

THE ROLE OF HISTONE MODIFICATIONS IN CHROMOSOME BREAKAGE AND TRANSLOCATION

A thesis

submitted by

BHARAT BURMAN

In partial fulfillment of the requirements
for the degree of

Doctor of Philosophy

in

Cell, Molecular and Developmental Biology

TUFTS UNIVERSITY

Sackler School of Graduate Biomedical Sciences

May, 2015

Advisors: Dr. Grace Gill (Tufts) and Dr. Tom Misteli (National Cancer Institute)

Thesis Chair: Dr. Gavin Schnitzler
Committee Member: Dr. Charlotte Kuperwasser
Committee Member: Dr. James Schwob

ABSTRACT

Chromosome translocations are genetic hallmarks of most cancer cells. Translocations require the formation of DNA double-strand breaks (DSBs) at two or more genomic loci, followed by the illegitimate joining of broken chromosomal ends through DNA repair. There is increasing evidence that translocations occur at non-random sites in the genome, suggesting that certain regions of the genome are more susceptible to DNA breakage than others. We hypothesize that altered chromatin properties predispose genomic sites to DNA breakage and translocations. Using large-scale computational analysis, we identified altered levels of specific histone modifications compared to baseline levels at common leukemia and lymphoma breakpoints in hematopoietic stem cells. To probe the physiological relevance of these modifications, we mapped histone modifications and chromatin structure at translocation-prone regions in anaplastic large cell lymphoma (ALCL) precursor cells. We find enrichment of histone marks associated with open chromatin and a decrease in marks associated with closed chromatin near frequent translocation breakpoints. In order to directly test the role of chromatin features in DNA breakage susceptibility, we developed a protein-DNA tethering system that allows us to create local chromatin domains at pre-defined sites in the genome containing inducible DSB sites *in vivo*. By measuring the amount of DSBs using ligation-mediated PCR, we find that histone modifying enzymes that create active chromatin marks generally increase breakage susceptibility. Finally, we developed a high-throughput break-apart FISH (hiBA-FISH) assay to detect low frequency chromosome breakage and translocation events in lymphocytic cells expressing chromatin modifying enzymes. Experimental elevation of H3K4 methylation promotes chromosome breaks and specific

translocations in response to genotoxic stress. Taken together, these experiments provide first insights into the role of histone modifications in the formation of nonrandom chromosomal breaks and the mechanisms that lead to translocations. Clarifying the role of epigenetic changes in translocations has significant clinical implications, particularly with regards to understanding cancer predisposition and susceptibility, and because reversal of aberrant epigenetic changes has emerged as a promising strategy for the treatment of cancer.

ACKNOWLEDGMENTS

This work could not be completed without the enduring support of my parents, Drs. Sneh and Sudhir Burman, my sister Deeya Burman, and my fiancé Alanna Windsor. My fiancé, especially, did not give up on me, traveling great distances to be with me despite my ongoing work in the lab, and for that I am forever grateful. Additionally, I would like to thank my aunts, uncles, cousins, and friends for their love and support.

I dedicate this work to my mom. As my workload increased in the lab, I was fortunate to see hers decrease. Her 35+ years of hard work as a physician, including 60+ hour weeks, overnight shifts, and of course, parenting of two children, is my greatest inspiration. Mom, I wish you the world in your retirement. Here is some reading material for your first trip!

Regarding the work itself, I thank all members, past and present, of the Misteli Lab at NCI for their guidance, technical assistance, and feedback. At the forefront of this talented group is Dr. Tom Misteli. I will never forget the day I met Tom and he told me “this is a project I would give a post-doc.” While it may have been my mistake to accept, I appreciated the faith Tom put in me from Day 1. This was the beginning of an adventure to say the least. For the next 4.5 years, Tom guided me through the lows and the highs of research. He taught me to move forward without looking back. In addition to science, he taught me how to write, speak, and manage, which may be the most important skills for a successful career in research. Tom, I thank you for training me, for guiding me, and for challenging me at every step of the PhD.

Tom was one of two people that “had my back” during the PhD. I am deeply indebted to Dr. Grace Gill, my co-advisor at Tufts. Grace did not have to do what she

did—advise a student 400 miles away—but I am very lucky that she volunteered for the task. Grace kept me afloat during the PhD through her bottomless support of my short and long term goals. She kept me in line with my PhD requirements, which may not sound like a big deal until you realize you are doing a post-doc level project. In other words, Grace reminded me that I was a PhD student, one of her students actually, with a deep support network in her, my thesis advisory committee, and the Tufts community. One thing I will cherish forever is that when experiments did not work, Grace would help me with my experiments, and when they did work, Grace would help me with my character. Grace, thank you for helping me get to where I am today.

Grace helped me put together an all-star thesis advisory committee at Tufts. This committee included Dr. Charlotte Kuperwasser, Dr. Gavin Schnitzler, Dr. Rick Van Etten (now at UC Irvine), and MD/PhD Director Dr. Jim Schwob, all of whom offered me support and helped me develop as a PhD student. I also thank the Clinical Director of the MD/PhD program, and my long time mentor, Dr. Gordon Huggins.

I also want to thank Katherine McKinnon (NCI) for help with FACS experiments, Dr. Tatiana Karpova (NCI Fluorescence Imaging Microscopy Facility) and Dr. Gianluca Pegoraro (NCI High-Throughput Imaging Facility) for help with microscopy, Dr. Zhuzhu Zhang (Salk Institute for Biological Sciences) and Dr. Jason Lieb (University of Chicago) for bioinformatics support, and members of the NCI Laboratory of Receptor Biology and Gene Expression for thoughtful advice.

Finally, my work was supported generously by the Intramural Research Program of NIH, NCI, Department of Defense, and by the HHMI-NIH medical student fellowship I was awarded for the first two years. Who said there was no money in science?

TABLE OF CONTENTS

	Abstract	i
	Acknowledgments	iii
	Table of contents	v
	List of tables	vi
	List of figures	vii
	List of abbreviations	ix
Chapter 1:	General introduction	1
	Introduction	
	Aims of thesis	
Chapter 2:	Computational analysis of chromatin features at recurrent translocation sites	13
	Introduction	14
	Results	16
	Materials and methods	30
Chapter 3:	Chromatin factor tethering and DSB formation by endonucleases	35
	Introduction	36
	Results	38
	Materials and methods	52
Chapter 4:	Effect of histone modifications on chromosome breakage and translocations in anaplastic large cell lymphoma	57
	Introduction	58
	Results	60
	Materials and methods	84
Chapter 5:	Quantitative detection of rare chromosome breaks and translocations by high-throughput imaging (Method development)	89
	Introduction	90
	Results	93
	Materials and methods	107
	Discussion of Chapter 5	111
Chapter 6:	General discussion and future directions	116
	Discussion of Chapters 1-4	117
	Future directions	125
Appendix 1:	Individual gene and breakpoint computational analyses	129
	References	153

LIST OF TABLES

Chapter 2:	page
1. Major breakpoint cluster regions of 15 translocation genes	27
Chapter 3:	
1. Histone modifying enzymes cloned into mCherry-Lac constructs	39
2. Size and texture of arrays tethered by histone modifying enzymes	45
Chapter 4:	
1. Histone modifications and variants studied in comparative ChIP experiments.	63
2. Lentiviral constructs of histone modifying enzymes used for stable cell lines.	73

LIST OF FIGURES

Chapter 2:

1. Available NIH Roadmap Epigenomics Project data for CD34+ cells.	18
2. Strategy for analysis of chromatin features.	19
3. Example of a translocation gene with one of its matched control genes.	19
4. Population analysis.	21
5. Examples of individual gene analysis.	22
6. Heatmap of individual analyses.	24
7. Heatmap of individual analyses ranked by gene expression.	25
8. Examples of breakpoint analysis.	29
9. Heatmap of breakpoint analyses.	30

Chapter 3:

1. Schematic representation of chromatin-protein tethering system.	39
2. Enrichment of histone modifications in LacR-chromatin domains: ASH2L.	41
3. Enrichment of histone modifications in LacR-chromatin domains: SET2.	41
4. Enrichment of histone modifications in LacR-chromatin domains: SUV3-9.	41
5. Enrichment of histone modifications in LacR-chromatin domains: TIP60.	42
6. Percentage of arrays with positive histone modification enrichments.	42
7. ChIP-qPCR of H3K4me1 in LacR-chromatin domains.	44
8. ChIP-qPCR of H3K4me3 in LacR-chromatin domains.	44
9. H3 density in LacR-chromatin domains.	44
10. Schematic of breakage by I-SceI endonuclease assay in chromatin domains.	46
11. Quantification of DSBs by LM-PCR in LacR vs. BRG1 chromatin domains.	48
12. Quantification of DSBs by LM-PCR in all LacR-chromatin domains.	48
13. Percentage of γ H2AX co-localizations in LacR vs. BRG1 chromatin domains.	50
14. Percentage of γ H2AX co-localizations in all LacR-chromatin domains.	50
15. Percentage of γ H2AX /53BP1 co-localizations in LacR vs. ASH2L domains.	51
16. LM-PCR in LacR vs. ASH2L domains.	51

Chapter 4:

1. Schematic of cell lines and comparative ChIP analysis strategy.	61
2. Schematic of <i>ALK</i> and <i>NPM1</i> breakpoint cluster regions and primers.	61
3. mRNA expression levels of <i>ALK</i> and <i>NPM1</i> translocation genes.	63
4. ChIP-qPCR of H3K4me1 at breakpoint and control regions.	65
5. ChIP-qPCR of H3K4me3 at breakpoint and control regions.	65
6. ChIP-qPCR of H3K36me3 at breakpoint and control regions.	66
7. ChIP-qPCR of H3K9me3 at breakpoint and control regions.	66
8. ChIP-qPCR of H3K27ac at breakpoint and control regions.	67
9. ChIP-qPCR of H3K79me2 at breakpoint and control regions.	67
10. ChIP-qPCR of H3K27me3 at breakpoint and control regions.	68

11. H3 density at breakpoint and control regions.	69
12. Schematic of experiment to test chromosome breakage and translocation.	70
13. Schematic representation of the hiBA-ALK probe set.	72
14. Example of merged maximum intensity projection image of hiBA-FISH.	72
15. <i>ALK</i> breakage screen of FE-PD stable cell lines.	75
16. <i>NPM1-ALK</i> translocation screen of FE-PD stable cell lines.	76
17. ChIP-qPCR of H3K4me1 in FE-PD stable cell lines.	77
18. ChIP-qPCR of H3K4me3 in FE-PD stable cell lines.	77
19. ChIP-qPCR of H3K9me3 in FE-PD stable cell lines.	78
20. Determination of <i>ALK</i> breakage frequencies in FE-PD stable cell lines.	79
21. Determination of <i>ALK</i> breakage frequencies in Jurkat stable cell lines.	80
22. Determination of <i>NPM1-ALK</i> translocation frequencies in FE-PD cell lines.	82
23. Schematic representation of the hiBA-Nanog probe set.	83
24. Determination of <i>Nanog</i> breakage frequencies in FE-PD stable cell lines.	83
25. Determination of <i>CycA-Nanog</i> translocation frequencies in FE-PD cell lines.	84

Chapter 5:

1. hiBA-FISH pipeline	93
2. Outline of hiBA-FISH event definitions	94
3. Schematic representation of hiBA-FISH probe sets.	95
4. Automated nucleus segmentation and FISH signal detection.	97
5. Qualitative identification of breakage and translocation events.	98
6. Automated hiBA-FISH signal detection metrics.	101
7. Experimental determination of a proximity threshold for FISH signals.	102
8. Automated determination of chromosomal breaks by hiBA-FISH.	104
9. Automated determination of chromosomal translocations by hiBA-FISH.	107

Appendix 1 contains extended individual gene (Chapter 2, Fig. 5) and breakpoint analyses (Chapter 2, Fig. 8)	129
---	-----

LIST OF ABBREVIATIONS

Note: this list does not include gene abbreviations.

3C	Chromosome conformation capture
3D	Three-dimensional
3'UTR	3' untranslated region
4C	Circularized chromosome conformation capture
5C	Carbon-copy chromosome conformation capture
5'UTR	5' untranslated region
A	Adenine nucleotide
Ac/Acetyl	Acetylated
AID	Activation-induced cytidine deaminase
ALCL	Anaplastic large cell lymphoma
ANOVA	Analysis of variance test
AR	Androgen receptor
BAC	Bacterial artificial chromosome
bcr	Breakpoint cluster region
BLESS	Direct in situ breaks labeling, enrichment on streptavidin and next-generation sequencing
bp	Basepairs
C	Cytosine nucleotide
Cas9	A CRISPR-associated (Cas) nuclease
CD34+	Cluster of differentiation 34 expressing
cDNA	Complementary DNA
CFP	Cyan fluorescent protein
CFS	Common fragile sites
ChIP	Chromatin immunoprecipitation
ChIP-Seq	Chromatin immunoprecipitation sequencing
CML	Chronic myelogenous leukemia
COSMIC	Catalogue of Somatic Mutations in Cancer
CpG islands	Regions of DNA with high frequency of cytosine-phosphate-guanine
CRISPR	Clustered regularly interspaced short palindromic repeat
CycA	Cyclophilin A (PPIA) gene
DAPI	4',6-diamidino-2-phenylindole (fluorescent nuclear DNA stain)
dbCRID	Database of Chromosomal Rearrangements in Disease
DDR	DNA damage response
DNA	Deoxyribonucleic acid
DNaseI-Seq	DNase I hypersensitive sites sequencing
dNTP	Deoxyribonucleotide
DSB	DNA double-strand break
ENCODE	Encyclopedia of DNA elements
ES cell	Embryonic stem cell
F	Forward (primer)
FACS	Fluorescence-activated cell sorting

FAIRE-Seq	Formaldehyde-assisted isolation of regulatory elements sequencing
Fig.	Figure
FISH	Fluorescence <i>in situ</i> hybridization
γ H2AX	Phosphorylated histone variant H2AX
G	Guanine nucleotide
GAPDH	Glyceraldehyde 3-phosphate dehydrogenase
GFP	Green fluorescent protein
GR	Glucocorticoid receptor
GSM	Gene expression omnibus SOFT format sample file
Gy	Grays
h	Hours
H1	Linker histone
H2A	Core histone H2A
H2B	Core histone H2B
H2AX	Histone variant H2AX
H2AZ	Histone variant H2AX
H3	Core histone H3
H4	Core histone H4
HAT	Histone acetyl transferase
HDAC	Histone deacetylase
hiBA-FISH	High-throughput break-apart FISH
hi-C	High resolution chromosome conformation capture
HME	Histone-modifying enzyme
HP1	Heterochromatin protein 1
HR	Homologous recombination
HTGTS	high-throughput, genome-wide translocation sequencing
I-SceI	Homing endonuclease that recognizes 18bp sequence not normally occurring in mouse or human genomes
IF	Immunofluorescence
Ig	Immunoglobulin
IHC	Immunohistochemistry
IP	Immunoprecipitation
IR	Ionizing radiation
K	When used after a histone, K = lysine
kb	Kilobases
LacO	Lac operator or operon
LacR	Lac repressor protein
mCh	mCherry protein
Me/ Methyl	Methylated
μ g	Microgram
μ m	Micrometer or micron
μ M	Micromolar
mg	Milligram
mM	Millimolar
ml	Milliliter
MNase-seq	Micrococcal nuclease digestion followed by sequencing

mRNA	Messenger RNA
NCI	National Cancer Institute
ng	Nanogram
NGS	Next generation sequencing
NHEJ	Non-homologous end joining
NIH	National Institutes of Health
nm	Nanometer
PCR	Polymerase chain reaction
px	Pixel
qPCR	Quantitative polymerase chain reaction
R	Reverse primer
RAG	Recombination-activating gene
RefSeq	Reference sequence
RNA	Ribonucleic acid
RNAi	RNA interference
RNA-seq	RNA sequencing
ROI	Region of interest
ROS	Reactive oxygen species
RPKM	Reads per kilobase of transcript per million mapped reads
RT-PCR/qRT-PCR	Quantitative reverse transcriptase polymerase chain reaction
SD	Standard deviation
SEM	Standard error of the mean
Seq	Sequencing
shRNA	Small/short hairpin RNA
SIM	Structured illumination microscopy
siRNA	Small interfering RNA
T	Thymine nucleotide
TALEN	Transcription activator-like effector nucleases
TC-seq	Translocation-capture sequencing
TetO	Tetracycline operator or operon
TetR	Tetracycline receptor
TICdb	Translocation Breakpoints in Cancer database
T _m	Melting temperature
TopoII/TOP2	Topoisomerase II
TSS	Transcription start site
UCSC	University of California, Santa Cruz
UV	Ultraviolet
V(D)J	V(D)J or somatic recombination

Chapter 1

General introduction

INTRODUCTION

The nucleus is a complex organelle that performs a wide array of critical functions. Within the nucleus the genome is highly organized (Spector, 2003). The linear genome sequence exists as intertwined strands of DNA and histone proteins called chromatin fibers. Higher levels of organization are achieved by the compaction of chromatin fibers hierarchically to form larger fibers with the support of structural non-histone proteins and RNA molecules. This compaction, on the order of 10,000 to 20,000-fold during metaphase, is essential for the entire genome to fit within the boundary of the cell nucleus (Woodcock and Ghosh, 2010). In turn, higher-order chromatin fibers give rise to 3D chromosomal domains and ultimately chromosomes (Felsenfeld and Groudine, 2003). It is now known that chromosomes persist as distinct nuclear subdomains, or chromosome territories, which occupy reproducible positions within the interphase nucleus (Boyle et al., 2001; Cremer and Cremer, 2001; Meaburn and Misteli, 2007).

In addition to a packaging role, higher order organization of the genome into chromatin fibers and chromosomes is crucial for genome integrity. This is highlighted by the findings that defects in chromatin organization are linked with disease (Zink et al., 2004; Misteli, 2010). For example, upon DNA damage, cells utilize a complex set of pathways referred to as the DNA damage response (DDR) to mediate repair of DNA lesions, and in the case of irreparable or extensive lesions, trigger senescence or apoptosis (Ciccio and Elledge, 2010). The DDR occurs in the context of higher order chromatin structure, and emerging evidence points to an important role of genome organization in the spatial and temporal coordination of distinct DDR events, including assembly of the repair

machinery, accessibility of the damaged areas to DDR factors, and restoration of the chromatin after repair (Soria et al., 2012). Deregulation of DDR pathways by defects in chromatin organization compromises genome instability and may lead to cancer-causing mutations or the formation of structural or numerical aberrations (Lukas et al., 2011; Hanahan and Weinberg, 2011).

In line with the crucial role of chromatin organization in maintaining genome stability, failure in DDR processes may lead to the illegitimate joining of non-homologous chromosome ends, resulting in chromosome translocations (Roukos et al., 2013). While translocations may arise in some cases as byproducts of global genomic instability in cancer cells, specific translocations can be causal of tumorigenesis by activating oncogenes or creating fusion transcripts that are responsible for malignant transformation (Rowley, 2001). Translocations play particularly prominent roles in hematologic malignancies where they are often the sole abnormality, but recent advances in whole genome sequencing have also identified complex translocation events in solid tumors (Presner and Chinnaiyan, 2009); however, these events are typically not the sole driver of tumor formation. In total, translocations are amongst the most prevalent genetic abnormalities found in human cancers and are estimated to contribute to 20% of cancer morbidity (Mitelman et al., 2007).

Despite the undisputed impact of translocations in cancer, our understanding of the molecular mechanisms that give rise to translocations remains limited (Zhang et al., 2009). It is clear that a translocation event requires DNA double-strand breaks (DSBs) on two or more heterologous chromosomal loci and the physical interaction of these

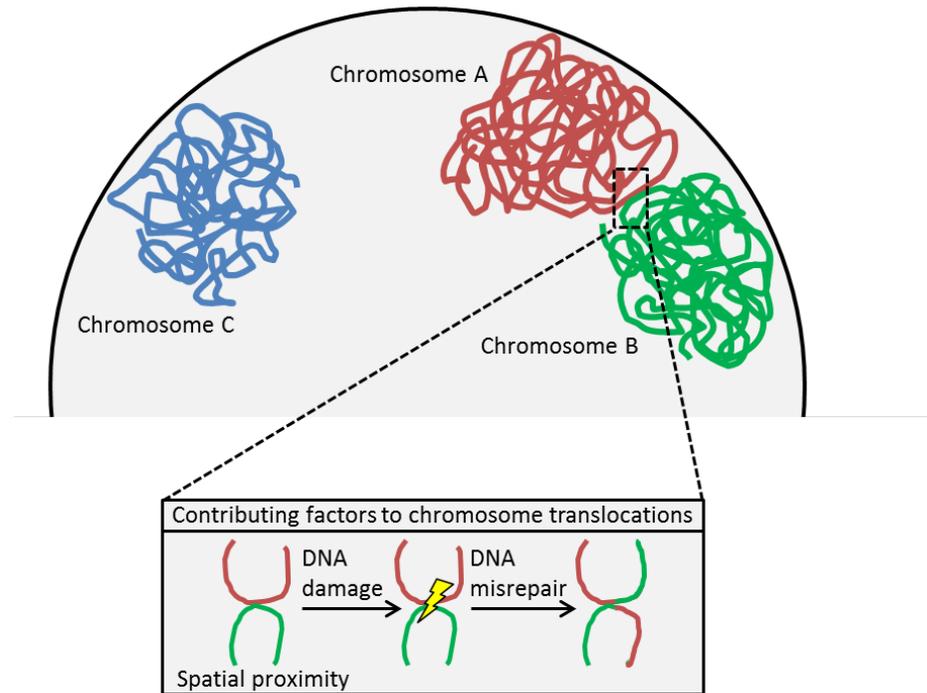


Figure 1-1: Contributing factors of chromosome translocations. Translocations are the end result of DSBs that are caused by cellular stress in the forms of genotoxic, oxidative, replicative, or transcriptional stress. Misrepair of DSBs can result in the formation of translocations. Proximal chromosomes such as A and B translocate at much higher frequency than distal ones, such as A and C, or B and C.

DSBs prior to repair (Fig. 1). Therefore, the spatial organization of the genome *in vivo* is a likely determinant of translocation frequency. In support of this view, numerous cytogenetic studies have pointed to a strong correlation between spatial proximity of chromosomes or genes and their translocation frequencies by showing that proximal genome sites are more prone to form translocations than distal ones (Roukos et al., 2013). For example, analysis of the spatial proximity of the *MYC* gene relatively to its possible translocation partners *IGH*, *IGK* and *IGL*, which form translocations in Burkitt's lymphoma, directly correlates with the observed frequency of these translocations in patients (Roix et al., 2003). Along the same lines, in anaplastic large-cell lymphoma (ALCL) cells the close proximity of *NPM1* and *ALK* facilitates the formation of the *NPM1-ALK* gene fusion upon irradiation (Mathas et al., 2009). Furthermore, translocation frequency correlates with the degree of intermingling between chromosomes, strongly suggesting the local arrangement and proximity of DSBs drives translocations (Pombo et al., 2006).

While these studies provide evidence for the contribution of the spatial arrangement of the genome as a determinant of the outcome of translocations, their correlative and retrospective analysis assumed that these regions form translocations without demonstrating it directly. Chromosome translocations are usually clonal and highly selected, and thus these correlations may not accurately mirror the contribution of spatial organization to translocation frequency. Moreover, most of these studies were limited to a few genes and control regions. Several recent studies overcame this limitation by capturing the genome-wide landscape of translocations in mouse B lymphocytes, which

were carried out before cellular selection affected the translocation detection frequency. Translocation frequencies across the genome were measured and mapped onto linear chromosomes (Chiarle et al., 2011; Klein et al., 2011) or mapped in comparison to chromosome conformation capture data to account for 3D genomic organization (Zhang et al., 2012; Hakim et al. 2012; Rocha et al. 2012). These studies, as well as time-lapse imaging of broken DNA ends in live cells (Roukos et al., 2013), confirmed the cytogenetic studies by demonstrating that translocation frequency is strongly determined by spatial proximity of genome regions.

While the organization of the genome clearly contributes to the formation of translocations, these studies also revealed that not all genomic loci are equally susceptible to translocation and that “hotspots” for DSBs exist (Chiarle et al., 2011; Klein et al., 2011; Hakim et al., 2012). Indeed, breakpoints observed in clinically relevant translocations tend to cluster in the translocating genes, often within introns (Zhang and Rowley, 2006). In addition, of all the genes reported in translocations, some are much more affected than others (Mitelman et al., 2007). It remains largely unclear what factors at the level of individual genes or breakpoints predispose these specific regions to translocations in the first place.

Circumstantial evidence suggests that DNA sequence features may facilitate breakage of genome regions. In support, RAG1/2 endonucleases involved in V(D)J recombination are thought to misrecognize sequences that resemble recombination signal sequences found in V(D)J regions, resulting in inappropriate DSBs and translocations (Raghavan et

al., 2001; Marculescu et al., 2002; Numata et al., 2010). Similarly, the AID endonuclease, involved in somatic hypermutation and class switch recombination, recognizes a single-stranded sequence motif at immunoglobulin (Ig) loci, but misrecognition at non-Ig targets may lead to translocations (Liu et al., 2008; Straszewski et al., 2011; Kato et al., 2012). Common fragile sites (CFS), which are enriched in AT-dinucleotide repeats, and CpG islands may promote genome instability by inducing secondary non-B DNA structures prone to breakage, particularly during replication (Inagaki et al., 2009; Kurahashi et al., 2010). However, large-scale sequencing of translocation junctions has shown that these DNA features are not universal markers of breakpoints and only apply to some translocation events (Talkowski et al., 2011). Therefore, while these studies re-affirm that breakpoints occur in defined regions of the genome, the presence of DNA features does not appear sufficient to promote translocations.

Given the integral role of chromatin in DNA accessibility and repair (Price and D'Andrea, 2013), and the fact that DSBs occur in the context of chromatin, it seems plausible that the local chromatin environment may contribute to predisposing genomic regions to breakage and translocations. In support of this view, the aforementioned genome-wide mapping studies found that translocations occur at higher frequency in transcriptionally active regions of the genome (Chiarle et al., 2011 and Klein et al., 2011). In prostate cancer, liganded androgen receptor (AR), a potent transcriptional activator, binds near breakpoints, promoting the formation of site-specific DSBs and prostate cancer-specific translocations (Lin et al., 2009). Similarly in anaplastic large cell

lymphoma (ALCL), genes near the endogenous breakpoint sites appear transcriptionally active prior to translocation formation (Mathas et al., 2009). These correlative studies point to the possibility that chromatin features, or chromatin remodeling through transcription factor binding, at or near breakpoints may influence susceptibility to translocation formation.

One potential mechanism for a contribution of chromatin to breakage and translocation susceptibility is via histone modifications. Core histone proteins (a H3-H4 heterotetramer and two H2A-H2B dimers) make up the most basic unit of chromatin, the nucleosome, around which 147 bp of DNA is wrapped. The highly basic amino terminal tails of histones project away from the nucleosome and are subject to post-translational modifications that include acetylation, methylation, phosphorylation and ubiquitylation. Histone modifications have been shown to play a key role in determining higher-order chromatin structure and modulating DNA transactions such as transcription, replication, DNA repair, and access to enzymes and regulatory proteins (Bannister and Kouzarides, 2011). In addition, histone modifications have been linked to chromatin status. For example, acetylation is generally indicative of decondensed, transcriptionally active chromatin, while condensed and transcriptionally silent regions are generally enriched in H3K9me3 or H3K27me3 marks (Rando, 2012).

Several histone modifications have been correlated to translocation breakpoints. In one of the genome-wide translocation sequencing studies, breakpoints were enriched for histone modifications associated with active chromatin, such as H3K4me3, H3K36me3,

and H3 acetylation (Klein et al., 2011). H3K4me3 has also been implicated in both RAG and AID endonuclease-mediated DSB mechanisms. The RAG2 plant homeodomain finger binds to H3K4me3 at the Ig locus in V(D)J recombination, and mutation of this domain greatly diminishes the efficiency of recombination (Liu et al., 2007; Matthews et al., 2007; Ramon-Maiques et al., 2007). In T-cells, H3K4me3 peaks at cryptic RAG binding sites in certain translocation breakpoints and it has been proposed that this binding promotes translocations in T-cell leukemias (Shimazaki et al., 2009). In prostate cancer cells treated with liganded AR, H3K79me2, a modification associated with DNA recombination, was found to be enriched near *TMPRSS2* and *ERG* breakpoints (Lin et al., 2009). Overexpression of the H3K79-specific methyltransferase DOT1L significantly increased translocation frequency (Lin et al., 2009). Along the same lines, genome-wide conversion to an H4K20 monomethylation state in mice led to defective DSB repair, Ig class-switch recombination, and IgH translocations (Schotta et al., 2008).

Finally, chromatin organization within the nucleus affects the recruitment of DDR factors (Lukas et al., 2011). After a DSB occurs, damaged chromatin around the break is thought to rapidly decondense to facilitate access of repair machineries, and then recondense as the repair process progresses (Khurana et al., 2014; Burgess et al., 2014). These events are orchestrated by chromatin remodelers that reposition nucleosomes, histone chaperone proteins that exchange core histones for specific histone variants, and histone modifying enzymes (Groth et al., 2007). Higher-order chromatin structure can drastically influence the progression of repair, possibly by impeding recruitment of these proteins (Soria et al., 2012). For example, radiation-induced DSBs in heterochromatin

were observed to repair more slowly than breaks in euchromatin (Goodarzi et al., 2008), and radiation-induced foci seemed to relocate from heterochromatin to euchromatin (Chiolo et al., 2013). These observations are highly relevant to understanding the translocation mechanism since the mis-joining of broken chromosome ends appears to be associated with deregulated DDR pathways (Zhang et al., 2010)

AIMS OF THESIS

The traditional view on translocation formation has been that translocations are spontaneous, albeit low frequency, events that occurred at random genomic loci, and their enrichment within a given tumor type is due to selection of proliferative or survival advantages provided by the translocation carrying cells (Mani and Chinnaiyan, 2010). However, the evidence presented above strongly suggests that translocation events occur at non-random sites in the genome. This leads to a key unresolved question in the field that serves as the basis for this thesis: why do chromosome breaks and translocations occur at particular sites of the genome?

Here, we study the role of histone modifications in predisposing genomic sites to breakage and translocations through three specific aims:

1. Identifying histone modifications linked to chromosomal breakage sites:

We hypothesized that specific sets of histone modifications mark translocation-prone genome regions. In Chapter 2, we developed a unique bioinformatics analysis to compare chromatin modifications at translocation-prone genome regions to control regions. In Chapter 3, we study the effect of these histone modifications on chromatin structure. In Chapter 4, we use chromatin-immunoprecipitation approaches to map a set of histone modifications at well-defined translocation breakpoints in a cancer cell line model.

2. Testing the role of histone modifications on chromosomal breakage susceptibility:

We hypothesized that specific patterns of histone modifications increase chromosome breakage susceptibility in response to genotoxic stress *in vivo*. In Chapter 3, we directly test the role of histone modifications on DSB formation by an endonuclease. In Chapter 4, we directly test the role of histone modifications on DSBs induced by ionizing radiation.

3. Establishing the role of histone modifications in chromosome translocation formation:

We hypothesized that specific patterns of histone modifications promote translocation formation. In Chapter 4, we directly test the role of histone modifications on translocation formation using a novel method to detect low frequency chromosome translocations in human cells. The development of this method is described in Chapter 5.

Taken together, the experiments described in this thesis provide key insights into the role of histone modifications in the formation of nonrandom chromosomal breaks and the mechanisms that lead to cancerous translocations.

Chapter 2

Computational analysis of chromatin features at recurrent translocation sites

Collaborators: Zhuzhu Z. Zhang (University of North Carolina/Salk Institute for Biological Studies) and Jason D. Lieb (University of North Carolina/University of Chicago)

While chromosome rearrangements are found in virtually every tumor type, translocations play a particularly prominent role in hematologic malignancies. The first discovery of a translocation event, the Philadelphia chromosome, was identified in cells from a patient with chronic myelogenous leukemia (CML) in 1960 (Nowell and Hungerford, 1985). This was later identified as a gene fusion between the *BCR* and *ABL1* genes, resulting from a balanced translocation between chromosomes 22 and 9 (Rowley, 1973). The observation that a specific translocation event could serve as a genetic hallmark of leukemia such as CML led to the search for other translocation events that may be causal, and potentially diagnostic, of cancers (Rabbitts, 1994). This was especially important for hematologic malignancies, which can present acutely, and during diagnosis, blood samples can be easily obtained.

As a result, since the discovery of *BCR-ABL1* in CML, hundreds of recurrent translocations have been found to be driver mutations in both leukemias and lymphomas, such as *IGH-MYC* in Burkitt's lymphoma, *PML-RARA* in acute promyelocytic leukemia, *RUNX1-RUNX1T1* in acute myeloid leukemia (Type M2), and *NPM1-ALK* in anaplastic large cell lymphoma (ALCL) (Mitelman et al., 2007). Interestingly, most of these translocations result from combinations of a relatively small number of genes meaning that certain genes participate in multiple translocation events (Mitelman et al., 2007). For example, the *MLL* gene has over 80 known translocation partners (Meyer et al., 2006). The repeated occurrence of certain genes in leukemia and lymphoma translocations

strongly suggests that translocation genes are a non-random set of genes that undergo breakage (Lin et al., 2012).

Therefore, as a first approach to studying chromatin structure at genome regions that are prone to chromosome breakage and translocation, we performed a systematic bioinformatics analysis of chromatin modifications in this well-defined set of translocation genes. We used data for CD34+ cells, an immature cell population that contains hematopoietic stem cells. CD34+ cells are ideally suited for this analysis since several early translocation events in leukemias and lymphomas are postulated to occur within this cell population or within early progenitors derived from CD34+ cells (Bonnet and Dick, 1997, Bernt and Armstrong, 2009). In addition, recent studies have suggested that chromatin patterns are inherited epigenetically from progenitor cells to differentiated cells (Dixon et al., 2015, Polak et al., 2015). As such, while some translocations are enriched only in lymphoid or myeloid lineage-derived cell types, it is reasonable to examine the chromatin structure in hematopoietic stem cells.

Our analysis takes advantage of recent advances in next-generation sequencing (NGS). An important outcome of the NGS revolution has been the development of methods that couple immunoprecipitation of chromatin modifications, such as histone modifications (ChIP-Seq) and nucleosome accessibility (DNase-I-Seq, MNase-Seq and FAIRE-Seq), to massively parallel DNA sequencing (Schones and Zhao, 2008). This has led to the creation of annotated “maps” of chromatin modifications across various genomes that include virtually every tissue and tumor type (Rivera and Ren, 2013). These maps are

deposited and organized into databases such as ENCODE and NIH Epigenomics Roadmap Project (Birney et al., 2007; Bernstein et al., 2010). From this data, the location of a specific chromatin modification, or combinations of modifications, has provided insight into the functional role of the modification. For example, the histone modifications H3K4me3 and H3 acetylation correlate strongly with transcribed regions of the genome, while H3K9me3 correlates strongly with silent genome regions (Rando 2007). The bivalent enrichment of H3K4me3 and H3K27me3 is seen at genes poised for transcription in embryonic stem cells (Bernstein et al., 2006). While these patterns make it possible to generalize chromatin status, ChIP-Seq data has also revealed that several exceptions to these generalizations apply (Rando 2012).

Here, we have computationally analyzed chromatin modification patterns in a large set of frequent translocation genes in CD34+ cells as a screening to identify candidate chromatin features that may mark human translocation genes. Our results complement and inform the biochemical analyses that we present in subsequent chapters.

RESULTS

Selection of datasets, translocation genes and control genes

To begin analyzing genome-wide chromatin data, we searched for publicly available datasets containing sequencing tracks for multiple chromatin features in CD34+ cells. The available data come largely from primary CD34+ cells mobilized from patients' bone marrow, as CD34+ cells differentiate rapidly in culture preventing immortalization and development of cell lines (Fig. 1). The sequencing of chromatin-immunoprecipitated

DNA from a limited number of primary hematopoietic cells was a major milestone in the field of epigenomics, and since its description several datasets have been deposited in the NIH Epigenomics Roadmap Project (Adli et al., 2010; Goren et al., 2010). We selected a primary CD34⁺ cell line generated from a 33 year-old female (Fig. 1). Among other CD34⁺ cell datasets available in the Epigenomics Roadmap Project, this primary CD34⁺ cell line had the most extensive datasets including ChIP-seq for histone modifications H3K4me1, H3K4me3, H3K27Ac, H3K36me3, H3K9me3, and H3K27me3, DNaseI-seq, and mRNA-seq data (Fig. 1).

To identify genome regions that are frequently involved in chromosomal translocations in hematologic malignancies, we mined the Mitelman Database of Chromosome Aberrations and Gene Fusions in Cancer (<http://cgap.nci.nih.gov/Chromosomes/Mitelman>), which catalogs all cases of chromosomal translocations reported in the clinical literature (Mitelman et al., 2007). Of the major databases that compile information on chromosome translocations, including Atlas of Genetic Oncology, Catalogue of Somatic Mutations in Cancer (COSMIC), Database of Chromosomal Rearrangements in Disease (dbCRID), and Translocation Breakpoints in Cancer (TICdb), the Mitelman Database is the only database that allowed sorting of translocation events by number of clinical cases (see Methods below). We set a threshold at a minimum of 10 clinical cases to identify genes from the most prevalent translocation events (range: 10-1,863 cases). Since the Mitelman database collects information on any chromosomal aberration, we filtered the list to only include translocations found in hematologic malignancies. From the filtered list, all of the translocation events involved a total of 84

Expand All
 Collapse All
 Search:

	Bisulfite-Seq	MeDIP-Seq	MRE-Seq	RRBS Signal	DNaseI	DGF	mRNA-Seq	smRNA-Seq	ChIP-Input	H3K27me3	H3K36me3	H3K4me1	H3K4me3	H3K9ac	H3K9me3	H3K27ac	H2AK5ac
Hematopoietic Stem																	
CD34, Primary 39y/M																	
CD34, Mobilized 23y/F																	
CD34, Mobilized 27y/F																	
CD34, Mobilized 30y/M																	
CD34, Mobilized 33y/F																	
CD34, Mobilized 36y/M																	
CD34, Mobilized 37y/M																	
CD34, Mobilized 39y/M																	
CD34, Mobilized 42y/F																	
CD34, Mobilized 43y/M																	
CD34, Mobilized 48y/M																	
CD34, Mobilized 50y/F																	

Figure 1: Available NIH Roadmap Epigenomics Project data for CD34+ cells. The 33 year-old female sample was chosen because of its extensive datasets.

translocation genes. Ten genes were further excluded because 7 genes had no mRNA-seq expression data, and 3 genes were redundant.

Our goal was to compare the levels of chromatin modifications between translocation-prone genes and “control genes” to test their hypothesis that regions that undergo chromosome breakage and translocation are marked by altered chromatin features. The most important criterion for defining control genes was that these genes had not been reported to form translocations in any cancer. At the time of the study, the number of genes involved in any translocation was 2,301 genes, so to generate a set of control genes, we began with a master list of 24,021 genes (26,322 total RefSeq genes – 2,301) genes. Since chromatin modifications are influenced by gene expression, and the expression status across the 74 translocation genes was variable, we decided to compare each translocation gene to control genes with similar expression (Fig. 2). In addition, the 74 translocation genes varied in size, exon/intron content, number of exons, and sequence

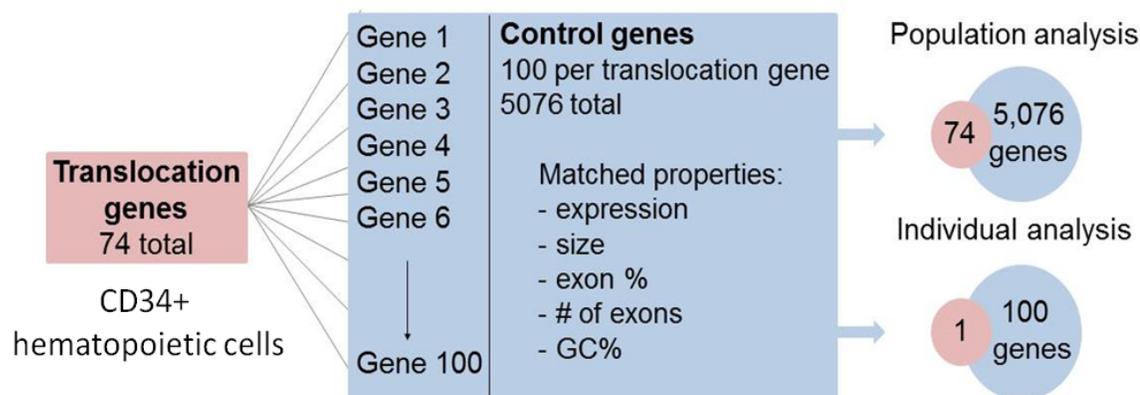


Figure 2: Strategy for analysis of chromatin features using a population or individual gene approach. 74 frequent translocation genes were identified and for each translocation gene, 100 control genes were selected with closely matched properties: expression, size, exon %, number of exons, and GC%.

	Gene	Exp	length	% of exon	# of exon	GC%
translocation	PRDM16	5.2	369,444	2.35%	17	47.1%
control	NOL4	5.2	371,371	1.00%	11	43.8%

Figure 3: An example of a translocation gene with one of its matched control genes.

characteristics, so we controlled for these properties as well (see Methods, Fig. 2; example, Fig. 3). Ultimately, a list of 100 control genes with similar properties, and no reported involvement in translocation events, was created for each translocation gene. Altogether 5,076 RefSeq genes (not 7,400 due to overlapping genes) were selected in total for the 74 translocation genes.

Population analysis

We first used a population comparison approach in which we compared the population of 74 translocation genes to the aggregate population of 5,076 control genes. The ChIP-seq and DNaseI-Seq tag densities over the 74 reported translocation genes, including 2kb upstream and downstream flanking regions, was measured and compared to the population of control genes (see Methods). Among the histone modifications examined, the population of translocation genes had a statistically significantly higher level of H3K4me1 than the population of control genes (Fig. 4A, Wilcoxon test p-value: 0.00008). There was no statistically significant difference of DNase-I hypersensitivity, H3K4me3, H3K9me3, H3K27ac, H3K27me3, or H3K36me3 between the translocation genes and control genes, although the shapes of the curves indicated that the two populations were not uniform in their enrichment levels (Fig. 4B-G).

Individual gene analysis

Therefore, to achieve higher sensitivity and to eliminate masking of enrichments in subpopulations of genes, each translocation gene was compared to its set of 100 matched control genes to determine whether subpopulations of genes were defined by altered levels of chromatin modifications. For each chromatin modification, the ChIP-seq and DNaseI-seq tag density of the translocation gene was ranked amongst its control genes from 0-100 based on the percentage of control genes with tag density below the translocation gene (examples Fig. 5, rest Appendix 1). This allowed us to study the relative level of a given chromatin modification across a translocation gene compared to

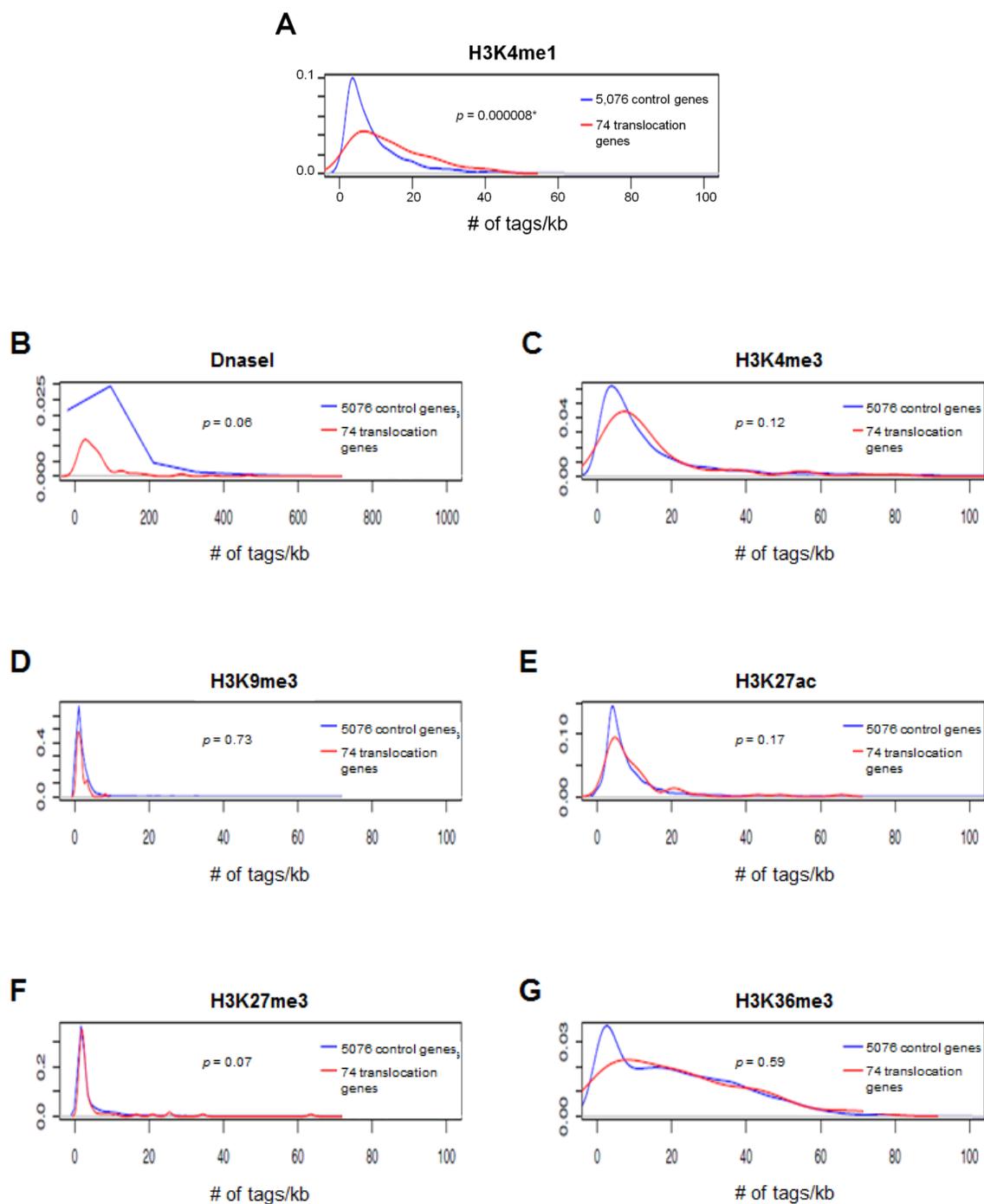


Figure 4: Population analysis: A-G: Kernel density plots comparing (A) H3K4me1 (B) DNaseI, (C) H3K4me3, (D) H3K9me3, (E) H3K27ac, (F) H3K27me3, and (G) H3K36me3 signal density distributions in the pooled set of 74 translocation genes versus 5,076 control genes. Signal density was measured over each annotated gene body \pm 2kb and the average read count (# of tags per kb per 1M reads) in each gene is represented on the x-axis. P-values calculated by Wilcoxon test, * $p < 0.01$.

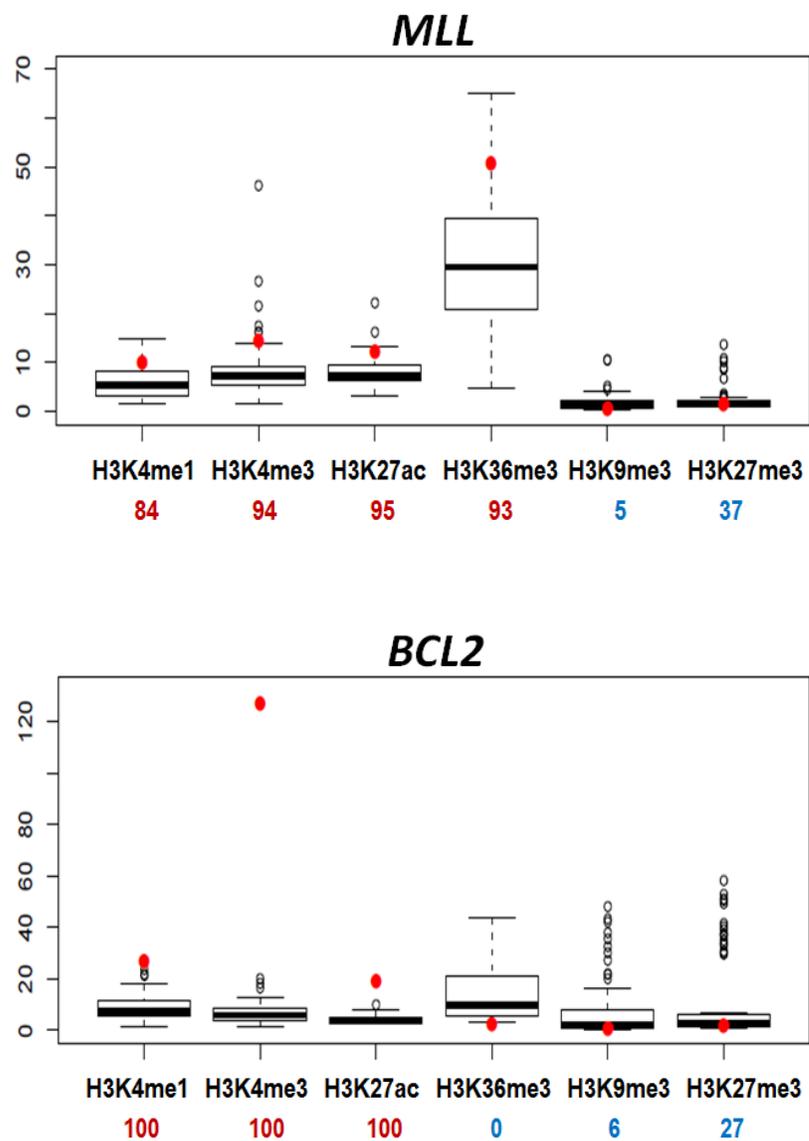


Figure 5: Examples of individual gene analysis (rest, Appendix 1). The average read count (# of tags per kb per 1M reads) for each histone modification of each translocation gene (red dot) was compared to its 100 control genes (Box plot). The percentage of control genes of which tag densities was lower than that of the translocation gene was calculated (rank listed below x-axis).

control genes with similar properties and provided insight on the variance of chromatin modifications across 101 genes with similar properties.

Next, the translocation genes were hierarchically clustered based on these rankings (Fig. 6). This analysis confirmed that H3K4me1 was strongly enriched in a majority of genes (41/74 genes ranked ≥ 75 ; average rank 71.1). In addition, most H3K4me1 marked translocation genes were mutually enriched in H3K4me3, H3K27Ac, and DNaseI hypersensitivity, such as *MLL*, *BCL2*, and *ETV6*. Importantly, these enrichments were not due to elevated transcriptional activity of the translocation genes since control genes were selected for comparable expression levels. Interestingly, 15 translocation genes, such as *RUNX1*, *CEBPA*, and *SEPT11*, featured bivalent enrichment of H3K4me3 and H3K27me3, which is commonly seen in genes poised for transcription in stem cells (Bernstein et al., 2006). This subset included four out of five *Hox* genes that we analyzed, confirming observed bivalent domains at the *Hox* gene cluster in embryonic stem cells (Bernstein et al., 2006).

Since the levels of histone modifications and chromatin accessibility are influenced by transcriptional activity, we also ordered the genes in the above heatmap by gene expression, which was determined from mRNA-seq data (Fig. 7). This revealed that a large cluster of genes in the bottom half of expression, including *BCL6*, *ELL*, and *RUNX1T1*, was characterized by depletion of H3K9me3 (14/37 ranked ≤ 25 ; average rank 37.8) as compared to control genes, despite their transcriptional silence based on RNAseq data.

