Advances in single-cell omics to study genome organization, regulation, and repair

Kim de Luca

The work described in this thesis was performed at the Hubrecht Institute for Developmental Biology and Stem Cell Research, affiliated with the Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht, within the Graduate School of Life Sciences at Utrecht University.

ISBN/EAN 978-90-393-7881-6
DOI https://doi.org/10.33540/2989
NUR-code 923

Cover artwork *Traveling Light IV* by Ad Arma, used with permission
Chapter artwork Midjourney and mixed media with manual editing
Layout and design Kim de Luca
Printing Groenprint Rotterdam
Paper Nautilus® SuperWhite 100% post-consumer recycled

Advances in single-cell omics to study genome organization, regulation, and repair

Voortgang in *omics*-technieken ter bestudering van genoom-organisatie, -regulatie, en -herstel op unicellulair niveau

(met een samenvatting in het Nederlands)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof. dr. ir. W. Hazeleger,
ingevolge het besluit van het College voor Promoties
in het openbaar te verdedigen op

maandag 30 juni 2025 des ochtends te 10.15

door

Kim Lai Keen de Luca

geboren op 27 december 1990
te Willemstad, Curaçao

Promotores:
Prof. dr. ir. A. van Oudenaarden
Prof. dr. J.H. Kind

Beoordelingscommissie:
Prof. dr. ir. E.P.J.G. Cuppen
Dr. Aniek Janssen
Prof. dr. G.J.P.L. Kops
Prof. dr. W.L. de Laat
Prof. dr. A. Pombo

Contents

Chapter 1	Overview and relevance	8
	Introduction	10
	Scope of dissertation	30
Chapter 2	Single-cell DamID to capture contacts between DNA and the nuclear lamina in individual mammalian cells	41
Chapter 3	Single-cell profiling of transcriptome and histone modifications with EpiDamID	59
Chapter 4	Retrospective and multi-factorial profiling resolves the order of chromatin dynamics in single cells	115
Chapter 5	Genome-wide profiling of DNA repair proteins in single cells	173
Chapter 6	Higher-order clustering and coordinated repair in response to DNA double-strand breaks	207
Chapter 7	Chromatin profiling with single-cell sequencing technologies	227
Chapter 8	Discussion	247
Addendum	Samenvatting in het Nederlands	278
	Resúmen kòrtiku na Papiamentu	280
	Curriculum vitae	281
	List of publications	282

Chapter 1

Introduction

Overview and relevance

At the basis of life lies proper cellular functioning. All of the complex information required to build a cell and, in turn, an organism, is encoded in the genome. In human cells, two meters of DNA are intricately packaged into the nucleus as a substrate called chromatin, and this compaction affects nearly all nuclear processes. Moreover, the genome must be decoded and maintained by specialized macromolecules, including proteins that read, regulate, replicate, recombine, and repair DNA. While spatial genome organization has been a topic of study since the 19th century, much remains unknown about its interplay with the encoded genetic information, associated proteins, and, ultimately, cellular functioning. A better understanding of these fundamental mechanisms can have applications across a broad spectrum of fields, from personalized medicine to improved fertility treatment to ecological conservation strategies.

The study of nuclear organization, chromatin, and spatial genome positioning has seen marked developments in the past few decades, with the advent of both microscopy and DNA sequencing at high resolution and throughput. Imaging-based methods initially revealed broad organizational features, for example that chromosomes occupy distinct nuclear territories, and that chromatin types are spatially segregated within the nucleus. Sequencing-based approaches to measure chromosome conformation further resolved this organization into various instances of DNA associating into hubs, loops, and stripes. Similarly, microscopy and sequencing form the basis of several powerful methods to profile where and how DNA interacts with functional proteins, generating diverse insights into genome regulation. High-throughput sequencing provides accurate measurements of genome positioning and/or protein binding in samples of several thousands to millions of cells. However, DNA-based processes occurring inside the nucleus are often highly dynamic and heterogeneous across cells. For instance, the same genetic information in each cell undergoes complex regulation to give rise to many different cell types and states. Such regulation is partly accomplished by the reorganization of the genome inside of the nucleus, or by additional, epigenetic, mechanisms that modify the DNA and its associated proteins. Genome repositioning and chromatin modification can also occur in response to breakage or other forms of lesions to the DNA. Both the occurrence and repair of damage are partially stochastic processes, highlighting the need for methods that can study these genomic locations at the level of individual cells.

Current efforts in genome-wide profiling of DNA positioning and associated nuclear features are largely focused on the following: 1) increasing spatial resolution, 2) incorporating temporal dynamics, 3) simultaneously measuring multiple cellular readouts, and 4) capturing cell-to-cell heterogeneity.

This thesis presents technological advancements in profiling DNA-binding proteins at the single-cell level, with a focus on genome positioning and epigenetic chromatin modifications. I describe novel multi-modal sequencing approaches to map regions of interest genome-wide at high resolution, including variations that feature simultaneous measurements of multiple chromatin proteins, or in combination with transcriptional output. These methods are applied to study cellular differentiation during organismal development, as well as the response to DNA damage. Lastly, I discuss the methodological considerations and biological lessons in the context of genome organization, regulation, and repair.

Introduction

1.1 Principles of genome organization
- 1.1.1 A brief historical perspective
- 1.1.2 Spatial organization of the nucleus
- 1.1.3 Chromatin structure
- 1.1.4 Genome topology across scales
- 1.1.5 Compartmentalization and motion

1.2 DNA repair in the context of nuclear organization
- 1.2.1 DNA double-strand break damage and repair
- 1.2.2 Structure of nuclear repair domains
- 1.2.3 Movement of DSBs within the nucleus
- 1.2.4 Nuclear organization in oncogenesis and treatment

1.3 Intercellular heterogeneity in form and function

1.4 Technological advances
- 1.4.1 Chromatin imaging
- 1.4.2 Genomics

Scope of dissertation

1.1 Principles of genome organization

1.1.1 A brief historical perspective

As an informal introduction to this thesis, I describe relevant discoveries in genome research, and the methodological avenues that divided scientific pursuit. On one hand, microscopists meticulously described cellular processes as deduced from visual observation. On the other, chemists sought to isolate cellular substances, rigorously characterizing their features and deducing rules for their provenance. Neither could conclude molecular functioning. It so happens that the state of the art, with all its drastic technological advances, is still largely divided along those lines. Only recently have the different fields started integrating morphology with molecular dynamics, approaching a more holistic view of cellular behavior. Lastly, this chronological digression serves to illustrate how even the earliest studies of the nucleus lay at the foundation of scientific knowledge across the spectrum of animal life, from embryo development to pathological growth.

Chromosomal inheritance

Beginning in the late 1700s, cellular behavior was studied by microscopic imaging of cell growth and multiplication during embryo development. The ability to visualize the cell and discern any structure within it relied on sophisticated staining techniques, which gave nuclear contents an especially enhanced contrast. So the term "chromatin"—*i.e.,* stainable substance—was coined. While the dynamics of chromatin were observable, it remained enigmatic what this substance represented and how it might be related to cellular behavior.

In the early 1800s, it was thought that nuclear contents were generated *de novo* each cell cycle from a mixture of chemical elements. DNA was discovered as a separate molecular class in 1869 by biochemist Friedrich Miescher[1]. Although Miescher could not have known that DNA was, in essence, the molecule of heredity, he found the substance in germ cells of many animals, and postulated that it could serve as the carrier of reproductive information[2].

In parallel, cytologists continued their analysis and tracking of chromatin (eventually also termed "chromosomes", *i.e.,* stainable bodies). The idea that chromatin and DNA were related molecular substances was initially proven, albeit indirectly, by showing their colocalization through staining[3]. The first evidence of chromosomal inheritance then came from the observation of long threads within the nucleus that assembled, split, and regressed as the cell progressed through cell division[4]. In his description of "karyomitosis", Flemming noted that chromatin threads separated into two groups, followed by the appearance of that same substance in what would become the daughter nuclei, thereby indicating continuity of the nuclear material.

Developmental biologists Boveri and Sutton independently expanded on the notion of chromosomes as genetic carriers, by concluding that chromosomes are individually persisting units originating from both parental sources[5,6]. Taken together, these findings formulated the chromosome theory of inheritance, substantiating what has become known as the Mendelian law of heredity[6].

In his pursuit to describe the first cell divisions after fertilization, Boveri noticed that some mitoses produced daughter cells with an unequal number of chromosomes, and that those embryos were "greatly abnormal"[5]. Based on this observation, he hypothesized that: 1) normal organismal development requires a balanced chromosome constitution, and 2) malignant tumor growth might be driven by a structurally disturbed set of chromosomes, *i.e.*, aneuploidy[7], thereby linking, for the first time, chromosomal abnormalities and cancer.

