

University of Massachusetts Medical School

eScholarship@UMMS

---

GSBS Dissertations and Theses

Graduate School of Biomedical Sciences

---

2016-02-11

## Characterization of Higher-order Chromatin Structure in Bone Differentiation and Breast Cancer: A Dissertation

Ahmet Rasim Barutcu

*University of Massachusetts Medical School*

Let us know how access to this document benefits you.

Follow this and additional works at: [https://escholarship.umassmed.edu/gsbs\\_diss](https://escholarship.umassmed.edu/gsbs_diss)



Part of the [Cancer Biology Commons](#), [Cell Biology Commons](#), [Genetics and Genomics Commons](#), and the [Neoplasms Commons](#)

---

### Repository Citation

Barutcu AR. (2016). Characterization of Higher-order Chromatin Structure in Bone Differentiation and Breast Cancer: A Dissertation. GSBS Dissertations and Theses. <https://doi.org/10.13028/M2T01Z>.

Retrieved from [https://escholarship.umassmed.edu/gsbs\\_diss/827](https://escholarship.umassmed.edu/gsbs_diss/827)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

*Characterization of higher-order chromatin structure in  
bone differentiation and breast cancer*

A Dissertation Presented

by

AHMET RASIM BARUTCU

Submitted to the Faculty of the

University of Massachusetts Graduate School of Biomedical Sciences, Worcester

in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

February 11<sup>th</sup>, 2016

Program in Cell Biology

***CHARACTERIZATION OF HIGHER-ORDER CHROMATIN STRUCTURE IN  
BONE DIFFERENTIATION AND BREAST CANCER***

A Dissertation Presented

By

Ahmet Rasim Barutcu

The signatures of the Dissertation Defense Committee signify  
completion and approval as to style and content of the Dissertation

\_\_\_\_\_  
Gary S. Stein, Ph.D. & Anthony N. Imbalzano, Ph.D., Thesis Advisors

\_\_\_\_\_  
Giovanni Bosco, Ph.D., Member of Committee

\_\_\_\_\_  
Thomas G. Fazio, Ph.D., Member of Committee

\_\_\_\_\_  
Paul D. Kaufman, Ph.D., Member of Committee

\_\_\_\_\_  
Scot A. Wolfe, Ph.D., Member of Committee

The signature of the Chair of the Committee signifies that the written dissertation  
meets the requirements of the Dissertation Committee

\_\_\_\_\_  
Craig L. Peterson, Ph.D., Chair of Committee

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that  
the student has met all graduation requirements of the school.

\_\_\_\_\_  
Anthony Carruthers, Ph.D.,  
Dean of the Graduate School of Biomedical Sciences

Program in Cell Biology

February 11<sup>th</sup>, 2016

## DEDICATION

This thesis is dedicated to my beloved wife, Seda Barutcu for her never-ending love and support; my father Lutfi Barutcu for being the exemplary person in my life and contributing to my well-being even if I had a chance to see him for a very short time in my life; my mother Ful Arun for her endless support and care; my brother Hayri Barutcu for his cheer and motivation; my grandmother Nesrin Dolek for always being there, my aunt Nur Civan for being the second mother for me, and all my relatives and loved ones who altogether are the inspiration for my work.

## ACKNOWLEDGEMENTS

My PhD life has been a long and arduous journey with many ups and downs, and many changes of paths. Therefore, there are several people whom I would like to thank.

I would like to thank Dr. Anthony Imbalzano for his support and scientific input for all these years. He has always brought humor to my daily life and turned my normally painful PhD experience into an enjoyable journey. It was really delightful working with him and I will always seek for his advice throughout my career.

I would like thank Dr. Gary Stein, Dr. Janet Stein and Dr. Jane Lian for their kind support and hospitality at their labs and their home. They provided me the freedom, resources and the environment in which doing science is fun. They always allowed me to pursue my own scientific interests in their lab. Working in their lab allowed me to mature into a well-rounded scientist with their enthusiasm and faith.

I would like to thank my thesis research and defense committee members Dr. Craig Peterson, Dr. Scot Wolfe, Dr. Sharon Cantor, Dr. Andre van Wijnen, Dr. Tom Fazio, Dr. Paul Kaufman and Dr. Giovanni Bosco for their guidance and input throughout my research and dissertation.

I would like to thank Dr. Job Dekker, who had significant contributions to my expertise and thesis research. Working in his lab at the start of my PhD greatly contributed to my scientific achievements. I also would like to thank Dr. Marian Walhout for teaching me several things which otherwise I could not have learned.

Apart from my mentors, my research would not have been possible by the input of many people in my labs. I would like to thank the previous and current members of

the Imbalzano Lab; Dr. Yu Jie Hu, Dr. Teresita Padilla, Dr. Scott LeBlanc and Dr. Qiong Wu, Dr. Brian Nasipak and Dr. Manuel Hernandez for their continuous support and companionship during my PhD. As the other half of my lab, I would like to thank the former and current members of the Stein Lab for critical discussions and motivation. Furthermore, I'd like to thank the former members of the Dekker Lab; Dr. Amartya Sanyal, Dr. Nynke van Berkum, Dr. Emily Smith, Dr. Jon Matthew Belton, Dr. Bryan Lajoie, Dr. Johan Gibcus, Ye Zhan, Gaurav Jain, Dr. Rachel Patton McCord, Dr. Natalia Naumova and Dr. Jennifer Phillips-Cremens for their friendship and cheerful support.

Finally, I would like to thank my friends, both in the US and abroad, for providing a family-like environment for me.

## ABSTRACT

Higher-order genome organization is important for the regulation of gene expression by bringing different cis-regulatory elements and promoters in proximity. The establishment and maintenance of long-range chromatin interactions occur in response to cellular and environmental cues with the binding of transcription factors and chromatin modifiers. Understanding the organization of the nucleus in differentiation and cancer has been a long standing challenge and is still not well-understood. In this thesis, I explore the dynamic changes in the higher-order chromatin structure in bone differentiation and breast cancer. First, we show dynamic chromatin contact between a distal regulatory element and the promoter of *Runx2* gene, which encodes the Runt-related transcription factor 2 (RUNX2) that is essential for bone development. Next, via using a genome-wide approach, we show that breast cancer cells have altered long-range chromatin contacts among small, gene-rich chromosomes and at telomeres when compared with mammary epithelial cells. Furthermore, we assess the changes in nuclear structure and gene expression of breast cancer cells following Runt-related transcription factor 1 (RUNX1) deficiency, an event frequently observed in breast cancer. Finally, I present the role of the central ATPase subunit of the SWI/SNF complex, SMARCA4 (BRG1), in mediating nuclear structure and gene expression. Taken together, the research presented in this thesis reveals novel insight and paradigm for the dynamic changes in disease and differentiation, as well as uncovers previously unidentified roles for two chromatin regulatory proteins, RUNX1 and SMARCA4.

## PREFACE

Certain parts of this dissertation appear in the following published work:

[1] **Barutcu, A. R.**; Tai, P. W.; Wu, H.; Gordon, J. A.; Whitfield, T. W.; Dobson, J. R.; Imbalzano, A. N.; Lian, J. B.; van Wijnen, A. J.; Stein, J. L. and others (2014). *The bone-specific Runx2-P1 promoter displays conserved three-dimensional chromatin structure with the syntenic Supt3h promoter*, Nucleic Acids Research 42 : 10360-10372.

[2] **Barutcu, A. R.**; Lajoie, B. R.; McCord, R. P.; Tye, C. E.; Hong, D.; Messier, T. L.; Browne, G.; van Wijnen, A. J.; Lian, J. B.; Stein, J. L. and others (2015). *Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells*, Genome Biology 16 : 1-14.

[3] **Barutcu, A. R.**; Fritz, A. J.; Zaidi, S. K.; van Wijnen, A. J.; Lian, J. B.; Stein, J. L.; Nickerson, J. A.; Imbalzano, A. N. and Stein, G. S. (2016). *C-ing the Genome: A Compendium of Chromosome Conformation Capture Methods to Study Higher-Order Chromatin Organization*, Journal of Cellular Physiology 231 : 31-35.

[4] Fritz, A. J.; **Barutcu, A. R.**; Martin-Buley, L.; van Wijnen, A. J.; Zaidi, S. K.; Imbalzano, A. N.; Lian, J. B.; Stein, J. L. and Stein, G. S. (2016). *Chromosomes at Work: Organization of Chromosome Territories in the Interphase Nucleus*, Journal of Cellular Biochemistry 117 : 9-19.

Other published work during graduate study that are not presented in this thesis:

[1] Reece-Hoyes, J. S.; **Barutcu, A. R.**; McCord, R. P.; Jeong, J. S.; Jiang, L.; MacWilliams, A.; Yang, X.; Salehi-Ashtiani, K.; Hill, D. E.; Blackshaw, S. and others (2011). *Yeast one-hybrid assays for gene-centered human gene regulatory network mapping*, Nature Methods 8 : 1050-1052.

[2] Medina, R.; Ghule, P. N.; Cruzat, F.; **Barutcu, A. R.**; Montecino, M.; Stein, J. L.; van Wijnen, A. J. and Stein, G. S. (2012). *Epigenetic control of cell cycle-dependent histone gene expression is a principal component of the abbreviated pluripotent cell cycle*, Molecular and Cellular Biology 32 : 3860-3871.

[3] LeBlanc, S. E.; Wu, Q.; **Barutcu, A. R.**; Xiao, H.; Ohkawa, Y. and Imbalzano, A. N. (2014). *The PPAR $\gamma$  locus makes long-range chromatin interactions with selected tissue-specific gene loci during adipocyte differentiation in a protein kinase A dependent manner*, PLoS One, 9(1):e86140.

[4] Tai, P. W.; Wu, H.; Gordon, J. A.; Whitfield, T. W.; **Barutcu, A. R.**; van Wijnen, A. J.; Lian, J. B.; Stein, G. S. and Stein, J. L. (2014). *Epigenetic landscape during osteoblastogenesis defines a differentiation-dependent Runx2 promoter region*, Gene 550 : 1-9.



## LIST OF THIRD PARTY COPYRIGHTED MATERIAL

The following materials were reproduced from the publishers below. No permission required unless the license number is noted.

<u>Material</u>	<u>Publisher</u>	
All Chapter 2 text and figures	Nucleic Acids Research (Oxford Journals)	
All Chapter 3 text and figures	Genome Biology (BioMed Central)	

<u>Material</u>	<u>Publisher</u>	<u>License Number</u>
Part of Chapter 1 and Figure1.1	John Wiley and Sons (JCP)	3786610502081
Part of Chapter 1	John Wiley and Sons (JCB)	3787081304246

## TABLE OF CONTENTS

DEDICATION .....	iii
ACKNOWLEDGEMENTS .....	iv
ABSTRACT .....	vi
PREFACE .....	vii
LIST OF THIRD PARTY COPYRIGHTED MATERIAL .....	viii
TABLE OF CONTENTS .....	ix
TABLE OF FIGURES .....	xiii
CHAPTER 1 : Introduction.....	1
1.1 A brief history of chromatin biology.....	1
1.2 An overview of chromatin structure.....	2
1.3 Methods to Study Chromatin Organization .....	7
1.3.1 Chromosome Conformation Capture based techniques .....	8
1.3.1.1 Chromosome Conformation Capture (3C).....	9
1.3.1.2 Circular chromosome conformation (4C) .....	10
1.3.1.3 Chromosome Conformation Capture Carbon Copy (5C) .....	11
1.3.1.4 Genome-wide chromosome conformation capture (Hi-C).....	12
1.3.1.5 ChIA-PET, 6C, and ChIP-Loop .....	13
1.3.2 Microscopy based techniques.....	14
1.4 Higher-order chromosome organization during development .....	15
1.5 Higher-order chromosome organization during tumorigenesis .....	18
CHAPTER 2 : The bone-specific Runx2-P1 promoter displays conserved three-dimensional chromatin structure with the syntenic Supt3h promoter.....	20
2.1 Introduction .....	20
2.2 Materials and Methods .....	22
2.2.1 WASHU epigenome and UCSC genome browser search.....	22
2.2.2 MC3T3-E1 Cell culture.....	22
2.2.3 Chromosome Conformation Capture (3C) of the <i>Runx2</i> locus .....	23
2.2.4 DNase-seq .....	25
2.2.5 Reverse-transcriptase qPCR.....	25
2.2.6 Reporter constructs.....	26
2.2.7 Co-transfections and Luciferase reporter assays .....	27
2.2.8 Interplasmid 3C .....	28

2.3 Results .....	28
2.3.1 Identification of a long-range interaction between the <i>Runx2-P1</i> and <i>Supt3h</i> promoters .....	28
2.3.2 Interaction frequency between <i>Runx2-P1</i> and <i>Supt3h</i> promoters is increased during osteoblast differentiation.....	36
2.3.3 Interactions between <i>Runx2-P1</i> and <i>Supt3h</i> are enriched for CTCF, RUNX2 and DHS during osteoblast differentiation.....	39
2.3.4 <i>Supt3h</i> expression levels remain constant during osteoblast differentiation..	41
2.3.5 <i>Runx2-P1</i> and <i>Supt3h</i> promoters can physically interact and regulate <i>Runx2-P1</i> expression <i>in-trans</i> .....	43
2.4 Discussion .....	50
CHAPTER3 : Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells.....	56
3.1 Introduction .....	56
3.2 Materials and Methods .....	58
3.2.1 Cell Culture .....	58
3.2.2 RNA-seq and Analysis .....	59
3.2.3 Preparation of Hi-C Libraries.....	60
3.2.4 Read Mapping / Binning / ICE correction .....	60
3.2.6 Calculation of Differential Interactions.....	61
3.2.7 Compartment Profiles .....	61
3.2.8 Identification of TAD Boundaries (Insulation Square Analysis).....	62
3.2.9 Identification of TAD Domains .....	63
3.3 Results .....	64
3.3.1 Small, Gene-Rich Chromosomes Interact Less Frequently in the MCF-7 Breast Cancer Genome.....	64
3.3.2 Decreased interaction frequency between small chromosomes in MCF-7 cells coincides with increased open chromatin compartmentalization.....	72
3.3.3 Decreased inter-chromosomal interactions and higher frequency of open compartmentalization on chr16-22 in MCF-7 are associated with WNT signaling related genes.....	77
3.3.4 Cell-Line Specific TAD Boundaries Are Conserved between MCF-10A and MCF-7 .....	83
3.3.5 MCF-7 TAD Boundaries Are Enriched for Several Oncoproteins .....	89
3.3.6 The telomeric / sub-telomeric regions in MCF-10A genome display stronger	

associations than those in MCF-7 .....	92
3.4 Discussion .....	97
CHAPTER4: BRG1 regulates gene expression and higher-order chromatin structure in proliferating mammary epithelial cells .....	101
4.1 Introduction .....	101
4.2 Materials and Methods .....	103
4.2.1 MCF-10A Cell Culture.....	103
4.2.2 RNA-seq and Analysis .....	104
4.2.3 Preparation of Hi-C Libraries.....	104
4.2.4 Read Mapping / Binning / ICE correction .....	104
4.2.5 Z-score Calculation .....	105
4.2.6 Calculation of Differential Interactions.....	105
4.2.7 Compartment Profiles .....	106
4.2.8 Identification of TAD Boundaries (Insulation Square Analysis).....	106
4.2.9 ChIP-seq Analysis .....	106
4.3 Results .....	107
4.3.1 BRG1 knockdown results in down-regulation of genes associated with extracellular matrix in MCF-10A cells .....	107
4.3.2 BRG1 ChIP-seq analysis reveals extensive binding to intergenic and intronic regions .....	116
4.3.3 Hi-C analysis of BRG1 knockdown and control cells .....	123
4.3.4 BRG1 knockdown results in extensive gain and loss of long-range chromatin interactions and altered genome organization .....	132
4.3.5 BRG1 occupancy is enriched at open compartment regions.....	138
4.3.6 BRG1 is associated with TAD boundaries and TAD boundary strength.....	142
4.4 Discussion .....	149
CHAPTER5: RUNX1-mediated control of higher-order chromatin organization and gene expression in breast cancer cells.....	153
5.1 Introduction .....	153
5.2 Materials and Methods .....	155
5.2.1 Generation of MCF-7 cell lines and cell culture .....	155
5.2.2 RNA-seq and Analysis .....	156
5.2.3 ChIP-seq and Analysis .....	156
5.2.4 Preparation of Hi-C Libraries.....	158

5.2.5 Hi-C read mapping/binning/ICE correction .....	158
5.2.6 Calculation of Genomic Compartment Profiles .....	158
5.2.7 Identification of TAD boundaries .....	159
5.2.8 Calculation of Differential Interactions.....	159
5.3 Results .....	160
5.3.1 RUNX1 knockdown results in aberrant gene regulation in MCF-7 cells ....	160
5.3.2 Genome-wide RUNX1 localization in parental MCF-7 cells .....	166
5.3.3 RUNX1-mediated genome-wide interaction analysis in MCF-7 cells.....	175
5.3.4 A sub-portion of genomic compartments are altered ulpon RUNX1 Knockdown .....	182
5.3.5 RUNX1 is enriched at TAD boundaries .....	188
5.3.6 RUNX1 knockdown results in emergence and disruption of many long-range interactions .....	192
5.4 Discussion .....	196
CHAPTER 6: Outlook and Future Directions .....	199
6.1 Long-range gene regulation and higher-order genome structure .....	199
6.2 Implications in bone biology and bone metastasis .....	200
6.3 Implications in breast cancer .....	203
6.4 Implications in the biology of TAD structures.....	206
REFERENCES .....	210
APPENDIX .....	241
1. Table of 3C primers in Chapter 2.....	241

## TABLE OF FIGURES

Figure 1.1. An overview of chromosome conformation capture techniques. ....	9
Figure 2.1. The long-range interactions in the <i>Runx2</i> locus in IMR90 Hi-C dataset. ....	30
Figure 2.2. The long-range interactions in the <i>Runx2</i> locus in IMR90 Hi-C replicates..	30
Figure 2.3. The long-range interactions in the <i>Runx2</i> locus in K562 ChIA-PET dataset. .....	31
Figure 2.4. The long-range interactions in the <i>Runx2</i> locus in K562 ChIA-PET biological replicates.....	31
Figure 2.5. The long-range interactions in the <i>Runx2</i> locus in the MCF7 ChIA-PET dataset.....	33
Figure 2.6. The long-range interactions in the <i>Runx2</i> locus in the MCF7 ChIA-PET biological replicates.....	33
Figure 2.7. RT-PCR expression levels of <i>Runx2</i> -P1 and <i>Runx2</i> -P2 in MC3T3 and RAW264.7 cells. ....	34
Figure 2.8. 3C analysis of the <i>Runx2</i> locus in the RAW264.7 cells. ....	35
Figure 2.9. 3C analysis of <i>Runx2</i> -P1 promoter in MC3T3 cells. ....	37
Figure 2.10. 3C analysis of d9 MC3T3 cultures with the anchor fragment located on <i>Runx2</i> -P2. ....	38
Figure 2.11. 3C performed with HindIII as validation of the interaction between <i>Runx2</i> - P1 and <i>Supt3h</i> promoters in d0 and d9 MC3T3s .....	39
Figure 2.12. The epigenetic landscape of the <i>Runx2</i> locus during differentiation.....	40
Figure 2.13. qPCR analysis of bone-specific genes during osteoblast differentiation....	42
Figure 2.14. Time course qPCR expression analysis of bone-related genes in MCF3T3 cells.....	43
Figure 2.15. Luciferase reporter assay of the <i>Supt3h</i> construct in d0 and d6 MC3T3-E1 cells.....	44
Figure 2.16a. Schematic representation of the fragments used in luciferase reporter assays.....	44
Figure 2.16b. Luciferase experiments for different <i>Supt3h</i> promoter constructs. ....	45
Figure 2.17. The schematic of the interplasmid-3C assay. ....	47
Figure 2.18. i3C analysis of co-transfected plasmids.....	48
Figure 2.19. In-trans luciferase assay.....	49
Table 3.1. Hi-C statistics of the MCF-7 and MCF-10A replicates. ....	65
Figure 3.1a. MCF-7 and MCF-10a Hi-C replicates are highly reproducible. ....	65
Figure 3.1b. Interactions in MCF-7 and MCF-10a Hi-C replicates are highly reproducible.....	66
Figure 3.2. Genome-wide all by all 1Mb Hi-C interaction heatmaps .....	67
Figure 3.3. Inter-chromosomal heatmaps showing the translocations in MCF-10A detected by a previous SKY-FISH study .....	68
Figure 3.4. Inter-chromosomal heatmaps showing the translocations in MCF-7 detected by a previous SKY-FISH study.....	69
Figure 3.5. Small chromosomes interact with each other more frequently in MCF-10A cells.....	70

Figure 3.6. Small chromosomes interact more frequently in the MCF-10A genome. ....	71
Figure 3.7. The positioning of chr18 is not altered in MCF-10A cells. ....	72
Figure 3.8. Replicate analysis of genomic compartments in MCF-10A and MCF-7 cells .....	73
Figure 3.9. Association of MCF-7 ENCODE datasets with genomic compartments. ....	73
Figure 3.10. Genomic compartments are altered between MCF-10A and MCF-7 genomes. ....	74
Figure 3.11. A subset of genomic compartments display switching between MCF-10A and MCF-7 genomes. ....	75
Figure 3.12. Compartment changes are distributed homogeneously throughout the genome. ....	75
Figure 3.13. Small chromosomes display enriched “B to A” switching. ....	77
Figure 3.14. Replicate correlation of MCF-10A and MCF-7 RNAseq datasets. ....	78
Figure 3.15. Differential RNA-seq gene expression between MCF-10A and MCF-7 cells. ....	78
Figure 3.16. The significant expression changes were enriched for the medium to highly expressed genes. ....	79
Table 3.2. GO terms of differentially expressed genes between MCF-10A and MCF-7 cells. ....	80
Figure 3.17. Compartmental changes are associated with changes in gene expression. .	81
Table 3.3. REACTOME pathways of the genes that are located on chr16-22 and at B → A compartment regions. ....	82
Figure 3.18. TAD boundary analysis for replicate Hi-C datasets. ....	84
Figure 3.19. Correlation plots of insulation scores across Hi-C biological replicates. ....	84
Figure 3.20. MCF-7 and MCF-10A cells have similar TADs. ....	85
Figure 3.21. TADs can be disrupted in cancer cells. ....	86
Figure 3.22 Gene expression is not associated with TAD boundaries. ....	87
Figure 3.23. Analysis of TAD domains. ....	87
Figure 3.24a. Analysis of TAD domain sizes. ....	88
Figure 3.24b. Gene expression analysis of TAD domains. ....	89
Figure 3.25. Analysis of factors enriched at MCF-7 TAD boundaries. ....	90
Figure 3.26. Analysis of MCF-7 Repli-seq data with TAD boundaries. ....	91
Figure 3.27. eQTL analysis across TAD boundaries. ....	91
Figure 3.28. Scaling plot of interaction frequencies against genomic distance for the MCF-7 and MCF-10A genomes. ....	93
Figure 3.29. Quantification of the interaction frequency between the telomeric regions	94
Figure 3.30. The scaling plots of MCF-10A and MCF-7 for chr1, chr2, chr7 and chr3.	94
Figure 3.31. The chromosome by chromosome scaling plots generated by using 250kb binned Hi-C data. ....	95
Figure 3.32. The scaling plots of individual chromosome arms for chr1, chr2 and chr7 did not show enriched telomeric / sub-telomeric interactions. ....	96
Figure 3.33. Overview of the changes in chromatin architecture between MCF-10A and MCF-7 cells. ....	100
Figure 4.1. BRG1 levels are reduced upon shRNA induction. ....	108

Figure 4.2. Correlation analysis of RNA-seq replicates.....	109
Figure 4.3. BRG1 depletion results in a lower mean gene expression.....	109
Figure 4.4. Scatterplot showing the log <sub>2</sub> gene expression values for shBRG1 and shSCRAM cells.....	110
Figure 4.5. Heatmap showing the transcripts per million (TPM) expression values of the differentially expressed genes between shSCRAM and shBRG1 for each biological replicate.....	110
Figure 4.6. BRG1 knockdown results in extensive transcriptional dysregulation.....	111
Figure 4.7. Locations of BRG1 regulated genes.....	112
Figure 4.8. BRG1-mediated gene regulation occurs throughout the genome.....	112
Figure 4.9. REACTOME terms of down-regulated genes upon BRG1 knockdown.....	113
Figure 4.10. REACTOME terms of up-regulated genes upon BRG1 knockdown.....	114
Figure 4.11. qRT-PCR validation of the RNA-seq data for 23 genes.....	115
Figure 4.12. Scatterplot showing the correlation of log <sub>2</sub> fold change values for the 23 genes from the RNA-seq and qRT-PCR analyses.....	115
Figure 4.13. Table showing the sequenced and mapped reads for each ChIP-seq biological replicate.....	116
Figure 4.14. Matrix showing the Pearson correlation of the signal intensity between the BRG1 pull down (ChIP) and input samples.....	117
Figure 4.15. BRG1 binds to defined locations in the genome.....	117
Figure 4.16. Annotation of BRG1 peaks.....	118
Figure 4.17. BRG1 is enriched at promoter regions.....	118
Figure 4.18. A significant portion of BRG1 overlaps with PolII.....	119
Figure 4.19. Motif analysis of BRG1 peaks.....	120
Figure 4.20. BRG1 is enriched at the promoter regions of down-regulated genes.....	120
Figure 4.21. Identification of BRG1 bound super-enhancers.....	122
Figure 4.22. BRG1 signal intensity of super-enhancers.....	122
Figure 4.23. Annotation of BRG1 bound super-enhancers.....	123
Table 4.1. Sequencing and mapping statistics of the shBRG1 and shSCRAM Hi-C biological replicates.....	124
Figure 4.24. shBRG1 and shSCRAM Hi-C data are reproducible.....	125
Figure 4.25. Scaling plot analysis of shBRG1 and shSCRAM Hi-C samples.....	125
Figure 4.26. Compartment analysis for shBRG1 and shSCRAM Hi-C replicates.....	126
Figure 4.27. Eigenvector analysis for shBRG1 and shSCRAM Hi-C replicates.....	126
Figure 4.28a. Correlation analysis of Hi-C insulation plot replicates.....	127
Figure 4.28b. Insulation plot analysis of Hi-C replicates.....	127
Figure 4.29. shBRG1 and shSCRAM Hi-C data shows high correlation with the parental MCF-10A cells.....	128
Figure 4.30. cis/trans ratio of the Hi-C replicates.....	128
Figure 4.31. Hi-C data reveals genomic compartments and TADs.....	130
Figure 4.32. Translocations in the MCF-10A genome.....	131
Figure 4.33. BRG1 knockdown results in extensive changes in long-range interactions.....	135
Figure 4.34. shBRG1 / shSCRAM interactions that are significantly different at 1Mb for	



each chromosome.....	136
Figure 4.35. Chromosome positioning is not altered upon BRG1 knockdown. ....	137
Figure 4.36. BRG1 knockdown results in alterations in higher-order genome organization.....	137
Figure 4.37. Compartment profiles (the first principal components) of shSCRAM and shBRG1 data for chromosome 2. ....	138
Figure 4.38. A subset of compartments are altered upon BRG1 knockdown. ....	139
Figure 4.39. Alterations in genomic compartments are associated with gene expression. ....	140
Figure 4.40. BRG1 is bound mostly on A-type compartments. ....	141
Figure 4.41. BRG1 is bound to compartment switch regions at similar levels. ....	142
Figure 4.42. Pile-up of interactions confirm TAD boundaries. ....	142
Figure 4.43. TADs are mostly stable upon BRG1 knockdown. ....	143
Figure 4.44. A large portion of TADs remain similar upon BRG1 knockdown. ....	144
Figure 4.45. TAD boundaries are enriched for PolII binding and gene frequency. ....	144
Figure 4.46. A significant portion of BRG1 binding is located on TAD boundaries. ..	146
Figure 4.47. BRG1 and BRG1-bound super-enhancers are enriched at TAD boundaries. ....	146
Figure 4.48. The frequency plot of BRG1 super-enhancers per 50kb for +/- 1Mb of every shBRG1 TAD boundary. ....	147
Figure 5.1. Knockdown of RUNX1 results in near complete depletion of the protein levels.....	161
Figure 5.2. Replicate correlation of gene expression (Transcript per million, TPM) for shNS biological triplicates. ....	162
Figure 5.3. Replicate correlation of gene expression (Transcript per million, TPM) for shRUNX1 biological triplicates. ....	163
Figure 5.4. RUNX1 knockdown results in extensive transcriptional changes.....	164
Figure 5.5. RUNX1 depletion results in up- and down-regulation of hundreds of genes. ....	164
Figure 5.6. REACTOME terms of down-regulated genes. ....	165
Figure 5.7. REACTOME terms of up-regulated genes. ....	166
Figure 5.8. Table showing the sequenced and mapped reads for each RUNX1 ChIP-seq biological replicate. ....	166
Figure 5.9. Matrix showing the Pearson correlation of the signal intensity between the pull down (ChIP) and input samples. ....	167
Figure 5.10. RUNX1 binds to the MALAT1 promoter.....	168
Figure 5.11. RUNX1 binds to the NEAT1 promoter. ....	168
Figure 5.12. MEME de novo motif analysis of the RUNX1 peaks. ....	169
Figure 5.13. Pie chart showing the distribution of RUNX1 ChIP-seq peak annotation for genic and intergenic regions.....	170
Figure 5.14. Normalized RUNX1 ChIP-seq signal intensity plot for all human UCSC genes +/- 2kb. ....	171
Figure 5.15. GO Terms of RUNX1 peaks.....	171
Figure 5.16. RUNX1 density across differentially expressed genes.....	172

Figure 5.17. RUNX1 associates with other chromatin modifiers. ....	173
Figure 5.18. RUNX1 associates with distinct classes of factors. ....	175
Table 5.1. The sequencing and mapping statistics of shNS and shRUNX1 Hi-C biological replicates. ....	176
Figure 5.19. Scatter plots comparing normalized interactions between pairs of 2.5Mb bins from two biological replicates ....	177
Figure 5.20. Example of the Hi-C interaction heatmaps at 100kb resolution.....	177
Figure 5.21. Reproducibility of the Hi-C data.....	178
Figure 5.22. Pearson correlation of the pooled replicate datasets.....	178
Figure 5.23. Bar graph showing the percentage of the percent cis ratios of the replicate and the pooled Hi-C libraries. ....	179
Figure 5.24. Scaling plot of the shNS and shRUNX1 pooled Hi-C libraries.....	179
Figure 5.25. The chromosome by chromosome interactions matrices for shNS and shRUNX1 datasets. ....	180
Figure 5.26. Hi-C interaction heatmaps of chr14 at increasing resolutions.....	181
Figure 5.27. Schematic representation of the genomic compartment switching at each non-overlapping bin. ....	183
Figure 5.28. Pie chart showing the genomic compartment changes between shNS and shRUNX1 datasets. ....	184
Figure 5.29. Bar graph showing the number of stable and altered compartments.....	185
Figure 5.30. Distribution of the 1st eigenvalues for each 250kb for shNS and shRUNX1 datasets ....	185
Figure 5.31. Transcriptional changes are correlated with compartment changes. ....	186
Figure 5.32. Pie chart showing the percentage of the compartment switch regions that are bound by RUNX1.....	187
Figure 5.33. Bar chart showing the percentage of stable and altered compartments that are RUNX1 bound.....	188
Figure 5.34. Venn diagram showing that the TAD boundaries are largely similar between shNS and shRUNX1 Hi-C datasets.....	189
Figure 5.35. Pie chart showing the percentage of RUNX1 localization at TAD boundaries. ....	189
Figure 5.36. The frequency plot of RUNX1 ChIP-seq peaks per 25kb for +/- 1Mb of every TAD boundary.....	190
Figure 5.37. An example region on chr10 showing three different TADs.....	190
Figure 5.38. A subset of RUNX1 peaks overlap with CTCF and cohesion at TAD boundaries. ....	191
Figure 5.39. The z-score differences of shRUNX1 / shNS Hi-C matrices for each chromosome at 1Mb bin size.....	193

## **CHAPTER 1: Introduction**

### **1.1 A brief history of chromatin biology**

How does the two meter DNA fit inside a  $\sim 10\mu\text{m}$  nucleus? Within such a condensed environment, how does the genome function in order to maintain the cellular functions, or to drive the differentiation process of a whole organism from a single cell? The fundamentals of chromatin folding and chromosome compaction have been a long-standing mystery in biology. The existence of chromosomes was noticed as early as late 19<sup>th</sup> century [1]. During that time, Walther Flemming identified that the chromatin corresponded to the thread-like structures, which were later named as "chromosome" (Greek: chroma – "color", soma – "body") by Heinrich Wilhelm Gottfried von Waldeyer-Hartz [2-5]. Towards the mid-1880s, it was discovered that chromosomes were the units for hereditary information [6]. In 1871, F. Miescher identified an acidic, phosphorous rich substance in leukocytes which he called "nuclein", and later in sperm cells, he identified the basic substance termed "protamin" [7]. In 1884, Hoppe-Seyler described the "histon" in the acidic fraction of the nucleus [8]. Later on, with the advancement of microscopic lenses with minimal optical aberrations and improved fixatives [9], in addition to identification of distinct nuclear substances [10], the chromatin biology field has achieved a deeper understanding about the structure of the nucleus.

There were several landmark discoveries in the first half of the 20<sup>th</sup> century including the rediscovery of Gregory Mendel's principles and the identification of genetic linkage. However, during this time, very little was identified regarding the

structure and the proteins associated with the chromatin. Many discoveries, such as the identification of the *Drosophila* puff chromosomes, relied on the enzymatic digestion and the visualization of the biological material by light microscopy. Towards the second half of the 20<sup>th</sup> century, some profound discoveries were made. The discovery of the structure of DNA [11-13], the demonstration that DNA is the backbone of the chromatin [14], the identification of the histones [15], and the association between histone modifications and gene expression (by histone acetylation and methylation) [16] can be considered as milestones of the era.

## **1.2 An overview of chromatin structure**

It is now very well established that many aspects of chromatin structure are intricately linked to several functions of the eukaryotic genome. The chromatin condenses/decondenses both locally and globally in the processes of cell division, replication, transcription, homologous recombination and DNA repair. Yet, the cell can dynamically regulate and fine-tune all of these nuclear functions during cell differentiation (including embryogenesis) and in response to intrinsic and extrinsic environmental cues including hormonal regulation, mechano-sensing, heat-shock, and hypoxia. Perturbations in the genomic processes are tightly associated with both developmental and pathologic diseases including cancer.

From yeast to mammalian genomes, the DNA is folded every ~146 base pairs (bp) into nucleosome complexes that consist of H2A, H2B, H3 and H4 core proteins, and H1 linker protein [17]. Nucleosomes are connected to each other by short DNA segments known as the "linker DNA". Electron micrographs of this structure are

visualized as the "beads on a string" model, which constitutes the primary structure of the chromatin. The linker DNA and the linker histone H1 associates with surrounding histone complexes and constitutes a degree of compaction, observed as the condensed 30nm chromatin, which has been observed *in vitro*. However, its *in vivo* presence is still being debated [18-20]. During cell division, the chromatin is further folded into mitotic chromosomes [21]. Even though the biology of the primary structure and the condensed structure (i.e. mitotic chromosomes) is rather well understood, our understanding about the intermediate folding characteristics of the chromatin has remained very limited.

During the interphase nucleus, each chromosome occupies a separate nuclear space, forming the chromosome territories [22-24]. It is known for a very long time that the chromosomes are radially positioned in the nucleus, where the larger, gene-poor chromosomes are located towards the periphery, and the small, gene-rich chromosomes are located towards the interior of the nucleus [25]. Because gene-dense chromosomes generally exhibit higher levels of transcription than gene-poor chromosomes [26], gene activity may be the underlying basis for the radial positioning of the genome.

Interestingly, this paradigm holds true not just for entire chromosome territories, but also for individual genes and gene complexes within chromosomes; as expressed alleles are generally found further from the nuclear periphery than alleles that are not expressed [27-30]. This could result from the preferential localization of tightly packed chromatin at the nuclear periphery [31; 32]. Although the chromosome territories are spatially separated, they cross-talk with each other by the intermingling of different chromosome territories, a phenomenon called "chromosome kissing" [33; 34]. One mechanism for

chromosome intermingling is "transvection" [35], where the DNA sequences of a chromatid can affect the homologous sequences on the other chromatid [36; 37]. Each chromosome has a preferred position with respect to the nuclear periphery and with respect to each other [38; 39]. Preferential inter-chromosomal associations measured between pairwise heterologous chromosome territories suggest an overall nonrandom organization of chromosome territories with respect to each other. This preferred positioning of chromosomes are cell-type specific, affecting the *trans*-interactions they make with other chromosomes, as revealed by the specific recurrent translocations between two chromosomes based on their proximity [40; 41]. Subsequently, unique chromosome territory profiles are detected in different cell types and tissues [41-45]. The localization of the chromosomes relative to the periphery and to each other dictates the transcriptional and regulatory outcome of the genes and other regulatory elements. For example, the ectopic transcriptional activation of a gene locus results in the movement of that locus from the nuclear periphery to the interior [46]. As cells differentiate, entire chromosome territories are repositioned [47; 48] and individual genes within chromosome territories exhibit altered inter-chromosomal interactions; the pluripotency-related gene *Nanog* is an example [49]. Within each chromosome territory, subdomains were proposed to occur via 6–12 chromatin loops of approximately 50–200 kb arranged in a rosette pattern composed of approximately 1 Mb of DNA [50-52]. The periphery of the nucleus contains nuclear lamins forming the nuclear lamina [53], a proteinaceous layer found at the interface between chromatin and the inner nuclear membrane [54]. Containing hundreds of lamin associated proteins

[55], lamin associated DNA regions (LADs) are largely consistent among the individual cells in a population [56]. Consistent with observations about gene activity and chromosome territory radial positioning, LADs are typically found in repressed heterochromatin microenvironments. These ideas are supported by the observation that disruption of the nuclear lamina in patients with progeria results in perturbation of telomere length [57], radial CT positioning [58], and the condensation and compartmentalization of peripheral heterochromatin [53; 59-62].

Beyond the chromosome territories and the nuclear lamina, even though the nucleus is not separated by a membrane-bound compartment, the regulatory machinery for transcription, DNA repair and replication are architecturally organized into sub-nuclear compartments with specific functions. There are prominent domains inside the nucleus, such as the nucleolus and heterochromatin, which were identified long before their functions were elucidated. Furthermore, other architectural components, including the nuclear matrix, nuclear lamina, Cajal bodies, speckles, transcription factories and the Barr body were identified once advances in microscopy became available and molecular approaches were introduced [63-68].

Nucleolus is one of the most extensively studied prominent domains impacting nuclear organization is the nucleolus. They are formed by the congregation of the acrocentric chromosomes 13, 14, 15, 21, and 22, specifically through the nucleolar organizer regions (NORs), which are comprised of tandem arrays of ribosomal RNA (rRNA) genes in the interphase chromatin. Detailed biochemical and DNA-sequencing analysis of nucleoli indicates that not only the acrocentric NOR-bearing chromosomes

interact with nucleoli, but all chromosomes have nucleolar-associating domains (NADs). These NADs share sequence similarity with LADs and localize to the perinucleolar heterochromatin [65]. This configuration of the chromatin does not persist through mitotic phase of the cells. During mitosis, the sub-nuclear compartments are diminished, the nuclear envelope is broken down, and the chromatin becomes tightly compacted forming the mitotic chromosomes [21]. However, as the cell exits mitosis, chromosomes need to de-compact and resume its transcriptional and other functions, recapitulating the cellular program inherited from the parent cells. This phenomenon is accomplished via a mechanism called "mitotic bookmarking", in which certain transcription factors and chromatin modifiers are retained on the chromosomes throughout the mitotic cell cycle [69]. Having gone through mitosis, the bookmarking mechanism enables the rapid resumption of cell-type specific gene expression [70].

With the advancement of recent molecular methods (discussed below), it was discovered that, similar to prior observations [50-52], megabase scaled genomic compartments were discovered. The A-type (open) compartments are gene-rich, transcriptionally active, early replicating, and active whereas the B-type (closed) compartments are gene-poor, transcriptionally more repressed, late replicating, and enriched in repressive histone modifications [71]. The localization of these compartments correlates very well with the timing of replication [72]. More recent studies identified the presence of clusters of 100kb to 1Mb sized interaction domains within the genomic compartments, which were named as topologically associating domains (TADs) [73-75]. The genes within a TAD displayed correlated gene



expression and transcriptional regulation. TADs are stable across species, during differentiation, and in response to hormonal regulation [76-78]. It was shown that insulator proteins, such as CTCF, cohesin and condensin, are enriched at TAD boundaries [73; 79; 80]. Very recently, it was shown that the orientation of CTCF (CCCTC-binding factor) binding was strongly correlated with TAD formation [81]. It is still debated whether TAD formation is primarily driven by the enrichment of genes and the transcription factors at the boundaries, or by the interactions within a TAD [82]. More studies are required to delineate the nuclear regulation in the context of TAD structures.

### **1.3 Methods to Study Chromatin Organization**

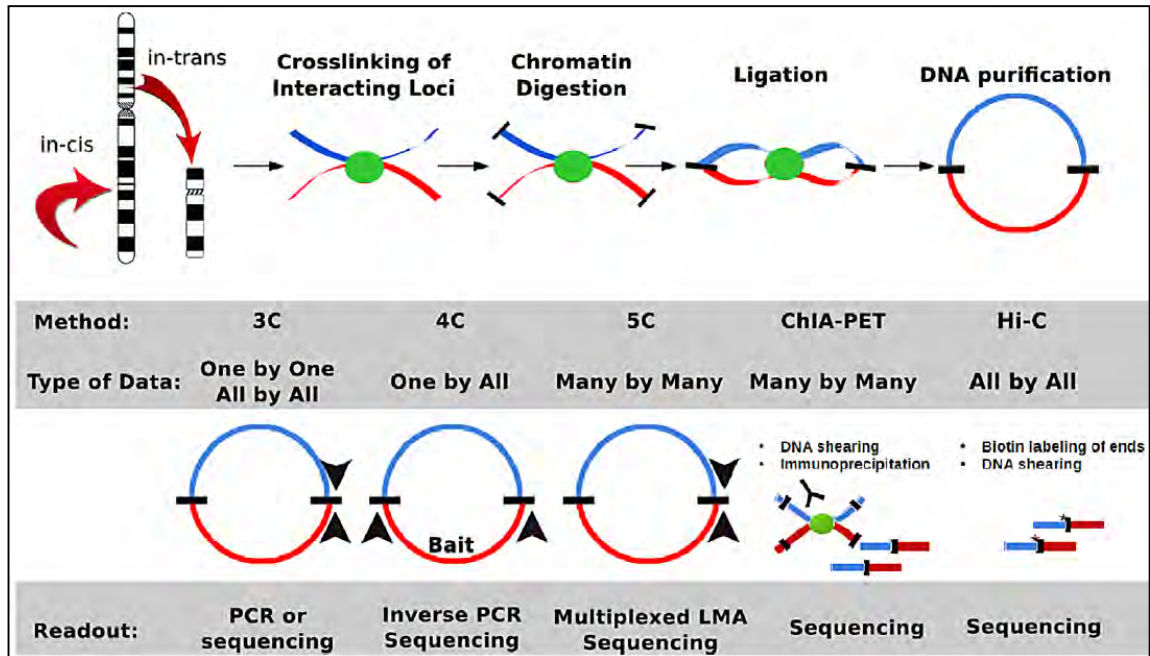
Chromatin organization is fundamental for biological processes that include transcriptional regulation, DNA replication and chromosome segregation. It has been long recognized that nuclear and chromatin organization is strongly associated with gene activity and chromatin state. This is illustrated by the demonstration that genes positioned near the nuclear periphery are often silenced, whereas genes in the interior of the nucleus are frequently active. Furthermore, changes in chromosome condensation and decondensation during the cell cycle were observed more than a century ago. Because it affects so many biological processes, understanding the principles of genome folding has been, and is still, an intense area of research.

The two major techniques used to observe the spatial organization of chromatin can be broadly categorized as microscopic and molecular assays. Light microscopy can

provide information about the shape and the distribution of chromosomes at a resolution of 50–100 nm in single cells. Electron microscopy, though it provides unprecedented resolution, does not provide sequence specific information about the structures observed. On the other hand, molecular assays provide a relative spatial-contact relationship among genomic loci for a population of cells.

### **1.3.1 Chromosome Conformation Capture based techniques**

Many of the molecular techniques rely on nuclear ligation [83], including the chromosome conformation capture methods (3C) [84]. The 3C technique captures the population-averaged interaction frequency of two loci based on their spatial proximity in the three-dimensional nucleus. Chromosome conformation capture and 3C-derived techniques rely on the same basic biochemical steps to capture chromatin interactions: cross-linking of the nucleus with formaldehyde, fragmenting chromatin by restriction digestion, re-ligating the digested ends in a dilute solution to favor intra- over inter-molecular interactions, and finally detecting the chimeric ligated products by PCR or deep sequencing. The resulting sequence information reveals the interactions of distant genomic fragments in the linear genome (Figure 1.1). Depending on the biological question, each approach presents advantages and disadvantages.



**Figure 1.1. An overview of chromosome conformation capture techniques.** They share a procedural front end that includes crosslinking, restriction-enzyme digestion, DNA ligation in dilute conditions, and DNA purification. Downstream, however, there are major differences in detection of these ligation products that yield significantly different types of data.

### 1.3.1.1 Chromosome Conformation Capture (3C)

Chromosome conformation capture is used to identify the interaction frequency of two selected fragments in the genome. At the end of the 3C procedure, a pool of genome-wide intra- and inter-chromosomal interactions is generated [85]. However, because the interaction frequency between any two fragments are analyzed in a pairwise manner (one by one) by PCR using specific primers for each fragment, researchers are limited to analyzing only a few loci or a genomic region within relatively small regions (10 kb to 1 Mb). Thus, 3C is considered a “hypothesis driven” technique, as a priori knowledge about the genomic locations of the elements to be tested is required. The drawbacks of the 3C techniques is that even though one can infer whether two fragments are in close proximity, the data do not necessarily infer functional relevance. Additional

experiments, such as luciferase-reporter assays, should be performed to assess the functional relevance of these looping interactions. Moreover, 3C does not provide information about the localization of these fragments in the nucleus (i.e., peripheral or interior). If this is of interest for the biological question, supplementing the 3C results with DNA-FISH is advised. Furthermore, 3C cannot determine the proximity of individual haplotype chromosomes—the data do not distinguish whether the paternal or the maternal chromosome, or both, make the long-range contact.

#### **1.3.1.2 Circular chromosome conformation (4C)**

A limitation of the 3C technique is that one can only look at a portion of the picture, which is constrained by the number of restriction fragments the 3C primers have been designed to query and the genomic distance (up to 1Mb). However, it is well established that promoters and enhancers can establish long-range interactions many megabases away with other regulatory regions. In order to circumvent this issue and probe the genome-wide interactions of a single fragment (one versus all), Simonis *et al.*, Wurtele *et al.*, and Zhao *et al.* [86-88] respectively developed chromosome conformation capture on ChIP, circular chromosome conformation capture and open-ended chromosome conformation capture techniques. Although these parallel methods answer the same biological question, they differ slightly in their experimental procedures. In the 4C technique, DNA fragments that are ligated to the “bait” fragment are amplified, and the amplified pool of interacting “prey” fragments are detected by either deep sequencing or microarray analysis. More recently a modification of 4C, called enhanced 4C (e4C), has been reported [89]. Using this technique, the interactions

of a bait fragment bound by a specific protein of interest can be evaluated. The advantage of 4C is that it can detect all of the intra- and inter-chromosomal interactions of a specific fragment (or an element) in high resolution. 4C has been widely used to study genes involved in development and disease [45; 90-92].

### **1.3.1.3 Chromosome Conformation Capture Carbon Copy (5C)**

Transcriptional regulation occurs in an orchestrated manner that typically involves several protein complexes and a congregation of cis-regulatory elements such as enhancers and insulators. In other words, a genomic region can, and many times does, participate in interactions with multiple elements. For instance, multiple enhancers can loop with a single developmental gene promoter to fine-tune its transcription. In addition, genes within a certain genomic region can be subject to similar modes of regulation. As a result, it can be important to chart the interactions between a large number of fragments in a genomic region. To address this need, Dostie *et al.* developed the chromosome conformation capture carbon copy (5C) technique [93]. Briefly the 5C method begins with preparation of a 3C library. Then, several to several hundred 5C primers are designed to span a large genomic region of interest such that the primers will anneal precisely at the ligation junctions of the restriction fragments in the 3C library. Next, the fragments are subjected to ligation mediated amplification (LMA), to simultaneously amplify thousands of 3C junctions in a single reaction. The resulting PCR amplicons are detected by either microarray analysis or deep sequencing. The 5C method has been successfully used to study the  $\beta$ -globin and HOX loci, and the embryonic stem cell gene loci, Klf4, Sox2, and Nanog and 1% of the genome [93-98].

As a result, 5C can provide in-depth information about the higher-order chromatin organization of a selected genomic region. 5C data from multiple samples or physiological conditions can be compared to infer biological function. Yet, the 5C approach is not genome-wide and requires a priori-defined region of interest.

#### **1.3.1.4 Genome-wide chromosome conformation capture (Hi-C)**

For many biological questions, understanding the overall interaction frequency of the genome in certain physiological conditions or disease states is desirable. In these cases, it is impractical to investigate interaction frequencies using probe based 3C approaches (i.e., 3C, 4C, or 5C). To capture an unbiased view of genomic interactions, Lieberman–Aiden and van Berkum et al. devised the Hi-C method [71]. A detailed protocol and guidelines for data interpretation has been extensively explained [99; 100]. Hi-C is very similar to 3C in terms of methodology, except that, after the restriction digestion, the digested ends are treated to incorporate biotin prior to the diluted ligation step. After ligation, all chromosome interactions can be captured genome-wide in an unbiased manner by recovering ligated fragments using streptavidin. The Hi-C method has been extensively used in studies that address a myriad of biological questions, related to disease, stem cell biology, evolution and cell cycle [62; 77; 78; 101]. Interaction frequencies in Hi-C data represent the population average of several million cells. Thus, the Hi-C technique cannot distinguish whether interactions are stable and present in some cells, and non-existent in others; or are dynamically present in all cells. DNA-FISH, on the other hand, can provide information about single cells, but it cannot be used to simultaneously incorporate information about the proximity of several loci. Used

as complementary methods, Hi-C and FISH can provide comprehensive and complementary information about chromosome conformation [97]. More recently, Hi-C has also been applied to single cells [102]. These analyses reveal that larger structures in the nucleus (such as chromosome territories) are stable among different cells; however, higher resolution structures (i.e., compartments) are more variable. Another modification of Hi-C is called Capture Hi-C, in which a pool of sequences of interest (i.e., promoter sequences or specific disease loci) is enriched in the library prior to sequencing. This targeted technique can provide a very high-resolution interaction map of enriched sequences [103-105]. Taken together, the Hi-C strategies for chromosome conformation analysis are promising tools to elucidate different roles of genome organization in the cell nucleus in physiologic and pathologic states.

#### **1.3.1.5 ChIA-PET, 6C, and ChIP-Loop**

All of the techniques explained thus far are used to study genome organization from a “DNA-centric” point of view. Because DNA organization is established and maintained by protein and RNA complexes, several methods have been devised to study genome structure from the protein perspective. One common approach is to combine chromatin immunoprecipitation with 3C. DNA is first cross-linked, the protein of interest is immunoprecipitated using an antibody, and 3C is then performed. These techniques not only reduce background, they also enrich for genomic interactions that are enriched for binding of the protein of interest. However, detecting chromosome interactions with these techniques does not necessarily mean that the interaction is mediated by the protein. Additional experiments, such as knocking down the protein of interest, are

required to assess the functional relevance of the data. 6C and ChIP-loop can query chromosome interactions in “one by one” manner, requiring the design of fragment specific probes [106; 107]. However, the ChIA-PET technique uses deep sequencing to assess the protein bound interactome genome-wide [108]. ChIA-PET has been thus far performed on the estrogen receptor, Pol2, and CTCF [108-110]. Although ChIA-PET gives a genome-wide view of chromatin interactions, this view is focused on the protein of interest under investigation.

### **1.3.2 Microscopy based techniques**

Fluorescence in situ hybridization (FISH) is a microscopic method to visualize specific DNA or RNA sequences within the nucleus by hybridizing complementary fluorescent-dye labeled DNA or RNA probes. There are a number of FISH approaches that can be applied to metaphase chromosomes, interphase chromosomes and oligonucleotide arrays [111; 112]. The two-dimensional FISH (2D-FISH), 3D-FISH, and cryo-FISH are variations of the FISH procedure that can directly visualize and measure the nuclear distance between DNA segments, the nuclear localization of DNA segments, whole chromosomes, and the localization of a DNA segment in relative to the rest of the chromosome. FISH has been used extensively to identify chromosomal aberrations (e.g., translocations or deletions), and characterize chromosome territories, and other nuclear bodies and microenvironments. Due to the minimum size limits of FISH probes, distances <100kb cannot be resolved with traditional FISH techniques. FISH is ideal for analysis of individual chromosome territories. However, the procedure takes time, and it is low throughput as the current protocols for microscopy and image analysis are time



consuming. With the microscopy based techniques, target sequence information is required for the design of the probes. With the most advanced equipment, it is theoretically possible to use five fluorescently labeled probes simultaneously; however, most studies query only two or three probes at a time to avoid technical difficulties. Conventional FISH is performed in fixed nuclei (data are from individual cells at the time they were prepared for microscopy). Recently, live-cell FISH has been achieved using catalytically inactive CRISPR/Cas9 labeled with EGFP and tiled sgRNAs [113]. FISH is ideal for determining the position of individual chromosomes territories in the nucleus and/or relative to nuclear compartments. Conventional microscopy resolution is constrained by the diffraction limit of light. Super resolution microscopy [114; 115], on the other hand, can increase resolution up to 20 fold over conventional microscopy [116]. Taken together, microscopy remains an important approach for high-throughput genome-wide studies. Used in combination with emerging molecular chromatin conformation capture techniques, and with more sophisticated and powerful analysis methods, it is likely that ongoing investigation will yield ever more precise and detailed information on the role of chromosome territory organization in biology.

#### **1.4 Higher-order chromosome organization during development**

During cellular differentiation, the cell responds to internal and external environmental and cellular signals. These responses are present in the form of gene regulatory networks, where the expression of a gene is regulated by other genes and proteins. The complexity of these gene regulatory networks are mediated by interactions between DNA-protein, RNA-protein, protein-protein and RNA-DNA

molecules. Harboring several layers of methods of regulation, including mRNA processing, translational regulation and post-translational modifications of proteins, one of the fundamental levels is the transcriptional regulation. Transcriptional regulation occurs when transcription factors, along with their co-factors including long non-coding RNAs, bind to the DNA at the regulatory regions such as promoters and enhancers, and affect the transcription of the regulated gene by modulation of the RNA polymerase complex. Binding of transcription factors often appear in clusters in an orchestrated way, presumably via making protein-protein interactions between the transcription factors at the regulatory regions. At a broader scale, many changes in chromosome territory positioning accompany the dynamic alterations in target genes' loci.

As explained previously, the chromatin is not a one dimensional entity. Genome organization is a fundamental effector of gene expression. Many genes are regulated by making long-range interactions between the regulatory elements and their cognate promoters. In certain instances, several promoters cluster together and are regulated by a single regulatory-element. Because cellular differentiation results in extensive alterations in transcriptional regulation, it is important to understand the differentiation-mediated higher-order chromatin structure changes to gain insight into the underlying biology of development.

One of the landmark examples of higher-order chromatin structure affecting gene expression is the X chromosome inactivation [117], where the X chromosome that is deemed to be inactivated is compacted into the Barr body near the nuclear periphery. Another remarkable example is found in the rod photoreceptor cells in the retina of

nocturnal and diurnal mammals [118]. The chromosome organization in the diurnal rod cells is similar to the conventional architecture present in nearly all other eukaryotic cells. However, the rods of nocturnal mammals possess an inverted organization, where the heterochromatin is localized at the nuclear center and the euchromatin is situated at the nuclear periphery. This type of organization allows the nocturnal rod photoreceptor to efficiently channel the light [118].

At the finer scale, it was shown that the interactions between the upstream locus control region (LCR) and the  $\beta$ -globin locus 80kb away make a long-range looping interaction at the onset of globin expression. All the globin genes in the locus, including the fetal globin and adult globin genes, are spatially clustered forming the active chromatin hub (ACH) [119]. During embryonic and adult development, the chromatin interactions between the globin genes and the LCR dynamically follow the transcriptional changes that accompany the order of globin gene expression, from fetal to adult globin genes [120]. Many transcription factors and chromatin remodeling factors were found to drive or stabilize this looping formation [121-123]. Similarly, the  $\alpha$ -globin gene promoter displays long-range interactions with its distal enhancer element [124].

Another developmental gene is the homeobox (Hox) gene cluster. There are 4 clusters in the mouse genome and in each cluster the individual genes are sequentially activated along the anterior-to-posterior body axis to organize the body plan during mammalian development [125]. Using 4C approaches, it was shown that there are dramatic architectural changes at the Hox gene cluster during embryonic development

[91].

.The  $\beta$ -globin and the Hox loci are only two examples of many, yet still highlights the importance of long-range interactions in the regulation of the loci. Transcription factors, co-factors, non-coding RNAs and epigenetic marks are all involved in this process. Understanding the underlying mechanisms of three-dimensional chromatin folding is an important challenge in order to possible allow manipulating cell differentiation for clinical and therapeutic purposes. In the second part of this thesis (Chapter 2), differentiation-dependent chromatin changes during bone-differentiation will be presented.

### **1.5 Higher-order chromosome organization during tumorigenesis**

Changes in nuclear morphology have been used for almost 140 years as a major diagnostic tool to detect cancer [126; 127]. One of the first examples of a microscopic observation of cancer was by Lionel S. Beale of King's College Hospital, London, in 1860. He studied the unstained sputum from a patient with pharynx cancer and identified nuclear size and shape variations in cancer cells [128]. The nuclei of cancer cells, irrespective of their tissue origin, are generally larger and more irregularly shaped than the healthy cell nuclei [126]. Certain types of cancer nuclei are characterized by an increased number and size of nucleoli [129; 130]. Several cancer types display partial loss of heterochromatin / euchromatin compartmentalization. The physical proximity of the chromosome territories dictate the frequency of the translocation events as assessed by microscopy [33; 38; 131; 132], as well as Hi-C [62].

Cases of aberrant chromosome territory movements are also observed in cancer. For example, in the majority of human pancreatic cancer cells, chromosome 8 moves toward the nuclear periphery [133]. Chromosomes 18 and 19 show position changes in multiple types of cancer [134; 135]. Beyond the movement of whole chromosomes, the centromere of chromosome 17 becomes more internally localized in cancer compared to normal nuclei [133]. In a finer scale, individual gene loci display nuclear re-positioning in cancer cells, a phenomenon which has been utilized as a diagnostic tool [136; 137]. Apart from whole-scale nuclear alterations, there are thousands of reports suggesting perturbations in the epigenetic landscape of cancer cells (reviewed in [138]). These changes include aberrant DNA methylation, histone modifications, binding of several transcription factors and chromatin modifiers to ectopic regions potentially resulting in perturbed gene expression.

The mechanisms of the altered higher-order nuclear structure to functional regulation, despite the wealth of data available, is not fully understood. The task of unraveling the principle consequences of altered genome structure represent one of the biggest challenges in molecular and cell biology. In the majority of this thesis (Chapters 3, 4 and 5), different studies to understand the molecular aspects of the nuclear structure of the breast cancer genome will be presented.

## **CHAPTER 2: The bone-specific Runx2-P1 promoter displays conserved three-dimensional chromatin structure with the syntenic Supt3h promoter.**

### **2.1 Introduction**

The Runt-related transcription factor 2 (Runx2/CBF $\alpha$ 1/AML3) is essential for osteoblastic differentiation and is required for bone and cartilage development [139-143]. A complete knockout of *Runx2* leads to embryonic lethality marked by an absence of bone development and ossification [141; 144; 145]. Moreover, Runx2 interacts with the nuclear matrix to affect histone modifications and chromatin remodeling [127; 146; 147].

The murine *Runx2* gene is located on chromosome 17 and spans a region of ~210-kb. Two predominant *Runx2* isoforms are transcribed from distinct promoters. The *Runx2 type-II* isoform controlled by the *P1* promoter is exclusively expressed in osteo-progenitor cells and is stimulated upon bone formation. The *Runx2 type-I* isoform is driven by the *P2* promoter and is expressed in both osteogenic and non-osteogenic mesenchymal tissues [148]. During embryonic development, *P1* promoter driven *Runx2 type-II* is the major isoform expressed in the developing skeleton [149]. Consistent with this pattern, the specific loss of expression from the *Runx2-P1* promoter in mice results in severe developmental defects with cleidocranial dysplasia (CCD)-like symptoms [149].

*Runx2* displays a syntenic relationship with the Suppressor of Ty3 homolog (*Supt3h*) gene, whose promoter resides within the first intron and ~38-kb downstream of the

*Runx2-P1* transcriptional start site (TSS). SUPT3H is a TAF associated protein that is a component of the human histone acetyl transferase STAGA complex (SPT3-TAF9-GCN5-acetylase) [150-153]. These two genes have different expression profiles; *Runx2-P1* is developmentally regulated, while *Supt3h* is ubiquitously expressed and is essential in all tissues. Interestingly, the syntenic relationship between *Runx2* and *Supt3h* is conserved among species from humans to sponges [154], which suggests the existence of an evolutionarily conserved selective pressure to preserve this syntenic relationship. This pressure may be due to a shared or linked regulatory control mechanism and a potential for crosstalk between these two genes [155; 156].

In this chapter, we investigated the higher-order organization of the *Runx2* locus in several cell types. Mining the ENCODE database through the WashU Epigenome Browser [157], we identified long-range associations between the *Runx2-P1* and *Supt3h* promoter regions. Carrying out chromosome conformation capture (3C) analyses in RAW 264.7 murine macrophage cells, where *Runx2-P1* is silent, we confirmed the existence of this interaction. As *Runx2-P1* activity is increased during osteoblastogenesis, we next asked whether this interaction is dynamic and functional. Interestingly, 3C analyses revealed an increase in the interaction frequency between the *Runx2-P1* and *Supt3h* promoters in MC3T3-E1 murine pre-osteoblast cultures throughout osteoblastic differentiation. The *Supt3h* promoter also showed enrichment for DNaseI hypersensitivity (DHS) and CTCF and RUNX2 localization with the onset of osteogenesis. Finally, we provide evidence that the *Supt3h* promoter can interact with *Runx2-P1 in-trans* and modulate its expression in a differentiation-dependent manner.

## **2.2 Materials and Methods**

### **2.2.1 WASHU epigenome and UCSC genome browser search**

Online interaction data was obtained from the WASHU epigenome browser ([20](#)) and the UCSC Genome Browser [158]. IMR90 HiC data [73] and the PolII ChIA-PET data [110] were extracted for the hg19: chr6:45,250,000-45,370,000 genomic coordinates. The DNaseI hypersensitivity tracks were obtained for University of Washington (UW) tracks for MCF7 and K562; and from Duke University tracks for IMR90. The MCF7 CTCF ChIP-seq data were extracted from University of Washington (UW) generated tracks. Both CTCF and PolII tracks were obtained from Stanford/Yale/Duke/Harvard (SYDH) tracks for K562 and IMR90 cells. The MCF7 PolII ChIP-seq data were obtained from UT Austin tracks. All RNA-seq data were gathered from the Cold Spring Harbor Laboratory ENCODE tracks.

### **2.2.2 MC3T3-E1 Cell culture**

The MC3T3-E1 clone-4 pre-osteoblastic murine cell-line [159] was obtained from the American Type Culture Collection (ATCC, Manassas, VA). Growth-phase cultures were maintained in  $\alpha$ -MEM without ascorbic acid (Hyclone, Thermo Fisher Scientific, Rochester, NY) and supplemented with 1% penicillin-streptomycin (Gibco, Life Technologies, Grand Island, NY), 2 mM L-glutamine (Gibco, Life Technologies, Grand Island, NY), and 10% Fetal Bovine Serum (Hyclone, Thermo Fisher Scientific, Rochester, NY). When cultures reached ~90% confluency, differentiation was initiated by the addition of 142  $\mu$ M ascorbic acid (Sigma-Aldrich, St. Louis, MO). After two



days, the ascorbic acid concentration was increased to 280  $\mu$ M and 5 mM  $\beta$ -Glycerophosphate (Sigma- Aldrich, St. Louis, MO) was added. Cultures were maintained at 37°C and at 5% CO<sub>2</sub>.

### **2.2.3 Chromosome Conformation Capture (3C) of the *Runx2* locus**

3C assays were performed as previously described [84; 85], with the following modifications: 3C restriction fragments were defined by *Bgl*III enzyme digestion. The anchor fragment used to query *Runx2-PI* chromosomal interaction spans from -975 to +1113 (mm9 chr17: 44,950,469-44,952,567).  $\sim 1 \times 10^8$  MC3T3-E1 cells were fixed with 1% formaldehyde in serum free  $\alpha$ MEM for 10 min at room temperature. Formaldehyde was quenched by the addition of 0.125 M glycine. Nuclei were released by dounce homogenization in ice-cold lysis buffer (10 mM Tris-HCl pH 8.0, 10 mM NaCl, 0.2% NP-40) containing cOmplete, Mini Protease Inhibitor Cocktail (Roche Applied Science, Indianapolis, IN). Nuclei were collected and subjected to overnight digestion with 400 U of *Bgl*III (New England BioLabs, Ipswich, MA). The enzyme reaction was halted by incubation at 65°C for 30 min in the presence of 10% SDS. Samples were aliquoted into 22 separate tubes and were diluted 40-fold in ligation buffer [85] and subjected to proximity mediated ligation with 10U of T4 DNA Ligase (Invitrogen, Life Technologies, Grand Island, NY) per reaction for 4 h at 16°C. Nuclear material was reverse cross-linked by overnight incubation with Proteinase K at 65°C. Ligated chromatin was extracted by phenol-chloroform extraction followed by ethanol precipitation. 3C primers that span the *Runx2* gene locus were designed by Primer3 software and are listed in Appendix 1. The annealing temperatures of all 3C primers

were  $60 \pm 1^\circ\text{C}$ . The PCR conditions were  $95^\circ\text{C}$  for 8 min followed by, 35 cycles of  $95^\circ\text{C}$  for 30 sec,  $60^\circ\text{C}$  for 30 sec, and  $72^\circ\text{C}$  for 30 sec, followed by  $72^\circ\text{C}$  for 8 min. All 3C PCR products were analyzed on 1% agarose gels stained with ethidium bromide. Gel quantifications were analyzed using the GEL-QUANT software ([www.gelquant.org](http://www.gelquant.org)).

Interaction frequencies were determined by assessing fold-change of 3C PCR amplification product of sample chromatin compared to randomly ligated *BglIII* digested bacterial artificial chromosomes (BACs) that span the *Runx2* locus. The following BAC clones, which span the *Runx2-Supt3h* locus and a gene desert region, were used: BACPAC CHORI (Children's Hospital Oakland Research Institute) catalog numbers RP23-22H7, RP23-92H18, RP23-356F5, RP23-443F11, and RP23-238O6. The ligation efficiencies of all 3C samples were normalized to each other by taking the  $\log_2$  average of the ligation frequencies of a gene desert region [160] for samples generated with *BglIII*, and ERCC3 locus for samples generated with *HindIII*. The BAC control template was prepared by mixing the different BACs in equimolar concentrations, followed by digestion and ligation. Then, the interaction frequency was calculated by dividing the amount of PCR product from the 3C template by the amount of PCR product from the BAC control template, thereby normalizing for differences in primer efficiencies. All of the 3C primer pairs yielded similar amounts of product with both the 3C and the BAC templates. Primers that gave very low PCR yields were discarded. 3C data represent the averaged ligation frequencies of two independent cultures quantified in three separate library preparations. Student's t-test was used to assess the p-values.

#### **2.2.4 DNase-seq**

Genome-wide DNase-hypersensitivity mapping of osteoblast cultures was performed by adapting the DNase-seq protocol from Song *et al.* [161] with slight modifications.

Approximately  $40 \times 10^6$  growth-phase (day 0, or d0) or matrix-deposition stage (day 9, or d9) MC3T3-E1 cells were harvested and were each subjected to 4, 12, and 40 U/ $\mu$ L of DNaseI for 15 min at 37°C. Steps involving the isolation of chromatin embedded in agarose, included a treatment with 10 U/mL  $\beta$ -agarase for 2 h at 37°C before extracting with phenol:chloroform:isoamyl alcohol (25:24:1 v/v) and ethanol precipitation. Peak signals in this report represent a single biological sample for each culture condition sequenced twice (combined technical duplicates) and normalized using align2rawsignal (Kundaje A., <http://code.google.com/p/align2rawsignal/>). DNase-seq analysis was validated by 4-fold representation (two biological replicates, each with technical duplicates) that pass ENCODE Consortium standards on F-seq called peaks [162] using IDR analysis (data not shown, Tai *et al.*, manuscript in preparation). The DNase-seq data was deposited under the accession GSE55046.

#### **2.2.5 Reverse-transcriptase qPCR**

Total RNA from cultures was extracted using TRIzol reagent (Invitrogen, Life Technologies, Grand Island, NY) followed by DNase treatment with DNA-Free RNA Kit (Zymo Research, Irvine, CA) according to manufacturer's instruction. cDNA was prepared using the SuperScriptIII First-Strand Synthesis System (Invitrogen, Life Technologies, Grand Island, NY). qPCR was performed with the iTaq SYBR Green

Supermix with ROX (Bio-Rad, Hercules, CA) and on the 7300 Sequence Detection System (Bio-Rad Laboratories, Hercules, CA). Relative transcript levels were determined by the  $\Delta\Delta C_t$  method and normalized to *gapdh*. Primer sequences for *runx2P1*, *runx2P2*, *osteocalcin*, *bone sialoprotein*, and *gapdh* are described elsewhere [149]. Primers for detection of *supt3h* message were designed using FoxPrimer ([www.foxprimer.org](http://www.foxprimer.org), Dobson *et al.*, manuscript in preparation) and are: forward, 5'-AAGGCATTGACGAGGATGAC-3' and reverse, 5'-TCTTCAAACATTGCCAGCAG-3'. Student's t-test was used to assess the p-values.

### 2.2.6 Reporter constructs

The design and preparation of the 3-kb (-2821 to -16) and 0.6-kb (-629 to -16) luciferase constructs are described elsewhere [163]. The *Runx2-P1* 0.9-kb-Luc construct was derived from the 3-kb luciferase construct by deleting sequence between -2821 and -966 using the quick-change method for large fragment deletion [163]. The murine 3.3-kb *Supt3h* promoter region (3315-bp) was PCR cloned from mouse C57BL/6 genomic DNA using Phusion High-Fidelity DNA Polymerase using the following forward and reverse primers: 5'-GCTCGCACTCAGCTTTGGGCA-3' and 5'-GGGAGAGACAGGCAAGGAGGGG-3'. The 3.3-kb *Supt3h* promoter region flanked by KpnI restriction sites was cloned upstream of 0.9-kb *Runx2-P1* pGL3 luciferase vector (GENEWIZ, Inc., South Plainfield, NJ). To generate the *Supt3h*-TOPO construct, the 3.3-kb *Supt3h* promoter region was subcloned into the pCR-4Blunt-TOPO vector using the Zero Blunt TOPO PCR Cloning Kit following manufacturers' recommended conditions. The *Supt3h*-DHS1 and *Supt3h*-DHS2 TOPO and 0.9-kb pGL3 constructs

described here were generated by similar methods using the following primer sets:

*Supt3h*-DHS1:FW,5'-GGA ACT TTG TAG AAA GGA ACG GGG G-3',RV,5'-CAT  
GCG CAC CCG GCT GGC C-3';

*Supt3h*-DHS2: FW, 5'-CGC TCT CGC CGC ACG GC-3',  
RV,5'-CTC CCA TAA ACC TGA GTT TTG AGC TAG G-3';

*Supt3h*-0.5kb: FW, 5'-GAT ATT AGT TGA GCA GAA TTT TAA T-3', RV, 5'-TAC  
TTC ATT AAT GTC TTG CCT ATG-3'

*Supt3h*-0.6kb:FW,5'-TAA CTT CAC AAG AGC TTC GTT TTC-3',RV 5' TAA ACA  
AAC AAA CAA ACA AAC TGC T-3';*Supt3h*-1.1kb:FW 5'TAA CTT CAC AAG  
AGC TTC GTT TTC-3', RV 5'-TAC TTC ATT AAT GTC TTG CCT ATG-3'

The empty-TOPO construct described in this report was generated by allowing the pCR-4Blunt-TOPO vector to self-circularize, a low-frequency event that occurs when reactions lack a blunt-ended fragment. The *TK*-pGL3 and SV40-renilla constructs were kind gifts from Dr. Stephen D. Hauschka.