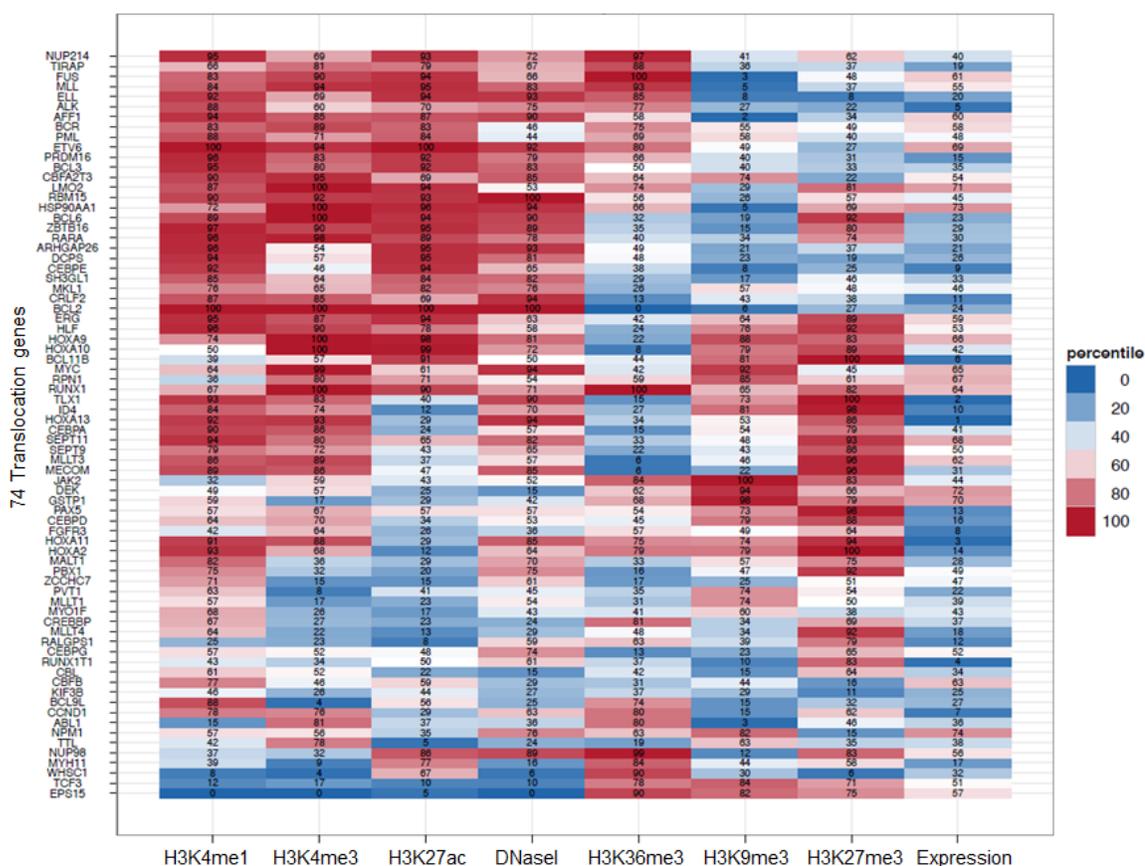


Figure 6: Heatmap of individual analyses of histone modifications and DNaseI hypersensitivity for 74 translocation genes. The percentiles of each translocation gene ranked amongst its control genes are indicated. The rank of expression for translocation genes from 1-74 is shown on the right. Translocation genes were hierarchically clustered using the overall ranking matrix.

Breakpoint analysis

Some chromatin modifications tend to localize to distinct regions of genes and have different shapes of signals (Wang et al., 2008). H3K4me3 is typically found near transcription start sites (TSSs) of actively transcribed genes. H3K27ac strongly co-localizes with enhancer elements. H3K36me3 and H3K27me3 signals can be found along the entire gene body, with H3K36me3 marking actively transcribed genes and

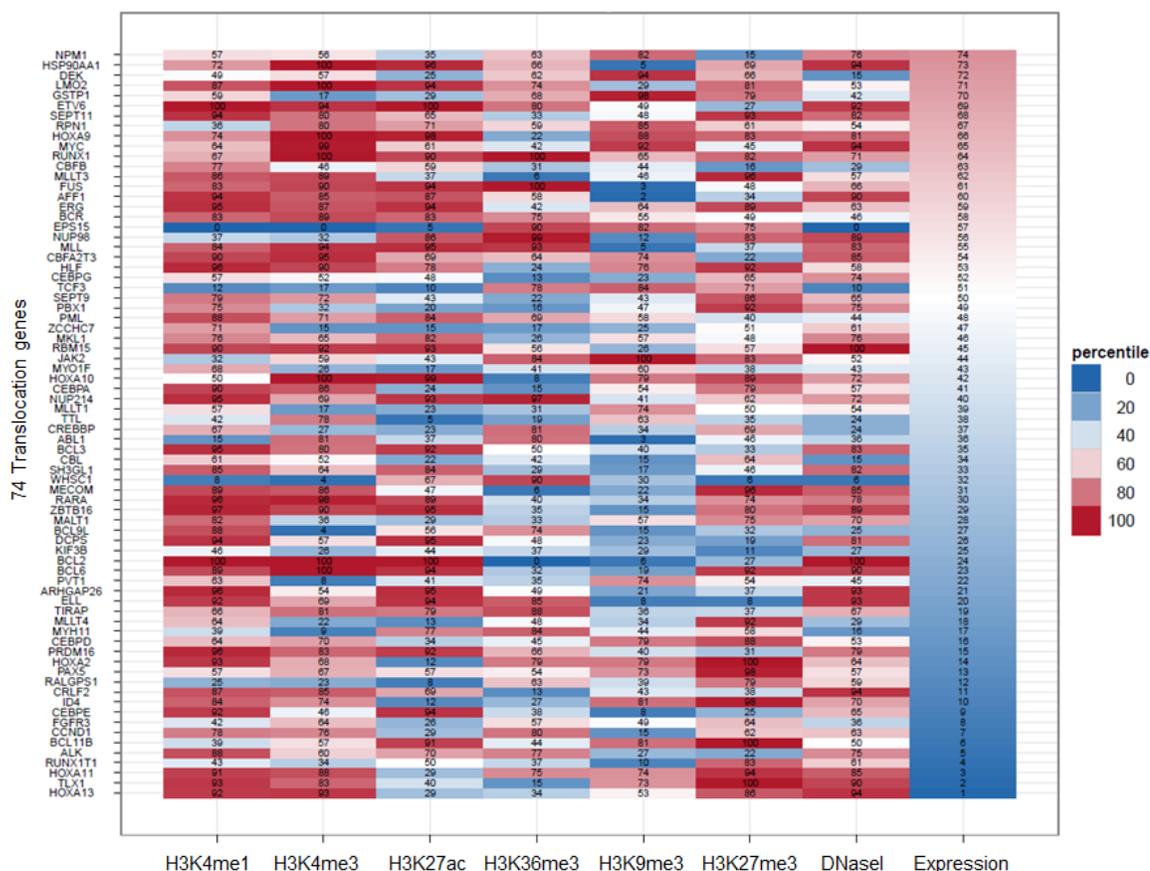


Figure 7: Same data as Fig. 6 but ranked by gene expression (right column). The percentiles of each translocation gene ranked amongst its control genes are indicated. Several non-transcribed and lowly-transcribed genes, including *BCL6*, *ELL* and *RUNX1T1*, show depletion of H3K9me3 compared to similarly expressed control genes.

H3K27me3 marking transcriptionally repressed genes. Interestingly, H3K27me3 can be found in domains along with H3K4me3, marking transcriptionally poised sites in embryonic stem cells (Bernstein et al., 2006). Finally, histone modifications may have unique patterns at intron-exon boundaries (Schwartz et al., 2009; Kolasinska-Zwierc et al., 2009).

Breakpoints are the positions within a chromosome that break recurrently to form translocations. Most breakpoints cluster within genes, particularly in introns, forming breakpoint cluster regions (bcr) (Zhang and Rowley, 2006). Therefore, to study the local chromatin structure at definitive sites of chromosome breakage and translocations, we attempted to perform a more focused comparative analysis of chromatin modifications between bcrs and matched control regions. A major limitation to this analysis is determining precise bcr information. For several translocation genes, the bcrs at the DNA level have not been sequenced or cloned. In addition, many genes feature multiple breakpoint regions scattered throughout the gene body in different introns, making it difficult to define bcrs. For example, the *BCR* gene (named for being the first gene identified with bcrs) has a major bcr (introns 12-16), minor bcr (intron 1), and micro bcr (intron 19) spread out across the gene (Zhang and Rowley, 2006). Finally, in some cases, the intron within which breakage occurs can be deduced if the exon-exon junction at the site of gene fusion is determined; however, the actual bcr within the intron is unknown.

Nevertheless, we were able to reliably identify the coordinates of bcrs in 15 of 74 translocation genes from the literature as well as the TICdb (<http://www.unav.es/genetica/TICdb/>) (Table 1). To confirm these sources, we also searched for mRNA annotations of gene fusions in the UCSC genome browser to see if the exon-exon junctions matched the reported bcr introns. Bcrs mapped in one or multiple introns in the translocation genes, resulting in regions of lengths from ~1kb to ~229kb. To increase our coverage for the analysis, we extended each bcr by 1kb on the 5' and 3' end.

Table 1: Major bcrs for 15 translocation genes reported in the literature. Coordinates map to human genome build 19.

Gene / chromosome	Major bcr	Breakpoint coordinates (hg19)
ABL1 / chr. 9	Intron 1	133589843-133729451
ALK / chr. 2	Intron 19	29446395-29448326
MLL / chr. 11	Introns 8-14	118353137-118359475
PBX1 / chr. 1	Intron 2	164532549-164761730
RUNX1 / chr. 21	Intron 6	36206899-36231770
TCF3 / chr. 19	Intron 16	1615821-1619109
MYH11 / chr. 16	Introns 28-32	15818850-15815278
PML / chr. 15	Intron 6	74325756-74326818
NPM1 / chr. 5	Intron 4	170818804-170819713
BCR / chr. 22	Introns 12-16	23630284-23637342
CBFB / chr. 16	Intron 5	67116212-67132612
ETV6 / chr. 12	Introns 4-5	12006496-12037378
RARA / chr. 17	Intron 2	38487649-38504567
DEK / chr. 6	Intron 9	18226474-18236682
BCL11B / chr. 14	Intron 3	99642533-99697681

In the individual gene analysis above, we defined control genes as non-translocating genes with similar properties. Using the same 100 genes, we selected control regions for the bcrs in two different ways. As mentioned above, each set of 100 control genes shared similar length to its translocation gene, although some variability is expected. Therefore, control regions of the same *relative* length compared to the whole gene (control regions I), and regions of the same *absolute* length (control regions II), were selected from the 100 control genes of each translocation gene (see Methods).

Each bcr and its corresponding control regions (I and II) were divided into 10 equal-sized, non-overlapping windows, and the chromatin modification signal density was calculated in each window and plotted (example Fig. 8 using control regions I, rest Appendix 1). The results from using control regions I and control regions II were comparable. Again, for each chromatin modification, the tag density (across all 10 windows) of the bcr was ranked amongst its control regions and hierarchically clustered based on these rankings (Fig. 9). This analysis revealed that H3K36me3 was strongly enriched at a majority of bcrs compared to control regions (14/15 ranked ≥ 75 ; average rank 89.9), followed by H3K27ac (9/15 ranked ≥ 75 ; average rank 76.7). Interestingly, H3K27me3 was depleted at a majority of bcrs (11/15 ranked ≤ 75 ; average rank 20.5).

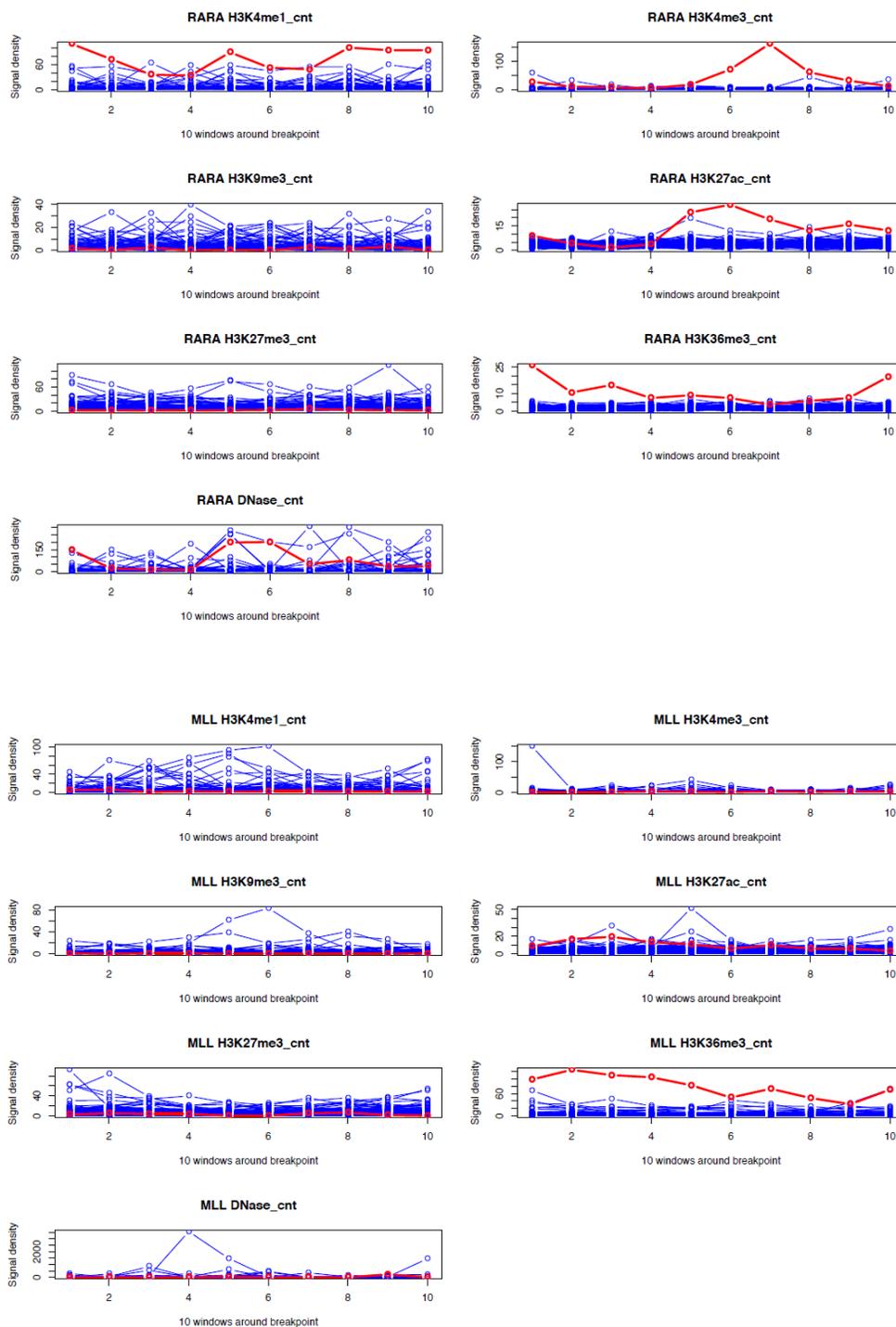


Figure 8: Examples of breakpoint analysis (rest, Appendix 1). The average read count (# of tags per kb per 1M reads) for each histone modification over the bcr (red line) was compared to its 100 control bcrs (blue lines) in ten equal-sized windows using Control Regions I (see Methods).

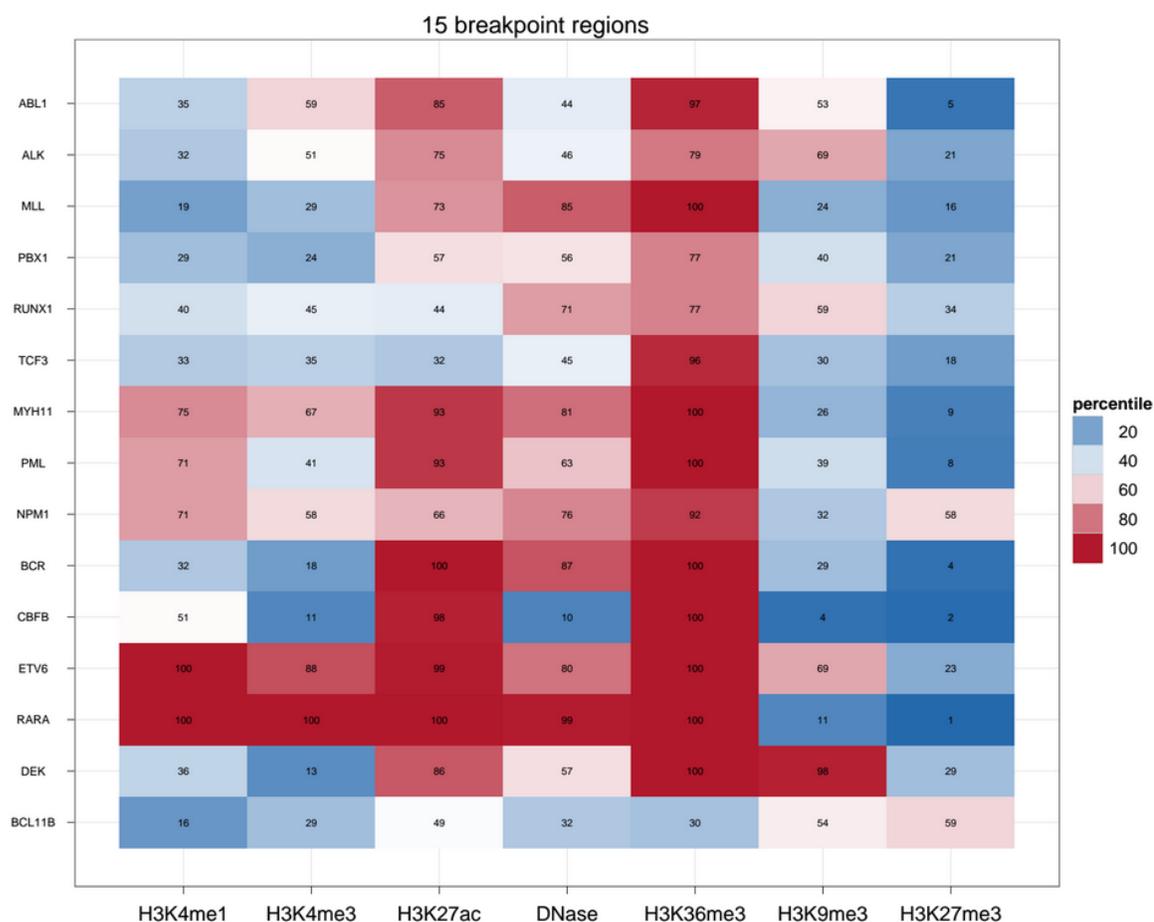


Figure 9: Heatmap of breakpoint analyses of histone modifications and DNaseI hypersensitivity for 15 translocation genes. The percentiles of each translocation gene bcr ranked amongst its control bcrs, using Control Regions 1 (see Methods) are indicated.

MATERIALS AND METHODS:

Datasets

RNA-, ChIP-, and DNase-seq data on a primary CD34+ cell line generated from a 33-year-old female (donor ID: RO 01549) was selected from the NIH Epigenomics Roadmap Project (Bernstein et al. 2010) (<http://www.ncbi.nlm.nih.gov/geo/roadmap/epigenomics/>). BED files of sequencing reads mapped to NCBI build

37/hg19 human reference genome from mRNA-seq (GSM909310), DNaseI-seq (GSM530657), and ChIP-seq data for active histone modifications H3K4me1 (GSM621451), H3K4me3 (GSM621439), H3K36me3 (GSM706843), H3K27ac (GSM772894) and repressive histone modifications H3K9me3 (GSM621436), H3K27me3 (GSM706844) were used in analyses.

Selection of genes

For selection of translocation genes, the Mitelman Database was downloaded from the Cancer Genome Anatomy Project in May, 2012 (<http://cgap.nci.nih.gov/chromosomes/mitelman>). The master list of reported chromosome rearrangements (recurrent_data.dat) was sorted by total cases. We set a threshold of ≥ 10 cases and identified 104 unique RefSeq genes involved in recurrent translocations (range 10-1863 cases). From this list, 20 genes were excluded because they occurred in ≤ 10 cases of hematopoietic malignancies, 7 genes were excluded because there was no expression data, and 3 genes were excluded due to redundancy. For each of the remaining 74 translocation genes, we selected a set of control genes from human RefSeq genes (Release 57) (Kent 2002; Pruitt et al. 2005) using the following criteria: 1) gene expression, 2) gene length, 3) percentage of transcript that was exonic, 4) number of exons, and 5) percentage of G/C of each control gene. RefSeq genes that ranked within ± 10 percentiles of the translocation gene for all five criteria were selected as control genes. Gene expression was measured by RPKM calculated from mRNA-seq data. GC% was calculated using R package Repitools v1.4.0 (Statham et al. 2010). If more than 100 genes fulfilled the criteria, 100 genes with the most similar gene expression levels were selected. Ten of 74 genes had less than 100

control genes: HSP90AA1 (91 genes), ETV6 (87), RUNX1 (76), TCF3 (73), BCL2 (71), RARA (70), BCR (35), PRDM16 (23), CBFA2T3 (20), and SEPT9 (15 control genes). For the population analysis (Fig. 1B, S1), 5,076 control genes were selected for all 74 translocation genes due to overlapping genes.

Measuring histone modification and chromatin accessibility levels

For each translocation gene and each control gene, the number of mapped sequencing tags in gene body (defined as transcribed region \pm 2kb for whole gene analyses and breakpoint cluster region \pm 1kb for breakpoint analysis) from each ChIP-seq and DNase-seq dataset were calculated and normalized by the length of the region and the number of all mapped tags in each dataset. This normalized number of mapped sequencing tags was referred to as tag density per kb and used to represent the level of histone modification or chromatin accessibility.

Whole gene analysis

Translocation genes and control genes were compared as both population (74 translocation genes vs. 5,076 control genes) and individual (1 translocation gene vs. 100 control genes) analyses (Fig. 2). For the population analysis, the distribution of tag densities for each histone modification and DNaseI hypersensitivity was compared between gene bodies (defined above) of 74 translocation genes and 5,076 control genes. Density plots were generated using function density available in R Project. For the individual analysis, the levels of histone modifications and chromatin accessibility in the gene body of each control gene were compared to those of its selected control genes. For each histone modification or DNaseI, the tag density of the translocation gene was

compared to the tag densities of its control genes in boxplots using function `boxplot` in R. The percentage of control genes of which tag densities were lower than that of the translocation gene was calculated and the translocation gene was given a rank (0-100) equal to this percentage. The 74 translocation genes were clustered by their ranks for each histone modification and DNaseI hypersensitivity using function `hclust` available in R and heatmaps were generated using R package `ggplot2` v0.8.9 (Wickham, 2009).

Breakpoint analysis

Coordinates of breakpoint regions in 15 translocation genes were collected from literature. Each breakpoint region was extended by on the 5' end and 1kb on the 3' end (referred as breakpoint +/- 1kb region). For each breakpoint +/- 1kb region, the breakpoint start position compared to the TSS, the fraction of the gene covered by the breakpoint region, and the breakpoint mid-point position compared to the TSS were calculated. For example, one breakpoint region could start at a position that was 10% (of the length of the gene) downstream of TSS with a length equal to 20% of the gene length and whose midpoint is 20% downstream of TSS. Control "breakpoint" regions from each control gene were selected in two different ways to ensure the robustness of the results: I) Control breakpoint region of same relative size compared to length of control gene: for each control gene, a breakpoint region was defined at the same relative position as in the true breakpoint gene. For the example above, the region would start at the base pair corresponding to 10% of the control gene length from the TSS, and the size of the region was of 20% of the control gene size. It should be noted that the length of each selected control breakpoint region was different and not equal to the size of the true breakpoint

region due to differences in control gene sizes. II) Control breakpoint region was the same size as true breakpoint region: for each control gene, a breakpoint region of the same absolute length as the true breakpoint region was selected. The selected region was centered at the same relative position in the control gene as the position of the center of the true breakpoint region within the translocation gene. For example, if the true breakpoint region was 5kb long and centered at a position 20% of the length of translocation gene from the TSS, the the control breakpoint region was defined as a 5kb region centered at a position 20% of the length of the control gene from the TSS. In this case, the lengths of breakpoint region and its control regions were the same.

Each breakpoint region or control region was divided into 10 equal-sized, non-overlapping windows. The tag density per kb (defined as above) of each histone modification and DNaseI was calculated in each window. Tag densities in windows of each breakpoint region were compared to those of control regions.

Chapter 3

Chromatin factor tethering and double strand break formation by endonucleases

Collaborators: Rebecca C. Burgess (National Cancer Institute)

ChIP data in CD34+ cells (Chapter 2) indicate altered levels of specific histone modifications at translocation prone genome regions. The enrichment of H3K4me1, H3K4me3, H3K27ac, and H3K36me3, and the depletion of H3K9me3, are patterns that are strongly associated with actively transcribing genes. However, translocation genes in CD34+ cells and ALCL cells showed altered levels of these modifications when compared to genomic regions with similar expression status, and included genes with both high and low transcriptional activities. This led to the hypothesis that translocation-prone genome regions are marked by aberrant accumulation or reduction of specific histone modifications, causing chromatin structure in the region to become more vulnerable to breakage.

Gross changes to chromatin structure can be visualized by microscopy. When tagged with a fluorescent protein, the transcriptional activators BRG1 and VP-16 open chromatin when tethered as lac repressor fusions to lac operator arrays (Tumbar et al, 1999; Burgess et al., 2014). The tethered arrays markedly increase the volume of the tethered chromatin. Similarly, tethering of heterochromatin protein 1 (HP1) isoforms causes compaction of the lac operator array (Danzer and Wallrath, 2004; Li et al., 2003; Verschure et al., 2005). All three constructs also affect the “texture” or appearance of the array, suggesting secondary structures induced by chromatin remodeling, with BRG1 and VP16 dramatically spreading chromatin fibers (Tumbar et al., 1999, Verschure et al., 2005). However, the effect of chromatin remodeling by specific histone modifications remains to be tested by microscopy. Understanding the effect of a local accumulation of

histone modifications on chromatin structure could have functional implications, since chromatin is the context for DNA-based events such as transcription, replication, DNA damage susceptibility, and the DNA damage response (DDR).

Histone modifications are regulated by histone modifying enzymes, which catalyze the addition (“writers”) or removal (“erasers”) of one or more covalent modifications at specific lysine residues on histone and non-histone proteins (Marmorstein and Trievel, 2009). Histone modifying enzymes vary in their catalytic mechanisms, substrate specificity, and regulation, with some enzymes writing or erasing only one modification and others writing or erasing several different modifications. Most histone modifying enzymes reside in complexes that regulate their activity, for example, by targeting them to specific genomic loci. Finally, deregulation of histone-modifying enzymes has been implicated in tumorigenesis, especially in hematologic malignancies, making these enzymes potential targets for cancer therapy (Waldmann and Schneider, 2013).

Here we sought to directly test in a controllable system the role of histone modifications on chromatin structure and DNA breakage. We use a combination of imaging and biochemical assays based on the tethering of histone modifying enzymes to a lac operator array that is upstream of a unique endonuclease site (Fig. 1). First, we visually detect and quantitate the effect of tethering on chromatin structure by microscopy. Second, we measure double strand break (DSB) formation within histone modification domains by challenging these domains with endonuclease. Endonuclease-mediated damage is highly relevant to translocation formation, since several translocation events in B-cells and T-

cells result from the off-target effect of endogenous endonucleases. Finally, we study the recruitment of DDR factors after histone modification domains are challenged by DSBs. Impairment of the DDR signaling cascade by chromatin structure could allow chromosome breakage to persist and result in translocation formation.

RESULTS

Characterization of histone modification domains

We created chromatin domains enriched in specific histone modifications using a previously characterized lac repressor/operator protein-chromatin tethering system (Tumbar et al., 1999; Nye et al., 2002; Verschure et al., 2005; Wang et al., 2005; Soutoglou and Misteli, 2008). The system consists of the stable integration of ~10 kb tandem arrays, each containing 256 copies of the lac operator (LacO) sequence and 96 copies of the tet operator (TetO) sequence flanking a unique I-SceI restriction endonuclease site, at two different chromosomal locations in human U2OS cells (Fig. 1). To these sites, we transiently tethered lac repressor proteins fused to various histone modifying enzymes and a mCherry tag for fluorescent detection of the modified chromatin domains (Fig. 1). Limiting the expression of these constructs to 24 hours after transient transfection prevents potential replication defects caused by extended tethering of fusion proteins to the lac array (Jacome and Fernandez-Capetillo, 2011, Burgess et al., 2014).

First, we identified histone modifying enzymes that modulate the modifications we found to be enriched or depleted at translocation sites in CD34+ and ALCL cells. The selection

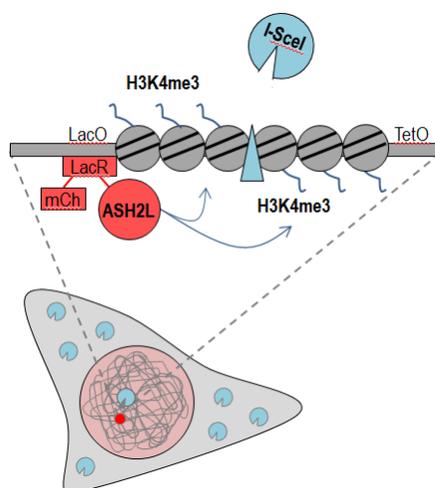


Figure 1: Schematic representation of chromatin-protein tethering system. Lac repressor (LacR) fusions to either mCherry alone or to histone-modifying enzymes are tethered to the LacO-I-SceI-TetO array in U2OS cells after transient expression. DSBs are formed after transient expression of the glucocorticoid receptor I-SceI (GR-I-SceI), which localizes to the cytoplasm with trace amounts in the nucleus due to leakiness of cytoplasmic retention.

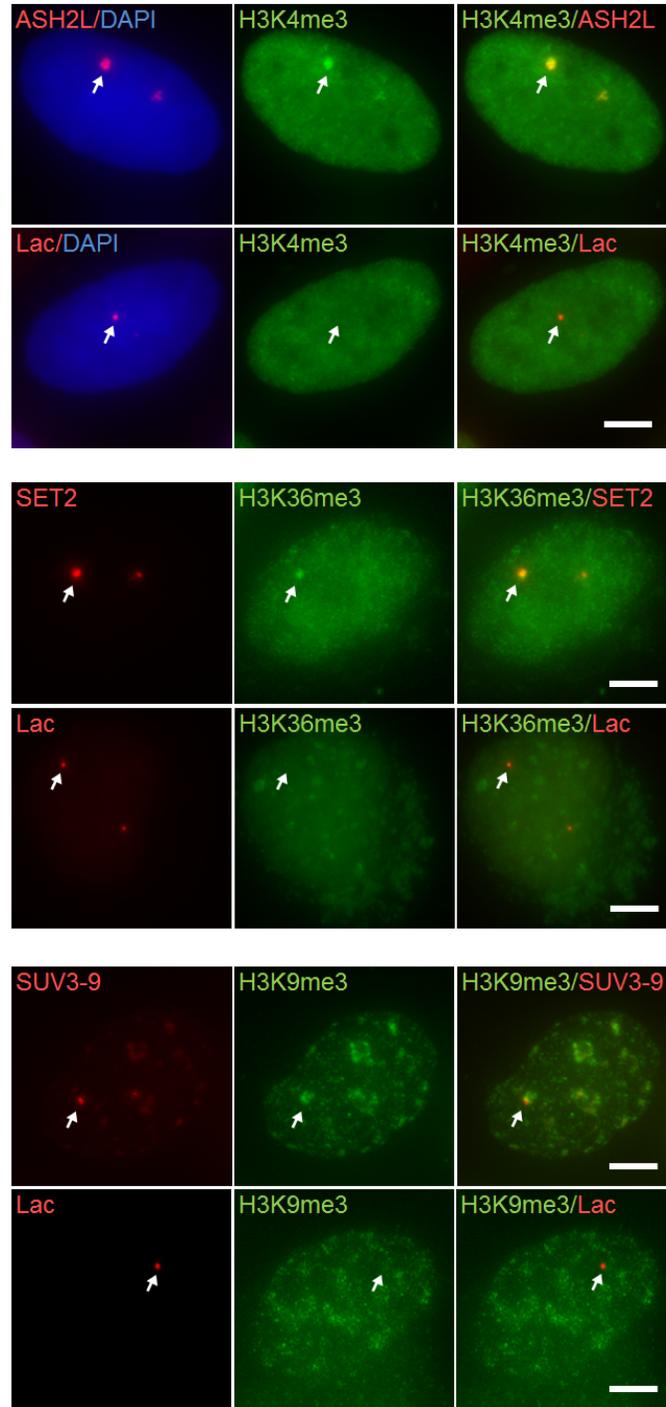
of enzymes was largely a result of enzymes already available and characterized in the Misteli laboratory (Table 1). We cloned the coding sequence of each enzyme into a mCherry-Lac expression vector.

Table 1: Histone modifying enzymes cloned into mCherry-Lac constructs.

Enzyme (cloned into mCherry-Lac)	Histone-modifying function
SET7/9	H3K4 mono-methyltransferase
ASH2L	H3K4 di- and tri-methyltransferase
SUV3-9	H3K9 tri-methyltransferase
EZH2	H3K27 mono-, di-, and tri-methyltransferase
ySET2* *yeast version of human SETD2	H3K36 mono-, di-, and tri-methyltransferase
TIP60	multiple H3 and H4 lysine acetyltransferase

Next, we confirmed the functional activity of each mCherry-Lac histone modifying enzyme construct. The mCherry-Lac with no enzyme (hereafter referred to as mCherry-Lac) was used as a control in all experiments. After transient transfection of each construct, cells were fixed at 20 hours. Histone modification deposition was first determined by indirect immunofluorescence using antibodies for specific modifications deposited by each enzyme: anti-H3K4me1 for mCherry-Lac-SET7/9, anti-H3K4me3 for mCherry-Lac-ASH2L, anti-H3K9me3 for mCherry-Lac-SUV3-9, anti-H3K27me3 for mCherry-Lac-EZH2, anti-H3K36me3 for mCherry-Lac-SET2, and anti-pentaAc (recognizes five acetylation residues) for mCherry-Lac-TIP60 (Fig. 2). Co-localizations of a bright immunofluorescence signal (Alexa488) with the tethered array (mCherry) were quantified (Fig. 3). These experiments indicated that histone modification domains are formed in a majority of cells 12 hours after transfection

Interestingly, no co-localizations were seen with the anti-H3K27me3 antibody after tethering of mCherry-Lac-EZH2 or the anti-H3K4me1 antibody after tethering of mCherry-Lac-SET7/9. We hypothesized that this could be due to the sensitivity of the antibodies, instability of these modifications, or the functionality of the constructs. Therefore, we first tested other antibodies and none resulted in co-localized signals for either construct. Next, we used ChIP with primers specific to the Lac array to confirm deposition of the histone modification. After repeated attempts, we could not confirm the presence of H3K27me3 at lac arrays tethered with mCherry-Lac-EZH2 (data not shown). However, we did see positive results with mCherry-Lac-SET7/9. In all replicates of these experiments, we also performed ChIP on cells expressing mCherry-Lac-ASH2L



Figures 2-4: Enrichment of histone modifications in LacR-tethered chromatin domains. Indirect immunofluorescence at Lac arrays of **(2)** anti-H3K4me3 upon mCherry-Lac-ASH2L tethering, **(3)** anti-H3K36me3 upon mCherry-LacR-SET2 tethering, and **(4)** anti-H3K9me3 upon mCherry-LacR-SUV3-9 tethering, compared to control mCherry-Lac tethered arrays (bottom panels) in U2OS cells (array shown by arrows). The indicated constructs were expressed for 20h, fixed, and immunostained. Images are maximum intensity projections of representative cells. Scale bars 5 μ m.

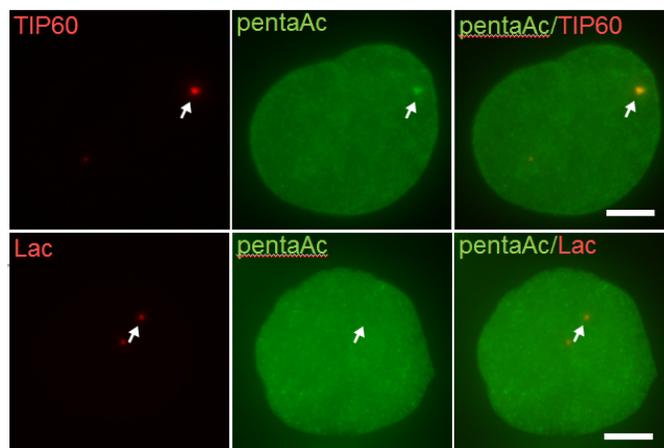


Figure 5: Enrichment of histone modifications in LacR-chromatin domains. Indirect immunofluorescence at Lac arrays of anti-pentaAc upon mCherry-Lac-TIP60 tethering compared to control mCherry-Lac tethered arrays (bottom panel) in U2OS cells (array shown by arrows). The construct were expressed for 20h, fixed, and immunostained. Images are maximum intensity projections of representative cells. Scale bars 5 μ m.

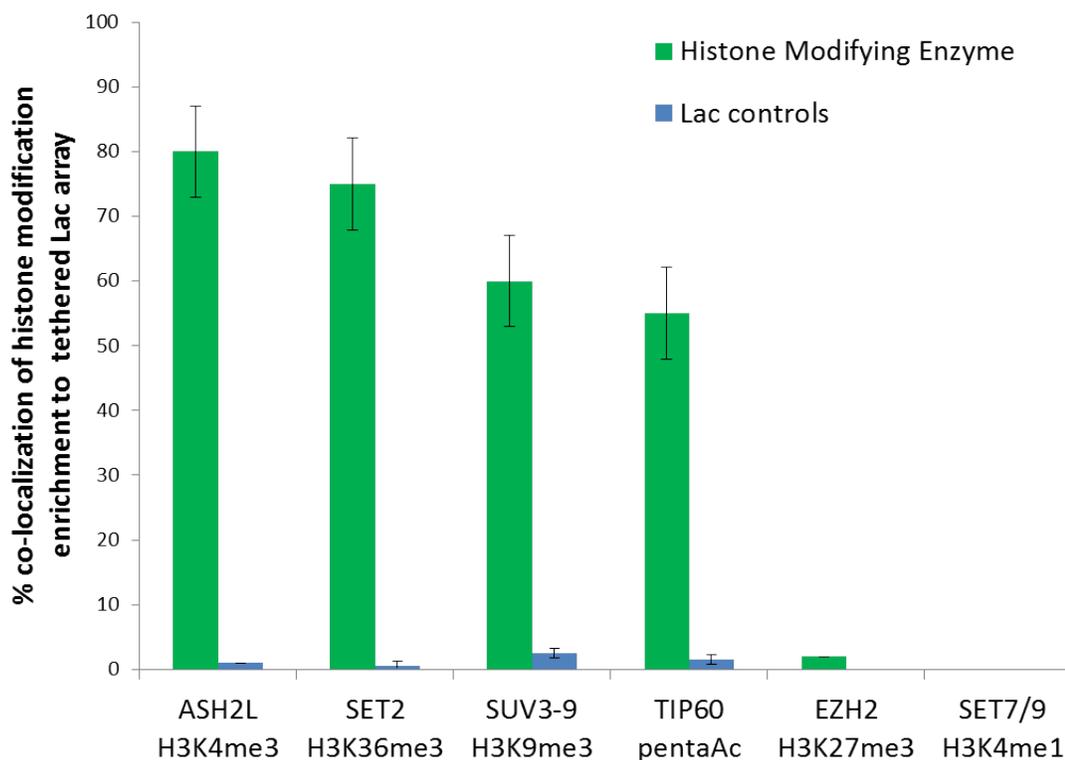
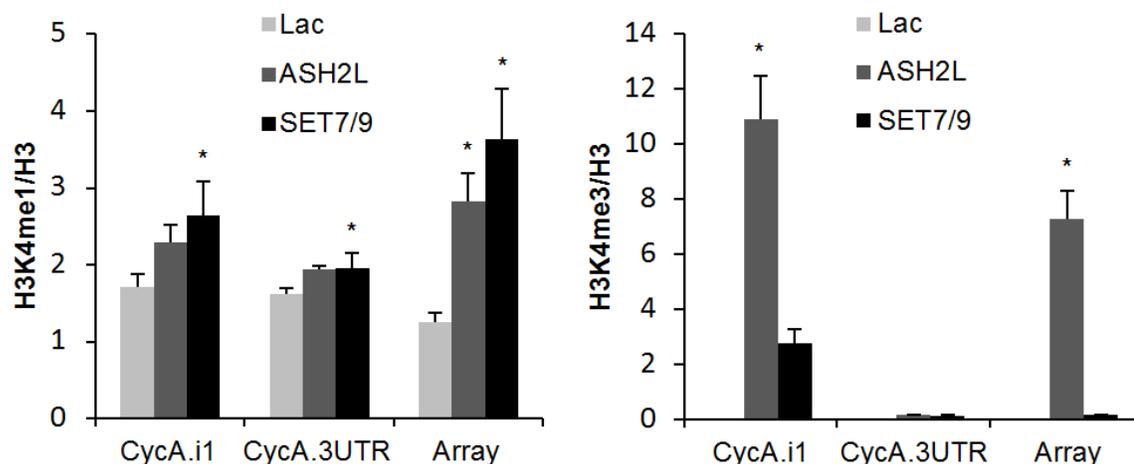


Figure 6: Percentage of arrays with positive histone modification enrichments (co-localizations) for the mCherry-Lac constructs containing the indicated histone-modifying enzyme. Data are from at least 3 experiments counting > 100 arrays each.

and performed an anti-H3K4me3 immunoprecipitation in parallel to anti-H3K4me1 (Figs. 7-8). These experiments revealed that mCherry-Lac-SET7/9 specifically induced a 2-fold increase in H3K4me1 at tethered arrays. Interestingly, mCherry-Lac-ASH2L also increased the levels of H3K4me1 by ~2-fold, although a much higher enrichment (~30 fold) was seen for H3K4me3 (Fig. 8). In these cells, H3K4me3 was also increased at intron 1 (~4-fold) of the highly expressed *Cyclophilin A* gene (*CycA*); however, it was not increased at transcriptionally silent loci such as *Nanog* and Sat-II repeats (SatII) on chromosome 1 (Fig. 8). This indicated that LacR-ASH2L acts globally like its untagged counterpart by preferentially acting at regions already marked by H3K4 methylation.

From these ChIP experiments, we also calculated the ratio of immunoprecipitated H3 DNA (ChIP for unmodified H3) to total DNA input (control ChIP with no antibody) at the lac array primers to see if tethering H3K4 methyltransferases influenced the relative nucleosome density across the lac array. Strikingly, the recovery of immunoprecipitated H3 DNA was less (~4.5-fold) in cells expressing mCherry-Lac-ASH2L or mCherry-Lac-SET7/9 than in cells expressing mCherry-Lac alone, despite comparable amounts of total DNA input (Fig. 9). This suggested that accumulation of H3K4 methylation modifications could remodel chromatin structure to a more decondensed state.

The fusion of an mCherry tag to the histone-modifying enzymes allowed us to visualize the texture and quantitate the size of chromatin at tethered arrays (Table 2). Noticeably, arrays tethered with mCherry-Lac-ASH2L and mCherry-Lac-SET7/9 were the largest in diameter and featured a coarse, “popcorn-like” appearance when visualized, similar to



Figures 7-8: ChIP-qPCR of (7) H3K4me1 and (8) H3K4me3 at the control *CycA* gene (intron 1 and 3' UTR) and LacO-I-SceI-TetO array (Array) in U2OS cells expressing mCherry-LacR-ASH2L (ASH2L, dark gray), mCherry-LacR-SET7/9 (SET7/9, black) or mCherry-LacR alone (Lac, light gray) for 20h. The percentage of input was normalized to unmodified H3. Values represent means \pm SEM from three independent experiments. * $p < 0.05$ to Lac, Student's *t* test.

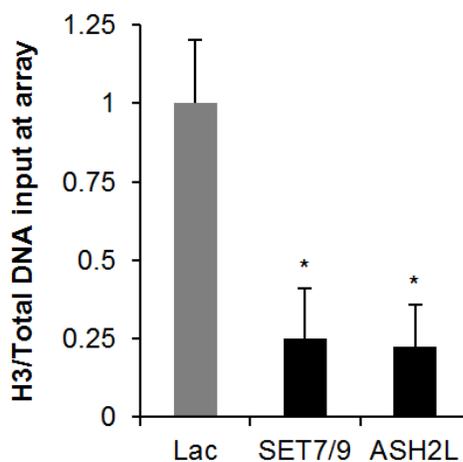


Figure 9: H3 density calculated from ChIP-qPCR of unmodified H3 relative to total DNA input at primers for the array, in cells expressing mCherry-LacR fusions to the indicated histone-modifying enzymes or mCherry-LacR alone (Lac). Values represent means \pm SD from three independent experiments. * $p < 0.05$, Student's *t* test.

the effects of tethering chromatin remodelers such as BRG1, while arrays tethered with mCherry-Lac-SUV3-9 and mCherry-Lac alone were compact and smooth-edged. Arrays tethered with mCherry-Lac-SET2 and mCherry-Lac-TIP60 developed coarse structures with sizes between that of mCherry-Lac-ASH2L and mCherry-Lac alone. These data also point to the remodeling of chromatin to a more decondensed state by H3K4 methyltransferases.

Table 2: Size and texture of arrays tethered by histone modifying enzymes.

Construct	Median array area \pm median absolute deviation (>50 nuclei*)	Schematic of array texture
mCherry-lac-BRG1	$2.43 \pm 0.75 \mu\text{m}^2$	
mCherry-lac-ASH2L	$1.24 \pm 0.38 \mu\text{m}^2$	
mCherry-lac-SET7/9	$1.16 \pm 0.45 \mu\text{m}^2$	
mCherry-lac-TIP60	$1.05 \pm 0.44 \mu\text{m}^2$	
mCherry-lac-SET2	$1.04 \pm 0.31 \mu\text{m}^2$	
mCherry-lac-SUV3-9	$0.45 \pm 0.20 \mu\text{m}^2$	
mCherry-Lac	$0.42 \pm 0.16 \mu\text{m}^2$	

*In nuclei with more than one array, the largest array was measured

Measurement of DSB formation within histone modification domains

To probe the effect of translocation-relevant histone modifications on chromatin accessibility and DSB formation, we challenged cells with histone modification domains with low levels of the I-SceI restriction enzyme. To do this, we expressed I-SceI fused to the glucocorticoid receptor and cyan fluorescent protein (CFP-GR-I-SceI) 12h after transfection of LacR fusion constructs, which caused its localization to the cytoplasm (Fig. 10). In previous experiments, we repeatedly observed that a trace amount of

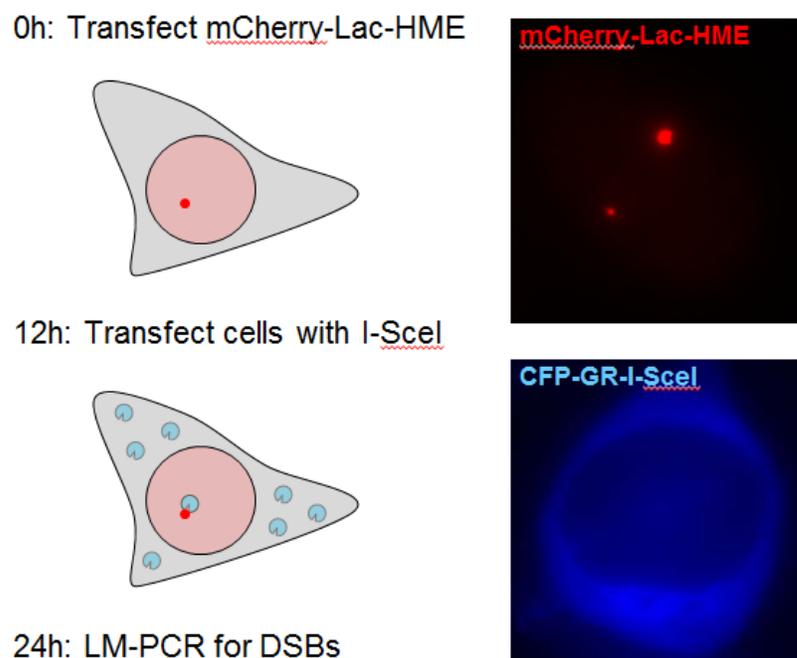


Figure 10: Schematic of assay to test the effect of specific histone modification domains on DNA breakage by I-SceI endonuclease. First, histone-modifying enzymes (HME) are tethered to the array as fusions to mCherry-LacR. Second, to induce low levels of I-SceI, the enzyme fused to the glucocorticoid receptor is transfected, causing its localization in the cytoplasm. A low amount of I-SceI leaks into the nucleus without the addition of dexamethasone. Third, DSBs are quantified by ligation-mediated PCR (LM-PCR).

CFP-GR-I-SceI enters the nucleus due to an intrinsic leakiness in U2OS cells without addition of GR ligands such as dexamethasone (Burgess, unpublished). Cells were harvested 12h after the transfection of CFP-GR-I-SceI (24h after transfection of histone-modifying enzymes) and DSBs at the array were detected by ligation-mediated PCR with one primer specific for the unique I-SceI cut site and one primer in the LacO repeats (Fig. 10). As a positive control for this assay, we challenged cells expressing mCherry-Lac-BRG1 with GR-I-SceI, which grossly decondenses the array upon tethering. We observed a twenty-fold increase in DSB formation between arrays tethered with

mCherry-Lac-BRG1 and mCherry-Lac alone in multiple experiments, indicating a dynamic range for the assay (Fig. 11).

In experiments where we tethered histone-modifying enzymes, H3K4-methylated domains created by mCherry-Lac-ASH2L and mCherry-Lac-SET7/9 allowed for 5-fold and 4-fold increases in DSB formation than mCherry-Lac alone (Fig. 12, $p < 0.05$). These domains were nearly two-fold more permissible to DSBs than H3K36 methylated domains created by mCherry-Lac-SET2 and hyperacetylated domains created by mCherry-Lac-TIP60 (Fig. 12, $p < 0.05$). In contrast, the mCherry-Lac-SUV3-9-mediated H3K9 methylated domain allowed two-fold less DSBs than mCherry-Lac alone, indicating that H3K9 methylation remodeled chromatin to a highly condensed state (Fig. 12, $p < 0.05$). Interestingly, similar results were seen when cells were treated with dexamethasone 20 min prior to harvesting, which causes translocation of CFP-GR-I-SceI into the nucleus, suggesting that histone modifications modulate DSB formation even in the presence of persistent endonuclease activity (Fig. 16; Burgess et al., 2014).