Nuclear organization

Beyond the observation of chromosomal dynamics during mitosis, it was debated whether chromosomes occupy distinct nuclear positions during interphase. Nuclear territories were first observed as small protrusions extending from the nuclear membrane, which seemed to correlate with the locations of chromosome ends[8], Figure 1.1. These protruding patterns remained relatively constant during interphase, indicating that chromosomes occupied stable positions within the nucleus. Moreover, while daughter cells shared nearly symmetric patterns of nuclear protrusions, these patterns were shuffled in subsequent mitoses, suggesting that chromosome positioning was fixed during interphase, but not rigidly inherited.

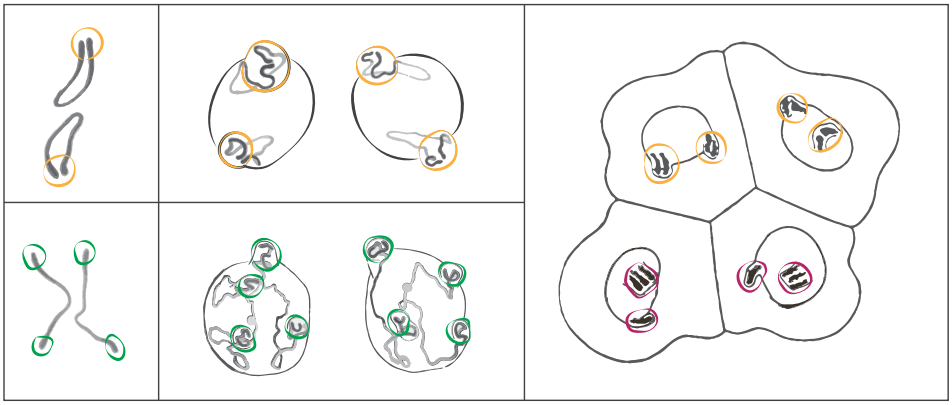


Figure 1.1. Nuclear organization and chromosomal alignment of sister cells in early embryonic divisions, re-interpreted from Boveri (1909)[8].
Left: chromosomal arrangement in the egg, prior to division. Middle: the resulting blastomeres. The top blastomeres each exhibit two nuclear protrusions (yellow), with each protrusion containing two chromosome ends, as in the egg (left). The bottom blastomeres exhibit four nuclear protrusions (green), each containing one chromosome end. *[cont]*

Right: 4-cell stage embryo, where the top nuclear pair exhibits similar protrusions, although it is unclear whether the two ends in each protrusion are from the same chromosome. The bottom nuclear pair has undergone rearrangement. Colored circles have been added for emphasis (not part of the original drawings).

By the 1920s, cytological techniques had improved such that chromosomal regions appeared to be stained to different degrees[9]. Heitz distinguished between lighter and more dispersed chromatin, which he termed "euchromatin", and darker and more condensed chromatin, referred to as "heterochromatin". Although Heitz did not explore three-dimensional segregation of chromatin types, he correctly inferred that "euchromatin is closely connected to gene activity during interphase; heterochromatin corresponds to genetically inert regions"[10], which are now known to correlate with nuclear positioning.

Conclusive evidence for the spatial organization of interphase nuclei came from experimental perturbation (rather than unbiased observation), by damaging and recording the DNA at particular positions within the nucleus. Damage at discrete nuclear regions (thus, a spatial record) was constrained to a few chromosomes (thus, a genomic record) rather than diffusely across chromosomes[11]. Subsequently, the technique of "chromosome painting" enabled the visual distinction of all individual chromosomes inside the nucleus[12,13], finally proving the theory of nuclear territories.

Although research on nuclear dynamics continued, attention shifted more towards biochemical encoding of genetic information, especially once the structure of the DNA molecule was discovered[14-16]. This endeavor culminated in the sequencing of the human genome in 2001[17]. Notably, however, only the euchromatic fraction was covered, since heterochromatic regions are often repetitive, complex, and therefore difficult to identify unambiguously. The final, end-to-end, completion of this sequence was achieved in 2022[18]. Nevertheless, it has become evident that DNA nucleotides provide but a one-dimensional view of genomic function. Thus, efforts to understand DNA-related processed such as transcription, replication, and repair have refocused on studying the nuclear environment in which DNA resides.

1.1.2 Spatial organization of the nucleus

The nucleus of eukaryotic cells not only serves as a repository for genetic material; it also presents a dynamic structure that is critical for genome function. Specialized compartments and components provide the framework for larger-scale organization of chromatin and nuclear bodies, to smaller-scale positioning of specific gene loci. Together, this regulates transcriptional activity, replication timing, and genomic stability during cellular processes such as lineage specification and disease progression.

Sub-nuclear structures and compartments

At the largest scale, nuclear architecture is defined by key structural components, including the nuclear envelope, nucleoplasm, and nucleolus. The nuclear envelope acts as a boundary between the nucleoplasm and the cytoplasm, ensuring regulated molecular transport via nuclear pores. Within the nucleoplasm, chromatin is spatially segregated at the periphery and the interior, creating environments with different genomic activities (see also *1.1.3 Chromatin structure*). Heterochromatin is condensed and transcriptionally repressed, with essential functions in maintaining genomic stability and regulating gene expression through silencing mechanisms. It is mainly located at the nuclear periphery and around the nucleolus. In contrast, euchromatin is more accessible, housing genes involved in active transcription and regulatory processes. The nucleolus, the largest structure nestled within the nuclear interior, is the site of ribosome biogenesis and contributes to cellular protein synthesis. Some of the nucleolus-associated domains (NADs) share features with regions tethered to the periphery, while others have hallmarks of facultative heterochromatin, such as higher levels of gene expression and early replication timing[19].

The nuclear periphery, adjacent to the inner nuclear membrane (INM), plays a crucial role in nuclear organization and genomic regulation. At the periphery lies the nuclear lamina (NL), a proteinaceous meshwork mainly composed of lamins and associated proteins such as the lamin B1 receptor (LBR). Functioning as a structural scaffold, the NL supports the nucleus and anchors chromatin, thereby influencing the transcriptional state of genomic regions. The NL involvement in gene regulation extends to processes like chromatin remodeling, DNA replication timing, and DNA repair. Genomic regions tethered to the NL, so-called lamina-associated domains (LADs), are dynamic in their positioning, e.g., changing during the cell cycle and in response to cell-extrinsic cues.

Biophysical and molecular processes

Organization of the nucleus arises from various processes, including protein-DNA interactions, compartmentalization, and higher-order chromatin folding. Genomic regions can be tethered to specific nuclear features, mediated by factors such as lamins and other nuclear matrix components, as described above. Second, chromosomes are folded into “loops” through the process of extrusion, driven by ATP-dependent molecular machinery (see also *1.1.4 Genome topology across scales*). Third, the biophysical mechanism of phase separation gives rise to distinct liquid-like compartments within the nucleoplasm, leading to segregation of chromatin types and formation of sub-nuclear bodies (see also *1.1.5. Compartmentalization and motion*). Nuclear organization and genome folding are ultimately the result of interplay between these (and potentially other) mechanisms.

State-of-the-art evidence posits that genomic regions in contact with the nuclear periphery (not limited to chromosome ends) are stochastically reshuffled during mitosis[20]. Notably, in the next G1 phase, distal chromosome ends localize to the periphery first, and gradually detach as the cell cycle progresses[21-23].

1.1.3 Chromatin structure: physical compaction and epigenetic modifications

Chromatin serves as the foundational structure for packaging of the genome inside the nucleus, and is the substrate for all DNA-related transactions. The core repeating unit of chromatin is the nucleosome, comprising DNA wrapped around a tetramer of histone proteins. Genomic regulation is partly achieved by tuning nuclear signaling according to various chemical marks on histone tails. These so-called histone post-translational modifications (HPTMS) are dynamic, which is reflected by the existence of enzymes responsible for both the deposition and removal of the modification[24]. Acetylation of lysine residues on the histone tail is associated with an active gene state[25]. The enzyme histone acetyltransferase (HAT) is responsible for the establishment of this mark and adds a negatively charged acetyl group onto a positively charged histone. Consequently, the positive charge of the histone is neutralized, which is believed to reduce its affinity to the negatively charged DNA. This causes genomic DNA to become more accessible to transcription factors, enabling transcriptional activation. Histone acetylation can be reversed by histone deacetylases (HDACs), which causes DNA to be tightly packaged with histones into closed chromatin[25]. This is the inactive state of chromatin and is thus mainly associated with gene silencing. Methylation of histones can lead to gene activation or inactivation, depending on the affected site, and can be reversed by histone demethylases (HDMs)[26]. The DNA fraction of chromatin can be modified by methylation, where methyl and hydroxymethyl groups are deposited on the cytosine within CpG dinucleotides. A CpG island is the most reliable predictor of presence of a gene and is associated with its 5' end, where methylation is believed to repress transcription by blocking the ability of TFs to bind their recognition sequences.

