B. Chromatin Domains: The Unit of Chromosome Organization. *Mol. Cell* **62**, 668–680 (2016).
31. Ma, H. *et al.* Spatial and Temporal Dynamics of DNA Replication Sites in Mammalian Cells. *J. Cell Biol.* **143**, 1415–1425 (1998).
32. Jackson, D. A. & Pombo, A. Replicon Clusters Are Stable Units of Chromosome Structure: Evidence That Nuclear Organization Contributes to the Efficient Activation and Propagation of S Phase in Human Cells. *J. Cell Biol.* **140**, 1285–1295 (1998).
33. Nora, E. P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381–385 (2012).
34. Sexton, T. *et al.* Three-Dimensional Folding and Functional Organization Principles of the Drosophila Genome. *Cell* **148**, 458–472 (2012).
35. Pope, B. D. *et al.* Topologically associating domains are stable units of replication-timing regulation. *Nature* **515**, 402–405 (2014).
36. Lieberman-Aiden, E. *et al.* Comprehensive Mapping of Long-Range Interactions Reveals Folding Principles of the Human Genome. *Science* **326**, 289–293 (2009).
37. Fortin, J.-P. & Hansen, K. D. Reconstructing A/B compartments as revealed by Hi-C using long-range correlations in epigenetic data. *Genome Biol.* **16**, (2015).
38. Dolan, M. The role of the Giemsa stain in cytogenetics. *Biotech. Histochem.* **86**, 94–97 (2011).
39. Ganji, M. *et al.* Real-time imaging of DNA loop extrusion by condensin. *Science* **360**, 102 (2018).
40. Takahashi, M. & Hirota, T. Folding the genome into mitotic chromosomes. *Curr. Opin. Cell Biol.* **60**, 19–26 (2019).

41. Hsiung, C. C.-S. *et al.* A hyperactive transcriptional state marks genome reactivation at the mitosis–G1 transition. *Genes Dev.* **30**, 1423–1439 (2016).
42. Festuccia, N., Gonzalez, I., Owens, N. & Navarro, P. Mitotic bookmarking in development and stem cells. *Development* **144**, 3633–3645 (2017).
43. Lodhi, N., Ji, Y. & Tulin, A. Mitotic Bookmarking: Maintaining Post-Mitotic Reprogramming of Transcription Reactivation. *Curr. Mol. Biol. Rep.* **2**, 10–15 (2016).
44. Liu, Y. *et al.* Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *Cell Rep.* **19**, 1283–1293 (2017).
45. Martínez-Balbás, M. A., Dey, A., Rabindran, S. K., Ozato, K. & Wu, C. Displacement of sequence-specific transcription factors from mitotic chromatin. *Cell* **83**, 29–38 (1995).
46. Kelly, T. K. *et al.* H2A.Z Maintenance during Mitosis Reveals Nucleosome Shifting on Mitotically Silenced Genes. *Mol. Cell* **39**, 901–911 (2010).
47. Wang, F. & Higgins, J. M. G. Histone modifications and mitosis: countermarks, landmarks, and bookmarks. *Trends Cell Biol.* **23**, 175–184 (2013).
48. Bellec, M., Radulescu, O. & Lagha, M. Remembering the past: Mitotic bookmarking in a developing embryo. *Curr. Opin. Syst. Biol.* **11**, 41–49 (2018).
49. Oomen, M. E. & Dekker, J. Epigenetic characteristics of the mitotic chromosome in 1D and 3D. *Crit. Rev. Biochem. Mol. Biol.* **52**, 185–204 (2017).
50. Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci.* **107**, 21931–21936 (2010).
51. Lupiáñez, D. G. *et al.* Disruptions of Topological Chromatin Domains Cause Pathogenic Rewiring of Gene-Enhancer Interactions. *Cell* **161**, 1012–1025 (2015).

52. Lupiáñez, D. G., Spielmann, M. & Mundlos, S. Breaking TADs: How Alterations of Chromatin Domains Result in Disease. *Trends Genet.* **32**, 225–237 (2016).
53. Winick-Ng, W. & Rylett, R. J. Into the Fourth Dimension: Dysregulation of Genome Architecture in Aging and Alzheimer's Disease. *Front. Mol. Neurosci.* **11**, (2018).
54. Guo, Y. *et al.* CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. *Cell* **162**, 900–910 (2015).
55. Giorgio, E. *et al.* A large genomic deletion leads to enhancer adoption by the lamin B1 gene: a second path to autosomal dominant adult-onset demyelinating leukodystrophy (ADLD). *Hum. Mol. Genet.* **24**, 3143–3154 (2015).
56. Spielmann, M. & Mundlos, S. Structural variations, the regulatory landscape of the genome and their alteration in human disease. *BioEssays* **35**, 533–543 (2013).
57. Flöttmann, R. *et al.* Microdeletions on 6p22.3 are associated with mesomelic dysplasia Savarirayan type. *J. Med. Genet.* **52**, 476–483 (2015).
58. Klein, H.-U. *et al.* Epigenome-wide study uncovers large-scale changes in histone acetylation driven by tau pathology in aging and Alzheimer's human brains. *Nat. Neurosci.* **22**, 37–46 (2019).
59. Huber, D., Voith von Voithenberg, L. & Kaigala, G. V. Fluorescence in situ hybridization (FISH): History, limitations and what to expect from micro-scale FISH? *Micro Nano Eng.* **1**, 15–24 (2018).
60. Kearney, L. Molecular cytogenetics. *Best Pract. Res. Clin. Haematol.* **14**, 645–668 (2001).
61. Arachchige, A. S., Samarabandu, J., Knoll, J., Khan, W. & Rogan, P. An image processing algorithm for accurate extraction of the centerline from human metaphase