Recruitment of DDR factors after DSB formation within histone modification domains

DSBs are often measured by detection of the phosphorylated form of the core histone H2A variant H2AX (γ H2AX), which has been observed to accumulate rapidly at sites of DNA damage (Rogakou et al., 1999). γ H2AX binds MDC1, which enhances MRN-ATM binding and causes a positive feedback loop that spreads γ H2AX along the chromatin surrounding the DSB (Kinner et al., 2008). The spreading of γ H2AX, in domains around one megabase of DNA, results in discrete foci formation that can be detected by indirect

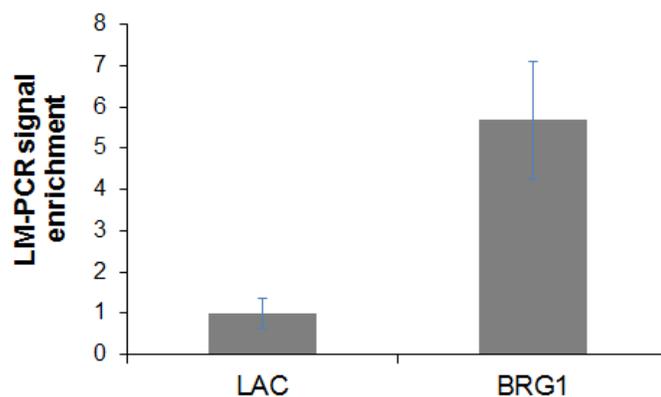


Figure 11: Quantification of DSBs by LM-PCR in cells expressing mCherry-Lac alone (LAC) or mCherry-Lac-BRG1 (BRG1) and GR-I-SceI. mCherry-Lac-BRG1, which grossly decondenses chromatin, reveals a dynamic range of DSB formation.

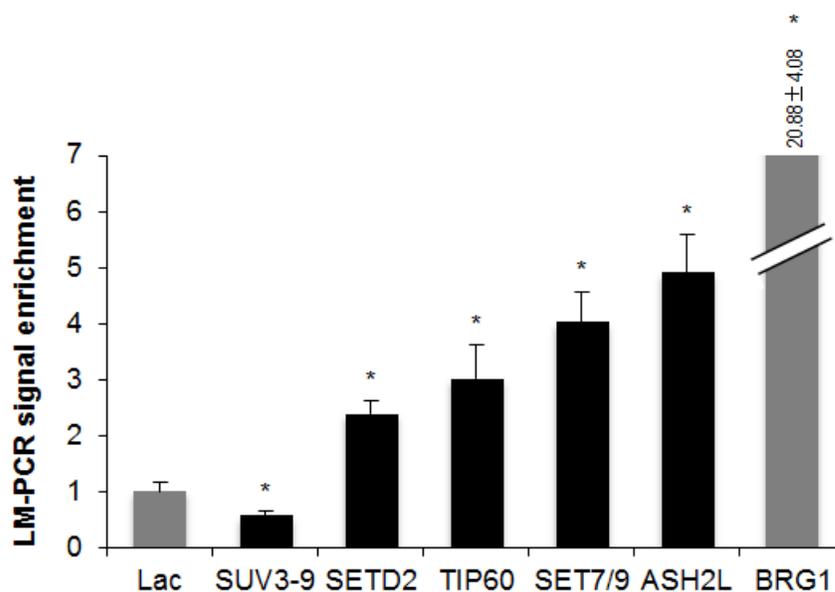


Figure 12: Quantification of DSBs by LM-PCR in cells expressing mCherry-LacR fusions to the indicated histone-modifying enzymes or mCherry-LacR alone (Lac) and GR-I-SceI. Values represent means \pm SD from three independent experiments. * $p < 0.05$, Student's t test.

immunofluorescence (Rogakou et al., 1999). γ H2AX is thought to form within seconds after DSB formation, however, discrete foci are usually visible 15-30 minutes later (Lobrich et al., 2010).

We studied the recruitment of γ H2AX as a marker for DSB formation and DDR activation by indirect immunofluorescence. The same timeline was used as in LM-PCR experiments with cells fixed 12 hours after GR-I-SceI transfection (no dexamethasone treatment). As with DSBs measured by LM-PCR, a dynamic range between γ H2AX foci formation between the negative control mCherry-Lac and positive control mCherry-Lac-BRG1 was first determined and we observed a two-fold increase between the two constructs (Fig. 13).

In experiments where we tethered histone-modifying enzymes, we found that H3K9-methylated domains created by mCherry-Lac-SUV3-9 had two-fold γ H2AX foci formation than mCherry-Lac alone, similar to mCherry-Lac-BRG1 domains. Intriguingly, H3K4-methylated domains created by mCherry-Lac-ASH2L and H3K36-methylated domains created by mCherry-Lac-SET2 had levels γ H2AX foci formation similar to mCherry-Lac alone in repeated experiments (Fig. 14). This was surprising since domains created by mCherry-Lac-ASH2L or mCherry-Lac-SET2 tethering were more than two-fold permissive to DSBs than mCherry-Lac alone (Fig. 12), and it indicated that γ H2AX foci may not be a reliable marker for DSB formation.

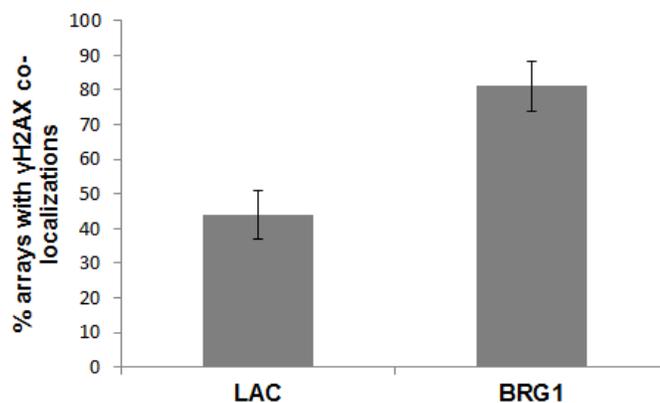


Figure 13: Percentage of γ H2AX co-localizations at tethered arrays in cells expressing mCherry-Lac (LAC) or mCherry-Lac-BRG1 and CFP-GR-I-SceI. mCherry-Lac-BRG1, which grossly decondenses chromatin, reveals a dynamic range of γ H2AX co-localization percentage. Data are from three experiments counting >100 arrays each. Values are means \pm standard deviations.

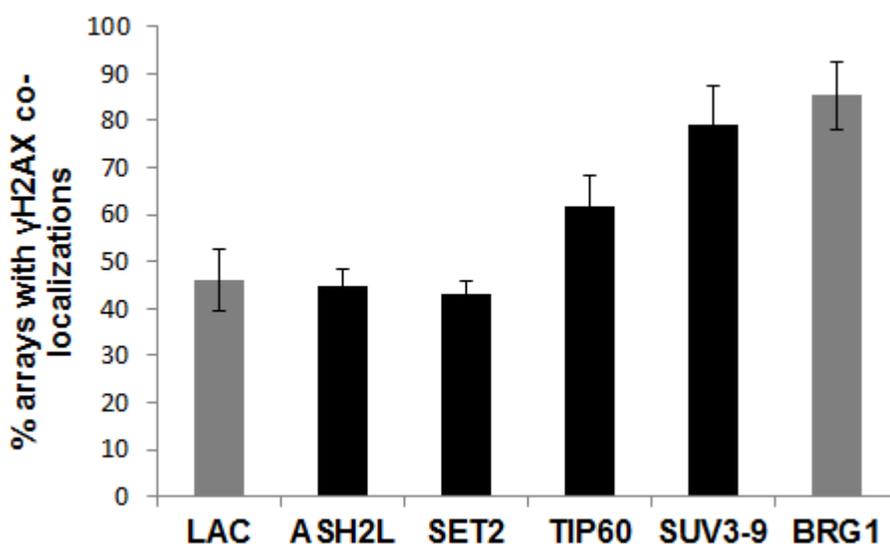
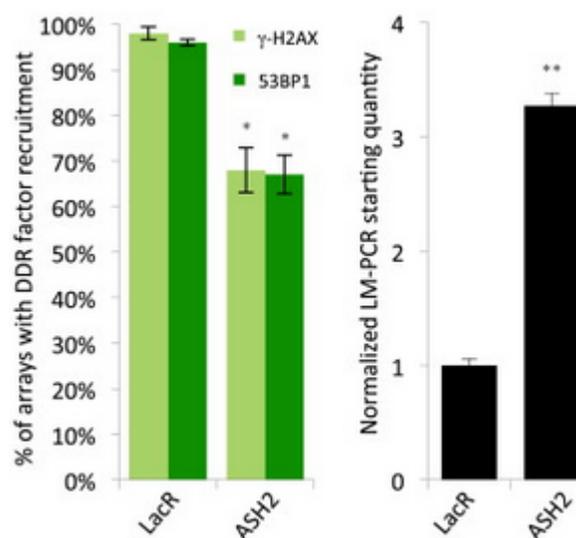


Figure 14: Percentage of γ H2AX co-localizations at tethered arrays in cells expressing mCherry-LacR fusions to the indicated histone-modifying enzymes or mCherry-LacR alone (LAC) and GR-I-SceI. Data are from three experiments counting >100 arrays each. Values are means \pm standard deviations.

Therefore, we performed additional experiments to confirm the inverse relationship between DSB formation and γ H2AX foci formation in domains tethered by mCherry-Lac-ASH2L. We repeated both LM-PCR and immunofluorescence experiments, but with cells treated with dexamethasone 20 min prior to fixation, which causes translocation of CFP-GR-I-SceI into the nucleus (Burgess et al., 2014). In addition to γ H2AX, we also performed immunofluorescence for 53BP1, another marker of DNA DSBs with similar kinetics to γ H2AX (Panier and Boulton, 2014), since results with two DDR markers would strengthen our findings. These experiments revealed that even in the presence of saturated I-SceI endonuclease activity at the array, and a 3-fold increase in DSB formation by LM-PCR ($p < 0.01$), recruitment of the DDR factors γ H2AX and 53BP1 were attenuated (Fig. 15-16, $p < 0.05$; Burgess et al., 2014).



Figures 15-16: (15) Percentage of mCherry-Lac (LacR) or mCherry-Lac-ASH2L (ASH2) tethered arrays with γ H2AX or 53BP1 co-localizations 20 min after CFP-GR-I-SceI induction by dexamethasone treatment. Data are from three experiments counting >100 arrays each. **(16)** LM-PCR detecting the quantity of DSBs from the same quantity of DSBs from the same experiments as **15**. Values are means \pm standard deviations, * $p < 0.05$, ** $p < 0.01$.

MATERIALS AND METHODS

Cell culture

The stable LacO-I-SceI-TetO U2OS cell line for tethering experiments (Soutoglou et al. 2007) was maintained in DMEM at 37°C and 5% CO₂. Culture medium was supplemented with 10% FBS (Atlanta Biologicals), 2 mM L-glutamine, 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin. Prior to transfection of CFP-GR-I-SceI, LacO-I-SceI-TetO U2OS cells were cultured in DMEM with 10% charcoal-dextran treated serum (Atlanta Biologicals) for 48 hours at 37°C and 5% CO₂.

Chromatin tethering

The mCherry-LacR-ASH2L, mCherry-LacR-SET2, and mCherry-LacR-TIP60 tethering constructs were generated by amplifying cDNAs from pcDNA3-Flag-hASH2L-HA (provided by Dr. D. Skalnik, IUPUI), yeast pN823-Flag-Set2 (provided by Dr. B. Strahl, UNC), and pCDNA3-HA-TIP60 (provided by Dr. J. Cote), respectively, using primers containing SmaI and XbaI sites and ligating them into the MCS of mCherry-LacR (Burgess et al., 2014). An EZH2-GFP plasmid from Alexander Tarakhovsky (Rockefeller University) was used for sequential subcloning to replace GFP with GFP-LacR-NLS using BsrG1 and MfeI sites, followed by replacement of the GFPLacR with mCherry-LacR excised with AgeI and EcoRV from the original mCherry-LacR repressor construct (Dundr et al., 2007). The mCherry-LacR-SET7/9 construct was made by amplifying SET7/9 cDNA from pcDNA3.1-SET7/9-Flag (Addgene plasmid 24084, gift of Dr. D. Reinberg) using primers containing BamHI and XbaI sites. The mCherry-LacR-SUV3-9H1, mCherry-LacR-BRG1, and CFP-I-SceI-GR constructs were previously

described (Burgess et al. 2014). All constructs were confirmed by sequencing. Transient transfections of LacR constructs were carried out by electroporation of 2 μ g of construct per 1 million LacO-I-SceI-TetO U2OS cells using the Amaxa Nucleofector Kit V (Lonza) according to the manufacturer's protocol. For IF and ChIP experiments, cells were fixed 20 hours after transfection. For I-SceI experiments without dexamethasone, 2 million cells were transfected first with 4 μ g of a chromatin tethering factor and 12 hours later cells were trypsinized, counted, and transfected with 5 μ g CFP-I-SceI-GR per 1 million cells using the same protocol. After 12 hours (24 hours total), cells were harvested for LM-PCR. For experiments with dexamethasone (Sigma), CFP-GR-I-SceI was activated with dexamethasone at a concentration of 100 nM for 20 min prior to harvesting. For all I-SceI experiments, cells were maintained in DMEM with 10% charcoal-dextran treated serum (Atlanta Biologicals) starting 48 hours prior the first transfection.

Immunofluorescence and imaging

Indirect immunofluorescence was performed on LacO-I-SceI-TetO U2OS cells as previously described (Soutoglou et al. 2007) after a 10 minute fixation with 1:1 methanol/acetone at 4°C. Images were captured on a DeltaVision workstation equipped with a CCD camera (CoolSNAP HQ; Photometrics) mounted on an Olympus IX70 microscope with a 60 \times 1.42 NA oil immersion objective. 20-50 focal planes were captured at 0.2-0.5 μ m intervals, and analyzed with the softWoRx package (Applied Precision). Antibodies that stained positively by immunofluorescence were as follows: rabbit anti-H3K4me3 (1:250 Upstate 05-745), rabbit anti-H3K9me3 (1:500 Upstate 07-442), rabbit anti-H3K36me3 (1:500 Abcam ab 9050), rabbit anti-hyperacetylated (penta)

H4 (1:500 Upstate 06-946), anti- γ H2AX-serine 139 (1:1000 Millipore 05-636), and anti-53BP1 (1:1000 Novus Biologicals NB100-304).

Chromatin immunoprecipitation (ChIP)

ChIP assay was performed as previously described (Luco et al., 2010). In brief, 2×10^6 cells per sample were crosslinked for 10 min in 1% formaldehyde at room temperature, quenched with 125 mM glycine, and swelled on ice for 10 min. Chromatin was sonicated (Bioruptor, Diagenode) to an average length of 200-500 bp and incubated overnight with pre-coated anti-IgG magnetic beads (Dynabeads M-280 Invitrogen) previously incubated with the antibodies for 6 h at 4°C. The antibodies used were: rabbit anti-H3K4me1 (5 μ g, Abcam ab8895), rabbit anti-H3K4me3 (5 μ g, Upstate 05-745), rabbit anti-H3K9Me3 (5 μ g, Abcam ab8898), mouse anti-H3K27me3 (10 μ l, provided by Dr. H. Kimura, Osaka) and rabbit anti-H3 (1 μ g, Abcam ab1791). Control immunoprecipitations were performed with no antibody and ChIP for histone modifications was normalized to anti-H3 ChIP. Subsequently, the beads were washed several times and eluted in 1% SDS and 100 mM NaHCO₃ buffer for 15 min at 65°C. The eluates were incubated at 65°C for 6 hours to reverse crosslinks. Chromatin was precipitated with 100% ethanol overnight, treated with proteinase K, and purified using a PCR purification kit (Qiagen). Immunoprecipitated DNA (1.5 μ l) and serial dilutions of the 10% input DNA (1:4, 1:20, 1:100 and 1:500) were analyzed with SYBR-Green (iQ Supermix Bio Rad) on a Bio Rad C1000 ThermoCycler. Primers used in this study are listed below.

hCycA1.i1	F: CCCCACCCACCTATGAGTGTAGT R: ACCCCTCCATTCTCATCAAGACCT	Tm: 60
hCycA.3utr	F: ATTCCTGGGTGATACCATTCAAT	Tm: 60

	R: ATGACAACGTGGTGAGGCTATTCT	
Lac_array	F: GAGTGGTAACTCGACATTACCCTG R: GAGGCGCCGAATTCCACAAAT	Tm: 60
hSatII	F: CATCGAATGGAAATGAAAGGAGTC R: ACCATTGGATGATTGCAGTCAA	Tm: 60

Ligation-mediated PCR

Genomic DNA was purified from LacO-I-SceI-TetO U2OS cells using the Blood and Tissue Kit (Qiagen). Ligation-mediated PCR was performed using an asymmetric adaptor ligatable to the unique I-SceI overhang as previously described (Soutoglou et al. 2007; Roukos et al. 2013b). Real-time PCR was performed using a primer to the adaptor and the TetO repeats with conditions that allowed amplification of a single product. PCR reactions were conducted at 98°C for 2 minutes, followed by 40 cycles at 98°C for 10 seconds and 55°C for 10 seconds on a C-1000 Thermal Cycler, CFX-96 (Biorad). Sequences of interest were amplified using the SsoFast Eva Green Super Mix (Biorad) using 2ng of DNA and 0.5 µM of each primer per reaction. Amplification of the I-SceI-TetO DNA was monitored after insertion of an additional heating step in the protocol before the plate reading (74°C, 5 sec), which was selected based on the melting curve and allowed the monitoring of the first full length I-SceI-TetO amplicon. Amplification of the genomic *GAPDH* locus was used for normalizing loading variability. Each reaction was carried out in triplicate and normalized to a standard curve made by dilution series of the sample showing the highest amplification (for both I-SceI-TetO and *GADPH* amplifications). Primers used in this study are listed below:

LM-I-SceI adaptor	F GCATCACTACGATGTAGGATG	Tm: 55
-------------------	-------------------------	--------

TetO repeats	R	TCGACTTTCACCTTTTCTCTAT	
hGAPDH	F:	CTGGGGAGGGACCTGGTATGTTC	Tm: 55
	R:	TGCCAGCTTCCTGTAGCACTCAAG	

Chapter 4

Effect of histone modifications on chromosome breakage and translocations in anaplastic large cell lymphoma

The presence of altered levels of histone modifications at translocation sites as compared to non-translocation sites in CD34⁺ cells (Chapter 2) indicates a possible role for chromatin structure in predisposing genomic sites to chromosome breakage and translocation. However, CD34⁺ hematopoietic stem cells do not proliferate and they differentiate quickly *in vitro*, making it challenging to use this system to study translocation formation (R. Van Etten, personal communication). The experiments in Chapter 3 demonstrate that chromatin modifications influence DSB formation by endonuclease. Although endonuclease-mediated translocations are highly relevant to B-cells and T-cells, the Lac-I-SceI-Tet arrays we used for chromatin tethering were integrated in U2OS cells for their ease of transfection and imaging, and the arrays do not resemble endogenous translocation breakpoints. Therefore, while these studies provided us with first correlations of histone modifications and translocations, we sought to directly test the role of histone modifications in chromosome breakage and translocation formation using a well-established cellular translocation system.

We used anaplastic large cell lymphoma (ALCL), which was previously used by the Misteli laboratory (Mathas et al., 2009). Most patients with ALCL have recurrent translocations between the 5' oligomerization domain of the nucleophosmin 1 (*NPM1*) gene on chromosome 5 (5q35) and the anaplastic lymphoma kinase (*ALK*) tyrosine kinase domain on chromosome 2 (2p23), leading to the NPM1-ALK fusion protein and constitutive activation of the ALK tyrosine kinase (Kinney et al., 2011). However, up to 40% of symptomatically indistinguishable patients with ALCL lacks t(2,5), indicating

that the translocation is not causal or required for the disease (Tabbo et al., 2013). In further support of *NPM1-ALK* being a bystander translocation event, NPM1-ALK protein expression is not sufficient to induce ALCL in animal models (Chiarle et al., 2008). Finally, both ALCL types share similar histopathological findings and molecular defects (Stein et al., 2000; Mathas et al., 2005; Janz et al., 2006).

Recently, the Misteli and Dorken laboratories showed that translocation-negative ALCL cells act as precursors to translocation-positive ALCL cells (Mathas et al., 2009). First, Mathas et al. found that several genes near the breakpoint regions on chromosomes 2 and 5 were upregulated in translocation-positive and translocation-negative ALCL cells, confirming that the two subtypes share similar molecular patterns despite the presence of *NPM1-ALK*. Second, they found that the *NPM1* and *ALK* genes in translocation-negative ALCL cells are in close spatial proximity in 3D nuclear space compared to non-ALCL T-cell lymphomas. Spatial proximity is a major contributing factor to translocation formation (Roukos et al., 2013; see Chapter 1). Finally, upon ionizing radiation, translocation-negative ALCL lines formed *NPM1-ALK* translocations at measurable frequency, while the non-ALCL cell lines did not (Mathas et al., 2009).

The characterization of translocation-negative ALCL cells as translocation precursor cells provides an opportunity to study well-defined translocation sites prior to translocation formation. In addition, the availability of translocation-negative ALCL, translocation-positive ALCL, and non-ALCL T-cell lymphoma lines allows for the modulation of histone modifications to study the formation of chromosome breaks and translocations.

Here, we first probe the histone modification landscape at *NPM1* and *ALK* breakpoints. After identification of specific modifications, we overexpress their corresponding histone-modifying enzymes and measure *NPM1* and *ALK* breakage and *NPM1-ALK* translocation frequency using a highly sensitive and specific method described extensively in Chapter 5.

RESULTS

Mapping histone modifications at ALCL translocation breakpoints

In order to map chromatin modifications at ALCL-specific translocation genes, *NPM1* and *ALK*, we decided to perform comparative ChIP studies coupled with quantitative real-time PCR (qPCR) between translocation-negative ALCL (i.e. *NPM1-ALK* translocation-prone) and non-ALCL control T-cell lymphoma (i.e. not *NPM1-ALK* translocation-prone) cell lines to determine if altered levels of histone modifications mark translocation-prone regions (Fig. 1).

The first step for these studies was to identify the translocation breakpoints to design primers for qPCR. The *NPM1* gene is ~25 kb, while the *ALK* gene is ~730 kb, so we decided to focus on the defined breakpoints within the genes, since each ChIP experiment yields a limited quantity of DNA for PCR reactions. From the literature, we learned that the recurrent breakpoints have been cloned and occur in 910-bp intron 4 for *NPM1* and 1,923-bp intron 19 for *ALK* (Ladanyi and Cavalchire, 1996; Sarris et al., 1997; Luthra et al., 1998). Since the breakpoint introns are relatively small, we developed primers within

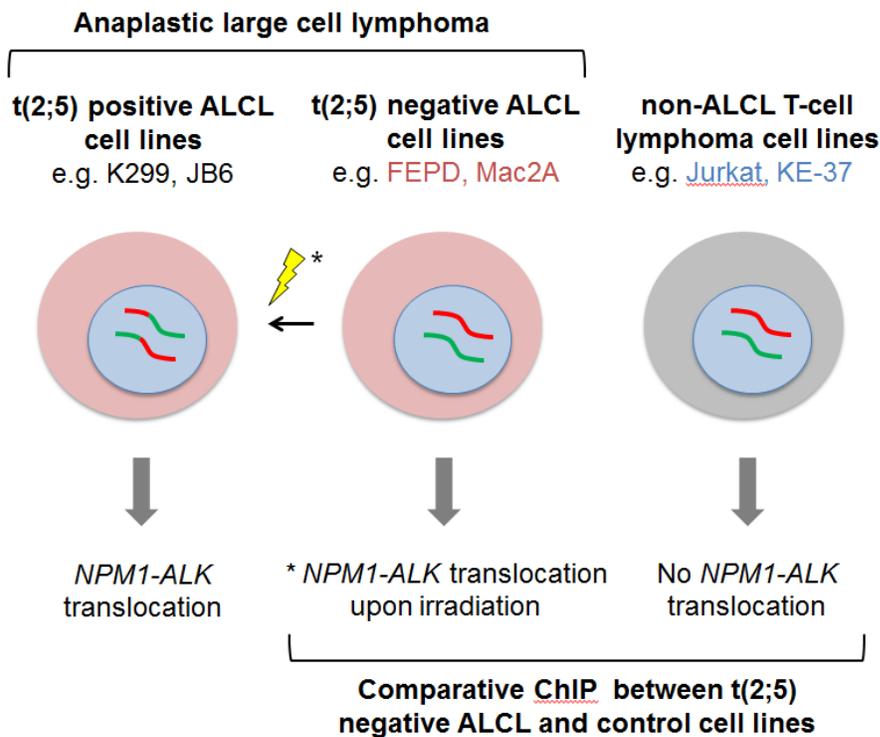


Figure 1: ALCL exists in translocation-positive [t(2,5)+ or *NPM1-ALK*] and translocation-negative [t(2,5)-] subtypes. Several lines of evidence support translocation-negative ALCL cells as precursors to t(2,5)+ including similar gene expression patterns, spatial proximity of breakpoint genes and the induction of specific *NPM1-ALK* translocations by irradiation of t(2,5)- ALCL cells (Mathas et al., 2009). Comparative ChIP was performed between translocation-negative ALCL precursors and non-ALCL T-cell lymphoma lines that do not form *NPM1-ALK* translocations upon irradiation (Mathas et al., 2009).

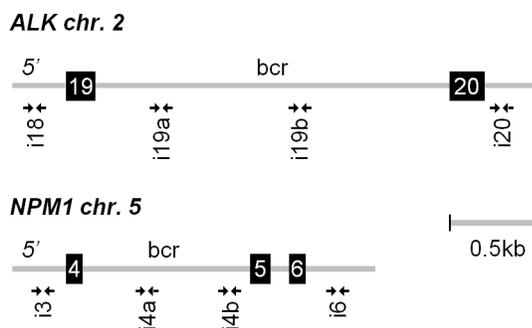


Figure 2: *ALK* and *NPM1* breakpoints (bcr) are localized in intron 19 and 4, respectively. Primer pairs used for ChIP studies are shown by double arrows.

the breakpoint and the neighboring introns (Fig. 2). While we were interested in studying the whole genes, we reasoned that if chromatin levels were altered at *NPM1* or *ALK*, we would observe these differences at the breakpoints of the genes. The *NPM1* and *ALK* genes were studied in comparison to other loci as well: intron 1 and the 3' untranslated region (3' UTR) of the constitutively active *Cyclophilin A* gene (*CycA*), transcription start site (TSS) of the constitutively inactive *Nanog* gene, TSS of the constitutively inactive *EVX1* gene, and pericentric heterochromatin-specific Sat-II repeat (SatII) on chromosome 1. These loci serve as positive and negative control regions to indicate whether or not a ChIP experiment was suitable for analysis.

Next, we selected cell lines for ChIP experiments. Initially, we performed experiments with one translocation-negative cell line (Mac2A) and one non-ALCL cell line (Jurkat), but we soon scaled up experiments with the addition of FE-PD and KE37 (Fig. 1). Results in two cell lines for each cell type would strengthen our claims. These cell lines were chosen because they had similar growth rates in culture and required the same media. Since histone modifications are associated with gene expression (Tessarz and Kouzarides, 2014), it was important for us to look at the expression levels of the translocation genes we would assay by ChIP. As previously reported, the *NPM1* gene was expressed at similar levels in all four cell lines, whereas the *ALK* gene transcript was undetectable in all lines by reverse transcriptase (RT)-qPCR for mRNA levels (Mathas et al., 2009; Fig. 3).

The CD34+ translocation gene analysis presented in Chapter 2 was limited by the available ChIP-Seq tracks for six histone modifications. While these are the most well-

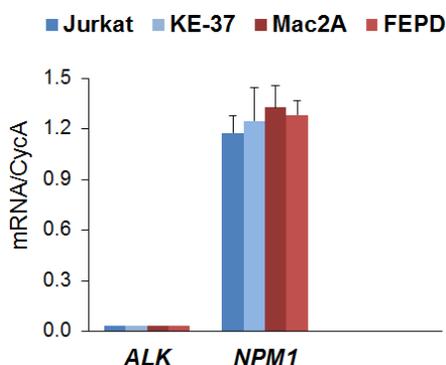


Figure 3: mRNA expression levels of translocation genes across t(2,5)-negative ALCL (red) and control T cell lines (blue).

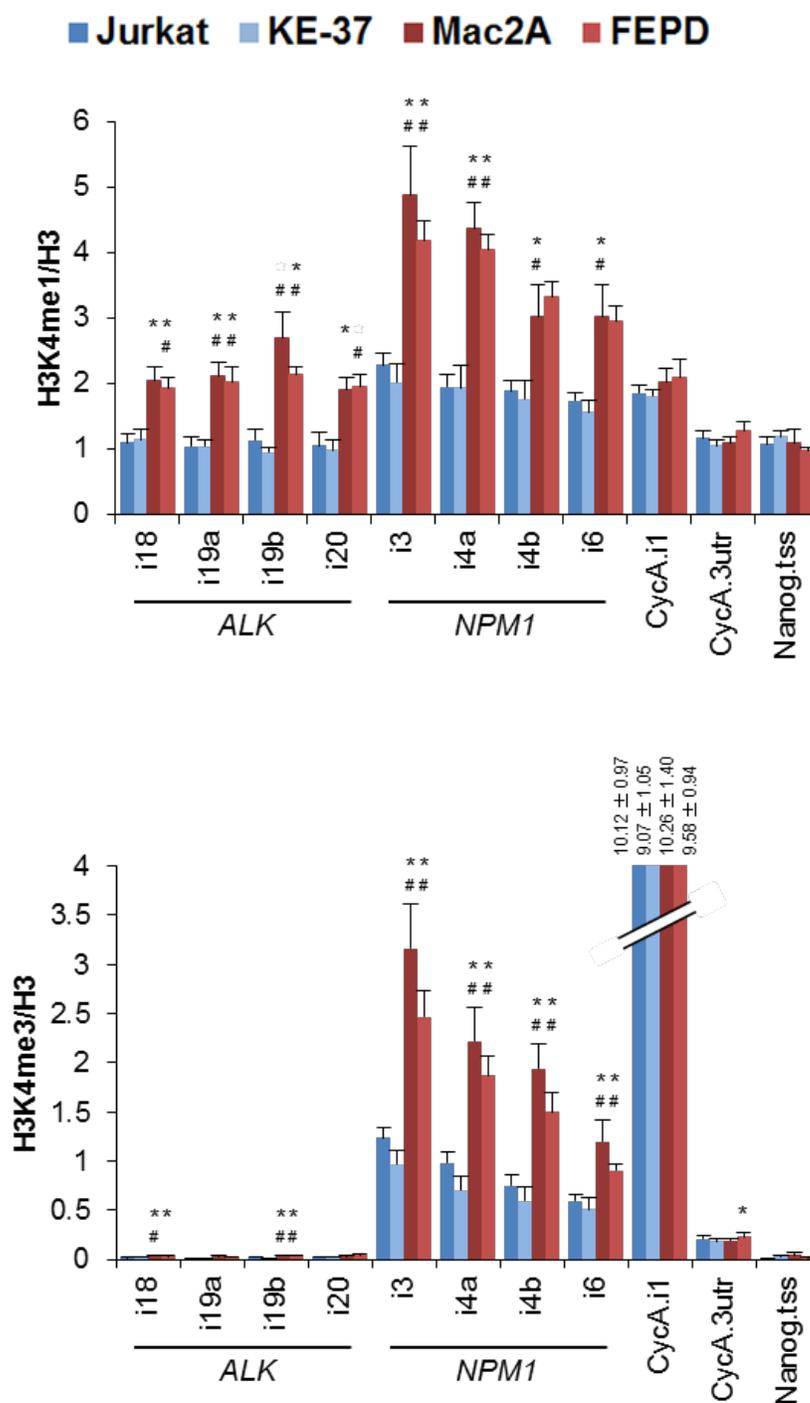
studied modifications, we wanted to expand our screening of chromatin modifications that could mark translocation-prone regions and possibly represent a unique signature of breakpoints. We performed ChIP studies on several histone modifications and histone variants; however, the antibodies for some modifications did not work, or there was no signal at any of the assayed genome region (Table 1; see Methods for antibodies).

Table 1: Histone modifications or variants studied in comparative ChIP experiments.

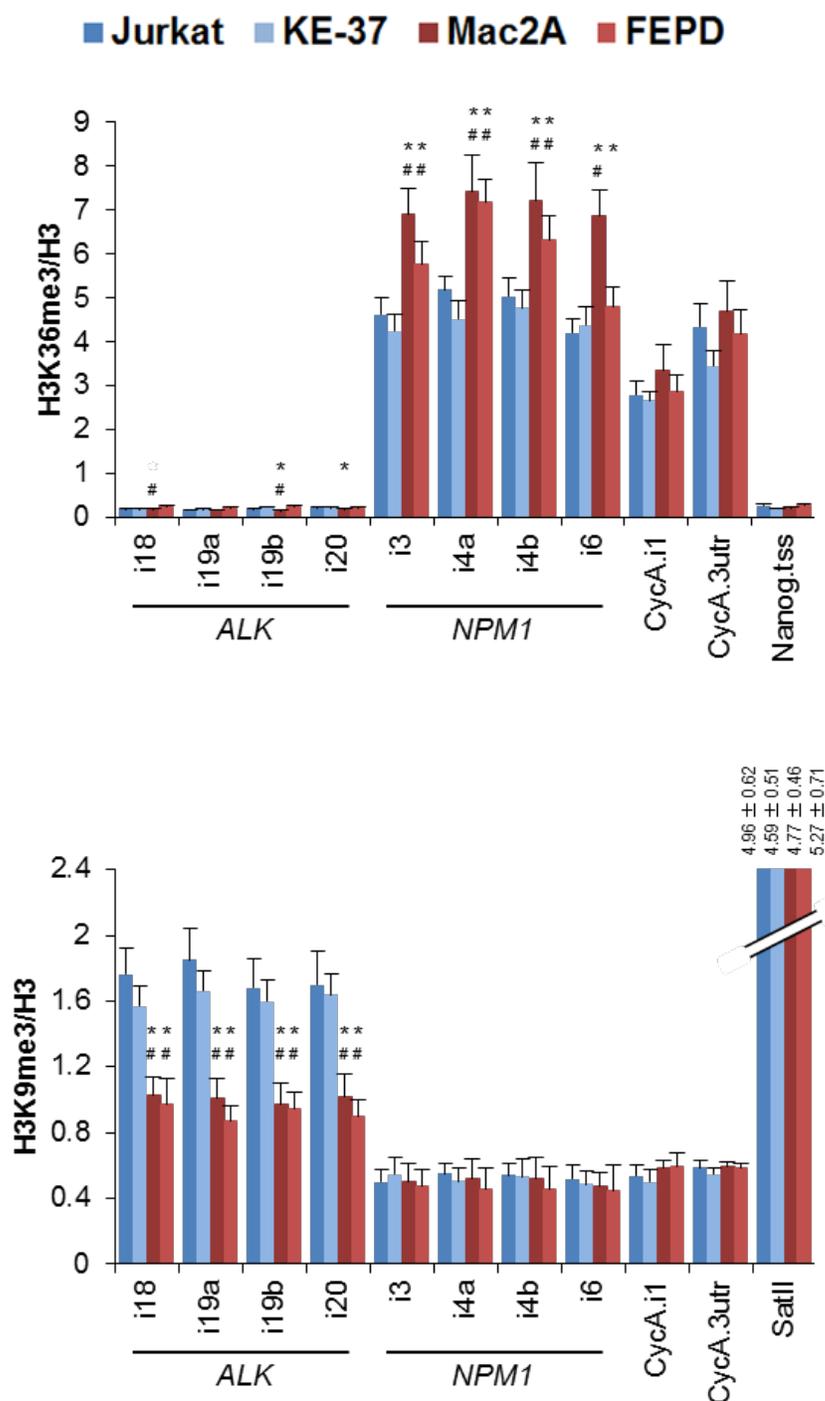
Histone modification or variant	Result
H3K4me1	Fig. X
H3K4me3	Fig. X
H3K36me3	Fig. X
H3K9me3	Fig. X
H3K27me3	Fig. X – no differences by ChIP
H3K27ac	Fig. X – no differences by ChIP
H3K79me2	Fig. X – no differences by ChIP
H3K9ac	(not shown) no differences by ChIP
H3K56ac	(not shown) no differences by ChIP
H3K9me1	(not shown) no differences by ChIP
H4K16ac	(not shown) no differences by ChIP
H4K20me1	(not shown) no differences by ChIP
H2A.Z	(not shown) no signals
H2A.X	(not shown) no signals
CENP-A	(not shown) no signals

Several histone modifications were found at altered levels over the *NPM1* and *ALK* translocation breakpoints in translocation-negative ALCL cells compared to control cells. At the *NPM1* locus, H3K4me1, H3K4me3, and H3K36me3 levels were on average two-fold higher across the breakpoint region compared to control cells (Figs. 4-6; $P < 0.05$). Importantly, as observed in the CD34+ translocation gene analysis, histone modification enrichments were not related to expression level as *NPM1* was similarly expressed across all four lines as assessed by real-time PCR (Fig. 3). The *ALK* locus was marked by an average 1.5-fold enrichment of H3K4me1 across the breakpoint region in translocation-negative ALCL lines compared to control cell lines (Fig. 4; $P < 0.05$). In addition, the *ALK* gene was also marked by an average 1.5-fold reduction in H3K9me3 in translocation-negative ALCL lines compared to control T-cell lines (Fig. 7), again independently of transcription status. Importantly, these alterations did not represent a global alteration of H3 methylation marks in ALCL cells, since the levels of these modifications were similar at *CycA*, *Nanog*, and *SatII* in all cell lines (Fig. 4-7).

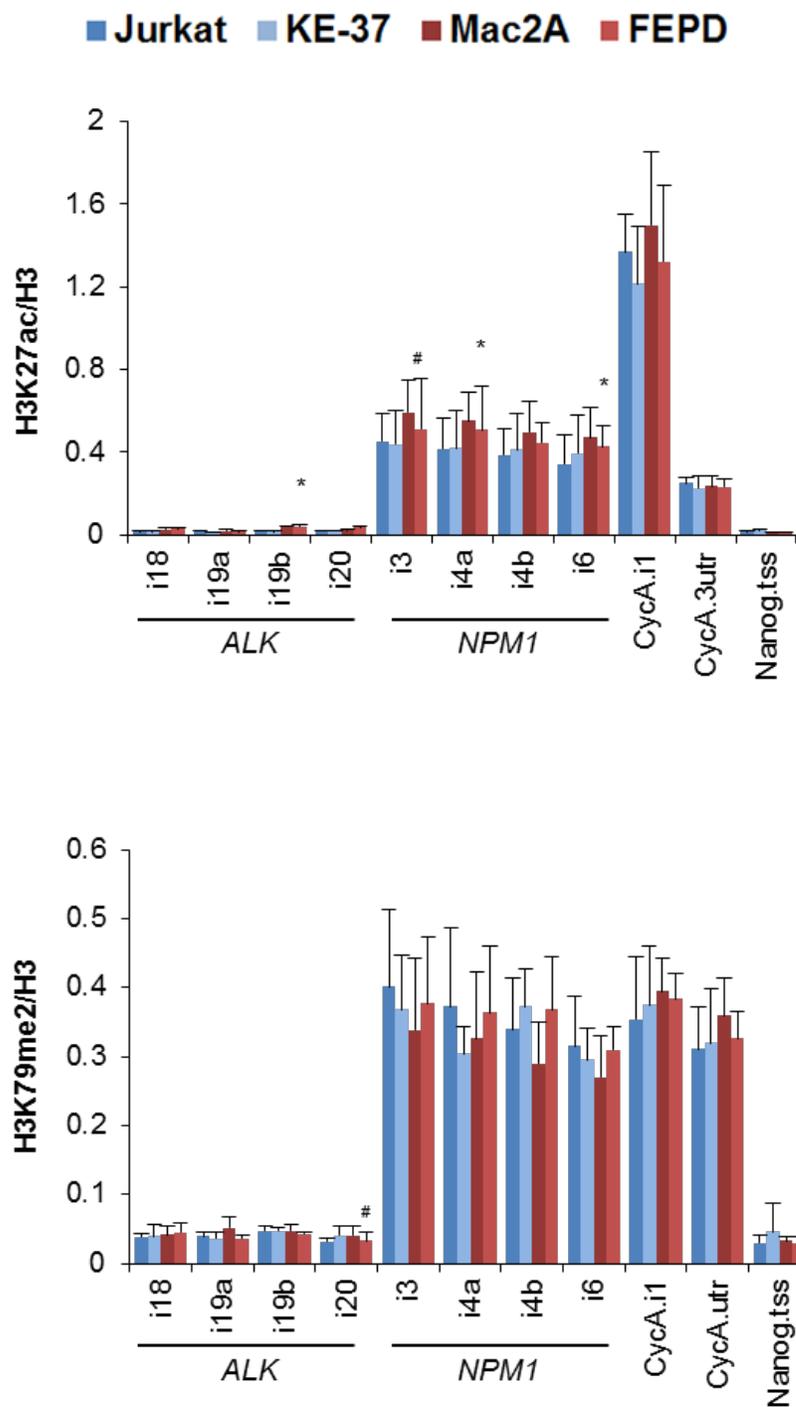
Several chromatin features including active (H3K9Ac, H3K27Ac, H3K79me2 and H4K16Ac) and inactive (H3K9me1 and H3K27me3) histone marks, as well as those found to mark both active and inactive genes (H3K56ac and H4K20me1 were unchanged between cell lines (examples Figs 8-10; rest not shown). These data further support the notion of enrichment of specific histone modifications at translocation prone genome regions.



Figures 4, 5: Mapping of (4) H3K4me1 and (5) H3K4me3 at breakpoint and control regions in t(2,5)-negative ALCL (red) and control T cell lines (blue) by ChIP-qPCR. The percentage of input was normalized to unmodified H3. Values represent means \pm SEM from three to four independent experiments. * $p < 0.05$ to Jurkat cells, # $p < 0.05$ to KE37 cells, Student's t test.



Figures 6, 7: Mapping of (6) H3K36me3 and (7) H3K9me3 at breakpoint and control regions in t(2,5)-negative ALCL (red) and control T cell lines (blue) by ChIP-qPCR. The percentage of input was normalized to unmodified H3. Values represent means \pm SEM from three to four independent experiments. * $p < 0.05$ to Jurkat cells, # $p < 0.05$ to KE37 cells, Student's t test.



Figures 8, 9: Mapping of (8) H3K27ac and (9) H379me2 at breakpoint and control regions in t(2,5)-negative ALCL (red) and control T cell lines (blue) by ChIP-qPCR. The percentage of input was normalized to unmodified H3. Values represent means \pm SEM from three to four independent experiments. * $p < 0.05$ to Jurkat cells, # $p < 0.05$ to KE37 cells, Student's t test.

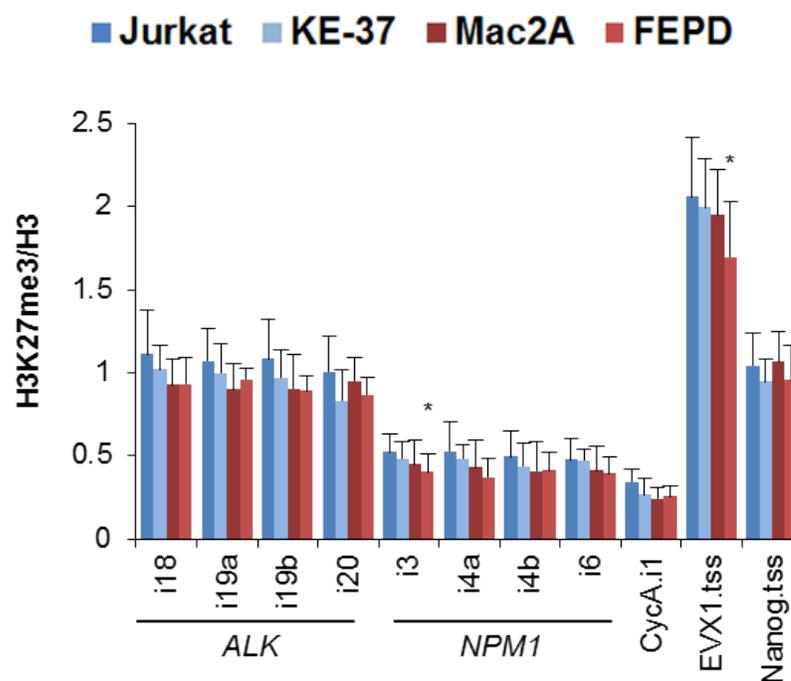


Figure 10: Mapping of H3K27me3 at breakpoint and control regions in t(2,5)-negative ALCL (red) and control T cell lines (blue) by ChIP-qPCR. The percentage of input was normalized to unmodified H3. Values represent means \pm SEM from three to four independent experiments. * $p < 0.05$ to Jurkat cells, # $p < 0.05$ to KE37 cells, Student's t test.

A further difference between translocation-negative cells and non-ALCL cells was the consistently lower level of nucleosome density across *NPM1* and *ALK* regions we repeatedly observed by the reduced ratio of immunoprecipitated H3 DNA to total DNA input (Fig. S3A; control immunoprecipitation with no antibody). In ALCL cell lines, the relative density of nucleosomes ranged from ~ 0.4 to 0.75 at *ALK* and ~ 0.65 to 0.95 at *NPM1* versus control cells despite comparable amounts of total DNA input (Fig. 1). In contrast, the recovery of immunoprecipitated H3 DNA was similar across all four cell lines at *CycA*, *Nanog*, and *SatIII* (Fig. 11).

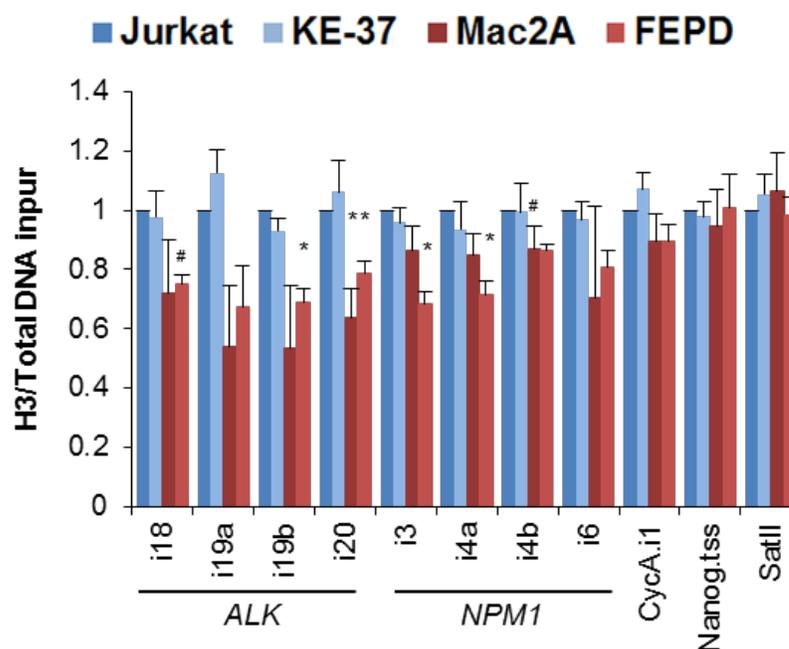


Figure 11: H3 density was calculated from ChIP-qPCR of unmodified H3 relative to total DNA input (no antibody) at breakpoint and control regions. Data is from six independent experiments. Jurkat cell line (dark blue) was set at 1.0, to which KE37 (light blue), MAC2A (dark red), and FE-PD (light red) were compared. Values represent means \pm SEM from a minimum of three independent experiments. * $p < 0.05$ to Jurkat cells, # $p < 0.05$ to KE37 cells, Student's t test.

Modulating histone modifications at ALCL translocation breakpoints - screening

Next, we set out to directly address whether modulation of breakpoint-associated histone modifications could affect *NPM1* and *ALK* breakage and *NPM1-ALK* translocation frequency. These studies would elucidate a possible role of histone modifications in predisposing genomic sites to chromosome breakage and translocation formation. Our goal was to overexpress histone-modifying enzymes in translocation-negative ALCL cells and non-ALCL T-cell lymphomas, irradiate cells to induce DNA damage, and measure breakage and translocation frequency (Fig. 12).

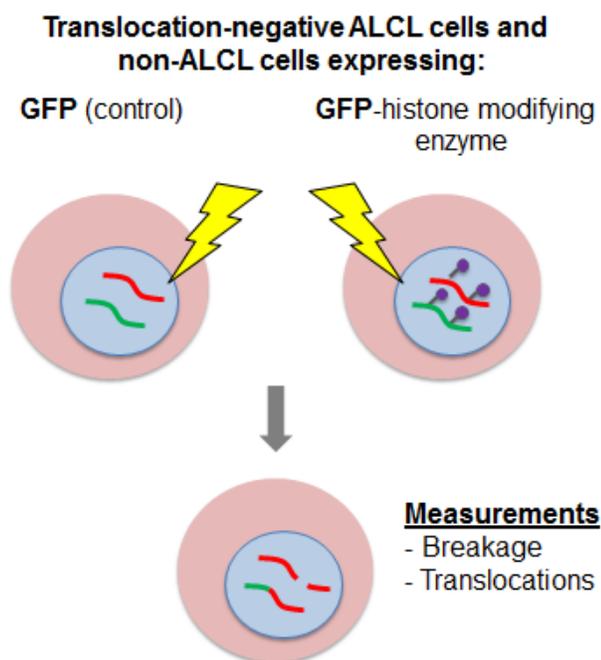


Figure 12: Schematic of experiment to test the effect of histone modifications on chromosome breakage and translocation formation in translocation-negative ALCL and non-ALCL cell lines. Briefly, stable cell lines expressing GFP-tagged histone modifying enzymes or GFP empty vector (control) were irradiated, and breaks and translocations were detected by hiBA-FISH (described in Chapter 5).

It was previously shown that irradiation of translocation-negative ALCL cells leads to a low, but measurable, frequency of *NPM1-ALK* translocations as determined by FISH on metaphase spreads (Mathas et al., 2009). A major limitation to using metaphase spreads is that they are generated at low efficiency, making it impractical to test many biological samples and measure statistically significant differences. Therefore, we first developed a method to quantitatively detect chromosome breaks and translocations with high sensitivity in interphase cells. This method, **high-throughput break-apart fluorescence in situ hybridization**, or hiBA-FISH, is described extensively in Chapter 5. Briefly, hiBA-FISH combines high-throughput imaging with the measurement of the spatial separation

of FISH probes flanking target genome regions of interest in interphase cells. We developed a two-color *ALK* break-apart probe consisting of a Green (Alexa488) probe targeting the 5' region upstream of the breakpoint and a Red (Alexa568) probe targeting the 3' region downstream of the breakpoint (Fig. 13). Used alone, this break-apart probe reports on intact *ALK* alleles by co-localization of Red and Green signals (Fig. 14I), and breakage by separation of the signals (Fig. 14B). When the break-apart probe is combined with a Far Red (Cy5) probe targeting the 5' region upstream of the *NPM1* breakpoint, *NPM1-ALK* translocations can be detected by co-localization of a separated Red signal and Far Red signal (Fig. 14T). It should be noted that this method was designed based on the FDA-approved *ALK* break-apart probe that is used by clinicians to diagnose translocations containing *ALK* in lung cancer.

In addition to developing a breakage and translocation detection method, a major optimization hurdle was the overexpression of plasmids containing histone-modifying enzymes in these lymphocytic cell lines. We attempted several transient transfection methods including nucleofection, lipofection, and non-liposomal transfection using a GFP construct. All of these methods resulted in low cell viability and low efficiency of transfection. Next, we attempted viral transduction methods. Using a lentiviral GFP-Lamin vector as a trial, we successfully transduced a majority of lymphocyte cells with no apparent effect on cell growth and survival using an adapted protocol adapted for retroviral transduction lymphocyte cell lines (Swift et al., 2001).

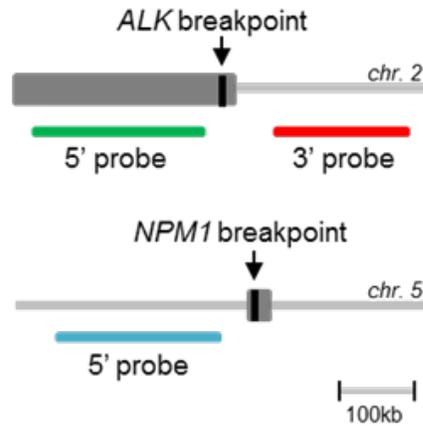


Figure 13: Schematic representation of the hiBA-ALK probe set. Green probe: 5' *ALK* region, Red probe: 3' *ALK* region, Far Red (represented as Cyan) probe: 5' *NPM1* region.