Together, modification of genomic DNA and associated proteins add epigenetic information to different layers of the chromatin substrate. As such, epigenetic modifications, alongside molecular processes such as phase separation and molecular motion, influence chromatin organization within the nucleus, affecting genome interactions and function. The implications of diverse chromatin states extend to various cellular processes, from governing gene expression to influencing DNA replication, repair mechanisms, and cellular differentiation.

1.1.4 Genome topology across scales

As discussed under *1.1.2 Spatial organization of the nucleus*, chromatin types are separated within nuclear space according to their respective levels of activity. The genomic sequences represented in those different chromatin types can largely be translated to compartments A and B, where A corresponds to more active euchromatin and B corresponds to inactive heterochromatin. These compartments are data-driven descriptors stemming from the observation that a distinct plaid- or checkerboard-like pattern emerges when visualizing pairwise contacts obtained with Hi-C. This technique is a genome-wide extension of chromosome conformation capture (3C) methods that measure contacts of DNA sequences based on their proximities in nuclear space. The checkerboard—squares of intermittently high and low pairwise contact frequencies—implies that the genome can be divided into two separate sets of regions (*i.e.*, compartments) that preferentially contact other sequences within the same set, rather than in the other set, even at distal regions[27]. Further sub-typing of compartments has been proposed based on specific combinations of epigenetic marks[28], and these chromosomal domains tend to correlate with nuclear regions[29]. Some regions may not interact with either A or B compartment[30,31], as is the case with Polycomb chromatin[32,33], especially during early mammalian development[34]. In addition, separate compartments may be formed upon genomic rearrangements in response to nuclear signaling, such as during DNA repair[35,36].

On a smaller scale, within the same compartment, regions along the chromosome tend to form more frequent contacts within a region than outside it. These consecutive regions, with a median size of ~1 Mb, are termed topologically associating domains (TADs)[37,38]. TADs are formed by the process of cohesin-mediated loop extrusion, where a DNA "loop" grows progressively until halted by boundary elements such as CTCF. Recent high-resolution microscopy studies have shown that intra-TAD spatial distances are smaller than inter-TAD ones, confirming proximity-based sequencing data[39,40]. Importantly, these distances vary greatly between cells, indicating that TADs (as we have come to describe them based on population-based Hi-C data) reflect an enrichment of contacts within bordered regions, not stable structures. Nevertheless, TADs can facilitate functional interactions between distal genetic elements, such as enhancer-promoter looping during gene activation, although the exact contribution of loop extrusion to transcriptional control is nuanced and context-dependent (reviewed in[41]). I will further elaborate on the particular role of TADs as functional units in the DNA damage response under *1.2.2 Structure of nuclear repair domains*.

1.1.5 Compartmentalization and motion

The larger phenomenon of compartmentalization occurs across all kingdoms of life and across different scales. Within the eukaryotic nucleus, compartmentalization

mainly refers to the observation that active and inactive chromatin types are spatially separated. Formation of sub-nuclear bodies, such as the nucleolus, is considered compartmentalization as well. Both chromatin and organelle compartments can be understood as being driven by the biophysical process of phase separation. It provides a possible explanation for spatial clustering of co-regulated sequences, thereby enabling quick, coordinated control of many loci over large genomic distances.

Compartmentalization by phase separation

Biological phase separation as referred to in the context of sub-nuclear compartments is the process of a solution separating into two (or more) immiscible liquids. Such liquid-liquid phase separation (LLPS) requires that at least one of the components have the tendency to interact more with itself than with the other component. In most biological systems that undergo LLPS, it is believed that protein-protein interactions are the driving component: when proteins have intrinsically disordered regions (IDRs) or domains of low complexity, they form multiple weak and short-lived interactions that can result in the emergence of a liquid-like phase, *i.e.*, formation of membrane-less compartments. This is in contrast to proteins with enzymatic functions and the majority of protein complex formation, wherein well-folded proteins interact specifically with one other. Intriguingly, most known examples of phase-separating proteins do not demix at physiological concentrations; rather, the measured endogenous concentration is much lower, suggesting that additional factors are required. Consistent with that hypothesis, many proteins capable of LLPS also have RNA-binding properties and are known to interact with RNA substrates[42]. Important reports of LLPS and subsequent compartmentalization are, among others: the formation of heterochromatin domains mediated by liquid demixing[43-45], droplet-like enhancer assembly and their function in gene control[46,47], induction of a separate liquid phase at sites of damaged chromatin[48-52]. For thorough reviews of phase separation in the context of chromatin, including careful discussion of functional implications, see [53-56].

Chromatin motion

Being a highly dynamic entity, chromatin undergoes both local and global movements. Local motion is often frequent and rapid, while longer-range motion is occasional and slow, likely in concordance with underlying biological processes such as transcriptional activity or repair of the DNA fiber (reviewed in [57,58]). If we consider chromosomes as polymeric chains, the local motion of genomic loci is mainly dictated by the elasticity of their neighboring segments[59]. See [60,61] for an introductory background to biophysics of chromatin motion and polymer modeling. Such models of the chromatin fiber can be experimentally tested by imaging chromatin in live cells (usually for seconds to minutes). Single-locus tracking studies give insights into the trajectory and type of motion undergone, for example by a gene in response to transcriptional perturbation, while genome-wide measurements (of fluorescently

tagged histones) provide information on the global constraints to nuclear territories. Recent technology to manipulate genomic loci in live cells showed that chromatin is more fluid than previously thought, its motion only somewhat hindered by topological effects and crowding[62].

The degree to which chromatin loci explore the nuclear environment can be distinguished into three classes: 1) free diffusion *i.e.,* Brownian motion; 2) confined sub-diffusion, where particles move within a limited space; and 3) directed motion, where movement exceeds diffusion in a particular direction. These classes of motion can be assessed according to mean-square displacement analysis of particles (see [63]).

In the case of sub-diffusion, chromatin mobility is largely constrained passively, by tethering of the chromatin fiber to structural components such as the NL. The formation of loops may also act as brief anchors for chromatin, such that their local motion is inhibited. In contrast, enhancement of chromatin mobility is largely driven by active processes. Any ATP-fueled process acting on the DNA, such as nucleosome remodeling, loop extrusion, transcription, replication, or repair, is expected to affect both local and global movements of chromatin. I will place the general concept of chromatin motion into the context of genome maintenance under *1.2.3 Movement of DSBs within the nucleus*.

In sum, these mechanisms are intrinsically connected: chromatin structure and nuclear organization can be both cause and consequence; compartments and TADs co-exist but are formed by different mechanisms; tethering and topology place constraints on genome function to various extents, altogether influencing nuclear processes and cellular outcomes.

1.2 DNA repair in the context of nuclear organization

As is the case for all other DNA-related processes, damage and repair of DNA occur within nuclear compartments; hence, chromatin type and topological conformation of the genomic region can influence repair outcome. In turn, genomic rearrangements and altered compartmentalization have also been described as a consequence of DNA damage and repair[35,36,64,65] (reviewed in a.o. [66,67]).

1.2.1 DNA double-strand break damage and repair

Damage to DNA poses a significant threat to genomic integrity. It necessitates intricate repair mechanisms to prevent mutations and structural anomalies in chromosomes, such as insertions, deletions, duplications, or translocations, which can drive oncogenesis and other neurodegenerative disorders. The work in this thesis focuses on double-stranded breaks (DSBs) that completely sever the DNA molecule and are

therefore particularly toxic to the cell. DSBs arise endogenously during processes such as replication and transcription, and can also be induced exogenously such as by ionizing radiation or chemotherapeutic agents. In eukaryotic cells, genotoxic lesions trigger a signaling cascade called the DNA damage response (DDR), leading to further recruitment of proteins that facilitate repair of damage and, as such, maintenance of genomic integrity. The DDR can also result in cell cycle arrest or cell death, e.g., if the damage load is too high, or if repair fails. It remains poorly understood which circumstances lead to particular cellular outcomes, in large part because the genomic locations of damage are difficult to map. To that end, *Chapter 5* describes the first genome-wide detection of DNA repair proteins by sequencing in single cells. This approach enables detailed analysis of DSB repair in the context of genome organization and chromatin.