### **2.2.7 Co-transfections and Luciferase reporter assays**

Growth-phase MC3T3-E1 clone-4 cells were grown to >90% confluency and co-transfected with Firefly Luciferase test constructs and SV40-Renilla constructs using Lipofectamine (Invitrogen, Life Technologies, Grand Island, NY) and Plus Reagent (Invitrogen, Life Technologies, Grand Island, NY) according to manufacturer's instructions. 2.5 µg of total plasmid DNA was transfected per 60 mm plate as described

in the Results section. At designated days post-switch, cultures were harvested and reporter activities were measured using the Dual-Luciferase Reporter Assay System (Promega, Madison, WI) on a VICTOR X4 Multilabel Plate Reader (Perkin Elmer, Waltham, MA), according to manufacturers' instructions. Each test condition described is represented by at least 9 replicate plates, and statistical significance values are reported where applicable.

### 2.2.8 Interplasmid 3C

We adapted a 3C-qPCR [164] approach for analyzing the interaction frequency between two transiently co-transfected plasmid constructs. We have named this method Interplasmid 3C (i3C). The following primers were designed to span *SphI* restriction sites: TOPO 5'- GCACGTACTCGGATGGAAG -3', pGL3 5'- CCGAGTGTAGTAAACATTCCAAAAC-3', Runx2 internal control FW 5'- CTCTTCATTTGCACTGGGTCACACG-3' and Runx2 internal control RV 5'- CCAGGGAAGTGGAGGGAAGGGTTG-3'. qPCR was performed as described above. Enrichment of ligation products were assessed by normalizing the Ct values of the 3C-ligation products to an internal loading control. Relative ligation frequency was obtained by normalizing the enrichment values to the empty TOPO - empty pGL3 control combination. Student's t-test was used to assess the p-values.

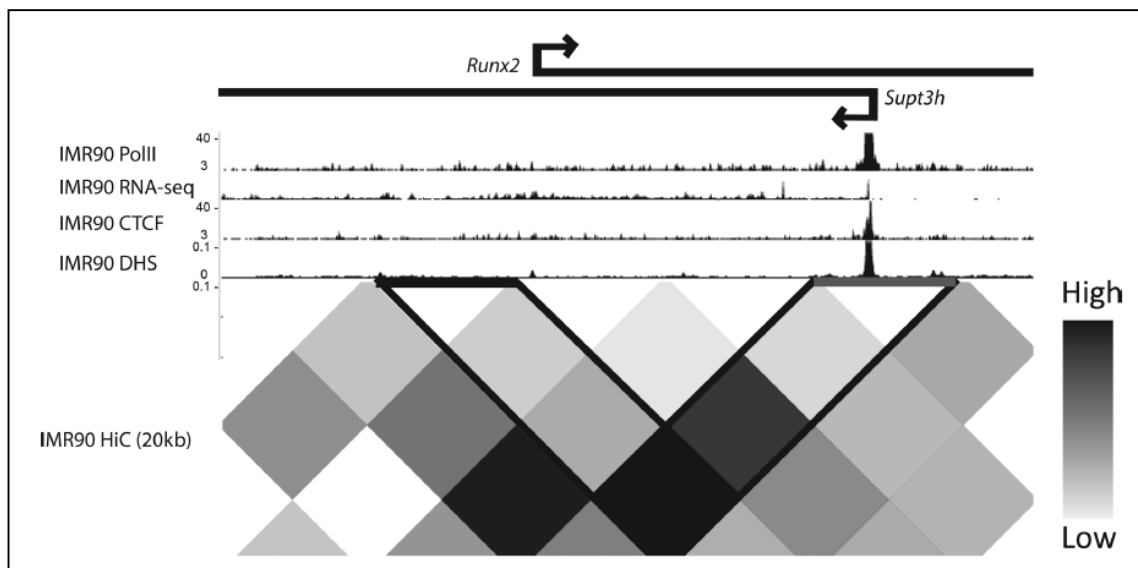
## 2.3 Results

### 2.3.1 Identification of a long-range interaction between the *Runx2-P1* and *Supt3h* promoters

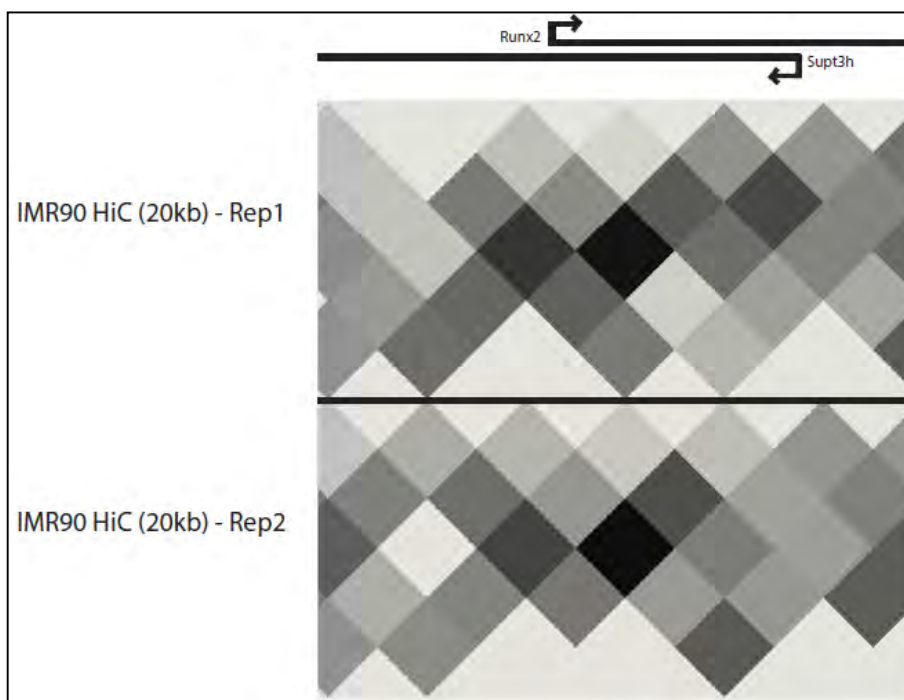
In order to assess the prevalence of long-range looping interactions between *Runx2-P1* and *Supt3h* promoters in a variety of cellular contexts, we examined (using the WashU

Epigenome Browser [157]) publicly available Hi-C and ChIA-PET datasets that are part of the ENCODE project. Because DHS and CTCF are strongly correlated with the presence of long-range interactions [94], we decided to include these marks in our analysis. In addition, RNA-seq and PolII ChIP-seq data were included as parameters of transcriptional activity. Hi-C (High-throughput chromosomal conformation capture) methodology probes long-range chromosomal interactions on a genome-wide scale [71]. ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag Sequencing), on the other hand, identifies chromatin interactions at regions associated with a transcription factor or a complex of interest via a combination of chromatin immunoprecipitation (ChIP) and 3C mapping [108].

IMR90 human lung fibroblasts display a long-range interaction between *Runx2-PI* and *Supt3h* promoter regions, which are ~38kb away from each other (Figure 2.1). Results from individual IMR90 Hi-C experimental replicate tracks show the same result (Figure 2.2). Moreover, RNA-seq data suggest that the *Supt3h* gene is transcribed. Consistent with the observation of the chromatin interaction, the *Supt3h* promoter also harbors DHS, CTCF and PolII binding. There is no detectable RNA-seq signal at the *Runx2-PI* promoter region, as expected from a non-osteogenic cell line (Figure 2.1).



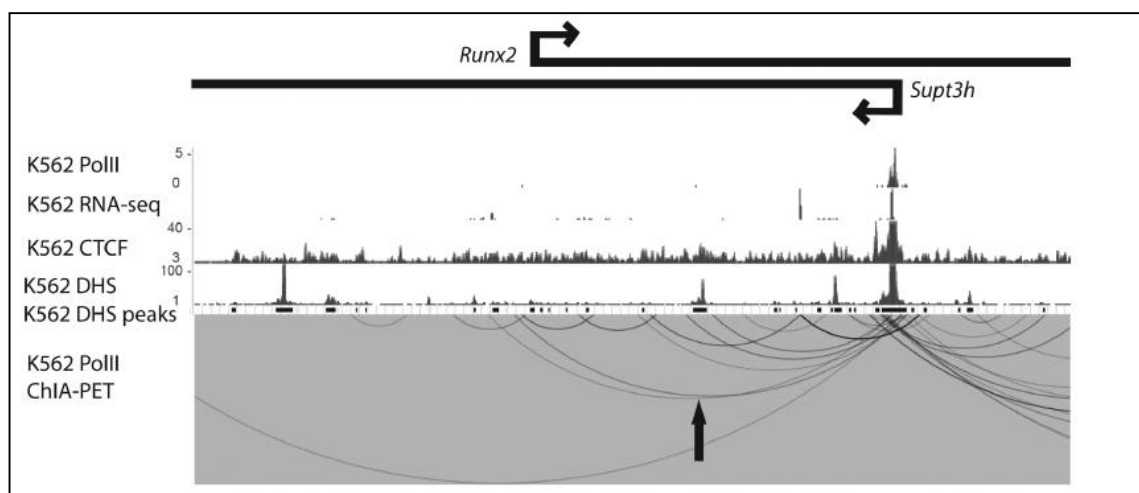
**Figure 2.1. The long-range interactions in the *Runx2* locus in IMR90 Hi-C dataset.** WashU Epigenome Browser snapshot of the Hi-C interaction frequencies in IMR90 cells between the *Runx2*-P1 and *Supt3h* promoter regions. The genes are diagrammed on top, and the transcriptional start sites are indicated by the arrows. The 20-kb regions encompassing the *Runx2*-P1 and *Supt3h* promoters are highlighted with black bars. In the Hi-C heatmap, darker colors represent higher interaction frequency. UCSC genome browser screenshots of ChIP-seq profiling signal tracks for PolIII, CTCF, DHS and RNA-seq data of IMR90 cells are labeled.



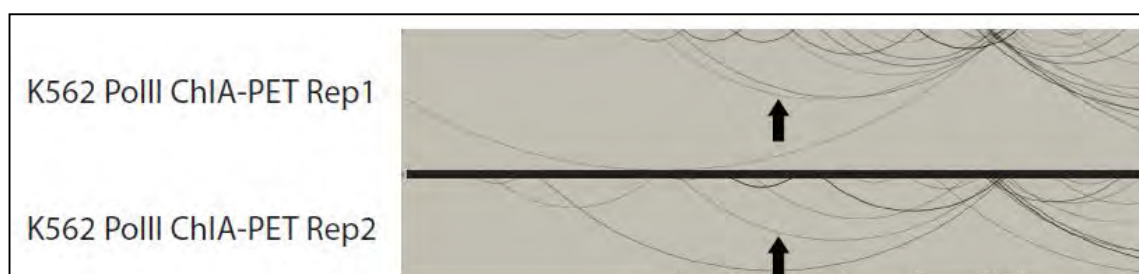
**Figure 2.2. The long-range interactions in the *Runx2* locus in IMR90 Hi-C replicates.** WashU Epigenome Browser snapshot of the Hi-C interaction frequencies in IMR90 cells for both of the replicates.



In K562 leukemia cells, ChIA-PET data for interactions bound by PolII demonstrate many looping interactions within the intervening sequences between *Runx2-P1* and *Supt3h* promoters, as indicated by the different sizes of arcs in Figure 2.3. Comparison of data from individual replicates suggests that the variability of detectable interaction events within this region is fairly high (Figure 2.4).



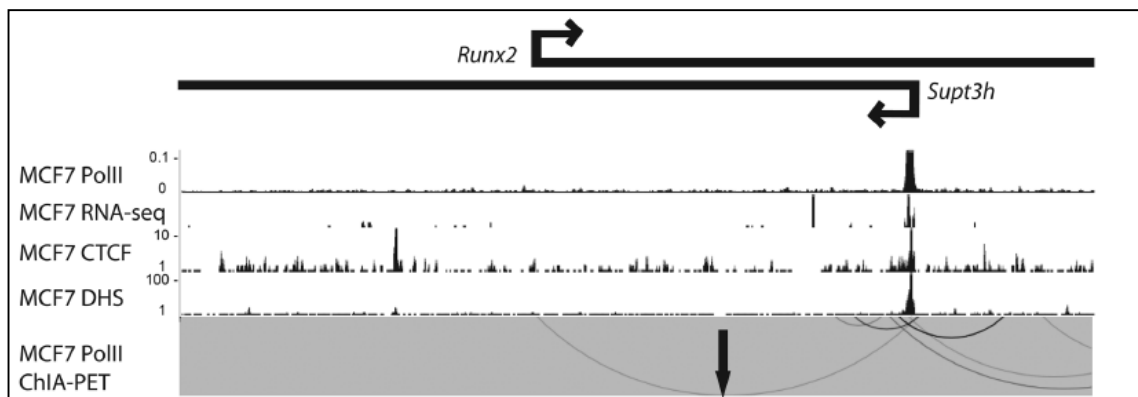
**Figure 2.3.** The long-range interactions in the *Runx2* locus in K562 ChIA-PET dataset. ChIA-PET interactions bound by PolII within the local *Runx2-P1* and *Supt3h* promoter regions, accompanied by UCSC genome browser screenshot of PolII and CTCF ChIP-seq, DHS and RNA-seq signals for K562 cells. The arrows indicate the specific interaction between *Runx2* and *Supt3h* promoters.



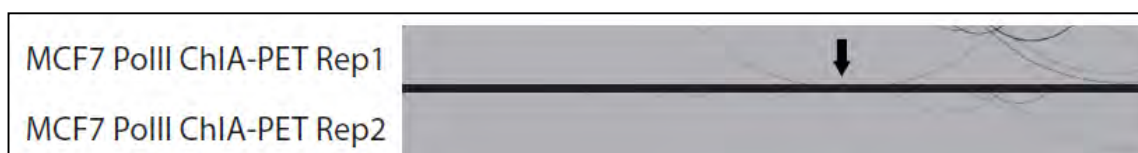
**Figure 2.4.** The long-range interactions in the *Runx2* locus in K562 ChIA-PET biological replicates. WashU Epigenome Browser snapshot of the ChIA-PET interactions for K562 cells for individual replicates.

The interaction between the *Runx2-P1* and *Supt3h* promoter regions is reproducible between these two replicates (Figure 2.4). The epigenetic marks correlating with higher-order chromatin organization such as the presence of DHS, enrichment of CTCF and PolII, and a strong RNA-seq peak are all observed in the *Supt3h* region. Similar to the case in IMR90 cells, there is no detectable ChIP-seq or RNA-seq signal in the *Runx2-P1* promoter (Figure 2.3). Interestingly, the intensity and locations of the looping interactions in K562 cells greatly differ between the *Runx2* and *Supt3h* promoter regions. This may be due to PolII tracking through the *Supt3h* gene body during active transcription. It is worth noting that the *Supt3h* promoter region also makes long-range interactions with regions other than the *Runx2-P1* promoter, suggesting a complex regulatory interaction network for these genes.

ChIA-PET data for MCF7 PolII also suggest a physical interaction between *Runx2* and *Supt3h* promoter regions, accompanied by a DHS, transcription of the *Supt3h* gene, and PolII and CTCF binding to the *Supt3h* promoter (Figure 2.5). There is no transcription detected from the *Runx2-P1* promoter. In comparison with the K562 PolII ChIA-PET data, looping between the *Supt3h* promoter and other regions is observed in addition to many local interactions in MCF7 cells. In this dataset, the interaction between *Runx2-Supt3h* is detected in only one of the two experimental replicates (Figure 2.6), which might indicate contacts are rare in MCF7 compared to K562 cells.



**Figure 2.5. The long-range interactions in the *Runx2* locus in the MCF7 ChIA-PET dataset.** ChIA-PET interactions bound by PolII within the local *Runx2*-P1 and *Supt3h* promoter regions, accompanied by UCSC genome browser screenshot of PolII and CTCF ChIP-seq, DHS and RNA-seq signals for MCF7 cells. The arrows indicate the specific interaction between *Runx2* and *Supt3h* promoters.

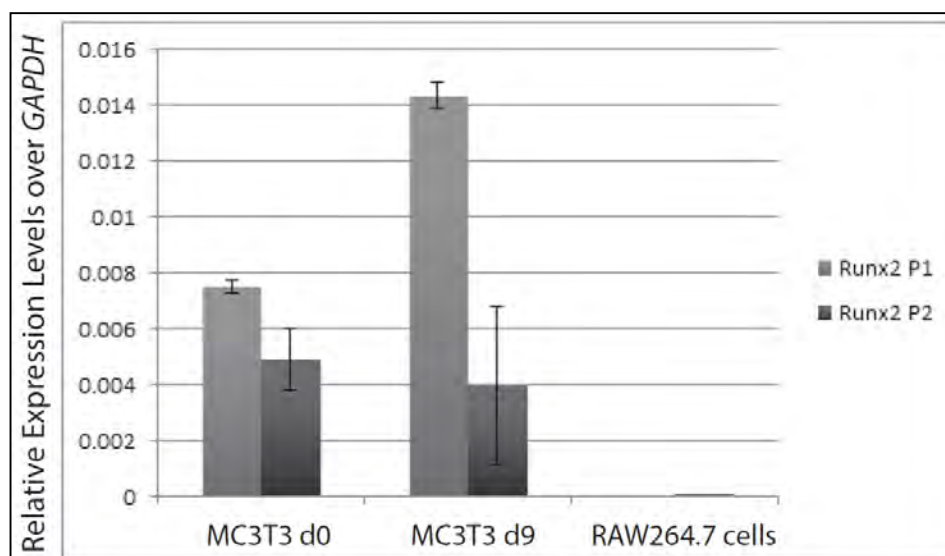


**Figure 2.6. The long-range interactions in the *Runx2* locus in the MCF7 ChIA-PET biological replicates.** WashU Epigenome Browser snapshot of the ChIA-PET interactions for MCF7 cells for individual replicates.

Because there is a long-range interaction between *Runx2*-*Supt3h* promoters in human cell lines that lack detectable *Runx2*-P1 driven transcription as assessed by RNA-seq tracks, we next asked whether this interaction also exists in a murine cell line where *Runx2* is silent. In order to address this question, we used the 3C approach [84; 85] to examine the RAW 264.7 murine macrophage cell line, which exhibits low levels of *Runx2* expression as assessed by RT-qPCR (Figure 2.7).

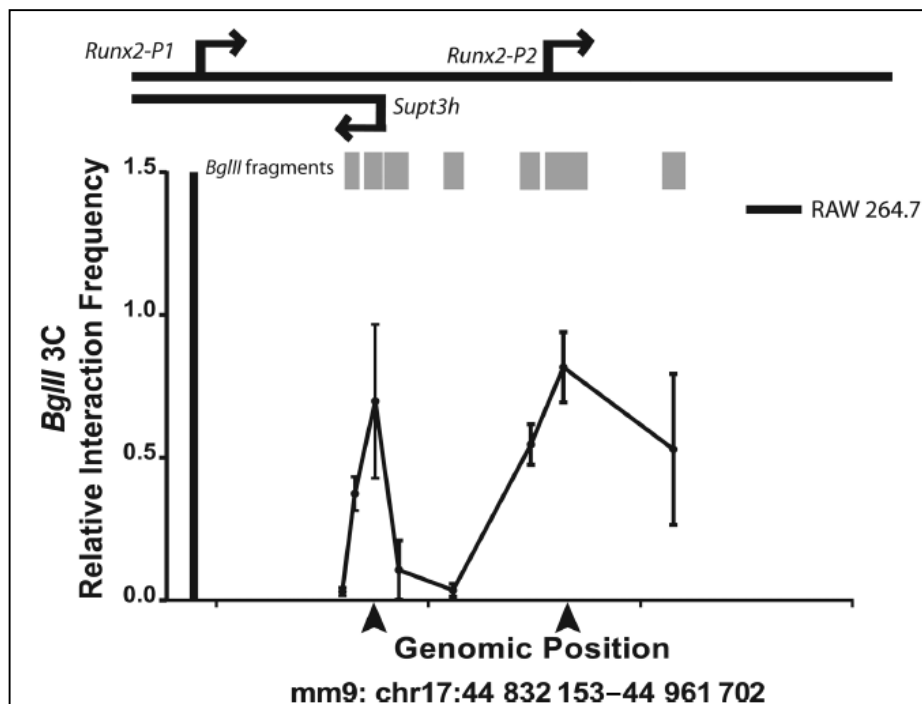
3C is a widely used method that employs the intra-molecular ligation of enzymatically digested cross-linked chromatin. Unique ligation junctions are quantified by PCR to assess the relative proximity of restriction fragments of interest to detect captured interacting chromosomal domains [84; 85]. We analyzed the interaction profile

of the *Runx2-P1* promoter with the BglIII restriction fragments encompassing the *Supt3h* and *Runx2-P2* promoters (Figure 2.8). The 3C primers were designed to query the ligation frequency between the anchor fragment located at the *P1* promoter and BglIII restriction fragments flanking the *Supt3h* and *Runx2-P2* promoter regions (see Appendix, Table 1 for the primer list).



**Figure 2.7. RT-PCR expression levels of *Runx2-P1* and *Runx2-P2* in MC3T3 and RAW264.7 cells.** Relative expression levels of *runx2-P1*, *runx2-P2*, d0 and d9 MC3T3 cultures and RAW 264.7 cells. Relative expression was normalized to *gapdh* levels.

Similar to the cases observed in different human cell lines that have minimal *Runx2* expression, we found long-range interactions between the *Runx2-P1* and *Runx2-P2* regions, and between *Runx2-P1* and the *Supt3h* promoters (Figure 2.8).



**Figure 2.8. 3C analysis of the *Runx2* locus in the RAW264.7 cells.** Chromosome conformation capture (3C) analysis of the *Runx2* gene locus in RAW264.7 macrophages. The x-axis represents the genomic position and the y-axis shows the relative interaction frequency. Anchor BglIII fragment at *Runx2-P1* is indicated with a black bar. Gray bars indicate the BglIII restriction fragments. Arrowheads point at the *Supt3h* and *Runx2-P2* interactions. Error bars: S.E.M.

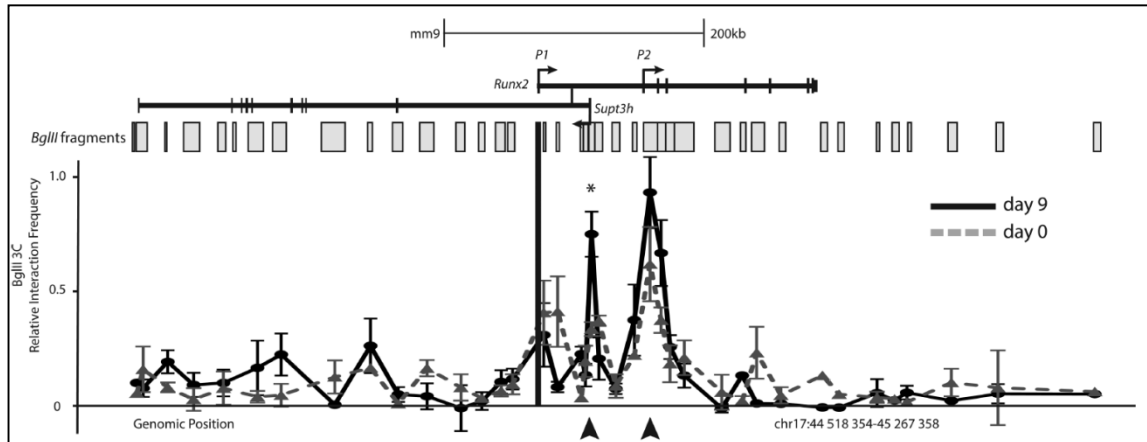
Taken together, these findings suggest that the *Runx2-P1* and *Supt3h* promoters are in close proximity in four different cell types of human and mouse origin. The human cell lines queried show the presence of DHS, and the enrichment of CTCF and PolIII at the *Supt3h* promoter. The presence of DHS and the enrichment of CTCF at this region correlate with the observation of the long-range interaction [94; 95]. In the mouse RAW 264.7 macrophage cells, in which the *Runx2* gene is transcriptionally silent, the same interaction is also observed via 3C analysis. These findings suggest that these interactions represent a static three-dimensional structure established between the *Runx2-P1* and *Supt3h* promoters in cells that have minimal levels of *Runx2-P1* driven transcription.

### **2.3.2 Interaction frequency between *Runx2-P1* and *Supt3h* promoters is increased during osteoblast differentiation**

*Runx2* is fundamental for bone formation and maintenance. Because *Runx2-P1* driven transcription increases during osteogenesis [149; 165], we asked whether the long range association observed between the *Runx2-P1* and *Supt3h* promoters is altered during osteoblast differentiation. We compared the MC3T3-E1 cell line at growth-phase, pre-osteoblasts (d0) vs. differentiating cultures at matrix-deposition stage (d9), since *Runx2* transcript levels have been shown to increase ~2-6-fold by the matrix-deposition stage of osteoblastogenesis [166; 167]. This marked increase in transcription occurs within the first 9 days of differentiation.

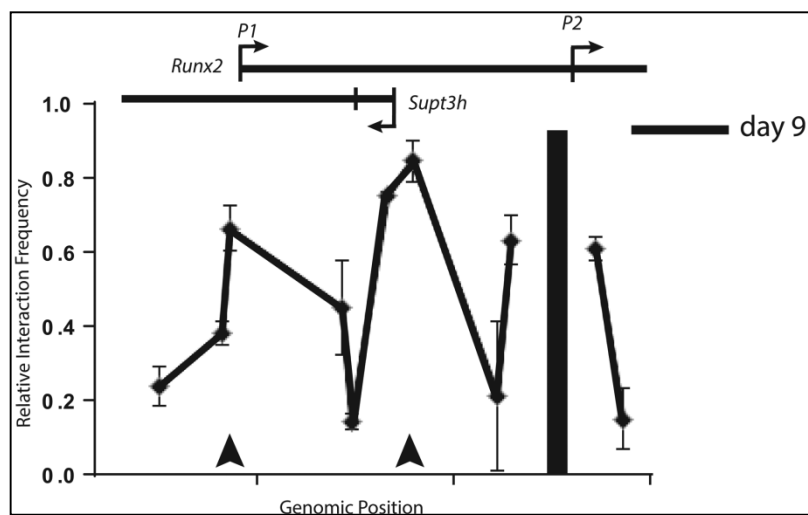
Using the 3C methodology, we queried the interaction profile of the *Runx2-P1* promoter with sequences flanking ~300kb 5' and 3' of this promoter in d0 and d9 cultures. The 3C results show that the *Runx2-P1* anchor fragment displays high-interaction frequency with the fragment encompassing the *Runx2-P2* promoter in both d0 and d9 cultures (Figure 2.9). Cells cultured in differentiation conditions for d9 show a modest, statistically insignificant increase in the interaction frequency between the *Runx2-P1* and *Runx2-P2* promoters. A notable interaction between the anchor fragment and the fragment encompassing the *Supt3h* promoter in d0 cultures was also observed. Interestingly, there was a statistically significant ~2-fold increase ( $p < 0.05$ ) in interaction frequency at this region in d9 vs. d0 cultures (Figure 2.9). Interaction with the *Supt3h* promoter region was among the most significantly changed throughout the entire 600kb *Runx2* locus during differentiation, suggesting a mechanistic link with the conserved

syntenic nature of these genes.



**Figure 2.9. 3C analysis of *Runx2*-P1 promoter in MC3T3 cells.** 3C analysis of the *Runx2* locus in proliferating (d0 and d9) post-differentiation MC3T3-E1 cultures. The genes are diagrammed on top, and the TSSs are designated by the arrows whereas the exons are represented by black bars. The anchor fragment is designated by a black bar at *Runx2*-P1. The x-axis represents the genomic position and the y-axis shows the relative 3C interaction frequency. The highest interaction frequency value (d9, *Runx2*-P2 peak) was normalized to 1. Error bars: S.E.M.

To confirm the interactions between the *Runx2*-P1, *Supt3h*, and *Runx2*-P2 promoter regions in differentiating (d9) osteoblasts, we utilized the 3C primer that is most proximal to the *Runx2*-P2 promoter as the anchor primer to probe for interactions between the *Runx2*-P2 promoter and flanking regions. When the BglII fragment spanning the *Runx2*-P2 promoter is used as the anchor, strong interaction frequencies with both *Runx2*-P1 and *Supt3h* promoters are observed (Figure 2.10).

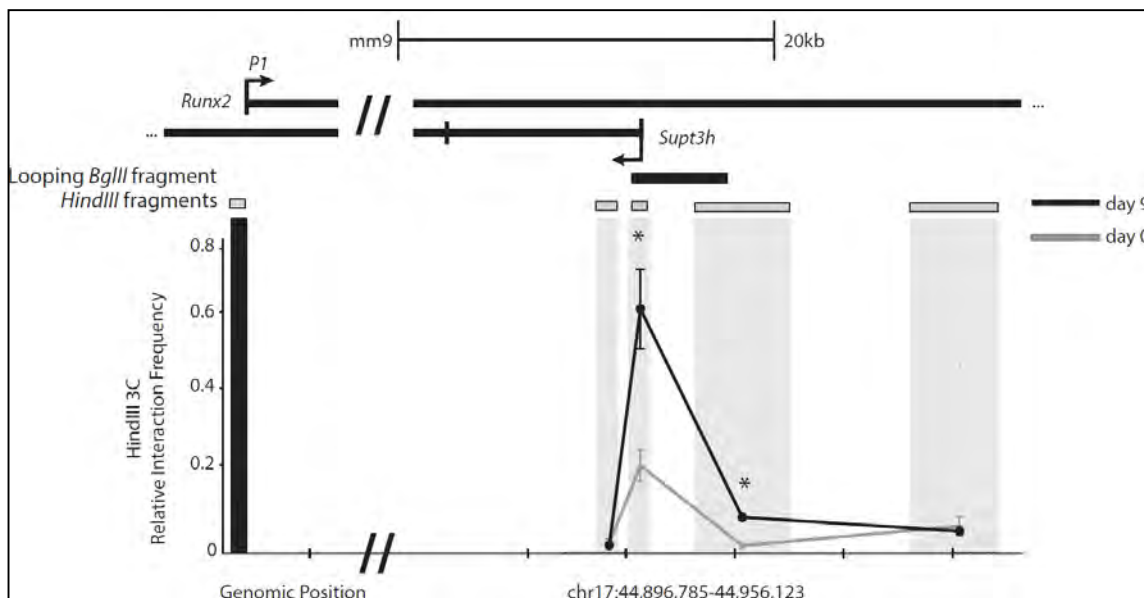


**Figure 2.10.** 3C analysis of d9 MC3T3 cultures with the anchor fragment located on *Runx2*-P2. The black arrows point at *Runx2*-P1 and *Supt3h* regions. Error bars: S.E.M.

To further validate these results, we repeated the 3C experiments with an alternative design, using the HindIII restriction enzyme instead of BglII. Interaction frequencies between the anchor *Runx2*-P1 fragment and the HindIII fragments at the *Supt3h* promoter region were analyzed in d0 and d9 MC3T3 cells. We observed that the HindIII fragment at the *Supt3h* TSS region has a significantly higher interaction frequency with the *Runx2*-P1 promoter in d9 cultures than in d0 cultures (Figure 2.11). It is also worth noting that the looping HindIII and BglII fragments (Figure 2.9) at the *Supt3h* promoter overlap with each other. HindIII fragments flanking the BglII fragment showed similar interaction frequencies in d0 and d9 cultures (Figure 2.11).

The increase in chromatin association between the *Runx2*-P1 and *Supt3h* promoters during differentiation suggests a possible regulatory relationship between these two regions, while other interaction events appear to remain constant.





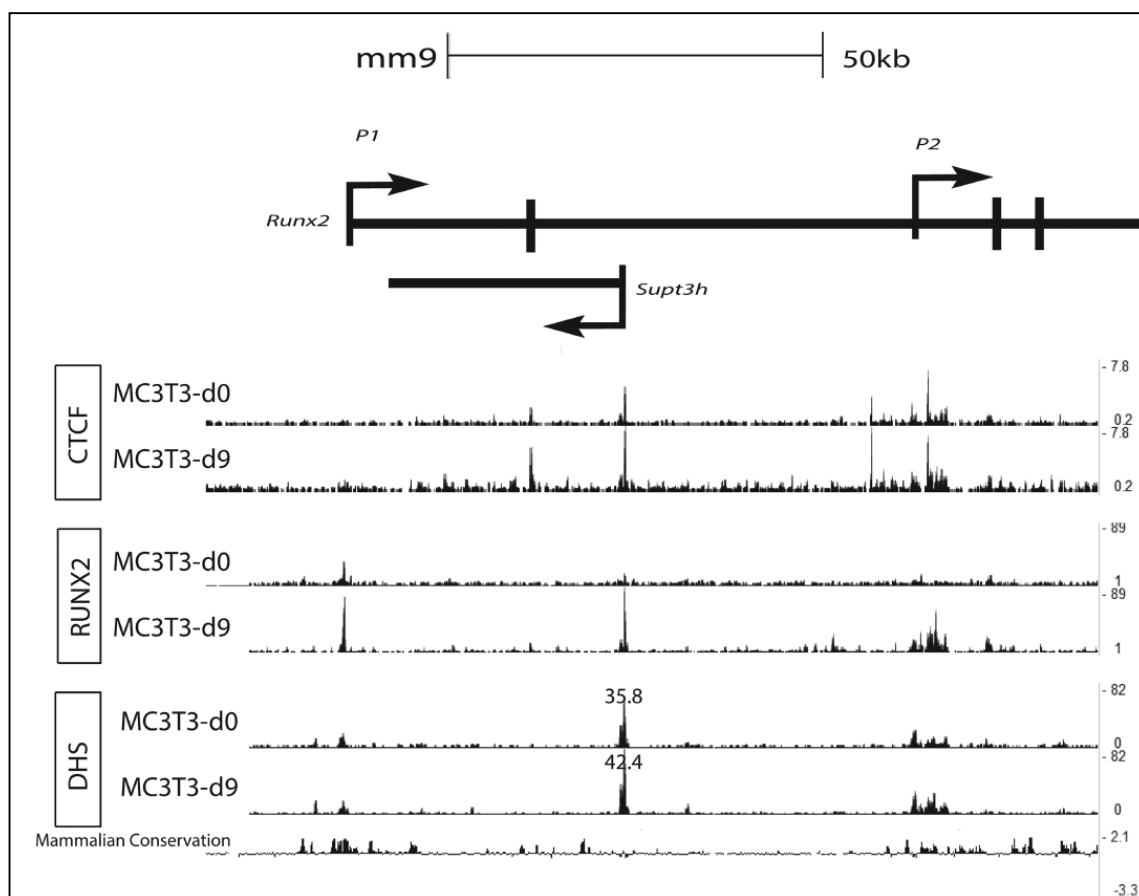
**Figure 2.11.** 3C performed with HindIII as validation of the interaction between *Runx2-P1* and *Supt3h* promoters in d0 and d9 MC3T3s. The anchor fragment is designated by a black bar at *Runx2-P1*. The looping *BglIII* fragment in Figure 2.9 is indicated by a black rectangle. Gray shades indicate the HindIII restriction fragments. Error bars: S.E.M. (\* $p < 0.05$  by Student's t-test).

### 2.3.3 Interactions between *Runx2-P1* and *Supt3h* are enriched for CTCF, RUNX2 and DHS during osteoblast differentiation

Genome-wide studies have recently shown that many developmentally regulated genes exhibiting long-range interactions are enriched for CTCF [94; 95]. Additionally, RUNX2 protein is shown to be a component of the nuclear matrix and to act as a nuclear scaffolding factor [143; 146; 147]. We asked if the local chromatin state is altered at the *Runx2-P1* and *Supt3h* promoter regions during differentiation, in which the looping intensity is significantly increased upon differentiation.

We have recently employed a comparative analysis of the genome wide enrichments of CTCF and RUNX2 via ChIP-seq during MC3T3-E1 differentiation [167]. To determine whether the interactions between the *Runx2-P1* and the *Supt3h* promoter

regions correlate with the recruitment of CTCF and/or RUNX2, we extracted the ChIP-seq data encompassing the *Runx2* locus. Additionally, to address whether there is altered nucleosome association at these regions due to a change in factor occupancy, we carried out DNase-seq experiments in differentiating MC3T3 cultures (Figure 2.12) (Tai *et al.*, manuscript in preparation).



**Figure 2.12. The epigenetic landscape of the *Runx2* locus during differentiation.** DNaseI-seq and ChIP-seq signal tracks for CTCF and RUNX2 enrichment in d0 and d9 MC3T3s. The DHS scores for the *Supt3h* promoter region are also shown.

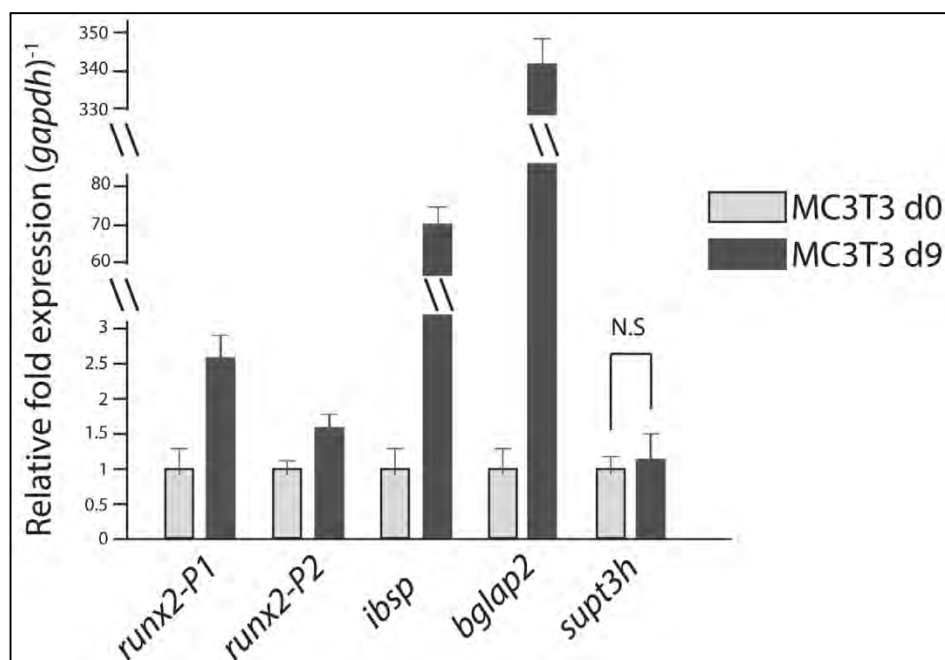
Interestingly, the DHS profile of the genomic region surrounding the *Runx2-Supt3h* gene locus shows the most pronounced peak at the *Supt3h* promoter region in both d0 and d9 cultures. ChIP-seq analysis demonstrates that CTCF, which is implicated in mediating long-range interactions, is enriched at the *Supt3h* promoter on d0, consistent with the basal level of interaction in cells that lack *Runx2-P1* activity. As pre-osteoblasts undergo osteoblast differentiation, we also observed a modest increase of CTCF enrichment at the *Supt3h* promoter (d9), coinciding with a similar increase in the DHS signal at this region (Figure 2.12). Interestingly, the timing of enrichment of CTCF at the *Supt3h* promoter overlaps with the increased looping frequency with *Runx2-P1* (Figure 2.9). Moreover, ChIP-seq data demonstrate that RUNX2 enrichment is substantially increased at the *Runx2-P1*, *Runx2-P2* and *Supt3h* promoters upon differentiation (Figure 2.12).

The increase of DHS and CTCF enrichment at the *Supt3h* promoter, together with the differentiation-dependent increased looping frequency, suggests a mechanistic link between the *Supt3h* and *Runx2-P1* promoters. The enrichment of RUNX2 observed at both *Runx2-P1* and *Supt3h* promoters is correlated with the increase in transcriptional activity of the bone-related *Runx2-P1* promoter [149], implying a regulatory role for sequences within the *Supt3h* promoter.

#### **2.3.4 *Supt3h* expression levels remain constant during osteoblast differentiation**

As the *Supt3h* promoter region undergoes alterations in chromatin conformation, DHS and factor binding profiles, we analyzed the expression of *Supt3h* throughout several time points during osteoblast differentiation. As expected [168], transcript levels of

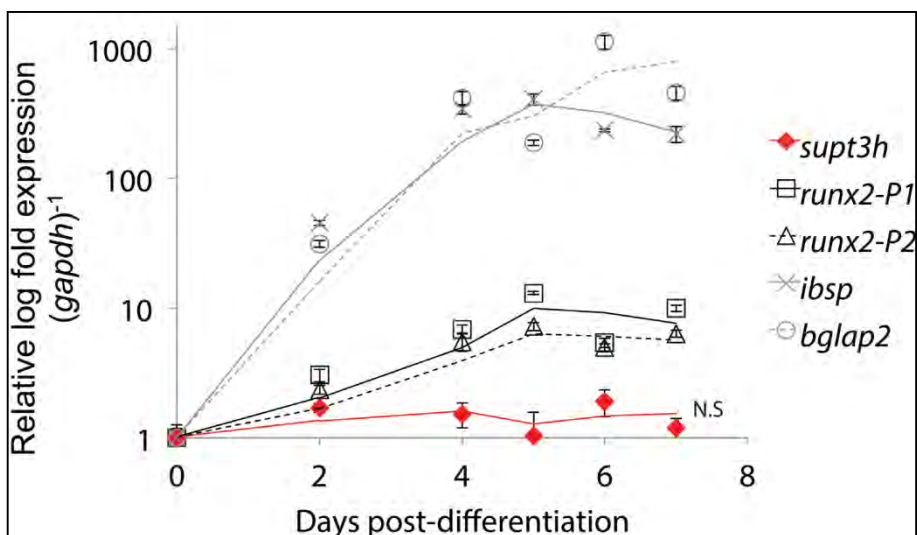
*Runx2-P1* and *Runx2-P2* increased between d0 and d9 cultures (2.5-fold and 1.6-fold, respectively) (Figure 2.13). mRNA levels of bone-sialoprotein (*Ibsp*) and osteocalcin (*Bglap2*), markers for osteoblast differentiation, were several-fold increased between the same two time-points (Figure 2.13) while the *Supt3h* mRNA levels were relatively unchanged.



**Figure 2.13. qPCR analysis of bone-specific genes during osteoblast differentiation.** Relative expression levels of *runx2-P1*, *runx2-P2*, *ibsp*, *bglap2* and *supt3h* in d0 and d9 cultures. Relative expression was normalized to 'd0' values.

In order to rule out the possibility that *Supt3h* mRNA levels might fluctuate between d0 and d9, we measured the *Supt3h* expression levels at additional time points during differentiation. Cultures harvested between d2 and d7 after initiation of differentiation showed no significant changes in *Supt3h* RNA levels (Figure 2.14). This lack of change in RNA levels is also true of cultures undergoing mineralization for 28d post-differentiation (data not shown). Therefore, despite changes in the local chromatin

architecture of its promoter (Figure 2.9 and Figure 2.12), *supt3h* expression was not changed during osteoblastic differentiation.



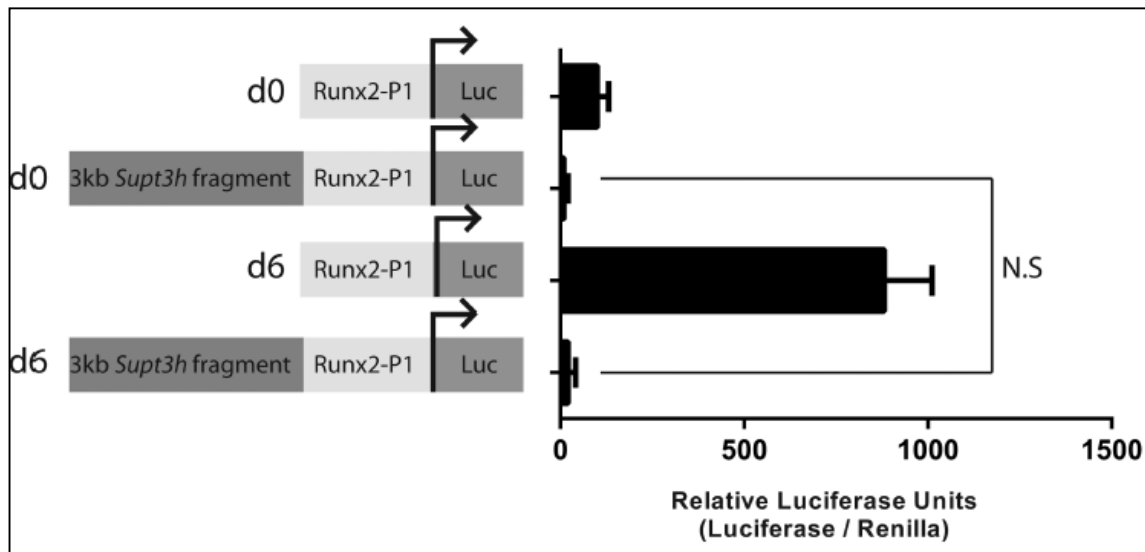
**Figure 2.14.** Time course qPCR expression analysis of bone-related genes in MCF3T3 cells. (d0, d2, d4, d5, d6 and d7).

### 2.3.5 *Runx2-P1* and *Supt3h* promoters can physically interact and regulate *Runx2-P1* expression *in-trans*

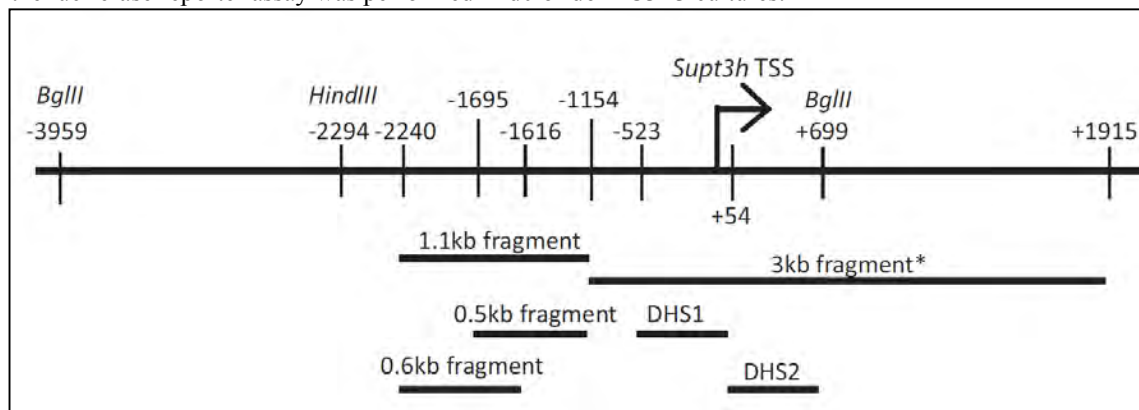
The increase of *Runx2-P1* and *Supt3h* interaction frequency as well as the enrichment of RUNX2 and presence of DHS at the *Supt3h* promoter, without affecting *Supt3h* expression, suggests that regulation of the *Runx2-P1* promoter includes chromatin alterations that do not affect *Supt3h* gene transcription.

To determine if the *Supt3h* promoter region can regulate the transcriptional activity of the *Runx2-P1* promoter, we generated a reporter construct by cloning a ~3-kb *Supt3h* fragment (-1154 to +1915 of the *Supt3h* TSS) upstream of a ~1-kb (-965 to -16) *Runx2-P1* promoter sequence that drives the luciferase reporter gene [169]. The 3kb *Supt3h*

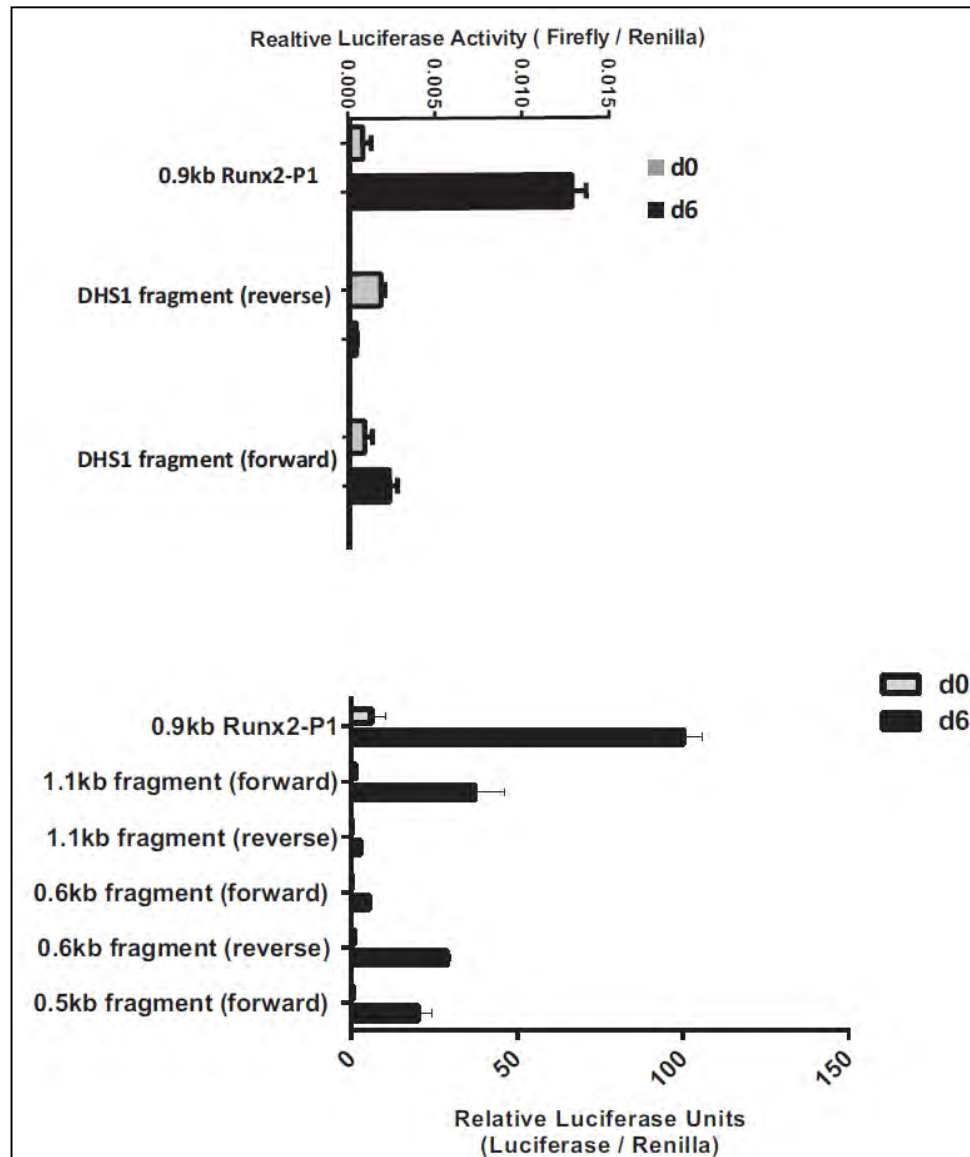
fragment partially overlaps with the looping BglII fragment (Figure 2.9). Luciferase reporter assays were conducted in undifferentiated (d0) and post-differentiated (d6) MC3T3-E1 cultures. We observe that the 3kb *Supt3h* promoter construct suppresses the transcriptional activity of the *Runx2-P1* promoter in both d0 and d6 cultures (Figure 2.15). Similar effects were observed when different fragments within the upstream 2.5kb of the *Supt3h* promoter were assayed for reporter gene expression (Figure 2.16a-b). Some of the regions tested also include the DNaseI hypersensitive site in Figure 2.12.



**Figure 2.15.** Luciferase reporter assay of the *Supt3h* construct in d0 and d6 MC3T3-E1 cells. The constructs shown were cloned upstream of the 0.9kb *Runx2-P1* promoter driving the luciferase gene, and the luciferase reporter assay was performed in d0 or d6 MC3T3 cultures.



**Figure 2.16a.** Schematic representation of the fragments used in luciferase reporter assays.

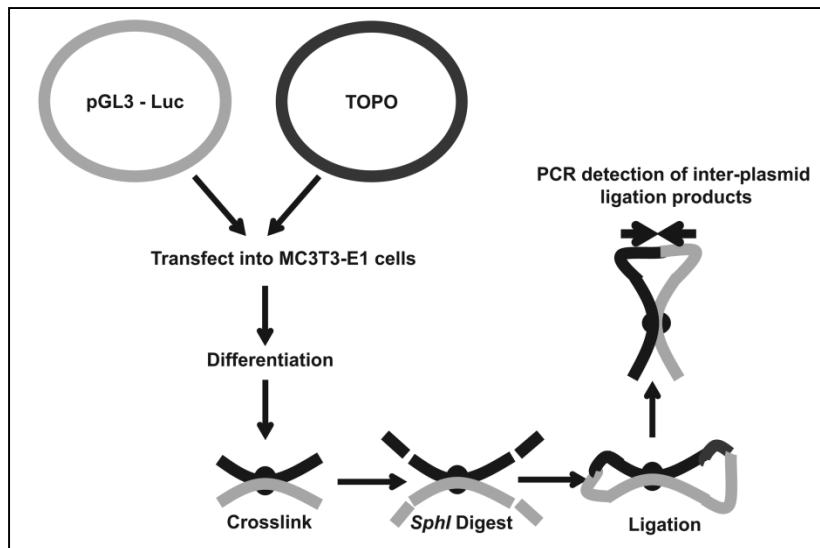


**Figure 2.16b. Luciferase experiments for different Supt3h promoter constructs.** The constructs shown were cloned upstream of the 0.9kb Runx2-P1 promoter driving the luciferase gene, and the luciferase reporter assay was performed in d0 or d6 MC3T3 cultures.

Taking into account the increase in *Runx2-P1* transcription by ~3 fold in d9 MC3T3 cultures (Figure 2.13) accompanied by the increase in looping interaction frequency (Figure 2.9), the suppressive effect on luciferase reporter activity of the *Supt3h* promoter was unexpected. We therefore reasoned that the transcriptional effect of *Runx2-P1* we observed with the different *Supt3h* promoter regions might be related to undefined spacing requirements for these regulatory sequences. When these regions are placed in tandem in the same plasmid (*in-cis*), they are in an artificial configuration removed from their endogenous context. Therefore, to better recapitulate the endogenous context, we asked whether *Runx2-P1* and *Supt3h* promoter regions could regulate *Runx2-P1* transcription while residing on different plasmid constructs via an *in-trans* association.

To test whether the *Supt3h* promoter region can physically interact with and regulate the *Runx2-P1* promoter activity *in-trans*, we employed a modified 3C-assay that we have named “interplasmid-3C” (i3C) (Figure 2.17). In the i3C method, as summarized in Figure 2.17, we used the (-965 to -16) *Runx2-P1* promoter pGL3 luciferase construct, and we also cloned the 3.3-kb *Supt3h* promoter region (-1154 to +1915) into a TOPO plasmid to generate the *Supt3h*-TOPO construct.



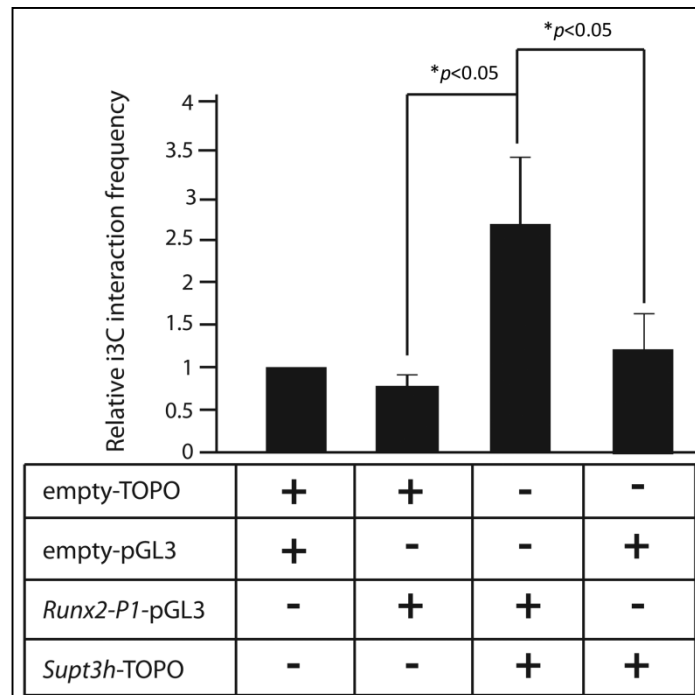


**Figure 2.17. The schematic of the interplasmid-3C assay.** After the co-transfection of pGL3 and TOPO vectors, chromosome conformation capture is performed, and the proximity of two plasmids is assessed via primers designed on the backbone of the vectors.

MC3T3-E1 cultures were co-transfected with the *Runx2-P1* pGL3 luciferase construct along with the *Supt3h*-TOPO construct to test if they associate with each other. Parallel cultures were co-transfected with the *Runx2-P1* pGL3 luciferase construct along with an empty-TOPO plasmid to serve as a negative control, and the empty pGL3 luciferase construct along with the empty-TOPO construct to serve as a normalization control for the i3C experiment. The *Supt3h*-TOPO vector combination was also added as a control. After co-transfection, cells were differentiated for 5 days and treated in the same manner as cultures subjected to the 3C methodology (see Materials and Methods). To assess the ligation frequency between the plasmids, PCR quantification was performed with i3C primers specific to either pGL3 or TOPO plasmid backbone sequences.

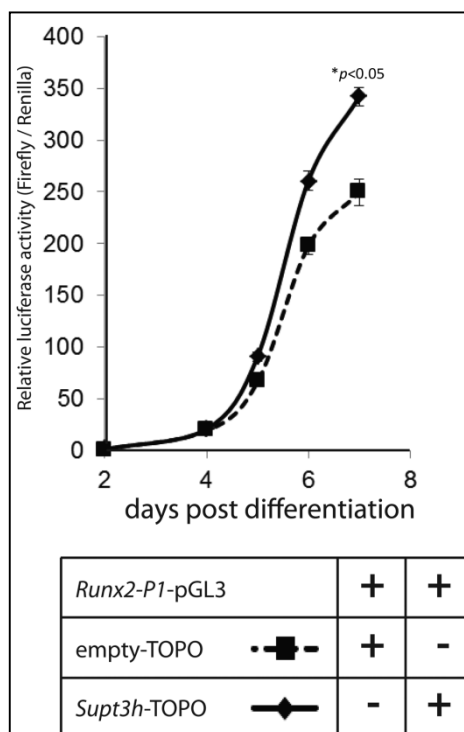
Our results reveal that when normalized to the empty pGL3 and empty-TOPO co-transfection control, d5 cultures transfected with the *Runx2-P1* pGL3 luciferase and *Supt3h*-TOPO constructs showed a ~2.5-fold increase in interaction frequency compared

to the *Runx2-P1* pGL3 and empty-TOPO constructs (Figure 2.18). In other words, pGL3 luciferase and TOPO constructs interact at a higher frequency only when the TOPO construct contains the 3.3-kb *Supt3h* region and the pGL3 constructs the *Runx2-P1* promoter region (Figure 2.18, 3<sup>rd</sup> and 4<sup>th</sup> lanes). Our i3C results demonstrate that co-transfected plasmids can physically interact. More importantly, we demonstrate that the *Supt3h* and the *Runx2-P1* promoters on separate plasmids can associate *in-trans* outside of their endogenous chromosomal context.



**Figure 2.18. i3C analysis of co-transfected plasmids.** The y-axis represents the relative ligation frequency between the plasmids. The co-transfection of plasmids was represented with a '+' below. Error bars: S.E.M.

We next tested whether the *Supt3h* promoter region can regulate the activity of the *Runx2-P1* promoter using the *in-trans* system described above. We co-transfected MC3T3-E1 cells with the *Runx2-P1* pGL3 luciferase construct together with either *Supt3h*-TOPO or empty-TOPO constructs. We then assayed for *Runx2-P1* promoter activity via luciferase reporter assay at daily intervals throughout differentiation (d2, d4, d5, d6, and d7) (Figure 2.19). Surprisingly, cultures transfected with *Supt3h*-TOPO displayed a nearly 40% increase in luciferase activity at d7 compared to cultures transfected with the empty-TOPO construct (Figure 2.19). The increase in *Runx2-P1* promoter activity after d6 suggests that the regulatory sequences within the *Supt3h* promoter, which exerted a suppressive effect *in-cis* (Figure 2.15), can positively regulate the *Runx2-P1* promoter activity *in-trans* in a differentiation dependent manner.



**Figure 2.19. In-trans luciferase assay.** Cells were transfected with either *Supt3h*-TOPO or empty-TOPO construct, and relative luciferase activity was measured at indicated time points.

## 2.4 Discussion

Recent mapping of genome-wide chromosomal interactions in both prokaryotes and eukaryotes suggests the regulatory importance of long-range associations to control gene expression [170; 171]. During development, the genome undergoes drastic structural and regulatory changes resulting in the alteration of cell identity. *Runx2* is an important regulator of bone formation and a key player in metastatic bone disease [139-143]. Perturbations to the *Runx2* gene and reduction of its transcript levels result in cleidocranial dysplasia [172]. Due to its importance in development and disease, understanding the structure and the regulation of the *Runx2* gene is relevant to many regulatory pathways.

Genetic evidence suggests that mammalian *Runx* genes acquired the utilization of two promoters (*P1* and *P2*) prior to their duplication event [173]. Therefore, if a regulatory relationship between different sequences within the *Runx* locus existed before its duplication, it is possible that this relationship may be conserved throughout evolution in paralogue *Runx* genes. In the light of these findings, the syntenic relationship between *Runx2* and *Supt3h* has prompted us to hypothesize that an architectural and regulatory relationship exists between these promoter regions. Apart from the syntenic relationship, the fact that the *Supt3h* promoter is embedded between *Runx2-P1* and *P2* promoters hints at the possibility that the *Supt3h* promoter sequence may have been co-opted by the *Runx2-P1* promoter as a cell specific enhancer. Promoters have been shown to act as enhancers for other genes [110].

During *Runx* gene duplication, although the syntenic relationship with *Supt3h* may

have been lost, the dependency of an intronic enhancer may have been retained in the *Runx1* and *Runx3* gene loci. Evidence for such conservation of cis-regulatory elements is observed in the *Runx1* gene, a paralogue of *Runx2*. *Runx1* has a similar gene structure to *Runx2*, with two isoforms transcribed from two distinct promoters. RUNX1 is required for hematopoietic cell development. Markova *et al.* reported that in human lymphoid and erythroid cell lines, there is a higher-order looping structure between the *Runx1-P1* and an intronic element ~35kb downstream of this promoter [174]. The distance and the localization of the intronic looping element in the *Runx1* locus coincide with the syntenic *Supt3h* promoter in the *Runx2* locus. This result is consistent with the fact that alternative promoter usage of *Runx* genes existed before their duplication [173], and it suggests the existence of a similar structural relationship in the *Runx2* gene locus.

When we queried the ENCODE database, we found that long-range chromatin interactions exist between *Runx2-P1* and *Supt3h* in human lung fibroblast (IMR90), breast cancer (MCF7), and leukemia (K562) cell lines (Figures 2.1-2.6). Interestingly, although the PolII ChIA-PET data in K562 and MCF7 cells suggested at a looping interaction between *Supt3h* and *Runx2* promoters, there was little to undetectable PolII ChIP-seq signal at the *Runx2-P1* promoter (Figures 2.1-2.6). A similar low-level PolII signal is also observed at *Runx2-P1* in IMR90 cells. These data indicate that a basal level of *Supt3h* – *Runx2-P1* interaction is present regardless of *Runx2-P1* expression.

Synteny results from selective evolutionary pressure. The selective pressure could be related to a requirement for looping events between the *Runx2-P1*, *Runx2-P2* and *Supt3h* promoter regions. These findings not only correlate with the evolutionary

relationship and the conservation of this synteny across many organisms, but are also consistent with the idea of a basal structural interaction between these promoters. Further evidence to support this idea comes from the 3C analysis with mouse macrophage RAW 264.7 cells (Figure 2.8) and undifferentiated d0 pre-osteoblastic cells (Figure 2.9). RAW 264.7 cells have minimal expression of *Runx2-P2*, and lack the expression of *Runx2-P1* isoforms (Figure 2.7); and d0 MC3T3 cells show low levels of *Runx2-P1* activity (Figure 2.13). However, we were able to observe an interaction between the *Supt3h* and *Runx2-P1* promoters in these cells. Moreover, we also detected a structural link between the *Runx2-P1* and *Runx2-P2* promoters (Figure 2.10), a phenomenon also observed in the *P1* and *P2* promoters of the *Runx1* gene locus [174].

Another interesting finding is that the *Supt3h* promoter region is epigenetically altered during differentiation, as indicated by the increase in DNaseI hypersensitivity and the increase of CTCF and RUNX2 enrichment (Figure 2.12). However, *Supt3h* expression levels remain unchanged throughout this process. The interaction frequency between the *Runx2-P1* and *Supt3h* promoters exhibits a striking increase during osteoblastic differentiation (Figure 2.9). When the 3C anchor is positioned on the *Runx2-P2* promoter, the data also suggest that three promoters (*Runx2-P1*, *Runx2-P2* and *Supt3h*) are in close proximity in d9 MC3T3 cultures (Figure 2.10). Taken together, our observations suggest that local changes occurring at the *Supt3h* promoter act by modulating *Runx2-P1* activity. Alternatively, the increase of CTCF enrichment may reflect the presence of an activated insulator element that flanks the *Supt3h* promoter region during differentiation. The recruitment of CTCF to the *Supt3h* promoter region

may also act to prevent the regulatory action of upstream sequences that may interfere with *Supt3h* transcription. Moreover, at d0, RUNX2 binds primarily to P2, but not to the P1 promoter. The RUNX2 protein is known to interact and co-bind to DNA with several other co-factors such as C/EBP $\beta$  [167; 175; 176]. It is possible that differential binding of co-factors to RUNX2 may change its affinity for its binding sites. Additionally, the *Runx2-P1* promoter contains binding sites for C/EBP $\beta$ , Oct1, AP-1, Runx2, Msx2/Dlx3/Dlx5, ATF, HLH/Twist, VDRE, LEF/TCF, NKX, NF-1, Sp1, and ETS. Combinatorial binding of these factors may also play roles in selectively recruiting RUNX2 to these promoters.

We also demonstrate that the interaction between the *Runx2-P1* and *Supt3h* promoters impacts the activity of the *Runx2-P1* promoter. We chose to include the -965/-16 region of the *Runx2-P1* promoter, as this region has been shown to adequately respond to differentiation conditions [169]. Also, there are 3 RUNX binding sites within the 5' UTR of the *Runx2-P1* promoter centered at +31,+39, and +49 base pairs downstream of the *Runx2-P1* transcriptional start site. These RUNX motifs have been shown to suppress *Runx2* expression as part of a negative feedback loop [163]. Luciferase reporter assays with different *Supt3h* constructs cloned *in-cis* upstream of the *Runx2-P1* promoter driving luciferase showed a reduction of *Runx2-P1* activity (Figures 2.15 and 2.16a-b). Because the *Supt3h* and *Runx2-P1* promoter regions reside more than 35kb away from each other in their endogenous context, we hypothesized that testing these regions in a “*trans*” configuration would better recapitulate the regulation occurring in the endogenous setting. We therefore measured the effect of the *Supt3h*

promoter on the *Runx2-P1* promoter while on separate plasmids. Although the ability for regulatory regions residing on separate plasmid constructs to modulate activity *in-trans* has been demonstrated previously [177; 178], the majority of *in vitro* assays that test for enhancer-promoter interactions are performed within the same DNA construct, *in cis*, where the enhancer is cloned 5' to the promoter. Functional assays aimed to validate long-range, *cis*-acting interactions are also performed in this manner. To test our hypothesis that the *Supt3h* promoter could interact with *Runx2-P1* when introduced on separate plasmids, we utilized a modified 3C protocol that we termed “interplasmid 3C” (i3C) (Figure 2.17). i3C results show that there is a ~2.5 fold higher interaction frequency between plasmids containing the *Runx2-P1* and *Supt3h* sequences than control plasmids (Figure 2.18). This result indicates that regulatory sequences of the *Supt3h* promoter need to be at a distance from the *Runx2-P1* promoter. Under these same conditions, when *Runx2-P1* promoter driven luciferase activity was assayed throughout differentiation, a significant increase is observed (Figure 2.19) when co-transfected with the *Supt3h*-TOPO construct but not the empty-TOPO construct, suggesting a differentiation-dependent activator role of the *Supt3h* promoter.

It is still not clear why background levels of structural interactions between these two promoters exist in cells which lack *Runx2-P1* transcription. However, we have shown a bone-differentiation specific regulatory function of the *Supt3h* promoter region on *Runx2-P1* promoter driven transcription. It is important to point out that in our 3C analysis; we only queried a ~+/-300kb genomic region surrounding the *Runx2* locus (Figure 2.9). Other studies have indicated that enhancers can exert their functions from



hundreds, or even thousands, of kilobases away [94]. It may be that additional distant regulatory regions which are located outside the ~600kb *Runx2* locus become associated with the *Runx2-P1* promoter region and contribute to its regulation.

Taken together, our results demonstrate a novel aspect of *Runx2* gene structure and regulation. We also demonstrate a role for *Supt3h* association with the *Runx2-P1* promoter in modulating the bone-specific activity of the *Runx2-P1* promoter. Further experiments such as deletion of the *Supt3h* promoter region will be needed to provide additional insight into the transcriptional control of the *Runx2-P1* promoter during osteogenesis.

## **CHAPTER3: Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells**

### **3.1 Introduction**

Three-dimensional genome organization is important for regulation of gene expression by bringing together distant promoter, enhancer and other cis-regulatory regions [179-181]. The development of cancer involves several genetic and epigenetic alterations that result in aberrant gene expression [182-185]. Moreover, cancer is a disease characterized by major morphological changes in the nucleus that are used as diagnostic markers [126; 186]. Even though the morphological features of cancer are well characterized, the molecular consequences of the aberrant nuclear morphology are still poorly understood.

The higher-order folding of chromatin within the nucleus involves hierarchical structures spanning different length scales [187]. Microscopic imaging shows that chromosomes are positioned within confined volumes known as chromosome territories [23]. In the nucleus, each chromosome has a preferred, but not fixed, position in which gene-dense chromosomes tend to be at the nuclear interior whereas the gene-poor chromosomes are found near the nuclear periphery [22-24; 39]. Increasing evidence highlights the importance of chromosome and gene positioning during breast cancer initiation [136; 188; 189]. Moreover, recent evidence demonstrates the influence of physical spatial proximity in the nucleus on recurrent translocations [131; 190; 191].

Several studies have revealed that chromosome territories consist of megabase-scale genomic compartments that are either euchromatic, gene-rich, highly transcribed (A-type compartments); or heterochromatic, gene-poor and silent (B-type compartments) [62; 71; 191; 192]. The open and closed compartments mostly interact with other open and closed compartments, respectively, whereas there are very few interactions between the two different types of compartments. The open (A-type) compartments preferentially and spatially cluster together in the nuclear interior, whereas the closed (B-type) compartments cluster together near the nuclear periphery [24].

Compartments are composed of 100kb to 1Mb scale topologically associated domains (TADs). TADs have been defined as clusters of interactions, in which the enhancers and promoters of co-regulated genes cross-talk with one another. The intra-TAD interactions are much more prevalent than inter-TAD interactions [24]. TADs have been shown to be largely invariant across different species, cell types and physiological conditions [73; 74] and may act as functional units for transcription regulation [75; 76; 193]. Recent work elucidated the role of TADs and transcription factor associated interactions at a genome-wide level in the context of hormonal regulation (i.e estrogen or progesterone treatment) [76; 108; 110; 194-198]. TADs are thought to facilitate transcriptional regulation by integrating the regulatory activities within the same domain [75; 187]. Within TADs, looping interactions at the 10 kb – 1 Mb scale bring together enhancers and promoters to regulate gene expression. Functional characterization of long-range interactions in breast cancer has been studied within certain candidate

regions [199-203]; or by examining the genome-wide interactions of a single locus using more unbiased approaches [45; 103; 204]. Probing chromatin structure in cancer has potential as a discovery tool for identifying candidate biomarkers [96], as the organization of chromatin is often perturbed at different hierarchical levels in cancer [205]. Despite the number of previous studies, differences in genome-wide chromatin structure between normal epithelial cells and tumorigenic breast cancer cells remain unknown.

In this study, in order to characterize different scales of genome organization during breast cancer development, we performed genome-wide chromosome conformation capture (Hi-C) analyses in MCF-10A mammary epithelial and MCF-7 tumorigenic breast cancer cells. Hi-C is a powerful molecular tool to probe genome-wide chromatin interactions in an unbiased way [67]. Our results uncovered fundamental differences of chromatin organization at different genomic scales between two commonly used mammary epithelial and tumorigenic breast cancer cell lines. This work provides an important foundation for understanding the relationship between the alterations in chromatin organization and gene expression in breast cancer.