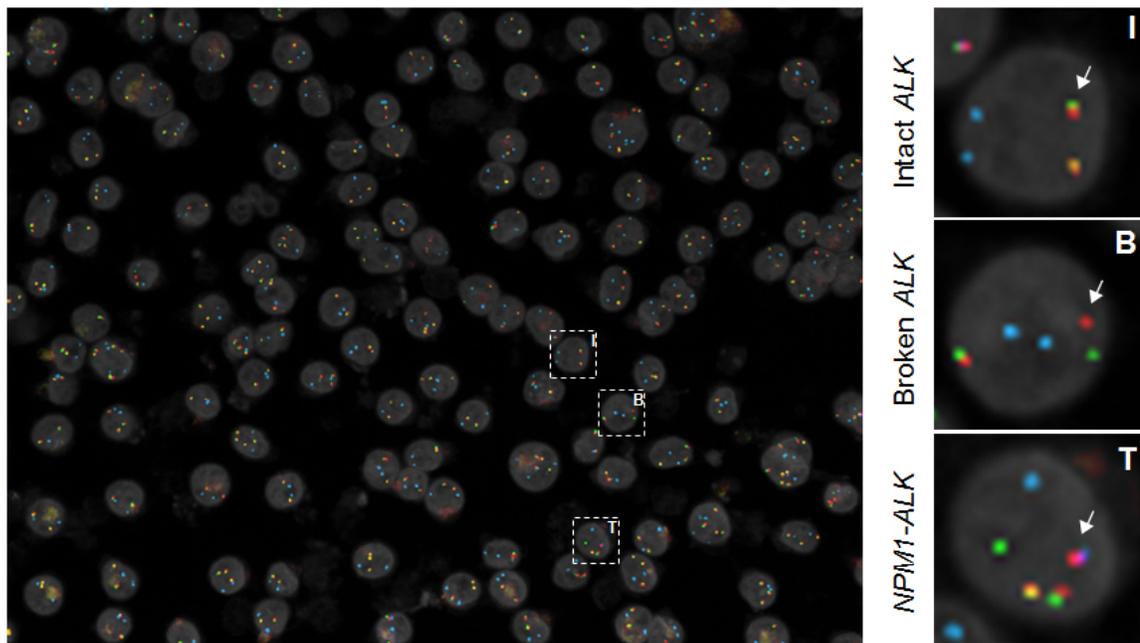


Figure 14: Example of merged maximum intensity projection image (7 Z-planes, 4 channels per plane) of irradiated FEPD cells stained with the *ALK* break-apart probe set using hiBA-FISH. Green: 5' *ALK*, Red: 3' *ALK*, Cyan: 5' *NPM1*, Gray: DAPI set to 30% transparency. The three dashed boxes indicate representative cells (magnified in the right panels) with **(I)** an intact *ALK* allele, **(B)** a broken *ALK* allele, or **(T)** an *NPM1-ALK* translocation, respectively (arrows). Scale bar: 10 μ m.

The success of this pilot experiment led us to clone histone-modifying enzymes in a lentiviral construct along with GFP. We focused on enzymes that modulated histone modifications H3K4me1, H3K4me3, H3K36me3, and H3K9me3, since altered levels of these modifications marked ALCL breakpoints (as well as several genes in the CD34+ analysis) and we had already used methyltransferases for these modifications in tethering experiments (Chapter 3). In addition, we cloned the corresponding demethylases (Table 2). Since translocation-negative ALCL cell lines are marked by enrichment of specific methylation marks, except H3K9me3, we hypothesized that demethylation would decrease translocation frequency in these cell lines, and vice versa for H3K9me3.

Table 2: Lentiviral (lenti) GFP-tagged histone-modifying enzyme constructs used to make stable cell lines.

Histone modification	Methyltransferase constructs	Demethylase constructs
H3K4me1	Lenti-GFP-SET7/9	Lenti-GFP-LSD1
H3K4me3	Lenti-GFP-ASH2L	Lenti-GFP-RBBP2
H3K9me3	Lenti-GFP-SUV3-9	Lenti-GFP-JMJD2B
H3K36me3	Lenti-GFP-ySET2* *yeast version of human SETD2	Lenti-GFP-KDM2A

Next, we transduced translocation-negative ALCL lines Mac2A and FE-PD and non-ALCL cell lines Jurkat and KE-37 with these constructs. A GFP-tagged empty vector was used as a negative control. The GFP tag in each construct allowed us to confirm transduction by fluorescence microscopy, and this led to the realization that our cell lines should be sorted by GFP fluorescence intensity to ensure comparable levels of expression across cell lines. This led to the development of stable cell lines which we also selected with antibiotics for 7 days.

In addition to transfection method optimization, several rounds of optimization were required for irradiation conditions, time course, plating for hiBA-FISH, and hiBA-FISH itself. We used the same irradiation conditions and time course as previously described (Mathas et al., 2009) and confirmed that 25 Gy of irradiation leads to on average 75% cell viability by trypan blue staining after 24 hours. For plating, we experienced considerable difficulty in plating cell lines in 96-well and 384-well plates designed for high-throughput imaging. This was primarily due to the fact that lymphocytes do not adhere well to plates, even in the presence of poly-D or poly-L lysine coatings, and the extensive washing required for FISH eradicated the cells. We soon adapted our high-throughput microscope to image coverslips (see Chapter 5) and decided that cells would be grown in suspension with or without exposure to irradiation and then spun onto poly-D-lysine coverslips prior to fixation and FISH.

We first performed a screening of 36 stable cell lines (Mac2A, FE-PD, KE-37, and Jurkat parental cell lines each expressing one of 4 methyltransferases, 4 demethylases, or GFP alone). As described in Chapter 5, cell lines were first analyzed without irradiation to determine background breakage and translocation frequency. Since hiBA-FISH is based on the distances between FISH probes, we were also able to determine whether or not histone-modifying enzymes affect spatial proximity of *NPM1* and *ALK* translocation genes. No significant differences were seen between histone-modifying enzymes (data not shown); however, in accordance with previous data, *NPM1* and *ALK* were more proximal in translocation-negative ALCL lines than non-ALCL cell lines (Mathas et al., 2009).

After irradiation, the frequency of *ALK* breakage and *NPM1-ALK* translocations increased in all cell lines (example Fig. 15-16: FE-PD cell lines; rest not shown). The results from this screen demonstrated that most histone-modifying enzymes do not considerably alter breakage or translocation frequency compared to the corresponding GFP control cell line, with some exceptions (examples Fig. 15-16: FE-PD cell lines; rest not shown). The most notable exceptions were increased *ALK* breakage and *NPM1-ALK* translocations in FE-PD, Mac2A, and Jurkat cell lines overexpressing the methyltransferases *ASH2L* and *SET7/9*, compared to the GFP control lines (examples Fig. 15-16: FE-PD cell lines; rest not shown). Interestingly, the corresponding demethylases, *RBBP2* and *LSD1* did not result in an inverse reduction in breakage or translocations compared to the GFP control lines (examples Fig. 15-16: FE-PD cell lines; rest not shown).

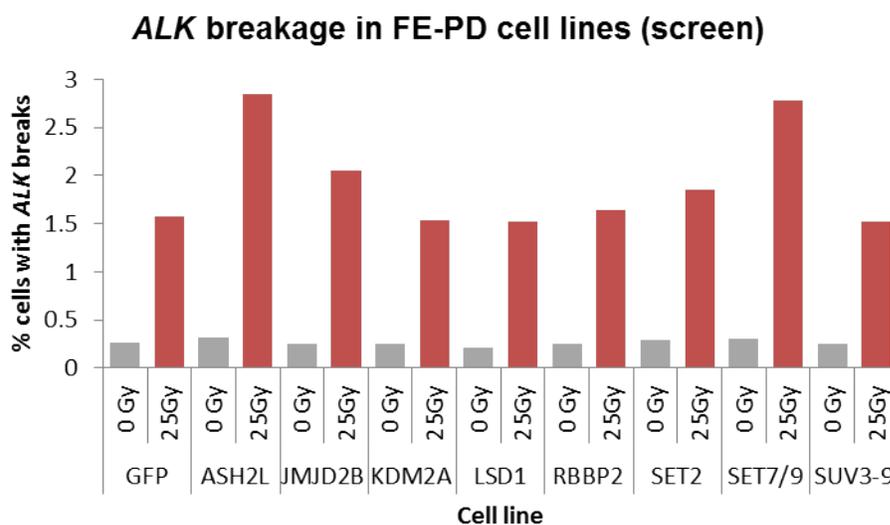


Figure 15: *ALK* breakage screen of FE-PD cell lines overexpressing histone modifying enzymes listed in Table 2, using hiBA-FISH. Values represent percentages of cells with at least one *ALK* breakage event in non-irradiated (gray) and irradiated (red) conditions from one experiment.

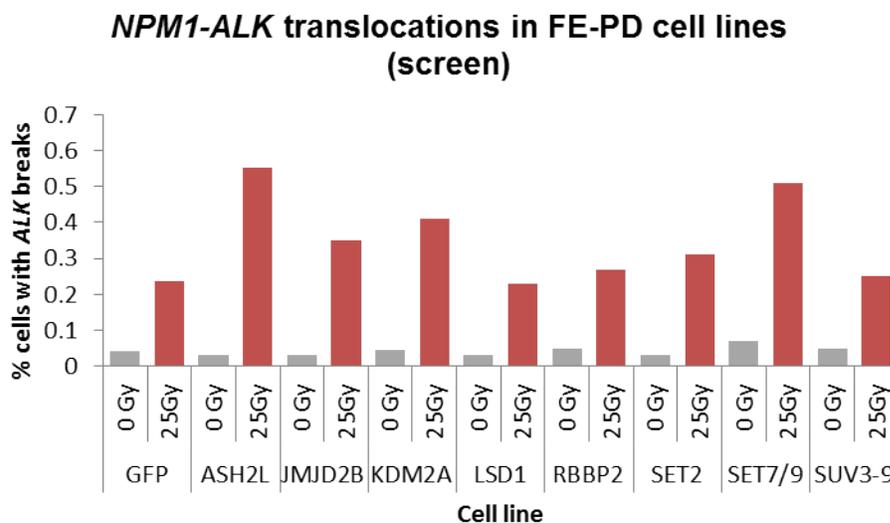
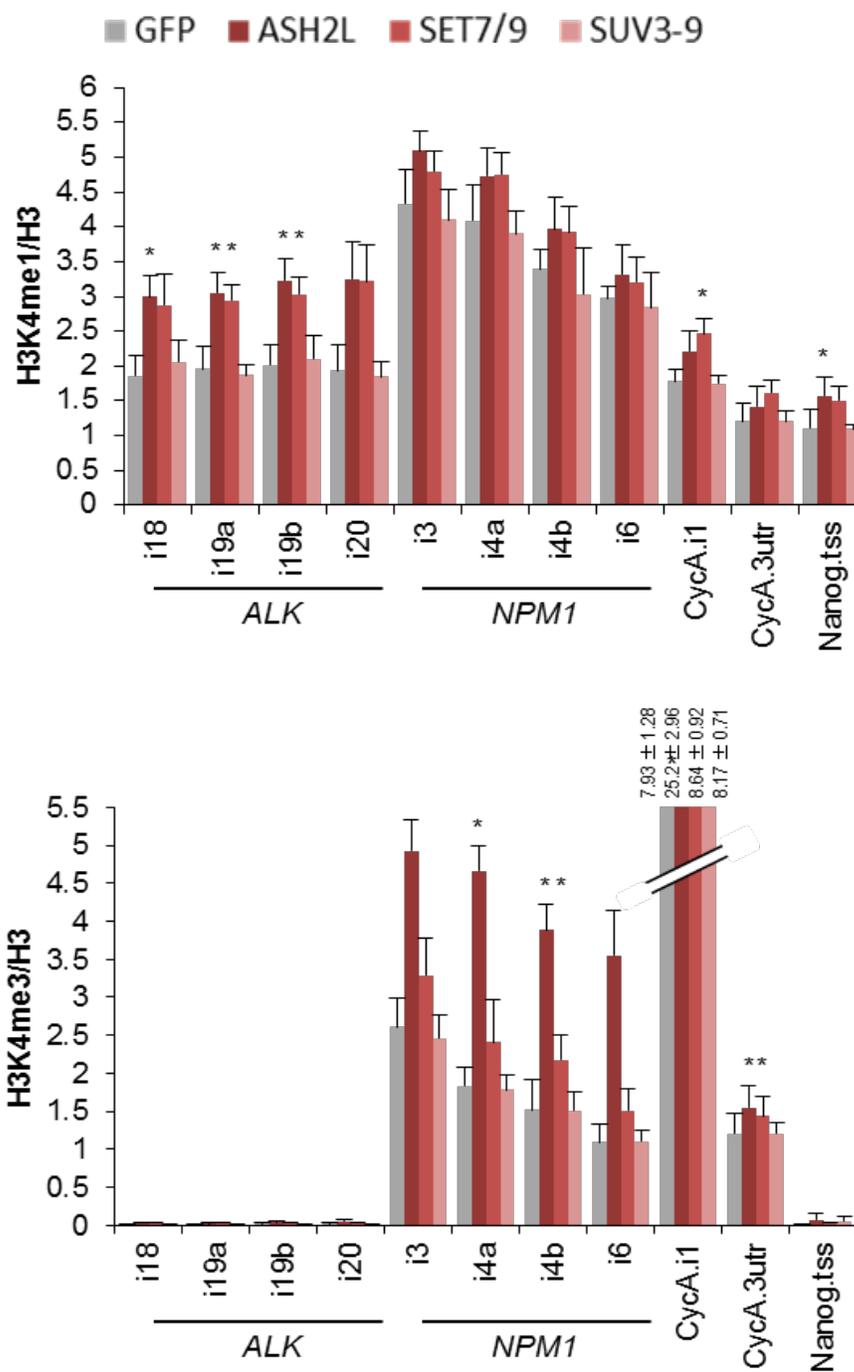


Figure 16: *NPM1-ALK* translocation screen of FE-PD cell lines overexpressing histone modifying enzymes listed in Table 2, using hiBA-FISH. Values represent percentages of cells with at least one *NPM1-ALK* translocation event in non-irradiated (gray) and irradiated (red) conditions from one experiment.

Modulating histone modifications at ALCL translocation breakpoints— H3K4 methylation

We narrowed our focus to the H3K4 methyltransferases ASH2L and SET7/9 in the translocation-negative cell line FE-PD and non-ALCL cell line Jurkat. As a control for a histone-modifying enzyme that did not significantly alter breakage and translocation frequency in the screen, we also continued experiments with FE-PD and Jurkat cells overexpressing SUV3-9. We first confirmed the modulation of histone modifications at *NPM1* and *ALK* breakpoints in FE-PD cell lines overexpressing these constructs by ChIP-qPCR (Figs. 17-19).



Figures 17-18: Mapping of (17) H3K4me1 and (18) H3K4me3 at breakpoint and control regions in FEPD derived cell lines stably expressing GFP alone (gray) or GFP-fusions to ASH2L (dark red), SET7/9 (light red), or SUV3-9H1 (pink) by ChIP-qPCR. The percentage of input was normalized to unmodified H3. Values represent means \pm SEM from three independent experiments. * $p < 0.05$ to GFP cells, Student's t test.

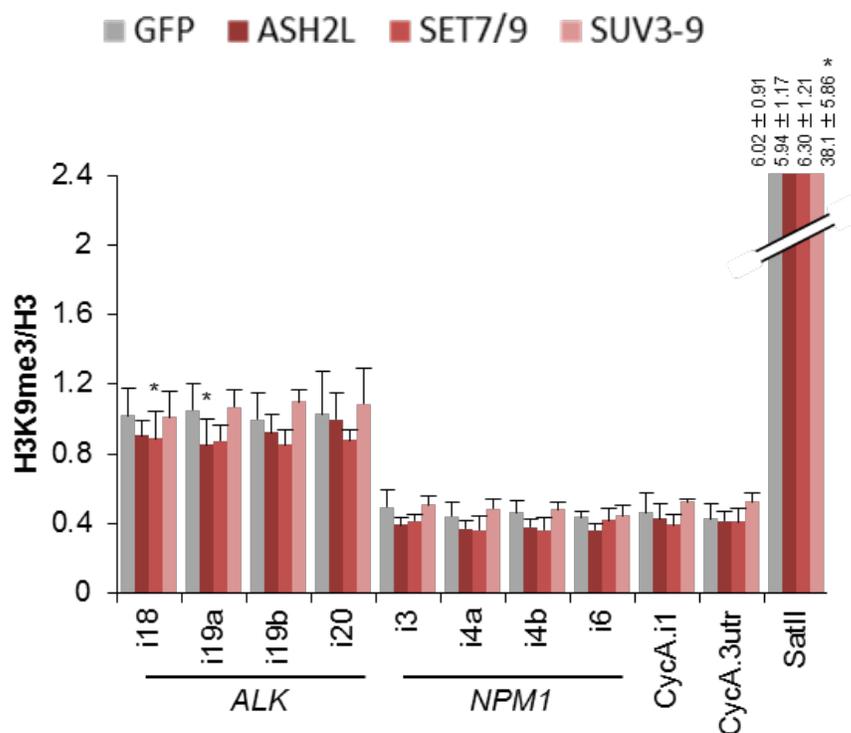


Figure 19: Mapping of H3K9me3 at breakpoint and control regions in FEPD derived cell lines stably expressing GFP alone (gray) or GFP-fusions to ASH2L (dark red), SET7/9 (light red), or SUV3-9H1 (pink) by ChIP-qPCR. The percentage of input was normalized to unmodified H3. Values represent means \pm SEM from three independent experiments. * $p < 0.05$ to GFP cells, Student's t test.

When challenged by irradiation, FEPD cells expressing GFP alone as a control developed *ALK* breaks in $1.53 \pm 0.19\%$ of cells (Fig. 20). In contrast, the breakage frequency was elevated ~ 2 -fold in the presence of ASH2L ($3.15 \pm 0.05\%$, $p < 0.01$) or SET7/9 ($3.02 \pm 0.49\%$, $p < 0.01$) versus GFP cells (Fig. 20). Cells expressing SUV3-9 showed no significant increase in breakage frequency ($1.64 \pm 0.22\%$) (Fig. 20). These frequencies were several-fold higher than the background frequency of breakage in non-irradiated cells ($0.29 \pm 0.03\%$ across cell lines) (Fig. 20). Along the same lines, expression of either H3K4 methyltransferase caused an increase in DNA breaks at *ALK* in Jurkat cell

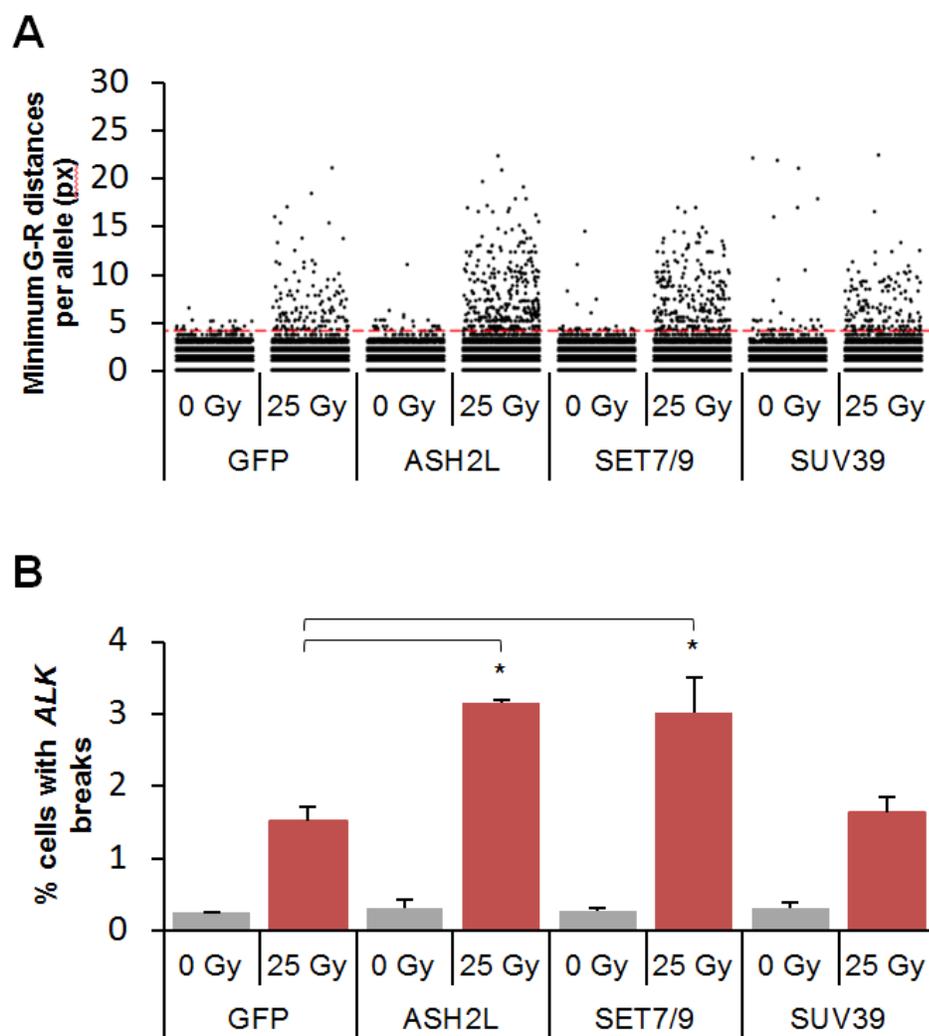


Figure 20: Determination of *ALK* breakage frequencies in FE-PD-derived cell lines by hiBA-FISH. **(A)** Jitter plot of minimum hiBA-FISH Green-Red (5'*ALK*-3'*ALK*) distances in FE-PD-derived cell lines stably expressing the indicated GFP fusion proteins. An *ALK* breakage event was defined as a minimum Green-Red distance > 4 pixels (dashed red line). For each experimental condition, at least 20,000 minimum distances are represented. **(B)** Percentages of cells with at least one *ALK* breakage event per cell. Values represent means \pm SD from three independent experiments. * $p < 0.05$, One-way ANOVA followed by Tukey's post-test for multiple comparisons between all irradiated samples.

lines, but to a lesser degree than in FE-PD cells (Fig. 21). This in line with the lower levels of H3K4 methylation at breakpoints in Jurkat cells (Figs. 4-5).

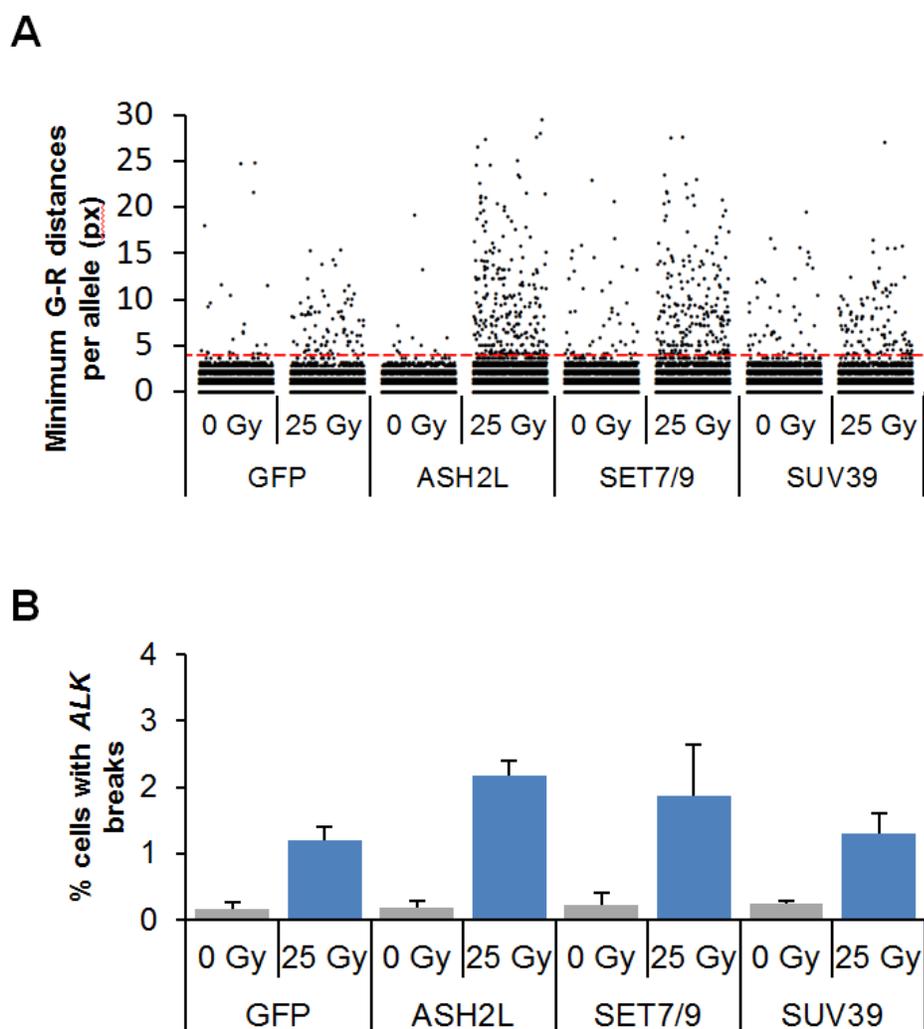


Figure 21: Determination of *ALK* breakage frequencies in Jurkat-derived cell lines by hiBA-FISH. **(A)** Jitter plot of minimum hiBA-FISH Green-Red (5' *ALK*-3' *ALK*) distances in Jurkat-derived cell lines stably expressing the indicated GFP fusion proteins. An *ALK* breakage event was defined as a minimum Green-Red distance > 4 pixels (dashed red line). For each experimental condition, at least 20,000 minimum distances are represented. **(B)** Percentages of cells with at least one *ALK* breakage event per cell. Values represent means \pm SD from three independent experiments. * $p < 0.05$, One-way ANOVA followed by Tukey's post-test for multiple comparisons between all irradiated samples.

The frequency of *NPM1-ALK* translocations also increased \sim 2-fold in FEPD cells expressing either ASH2L ($0.53 \pm 0.09\%$ cells with translocations, $p < 0.01$) or SET7/9

($0.51 \pm 0.03\%$, $p < 0.01$) compared to cells expressing GFP ($0.28 \pm 0.03\%$) alone or SUV3-9 ($0.25 \pm 0.02\%$) (Fig. 22). These frequencies were several-fold higher than the background frequency of translocations in non-irradiated cells ($0.03 \pm 0.008\%$ across cell lines) (Fig. 22). The increased translocation frequency in ASH2L- and SET7/9-expressing cells is consistent with the increased number of DNA breaks seen in these cells (Fig. 20). *NPM1-ALK* translocations increased only slightly in Jurkat cells expressing either H3K4 methyltransferase (data not shown). In all conditions, Jurkat cells had less *NPM1-ALK* translocations than FE-PD cells in accordance with previous data (Mathas et al., 2009).

To test whether the effect of ASH2L and SET7/9 on breakage in FE-PD cells was specific to the *ALK* gene or reflected genome-wide changes, we developed a break-apart probe for the *Nanog* locus as a control (Fig. 23). Like *ALK*, *Nanog* is not expressed in translocation-negative ALCL cells, but has not been reported to translocate in ALCL. In addition, we used this locus as a control region for all of our ChIP experiments. After irradiation, ASH2L and SET7/9 increased *Nanog* breakage moderately over background levels (Fig. 24). Importantly, the extent of breakage correlated with H3K4 methylation levels, which were considerably higher at *ALK* than at *Nanog* (Figs. 17-18). As a control for the *NPM1-ALK* translocation event, we analyzed the level of translocations between the normally non-translocating 3' *Nanog* and 5' *CycA* (Fig. 25). Translocations between these two loci was significantly less than *NPM1-ALK* translocations for FE-PD cells expressing ASH2L ($0.25 \pm 0.02\%$, $p < 0.01$) or SET7/9 ($0.26 \pm 0.03\%$, $p < 0.01$) (Fig. 25).

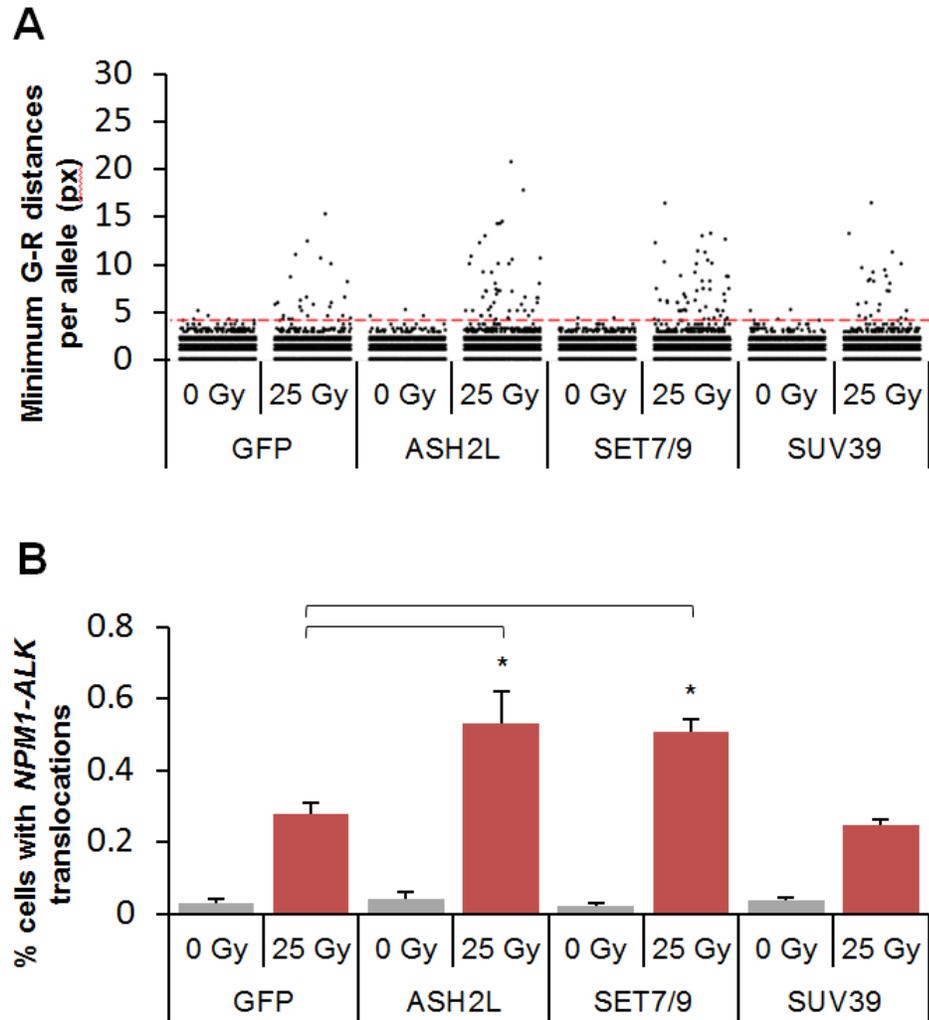


Figure 22: Determination of *NPM1-ALK* translocation frequencies in FE-PD-derived cell lines by hiBA-FISH. **(A)** Jitter plot of minimum hiFISH Green-Red (5'*ALK*-3'*ALK*) distances in the subset of cells from Fig. 21 that also contain proximal events between 5'*NPM1* and 3'*ALK*. An *NPM1-ALK* translocation event was defined on a per Red allele basis as a minimum Red-Far Red distance ≤ 4 pixels and a minimum Green-Red distance > 4 pixels (dashed red line). For each experimental condition, at least 3,000 minimum distances are represented. **(B)** Percentages of cells with at least one *NPM1-ALK* translocation event per cell. Values represent means \pm SD from three independent experiments. * $p < 0.05$, One-way ANOVA followed by Tukey's post-test for multiple comparisons between all irradiated samples.

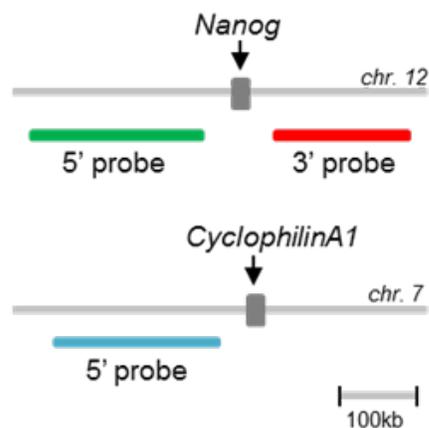


Figure 23: Schematic representation of the hiBA-Nanog probe set. Green probe: 5' *Nanog* region, Red probe: 3' *Nanog* region, Far Red (represented as Cyan) probe: 5' *CycA* region.

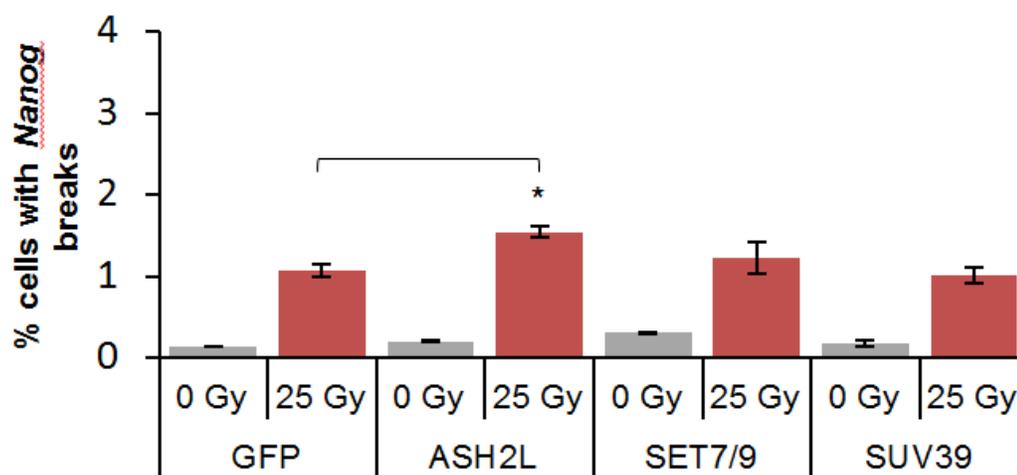


Figure 24: Determination of *Nanog* breakage frequencies in FE-PD-derived cell lines by hiBA-FISH. Percentages of cells with at least one *Nanog* breakage event per cell are shown. Values represent means \pm SD from three independent experiments. * $p < 0.05$, One-way ANOVA followed by Tukey's post-test for multiple comparisons between all irradiated samples.

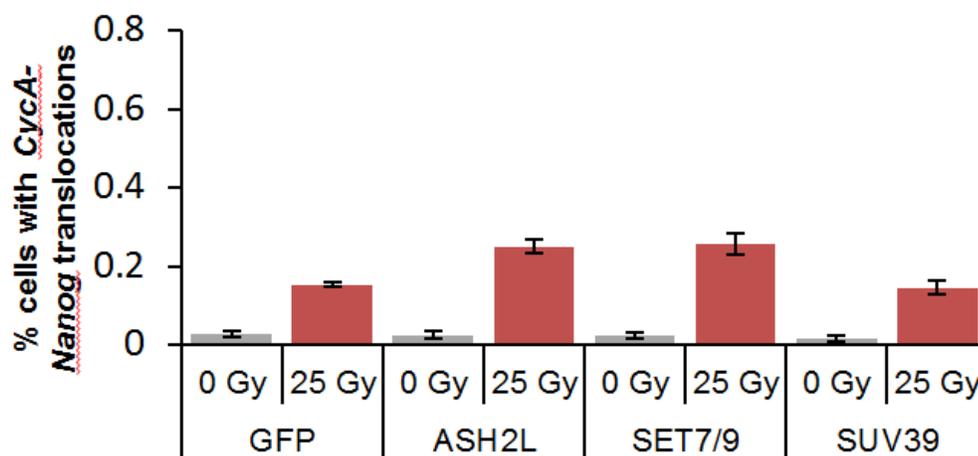


Figure 25: Determination of *CycA-Nanog* translocation frequencies in FE-PD-derived cell lines by hiBA-FISH. Percentages of cells with at least one *CycA-Nanog* translocation event per cell are shown. Values represent means \pm SD from three independent experiments. * $p < 0.05$, One-way ANOVA followed by Tukey's post-test for multiple comparisons between all irradiated samples.

MATERIALS AND METHODS

Cell culture

Translocation-negative ALCL (FEPD and Mac2A) and non-ALCL T cell lymphoma derived (Jurkat and KE-37) cell lines were kindly provided by Dr. S. Mathas (Charite-Berlin) and maintained in RPMI-1640 at 37°C and 5% CO₂. The 293FT cell line for lentivirus production (Life Technologies) was maintained in DMEM at 37°C and 5% CO₂. All culture medium was supplemented with 10% FBS (Atlanta Biologicals), 2 mM L-glutamine, 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin.

RNA Isolation and RT-PCR

Total RNA was isolated using the RNeasy Mini Kit (Qiagen). One μg RNA was transcribed into cDNA using Multiscribe reverse transcriptase, random primers and dNTPs (Applied Biosystems) in a final volume of 20 μl according to the manufacturer's instructions. RT-qPCRs were carried out on the selected genes in triplicates with SYBR-Green (iQ Supermix Bio Rad) on a Bio Rad C1000 ThermoCycler. The relative expression was determined using a standard curve for each gene and expression levels are shown as ratios to cyclophilin A. Primers used in this study are listed below:

hALK.mRNA	F: TGCCGCGGAAAAACATCAC	Tm: 60
	R: GCAGCGTCTTCACAGCCACTT	
hNPM1.mRNA	F: CGCCACCCGATGGAAGATTC	Tm: 60
	R: ACCAGCCCCTAAACTGACCGTTC	
hCycA.mRNA	F: GTCAACCCACCGTGTCTT	Tm: 60
	R: CTGCTGTCTTTGGGACCTTGT	

Chromatin immunoprecipitation (ChIP)

ChIP assay was performed as previously described (Luco et al., 2010). In brief, 2×10^6 cells per sample were crosslinked for 10 min in 1% formaldehyde at room temperature, quenched with 125 mM glycine, and swelled on ice for 10 min. Chromatin was sonicated (Bioruptor, Diagenode) to an average length of 200-500 bp and incubated overnight with pre-coated anti-IgG magnetic beads (Dynabeads M-280 Invitrogen) previously incubated with the antibodies for 6 h at 4°C. The antibodies used were: rabbit anti-H3K4me1 (5 μg , Abcam ab8895), rabbit anti-H3K4me3 (5 μg , Upstate 05-745), rabbit anti-H3K9me1 (5 μg , Abcam ab9045), rabbit anti-H3K9Me3 (5 μg , Abcam ab8898), rabbit anti-H3K9ac (7.5 μg , Upstate 06-942), mouse anti-H3K27me3 (10 μl , provided by Dr. H. Kimura,

Osaka), rabbit-anti H3K27ac (10 μ l, Dr. H. Kimura), rabbit anti-H3K36me3 (4 μ g, Abcam 9050), rabbit anti-H3K56ac (5 μ g, Upstate 07-677), rabbit H3K79me2 (5 μ g, Abcam 3594), rabbit H4K20me1 (5 μ g, Abcam 9051), and rabbit anti-H3 (1 μ g, Abcam ab1791). Control immunoprecipitations were performed with no antibody and ChIP for histone modifications was normalized to anti-H3 ChIP. Subsequently, the beads were washed several times and eluted in 1% SDS and 100 mM NaHCO₃ buffer for 15 min at 65°C. The eluates were incubated at 65°C for 6 hours to reverse crosslinks. Chromatin was precipitated with 100% ethanol overnight, treated with proteinase K, and purified using a PCR purification kit (Qiagen). Immunoprecipitated DNA (1.5 μ l) and serial dilutions of the 10% input DNA (1:4, 1:20, 1:100 and 1:500) were analyzed with SYBR-Green (iQ Supermix Bio Rad) on a Bio Rad C1000 ThermoCycler. Primers used in this study are listed below.

hCycA1.i1	F: CCCACCCACCTATGAGTGTAGT R: ACCCTCCATTCTCATCAAGACCT	Tm: 60
hCycA.3utr	F: ATTCCCTGGGTGATACCATTCAAT R: ATGACAACGTGGTGAGGCTATTCT	Tm: 60
hNanog.tss	F: TCTTGAATGTTGGGTTTGGGAATAG R: TCTGGGGGAAGGGAGGTGTTT	Tm: 60
hSatII	F: CATCGAATGGAAATGAAAGGAGTC R: ACCATTGGATGATTGCAGTCAA	Tm: 60
hEVX1.tss	F: TTCGCTGTGGCAGACGTTTCTATT R: AAGCCCCATTGCCCTCTTCTTT	Tm: 60
hALK.i18	F: AACTGGAAGACAGGTCCCA R: AGCCTGTCAAATCGGGATGAG	Tm: 65
hALK.i19a	F: CAGCCAGGAGGATACACACG R: GAATTGGGTGGGGTGGTGAT	Tm: 65
hALK.i19b	F: AGGCAGGGATGGTAACTCCT	Tm: 58

	R: GCCTTCCAGAACATCCTCACACA	
hALK.i20	F: TCATGCTCCTTGGGGAGAGA R: AAGTCTACCTGGCTCCCCTT	Tm: 65
hNPM1.i3	F: TACAGCCAACGGTAAGGGCA R: GCAGAAGCCCCCATGTTCAA	Tm: 58
hNPM1.i4a	F: CCCTTGGGGGCTTTGAAATAAC R: GCAACTGCACTAAAGAGGACAA	Tm: 65
hNPM1.i4b	F: ACCTGGCAGTGAACATTAGG R: AACTCAAGCAGCAAGAAGTCA	Tm: 65
hNPM1.i6	F: GGATATTGGGTCTGTGAGCCT R: AAACATACCAGGCTTGATGGG	Tm: 58

Lentiviral constructs and stable cell lines

The eGFP cDNA was first subcloned into pCDH1-MCS1-EF1-blasticidin (Pegoraro et al. 2009) using NheI and XhoI. ASH2L (provided by Dr. D. Skalnik, IUPUI), SET7/9 (Addgene plasmid 24084, provided by Dr. D. Reinberg, NYU) and SUV3-9H1 (provided by Dr. T. Jenuwein, Max Planck Institute, Freiburg) cDNAs were amplified using primers containing unique XhoI and NotI sites. The digested product was ligated in-frame into the pCDH1-MCS1-EF1-eGFP vector. Lentivirus was packaged by co-transfection of constructs (4.9 µg) with pMD2.G (1.5 µg) and pSpax2 (3.6 µg) with XtremeGENE HP (Roche) in 10cm plates of 293FT cells. Medium was changed from DMEM to RPMI-1640 at 12 hours after transfection, and supernatant was collected at 60 hours and filtered through a 0.45 µm filter. 500,000 cells were plated in 100ul in 24-well plates. One ml of viral supernatant with 4 µg ml⁻¹ polybrene was added per well and the plate was spun at 670g for 60 minutes. Cells were placed in an incubator at 32°C and 5% CO₂ for 16 hours, pelleted, resuspended in fresh RPMI-1640, and transferred to a 37°C

incubator. Blasticidin was added at a final concentration of $7.5 \mu\text{g ml}^{-1}$ 48 hours after transduction and continued for 7 days. Stable cell lines were then sorted for GFP expression to generate populations with comparable transgene expression levels (FACSVantage, BD Biosciences).

Irradiation

To induce DNA breaks and translocations, cells were irradiated using a cesium Mark-1 irradiator. A dose of 25 Gy was optimized previously to ensure comparable induction of H2AX-phosphorylation and cell viability after IR (Mathas et al., 2009).

Note: hiBA-FISH methods are described in Chapter 5

Chapter 5

Quantitative detection of rare interphase chromosome breaks and translocations by high-throughput imaging

Collaborator: Gianluca Pegoraro (NCI High Throughput Imaging Facility)

Balanced chromosome translocations are often causally associated with disease and are commonly used for diagnostic purposes (Mitelman et al., 2007; Frohling and Dohner, 2008). In clinical practice, translocations are routinely detected by cytogenetic and polymerase chain reaction (PCR)-based methods. Reverse-transcription PCR (RT-PCR) is widely used for detecting oncogenic gene fusions since primers spanning exon-exon junctions of the chimeric mRNA result in high sensitivity and specificity. PCR of genomic DNA is also suitable for detecting translocations that involve the fusion of an oncogene to a gene's regulatory region without the formation of a chimeric mRNA. However, PCR detection of translocations requires relatively precise knowledge of the break sites and translocation breakpoints in a given translocation partner are frequently found over large genomic distances requiring multiplexing with extensive primer sets or nested approaches making their detection by PCR often impractical. More recently, genome-wide sequencing approaches have enabled detection of translocations in an unbiased fashion but with reduced sensitivity compared to PCR-based methods (Chiarle et al., 2011; Klein et al., 2011; Talkowski et al., 2011).

A prominent cytological method for detection of translocations is Fluorescence in Situ Hybridization (FISH), which is commonly used in biological and clinical applications. The physical pairing of translocation partners can be detected as the co-localization of FISH probes targeted to the involved translocation genes in metaphase chromosome spreads (Wolff et al., 2007). Specificity is greatly increased by the use of break-apart FISH probes consisting of two differentially labeled probes placed upstream and

downstream of the putative breakpoint region, which can be used to detect chromosome breakage in metaphase or interphase preparations (Gozetti and Le Beau, 2000). Chromosome breakage is indicated by separation of the two probes and, if combined with a third probe targeted towards a putative translocation partner, translocations can be detected by co-localization of a separated break-apart probe with the translocation partner. A major advantage of using break-apart FISH probes over PCR analysis to detect translocations is that precise knowledge of the translocation partner or chromosome breakpoint is not required and probes can be designed so that large regions, up to 500kb can be interrogated (Wolff et al., 2007; Gozetti and Le Beau, 2000; Kearney, 2001; De Melo et al; 2008).

A major limitation of standard or break-apart FISH is that it requires visual inspection of a large number of cells to detect a sufficient number of chromosome breakage or translocation events, and determining a split signal can be biased by user subjectivity. For practical reasons, the number of cells analyzed by traditional FISH is typically limited to a few hundred and as such FISH is well suited for analysis of cell populations that contain frequent translocations, but detection of rare translocations is often prohibitive. In addition, visual inspection of relatively small cell numbers makes it difficult to measure statistically significant differences between biological samples containing low-frequency chromosome breakage and translocation events (Wolff et al., 2007; Gozetti and Le Beau, 2000; Kearney, 2001).

Considerable progress has recently been made in High-Throughput Imaging (HTI) and automated image analysis (Conrad and Gerlich, 2010; Elceiri et al., 2012; Roukos and Misteli, 2014; Liberali et al., 2015). We describe here the development of a systematic and unbiased method for the quantitative detection of rare chromosome breakage and translocation events in interphase cells by combining break-apart FISH with HTI. We implement a **high-throughput break-apart FISH** (hiBA-FISH) pipeline based on the physical separation in 3D space of break-apart probes flanking putative translocation breakpoint regions (Fig. 1). hiBA-FISH consists of fixation of interphase cells on coverslips, followed by DNA FISH using translocation gene-specific break-apart probes. Large image datasets containing thousands of cells per experimental condition are acquired using automated 3D confocal high-throughput microscopy and analyzed using high-content image analysis software to determine the spatial positioning of FISH signals in three separate channels and calculates distances between them. To detect chromosome breakage and translocation events, FISH signal distance datasets are analyzed using statistical analysis software and frequencies of chromosome breakage and translocation events are measured by establishing distance thresholds for the FISH probes.

As a proof-of-principle, we applied hiBA-FISH to measure the number of chromosome breaks at the *NPM1* and *ALK1* gene loci and of the frequency of the anaplastic large cell lymphoma-specific *NPM1-ALK* translocation upon irradiation (Tabbo et al., 2013). We demonstrate sensitive detection of rare chromosome breakage and translocation events by hiBA-FISH. Application of hiBA-FISH to cell lines with several different biological treatments is described in Chapter 4.

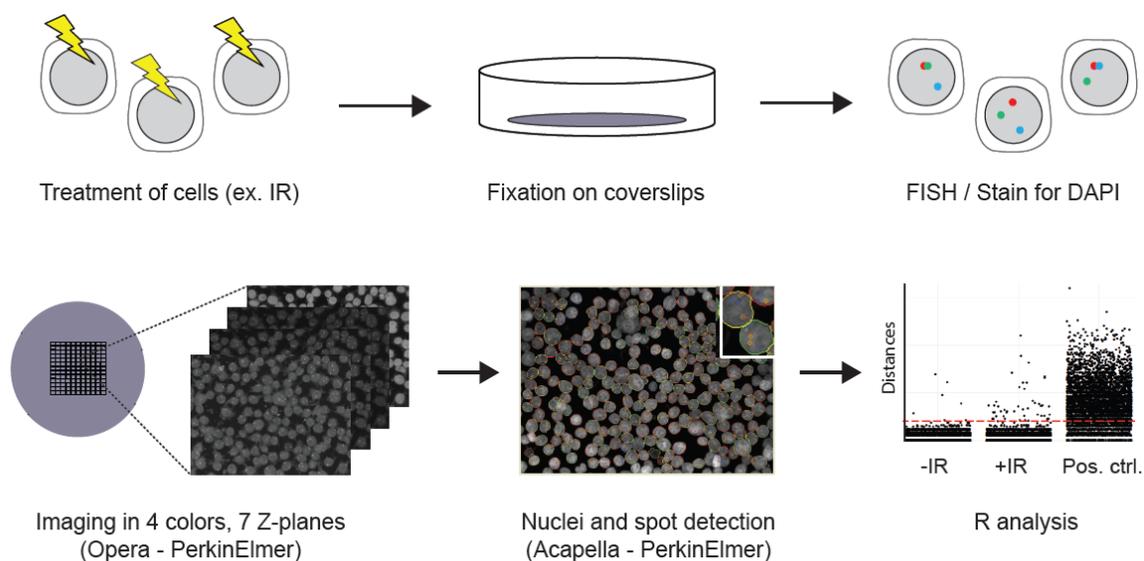


Figure 1: hiBA-FISH pipeline. The green, red and blue dots represent FISH signals in fixed interphase cell nuclei. FISH: Fluorescence in Situ Hybridization. DAPI: 4',6-diamidino-2-phenylindole. HTI: High-throughput Imaging.

RESULTS

Break-apart probe design

hiBA-FISH is based on the combinatorial use of break-apart probes that flank known or putative translocation breakpoints (Fig. 2). Several commercial, quality-controlled break-apart probes are readily available and can be used for hiBA-FISH, or break-apart probes can be generated for virtually any region of the genome by incorporation of fluorescent nucleotides into bacterial artificial chromosome (BAC) DNAs using standard nick translation (Ventura et al., 2006). Suitable BAC DNAs upstream and downstream of the target breakpoints (up to a few hundred kb) are readily identified using the UCSC genome browser. Ideally, BAC DNAs should be selected with similar sequence lengths

to generate similar FISH signal sizes, although signal size may be influenced by secondary DNA structure and should be optimized by visualization (Ventura et al., 2006).

When used alone in interphase cells, two-color break-apart probes report on chromosome breakage (Fig. 2). The two signals of a break-apart probe pair are in proximity at the intact allele (Fig. 2A). Chromosomal breakage and separation of the chromosomal region between the two probes can be detected by their separation (Fig. 2B). In addition, the combination of a two-color break-apart probe with a third color probe targeted a translocation partner can be used to identify translocation events, which are detected by the concomitant proximity of a separated break-apart probes with the probe targeted against a known translocation partner (Fig. 2C).