DSBs can be processed by two primary repair pathways: non-homologous end-joining (NHEJ) and homology-directed repair (HDR). NHEJ is a flexible pathway that ligates minimally processed broken ends, while HDR uses the intact (post-replicative) sister chromatid as a template for precise repair after extensive end resection and homology searching. Additionally, microhomology-mediated end joining (MMEJ) represents an alternative repair pathway that utilizes short homologous sequences flanking the DSBs for alignment and repair, often resulting in small deletions at the junction site.

See e.g., [68-71] for comprehensive reviews on DDR signaling and DSB repair. Which pathway is used to repair a broken locus is dependent on and regulated by various factors, such as the type of lesion and its DNA end structure; nuclear compartment, local chromatin modifications, and transcriptional status of the locus; and the phase of the cell cycle.

Chromatin restructuring during the DDR therefore involves multiple histone-modifying enzymes and chromatin remodelers that orchestrate histone variant exchange, PTMs, and nucleosome sliding or eviction. Recent studies have described approaches to more systematically investigate DSB repair in different chromatin contexts, using genome-wide CRISPR/Cas9-mediated cutting[72-74] and targeting of specific chromatin features[75]. These works concluded that NHEJ is generally favored in euchromatin, while MMEJ is more active in specific heterochromatin contexts, both constitutive (H3K9me2/3) and facultative (H3K27me3). The balance between NHEJ and MMEJ is particularly fine-tuned, and may vary more than five-fold across different local chromatin features. Perhaps counterintuitively, it appears that the toxicity of a DSB does not depend on chromatin; rather, the number of DSBs (including off-target effects) most significantly affects cell proliferation[75]. See [76-78] for comprehensive reviews on the biochemical modifications to chromatin in response to DNA damage.

1.2.2 Structure of nuclear repair domains

Protein foci and associated signaling cascade

Repair proteins bind to DNA at or surrounding a damaged locus in order to exert their functions. I will use the term "repair domain" to refer to a segment of DNA that is occupied by repair protein molecules. Repair domains (without consideration of underlying genomic sequences) were first observed as discrete "foci" of protein accumulation inside the nucleus[79]. The formation of such foci occurs in a highly ordered manner, in terms of molecular signaling as well as spatial assembly. Briefly: early DDR signaling events trigger phosphorylation of histone H2AX (denoted as γH2AX), which serves as the specific binding platform for MDC1 (mediator of DNA damage checkpoint protein 1). MDC1 then coordinates the recruitment of downstream DDR factors, including 53BP1 and BRCA1, which also accumulate in foci but have antagonizing roles in DSB repair pathway usage (reviewed in e.g., [80-82]).

Formation and function of genomic repair domains

The advent of high-throughput sequencing enabled genome-wide mapping of DNA-associated proteins, providing first insights into the binding patterns of various repair machineries. One particularly intriguing aspect is the propagation of DSB hallmarks such as γH2AX over considerable distances from the break site (up to ~1 Mb in either direction), forming extensive repair foci. It has been proposed that γH2AX spreads across entire TADs, supported by observations of genome-wide increases in cohesin binding[83], and strengthened TAD borders[65] in response to DNA damage. Concordantly, γH2AX flanks the genomic regions surrounding a DSB in a manner that is constrained by cohesin [84] and dependent on chromosomal contacts[85]. These findings were generalized in a model proposing that TADs are functional units of repair activity wherein proteins accumulate until a boundary element (such as CTCF) is reached[86]. Anchored by a DSB on one side, cohesin-mediated loop extrusion is thus a process that actively contributes to spreading of repair-associated molecules and the establishment of a repair-prone chromatin environment. In *Chapter 5*, I describe different scenarios of repair protein spreading. At a given DSB location, spreading can vary extensively among cells, likely following heterogeneity in pre-existing genome topology. In addition, I report mapping of structural and repair proteins in the same cell, showing that boundary elements demarcate individual repair domains.

Additionally, super-resolution microscopy has revealed that repair foci are not uniformly occupied by protein. For instance, 53BP1 forms a distinct ring-shaped structure, consisting of several nano-domains[87]. Similarly, γH2AX signal around a DSB site is organized into distinct nano-domains that together form a spatial cluster (conventionally seen as one focus)[88,89]. Importantly, each nano-domain corresponds to a single TAD, establishing the relationship between DNA repair and genome topology.

The circular organization of protein nano-domains is proposed to have a role in limiting DNA end resection, which I consider further under *Chapter 8 – Discussion*. In sum, pre-existing genome topology has a profound impact on the spatial organization of repair proteins and, ultimately, the efficiency of repair processes.

Phase separation and other compartments

Membrane-less compartments formed by LLPS also have a role in stabilizing repair domains. LLPS-driven condensates aid in the assembly of repair proteins like 53BP1 that have disordered regions and self-associate[49]. Heterochromatin-associated proteins such as HP1 and Polycomb complexes also have phase separation properties and can be recruited to DSBs[90,91], potentially playing a role in mobility of damaged chromatin. The physiological function of such localized protein concentrations in the form of foci remains an active area of study, with speculations that they might aid repair efficiency or shield aberrant recombination.

Understanding the coordination between repair domains and other compartments, such as nuclear speckles and nucleoli, is becoming increasingly essential. These domains overlap both spatially and functionally, indicating potential interactions and shared mechanisms between DNA repair, transcription, and replication. Further exploration of the dynamic assembly and coordination of these multifunctional domains within the 3D nuclear landscape will prove critical for understanding genome maintenance in both normal and pathological contexts.

1.2.3 Movement of DSBs within the nucleus

Damaged genomic loci display altered mobility compared to the intact genome. Experimental evidence regarding DSB (im)mobility is highly varied and seemingly contradictory across model systems, pathways, cell cycle, and nature of the break. In this section I highlight different types of movement, focusing on observations of chromatin dynamics in mammals, with findings in yeast and fly indicated where relevant. Generally, increased mobility is associated with HDR, while the NHEJ machinery may constrain movement.

Homology search

During HDR, broken DNA ends are resected to generate 3' single-stranded (ss) DNA overhangs. These become coated with Rad51, forming filaments that facilitate the search for homologous sequences in the nuclear environment. While a suitable sequence may be found on the homologous chromosome, the intact sister chromatid is the preferred template due to its proximity to the damaged site. Nevertheless, DSBs explore an area more than two-fold larger than unbroken loci[92], displaying both directed and confined motion[93].

DSB relocalization

Chromosomal movements may relocalize broken DNA to sites (more) permissive for their repair, particularly when DSBs arise in repetitive sequences and need to migrate further distances to avoid aberrant recombination. As described under *1.1.3. Chromatin structure*, repetitive regions are often heterochromatic, and relocalization therefore often constitutes exclusion from structures that harbor constitutive heterochromatin. An early study in *Drosophila* showed that heterochromatic breaks rapidly move outside of their HP1a domain, after which they associate with HDR protein Rad51[94]. Similarly, loss of H3K9me3 ensures that breaks in heterochromatin may move outside the domain. DSBs in facultative heterochromatin relocate outside of their Polycomb bodies, concomitantly (or after) losing the H3K27me3 mark around the break[95]. It was recently proposed that local accumulation of H3K9ac (a hallmark of accessible chromatin) promotes movement and HDR of heterochromatic breaks[96]. Most importantly, this study highlights the importance of the pre-existing chromatin context in which damage is induced. DSBs induced in mouse pericentromeres move to the very edge of the chromocenter before Rad51 recruitment and further repair[97]. Notably, these breaks only move when undergoing HDR in G2; breaks repaired by NHEJ in G1 remain positionally stable within the heterochromatin.

In some instances, chromatin movement towards the nuclear periphery may position DSBs near nuclear pores, which act as hubs for repair in yeast[98,99], fly[100], and mammalian cells[101]. The latter study showed that DSBs at the NL are preferentially repaired by (non-canonical) NHEJ, while the nuclear pore environment favors HDR. Notably, heterochromatic DSBs generated at the NL fail to migrate to nuclear pores in human cells[101]. Similarly, DSBs on ribosomal DNA are rearranged such that the breaks are situated at the nucleolar periphery to undergo HDR, regardless of the cell cycle phase[102]. The interplay between rDNA damage, subsequent transcriptional inhibition, nucleolar reorganization, and repair pathway usage remains a key area for future research. For extensive reviews of nuclear reorganization in response to heterochromatic damage, see [103-105] and [106] focusing on rDNA.

DSB clustering and compartmentalization

In response to damage, nuclear rearrangements may also occur on a more global, genome-wide, scale. DSBs are not only mobile, they are also capable of "clustering" in the nucleus, forming common repair foci. The phenomenon of repair foci clustering was first observed by microscopy[107]: over time, a reduction in the number of γH2AX foci and an increase in their size supported the "breakage-first" theory of chromosomal translocations. Live-cell imaging further showed fusion of distinct repair foci by fluorescent tagging of repair proteins 53BP1[36,92,108] and Rad52[109]. High-throughput sequencing experiments showed clustering on the genomic (rather than protein) level as well, by measuring chromosomal contacts in damaged and undamaged conditions[35,36,64]).