- chromosomes. in *2010 IEEE International Conference on Image Processing* 3613–3616 (2010). doi:10.1109/ICIP.2010.5652017.
62. Perez, J. D., Rubinstein, N. D. & Dulac, C. New Perspectives on Genomic Imprinting, an Essential and Multifaceted Mode of Epigenetic Control in the Developing and Adult Brain. *Annu. Rev. Neurosci.* **39**, 347–384 (2016).
63. Ishida, R. *et al.* Inhibition of DNA topoisomerase II by ICRF-193 induces polyploidization by uncoupling chromosome dynamics from other cell cycle events. *J. Cell Biol.* **126**, 1341 (1994).
64. Khrustaleva, L., Jiang, J. & Havey, M. J. High-resolution tyramide-FISH mapping of markers tightly linked to the male-fertility restoration (Ms) locus of onion. *Theor. Appl. Genet.* **129**, 535–545 (2016).
65. Dorman, S. N., Shirley, B. C., Knoll, J. H. M. & Rogan, P. K. Expanding probe repertoire and improving reproducibility in human genomic hybridization. *Nucleic Acids Res.* **41**, e81–e81 (2013).
66. Ye, J. *et al.* Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, (2012).
67. UCSC Genome Browser Home. <https://genome.ucsc.edu/index.html>.
68. NCBI Resource Coordinators *et al.* Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **46**, D8–D13 (2018).
69. Knoll, J. H. M., Lichter, P., Bakdounes, K. & Eltoum, I.-E. A. In Situ Hybridization and Detection Using Nonisotopic Probes. *Curr. Protoc. Mol. Biol.* **79**, 14.7.1-14.7.17 (2007).
70. Dunnen, J. T. den & Antonarakis, S. E. Mutation nomenclature extensions and suggestions to describe complex mutations: A discussion. *Hum. Mutat.* **15**, 7–12 (2000).

71. Wang, Y. X., Hill, S. L., Rogan, P. K. & Knoll, J. H. Expanding the catalogue of differentially compacted loci on homologous chromosomes in the human genome. (2018).
72. Rytönen, E. M. *et al.* The human gene for xanthine dehydrogenase (XDH) is localized on chromosome band 2q22. *Cytogenet. Cell Genet.* **68**, 61–63 (1995).
73. Rogalla, P. *et al.* Mapping and molecular characterization of five HMG1-related DNA sequences. *Cytogenet. Genome Res.* **83**, 124–129 (1998).
74. Chaffanet, M. *et al.* Mapping of an ordered set of 14 cosmids to human chromosome 12p by two-color in situ hybridization. *Cytogenet. Genome Res.* **69**, 27–32 (1995).
75. Lee, F. K., Cheung, M. C. & Chung, S. The Human Sorbitol Dehydrogenase Gene: cDNA Cloning, Sequence Determination, and Mapping by Fluorescence in Situ Hybridization. *Genomics* **21**, 354–358 (1994).
76. Hofmann, S., Lichtner, P., Schuffenhauer, S., Gerbitz, K.-D. & Meitinger, T. Assignment¹ of the human genes coding for cytochrome c oxidase subunits Va (COX5A), VIc (COX6C) and VIIc (COX7C) to chromosome bands 15q25, 8q22→q23 and 5q14 and of three pseudogenes (COX5AP1, COX6CP1, COX7CP1) to 14q22, 16p12 and 13q14→q21 by FISH and radiation hybrid mapping. *Cytogenet. Genome Res.* **83**, 226–227 (1998).
77. Eyre, H. *et al.* Assignment of the human skeletal muscle α -tropomyosin gene (TPM1) to band 15q22 by fluorescence in situ hybridization. *Cytogenet. Genome Res.* **69**, 15–17 (1995).
78. Lonsdale, J. *et al.* The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
79. Uhlén, M. *et al.* Tissue-based map of the human proteome. *Science* **347**, 1260419 (2015).

80. E-MTAB-2836 < Browse < ArrayExpress < EMBL-EBI.
<https://www.ebi.ac.uk/arrayexpress/experiments/E-MTAB-2836/>.
81. Brown, M. G. & Jennings Lawce, H. Chapter 3 Peripheral Blood Cytogenetic Methods. in *The AGT cytogenetics laboratory manual*. (eds. Barch, M. J., Knutsen, Turid., Spurbeck, J. L. & Association of Genetic Technologists.) (Lippincott-Raven Publishers, 1997).
82. Roulston, D. & Le Beau, M. M. Chapter 7 Cytogenetic Analysis of Hematologic Malignant Diseases. in *The AGT cytogenetics laboratory manual*. (eds. Barch, M. J., Knutsen, Turid., Spurbeck, J. L. & Association of Genetic Technologists.) (Lippincott-Raven Publishers, 1997).
83. Priest, J. H. Chapter 4 General Cell Culture Principles and Fibroblast Culture. in *The AGT cytogenetics laboratory manual*. (eds. Barch, M. J., Knutsen, Turid., Spurbeck, J. L. & Association of Genetic Technologists.) (Lippincott-Raven Publishers, 1997).
84. Johnson, G. D. *et al.* Fading of immunofluorescence during microscopy: a study of the phenomenon and its remedy. *J. Immunol. Methods* **55**, 231–242 (1982).
85. The ENCODE Project Consortium. A User’s Guide to the Encyclopedia of DNA Elements (ENCODE). *PLoS Biol.* **9**, e1001046 (2011).
86. Wang, Y. *et al.* The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol.* **19**, (2018).
87. Lu, R., Mucaki, E. J. & Rogan, P. K. Discovery and validation of information theory-based transcription factor and cofactor binding site motifs. *Nucleic Acids Res.* **45**, e27–e27 (2017).

88. Thomson, J. A. Embryonic Stem Cell Lines Derived from Human Blastocysts. *Science* **282**, 1145–1147 (1998).
89. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944.e22 (2017).
90. Oomen, M. E., Hansen, A. S., Liu, Y., Darzacq, X. & Dekker, J. CTCF sites display cell cycle–dependent dynamics in factor binding and nucleosome positioning. *Genome Res.* **29**, 236–249 (2019).
91. Nakajima, M. *et al.* The complete removal of cohesin from chromosome arms depends on separase. *J. Cell Sci.* **120**, 4188–4196 (2007).
92. Ke, Y. *et al.* 3D Chromatin Structures of Mature Gametes and Structural Reprogramming during Mammalian Embryogenesis. *Cell* **170**, 367-381.e20 (2017).
93. Kieffer-Kwon, K.-R. *et al.* Myc Regulates Chromatin Decompaction and Nuclear Architecture during B Cell Activation. *Mol. Cell* **67**, 566-578.e10 (2017).
94. Rawlings, J. S., Gatzka, M., Thomas, P. G. & Ihle, J. N. Chromatin condensation via the condensin II complex is required for peripheral T-cell quiescence: Chromatin condensation is required for quiescence. *EMBO J.* **30**, 263–276 (2011).