As a model system to develop and test hiBA-FISH, we designed probes for the well-characterized recurrent translocation between the 5' region upstream of the *NPM1* breakpoint in intron 4 (Chr. 5q35) and the 3' region downstream of the *ALK* breakpoint in intron 19 (Chr. 2p23) in anaplastic large cell lymphoma (ALCL) (Duyster et al., 2001). We created two separate three-color probe sets named after the break-apart probes they contain (Fig. 3). The hiBA-*ALK* probe set consists of an Alexa488-labeled probe (Green) targeting the 5'-region upstream of the *ALK* breakpoint, an Alexa568-labeled probe targeting the 3'-region downstream of the *ALK* breakpoint, and a Cy5-labeled probe (Far Red) targeting the 5' region upstream of the *NPM1* breakpoint (Fig. 3). The hiBA-*NPM1* probe set was designed analogously (Fig. 3). This design strategy allowed us to generate probe sets that report separately on *NPM1* or *ALK* breakage in two colors,

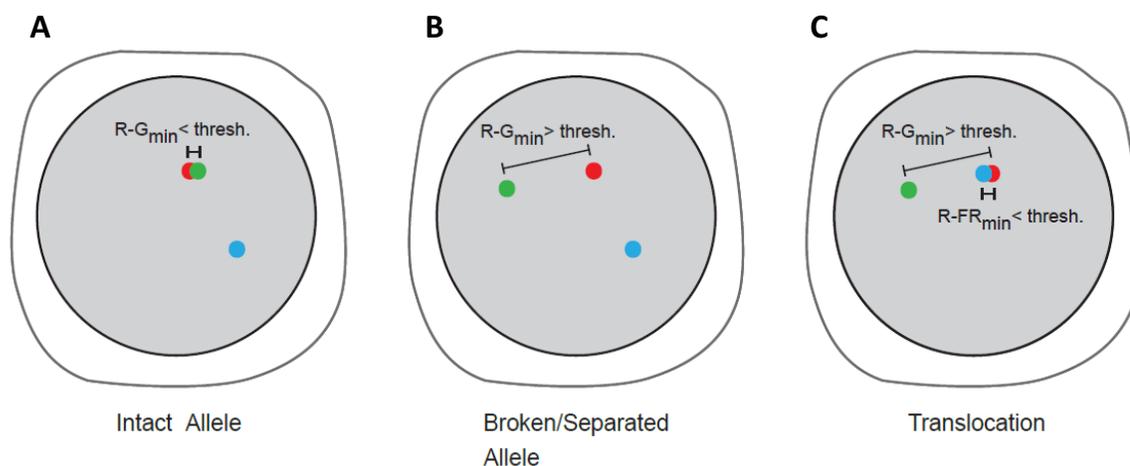


Figure 2: Outline of hiBA-FISH event definitions based on the thresholding of relative Euclidean distances of FISH signals in different colors. (A) intact allele, (B) broken allele, and (C) translocation. $R-G_{\min}$ and $R-FR_{\min}$ indicate the per Red signal minimum Red/Green and Red/FarRed distances, respectively.

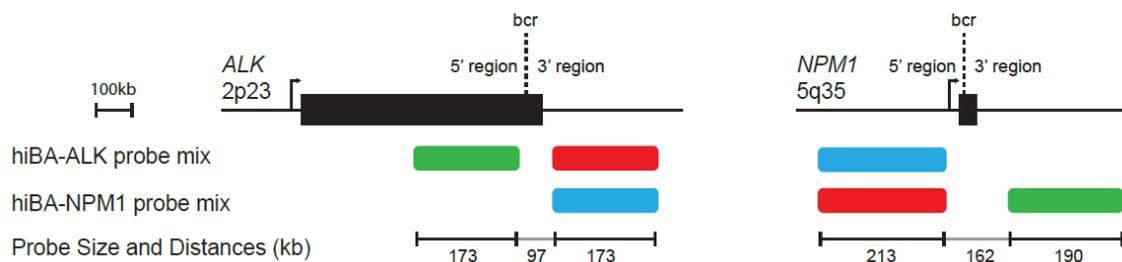


Figure 3: Schematic representation of the size and location of the chromosomal breakpoint regions recognized by the two different hiBA-FISH probe sets used in this study. bcr: breakpoint cluster region.

and *NPM1-ALK* translocations in three colors use only four BAC DNAs.

Automated hiBA-FISH signal detection

For hiBA-FISH, cells were grown in suspension and plated on poly-D-lysine coverslips. After fixation with 4% paraformaldehyde/PBS, nuclei were FISH stained using hiBA-ALK or hiBA-NPM1 probe sets (see Methods for details). Coverslips were mounted using DAPI-containing medium on glass slides and imaged using a high-throughput confocal microscope with a 40X water objective and a camera pixel binning of 2 (pixel size: 320nm). For each field of view, samples were imaged in 4 channels and 7 Z-planes. A typical imaging field contained 100-200 cells (Fig. 4) and more than 50 fields were routinely imaged per coverslip to acquire 5-10,000 cells per experimental condition.

In order to localize FISH signals and measure relative distances between them, we adapted a previously described custom image analysis script (Roukos et al., 2013). Briefly, image stacks of each channel for each field of view were maximally projected in 2D and nuclei were segmented based on the DAPI channel (Fig. 4A-B). Irregularly shaped segmented nuclei, often due to segmentation errors, and nuclei touching the image border were excluded from further analysis. The nucleus Region of Interest (ROI) was then used as the search region for FISH spot detection in the Green, Red and FarRed channels (Figs.4D, 4F and 4H). To determine the relative position of FISH signals, center-to-center Euclidean distances between each Red FISH signal and all the Green and FarRed signals in the same cell were measured (see Methods). Output attributes of the hiBA-FISH image analysis pipeline include the number of nuclei, the number of FISH spots detected per cell in each channel, as well as complete Red/Green and Red/FarRed distance datasets for all detected Red FISH signals (see Methods).

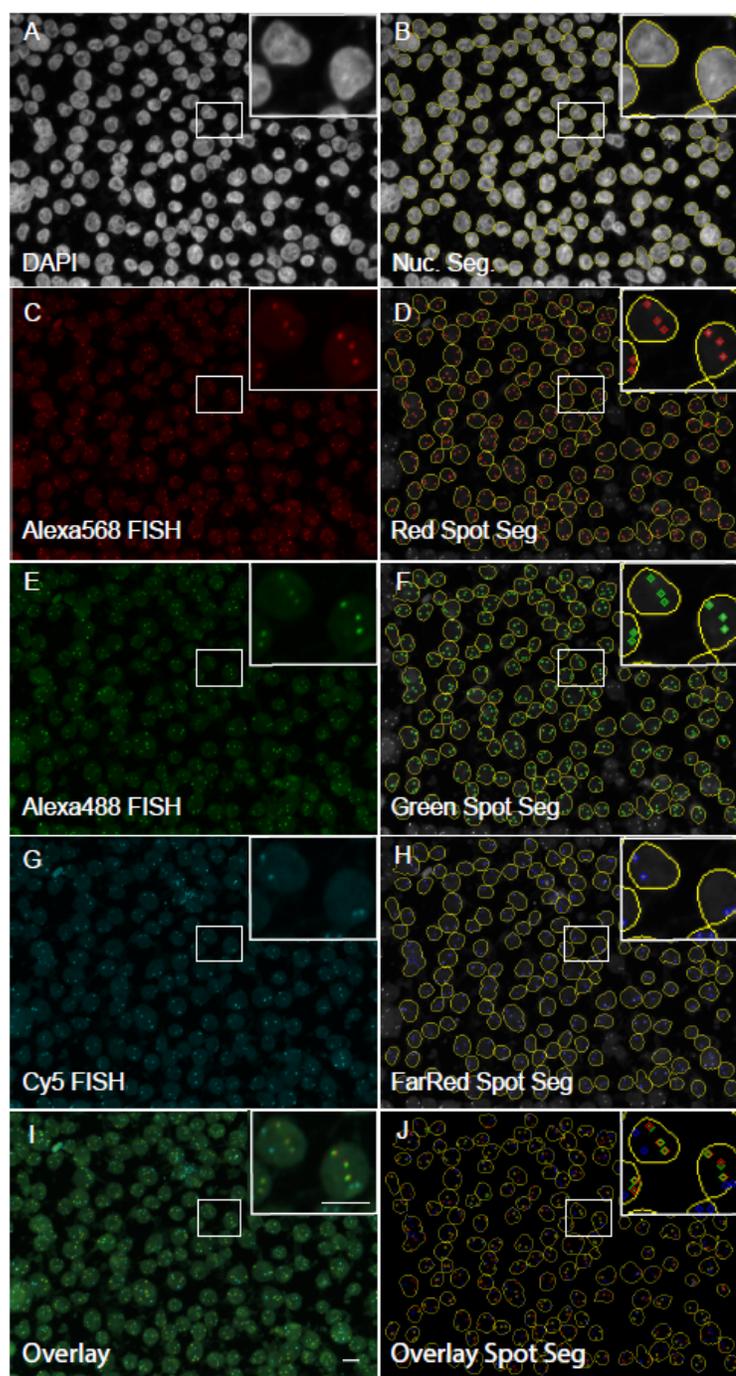


Figure 4: Automated nucleus segmentation and FISH signal detection. **(A, C, E, G, I)** Maximal projections of 40X confocal image Z-stacks of Mac2A cells stained with the hiBA-ALK probe set. The overlay represents a composite image of the Green (Alexa488), Red (Alexa568) and Far Red (Cy5) channels. The inset in each panel represents a magnified image of two representative Mac2A cells, DAPI: 4',6-diamidino-2-phenylindole. Scale bars: 10 μm . **(B, D, F, H, J)** *In silico* generated images representing the detected nucleus ROI (in yellow) and the FISH signal ROI's (in Green, Red and Blue).

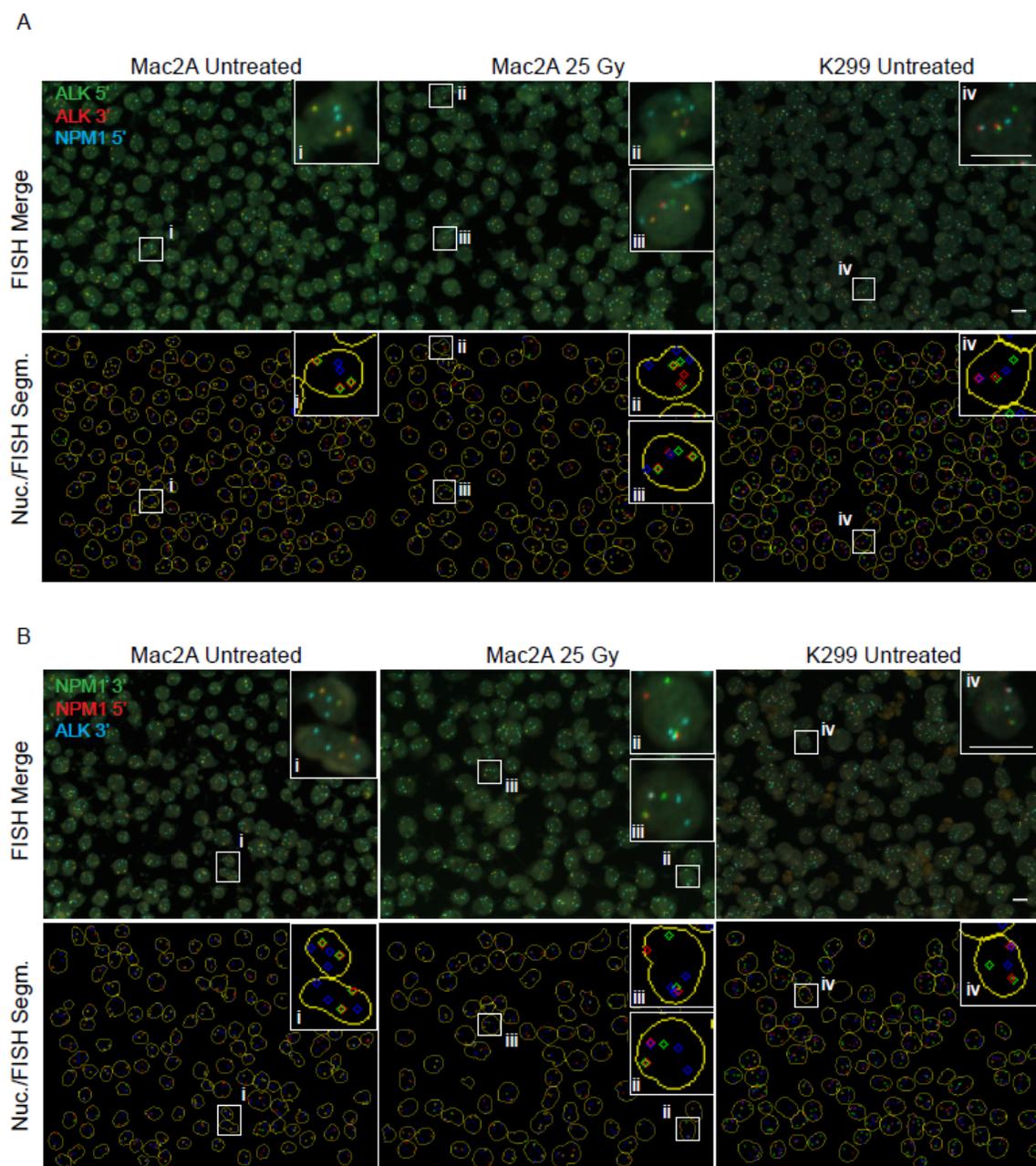


Figure 5: Qualitative identification of breakage and translocation events with hiBA-FISH. **(A)** Representative maximal projections of 40X confocal images of *NPM1-ALK* translocation-negative Mac2A cells, untreated or treated with 25Gy of ionizing radiation, and untreated *NPM1-ALK* translocation-positive K299 cells stained with the hiBA-*ALK* probe set. Overlays of the three FISH probe channels images (top), the nucleus segmentation (yellow) and FISH spot detection (bottom). Ai: Magnification of an intact *ALK* allele event. Aii: Broken *ALK* allele events. Aiii and Aiv: *NPM1-ALK* translocation events. Scale bar: 10 μ m. **(B)** Same as A, but cells were stained with the hiBA-*ALK* probe set. Ai: Magnification of an intact *NPM1* allele event. Aii: Broken *NPM1* allele events. Aiii and Aiv: *NPM1-ALK* translocation events. Scale bars: 10 μ m.

Qualitative visual inspection of *NPM1-ALK* translocation-negative Mac2A and *NPM1-ALK* translocation-positive K299 cells (Mathas et al., 2009) confirmed the predicted spatial positioning patterns for the FISH probe sets (Fig. 5). As expected, in Mac2A cells, Red and Green break-apart probe signals for both probe sets were almost exclusively in spatial proximity of each other, indicating intact *ALK* and *NPM1* alleles in almost all-cells (Fig. 5Ai and Fig. 5Bi). Separation of Red and Green signals indicating *ALK* or *NPM1* breakage (Fig. 5Aii and Fig. 5Bii), and concomitant spatial proximity of a separated Red signal with a FarRed signals indicating *NPM1-ALK* translocations was found in a small number of Mac2A cells after irradiation of cells with 25 Gy to induce global DNA damage (Fig. 5Aiii and Fig. 5Biii). As expected, almost all *NPM1-ALK* translocation-positive K299 cells showed separation of at least one break-apart probe and spatial proximity of at least one separated Red signal with a FarRed signal (Fig. 5Aiv and Fig. 5Biv).

hiBA-FISH signal quantitation

Automated hiBA-FISH image analysis was used to quantitate FISH signal positioning patterns and inter-signal distances in the cell population. Quantitative signal data was generated by automated analysis of typically over 5,000 individual *NPM1-ALK* translocation-negative Mac2A cells and over 2,000 *NPM1-ALK* translocation-containing K299 cells. FISH signals were detected with greater than 99% accuracy based on comparison of visual and automated detection of FISH signals (data not shown). In agreement with previous visual counting of FISH spots (Mathas et al., 2009), most Mac2A cells had 3 *ALK* alleles (71.40% using the hiBA-*ALK* Green probe, 72.91%

hiBA-ALK Red, 69.86% hiBA-NPM1 FarRed) and 2 *NPM1* alleles (82.73% hiBA-NPM1 Green, 82.52% hiBA-NPM1 Red, 81.55% hiBA-ALK FarRed) (Fig. 6A, 6B and 6C). In K299 cells, subpopulations of cells with 2, 3 or 4 *NPM1* and *ALK* alleles were detected with the major subpopulation containing 4 alleles for both genes using both hiBA-FISH probe sets (Fig. 6A, 6B and 6C) (Gogusev et al., 2002). Irradiation of Mac2A cells did not significantly alter the FISH detection efficiency when compared to untreated samples (Fig. 6A, 6B and 6C). Altogether, these results indicate that hiBA-FISH can be used for the precise, robust and high-throughput detection of FISH signals and of their spatial arrangement in interphase nuclei in multiple channels.

Determination of FISH signal separation and proximity thresholds

To determine the threshold of separation of the break-apart probes that indicates breakage in the target region, we measured Red/Green and Red/FarRed distances in a large number of cells to define the distribution of distances of break-apart probes at intact sites. Only distances from cells that had at least 2 FISH signals in all three channels and had the same number of Green and Red spots were considered to eliminate cells with missed or spurious FISH spot detection events. To establish a threshold for the separation between break-apart probes, indicating chromosome breakage, we plotted the distribution of minimum Red/Green distances in non-irradiated Mac2A cells, which are not expected to contain breaks. Using hiBA-ALK and hiBA-NPM1 probe sets, 99.48% and 99.77% of the Red FISH signals were separated by ≤ 4 pixels (1.28 μm) from the closest Green FISH signal, respectively (Fig. 7A), with a median Red/Green distance of 1.0 pixel for

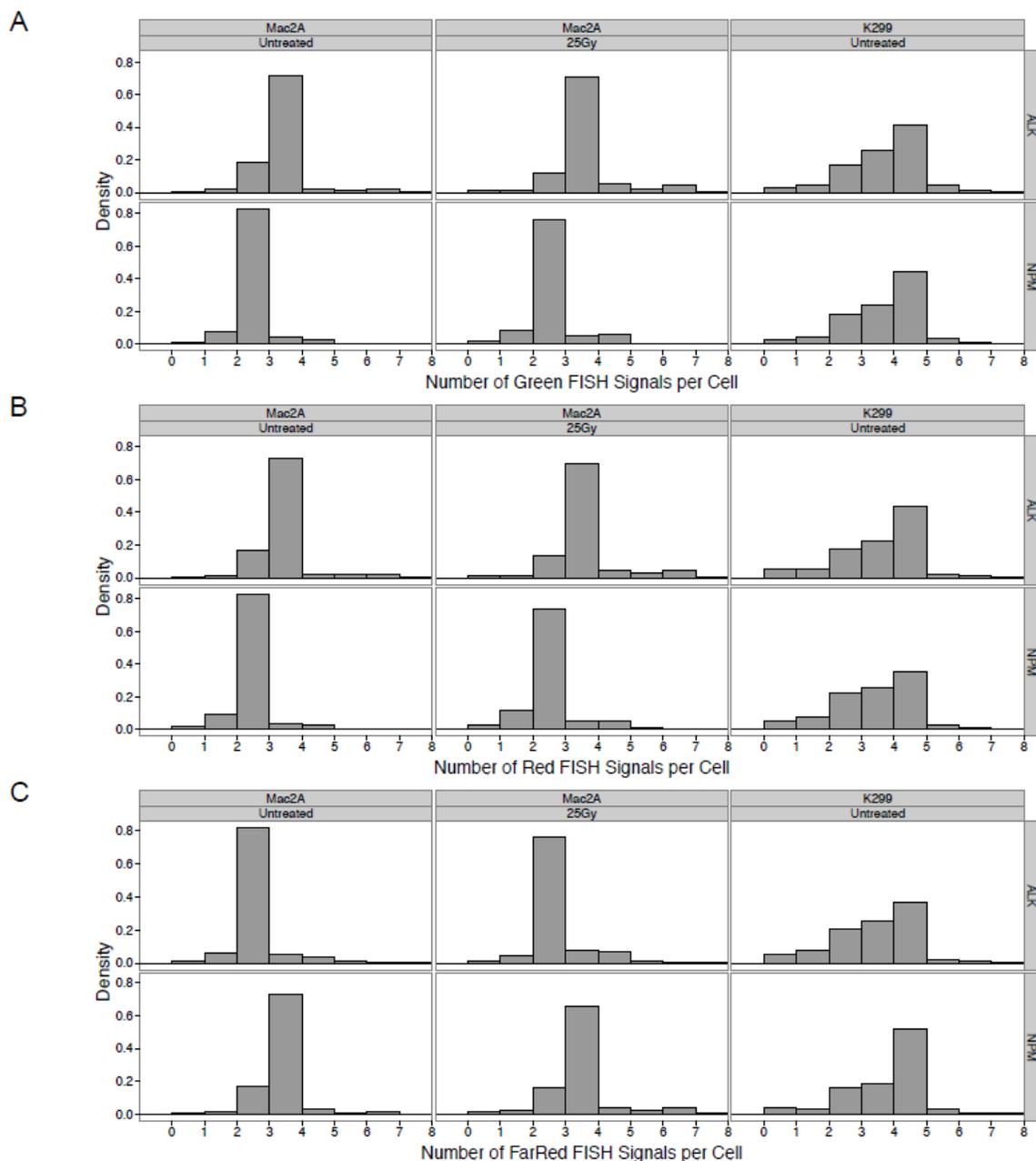


Figure 6: Automated hiBA-FISH signal detection metrics. **(A)** Histograms of the distributions of Green FISH signal number per cell as measured by automated image analysis in the indicated cell lines and experimental conditions. ALK: hiBA-ALK probe set, NPM: hiBA-NPM1 probe set. Bin size = 1 FISH signal per cell. The first bin includes cells with 0 spots. **(B)** Same as **A**, but for Red FISH signals. **(C)** Same as **A**, but for Far Red Signals.

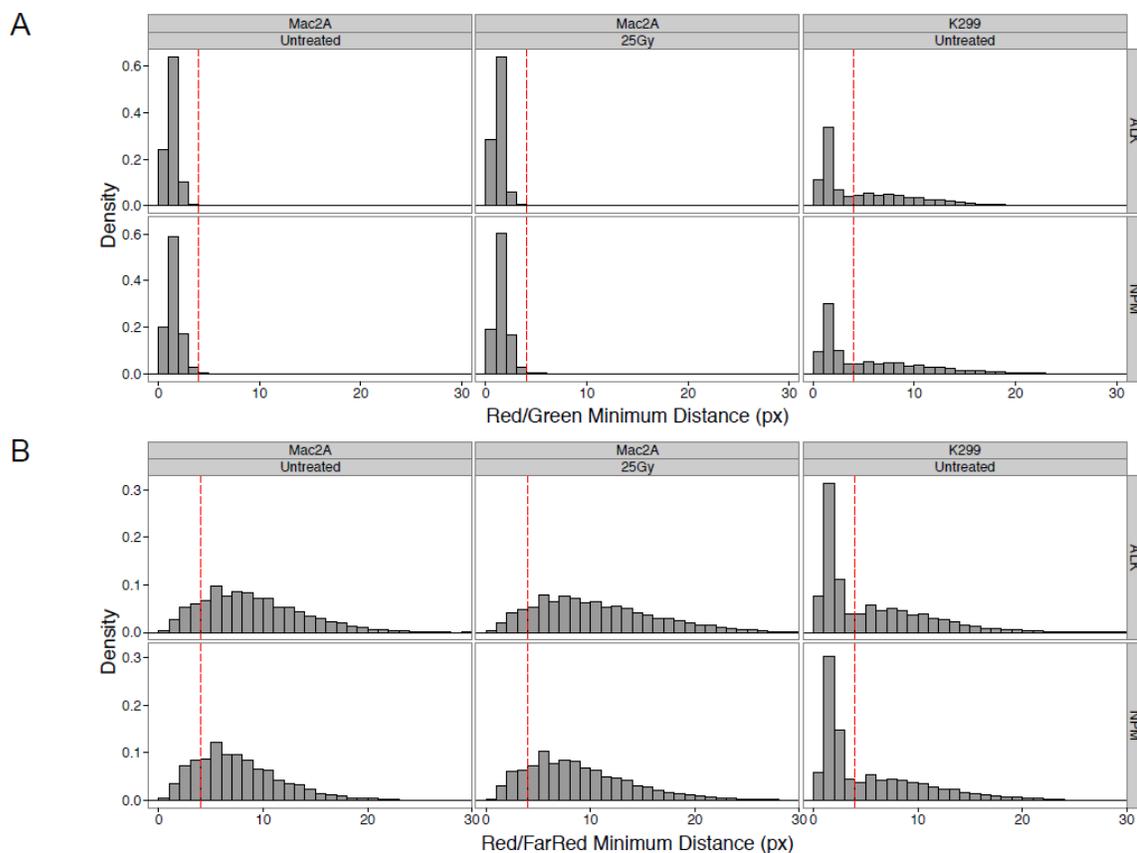


Figure 7: Experimental determination of a proximity threshold for FISH signals based on hiBA-FISH distances. **(A)** Distance distribution histograms of minimum Red/Green FISH signal distances per Red allele in the indicated cell lines and experimental conditions. ALK: hiBA-ALK probe set, NPM: hiBA-NPM1 probe set. 1 pixel = 320nm. Bin size = 1 pixel. The first bin includes distances between 0 and less than 1 pixel. The red dashed line represents a threshold of 4 pixels (1.28 μm). **(B)** Same as **A**, but for minimum Red/FarRed FISH signal distances.

both probe sets. Based on these data, we chose separation between break-apart probes by more than 4 pixels as an indicator of chromosome breakage. The accuracy of this threshold was validated in translocation-positive K299 cells, where 56.53% and 54.99% of Red FISH signals were separated by more than 4 pixels from the closest Green FISH signal using hiBA-ALK and hiBA-NPM1 probe sets, respectively, consistent with the

presence of at least one *NPM1-ALK* translocation per cell in the vast majority of cells (Fig. 7A).

In line with a threshold of ≤ 4 pixels as an indicator of an intact locus, we also defined co-localization of 5' *NPM1* and 3' *ALK* probes in translocation events as a distance of ≤ 4 pixels (Fig. 7B). In *NPM1-ALK* translocation-negative Mac2A cells, 15.09% and 20.87% of 5' *NPM1* and 3' *ALK* pairs had distances ≤ 4 pixels, when detected using the hiBA-NPM1 or hiBA-ALK probe sets, respectively. The higher percentage of proximal 5' *NPM1* and 3' *ALK* pairs for the hiBA-ALK probe is accounted for by the presence of 3 *ALK* alleles compared to 2 *NPM1* alleles in Mac2A cells (Mathas et al., 2009). In contrast, in K299 cells, 54.50% and 55.89% of 5' *NPM1* and 3' *ALK* pairs were in spatial proximity (≤ 4 pixels) using hiBA-ALK and hiBA-NPM1 probes, respectively (Fig. 7B), in accordance with the observation that half of *ALK* and *NPM1* alleles are translocated in these cells (Fischer et al., 1988).

Quantitation of chromosome breaks

Using the thresholds determined above, we defined an *ALK* or *NPM1* breakage event as separation of the break-apart probes by > 4 pixels. As predicted, an overwhelming majority of control K299 cells possessed at least one DNA break event in the *ALK* gene (hiBA-ALK, 2571/2695, 95.40%, 95% CI: 94.53% – 96.12%) or in the *NPM1* gene (hiBA-NPM1, 2352/2448, 96.08%, 95% CI: 95.23% – 96.78%) per cell (Fig. 8A-C). In contrast, in untreated Mac2A cells, where *ALK* and *NPM1* breakage is not present, the percentage of cells carrying at least one *ALK* break or one *NPM1* break was 0.66%

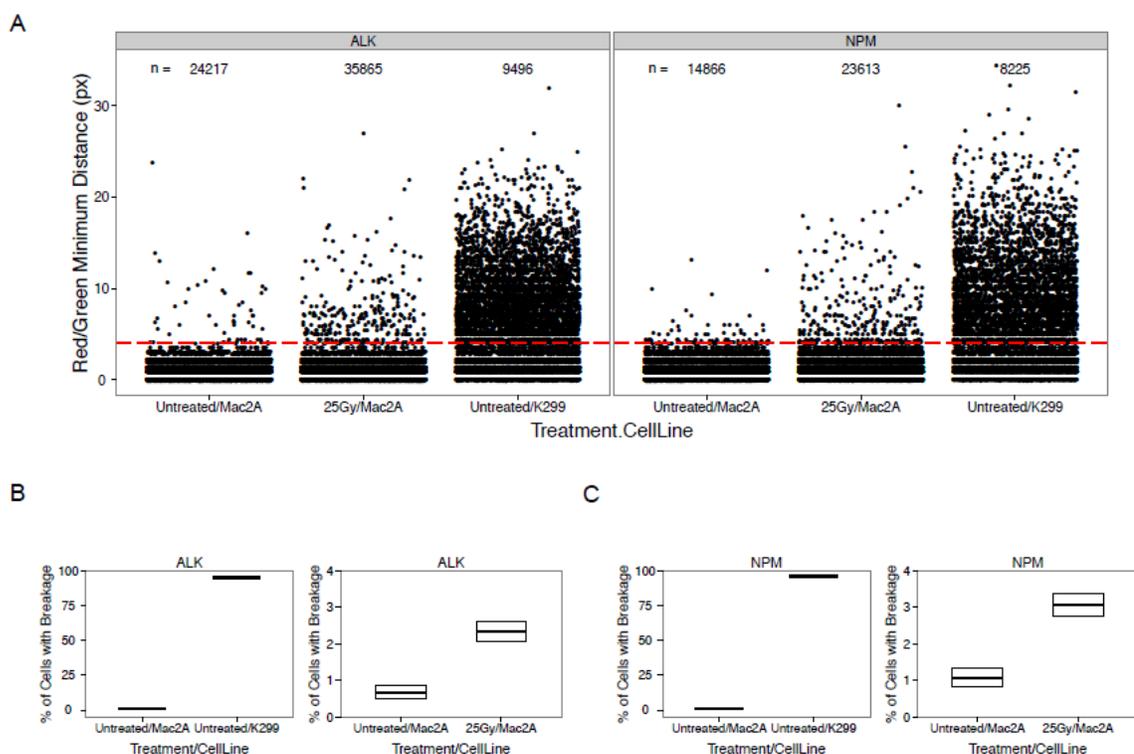


Figure 8: Automated determination of chromosomal breaks by hiBA-FISH. **(A)** Jitter-plot of the minimum Red/Green FISH per Red allele signal distances in the indicated cell lines and experimental conditions. ALK: hiBA-ALK probe set, NPM: hiBA-NPM1 set. 1 pixel = 320nm. Each dot represents a measured Red/Green Distance. The number of Red/Green distances plotted for each single experimental condition is indicated. Breakage events have a Red/Green minimum distance of > 4 pixels (1.28 μm , red dashed line). **(B)** Crossbar plot indicates the frequency of cells with at least one breakage event (middle line) and its relative 95% CI (from top to bottom line) expressed as percentages for the indicated cell lines and treatments as measured with the hiBA-ALK probe set for the datasets shown in **A**. The same frequency values relative to Untreated/Mac2A cells were plotted in both the left and right panels for comparison purposes. **(C)** Same as **B**, but for the hiBA-ALK set.

(hiBA-ALK, 53/7984, 95% CI: 0.50% – 0.86%) and 1.05% (hiBA-NPM, 75/7089, 95% CI: 0.84% – 1.32%), respectively (Fig. 8A-C).

To determine the sensitivity of hiBA-FISH, we induced chromosome breaks and translocation by irradiation of Mac2A cells with 25 Gy, generating 500-750 DNA breaks per diploid genome (Dikomey et al., 1993). Considering that the break-apart probes are separated by ~100kb of DNA, a breakage event in the probed *ALK* and *NPM1* region is expected in ~ 1 out of 50 cells. In line with this estimate when compared to untreated cells, the percentage of detected chromosome breaks upon irradiation increased 3.6 fold from 0.66% (53/7984, 95% CI: 0.50% – 0.86%) to 2.36% (277/11753; 95% CI: 2.10% – 2.65%; Fisher's Exact Test p-value < 2.2e-16) for the hiBA-ALK probe (Fig. 8B) and 2.9 fold from 1.05% (75/7089, 95% CI: 0.84% – 1.32%) to 3.05% (343/11230; 95% CI: 2.75% – 3.39%; Fisher's Exact Test p-value < 2.2e-16) for the hiBA-NPM1 probe (Fig. 8C). We conclude that hiBA-FISH is capable of detecting chromosome breakage on a per cell basis with high sensitivity.

Sensitive detection of rare NPM1-ALK translocations

Finally, we analyzed the percentage of cells carrying at least one translocation event, defined as separation of the break-apart probe by > 4 pixels with simultaneous proximity of 5' *NPM1* and 3' *ALK* within a distance of ≤ 4 pixels. hiBA-FISH identified at least one *NPM1-ALK* translocation event in 93.06% (2508/2695, 95% CI: 92.04% - 93.96%) of the control K299 cells using the hiBA-ALK probe set and in 94.61% (2316/2448, 95% CI: 93.64% - 95.44%) using the hiBA-NPM probe set (Fig. 9A-C). This puts the false negative rate for the detection of translocations by hiBA-FISH between 5% and 10% assuming that the K299 cells are homogeneously *NPM1-ALK* positive. More importantly, a statistically significant difference between the percentage of *NPM1-ALK*

translocations in untreated Mac2A cells and irradiated Mac2A cells was measured using both hiBA-FISH probe sets (Fig. 9B and 9C). Upon irradiation of Mac2A cells, the percentage of cells carrying at least one *NPM1-ALK* translocation increased 4.3 fold from 0.088% (7/7984; 95% CI: 0.038% – 0.18%) to 0.38% (45/11753; 95% CI: 0.29% - 0.51%; Fisher's Exact Test p-value = 4.674e-5) as measured by the hiBA-ALK probe set (Fig. 9B); and 2 fold from 0.27% (19/7089; 95% CI: 0.17% – 0.42%) to 0.54% (61/11230; 95% CI: 0.42% - 0.69%; Fisher's Exact Test p-value = 0.00562) for the hiBA-NPM1 probe set (Fig. 9C).

Taken together, these results confirm that hiBA-FISH, through a combination of high-throughput imaging and single cell analysis of FISH signals, is a sensitive method for the detection and quantitative measurement of low-frequency breakage and translocation events in interphase cells.

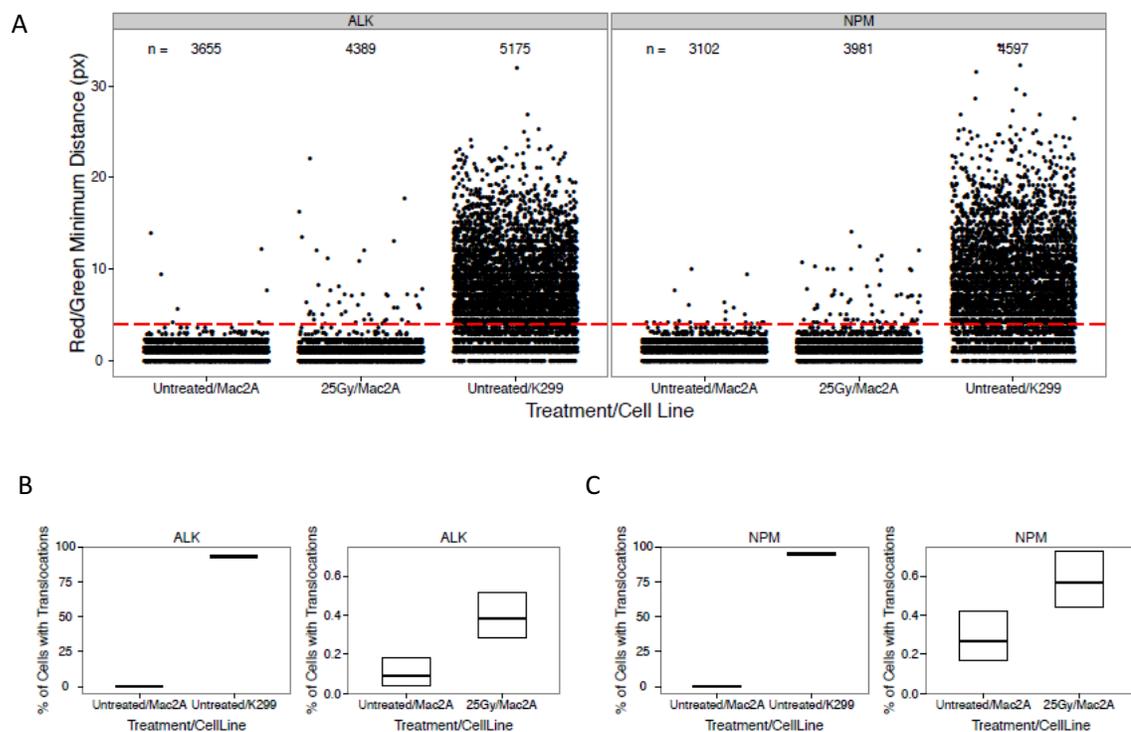


Figure 9: Automated determination of chromosomal translocations by hiBA-FISH. (A) Jitter-plot of the subset of Red FISH signals in proximity of a FarRed signal (Red/FarRed minimum distance ≤ 4 pixels). Translocation events have a Red/Green minimum distance of > 4 pixels (red dashed line). Each dot represents a Red/Green Distance. The number of Red/Green distances plotted for each single experimental condition is indicated. (B) Crossbar plot but for the frequency of cells with at least one *NPM1-ALK* translocation event as measured by the hiBA-ALK probe set. The same frequency values relative to untreated/Mac2A cells were plotted in both the left and right panels for comparison purposes. (C) Same as B, but for the hiBA-NPM1 probe set.

MATERIALS AND METHODS

Cell culture

Translocation-positive ALCL K299 and translocation-negative ALCL Mac2A cell lines were kindly provided by Dr. S. Mathas (Charite-Berlin) and maintained in RPMI-1640 at

37°C and 5% CO₂. Culture medium was supplemented with 10% FBS (Atlanta Biologicals), 2 mM L-glutamine, 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin.

Irradiation

To induce DNA breaks and translocations, cells were irradiated using a Cesium Mark-1 irradiator. A dose of 25 Gy was optimized previously to ensure comparable induction of H2AX-phosphorylation and cell viability across T cell lines after irradiation (Mathas et al., 2009).

FISH

To produce probes for 3D FISH, bacterial artificial chromosomes (BACs; BACPAC Resources Center) were directly labeled by nick translation with fluorescently labeled dUTPs (Chromatide AlexaFluor 488-5-dUTP and 568-5-dUTP from Life Technologies; Cy5 dUTP from Fisher Scientific) using a nick translation kit (Abbott Molecular). The following BAC clones were used:

hiBA-ALK Probe Set:

<i>ALK</i> 5' break-apart (labeled in 488):	RP11-119L19
<i>ALK</i> 3' break-apart (labeled in 568):	RP11-100C1
<i>NPM1</i> 5' (labeled in Cy5):	RP11-1072I20

hiBA-NPM1 Probe Set:

<i>NPM1</i> 5' break-apart (labeled in 568):	RP11-1072I20
<i>NPM1</i> 3' break-apart (labeled in 488):	RP11-145P20
<i>ALK</i> 3' (labeled in Cy5):	RP11-100C1

hiBA-Nanog Probe Set (described in Chapter 4):

<i>Nanog</i> 5' break apart (labeled in 488):	RP11-298G5
---	------------

<i>Nanog</i> 3' break apart (labeled in 568):	RP11-141A17
<i>CycA</i> 5' (labeled with Cy5):	RP11-829C16

The *ALK* break-apart probes were located 32 kb upstream and 65 kb downstream of the *ALK* breakpoint in intron 19 (Fig. 1C). The *NPM1* break-apart probes mapped 55kb upstream and 107kb downstream of the known *NPM1* breakpoint in intron 4 (Fig. 1C). The *Nanog* break-apart probes were upstream and downstream of the *Nanog* gene and were chosen with a distance of 95.5 kb in between them to resemble the *ALK* break apart probe. The 5' *CycA* probe mapped upstream of the *CycA* gene. The sequence specificity of all probes was verified by PCR.

For 3D FISH, cell suspensions were plated on 170 μ m thick glass poly-D-lysine coated coverslips (Neuvitro) in a 24-well plate at a concentration of 1 million cells per well, spun at 170g for 5 minutes, and then incubated at 37°C for 30 minutes. Cells were fixed in 4% PFA/PBS for 15 minutes, permeabilized for 20 minutes in 0.5% saponin (Sigma Aldrich)/0.5% Triton X-100/PBS and incubated in 0.1N HCl for 15 minutes with PBS washes between each step. After a 2x SSC wash, cells were equilibrated in 50% formamide/2x SSC buffer for at least 30 minutes. For each coverslip, a probe mix containing 80ng of each probe, 3 μ g COT1 DNA (Roche) and 20 μ g tRNA (Ambion) were ethanol precipitated and resuspended in 7.5 μ l hybridization buffer (10% dextran sulfate, 50% formamide, 2x SSC, and 1% Tween-20). Nuclei and probe were co-denatured at 85°C for 5 minutes and left to hybridize at 37°C in a humidified chamber overnight. To remove excess probe, cells were washed with 1x SSC at 45°C three times for 5 minutes each, followed by three washes with 0.1x SSC at 45°C for 5 minutes each. Coverslips

were then mounted in DAPI-containing Vectashield (Vector, Burlingame, CA) on glass slides optimized for high-throughput imaging (Tekdon, Myakka City, FL).

High-throughput imaging

Mounted coverslips were imaged using an Opera QEHS High-throughput confocal microscope (PerkinElmer, Waltham, MA) using a slide holder adaptor and running the Opera 1.8.1 software. All the image acquisition was performed using a Planar Achromatic 40X water immersion lens (Olympus, NA = 0.9), using 1.3 MegaPixel CCD cameras with pixel binning of 2. The pixel size in this imaging configuration is 320 nm. 4 channels (DAPI, Alexa488, Alexa568 and Cy5) were acquired in three separate exposures. For each coverslip, multi-channel images were acquired in 7 Z-planes (1.5 μ m apart) per field, over more than 50 fields per coverslip. For Mac2A cells, more than 5000 cells were imaged per experimental condition. For K299 cells more than 2000 cells were imaged per condition.

Automated image analysis

Image analysis was performed using Acapella 2.6 (PerkinElmer, Waltham, MA) using a modified version of previously described custom Acapella script (Roukos et al., 2013). Briefly, for each field images from different Z-planes were maximally projected. For each field, the image in the maximally projected DAPI channel was used for nuclear segmentation. Nuclear area and roundness were then calculated. These two nuclear attributes were used to filter out small or irregular nuclear objects, likely representing nuclear debris and/or nuclear segmentation errors. The filtered population of nuclei

ROI's was then used to sequentially detect FISH signals in different channels using a previously described spot detection algorithm (Roukos et al., 2013). Euclidean 2D distances in pixel units between the spot ROI centers of all the possible combinations of FISH signals of different colors (Alexa568/Alexa488: Red/Green; Alexa568/Cy5: Red/FarRed) in the same nucleus were then calculated. All the single-cell and single-spot-distance level data was individually indexed and exported as independent text files. The Acapella script and parameter files used for the analysis are available upon request.

Data analysis

hiBA-FISH data analysis was performed using the statistical analysis software R (Version 3.1.3, <http://www.R-project.org/>.) and RStudio (www.rstudio.com) using the knitr, plyr, data.table, ggplot2 and stringr packages. Single-cell and single-distance level data was read from the text files, concatenated, and experimental annotations (Cell line, IR treatment, Transfected construct, Experiment name, FISH probe mix) were extracted from the files names. Single-cell level information, such as the number of detected FISH signals per cell in each channel, was added to the single-spot-distance level data by an inner join operation based on common indexes in the two datasets. In order to exclude possible FISH and spot detection artifacts from the subsequent analysis, the single-spot-distance level dataset was filtered to retain only spot distances relative to cells that conformed to both these two criteria: a) at least 2 Green, 2 Red and 2 Far Red FISH signals and b) the same number of Green and Red FISH signals. Minimum per Red allele spot distances were calculated for the Red/Green and Red/FarRed datasets. The minimum Red/FarRed distances were then added to the corresponding minimum

Red/Green distances by an inner join using matching Red Spot indexes. A spot center-to-center distance proximity threshold of 4 pixels (1.28 μm) was empirically determined based on Mac2A cells. A breakage event on a per Red spot basis was defined as a minimum Red/Green distance of > 4 pixels. A translocation event on a per Red spot basis was defined as a minimum Red/Green distance of > 4 pixels and a minimum Red/FarRed distance of ≤ 4 pixels. Cells containing at least one breakage or translocation event were classified as positive for the respective event class. Confidence intervals for the percentages shown in Fig. 8 and 9 were calculated based on the modified Wald method (Agresti and Coull, 1998). Pairwise comparison of translocation or breakage events percentages between untreated and irradiated Mac2A cells was performed by applying the Fisher's exact test for count data. The original single-cell datasets, single-spot distance datasets, and R analysis script .rmd files are available upon request.

DISCUSSION

Here, we describe an unbiased, quantitative method for the detection of rare chromosome breaks and translocations in interphase cells with high sensitivity. hiBA-FISH is based on the high-throughput measurement of the spatial separation of break-apart FISH probes and analysis of large distance distributions datasets. The method allows for the visualization of individual, allele-specific breakage and translocation events and generates single-cell data statistics across large cell populations.

Break-apart FISH offers several advantages over conventional FISH. First, probes can be easily and rapidly designed to flank virtually any region of the genome with intervening

distances ranging from a few to several hundred kilobases. In most cases, existing BAC probes, often commercially available, can be used as probes, or genomic DNA clones can be generated for regions which are not covered by available BACs. This allows, for example, design of break-apart probes that flank chromosomal breakpoint sites in non-coding regulatory regions, which are involved in many B- and T-cell translocations (Markey et al., 2014; Semrau et al., 2014). Second, break-apart FISH probes can be used with multiple translocation partners to determine the relative frequency of each translocation within a population of cells. For example, the 5' *NPM1* probe in our hiBA-ALK probe mix could be substituted with any of several known, or hypothesized, *ALK* fusion partners (Marino-Enriquez and Dal Cin, 2013). Third, the use of reciprocal break-apart probe sets for each translocation partner, as shown here for the hiBA-ALK and hiBA-NPM1 probe sets, increases the accuracy of breakage and translocation detection. Finally, in addition to quantitation of chromosome breaks and translocations, hiBA-FISH provides information on allele copy number and spatial positioning of genome regions in intact nuclei via counting of the number of FISH signals and measurement of FISH signal distances.

Due to its high-throughput nature, hiBA-FISH is a highly sensitive method and is suited to analyze several of thousands of cells per experimental condition, in contrast to most FISH approaches that rely on visual inspection of relatively small sample numbers, or dual-fusion FISH probes which require metaphase spreads to distinguish between pairing and translocations. The combination of break-apart probes with a third probe to detect a potential translocation partner generates a highly specific and sensitive detection system,

since it requires a concomitant separation and a spatial proximity event to define a translocation. Imaging of thousands of cells per sample ensures a precise estimation of the frequency of these events. For the hiBA-ALK probe set, we detected here at least one *ALK* break in ~100 irradiated Mac2A cells and at least one *NPM1-ALK* translocation in ~300 irradiated Mac2A cells. While it seems plausible to detect even less frequent events by increasing the number of imaged cells, the imaging time required in these conditions may become prohibitive.

hiBA-FISH is a versatile experimental tool to probe the effect of biological, chemical, or physical treatments on chromosome breakage and translocation formation. While we demonstrate here hiBA-FISH in suspension cells plated on poly-D-lysine-coated coverslips, hiBA-FISH should be equally applicable to adherent cells grown directly on coverslips. In addition, hiBA-FISH can be scaled up considerably to assess multiple treatment conditions and probe combinations in the same experiment by use of multi-well plates (e.g. 96- or 384-well formats).

hiBA-FISH may also have clinical applications. DNA FISH, using both two-color fusion and break-apart probes, is an established tool in the clinical setting in the diagnosis and monitoring of patients with chromosome translocations. While RT-PCR remains the gold standard for detecting known gene fusions, FISH is commonly used as a complementary cytological method to validate RT-PCR results or when one of the translocation partners is highly variable. In addition, DNA FISH is the method of choice when RT-PCR primer sets are not available for a given fusion, when a fusion involves a non-coding region, or

when there is considerable breakpoint heterogeneity (Wolff et al., 2007). Several FDA-approved break-apart probe sets are already available to pathologists for the application to clinical samples. However, break-apart FISH is currently limited to manual or semi-automated analysis and is thus only useful for detection of translocations that occur with high frequency in a cell population. hi-BA FISH overcomes this limitation and may therefore be a useful complementary tool for diagnosticians to detect rare breakage and translocation events in highly heterogeneous samples. A particularly suitable application may be the cytogenetic detection of minimal residual disease in cell populations following treatment regimes. Finally, as HTI instrumentation and image analysis tools evolve, we envision that hiBA-FISH will be applicable to tissue samples and analysis of clinical biopsy samples from solid organs.

Chapter 6

General discussion and future directions

DISCUSSION

Chromosome breakage is a requisite early event in the formation of cancerous translocations. Here, we provide evidence that altered levels of histone modifications at translocation sites influence chromosome breakage and translocation frequency. Using large-scale computational analysis of the histone modification landscape of 74 frequent translocation genes, we identified, in an unbiased fashion, several individual and combinations of histone modifications that are frequently enriched over translocation prone genes (Chapter 2). To probe the physiological relevance of these modifications, we biochemically characterized by ChIP in detail the histone modifications patterns in the well-defined and clinically relevant *NPM1* and *ALK* translocation regions in anaplastic large cell lymphoma (Chapter 4). The functional role of histone modifications in breakage formation and translocations is directly demonstrated in several sets of experiments in which we modulate the histone modification landscape, either locally using a tethering approach (Chapter 3) or globally (Chapter 4), and we demonstrate increased susceptibility to breakage and higher frequency of translocation formation in chromatin domains marked by specific histone modifications, particularly H3K4 methylation.

Using both genome-wide (Chapter 2) and locus-specific (Chapter 4) approaches to map histone modifications, we identified H3K4 mono- and tri-methylation as the most abundant histone modifications in translocation-prone genome regions. These modifications are generally associated with transcriptionally active, open genome regions (Bannister and Kouzarides, 2011). This finding is in line with recent correlative

observations from genome-wide studies that mapped translocation junctions after DSBs were experimentally induced by I-SceI in B cells (Chiarle et al., 2011 and Klein et al., 2011). The majority of translocation junctions were mapped within or near transcriptionally active regions of the genome that were enriched in H3K4me3, H3K36me3, and H3 acetylation (Klein et al., 2011). In addition, in translocation-negative prostate cancer cells, liganded androgen receptor (AR) binds near the *TMPRSS2* translocation gene and increases its transcriptional activity, consequently leading to elevated H3K4me3 levels across the breakpoint region (Lin et al., 2009, Yu et al., 2010). As observed in our experiments in ALCL cells, upon ionizing radiation AR-treated prostate cancer cells form increased DSBs and *TMPRSS2-ERG* fusions when compared to untreated cells (Lin et al., 2009). The striking difference between ALCL and prostate cancer is that neither *NPM1* nor *ALK* was found to be more transcriptionally active in translocation-negative ALCL cells than in non-ALCL cells; however, the breakpoint regions still featured higher levels of histone modifications associated with open chromatin (H3K4me1, H3K4me3, and H3K36me3) and lower levels of the repressive modification H3K9me3.