However, no experimental evidence has been presented that directly couples repair foci clustering and positional information of DNA breaks. In *Chapters 5 and 6*, I set out to map repair proteins at know DSB loci, and show that cluster-forming breaks are occupied in a coordinated fashion. Furthermore, the genome-wide coalescence of DSBs in *cis* as well as in *trans* is associated with compartment switching, whereby damaged chromatin switches from B to A compartment. A novel damage (D) compartment may even be formed in G1, harboring mostly DDR genes[36].

Mechanical forces involved in DSB mobility

Mobility of damaged chromatin at local and global scales is driven by various forces, including mechanical energy (converted from chemical energy by motor proteins). Several studies have explored the role of cellular and/or nuclear skeleton proteins in DSB movement, thereby impacting repair efficiency. Following genotoxic treatment, nuclear actin and tubulin can polymerize into filaments[110,111] and promote DSB clustering in mammalian cells[64,109]. The relocalization of heterochromatin breaks to the nuclear periphery described above is also dependent on actin and myosin[112]. In addition to nuclear forces, DSB dynamics are influenced by cytoskeletal elements. The linker of nucleoskeleton and cytoskeleton (LINC) complex is situated in the nuclear membrane and transmits forces exerted by cytoplasmic microtubules and actin filaments to the chromatin in the nucleus. DSB movement in yeast, fly, mouse, and human is partly dependent on LINC subunits[64,99,113-115], although the mechanisms appear to differ and remain poorly understood. More broadly, active (directed) and passive (diffusive) motion are yet to be conclusively assigned to the multi-scale chromatin reorganizations in response to damage (see further under *Chapter 8 – Discussion*).

1.2.4 Nuclear organization in oncogenesis and treatment

The implication of nuclear organization in diseased states underscores its critical role in maintaining genomic stability and cellular homeostasis. As illustration, the position of a chromosome inside the nucleus is directly related to segregation error frequency and micronucleus entrapment[116]. Peripheral chromosomes missegregate more often, resulting in non-random aneuploidy. Since such errors are hallmarks of cancer, this establishes a link between nuclear organization, mitotic fidelity, and pathological outcomes. Relatedly, genome reorganization in response to damage—as is known to occur during oncogenesis—can lead to acquired resistance to chemo-therapeutic drugs[117]. In this setting, detachment from the nuclear periphery induces insensitivity to Taxol, which is a microtubule stabilizing agent that promotes mitotic arrest and cell death[118].

Notably, lamina-associated regions have a higher somatic mutation rate compared to the nuclear interior, along with distinct mutational signatures[119]. However, the relationship between these mutations and repair proteins remains unclear. A recent study reported a direct interaction between nuclear lamina component Lamin B1 and 53BP1[120]. This interaction appears to be dynamic, with Lamin B1 sequestering 53BP1 until damage occurs. When (highly) overexpressed, Lamin B1 maintains its binding with 53BP1, preventing recruitment to sites of damage and delaying repair. In another study, Lamin A/C was found to promote repair by binding 53BP1 and ensuring a nucleoplasmic pool[121]. Further, post-translational modifications of Lamin B1 can regulate its function, as is the case for K134 acetylation, which inhibits 53BP1 recruitment to DSBs[122]. Although no direct interaction has been reported between 53BP1 and the LINC complex, mechanical forces (mediated by LINC) still influence chromatin mobility. These forces may have a larger effect on damaged loci, because their association with 53BP1 changes the flexibility of the chromatin fiber. Such findings are just beginning to uncover the links between repair condensates, the nuclear cytoskeleton, and mechanobiology.

Hence, deciphering the role of nuclear organization has implications for our fundamental understanding of DNA damage and repair and subsequent disease progression, as well as exploitation of genotoxic mechanisms for therapeutic ends.

1.3 Intercellular heterogeneity in form and function

Evolution from unicellular organisms into much larger and complex multicellular organisms is enabled by specialization in cellular function. Many interconnected layers of regulation drive the emergence of highly variable cellular dynamics, eventually fostering functional diversity and adaptive potential.

Characterizing cellular state

When studying a process at the cellular level, there are multiple aspects of cellular identity that can be considered. Cellular state can be described as existing along a continuum of phenotypes that progress over time and space[123]. Measurable phenotypes could include factors like physical morphology, spatial position, movement, and the micro-environment. Additionally, molecular features—e.g., a cell's genetic and epigenetic make-up, transcriptional output, and protein abundance—can aid in characterizing the regulatory mechanisms at play during processes such as cellular signaling and metabolism. In turn, those dynamics inform more coordinated undertakings in development, homeostasis, and disease.

Epigenetic regulation during differentiation

One of the most illustrative examples of intercellular variation is the notion that distinct gene expression programs are required for the proper establishment of cell

fate and lineage specification. Transcription is thus highly heterogeneous among cells, and this variability is commonly used to assign cellular identity. However, in order to fine-tune transcriptional output, genes (largely identical between cells) are subject to epigenetic regulation, which is a heterogeneous process itself. *Chapter 3* describes our approach to jointly measure gene expression and histone modifications during cellular differentiation. To further delve into epigenetic regulation, *Chapter 4* explores formation of different types of heterochromatin in relation to the nuclear periphery.

Phenotypic variation among malignant cells

Relatedly, transcriptional changes can also signify cellular states in a pathological scenario. For instance, after chemotherapy, cancer cells exhibit heterogeneous transcriptional responses that correspond to distinct cellular fates: apoptosis, cell cycle arrest, or stress resistance[124]. On a deeper level, such outcomes are responses to DNA damage. Genotoxicity is induced in cells with the aim of overloading the DNA repair machinery to the extent of causing cell death. Why cells react differently to presumably equal loads of damage remains an important outstanding question. While several sources of heterogeneity are likely to be involved, one unexplored aspect is the location of damage in the genome. Many break-inducing agents are blind to DNA sequence, leading to the assumption that the damaged positions are stochastic.

The single cell as origin of disease

When DNA damage remains unrepaired or is incorrectly restored, this can lead to many different diseases, mainly cancer and neurodegenerative disorders. The concept that cancer originates from the abnormal proliferation of individual cells rather than whole tissues or organs was proposed in the mid-19th century[125]. Building upon this pioneering work, Boveri put forward the hypothesis of chromosomal imbalance as the primary driver of tumorigenesis[7] (see also *1.1.1 Historical perspective*). Together, this laid the foundation for our modern understanding of cancer as a disease rooted in genetic alterations at the level of a single cell. In accordance, studies in the past 15 years have shown that tumor evolution is reflected in single-cell DNA mutations[126], and have revealed substantial genetic mosaicism among individual cancer cells[127,128].

Improved understanding of the mechanisms underlying intercellular heterogeneity—from gene regulation to the DNA damage response—provides critical insights into both normal cellular function and the origins of disease, setting the stage for the technological advancements discussed in the next section. Against this backdrop, the next section introduces the methods and technologies that now enable us to dissect these cellular differences in unprecedented detail.

1.4 Technological advances

The various sources of heterogeneity described above pose challenges for traditional methods that are based on large populations of cells. This necessitates the adoption of single-cell measurements to unravel the complexities inherent in cellular functionality. Below I describe an overview of the state of the art in interrogating chromatin at the level of individual cells. This section is not meant to be exhaustive; it focuses on "gold standard" approaches, recent developments in the field, and background necessary to evaluate the technical innovations presented in this thesis. For extensive reviews on methods to assay 3D genome organization see [129] (up until 2021) and [130] (with a focus on imaging). In *Chapter 7*, I detail recent progress in sequencing technologies to profile chromatin in single cells.

1.4.1 Chromatin imaging

As mentioned under *1.1.1 Historical perspective*, the organization of chromatin types and individual territories had long been observed by light microscopy, but could only be decisively proven using fluorescent *in situ* hybridization (FISH) and imaging of known DNA sequences. Recent advances in super-resolution light microscopy now enable genome-scale imaging at a resolution of 1-3 Mb, chromosome-specific scale of 25-50 kb, and 2 kb for some regulatory regions (ORCA[131], DNA-MERFISH[132], DNA seqFISH[133], among others). This allows for measurements of position, shape and distance between genomic loci of interest. An alternative FISH-based approach is OligoFISSEQ[134], which couples *in situ* sequencing (originally designed for transcript detection[135]) with Oligopaint probes, potentially imaging thousands of cells genome-wide. To increase sensitivity of measuring DNA loci in close proximity without the need for super-resolution microscopy, DNA FISH has been coupled with Förster resonance energy transfer (FRET)[136], since FRET relies on signal detection within 1-10 nm. Resolution can also be improved by enlarging the sample compared to the optical setup, as applied in expansion microscopy (ExM[137]) combined with DNA FISH at single gene loci[138]. In addition to approaches relying on DNA sequence, general organizational principles of fixed chromatin can now be revealed by confocal microscopy, achieving a resolution of 15 nm. By combining chromatin expansion microscopy (ChromExM) with stimulated emission depletion (STED) imaging, resolution approaches <3 nm, visualizing even individual nucleosomes of 10 nm in size[139].