## **3.2 Materials and Methods**

### **3.2.1 Cell Culture**

MCF-10A cells were obtained from the Barbara Ann Karmanos Cancer Institute (Detroit, MI). The cells were maintained in monolayer in Dulbecco's modified Eagle's medium-F12 (DMEM/F12) (Invitrogen, 21041025) supplemented with 5% horse serum

(Invitrogen, 16050122), 1% penicillin/streptomycin (Invitrogen, 15140122), 0.5 µg/ml hydrocortisone (Sigma, H-0888), 100 ng/ml cholera toxin (Sigma, C-8052), 10 µg/ml insulin (Sigma, I-1882), and 20 ng/ml recombinant human EGF (Peprotech, 100-15) as previously described [206]. MCF-7 cells were obtained from ATCC and were cultured in DMEM supplemented with 10% FBS and pen-strep.

### **3.2.2 RNA-seq and Analysis**

The RNA-seq libraries were generated with TruSeq Stranded Total RNA with Ribo-Zero Gold Kit and the samples were sequenced 100bp single-end using a Hi-Seq 2000 instrument. To do RNA-Seq analysis, first, the adapter sequences were removed from the RNA-seq reads. Ribosomal RNA reads, if any, were filtered out using Bowtie [207]. After quality filtering and adapter removal steps, the reads were aligned to a transcriptome and quantified using RSEM v1.2.7 [208]. The annotation file was downloaded from University of California, Santa Cruz (UCSC) genome browser, human hg19 assembly. To quantify the gene expression, the gene counts and transcripts per million (TPM) values were calculated by using the RSEM tool. Differential gene expression was calculated by using the Deseq2 version 1.4.5 package in R 3.1.0 using the mean value of gene-wise dispersion estimates [209]. To find significant differentially expressed genes, we used 0.01 for p-adjusted value and >1 log<sub>2</sub> fold change. Gene Ontology analysis was performed with the FuncAssociate software [210]. The RNA-seq plots were confirmed using the ngs.plot software [211].

### 3.2.3 Preparation of Hi-C Libraries

Hi-C was performed as previously described with minor modifications [99]. The modified part of the protocol was in the biotin incorporation step, where the mixture was incubated at 37°C for 40 minutes with continuous shaking and tapping the tube every 10 minutes. The MCF-10A and MCF-7 Hi-C samples displayed a range of 40% to 85% biotin incorporation efficiency. At the end of Hi-C sample preparation, the libraries were sequenced using PE100 read with a Hi-Seq 2000 instrument.

### 3.2.4 Read Mapping / Binning / ICE correction

Table 3.1 summarizes the mapping results and different classes of reads and interactions observed [100]. The data were binned at 6.5Mb, 1Mb, 250kb, 100kb and 40kb non-overlapping genomic intervals. In our Hi-C analyses of the near diploid MCF-10A and aneuploidy MCF-7 cells, we utilized the iterative correction and eigenvector decomposition (ICE) method [212], which corrects for differences in copy number. A tetraploid chromosome may have twice as many sequenced interactions as a diploid chromosome, but the ICE method divides its final interaction counts by the total sum of all interactions and thus normalizes this difference. Iterative mapping and correction of Hi-C data were performed as previously described [101]. Biological replicates showed high reproducibility (Pearson's correlation coefficient  $> 0.9$  for 1Mb resolution data). Similarly, the first eigenvector comparison of the replicates showed high reproducibility. For the downstream analyses, sequences obtained from both biological replicates were pooled and ICE-corrected to serve as a combined dataset.

### **3.2.5 Z-score Calculation**

We modeled the overall Hi-C decay with distance using a modified LOWESS method (alpha = 1%, IQR filter), as described previously [94]. LOWESS calculates the weighted-average and weighted-standard deviation for every genomic distance and therefore normalizes for genomic distance signal bias.

### **3.2.6 Calculation of Differential Interactions**

To capture the differences between MCF-10A and MCF-7 interactions, we first transformed the 6.5Mb Hi-C data into Z-score matrices for all 4 replicate datasets (MCF-7-R1, MCF-7-R2, MCF-10A-R1, and MCF-10A-R2). For each interaction, the mean sample:sample (between samples) Z-score difference was calculated from all pairwise combinations of the four datasets (MCF-7-R1 – MCF-10A-R1, MCF-7-R1 – MCF-10A-R2, MCF-7-R2 – MCF-10A-R1, MCF-7-R2 – MCF-10A-R2). The replicate:replicate Z-score difference (within samples) was also calculated for a random set of 500,000 interactions. These random replicate-replicate Z-score differences were then used to build an expected distribution of Z-score differences. The resulting Z-score difference matrix was then derived by calculating for each bin the ratio of the mean of the set of 4 possible sample:sample Z-score differences minus the genome-wide mean of the replicate:replicate Z-score difference, divided by the genome-wide standard error of the replicate:replicate Z-score differences.

### **3.2.7 Compartment Profiles**

First, the z-scores of the interaction matrices at 250kb resolution were generated as described previously [191]. Then, Pearson Correlation on the Z-score matrices was

calculated. In performing principal component analysis [71; 191], the first principle component usually detects the patterns of increased and decreased interaction across the genome that appear as a “plaid pattern” in the heatmap. Each genomic region tends to match this prominent interaction pattern (positive eigenvector value) or its opposite (negative eigenvector value) and these represent the two spatially segregated compartments. In any given analysis, though, the generally open, gene rich “A-type” compartment may end up with either a positive or negative eigenvector. To detect which compartment is the open “A-type” and which is the closed “B-type”, the genome wide gene density was calculated to assign the “A-type” and “B-type” compartmentalization.

### **3.2.8 Identification of TAD Boundaries (Insulation Square Analysis)**

TAD calling was performed as calculating the insulation score of each bin using the 40kb resolution combined Hi-C data [213]. The mean of the interactions across each bin were calculated. By sliding a 1Mb x 1Mb (25bins x 25bins) square along the diagonal of the interaction matrix for every chromosome, we obtained the insulation score of the interaction matrix. Valleys in the insulation score indicated the depletion of Hi-C interactions occurring across a bin. These 40kb valleys represent the TAD boundaries. Based on the variation of boundaries between replicates, we chose to add a total of 160kb (80kb to each side) to the boundary to account for replicate variation. The final boundaries span a 200kb region. All boundaries with a boundary strength  $< 0.15$  were excluded as they were considered weak and non-reproducible. The insulation plots for the biological replicates showed high reproducibility (Pearson correlation coefficient =



0.80 for MCF-7 and 0.90 for MCF-10A replicates), suggesting the robustness of the method. Similarly, the overlap of detected boundaries also showed high reproducibility between the biological replicates (~85% TAD boundary overlap for MCF-7 and ~91% for MCF-10A). Therefore, we used the combined Hi-C replicates for the TAD analyses.

### **3.2.9 Identification of TAD Domains**

The cell line specific TAD domains were identified by using the bedtools suite [214]. First the boundaries on all chromosomes for both MCF-10A and MCF-7 were merged. The boundaries that overlapped were categorized as “all overlapping TAD boundaries”. Then, the regions outside of the boundaries were extracted using the “complementBed” function. The telomere/centromere regions were filtered by using the “intersectBed -v” option. The resulting regions constituted the “all overlapping TAD domains”. Next, the TAD boundaries identified in MCF-10A and MCF-7 datasets were independently subtracted (by using the subtractBed function) from the “all overlapping TAD boundaries”. Within these two independently subtracted datasets, the TAD domains that have at least 90% overlap ( $-f 0.90 -r$ ) were considered as “overlapping TAD domains”, TAD domains that were found only in MCF-7 were categorized as “MCF-7 specific TAD domains”, and the domains that were only found in MCF-10A subtracted datasets was categorized as “MCF-10A specific TAD domains”.

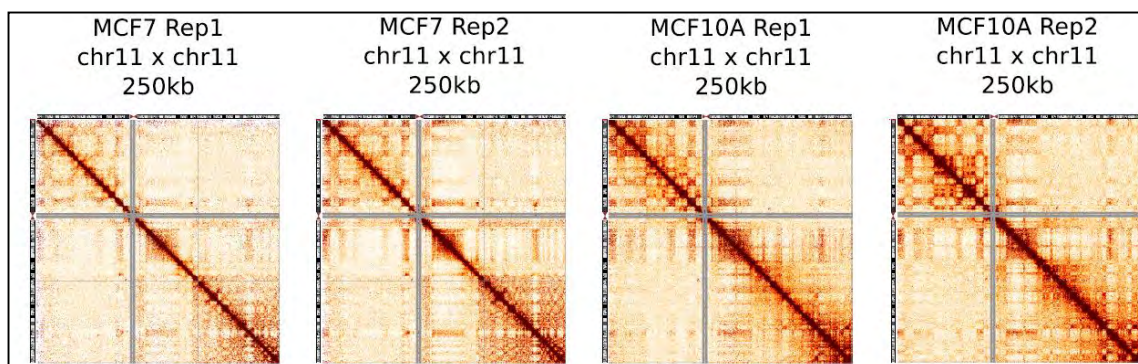
### **3.3 Results**

#### **3.3.1 Small, Gene-Rich Chromosomes Interact Less Frequently in the MCF-7 Breast Cancer Genome**

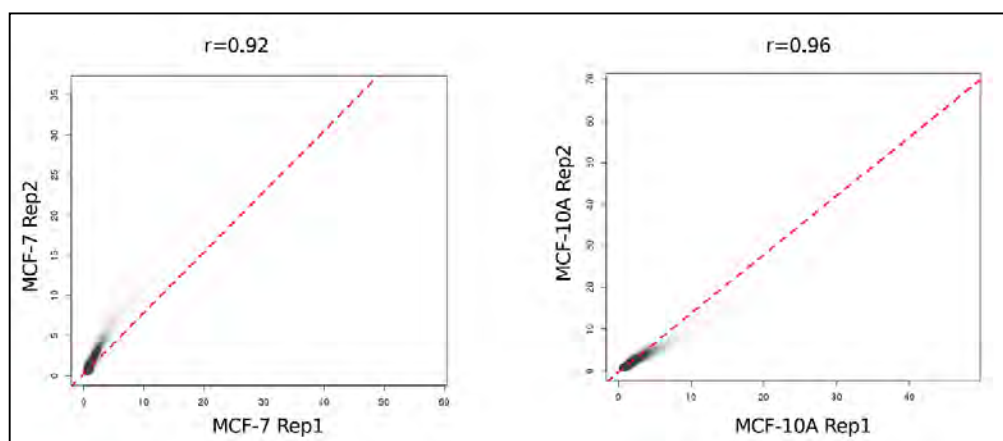
In order to probe the genome-wide chromatin structure of mammary epithelial and breast cancer cells, we generated Hi-C libraries from two independent biological replicates for the MCF-10A and MCF-7 cell lines. After sequence filtering [100], a total of ~152 and ~143 million interactions were obtained from the MCF-10A and MCF-7 combined replicate Hi-C libraries, respectively (Table 3.1), with high reproducibility between the biological replicates (Figure 3.1a-b). For the initial Hi-C analyses, we used the Iterative Correction Method (ICE) [212] to correct for systematic biases, including copy number differences.

Mapping Statistics	MCF-10A-Rep1		MCF-10A-Rep2		MCF-7-Rep1		MCF7-Rep2	
side1TotalReads	159334130		130640941		156841886		129444410	
side1NoMap	33272095	(20.88%)	19171967	(14.68%)	35140925	(22.41%)	19465398	(15.04%)
side1MultiMap	3830572	(2.40%)	4112221	(3.15%)	3957593	(2.52%)	3749200	(2.90%)
side1UniqueMap	122231463	(76.71%)	107356753	(82.18%)	117743368	(75.07%)	106229812	(82.07%)
side2TotalReads	159334130		130640941		156841886		129444410	
side2NoMap	35079522	(22.02%)	20653438	(15.81%)	36420678	(23.22%)	20566318	(15.89%)
side2MultiMap	3708345	(2.33%)	3967826	(3.04%)	3873238	(2.47%)	3660077	(2.83%)
side2UniqueMap	120546263	(75.66%)	106019677	(81.15%)	116547970	(74.31%)	105218015	(81.28%)
<b>Hi-C Library Quality Metrics</b>								
totalReads	159334130		130640941		156841886		129444410	
unMapped	5317280	(3.34%)	3862425	(2.96%)	9928451	(6.33%)	3676382	(2.84%)
singleSided	65255974	(40.96%)	40180602	(30.76%)	59535532	(37.96%)	40088229	(30.97%)
bothSideMapped	88760876	(55.71%)	86597914	(66.29%)	87377903	(55.71%)	85679799	(66.19%)
sameFragment	6375569	(7.18%)	8797783	(10.16%)	18357109	(21.01%)	10456595	(12.20%)
selfCircle	4263783	(4.80%)	7128600	(8.23%)	10690007	(12.23%)	7864717	(9.18%)
danglingEnd	2085755	(2.35%)	1522966	(1.76%)	7650191	(8.76%)	2562991	(2.99%)
bounded	23114	(1.11%)	11650	(0.76%)	9188	(0.12%)	6710	(0.26%)
internal	2062641	(98.89%)	1511316	(99.24%)	7641003	(99.88%)	2556281	(99.74%)
error	26031	(0.03%)	146217	(0.17%)	16911	(0.02%)	28887	(0.03%)
validPair	82385307	(92.82%)	77800131	(89.84%)	69020794	(78.99%)	75223204	(87.80%)
cis	16291746	(19.78%)	19266682	(24.76%)	12377949	(17.93%)	12489624	(16.60%)
trans	66093561	(80.22%)	58533449	(75.24%)	56642845	(82.07%)	62733580	(83.40%)
<b>Hi-C Library Redundancy Metrics</b>								
validPair	82385307	(100%)	77800131	(100%)	69020794	(100%)	75223204	(100%)
totalMolecules	82385307	(100%)	77800131	(100%)	69020794	(100%)	75223204	(100%)
redundantInteractions	11540337	(14.01%)	1234095	(1.59%)	972494	(1.41%)	806246	(1.07%)
nonRedundantInteractions	70844970	(85.99%)	76566036	(98.41%)	68048300	(98.59%)	74416958	(98.93%)
percentRedundant	14.01%		1.59%		1.41%		1.07%	

**Table 3.1. Hi-C statistics of the MCF-7 and MCF-10A replicates.** The sequencing, mapping and filtering statistics for the Hi-C replicates analyzed as previously described [100].

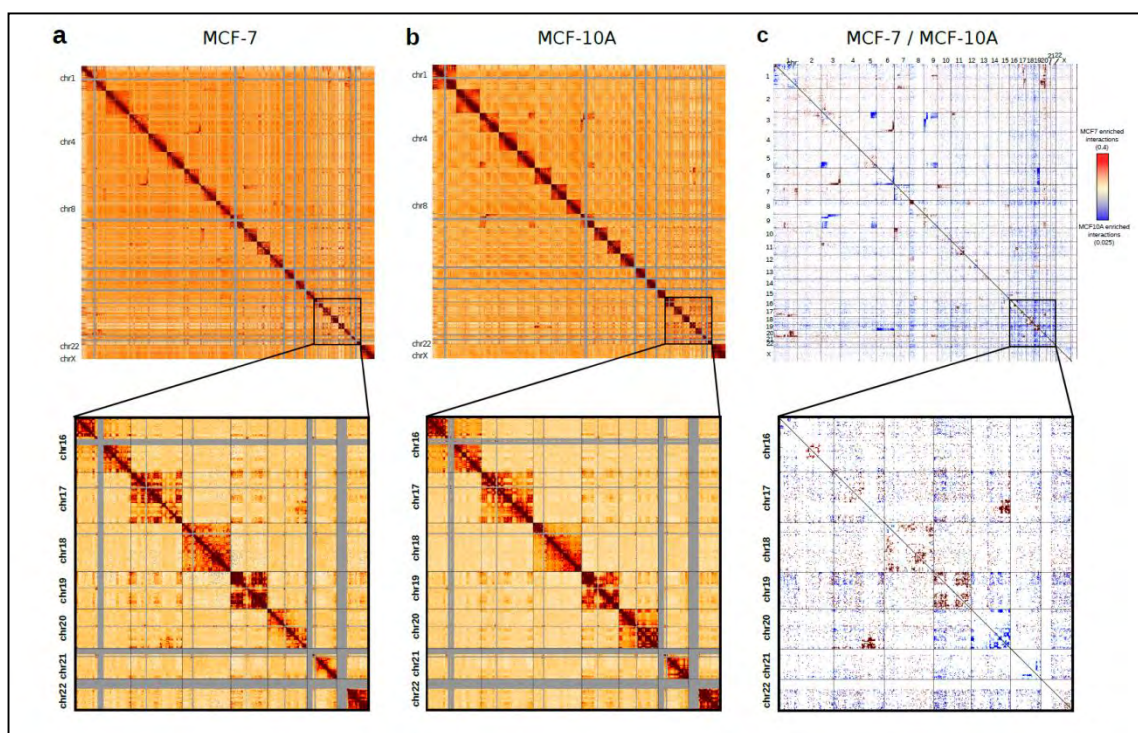


**Figure 3.1a. MCF-7 and MCF-10a Hi-C replicates are highly reproducible.** Heatmaps of 250kb interaction matrices of chr11 for each biological replicate.

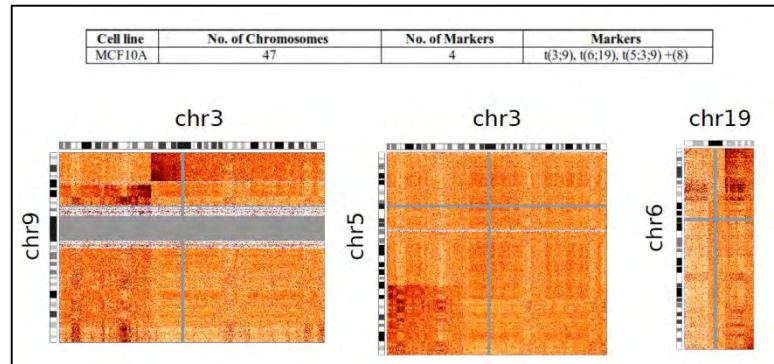


**Figure 3.1b. Interactions in MCF-7 and MCF-10a Hi-C replicates are highly reproducible.** Pearson correlation coefficient of genome-wide interactions of the biological replicates for each cell line.

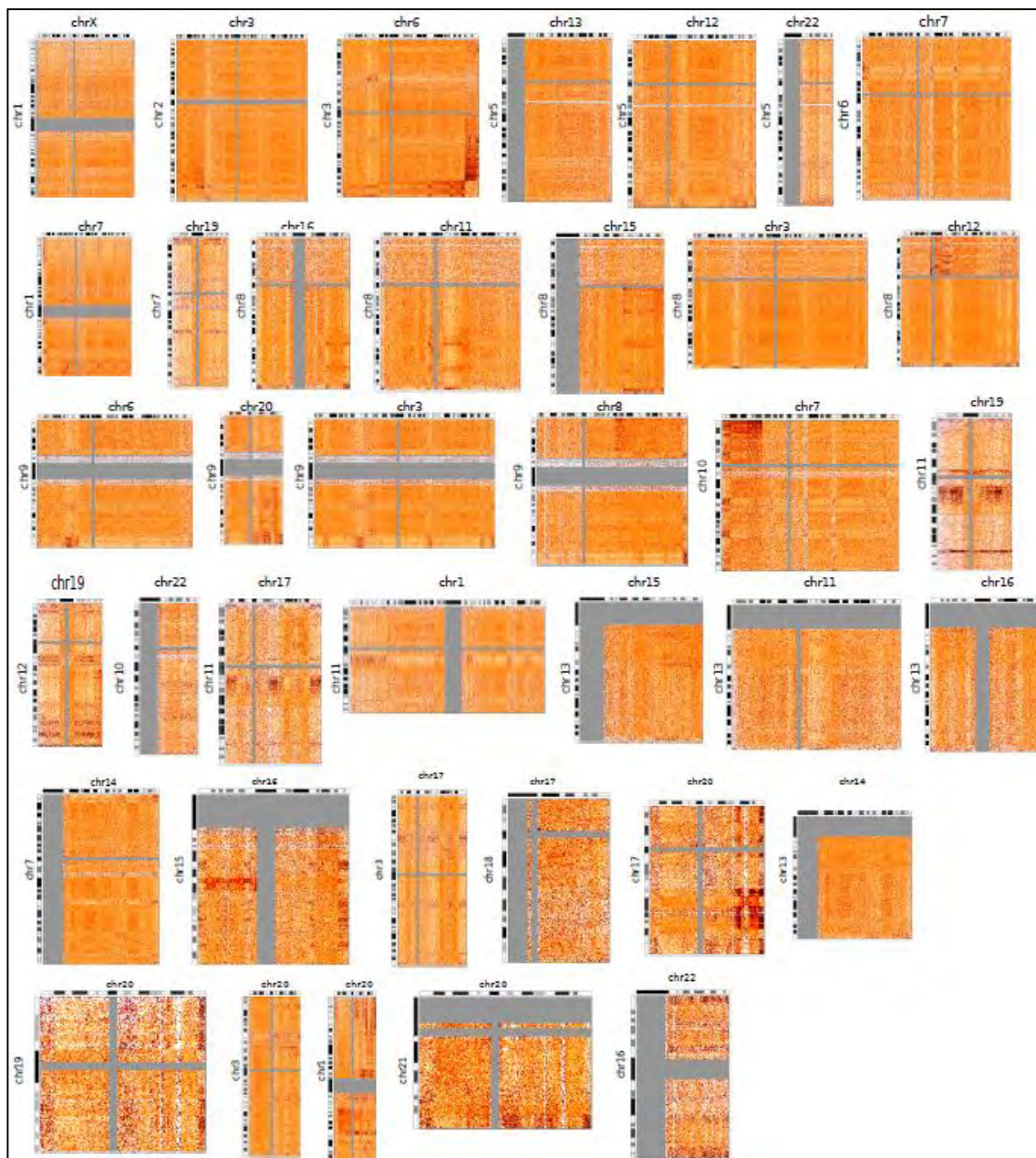
Genome-wide interaction data were visualized as chromosome versus chromosome heat maps, where darker colors represent more frequent interactions events (Figure 3.2). The heat maps revealed two aspects of large scale genome organization in the MCF-10A and MCF-7 cells. First, consistent with the notion of chromosome territories [23], *intra*-chromosomal interactions (visualized as darker boxes along the diagonal) were much more frequent than *inter*-chromosomal interactions (Figure 3.2). Second, we observed a number of large blocks of *inter*-chromosomal interactions representing the translocation events in these cell lines. Comparing the translocated regions in the Hi-C data with previously published MCF-10A and MCF-7 spectral karyotyping (SKY) and multiplex fluorescence in situ hybridization (M-FISH) data [44; 215], we observed that the majority of the translocated regions identified by SKY/M-FISH are also identified by Hi-C (Figure 3.3 and Figure 3.4).



**Figure 3.2. Genome-wide all by all 1Mb Hi-C interaction heatmaps of a) MCF-10A and b) MCF-7 cells.** The chromosomes in all by all heatmaps are stacked from top-left to bottom-right in order (chr1, chr2...chr22 and chrX). The gray regions indicate repetitive regions (such as centromeres) in which the sequencing reads could not be mapped. The *intra*-chromosomal interactions were much more frequent than *inter*-chromosomal interactions. The blocks of enriched *inter*-chromosomal interactions represent the translocated regions. On the lower panel, an enlargement of the *cis*- and *trans*-interactions for chr16 through chr22 is shown. **c) Genome-wide heatmap of significant differential interactions between MCF-10A and MCF-7.** Each dot denotes a genomic region of 6.5Mb. Chromosomes are stacked from top-left to bottom-right from chr1 through chr22 and chrX. The red color indicates MCF-7 enriched and the blue color indicates MCF-10A enriched interactions, respectively. The white regions denote interacting regions that are not significantly changed between the cell lines. On the lower panel, significant interactions within and in between chr16-22 are shown.

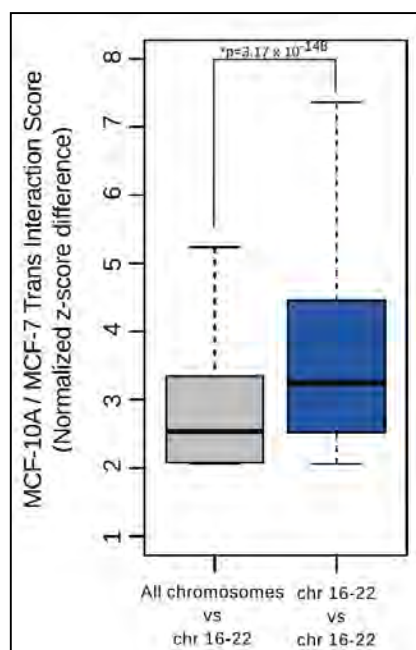


**Figure 3.3.** Inter-chromosomal heatmaps showing the translocations in MCF-10A detected by a previous SKY-FISH study [44].



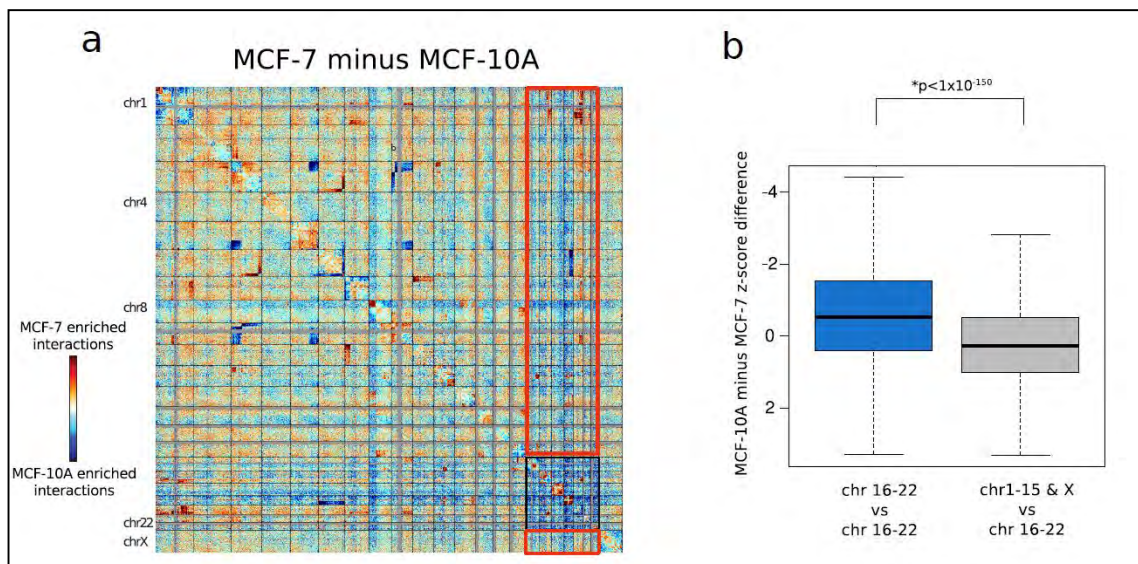
**Figure 3.4.** Inter-chromosomal heatmaps showing the translocations in MCF-7 detected by a previous SKY-FISH study [215].

In order to assess whether the clustering of chromosomes is altered between MCF-10A and MCF-7 cells, we compared the genome-wide interaction differences (see Methods, and Figure 3.2c). Strikingly, we observed a strong physical proximity of gene-rich, small chromosomes (chr16 to chr22) in MCF-10A compared to MCF-7 (Figure 3.2a-c, lower panels). This interaction network of small chromosomes also included the p-arm of chr8 (Figure 3.2c). Quantification of the *inter*-chromosomal interactions between chr16-22, and between chr16-22 and the rest of the genome revealed that there is a significant increase of *inter*-chromosomal associations between chr16-22 in the MCF-10A genome (Figure 3.5). The same result was also observed when, as an alternative approach, a direct subtraction of the MCF-10A and MCF-7 interaction matrices was performed (Figure 3.6).



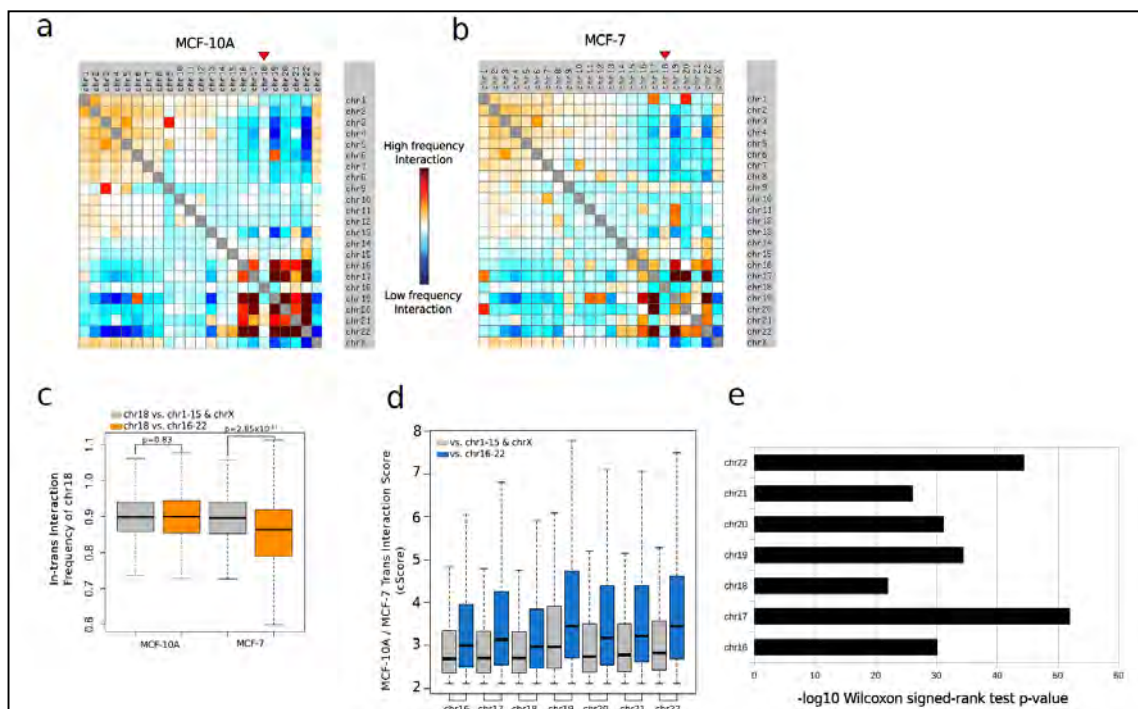
**Figure 3.5. Small chromosomes interact with each other more frequently in MCF-10A cells.** Boxplot showing the MCF-10A/MCF-7 *inter*-chromosomal interaction frequency differences between chr16 through chr22 and all the other chromosomes (grey); or in between chr16 through chr22 (blue). p-value: Wilcoxon rank-sum test





**Figure 3.6. Small chromosomes interact more frequently in the MCF-10A genome.** **a)** Subtraction of the MCF-10A z-score matrix from that of MCF-7 at 2.5Mb resolution. The blue and the red squares represent interactions enriched in the MCF-10A and MCF-7 genomes, respectively. The black rectangle denotes the interactions among the small chromosomes (chr16-22) and the red rectangles show the interactions between the small chromosomes and chr1-15 and chrX. **b)** Box plot showing the quantification of interactions among chr16-22 (blue), or with with the rest of the genome (grey). P-value: Wilcoxon rank-sum test.

Moreover, the larger chromosomes (chr1-15 & X) in the MCF-10A genome showed similar levels of differential interaction frequency with other large chromosomes or chr16-22. Consistent with this observation, the positioning of chr18 with other small chromosomes was not prevalent in the raw Hi-C interaction matrices (Figure 3.7a-c). However, the relative (MCF-10A / MCF-7) interaction frequency of chr18 with other small chromosomes was significantly increased in the MCF-10A cells (Figure 3.7d-e), which suggests that all of the small chromosomes in MCF-10A cells show increased proximity to each other compared to the relative proximity in the MCF-7 cancer cell line.

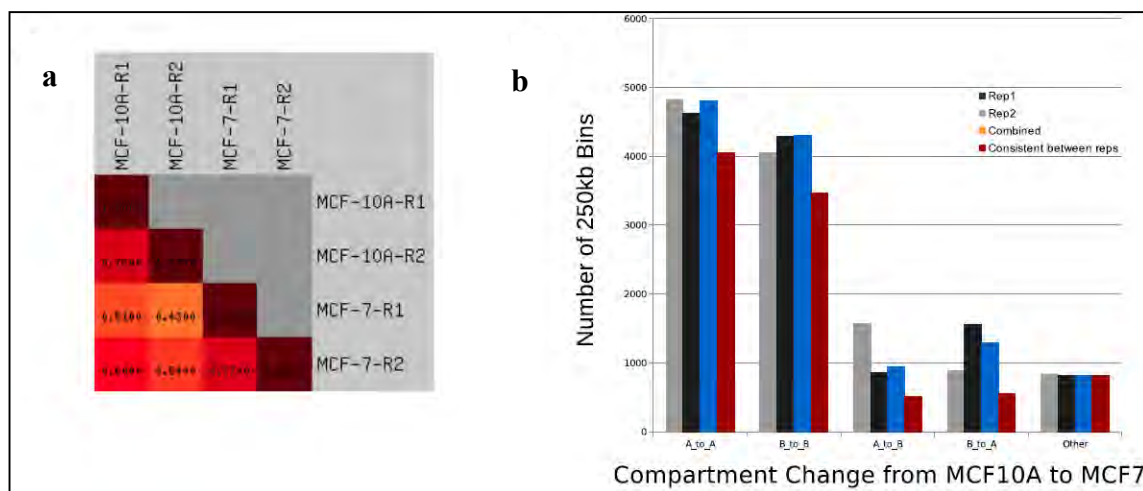


**Figure 3.7. The positioning of chr18 is not altered in MCF-10A cells.** All by all raw interaction matrix for a) MCF-10A and b) MCF-7 genomes. The red arrow denotes chr18. c) Interaction frequency of chr18 either with large chromosomes (grey), or the small chromosomes (orange) in MCF-10A and MCF-7 cells. The MCF-7 chr18 displays less interaction frequency with other small chromosomes compared to MCF-10A chr18. d) Box plots showing the relative interaction frequency (MCF-10A / MCF-7) of each small chromosome. e) p-values of the differences between small and large chromosome clustering of each small chromosome in Panel D, assessed by Wilcoxon rank sum test.

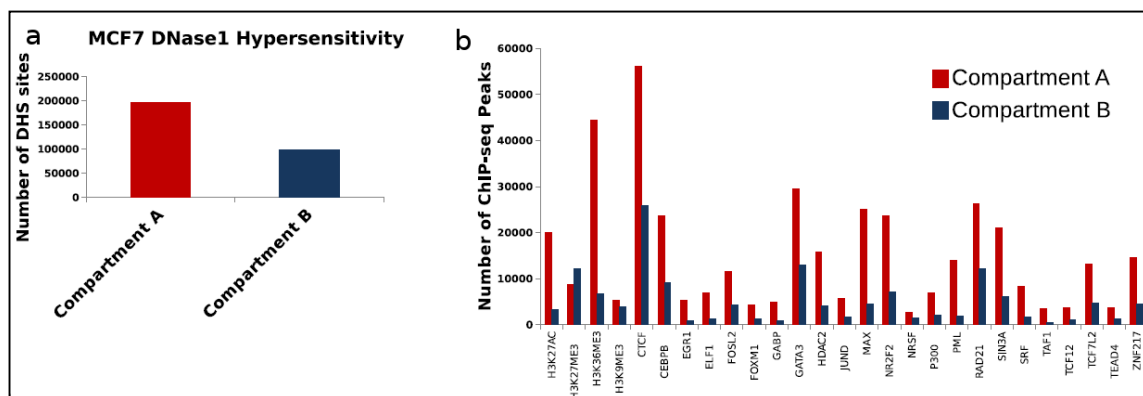
### 3.3.2 Decreased interaction frequency between small chromosomes in MCF-7 cells coincides with increased open chromatin compartmentalization

Previous evidence [71] has shown there are two unique patterns of interactions in the genome, representing the open (A-type) and closed (B-type) genomic compartments. We identified the two patterns of compartmentalization in both genomes with high reproducibility among the biological replicates (see Methods) (Figure 3.8).

Associating the MCF-7 ENCODE ChIP-seq datasets with the genomic compartments revealed the known features of genomic compartmentalization, including increased DNase1 hypersensitivity, and higher levels of transcription factor binding in open (A-type) compartments in the MCF-7 genome (Figure 3.9).

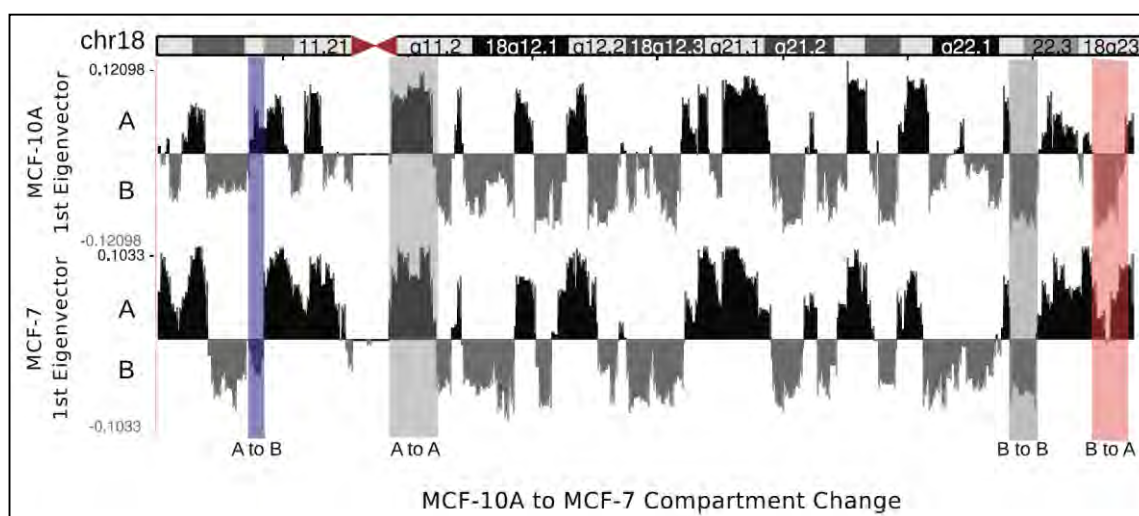


**Figure 3.8. Replicate analysis of genomic compartments in MCF-10A and MCF-7 cells. a)** Pearson correlations of the 1<sup>st</sup> eigen values for each Hi-C biological replicate. **b)** Bar graph showing the number of reproducible compartments between the replicates. “Other” indicated repetitive regions that are masked. The compartments that are consistent between the biological replicates were used for downstream analysis.

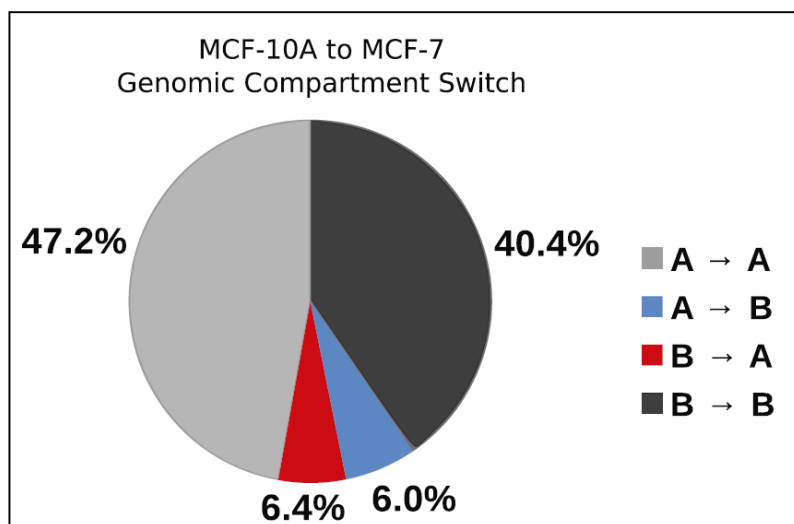


**Figure 3.9. Association of MCF-7 ENCODE datasets with genomic compartments. a)** Number of DNase1 hypersensitive sites in MCF-7 cells within open or closed compartments. **b)** Number of MCF-7 ENCODE transcription factor ChIP-seq peaks per compartment type.

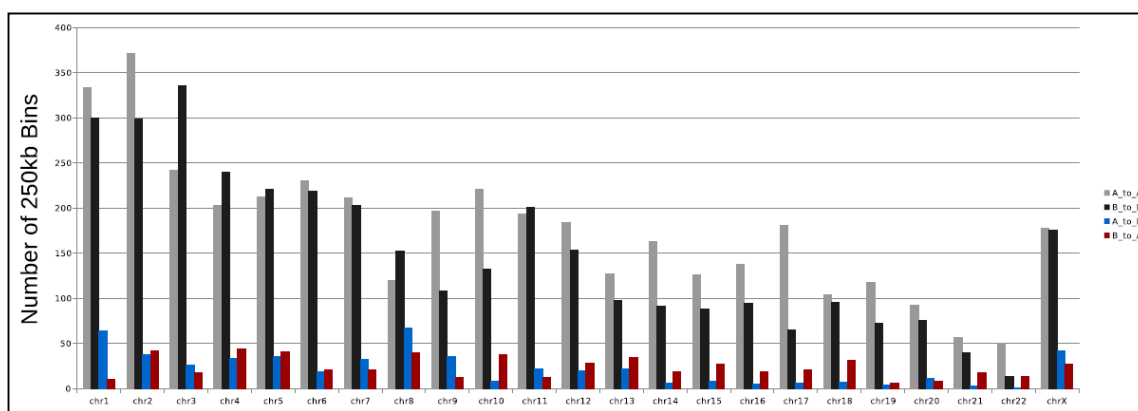
To determine whether there are any differences in the compartmentalization between the MCF-10A and MCF-7 genomes, we compared the compartments throughout the genome at 250kb resolution. The MCF-10A and MCF-7 genomes displayed similar distribution of open and closed compartments, with certain regions showing a change in genomic compartmentalization from A-type to B-type and vice versa (Figure 3.10 and Figure 3.11). The majority of compartments were the same in both cell lines, where 47% of all compartments constituted the A-type compartments and 40% constituted the B-type compartments (Figure 3.11). Importantly, 12% of all compartments in the MCF-10A genome transitioned to the opposite compartment (A-type to B-type and vice versa) in MCF-7 cells (Figure 3.11). Compartment switching was homogeneous throughout the chromosomes, rather than in a few hot spots (Figure 3.12).



**Figure 3.10. Genomic compartments are altered between MCF-10A and MCF-7 genomes.** First principal component of chr18, representing the open A-type (black) and closed B-type (grey) compartmentalization. Highlighted bars represent examples of regions with either stable or differential compartmentalization. The differential compartments are defined as genomic regions in which one type of compartmentalization is observed in one cell line, and the other compartment type in the second cell line.

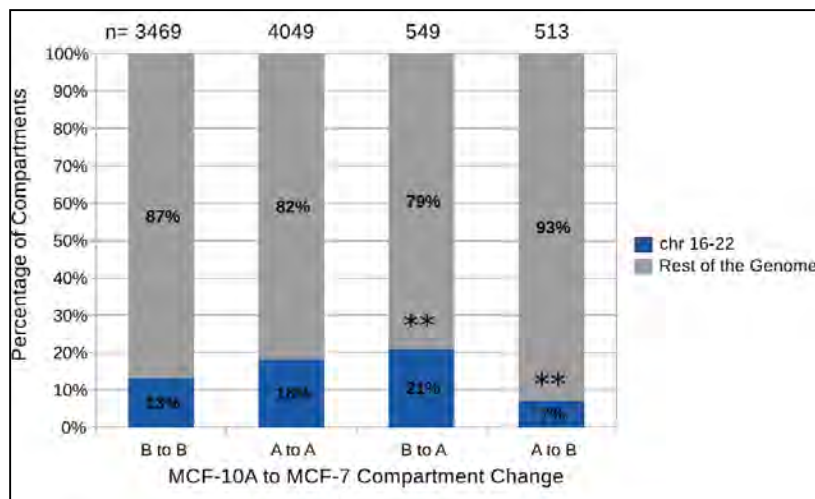


**Figure 3.11. A subset of genomic compartments display switching between MCF-10A and MCF-7 genomes.** Pie chart showing the genomic compartment changes between MCF-10A and MCF-7 genomes. “A” and “B” denotes the open and closed compartments, respectively. “A → A” represents compartments that are open in both cell lines, “B → B” represents compartments that are closed in both cell lines, “A → B” denotes compartments that are open in MCF-10A but closed in MCF-7, and “B → A” denotes compartments that are closed in MCF-10A and open in MCF-7.



**Figure 3.12. Compartment changes are distributed homogeneously throughout the genome.** Bar graph displaying the compartmental changes for each chromosomes. Compartmental alterations are distributed homogeneously throughout all the chromosomes, rather than localized to a few hotspots.

To understand if the inter-chromosomal interaction changes we observed between small chromosomes were related to any compartment change, we asked whether there was an enrichment in transition of genomic compartments on small chromosomes (chr16-22). We found a significant enrichment of genomic regions on chr16-22 that switched to the A-type compartment in MCF-7 cells from the B-type compartment in MCF-10A cells (Figure 3.13). Conversely, we also observed a significant decrease of compartment transition from A-type in MCF-10A to B-type in MCF-7 on small chromosomes (Figure 3.13). These findings show that there is a higher frequency of open compartments on small chromosomes in the MCF-7 genome, which suggests a relationship between *inter*-chromosomal clustering, compartmentalization and phenotypic gene expression.



**Figure 3.13. Small chromosomes display enriched “B to A” switching.** Bargraph showing the percentage of compartments that have switched ( $A \rightarrow B$  or  $B \rightarrow A$ ) or remained similar ( $A \rightarrow A$  or  $B \rightarrow B$ ) between MCF-10A and MCF-7 genomes for chr16 through chr22 (blue) and rest of the genome (grey). chr16 – chr22 displays a higher percentage of  $B \rightarrow A$  compartment switching, and a lower percentage of  $A \rightarrow B$  compartment switching between MCF-10A and MCF-7, suggesting a more open compartmentalization in MCF-7. \*\*p-value<0.001: Chi-square with Yates' correction

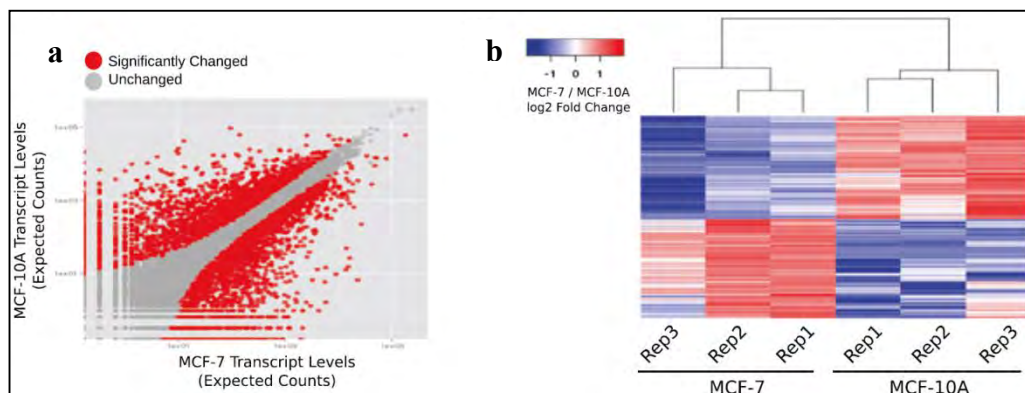
### 3.3.3 Decreased inter-chromosomal interactions and higher frequency of open compartmentalization on chr16-22 in MCF-7 are associated with WNT signaling related genes

Open compartmentalization is correlated with increased gene expression. We asked if the differential interaction network and compartmentalization of chr16 through chr22 between MCF-10A and MCF-7 are associated with differential gene expression. First, to characterize the gene expression differences between MCF-10A and MCF-7 cells, we performed RNA-seq with ribosomal RNA-depleted RNA from MCF-10A and MCF-7 cells with biological triplicates (Figure 3.14).

a MCF--10A Replicates Transcript Per Million (TPM) Pearson Correlation R <sup>2</sup> Values				b MCF--7 Replicates Transcript Per Million (TPM) Pearson Correlation R <sup>2</sup> Values			
	Replicate1	Replicate2	Replicate3		Replicate1	Replicate2	Replicate3
Replicate1				Replicate1			
Replicate2	0.94			Replicate2	0.99		
Replicate3	0.97	0.93		Replicate3	0.99	0.99	

**Figure 3.14. Replicate correlation of MCF-10A and MCF-7 RNAseq datasets.** Pearson correlations of the transcript per million (TPM) values for each biological replicate of a) MCF-10A and b) MCF-7 cells.

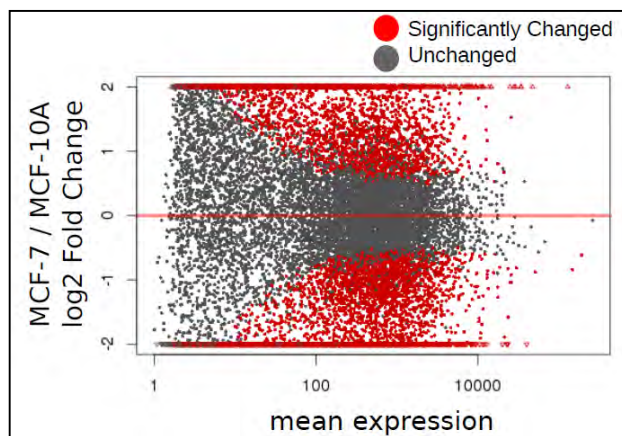
Differential expression analyses identified 2437 MCF-7 upregulated and 2427 MCF-7 down-regulated genes ( $\log_2 FC > 1$ ,  $p < 0.01$ ) with high reproducibility (Figure 3.15). The number of differentially expressed genes identified in this study is comparable to previously published microarray studies [216].



**Figure 3.15. Differential RNA-seq gene expression between MCF-10A and MCF-7 cells.** a) Scatter plot showing differential gene expression between MCF-10A and MCF-7 cells. The axes represent normalized RNA-seq  $\log_2$  gene expression counts. Red and grey dots denote significantly expressed and unchanged genes, respectively. b) Heatmap showing the MCF-7 up and down-regulated genes for each biological replicate. Differential expression analyses identified 2437 MCF-7 upregulated and 2427 MCF-7 down-regulated genes ( $\log_2 FC > 1$ ,  $p < 0.01$ ) with high reproducibility.



The significant expression changes were enriched for the medium to highly expressed genes (Figure 3.16).



**Figure 3.16. The significant expression changes were enriched for the medium to highly expressed genes.** MA-plot showing the log<sub>2</sub> fold change versus the mean expression values of significantly expressed genes.

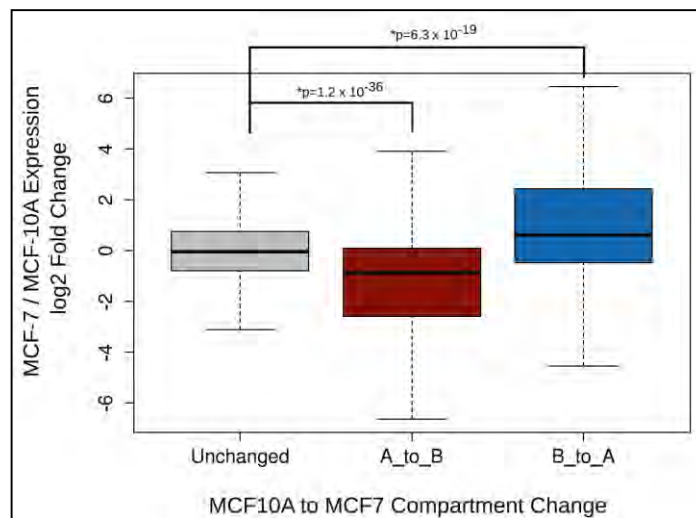
The gene ontology terms associated with MCF-7 downregulated (i.e. MCF-10A over-expressed) genes included terms such as “hemidesmosome assembly”, “focal adhesion”, and “neutral lipid biosynthetic process” (Table 3.2). On the other hand, gene ontology terms associated with MCF-7 upregulated genes included terms such as “calcium-dependent cell adhesion” (Table 3.2).

MCF-7 Downregulated Genes		
Attribute Name	GO_ID	P_adj
hemidesmosome assembly	GO:0031581	<0.001
neutral lipid biosynthetic process	GO:0046460	0.008
acylglycerol biosynthetic process	GO:0046463	0.008
NAD binding	GO:0051287	0.003
O-acyltransferase activity	GO:0008374	0.046
fatty acid beta-oxidation	GO:0006635	0.022

MCF-7 Upregulated Genes		
Attribute Name	GO_ID	P_adj
calcium-dependent cell-cell adhesion	GO:0016339	0.049
synapse assembly	GO:0007416	0.006
homophilic cell adhesion	GO:0007156	<0.001
nervous system development	GO:0007399	0.001
cell-cell adhesion	GO:0016337	0.001

**Table 3.2. GO terms of differentially expressed genes between MCF-10A and MCF-7 cells.** Top 5 gene ontology terms of the MCF-7 down- and up-regulated genes (over MCF-10A).

To test the link between genome-wide open spatial compartmentalization and increased gene expression more directly, we analyzed the frequency of differentially expressed genes at regions where a compartment transition is observed. In agreement with previous findings [77], MCF-7 downregulated genes were enriched in regions where the open A-type compartment in MCF-10A transitioned to a closed B-type compartment in MCF-7 (Figure 2C). Conversely, there was an enrichment of MCF-7 upregulated genes in regions with a B-type compartment in MCF-10A that switched to an A-type compartment in MCF-7 (Figure 3.17). In other words, when the MCF-7 / MCF-10A log<sub>2</sub> fold change expression levels were plotted for each compartment change category, we observed a down-regulation of MCF-7 genes in A-type to B-type compartment switch regions and an up-regulation of MCF-7 genes in B-type to A-type switch regions, respectively (Figure 3.17). These results show that compartment changes in the genome reflect differential gene expression.



**Figure 3.17. Compartmental changes are associated with changes in gene expression.** MCF-7 / MCF-10A log<sub>2</sub> fold change expression box plot of all the genes residing at regions for different compartmental switch categories. The compartments that are A → B and B → A show significantly decreased and increased expression levels, respectively. p-value: Wilcoxon rank-sum test

Finally, to assess whether the differences in interactions and genomic compartments among the small chromosomes are associated with altered gene expression, we focused on the MCF-7 upregulated genes on small chromosomes where the compartmentalization was switched from the B-type to A-type (MCF-10A to MCF-7). REACTOME pathway analysis of these genes revealed well known oncogenic pathways including “Repression of WNT target genes” and “TCF/LEF binding to gene promoters” (Table 3.3). Even though the pathway terms are deemed significant, the total number of genes and the number of genes identified in our dataset are relatively low. Therefore, studying these genes individually may yield more insight into the phenotypic transcriptional changes at these regions with altered interactions.

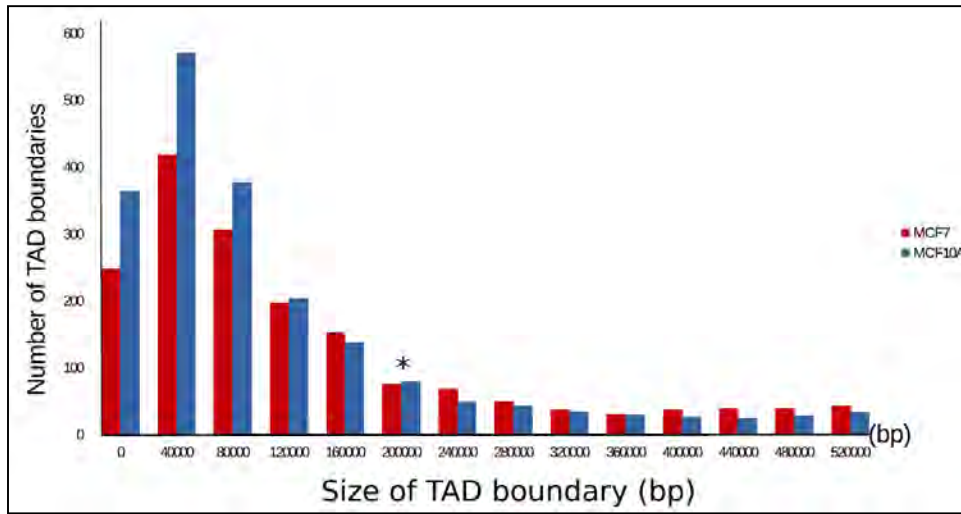
Pathway name	Entities found	Entities Total	Entities pValue
DSCAM interactions	3	15	0.00104
binding of TCF/LEF:CTNNB1 to target gene promoters	2	9	0.00625
ATF6-alpha activates chaperone genes	2	15	0.0165
repression of WNT target genes	2	15	0.0165
Ca2+ pathway	4	78	0.0191
ATF6-alpha activates chaperones	2	17	0.0208
Defective ACTH causes Obesity and Pro-opiomelanocortin deficiency (POMCD)	1	2	0.0256
RHO GTPases activate CIT	2	20	0.0281
RHO GTPases Activate ROCKs	2	22	0.0335
PI and PC transport between ER and Golgi membranes	1	3	0.0381
RHO GTPases activate PAKs	2	27	0.0484

Table 3.3. REACTOME pathways of the genes that are located on chr16-22 and at B → A compartment regions.

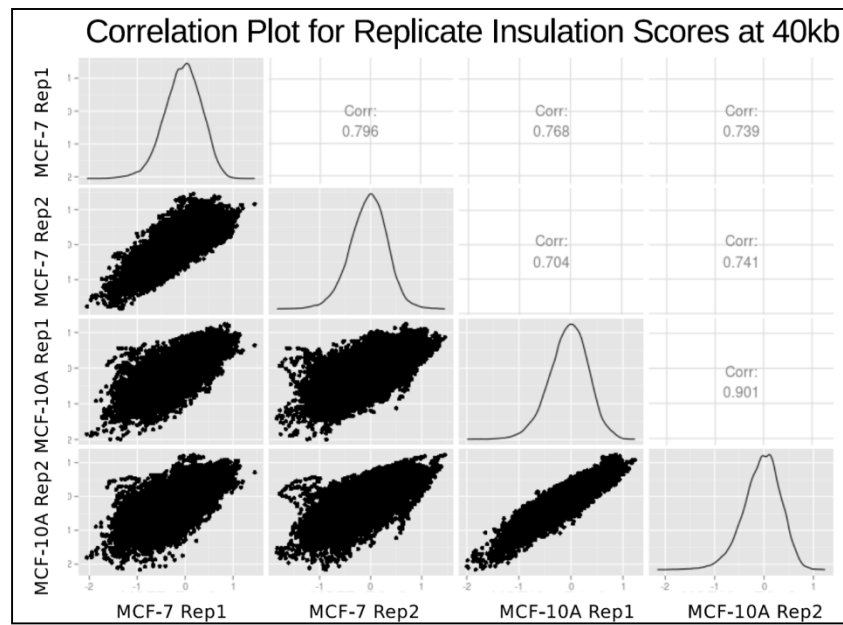
Taken together, these results suggest that the decrease of *inter*-chromosomal associations of small chromosomes in the MCF-7 genome is associated with a higher open compartmentalization in MCF-7 and expression of genes related to WNT signaling pathway, which is frequently implicated in tumorigenesis.

### **3.3.4 Cell-Line Specific TAD Boundaries Are Conserved between MCF-10A and MCF-7**

Chromosome conformation capture based studies revealed that A-type and B-type compartments are composed of topologically associating domains (TADs), in which the expression levels of the genes in a single TAD can be co-regulated [73; 76; 217]. TADs have been shown to be stable units in different species, cell types and physiological conditions [76; 77]. However, whether the large-scale chromosomal interactions and altered genomic compartments observed between MCF-10A and MCF-7 genomes have an effect on the structure of the underlying TAD formation and ultimately on gene expression is unknown. To address this question, we identified the TAD boundaries by calculating the insulation plot of the 40kb resolution genome-wide interaction maps (see Methods, and Figure 3.18), with high reproducibility between the biological replicates (Figure 3.19).

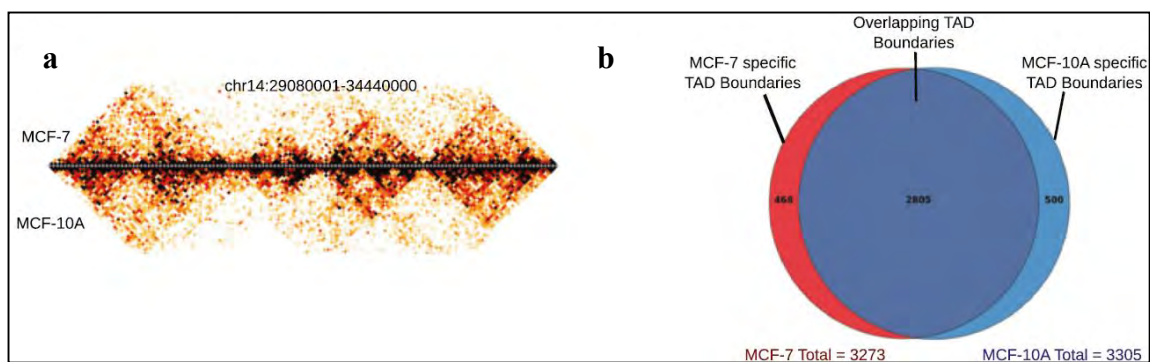


**Figure 3.18. TAD boundary analysis for replicate Hi-C datasets.** Bar graph showing the number of shared TAD boundaries in relation to their set size (see Methods). **b)** Scatter plots showing the Pearson correlation of the genome-wide insulation plots across the biological replicates.



**Figure 3.19. Correlation plots of insulation scores across Hi-C biological replicates.** Scatter plots showing the Pearson correlation of the genome-wide insulation plots across the biological replicates.

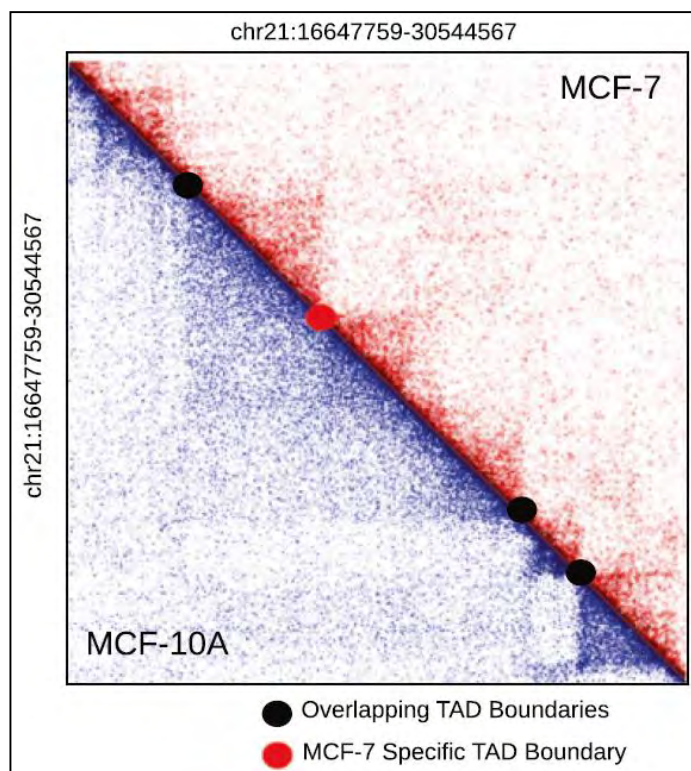
We detected 3305 and 3272 TAD boundaries in MCF-10A and MCF-7 genomes, respectively. Despite the differences in chromosomal structure and changes in compartmentalization and gene expression, ~85% (2805) of the TAD boundaries were common between the cell lines (Figure 3.20). This rate of TAD boundary overlap is consistent with previous comparisons in different cell types and conditions [76; 77]. This result suggests that despite having cell-type specific translocations and large-scale structural differences, TAD boundaries are consistent between non-tumorigenic and tumorigenic cells.



**Figure 3.20. MCF-7 and MCF-10A cells have similar TADs.** **a)** TADs are similar between MCF-10A and MCF-7 genomes. An example heatmap of a portion of MCF-10A chr14 at 40kb resolution, where the upper part of the heatmap shows the MCF-7 TADs, and the bottom panel showing the MCF-10A TADs. **b)** Venn diagram showing that majority (~85%) of all the TAD boundaries between MCF7 and MCF10A are conserved.

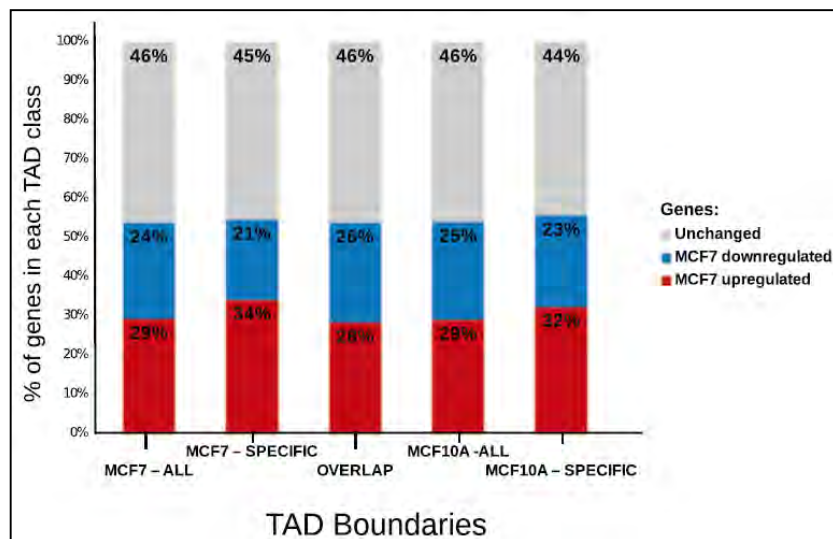
Closer examination of TAD boundaries revealed that there were several TADs that were “broken” into multiple sub-TADs between the cell lines. The boundaries that were shared among the larger and smaller TADs between the cell lines were categorized as “overlapping”, and the boundaries that were unique to a cell line were categorized as “cell-line specific” boundaries (Figure 3.21). We asked whether the genes residing at the cell-specific boundaries showed cell-specific differential gene

expression. When the percentages of unchanged and MCF-7 up- and down-regulated genes were plotted per TAD boundary category, we did not find a strong correlation between cell-type specific TAD boundaries and differential gene expression (Figure 3.22).



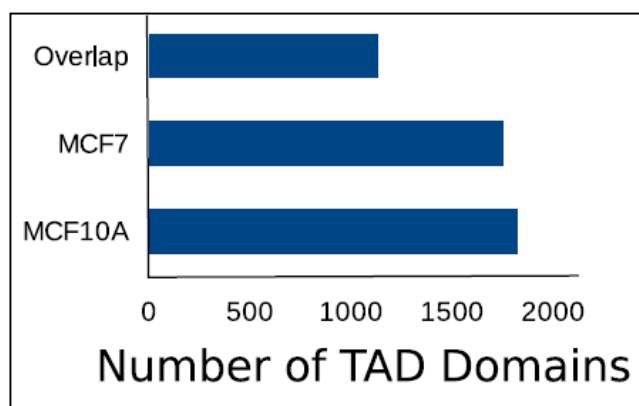
**Figure 3.21. TADs can be disrupted in cancer cells.** An example heatmap showing an example of a differential TAD between MCF-10A (blue) and MCF-7 (red) on chr21 (chr21:16647759-30544567). The black dots represent the overlapping boundaries that are present in both cell lines, and the red dot denotes the MCF7-specific TAD boundary.



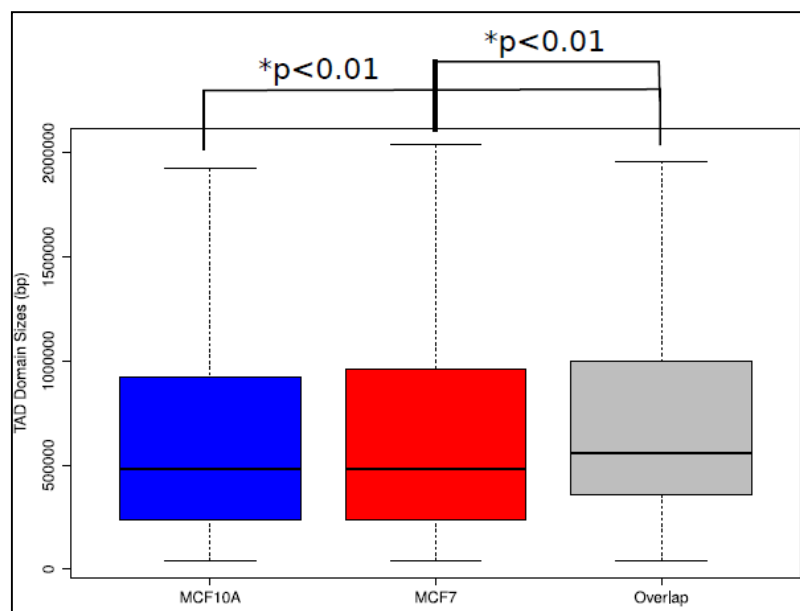


**Figure 3.22 Gene expression is not associated with TAD boundaries.** The percentage of unchanged (grey), MCF7 downregulated (blue) and MCF7 upregulated (red) genes located at each TAD boundary category.

Apart from the TAD boundaries, we next analyzed the TAD domains. We categorized the TAD domains as overlapping (>90% overlap), MCF-7 specific or MCF-10A specific (see Methods) (Figure 3.23).



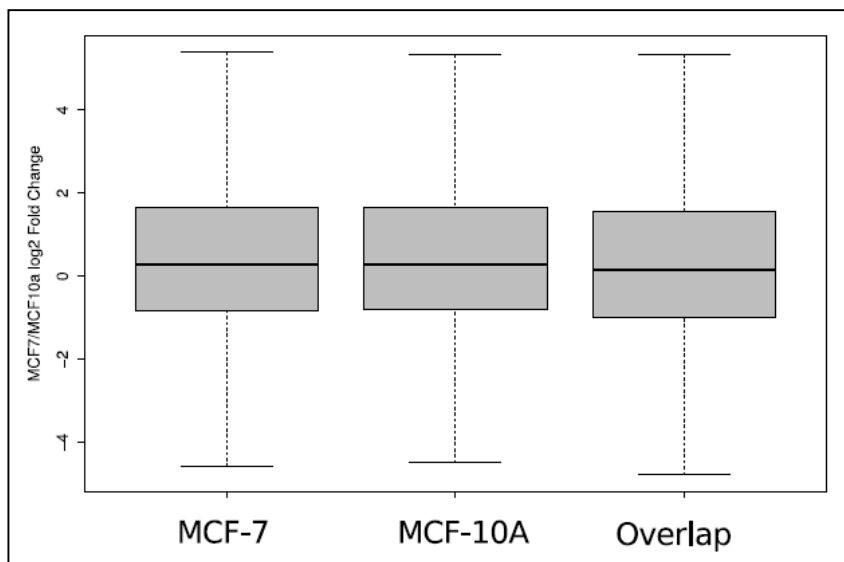
**Figure 3.23. Analysis of TAD domains.** Number of overlapping and cell-type specific TAD domains.



**Figure 3.24a. Analysis of TAD domain sizes.** Box plots showing the sizes of MCF-10A and MCF-7 specific, as well as overlapping TADs. p-value : Wilcoxon rank sum test.

The sizes of the overlapping TAD domains were slightly larger than the cell-line specific TADs (Figure 3.24).

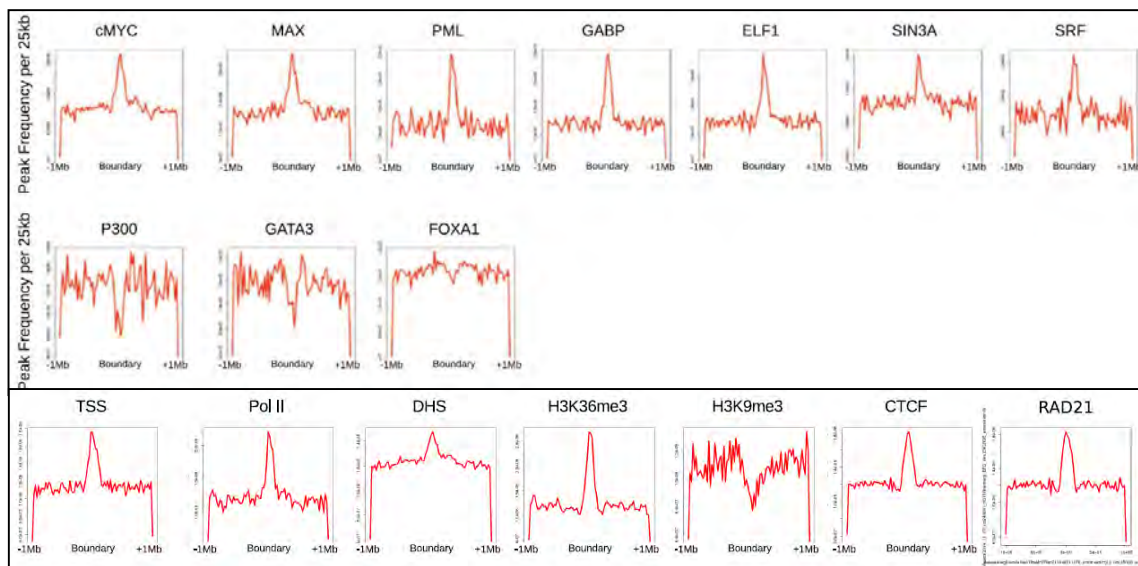
We then asked whether cell-line specific TAD domains showed differential gene expression. Analysis of differential gene expression for each TAD domain category showed that cell-type specificity of the TAD domains was not correlated with cell-type specific gene expression (Figure 3.25).