While we eventually focused on H3K4 methylation to test the modulation of histone modifications on translocation formation (Chapter 4), it is likely that this modification does not act alone but in combination with other modifications. For example, our computational analysis revealed that H3K27ac accompanies H3K4me1 and H3K4me3 at translocation sites, supporting a well-established correlation between hyperacetylation and H3K4 methylation (Ruthenberg et al., 2007). Similarly, the co-occurrence of

H3K4me1 and H3K27Ac is a hallmark of intergenic enhancers, which reside in domains of open chromatin to facilitate gene regulatory interactions, although enhancer regions are typically devoid of H3K4me3 (Smallwood and Ren, 2013). Additionally, H3K4 methylation was inversely correlated with H3K9me3 at several translocation sites. Several *in vitro* and *in vivo* studies have reported that H3K4 methylation precludes H3K9 methyltransferase activity and vice versa, and the counterbalance between these two methylation states may be an evolved mechanism to separate active chromatin from inactive chromatin (Binda, 2013). While the predominance of H3K4me3 and lack of H3K9me3 was expected at the actively transcribed *NPM1* gene in our ChIP studies, we were surprised that the silent *ALK* gene exhibited higher H3K4me1 and lower H3K9me3 in ALCL cells compared to control cells. These observations support the well established cross-talk between histone modifications (Kimura, 2013).

In addition, the significant lack of H3K9me3 at *NPM1* and *ALK* breakpoints in ALCL cells, coupled with high levels of H3K4 methylation at baseline, may have allowed for the increase in H3K4 methylation induced by SET7/9 and ASH2L overexpression (Chapter 4). Interestingly, overexpression of SUV3-9 did not lead to any increase in the level of H3K9me3 at breakpoints or a significant decrease in H3K4 methylation, suggesting that H3K4 methylation precludes SUV3-9 activity at breakpoints. This may be one reason why SUV3-9-expressing cells exhibited similar break and translocation frequency to GFP-expressing cells, rather than causing a decrease as expected by the decreased breakage frequency by an endonuclease (Chapter 3). Indeed, SET7/9 was recently shown to methylate SUV3-9 *in vivo*, resulting in its down-regulation of H3K9

methyltransferase activity, heterochromatin relaxation, and genomic instability (Peters et al., 2001; Wang et al., 2013).

We tested the effect of histone modifications on DSB formation in response to genotoxic stress. When challenged by an endonuclease (Chapter 3) we find increased DSB formation in H3K4me1 and H3K4me3 domains. H3K4me3 has been implicated in DSBs generated by endonucleases during class switch and V(D)J recombination processes in lymphocytes (Daniel and Nussenzweig, 2012). Off-target effects of the RAG and AID endonucleases have been suggested to be responsible for several tumor translocations in naïve B- and T-cells, and AID has been implicated in prostate cancer translocations (Lin et al., 2012). Both RAG2 and AID have been shown to target extrachromosomal sites highly enriched in H3K4me3 in a sequence-independent manner (Ji et al., 2010; Shimazaki et al., 2009; Stanlie et al., 2010). Interestingly, these studies found that the levels of H3K4me3 at endogenous Ig gene segments as well as extrachromosomal sites were even higher than in promoter regions of transcribed genes, suggesting aberrant accumulation of H3K4me3. While the RAG2 endonuclease has a PHD domain that enhances its interaction with H3K4me3, AID does not, and the most likely mode of action for elevated H3K4me3 to facilitate DSBs and translocations is via chromatin decondensation allowing access of the endonuclease (Mostoslavsky et al., 2003).

In line with this interpretation, tethering of ASH2L or SETD7 to the lac array led to its visible decondensation, along the lines of tethering chromatin remodellers such as BRG1 or VP16 (Tumbar et al., 1999). Decondensation of chromatin by H3K4me3 may also

facilitate DSB formation by irradiation. Indirect evidence for this comes from studies that found increased DSBs after globally decondensing chromatin by histone deacetylase inhibition (Camphausen and Tofilon, 2007) or by placing nuclei in hypotonic conditions and subjecting them to irradiation (Warters and Lyons, 1992; Takata et al., 2013). It has been suggested that chromatin decondensation results in more hydrated chromatin, allowing for more free radical formation and DSBs (Falk et al., 2010). A major limitation to assessing whether local chromatin domains at a gene level could be more or less sensitive to radiation-induced DSBs is the lack of suitable markers for DSBs *in vivo* as well as the observation that γ H2AX preferentially localizes to euchromatic regions minutes after irradiation (Falk et al., 2010). Of note, one study used immuno-FISH to quantitate γ H2AX foci at a defined region of transcribed genes, presumably with high H3K4 methylation content, and a gene-poor region and found more γ H2AX formation at the transcribed locus after irradiation (Falk et al., 2008). Taken together, these observations suggest that H3K4 methylation may promote DSB formation by irradiation by influencing chromatin decondensation.

Histone modifications may affect translocation formation at various stages of the process. Our observation of increased breakage frequency of multiple gene loci in several cell lines in response to elevated H3K4 methylation levels, as measured by the separation of break-apart DNA FISH probes (Chapter 4), suggests that a major contribution of histone modifications is in the formation of persistent breaks. This may occur via increased susceptibility of decondensed chromatin to DNA damage agents as discussed above. Alternatively, histone modifications and chromatin decondensation may affect the DNA

damage response (DDR) resulting in accumulation of persistent DNA breaks. Chromatin structure changes dynamically in response to DSBs and it has been proposed that chromatin condensation is necessary for efficient DDR activation (Ayrapetov et al., 2014; Khurana et al., 2014; Burgess et al., 2014). Along with others, we recently showed that chromatin rapidly decondenses prior to this condensation step in the response to DNA damage (Kruhlak et al., 2006; Delleire et al., 2009; Ayrapetov et al., 2014; Khurana et al., 2014; Burgess et al., 2014). However, in ASH2L-expressing cells, condensation at the site of DNA damage as well as upstream DDR signaling is dampened (Burgess et al., 2014) and so is the recruitment of upstream DDR factors such as γ H2AX and 53BP1 (Chapter 3). This is line with the observation that γ H2AX foci formation is associated with H3K4 hypomethylation after irradiation (Seiler et al., 2011; Lafon-Hughes et al., 2013). The implication of these findings is that H3K4 hypermethylation slows down DDR by counteracting condensation required for efficient DDR, and this leads to the persistence of DSBs. However, DDR impairment by H3K4 hypermethylation is likely only attenuated and not total, since, since the illegitimate joining of DSBs to form translocations requires DNA repair.

Several lines of evidence have established that the spatial proximity of translocation partners influences translocation probability (Schwartz and Hakim, 2014). In cytogenetic studies, numerous correlations exist between the spatial proximity of chromosome and genes and their translocation frequency in a cell-type and tissue-type specific manner (Roukos et al., 2013). These correlations were substantiated by genome-wide studies that mapped chromosome breaks and translocations in the absence of selection, and found

that translocations occur primarily between proximal loci (Chiarle et al., 2011; Klein et al., 2011; Zhang et al., 2012; Rocha et al., 2012). In line with a role of spatial proximity, we find in translocation-negative ALCL that *NPM1* and *ALK* co-localize at a higher frequency than in non-ALCL T-cell lymphomas, and in comparison to several control genes (Mathas et al., 2009). Upon elevation of H3K4 methylation we find significantly increased levels of translocations between *NPM1* and *ALK* in FEPD cells, where the two loci are in close spatial proximity, whereas no such increase was observed in Jurkat cells (Chapter 4), where the two loci are more distally located (Mathas et al., 2009). Of course, not all proximal genes translocate with each other and *NPM1-ALK* remains the most prevalent translocation in ALCL. Using hiBA-FISH (Chapter 5), we observed that *Nanog* and *CycA* genes were more proximal to each other in FEPD cells as *NPM1* and *ALK* (data not shown); however, the translocation frequency between these two genes was significantly less. This was likely due to the lower incidence of breaks at *Nanog*, as indicated by separation of break-apart probe. While the proportion of breaks forming translocations was higher for *Nanog* due to spatial proximity, translocation frequency remained low because break frequency was significantly lower (Chapter 4).

Histone modifications may also affect the localization of potential translocation partners in 3D space or influence the mobility of persistently broken chromosome ends and thus their ability to form translocations (Roukos et al., 2014). Chromatin decondensation by VP16 at lac arrays (Chuang et al., 2006), and by INO80 targeting to a specific locus (Neumann et al., 2012), were shown to induce long-range motion in mammalian cells. In contrast, DSB-containing chromatin domains experienced reduced motion when cells

were treated with histone deacetylase (HDAC) inhibitors or histone acetylase inhibitors (Krawczyk et al., 2012). In our experiments using hiBA-FISH, we did not observe a significant difference in the proximity between translocation partners *NPM1* and *ALK* in cell lines overexpressing histone-modifying enzymes (Chapter 4). This suggests that histone modifications do not significantly affect the localization or mobility of the translocation partners and further supports the notion that the major effect of histone modifications is on influencing breakage frequency.

A long-standing view of translocation formation has been that translocations are spontaneous, albeit low frequency, events that occur at random genomic loci, and their enrichment in a given tumor type is due to selection of proliferative or survival advantages. While *in vivo* selection clearly contributes to the observed prevalence of recurrent translocations, it is likely that cell-intrinsic properties contribute to the susceptibility of a given chromosome or genome region to break and undergo translocations. Here, we suggest that the local chromatin environment, specifically histone modifications, predisposes genome regions to breakage and translocation formation. Understanding the mechanism by which altered levels of histone modifications occur will have significant impact on our still remarkably limited understanding of how chromosome translocations form and may have important clinical implications, particularly with regards to understanding cancer susceptibility and predisposition.

FUTURE DIRECTIONS

Our data make several general points. First, they identify the epigenetic landscape as a key molecular player in chromosome breakage, translocation formation and genome stability. Second, they begin to address the long-standing question of what molecular features determine the non-randomness of chromosomal break sites. Third, the involvement of epigenetic marks as determinants of breakage suggests a significant influence of intrinsic and extrinsic cellular stimuli on susceptibility of a genome region to breakage and translocation. Overall, our results represent the first direct evidence that histone modifications are a driving factor in determining genomic break- and translocation-sites and as such have considerable implications for our understanding of genome stability and translocation formation. These results should be extended by more in-depth investigation along the following lines:

1. Utilization of new cell-based technologies to study formation of recurrent translocations: In Chapters 3 and 4, we utilize cell-based assays to provide the first direct evidence that histone modifications influence chromosome breakage and translocations. The lac array system in U2OS cells allowed us to study breakage within specific histone modification domains without the capacity to study translocation formation. The overexpression experiments in Chapter 4 allowed us to study translocation formation in ALCL and non-ALCL T-cells through global overexpression. These experiments could be improved with a system that allows for testing of both breakage and translocation formation *in vivo* in the same cells. Recently, TALEN and CRISPR/Cas9 technology has been used to induce DSBs at the recurrent breakpoints in

NPM1 and *ALK* genes (Piganeau et al., 2013; Ghezraoui et al., 2014). The advantage of these systems is that chromatin modifying enzymes can be overexpressed or knocked down globally or in a site-specific manner by CRISPR/Cas9-mediated tethering (Hsu et al., 2014). Moreover, breakpoint-specific DSBs can be induced by CRISPR/Cas9 nucleases, and primers specific for nuclease breaks can be developed to measure translocation frequency by quantitative PCR analysis. Finally, new droplet digital PCR methods, which are able to detect very low levels of fusion transcripts (Shuga et al., 2013), can be used to assay for rare *NPM1-ALK* translocation events.

2. Investigate the role of histone modifications in translocation-associated DSB repair: Here, we studied the one of the most upstream factors in translocation formation, chromosome breakage. Equally as important to the formation of a translocation is the repair of broken chromosome ends. It is likely that chromatin modulation by histone-modifying enzymes influences DSB repair pathway choice, kinetics, and fidelity, as suggested by our preliminary data in *yH2AX* and *53BP1* recruitment at DSB sites in Chapter 3. Translocations are thought to be the result of deregulated non-homologous end joining (NHEJ), rather than homologous recombination (HR), repair pathways (Zhang et al., 2010). We can test repair pathway choice by assessing the recruitment of pathway-specific repair proteins in cells with overexpression or knockdown of histone-modifying enzymes (Jakob et al., 2011). Possible experiments include the immunostaining of repair factors and live-cell imaging with tagged-repair factors, for example after laser microirradiation (Burgess et al., 2014). These experiments should be

performed with dead enzyme mutant controls to rule out non-chromatin driven effects of these histone-modifying enzymes.

3. Genome-wide mapping of chromosome breaks and translocations: In our breakage and translocation assay, we specifically assayed for *NPM1* and *ALK* breakage and *NPM1-ALK* translocations, as well as a set of control genes. However, the global overexpression of histone-modifying enzymes likely affects the genome-wide landscape of breakage and translocations. During the course of this study, several genome-wide techniques such as ligation-mediated break quantification technology (BLESS; Crosetto et al., 2013) and translocation sequencing (HTGTS or TC-seq; Chiarle et al., 2011; Klein et al., 2011) were developed. The advantage of these techniques is that they can survey breaks and translocations occurring in different DNA sequence and chromatin environments. Therefore, the relative role of DNA sequence and chromatin—such as heterochromatin, euchromatin, and transcriptionally active or inactive regions—in breakage and translocation formation can be studied. The super-imposition of chromatin conformation capture studies (3C, 4C, 5C, or hiC) would also provide 3D genomic context to breakage and translocations by mapping of these events to defined topological domains (Zhang et al., 2012; Hakim et al. 2012; Rocha et al. 2012).

4. High-throughput screening of enzymes and chemical inhibitors that modulate breakage and translocation frequency: Here, we provide evidence that overexpression of H3K4 methyltransferases increases chromosome breakage and translocation frequency using a high-throughput assay (hiBA-FISH, Chapter 5). This assay can be adapted to systematic high-throughput screenings of siRNAs/shRNAs or chemical inhibitors using

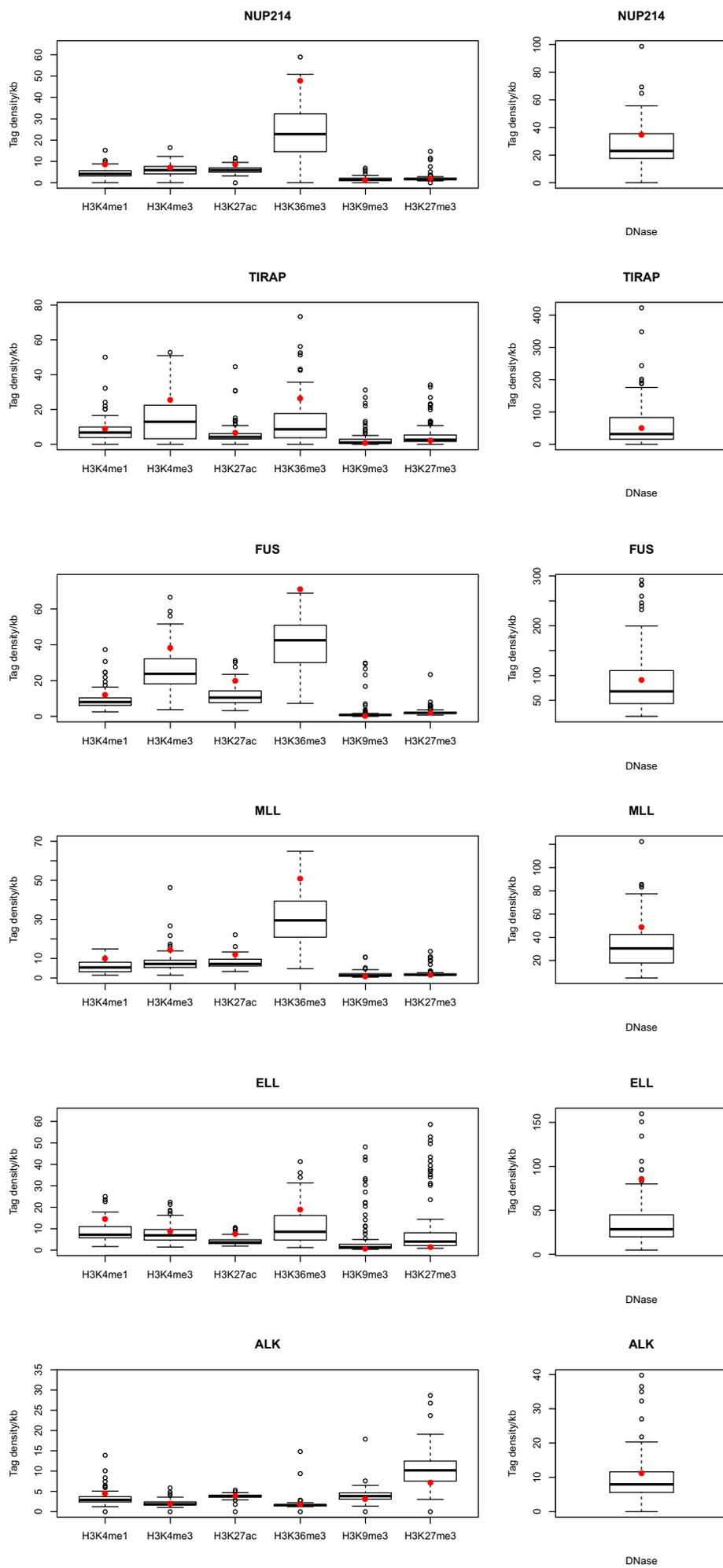
96- or 384-well plates. For example, there are chemical libraries containing ~80 chemical epigenetic inhibitors (Cayman Chemicals) and chromatin-focused siRNA libraries of ~300 siRNAs (Dharmacon) consisting of histone-modifying enzymes, chromatin remodelers, histone chaperones and chromatin adaptor proteins. Since many chromatin-modifying enzymes have analogous or redundant counterparts, these studies would help establish whether families of enzymes or specific enzymes modulate breakage and translocation frequency. The identification of factors can also guide the development of assays to test chemical inhibitors as possible cancer therapies that may be useful in decreasing secondary translocations and tumorigenicity, a major problem in radiation and chemo-therapy. Interestingly, at the time of this study, an inhibitor to SET7/9, one of the enzymes implicated in translocation formation using our assay, was described (Wagner and Jung, 2012). Finally, the proposed high-throughput studies can be extended to non-chromatin related factors.

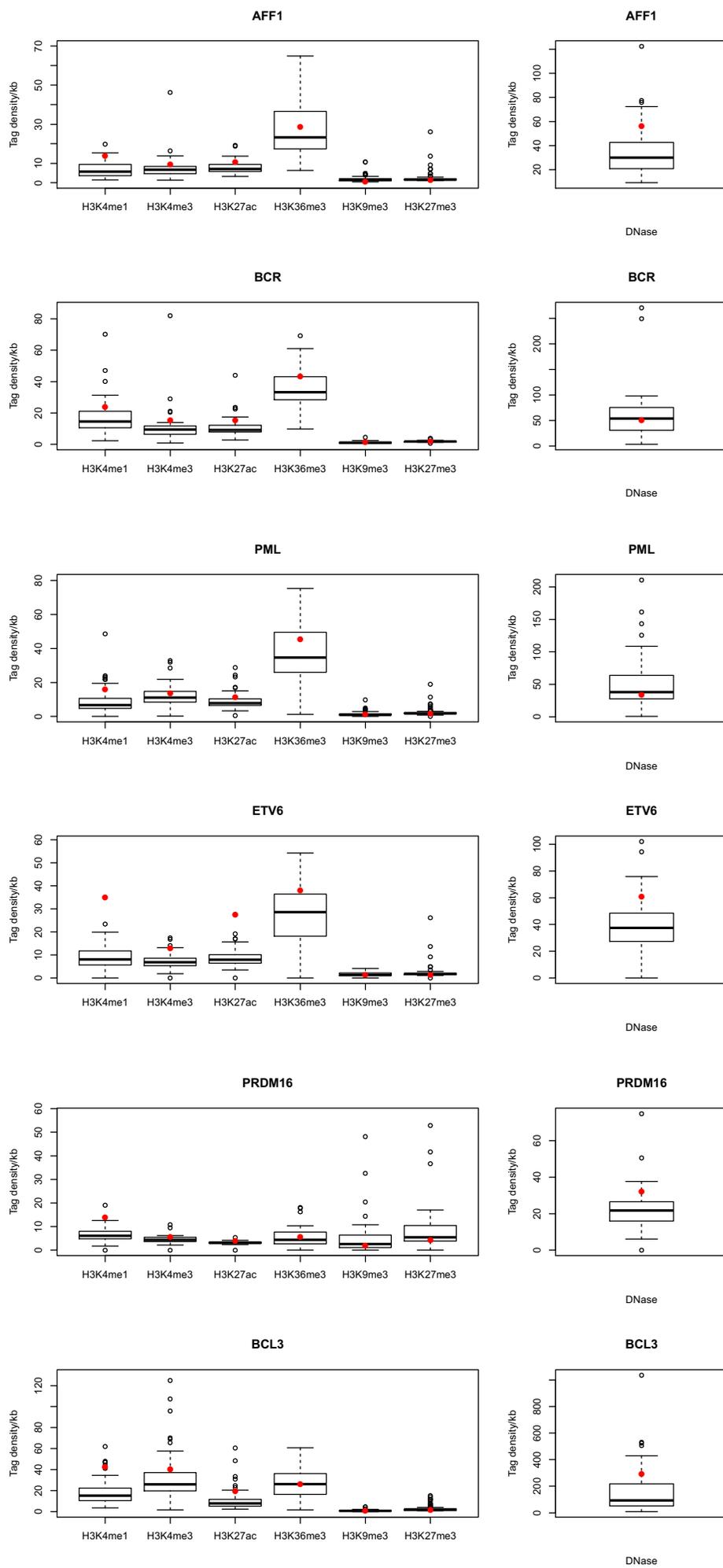
Appendix 1

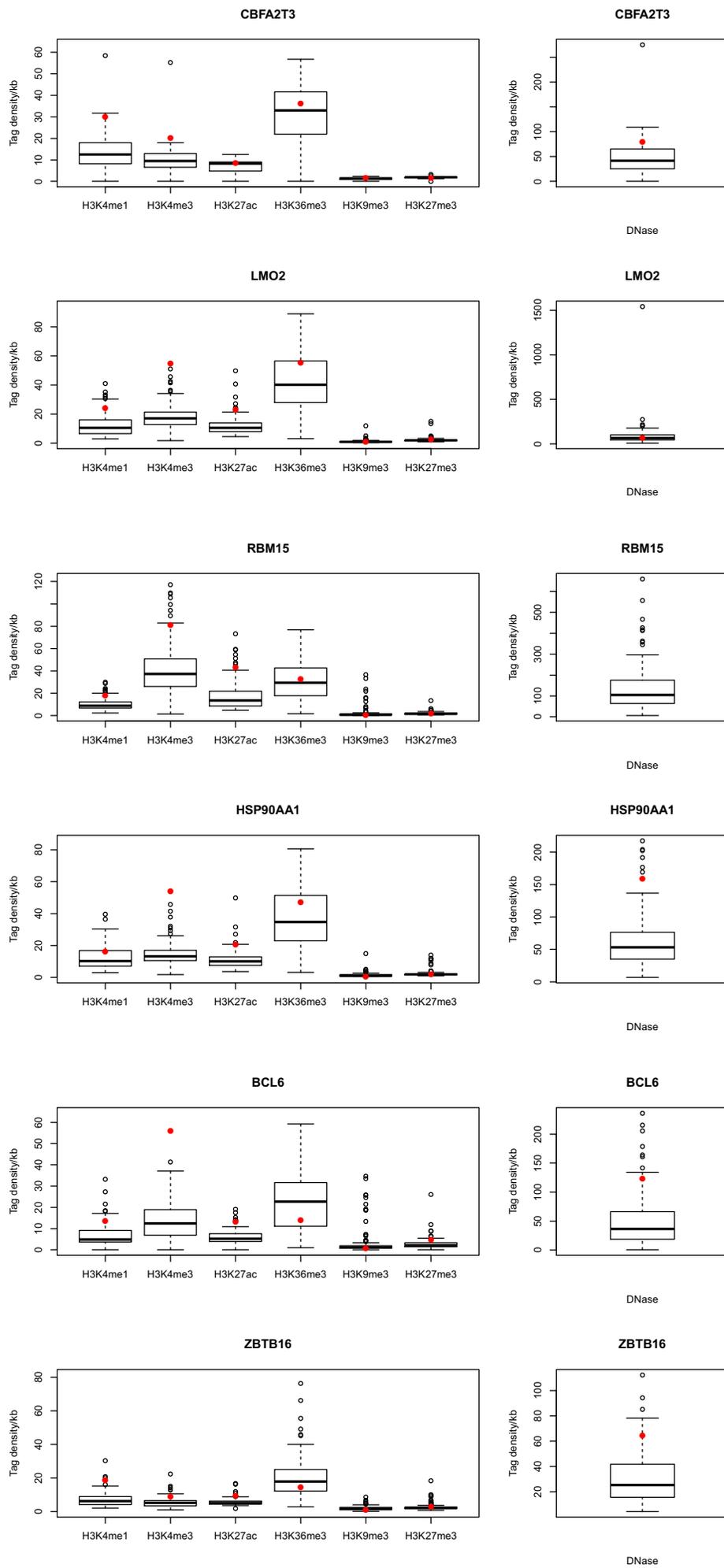
Individual gene and breakpoint computational analyses

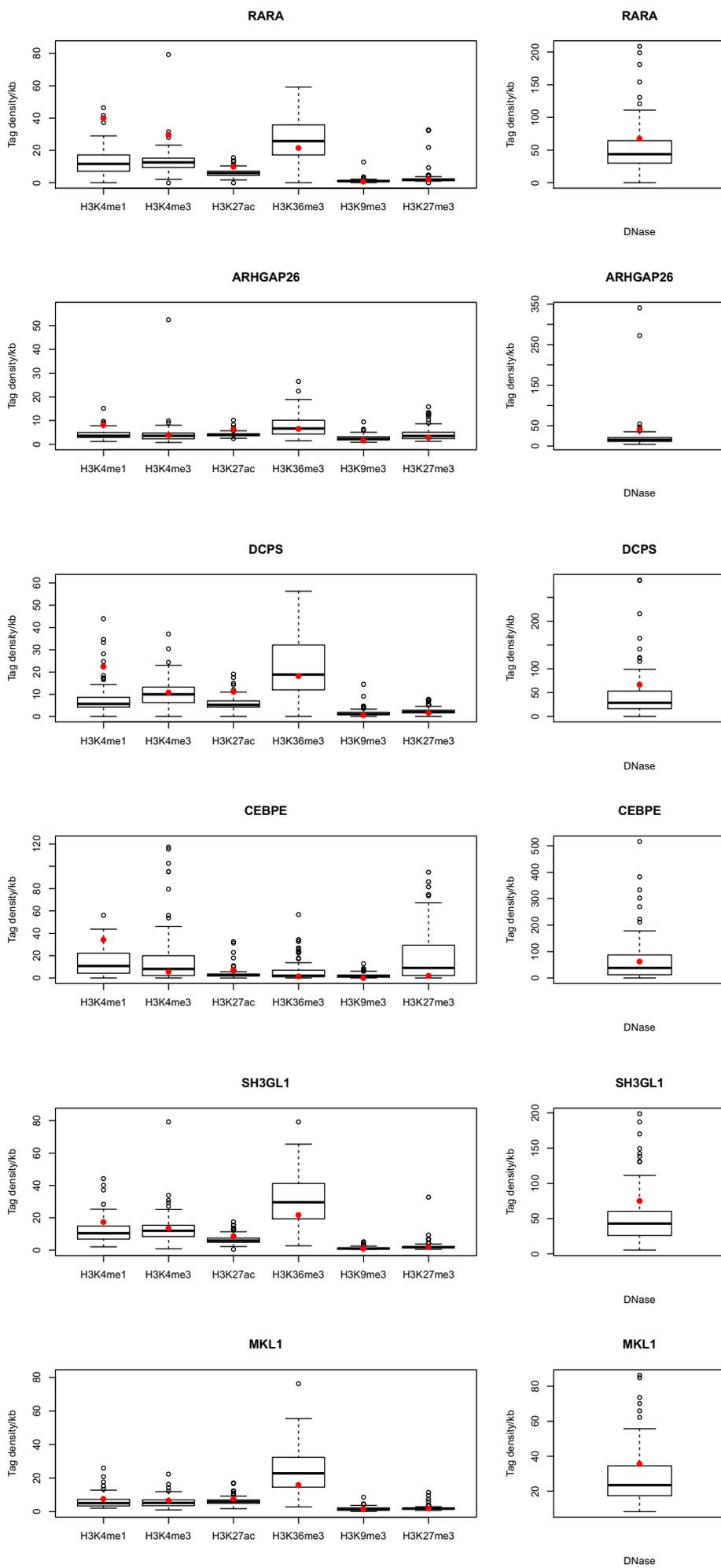
Collaborators: Zhuzhu Z. Zhang (University of North Carolina/Salk Institute for Biological Studies) and Jason D. Lieb (University of North Carolina/University of Chicago)

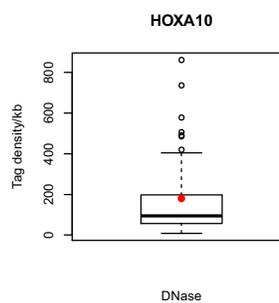
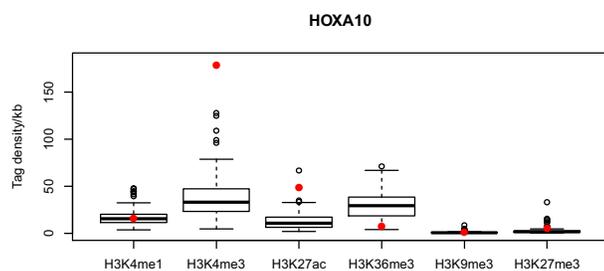
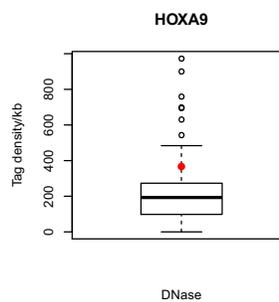
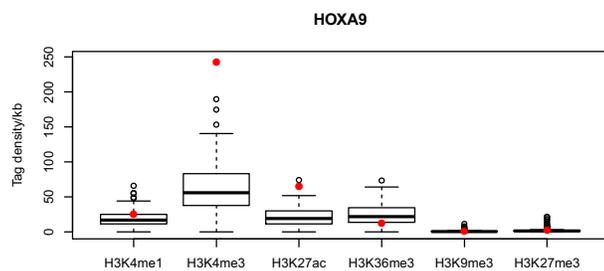
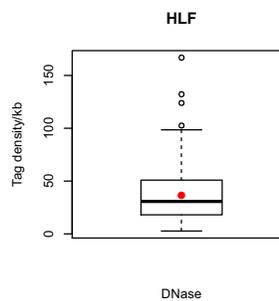
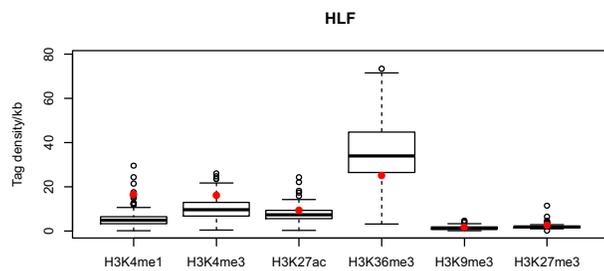
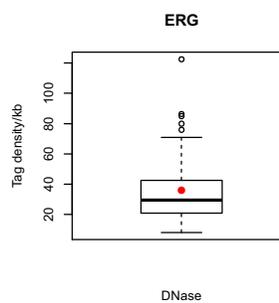
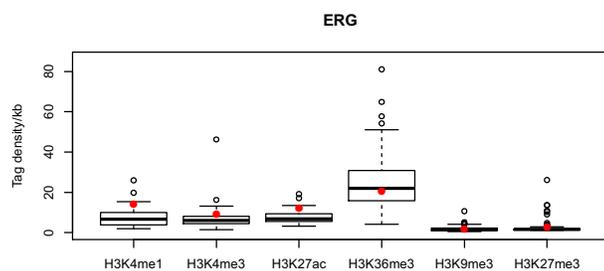
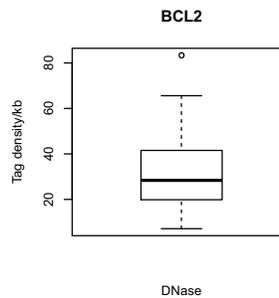
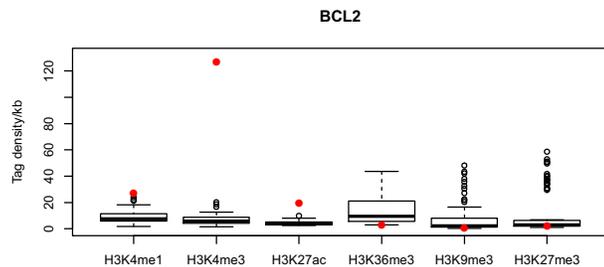
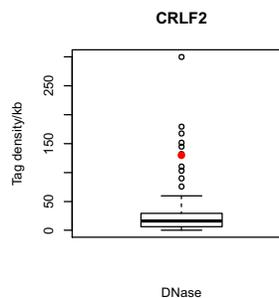
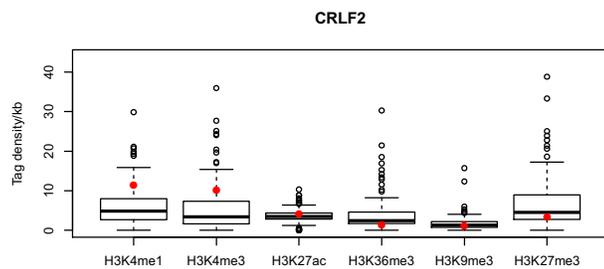
This appendix provides extended analyses for Chapter 2 Figures 5 and 8.

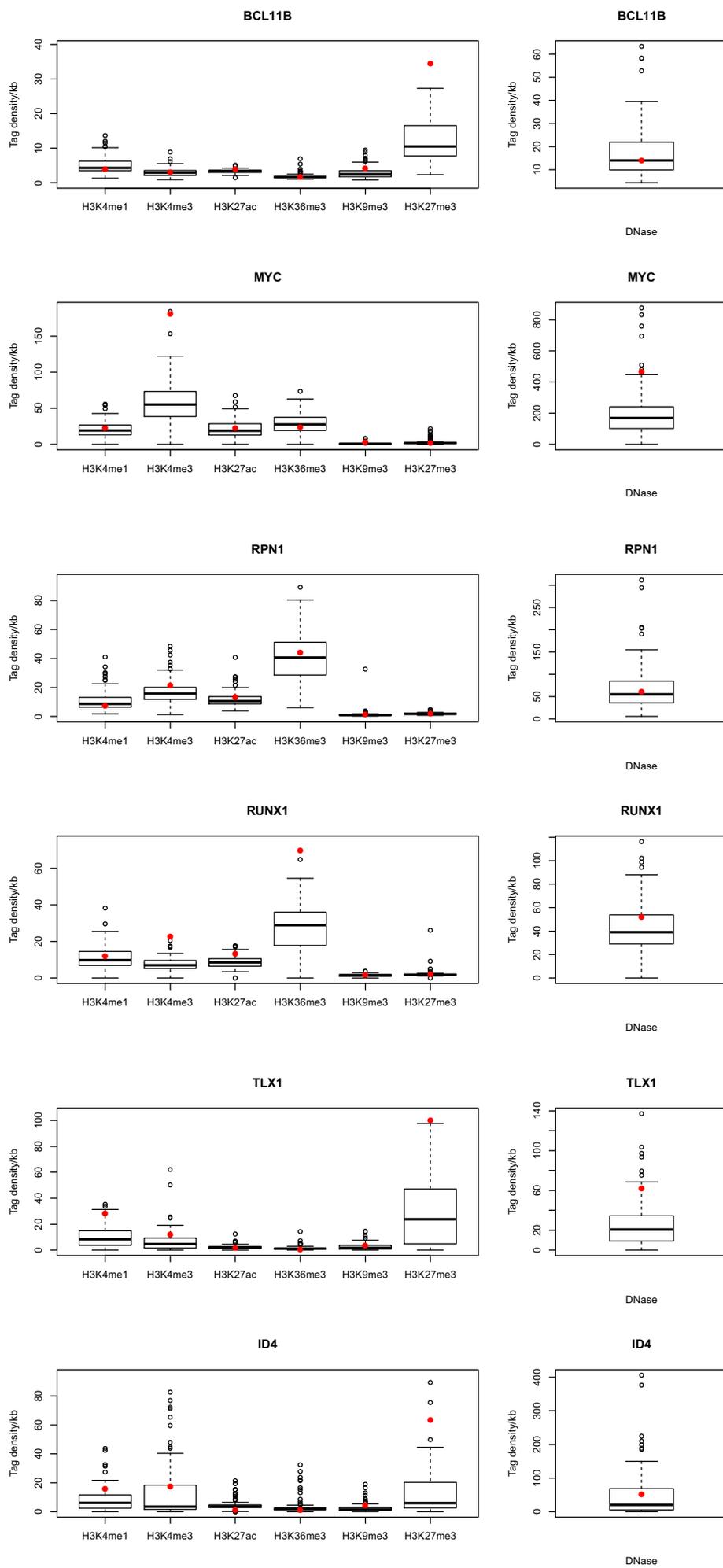


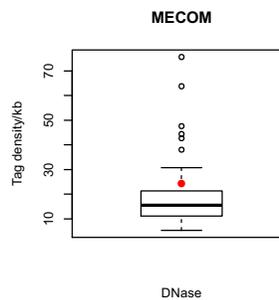
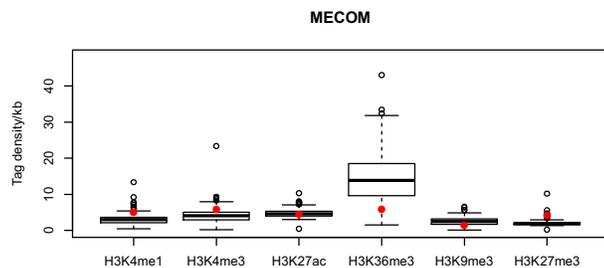
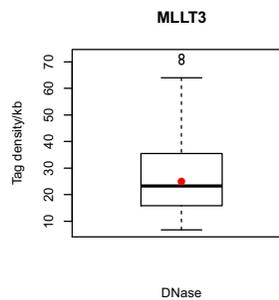
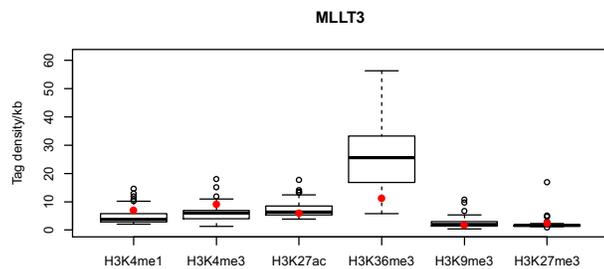
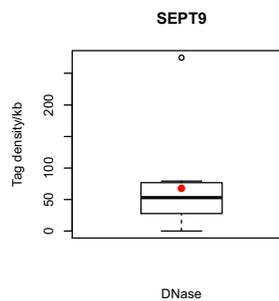
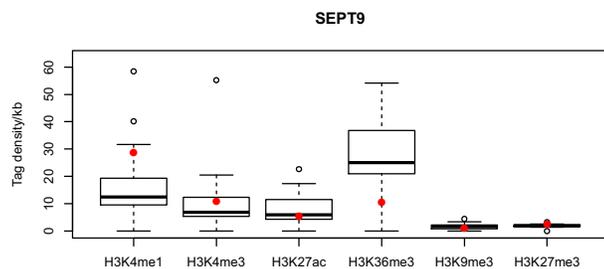
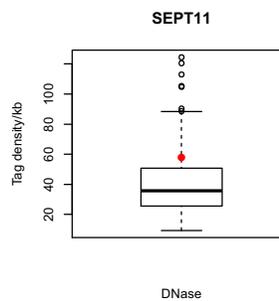
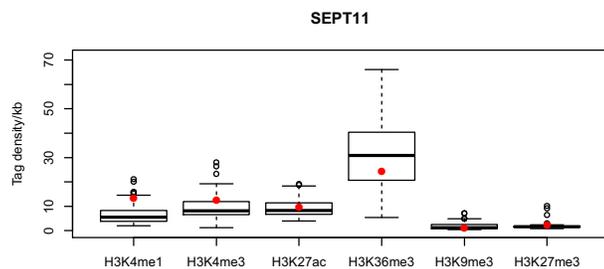
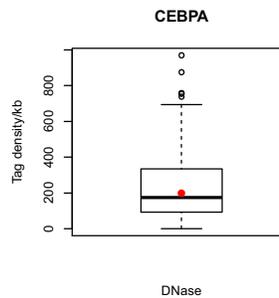
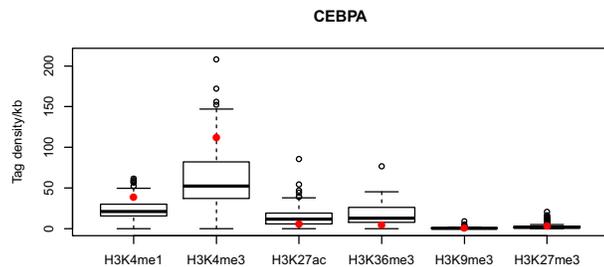
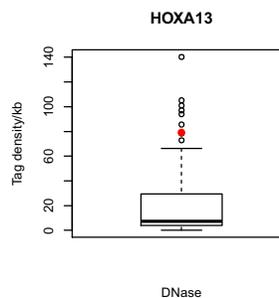
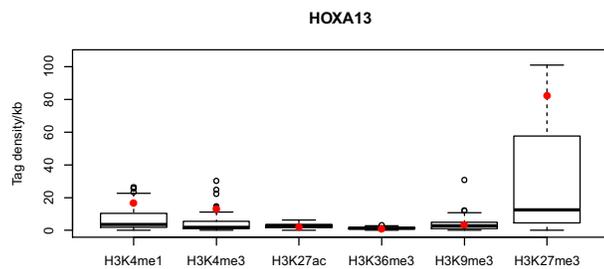


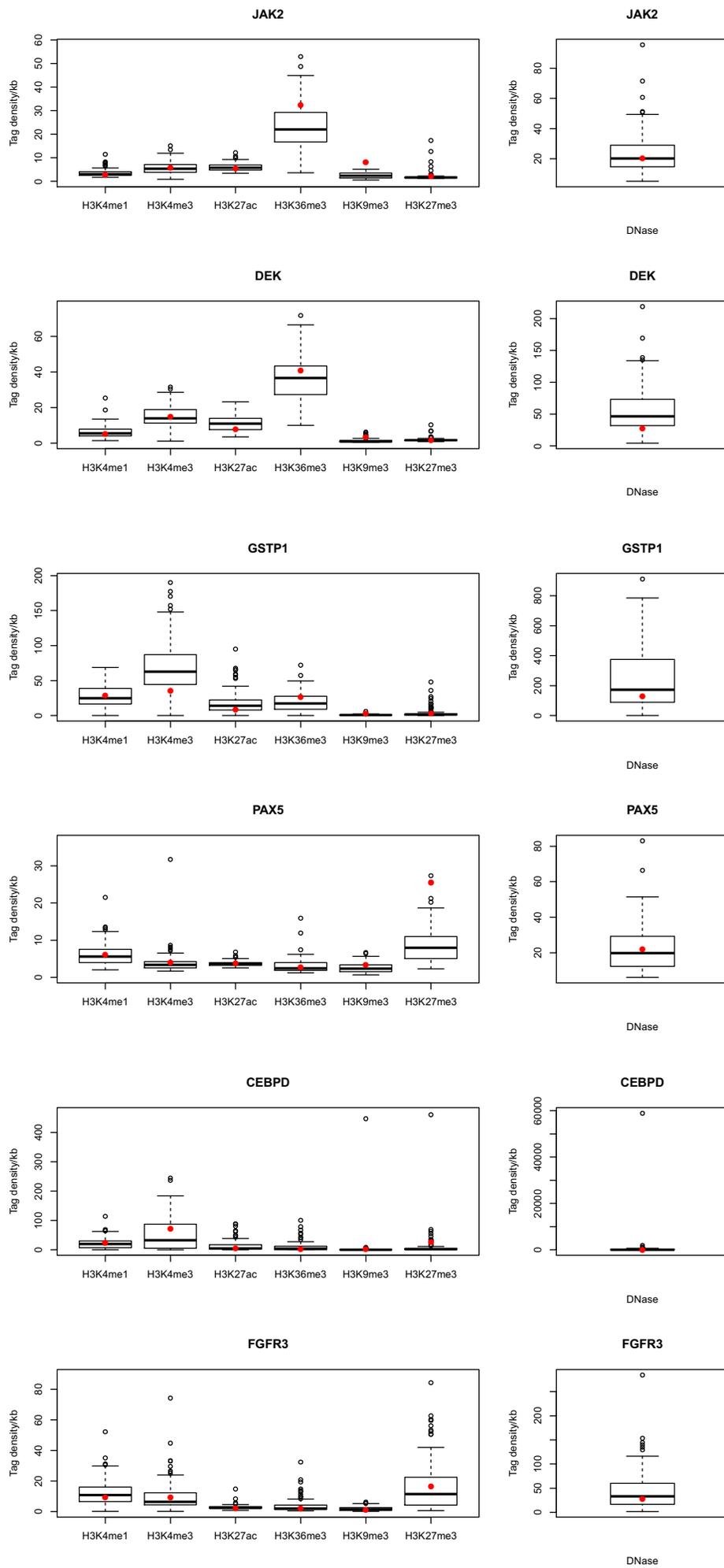


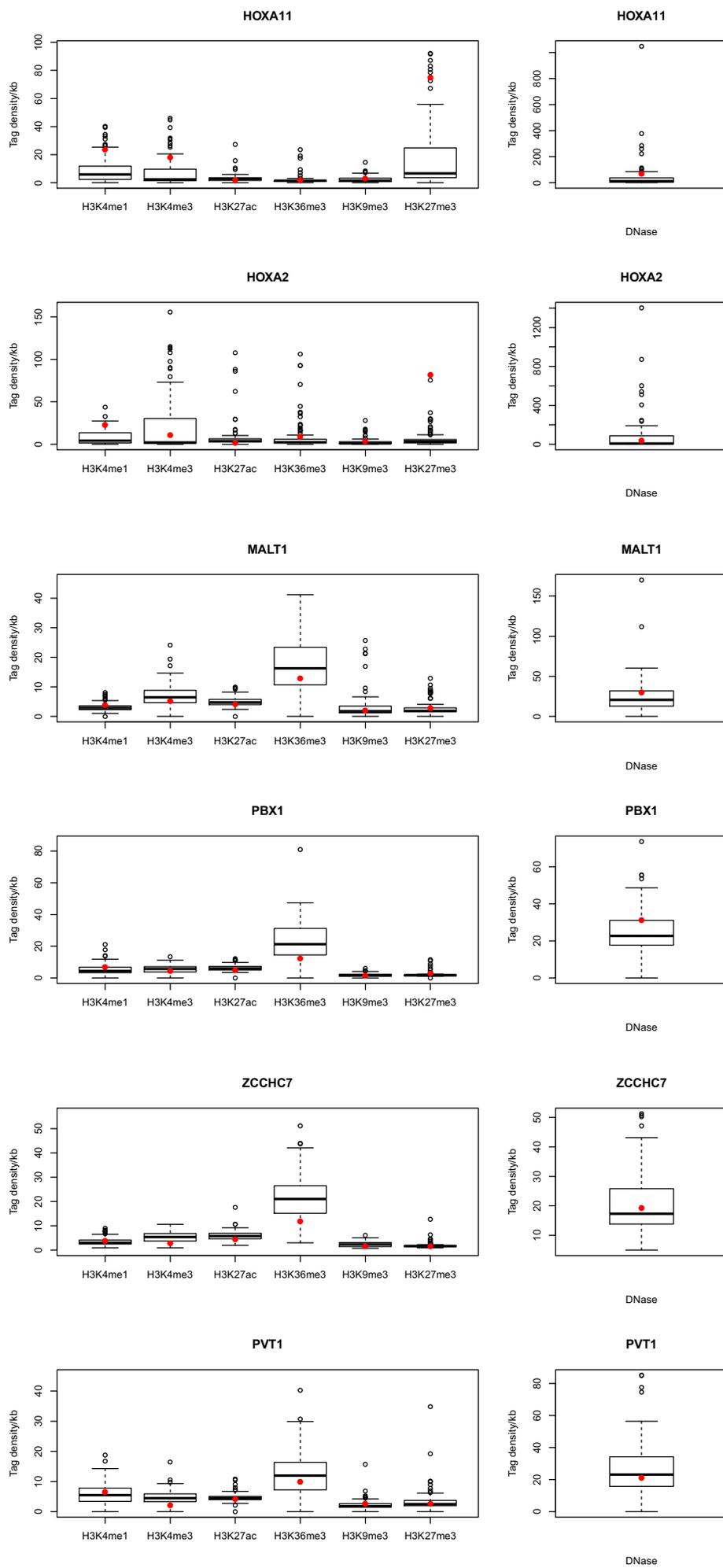


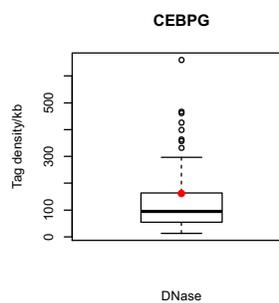
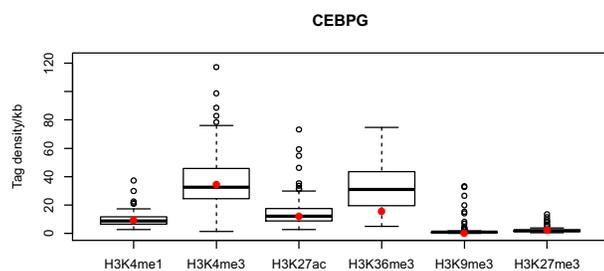
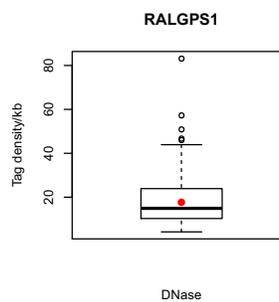
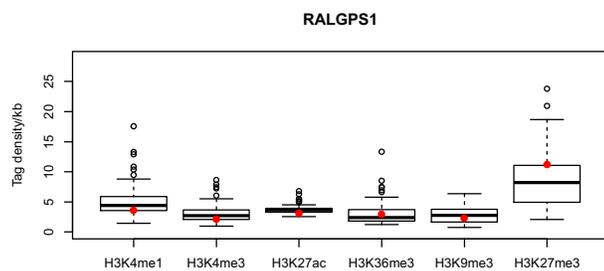
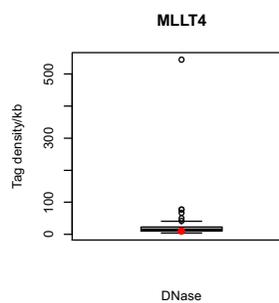
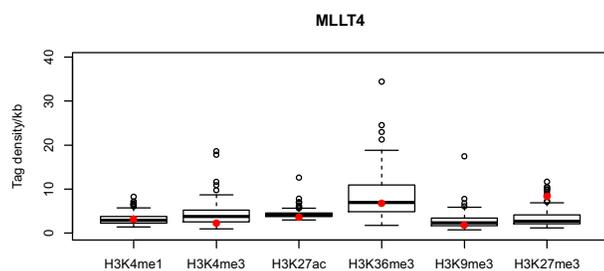
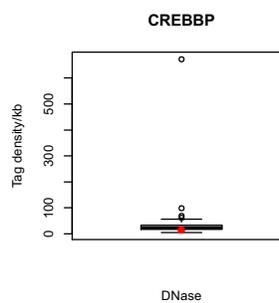
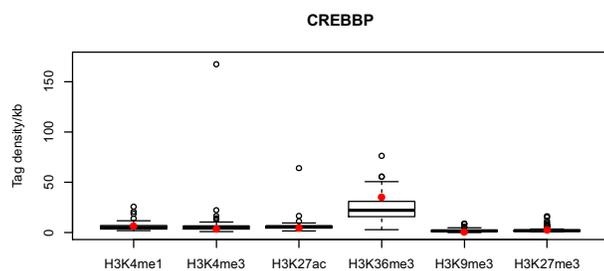
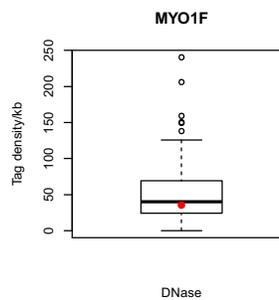
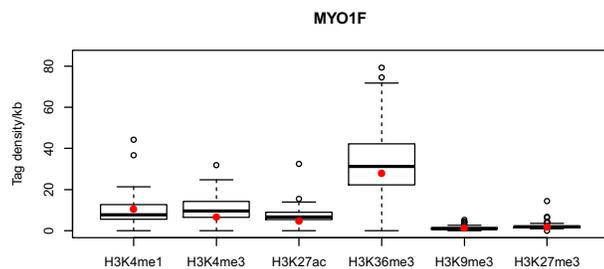
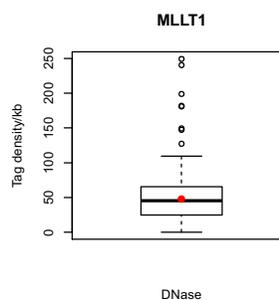
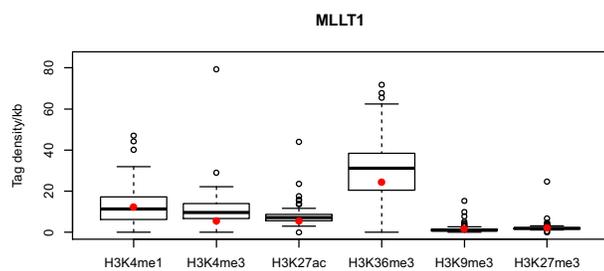