Live-cell imaging presents opportunities and limitations of its own. In the context of chromatin dynamics, notable developments have been the integration of repressor-operator systems at specific loci, and, more recently, CRISPR/Cas9-based targeting systems. Such tools were used to study mobility of tagged genes and regulatory elements as well as transcription-unrelated chromatin looping.

On a more global scale, DNA staining or tagging of pan-nuclear proteins can be used in combination with imaging approaches to measure kinetics and directionality of movement (see also *1.1.5 Compartmentalization and motion*). For instance, high-resolution diffusion mapping (Hi-D) showed sub-diffusive movement of DNA within domains of 0.3-3 μm, with notable heterogeneity in motion among the domains[140]. Fluorescence lifetime imaging (FLIM) coupled with FRET of histones generates maps of chromatin dynamics, showing (de)compaction at a spatial scale of 1-10 nm[141,142].

DNA FISH approaches have been integrated with other modalities, namely antibody-based protein localization and RNA FISH, thereby exploring genome topology, nuclear organization, epigenetic modifications, and transcription in the same cell. Similarly, FLIM-FRET can be multiplexed with immunofluorescence[143] to probe various combinations of proteins across cellular processes.

1.4.2 Genomics

High-throughput sequencing is the gold standard for measuring the genome and its associated features. Distinctions can be made on the basis of i.a., what aspect of the DNA is quantified, whether there is targeted capture (or enrichment) of certain regions, and (capacity for) single-cell implementation.

Chromatin conformation

Genome folding across scales is most commonly studied with so-called C-based techniques, which originated with chromosome conformation capture (3C)[144]. The assay relies on proximity ligation of fixed chromatin: genomic loci in close contact will be preferentially detected. Numerous extensions and variations of C-based methods exist, most importantly 4C[145] (which measures contacts of one chosen locus versus all) and Hi-C[37] (measuring all versus all, thus, genome-wide). Chromatin conformation driven by specific proteins of interest can be interrogated by combining a C-based approach with immunoprecipitation, as done in ChIA-PET[146] or HiChIP[147]. Enrichment at selected loci of interest can be achieved with probe-based capture[148,149]. Because these methods quantify signal averaged over many cells, the contacts can only be called between two loci at once, *i.e.*, pairwise. Multiway or higher-order contacts—formed between more than two genomic loci—are not assayed effectively with the standard C-based methods. Adaptations have therefore been introduced that intend to retain single-molecule or allele-specific resolution, namely concatemer ligation assay (COLA)[150], Tri-C[151], and multi-contact 4C[152], albeit in a locus-specific manner. Genome-wide, higher-order, chromatin conformation can be measured with ligation-free methods, namely genome architecture mapping (GAM)[153], split-pool recognition of interactions by tag extension (SPRITE)[154] and chromatin interaction analysis via droplet-based and barcode-linked sequencing

(ChIA-Drop)[155]. In addition, the radial positioning of DNA inside the nucleus can be mapped with genomic loci positioning by sequencing (GPSeq)[33], which uses gradual chromatin digestion from the nuclear periphery towards the center. DNA positioned at the lamina or nucleolus can be detected with DamID[156], which confers specificity via DNA-bound proteins, in this case subunits of the respective structures. GAM inherently extracts information from individual nuclear slices, and single-cell adaptations exist of Hi-C[157], SPRITE[158], and DamID[159], with varying limitations (see *Chapter 7 – Technology Review* and *Chapter 8 – Discussion*). Below, I briefly introduce DamID, and refer to *Chapter 2* for its implementation to study LADs in individual mammalian cells.

Epigenetic modifications and DNA-binding proteins: single-cell approaches

The principle of DamID to map contacts between the genome and sub-nuclear structures applies to DNA-binding proteins more broadly, including chromatin proteins and associated regulatory factors. DamID relies on the tethering of bacterial methyltransferase Dam to a protein of interest (POI)[156]. *In vivo* expression of the fusion protein thereby deposits exogenous methylation on proximal genomic loci, which are then selectively amplified and detected by sequencing. As such, DamID is an alternative to techniques that use antibodies to confer specificity, such as the gold standard chromatin immunoprecipitation (ChIP). As part of the epigenetic repertoire, histone modifications have been thoroughly assayed by ChIP, classifying the genome into functional elements based on their chromatin states[160]. Previously, DamID was inadequate for studying chromatin marks, because post-translational modifications themselves cannot be genetically engineered and introduced *in vivo*. *Chapter 3* presents an adaptation wherein Dam is fused to protein domains that selectively recognize hPTMs, making the technique suitable for epigenetic studies (EpiDamID)[161].

Single-cell implementations have been developed of IP as well as nuclease- or transposase-based approaches: scChIP-seq[162,163], CUT&RUN[164], scChIC[165] and sortChIC[166], CoBATCH[167] and CUT&Tag[168-170], and ChIL-seq[171]. Immunocleavage is especially successful in ultra-low-input settings since it does not rely on pulldown of material, thereby minimizing sample loss. Recently, efforts have been made to quantify many epigenetic factors simultaneously in single cells, although this remains highly challenging. Approaches that measure multiple chromatin-associated proteins currently all rely on antibodies, mostly in combination with tagmentation: Multi-CUT&Tag[172], MulTI-Tag[173], nano-CT[174], NTT-seq[175], and uCoTarget[176]. Computational deconvolution of two multiplexed hPTMs was demonstrated with scChIX-seq[177]. We developed MAbID, an orthogonal approach relying on restriction enzyme digestion that can measure up to six targets simultaneously[178]. Additionally, as described in *Chapter 4*, we integrated single-cell DamID and sortChIC, to profile two distinct chromatin proteins or measure the same target with different approaches.

Single-cell multi-modal epigenomics

Across the central dogma of molecular biology, various molecular characteristics (*i.e.,* modalities) are hierarchically interconnected: among others, one can measure the genome (including its many variants, extrachromosomal copies, and replication timing), epigenome (including DNA accessibility, DNA modifications, and chromatin marks), (epi)transcriptome, translatome, (phospho)proteome, and metabolome. These modalities are variable across cells, which warrants the need for single-cell approaches to study each modality individually. However, to fully understand the interplay and potential dependencies between them, multi-modal measurements at the single-cell level are necessary.

While methods combining DNA accessibility and/or RNA are well established (and even commercialized (10x Multiome), simultaneous measurement of DNA-associated proteins with other analytes presents significant hurdles, mainly due to technical incompatibilities in extraction and amplification. We developed the first sequencing-based approach to quantify chromatin proteins and transcriptional output in single cells, termed scDam&T-seq [179]. With the establishment of EpiDamID described in *Chapter 3*, we extended the multi-modal readout to include post-translational histone modifications and RNA. *Chapters 3* and *7* give a comparison of experimental design and output of similar methods. Notably, a more recent approach reports simultaneous measurement of multiple hPTMs and RNA (uCoTargetX [176]).

For extensive (but not exhaustive) reviews of single-cell multi-modal and spatial omics, see [180,181]. For a detailed overview of single-cell epigenomics including progress since publication of those reviews, I refer to *Chapter 7*. In *Chapter 8*, I further consider technological adaptations and implementations of single-cell and single-molecule approaches as relevant for the research in this thesis.

Scope of dissertation

This dissertation explores intercellular variation in genome organization, regulation, and repair. My research advances two areas in particular: (1) technological innovation of single-cell sequencing tools for chromatin profiling and analysis of DNA-binding proteins, and (2) investigation of heterogeneity in DNA repair and its interplay with genome topology.

The work focuses on technology- and discovery-driven insights: it addresses fundamental regulatory principles using molecular biology and engineering. While highly relevant for the development of therapeutic strategies and companion diagnostics, it does not aim to provide direct clinical applications. Additionally, although I incorporate microscopy-based approaches throughout the thesis, a workflow for fully integrated single-cell imaging and sequencing remains a future goal.