APPENDICES

Appendix I: Gene Names

COX5A – Cytochrome C Oxidase Subunit 5A

CTCF – CCCTC-Binding Factor Like

DUOX1 – Dual Oxidase 1

FGF6 – Fibroblast Growth Factor 6

HMGB1P1 – High Mobility Group Box 1 Pseudogene 1

HMGB1P5 – High Mobility Group Box 1 Pseudogene 5

PCK1 – Phosphoenolpyruvate Carboxykinase 1

RPM38 – RNA Binding Motif Protein 38

SCAMP2 – Secretory Carrier Membrane Protein 2

TPM1 – Tropomyosin 1

XDH – Xanthine Dehydrogenase

ZNF385D – Zinc Finger Protein 385D

Appendix II: Copyright Release for Use of Image

THE AMERICAN ASSOCIATION FOR THE ADVANCEMENT OF SCIENCE LICENSE TERMS AND CONDITIONS

Dec 16, 2019

This Agreement between Seana Hill ("You") and The American Association for the Advancement of Science ("The American Association for the Advancement of Science") consists of your license details and the terms and conditions provided by The American Association for the Advancement of Science and Copyright Clearance Center.

License Number	4723131041490
License date	Dec 06, 2019
Licensed Content Publisher	The American Association for the Advancement of Science
Licensed Content Publication	Science
Licensed Content Title	A pathway for mitotic chromosome formation
Licensed Content Author	Johan H. Gibcus, Kumiko Samejima, Anton Goloborodko, Itaru Samejima, Natalia Naumova, Johannes Nuebler, Masato T. Kanemaki, Linfeng Xie, James R. Paulson, William C. Earnshaw, Leonid A. Mirny, Job Dekker
Licensed Content Date	Feb 9, 2018
Licensed Content Volume	359
Licensed Content Issue	6376
Volume number	359
Issue number	6376
Type of Use	Thesis / Dissertation
Requestor type	Scientist/individual at a research institution
Format	Print and electronic
Portion	Figure
Number of figures/tables	1

Order reference number	
Title of your thesis / dissertation	For Chapter 1: Introduction; Organization of Chromatin into Metaphase Chromosomes
Expected completion date	Mar 2020
Estimated size(pages)	130
Requestor Location	Seana Hill 4044 Dental Sciences Building 1151 Richmond Street London, ON N6A5C1 Canada Attn: Seana Hill
Total	0.00 USD

Appendix III: Python script reporting mean expression values of RefSeq genes in lymphocytes, bone marrow, and fibroblast cells from the Human Protein Atlas (HPA) and Genotype-Tissue Expression (GTEx) databases

```

import csv
import sys
import statistics

# list of keywords for tissues of interest, general and specific to search through files, taken
from list of tissues (GTEx, Protein Atlas)

tissues_of_interest = ['Blood', 'blood', 'Whole Blood', 'whole blood', 'Cells - EBV-transformed
lymphocytes',
'Skin', 'skin', 'Hematopoietic Stem Cells', 'hematopoietic stem cells', 'Bone marrow', 'bone
marrow',
'Placenta', 'placenta', 'Transformed fibroblasts', 'transformed fibroblasts', 'Cells - Transformed
fibroblasts',
'Skin - Not Sun Exposed (Suprapubic)', 'Skin - Sun Exposed (Lower leg)']
]

lymphocyte = ['Cells - EBV-transformed lymphocytes']
wholeblood = ['Blood', 'blood', 'Whole Blood', 'whole blood']
skin = [ 'Skin', 'skin', 'Skin - Not Sun Exposed (Suprapubic)', 'Skin - Sun Exposed (Lower
leg)']
fibroblast = ['Transformed fibroblasts', 'transformed fibroblasts', 'Cells - Transformed
fibroblasts']
hematopoietic = [ 'Hematopoietic Stem Cells', 'hematopietic stem cells']
bone_marrow = ['Bone marrow', 'bone marrow']
prenatal = ['Placenta', 'placenta']

"""
Make dictionary to allow conversion of sample name to their tissue of origin
"""
handle1 = open('E-MTAB-5214.sdrf.tsv') #open information of tissue type corresponding
to sample name
line = handle1.readline() #skip header

sample = []
tissue_type = []

for line in handle1:
    category = line.strip('\n').split('\t')
    sample.append(category[0]) #keys = sample name
    tissue_type.append(category[5]) #values = specific tissue type information

specific_tissuetypes = dict(zip(sample, tissue_type)) #make dictionary to interpret samples as
their tissue types

```

```

handle1.close()

def output_TPM_GTEEx(input, output):
    """
    Input file formatted by GTEEx; extract and report data points only of tissues and genes of
    interest.
    Two file types anticipated:
    1. full data set with all samples where need to identify the tissue of origin of each sample
    reported and
        calculate the mean expression in transcripts per million (TPM) for genes of interest
    2. summarized data with tissue types given where need to isolate and report tissues and
    genes of interest
    """
    handle = open(input) #GTEEx date set
    _ = handle.readline()
    _ = handle.readline()

    samples = handle.readline().strip("\n").split("\t") #isolate each sample name in 3rd line

#lists of each column position of target tissue type samples
lymph = []
blood = []
fibro = []
peau = [] #list for skin
hemato = []
marrow = []
pren = []

for sample in samples: #add positions of target tissues into specific lists
    if sample in tissues_of_interest: #if sample given as tissue types make lists for each
category of tissue/cell
        if sample in wholeblood:
            blood.append(samples.index(sample))
        if sample in lymphocyte:
            lymph.append(samples.index(sample))
        if sample in fibroblast:
            fibro.append(samples.index(sample))
        if sample in skin:
            peau.append(samples.index(sample))
        if sample in hematopoietic:
            hemtao.append(samples.index(sample))
        if sample in bone_marrow:
            marrow.append(samples.index(sample))
        if sample in prenatal:
            pren.append(samples.index(sample))