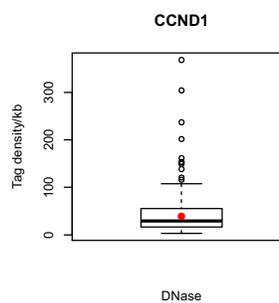
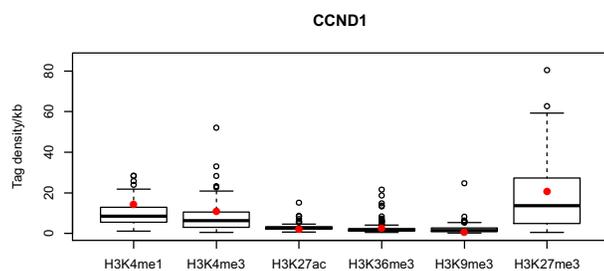
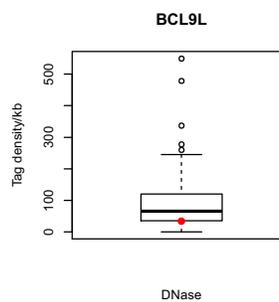
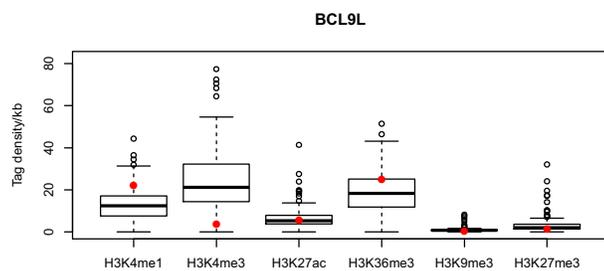
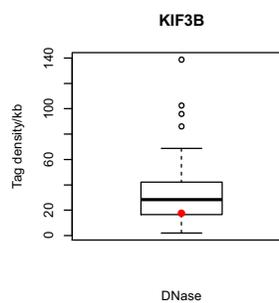
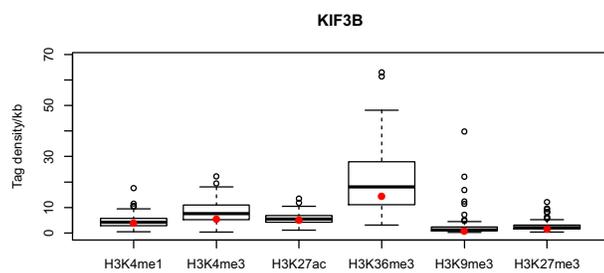
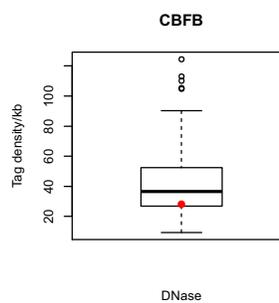
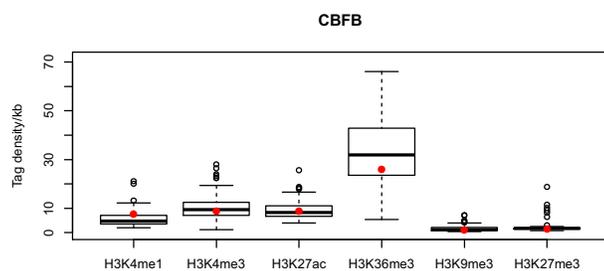
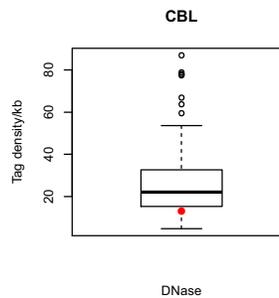
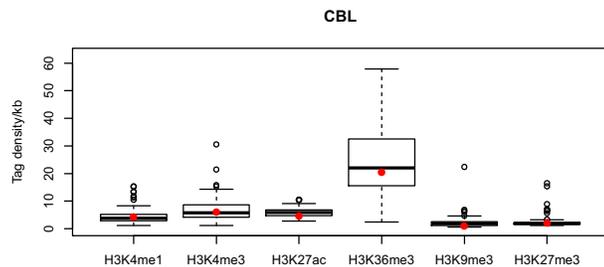
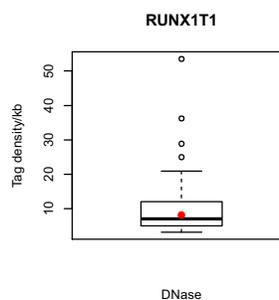
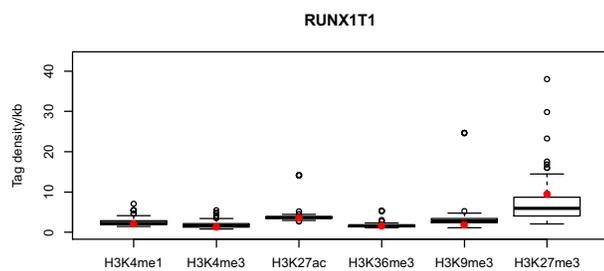


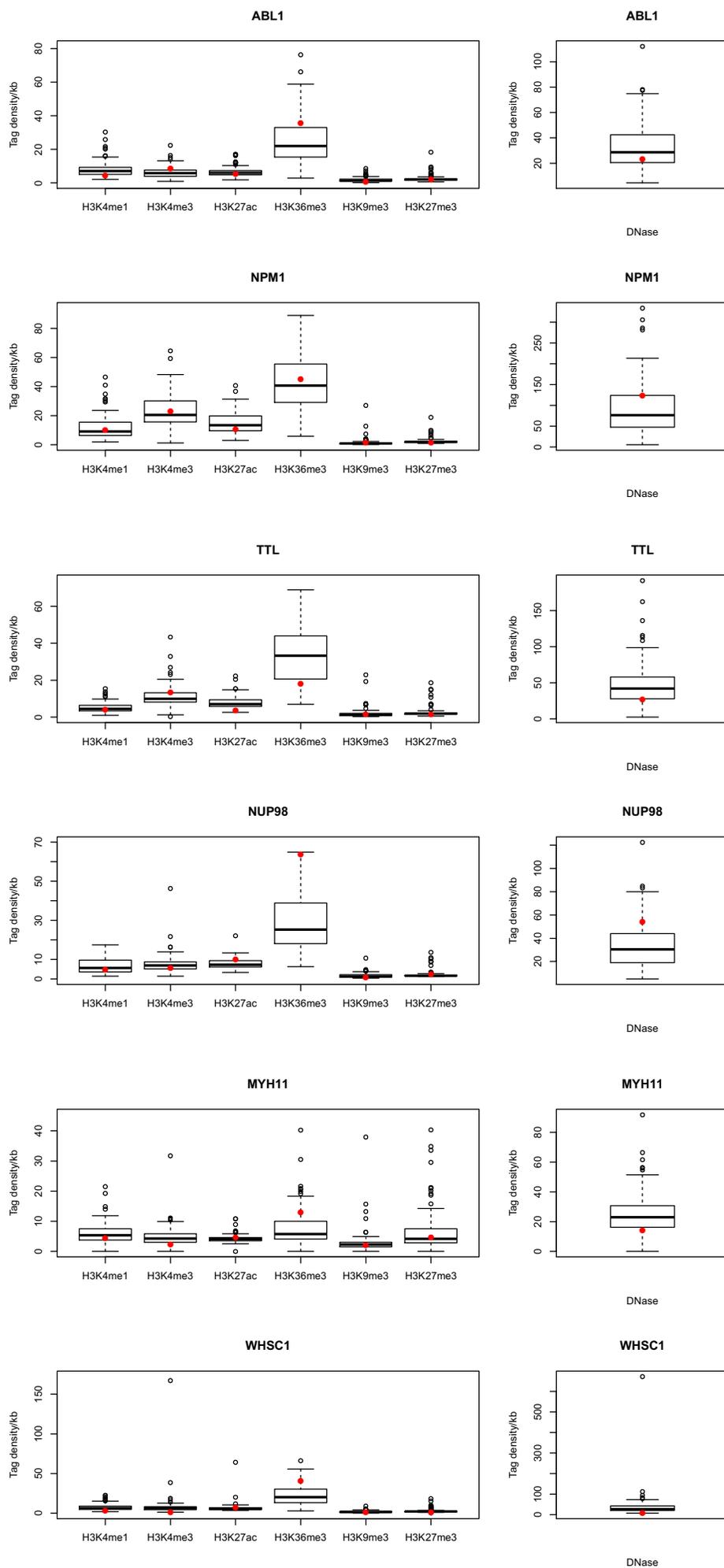


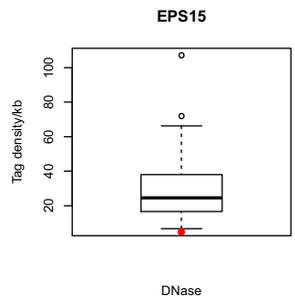
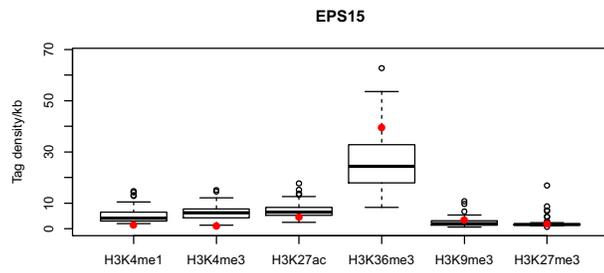
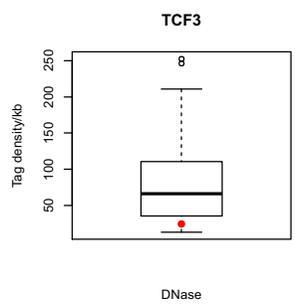
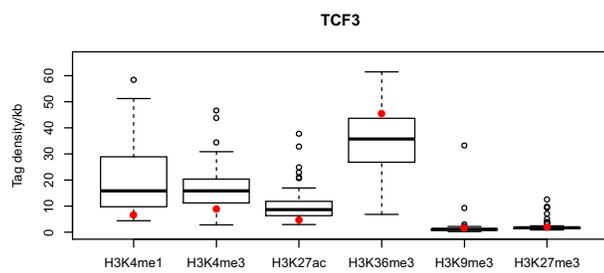


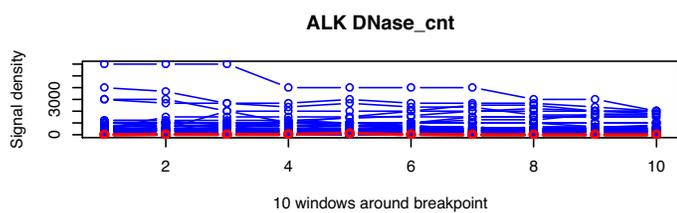
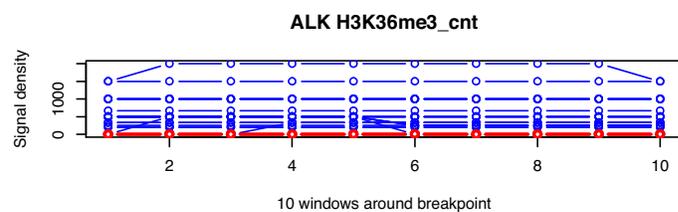
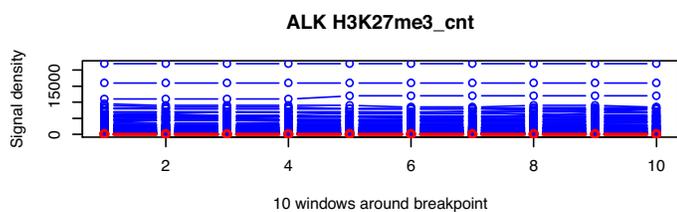
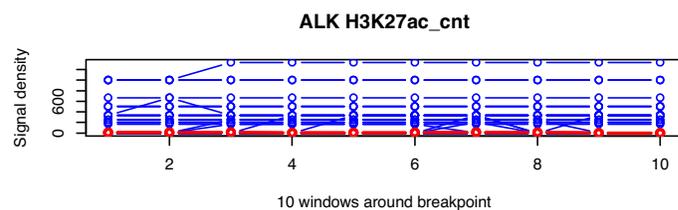
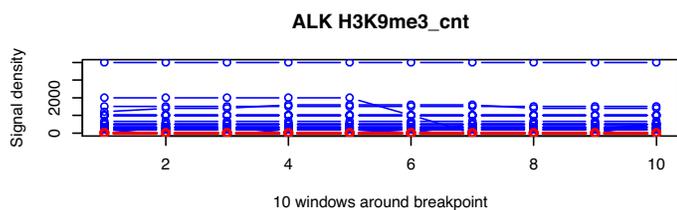
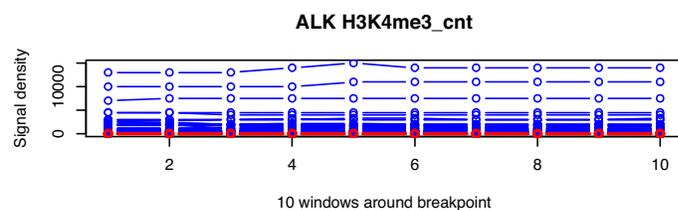
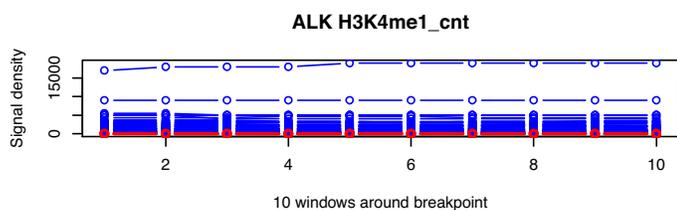
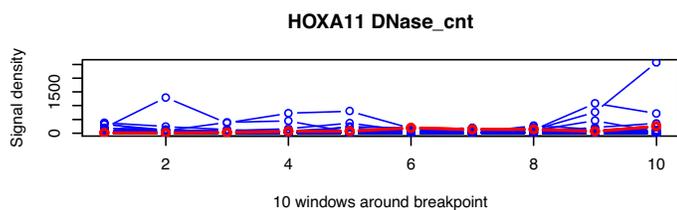
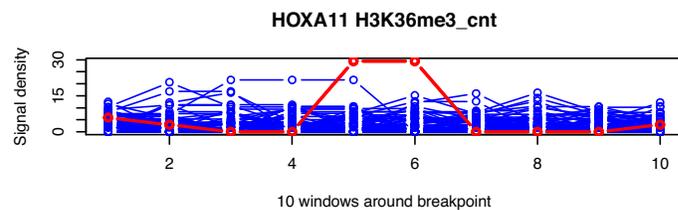
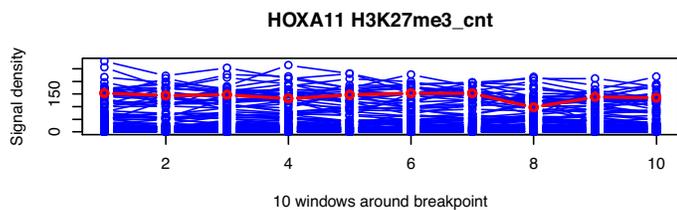
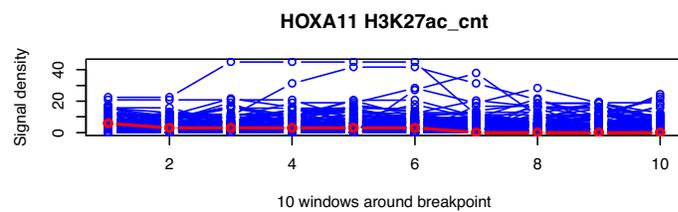
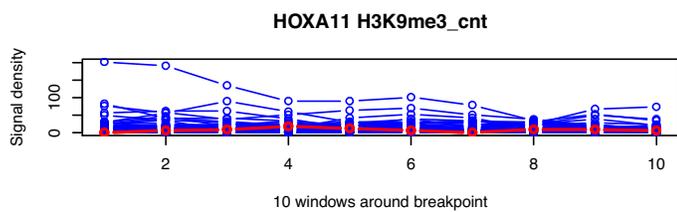
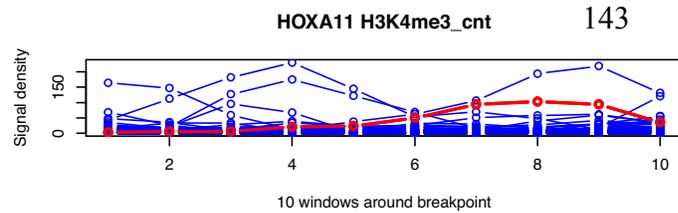
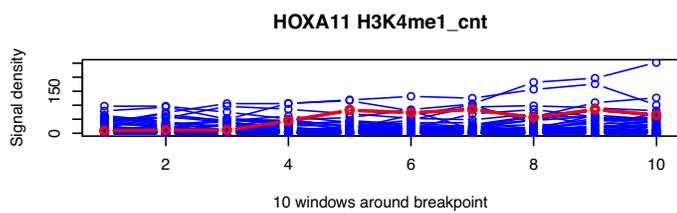


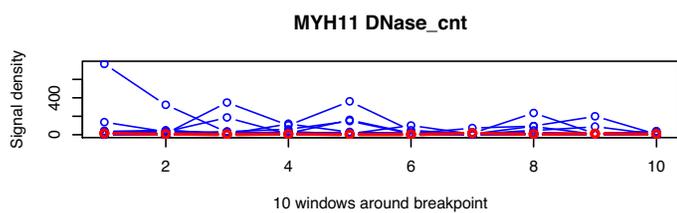
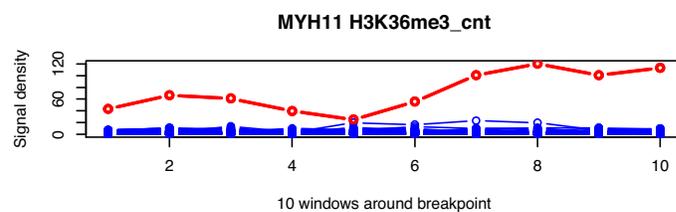
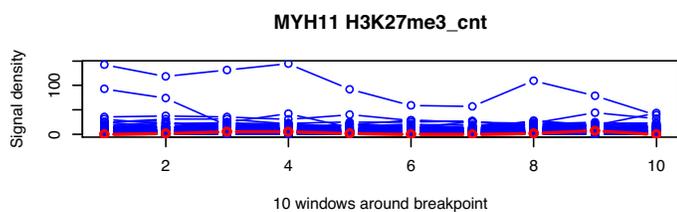
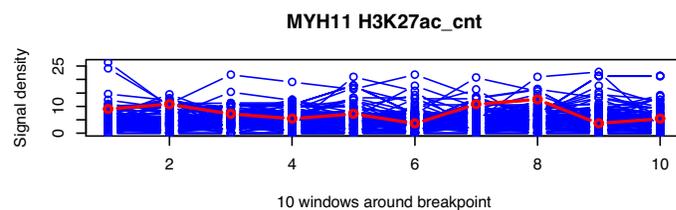
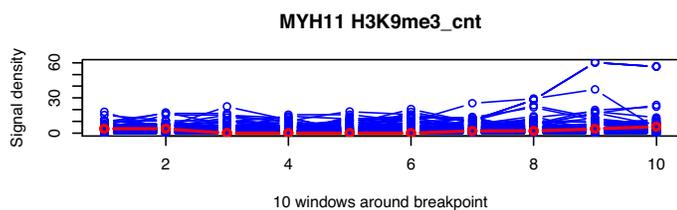
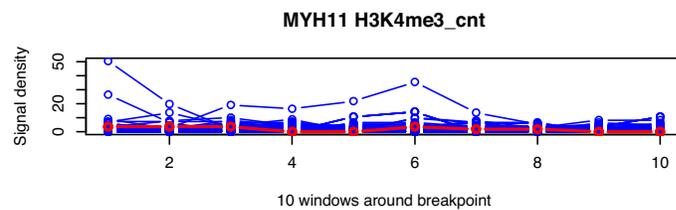
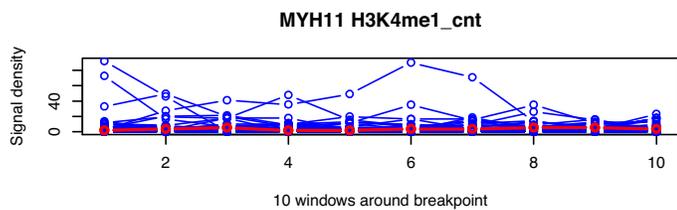
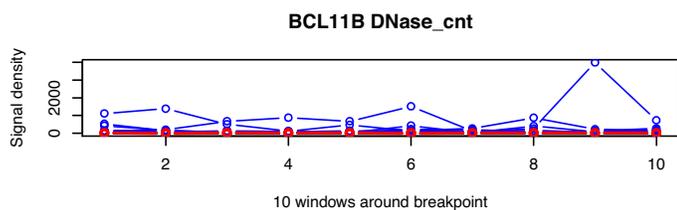
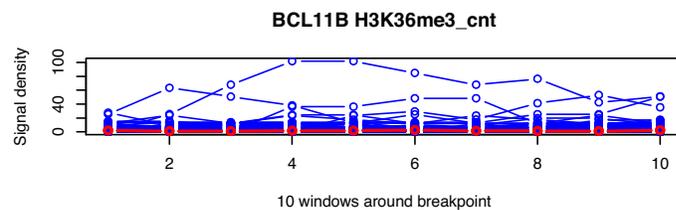
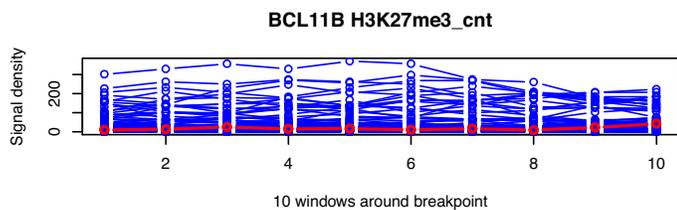
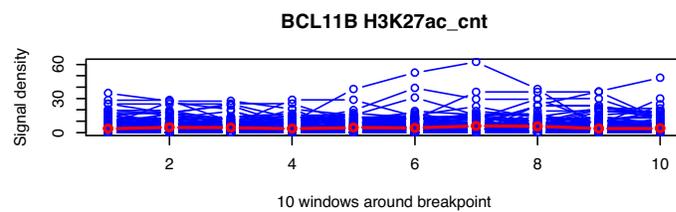
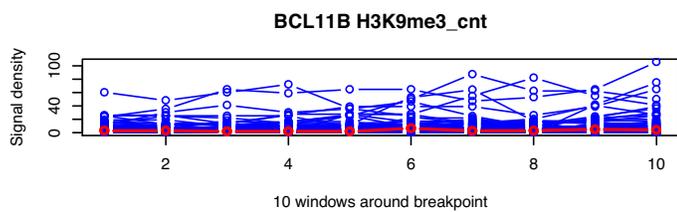
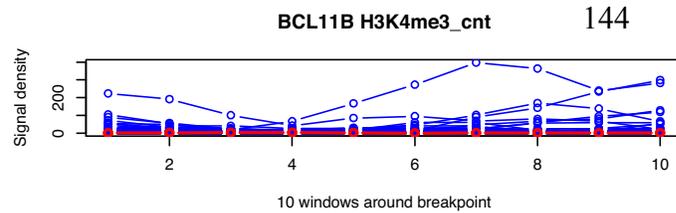
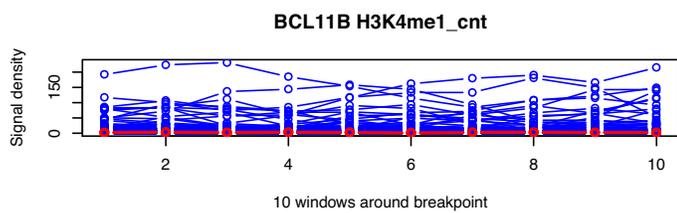


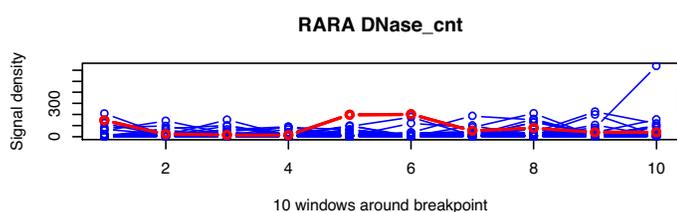
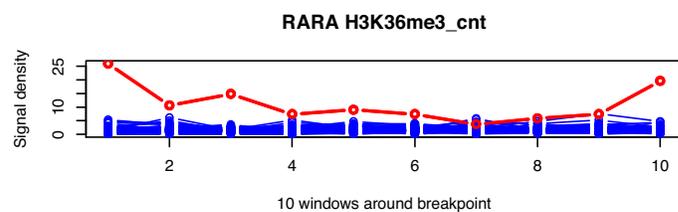
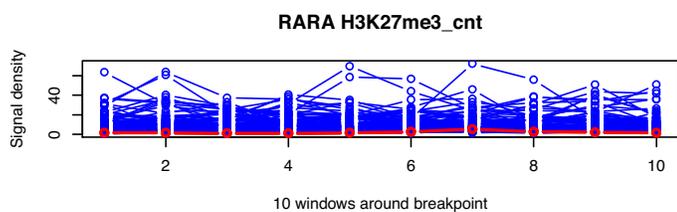
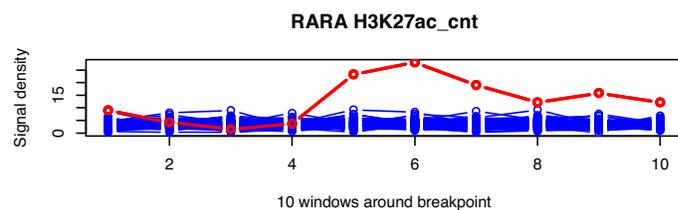
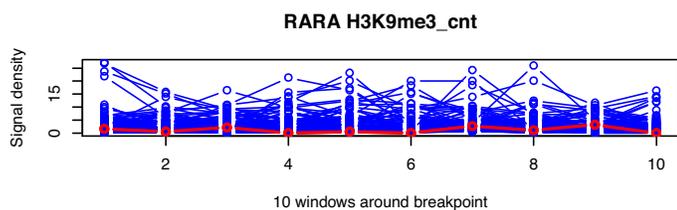
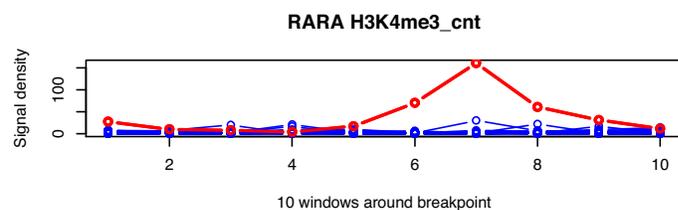
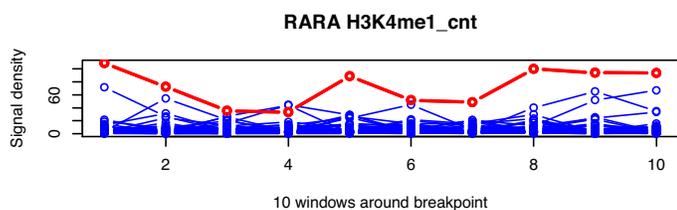
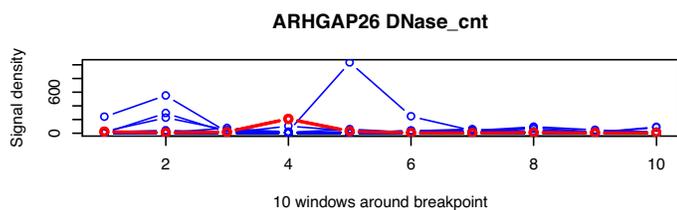
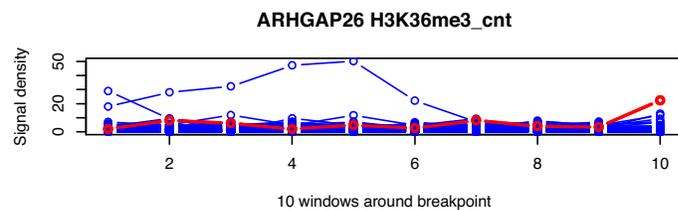
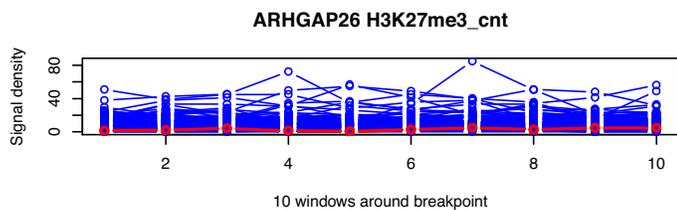
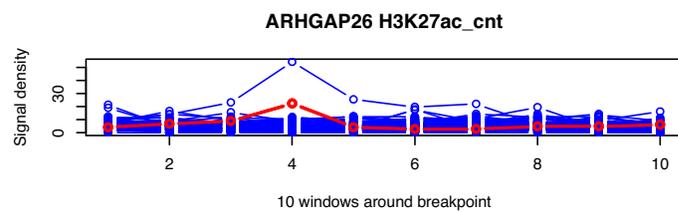
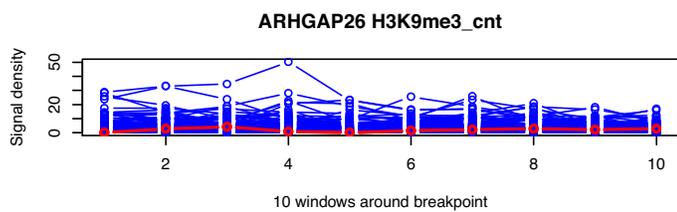
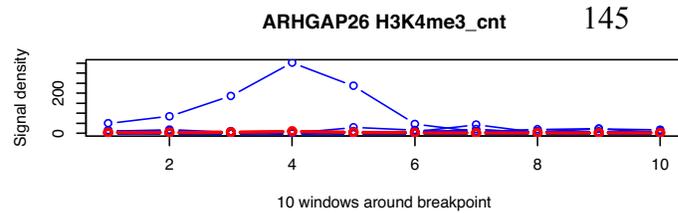
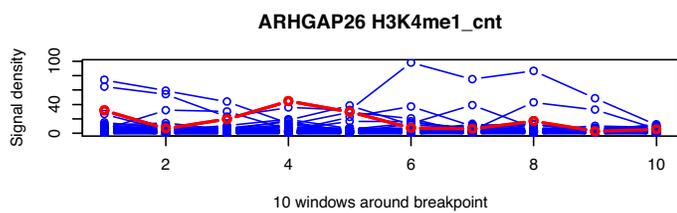


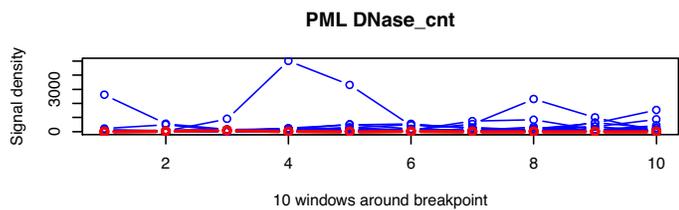
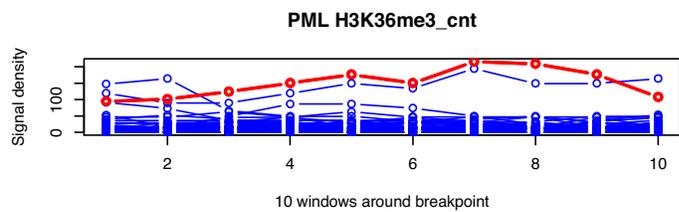
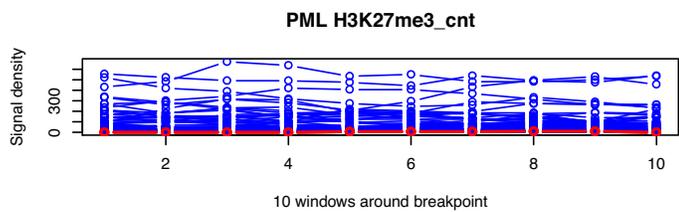
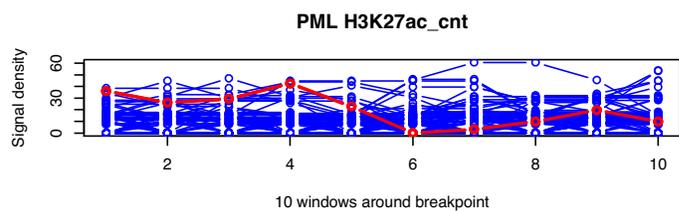
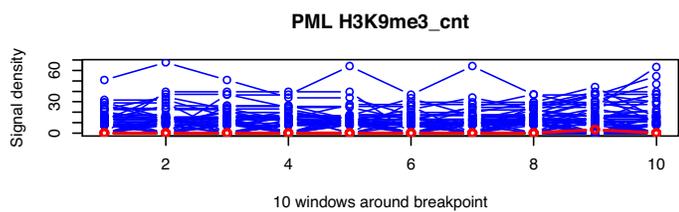
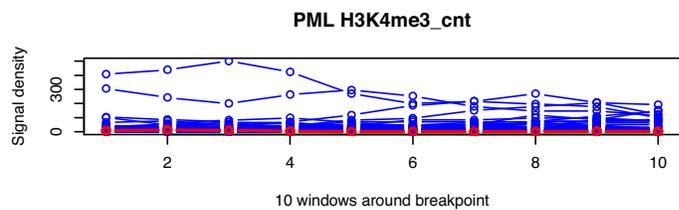
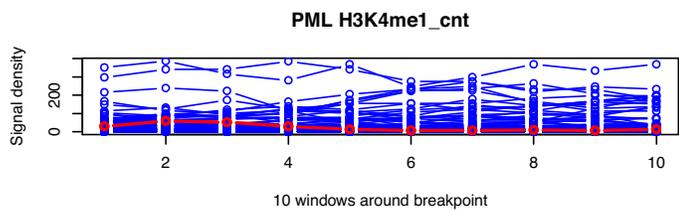
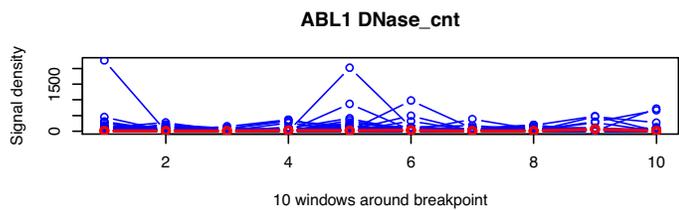
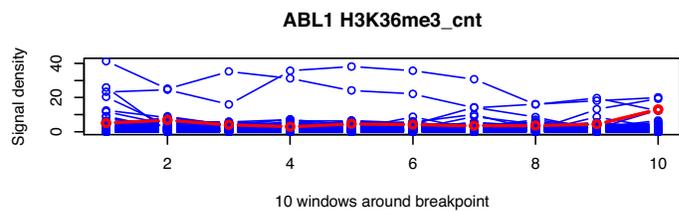
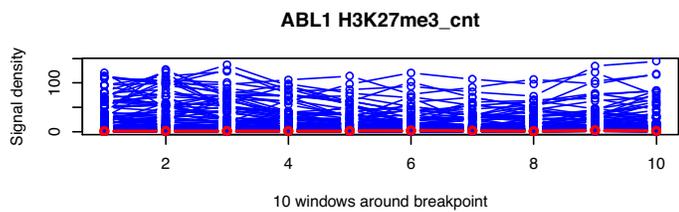
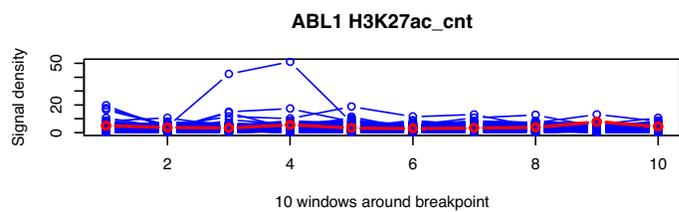
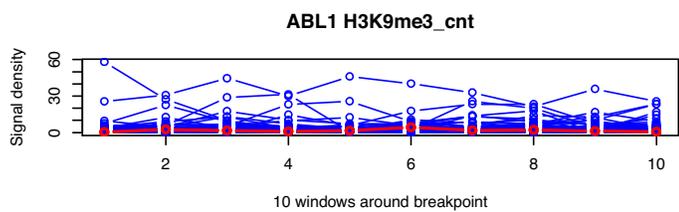
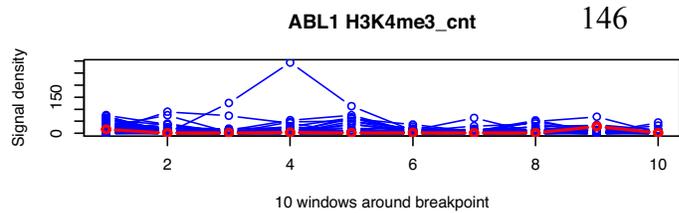
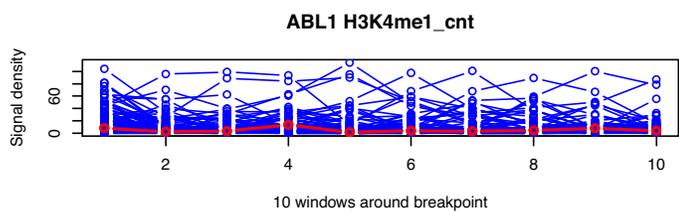


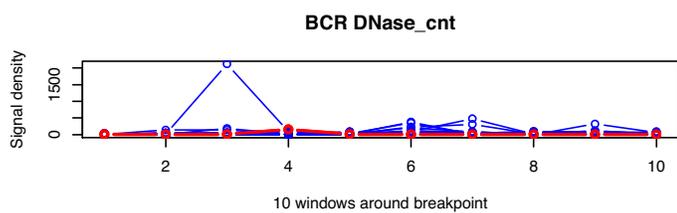
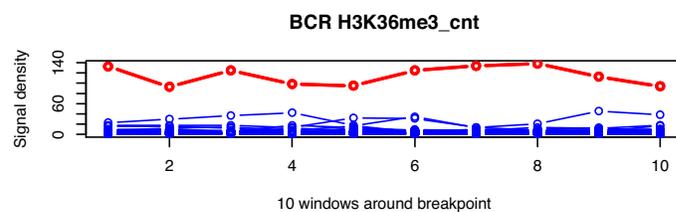
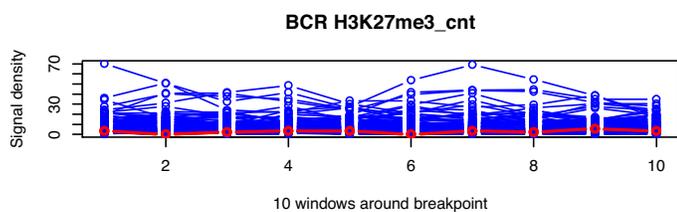
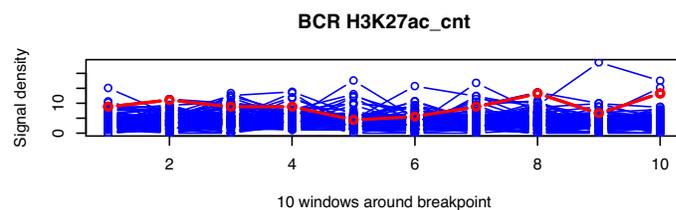
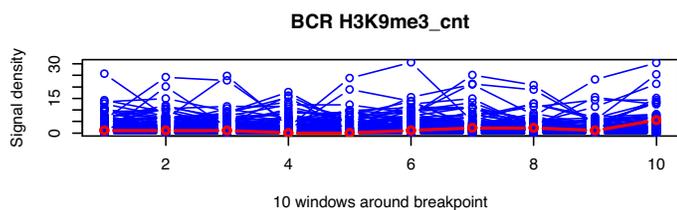
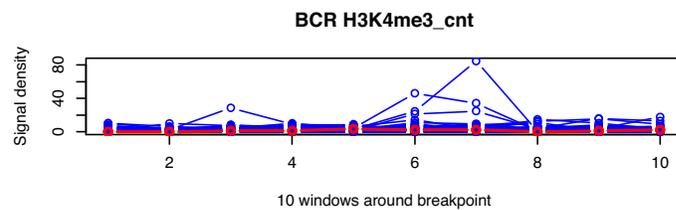
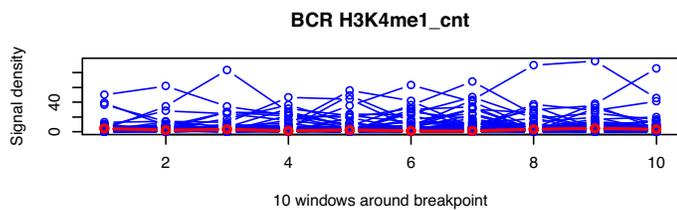
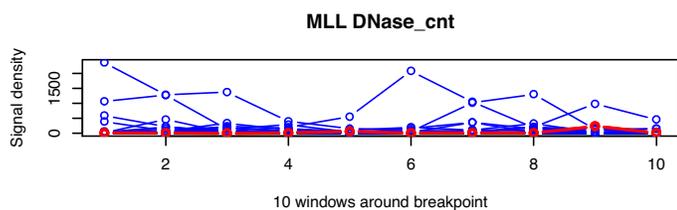
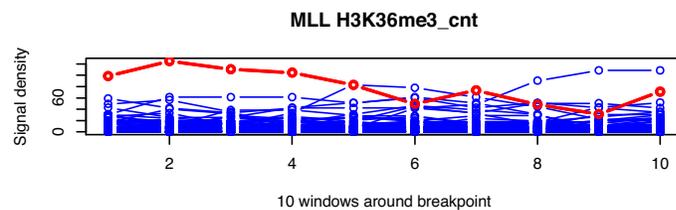
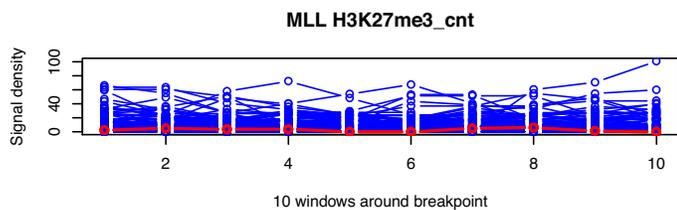
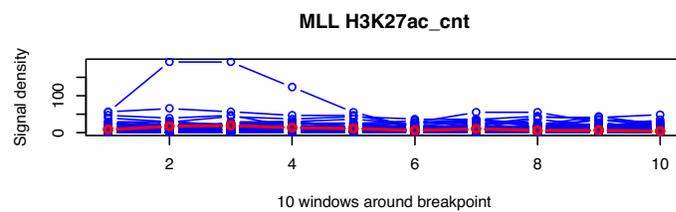
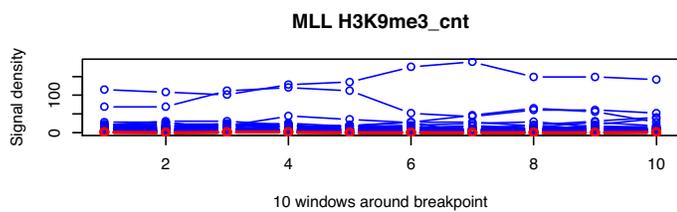
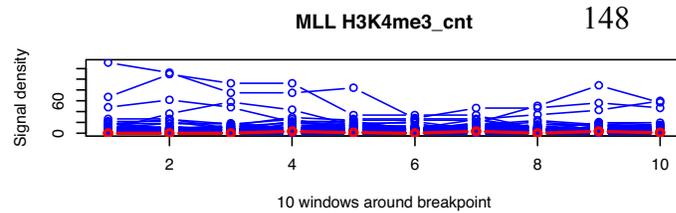
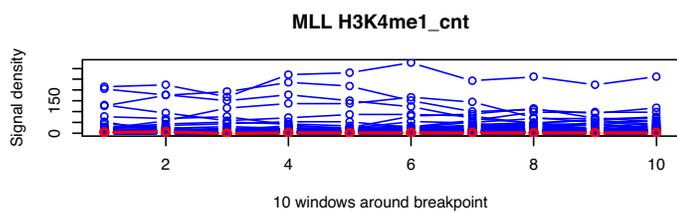


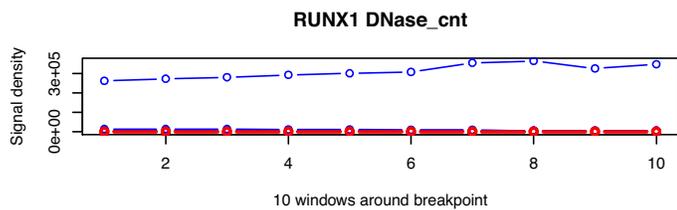
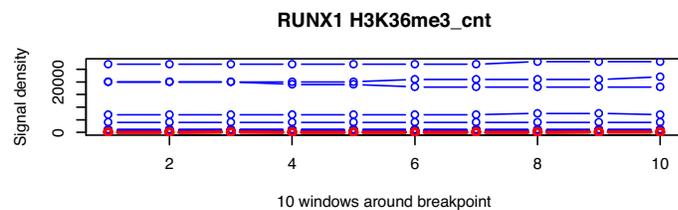
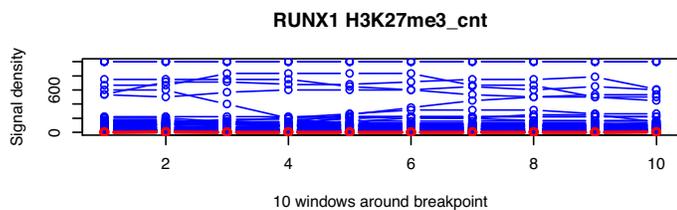
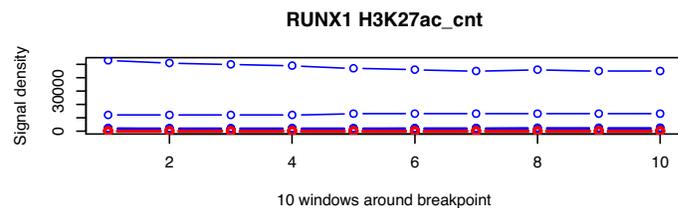
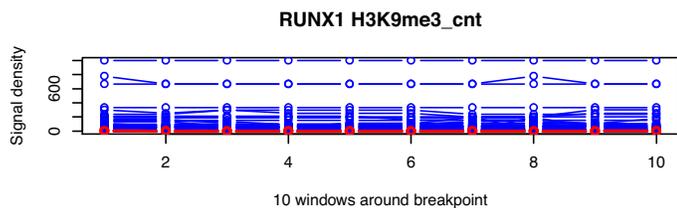
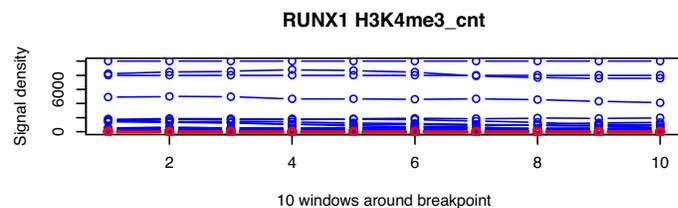
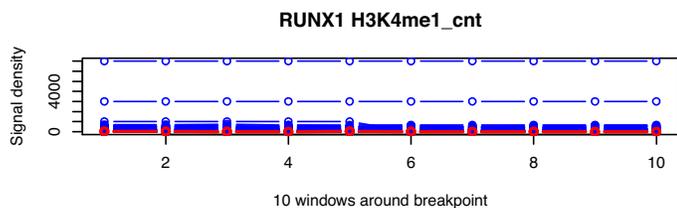
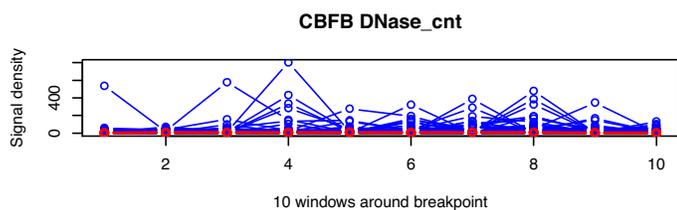
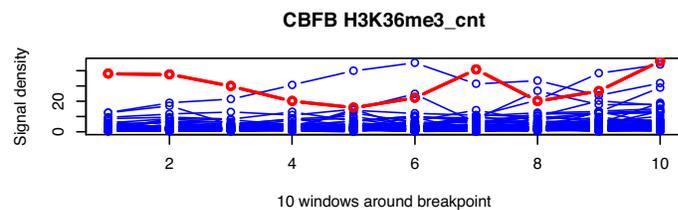
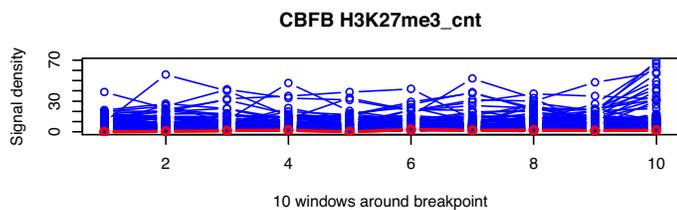
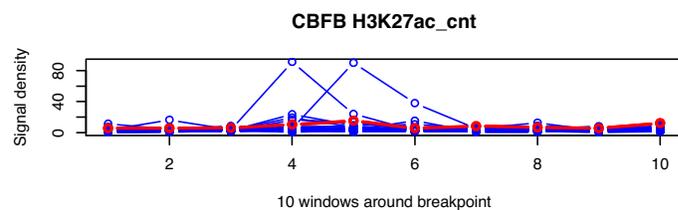
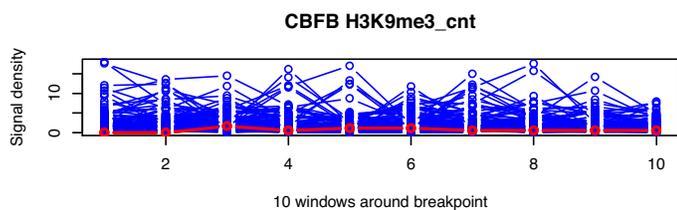
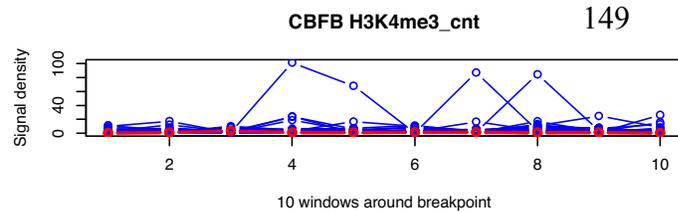
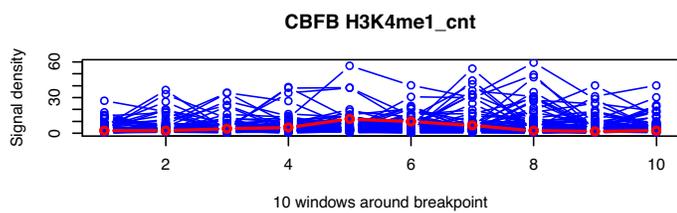