**Figure 3.24b. Gene expression analysis of TAD domains.** Box plots showing MCF-7 / MCF-10A log<sub>2</sub> fold change values for the genes located at MCF-7 and MCF-10A specific, and overlapping TAD domains. There is not significant difference. p-value : Wilcoxon rank sum test.

### 3.3.5 MCF-7 TAD Boundaries Are Enriched for Several Oncoproteins

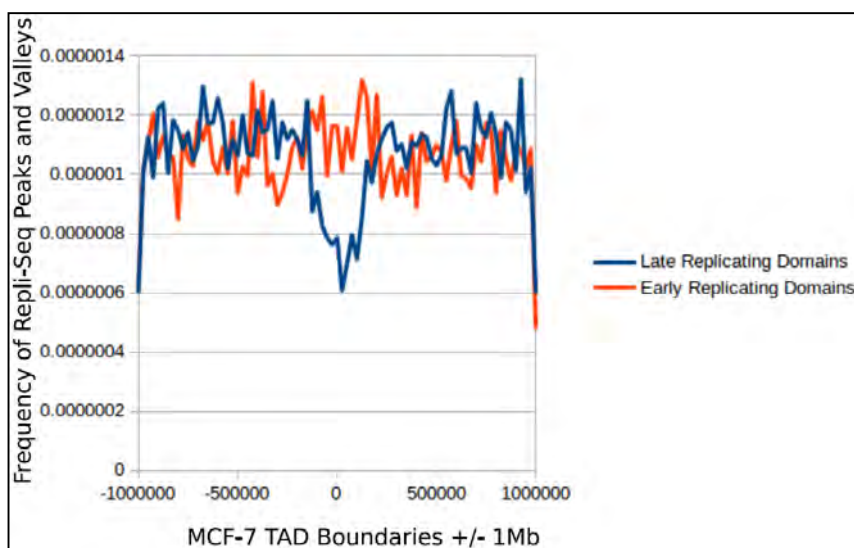
TAD boundaries are bound by multiple factors [73; 79]. To investigate the chromatin states of the boundaries, we calculated the enrichment of factors characterized by MCF-7 ENCODE datasets at the MCF-7 TAD boundaries (Figure 3.25). The known features of the TAD boundaries, such as the enrichment of H3K36me<sub>3</sub>, CTCF, RAD21, TSS, POL2 and DHS, and the depletion of H3K9me<sub>3</sub>, were observed at the MCF-7 TAD boundaries (Figure 3.25, lower panel). Interestingly, we observed a strong association of GABP, ELF1, PML, SIN3A, SRF and the oncogenic drivers cMYC and MAX at MCF-7 TAD boundaries, and a depletion of GATA3 and FOXA1 (Figure 3.25). Consistent with previous work [73], P300 was depleted at the MCF-7 boundary regions. The rest of the MCF-7 ENCODE datasets did not show any enrichment.



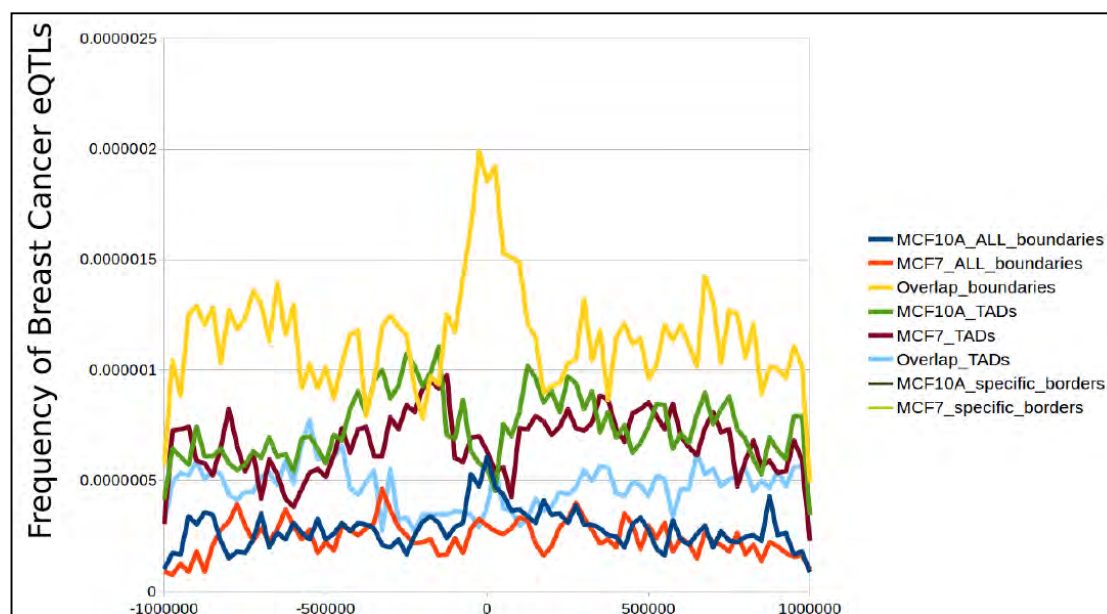
**Figure 3.25. Analysis of factors enriched at MCF-7 TAD boundaries.** The frequency plots of factors enriched at MCF-7 TAD boundaries per 25kb for plus/minus 1Mb of every MCF-7 TAD boundary.

Recent evidence suggested that TADs may act as stable units of replication domains [72]. Therefore, we intersected the previously published MCF-7 Repli-seq dataset [72] with MCF-7 TAD boundaries and, consistent with the literature, we determined that late replicating regions were depleted at TAD boundary regions (Figure 3.26).

Moreover, eQTLs have been shown to be preferentially located at TAD boundaries [54]. Integrating the breast cancer eQTL data [110] with MCF-7 TAD boundaries, we determined that breast cancer associated eQTLs were enriched in overlapping TAD boundaries (Figure 3.27). Altogether, these results uncover previously unidentified transcription factors and chromatin states that may potentially play roles at the TAD boundaries.



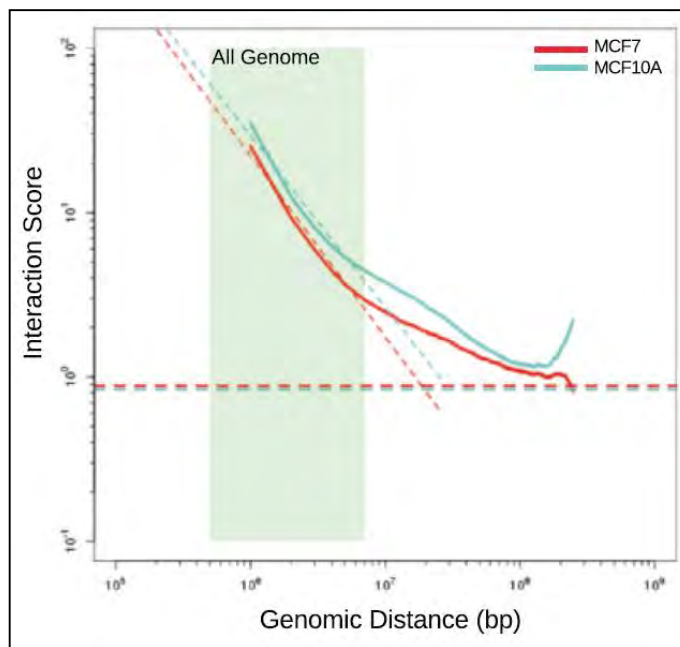
**Figure 3.26. Analysis of MCF-7 Repli-seq data with TAD boundaries.** MCF-7 Repli-seq data [72] frequency across MCF-7 TAD boundaries.



**Figure 3.27. eQTL analysis across TAD boundaries.** Enrichment of breast cancer associated eQTLs [218] with MCF-7 TAD boundaries and domains.

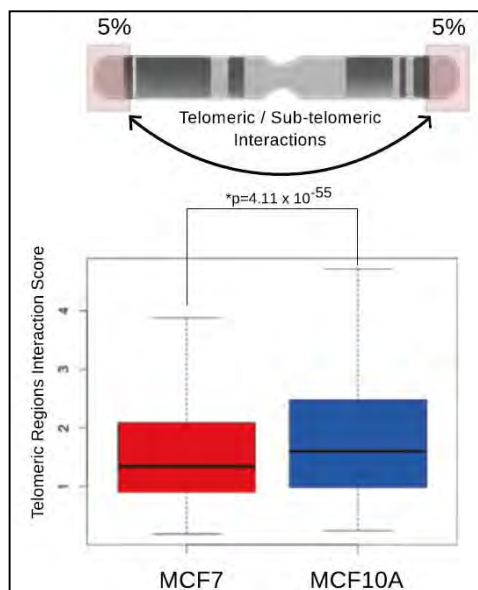
### **3.3.6 The telomeric / sub-telomeric regions in MCF-10A genome display stronger associations than those in MCF-7**

Previous evidence has shown that interaction frequency decreases as a function of genomic distance [71]. This phenomenon represents the nature of the chromatin fiber and is a reflection of the folding status of the underlying chromatin [219]. We first asked whether the fiber characteristics of the MCF-10A and MCF-7 genomes were similar. Scaling plots of 1Mb binned genome-wide *intra*-chromosomal interactions displayed the expected exponential decrease of contact probability as a function of increasing genomic distance in both MCF-10A and MCF-7 cells (Figure 3.28). Surprisingly, and in contrast to all previously published human Hi-C datasets, the frequency of interactions in MCF-10A showed an increase at very large genomic distances (>200Mb) (Figure 3.28). This suggests that very distant (i.e telomeric / sub-telomeric) regions of chromosomes show a higher interaction frequency on the same chromosome.



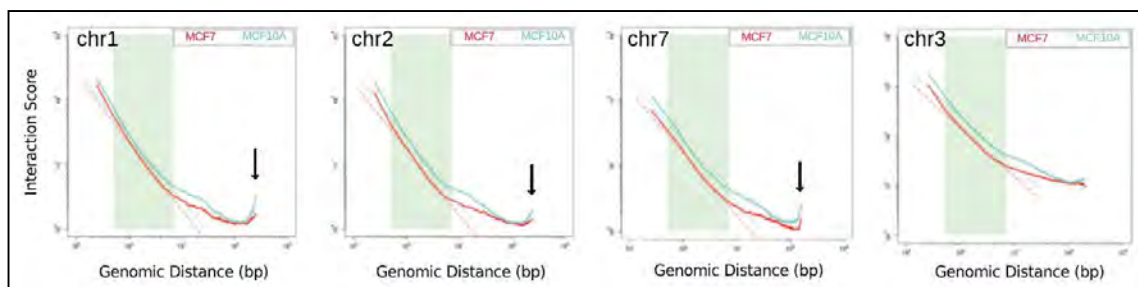
**Figure 3.28. Scaling plot of interaction frequencies against genomic distance for the MCF-7 and MCF-10A genomes.** The MCF-10A genome showed increased interaction frequency at genomic distances that are larger than 200 megabases, suggesting telomere/sub-telomere associations.

To assess whether the telomeric ends of the chromosomes in MCF-10A indeed have higher frequency of interactions compared to MCF-7, we calculated the *intra*-chromosomal interaction frequency of the ends of each chromosome (5% by length) in MCF-10A and MCF-7. We observed a significant increase in telomeric / sub-telomeric interaction frequency in the MCF-10A genome (Figure 3.29), which supports the observation that *intra*-chromosomal telomeric interactions are more frequent in MCF-10A cells.



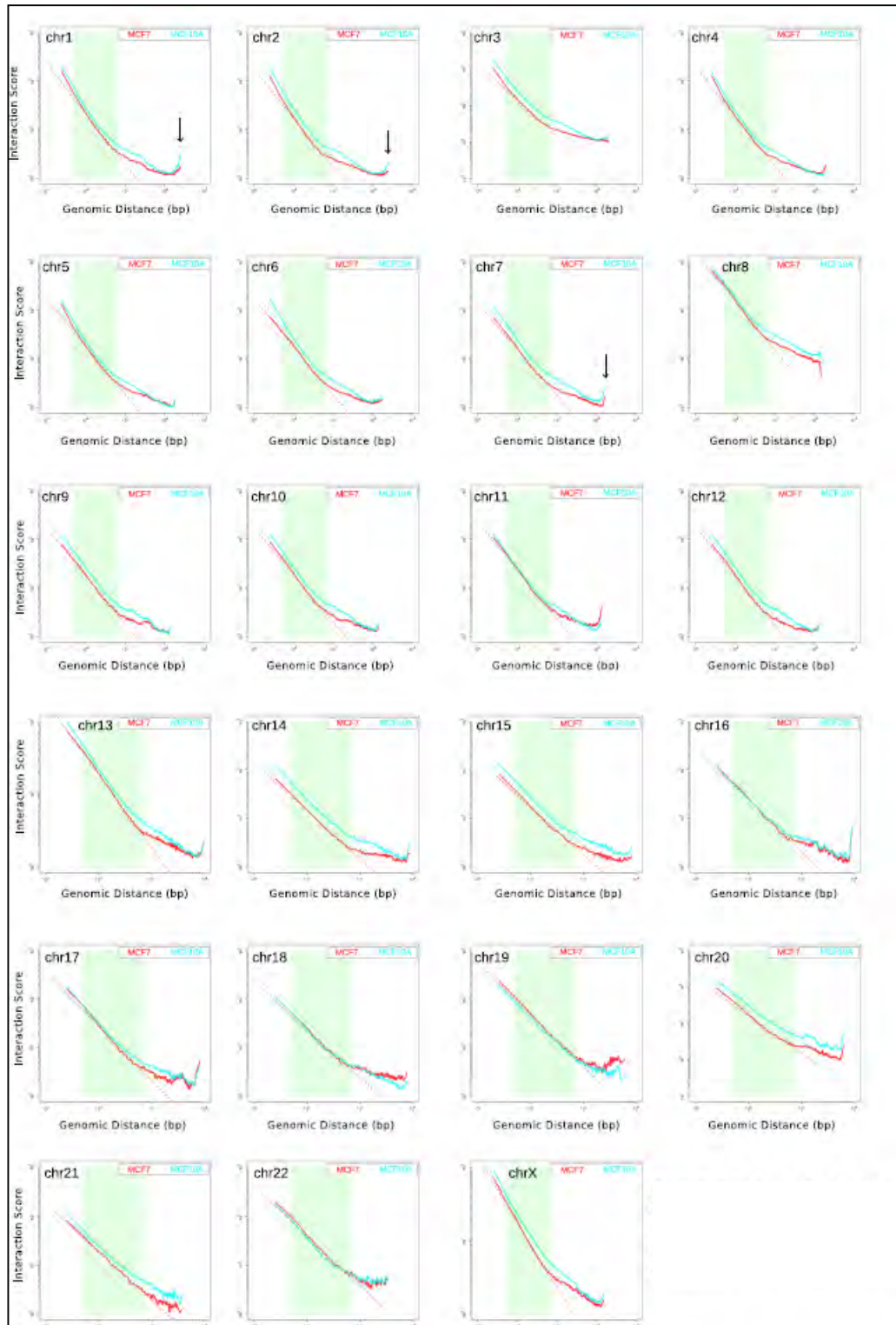
**Figure 3.29. Quantification of the interaction frequency between the telomeric regions (5% of the ends by length) of each chromosome in MCF7 and MCF10A.** p-value: Wilcoxon rank-sum test.

Scaling plots of each chromosome individually at 250kb resolution indicate that the increase in telomeric / sub-telomeric interactions seemed to be driven by chr1, chr2 and chr7 in the MCF-10A genome (Figure 3.30 and Figure 3.31). However, this phenomenon was not observed on other large chromosomes in MCF-10A cells, such as chr3 (Figure 3.30).



**Figure 3.30. The scaling plots of MCF-10A and MCF-7 for chr1, chr2, chr7 and chr3.** Chromosomes 1, 2 and 7 displayed an increased interaction frequency at large distances in MCF-10A, however chromosome 3 did not.

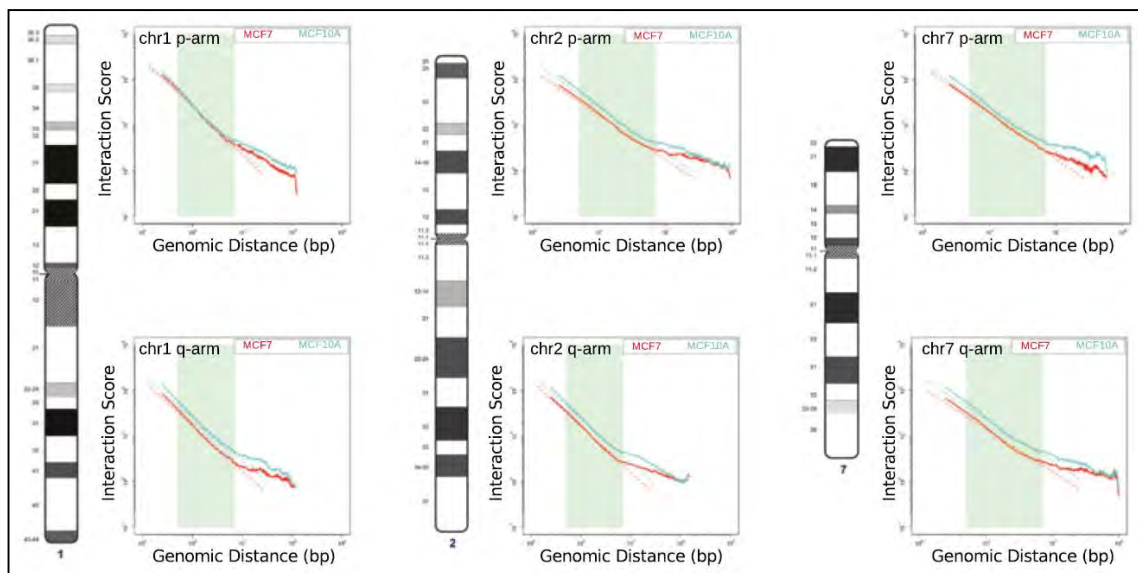




**Figure 3.31.** The chromosome by chromosome scaling plots generated by using 250kb binned Hi-C data.

Certain chromosomes, such as chr11 and chr16, showed increased interaction frequency at large distances both in the MCF-10A and MCF-7 genomes even though their lengths did not span 200Mb (Figure 3.31). As expected, this observation was lost when the scaling plots for individual chromosomal arms were analyzed (Figure 3.32).

These results suggest that the telomeric ends of the chromosomes, especially chr1, 2 and 7, in the MCF-10A genome are in closer proximity than those in MCF-7. Taken together, we identified large-scale differences in both *cis*- and *trans*-chromosomal interactions between two commonly used cell lines in breast cancer research.



**Figure 3.32. The scaling plots of individual chromosome arms for chr1, chr2 and chr7 did not show enriched telomeric / sub-telomeric interactions.**

### 3.4 Discussion

Cancer is a disease characterized by major morphological changes in the nucleus [126; 186]. Although individual gene positioning may differ [189], relative arrangement of chromosomes in the interphase nucleus can be conserved between normal and cancer cells [38]. Furthermore, extensive epigenetic dysregulation is observed in the cancer state. In order to map the genome-wide interactions and perform a comparative analysis, we performed Hi-C in the MCF-10A and MCF-7 cell lines. We observed a higher background interaction frequency in the MCF-7 genome compared to MCF-10A (Figure 3.2). This background could be the result of a technical source (i.e the ligation step in the Hi-C procedure), or because of increased background interaction frequency in the MCF-7 genome due to the probabilistic positioning of the chromosomes inside the aneuploid nucleus and increased diversity of interactions within this genome.

Comparison of MCF-7 and MCF-10A Hi-C data revealed a significant depletion of *inter*-chromosomal associations between small, gene-rich chromosomes (chr16-22) in the MCF-7 genome. One possibility for the loss of interactions among the small chromosomes in MCF-7 compared to MCF-10A cells is that randomization (*i.e* loss of specificity) of contacts within the MCF-7 genome could lead to lower frequencies of individual contacts, and hence to an apparent loss of interaction. However, loss of specific contacts does not itself cause a difference in overall chromosome contacts. Two whole chromosomes that tend to be close together in a cell will overall show more inter-chromosomal interactions with each other by Hi-C than will two distant chromosomes,

even if they have no specific interactions that are consistent across the population of cells. If each cell in the population has a different arrangement of chromosome territories, this will look on average like less clustering of small chromosomes. But, this scenario should also reveal more interactions between large and small chromosomes and less clustering of large chromosomes. In Figure 3.22 and Figure 3.6, in contrast, we do not observe a compensating increase in interactions between the small and large chromosomes, suggesting that this is not just a randomization of interactions. Moreover, it should be kept in mind that there are several extensive rearrangements in the MCF-7 genome, and it could be that only the re-arranged copies of a highly aneuploid chromosome may show a particular three-dimensional conformation.

The decreased clustering of small chromosomes and the differentially open compartmentalized regions in MCF-7 are associated with increased expression of genes related to tumorigenesis. The correlation between increased gene expression in B-type to A-type compartment switch regions and a higher number of A-type compartments on MCF-7 chr16-22 suggests that the underlying mechanism for this phenomenon is most likely due to transcriptional differences, rather than chromosomal copy number changes between the cell lines. The loss of small chromosome clustering may also be interpreted as a reflection of mis-organization of the chromosome territories in cancer.

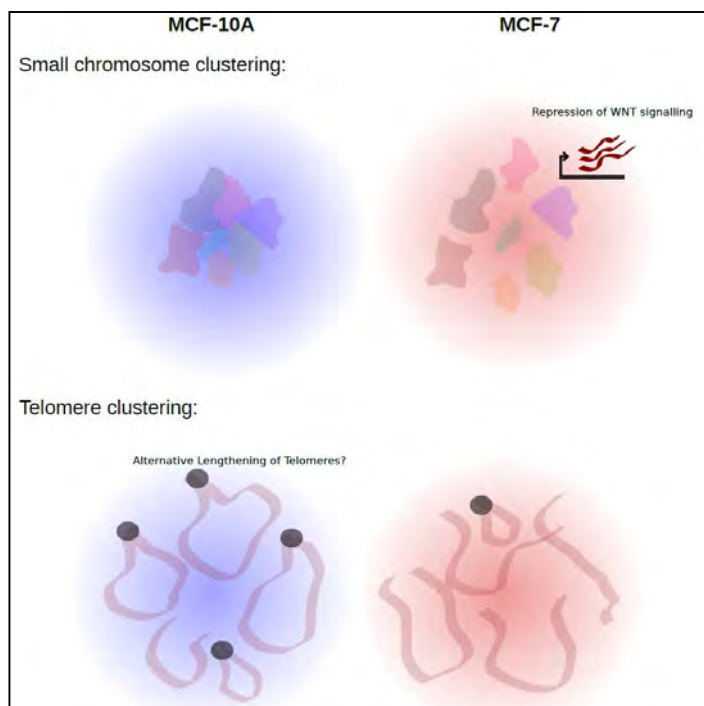
Genomic compartmentalization has been shown to be associated with gene expression [71; 77]. One hypothesis for the clustering, compartmental and transcriptional changes we observe in small chromosomes would be that once a gene is

activated/repressed in the process of tumorigenesis, its position in the three-dimensional nuclear space is changed, with movement towards the open/closed compartment regions. Such a phenomenon has been previously shown by microscopic studies [220]. An alternative hypothesis is that chromosomes change compartments before gene expression changes. A recent study supports the alternative hypothesis in which chromatin decondensation plays a major role in cell differentiation [221].

Scaling plot analysis (Figures 3.28-3.32) suggested that distinct types of chromatin folding states might exist between MCF-10A and MCF-7 cells, both genome-wide, and at individual chromosomes [219]. Surprisingly, and in contrast to all previous human Hi-C datasets, there was an increased frequency of interactions at distances >200Mb away in MCF-10A cells, suggesting interactions between telomeric and sub-telomeric regions on the same chromosome. It has been suggested that telomere clustering is associated with the Alternate Telomere Lengthening (ALT) mechanism [222]. ALT is a mechanism in which telomere length is maintained through a homologous recombination-dependent process. It could be possible that the MCF-10A and MCF-7 cells have different mechanisms of telomere maintenance, and the proximity of telomeric ends in the MCF-10A genome might suggest an effect of increased ALT regulation. Increased telomere interactions were observed in chr1, chr2 and chr7, and on some smaller chromosomes (Figure 3.31), but not in individual chromosomal arms (Figure 3.32). A recent report suggests that 10% of all cancers and immortalized cell lines display the ALT mechanism [223]. Our results are consistent with previous

findings that the presence of an ALT mechanism results in clustering of telomeres, which is observed in epithelial MCF-10A cells but not in tumorigenic MCF-7 cells.

Overall, in this study we charted the chromatin structure of mammary epithelial and breast cancer cells at different chromosomal scales, from large-scale chromosomal cis- and trans-interactions to genomic compartmentalization and TAD formation (Figure 3.33). Further studies on normal and cancer genomes and primary cells will provide additional insight into the functional role of chromatin organization in transcriptional regulation and tumorigenesis.



**Figure 3.33. Overview of the changes in chromatin architecture between MCF-10A and MCF-7 cells.** There are two large-scale changes between the genomes of MCF-10A and MCF-7. First, the inter-chromosomal interactions between the small, gene rich chromosomes are drastically weaker in the MCF-7 breast cancer genome. In concordance with this, there is higher frequency of open genomic compartments, and higher gene expression on chromosomes 16 through 22, especially of the genes related with pathways reflecting the phenotype of the MCF-7 cells. Secondly, the intra-chromosomal associations of the telomeric ends of chromosomes are lost in the MCF-7 genome. This phenomenon may reflect differential telomeric maintenance mechanisms.

## **CHAPTER4: BRG1 regulates gene expression and higher-order chromatin structure in proliferating mammary epithelial cells**

### **4.1 Introduction**

Organization of chromatin is essential for many biological processes. Packaging of the DNA around the nucleosomes acts to tightly condense the genome [224]. At the same time, the cell has to regulate the accessibility of chromatin to many enzymes for the regulation of gene expression, DNA replication and repair. Maintaining a balance between tight packaging and accessibility of chromatin is an important function of the eukaryotic nucleus. This balance is achieved by multiple specialized protein complexes that dynamically alter chromatin structure in an ATP-dependent manner [225]. Four families of ATP-dependent chromatin remodelers exist: SWI/SNF, ISWI, INO80 and CHD (reviewed in [226; 227]). The ATPase subunits of each family have a conserved helicase-like ATPase domain that uses the energy from ATP hydrolysis to evict, reposition or modify nucleosomes [228; 229]. These structural changes in chromatin result in local alteration of chromatin followed by the binding of regulatory proteins. Different families of remodelers work in a dynamic and orchestrated way in cells to fine tune DNA accessibility [230].

The mammalian SWI/SNF (mating-type switching – SWI / and sucrose non fermenting – SNF) complexes contain one of two distinct ATPase subunits, BRM (Brahma, or SMARCA2) or BRG1 (Brahma-related Gene 1, or SMARCA4) [231-233]. BRM is dispensable, as BRM null mice can properly develop to adulthood [234]. On the other hand, BRG1 is essential, as BRG1 null mice are embryonic lethal and BRG1 heterozygous mice are prone to mammary tumor formation [235; 236]. BRG1 has been

shown to be involved in many developmental processes and in transcriptional regulation, DNA repair, cell cycle control, and cancer [237; 238]. The role of BRG1 in gene regulation is contextual, as it can activate some promoters while repressing others. In addition, extensive dysregulation and mutations of BRG1 have been implicated in many different cancer types, making BRG1 a potential therapeutic target for cancer [239; 240].

The organization of the nucleus occurs in a hierarchical manner. First, chromosomes are positioned in distinct volumes forming the chromosome territories [23] , which consist of open (A-type) and closed (B-type) genomic compartments [71]. The genomic compartments are further folded into sub-megabase scaled structures called topologically associated domains (TADs)[73; 74], where local looping interactions between promoters and enhancers occur [94] . Through interactions with many different protein partners [237; 241], BRG1 is involved in nuclear structure and mediating long-range chromatin interactions between genes expressed at the same time in response to differentiation signaling [242], and between the enhancers and their cognate promoters at many gene loci, including the beta and alpha-globin genes [243; 244] , the IgH locus [245] , and the class II major histocompatibility complex gene (CIITA) [246]. BRG1 binds to poised developmental enhancers in embryonic stem cells [247; 248] and B-cells [245] and colocalizes with pluripotency factors [249], suggesting important roles in enhancer function. Furthermore, previous work classifying genome-wide interactions according to their histone modifications and transcription factor binding revealed BRG1 enrichment at open chromatin regions, indicating a possible



structural role for BRG1 [250]. In addition, BRG1 regulates nuclear size [251] and the integrity of nuclear shape via a mechanism independent of cytoskeletal connections [252]. Recently, it has been shown that BRG1 is involved in the lncRNA-dependent assembly of nuclear bodies [253].

Taken together, apart from its chromatin remodeling activity at the regulatory regions of target genes, emerging evidence suggests a possible important role for BRG1 in maintaining the structural integrity of the nucleus by regulating global chromatin structure [254]. To date, very little is known about the role of BRG1 in global higher-order genome organization. In order to gain insight into the role of BRG1 in nuclear organization at a genome-wide level, we performed RNA-seq and Hi-C in BRG1-depleted and control MCF-10A human mammary epithelial cells. Furthermore, to map the localization of BRG1 in the genome, we performed BRG1 ChIP-seq in the parental MCF-10A cells. We show that BRG1 depletion is associated with extensive changes in gene expression and higher-order chromatin structure at multiple levels.

## **4.2 Materials and Methods**

### **4.2.1 MCF-10A Cell Culture**

MCF-10A cells expressing control shRNA and shRNA targeting BRG1 were generated as previously described [255]. The cells were maintained in monolayer in Dulbecco's modified Eagle's medium-F12 (DMEM/F12) (Invitrogen, 21041025) supplemented with 5% horse serum (Invitrogen, 16050122), 1% penicillin/streptomycin (Invitrogen, 15140122), 0.5 µg/ml hydrocortisone (Sigma, H-0888), 100 ng/ml cholera toxin (Sigma, C-8052), 10 µg/ml insulin (Sigma, I-1882), and 20 ng/ml recombinant human EGF

(Peprtech, 100-15) as previously described [206]. The doxocycline induction was performed by the addition of 0.05 microgram per milliliter DOX to the cells and incubating them for 3 to 4 days.

#### **4.2.2 RNA-seq and Analysis**

RNA was isolated from MCF-10A cells at ~75% confluence using the TRizol Reagent (Life Technologies #15596-026) treated with DNase1. The poly(A)-selected RNA-seq libraries were generated with TruSeq RNA Sample Preparation Kit v2 and SE100 sequencing was performed using a Hi-Seq 2000 instrument. RNA-seq analysis was performed by filtering and mapping the reads by Bowtie [269], quantifying the transcripts by RSEM v1.2.7 [208] and finding the differentially expressed genes ( $\log_2$  fold change  $>1$ ,  $p < 0.01$ ) by DeSeq2 [209].

#### **4.2.3 Preparation of Hi-C Libraries**

Hi-C was performed as previously described with minor modifications [99]. The modification was in the biotin incorporation step, where the mixture was incubated at 37°C for 40 minutes. The MCF-10A shSCRAM and shBRG1 samples displayed a range of 25% to 50% biotin incorporation efficiency. At the end of Hi-C sample preparation, the libraries were sequenced using PE100 read with a Hi-Seq 2000 instrument.

#### **4.2.4 Read Mapping / Binning / ICE correction**

Table 4.1 summarizes the mapping results and different classes of reads and interactions observed [100]. The data were binned at 2.5Mb, 1Mb, 250kb, 100kb and 40kb non-overlapping genomic intervals. In our Hi-C analyses, we utilized the iterative correction and eigenvector decomposition (ICE) method [212]. The replicates showed high

correlation at multiple scales. For the downstream analyses, sequences obtained from both biological replicates were pooled and ICE-corrected to serve as a combined dataset.

#### **4.2.5 Z-score Calculation**

We calculated the z-scores by modeling the overall Hi-C decay with distance using a modified LOWESS method (alpha = 1%, IQR filter), as described previously [94].

LOWESS calculates the weighted-average and weighted-standard deviation for every genomic distance and therefore normalizes for genomic distance signal bias.

#### **4.2.6 Calculation of Differential Interactions**

To capture the differences between shSCRAM and shBRG1 interactions, we used a method previously described [263]. Briefly, we first transformed the Hi-C data into Z-score matrices for all 4 replicate datasets (shBRG1-R1, shBRG1-R2, shSCRAM-R1, and shSCRAM-R2). For each interaction, the mean sample:sample (between samples) Z-score difference was calculated from all pairwise combinations of the four datasets (shBRG1-R1 – shSCRAM-R1, shBRG1-R1 – shSCRAM-R2, shBRG1-R2 – shSCRAM-R1, shBRG1-R2 – shSCRAM-R2). The replicate:replicate Z-score difference (within samples) was also calculated for a random set of 500,000 interactions. These random replicate-replicate Z-score differences were then used to build an expected distribution of Z-score differences. The resulting Z-score difference matrix was then derived by calculating for each bin the ratio of the mean of the set of 4 possible sample:sample Z-score differences minus the genome-wide mean of the replicate:replicate Z-score difference, divided by the genome-wide standard error of the replicate:replicate Z-score differences.

#### **4.2.7 Compartment Profiles**

To detect the genomic compartments, first, Pearson Correlation of the Z-score matrices was calculated. In performing principal component analysis [71; 191], the first principle component detects the patterns of increased and decreased interaction across the genome that appear as a “plaid pattern” in the heatmap. Each genomic region matches this prominent interaction pattern (positive eigenvector value) or its opposite (negative eigenvector value) and these represent the two spatially segregated compartments. The open, gene rich “A-type” compartment may end up with either a positive or negative eigenvector. To detect which compartment is the open “A-type” and which is the closed “B-type”, the genome wide gene density was calculated to assign the “A-type” and “B-type” compartmentalization.

#### **4.2.8 Identification of TAD Boundaries (Insulation Square Analysis)**

TAD calling was performed as calculating the “insulation” score of each bin using the 40kb resolution combined Hi-C data as previously described [213; 263] .

#### **4.2.9 ChIP-seq Analysis**

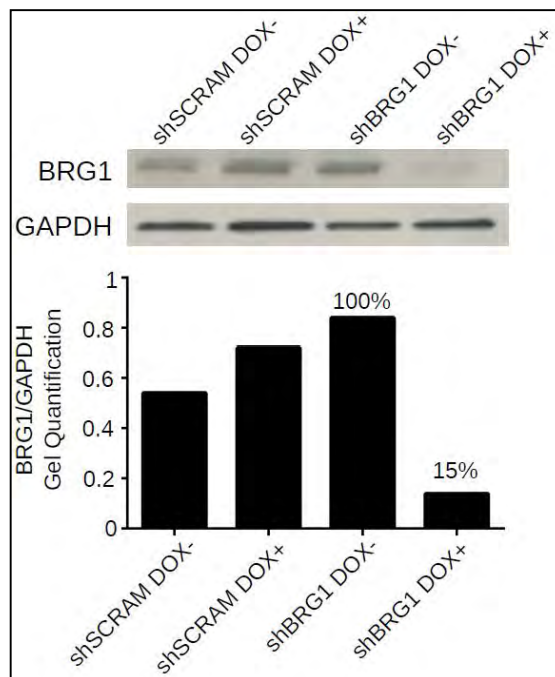
The ChIP assay was performed as previously described [270]. Briefly,  $\sim 1 \times 10^7$  parental MCF-10A cells were crosslinked with formaldehyde at room temperature for 10 minutes. Then, the cells were lysed using lysis buffer A (50mM HEPES, 140mM NaCl, 1mM EDTA pH=8, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100) and the residual cytoplasmic protein was removed using lysis buffer B (10mM Tris-HCl pH=8, 200mM NaCl, 1mM EDTA, 1mM EGTA). The nuclear fraction was released using lysis buffer C (10mM Tris-Hcl pH=8, 100mM NaCl, 1mM EDTA, 1mM EGTA, 0.1% Sodium

Deoxycholate, 0.5% N-lauroylsarcosine). Chromatin was then sheared by using a Bioruptor instrument on high setting, 30' on , 30' off, for 5 minutes for 5 cycles. The pull-down was performed using a BRG1 antibody (Santa Cruz #G-7). Samples were washed three times with RIPA buffer (Tris-HCl pH=8, 150mM NaCl, 1mM EDTA, 1% NP-40, 0.25% Sodium deoxycholate, 0.1% SDS) and were eluted. The pull-down and input control sequencing libraries were generated using the NEXTflex Rapid DNA Sequencing Kit (Bioo Scientific #5144-02) and were sequenced by using SE100 reads with a HiSeq 2000 instrument. The adapters were trimmed from the sequencing reads, and the reads were aligned to the hg19 human genome using the Bowtie2 tool [269]. Quality controls, peak calling, motif analysis and peak annotation were performed using the HOMER suite [271]. As the ChIP signal across the biological replicates showed high correlation , we performed ChIP-seq peak calling on the pooled replicates.

### **4.3 Results**

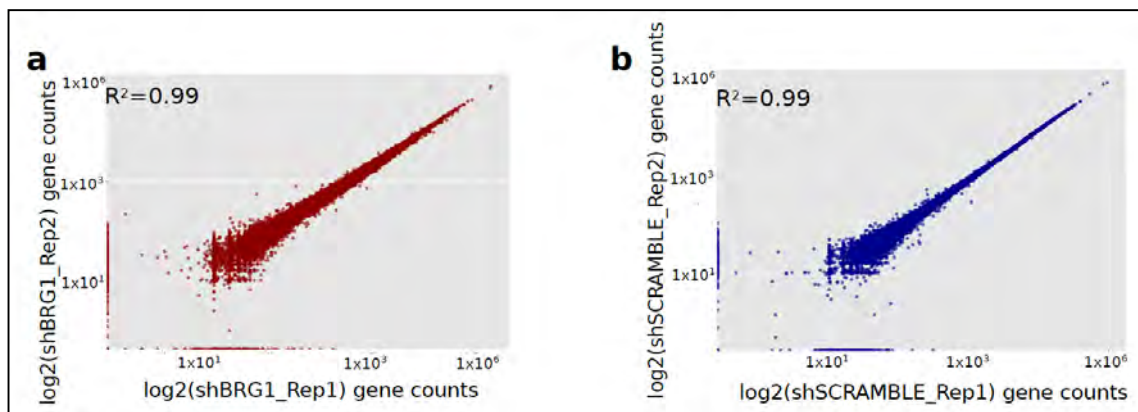
#### **4.3.1 BRG1 knockdown results in down-regulation of genes associated with extracellular matrix in MCF-10A cells**

BRG1 plays an extensive role in regulating gene expression. In order to investigate the transcriptional effects of BRG1 depletion in mammary epithelial cells, we used previously described doxocycline-inducible MCF-10A mammary epithelial cells expressing either a non-specific (scrambled) shRNA (shSCRAM), or shRNA against BRG1 (shBRG1) [255]. BRG1 knockdown was confirmed by western blot analysis. Quantification of the western blot demonstrated ~85% reduction of BRG1 protein levels (Figure 4.1).

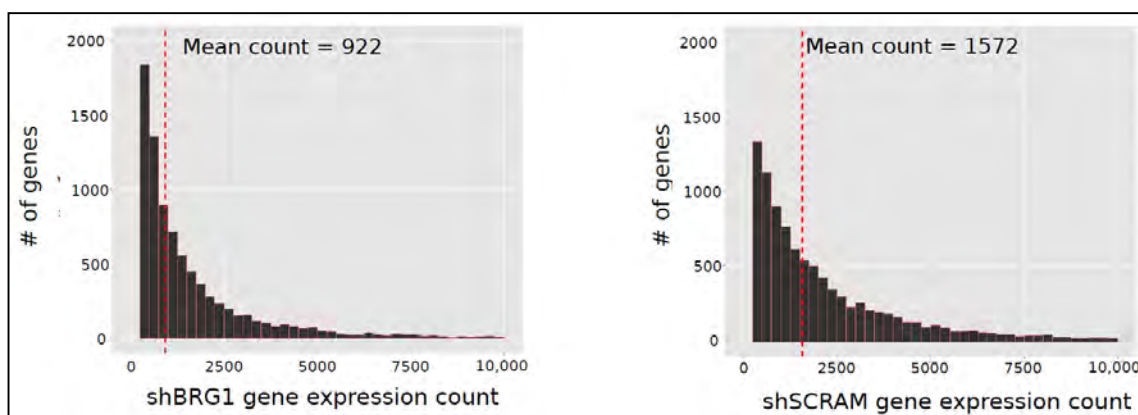


**Figure 4.1. BRG1 levels are reduced upon shRNA induction.** Western blot of the BRG1 protein levels of shSCRAM and shBRG1 MCF10A cells in the non-induced (DOX -) and induced (DOX +) conditions. Lower panel: Quantification of the western blot showing ~85% reduction of BRG1 protein levels upon doxocycline induction.

We performed polyA RNA-seq in doxocycline-induced shBRG1 and shSCRAM MCF-10A cells with two biological replicates. The mean log<sub>2</sub> gene expression values between biological replicates showed a very strong correlation (Pearson Correlation Coefficient = 0.99) both for shBRG1 and shSCRAM MCF-10A RNA-seq samples (Figure 4.2a-b). BRG1 depletion in MCF-10A cells resulted in a lower mean expression compared to control cells (Figure 4.3).

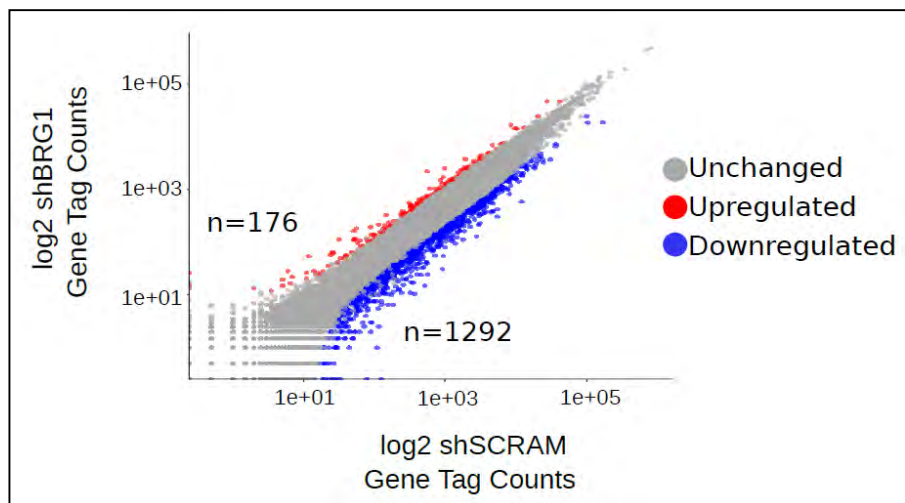


**Figure 4.2. Correlation analysis of RNA-seq replicates.** Scatterplots showing the gene expression correlation between the RNA-seq biological replicates for a) shBRG1 and b) shSCRAM samples.

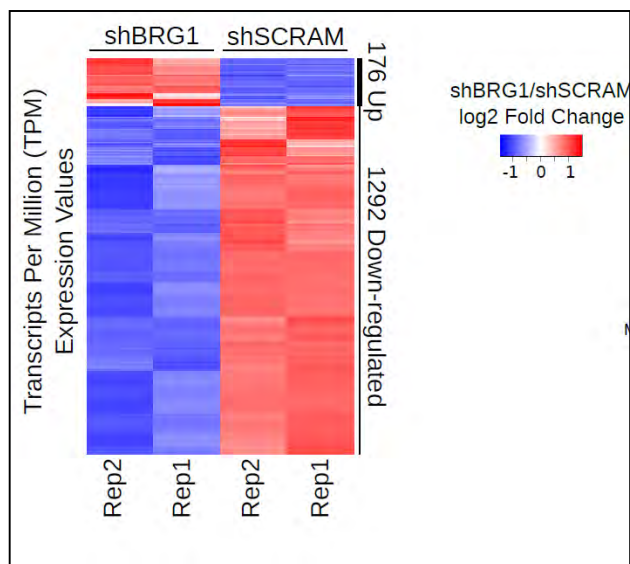


**Figure 4.3. BRG1 depletion results in a lower mean gene expression.** Histograms showing the number of genes and the expression levels for shBRG1 (left panel) and shSCRAM (right panel) combined datasets.

Consistent with this, by performing a differential gene expression analyses [209] , we identified 176 up-regulated and 1292 down-regulated genes upon BRG1 depletion ( $\log_2FC > 1$ ,  $p < 0.01$ ) that showed high reproducibility between the biological replicates (Figure 4.4 and Figure 4.5). The average expression level range of down-regulated genes was lower when compared to the average expression level range of up-regulated genes (Figure 4.6).

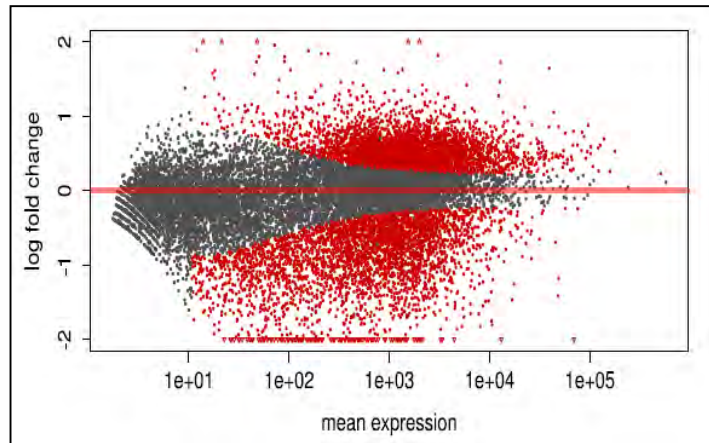


**Figure 4.4.** Scatterplot showing the  $\log_2$  gene expression values for shBRG1 and shSCRAM cells. The red and blue dots denote the up and down-regulated genes between the two conditions, respectively.



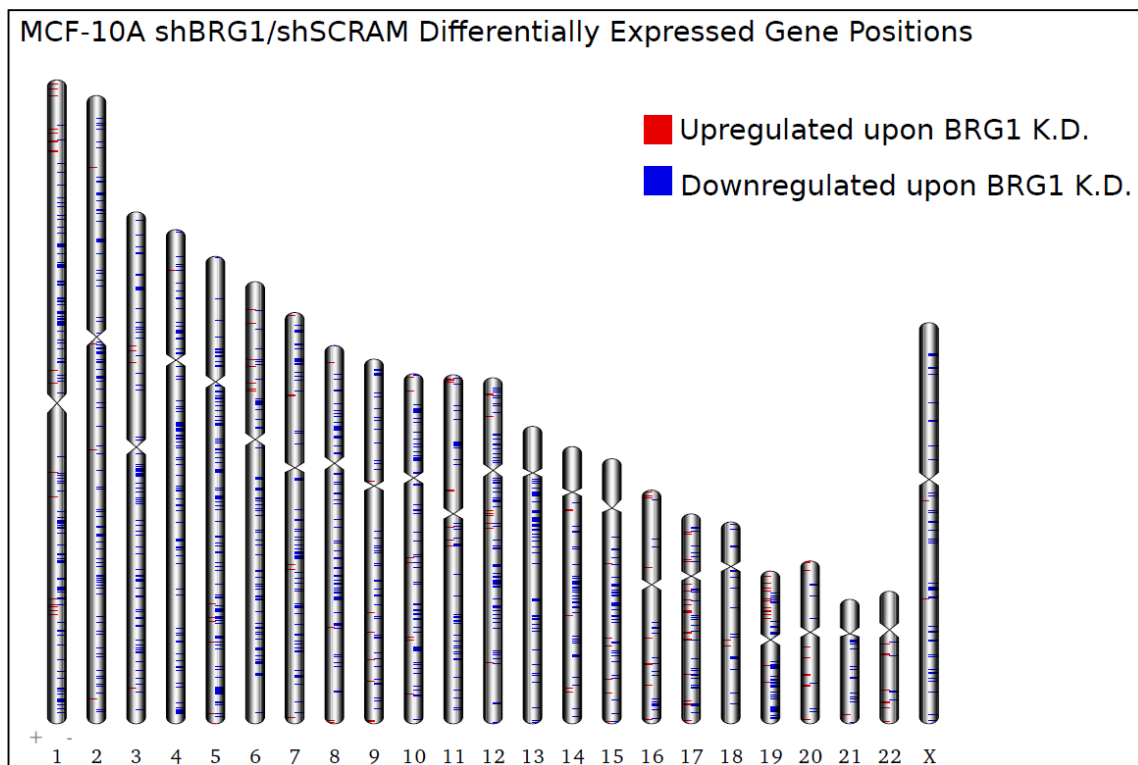
**Figure 4.5.** Heatmap showing the transcripts per million (TPM) expression values of the differentially expressed genes between shSCRAM and shBRG1 for each biological replicate.



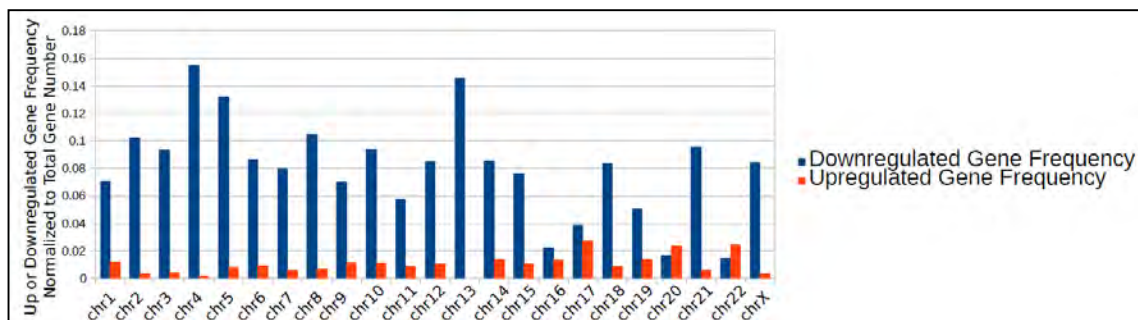


**Figure 4.6. BRG1 knockdown results in extensive transcriptional dysregulation.** MA-plot showing the log<sub>2</sub> fold change and the mean expression of significantly altered (red) and unaltered genes (black).

The genomic localization of the differentially expressed genes was similar across the chromosomes (Figure 4.7). The frequency of up and down-regulated genes showed a widespread distribution between the chromosomes, rather than localizing to a few hot spots in the genome (Figure 4.8).



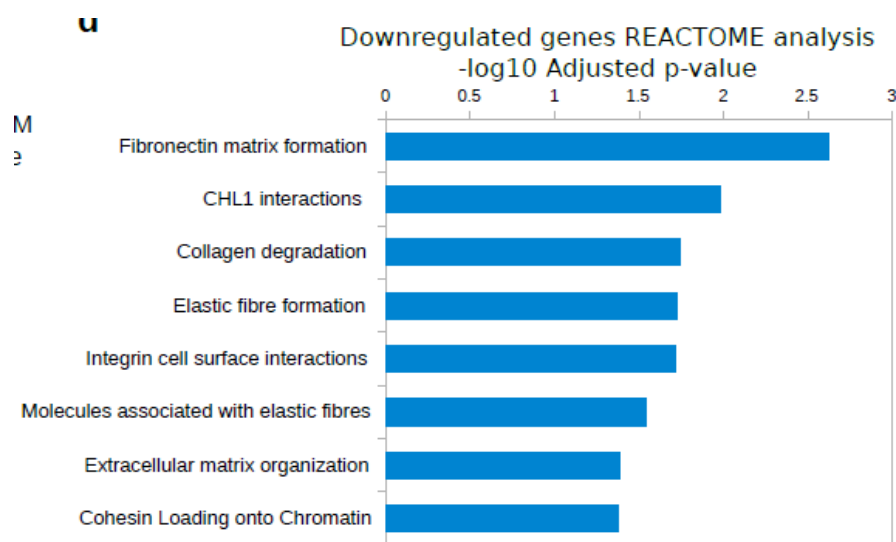
**Figure 4.7. Locations of BRG1 regulated genes.** Chromosome ideograms showing the locations of up (red) and down (blue) regulated genes upon BRG1 K.D.



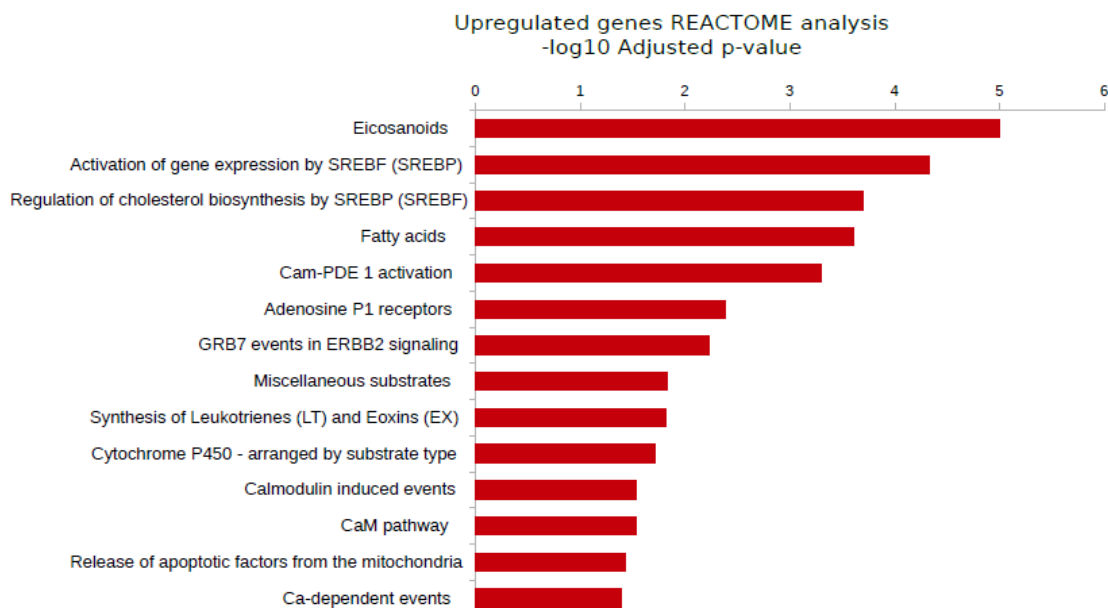
**Figure 4.8. BRG1-mediated gene regulation occurs throughout the genome.** Bargraph showing the gene frequencies of up (red) and down (blue) regulated genes for each chromosome.

To assess the biological pathways associated with the differentially expressed genes, we performed REACTOME pathway analysis [256; 257]. Pathway analysis of down-regulated genes identified pathways related to “extracellular matrix organization”, “collagen degradation”, “cell adhesion molecule L1-like (CHL1) interactions” and

“cohesin loading onto chromatin” (Figure 4.9). A significant portion of the down-regulated genes were associated with cell adhesion, including many proteoglycans, integrins and laminins. On the other hand, the pathways of up-regulated genes were associated with lipid metabolism, including pathways such as “regulation of cholesterol biosynthesis by SREBP”, “fatty acids” and “eicosanoids”, which are the byproducts of fatty acid oxidation (Figure 4.10). These pathways point out to a BRG1-mediated effect on cell metabolism.



**Figure 4.9. REACTOME terms of down-regulated genes upon BRG1 knockdown.** Bar graph showing the -log<sub>10</sub> p-values for the REACTOME terms of the 1292 genes that are down-regulated upon BRG1 knockdown.

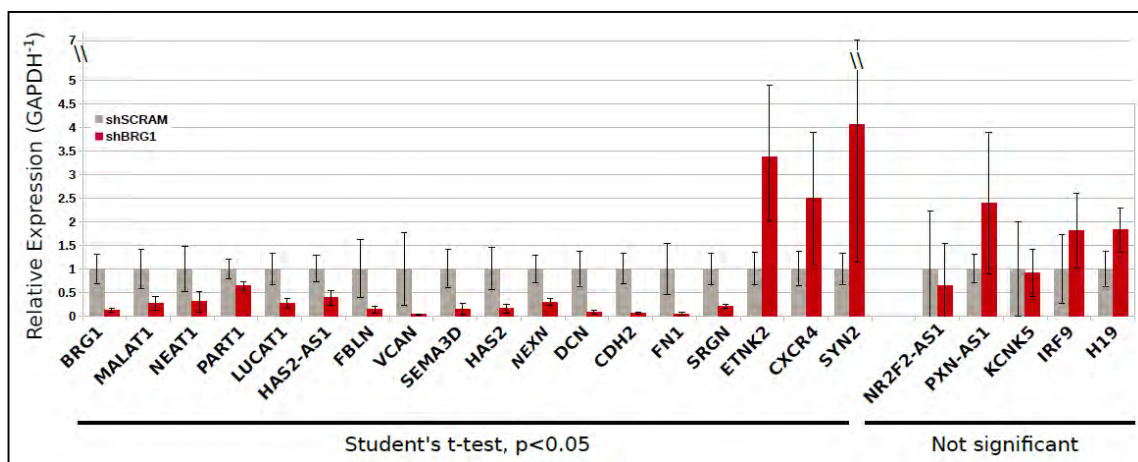


**Figure 4.10. REACTOME terms of up-regulated genes upon BRG1 knockdown.** Bar graph showing the -log<sub>10</sub> p-values for the REACTOME terms of the 176 genes that are up-regulated upon BRG1 knockdown.

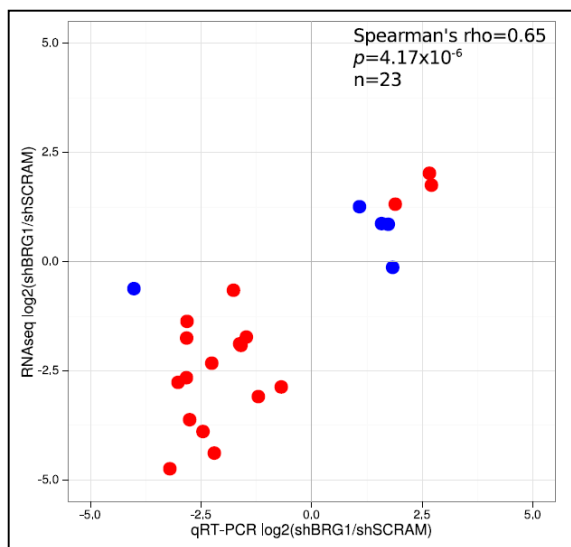
Furthermore, in order to separately assess the transcriptional changes of polyadenylated long non-coding RNAs (lncRNAs), we analyzed the differential expression between the control and BRG1 knockdown MCF-10A cells using the GENCODE v19 lncRNA gene annotation [258]. We identified 88 down-regulated and 64 up-regulated polyA lncRNAs upon BRG1 depletion, suggesting a widespread role of BRG1 in the transcriptional regulation of both coding and non-coding genes. Even though the majority of the differentially expressed lncRNAs were unannotated, we observed down-regulation of several well-known lncRNAs, including XIST, NEAT1 and MALAT1, and up-regulation of the imprinted lncRNA H19.

In order to validate the RNA-seq results, we performed qRT-PCR on 23 coding and non-coding genes of interest. The RNA-seq expression profile of 18 of 23 genes (~78%) was validated by qRT-PCR with statistical significance (Figure 4.11). Plotting

the log<sub>2</sub> ratios of qRT-PCR and RNA-seq shBRG1/shSCRAM expression levels revealed a significant correlation (Spearman's rho=0.65,  $p=4.17 \times 10^{-6}$ ) between the RNA-seq and qRT-PCR experiments (Figure 4.12).



**Figure 4.11. qRT-PCR validation of the RNA-seq data for 23 genes.** The y-axis shows the relative expression level of each gene compared to GAPDH. 18 of 23 genes showed significant differential expression.



**Figure 4.12. Scatterplot showing the correlation of log<sub>2</sub> fold change values for the 23 genes from the RNA-seq and qRT-PCR analyses.** There is a significant positive correlation between the RNA-seq and qRT-PCR data (Spearman's rho=0.65).

We conclude that BRG1 positively and negatively regulates many coding and non-coding genes and plays a role in the positive regulation of the expression of several

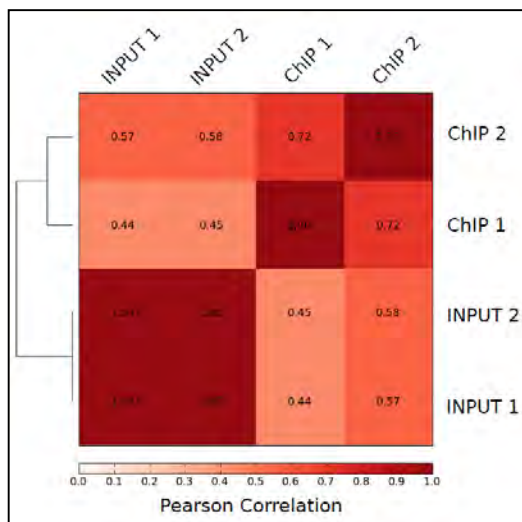
genes associated with extracellular matrix and cell adhesion in mammary epithelial cells, which supports the role of BRG1 in regulation of extracellular matrix protein coding genes [259-261]. BRG1 is also involved in the negative regulation of several pathways involved in lipid metabolism.

#### **4.3.2 BRG1 ChIP-seq analysis reveals extensive binding to intergenic and intronic regions**

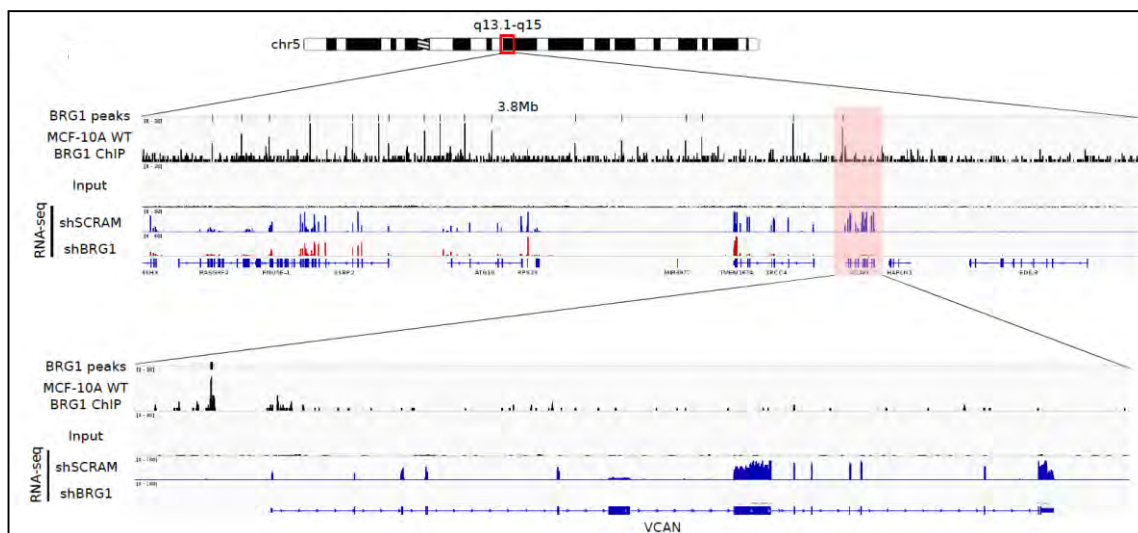
In order to expand our understanding about the function of BRG1 in mammary epithelial cells, we mapped its genome-wide localization by performing chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) in the parental MCF-10A cell line (Figure 4.13, 4.14 and 4.15). We identified 15,046 BRG1 bound regions in the genome. The binding profile of BRG1 demonstrated that BRG1 binds to defined locations in the genome (Figure 4.15).

<b>Sample Name</b>	<b>Biological Replicate</b>	<b>Raw Reads</b>	<b>Mapped Reads</b>
BRG1 ChIP 1	1	14,642,947	9,057,272
BRG1 ChIP 2	2	25,641,108	20,088,204
INPUT 1	1	25,624,664	22,301,264
INPUT 2	2	37,688,445	32,712,701

**Figure 4.13. Table showing the sequenced and mapped reads for each ChIP-seq biological replicate.**



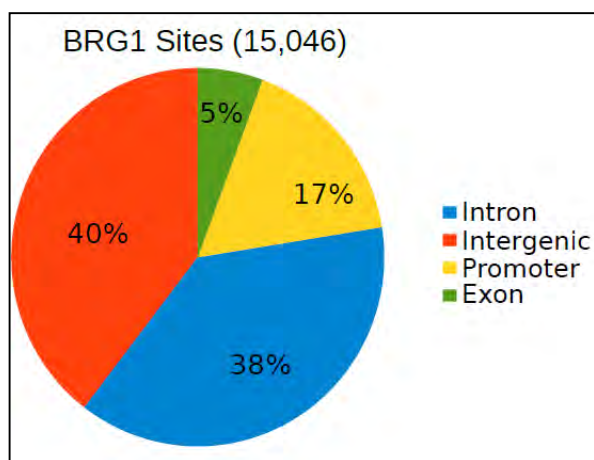
**Figure 4.14. Matrix showing the Pearson correlation of the signal intensity between the BRG1 pull down (ChIP) and input samples.**



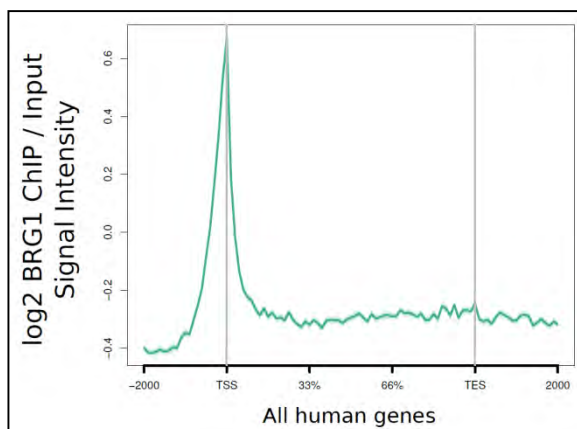
**Figure 4.15. BRG1 binds to defined locations in the genome.** Example of a ChIP-seq genome browser view of BRG1 binding and the input control, as well as the shSCRAM and shBRG1 RNA-seq on chr5, and a zoom-in on the VCAN (V-cadherin) gene in the lower panel. The y-axis represents the tag densities relative to hg19 genomic coordinates.

Annotation of BRG1 ChIP-seq peaks revealed that 60% of the binding sites were localized in promoters, introns and exons (gene bodies), whereas 40% of the sites were bound to intergenic regions (Figure 4.16). Consistent with the annotation of BRG1 peaks, normalized BRG1 binding signal across all the human genes was highest at

transcriptional start site (TSS) and promoter regions (Figure 4.17). Since BRG1 was enriched at the promoters, we asked whether its localization overlapped with Pol II binding. When the BRG1 peaks were intersected a publicly available MCF-10A Pol II ChIP-seq dataset (GSM935588), we determined that 27% of all BRG1 bound sites were also bound by Pol II (Figure 4.18).

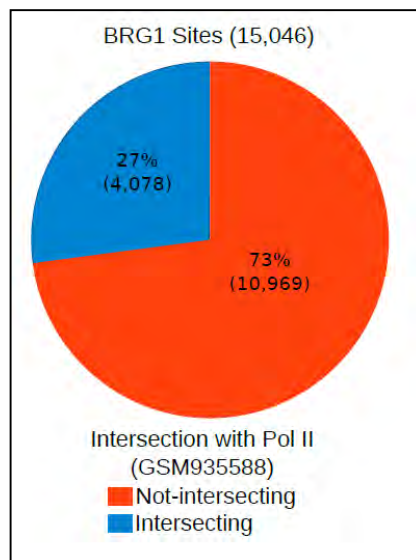


**Figure 4.16. Annotation of BRG1 peaks.** Distribution of BRG1 ChIP-seq peak annotation for genic and intergenic regions.



**Figure 4.17. BRG1 is enriched at promoter regions.** Normalized BRG1 ChIP-seq signal intensity plot for all human UCSC genes +/- 2kb. BRG1 binding is enriched at the promoter regions.

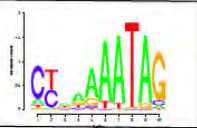
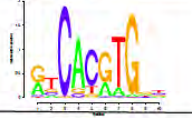
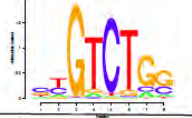
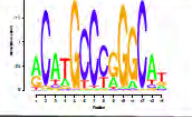
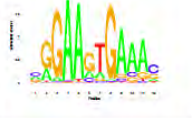




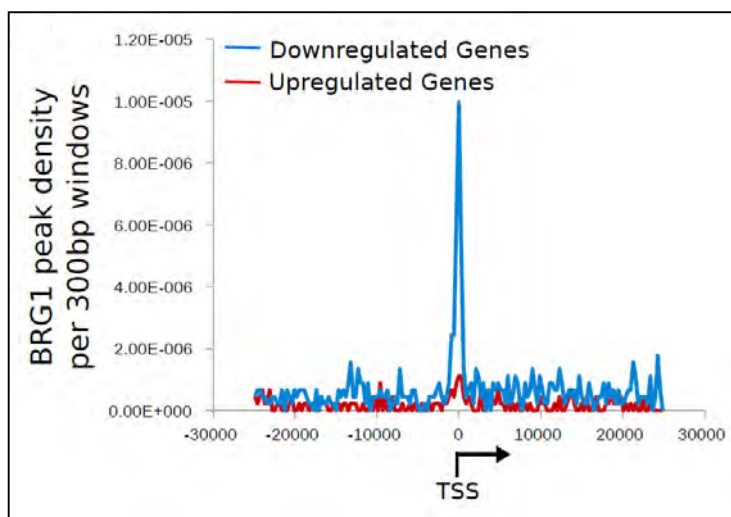
**Figure 4.18. A significant portion of BRG1 overlaps with PolII.** Pie chart showing intersections of BRG1 peaks with publicly available Pol2 peaks from MCF-10A cells.

BRG1 binds to DNA by complexing with different partner proteins. To obtain information about the context of the underlying DNA sequences of BRG1-bound regions, we performed a motif analysis on BRG1 peaks. We identified many binding motifs significantly enriched at BRG1 peaks. Among those, the five most significantly enriched motifs were MEF2A/C, USF2, SMAD2/4, p53 and PU.1 (Figure 4.19).

Next, we asked whether BRG1 binding at the promoters was associated with differentially expressed genes. To address this, we analyzed the BRG1 peak frequency at the promoters of up- and down-regulated genes. We observed a dramatic increase in the frequency of BRG1 binding at the promoters of down-regulated genes, whereas there was minimal association of BRG1 with the promoters of up-regulated genes (Figure 4.20). This result suggests a more direct role for BRG1 in the positive regulation of gene expression.

Motif	Factor	p-value
	MEF2A MEF2C	1e-642
	USF2	1e-58
	SMAD2 SMAD4	1e-53
	p53	1e-46
	PU.1	1e-26

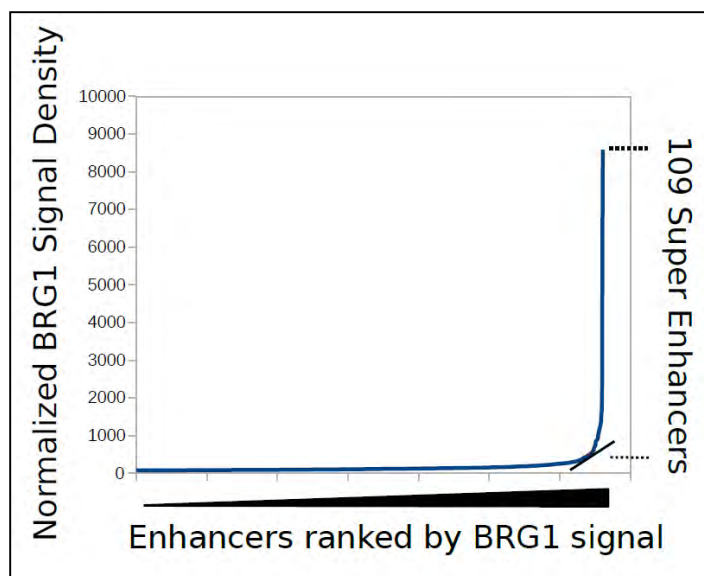
**Figure 4.19. Motif analysis of BRG1 peaks.** Top 5 sequence motifs of the BRG1 peaks.