Chapter 2 presents a detailed experimental protocol to perform single-cell DamID sequencing. It focuses on studying nuclear organization through genome-wide mapping of nuclear lamina-associated domains. The protocol implements multiplexing to improve throughput and reduce technical variability, making large-scale single-cell chromatin mapping more efficient and cost-effective.
Chapter 3 describes EpiDamID, a novel approach for detecting post-translational histone modifications using genetically engineered chromatin binders. In combination with scDam&T-seq, this enables joint profiling of chromatin state and transcriptional output in single cells, and remains the only antibody-free technique to do so.
Chapter 4 details the development of a dual readout system combining DamID and ChIC, with two distinct manifestations: 1) As a multi-factorial method, two different DNA-binding proteins can be simultaneously profiled. 2) As a multi-modal method, the same protein can be measured with two different temporal resolutions, namely the cumulative signal of DamID versus the snapshot signal of ChIC, contributing a unique view on binding dynamics.
Chapter 5 presents the first genome-wide single-cell maps of DNA repair protein localization at double-strand breaks. By applying the techniques described in the previous chapters to this novel context, I present a single-cell genomics toolbox to study heterogeneity in DNA repair, from epigenetic chromatin context to transcriptional changes in response to damage.
Chapter 6 builds upon the data generated in Chapter 5 and investigates single-cell coordination of repair proteins at sites of damage. I present evidence that damaged loci cluster in space, forming higher-order contacts that are coordinately bound by repair protein, indicating cooperative behavior.

Chapter 7 provides a comprehensive overview of recent progress in single-cell and single-molecule chromatin profiling, positioning the methodologies developed in this dissertation within the broader landscape of genome regulation research.
Chapter 8 contextualizes above findings, first addressing the technical features of current single-cell techniques. In relation to Chapter 3, I briefly discuss epigenetic regulation with a focus on Polycomb repressive complexes in development and disease. Then I consider my work on DNA repair profiling and genome organization, highlighting recent insights from other work and proposing future directions. Lastly, I reflect on the technological advancements presented in this dissertation, outlining potential short-term applications and long-term innovations.

In sum, by advancing the scope of single-cell sequencing methodologies, this thesis provides new tools and insights into genome organization and DNA repair, laying the foundation for future studies into the chromatin-based regulation of genome stability.

References

1. Miescher, J.F. *Ueber die chemische Zusammensetzung der Eiterzellen*, (1871).
2. Miescher, J.F. *Die Spermatozoen einiger Wirbelthiere: ein Beitrag zur Histochemie*, (1874).
3. Zacharias, E. *Über die chemische Beschaffenheit des Zellkerns*, (Arthur Felix, 1881).
4. Flemming, W. *Zellsubstanz, kern und zelltheilung*, (Vogel, 1882).
5. Boveri, T. Über mehrpolige mitosen als mittel zur analyse des zellkerns. *Verhandlungen der physikalisch-medizinischen Gesellschaft zu Würzburg* **35**, 67-90 (1902).
6. Sutton, W.S. On the morphology of the chromoso group in Brachystola magna. *The Biological Bulletin* **4**, 24-39 (1902).
7. Boveri, T. *Zur frage der entstehung maligner tumoren*, (Fischer, 1914).
8. Boveri, T. Die Blastomerenkerne von Ascaris megalocephala und die Theorie der Chromosomenindividualitat. *Archiv fur Zellforschung* **3**, 181-268 (1909).
9. Heitz, E. *Das heterochromatin der moose. I Jahrb. wiss. Bot. 69: 762–818*, (1928).
10. Heitz, E. Chromosomenstruktur und Gene. *Zeitschrift für Induktive Abstammungs-und Vererbungslehre* **70**, 402-447 (1935).
11. Cremer, T. *et al.* Analysis of chromosome positions in the interphase nucleus of Chinese hamster cells by laser-UV-microirradiation experiments. *Human genetics* **62**, 201-209 (1982).
12. Speicher, M.R., Ballard, S.G. & Ward, D.C. Karyotyping human chromosomes by combinatorial multi-fluor FISH. *Nature Genetics* **12**, 368-375 (1996).
13. Schröck, E. *et al.* Multicolor spectral karyotyping of human chromosomes. *Science* **273**, 494-497 (1996).
14. Levene, P.A. & Bass, L. Nucleic Acids. The Chemical Catalog Company. *Inc., New York*, 294-308 (1931).
15. Chargaff, E. Chemical specificity of nucleic acids and mechanism of their enzymatic degradation. *Experientia* **6**, 201-209 (1950).
16. Watson, J.D. & Crick, F.H. Molecular structure of nucleic acids: a structure for deoxyribose nucleic acid. *Nature* **171**, 737-738 (1953).
17. Venter, J.C. *et al.* The sequence of the human genome. *Science* **291**, 1304-1351 (2001).
18. Nurk, S. *et al.* The complete sequence of a human genome. *Science* **376**, 44-53 (2022).
19. Vertii, A. *et al.* Two contrasting classes of nucleolus-associated domains in mouse fibroblast heterochromatin. *Genome Res* **29**, 1235-1249 (2019).
20. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178-92 (2013).
21. Dechat, T. *et al.* LAP2alpha and BAF transiently localize to telomeres and specific regions on chromatin during nuclear assembly. *J Cell Sci* **117**, 6117-28 (2004).
22. Crabbe, L., Cesare, A.J., Kasuboski, J.M., Fitzpatrick, J.A. & Karlseder, J. Human telomeres are tethered to the nuclear envelope during postmitotic nuclear assembly. *Cell Rep* **2**, 1521-9 (2012).
23. van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P. & van Steensel, B. Cell cycle dynamics of lamina-associated DNA. *EMBO Rep* **21**, e50636 (2020).
24. Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. *Nature* **403**, 41-5 (2000).
25. Grunstein, M. Histone acetylation in chromatin structure and transcription. *Nature* **389**, 349-52 (1997).
26. Agger, K., Christensen, J., Cloos, P.A. & Helin, K. The emerging functions of histone demethylases. *Curr Opin Genet Dev* **18**, 159-68 (2008).
27. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
28. Rao, S.S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665-80 (2014).
29. Belaghzal, H. *et al.* Liquid chromatin Hi-C characterizes compartment-dependent chromatin interaction dynamics. *Nat Genet* **53**, 367-378 (2021).
30. Schwarzer, W. *et al.* Two independent modes of chromatin organization revealed by cohesin removal. *Nature* **551**, 51-56 (2017).
31. Johnstone, S.E. *et al.* Large-Scale Topological Changes Restrain Malignant Progression in Colorectal Cancer. *Cell* **182**, 1474-1489.e23 (2020).
32. Boyle, S. *et al.* A central role for canonical PRC1 in shaping the 3D nuclear landscape. *Genes Dev* **34**, 931-949 (2020).
33. Girelli, G. *et al.* GPSeq reveals the radial

organization of chromatin in the cell nucleus. *Nat Biotechnol* **38**, 1184-1193 (2020).