```

```

else: #if not, convert sample name to tissue type and collect positions of targeted tissues
    tissuetype = specific_tissuetypes.get(sample, None)
    if tissuetype in tissues_of_interest:
        if specific_tissuetypes is None:
            continue
        if tissuetype in wholeblood:
            blood.append(samples.index(sample))
        if tissuetype in lymphocyte:
            lymph.append(samples.index(sample))
        if tissuetype in fibroblast:
            fibro.append(samples.index(sample))
        if tissuetype in skin:
            peau.append(samples.index(sample))
        if tissuetype in hematopoietic:
            hemtao.append(samples.index(sample))
        if tissuetype in bone_marrow:
            marrow.append(samples.index(sample))
        if tissuetype in prenatal:
            pren.append(samples.index(sample))

#list of tissuetype lists to iterate over
tissue_lists = [lymph, blood, fibro, peau, hemato, marrow, pren]

#dictionary of the position lists and corresponding tissue to allow identification of tissue
types in new file
tissue_types = {str(lymph):'Lymphocytes', str(blood):'Whole Blood',
str(fibro):'Fibroblasts',
str(peau):'Skin', str(hemato):'Hematopoietic Stem Cells', str(marrow):'Bone Marrow',
str(pren):'Prenatal'
}

outfile = open (output, 'w') #open new file with results
reader = csv.reader(handle, delimiter='\t', quotechar='')
writer = csv.writer(outfile, delimiter='\t', quotechar='')

results = []
for row in reader:
    gene = row[1] #gene name in 2nd column
    if gene in sys.argv[1:]: #iterate over the gene names within the file, match with those
entered in terminal
        total = 0
        for item in tissue_lists:

            if len(item) == 1: #if only one data point per tissue assume summarized data,
report as given
                for position in item:

```

```

        total = float(row[int(position)])
        tissue = tissue_types.get(str(item), None) #corresponding tissue type to list
item
        result = gene, tissue, total
        results.append(result)

    if len(item) != 0: #if there is a sample that matches a tissue of interest
        listTPM = []
        for position in item:
            TPM = float(row[int(position)])
            tissue = tissue_types.get(str(item), None)
            listTPM.append(TPM)
        result = gene, tissue, total, statistics.mean(listTPM), statistics.stdev(listTPM)
        writer.writerow(result) #write the result to a new file with gene name and tissue
type
    for result in sorted(results):
        writer.writerow(result) #write the result to a new file with gene name and tissue type
in alphabetical order

    handle.close()
    outfile.close()

def output_TPM_HPA(input, output):
    """
    Input data from Human Protien Atlas (HPA) and select transcription data for tissues and
genes of interest
    Downloaded as previously summarized data files
    """

    handle = open(input)
    outfile = open(output, 'w')
    writer = csv.writer(outfile, delimiter='\t', quotechar='')

    labels = handle.readline().strip('\n').split('\t') # column labels in first line
    results = []

    for row in handle:
        column = row.strip('\n').split('\t') #separate each column
        gene_name = column[1]
        tissue_sample = column[2]
        TPM = column[3]

        if gene_name in sys.argv[1:]: #if gene of interest
            if tissue_sample in tissues_of_interest: #if tissue type corresponds to tisses of interst
                result = gene_name, tissue_sample, TPM

```

```

        results.append(result) # add results to results list

for result in sorted(results):
    writer.writerow(result) #write results in alphabetical order

handle.close()
outfile.close()

def compile_results(results, output):
    """
    Takes list of data files generated from above functions and write all to one file.
    """
    outfile = open(output, 'w')

    combined_rows = []
    for file in results: #iterate over files in list of resulting files
        new_file = open(file)
        rows = []
        file_name = str(new_file)+'\n'
        rows.append(file_name)
        for row in new_file:
            rows.append(row)
        combined_rows.extend(rows) #add each individual row from file to new file
    for item in combined_rows:
        outfile.write(item)

    new_file.close()
    outfile.close()

output_TPM_GTEEx('GTEEx_Analysis_2016-01-
15_v7_RNASeQCv1.1.8_gene_tpm.tsv','GTEEx_gene_mean_tpm.tsv')
output_TPM_HPA('rna_tissue.tsv', 'HPA_gene_mean_tpm.tsv')

results = ['GTEEx_gene_mean_tpm.tsv', 'HPA_gene_mean_tpm.tsv'] #list of resulting files to
compile

compile_results(results, 'compiled_tissue_TPMS_Nov6.tsv') #bring all results together into
one file

```

Appendix IV: Details for the production of all new ScFISH probes developed in this study. This includes chromosome band, probe name, probe coordinates and size, with the forward and reverse primers used to make each probe, and the probe's CG and CpG content.