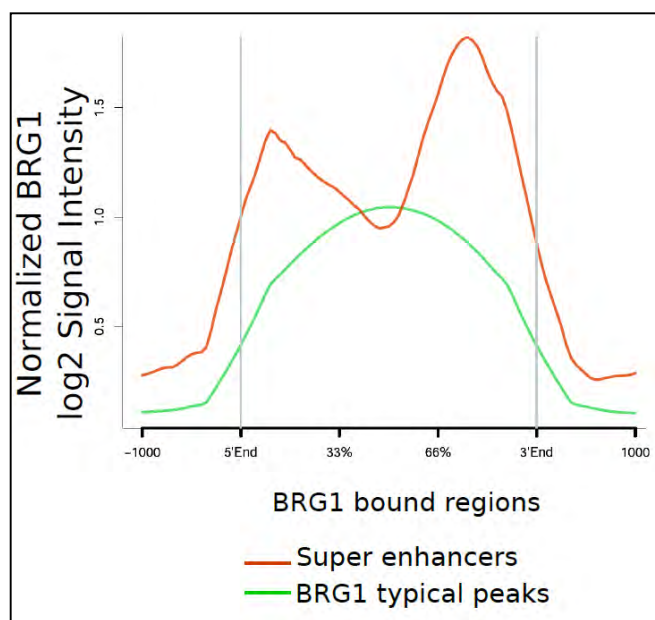
**Figure 4.20. BRG1 is enriched at the promoter regions of down-regulated genes.** BRG1 peak density within +/- 20kb of the TSS of significantly downregulated (blue), or upregulated genes (red).

Recently, super-enhancers, a novel type of regulatory regions having an unusual enrichment of transcription factors, were described [262]. They are identified by ranking the ChIP-seq peaks by signal, and taking most enriched peaks that are “stitched” over a 12kb windows. Super-enhancers are mostly associated with developmentally regulated genes. BRG1 is localized at super-enhancers in leukemic and in normal B-cells [245]. Using the approach previously published [262], we identified 109 BRG1-bound super-enhancers in the MCF-10A genome (Figure 4.21). BRG1 signal intensity was greater at super-enhancers than at other BRG1 ChIP-seq peak regions (Figure 4.22). Annotation of the super-enhancers revealed ~60% localization at intergenic regions and ~40% localization at gene bodies, but not at promoters (Figure 4.23). This localization pattern is opposite of the annotation of typical BRG1 peaks (Figure 4.16).

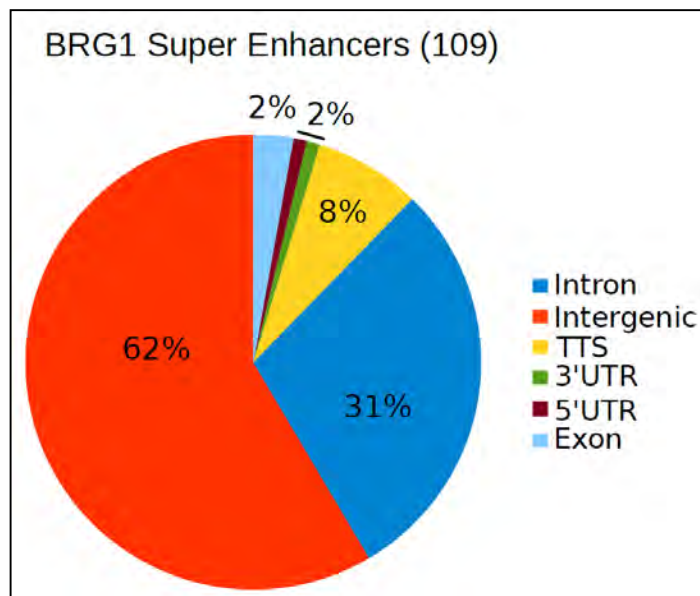
Altogether, these results suggest that BRG1 is bound to specific loci in the genome, mostly (~60%) at gene bodies. It is likely co-bound with other factors, possibly including MEF2C, USF2, SMAD2 and p53, as it exerts its remodeling function. BRG1-bound super-enhancers, on the other hand, are located mostly in intergenic regions and introns, and are not found in promoter regions.



**Figure 4.21. Identification of BRG1 bound super-enhancers.** Distribution of BRG1 ChIP-seq signal across the MCF-10A enhancers. BRG1 binding is not uniformly distributed across the enhancers, as 109 super-enhancers display high amounts of BRG1 binding.



**Figure 4.22. BRG1 signal intensity of super-enhancers.** BRG1 signal is higher over super-enhancers (red) than typical enhancers (green).



**Figure 4.23. Annotation of BRG1 bound super-enhancers.** Distribution of BRG1 bound super-enhancer annotation for genic and intergenic regions.

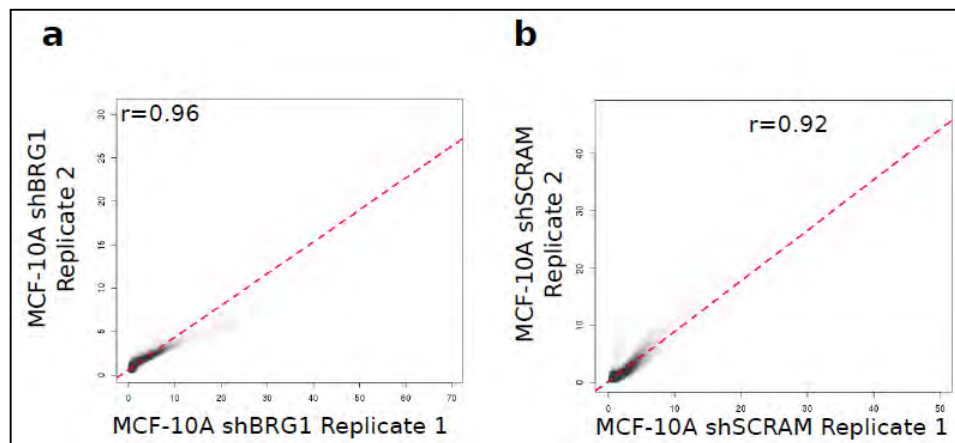
### 4.3.3 Hi-C analysis of BRG1 knockdown and control cells

To probe changes in higher-order chromatin structure upon BRG1 depletion, we performed Hi-C in doxycycline-induced shSCRAM and shBRG1 MCF-10A cells. Two independently grown and fixed batches of cells were sequenced to an average depth of ~115 million reads per replicate (Table 4.1). The sequence mapping and the initial Hi-C analysis was performed as described previously [100; 213; 263], using the Iterative Correction Method (ICE) [212] to correct for systematic biases.

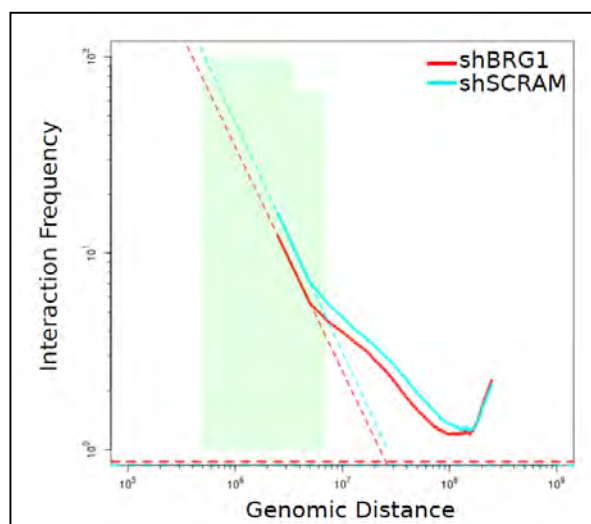
	MCF10A shBRG1 Rep1		MCF10A shBRG1 Rep2		MCF10A shSCRAM Rep1		MCF10A shSCRAM Rep2	
	Number of Reads	Percentage	Number of Reads	Percentage	Number of Reads	Percentage	Number of Reads	Percentage
<b>Mapping Statistics</b>								
side1TotalReads	126112222		139322117		68005545		129704939	
side1NoMap	21517521	(17.06%)	22969481	(16.51%)	11500278	(16.91%)	22645804	(17.46%)
side1MultiMap	4087114	(3.24%)	4293707	(3.08%)	2830203	(4.16%)	3858068	(2.97%)
side1UniqueMap	100507587	(79.70%)	112031929	(80.41%)	53675064	(78.93%)	103201067	(79.57%)
side2TotalReads	126112222		139322117		68005545		129704939	
side2NoMap	23200302	(18.40%)	25414327	(18.24%)	21318997	(31.35%)	29351593	(22.63%)
side2MultiMap	3910177	(3.10%)	4088474	(2.93%)	2068855	(3.04%)	3381613	(2.61%)
side2UniqueMap	99001743	(78.50%)	109819316	(78.82%)	44619693	(65.61%)	96971733	(74.76%)
<b>Hi-C Library Quality Metrics</b>								
totalReads	126112222		139322117		68005545		129704939	
unMapped	4625491	(3.67%)	4444297	(3.19%)	5700459	(8.38%)	5049870	(3.89%)
singleSided	43464132	(34.46%)	47904395	(34.38%)	26315415	(38.70%)	49137338	(37.88%)
bothSideMapped	78022599	(61.87%)	86973425	(62.43%)	35989671	(52.92%)	75517731	(58.22%)
sameFragment	8625171	(11.05%)	10540171	(12.12%)	7189949	(19.98%)	14374837	(19.04%)
selfCircle	4552732	(5.84%)	7430909	(8.54%)	1001252	(2.78%)	11074896	(14.67%)
danglingEnd	4033674	(5.17%)	3068729	(3.55%)	5888191	(16.36%)	3184759	(4.22%)
bounded	55362	(1.37%)	16988	(0.55%)	20229	(0.34%)	12640	(0.40%)
internal	3978312	(98.63%)	3071741	(99.45%)	5867962	(99.66%)	3172119	(99.60%)
error	38765	(0.05%)	20533	(0.02%)	300506	(0.83%)	115182	(0.15%)
validPair	69397428	(88.95%)	76433254	(87.88%)	28799722	(80.02%)	61142894	(80.96%)
cis	20629873	(29.73%)	11331458	(14.83%)	6632107	(23.03%)	8420921	(13.77%)
trans	48767555	(70.27%)	65101796	(85.17%)	22167615	(76.97%)	52721973	(86.23%)
<b>Hi-C Library Redundancy</b>								
validPair	69397428	(100%)	76433254	(100%)	28799722	(100%)	61142894	(100%)
totalMolecules	69397428	(100%)	76433254	(100%)	28799722	(100%)	61142894	(100%)
redundantInteractions	27456057	(39.56%)	7727938	(10.11%)	12573823	(43.66%)	20639957	(33.76%)
<b>nonRedundantInteractions</b>	<b>41941371</b>	<b>(60.44%)</b>	<b>68705316</b>	<b>(89.89%)</b>	<b>16225899</b>	<b>(56.34%)</b>	<b>40502937</b>	<b>(66.24%)</b>
percentRedundant	39.56%		10.11%		43.66%		33.76%	

**Table 4.1. Sequencing and mapping statistics of the shBRG1 and shSCRAM Hi-C biological replicates.**

The biological replicates showed a high degree of correlation (Figure 4.24). The scaling plot curves of genomic interaction frequencies along genomic distance showed similar trends of decay, and an increase at distances larger than >200Mb, which we previously showed to be a characteristic of the MCF-10A genome [263] (Figure 4.25).

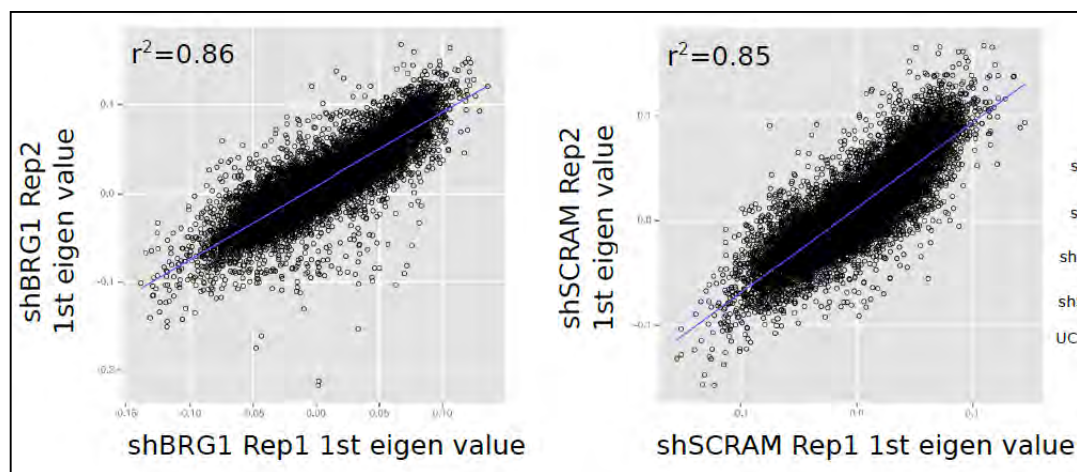


**Figure 4.24. shBRG1 and shSCRAM Hi-C data are reproducible.** Scatter plots comparing normalized interactions between pairs of 2.5Mb bins in the two biological replicates from **a)** shBRG1 and **b)** shSCRAM datasets.

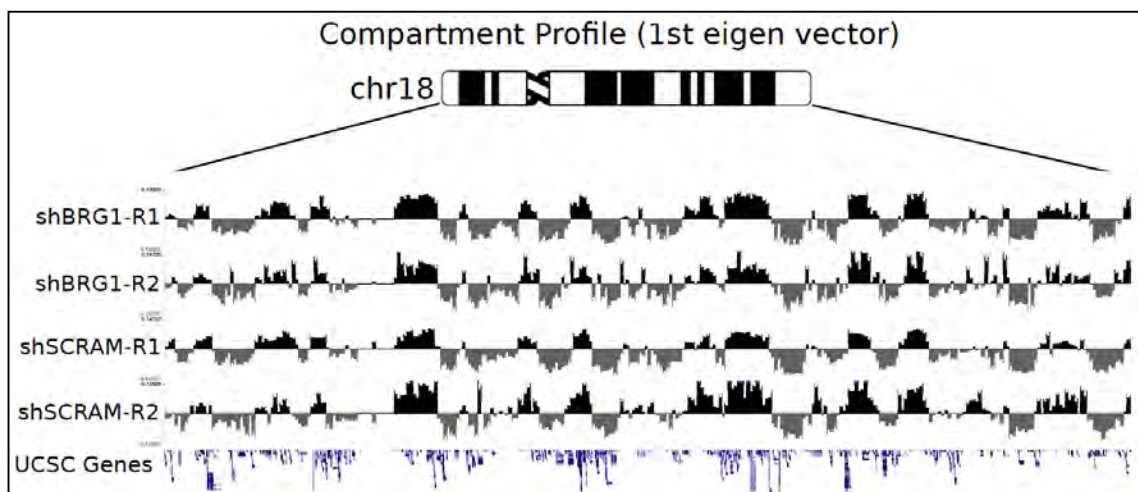


**Figure 4.25. Scaling plot analysis of shBRG1 and shSCRAM Hi-C samples.** Interaction frequency decreases similarly as a function of genomic distance. The difference between the shBRG1 and shSCRAM curves is due to the differences in % cis interactions.

Moreover, correlation between the genomic compartments (Figure 4.26 and Figure 4.27) and insulation plots to assess TADs also showed high correlation (Figure 4.28a and Figure 4.28b).

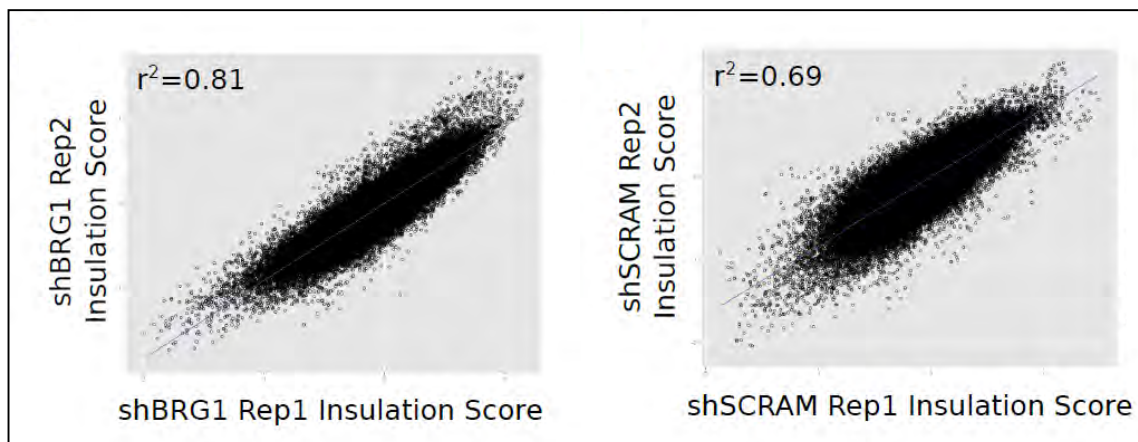


**Figure 4.26. Compartment analysis for shBRG1 and shSCRAM Hi-C replicates.** Scatter plots showing the correlation of the 1st eigenvector values for each 250kb bin from the compartment analysis for each biological replicate in shBRG1 (left panel) and shSCRAM (right panel) datasets.

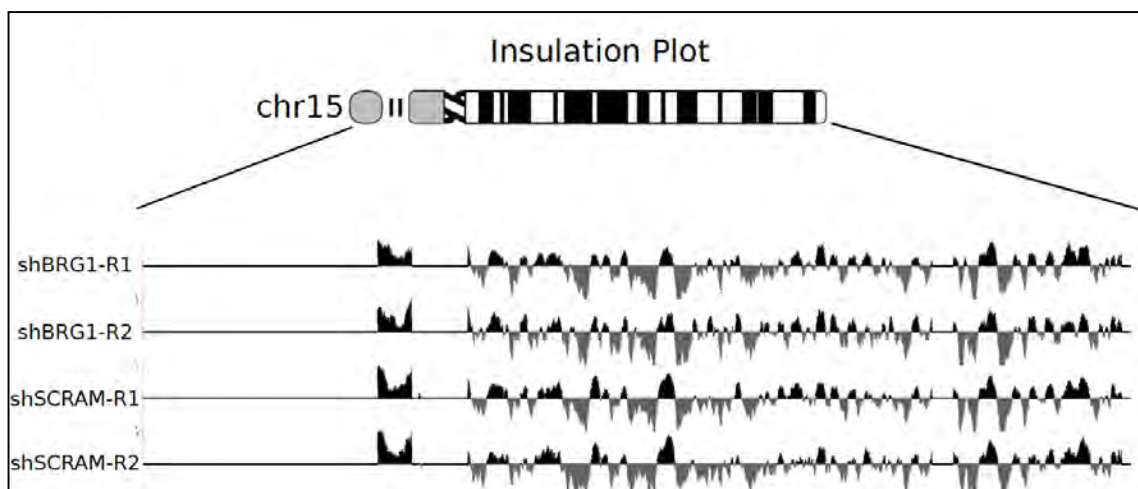


**Figure 4.27. Eigenvector analysis for shBRG1 and shSCRAM Hi-C replicates.** Example of the first eigen values for each biological replicate across chromosome 18. The replicates show high correlation in genomic compartmentalization.





**Figure 4.28a. Correlation analysis of Hi-C insulation plot replicates.** Insulation score correlation among the biological replicates for each 40kb bin for shBRG1 (left) and shSCRAM (right) datasets.

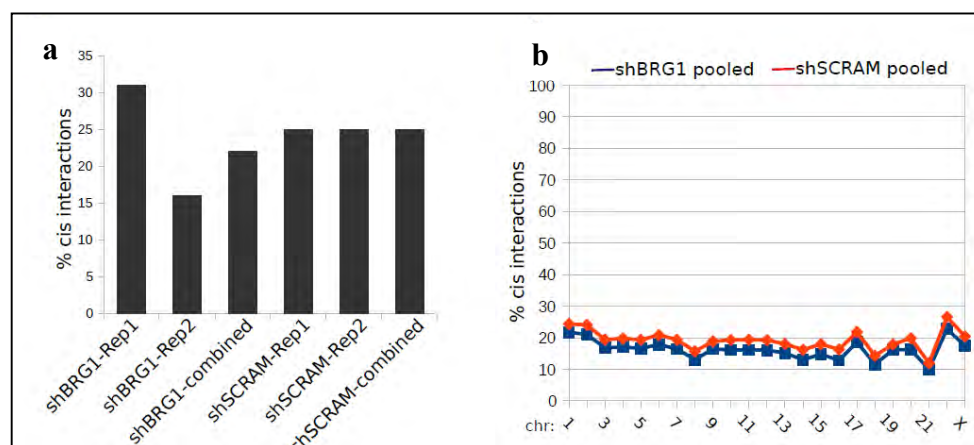


**Figure 4.28b. Insulation plot analysis of Hi-C replicates.** Example of the insulation plot across chromosome 15 for each biological replicate.

Therefore, we pooled the biological replicates for further downstream analyses. The pooled shSCRAM and shBRG1 Hi-C datasets also showed a high correlation with the parental (wildtype) MCF-10A Hi-C data [263] (Figure 4.29). The pooled Hi-C datasets displayed similar cis/trans interaction ratios (Figure 4.30).

Pearson Correlation Between Pooled Replicates (R-value)	
shSCRAM vs. shBRG1	0.984
shSCRAM vs. parental (Barutcu et al. 2015)	0.959
shBRG1 vs. parental (Barutcu et al. 2015)	0.973

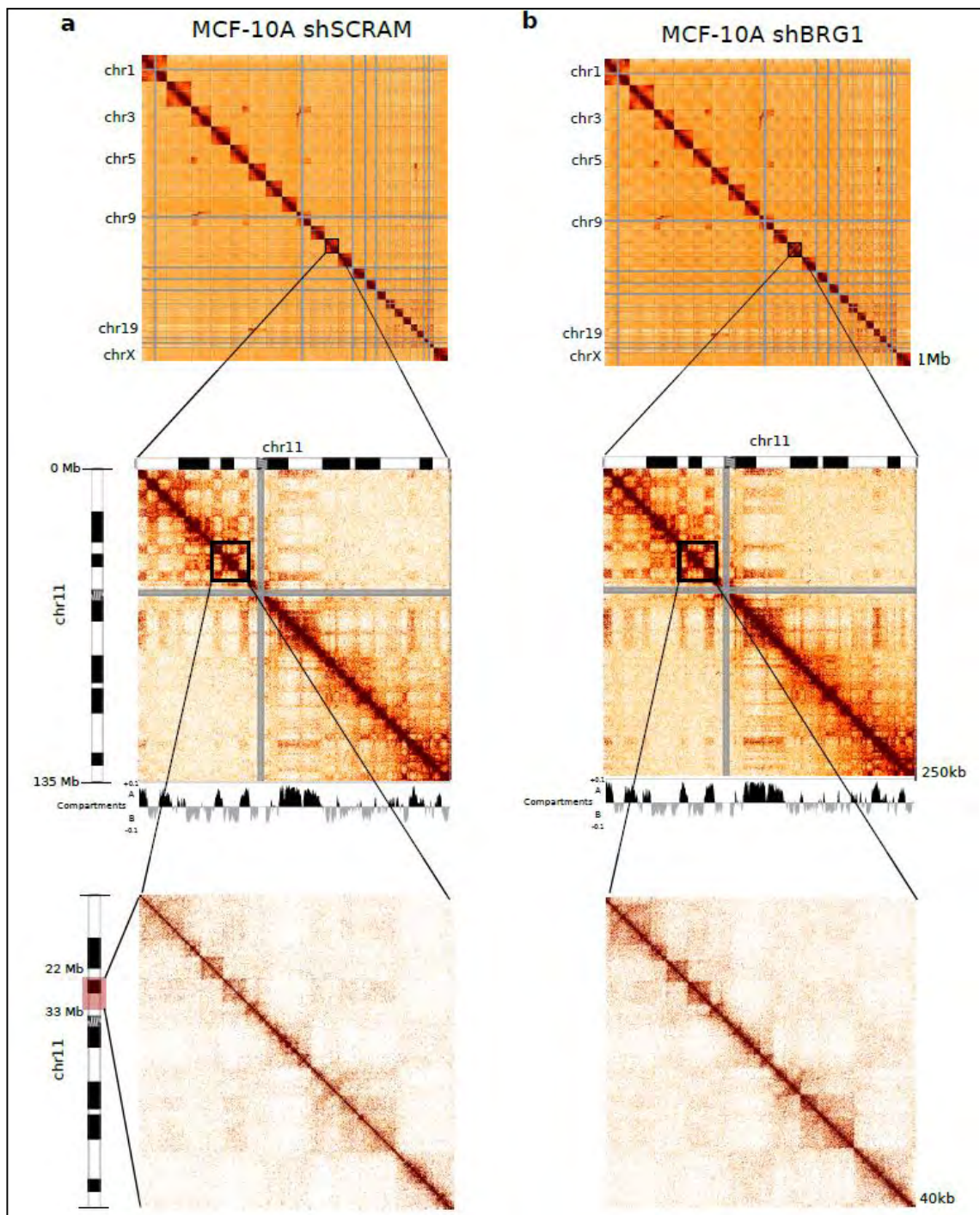
**Figure 4.29. shBRG1 and shSCRAM Hi-C data shows high correlation with the parental MCF-10A cells.** Pearson correlations of the pooled Hi-C replicates between shSCRAM, shBRG1, and previously published parental (wildtype) MCF-10A cells.



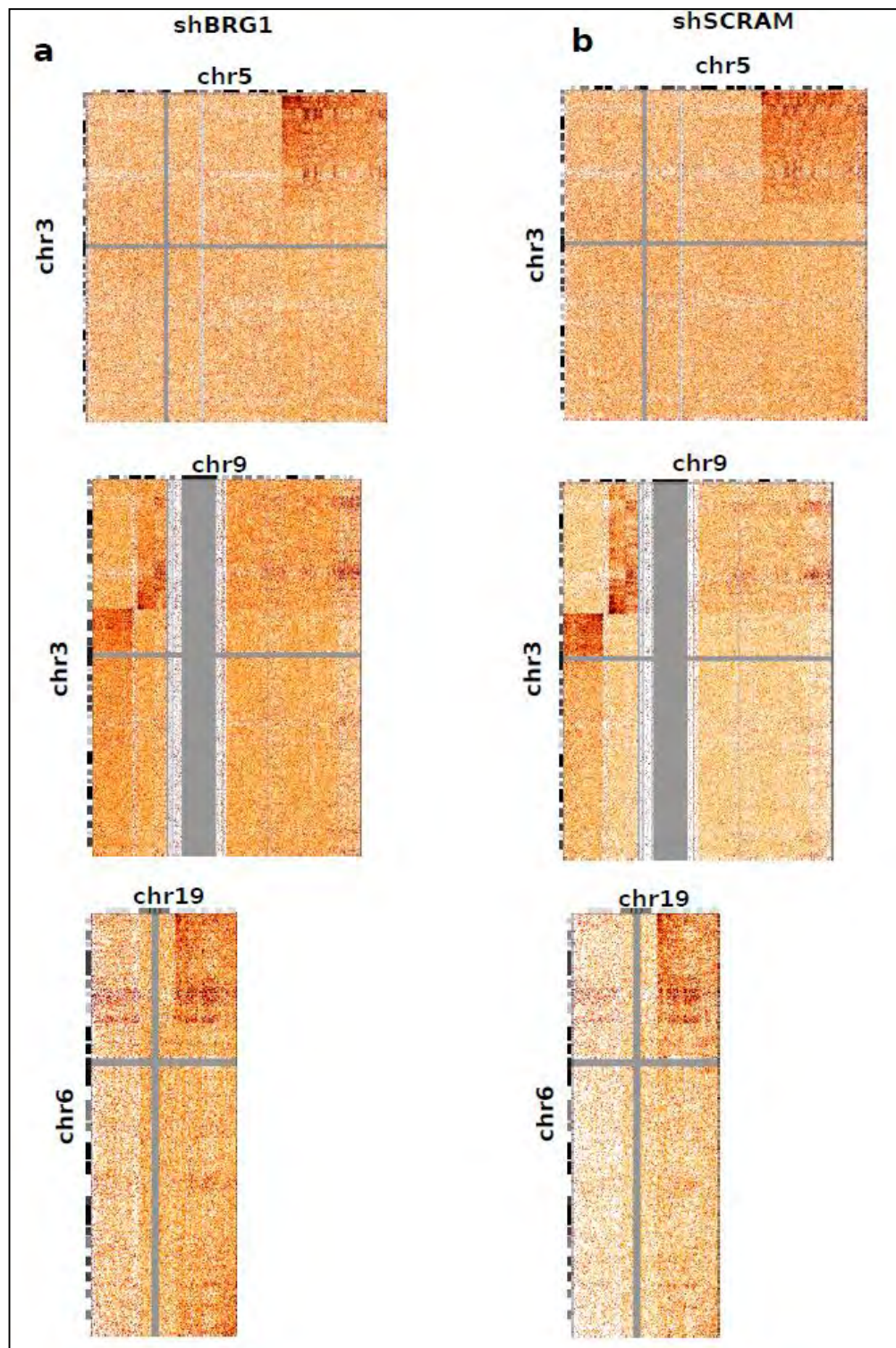
**Figure 4.30. cis/trans ratio of the Hi-C replicates. a)** Bargraph showing the cis interaction percentage of individual and pooled shBRG1 and shSCRAM Hi-C replicates. **b)** The percentage of cis/trans interaction frequencies of pooled shBRG1 and shSCRAM datasets for each chromosome.

Genome-wide Hi-C interaction heatmaps showed that, consistent with previous Hi-C studies and the notion of chromosome territories [23], the *intra*-chromosomal interactions are more frequent than *inter*-chromosomal interactions, which are visualized as black boxes along the diagonal (Figure 4.31). Moreover, we identified large blocks

of *inter*-chromosomal interactions between chromosomes 3 and 5, chromosomes 3 and 9, and chromosomes 6 and 19, which represent previously known translocations in the MCF-10A genome [44; 263] (Figure 4.32). Visualizing the Hi-C heatmaps from the shSCRAM and shBRG1 datasets at increasing resolutions revealed hierarchical higher-order chromatin structures such as genomic compartments and TADs (Figure 4.31).



**Figure 4.31. Hi-C data reveals genomic compartments and TADs.** Genome-wide all by all Hi-C interaction heatmaps at 1Mb resolution and a zoom-in of chr11 at 250kb resolution (middle panel) and 40kb resolution (lower panel) in **a**) MCF-10A shSCRAM and **b**) MCF-10A shBRG1 cells. For the genome-wide heatmaps, the chromosomes are stacked from top-left to bottom-right in order (chr1, chr2...chr22 and chrX). The gray regions indicate repetitive regions (such as centromeres) in which the sequencing reads could not be mapped.



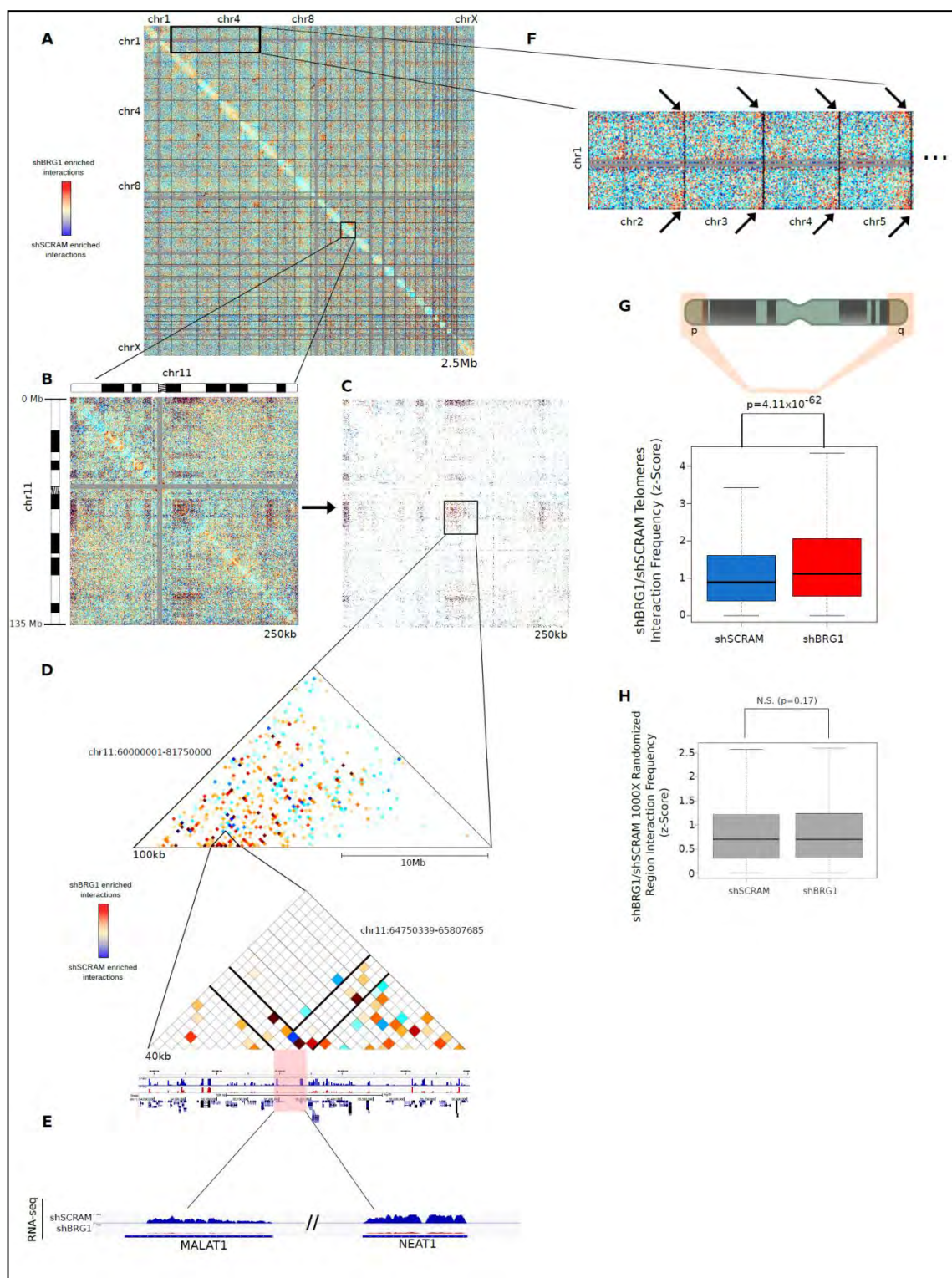
**Figure 4.32. Translocations in the MCF-10A genome.** Inter-chromosomal interaction heatmaps showing the translocated regions in the MCF-10A genome for **a)** shBRG1 and **b)** shSCRAM Hi-C datasets.

#### **4.3.4 BRG1 knockdown results in extensive gain and loss of long-range chromatin interactions and altered genome organization**

To map BRG1-mediated alterations in higher-order chromatin structure, we compared the genome-wide interactions as previously described [213; 263] (Figure 4.33a-e). At multiple resolutions, BRG1 depletion resulted in the widespread disruption of existing interactions and the emergence of novel specific contacts (Figure 4.33a-e) throughout each chromosome (Figure 4.33b and Figure 4.34). Mapping the high-confidence interactions that are depleted/enriched upon BRG1 knockdown revealed specific regions potentially important for transcriptional regulation (Figure 4.33d-e). For instance, a zoom-in view of the differential interactions at the promoter regions of MALAT1 and NEAT1 lncRNAs, which are down-regulated upon BRG1 loss, displayed several differentially interacting regions upon BRG1 knockdown (Figure 4.33d-e).

To analyze the global effects of BRG1 depletion in higher-order chromatin organization, we visualized the genome-wide interaction matrices in a chromosome by chromosome manner and asked whether chromosomal organization was similar in BRG1 depleted cells. Consistent with earlier work [23], the smaller, gene-rich chromosomes interacted more frequently with each other in both shBRG1 and shSCRAM samples (Figure 4.35a-b). Therefore, BRG1 knockdown does not result in extensive chromosomal reorganization in MCF-10A mammary epithelial cells (Figure 4.35a-b). Analyzing the *intra* and *inter*-chromosomal interaction frequencies, we observed that the frequency of significant *inter*-chromosomal interactions was higher in shBRG1 cells (Figure 4.36-a-b). Moreover, in the context of *cis*-interactions, BRG1 knockdown resulted in significant interactions that are more frequent between *intra*-

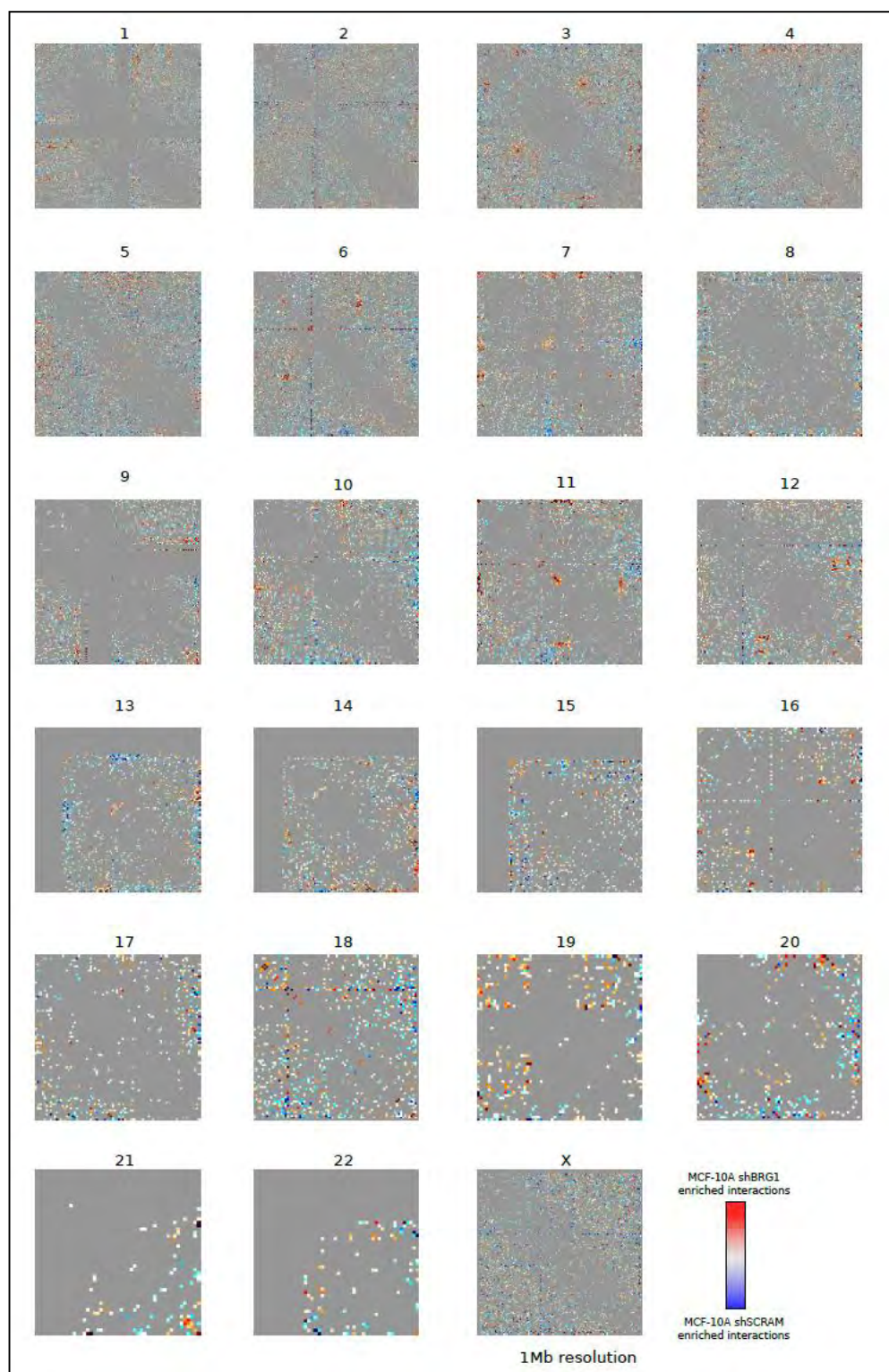
chromosomal arm and *inter*-chromosomal arms (Figure 4.36-a-b).



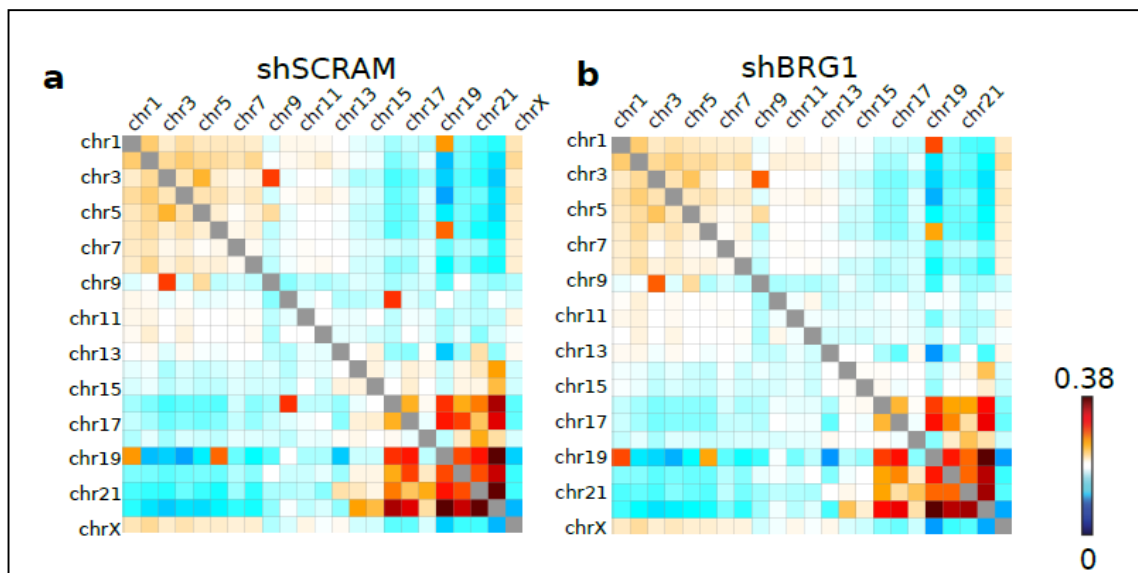


**Figure 4.33. BRG1 knockdown results in extensive changes in long-range interactions.** **a)** Genome wide interaction heatmap at 2.5Mb resolution showing the differences between interactions that are gained and lost upon BRG1 depletion. The chromosomes are stacked from top-left to bottom-right in order (chr1, chr2...chr22 and chrX). **b)** A zoom in of chr1 1 at 250kb resolution showing the all differential interactions, and **c)** the interactions that are altered with significance. **d)** A further zoom in view of a genomic region on chr1 1 (chr1 1:60000001-81750000) (top panel), where the MALAT1 and NEAT1 loci reside (chr1 1:64750339-65807685), showing the differential interactions. **e)** RNA-seq tracks of shBRG1 and shSCRAM cells showing a reduction of expression in NEAT1 and MALAT1 lncRNA genes upon BRG1 loss. **f)** A zoom-in of the *inter*-chromosomal interactions between chr1 and chr2 through chr5, with arrows indicating the enriched telomeric interactions in the shBRG1 cells. This pattern of telomeric / sub-telomeric interaction occurs throughout the genome. **g)** Quantification of the interactions among subtelomeric ends for shSCRAM and shBRG1 Hi-C datasets. The subtelomeric ends show significantly higher frequency of interactions, compared to **h)** 1000X randomized set of regions. p-values: Wilcoxon rank-sum test.

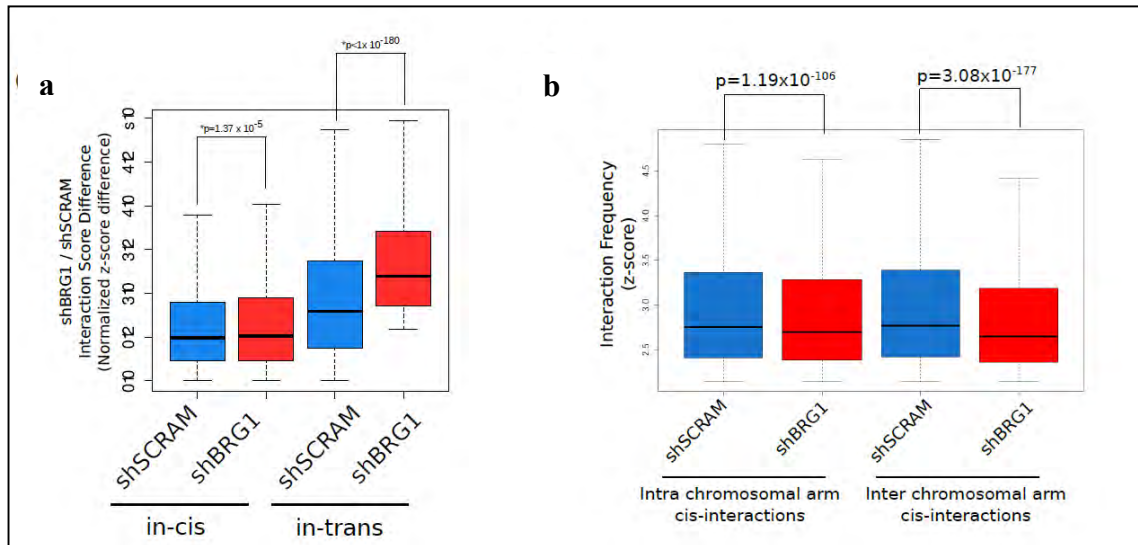
Visualizing the chromosomal interactions in higher resolution (Figure 4.33f), we observed a systematic pattern of increased interactions in shBRG1 cells, both *in-cis* and *in-trans*, among the sub-telomeric regions of the chromosomes (Figure 4.33f). In other words, the sub-telomeric regions of each chromosome displayed a striking enrichment of interactions with each other upon BRG1 knockdown compared to shSCRAM control (Figure 4.33f). Quantification of the sub-telomeric interactions suggested a significant increase in both *intra*- and *inter*-chromosomal associations in shBRG1 cells compared to shSCRAM (Figure 4.33g). However, this was not the case when interactions in the same regions were randomized (Figure 4.33h). Taken together, these results indicate a novel role for BRG1 in telomere structure and suggest that disruption of BRG1 levels results in altered three-dimensional organization of telomeric regions of the genome.



**Figure 4.34. shBRG1 / shSCRAM interactions that are significantly different at 1Mb for each chromosome.**



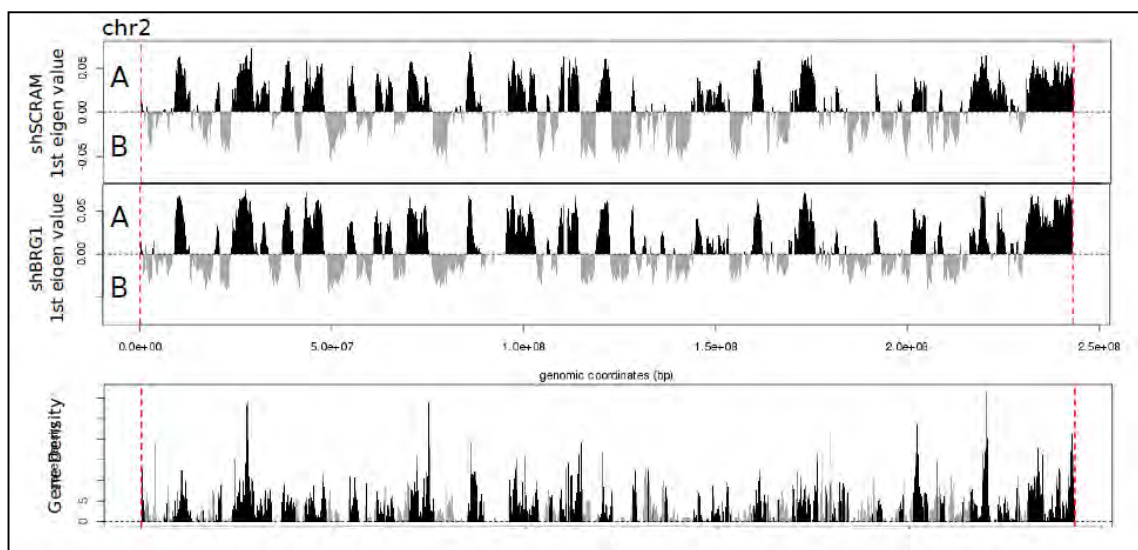
**Figure 4.35. Chromosome positioning is not altered upon BRG1 knockdown.** Chromosome by chromosome interaction heatmaps for a) shSCRAM and b) shBRG1 Hi-C datasets.



**Figure 4.36. BRG1 knockdown results in alterations in higher-order genome organization.** a) Boxplot showing the shBRG1 / shSCRAM interaction difference scores for cis and trans-interactions. shBRG1 MCF-10A cells show higher frequency of trans-interactions. b) Boxplot showing the intra-arm and inter-arm chromosomal interaction frequencies for shBRG1 and shSCRAM datasets. p-values: Wilcoxon rank-sum test.

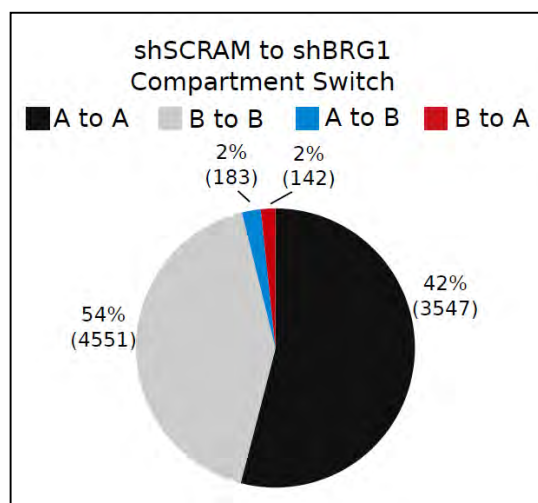
### 4.3.5 BRG1 occupancy is enriched at open compartment regions

Each chromosome territory is composed of megabase-scale genomic compartments that are either A-type (i.e open, gene rich) or B-type (i.e closed, gene poor). The frequency of interactions within one compartment occurs much more frequently than the interactions in between compartments [71]. We asked whether BRG1 depletion resulted in any compartment change. To address this, we binned the genome at 250kb non-overlapping intervals and compared the type of compartmentalization for each bin (Figure 4.37).



**Figure 4.37. Compartment profiles (the first principal components) of shSCRAM and shBRG1 data for chromosome 2.** The A-type (open) compartments are shown in black, and the B-type (closed) compartments are shown in grey. The same color scheme was used for the gene density plot for chromosome 2 in the lower panel.

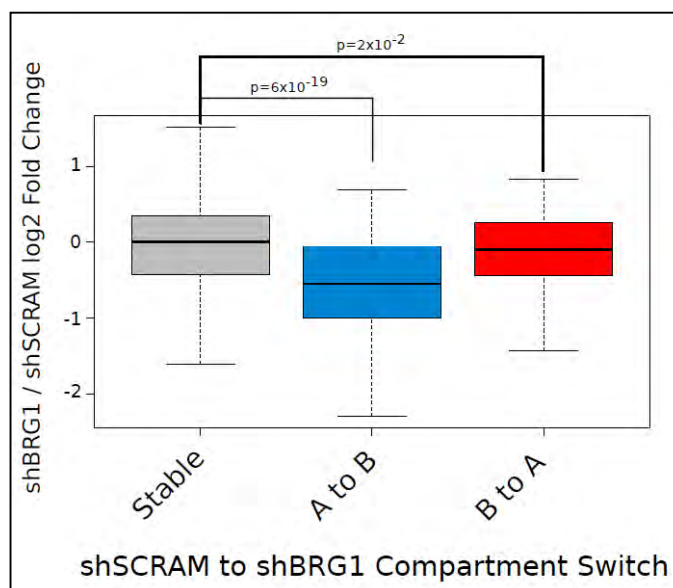
The majority of the compartmentalization was similar in shSCRAM and shBRG1 MCF-10A cells, with 42% of the genome consisting of A-type compartments and 54% consisting of B-type compartments (Figure 4.38). Upon BRG1 depletion, a total of 2% of the genome altered its compartmentalization from A-type to B-type and 2% showed alteration from B-type to A-type (Figure 4.38). Compartmentalization of the genome is correlated with gene expression [77; 263].



**Figure 4.38. A subset of compartments are altered upon BRG1 knockdown.** Pie chart showing the genomic compartment changes between shSCRAM and shBRG1 datasets. “A” and “B” denotes the open and closed compartments, respectively. “A to A” represents compartments that are open in both cell lines, “B to B” represents compartments that are closed in both cell lines, “A to B” denotes compartments that are open in shSCRAM but closed in shBRG1, and “B to A” denotes compartments that are closed in shSCRAM and open in shBRG1.

To understand the link between compartment switching and gene expression upon BRG1 depletion, we plotted the shBRG1 / shSCRAM log<sub>2</sub> fold change RNA-seq expression levels of the genes that were located either within unchanged compartments or within compartment switch regions (Figure 4.39). The genes located in regions with a compartment switch from A-type to B-type upon BRG1 knockdown showed significantly lower expression levels than the genes within unaltered compartment

regions. Conversely, genes within B-type to A-type compartment switch regions (shSCRAM to shBRG1), even though significant, showed similar expression levels upon BRG1 knockdown (Figure 4.39). These results, consistent with previous work [77; 263], suggests a prominent correlation between differential compartmentalization on gene expression.

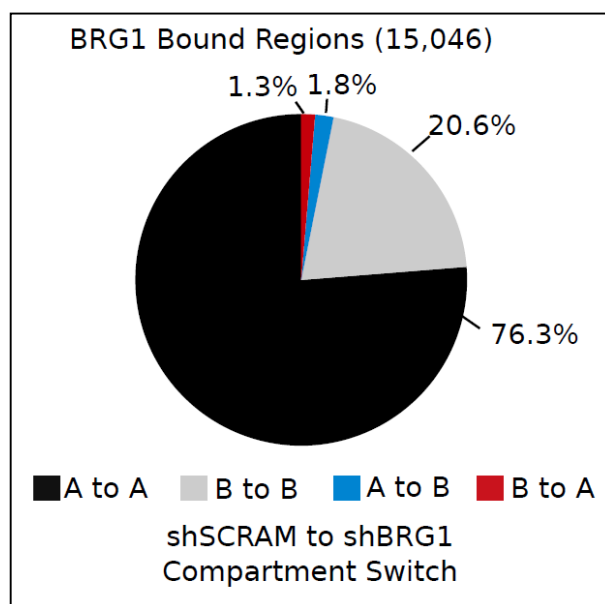


**Figure 4.39. Alterations in genomic compartments are associated with gene expression.** shBRG1 / shSCRAM log<sub>2</sub> fold change RNA-seq expression boxplot of all the genes residing at regions for different compartmental switch categories. The compartments that are switched from A to B and from B to A show significantly decreased and increased expression levels, respectively. p-value: Wilcoxon rank-sum test

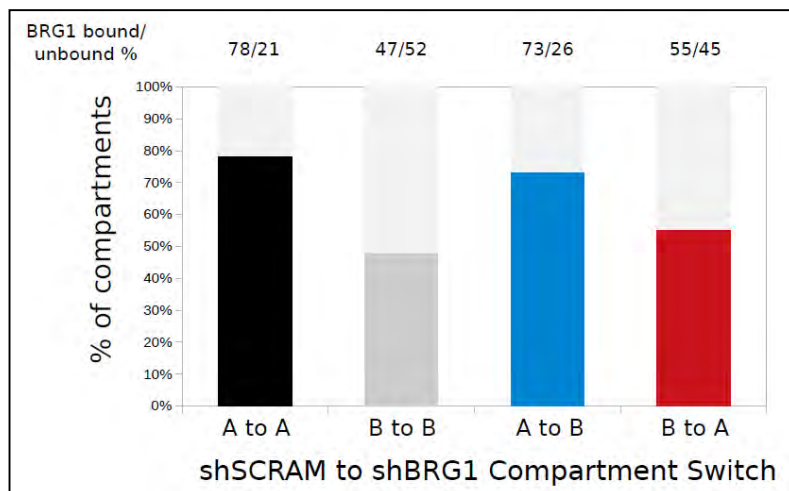
Next, we asked whether the regions with BRG1 binding were associated with compartment switching. The majority (76%) of all BRG1-bound sites were located within the constitutive open A-type compartments. In contrast, a smaller fraction (21%) of BRG1 ChIP-seq peaks were found in the closed B-type compartment regions (Figure 4.40). Only a small percentage of BRG1 peaks were in regions showing compartment switching (Figure 4.40). We then assessed the percentage of genomic compartment

switching regions that were either bound or not bound by BRG1. We observed that BRG1 was bound to ~80% of constitutive A-type compartments (A to A) and ~50% of constitutive B-type compartments (Figure 4.41). The frequency of BRG1 binding to altered compartment regions were similar as; 75% of “A to B” and ~55% of “B to A” compartment switch regions showed BRG1 binding.

Taken together, these results suggest that BRG1 binding is more enriched at open compartments. In addition, this also suggests that the depletion of BRG1 results in a modest level of compartment switching from both A-type to B-type and B-type to A-type for a subset of regions.



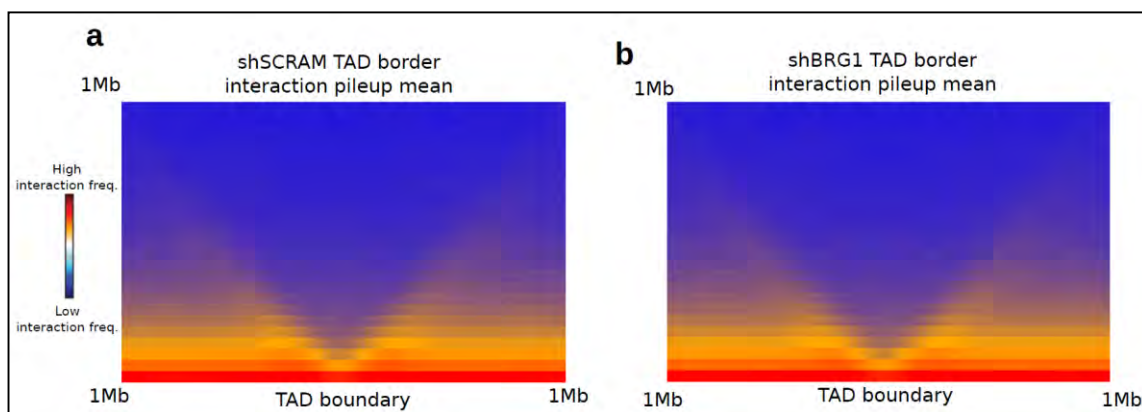
**Figure 4.40. BRG1 is bound mostly on A-type compartments.** Pie chart showing the compartment change profiles of BRG1 bound regions.



**Figure 4.41. BRG1 is bound to compartment switch regions at similar levels.** Bargraph showing the percentage of the compartment switch regions that are bound by BRG1. The colored portions of the graph denotes the BRG1 bound percentage of each compartment change category.

#### 4.3.6 BRG1 is associated with TAD boundaries and TAD boundary strength

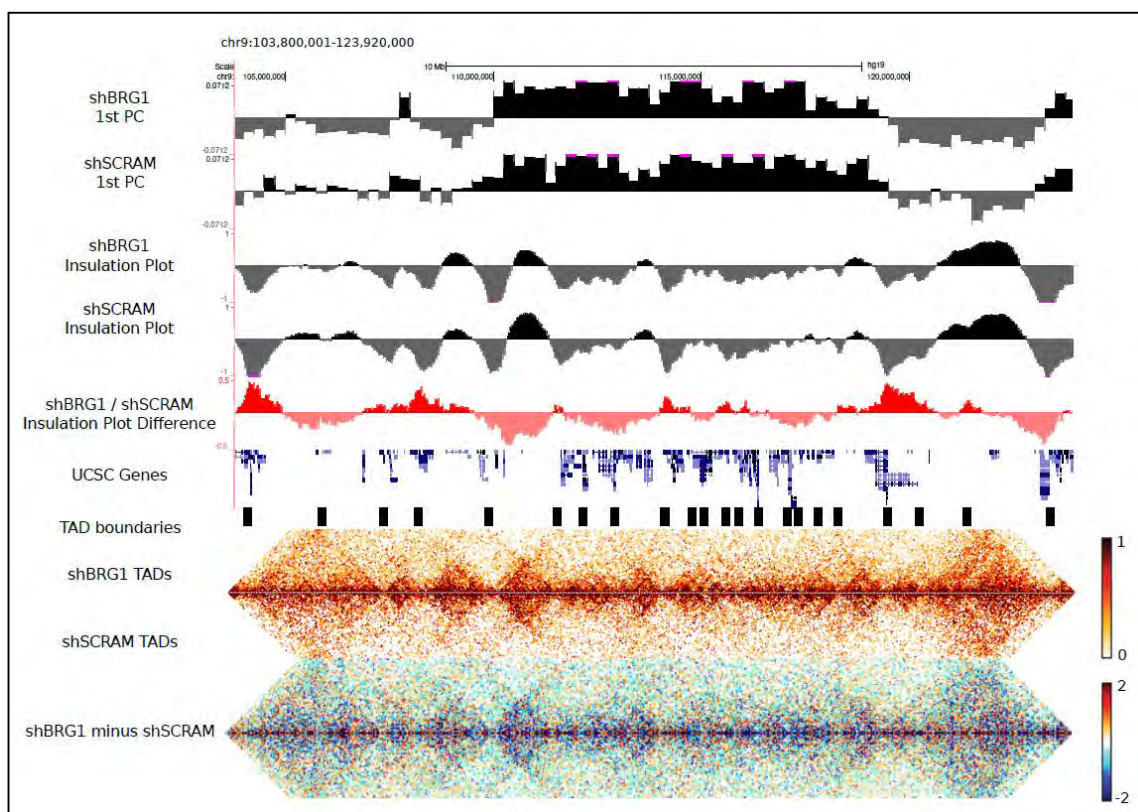
Each compartment is composed of TADs, which are sub-megabase scale structures constituting a confined nuclear micro-environment for the proper association and regulation of promoters and enhancers [75]. The 40kb resolution MCF-10A shSCRAM and shBRG1 interaction maps revealed these sub-megabase scale interaction domains on all chromosomes, suggesting that BRG1 depletion does not result in a loss of TAD formation (Figure 4.42a-b and Figure 4.43).



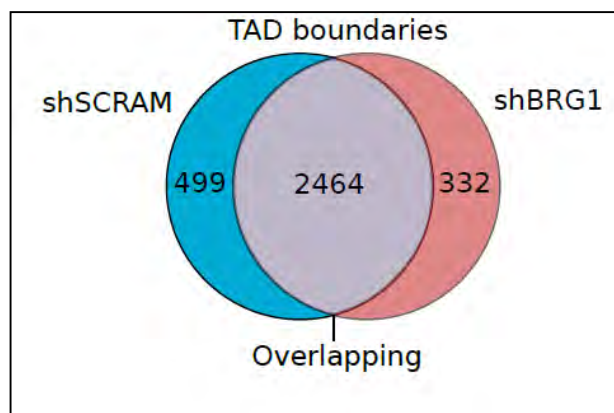
**Figure 4.42. Pile-up of interactions confirm TAD boundaries.** Mean of pile up interaction frequencies 1Mb around the TAD boundaries for **a)** shSCRAM and **b)** shBRG1 cells



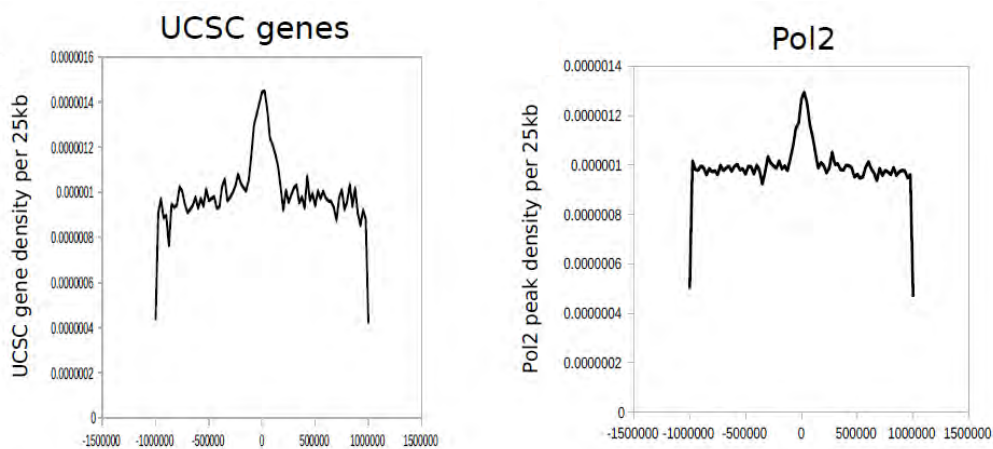
In order to quantify the TAD boundary scores and identify specific TADs, we assigned an insulation score for each genomic interval along the genome using a method that was described previously [213; 263]]. We identified 2963 and 2796 TAD boundaries in shSCRAM and shBRG1 MCF-10A cells, respectively (Figure 4.44). The identified TADs showed known characteristics, such as enrichment of genes and Pol II binding at the boundaries compared to the surrounding regions (Figure 4.45a-b).



**Figure 4.43. TADs are mostly stable upon BRG1 knockdown.** An example of a region on chromosome 9 (chr9:103800001-123920000) showing (from top to bottom) the compartment profiles of shBRG1 and shSCRAM at 250kb intervals, the insulation plot profiles at 40kb intervals (see Methods), the insulation plot difference between shBRG1 and shSCRAM, hg19 UCSC genes, TAD boundaries, shBRG1 and shSCRAM contact heatmaps showing the TADs, and a subtraction of the shSCRAM from the shBRG1 contact heatmap.



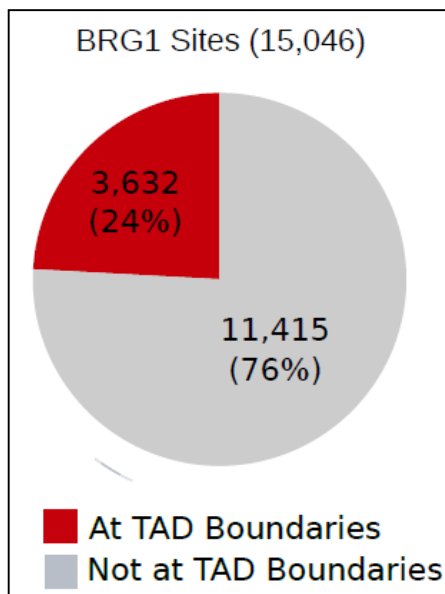
**Figure 4.44. A large portion of TADs remain similar upon BRG1 knockdown.** Venn diagram showing that the TAD boundaries are largely similar between shSCRAM and shBRG1 Hi-C datasets.



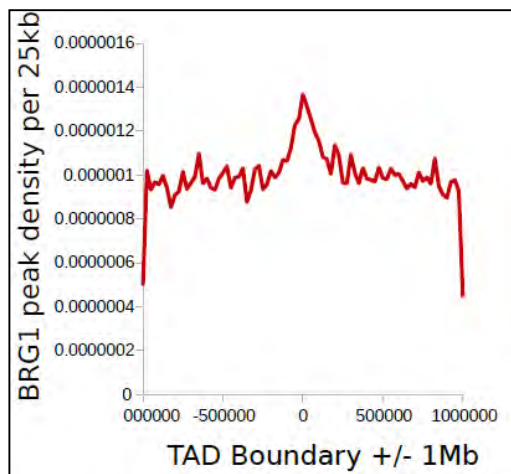
**Figure 4.45. TAD boundaries are enriched for PolII binding and gene frequency.** Frequency of UCSC genes (left) and Pol2 binding (right) at and +/- 1Mb of TAD boundaries

Consistent with the notion that TADs are stable across different cell types, species and different biological contexts [73-76], BRG1 knockdown did not significantly alter the localization of the TAD boundaries, as the majority of TAD boundaries (83% of shSCRAM and 88% of shBRG1) were overlapping between the

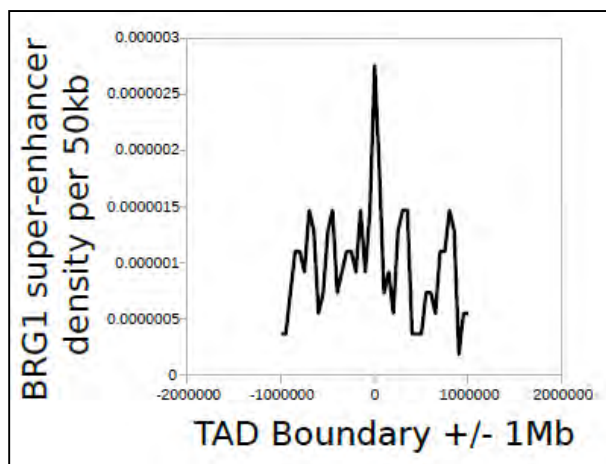
control and the BRG1 knockdown cells (Figure 4.44). TAD boundaries are bound by proteins such as CTCF and cohesin in vertebrates [73; 78] and by several architectural binding proteins in flies [79]. Therefore, we asked whether BRG1 plays any role in TAD boundaries and assessed whether BRG1 localization was enriched at TAD boundaries. Intersection of BRG1 ChIP-seq peaks with both the shBRG1 and shSCRAM TAD boundary definitions yielded similar results, where ~25% of all BRG1 binding was located at TAD boundaries (Figure 4.46). Surprisingly, we observed an enrichment of BRG1 binding at the boundaries when the frequency of BRG1 binding was plotted around the TAD borders (Figure 4.47). 66% of all TAD boundaries were bound by BRG1. Thus, TAD boundaries are enriched for BRG1 binding. A similar phenomenon was observed when BRG1-bound super-enhancers were plotted across the TAD boundaries (Figure 4.48).



**Figure 4.46. A significant portion of BRG1 binding is located on TAD boundaries.** Pie chart showing the percentage of BRG1 localization at TAD boundaries.

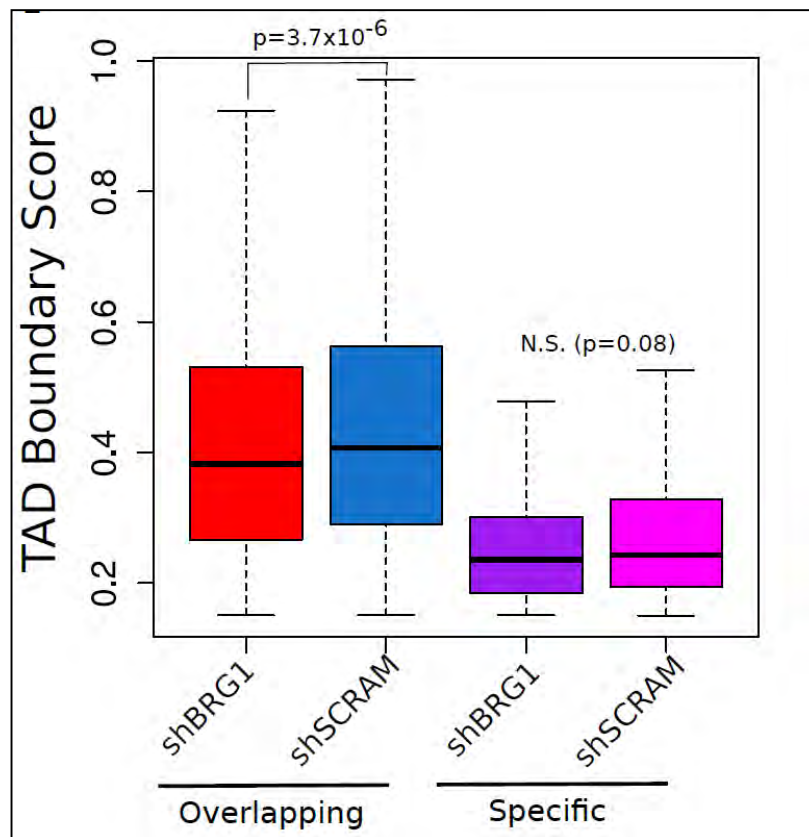


**Figure 4.47. BRG1 and BRG1-bound super-enhancers are enriched at TAD boundaries.** The frequency plot of BRG1 ChIP-seq peaks per 25kb for +/- 1Mb of every shBRG1 TAD boundary.