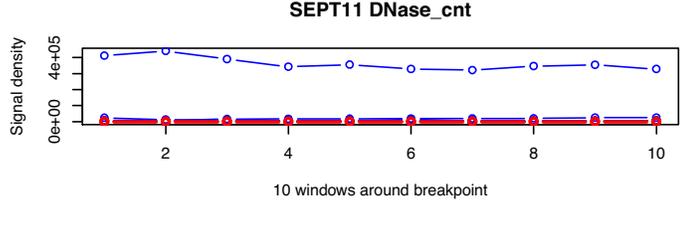
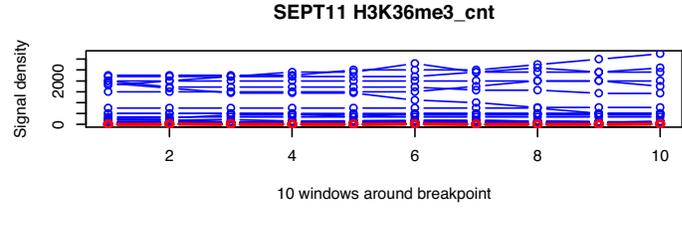
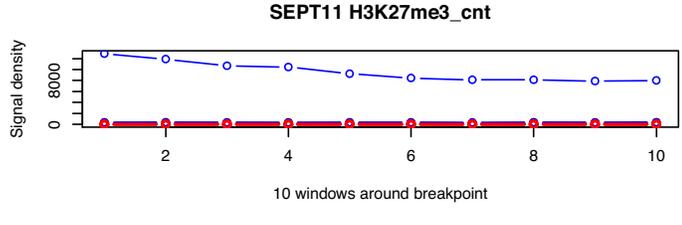
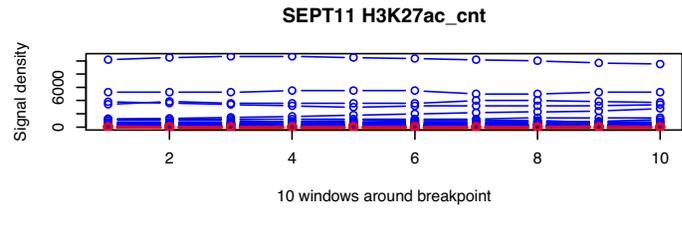
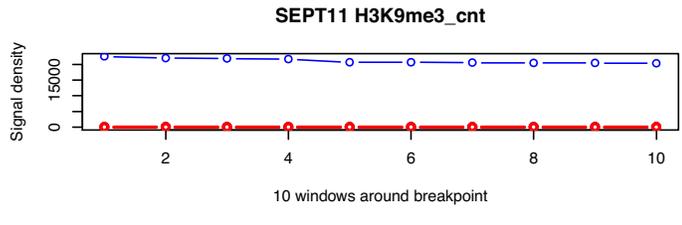
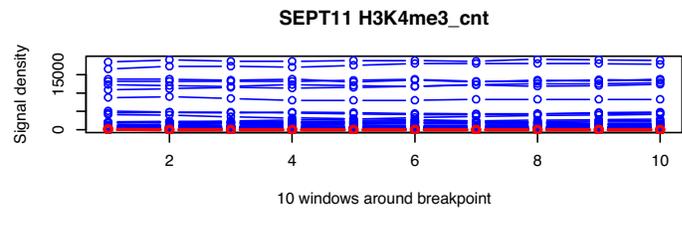
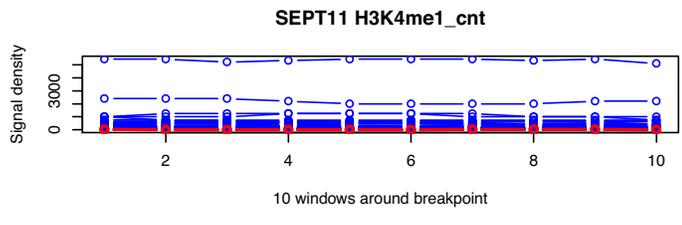
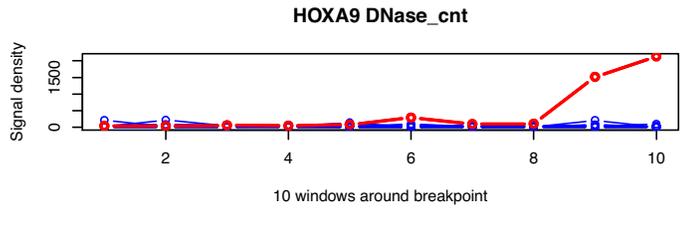
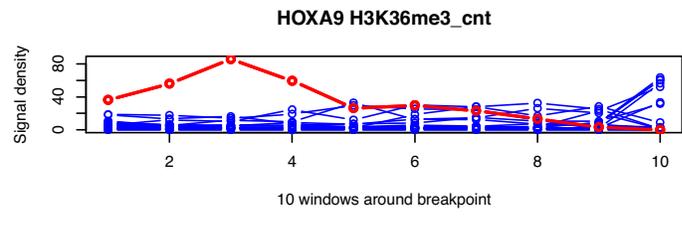
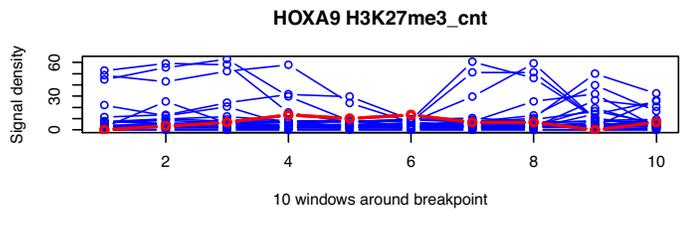
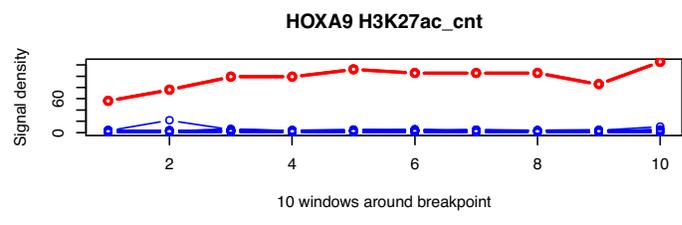
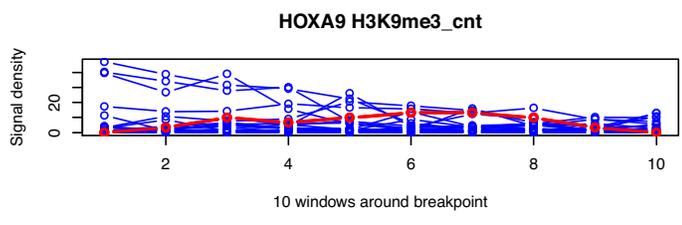
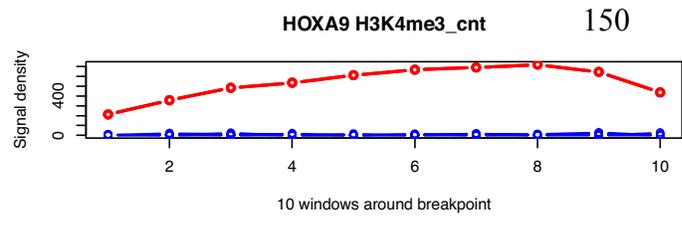
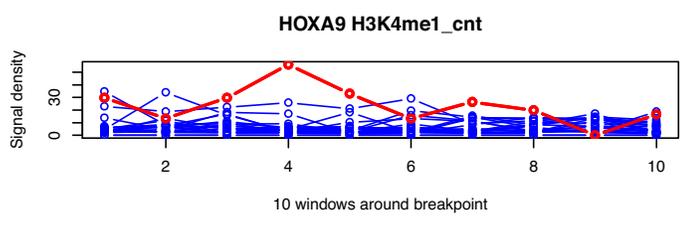


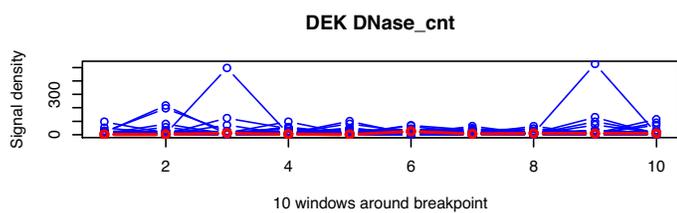
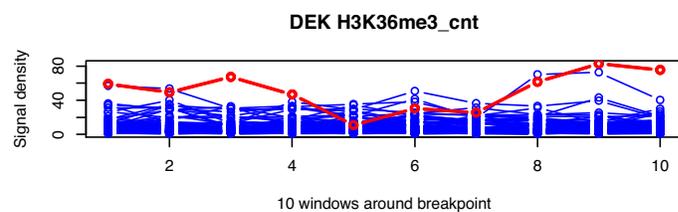
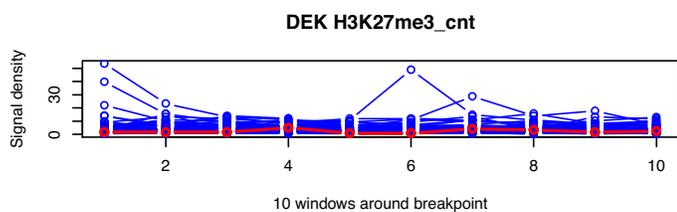
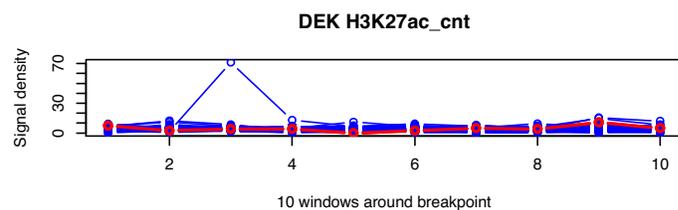
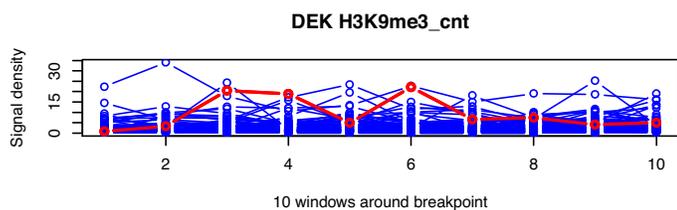
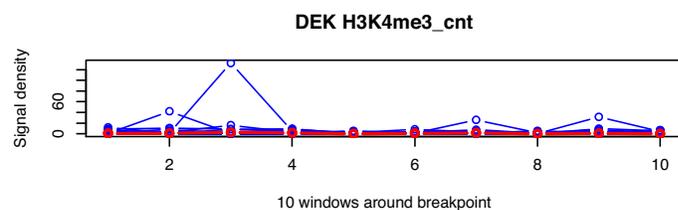
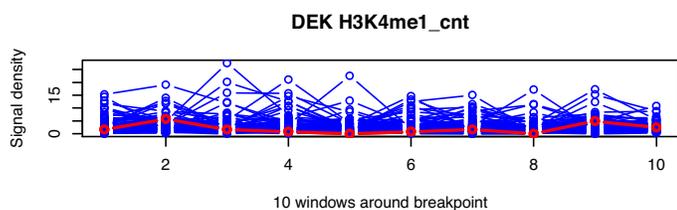
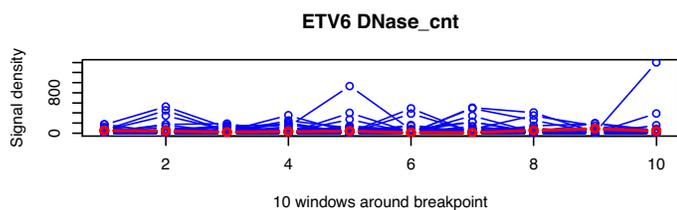
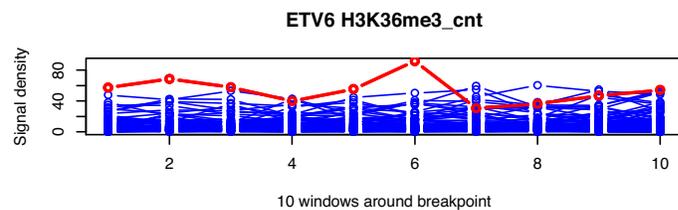
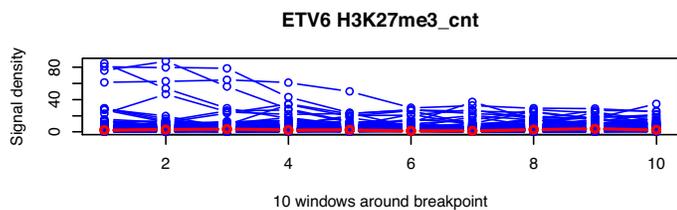
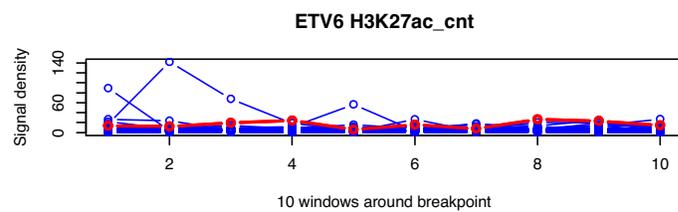
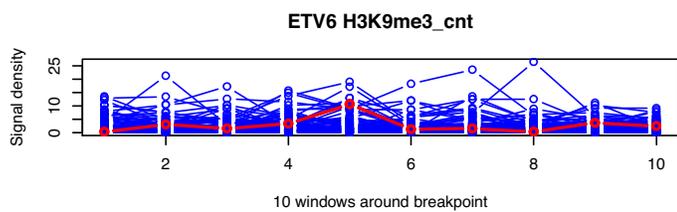
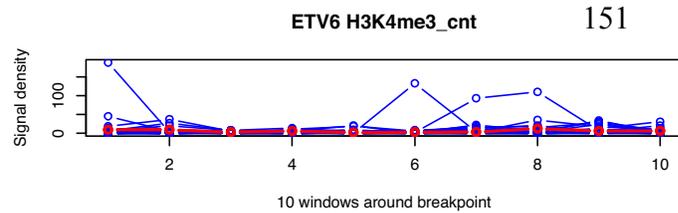
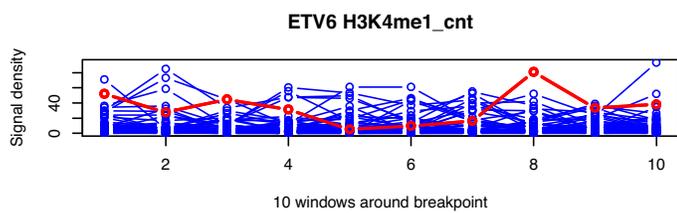


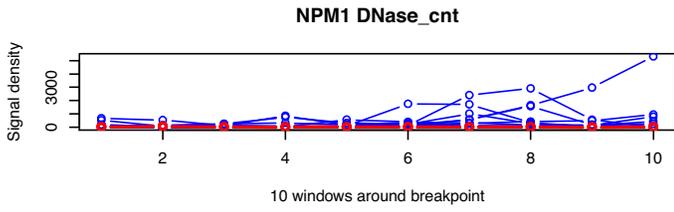
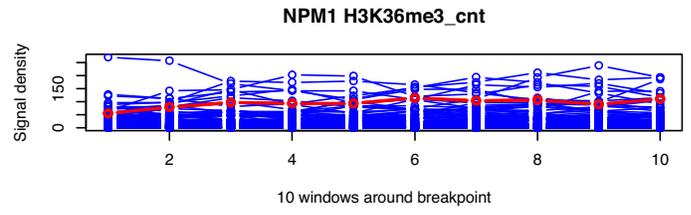
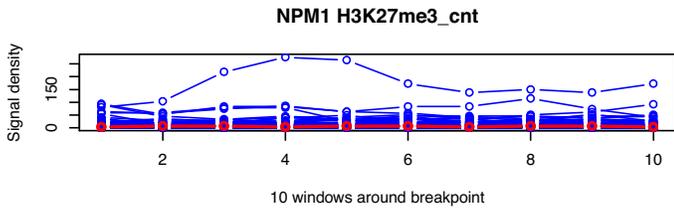
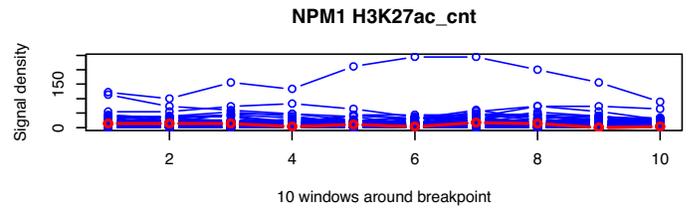
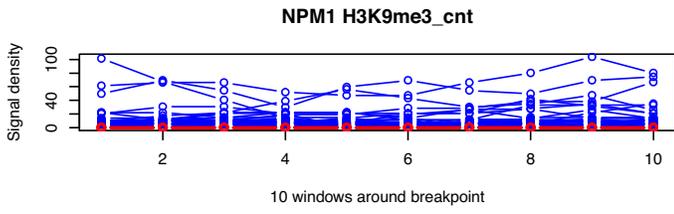
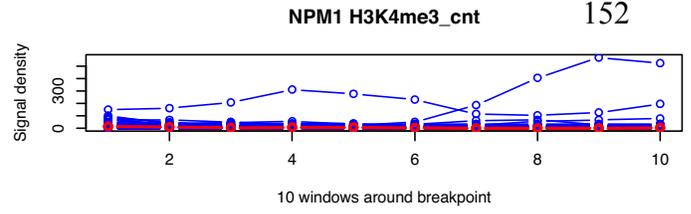
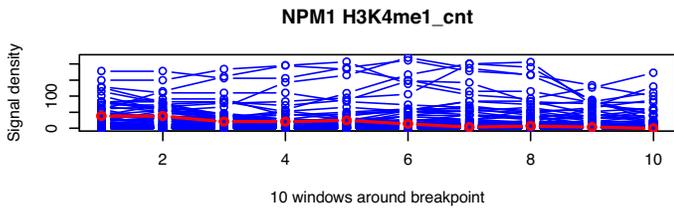












REFERENCES

- Adli M, Zhu J, Bernstein BE. 2010. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nature methods* **7**: 615-618.
- Agresti A, Coull BA. 1998. Approximate is better than "exact" for interval estimation of binomial proportions. *Am Stat* **52**: 119-126.
- Ayrapetov MK, Gursoy-Yuzugullu O, Xu C, Xu Y, Price BD. 2014. DNA double-strand breaks promote methylation of histone H3 on lysine 9 and transient formation of repressive chromatin. *Proceedings of the National Academy of Sciences of the United States of America* **111**: 9169-9174.
- Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell research* **21**: 381-395.
- Bernstein BE, Mikkelsen TS, Xie X, Kamal M, Huebert DJ, Cuff J, Fry B, Meissner A, Wernig M, Plath K et al. 2006. A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**: 315-326.
- Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, Kellis M, Marra MA, Beaudet AL, Ecker JR et al. 2010. The NIH Roadmap Epigenomics Mapping Consortium. *Nature biotechnology* **28**: 1045-1048.
- Bernt KM, Armstrong SA. 2009. Leukemia stem cells and human acute lymphoblastic leukemia. *Seminars in hematology* **46**: 33-38.
- Binda O. 2013. On your histone mark, SET, methylate! *Epigenetics : official journal of the DNA Methylation Society* **8**: 457-463.
- Birney E, Stamatoyannopoulos JA, Dutta A, Guigo R, Gingeras TR, Margulies EH, Weng Z, Snyder M, Dermitzakis ET, Thurman RE et al. 2007. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**: 799-816.
- Bonnet D, Dick JE. 1997. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature medicine* **3**: 730-737.
- Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human molecular genetics* **10**: 211-219.
- Branco MR, Pombo A. 2006. Intermingling of chromosome territories in interphase suggests role in translocations and transcription-dependent associations. *PLoS biology* **4**: e138.

- Burgess RC, Burman B, Kruhlak MJ, Misteli T. 2014. Activation of DNA damage response signaling by condensed chromatin. *Cell reports* **9**: 1703-1717.
- Camphausen K, Tofilon PJ. 2007. Inhibition of histone deacetylation: a strategy for tumor radiosensitization. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **25**: 4051-4056.
- Chiarle R, Voena C, Ambrogio C, Piva R, Inghirami G. 2008. The anaplastic lymphoma kinase in the pathogenesis of cancer. *Nature reviews Cancer* **8**: 11-23.
- Chiarle R, Zhang Y, Frock RL, Lewis SM, Molinie B, Ho YJ, Myers DR, Choi VW, Compagno M, Malkin DJ et al. 2011. Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**: 107-119.
- Chuang CH, Carpenter AE, Fuchsova B, Johnson T, de Lanerolle P, Belmont AS. 2006. Long-range directional movement of an interphase chromosome site. *Current biology : CB* **16**: 825-831.
- Ciccia A, Elledge SJ. 2010. The DNA damage response: making it safe to play with knives. *Molecular cell* **40**: 179-204.
- Conrad C, Gerlich DW. 2010. Automated microscopy for high-content RNAi screening. *The Journal of cell biology* **188**: 453-461.
- Cremer T, Cremer C. 2001. Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nature reviews Genetics* **2**: 292-301.
- Crosetto N, Mitra A, Silva MJ, Bienko M, Dojer N, Wang Q, Karaca E, Chiarle R, Skrzypczak M, Ginalski K et al. 2013. Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nature methods* **10**: 361-365.
- Daniel JA, Nussenzweig A. 2012. Roles for histone H3K4 methyltransferase activities during immunoglobulin class-switch recombination. *Biochimica et biophysica acta* **1819**: 733-738.
- Danzer JR, Wallrath LL. 2004. Mechanisms of HP1-mediated gene silencing in *Drosophila*. *Development* **131**: 3571-3580.
- De Melo V, Vetter M, Mazzullo H, Howard JD, Betts DR, Nacheva EP, Apperley JF, Reid AG. 2008. A simple FISH assay for the detection of 3q26 rearrangements in myeloid malignancy. *Leukemia* **22**: 434-437.
- Dellaire G, Kepkay R, Bazett-Jones DP. 2009. High resolution imaging of changes in the structure and spatial organization of chromatin, gamma-H2A.X and the MRN complex within etoposide-induced DNA repair foci. *Cell Cycle* **8**: 3750-3769.

- Dixon JR, Jung I, Selvaraj S, Shen Y, Antosiewicz-Bourget JE, Lee AY, Ye Z, Kim A, Rajagopal N, Xie W et al. 2015. Chromatin architecture reorganization during stem cell differentiation. *Nature* **518**: 331-336.
- Dundr M, Ospina JK, Sung MH, John S, Upender M, Ried T, Hager GL, Matera AG. 2007. Actin-dependent intranuclear repositioning of an active gene locus in vivo. *The Journal of cell biology* **179**: 1095-1103.
- Duyster J, Bai RY, Morris SW. 2001. Translocations involving anaplastic lymphoma kinase (ALK). *Oncogene* **20**: 5623-5637.
- Eliceiri KW, Berthold MR, Goldberg IG, Ibanez L, Manjunath BS, Martone ME, Murphy RF, Peng H, Plant AL, Roysam B et al. 2012. Biological imaging software tools. *Nature methods* **9**: 697-710.
- Falk M, Lukasova E, Kozubek S. 2008. Chromatin structure influences the sensitivity of DNA to gamma-radiation. *Biochimica et biophysica acta* **1783**: 2398-2414.
- . 2010. Higher-order chromatin structure in DSB induction, repair and misrepair. *Mutation research* **704**: 88-100.
- Felsenfeld G, Groudine M. 2003. Controlling the double helix. *Nature* **421**: 448-453.
- Fischer P, Nacheva E, Mason DY, Sherrington PD, Hoyle C, Hayhoe FG, Karpas A. 1988. A Ki-1 (CD30)-positive human cell line (Karpas 299) established from a high-grade non-Hodgkin's lymphoma, showing a 2;5 translocation and rearrangement of the T-cell receptor beta-chain gene. *Blood* **72**: 234-240.
- Frohling S, Dohner H. 2008. Chromosomal abnormalities in cancer. *The New England journal of medicine* **359**: 722-734.
- Ghezraoui H, Piganeau M, Renouf B, Renaud JB, Sallmyr A, Ruis B, Oh S, Tomkinson AE, Hendrickson EA, Giovannangeli C et al. 2014. Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Molecular cell* **55**: 829-842.
- Gogusev J, Telvi L, Nezelof C. 2002. Molecular cytogenetic aberrations in CD30+ anaplastic large cell lymphoma cell lines. *Cancer genetics and cytogenetics* **138**: 95-101.
- Goodarzi AA, Noon AT, Deckbar D, Ziv Y, Shiloh Y, Lobrich M, Jeggo PA. 2008. ATM signaling facilitates repair of DNA double-strand breaks associated with heterochromatin. *Molecular cell* **31**: 167-177.
- Goren A, Oszolak F, Shores N, Ku M, Adli M, Hart C, Gymrek M, Zuk O, Regev A, Milos PM et al. 2010. Chromatin profiling by directly sequencing small quantities of immunoprecipitated DNA. *Nature methods* **7**: 47-49.

- Gozzetti A, Le Beau MM. 2000. Fluorescence in situ hybridization: uses and limitations. *Seminars in hematology* **37**: 320-333.
- Groth A, Rocha W, Verreault A, Almouzni G. 2007. Chromatin challenges during DNA replication and repair. *Cell* **128**: 721-733.
- Hakim O, Resch W, Yamane A, Klein I, Kieffer-Kwon KR, Jankovic M, Oliveira T, Bothmer A, Voss TC, Ansarah-Sobrinho C et al. 2012. DNA damage defines sites of recurrent chromosomal translocations in B lymphocytes. *Nature* **484**: 69-74.
- Hanahan D, Weinberg RA. 2011. Hallmarks of cancer: the next generation. *Cell* **144**: 646-674.
- Hsu PD, Lander ES, Zhang F. 2014. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**: 1262-1278.
- Inagaki H, Ohye T, Kogo H, Kato T, Bolor H, Taniguchi M, Shaikh TH, Emanuel BS, Kurahashi H. 2009. Chromosomal instability mediated by non-B DNA: cruciform conformation and not DNA sequence is responsible for recurrent translocation in humans. *Genome research* **19**: 191-198.
- Jacome A, Fernandez-Capetillo O. 2011. Lac operator repeats generate a traceable fragile site in mammalian cells. *EMBO reports* **12**: 1032-1038.
- Jakob B, Splinter J, Conrad S, Voss KO, Zink D, Durante M, Lohrlich M, Taucher-Scholz G. 2011. DNA double-strand breaks in heterochromatin elicit fast repair protein recruitment, histone H2AX phosphorylation and relocation to euchromatin. *Nucleic acids research* **39**: 6489-6499.
- Janz M, Hummel M, Truss M, Wollert-Wulf B, Mathas S, Johrens K, Hagemeyer C, Bommert K, Stein H, Dorken B et al. 2006. Classical Hodgkin lymphoma is characterized by high constitutive expression of activating transcription factor 3 (ATF3), which promotes viability of Hodgkin/Reed-Sternberg cells. *Blood* **107**: 2536-2539.
- Ji Y, Resch W, Corbett E, Yamane A, Casellas R, Schatz DG. 2010. The in vivo pattern of binding of RAG1 and RAG2 to antigen receptor loci. *Cell* **141**: 419-431.
- Kato L, Begum NA, Burroughs AM, Doi T, Kawai J, Daub CO, Kawaguchi T, Matsuda F, Hayashizaki Y, Honjo T. 2012. Nonimmunoglobulin target loci of activation-induced cytidine deaminase (AID) share unique features with immunoglobulin genes. *Proceedings of the National Academy of Sciences of the United States of America* **109**: 2479-2484.
- Kearney L. 2001. Molecular cytogenetics. *Best practice & research Clinical haematology* **14**: 645-669.

- Kent WJ. 2002. BLAT--the BLAST-like alignment tool. *Genome research* **12**: 656-664.
- Khurana S, Kruhlak MJ, Kim J, Tran AD, Liu J, Nyswaner K, Shi L, Jailwala P, Sung MH, Hakim O et al. 2014. A macrohistone variant links dynamic chromatin compaction to BRCA1-dependent genome maintenance. *Cell reports* **8**: 1049-1062.
- Kimura H. 2013. Histone modifications for human epigenome analysis. *Journal of human genetics* **58**: 439-445.
- Kinner A, Wu W, Staudt C, Iliakis G. 2008. Gamma-H2AX in recognition and signaling of DNA double-strand breaks in the context of chromatin. *Nucleic acids research* **36**: 5678-5694.
- Kinney MC, Higgins RA, Medina EA. 2011. Anaplastic large cell lymphoma: twenty-five years of discovery. *Archives of pathology & laboratory medicine* **135**: 19-43.
- Klein IA, Resch W, Jankovic M, Oliveira T, Yamane A, Nakahashi H, Di Virgilio M, Bothmer A, Nussenzweig A, Robbiani DF et al. 2011. Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* **147**: 95-106.
- Kolasinska-Zwierz P, Down T, Latorre I, Liu T, Liu XS, Ahringer J. 2009. Differential chromatin marking of introns and expressed exons by H3K36me3. *Nature genetics* **41**: 376-381.
- Krawczyk PM, Borovski T, Stap J, Cijssouw T, ten Cate R, Medema JP, Kanaar R, Franken NA, Aten JA. 2012. Chromatin mobility is increased at sites of DNA double-strand breaks. *Journal of cell science* **125**: 2127-2133.
- Kruhlak MJ, Celeste A, Dellaire G, Fernandez-Capetillo O, Muller WG, McNally JG, Bazett-Jones DP, Nussenzweig A. 2006. Changes in chromatin structure and mobility in living cells at sites of DNA double-strand breaks. *The Journal of cell biology* **172**: 823-834.
- Kurahashi H, Inagaki H, Ohye T, Kogo H, Tsutsumi M, Kato T, Tong M, Emanuel BS. 2010. The constitutional t(11;22): implications for a novel mechanism responsible for gross chromosomal rearrangements. *Clinical genetics* **78**: 299-309.
- Ladanyi M, Cavalchire G. 1996. Detection of the NPM-ALK genomic rearrangement of Ki-1 lymphoma and isolation of the involved NPM and ALK introns. *Diagnostic molecular pathology : the American journal of surgical pathology, part B* **5**: 154-158.
- Lafon-Hughes L, Di Tomaso MV, Liddle P, Toledo A, Reyes-Abalos AL, Folle GA. 2013. Preferential localization of gammaH2AX foci in euchromatin of retina rod cells after DNA damage induction. *Chromosome research : an international*

journal on the molecular, supramolecular and evolutionary aspects of chromosome biology **21**: 789-803.

- Li Y, Danzer JR, Alvarez P, Belmont AS, Wallrath LL. 2003. Effects of tethering HP1 to euchromatic regions of the *Drosophila* genome. *Development* **130**: 1817-1824.
- Liberali P, Snijder B, Pelkmans L. 2015. Single-cell and multivariate approaches in genetic perturbation screens. *Nature reviews Genetics* **16**: 18-32.
- Lin C, Yang L, Rosenfeld MG. 2012. Molecular logic underlying chromosomal translocations, random or non-random? *Advances in cancer research* **113**: 241-279.
- Lin C, Yang L, Tanasa B, Hutt K, Ju BG, Ohgi K, Zhang J, Rose DW, Fu XD, Glass CK et al. 2009. Nuclear receptor-induced chromosomal proximity and DNA breaks underlie specific translocations in cancer. *Cell* **139**: 1069-1083.
- Liu M, Duke JL, Richter DJ, Vinuesa CG, Goodnow CC, Kleinstein SH, Schatz DG. 2008. Two levels of protection for the B cell genome during somatic hypermutation. *Nature* **451**: 841-845.
- Liu Y, Subrahmanyam R, Chakraborty T, Sen R, Desiderio S. 2007. A plant homeodomain in RAG-2 that binds Hypermethylated lysine 4 of histone H3 is necessary for efficient antigen-receptor-gene rearrangement. *Immunity* **27**: 561-571.
- Lobrich M, Shibata A, Beucher A, Fisher A, Ensminger M, Goodarzi AA, Barton O, Jeggo PA. 2010. gammaH2AX foci analysis for monitoring DNA double-strand break repair: strengths, limitations and optimization. *Cell Cycle* **9**: 662-669.
- Luco RF, Pan Q, Tominaga K, Blencowe BJ, Pereira-Smith OM, Misteli T. 2010. Regulation of alternative splicing by histone modifications. *Science* **327**: 996-1000.
- Lukas J, Lukas C, Bartek J. 2011. More than just a focus: The chromatin response to DNA damage and its role in genome integrity maintenance. *Nature cell biology* **13**: 1161-1169.
- Luthra R, Pugh WC, Waasdorp M, Morris W, Cabanillas F, Chan PK, Sarris AH. 1998. Mapping of genomic t(2;5)(p23;q35) break points in patients with anaplastic large cell lymphoma by sequencing long-range PCR products. *Hematopathology and molecular hematology* **11**: 173-183.
- Mani RS, Chinnaiyan AM. 2010. Triggers for genomic rearrangements: insights into genomic, cellular and environmental influences. *Nature reviews Genetics* **11**: 819-829.

- Marculescu R, Le T, Simon P, Jaeger U, Nadel B. 2002. V(D)J-mediated translocations in lymphoid neoplasms: a functional assessment of genomic instability by cryptic sites. *The Journal of experimental medicine* **195**: 85-98.
- Marino-Enriquez A, Dal Cin P. 2013. ALK as a paradigm of oncogenic promiscuity: different mechanisms of activation and different fusion partners drive tumors of different lineages. *Cancer genetics* **206**: 357-373.
- Markey FB, Ruezinsky W, Tyagi S, Batish M. 2014. Fusion FISH imaging: single-molecule detection of gene fusion transcripts in situ. *PloS one* **9**: e93488.
- Marmorstein R, Trievel RC. 2009. Histone modifying enzymes: structures, mechanisms, and specificities. *Biochimica et biophysica acta* **1789**: 58-68.
- Maroschik B, Gurtler A, Kramer A, Rossler U, Gomolka M, Hornhardt S, Mortl S, Friedl AA. 2014. Radiation-induced alterations of histone post-translational modification levels in lymphoblastoid cell lines. *Radiat Oncol* **9**: 15.
- Mathas S, Johrens K, Joos S, Lietz A, Hummel F, Janz M, Jundt F, Anagnostopoulos I, Bommert K, Lichter P et al. 2005. Elevated NF-kappaB p50 complex formation and Bcl-3 expression in classical Hodgkin, anaplastic large-cell, and other peripheral T-cell lymphomas. *Blood* **106**: 4287-4293.
- Mathas S, Kreher S, Meaburn KJ, Johrens K, Lamprecht B, Assaf C, Sterry W, Kadin ME, Daibata M, Joos S et al. 2009. Gene deregulation and spatial genome reorganization near breakpoints prior to formation of translocations in anaplastic large cell lymphoma. *Proceedings of the National Academy of Sciences of the United States of America* **106**: 5831-5836.
- Matthews AG, Kuo AJ, Ramon-Maiques S, Han S, Champagne KS, Ivanov D, Gallardo M, Carney D, Cheung P, Ciccone DN et al. 2007. RAG2 PHD finger couples histone H3 lysine 4 trimethylation with V(D)J recombination. *Nature* **450**: 1106-1110.
- Meaburn KJ, Misteli T. 2007. Cell biology: chromosome territories. *Nature* **445**: 379-781.
- Meyer C, Schneider B, Jakob S, Strehl S, Attarbaschi A, Schnittger S, Schoch C, Jansen MW, van Dongen JJ, den Boer ML et al. 2006. The MLL recombinome of acute leukemias. *Leukemia* **20**: 777-784.
- Misteli T. 2010. Higher-order genome organization in human disease. *Cold Spring Harbor perspectives in biology* **2**: a000794.
- Mitelman F, Johansson B, Mertens F. 2007. The impact of translocations and gene fusions on cancer causation. *Nature reviews Cancer* **7**: 233-245.

- Mostoslavsky R, Alt FW, Bassing CH. 2003. Chromatin dynamics and locus accessibility in the immune system. *Nature immunology* **4**: 603-606.
- Nambiar M, Raghavan SC. 2011. How does DNA break during chromosomal translocations? *Nucleic acids research* **39**: 5813-5825.
- Neumann FR, Dion V, Gehlen LR, Tsai-Pflugfelder M, Schmid R, Taddei A, Gasser SM. 2012. Targeted INO80 enhances subnuclear chromatin movement and ectopic homologous recombination. *Genes & development* **26**: 369-383.
- Nowell PC. 1985. Citation Classic - a Minute Chromosome in Human Chronic Granulocytic-Leukemia. *Cc/Life Sci*: 19-19.
- Numata M, Saito S, Nagata K. 2010. RAG-dependent recombination at cryptic RSSs within TEL-AML1 t(12;21)(p13;q22) chromosomal translocation region. *Biochemical and biophysical research communications* **402**: 718-724.
- Nye AC, Rajendran RR, Stenoien DL, Mancini MA, Katzenellenbogen BS, Belmont AS. 2002. Alteration of large-scale chromatin structure by estrogen receptor. *Molecular and cellular biology* **22**: 3437-3449.
- Panier S, Boulton SJ. 2014. Double-strand break repair: 53BP1 comes into focus. *Nature reviews Molecular cell biology* **15**: 7-18.
- Pegoraro G, Kubben N, Wickert U, Gohler H, Hoffmann K, Misteli T. 2009. Ageing-related chromatin defects through loss of the NURD complex. *Nature cell biology* **11**: 1261-1267.
- Peters AH, O'Carroll D, Scherthan H, Mechtler K, Sauer S, Schofer C, Weipoltshammer K, Pagani M, Lachner M, Kohlmaier A et al. 2001. Loss of the Suv39h histone methyltransferases impairs mammalian heterochromatin and genome stability. *Cell* **107**: 323-337.
- Piganeau M, Ghezraoui H, De Cian A, Guittat L, Tomishima M, Perrouault L, Rene O, Katibah GE, Zhang L, Holmes MC et al. 2013. Cancer translocations in human cells induced by zinc finger and TALE nucleases. *Genome research* **23**: 1182-1193.
- Polak P, Karlic R, Koren A, Thurman R, Sandstrom R, Lawrence MS, Reynolds A, Rynes E, Vlahovicek K, Stamatoyannopoulos JA et al. 2015. Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**: 360-364.
- Prensner JR, Chinnaiyan AM. 2009. Oncogenic gene fusions in epithelial carcinomas. *Current opinion in genetics & development* **19**: 82-91.

- Price BD, D'Andrea AD. 2013. Chromatin remodeling at DNA double-strand breaks. *Cell* **152**: 1344-1354.
- Pruitt KD, Tatusova T, Maglott DR. 2005. NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic acids research* **33**: D501-504.
- Rabbitts TH. 1994. Chromosomal translocations in human cancer. *Nature* **372**: 143-149.
- Raghavan SC, Kirsch IR, Lieber MR. 2001. Analysis of the V(D)J recombination efficiency at lymphoid chromosomal translocation breakpoints. *The Journal of biological chemistry* **276**: 29126-29133.
- Ramon-Maiques S, Kuo AJ, Carney D, Matthews AG, Oettinger MA, Gozani O, Yang W. 2007. The plant homeodomain finger of RAG2 recognizes histone H3 methylated at both lysine-4 and arginine-2. *Proceedings of the National Academy of Sciences of the United States of America* **104**: 18993-18998.
- Rando OJ. 2007. Global patterns of histone modifications. *Current opinion in genetics & development* **17**: 94-99.
- . 2012. Combinatorial complexity in chromatin structure and function: revisiting the histone code. *Current opinion in genetics & development* **22**: 148-155.
- Rivera CM, Ren B. 2013. Mapping human epigenomes. *Cell* **155**: 39-55.
- Rocha PP, Micsinai M, Kim JR, Hewitt SL, Souza PP, Trimarchi T, Strino F, Parisi F, Kluger Y, Skok JA. 2012. Close proximity to Igh is a contributing factor to AID-mediated translocations. *Molecular cell* **47**: 873-885.
- Rogakou EP, Boon C, Redon C, Bonner WM. 1999. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *The Journal of cell biology* **146**: 905-916.
- Roix JJ, McQueen PG, Munson PJ, Parada LA, Misteli T. 2003. Spatial proximity of translocation-prone gene loci in human lymphomas. *Nature genetics* **34**: 287-291.
- Roukos V, Burman B, Misteli T. 2013a. The cellular etiology of chromosome translocations. *Current opinion in cell biology* **25**: 357-364.
- Roukos V, Misteli T. 2014a. The biogenesis of chromosome translocations. *Nature cell biology* **16**: 293-300.
- . 2014b. Deep Imaging: the next frontier in microscopy. *Histochemistry and cell biology* **142**: 125-131.

- Roukos V, Voss TC, Schmidt CK, Lee S, Wangsa D, Misteli T. 2013b. Spatial dynamics of chromosome translocations in living cells. *Science* **341**: 660-664.
- Rowley JD. 1973. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**: 290-293.
- . 2001. Chromosome translocations: dangerous liaisons revisited. *Nature reviews Cancer* **1**: 245-250.
- Ruthenburg AJ, Allis CD, Wysocka J. 2007. Methylation of lysine 4 on histone H3: intricacy of writing and reading a single epigenetic mark. *Molecular cell* **25**: 15-30.
- Sarris AH, Luthra R, Cabanillas F, Morris SW, Pugh WC. 1998. Genomic DNA amplification and the detection of t(2;5)(p23;q35) in lymphoid neoplasms. *Leukemia & lymphoma* **29**: 507-514.
- Schones DE, Zhao K. 2008. Genome-wide approaches to studying chromatin modifications. *Nature reviews Genetics* **9**: 179-191.
- Schotta G, Sengupta R, Kubicek S, Malin S, Kauer M, Callen E, Celeste A, Pagani M, Opravil S, De La Rosa-Velazquez IA et al. 2008. A chromatin-wide transition to H4K20 monomethylation impairs genome integrity and programmed DNA rearrangements in the mouse. *Genes & development* **22**: 2048-2061.
- Schwartz M, Hakim O. 2014. 3D view of chromosomes, DNA damage, and translocations. *Current opinion in genetics & development* **25**: 118-125.
- Schwartz S, Meshorer E, Ast G. 2009. Chromatin organization marks exon-intron structure. *Nature structural & molecular biology* **16**: 990-995.
- Seiler DM, Rouquette J, Schmid VJ, Strickfaden H, Ottmann C, Drexler GA, Mazurek B, Greubel C, Hable V, Dollinger G et al. 2011. Double-strand break-induced transcriptional silencing is associated with loss of tri-methylation at H3K4. *Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology* **19**: 883-899.
- Semrau S, Crosetto N, Bienko M, Boni M, Bernasconi P, Chiarle R, van Oudenaarden A. 2014. FuseFISH: robust detection of transcribed gene fusions in single cells. *Cell reports* **6**: 18-23.
- Shimazaki N, Tsai AG, Lieber MR. 2009. H3K4me3 stimulates the V(D)J RAG complex for both nicking and hairpinning in trans in addition to tethering in cis: implications for translocations. *Molecular cell* **34**: 535-544.

- Shuga J, Zeng Y, Novak R, Lan Q, Tang X, Rothman N, Vermeulen R, Li L, Hubbard A, Zhang L et al. 2013. Single molecule quantitation and sequencing of rare translocations using microfluidic nested digital PCR. *Nucleic acids research* **41**: e159.
- Smallwood A, Ren B. 2013. Genome organization and long-range regulation of gene expression by enhancers. *Current opinion in cell biology* **25**: 387-394.
- Soria G, Polo SE, Almouzni G. 2012. Prime, repair, restore: the active role of chromatin in the DNA damage response. *Molecular cell* **46**: 722-734.
- Soutoglou E, Dorn JF, Sengupta K, Jasin M, Nussenzweig A, Ried T, Danuser G, Misteli T. 2007. Positional stability of single double-strand breaks in mammalian cells. *Nature cell biology* **9**: 675-682.
- Soutoglou E, Misteli T. 2008. Activation of the cellular DNA damage response in the absence of DNA lesions. *Science* **320**: 1507-1510.
- Spector DL. 2003. The dynamics of chromosome organization and gene regulation. *Annual review of biochemistry* **72**: 573-608.
- Stanlie A, Aida M, Muramatsu M, Honjo T, Begum NA. 2010. Histone3 lysine4 trimethylation regulated by the facilitates chromatin transcription complex is critical for DNA cleavage in class switch recombination. *Proceedings of the National Academy of Sciences of the United States of America* **107**: 22190-22195.
- Staszewski O, Baker RE, Ucher AJ, Martier R, Stavnezer J, Guikema JE. 2011. Activation-induced cytidine deaminase induces reproducible DNA breaks at many non-Ig Loci in activated B cells. *Molecular cell* **41**: 232-242.
- Statham AL, Strbenac D, Coolen MW, Stirzaker C, Clark SJ, Robinson MD. 2010. Repitools: an R package for the analysis of enrichment-based epigenomic data. *Bioinformatics* **26**: 1662-1663.
- Stein H, Foss HD, Durkop H, Marafioti T, Delsol G, Pulford K, Pileri S, Falini B. 2000. CD30(+) anaplastic large cell lymphoma: a review of its histopathologic, genetic, and clinical features. *Blood* **96**: 3681-3695.
- Swift S, Lorens J, Achacoso P, Nolan GP. 2001. Rapid production of retroviruses for efficient gene delivery to mammalian cells using 293T cell-based systems. *Current protocols in immunology / edited by John E Coligan [et al]* **Chapter 10**: Unit 10 17C.
- Tabbo F, Ponzoni M, Rabadan R, Bertoni F, Inghirami G. 2013. Beyond NPM-anaplastic lymphoma kinase driven lymphomagenesis: alternative drivers in anaplastic large cell lymphoma. *Current opinion in hematology* **20**: 374-381.

- Takata H, Hanafusa T, Mori T, Shimura M, Iida Y, Ishikawa K, Yoshikawa K, Yoshikawa Y, Maeshima K. 2013. Chromatin compaction protects genomic DNA from radiation damage. *PLoS one* **8**: e75622.
- Talkowski ME, Ernst C, Heilbut A, Chiang C, Hanscom C, Lindgren A, Kirby A, Liu S, Muddukrishna B, Ohsumi TK et al. 2011. Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. *American journal of human genetics* **88**: 469-481.
- Tumbar T, Sudlow G, Belmont AS. 1999. Large-scale chromatin unfolding and remodeling induced by VP16 acidic activation domain. *The Journal of cell biology* **145**: 1341-1354.
- Ventura RA, Martin-Subero JI, Jones M, McParland J, Gesk S, Mason DY, Siebert R. 2006. FISH analysis for the detection of lymphoma-associated chromosomal abnormalities in routine paraffin-embedded tissue. *The Journal of molecular diagnostics : JMD* **8**: 141-151.
- Verschure PJ, van der Kraan I, de Leeuw W, van der Vlag J, Carpenter AE, Belmont AS, van Driel R. 2005. In vivo HP1 targeting causes large-scale chromatin condensation and enhanced histone lysine methylation. *Molecular and cellular biology* **25**: 4552-4564.
- Wagner T, Jung M. 2012. New lysine methyltransferase drug targets in cancer. *Nature biotechnology* **30**: 622-623.
- Waldmann T, Schneider R. 2013. Targeting histone modifications--epigenetics in cancer. *Current opinion in cell biology* **25**: 184-189.
- Wang D, Zhou J, Liu X, Lu D, Shen C, Du Y, Wei FZ, Song B, Lu X, Yu Y et al. 2013. Methylation of SUV39H1 by SET7/9 results in heterochromatin relaxation and genome instability. *Proceedings of the National Academy of Sciences of the United States of America* **110**: 5516-5521.
- Wang YM, Tegenfeldt JO, Reisner W, Riehn R, Guan XJ, Guo L, Golding I, Cox EC, Sturm J, Austin RH. 2005. Single-molecule studies of repressor-DNA interactions show long-range interactions. *Proceedings of the National Academy of Sciences of the United States of America* **102**: 9796-9801.
- Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, Cui K, Roh TY, Peng W, Zhang MQ et al. 2008. Combinatorial patterns of histone acetylations and methylations in the human genome. *Nature genetics* **40**: 897-903.
- Warters RL, Lyons BW. 1992. Variation in radiation-induced formation of DNA double-strand breaks as a function of chromatin structure. *Radiation research* **130**: 309-318.

- Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. *Use R*: 1-212.
- Wolff DJ, Bagg A, Cooley LD, Dewald GW, Hirsch BA, Jacky PB, Rao KW, Rao PN. 2007. Guidance for fluorescence in situ hybridization testing in hematologic disorders. *The Journal of molecular diagnostics : JMD* **9**: 134-143.
- Woodcock CL, Ghosh RP. 2010. Chromatin higher-order structure and dynamics. *Cold Spring Harbor perspectives in biology* **2**: a000596.
- Yu J, Mani RS, Cao Q, Brenner CJ, Cao X, Wang X, Wu L, Li J, Hu M, Gong Y et al. 2010. An integrated network of androgen receptor, polycomb, and TMPRSS2-ERG gene fusions in prostate cancer progression. *Cancer cell* **17**: 443-454.
- Zhang F, Carvalho CM, Lupski JR. 2009. Complex human chromosomal and genomic rearrangements. *Trends in genetics : TIG* **25**: 298-307.
- Zhang Y, Gostissa M, Hildebrand DG, Becker MS, Boboila C, Chiarle R, Lewis S, Alt FW. 2010. The role of mechanistic factors in promoting chromosomal translocations found in lymphoid and other cancers. *Advances in immunology* **106**: 93-133.
- Zhang Y, McCord RP, Ho YJ, Lajoie BR, Hildebrand DG, Simon AC, Becker MS, Alt FW, Dekker J. 2012. Spatial organization of the mouse genome and its role in recurrent chromosomal translocations. *Cell* **148**: 908-921.
- Zhang Y, Rowley JD. 2006. Chromatin structural elements and chromosomal translocations in leukemia. *DNA repair* **5**: 1282-1297.
- Zink D, Fischer AH, Nickerson JA. 2004. Nuclear structure in cancer cells. *Nature reviews Cancer* **4**: 677-687.