Chr Band	Probe Name	Coordinates	Length (bp)	Forward Primer (5' - 3')	Reverse Primer (5' - 3')	CG(%)	CpG(%)
2p23.1	<i>XDH_tel</i> 9264	chr2:31,545,815-31,547,924	2110	AAGGCAATAAAAGC ACAGAAGACACA	AGAGCTTTCCTCTAA GGCACATAGC	36.0	1.80
	<i>XDH_tel</i> 2387	chr2:31,551,816-31,554,801	2986	GAAAACCTGATTTTG GGACTTAGGAACA	GGATTAGACTGAGG CATTAAAGGTAGTGA	48.0	1.74
	<i>XDH_IVS</i> 30- <i>IVS</i> 27	chr2:31,568,769-31,571,269	2501	AATCCATGCTAAAAC CCAAC TACCAAAA	TTTACTGAATTTTGT GTCTGGCATCCTT	50.0	2.56
3p24.3	<i>ZNF385D_tel</i> 640535	chr3:22,433,351-22,436,333	3968	CATATTTGCCTTTCA TGCTTAGAGACCA	AATACAGGCTCTAAT GCAGAACCATTTC	32.0	0.50
	<i>ZNF385D_tel</i> 678016	chr3:22,470,832-22,473,082	2251	AACCGAGCAAATTC CATA CCAAACC	GCCTGTCACCTTCTG TTATGTGCTC	34.0	0.98
12p12.3	<i>FGF6_cen</i> 4492	chr12:4,537,157-4,538,816	1660	TTAATGTCTGTCTTC CTTGCCAGTTTAT	AACTCCTTCATCCTT CTATCCCAATGTT	47.0	3.49
	<i>FGF6_IVS</i> 2	chr12:4,549776-4,553,205	2251	AGCACACAATTGAA CTGTTGATTTAGGA	CTATCTAAATGGGAA AGGAAAGGCTGTG	52.0	5.69
15q21.1	<i>DUOX1_IVS</i> 1- <i>IVS</i> 3	chr15:45,422,890-45,424,597	1708	GTTATTAAGCATGTC CACCTTCCTCTTC	CAATCTGACCCAAG ATCCTTTATTGTC	49.0	1.29
15q22.2	<i>TPMI_IVS</i> 5- <i>IVS</i> 8	chr15:63,353,573-63,355,980	2408	TTCACCCTCTGCTAT TTATATCTTGCTT	TAAATTGTAAACGA AACCACAGAAGGCT	40.0	1.25
	<i>TPMI_IVS</i> 8	chr15:63,357,346-63,360,645	3300	TTTACTGTTAGGCAG TAAGAGTTGGAGT	TATGTTTTCTCGCTT GGAATATCTGCTG	41.0	1.45
	<i>TPMI_tel</i> 3200	chr15:63,367,314-63,369,607	2294	AGGGTTTT CAGAGCT ACATTCCTCC	TGTCCTATAACCCCC AACCTTGAAA	48.0	3.05
15q24.1	<i>SCAMP2_IV</i> S7- <i>IVS</i> 4	chr15:75,142,349-75,145,350	3002	CATAAACCTGAGGT ACAATTCCAGAAC	TGACCAGTTTGAGAT CTCTCCCTAAAAA	55.0	2.07

	<i>SCAMP2_IV</i> S1	chr15:75,161,78 3-75,163,308	1526	ACTCTCAAGAATGGT GACAACCTGTG	CATTAACCTCACCTG CTTGTCCTTT	45.0	1.18
15q24. 2	<i>COX5A_tel2</i> 0100	chr15:75,250,59 5-75,252,319	1725	GAAGCATTAGAGAA AGATGAGATTGGGG	CAGTGTGCTGTTTTTC TTGCAGATATTTT	59.0	9.51
20q13. 3	<i>RBM38_tel25</i> 076	chr20:56,009,46 5-56,011,485	2021	TGACCCATGAGAGCT GTAAACTCAA	TCAGCCGAGAAGAA GAGAACAATCA	55.0	5.24
	<i>CTCFI_cen3</i> 4302	chr20:56,033,16 7-56,036,719	3553	CACTGCACCATGCAT TTCATTCATTATC	ATTCCCAAATTGTTG GTAAGAAACCCAT	55.0	3.49
	<i>PCK1_cen13</i> 036	chr20:56,119,56 9-56,123,101	3533	TAAATGTGACTGCTG GAAAGAGTTTGAG	TTGAGAAGGTTCTGG GTTTTATTTGCAT	44.0	2.26
	<i>PCK1_cen18</i> 0-IVS6	chr20:56,135,95 7-56,139,048	3092	CGAAGTCTCCAGCA TTCATTAACA	TCTTGCTTCCTCCAC AATTTTCCAC	51.0	5.30

Appendix V: Frequency of copy number variants (CNVs) that overlap DA intervals used for scFISH probes. Data collected from independent microarray datasets of control populations in Database of Genomic Variants (DGV) and Healthy sample population set (Affymetrix, Inc).

DA interval	Common CNVs*
<i>XDH_tel9264</i>	DGV: 89kb gain Freq G: 0.090%, Freq L: 0%, gains: 1, losses: 0, sample size: 1109
<i>XDH_tel2387</i>	DGV: same as above
<i>XDH_IVS30-IVS27</i>	None
<i>ZNF385D_tel640535</i>	None
<i>ZNF385D_tel678016</i>	DGV: 154 kb gain Freq G: 1.052%, Freq L: 0%, gains: 1, losses: 0, sample size: 95; 2100 kb gain Freq G: 0.003%, Freq L: 0%, gains: 1, losses: 0, sample size: 29084
<i>FGF6_cen4492</i>	DGV: gain .003-.005% G with sample size: 17,000 or 29,084
<i>FGF6_IVS2</i>	DGV: 1659 kb gain - Freq G: 0.003%, Freq L: 0%, gains: 1, losses: 0, sample size: 29084; 300 kb gain Freq G: 0.005%, Freq L: 0%, gains: 1, losses: 0, sample size: 17421; 45 kb loss; 2100 kb LOSS Freq G: 0%, Freq L: 2.564%, gains: 0, losses: 1, sample size: 39
<i>DUOX1_IVS1-IVS3</i>	DGV: 30kb 1 loss (0.064%) 0 gains sample size 1557 & 2 losses (.011%) 0 gains sample size 17421//43kb gain 1 gain (0.049%) 0 loss sample size 2026//78kb loss 1 loss (.005%) and 0 gains of 17421 samples//44kb gain 1 gain (.005%) and 0 loss of 17421 samples
<i>TPM1_IVS5-IVS8</i>	DGV: 282kb gain Freq G: 1.612%, Freq L: 0%, gains: 1, losses: 0, sample size: 62; 944 kb gain Freq G: 0.003%, Freq L: 0%, gains: 1, losses: 0, sample size: 2908
<i>TPM1_IVS8</i>	None
<i>TPM1_tel3200</i>	DGV: 282kb gain Freq G: 1.612%, Freq L: 0%, gains: 1, losses: 0, sample size: 62; 944 kb gain Freq G: 0.003%, Freq L: 0%, gains: 1, losses: 0, sample size: 2908
<i>SCAMP2_IVS7-IVS4</i>	DGV 178kb loss with 1 loss (1.052%) no gains of 95 samples
<i>SCAMP2_IVS1</i>	DGV: 178KB loss Freq G: 0%, Freq L: 1.052%, gains: 0, losses: 1, sample size: 95
<i>COX5A_tel20100</i>	DGV: 9; 242Kb gain Freq G: 0.005%, Freq L: 0%, gains: 1, losses: 0, sample size: 17421
<i>RBM38_tel25076</i>	DGV: 55kb Freq G: 0%, Freq L: 0.049%, gains: 0, losses: 1, sample size: 2026; Freq G: 0.003%, Freq L: 0%, gains: 1, losses: 0, sample size: 29084
<i>CTCF_L_cen34302</i>	DGV: 259 kb gain with 1 gain (.003%) no losses of 29084 samples