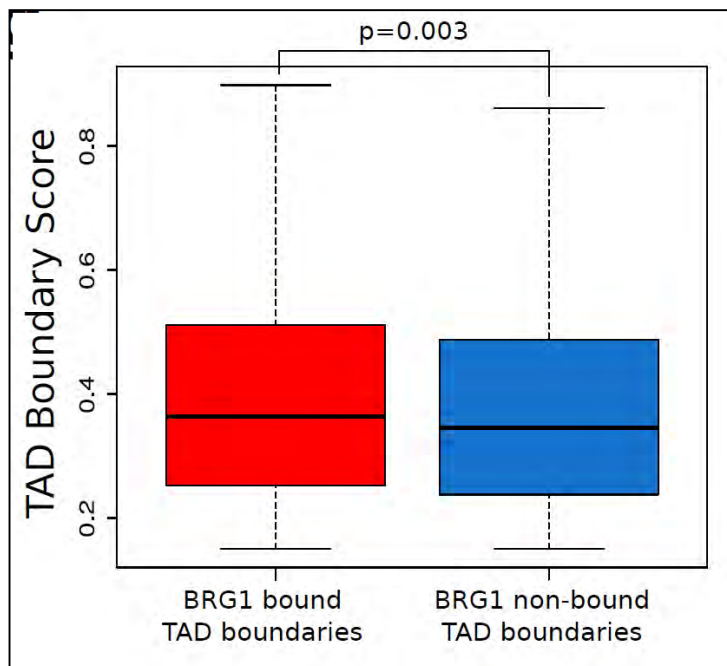


**Figure 4.48.** The frequency plot of BRG1 super-enhancers per 50kb for +/- 1Mb of every shBRG1 TAD boundary.

The strength of a TAD boundary is a measure of the allowance of inter-TAD interactions across the boundary [79; 80]. Even though our data indicate that the majority of TAD boundaries are similar, we wondered whether BRG1 depletion resulted in a change in TAD boundary strength. Interestingly, BRG1 knockdown resulted in an overall decrease in the overlapping TAD boundary strength, as shown by plotting the boundary scores for the overlapping and unique TAD boundaries for the shBRG1 and shSCRAM cells (Figure 4.49). Furthermore, to address whether this decrease is related to BRG1 binding, we compared the TAD boundary scores of the TAD boundaries that were either BRG1-bound or not. We observed a significant decrease of TAD boundary strength at borders that lacked BRG1 binding (Figure 4.50). As a result, the data suggests that BRG1 binding is associated with stronger TAD boundaries.



**Figure 4.49. BRG1 knockdown results in a lower TAD boundary score.** Boxplot showing the TAD boundary score distribution for the overlapping, shSCRAM and shBRG1 specific TAD boundaries.



**Figure 4.50. BRG1 binding is associated with higher TAD boundary score.** Box plot showing the TAD boundary scores for BRG1 bound and unbound TAD boundaries.

#### 4.4 Discussion

In this study, we characterized BRG1-dependent alterations in gene expression and higher order chromatin structure in mammary epithelial MCF-10A cells. Modification of chromatin structure by ATP-dependent remodeling complexes is an essential process in transcriptional regulation. The role of BRG1 involves both the activation and repression of many genes through its interactions with transcription factors and other cofactors [237]. RNA-seq analysis in the control (shSCRAM) and BRG1 knockdown (shBRG1) cells showed an extensive down-regulation of gene expression (Figure 4.4). ChIP-seq analysis indicated that BRG1 is mostly bound (~60%) to gene bodies and ~40% is bound to intergenic regions (Figure 4.16), which is similar to the case in murine mammary epithelial cells [230]. Differentially expressed

genes that were positively regulated by BRG1 generally showed BRG1 binding at the locus, whereas genes negatively regulated by BRG1 generally did not show binding by BRG1, suggesting that negative regulation by BRG1 occurs via indirect mechanisms.

Growing evidence suggests that the shape of the nucleus, the stiffest organelle in the cell, might partly be affected by force-induced changes, a phenomenon known as nuclear mechanotransduction [264]. It is becoming well established that the cell surface adhesion receptors, such as integrins and cadherins, can exert the mechanical forces to the nucleus and can potentially cause gene activation and/or chromatin reorganization [265]. However, changes in nuclear shape can be induced from either external forces exerted by the cytoskeleton, or via internal nuclear forces. Previous work showed that BRG1 depletion results in nuclear shape alterations in MCF-10A cells [252]. However, the disruption of the cytoplasmic filaments (actin, tubulin and cytokeratins) did not alter BRG1-dependent structural changes observed in the MCF-10A cells [252]. This implies that BRG1, apart from its chromatin remodeling function, might have additional roles in maintaining the structural integrity of the nucleus [254]. Interestingly, in this study, we identified that the majority of the genes that are down-regulated upon BRG1 knockdown were associated with the extracellular matrix (ECM). These findings suggest an additional mechanism for BRG1-mediated regulation of nuclear integrity in MCF-10A cells via regulation of ECM genes and possible alteration of cell surface connections and mechanotransducing forces to the nucleus. In contrast, the up-regulated genes were associated with the lipid synthesis pathways (Figure 4.10), which suggests a role for BRG1 in regulating cell metabolism.



Hi-C analysis of BRG1 knockdown and control MCF-10A cells revealed a significant enrichment of sub-telomeric interactions in the shBRG1 cells (Figure 4.33f-h). Recently, Guidi *et al.* showed that the telomeres of yeast cells undergo spatial reorganization upon switching to different metabolic states [266]. Since our RNA-seq analysis showed up-regulation of genes related to lipid synthesis and since Hi-C analysis showed alterations in telomeric interactions in shBRG1 cells, it is very tempting to speculate that the increased telomeric interactions (Figure 4.33f-h) may be associated with changes in the metabolic state of the shBRG1 cells.

The formation of topologically associated domain boundaries is dependent on many factors [79; 267] ], especially insulators such as CTCF and cohesin [73; 78]. Here, we report the rather remarkable observation that BRG1 peaks and super-enhancers are also enriched at TAD boundaries (Figure 4.46-3.48), and that the loss of BRG1 results in lower TAD boundary scores genome-wide (Figure 4.49), implicating BRG1 as a co-regulator of the integrity of topological associated domains. Furthermore, when the TAD boundaries of the control (shSCRAM) cells are considered, BRG1-bound boundaries exhibited a stronger boundary score than the boundaries not bound by BRG1 (Figure 4.50). There are ~1400 sequence specific transcription factors in the human genome [268]. Therefore, apart from well-studied insulators such as CTCF and cohesin, maintenance of TAD boundaries may at least in part be regulated by other chromatin regulators such as BRG1, and by extension, the SWI/SNF enzyme.

Taken together, we identify novel roles for BRG1 in regulating higher-order chromatin structure by affecting telomere organization, TAD boundary strength and the

frequency and specificity of long-range chromatin interactions. In conjunction with our earlier studies indicating that BRG1 regulates the aspect of nuclear size and shape [252; 254], these data provide possible mechanisms by which changes in global chromatin organization may modulate the integrity of the nucleus.

## **CHAPTER5: RUNX1-mediated control of higher-order chromatin organization and gene expression in breast cancer cells**

### **5.1 Introduction**

The mammalian RUNX family of transcription factors consists of three members, RUNX1, RUNX2 and RUNX3, which show distinct tissue-specific expression patterns and cell-context dependent functions through interactions with their common partner core binding factor beta (CBF $\beta$ ) [272]. Each RUNX member has a divergent physiological role. RUNX1 is involved in hematopoiesis [273], RUNX2 is required for bone formation [274], and RUNX3 is associated with gastrointestinal and neuronal development [275; 276]. Therefore, RUNX proteins are considered master regulators of multiple signaling pathways in both normal and abnormal physiological conditions, and hence have been shown to play key roles in many different types of cancer (reviewed in [277-280]). In particular, RUNX1 is involved in the RAS, ERK, TGF $\beta$  and WNT signaling pathways in a variety of cellular contexts including cancer [281-284].

Several translocations and mutations at the *Runx1* gene locus are associated with the onset of human leukemia types [285; 286]. However, the role of RUNX1 is not confined to the hematopoietic lineage. In the last decade, accumulating evidence suggests a pivotal role for RUNX1 in the development of breast cancer both as an oncogene and a tumor suppressor [287-292]. RUNX1 has been shown to be down-regulated [287; 292; 293] as well as up-regulated [292] in breast cancer. In addition, recent whole-genome sequencing studies identified deletions and point mutations in the

*Runx1* gene in a significant portion of human luminal breast tumors [182; 294; 295]. Another study based on a meta-analysis of microarray studies comparing *Runx1* expression in normal versus tumor tissues reported that *Runx1* is in the top 1% overexpressed genes in breast cancer [296].

RUNX1 exerts its function via both activating and repressing target gene expression depending on its interaction partners [276; 277]. The RUNX1 protein assembles into subnuclear domains and associates with the nuclear matrix [297]. RUNX1 has been shown to tether estrogen receptor alpha (ERalpha), an important regulator in breast cancer, to chromatin [298]. Similar to tethering ERalpha to its target sites, RUNX1 also interacts with the polycomb repressive complex 1 (PRC1) and regulates its recruitment to the chromatin [299; 300]. Furthermore, in the context of hematopoiesis, chromosome conformation capture studies showed RUNX1 involvement in mediating locus-specific, long-range interactions to regulate gene expression [301; 302].

Cancer is a disease characterized by large scale changes in the nucleus [126]. The organization of the genome involves multiple hierarchical structures [187]. First, each chromosome is positioned within a confined volume in the nucleus forming the chromosome territories [60]. Each chromosome is partitioned into megabase scaled genomic compartments [71] and further folded into discrete chromosomal neighborhoods called topologically associating domains (TADs) [73; 74]. Being mostly invariant across tissue types and species [73-78], TADs are defined as clusters of interaction domains in which the enhancers and promoters inside a single TAD cross-

talk with one another. The expression of the genes inside a single TAD is co-regulated, and two neighboring TADs can have different modes of regulation [76]. Alteration in higher-order chromatin organization is a frequently observed phenomenon in breast cancer [126]. Recently, we demonstrated that breast cancer cells have extensively altered long-range chromatin contacts among small, gene-rich chromosomes and at telomeres when compared with mammary epithelial cells [263]. Given the importance of higher-order genome folding and the involvement of RUNX1 in the context of breast cancer, the relationship between the function of RUNX1 involvement in chromatin organization and structure and its functional role in breast cancer requires investigation.

In this study, we characterized the genome-wide alterations in higher-order chromatin structure and gene expression upon RUNX1 knockdown in the tumorigenic human breast cancer MCF-7 cell line by using Hi-C and RNA-seq technologies. In addition, to gain insight into RUNX1-mediated regulation of the chromatin, we probed RUNX1 localization by performing ChIP-seq analysis. We observed that RUNX1 is involved in significant alterations in gene expression and higher-order genome structure, particularly at TAD boundaries. As a result, we show a previously unidentified role for RUNX1 in genome architecture, and we provide additional insight into the underlying consequences of RUNX1 perturbation in breast cancer.

## **5.2 Materials and Methods**

### **5.2.1 Generation of MCF-7 cell lines and cell culture**

The MCF-7 cells were obtained from ATCC and were cultured in DMEM supplemented with 10% fetal bovine serum and 5% penicillin/streptomycin. For the shRNA-mediated

knockdown of RUNX1, MCF-7 cells were plated in six-well plates ( $1 \times 10^5$  cells per well) and infected 24h later with lentivirus expressing shRunx1 (5'-GATCATCTAGTTTCTGCCG-3') or nonspecific shRNA (shNS) (Thermo Scientific). Briefly, cells were treated with 0.5 ml of lentivirus and 1.5 ml complete fresh DMEM high glucose per well with a final concentration of 4  $\mu\text{g/ml}$  polybrene. Plates were centrifuged upon addition of the virus at  $1460 \times g$  at 37 C for 30 min. Infection efficiency was monitored by GFP co-expression 2 days after the infection. Cells were selected with 2  $\mu\text{g/ml}$  puromycin for at least two additional days. After removal of the floating cells, the remaining attached cells were passed and analyzed.

### **5.2.2 RNA-seq and Analysis**

The RNA-seq libraries were generated with TruSeq Stranded Total RNA with Ribo-Zero Gold Kit and the samples were sequenced as 100-bp single-end reads using a Hi-Seq 2000 instrument. For the RNA-seq analysis, first the adapter sequences were removed from the RNA-seq reads. Next, any ribosomal RNA reads were filtered using Bowtie [269]. The reads were then aligned to the hg19 transcriptome and the gene counts and transcript per million (TPM) values were quantified using the RSEM v.1.2.7 tool [208]. Differential gene expression was calculated using the DeSeq2 v.1.4.5 package in R 3.1.0 using the mean value of gene-wise dispersion estimates [209]. Genes with adjusted p-value less than 0.01 and log<sub>2</sub> fold change >1 were considered as differentially expressed.

### **5.2.3 ChIP-seq and Analysis**

The ChIP assay was performed as previously described [270]. Briefly,  $\sim 1 \times 10^7$  parental (wildtype) MCF-7 cells were crosslinked with formaldehyde at room temperature for 10 minutes. Then, the cells were lysed using lysis buffer A (50mM HEPES, 140mM NaCl, 1mM EDTA pH=8, 10% Glycerol, 0.5% NP-40, 0.25% Triton X-100), and the residual cytoplasmic protein was removed using lysis buffer B (10mM Tris-HCl pH=8, 200mM NaCl, 1mM EDTA, 1mM EGTA). The nuclear fraction was released using lysis buffer C (10mM Tris-Hcl pH=8, 100mM NaCl, 1mM EDTA, 1mM EGTA, 0.1% Sodium Deoxycholate, 0.5% N-lauroylsarcosine). The chromatin was then sheared using a Covaris S2 instrument with 10% duty cycle, 5 intensity, 200 cycles per burst, frequency sweeping mode, 60 second process time and for 4 cycles. The pull-down was performed using a RUNX1 antibody (Cell Signaling #4334). Samples were washed three times with RIPA buffer (Tris-HCl pH=8, 150mM NaCl, 1mM EDTA, 1% NP-40, 0.25% Sodium deoxycholate, 0.1% SDS) and were eluted. The pull-down and input control sequencing libraries were generated by using the NEXTflex Rapid DNA Sequencing Kit (Bioo Scientific #5144-02) and were sequenced using SE100 reads with a HiSeq 2000 instrument. The adapters were trimmed from the sequencing reads, and the reads were aligned to the hg19 human genome using the Bowtie2 tool [269]. Quality controls, peak calling, and peak annotation was performed using the HOMER suite [271]. The ChIP-seq peaks that were reproducible across the biological replicates were used for downstream analysis. *De novo* motif analysis was performed using the MEME-ChIP suite [306]. The geneset enrichment of ChIP-seq peaks was analyzed with the ChIP-Enrich tool [310].

#### **5.2.4 Preparation of Hi-C Libraries**

The Hi-C experiments were performed with two biological replicates as previously described with minor modifications [99]. The altered sections of the protocol was the biotin incorporation step, where the mixture was incubated at 37°C for 40 minutes with continuous shaking. The MCF-7 shNS and shRUNX1 replicates displayed a 40-90% biotin incorporation efficiency. The Hi-C libraries were sequenced using PE100 reads with a HiSeq 2000 instrument.

#### **5.2.5 Hi-C read mapping/binning/ICE correction**

Initial Hi-C analysis was performed as previously described [100]. Table 3 summarizes the mapping statistics of different classes of reads and interactions observed for each biological replicate [100]. The data was binned at 2.5Mb, 1Mb, 250kb, 100kb and 40kb non-overlapping genomic intervals. The iterative correction and eigenvector decomposition (ICE) method was used to correct for systematic biases [212]. The biological replicates showed high reproducibility. For the downstream analyses, sequences from both biological replicates were pooled and ICE-corrected to serve as a combined dataset.

#### **5.2.6 Calculation of Genomic Compartment Profiles**

First, we modeled the overall Hi-C decay with distance using a modified LOWESS method ( $\alpha=1\%$ , interquartile range filter) as described previously [94]. LOWESS is used for calculating the weighted average and weighted standard deviation for every



genomic distance and normalizing for genomic distance bias. To calculate the compartment profiles, the z-scores of the interaction matrices at 250kb resolution were generated. Then, the Pearson correlation of the z-score matrices was calculated. By performing principal component analysis, the first principal component detects the patterns of increased and decreased interaction frequency, which appear as a “plaid pattern” in the heatmap. Each region has a positive or a negative first eigen value. To detect which compartment (positive or the negative eigen value) is A-type, the gene density was calculated to assign the A-type and the B-type compartmentalization.

### **5.2.7 Identification of TAD boundaries**

TAD calling was performed by using the insulation score of each bin using the 40kb resolution combined data, as described previously [213; 263]. By sliding a 1Mb x 1Mb square along the diagonal of the interaction matrix for each chromosome, we obtained the “insulation plot” of the matrix. The insulation plots showed high correlation between the biological replicates. Valleys within the insulation plots indicate the depletion of interaction across two TADs. Based on the variation of boundaries between replicates, we added a total of 160kb (80kb to each side) to the boundary to account for replicate variation, thereby making the final boundary size 200kb. All boundaries with a strength  $<0.15$  were excluded.

### **5.2.8 Calculation of Differential Interactions**

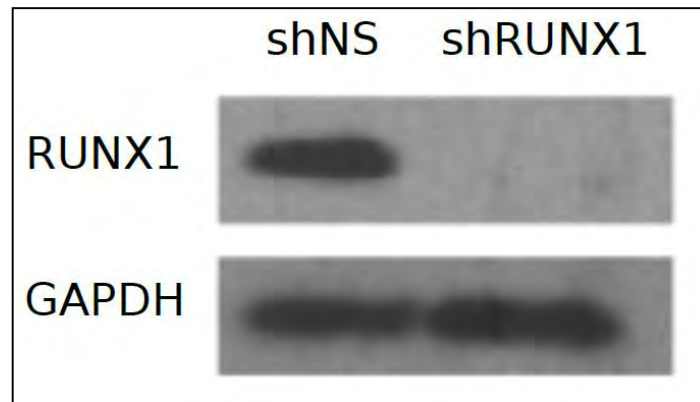
In order to probe the differential interactions between MCF-7 shNS and shRUNX1 samples, we used a method previously described [263]. We first transformed the Hi-C data into Z-score matrices for all biological replicates. For each interaction, the mean

sample to sample Z-score was calculated between all pairwise combinations of the four replicate datasets (shNS-R1 vs. shNS-R2, shNS-R1 vs. shRUNX1-R1, shNS-R1 vs. shRUNX1-R2, shRUNX1-R1 vs. shRUNX1-R2, and shRUNX1-R2 vs. shNS-R2). A randomized set of 500,000 interactions were also calculated. These random interaction Z-score differences were used to build an expected distribution of the Z-score differences. The resulting Z-score difference matrix was then derived by calculating the ratio of the mean of the set of four possible sample:sample Z-score differences minus the genome-wide mean of the replicate:replicate Z-score difference, divided by the genome-wide standard error of the replicate:replicate Z-score differences for each bin.

### **5.3 Results**

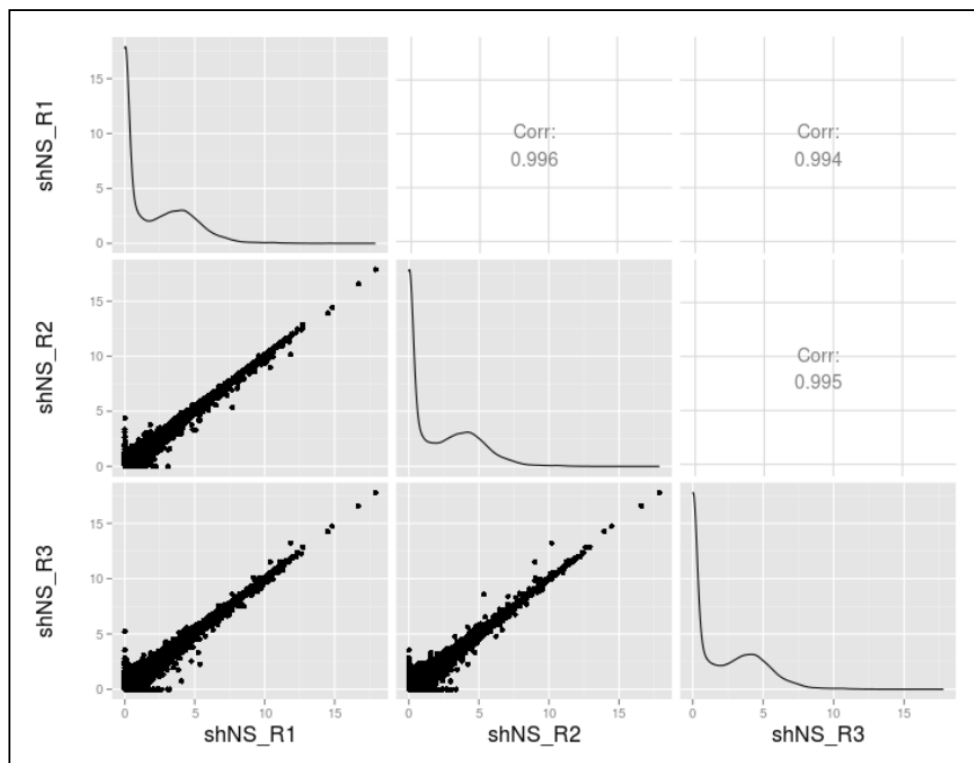
#### **5.3.1 RUNX1 knockdown results in aberrant gene regulation in MCF-7 cells**

To investigate the transcriptional effects of RUNX1 knockdown, we generated MCF-7 cell lines stably expressing either a non-silencing shRNA (shNS), or shRNA against RUNX1 (shRUNX1). Western blot analysis of induced cells showed no detectable levels of RUNX1, confirming a near complete knockdown (Figure 5.1).

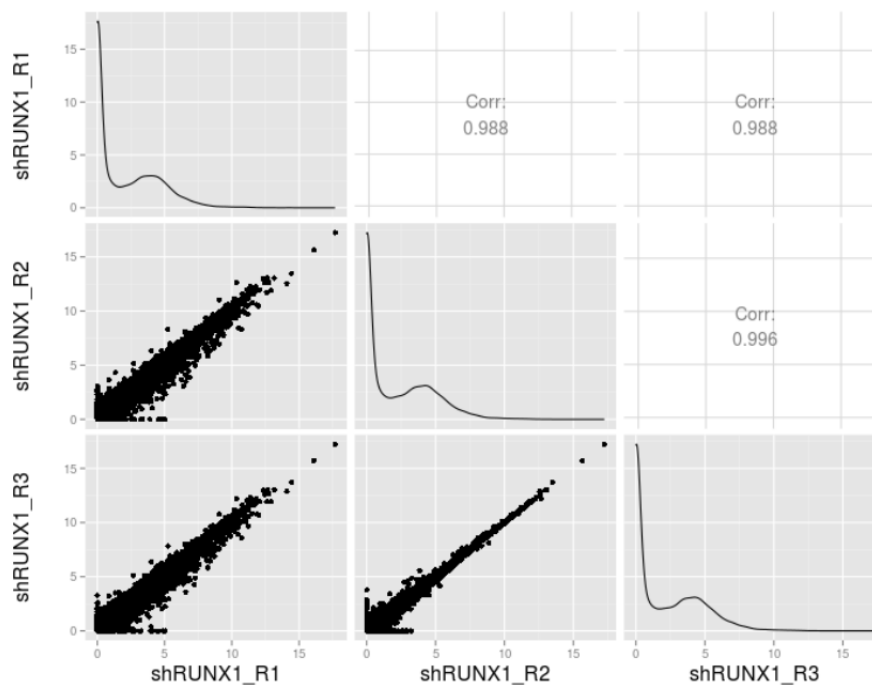


**Figure 5.1. Knockdown of RUNX1 results in near complete depletion of the protein levels.** Western blot of the RUNX1 protein levels of shNS and shRNX1 MCF-7 cells. There is a near complete knock down of RUNX1 protein levels.

In order to analyze the effect of RUNX1 depletion on global gene expression, we performed RNA-seq from ribosomal-RNA depleted samples in shNS and shRUNX1 MCF-7 cells in triplicates, which showed high correlation among each other (Figure 5.2 and Figure 5.3).

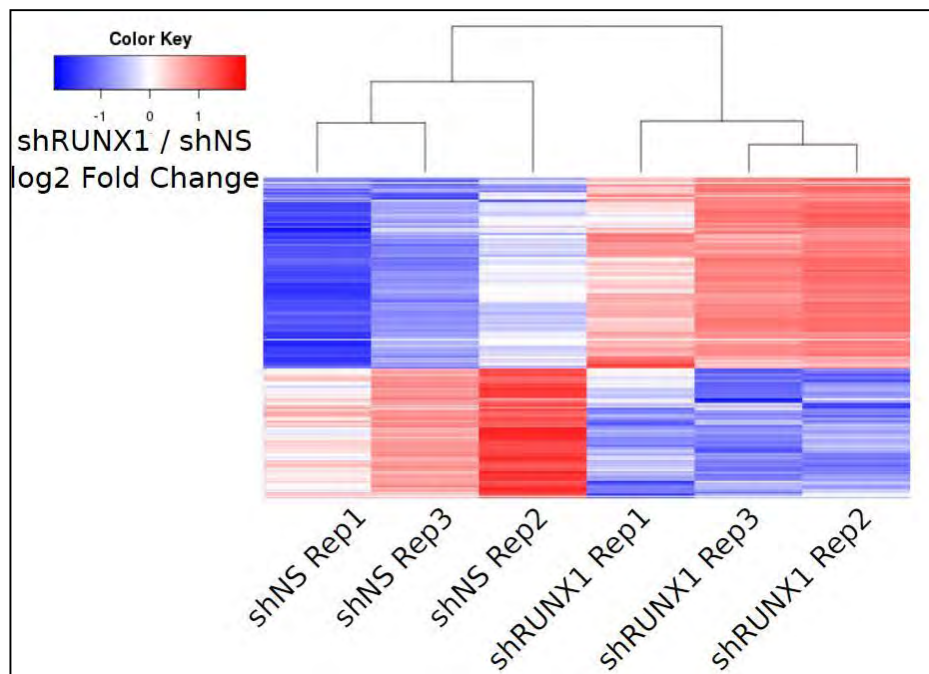


**Figure 5.2. Replicate correlation of gene expression (Transcript per million, TPM) for shNS biological triplicates.** The Pearson Correlation values for each pairwise replicate are on the right-top side of the matrices.

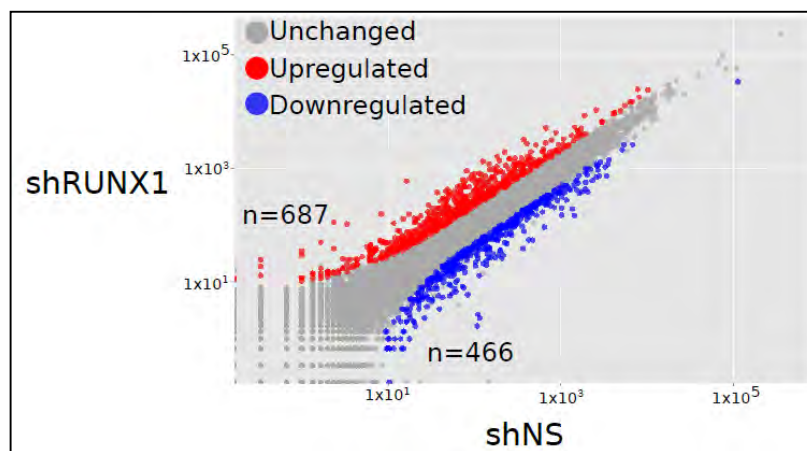


**Figure 5.3. Replicate correlation of gene expression (Transcript per million, TPM) for shRUNX1 biological triplicates.** The Pearson Correlation values for each pairwise replicate are on the right-top side of the matrices.

By performing a differential gene expression analysis [209], we identified 466 down-regulated and 687 up-regulated genes upon RUNX1 knockdown (Figure 5.4 and Figure 5.5).



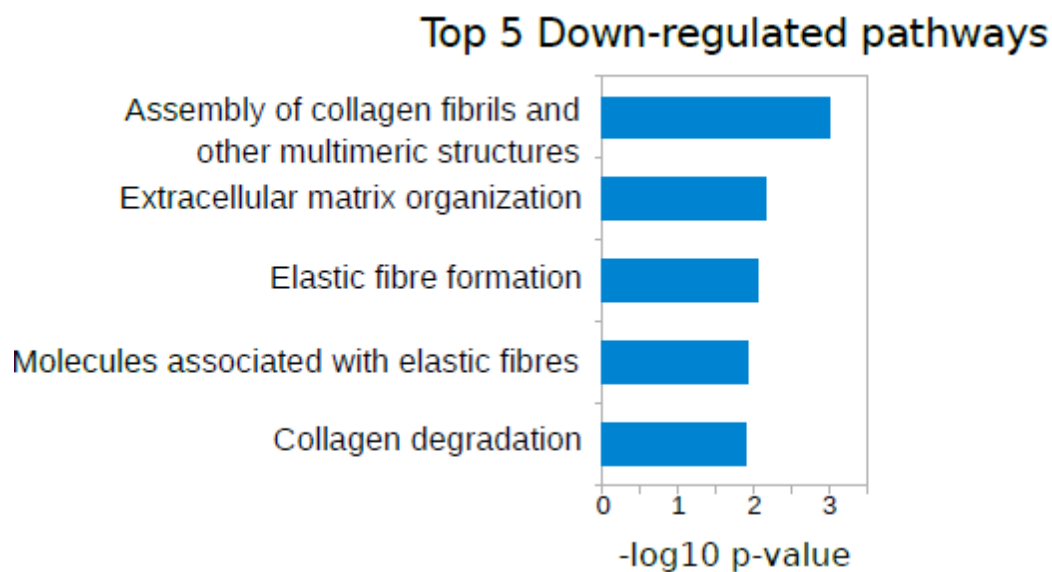
**Figure 5.4. RUNX1 knockdown results in extensive transcriptional changes.** Heatmap showing the transcripts per million (TPM) expression values of the differentially expressed genes between shNS and shRUNX1 samples for each biological replicate.



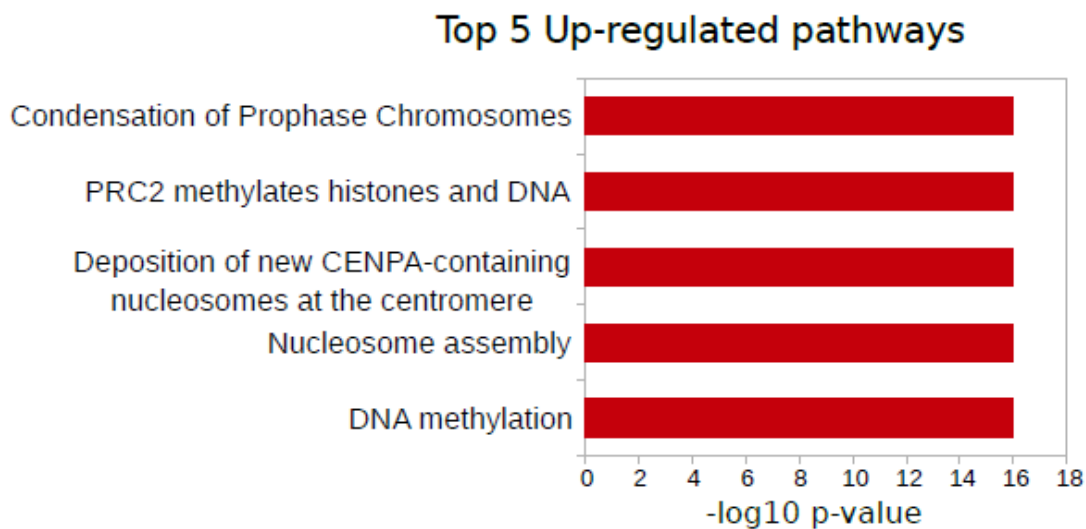
**Figure 5.5. RUNX1 depletion results in up- and down-regulation of hundreds of genes.** Scatter plot showing the log<sub>2</sub> pooled replicate gene expression values for shNS and shRUNX1 cells. The red and blue dots denote the significantly up- and down-regulated genes, respectively.

REACTOME pathway analysis [256; 257] of the down-regulated genes identified pathways related to extracellular matrix (ECM) constituents, especially involving collagen fibers and collagen assembly (Figure 5.6). The down-regulated genes included certain well-studied lncRNAs, such as NEAT1 and MALAT1, that are involved in nuclear organization. In contrast, the up-regulated genes were associated with chromatin regulation, including condensation of chromosomes, as well as nucleosome and centromere assembly (Figure 5.7).

Collectively, these results confirm a role for RUNX1 as a both activator and a repressor of gene expression in breast cancer cells. Furthermore, knockdown of RUNX1 results in dysregulation of several genes related with ECM biology and chromatin structure, both of which have been frequently implicated in breast cancer [263; 303]].



**Figure 5.6. REACTOME terms of down-regulated genes.** Bar graph showing the  $-\log_{10}$  p-values for the top 5 REACTOME terms for the genes that are down-regulated upon RUNX1 knockdown.



**Figure 5.7. REACTOME terms of up-regulated genes.** Bar graph showing the  $-\log_{10}$  p-values for the top 5 REACTOME terms for the genes that are up-regulated upon RUNX1 knockdown.

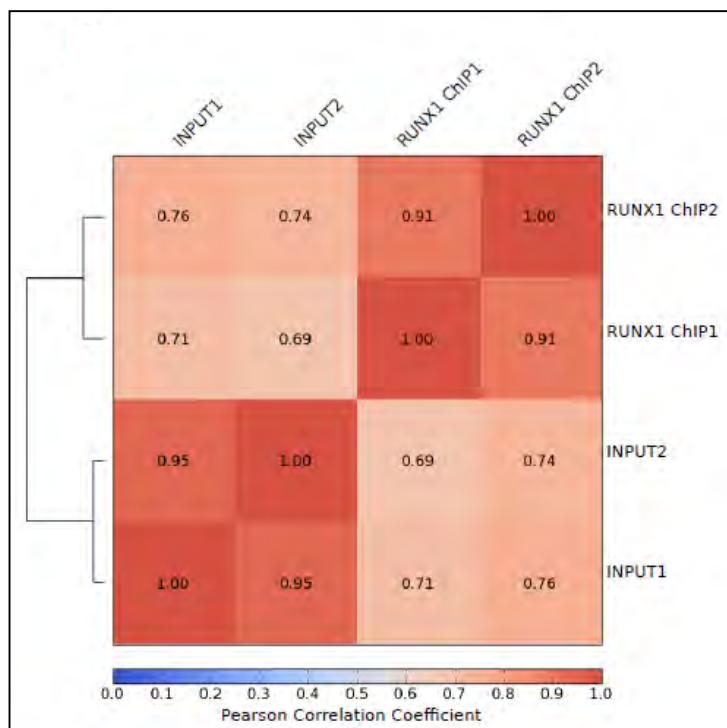
### 5.3.2 Genome-wide RUNX1 localization in parental MCF-7 cells

To expand our understanding about the role of RUNX1 in breast cancer cells, we probed its genome-wide occupancy by performing chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) in the parental MCF-7 cells (Figure 5.8 and Figure 5.9).

Sample Name	Biological Replicate	Raw Reads	Mapped Reads
RUNX1 ChIP 1	1	29,734,121	25,949,243
RUNX1 ChIP 2	2	29,988,988	22,957,545
INPUT 1	1	30,075,142	19,047,821
INPUT 2	2	10,900,442	7,699,867

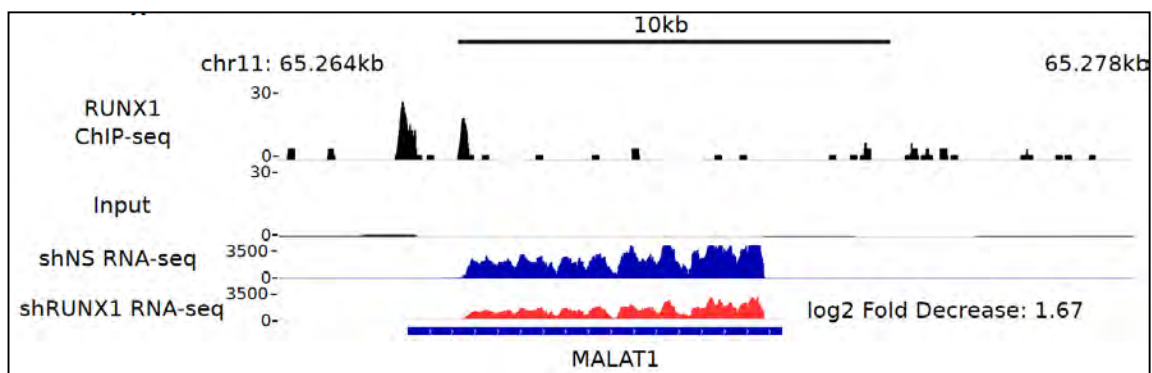
**Figure 5.8. Table showing the sequenced and mapped reads for each RUNX1 ChIP-seq biological replicate.**



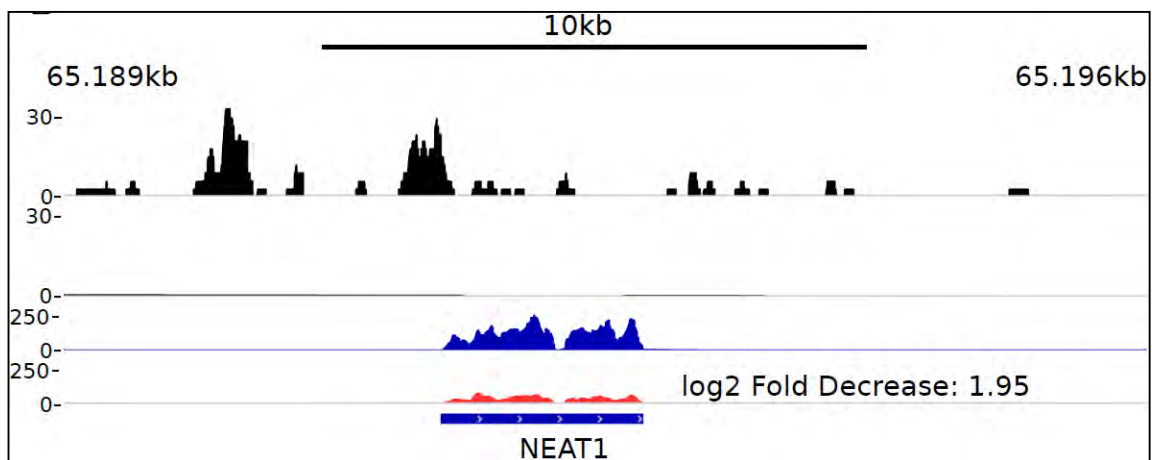


**Figure 5.9.** Matrix showing the Pearson correlation of the signal intensity between the pull down (ChIP) and input samples.

We identified 3466 high-confidence RUNX1 ChIP-seq peaks. Visualization of the RUNX1 peaks revealed binding to defined locations in the genome as expected from a sequence specific transcription factor (Figure 5.10 and Figure 5.11). Interestingly, RUNX1 was bound to the promoters of NEAT1 and MALAT1, the two lncRNAs whose transcription was significantly reduced upon RUNX1 knockdown, suggesting a direct transcriptional role of RUNX1 at the regulation of these lncRNAs that are important for cancer progression (Figure 5.10 and Figure 5.11) [304; 305].









**Figure 5.10. RUNX1 binds to the MALAT1 promoter.** Example of a ChIP-seq genome browser view of RUNX1 binding and the input control, along with shNS and shRUNX1 RNA-seq tracks for MALAT1 (chr11:65182867-65201926).



**Figure 5.11. RUNX1 binds to the NEAT1 promoter.** Example of a ChIP-seq genome browser view of RUNX1 binding and the input control, along with shNS and shRUNX1 RNA-seq tracks for NEAT1 (chr11:65255544-65285081).

To obtain information about the context of the underlying DNA sequences of RUNX1 bound regions, we performed a motif analysis on RUNX1 peaks using the *de novo* DNA motif discovery software MEME [306]. Our analysis identified several binding motifs significantly enriched at RUNX1 peaks (Figure 5.12). The most

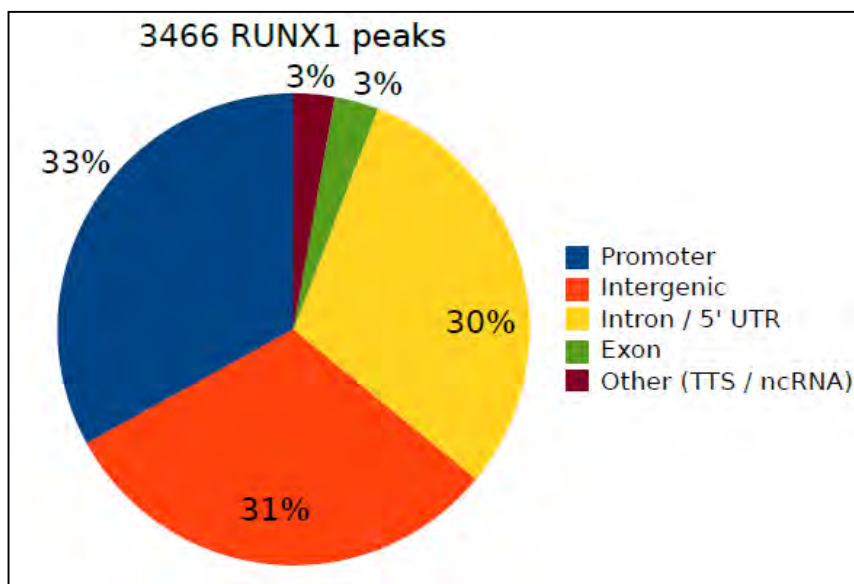
significantly enriched motif was the RUNX1 binding sequence motif. We additionally identified binding motifs for the factors EBF1, STAT3, FOXP3, EGR1, FOS and ESRR. These results are intriguing as RUNX1 has been shown to be a co-regulator for these factors [296; 307-309], as well as being a direct interaction partner of estrogen receptor (ESRR) [298].

Motif	Factor	E-value
	RUNX1	8.6e-173
	EBF1 STAT3	8.6e-41
	FOXP3 FOXD2 FOXL1	1.6e-29
	EGR1 SP1 KLF5	2.3e-29
	FOS JUN FOSL1	2.6e-15
	ESRRG ESRRA	7.9e-15

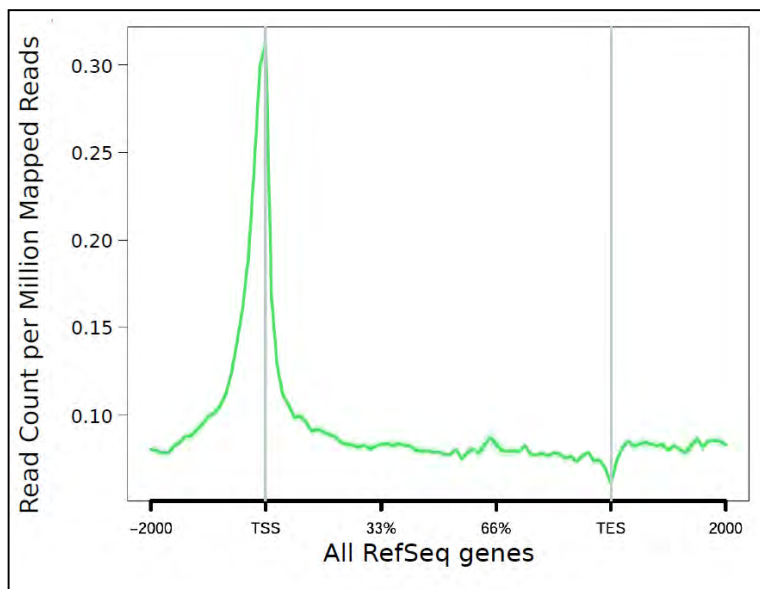
**Figure 5.12. MEME de novo motif analysis of the RUNX1 peaks.** The peaks are ordered by significance from top to bottom.

Annotation of the RUNX1 bound regions revealed ~70% binding at gene bodies and ~30% binding at intergenic regions (Figure 5.13). This suggests that the regulation of RUNX1 occurs mostly through the gene bodies, and possibly through intergenic *cis*-

regulatory regions. The strongest RUNX1 signals at known genes were observed at gene promoter regions (Figure 5.14).



**Figure 5.13.** Pie chart showing the distribution of RUNX1 ChIP-seq peak annotation for genic and intergenic regions.



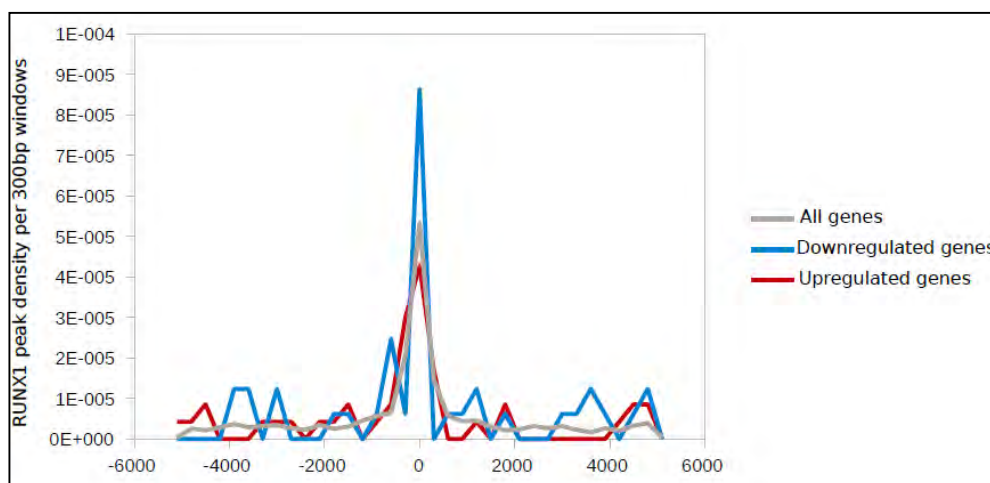
**Figure 5.14. Normalized RUNX1 ChIP-seq signal intensity plot for all human UCSC genes +/- 2kb.** RUNX1 binding is enriched at the promoter regions.

In order to assess the functional significance of RUNX1 binding, we analyzed the enrichment of biological pathways of the genes with RUNX1 binding within 10kb of their transcriptional start site (TSS) [310]. We identified pathways such as G-protein coupled receptor pathway and RNA binding, which is implicated with RUNX1 binding in leukemia as well [311], suggesting RUNX1 mediated regulation of intra-cellular signaling and RNA processing pathways (Figure 5.15).

	GO Molecular Function	pvalue
GO:0004888	transmembrane signaling receptor activity	4.410E-008
GO:0038023	signaling receptor activity	1.325E-006
GO:0004930	G-protein coupled receptor activity	1.021E-005
GO:0003723	RNA binding	1.115E-005
GO:0015267	channel activity	2.776E-005

**Figure 5.15. GO Terms of RUNX1 peaks.** Table showing the gene ontology terms of the genes associated with RUNX1 peaks.

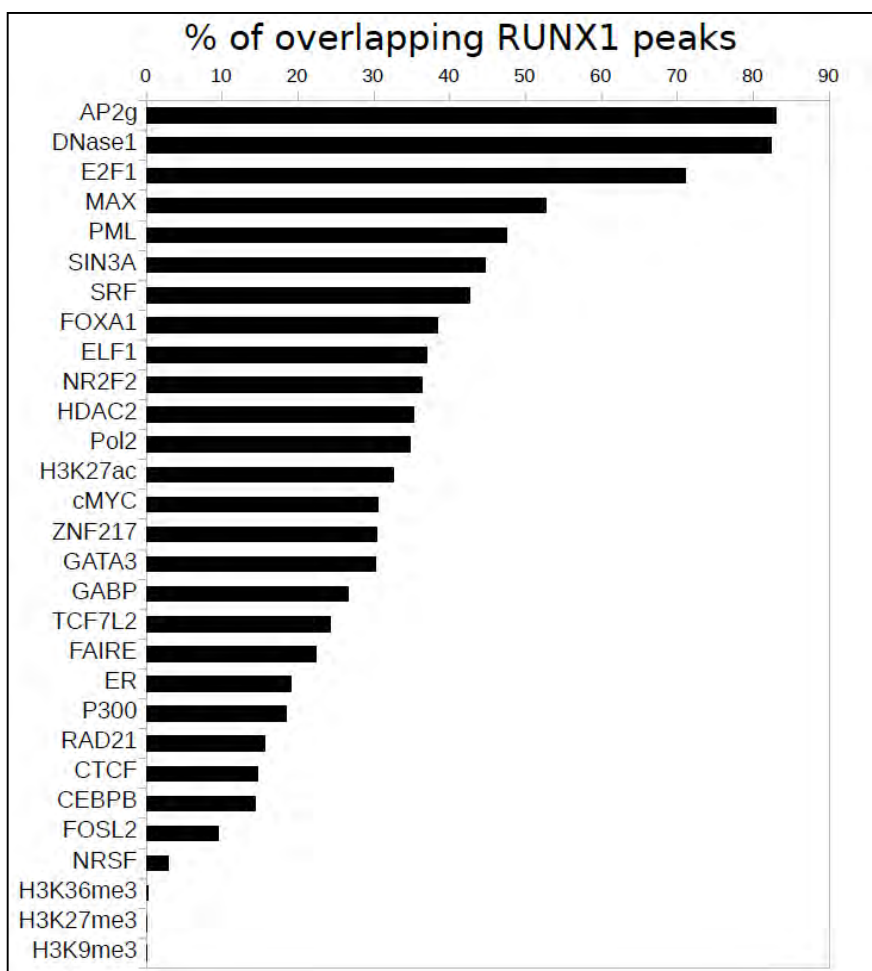
Furthermore, we checked whether RUNX1 peaks were associated with RUNX1-dependent differentially expressed genes. Therefore we plotted the RUNX1 peak density  $\pm$  5kb of the differentially expressed genes' promoters (Figure 5.16).  $\sim$ 10% of all differentially expressed gene promoters displayed RUNX1 binding. The promoters of down-regulated genes displayed a modest enrichment of RUNX1 binding compared to the up-regulated genes, suggesting a more direct role for RUNX1 in differential gene expression.



**Figure 5.16. RUNX1 density across differentially expressed genes.** RUNX1 peak density within  $\pm$  5kb of the TSS of all genes (grey), significantly downregulated genes (blue), or upregulated genes (red).

RUNX1 has been shown to directly interact and be involved with several regulatory complexes [276] and signaling pathways [280]. We asked whether RUNX1 localization overlapped with other MCF-7 ENCODE factors. To address this question, we intersected the MCF-7 ENCODE data [312] for the binding of transcription factors, histone modifications (by ChIP-seq) and open chromatin regions (by FAIRE and DNase1 hypersensitivity) with RUNX1 binding (Figure 5.17). Interestingly, RUNX1

showed the most overlap with the activating enhancer binding protein 2 gamma (AP2g, also known as TFAP2C), which is an important developmental regulator [313]. RUNX1 also was bound at DNase1 hypersensitive regions. Furthermore, RUNX1 showed >50% overlap with HAE2F1 and MAX, whereas several other factors, including ERalpha, demonstrated varying levels of overlap with RUNX1 binding. The lowest amount of overlap was observed with two repressive histone marks, H3K27me3 and H3K9me3 (Figure 5.17). These findings suggest a role for the co-occupancy of RUNX1 with other factors in gene regulation.



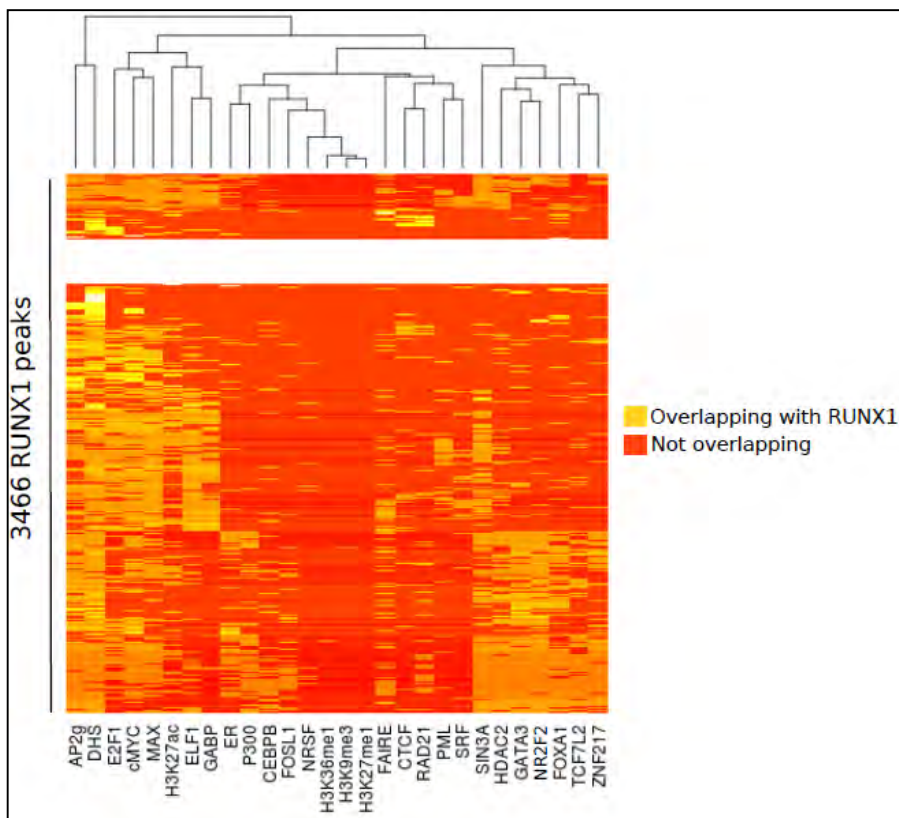
**Figure 5.17. RUNX1 associates with other chromatin modifiers.** Bar graph showing the percent overlap of RUNX1 peaks with open chromatin, transcription factor binding and histone modification

peaks from the ENCODE project.

To gain further insight into the co-binding of RUNX1 with other factors, we analyzed the overlap of each RUNX1 peak with the ENCODE factors and performed a hierarchical clustering (Figure 5.18). We identified two groups of factors that distinctly overlapped with RUNX1. One of the groups consisted of AP2g, HAE2F1, cMYC, MAX, active histone mark H3K27ac, DNase1 hypersensitivity, ELF1 and GABP; whereas the second group consisted of SIN3A, HDAC2, GATA3, NR2F2, FOXA1, TCF7L2 and ZNF217 (Figure 5.18). From this pattern of overlap, it can be inferred that RUNX1 binds to the genome within differentially interacting complexes, and possibly exerts differential (i.e activating or repressive) functions.

Taken together, profiling the genome-wide localization of RUNX1 revealed binding mostly within gene bodies, and related to differential gene expression. Moreover, motif and overlap analysis revealed RUNX1 association with distinct classes of co-regulators.





**Figure 5.18. RUNX1 associates with distinct classes of factors.** A hierarchically clustered heatmap showing the co-overlap of RUNX1 with open chromatin regions FAIRE and DHS, and available ChIP-seq peaks from MCF-7 ENCODE data.

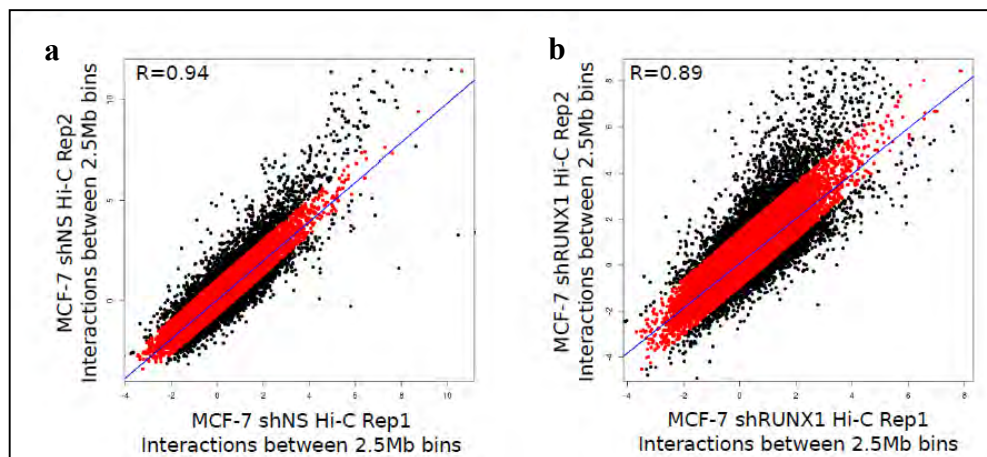
### 5.3.3 RUNX1-mediated genome-wide interaction analysis in MCF-7 cells

In order to probe the higher-order chromatin structure upon RUNX1 knockdown in breast cancer cells, we performed Hi-C in shNS and shRUNX1 MCF-7 cells. Two independently grown and fixed batches of cells were sequenced to an average of depth of ~183 million reads. Sequence mapping and the initial Hi-C analysis were performed as described previously [100; 213; 263], using the Iterative Correction Method (ICE) [212] to correct for systematic biases (Table 5.1).

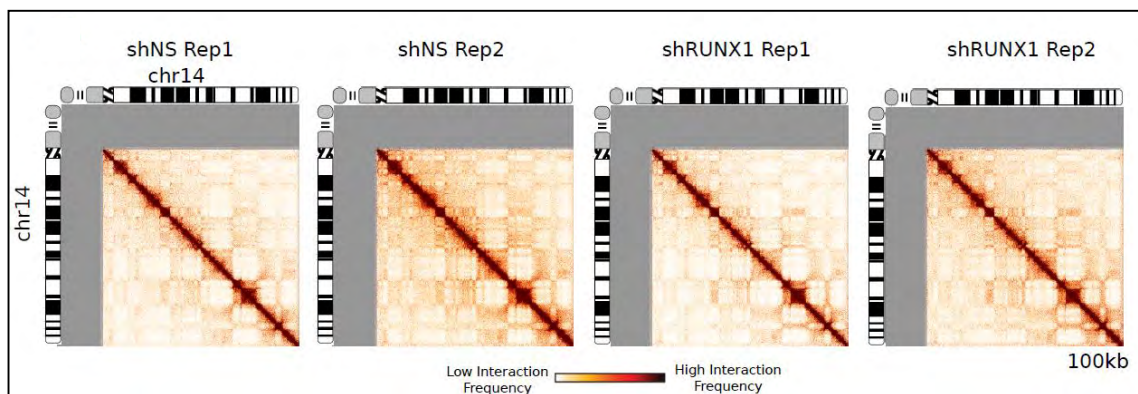
	MCF7 shNS Rep1		MCF7 shNSRep2		MCF7 shRUNX1 Rep1		MCF7 shRUNX1 Rep2	
	Number of Reads	Percentage	Number of Reads	Percentage	Number of Reads	Percentage	Number of Reads	Percentage
<b>Mapping Statistics</b>								
side1TotalReads	187494360		175157276		186989017		183032839	
side1NoMap	16015472	(8.54%)	27162951	(15.51%)	15707167	(8.40%)	30302401	(16.56%)
side1MultiMap	9233932	(4.92%)	5314324	(3.03%)	9752993	(5.22%)	5100587	(2.79%)
side1UniqueMap	162244956	(86.53%)	142880001	(81.46%)	161528857	(86.38%)	147629851	(80.66%)
side2TotalReads	187494360		175157276		186989017		183032839	
side2NoMap	18961810	(10.11%)	32116869	(18.34%)	19255892	(10.30%)	36078375	(19.71%)
side2MultiMap	8934512	(4.77%)	4981817	(2.84%)	9398288	(5.03%)	4671441	(2.55%)
side2UniqueMap	159598238	(85.12%)	138058790	(78.82%)	158334837	(84.68%)	142283023	(77.74%)
<b>Hi-C Library Quality Metrics</b>								
totalReads	187494360		175157276		186989017		183032839	
unMapped	9774342	(5.21%)	6458744	(3.69%)	9969310	(5.33%)	6721909	(3.67%)
singleSided	33598842	(17.92%)	56658273	(32.35%)	34175720	(18.28%)	62709596	(34.26%)
bothSideMapped	144123176	(76.87%)	112040259	(63.97%)	142843987	(76.39%)	113601644	(62.07%)
sameFragment	90338538	(48.18%)	18805876	(10.74%)	82919849	(44.35%)	16833025	(9.19%)
selfCircle	3490566	(1.86%)	9306391	(5.32%)	2435027	(1.30%)	10363891	(5.66%)
danglingEnd	88561769	(47.24%)	9003060	(5.14%)	80218389	(42.95%)	6174470	(3.37%)
bounded	491701	(0.26%)	42551	(0.24%)	522912	(0.28%)	39804	(0.22%)
internal	86070068	(45.92%)	8960509	(5.11%)	79695477	(42.67%)	6134666	(3.35%)
error	286203	(0.15%)	496425	(0.28%)	266433	(0.14%)	294664	(0.16%)
validPair	53784638	(28.69%)	93234383	(53.24%)	59924138	(32.11%)	96768619	(52.88%)
cis	24312821	(12.98%)	25381153	(14.49%)	30131169	(16.13%)	21466406	(11.73%)
trans	29471817	(15.75%)	67853230	(38.75%)	29792969	(16.00%)	75302213	(41.15%)
<b>Hi-C Library Redundancy Metrics</b>								
validPair	53784638	(100%)	93234383	(100%)	59924138	(100%)	96768619	(100%)
totalMolecules	53784638	(100%)	93234383	(100%)	59924138	(100%)	96768619	(100%)
redundantInteractions	1124163	(2.09%)	6620329	(7.10%)	3032197	(5.06%)	8753530	(9.05%)
<b>nonRedundantInteractions</b>	<b>52660475</b>	<b>(97.91%)</b>	<b>86614054</b>	<b>(92.90%)</b>	<b>56891941</b>	<b>(94.94%)</b>	<b>88015089</b>	<b>(90.95%)</b>
percentRedundant	2.09%		7.10%		5.06%		9.05%	

**Table 5.1. The sequencing and mapping statistics of shNS and shRUNX1 Hi-C biological replicates.**

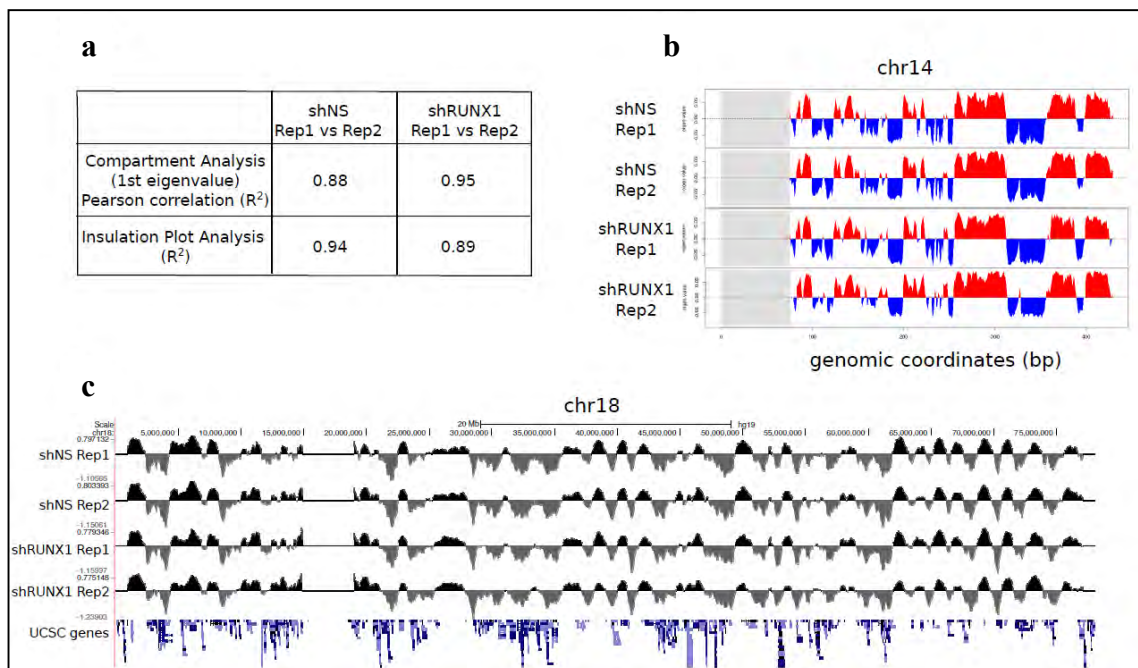
There was a high correlation between the biological replicates at multiple resolutions (Figure 5.19 and Figure 5.20). The compartments and the TAD analysis (discussed below) also showed high reproducibility (Figure 5.21a-c). Therefore, we combined the replicate datasets for downstream analysis. The pooled shNS and shRUNX1 datasets also showed high correlation with the previously published parental MCF-7 Hi-C dataset (Figure 5.22, [263]). The Hi-C datasets displayed similar cis/trans ratios with each other (Figure 5.23). Interaction frequency decreases as a function of genomic distance. Both the shNS and shRUNX1 datasets showed similar rates of decay of interaction frequency with increasing distance (Figure 5.24). Collectively these analyses reflect the high quality and reproducibility of the Hi-C datasets.



**Figure 5.19. Scatter plots comparing normalized interactions between pairs of 2.5Mb bins from two biological replicates** from a) MCF-7 shNS b) MCF-7 shRUNX1. There is a strong correlation between the biological replicates. The black dots denote the outliers whereas the red dots denote the reproducible interactions.



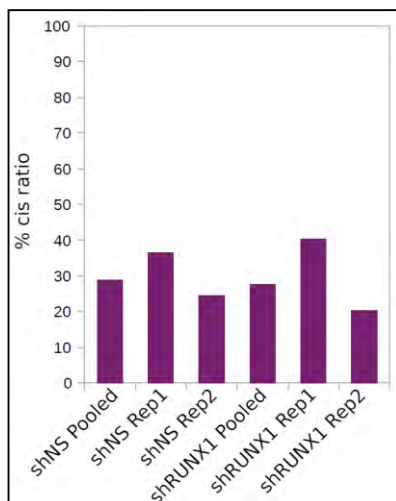
**Figure 5.20. Example of the Hi-C interaction heatmaps at 100kb resolution** for chr14 of the shNS and shRUNX1 biological replicates. The grey area denotes the acrocentric repetitive regions that could not be mapped.



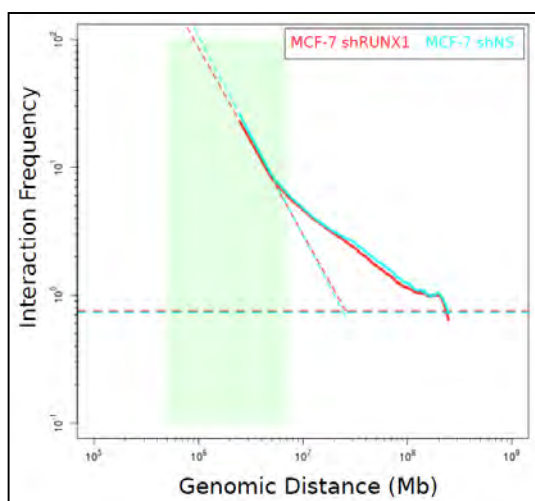
**Figure 5.21. Reproducibility of the Hi-C data** a) Pearson correlation of the compartment analysis (first eigenvector) at 250kb resolution and insulation plot analysis at 40kb resolution between the biological replicates b) An example graph showing the first eigenvector (compartmentalization) of chr14 in shNS and shRUNX1 biological replicates. c) UCSC snapshot showing the insulation plot (TAD analysis) of chr18 in shNS and shRUNX1 biological replicates.

Pearson Correlation Between Pooled Replicates (R-value)	
shNS vs. shRUNX1	0.994
shNS vs. parental (Barutcu et al. 2015)	0.962
shRUNX1 vs. parental (Barutcu et al. 2015)	0.969

**Figure 5.22. Pearson correlation of the pooled replicate datasets** between shNS and shRUNX1, and between the parental MCF-7 cells (from Barutcu et al., 2015)

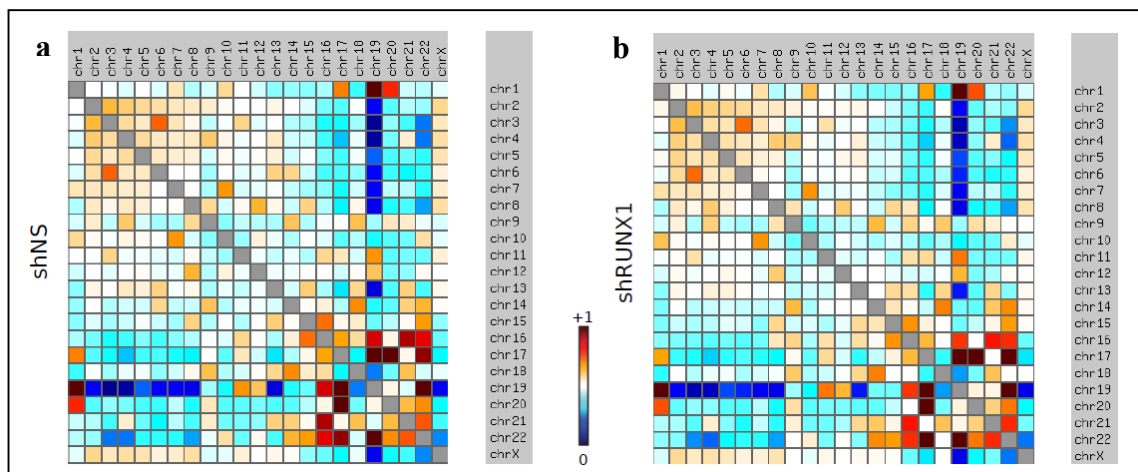


**Figure 5.23.** Bar graph showing the percentage of the percent cis ratios of the replicate and the pooled Hi-C libraries.



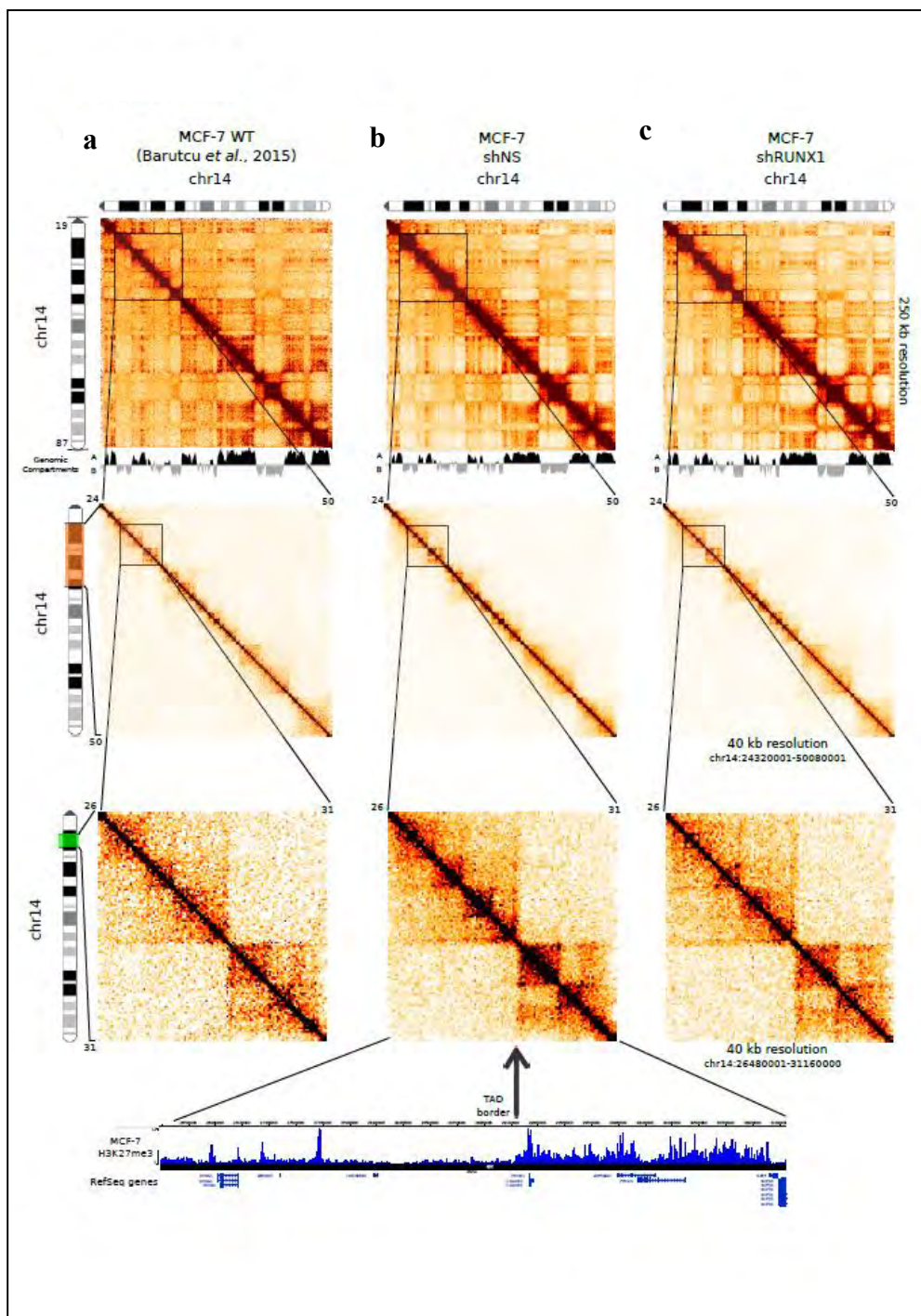
**Figure 5.24.** Scaling plot of the shNS and shRUNX1 pooled Hi-C libraries. The frequency of interaction decrease as a function of genomic distance.

Interaction heatmaps of the individual chromosomes showed no drastic changes in the chromosomal organization of MCF-7 cells upon RUNX1 knockdown (Figure 5.25).



**Figure 5.25. The chromosome by chromosome interactions matrices for shNS and shRUNX1 datasets.**

Visualizing the MCF-7 Hi-C interaction matrices at increasing resolutions revealed the genomic compartments, interaction topology and the TAD structures within the genome (Figure 5.26). The MCF-7 wildtype, shNS and RUNX1 knockdown cells showed similar genomic compartmentalization and TAD structures (Figure 5.26). Consistent with earlier reports [73; 74], two neighboring TADs can display a dramatic difference in repressive histone modification H3K27me3 (Figure 5.26, lower panel), suggesting a differential mode of gene regulation in adjacent TADs.



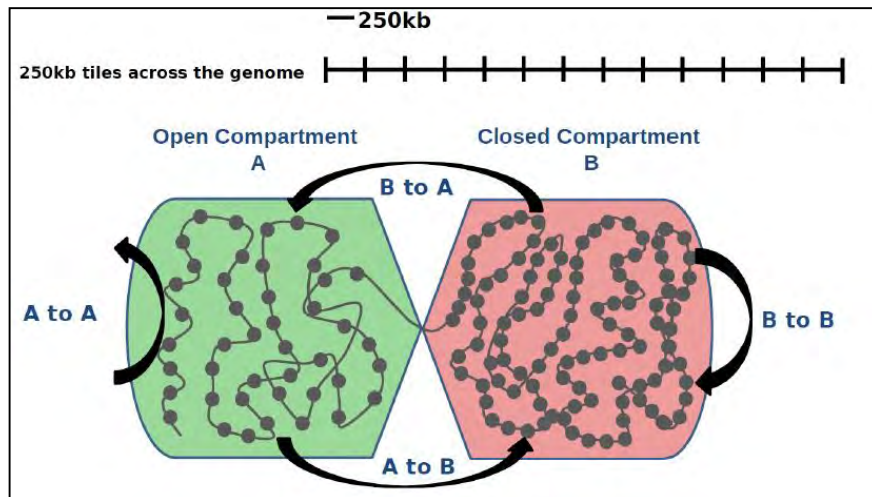
**Figure 5.26. Hi-C interaction heatmaps of chr14 at increasing resolutions** that reveal the genomic compartments and the TAD structures for a) wildtype MCF-7 cells from Barutcu et al, 2015. b) shNS and c) shRUNX1 pooled Hi-C datasets, showing the genomic compartments and the TAD structures. Bottom panel: A browser shot of the MCF-7 H3K27me3 ChIP-seq signal from the ENCODE database showing differential histone methylation across a TAD boundary.

### **5.3.4 A sub-portion of genomic compartments are altered upon RUNX1**

#### **Knockdown**

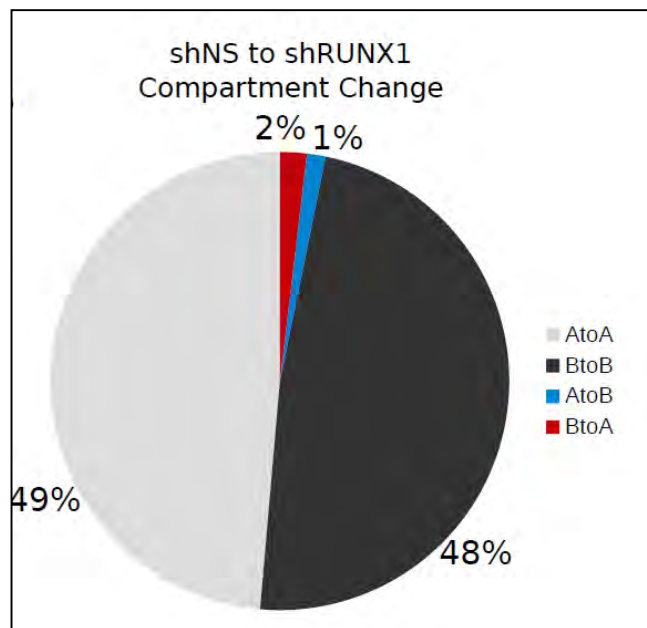
Decades of microscopy studies revealed that the open euchromatic regions occupy a distinct subnuclear region different than the closed heterochromatic regions within each chromosome territory [23]. More recently, Hi-C studies revealed the presence of megabase scale sub-chromosomal structures named A-type and B-type genomic compartments, where the A-type compartments are gene-rich, early replicating, and active whereas the B-type compartments are gene-poor, late replicating and repressed [71]. It was previously shown that compartmental alterations are associated with differential gene expression [77; 263]. Since RUNX1 knockdown in MCF-7 cells was associated with the differential expression of many genes, we asked whether depletion of RUNX1 resulted in any change in genomic compartments. To address this question, we calculated the compartmentalization of each 250kb non-overlapping bin throughout all of the chromosomes and compared the compartment type between shNS and shRUNX1 Hi-C datasets (Figure 5.27).





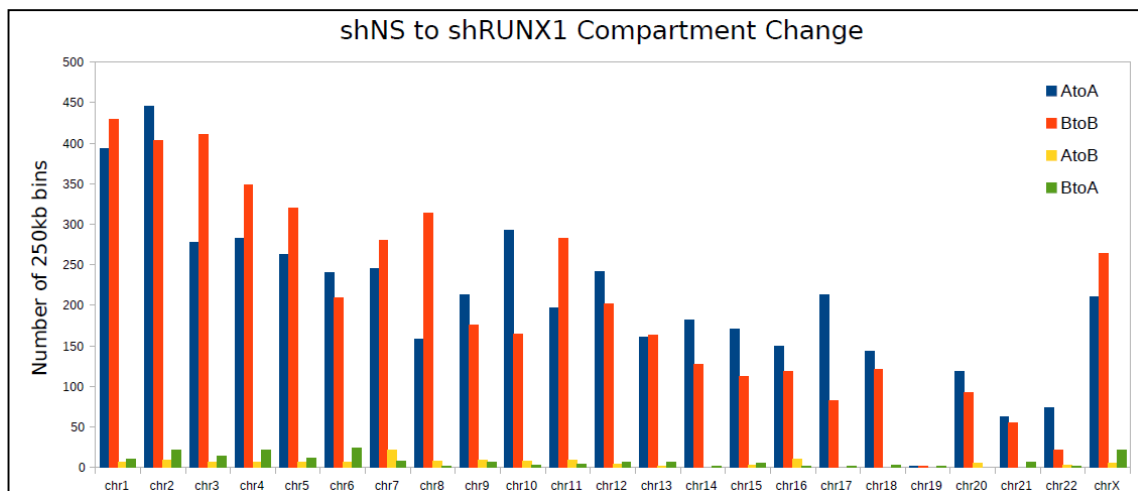
**Figure 5.27.** Schematic representation of the genomic compartment switching at each non-overlapping bin.

Even though RUNX1 depletion resulted in hundreds of differentially expressed genes, the majority of the compartments were similar (stable) between the shNS and shRUNX1 datasets (Figure 5.28). Only a few percent of the genome showed a compartmental alteration from A-type to B-type and vice versa upon RUNX1 knockdown (Figure 5.28).

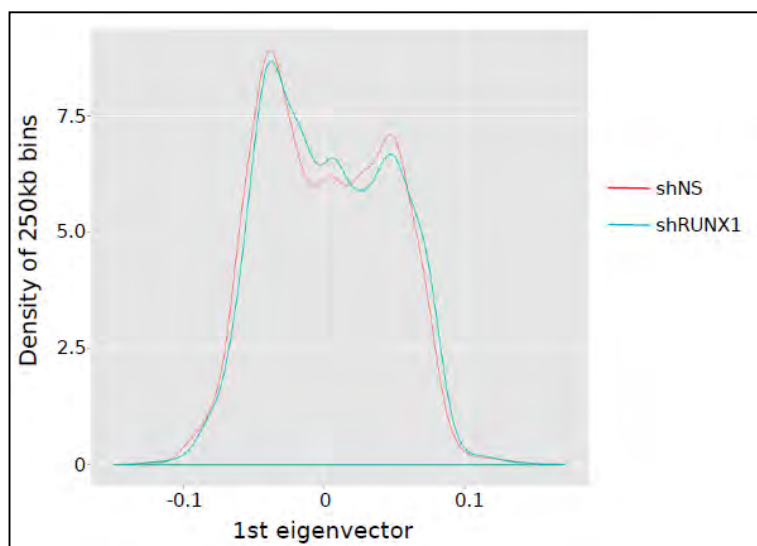


**Figure 5.28. Pie chart showing the genomic compartment changes between shNS and shRUNX1 datasets.** “A” and “B” denotes the open and closed compartments, respectively. “A to A” represents compartments that are open in both cell lines, “B to B” represents compartments that are closed in both cell lines, “A to B” denote compartments that are open in shNS but closed in shRUNX1, and “B to A” denotes compartments that are closed in shNS and open in shRUNX1.

The change in compartmentalization was homogeneous throughout the chromosomes, rather than being localized to a few hot spots (Figure 5.29). The distribution of the compartment strength (the 1<sup>st</sup> eigenvector) was similar in shNS and shRUNX1 datasets, suggesting that the observed differences are indeed due to changes in specific interactions (Figure 5.30).



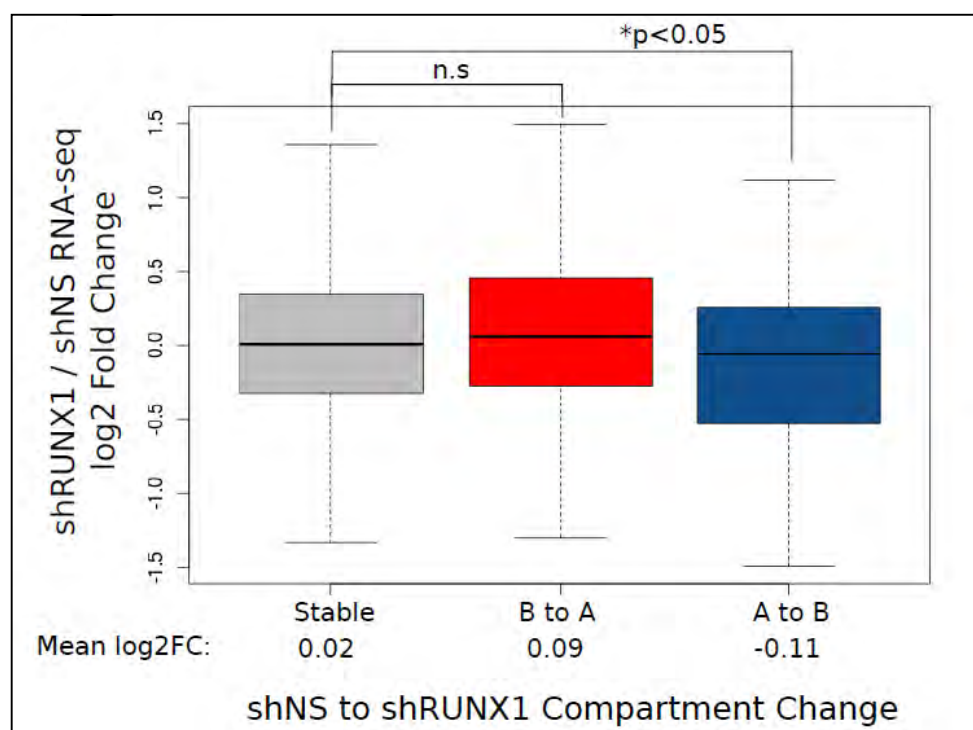
**Figure 5.29.** Bar graph showing the number of stable and altered compartments for each chromosome between shNS and shRUNX1 Hi-C datasets. The switch in compartmentalization is spread throughout the chromosomes, instead of localizing in a few hotspots.



**Figure 5.30.** Distribution of the 1st eigenvalues for each 250kb for shNS and shRUNX1 datasets, showing similar distributions.

Although the majority of the compartments were similar, changes in gene expression were correlated with altered compartment regions. The regions that switched from an open A-type compartment to a closed B-type compartment upon RUNX1 depletion

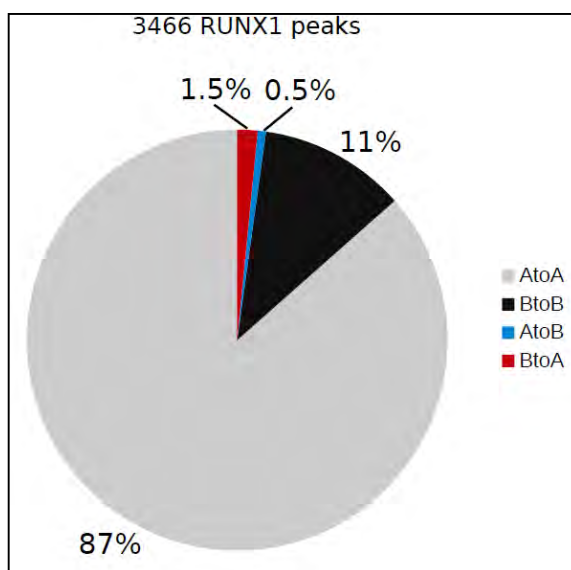
showed significantly more down-regulation of gene expression than the genes in the stable compartments (A-type to A-type and B-type to B-type) (Figure 5.31). Even though not significant, genes in compartments that switched from the B-type to A-type compartment switch regions upon RUNX1 loss showed a trend for up-regulation of gene expression as assessed by the mean log<sub>2</sub> fold change values (Figure 5.31).



**Figure 5.31. Transcriptional changes are correlated with compartment changes.** shRUNX1 / shNS log<sub>2</sub> fold change RNA-seq expression box plot of all the genes residing at regions for different compartmental switch categories. The compartments that are switched from A to B and from B to A show decreased and increased expression levels, respectively. p-value: Wilcoxon rank-sum test.

Next, we assessed whether RUNX1 binding was associated with the compartment switch regions. We intersected the RUNX1 peaks and observed that the majority (87%) of the RUNX1 bound regions were in the stable A-type open compartments (Figure 5.32). In contrast, only 11% of the RUNX1 peaks were in the

stable closed B-type compartment regions. Only ~3% of the RUNX1 binding sites were associated with altered compartments (Figure 5.32).

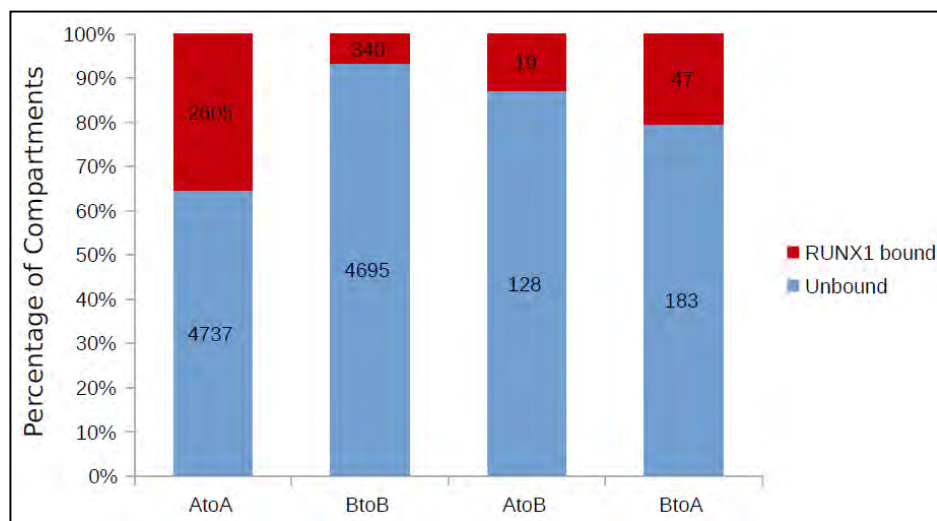


**Figure 5.32. Pie chart showing the percentage of the compartment switch regions that are bound by RUNX1.** The colored portions of the graph denote the RUNX1 bound percentage of each compartment change category.

Furthermore, we analyzed the percentage of each type of compartment that contained one or more RUNX1 ChIP-seq peaks (Figure 5.33). RUNX1 binding was most pronounced at “A to A” stable open compartments, and least enriched in “B to B” stable closed compartments. Interestingly, even though the regions showing “A to B” and “B to A” switching represented only a small percentage of the total (Figure 5.28), we observed comparable relative enrichment of RUNX1 binding in the regions with a compartment switch (Figure 5.33).

Taken together, these results suggest that RUNX1 depletion results in a compartment switch of a small subset of genomic regions that is associated with

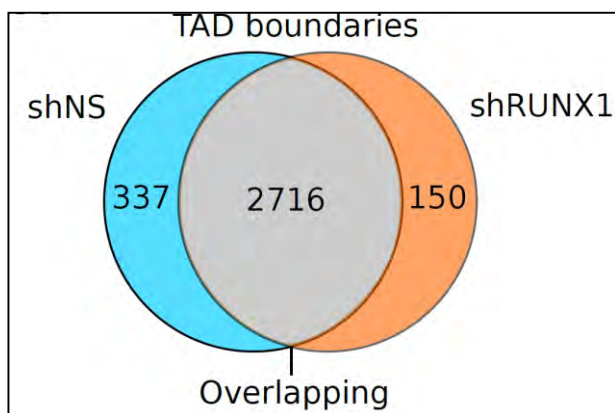
RUNX1-dependent differential gene expression. Moreover, RUNX1 binding is similarly associated with each compartment switch type, and is more enriched at open compartments than at closed compartments.



**Figure 5.33. Bar chart showing the percentage of stable and altered compartments that are RUNX1 bound.**

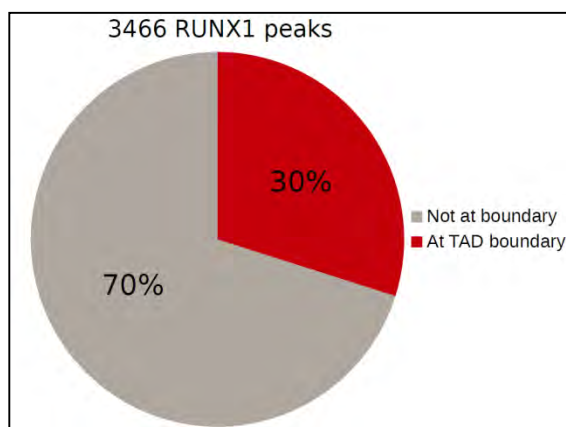
### 5.3.5 RUNX1 is enriched at TAD boundaries

Each genomic compartment consists of 100kb to 1Mb scale TADs. TADs have been shown to be largely invariant across tissues, species and in tumorigenesis [73; 74; 78; 263]. We asked if RUNX1 depletion resulted in any differential TAD formation. To address this, we generated insulation plots for each chromosome and calculated the TAD boundaries using the 40kb shNS and shRUNX1 Hi-C datasets (see Methods and [100; 263]). The majority (~92%) of the TAD boundaries overlapped when RUNX1 was knocked down (Figure 5.34).



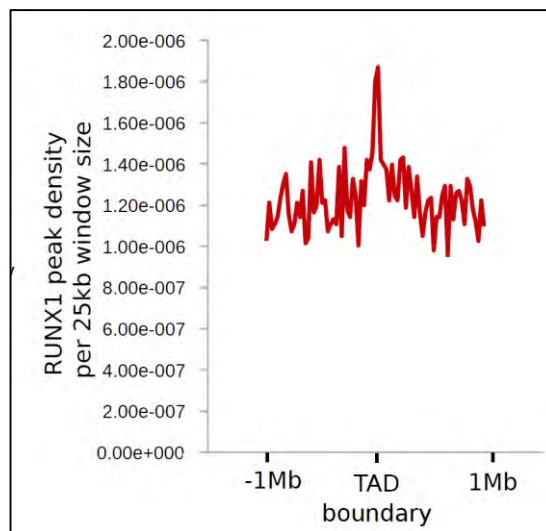
**Figure 5.34.** Venn diagram showing that the TAD boundaries are largely similar between shNS and shRUNX1 Hi-C datasets.

In order to determine whether RUNX1 binding was associated with TAD borders, we calculated the RUNX1 peak overlap with TAD boundaries. Thirty percent of all RUNX1 bound sites were located on TAD boundaries (Figure 5.35).

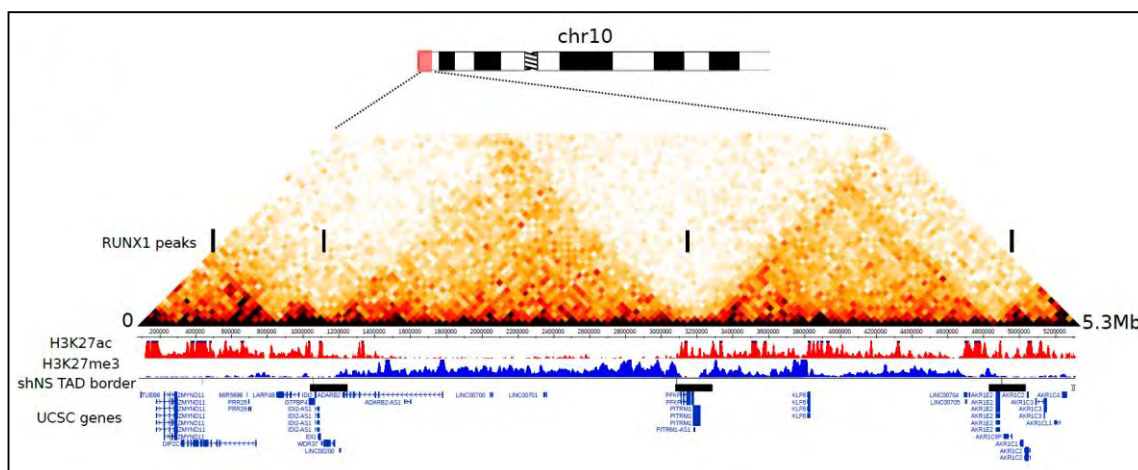


**Figure 5.35.** Pie chart showing the percentage of RUNX1 localization at TAD boundaries.

When we analyzed the RUNX1 peak density across ( $\pm$  1Mb) the TAD boundaries, we observed an enrichment of RUNX1 binding at TAD boundaries, suggesting that RUNX1 might function at TAD boundaries (Figure 5.36 and Figure 5.37).



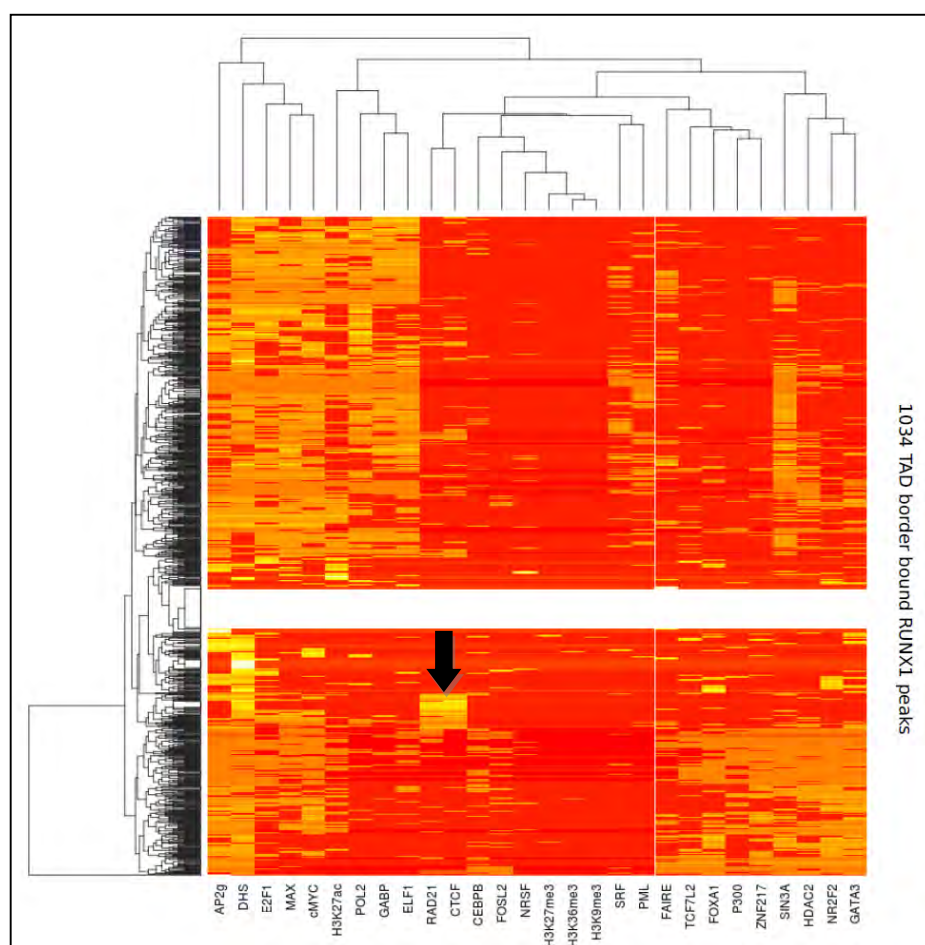
**Figure 5.36.** The frequency plot of RUNX1 ChIP-seq peaks per 25kb for +/- 1Mb of every TAD boundary.



**Figure 5.37.** An example region on chr10 showing three different TADs, along with their chromatin states assessed by H3K27ac and H3K27me3 ENCODE ChIP-seq tracks, shNS TAD borders and RUNX1 ChIP-seq peak binding at the TAD border regions.



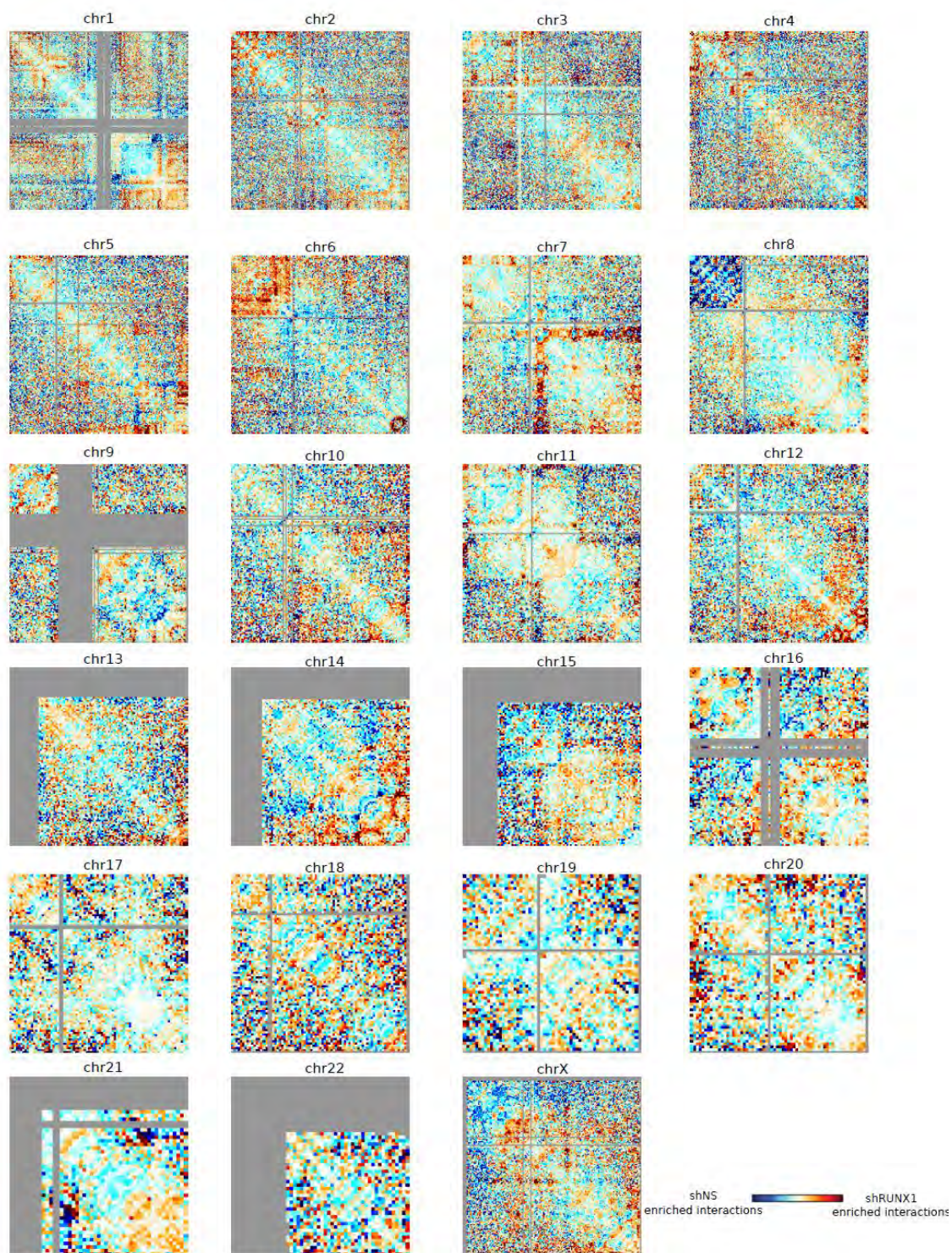
Overlapping the RUNX1 peaks that were located on TAD boundaries with the MCF-7 ENCODE data revealed clustering of several factors (Figure 5.38). Interestingly, unlike the clustering of all the RUNX1 peaks overlaps (Figure 5.18), RAD21 (cohesin) and CTCF were clustered together, overlapping with a subset of RUNX1 peaks (Figure 5.38). This suggests a combinatorial role of RUNX1 with other co-regulators previously implicated as regulators of genome structure at TAD boundaries.



**Figure 5.38. A subset of RUNX1 peaks overlap with CTCF and cohesin at TAD boundaries.** A hierarchically clustered heatmap showing the co-overlap TAD border bound RUNX1 peaks with open chromatin regions (FAIRE and DHS) and available ChIP-seq peaks from MCF-7 ENCODE data. A portion of RUNX1 peaks overlap with CTCF and cohesin (black arrow).

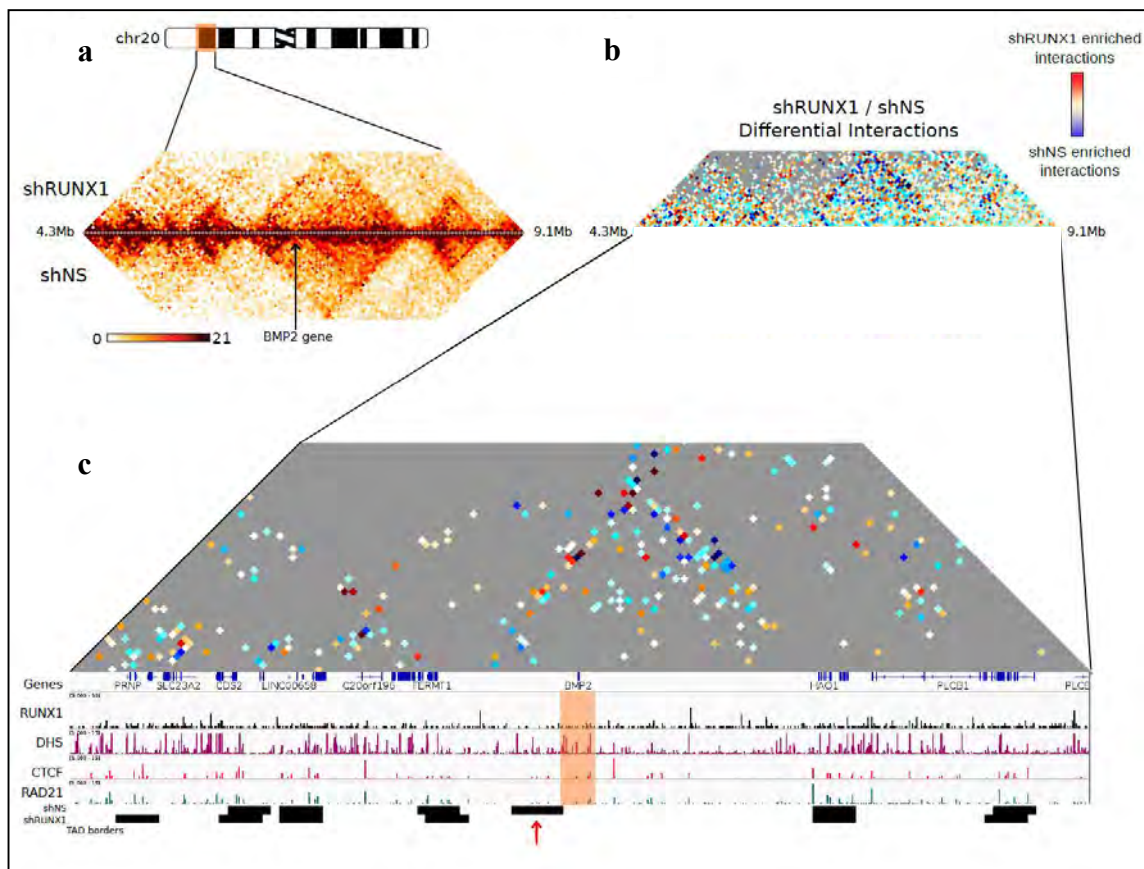
### **5.3.6 RUNX1 knockdown results in emergence and disruption of many long-range interactions**

To assess the alterations in long-range chromatin interactions upon RUNX1 knockdown, we compared the genome-wide interaction differences across all chromosomes (see Methods). Analysis of the interactions that are altered upon RUNX1 knockdown revealed an extensive network of disruption and emergence of novel interactions on each chromosome (Figure 5.39).



**Figure 5.39.** The z-score differences of shRUNX1 / shNS Hi-C matrices for each chromosome at 1Mb bin size.

For instance, the BMP2 gene, which is important for epithelial cell transformation [314], is down-regulated by log2 fold change of -1.55 upon RUNX1 knockdown. The BMP2 gene resides inside a sub-TAD, near a TAD border on chromosome 20 (Figure 5.40a). Upon RUNX1 knockdown, we observed extensive long-range chromatin interaction alterations within +/- 5Mb of the BMP2 locus (Figure 5.40b). Visualizing the significantly altered interactions revealed interesting changes in the higher-order chromatin structure of this locus. The BMP2 locus is flanked by two RUNX1 bound regions, harbors several DNase1 hypersensitive sites, and is flanked by CTCF and RAD21 (cohesin) binding (Figure 5.40c). However, upon RUNX1 knockdown, the TAD boundary adjacent to the BMP2 gene region is lost (Figure 5.40c), accompanied by a loss of long-range interactions between the two TAD borders flanking the TAD border that was lost. Taken together, these results suggest a prominent role for RUNX1 in genome architecture.



**Figure 5.40. An example of interaction changes upon RUNX1 knockdown.** **a)** Genome wide interaction heatmap of the BMP2 gene locus indicated by an arrow. +/- 5Mb on chr20 chr20:4388564-9118290. The top and the bottom parts of the heatmap show the interaction frequency of the shRUNX1 and shNS cells, respectively. **b)** Heatmap of the same region showing the differential interactions between shRUNX1 and shNS. The red and blue pixels denote interactions that are gained and lost upon RUNX1 knockdown, respectively. **c)** Heatmap showing the significantly altered interactions around the BMP2 gene locus. In the lower panel, the UCSC genes, MCF-7 ChIP-seq signal for RUNX1 from this study, ChIP-seq for CTCF, RAD21 as well as Dnase1 hypersensitivity from the ENCODE database, and the TAD borders for shNS and shRUNX1 cells are shown. The highlighted region indicates the BMP2 gene. The TAD border near the BMP2 gene is lost upon RUNX1 knockdown (red arrow), and the interactions between the surrounding borders are decreased.

## 5.4 Discussion

RUNX1 protein can act as a both tumor suppressor and an oncogene depending on the cellular context (reviewed in [315; 316]). Although much is known about the role of RUNX1 in breast tumor prognosis [291; 292], its role in mediating gene expression and affecting higher-order genome architecture remains unknown. Here, we probed RUNX1-mediated genome-wide transcriptional and architectural changes in the MCF-7 breast cancer cells. Understanding the molecular consequences of RUNX1 loss is important in delineating the cell context-specific roles of RUNX1 in breast cancer.

RUNX1 depletion in the MCF-7 cell line resulted in down- and up-regulation of many genes, suggesting both an activator and a repressor role for RUNX1. Associating RUNX1 localization with other MCF-7 ENCODE datasets suggested that RUNX1 binds to the chromatin with different transcriptional regulators and cofactors, likely as part of multiple distinct complexes (Figure 5.17). Therefore, the dual transcriptional role of RUNX1 most likely depends on its interaction partners, which is consistent with prior findings that the RUNX1 protein harbors many partner proteins [308]. RUNX1 is primarily bound to gene bodies (~70%) and to a lesser extent to intergenic regions (Figure 5.13). This suggests a direct regulation of RUNX1-mediated gene expression (Figure 5.16). RUNX1 depletion resulted in the down-regulation cell surface connections and extracellular matrix (Figure 5.6); and up-regulation of several chromatin modifying genes related to DNA methylation, nucleosome and centrosome assembly (Figure 5.7). All of these categories are tightly associated with tumorigenesis. As a result, it can be hypothesized that RUNX1 loss in different cellular contexts may

have distinct molecular consequences, therefore distinct phenotypic outcomes. This can partly explain the dichotomous role of RUNX1 in different contexts.

From an architectural point of view, Hi-C analysis revealed that the vast majority of both the compartments and the TADs are maintained upon RUNX1 depletion (Figure 5.34). Even though only a small fraction of the genome displayed an alteration in genomic compartments, the changes in gene expression were correlated with the compartmental changes for a subset of genes and regions. This implies a strong relationship between genomic compartmentalization and gene expression. It is still not very well understood whether compartment switching affects gene expression or vice versa. There is literature evidence favoring both cases, where the relocation of a gene inside the nucleus precedes its transcription [317], or alternatively chromatin decondensation (i.e gene positioning) plays a major role in gene expression [221].

In the three dimensional genome, TADs are regarded as the functional transcriptional units where the genes inside a given TAD are regulated differently compared to the genes inside another TAD [75]. Remarkably, we identified an enrichment of RUNX1 binding at TAD boundaries (Figure 5.36). This is consistent with the fact that TAD boundaries are also enriched for genes [73]. It has been shown that, in flies, combinatorial binding of multiple architectural proteins are associated with TAD boundary strength [79]. In vertebrates, the insulators cohesin, condensin, and especially CTCF, were shown to be strongly associated with TAD boundaries and looping formation [78; 81]. There are ~1400 sequence specific transcription factors in the human genome [268]. Therefore, it is possible that, either directly or indirectly,

RUNX1 may also contribute to the TAD structure as a genome regulator along with these other important insulators. Finally, RUNX1 loss resulted in the disruption and emergence of thousands of individual interactions, demonstrating a structural role for RUNX1 in the organization of chromatin structure throughout the genome (Figure 5.39 and Figure 5.40).

Taken together, these results provide insight into how several cellular processes, including transcription and genomic organization, are perturbed upon RUNX1 loss. Further studies assessing the architectural consequences of RUNX1 loss in different tumor subtypes will shed light into the underlying mechanisms of RUNX1 action in breast cancer.



## **CHAPTER 6: Outlook and Future Directions**

### **6.1 Long-range gene regulation and higher-order genome structure**

The transcriptional programs that establish and maintain specific cell states are orchestrated by the binding of regulatory proteins to specific genomic elements [318]. There are thousands of enhancers within a given cell [312; 319]. The mechanisms of how individual enhancers regulate the expression of their target genes in a specific manner is not fully understood. Insulator proteins are known to block enhancer function across an insulator element [320]. Apart from regulatory proteins, it is now widely accepted that long non-coding RNAs are also a fundamental part of transcriptional regulation and genome organization [321; 322]. For instance, the expression of non-coding RNAs from enhancer regions, known as eRNAs [323], orchestrated with the binding of chromatin modifiers and transcription factors (reviewed in [324]), result in the long-range looping interaction with their cognate gene promoter [94]. Several recent studies suggest that the transcriptional regulation occurs in the context of three-dimensional chromatin structures. Understanding the patterns of mechanisms and underlying principles in this molecular “lego” will be a long-lasting vocation for many scientists. It is becoming increasingly evident that studying a physiological process in the context of a linear genome is not sufficient to elucidate the full picture of the nuclear processes. Therefore, it becomes vital to understand the higher-order chromatin structure during physiological and disease states. The key question is: What is the difference in the genome structure that results in massive changes in cellular physiology,

including the morphology and phenotype of the cells, the transcriptional landscape and the different characteristics that the cell acquire?

In this thesis, I aimed to understand the higher-order structural chromatin structure alterations using bone differentiation and breast cancer as model systems (Chapters 2 and 3). Moreover, in the context of breast cancer, I studied the architectural role of the ATPase subunit of the SWI/SNF chromatin remodeling complex, BRG1 (Chapter 4); and a transcription factor, RUNX1 (Chapter 5), both of which have been implicated in breast tumorigenesis, and identified previously unknown effects for these factors in three-dimensional genome organization.

## **6.2 Implications in bone biology and bone metastasis**

The findings in Chapter 2 provides an important foundation for future studies regarding bone-related diseases and cancer metastasis to bone. Our 3C analysis have identified dynamic three-dimensional interactions between the *Runx2-P1* and the syntenic *Supt3h* promoters. I have furthermore presented a mechanistic insight into the effect of this looping interaction on *Runx2-P1* expression. These results provide a model for many dimensions in higher-order genome structure. Firstly, from the developmental point of view, the *Runx2* gene expression stands high in the hierarchy in the transcriptional cascade, as *Runx2* null embryos are impaired in bone development. Our 3C work provides the first framework in which the higher-order genome structure affects the gene expression. The osteogenic progenitor cells appear in fusiform, fibroblast cell-like morphology; however upon osteogenic differentiation they start to mineralize and their size is significantly reduced, forming the cobblestone-like

structures. It is of great interest to find out what nuclear structural changes accompany these dramatic morphological changes. This study, in the future, can be followed by performing more high-throughput chromosome conformation capture approaches (4C, 5C and Hi-C) to obtain a bigger picture of the nuclear architecture in conjunction with transcriptional changes. To obtain an alternate view of genomic structure during osteogenesis, ChIA-PET experiments using an antibody against the RUNX2 protein can be performed, thereby providing a RUNX2-centric point of view of the genome at high resolution. The resulting RUNX2-bound interactome dataset can be combined with other available ChIP-seq and DNase1 hypersensitivity data [167] to infer the underlying mechanisms of bone differentiation-dependent changes in the transcriptional cascade and chromatin structure. These approaches need not only be followed during osteoblastogenesis. Several bone related diseases, such as cleidocranial dysplasia, harbor mutations or contain single nucleotide polymorphisms in the *Runx2* gene locus. Even though certain mutations fall in the exonic regions and cause a change in the RUNX2 protein structure leading to perturbed protein functions, many other mutations are either in the introns or in other cis-regulatory regions of the *Runx2* gene. By performing 3C-based approaches in patient cells and utilizing the current findings in this thesis as a reference point may help us decipher the underlying causative effects of these mutations on the disease prognosis. With the emerging more precise CRISPR-Cas9 genome editing techniques [325; 326], there is a therapeutic potential in correcting the malfunctioning genomic regions of the *Runx2* locus.

These possible approaches outlined are not only deemed for the contexts of differentiation and disease. RUNX2 is significantly implicated in several types of cancer, and especially cancer metastasis to bone. A part of my study in Chapter 2, I probed the long-range interactions of *Runx2-P1* *in silico* in various cancer cell lines. Further experiments comparing the interactions between non-metastatic cancer cells and cells metastasizing to the bone, where *Runx2* is frequently over-expressed, may yield prognostic biomarkers in the future.

From an evolutionary point of view, the *Runx2* gene locus is an interesting locus, where the spatial distance between *Runx2-P1*, *Runx2-P2* and *Supt3h* promoters are conserved from sponges to humans. In this study we show that, at least in mice and humans, there is a steady state detectable interaction frequency between the *Runx2-P1* and *Supt3h* promoters. It is very possible that this spatial proximity appeared early in the evolutionary tree, which can be followed up by performing 3C between these two regions in a myriad of species not limited to model organisms. This finding also hints at the higher-order structural relationships between the homologous genomic regions in different species. It is known that the TAD structures are conserved in homologous regions in different species [78]. Therefore, the outstanding question is: can we say whether the local looping interactions at tightly linked loci stable across evolution? A supporting evidence comes from a recent study applying the 4C technique in flies and humans showing that the enhancer-promoter looping interactions are more dynamic and cell-type specific in developmentally regulated genes (example: the  $\beta$ -globin gene), but more stable in housekeeping gene loci [327].

### 6.3 Implications in breast cancer

Cancer is pathologically characterized by irregularly shaped nuclei, and the pattern of the heterochromatin / euchromatin distribution inside the nucleus is perturbed. Understanding the molecular alterations in the cancer nucleus has been and still is a long-standing challenge involving scientists from several fields of biology and medicine. There are numerous studies charting the mutational, transcriptional and epigenomic landscape between the normal and cancer cells. Emerging trends and patterns of regulations from these studies result in many therapeutic agents to increase the prognosis of the disease. For instance, histone demethylase or bromo-domain inhibitors have been successfully used to treat a subset of cancer patients [328; 329]. Thus, a deeper understanding of the cancer-specific alterations in the nucleus is of great importance in developing new treatments.

In Chapter 3 of this thesis, I present for the first time in the literature, a genome-wide molecular view of differential genomic-organization between normal and cancer cells. We observed extensive differential interactions in small chromosomes, and at telomeres between normal and breast cancer cells. These changes were accompanied by the transcriptional differences, thereby providing a nice evidence of the connection between gene expression and nuclear architecture. Even though this study involves two cell lines commonly used to study the progression of breast cancer, further experiments including more cell lines with diverse hormonal backgrounds (ER, PR and HER markers) will yield insight into biology of breast cancer models. The Hi-C technique allows us to obtain interaction data from as low as ~5 million cells, in which the

resulting data will have enough resolution to obtain information about the chromosomal interactions and the genomic compartments. Therefore, further extending these findings with primary tumors, and their matched controls will no doubt shed light on the common architectural abnormalities in breast cancer. Another finding in Chapter 3 is the differential telomeric interactions between normal and cancer cells. Telomere biology is a growing field with severe implications in cellular transformation, and have been frequently implicated in cancer. In this thesis, alterations in telomeric interactions provide a fundamental resource for future studies on a number of aspects, including telomere length and prognosis, the effect of telomere length in telomere clustering, and different telomere maintenance mechanisms including the alternative lengthening of telomere (ALT) mechanism.

More importantly, the Hi-C datasets provided in Chapter 3 can be utilized in a myriad of ways. First, one can integrate SNP and copy number variation analysis of these cells and examine whether higher-order structure is associated with common or unique genomic signatures. Furthermore, the MCF-7 breast cancer cell is part of the ENCODE consortium, and there are many different types of genome-wide epigenomic marks readily available in these cells. Further meta-analysis of the epigenomic signatures integrated with the expression and chromatin interaction datasets would provide mechanistic perspective about how the binding of different factors, the presence of different histone modifications and the chromatin accessibility is affected by or affect higher-order chromatin structure.

In Chapter 5, we characterize the consequences of RUNX1 perturbation in breast cancer cells. The RUNX1 transcription factor is most studied in the context of leukemia. However, RUNX1 is also a significant contributor to breast cancer and the molecular effects of its perturbation is not well-understood. Here, I presented RUNX1-dependent transcriptional and structural changes, and how these changes associated with RUNX1 binding in breast cancer cells. Even though there are some studies which demonstrate the effects of RUNX1 in the tissue or the organismal level in the context of breast cancer, this study stands out as one of the first molecular characterization of RUNX1 deficiency in breast cancer. Future studies dissecting the causative effects of ectopic RUNX1 binding, or disruption of RUNX1 binding at target loci, especially in relationship with its co-regulatory interaction partners, will help us understand the exact roles of RUNX1. For instance, performing a mass spectrometry analysis of RUNX1, followed by ChIP-seq analysis of the RUNX1-interacting partners that bind to DNA would yield significant views in transcriptional regulation and chromatin structure that is RUNX1 dependent. Finally, by combining the Hi-C data from the MCF-7 cells in Chapter 3 and Chapter 5, one can achieve a very high resolution (<20kb) interaction map of the MCF-7 cells. A similar approach has been used previously by combining the Hi-C data from TNFalpha treated and control fibroblast cells [170]. The same approach for achieving high resolution Hi-C data can also be followed for BRG1 and MCF-10A cells. As a result, the findings in this thesis open up new avenues for many different types of basic and therapeutic research projects.

#### **6.4 Implications in the biology of TAD structures**

One of the main results in Chapters 3, 4 and 5 of this thesis is that the TAD boundaries are largely invariant between normal epithelial and breast cancer cells following BRG1 and RUNX1 depletion, respectively. These results are consistent with earlier reports that the TAD boundaries are largely invariant across species, physiological cues, tumorigenesis, and factor depletion including histone H1 [73; 76; 77; 98; 263; 330]. Even though the genes inside a single TAD are co-regulated, the formation and disappearance of TAD boundaries is not necessarily correlated with gene expression [73; 74; 98; 263]. A recent study in fly cells in which the global gene expression has been shut off by heat shock demonstrates that the localization of the TAD boundaries is not altered; rather the strength of the boundaries is decreased, allowing more inter-TAD interactions [80]. It has been suggested that the TADs are the principle units of DNA replication [72]. The biology of TAD structures is still not very much understood. In flies, it was shown that the localization of architectural binding proteins at TAD boundaries affected the strength of a TAD boundary, as demonstrated by the presence of increased number of architectural binding proteins at stronger TAD boundaries [79; 267]. The definition of a TAD boundary is a stretch of a genomic region, ranging from a few to tens of kilobases, rather than a localized spot. Thus, it is still unclear what underlying characteristics establish a TAD boundary. It is known that the presence of genes, active histone marks, and certain transcription factors are enriched across the TAD boundaries. Is the enrichment of these characteristics at localized regions sufficient enough to generate a TAD boundary? If one used the



CRISPR/Cas9 technique to cut a defined TAD boundary and paste it into a non-boundary region, would it create a new TAD boundary? How would the expression of the genes proximal to the new TAD boundary be affected? What is the localization (i.e. radial positioning) of the TAD boundaries in the nucleus? Is the separation of two neighboring TADs a result of an active biological consequence of protein (and lncRNA) binding at/near gene bodies, or a passive biophysical separation of different genomic regions, similar to the notion of nuclear bodies? These are all outstanding questions regarding TAD boundaries and their biological roles as they are conserved in evolution, evidenced by the presence of TADs, chromatin interaction domains (CIDs) and gene crumples in worm, bacteria and yeast, respectively [171; 213; 331].

A growing evidence of TAD boundary formation comes from the meta-analysis of CTCF bound sites in the genome. It was shown by multiple groups that the convergent binding of CTCF is very strongly associated with TAD boundaries [81; 217; 332; 333]. In this thesis, I showed that RUNX1 and BRG1 are enriched at TAD boundaries. Following the depletion of BRG1, a reduction in TAD boundary strength was observed (Chapter 4). Given complexity of the protein repertoire of the nucleus at any given time, and given the regulatory mechanisms of many proteins at the post-translational level, the complex architectural regulation at TAD boundaries likely involves binding of several proteins. This can be achieved dependent or independent of the recruitment of CTCF to TAD boundaries by co-regulatory proteins. The multiplicity and the redundancy in transcription factor binding especially at TAD boundaries may explain why the localization of TADs are not that easy to disrupt.

Rao *et al.* suggests an alternative view of TADs, claiming that each individual TAD is a single chromatin looping event and the intervening enriched interactions inside a TAD appear as a consequence of closer proximity resulting from the looping of the two elements located at the boundaries. They also suggest that instead of millions, there are only ~10,000 interaction events in the human genome [217]. An appropriate analogy would be to fold the ends of a piece of paper, where the ends represent the TAD boundaries, and as a result all the mid-sections of the piece of paper are now in proximity in 3D space. This phenomenon is evidenced by the fact that in certain TADs, there is a significantly elevated single interaction event between the boundaries of the TAD, and the rest of the interactions inside the same TAD are seen as an averaged interaction pattern, rather than sharp, discrete spots [217]. Even though this hypothesis may be true for a subset of TADs, there are several TADs without such a prominent discrete boundary-to-boundary interactions [217; 263]. Therefore, the proposed hypothesis does not necessarily explain the second type of TAD structures. It may be that the genome contain two different types of TAD structures. It was recently predicted by computational modeling that the chromatin acquires diverse fluctuating configurations inside a TAD rather than displaying stable structures [334]. A more recent study investigated the higher-order interactions that the TADs make with each other, and identified hierarchical structures with domains-within-domains which they termed “meta-TADs” [335]. The meta-TADs are stable during differentiation and are correlated with epigenetic signatures and gene expression, and the tight packaging of the genome without losing contact-specificity [335]. With all these important findings at

hand, it is still under discussion whether it is the dynamic interactions inside a TAD that form the topologically associating domains, or is it the molecular and physical barriers at the TAD boundaries that form them [82]. I propose a possible alternative view for TADs, that each TAD, in conjunction with the interactions the TAD makes with other TADs (meta-TADs), may represent different nuclear bodies in the nucleus. A testable hypothesis is provided by two recent studies where the authors determine the viscosity (or phasing) of the nuclear body emanating from the introduction of an RNA-protein complex [336; 337]. Therefore, it could be possible that each TAD, and its interacting TAD partners, may constitute a separate nuclear body in the nucleus. This hypothesis would be consistent with the correlated expression of genes inside a TAD (and other TADs), and the notion of transcription factories [338], splicing sites [339], and many other nuclear bodies such as the RUNX2 nuclear bodies, or any protein that has a “puncta” pattern of immunofluorescence staining. The findings in this thesis add significant insights into the missing TAD puzzle, by providing evidence that chromatin modifiers other than well-studied insulators can also in fact affect TAD boundaries. This important finding opens up new exploratory avenues for other factors that potentially affect TAD boundaries.

## REFERENCES

- [1] Schwann, T. (1839). *Microscopical Researches into the Accordance in the Structure and Growth of Animals and Plants*, H. Smith. London .
- [2] Flemming, W., 1882. *Zellsubstanz, kern und zelltheilung*. Vogel, .
- [3] Hughes, A., 1959. *A History of Cytology: Arthur Hughes....* Abelard-Schuman, .
- [4] Paweletz, N. (2001). *Walther Flemming: pioneer of mitosis research*, Nature Reviews Molecular Cell Biology 2 : 72-76.
- [5] Winkelmann, A. (2007). *Wilhelm von Waldeyer-Hartz (1836-1921): An anatomist who left his mark*, Clinical Anatomy 20 : 231-234.
- [6] Avery, O. T.; MacLeod, C. M. and McCarty, M. (1944). *Studies on the chemical nature of the substance inducing transformation of pneumococcal types induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III*, The Journal of experimental medicine 79 : 137-158.
- [7] Miescher-Rüsch, F., 1871. *Ueber die chemische Zusammensetzung der Eiterzellen.* ,
- [8] Kossel, A., 1911. *Ueber die chemische Beschaffenheit des Zellkerns..* , .
- [9] Bradbury, S. (1989). *Landmarks in biological light microscopy*, Journal of Microscopy 155 : 281-305.
- [10] Van Holde, K. E., 2012. *Chromatin*. Springer Science & Business Media, .
- [11] Franklin, R. E. and Gosling, R. G. (1953). *Molecular configuration in sodium thymonucleate*, Nature 171 : 740-741.
- [12] Watson, J. D.; Crick, F. H. and others (1953). *Molecular structure of nucleic acids*, Nature 171 : 737-738.
- [13] Wilkins, M. H. F.; Stokes, A. R. and Wilson, H. R. (1953). *Molecular structure of nucleic acids: molecular structure of deoxypentose nucleic acids*, Nature 171 : 738-740.
- [14] Gall, J. G. (1963). *Kinetics of deoxyribonuclease action on chromosomes*, Nature 198 : 36-38.
- [15] Johns, E. (1969). *The histones, their interactions with DNA, and some aspects of gene control*, Ciba Foundation Symposium - Homeostatic Regulators: 128-143.
- [16] Allfrey, V.; Faulkner, R. and Mirsky, A. (1964). *Acetylation and methylation of*

*histones and their possible role in the regulation of RNA synthesis*, Proceedings of the National Academy of Sciences of the United States of America 51 : 786.

[17] Luger, K.; Mäder, A. W.; Richmond, R. K.; Sargent, D. F. and Richmond, T. J. (1997). *Crystal structure of the nucleosome core particle at 2.8 Å resolution*, Nature 389 : 251-260.

[18] Woodcock, C. L. and Dimitrov, S. (2001). *Higher-order structure of chromatin and chromosomes*, Current opinion in genetics & development 11 : 130-135.

[19] Tremethick, D. J. (2007). *Higher-order structures of chromatin: the elusive 30 nm fiber*, Cell 128 : 651-654.

[20] Hansen, J. C. (2012). *Human mitotic chromosome structure: what happened to the 30-nm fibre?*, The EMBO journal 31 : 1621-1623.

[21] Koshland, D. and Strunnikov, A. (1996). *Mitotic chromosome condensation*, Annual review of cell and developmental biology 12 : 305-333.

[22] Cremer, M.; von Hase, J.; Volm, T.; Brero, A.; Kreth, G. and Walter, J. (2001). *Non-random radial higher-order chromatin arrangements in nuclei of diploid human cells*, Chromosome Res. 9 : 541-567.

[23] Cremer, T.; Cremer, M.; Dietzel, S.; Muller, S.; Solovei, I. and Fakan, S. (2006). *Chromosome territories--a functional nuclear landscape*, Curr Opin Cell Biol. 18 : 307-316.

[24] Cremer, T. and Cremer, M. (2010). *Chromosome territories*, Cold Spring Harb Perspect Biol. 2 : a003889.

[25] Kreth, G.; Finsterle, J.; Von Hase, J.; Cremer, M. and Cremer, C. (2004). *Radial arrangement of chromosome territories in human cell nuclei: a computer model approach based on gene density indicates a probabilistic global positioning code*, Biophysical journal 86 : 2803-2812.

[26] Versteeg, R.; van Schaik, B. D.; van Batenburg, M. F.; Roos, M.; Monajemi, R.; Caron, H.; Bussemaker, H. J. and van Kampen, A. H. (2003). *The human transcriptome map reveals extremes in gene density, intron length, GC content, and repeat pattern for domains of highly and weakly expressed genes*, Genome research 13 : 1998-2004.

[27] Dietzel, S.; Zolghadr, K.; Hepperger, C. and Belmont, A. S. (2004). *Differential large-scale chromatin compaction and intranuclear positioning of transcribed versus non-transcribed transgene arrays containing  $\beta$ -globin regulatory sequences*, Journal of

cell science 117 : 4603-4614.

[28] Pickersgill, H.; Kalverda, B.; de Wit, E.; Talhout, W.; Fornerod, M. and van Steensel, B. (2006). *Characterization of the Drosophila melanogaster genome at the nuclear lamina*, Nature genetics 38 : 1005-1014.

[29] Fraser, P. and Bickmore, W. (2007). *Nuclear organization of the genome and the potential for gene regulation*, Nature 447 : 413-417.

[30] Finlan, L. E.; Sproul, D.; Thomson, I.; Boyle, S.; Kerr, E.; Perry, P.; Ylstra, B.; Chubb, J. R.; Bickmore, W. A. and others (2008). *Recruitment to the nuclear periphery can alter expression of genes in human cells*, PLoS Genet 4 : e1000039.

[31] Belmont, A. S.; Zhai, Y. and Thilenius, A. (1993). *Lamin B distribution and association with peripheral chromatin revealed by optical sectioning and electron microscopy tomography.*, The Journal of cell biology 123 : 1671-1685.

[32] Busch, H., 2012. *The cell nucleus*. Elsevier, .

[33] Branco, M. R. and Pombo, A. (2007). *Chromosome organization: new facts, new models*, Trends in cell biology 17 : 127-134.

[34] Cavalli, G. (2007). *Chromosome kissing*, Current opinion in genetics & development 17 : 443-450.

[35] Lewis, E. (1954). *The theory and application of a new method of detecting chromosomal rearrangements in Drosophila melanogaster*, American Naturalist : 225-239.

[36] Wu, C. (1993). *Transvection, nuclear structure, and chromatin proteins.*, The Journal of cell biology 120 : 587-590.

[37] Morris, J. R.; Chen, J.-I.; Geyer, P. K. and others (1998). *Two modes of transvection: enhancer action in trans and bypass of a chromatin insulator in cis*, Proceedings of the National Academy of Sciences 95 : 10740-10745.

[38] Parada, L.; McQueen, P.; Munson, P. and Misteli, T. (2002). *Conservation of relative chromosome positioning in normal and cancer cells*, Curr Biol. 12 : 1692-1697.

[39] Bolzer, A.; Kreth, G.; Solovei, I.; Koehler, D.; Saracoglu, K. and Fauth, C. (2005). *Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes*, PLoS Biol. 3 : e157.

[40] Bickmore, W. A. and Teague, P. (2002). *Influences of chromosome size, gene*

*density and nuclear position on the frequency of constitutional translocations in the human population*, Chromosome research 10 : 707-715.

- [41] Parada, L. A.; McQueen, P. G. and Misteli, T. (2004). *Tissue-specific spatial organization of genomes*, Genome biology 5 : R44.
- [42] Tanabe, H.; Habermann, F. A.; Solovei, I.; Cremer, M. and Cremer, T. (2002). *Non-random radial arrangements of interphase chromosome territories: evolutionary considerations and functional implications*, Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis 504 : 37-45.
- [43] Mayer, R.; Brero, A.; Von Hase, J.; Schroeder, T.; Cremer, T. and Dietzel, S. (2005). *Common themes and cell type specific variations of higher order chromatin arrangements in the mouse*, BMC cell biology 6 : 44.
- [44] Marella, N.; Malyavantham, K.; Wang, J.; Matsui, S.; Liang, P. and Berezney, R. (2009). *Cytogenetic and cDNA microarray expression analysis of MCF10 human breast cancer progression cell lines*, Cancer Res. 69 : 5946-5953.
- [45] Zeitz, M.; Ay, F.; Heidmann, J.; Lerner, P.; Noble, W. and Steelman, B. (2013). *Genomic interaction profiles in breast cancer reveal altered chromatin architecture*, PLoS One. 8 : e73974.
- [46] Tumber, T. and Belmont, A. S. (2001). *Interphase movements of a DNA chromosome region modulated by VPI6 transcriptional activator*, Nature cell biology 3 : 134-139.
- [47] Kuroda, M.; Tanabe, H.; Yoshida, K.; Oikawa, K.; Saito, A.; Kiyuna, T.; Mizusawa, H. and Mukai, K. (2004). *Alteration of chromosome positioning during adipocyte differentiation*, Journal of cell science 117 : 5897-5903.
- [48] Marella, N. V.; Seifert, B.; Nagarajan, P.; Sinha, S. and Berezney, R. (2009). *Chromosomal rearrangements during human epidermal keratinocyte differentiation*, Journal of cellular physiology 221 : 139-146.
- [49] Apostolou, E.; Ferrari, F.; Walsh, R. M.; Bar-Nur, O.; Stadtfeld, M.; Cheloufi, S.; Stuart, H. T.; Polo, J. M.; Ohsumi, T. K.; Borowsky, M. L. and others (2013). *Genome-wide chromatin interactions of the Nanog locus in pluripotency, differentiation, and reprogramming*, Cell stem cell 12 : 699-712.
- [50] Jackson, D. A. and Pombo, A. (1998). *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*, The Journal of cell biology 140 : 1285-1295.

[51] Ma, H.; Samarabandu, J.; Devdhar, R. S.; Acharya, R.; Cheng, P.-c.; Meng, C. and Berezney, R. (1998). *Spatial and temporal dynamics of DNA replication sites in mammalian cells*, The Journal of cell biology 143 : 1415-1425.

[52] Berezney, R. (2002). *Regulating the mammalian genome: the role of nuclear architecture*, Advances in enzyme regulation 42 : 39-52.

[53] Goldman, R. D.; Gruenbaum, Y.; Moir, R. D.; Shumaker, D. K. and Spann, T. P. (2002). *Nuclear lamins: building blocks of nuclear architecture*, Genes & development 16 : 533-547.

[54] Fawcett, D. W. (1966). *On the occurrence of a fibrous lamina on the inner aspect of the nuclear envelope in certain cells of vertebrates*, American Journal of Anatomy 119 : 129-145.

[55] Wilson, K. L. and Foisner, R. (2010). *Lamin-binding proteins*, Cold Spring Harbor perspectives in biology 2 : a000554.

[56] Kind, J.; Pagie, L.; de Vries, S. S.; Nahidiazar, L.; Dey, S. S.; Bienko, M.; Zhan, Y.; Lajoie, B.; de Graaf, C. A.; Amendola, M. and others (2015). *Genome-wide maps of nuclear lamina interactions in single human cells*, Cell 163 : 134-147.

[57] Gonzalez-Suarez, I.; Redwood, A. B.; Perkins, S. M.; Vermolen, B.; Lichtensztejin, D.; Grotzky, D. A.; Morgado-Palacin, L.; Gapud, E. J.; Sleckman, B. P.; Sullivan, T. and others (2009). *Novel roles for A-type lamins in telomere biology and the DNA damage response pathway*, The EMBO journal 28 : 2414-2427.

[58] Mehta, I. S.; Amira, M.; Harvey, A. J. and Bridger, J. M. (2010). *Rapid chromosome territory relocation by nuclear motor activity in response to serum removal in primary human fibroblasts*, Genome Biology .

[59] Goldman, R. D.; Shumaker, D. K.; Erdos, M. R.; Eriksson, M.; Goldman, A. E.; Gordon, L. B.; Gruenbaum, Y.; Khuon, S.; Mendez, M.; Varga, R. and others (2004). *Accumulation of mutant lamin A causes progressive changes in nuclear architecture in Hutchinson--Gilford progeria syndrome*, Proceedings of the National Academy of Sciences of the United States of America 101 : 8963-8968.

[60] Nikolova, V.; Leimena, C.; McMahan, A. C.; Tan, J. C.; Chandar, S.; Jogia, D.; Kesteven, S. H.; Michalick, J.; Otway, R.; Verheyen, F. and others (2004). *Defects in*



*nuclear structure and function promote dilated cardiomyopathy in lamin A/C--deficient mice*, Journal of Clinical Investigation 113 : 357.

[61] Taimen, P.; Pflieger, K.; Shimi, T.; Möller, D.; Ben-Harush, K.; Erdos, M. R.; Adam, S. A.; Herrmann, H.; Medalia, O.; Collins, F. S. and others (2009). *A progeria mutation reveals functions for lamin A in nuclear assembly, architecture, and chromosome organization*, Proceedings of the National Academy of Sciences 106 : 20788-20793.

[62] McCord, R.; Nazario-Toole, A.; Zhang, H.; Chines, P.; Zhan, Y. and Erdos, M. (2013). *Correlated alterations in genome organization, histone methylation, and DNA-lamin A/C interactions in Hutchinson-Gilford progeria syndrome*, Genome Res. 23 : 260-269.

[63] Carmo-Fonseca, M.; Mendes-Soares, L. and Campos, I. (2000). *To be or not to be in the nucleolus*, Nature cell biology 2 : E107-E112.

[64] Prieto, J. and McStay, B. (2005). *Nucleolar biogenesis: the first small steps.*, Biochemical Society Transactions 33 : 1441-1443.

[65] Németh, A.; Conesa, A.; Santoyo-Lopez, J.; Medina, I.; Montaner, D.; Péterfia, Bá.; Solovei, I.; Cremer, T.; Dopazo, J. and Langst, G. (2010). *Initial genomics of the human nucleolus*, PLoS Genet 6 : e1000889.

[66] Mehta, I. S.; Figgitt, M.; Clements, C. S.; Kill, I. R. and Bridger, J. M. (2007). *Alterations to nuclear architecture and genome behavior in senescent cells*, Annals of the New York Academy of Sciences 1100 : 250-263.

[67] Hemmerich, P.; Schmiedeberg, L. and Diekmann, S. (2011). *Dynamic as well as stable protein interactions contribute to genome function and maintenance*, Chromosome research 19 : 131-151.

[68] Zhu, L. and Brangwynne, C. P. (2015). *Nuclear bodies: the emerging biophysics of nucleoplasmic phases*, Current opinion in cell biology 34 : 23-30.

[69] Delcuve, G. P.; He, S. and Davie, J. R. (2008). *Mitotic partitioning of transcription factors*, Journal of cellular biochemistry 105 : 1-8.

[70] Kadauke, S. and Blobel, G. A. (2009). *Chromatin loops in gene regulation*, Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1789 : 17-25.

[71] Lieberman-Aiden, E.; van Berkum, N.; Williams, L.; Imakaev, M.; Ragoczy, T. and Telling, A. (2009). *Comprehensive mapping of long-range interactions reveals folding*

*principles of the human genome*, Science. 326 : 289-293.

- [72] Pope, B.; Ryba, T.; Dileep, V.; Yue, F.; Wu, W. and Denas, O. (2014). *Topologically associating domains are stable units of replication-timing regulation*, Nature. 515 : 402-405.
- [73] Dixon, J. R.; Selvaraj, S.; Yue, F.; Kim, A.; Li, Y.; Shen, Y.; Hu, M.; Liu, J. S. and Ren, B. (2012). *Topological domains in mammalian genomes identified by analysis of chromatin interactions*, Nature 485 : 376-380.
- [74] Nora, E.; Lajoie, B.; Schulz, E.; Giorgetti, L.; Okamoto, I. and Servant, N. (2012). *Spatial partitioning of the regulatory landscape of the X-inactivation centre*, Nature. 485 : 381-385.
- [75] Nora, E.; Dekker, J. and Heard, E. (2013). *Segmental folding of chromosomes: a basis for structural and regulatory chromosomal neighborhoods?*, Bioessays. 35 : 818-828.
- [76] Le Dily, F.; Bau, D.; Pohl, A.; Vicent, G.; Serra, F. and Soronellas, D. (2014). *Distinct structural transitions of chromatin topological domains correlate with coordinated hormone-induced gene regulation*, Genes Dev. 28 : 2151-2162.
- [77] Dixon, J.; Jung, I.; Selvaraj, S.; Shen, Y.; Antosiewicz-Bourget, J. and Lee, A. (2015). *Chromatin architecture reorganization during stem cell differentiation*, Nature. 518 : 331-336.
- [78] Rudan, M. V.; Barrington, C.; Henderson, S.; Ernst, C.; Odom, D. T.; Tanay, A. and Hadjur, S. (2015). *Comparative Hi-C reveals that CTCF underlies evolution of chromosomal domain architecture*, Cell reports 10 : 1297-1309.
- [79] Van Bortle, K.; Nichols, M.; Li, L.; Ong, C.; Takenaka, N. and Qin, Z. (2014). *Insulator function and topological domain border strength scale with architectural protein occupancy*, Genome Biol. 15 : R82.
- [80] Li, L.; Lyu, X.; Hou, C.; Takenaka, N.; Nguyen, H. Q.; Ong, C.-T.; Cubeñas-Potts, C.; Hu, M.; Lei, E. P.; Bosco, G. and others (2015). *Widespread Rearrangement of 3D Chromatin Organization Underlies Polycomb-Mediated Stress-Induced Silencing*, Molecular cell 58 : 216-231.
- [81] Guo, Y.; Xu, Q.; Canzio, D.; Shou, J.; Li, J.; Gorkin, D. U.; Jung, I.; Wu, H.; Zhai, Y.; Tang, Y. and others (2015). *CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function*, Cell 162 : 900-910.