<i>PCK1_cen13036</i>	one 37kb gain Healthy samples - healthy samples DGV: PMID 18451855 One gain by Kidd et al 2008 FISH
<i>PCK1_cen180-IVS6</i>	None
*G = gain, L= loss, Freq = frequency	

Appendix VI: Comparison between DA domain location in metaphase and interphase TAD and sub-TAD structures for the XDH, FGF6, and TPM1 domains

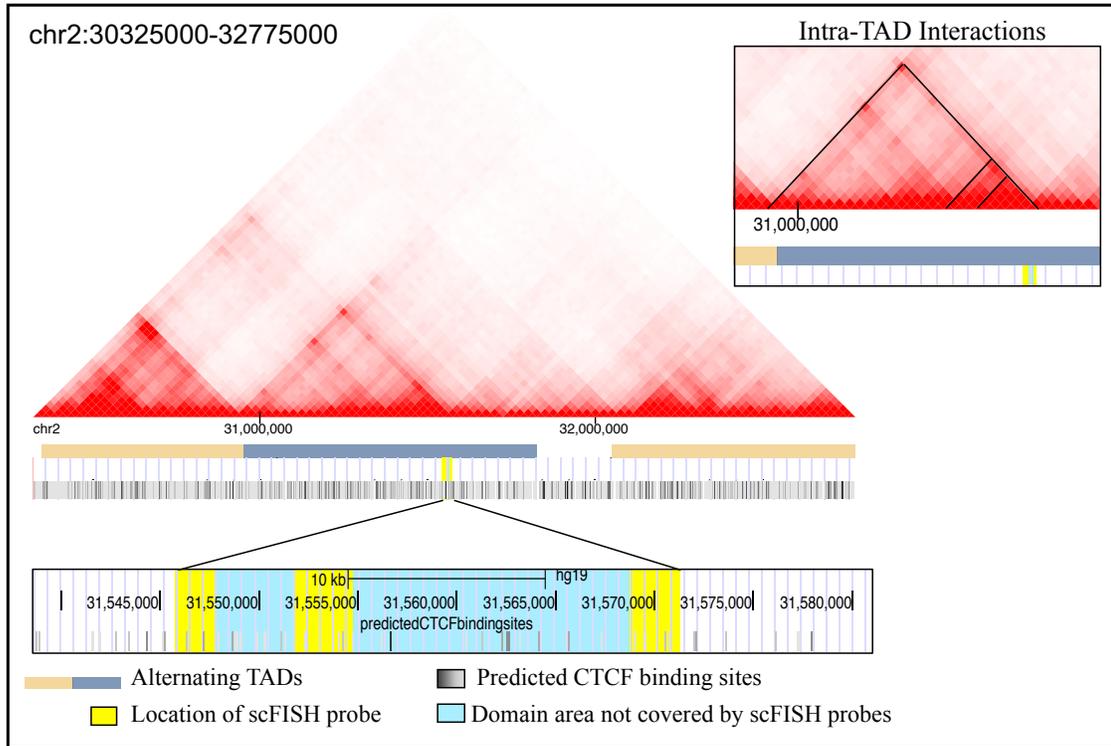


Figure 1: Localization of the XDH domain in metaphase cells relative to TAD structures present in interphase cells. The heat map (3D Genome Browser)⁸⁶ shows interaction frequencies between chromatin within the XDH domain (indicated by yellow-light blue area in UCSC genome browser image⁶⁷) and surrounding chromatin. Alternating blue-grey and light tan bars represent alternating TADs. The XDH domain localizes within a single blue-grey TAD. Intensity of red increases with increased frequency of interaction. The scale in the top left measures the intensity of contact frequency as the normalized number of contacts between 2 points. The XDH domain is within a section of genome with multiple levels of intra-TAD interactions observed by multiple areas of high contact frequency at varying distances. Three of these intra-TAD interactions over the domain are magnified in the top right corner. Predicted CTCF binding sites⁸⁷ are presented in grey-scale, with the colour darkening with increased site binding affinity. Magnified image of DA domain (UCSC Genome Browser; GRCh37/hg19 assembly) shows full domain in light blue with yellow indicating location of hybridized scFISH probes.

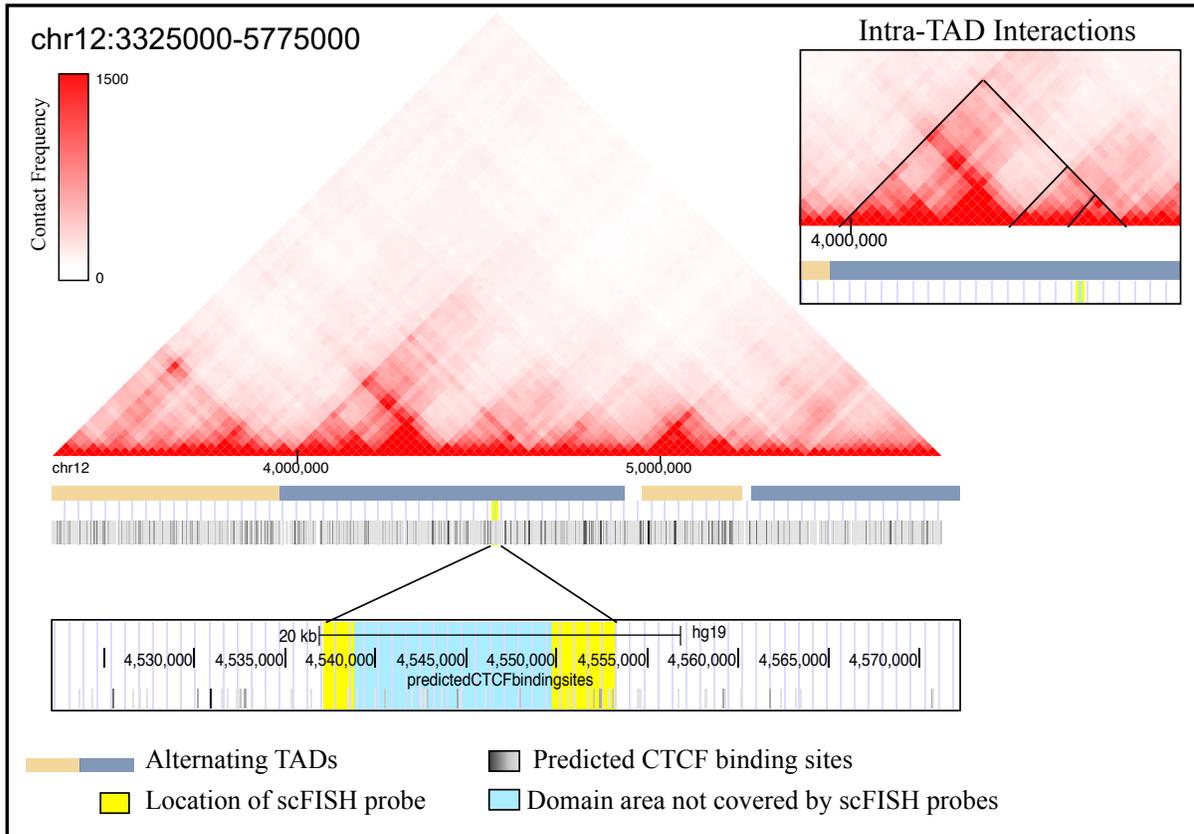


Figure 2: Localization of the FGF6 domain in metaphase cells relative to TAD structures present in interphase cells. The heat map (3D Genome Browser)⁸⁶ shows interaction frequencies between chromatin within the FGF6 domain (indicated by yellow-light blue area in UCSC genome browser image⁶⁷) and surrounding chromatin. Alternating blue-grey and light tan bars represent alternating TADs. The FGF6 domain localizes within a single blue-grey TAD. Intensity of red increases with increased frequency of interaction. The scale in the top left measures the intensity of contact frequency as the normalized number of contacts between 2 points. There is a high frequency of intra-TAD interactions within and around this domain, interacting with multiple points within the TAD. Three of these intra-TAD interactions over the FGF6 domain are magnified in the top right corner. Predicted CTCF binding sites⁸⁷ are presented in grey-scale, with the colour darkening with increased site binding affinity. Magnified image of DA domain (UCSC Genome Browser; GRCh37/hg19 assembly) shows full domain in light blue with yellow indicating location of hybridized scFISH probes.

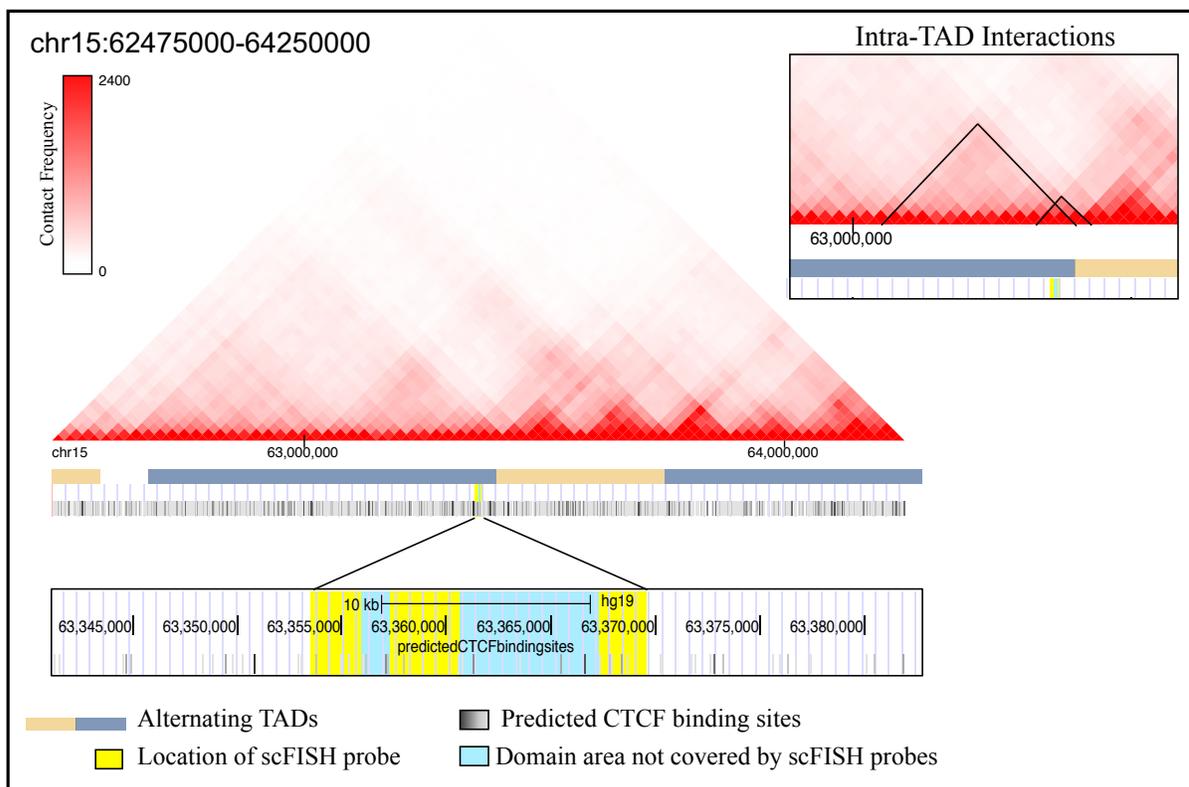


Figure 3: Localization of the TPM1 domain in metaphase cells relative to TAD structures present in interphase cells. The heat map (3D Genome Browser)⁸⁶ shows interaction frequencies between chromatin within the TPM1 domain (indicated by yellow-light blue area in UCSC genome browser image⁶⁷) and surrounding chromatin. Alternating blue-grey and light tan bars represent alternating TADs. The TPM1 domain localizes within a single blue-grey TAD approaching the boundary with the adjacent TAD. Intensity of red increases with increased frequency of interaction. The scale in the top left measures the intensity of contact frequency as the normalized number of contacts between 2 points. There are some intra-TAD interactions with sequences close to the domain, however compared to other domains, there are fewer areas of high frequency intra-TAD interactions. Two intra-TAD interactions over the TPM1 domain are outlined in the top right corner. Predicted CTCF binding sites⁸⁷ are presented in grey-scale, with the colour darkening with increased site binding affinity. Magnified image of DA domain (UCSC Genome Browser; GRCh37/hg19 assembly) shows full domain in light blue with yellow indicating location of hybridized scFISH probes.