- [82] Dekker, J. and Heard, E. (2015). *Structural and functional diversity of Topologically Associating Domains*, FEBS letters 589 : 2877-2884.
- [83] Cullen, K. E.; Kladde, M. P. and Seyfred, M. A. (1993). *Interaction between transcription regulatory regions of prolactin chromatin*, Science 261 : 203-206.
- [84] Dekker, J.; Rippe, K.; Dekker, M. and Kleckner, N. (2002). *Capturing chromosome conformation*, Science 295 : 1306-1311.
- [85] Naumova, N.; Smith, E. M.; Zhan, Y. and Dekker, J. (2012). *Analysis of long-range chromatin interactions using Chromosome Conformation Capture*, Methods 58 : 192-203.
- [86] Simonis, M.; Klous, P.; Splinter, E.; Moshkin, Y.; Willemsen, R.; de Wit, E.; van Steensel, B. and de Laat, W. (2006). *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture--on-chip (4C)*, Nature genetics 38 : 1348-1354.
- [87] Würtele, H. and Chartrand, P. (2006). *Genome-wide scanning of HoxB1-associated loci in mouse ES cells using an open-ended Chromosome Conformation Capture methodology*, Chromosome Research 14 : 477-495.
- [88] Zhao, Z.; Tavoosidana, G.; Sjölander, M.; Göndör, A.; Mariano, P.; Wang, S.; Kanduri, C.; Lezcano, M.; Sandhu, K. S.; Singh, U. and others (2006). *Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra-and interchromosomal interactions*, Nature genetics 38 : 1341-1347.
- [89] Sexton, T.; Kurukuti, S.; Mitchell, J. A.; Umlauf, D.; Nagano, T. and Fraser, P. (2012). *Sensitive detection of chromatin coassociations using enhanced chromosome conformation capture on chip*, Nature protocols 7 : 1335-1350.
- [90] Montavon, T.; Soshnikova, N.; Mascrez, Bé.; Joye, E.; Thevenet, L.; Splinter, E.; de Laat, W.; Spitz, F. and Duboule, D. (2011). *A regulatory archipelago controls Hox genes transcription in digits*, Cell 147 : 1132-1145.
- [91] Noordermeer, D.; Leleu, M.; Splinter, E.; Rougemont, J.; De Laat, W. and Duboule, D. (2011). *The dynamic architecture of Hox gene clusters*, Science 334 : 222-225.
- [92] Wei, Z.; Gao, F.; Kim, S.; Yang, H.; Lyu, J.; An, W.; Wang, K. and Lu, W. (2013). *Klf4 organizes long-range chromosomal interactions with the oct4 locus in reprogramming and pluripotency*, Cell stem cell 13 : 36-47.

- [93] Dostie, J.; Richmond, T. A.; Arnaout, R. A.; Selzer, R. R.; Lee, W. L.; Honan, T. A.; Rubio, E. D.; Krumm, A.; Lamb, J.; Nusbaum, C. and others (2006). *Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements*, Genome research 16 : 1299-1309.
- [94] Sanyal, A.; Lajoie, B. R.; Jain, G. and Dekker, J. (2012). *The long-range interaction landscape of gene promoters*, Nature 489 : 109-113.
- [95] Phillips-Cremins, J. E.; Sauria, M. E.; Sanyal, A.; Gerasimova, T. I.; Lajoie, B. R.; Bell, J. S.; Ong, C.-T.; Hookway, T. A.; Guo, C.; Sun, Y. and others (2013). *Architectural protein subclasses shape 3D organization of genomes during lineage commitment*, Cell 153 : 1281-1295.
- [96] Rousseau, M.; Ferraiuolo, M.; Crutchley, J.; Wang, X.; Miura, H. and Blanchette, M. (2014). *Classifying leukemia types with chromatin conformation data*, Genome Biol. 15 : R60.
- [97] Williamson, I.; Berlivet, S.; Eskeland, R.; Boyle, S.; Illingworth, R. S.; Paquette, D.; Dostie, J. and Bickmore, W. A. (2014). *Spatial genome organization: contrasting views from chromosome conformation capture and fluorescence in situ hybridization*, Genes & development 28 : 2778-2791.
- [98] Smith EM Lajoie BR, J. G. D. J. (2016). *Invariant TAD Boundaries Constrain Cell-Type-Specific Looping Interactions between Promoters and Distal Elements around the CFTR Locus*, Am J Hum Genet 98 : 185.
- [99] Belton, J.; McCord, R.; Gibcus, J.; Naumova, N.; Zhan, Y. and Dekker, J. (2012). *Hi-C: a comprehensive technique to capture the conformation of genomes*, Methods. 58 : 268-276.
- [100] Lajoie, B.; Dekker, J. and Kaplan, N. (2015). *The Hitchhiker's guide to Hi-C analysis: Practical guidelines*, Methods. 72 : 65-75.
- [101] Naumova, N.; Imakaev, M.; Fudenberg, G.; Zhan, Y.; Lajoie, B. R.; Mirny, L. A. and Dekker, J. (2013). *Organization of the mitotic chromosome*, Science 342 : 948-953.
- [102] Nagano, T.; Lubling, Y.; Stevens, T. J.; Schoenfelder, S.; Yaffe, E.; Dean, W.; Laue, E. D.; Tanay, A. and Fraser, P. (2013). *Single-cell Hi-C reveals cell-to-cell variability in chromosome structure*, Nature 502 : 59-64.
- [103] Dryden, N.; Broome, L.; Dudbridge, F.; Johnson, N.; Orr, N. and Schoenfelder, S. (2014). *Unbiased analysis of potential targets of breast cancer susceptibility loci by*

*Capture Hi-C*, Genome Res. 24 : 1854-1868.

[104] Schoenfelder, S.; Furlan-Magaril, M.; Mifsud, B.; Tavares-Cadete, F.; Sugar, R.; Javierre, B.-M.; Nagano, T.; Katsman, Y.; Sakthidevi, M.; Wingett, S. W. and others (2015). *The pluripotent regulatory circuitry connecting promoters to their long-range interacting elements*, Genome research 25 : 582-597.

[105] Schoenfelder, S.; Sugar, R.; Dimond, A.; Javierre, B.-M.; Armstrong, H.; Mifsud, B.; Dimitrova, E.; Matheson, L.; Tavares-Cadete, F.; Furlan-Magaril, M. and others (2015). *Polycomb repressive complex PRC1 spatially constrains the mouse embryonic stem cell genome*, Nature genetics .

[106] Simonis, M.; Kooren, J. and De Laat, W. (2007). *An evaluation of 3C-based methods to capture DNA interactions*, Nature methods 4 : 895-901.

[107] Tiwari, V. K.; Cope, L.; McGarvey, K. M.; Ohm, J. E. and Baylin, S. B. (2008). *A novel 6C assay uncovers Polycomb-mediated higher order chromatin conformations*, Genome research 18 : 1171-1179.

[108] Fullwood, M.; Liu, M.; Pan, Y.; Liu, J.; Xu, H. and Mohamed, Y. (2009). *An oestrogen-receptor-alpha-bound human chromatin interactome*, Nature. 462 : 58-64.

[109] Handoko, L.; Xu, H.; Li, G.; Ngan, C. Y.; Chew, E.; Schnapp, M.; Lee, C. W. H.; Ye, C.; Ping, J. L. H.; Mulawadi, F. and others (2011). *CTCF-mediated functional chromatin interactome in pluripotent cells*, Nature genetics 43 : 630-638.

[110] Li, G.; Ruan, X.; Auerbach, R.; Sandhu, K.; Zheng, M. and Wang, P. (2012). *Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation*, Cell. 148 : 84-98.

[111] Speicher, M. R. and Carter, N. P. (2005). *The new cytogenetics: blurring the boundaries with molecular biology*, Nature Reviews Genetics 6 : 782-792.

[112] Volpi, E. V. and Bridger, J. M. (2008). *FISH glossary: an overview of the fluorescence in situ hybridization technique*, Biotechniques 45 : 385-386.

[113] Chen, B.; Gilbert, L. A.; Cimini, B. A.; Schnitzbauer, J.; Zhang, W.; Li, G.-W.; Park, J.; Blackburn, E. H.; Weissman, J. S.; Qi, L. S. and others (2013). *Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system*, Cell 155 : 1479-1491.

[114] Markaki, Y.; Smeets, D.; Fiedler, S.; Schmid, V. J.; Schermelleh, L.; Cremer, T. and Cremer, M. (2012). *The potential of 3D-FISH and super-resolution structured*

*illumination microscopy for studies of 3D nuclear architecture*, Bioessays 34 : 412-426.

[115] Lakadamyali, M. and Cosma, M. P. (2015). *Advanced microscopy methods for visualizing chromatin structure*, FEBS letters .

[116] Long, B. R.; Robinson, D. C. and Zhong, H. (2014). *Subdiffractive microscopy: techniques, applications, and challenges*, Wiley Interdisciplinary Reviews: Systems Biology and Medicine 6 : 151-168.

[117] Hall, L. L. and Lawrence, J. B. (2011). *XIST RNA and Architecture of the Inactive X Chromosome Implications for the Repeat Genome*, : sqb-2010.

[118] Solovei, I.; Kreysing, M.; Lanctôt, C.; Kösem, S.; Peichl, L.; Cremer, T.; Guck, J. and Joffe, B. (2009). *Nuclear architecture of rod photoreceptor cells adapts to vision in mammalian evolution*, Cell 137 : 356-368.

[119] Tolhuis, B.; Palstra, R.-J.; Splinter, E.; Grosveld, F. and de Laat, W. (2002). *Looping and interaction between hypersensitive sites in the active  $\beta$ -globin locus*, Molecular cell 10 : 1453-1465.

[120] Palstra, R.-J.; Tolhuis, B.; Splinter, E.; Nijmeijer, R.; Grosveld, F. and de Laat, W. (2003). *The  $\beta$ -globin nuclear compartment in development and erythroid differentiation*, Nature genetics 35 : 190-194.

[121] Drissen, R.; Palstra, R.-J.; Gillemans, N.; Splinter, E.; Grosveld, F.; Philipsen, S. and de Laat, W. (2004). *The active spatial organization of the  $\beta$ -globin locus requires the transcription factor EKLF*, Genes & development 18 : 2485-2490.

[122] Vakoc, C. R.; Letting, D. L.; Gheldof, N.; Sawado, T.; Bender, M.; Groudine, M.; Weiss, M. J.; Dekker, J. and Blobel, G. A. (2005). *Proximity among distant regulatory elements at the  $\beta$ -globin locus requires GATA-1 and FOG-1*, Molecular cell 17 : 453-462.

[123] Splinter, E.; Heath, H.; Kooren, J.; Palstra, R.-J.; Klous, P.; Grosveld, F.; Galjart, N. and de Laat, W. (2006). *CTCF mediates long-range chromatin looping and local histone modification in the  $\beta$ -globin locus*, Genes & development 20 : 2349-2354.

[124] Vernimmen, D.; De Gobbi, M.; Sloane-Stanley, J. A.; Wood, W. G. and Higgs, D. R. (2007). *Long-range chromosomal interactions regulate the timing of the transition between poised and active gene expression*, The EMBO journal 26 : 2041-2051.

[125] Kmita, M. and Duboule, D. (2003). *Organizing axes in time and space; 25 years of colinear tinkering*, Science 301 : 331-333.

- [126] Zink, D.; Fischer, A. and Nickerson, J. (2004). *Nuclear structure in cancer cells*, Nat Rev Cancer. 4 : 677-687.
- [127] Zaidi, S. K.; Young, D. W.; Montecino, M. A.; Lian, J. B.; Van Wijnen, A. J.; Stein, J. L. and Stein, G. S. (2010). *Mitotic bookmarking of genes: a novel dimension to epigenetic control*, Nature Reviews Genetics 11 : 583-589.
- [128] Beale, L. (1860). *Examination of sputum from a case of cancer of the pharynx and the adjacent parts*, Arch Med 2 : 1860-61.
- [129] Derenzini, M.; Trerè, D.; Pession, A.; Govoni, M.; Sirri, V. and Chieco, P. (2000). *Nucleolar size indicates the rapidity of cell proliferation in cancer tissues*, The Journal of pathology 191 : 181-186.
- [130] Maggi Jr, L. B. and Weber, J. D. (2005). *Nucleolar adaptation in human cancer*, Cancer investigation 23 : 599-608.
- [131] Roix, J.; McQueen, P.; Munson, P.; Parada, L. and Misteli, T. (2003). *Spatial proximity of translocation-prone gene loci in human lymphomas*, Nat Genet. 34 : 287-291.
- [132] Soutoglou, E.; Dorn, J. F.; Sengupta, K.; Jasin, M.; Nussenzweig, A.; Ried, T.; Danuser, G. and Misteli, T. (2007). *Positional stability of single double-strand breaks in mammalian cells*, Nature cell biology 9 : 675-682.
- [133] Wiech, T.; Timme, S.; Riede, F.; Stein, S.; Schuricke, M.; Cremer, C.; Werner, M.; Hausmann, M. and Walch, A. (2005). *Human archival tissues provide a valuable source for the analysis of spatial genome organization*, Histochemistry and cell biology 123 : 229-238.
- [134] Cremer, M.; Küpper, K.; Wagler, B.; Wizelman, L.; Hase, J. v.; Weiland, Y.; Kreja, L.; Diebold, J.; Speicher, M. R. and Cremer, T. (2003). *Inheritance of gene density--related higher order chromatin arrangements in normal and tumor cell nuclei*, The Journal of cell biology 162 : 809-820.
- [135] Wiech, T.; Stein, S.; Lachenmaier, V.; Schmitt, E.; Schwarz-Finsterle, J.; Wiech, E.; Hildenbrand, G.; Werner, M. and Hausmann, M. (2009). *Spatial allelic imbalance of BCL2 genes and chromosome 18 territories in nonneoplastic and neoplastic cervical squamous epithelium*, European Biophysics Journal 38 : 793-806.
- [136] Meaburn, K. and Misteli, T. (2008). *Locus-specific and activity-independent gene repositioning during early tumorigenesis*, J Cell Biol. 180 : 39-50.

- [137] Ferrai, C.; de Castro, I. J.; Lavitas, L.; Chotalia, M. and Pombo, A. (2010). *Gene positioning*, Cold Spring Harbor perspectives in biology 2 : a000588.
- [138] Berdasco, M. and Esteller, M. (2010). *Aberrant epigenetic landscape in cancer: how cellular identity goes awry*, Developmental cell 19 : 698-711.
- [139] Ducy, P.; Zhang, R.; Geoffroy, V.; Ridall, A. L. and Karsenty, G. (1997). *Osf2/Cbfa1: a transcriptional activator of osteoblast differentiation*, Cell 89 : 747-754.
- [140] Komori, T.; Yagi, H.; Nomura, S.; Yamaguchi, A.; Sasaki, K.; Deguchi, K.; Shimizu, Y.; Bronson, R.; Gao, Y.-H.; Inada, M. and others (1997). *Targeted disruption of Cbfa1 results in a complete lack of bone formation owing to maturational arrest of osteoblasts*, Cell 89 : 755-764.
- [141] Otto, F.; Thornell, A. P.; Crompton, T.; Denzel, A.; Gilmour, K. C.; Rosewell, I. R.; Stamp, G. W.; Beddington, R. S.; Mundlos, S.; Olsen, B. R. and others (1997). *Cbfa1, a candidate gene for cleidocranial dysplasia syndrome, is essential for osteoblast differentiation and bone development*, Cell 89 : 765-771.
- [142] Ducy, P.; Starbuck, M.; Priemel, M.; Shen, J.; Pinero, G.; Geoffroy, V.; Amling, M. and Karsenty, G. (1999). *A Cbfa1-dependent genetic pathway controls bone formation beyond embryonic development*, Genes & Development 13 : 1025-1036.
- [143] Stein, G. S.; Lian, J. B.; Van Wijnen, A. J.; Stein, J. L.; Montecino, M.; Javed, A.; Zaidi, S. K.; Young, D. W.; Choi, J.-Y. and Pockwinse, S. M. (2004). *Runx2 control of organization, assembly and activity of the regulatory machinery for skeletal gene expression*, Oncogene 23 : 4315-4329.
- [144] Choi, J.-Y.; Pratap, J.; Javed, A.; Zaidi, S. K.; Xing, L.; Balint, E.; Dalamangas, S.; Boyce, B.; Van Wijnen, A. J.; Lian, J. B. and others (2001). *Subnuclear targeting of Runx/Cbfa/AML factors is essential for tissue-specific differentiation during embryonic development*, Proceedings of the National Academy of Sciences 98 : 8650-8655.
- [145] Lou, Y.; Javed, A.; Hussain, S.; Colby, J.; Frederick, D.; Pratap, J.; Xie, R.; Gaur, T.; Van Wijnen, A. J.; Jones, S. N. and others (2009). *A Runx2 threshold for the cleidocranial dysplasia phenotype*, Human molecular genetics 18 : 556-568.
- [146] Pockwinse, S. M.; Kota, K. P.; Quaresma, A. J.; Imbalzano, A. N.; Lian, J. B.; Van Wijnen, A. J.; Stein, J. L.; Stein, G. S. and Nickerson, J. A. (2011). *Live cell imaging of the cancer-related transcription factor RUNX2 during mitotic progression*, Journal of cellular physiology 226 : 1383-1389.



- [147] Ali, S. A.; Dobson, J. R.; Lian, J. B.; Stein, J. L.; Van Wijnen, A. J.; Zaidi, S. K. and Stein, G. S. (2012). *A RUNX2--HDAC1 co-repressor complex regulates rRNA gene expression by modulating UBF acetylation*, Journal of cell science 125 : 2732-2739.
- [148] Harada, H.; Tagashira, S.; Fujiwara, M.; Ogawa, S.; Katsumata, T.; Yamaguchi, A.; Komori, T. and Nakatsuka, M. (1999). *Cbfa1 isoforms exert functional differences in osteoblast differentiation*, Journal of Biological Chemistry 274 : 6972-6978.
- [149] Liu, J. C.; Lengner, C. J.; Gaur, T.; Lou, Y.; Hussain, S.; Jones, M. D.; Borodic, B.; Colby, J. L.; Steinman, H. A.; Van Wijnen, A. J. and others (2011). *Runx2 protein expression utilizes the Runx2 P1 promoter to establish osteoprogenitor cell number for normal bone formation*, Journal of Biological Chemistry 286 : 30057-30070.
- [150] Martinez, E.; Kundu, T. K.; Fu, J. and Roeder, R. G. (1998). *A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID*, Journal of Biological Chemistry 273 : 23781-23785.
- [151] Yu, J.; Madison, J. M.; Mundlos, S.; Winston, F. and Olsen, B. R. (1998). *Characterization of a Human Homologue of the Saccharomyces cerevisiae Transcription Factor Spt3 (SUPT3H)*, Genomics 53 : 90-96.
- [152] Barbaric, S.; Reinke, H. and Hörz, W. (2003). *Multiple mechanistically distinct functions of SAGA at the PHO5 promoter*, Molecular and cellular biology 23 : 3468-3476.
- [153] Liu, X.; Vorontchikhina, M.; Wang, Y.-L.; Faiola, F. and Martinez, E. (2008). *STAGA recruits Mediator to the MYC oncoprotein to stimulate transcription and cell proliferation*, Molecular and cellular biology 28 : 108-121.
- [154] Robertson, A. J.; Larroux, C.; Degnan, B. M. and Coffman, J. A. (2009). *The evolution of Runx genes II. The C-terminal Groucho recruitment motif is present in both eumetazoans and homoscleromorphs but absent in a haplosclerid demosponge*, BMC research notes 2 : 59.
- [155] Kikuta, H.; Laplante, M.; Navratilova, P.; Komisarczuk, A. Z.; Engström, P. G.; Fredman, D.; Akalin, A.; Caccamo, M.; Sealy, I.; Howe, K. and others (2007). *Genomic regulatory blocks encompass multiple neighboring genes and maintain conserved synteny in vertebrates*, Genome research 17 : 545-555.
- [156] Navratilova, P. and Becker, T. S. (2009). *Genomic regulatory blocks in vertebrates and implications in human disease*, Briefings in functional genomics & proteomics 8 : 333-342.

- [157] Zhou, X.; Lowdon, R. F.; Li, D.; Lawson, H. A.; Madden, P. A.; Costello, J. F. and Wang, T. (2013). *Exploring long-range genome interactions using the WashU Epigenome Browser*, Nature methods 10 : 375-376.
- [158] Kent, W. J.; Sugnet, C. W.; Furey, T. S.; Roskin, K. M.; Pringle, T. H.; Zahler, A. M. and Haussler, D. (2002). *The human genome browser at UCSC*, Genome research 12 : 996-1006.
- [159] Wang, D.; Christensen, K.; Chawla, K.; Xiao, G.; Krebsbach, P. H. and Franceschi, R. T. (1999). *Isolation and characterization of MC3T3-E1 preosteoblast subclones with distinct in vitro and in vivo differentiation/mineralization potential*, Journal of Bone and Mineral Research 14 : 893-903.
- [160] Kagey, M. H.; Newman, J. J.; Bilodeau, S.; Zhan, Y.; Orlando, D. A.; van Berkum, N. L.; Ebmeier, C. C.; Goossens, J.; Rahl, P. B.; Levine, S. S. and others (2010). *Mediator and cohesin connect gene expression and chromatin architecture*, Nature 467 : 430-435.
- [161] Song, L. and Crawford, G. E. (2010). *DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells*, Cold Spring Harbor Protocols 2010 : pdb-prot5384.
- [162] Boyle, A. P.; Guinney, J.; Crawford, G. E. and Furey, T. S. (2008). *F-Seq: a feature density estimator for high-throughput sequence tags*, Bioinformatics 24 : 2537-2538.
- [163] Drissi, H.; Luc, Q.; Shakoori, R.; Chuva De Sousa Lopes, S.; Choi, J.-Y.; Terry, A.; Hu, M.; Jones, S.; Neil, J. C.; Lian, J. B. and others (2000). *Transcriptional autoregulation of the bone related CBF1/RUNX2 gene*, Journal of cellular physiology 184 : 341-350.
- [164] Hagège, Hé.; Klous, P.; Braem, C.; Splinter, E.; Dekker, J.; Cathala, G.; de Laat, W. and Forné, T. (2007). *Quantitative analysis of chromosome conformation capture assays (3C-qPCR)*, Nature protocols 2 : 1722-1733.
- [165] Park, M.-H.; Shin, H.-I.; Choi, J.-Y.; Nam, S.-H.; Kim, Y.-J.; Kim, H.-J. and Ryoo, H.-M. (2001). *Differential expression patterns of Runx2 isoforms in cranial suture morphogenesis*, Journal of Bone and Mineral Research 16 : 885-892.
- [166] Lian, J. B.; Javed, A.; Zaidi, S. K.; Lengner, C.; Montecino, M.; Van Wijnen, A. J.; Stein, J. L. and Stein, G. (2004). *Regulatory controls for osteoblast growth and differentiation: role of Runx/Cbfa/AML factors*, Critical Reviews in Eukaryotic Gene

## Expression 14.

- [167] Wu, H.; Whitfield, T. W.; Gordon, J.; Dobson, J. R.; Tai, P.; van Wijnen, A. J.; Stein, J. L.; Stein, G. S. and Lian, J. B. (2014). *Genomic occupancy of Runx2 with global expression profiling identifies a novel dimension to control of osteoblastogenesis*, Genome Biol 15 : R52.
- [168] Quarles, L. D.; Yohay, D. A.; Lever, L. W.; Caton, R. and Wenstrup, R. J. (1992). *Distinct proliferative and differentiated stages of murine MC3T3-E1 cells in culture: An in vitro model of osteoblast development*, Journal of Bone and Mineral Research 7 : 683-692.
- [169] Tai, P. W.; Wu, H.; Gordon, J. A.; Whitfield, T. W.; Barutcu, A. R.; van Wijnen, A. J.; Lian, J. B.; Stein, G. S. and Stein, J. L. (2014). *Epigenetic landscape during osteoblastogenesis defines a differentiation-dependent Runx2 promoter region*, Gene 550 : 1-9.
- [170] Jin, F.; Li, Y.; Dixon, J. R.; Selvaraj, S.; Ye, Z.; Lee, A. Y.; Yen, C.-A.; Schmitt, A. D.; Espinoza, C. A. and Ren, B. (2013). *A high-resolution map of the three-dimensional chromatin interactome in human cells*, Nature 503 : 290-294.
- [171] Le, T. B.; Imakaev, M. V.; Mirny, L. A. and Laub, M. T. (2013). *High-resolution mapping of the spatial organization of a bacterial chromosome*, Science 342 : 731-734.
- [172] Mundlos, S.; Otto, F.; Mundlos, C.; Mulliken, J.; Aylsworth, A.; Albright, S.; Lindhout, D.; Cole, W.; Henn, W.; Knoll, J. and others (1997). *Mutations involving the transcription factor CBF1 cause cleidocranial dysplasia*, Cell 89 : 773-779.
- [173] Rennert, J.; Coffman, J. A.; Mushegian, A. R. and Robertson, A. J. (2003). *The evolution of Runx genes I. A comparative study of sequences from phylogenetically diverse model organisms*, BMC evolutionary biology 3 : 4.
- [174] Markova, E. N.; Kantidze, O. L. and Razin, S. V. (2011). *Transcriptional regulation and spatial organisation of the human AML1/RUNX1 gene*, Journal of cellular biochemistry 112 : 1997-2005.
- [175] Li, X.; Decker, M. and Westendorf, J. J. (2010). *TETHerred to Runx: novel binding partners for runx factors*, Blood Cells, Molecules, and Diseases 45 : 82-85.
- [176] Meyer, M. B.; Benkusky, N. A. and Pike, J. W. (2014). *The RUNX2 Cistrome in Osteoblasts Characterization, Down-Regulation Following Differentiation, and Relationship to Gene Expression*, Journal of Biological Chemistry 289 : 16016-16031.

- [177] Xu, M. and Cook, P. R. (2008). *Similar active genes cluster in specialized transcription factories*, *The Journal of cell biology* 181 : 615-623.
- [178] Zullo, J. M.; Demarco, I. A.; Piqué-Regi, R.; Gaffney, D. J.; Epstein, C. B.; Spooner, C. J.; Luperchio, T. R.; Bernstein, B. E.; Pritchard, J. K.; Reddy, K. L. and others (2012). *DNA sequence-dependent compartmentalization and silencing of chromatin at the nuclear lamina*, *Cell* 149 : 1474-1487.
- [179] Bickmore, W. A. (2013). *The spatial organization of the human genome*, *Annual review of genomics and human genetics* 14 : 67-84.
- [180] de Laat, W. and Duboule, D. (2013). *Topology of mammalian developmental enhancers and their regulatory landscapes*, *Nature*. 502 : 499-506.
- [181] Bonora, G.; Plath, K. and Denholtz, M. (2014). *A mechanistic link between gene regulation and genome architecture in mammalian development*, *Current opinion in genetics & development* 27 : 92-101.
- [182] Network, C. G. A. (2012). *Comprehensive molecular portraits of human breast tumours*, *Nature* 490 : 61-70.
- [183] Stephens, P.; Tarpey, P.; Davies, H.; Van Loo, P.; Greenman, C. and Wedge, D. (2012). *The landscape of cancer genes and mutational processes in breast cancer*, *Nature*. 486 : 400-404.
- [184] Kandoth, C.; McLellan, M.; Vandin, F.; Ye, K.; Niu, B. and Lu, C. (2013). *Mutational landscape and significance across 12 major cancer types*, *Nature*. 502 : 333-339.
- [185] Suva, M.; Riggi, N. and Bernstein, B. (2013). *Epigenetic reprogramming in cancer*, *Science*. 339 : 1567-1570.
- [186] Dey, P. (2010). *Cancer nucleus: morphology and beyond*, *Diagn Cytopathol*. 38 : 382-390.
- [187] Gibcus, J. and Dekker, J. (2013). *The hierarchy of the 3D genome*, *Mol Cell*. 49 : 773-782.
- [188] Nye, A.; Rajendran, R.; Stenoien, D.; Mancini, M.; Katzenellenbogen, B. and Belmont, A. (2002). *Alteration of large-scale chromatin structure by estrogen receptor*, *Mol Cell Biol*. 22 : 3437-3449.
- [189] Meaburn, K.; Gudla, P.; Khan, S.; Lockett, S. and Misteli, T. (2009). *Disease-specific gene repositioning in breast cancer*, *J Cell Biol*. 187 : 801-812.

- [190] Rocha, P.; Micsinai, M.; Kim, J.; Hewitt, S.; Souza, P. and Trimarchi, T. (2012). *Close proximity to Igh is a contributing factor to AID-mediated translocations*, Mol Cell. 47 : 873-885.
- [191] Zhang, Y.; McCord, R.; Ho, Y.; Lajoie, B.; Hildebrand, D. and Simon, A. (2012). *Spatial organization of the mouse genome and its role in recurrent chromosomal translocations*, Cell. 148 : 908-921.
- [192] Seitan, V.; Faure, A.; Zhan, Y.; McCord, R.; Lajoie, B. and Ing-Simmons, E. (2013). *Cohesin-based chromatin interactions enable regulated gene expression within preexisting architectural compartments*, Genome Res. 23 : 2066-2077.
- [193] Giorgetti, L.; Galupa, R.; Nora, E.; Piolot, T.; Lam, F. and Dekker, J. (2014). *Predictive polymer modeling reveals coupled fluctuations in chromosome conformation and transcription*, Cell. 157 : 950-963.
- [194] Hah, N.; Murakami, S.; Nagari, A.; Danko, C. and Kraus, W. (2013). *Enhancer transcripts mark active estrogen receptor binding sites*, Genome Res. 23 : 1210-1223.
- [195] Hsu, P.; Hsu, H.; Lan, X.; Juan, L.; Yan, P. and Labanowska, J. (2013). *Amplification of distant estrogen response elements deregulates target genes associated with tamoxifen resistance in breast cancer*, Cancer Cell. 24 : 197-212.
- [196] Osmanbeyoglu, H.; Lu, K.; Oesterreich, S.; Day, R.; Benos, P. and Coronello, C. (2013). *Estrogen represses gene expression through reconfiguring chromatin structures*, Nucleic Acids Res. 41 : 8061-8071.
- [197] Wang, J.; Lan, X.; Hsu, P.; Hsu, H.; Huang, K. and Parvin, J. (2013). *Genome-wide analysis uncovers high frequency, strong differential chromosomal interactions and their associated epigenetic patterns in E2-mediated gene regulation*, BMC Genomics. 14 : 70.
- [198] Mourad, R.; Hsu, P.; Juan, L.; Shen, C.; Koneru, P. and Lin, H. (2014). *Estrogen induces global reorganization of chromatin structure in human breast cancer cells*, PLoS One. 9 : e113354.
- [199] Barnett, D.; Sheng, S.; Charn, T.; Waheed, A.; Sly, W. and Lin, C. (2008). *Estrogen receptor regulation of carbonic anhydrase XII through a distal enhancer in breast cancer*, Cancer Res. 68 : 3505-3515.
- [200] Bretschneider, N.; Kangaspeska, S.; Seifert, M.; Reid, G.; Gannon, F. and Denger, S. (2008). *E2-mediated cathepsin D (CTSD) activation involves looping of distal*

*enhancer elements*, Mol Oncol. 2 : 182-190.

[201] Saramaki, A.; Diermeier, S.; Kellner, R.; Laitinen, H.; Vaisanen, S. and Carlberg, C. (2009). *Cyclical chromatin looping and transcription factor association on the regulatory regions of the p21 (CDKN1A) gene in response to 1alpha,25-dihydroxyvitamin D3*, J Biol Chem. 284 : 8073-8082.

[202] Matilainen, J.; Malinen, M.; Turunen, M.; Carlberg, C. and Vaisanen, S. (2010). *The number of vitamin D receptor binding sites defines the different vitamin D responsiveness of the CYP24 gene in malignant and normal mammary cells*, J Biol Chem. 285 : 24174-24183.

[203] Wright, J.; Brown, S. and Cole, M. (2010). *Upregulation of c-MYC in cis through a large chromatin loop linked to a cancer risk-associated single-nucleotide polymorphism in colorectal cancer cells*, Mol Cell Biol. 30 : 1411-1420.

[204] Hughes, J.; Roberts, N.; McGowan, S.; Hay, D.; Giannoulatou, E. and Lynch, M. (2014). *Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment*, Nat Genet. 46 : 205-212.

[205] Misteli, T. (2010). *Higher-order genome organization in human disease*, Cold Spring Harb Perspect Biol. 2 : a000794.

[206] Debnath, J.; Muthuswamy, S. and Brugge, J. (2003). *Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures*, Methods. 30 : 256-268.

[207] Song, L.; Florea, L. and Langmead, B. (2014). *Lighter: fast and memory-efficient sequencing error correction without counting*, Genome Biol. 15 : 509.

[208] Li, B. and Dewey, C. (2011). *RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome*, BMC Bioinformatics. 12 : 323.

[209] Love, M.; Huber, W. and Anders, S. (2014). *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2*, Genome Biol. 15 : 550.

[210] Berriz, G. F.; Beaver, J. E.; Cenik, C.; Tasan, M. and Roth, F. P. (2009). *Next generation software for functional trend analysis*, Bioinformatics 25 : 3043-3044.

[211] Shen, L.; Shao, N.; Liu, X. and Nestler, E. (2014). *ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases*, BMC Genomics. 15 : 284.

[212] Imakaev, M.; Fudenberg, G.; McCord, R.; Naumova, N.; Goloborodko, A. and

- Lajoie, B. (2012). *Iterative correction of Hi-C data reveals hallmarks of chromosome organization*, Nat Methods. 9 : 999-1003.
- [213] Crane, E.; Bian, Q.; Mccord, R. P.; Lajoie, B. R.; Wheeler, B. S.; Ralston, E. J.; Uzawa, S.; Dekker, J. and Meyer, B. J. (2015). *Condensin-driven remodelling of X chromosome topology during dosage compensation*, Nature : 240–244.
- [214] Quinlan, A. (2014). *BEDTools: The Swiss-army tool for genome feature analysis*, Curr Protoc Bioinformatics 47 : 11 12 11-34.
- [215] Davidson, J.; Gorringer, K.; Chin, S.; Orsetti, B.; Besret, C. and Courtay-Cahen, C. (2000). *Molecular cytogenetic analysis of breast cancer cell lines*, Br J Cancer. 83 : 1309-1317.
- [216] Kao, J.; Salari, K.; Bocanegra, M.; Choi, Y.; Girard, L. and Gandhi, J. (2009). *Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery*, PLoS One. 4 : e6146.
- [217] Rao, S.; Huntley, M.; Durand, N.; Stamenova, E.; Bochkov, I. and Robinson, J. (2014). *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping*, Cell. 159 : 1665-1680.
- [218] Li, Q.; Seo, J.; Stranger, B.; McKenna, A.; Pe'er, I. and Laframboise, T. (2013). *Integrative eQTL-based analyses reveal the biology of breast cancer risk loci*, Cell. 152 : 633-641.
- [219] Barbieri, M.; Chotalia, M.; Fraser, J.; Lavitas, L.; Dostie, J. and Pombo, A. (2013). *A model of the large-scale organization of chromatin*, Biochem Soc Trans. 41 : 508-512.
- [220] Chuang, C. and Belmont, A. (2007). *Moving chromatin within the interphase nucleus-controlled transitions?*, Semin Cell Dev Biol. 18 : 698-706.
- [221] Therizols, P.; Illingworth, R.; Courilleau, C.; Boyle, S.; Wood, A. and Bickmore, W. (2014). *Chromatin decondensation is sufficient to alter nuclear organization in embryonic stem cells*, Science. 346 : 1238-1242.
- [222] Draskovic, I.; Arnoult, N.; Steiner, V.; Bacchetti, S.; Lomonte, P. and Londono-Vallejo, A. (2009). *Probing PML body function in ALT cells reveals spatiotemporal requirements for telomere recombination*, Proc Natl Acad Sci U S A. 106 : 15726-15731.
- [223] Heaphy, C.; Subhawong, A.; Hong, S.; Goggins, M.; Montgomery, E. and Gabrielson, E. (2011). *Prevalence of the alternative lengthening of telomeres telomere*

*maintenance mechanism in human cancer subtypes*, Am J Pathol. 179 : 1608-1615.

[224] Cutter, A. R. and Hayes, J. J. (2015). *A brief review of nucleosome structure*, FEBS letters .

[225] Clapier, C. R. and Cairns, B. R. (2009). *The biology of chromatin remodeling complexes*, Annual review of biochemistry 78 : 273-304.

[226] Flaus, A. and Owen-Hughes, T. (2001). *Mechanisms for ATP-dependent chromatin remodelling*, Current opinion in genetics & development 11 : 148-154.

[227] Varga-Weisz, P. (2001). *ATP-dependent chromatin remodeling factors: nucleosome shufflers with many missions.*, Oncogene 20 : 3076-3085.

[228] Sala, A.; Toto, M.; Pinello, L.; Gabriele, A.; Di Benedetto, V.; Ingrassia, A. M.; Bosco, G. L.; Di Gesù, V.; Giancarlo, R. and Corona, D. F. (2011). *Genome-wide characterization of chromatin binding and nucleosome spacing activity of the nucleosome remodelling ATPase ISWI*, The EMBO journal 30 : 1766-1777.

[229] Narlikar, G. J.; Sundaramoorthy, R. and Owen-Hughes, T. (2013). *Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes*, Cell 154 : 490-503.

[230] Morris, S. A.; Baek, S.; Sung, M.-H.; John, S.; Wiench, M.; Johnson, T. A.; Schiltz, R. L. and Hager, G. L. (2014). *Overlapping chromatin-remodeling systems collaborate genome wide at dynamic chromatin transitions*, Nature structural & molecular biology 21 : 73-81.

[231] Khavari, P. A.; Peterson, C. L.; Tamkun, J. W.; Mendel, D. B. and Crabtree, G. R. (1993). *BRG1 contains a conserved domain of the SWI2/SNF2 family necessary for normal mitotic growth and transcription*, Nature 366 : 170-174.

[232] Muchardt, C. and Yaniv, M. (1993). *A human homologue of Saccharomyces cerevisiae SNF2/SWI2 and Drosophila brm genes potentiates transcriptional activation by the glucocorticoid receptor.*, The EMBO Journal 12 : 4279.

[233] Wang, W.; Cote, J.; Xue, Y.; Zhou, S.; Khavari, P.; Biggar, S.; Muchardt, C.; Kalpana, G.; Goff, S.; Yaniv, M. and others (1996). *Purification and biochemical heterogeneity of the mammalian SWI-SNF complex.*, The EMBO journal 15 : 5370.

[234] Reyes, J.; Barra, J.; Muchardt, C.; Camus, A.; Babinet, C. and Yaniv, M. (1998). *Altered control of cellular proliferation in the absence of mammalian brahma (SNF2 $\beta$ )*, The EMBO journal 17 : 6979-6991.

[235] Bultman, S.; Gebuhr, T.; Yee, D.; Mantia, C. L.; Nicholson, J.; Gilliam, A.;



- Randazzo, F.; Metzger, D.; Chambon, P.; Crabtree, G. and et al. (2000). *A Brg1 Null Mutation in the Mouse Reveals Functional Differences among Mammalian SWI/SNF Complexes*, Molecular Cell : 1287–1295.
- [236] Bultman, S.; Herschkowitz, J.; Godfrey, V.; Gebuhr, T.; Yaniv, M.; Perou, C. and Magnuson, T. (2008). *Characterization of mammary tumors from Brg1 heterozygous mice*, Oncogene 27 : 460-468.
- [237] Trotter, K. W. and Archer, T. K. (2008). *The BRG1 transcriptional coregulator*, Nuclear receptor signaling 6.
- [238] King, H. A.; Trotter, K. W. and Archer, T. K. (2012). *Chromatin remodeling during glucocorticoid receptor regulated transactivation*, Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms 1819 : 716-726.
- [239] Kadoch, C.; Hargreaves, D. C.; Hodges, C.; Elias, L.; Ho, L.; Ranish, J. and Crabtree, G. R. (2013). *Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy*, Nature genetics 45 : 592-601.
- [240] Shain, A. H. and Pollack, J. R. (2013). *The spectrum of SWI/SNF mutations, ubiquitous in human cancers*, PLoS one 8 : e55119.
- [241] Euskirchen, G. M.; Auerbach, R. K.; Davidov, E.; Gianoulis, T. A.; Zhong, G.; Rozowsky, J.; Bhardwaj, N.; Gerstein, M. B. and Snyder, M. (2011). *Diverse roles and interactions of the SWI/SNF chromatin remodeling complex revealed using global approaches*, PLoS Genet 7 : e1002008-e1002008.
- [242] Harada, A.; Mallappa, C.; Okada, S.; Butler, J. T.; Baker, S. P.; Lawrence, J. B.; Ohkawa, Y. and Imbalzano, A. N. (2015). *Spatial re-organization of myogenic regulatory sequences temporally controls gene expression*, Nucleic acids research : gkv046.
- [243] Kim, S.-I.; Bresnick, E. H. and Bultman, S. J. (2009). *BRG1 directly regulates nucleosome structure and chromatin looping of the  $\beta$  globin locus to activate transcription*, Nucleic acids research 37 : 6019-6027.
- [244] Kim, S.-I.; Bultman, S. J.; Kiefer, C. M.; Dean, A. and Bresnick, E. H. (2009). *BRG1 requirement for long-range interaction of a locus control region with a downstream promoter*, Proceedings of the National Academy of Sciences 106 : 2259-2264.
- [245] Bossen, C.; Murre, C. S.; Chang, A. N.; Mansson, R.; Rodewald, H.-R. and

Murre, C. (2015). *The chromatin remodeler Brg1 activates enhancer repertoires to establish B cell identity and modulate cell growth*, Nature Immunology .

[246] Ni, Z.; El Hassan, M. A.; Xu, Z.; Yu, T. and Bremner, R. (2008). *The chromatin-remodeling enzyme BRG1 coordinates CIITA induction through many interdependent distal enhancers*, Nature Immunology 9 : 785-793.

[247] Hu, G.; Schones, D. E.; Cui, K.; Ybarra, R.; Northrup, D.; Tang, Q.; Gattinoni, L.; Restifo, N. P.; Huang, S. and Zhao, K. (2011). *Regulation of nucleosome landscape and transcription factor targeting at tissue-specific enhancers by BRG1*, Genome Research 21 : 1650-1658.

[248] Rada-Iglesias, A.; Bajpai, R.; Swigut, T.; Brugmann, S. A.; Flynn, R. A. and Wysocka, J. (2011). *A unique chromatin signature uncovers early developmental enhancers in humans*, Nature 470 : 279-283.

[249] Ho, L.; Jothi, R.; Ronan, J. L.; Cui, K.; Zhao, K. and Crabtree, G. R. (2009). *An embryonic stem cell chromatin remodeling complex, esBAF, is an essential component of the core pluripotency transcriptional network*, Proceedings of the National Academy of Sciences 106 : 5187-5191.

[250] Lan, X.; Witt, H.; Katsumura, K.; Ye, Z.; Wang, Q.; Bresnick, E. H.; Farnham, P. J. and Jin, V. X. (2012). *Integration of Hi-C and ChIP-seq data reveals distinct types of chromatin linkages*, Nucleic acids research : gks501.

[251] Hill, D. A.; Chiosea, S.; Jamaluddin, S.; Roy, K.; Fischer, A. H.; Boyd, D. D.; Nickerson, J. A. and Imbalzano, A. N. (2004). *Inducible changes in cell size and attachment area due to expression of a mutant SWI/SNF chromatin remodeling enzyme*, Journal of cell science 117 : 5847-5854.

[252] Imbalzano, K. M.; Cohet, N.; Wu, Q.; Underwood, J. M.; Imbalzano, A. N. and Nickerson, J. A. (2013). *Nuclear shape changes are induced by knockdown of the SWI/SNF ATPase BRG1 and are independent of cytoskeletal connections*, PLoS One .

[253] Kawaguchi, T.; Tanigawa, A.; Naganuma, T.; Ohkawa, Y.; Souquere, S.; Pierron, G. and Hirose, T. (2015). *SWI/SNF chromatin-remodeling complexes function in noncoding RNA-dependent assembly of nuclear bodies*, Proceedings of the National Academy of Sciences 112 : 4304-4309.

[254] Imbalzano, A.; Imbalzano, K. and Nickerson, J. A. (2013). *BRG1, a SWI/SNF chromatin remodeling enzyme ATPase, is required for maintenance of nuclear shape and integrity*, Communicative & integrative biology 6 : e55628.

- [255] Cohet, N.; Stewart, K. M.; Mudhasani, R.; Asirvatham, A. J.; Mallappa, C.; Imbalzano, K. M.; Weaver, V. M.; Imbalzano, A. N. and Nickerson, J. A. (2010). *SWI/SNF chromatin remodeling enzyme ATPases promote cell proliferation in normal mammary epithelial cells*, *Journal of cellular physiology* 223 : 667-678.
- [256] Milacic, M.; Haw, R.; Rothfels, K.; Wu, G.; Croft, D.; Hermjakob, H.; D'Eustachio, P. and Stein, L. (2012). *Annotating cancer variants and anti-cancer therapeutics in reactome*, *Cancers* 4 : 1180-1211.
- [257] Croft, D.; Mundo, A. F.; Haw, R.; Milacic, M.; Weiser, J.; Wu, G.; Caudy, M.; Garapati, P.; Gillespie, M.; Kamdar, M. R. and others (2014). *The Reactome pathway knowledgebase*, *Nucleic acids research* 42 : D472-D477.
- [258] Trapnell, C.; Hendrickson, D. G.; Sauvageau, M.; Goff, L.; Rinn, J. L. and Pachter, L. (2013). *Differential analysis of gene regulation at transcript resolution with RNA-seq*, *Nature biotechnology* 31 : 46-53.
- [259] Xu, R.; Spencer, V. A. and Bissell, M. J. (2007). *Extracellular matrix-regulated gene expression requires cooperation of SWI/SNF and transcription factors*, *Journal of biological chemistry* 282 : 14992-14999.
- [260] Stankunas, K.; Hang, C. T.; Tsun, Z.-Y.; Chen, H.; Lee, N. V.; Wu, J. I.; Shang, C.; Bayle, J. H.; Shou, W.; Iruela-Arispe, M. L. and others (2008). *Endocardial Brg1 represses ADAMTS1 to maintain the microenvironment for myocardial morphogenesis*, *Developmental cell* 14 : 298-311.
- [261] Saladi, S. V.; Keenen, B.; Marathe, H. G.; Qi, H.; Chin, K.-V. and de la Serna, I. L. (2010). *Modulation of extracellular matrix/adhesion molecule expression by BRG1 is associated with increased melanoma invasiveness*, *Molecular cancer* 9 : 280.
- [262] Whyte, W. A.; Orlando, D. A.; Hnisz, D.; Abraham, B. J.; Lin, C. Y.; Kagey, M. H.; Rahl, P. B.; Lee, T. I. and Young, R. A. (2013). *Master transcription factors and mediator establish super-enhancers at key cell identity genes*, *Cell* 153 : 307-319.
- [263] Barutcu, A. R.; Lajoie, B. R.; McCord, R. P.; Tye, C. E.; Hong, D.; Messier, T. L.; Browne, G.; van Wijnen, A. J.; Lian, J. B.; Stein, J. L. and others (2015). *Chromatin interaction analysis reveals changes in small chromosome and telomere clustering between epithelial and breast cancer cells*, *Genome biology* 16 : 1-14.
- [264] Dahl, K. N.; Ribeiro, A. J. and Lammerding, J. (2008). *Nuclear shape, mechanics, and mechanotransduction*, *Circulation research* 102 : 1307-1318.

- [265] Wang, N.; Tytell, J. D. and Ingber, D. E. (2009). *Mechanotransduction at a distance: mechanically coupling the extracellular matrix with the nucleus*, Nature reviews Molecular cell biology 10 : 75-82.
- [266] Guidi, M.; Ruault, M.; Marbouty, M.; Loiodice, I.; Cournac, A.; Billaudeau, C.; Hocher, A.; Mozziconacci, J.; Koszul, R. and Taddei, A. (2015). *Spatial reorganization of telomeres in long-lived quiescent cells*, Genome biology 16 : 1-15.
- [267] Cubeñas-Potts, C. and Corces, V. G. (2015). *Architectural proteins, transcription, and the three-dimensional organization of the genome*, FEBS letters .
- [268] Vaquerizas, J. M.; Kummerfeld, S. K.; Teichmann, S. A. and Luscombe, N. M. (2009). *A census of human transcription factors: function, expression and evolution*, Nature Reviews Genetics 10 : 252-263.
- [269] Langmead, B. and Salzberg, S. L. (2012). *Fast gapped-read alignment with Bowtie 2*, Nature methods 9 : 357-359.
- [270] Lee, T. I.; Johnstone, S. E. and Young, R. A. (2006). *Chromatin immunoprecipitation and microarray-based analysis of protein location*, Nature protocols 1 : 729-748.
- [271] Heinz, S.; Benner, C.; Spann, N.; Bertolino, E.; Lin, Y. C.; Laslo, P.; Cheng, J. X.; Murre, C.; Singh, H. and Glass, C. K. (2010). *Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities*, Molecular cell 38 : 576-589.
- [272] Speck, N. A. and Stacy, T. (1995). *A new transcription factor family associated with human leukemias*, Critical Reviews™ in Eukaryotic Gene Expression 5.
- [273] Imperato, M. R.; Cauchy, P.; Obier, N. and Bonifer, C. (2015). *The RUNX1--PU.1 axis in the control of hematopoiesis*, International journal of hematology 101 : 319-329.
- [274] Lian, J. B.; Stein, G. S.; Javed, A.; Van Wijnen, A. J.; Stein, J. L.; Montecino, M.; Hassan, M. Q.; Gaur, T.; Lengner, C. J. and Young, D. W. (2006). *Networks and hubs for the transcriptional control of osteoblastogenesis*, Reviews in Endocrine and Metabolic Disorders 7 : 1-16.
- [275] Inoue, K.-i.; Shiga, T. and Ito, Y. (2008). *Runx transcription factors in neuronal development*, Neural Dev 3 : 8104-3.
- [276] Chuang, L. S. H.; Ito, K. and Ito, Y. (2013). *RUNX family: regulation and*

*diversification of roles through interacting proteins*, International Journal of Cancer 132 : 1260-1271.

[277] Durst, K. L. and Hiebert, S. W. (2004). *Role of RUNX family members in transcriptional repression and gene silencing*, Oncogene 23 : 4220-4224.

[278] Ito, Y. (2004). *Oncogenic potential of the RUNX gene family: 'overview'*, Oncogene 23 : 4198-4208.

[279] Blyth, K.; Cameron, E. R. and Neil, J. C. (2005). *The RUNX genes: gain or loss of function in cancer*, Nature Reviews Cancer 5 : 376-387.

[280] Ito, Y.; Bae, S.-C. and Chuang, L. S. H. (2015). *The RUNX family: developmental regulators in cancer*, Nature Reviews Cancer 15 : 81-95.

[281] Tanaka, T.; Kurokawa, M.; Ueki, K.; Tanaka, K.; Imai, Y.; Mitani, K.; Okazaki, K.; Sagata, N.; Yazaki, Y.; Shibata, Y. and others (1996). *The extracellular signal-regulated kinase pathway phosphorylates AML1, an acute myeloid leukemia gene product, and potentially regulates its transactivation ability.*, Molecular and cellular biology 16 : 3967-3979.

[282] Niimi, H.; Harada, H.; Harada, Y.; Ding, Y.; Imagawa, J.; Inaba, T.; Kyo, T. and Kimura, A. (2006). *Hyperactivation of the RAS signaling pathway in myelodysplastic syndrome with AML1/RUNX1 point mutations*, Leukemia 20 : 635-644.

[283] Osorio, K. M.; Lilja, K. C. and Tumbar, T. (2011). *Runx1 modulates adult hair follicle stem cell emergence and maintenance from distinct embryonic skin compartments*, The Journal of cell biology 193 : 235-250.

[284] Logan, T. T.; Villapol, S. and Symes, A. J. (2013). *TGF-beta superfamily gene expression and induction of the Runx1 transcription factor in adult neurogenic regions after brain injury*, PloS one 8 : e59250.

[285] Osato, M. (2004). *Point mutations in the RUNX1/AML1 gene: another actor in RUNX leukemia*, Oncogene 23 : 4284-4296.

[286] Koh, C.; Wang, C. Q.; Ng, C. E. L.; Ito, Y.; Araki, M.; Tergaonkar, V.; Huang, G. and Osato, M. (2013). *RUNX1 meets MLL: epigenetic regulation of hematopoiesis by two leukemia genes*, Leukemia 27 : 1793-1802.

[287] Kadota, M.; Yang, H. H.; Gomez, B.; Sato, M.; Clifford, R. J.; Meerzaman, D.; Dunn, B. K.; Wakefield, L. M. and Lee, M. P. (2010). *Delineating genetic alterations for tumor progression in the MCF10A series of breast cancer cell lines*, PloS one 5 :

e9201.

[288] Jemal, A.; Bray, F.; Center, M. M.; Ferlay, J.; Ward, E. and Forman, D. (2011). *Global cancer statistics*, CA: a cancer journal for clinicians 61 : 69-90.

[289] Taniuchi, I.; Osato, M. and Ito, Y. (2012). *Runx1: no longer just for leukemia*, The EMBO journal 31 : 4098-4099.

[290] Ferrari, N.; Mohammed, Z. M.; Nixon, C.; Mason, S. M.; Mallon, E.; McMillan, D. C.; Morris, J. S.; Cameron, E. R.; Edwards, J. and Blyth, K. (2014). *Expression of RUNX1 correlates with poor patient prognosis in triple negative breast cancer*, .

[291] van Bragt, M. P.; Hu, X.; Xie, Y. and Li, Z. (2015). *RUNX1, a transcription factor mutated in breast cancer, controls the fate of ER-positive mammary luminal cells*, Elife 3 : e03881.

[292] Browne, G.; Taipaleenmäki, H.; Bishop, N. M.; Madasu, S. C.; Shaw, L. M.; van Wijnen, A. J.; Stein, J. L.; Stein, G. S. and Lian, J. B. (2015). *Runx1 is associated with breast cancer progression in MMTV-PyMT transgenic mice and its depletion in vitro inhibits migration and invasion*, Journal of cellular physiology .

[293] Ramaswamy, S.; Ross, K. N.; Lander, E. S. and Golub, T. R. (2003). *A molecular signature of metastasis in primary solid tumors*, Nature genetics 33 : 49-54.

[294] Banerji, S.; Cibulskis, K.; Rangel-Escareno, C.; Brown, K. K.; Carter, S. L.; Frederick, A. M.; Lawrence, M. S.; Sivachenko, A. Y.; Sougnez, C.; Zou, L. and others (2012). *Sequence analysis of mutations and translocations across breast cancer subtypes*, Nature 486 : 405-409.

[295] Curtis, C.; Shah, S. P.; Chin, S.-F.; Turashvili, G.; Rueda, O. M.; Dunning, M. J.; Speed, D.; Lynch, A. G.; Samarajiwa, S.; Yuan, Y. and others (2012). *The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups*, Nature 486 : 346-352.

[296] Scheitz, C. J. F.; Lee, T. S.; McDermitt, D. J. and Tumber, T. (2012). *Defining a tissue stem cell-driven Runx1/Stat3 signalling axis in epithelial cancer*, The EMBO journal 31 : 4124-4139.

[297] Harrington, K. S.; Javed, A.; Drissi, H.; McNeil, S.; Lian, J. B.; Stein, J. L.; Van Wijnen, A. J.; Wang, Y.-L. and Stein, G. S. (2002). *Transcription factors RUNX1/AML1 and RUNX2/Cbfa1 dynamically associate with stationary subnuclear domains*, Journal of cell science 115 : 4167-4176.

- [298] Stender, J. D.; Kim, K.; Charn, T. H.; Komm, B.; Chang, K. C.; Kraus, W. L.; Benner, C.; Glass, C. K. and Katzenellenbogen, B. S. (2010). *Genome-wide analysis of estrogen receptor  $\beta$  DNA binding and tethering mechanisms identifies Runx1 as a novel tethering factor in receptor-mediated transcriptional activation*, Molecular and cellular biology 30 : 3943-3955.
- [299] Ross, K.; Sedello, A. K.; Todd, G. P.; Paszkowski-Rogacz, M.; Bird, A. W.; Ding, L.; Grinenko, T.; Behrens, K.; Hubner, N.; Mann, M. and others (2012). *Polycomb group ring finger 1 cooperates with Runx1 in regulating differentiation and self-renewal of hematopoietic cells*, Blood 119 : 4152-4161.
- [300] Yu, M.; Mazor, T.; Huang, H.; Huang, H.-T.; Kathrein, K. L.; Woo, A. J.; Chouinard, C. R.; Labadorf, A.; Akie, T. E.; Moran, T. B. and others (2012). *Direct recruitment of polycomb repressive complex 1 to chromatin by core binding transcription factors*, Molecular cell 45 : 330-343.
- [301] Jiang, H. and Peterlin, B. M. (2008). *Differential chromatin looping regulates CD4 expression in immature thymocytes*, Molecular and cellular biology 28 : 907-912.
- [302] Levantini, E.; Lee, S.; Radomska, H. S.; Hetherington, C. J.; Alberich-Jorda, M.; Amabile, G.; Zhang, P.; Gonzalez, D. A.; Zhang, J.; Basseres, D. S. and others (2011). *RUNX1 regulates the CD34 gene in haematopoietic stem cells by mediating interactions with a distal regulatory element*, The EMBO journal 30 : 4059-4070.
- [303] Lu, P.; Weaver, V. M. and Werb, Z. (2012). *The extracellular matrix: a dynamic niche in cancer progression*, The Journal of cell biology 196 : 395-406.
- [304] Gutschner, T.; Hämmerle, M.; Eißmann, M.; Hsu, J.; Kim, Y.; Hung, G.; Revenko, A.; Arun, G.; Stentrup, M.; Groß, M. and others (2013). *The noncoding RNA MALAT1 is a critical regulator of the metastasis phenotype of lung cancer cells*, Cancer research 73 : 1180-1189.
- [305] Chakravarty, D.; Sboner, A.; Nair, S. S.; Giannopoulou, E.; Li, R.; Hennig, S.; Mosquera, J. M.; Pauwels, J.; Park, K.; Kossai, M. and others (2014). *The oestrogen receptor alpha-regulated lncRNA NEAT1 is a critical modulator of prostate cancer*, Nature communications 5.
- [306] Machanick, P. and Bailey, T. L. (2011). *MEME-ChIP: motif analysis of large DNA datasets*, Bioinformatics 27 : 1696-1697.
- [307] Ono, M.; Yaguchi, H.; Ohkura, N.; Kitabayashi, I.; Nagamura, Y.; Nomura, T.; Miyachi, Y.; Tsukada, T. and Sakaguchi, S. (2007). *Foxp3 controls regulatory T-cell*

*function by interacting with AML1/Runx1*, Nature 446 : 685-689.

- [308] Pencovich, N.; Jaschek, R.; Tanay, A. and Groner, Y. (2011). *Dynamic combinatorial interactions of RUNX1 and cooperating partners regulates megakaryocytic differentiation in cell line models*, Blood 117 : e1-e14.
- [309] Seo, W.; Ikawa, T.; Kawamoto, H. and Taniuchi, I. (2012). *Runx1--Cbf $\beta$  facilitates early B lymphocyte development by regulating expression of Ebf1*, The Journal of experimental medicine 209 : 1255-1262.
- [310] Welch, R. P.; Lee, C.; Imbriano, P. M.; Patil, S.; Weymouth, T. E.; Smith, R. A.; Scott, L. J. and Sartor, M. A. (2014). *ChIP-Enrich: gene set enrichment testing for ChIP-seq data*, Nucleic acids research : gku463.
- [311] Trombly, D. J.; Whitfield, T. W.; Padmanabhan, S.; Gordon, J. A.; Lian, J. B.; van Wijnen, A. J.; Zaidi, S. K.; Stein, J. L. and Stein, G. S. (2015). *Genome-wide co-occupancy of AML1-ETO and N-CoR defines the t (8; 21) AML signature in leukemic cells*, BMC genomics 16 : 309.
- [312] ENCODE Consortium (2012). *An integrated encyclopedia of DNA elements in the human genome*, Nature 489 : 57-74.
- [313] Eckert, D.; Buhl, S.; Weber, S.; Jager, R. and Schorle, H. (2005). *The AP-2 family of transcription factors*, Genome Biol 6 : 246.
- [314] Chapellier, M.; Bachelard-Cascales, E.; Schmidt, X.; Clément, F.; Treilleux, I.; Delay, E.; Jammot, A.; Ménétrier-Caux, C.; Pochon, G.; Besançon, R. and others (2015). *Disequilibrium of BMP2 Levels in the Breast Stem Cell Niche Launches Epithelial Transformation by Overamplifying BMPR1B Cell Response*, Stem cell reports 4 : 239-254.
- [315] Wotton, S.; Stewart, M.; Blyth, K.; Vaillant, F.; Kilbey, A.; Neil, J. C. and Cameron, E. R. (2002). *Proviral insertion indicates a dominant oncogenic role for Runx1/AML-1 in T-cell lymphoma*, Cancer research 62 : 7181-7185.
- [316] Chimge, N. and Frenkel, B. (2013). *The RUNX family in breast cancer: relationships with estrogen signaling*, Oncogene 32 : 2121-2130.
- [317] Chuang, C.-H.; Carpenter, A. E.; Fuchsova, B.; Johnson, T.; de Lanerolle, P. and Belmont, A. S. (2006). *Long-range directional movement of an interphase chromosome site*, Current Biology 16 : 825-831.
- [318] Smith, E. and Shilatifard, A. (2014). *Enhancer biology and enhanceropathies*,



Nature structural & molecular biology 21 : 210-219.

[319] Thurman, R. E.; Rynes, E.; Humbert, R.; Vierstra, J.; Maurano, M. T.; Haugen, E.; Sheffield, N. C.; Stergachis, A. B.; Wang, H.; Vernot, B. and others (2012). *The accessible chromatin landscape of the human genome*, Nature 489 : 75-82.

[320] Bell, A. C.; West, A. G. and Felsenfeld, G. (1999). *The protein CTCF is required for the enhancer blocking activity of vertebrate insulators*, Cell 98 : 387-396.

[321] Rinn, J. L. and Chang, H. Y. (2012). *Genome regulation by long noncoding RNAs*, Annual review of biochemistry 81.

[322] Batista, P. J. and Chang, H. Y. (2013). *Long noncoding RNAs: cellular address codes in development and disease*, Cell 152 : 1298-1307.

[323] Lam, M. T.; Li, W.; Rosenfeld, M. G. and Glass, C. K. (2014). *Enhancer RNAs and regulated transcriptional programs*, Trends in Biochemical Sciences 39 : 170–182.

[324] Spitz, F. and Furlong, E. E. M. (2012). *Transcription factors: from enhancer binding to developmental control*, Nature Reviews Genetics 13 : 613–626.

[325] Bolukbasi, M. F.; Gupta, A.; Oikemus, S.; Derr, A. G.; Garber, M.; Brodsky, M. H.; Zhu, L. J. and Wolfe, S. A. (2015). *DNA-binding-domain fusions enhance the targeting range and precision of Cas9*, Nat Meth 12 : 1150–1156.

[326] Bolukbasi, M. F.; Gupta, A. and Wolfe, S. A. (2015). *Creating and evaluating accurate CRISPR-Cas9~scalpels for genomic surgery*, Nat Meth 13 : 41–50.

[327] Ghavi-Helm, Y.; Klein, F. A.; Pakozdi, T.; Ciglar, L.; Noordermeer, D.; Huber, W. and Furlong, E. E. (2014). *Enhancer loops appear stable during development and are associated with paused polymerase*, Nature 512 : 96-100.

[328] Muller, S.; Filippakopoulos, P. and Knapp, S. (2011). *Bromodomains as therapeutic targets*, Expert Reviews in Molecular Medicine 13.

[329] Rotili, D. and Mai, A. (2011). *Targeting Histone Demethylases: A New Avenue for the Fight against Cancer*, Genes & Cancer 2 : 663–679.

[330] Geeven, G.; Zhu, Y.; Kim, B. J.; Bartholdy, B. A.; Yang, S.-M.; Macfarlan, T. S.; Gifford, W. D.; Pfaff, S. L.; Versteegen, M. J. A. M.; Pinto, H. and et al. (2015). *Local compartment changes and regulatory landscape alterations in histone H1-depleted cells*, Genome Biology 16.

[331] Hsieh, T.-H.; Weiner, A.; Lajoie, B.; Dekker, J.; Friedman, N. and Rando, O.

(2015). *Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C*, Cell 162 : 108–119.

[332] Tang, Z.; Luo, O.; Li, X.; Zheng, M.; Zhu, J.; Szalaj, P.; Trzaskoma, P.; Magalska, A.; Wlodarczyk, J.; Ruszczycki, B. and et al. (2015). *CTCF-Mediated Human 3D Genome Architecture Reveals Chromatin Topology for Transcription*, Cell 163 : 1611–1627.

[333] de Wit, E.; Vos, E.; Holwerda, S.; Valdes-Quezada, C.; Verstegen, M.; Teunissen, H.; Splinter, E.; Wijchers, P.; Krijger, P. and de Laat, W. (2015). *CTCF Binding Polarity Determines Chromatin Looping*, Molecular Cell 60 : 676–684.

[334] Giorgetti, L.; Galupa, R.; Nora, E.; Piolot, T.; Lam, F.; Dekker, J.; Tiana, G. and Heard, E. (2014). *Predictive Polymer Modeling Reveals Coupled Fluctuations in Chromosome Conformation and Transcription*, Cell 157 : 950–963.

[335] Fraser, J.; Ferrai, C.; Chiariello, A. M.; Schueler, M.; Rito, T.; Laudanno, G.; Barbieri, M.; Moore, B. L.; Kraemer, D. C.; Aitken, S. and et al. (2015). *Hierarchical folding and reorganization of chromosomes are linked to transcriptional changes in cellular differentiation*, Molecular Systems Biology 11 : 852–852.

[336] Lin, Y.; Protter, D.; Rosen, M. and Parker, R. (2015). *Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins*, Molecular Cell 60 : 208–219.

[337] Zhang, H.; Elbaum-Garfinkle, S.; Langdon, E. M.; Taylor, N.; Occhipinti, P.; Bridges, A.; Brangwynne, C. and Gladfelter, A. (2015). *RNA Controls PolyQ Protein Phase Transitions*, Molecular Cell 60 : 220–230.

[338] Sutherland, H. and Bickmore, W. A. (2009). *Transcription factories: gene expression in unions?*, Nature Reviews Genetics 10 : 457–466.

[339] Lamond, A. I. and Spector, D. L. (2003). *Nuclear speckles: a model for nuclear organelles*, Nat Rev Mol Cell Biol 4 : 605–612.

## APPENDIX

## 1. Table of 3C primers in Chapter 2.

BglIII Restriction fragment number for chr17:44,518,354-45,267,358 (mm9)	BglIII Restriction fragment start (mm9)	BglIII Restriction fragment end (mm9)	BglIII Restriction fragment size	Primer sequence	Primer start position (mm9)	Primer end position (mm9)	Primer Size
4	44527462	44533934	6472	CCATTCCTCCAGCGTTAGCATCCCAAGC	44533863	44533890	28
30	44602154	44606505	4351	AGGGCTTGCTCCGGTTGCTTCTGCAGG	44606400	44606427	28
41	44636355	44643107	6752	GCTGTCTGACCTACCTATGCTTTCTCT	44643011	44643038	28
47	44672542	44674466	1924	ACACACACACCTCACTAACTCATTC	44674362	44674387	26
50	44680475	44685278	4803	CAATGAAGTGAATCTACCTTGCTAAC	44685217	44685244	28
52	44695157	44697297	2140	AGCTGGGAGAAGCGAATTTCCCTCCA	44697221	44697248	28
62	44722984	44727138	4154	GCCATTTTAAACGATTACCCCTGGGAC	44727047	44727075	29
66	44734344	44739983	5639	GGCTGGGAAGGACGGGTCTTTGGCAT	44739902	44739929	28
78	44766656	44770765	4109	AAACCATACAGGAAAACGTGTGACCACTG	44770697	44770726	30
82	44781635	44792520	10885	AAAACAGATACACAGAAGCTGAGAAAAC	44792419	44792445	27
85	44796302	44799957	3655	TCTGAGTGTCTGGCGTGTCTCTCTGA	44799851	44799878	28
89	44807895	44819268	11373	TATAAGAACATGGTTGCTTTCAGAGAT	44819177	44819204	28
101	44834989	44849318	14329	TTCTCTCTATGGATATTGGTGGCTCC	44849219	44849248	30
102	44849319	44856523	7204	AAGATAGTACATGGGTTTAAAGTCATTTG	44856484	44856511	28
103	44856524	44862933	6409	AACTCTAGTCTGGCTGAAATACT	44862903	44862927	25
104 - Runx2-P2 promoter	44862934	44873350	10416	GGCCCCCTCGCTTTCAAGGTGCGGG	44873282	44873309	28
107	44881727	44886668	4941	TTTTCTTCTGAGAAATAGTCAGGTATT	44886536	44886563	28
109	44891526	44896782	5256	TTCTGCTTGAAGTATGATTTGTATTTT	44896682	44896709	28
116	44903712	44910160	6448	GCTTGCTGTGAGCGCTCTTTGGACCTC	44910058	44910085	28
117 - Supt3h promoter	44910161	44914813	4652	AATCATTTAGCCGGTGGTGTGGGGGGTA	44914736	44914764	29
119	44915581	44919033	3452	GATGGGACTTAACGTTACTGCCAGTGATA	44918944	44918973	30
120	44919034	44921519	2485	GGATAATGGAAGCAACGAGAAGTGAAGTAA	44921417	44921446	30
126	44937252	44939514	2262	ACCTCTTTAGACGTTTAAAGTTGCTAGCC	44939428	44939457	30
128	44948362	44949365	1003	CTAGAAGATGGCCTTGATTTATCTCT	44949327	44949354	28
130 - Runx2-P1 promoter	44950463	44952567	2104	TTTTATTACGTGGCGCTCTACAATAAAG	44952479	44952508	30
133	44969008	44976001	6993	GTCCTGTCAGTGGTCTAAGTCTTGTAC	44975893	44975920	28
135	44977679	44984844	7165	GTGTTGGTGTTCAGTTTTACATAGTTT	44984737	44984764	28
142	44993429	44998418	4989	ATACCTGCCTTTTTACTCTGTGTTG	44998384	44998409	26
147	45008716	45014815	6099	CTGATGGTCAGAGCGTCAAGGGGTATCTC	45014730	45014758	29
152	45032347	45042871	10524	GCACACTTGAGAGGTGATGTCAGGGTTTC	45042776	45042804	29
162	45055719	45063155	7436	AACTGTTAAAACCACTAAGGAAATCAAT	45063101	45063128	28
171	45078575	45082558	3983	GTTCTTTGGGTCTCGAGATAACCCCATG	45082451	45082479	29
174	45098520	45116826	18306	GTCCTTCATCCTAATCAAGTCCCTCTCAGC	45116742	45116771	30
185	45143242	45153503	10261	CCTTGAGTTATCCTAGTCAACCTAAACT	45153272	45153299	28
190	45160642	45172354	11712	CCACAGATGTTGAACAAGCGTTTGTGAGCC	45172286	45172315	30
197	45189172	45195502	6330	ACTGAAGCAGCGTCCGTTCTAGAAGC	45195412	45195440	29
203	45208917	45220903	11986	ACTGAGCCAGATCATTGGCCCAAGGACAT	45220815	45220843	29
206	45234393	45234743	350	GTACCATATTATAGGTTCCCTAGATACC	45234660	45234688	29
214	45249328	45257385	8057	ATTGCCCTGGCGTCCCTAATCATGGGAAA	45257284	45257313	30
216	45257409	45258782	1373	GTGGGTAAGAGCACCAATTGCTCTCCA	45258698	45258726	29

**BglII Gene Desert Region Primers:**

Chromosome	Fragment Start	Fragment End	BglII fragment number	Primer Sequence
chr3	147368198	147372443	4	TTTGCTACTTTGTTTTGCTTTAGAGATA
chr3	147372444	147374135	5	GACTGAAGTCGGGAAAACCTGTGGTTGA
chr3	147377816	147385393	7	TTCGTCTTGCACTTACTAAGGTATTTAA
chr3	147386300	147396851	9	AACAAACAAACAAAATGCAAGTACTAAG
chr3	147407486	147410643	11	GCCACCCATATTCTCAGACTCATTAA
chr3	147412891	147415729	13	GGCAGCCAGAGAAGTACGTGCTGCTCTTT
chr3	147453705	147456495	29	GGAGGTAAGGAACGTACTTTTGTCTCAG
chr3	147468681	147469955	32	TCCTTCTAGAAGAAGTCTCAGGGGTCAG
chr3	147365717	147367097	0	GCAATATTTACATGTGATAAAGGTCATT

**HindIII 3C Primers:**

Chromosome	Fragment Start	Fragment End	HindIII fragment number	Primer Sequence
chr17	44951598	44952311	HindIII_RX2_83	TGCAAGGCTGCTTCGGATTCCCAGGG
chr17	44897649	44901627	HindIII_RX2_57	GCCATAGTGTACGTACTTATAATCTC
chr17	44907410	44911873	HindIII_RX2_60	TTTGGTGGTGTGATGTCACCTGGC
chr17	44,913,822	44,914,828	HindIII_RX2_63	TAAATCATTTAGCCGGTGGTGTCCGG
chr17	44915240	44916274	HindIII_RX2_65	GTATCCAGAGTTCGCCCTGATATTTG

**HindIII ERCC3 locus Primers:**

Chromosome	Fragment Start	Fragment End	HindIII fragment number	Primer Sequence
chr18	32396743	32401728	GDR0	GTGACAGTCTAGTGGTGAAGAGTC
chr18	32403345	32407277	GDR2	ATTCATCTGGGAACGCCACTCCAGACA
chr18	32420831	32421376	GDR6	GCTCTGGCAAAGCGGTATTAGATG
chr18	32433479	32436646	GDR9	TGACTGTCCCAAGCGAAGACTGTTCA
chr18	32440591	32443799	GDR12	GTGCTAGGAGTTATGCTAATATCACC