Appendix VII: DA and EA cell count for each probe and tissue type examined with results of significance testing between DA and EA proportions per probe and between individuals. The number of individuals sampled for each probe investigated is given with p value results from the two-proportion Z-test testing if there is a statistical difference between individuals and a two-tailed binomial test with normal approximation testing statistical difference between proportion of DA and EA cells scored overall for each probe

Probe Name	Tissue Type	DA	EA	# of individuals tested	p-value ($\alpha = 0.05$)	
					Two proportion Z-test between samples	Two-tailed binomial test between DA and EA*
3.3_1p36	Lymph	13	41	2	0.55	1.4E-04
	BM	32	53	2	0.75	2.3E-02
	Fibro	14	34	2	0.91	3.9E-03
<i>XDH</i> IVS 30-IVS27	Lymph	48	9	2	0.38	2.4E-07
	BM	44	9	2	0.36	1.5E-06
	Fibro	21	4	1	N/A	6.7E-04
<i>DUOX1</i> I VS1-IVS3	Lymph	95	19	2	0.67	1.1E-12
	BM	50	13	2	0.82	3.1E-06
	Fibro	32	9	1	N/A	3.3E-04
<i>PCK1</i> _cen 180-IVS6	Lymph	49	18	2	0.80	1.5E-04
	BM	79	21	2	0.85	6.6E-09
	Fibro	86	9	2	0.75	2.9E-15
<i>TPMI</i> _tel 3200	Lymph	55	13	2	0.43	3.5E-07
	BM	23	3	1	N/A	8.8E-05
	Fibro	31	9	1	N/A	5.0E-04
<i>CTCF</i> L_c en34302	Lymph	52	14	2	0.21	2.9E-06
	BM	81	12	2	0.0038	8.4E-13
	Fibro	37	11	2	0.98	1.7E-04
+when applicable						
* in cases with more than 2 samples, 2 samples with largest number of cells selected						

Curriculum Vitae

Post-Secondary Education

MSc in Pathology and Laboratory Medicine 2017-2020

University of Western Ontario, London, ON

Scholarships and Awards

- Western Graduate Research Scholarship (WGRS) 2017-2019
- Ontario Graduate Scholarship (OGS) 2018-2019
- Northeast Regional Chromosome Pairing Conference Travel Award 2018

Honours BSc in Forensic Science 2013-2017

Laurentian University, Sudbury, ON

Scholarships and Awards

- Alumni Association Entrance Scholarship 2013
- Academic Excellence Scholarship 2013-2017
- Trevor Sullivan Memorial Forensic Science Scholarship 2014
- Laurentian University Administrative and Professional Staff Association Scholarship 2015
- Wah Chung Scholarship 2016
- Natural Sciences and Engineering Research Council of Canada Undergraduate Student Research Award (NSERC-USRA) 2016
- Dean's Honour List 2013-2017

Related Work Experience

Graduate Research Assistant 2017-2019

University of Western Ontario

Teaching Assistant (Seminar and Research Project, PATH4980) 2018

University of Western Ontario

Student Research Assistant 2016-2017

Laurentian University

Teaching Assistant (Intro to Forensic Science/Biology; Lab) 2016-2017

Laurentian University

Related Volunteer Experience

Western Pathology Association (WPA) – VP Finance 2018-2019

University of Western Ontario

WPA Mentor-Mentee Program – Mentee (2017-2018), Mentor (2018-2019)

University of Western Ontario

Laurentian Association of Forensic Scientists – Secretary 2016-2017

Laurentian University

Publications

Hill, S.L. (2017) *Development and validation of a method for the determination of nine benzodiazepines and metabolites in dried blood spots (DBS) using UPLC-QTOF-MS*. (Undergraduate Thesis, Laurentian University, Sudbury ON). Retrieved from <https://zone.biblio.laurentian.ca/handle/10219/2775>.

Presentations

Title	Audience	Date
Defining domains of differential chromatin compaction between human metaphase homologues: Exploring their link to higher-order chromosome structure <i>S Hill, Y Wang, PK Rogan, JH Knoll</i>	London Health Research Day (Poster Presentation)	April 2019
Defining domains of differential chromatin compaction between human metaphase homologues: A higher-order chromosome structure link? <i>S Hill, Y Wang, PK Rogan, JH Knoll</i>	Pathology and Laboratory Medicine Research Day (Poster Presentation)	March 2019
Defining differentially accessible domains within human metaphase chromosomes <i>S Hill, Y Wang, PK Rogan, JH Knoll</i>	Northeast Regional Chromosome Pairing Conference (Platform Presentation) *Selected as Chair for one session	October 2018
Expanded census of human genomic regions with chromatin accessibility differences between homologous metaphase chromosomes. <i>S Hill, Y Wang, PK Rogan, JH Knoll</i>	London Health Research Day (Poster Presentation)	May 2018
Defining domains of chromatin accessibility in human homologous metaphase chromosomes. <i>S Hill, Y Wang, PK Rogan, JH Knoll</i>	Pathology and Laboratory Medicine Research Day (Poster Presentation)	April 2018
Analysis of Benzodiazepines in Dried Blood Spots (DBS) by Protein Filtration and Ultra Performance Liquid Chromatography/ quadrupole Time of Flight/ Mass Spectrometry (UPLC/qTOF/MS). <i>S Hill & J Watterson</i>	American Academy of Forensic Sciences Annual Meeting Young Forensic Scientists' Forum (Poster Presentation)	February 2017