34. Guerreiro, I. *et al.* Antagonism between H3K27me3 and genome-lamina association drives atypical spatial genome organization in the totipotent embryo. *Nat Genet* **56**, 2228-2237 (2024).
35. Zagelbaum, J. *et al.* Multiscale reorganization of the genome following DNA damage facilitates chromosome translocations via nuclear actin polymerization. *Nat Struct Mol Biol* **30**, 99-106 (2023).
36. Arnould, C. *et al.* Chromatin compartmentalization regulates the response to DNA damage. *Nature* **623**, 183-192 (2023).
37. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).
38. Nora, E.P. *et al.* Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* **485**, 381-5 (2012).
39. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**(2018).
40. Finn, E.H. & Misteli, T. Molecular basis and biological function of variability in spatial genome organization. *Science* **365**(2019).
41. Karpinska, M.A. & Oudelaar, A.M. The role of loop extrusion in enhancer-mediated gene activation. *Curr Opin Genet Dev* **79**, 102022 (2023).
42. Lin, Y., Protter, D.S., Rosen, M.K. & Parker, R. Formation and Maturation of Phase-Separated Liquid Droplets by RNA-Binding Proteins. *Mol Cell* **60**, 208-19 (2015).
43. Larson, A.G. *et al.* Liquid droplet formation by HP1α suggests a role for phase separation in heterochromatin. *Nature* **547**, 236-240 (2017).
44. Strom, A.R. *et al.* Phase separation drives heterochromatin domain formation. *Nature* **547**, 241-245 (2017).
45. Erdel, F. & Rippe, K. Formation of Chromatin Subcompartments by Phase Separation. *Biophys J* **114**, 2262-2270 (2018).
46. Hnisz, D., Shrinivas, K., Young, R.A., Chakraborty, A.K. & Sharp, P.A. A Phase Separation Model for Transcriptional Control. *Cell* **169**, 13-23 (2017).
47. Sabari, B.R. *et al.* Coactivator condensation at super-enhancers links phase separation and gene control. *Science* **361**(2018).
48. Altmeyer, M. *et al.* Liquid demixing of intrinsically disordered proteins is seeded by poly(ADP-ribose). *Nat Commun* **6**, 8088 (2015).
49. Kilic, S. *et al.* Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments. *EMBO J* **38**, e101379 (2019).
50. Frattini, C. *et al.* TopBP1 assembles nuclear condensates to switch on ATR signaling. *Mol Cell* **81**, 1231-1245.e8 (2021).
51. Oshidari, R. *et al.* DNA repair by Rad52 liquid droplets. *Nat Commun* **11**, 695 (2020).
52. Wang, Y.L. *et al.* MRNIP condensates promote DNA double-strand break sensing and end resection. *Nat Commun* **13**, 2638 (2022).
53. Banani, S.F., Lee, H.O., Hyman, A.A. & Rosen, M.K. Biomolecular condensates: organizers of cellular biochemistry. *Nat Rev Mol Cell Biol* **18**, 285-298 (2017).
54. Rippe, K. & Papantonis, A. RNA polymerase II transcription compartments: from multivalent chromatin binding to liquid droplet formation? *Nat Rev Mol Cell Biol* **22**, 645-646 (2021).
55. Hildebrand, E.M. & Dekker, J. Mechanisms and Functions of Chromosome Compartmentalization. *Trends Biochem Sci* **45**, 385-396 (2020).
56. McSwiggen, D.T., Mir, M., Darzacq, X. & Tjian, R. Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. *Genes Dev* **33**, 1619-1634 (2019).
57. Soutoglou, E. & Misteli, T. Mobility and immobility of chromatin in transcription and genome stability. *Curr Opin Genet Dev* **17**, 435-42 (2007).
58. Hinde, E. & Cardarelli, F. Measuring the flow of molecules in cells. *Biophys Rev* **3**, 119 (2011).
59. Doi, M., Edwards, S.F. & Edwards, S.F. *The theory of polymer dynamics*, (Oxford University Press, 1988).
60. Fudenberg, G. & Mirny, L.A. Higher-order chromatin structure: bridging physics and biology. *Curr Opin Genet Dev* **22**, 115-24 (2012).
61. Tortora, M.M., Salari, H. & Jost, D. Chromosome dynamics during interphase: a biophysical perspective. *Curr Opin Genet Dev* **61**, 37-43 (2020).
62. Keizer, V.I.P. *et al.* Live-cell micromanipulation of a genomic locus reveals interphase chromatin mechanics. *Science* **377**, 489-495 (2022).
63. Miné-Hattab, J. & Chiolo, I. Complex Chromatin Motions for DNA Repair. *Front Genet* **11**, 800 (2020).
64. Aymard, F. *et al.* Genome-wide mapping

of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nat Struct Mol Biol* **24**, 353-361 (2017).

65. Sanders, J.T. *et al.* Radiation-induced DNA damage and repair effects on 3D genome organization. *Nat Commun* **11**, 6178 (2020).
66. Carre-Simon, A. & Fabre, E. 3D Genome Organization: Causes and Consequences for DNA Damage and Repair. *Genes (Basel)* **13**(2021).
67. Zagelbaum, J. & Gautier, J. Double-strand break repair and mis-repair in 3D. *DNA Repair (Amst)* **121**, 103430 (2023).
68. Scully, R., Panday, A., Elango, R. & Willis, N.A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* **20**, 698-714 (2019).
69. Hustedt, N. & Durocher, D. The control of DNA repair by the cell cycle. *Nat Cell Biol* **19**, 1-9 (2016).
70. Schwertman, P., Bekker-Jensen, S. & Mailand, N. Regulation of DNA double-strand break repair by ubiquitin and ubiquitin-like modifiers. *Nat Rev Mol Cell Biol* **17**, 379-94 (2016).
71. Panier, S. & Durocher, D. Push back to respond better: regulatory inhibition of the DNA double-strand break response. *Nat Rev Mol Cell Biol* **14**, 661-72 (2013).
72. Brinkman, E.K. *et al.* Kinetics and Fidelity of the Repair of Cas9-Induced Double-Strand DNA Breaks. *Mol Cell* **70**, 801-813.e6 (2018).
73. Schep, R. *et al.* Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol Cell* **81**, 2216-2230 e10 (2021).
74. Vergara, X. *et al.* Widespread chromatin context-dependencies of DNA double-strand break repair proteins. *Nat Commun* **15**, 5334 (2024).
75. Friskes, A. *et al.* Double-strand break toxicity is chromatin context independent. *Nucleic Acids Res* **50**, 9930-9947 (2022).
76. Jeggo, P.A., Downs, J.A. & Gasser, S.M. Chromatin modifiers and remodellers in DNA repair and signalling. *Philos Trans R Soc Lond B Biol Sci* **372**(2017).
77. Clouaire, T. & Legube, G. A Snapshot on the Cis Chromatin Response to DNA Double-Strand Breaks. *Trends Genet* **35**, 330-345 (2019).
78. Caron, P., Pobega, E. & Polo, S.E. DNA Double-Strand Break Repair: All Roads Lead to HeterochROMAtin Marks. *Front Genet* **12**, 730696 (2021).
79. Rogakou, E.P., Boon, C., Redon, C. & Bonner, W.M. Megabase chromatin domains involved in DNA double-strand breaks in vivo. *J Cell Biol* **146**, 905-16 (1999).
80. Panier, S. & Boulton, S.J. Double-strand break repair: 53BP1 comes into focus. *Nat Rev Mol Cell Biol* **15**, 7-18 (2014).
81. Zimmermann, M. & de Lange, T. 53BP1: pro choice in DNA repair. *Trends Cell Biol* **24**, 108-17 (2014).
82. Densham, R.M. & Morris, J.R. Moving Mountains-The BRCA1 Promotion of DNA Resection. *Front Mol Biosci* **6**, 79 (2019).
83. Kim, B.J. *et al.* Genome-wide reinforcement of cohesin binding at pre-existing cohesin sites in response to ionizing radiation in human cells. *J Biol Chem* **285**, 22784-92 (2010).
84. Caron, P. *et al.* Cohesin protects genes against γH2AX Induced by DNA double-strand breaks. *PLoS Genet* **8**, e1002460 (2012).
85. Collins, P.L. *et al.* DNA double-strand breaks induce H2Ax phosphorylation domains in a contact-dependent manner. *Nat Commun* **11**, 3158 (2020).
86. Arnould, C. *et al.* Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660-665 (2021).
87. Ochs, F. *et al.* Stabilization of chromatin topology safeguards genome integrity. *Nature* **574**, 571-574 (2019).
88. Natale, F. *et al.* Identification of the elementary structural units of the DNA damage response. *Nat Commun* **8**, 15760 (2017).
89. Varga, D., Majoros, H., Ujfaludi, Z., Erdélyi, M. & Pankotai, T. Quantification of DNA damage induced repair focus formation via super-resolution dSTORM localization microscopy. *Nanoscale* **11**, 14226-14236 (2019).
90. Luijsterburg, M.S. *et al.* Heterochromatin protein 1 is recruited to various types of DNA damage. *J Cell Biol* **185**, 577-86 (2009).
91. Chou, D.M. *et al.* A chromatin localization screen reveals poly (ADP ribose)-regulated recruitment of the repressive polycomb and NuRD complexes to sites of DNA damage. *Proc Natl Acad Sci U S A* **107**, 18475-80 (2010).
92. Krawczyk, P.M. *et al.* Chromatin mobility is increased at sites of DNA double-strand breaks. *J Cell Sci* **125**, 2127-33 (2012).
93. Cho, N.W., Dilley, R.L., Lampson, M.A. & Greenberg, R.A. Interchromosomal homology searches drive directional ALT telomere movement and synapsis. *Cell* **159**, 108-121 (2014).

94. Chiolo, I. *et al.* Double-strand breaks in heterochromatin move outside of a dynamic HP1a domain to complete recombinational repair. *Cell* **144**, 732-44 (2011).
95. Wensveen, M.R. *et al.* Double-strand breaks in facultative heterochromatin require specific movements and chromatin changes for efficient repair. *Nat Commun* **15**, 8984 (2024).
96. Kendek, A. *et al.* DNA double-strand break movement in heterochromatin depends on the histone acetyltransferase dGcn5. *Nucleic Acids Res* **52**, 11753-11767 (2024).
97. Tsouroula, K. *et al.* Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol Cell* **63**, 293-305 (2016).
98. Nagai, S. *et al.* Functional targeting of DNA damage to a nuclear pore-associated SUMO-dependent ubiquitin ligase. *Science* **322**, 597-602 (2008).
99. Oza, P., Jaspersen, S.L., Miele, A., Dekker, J. & Peterson, C.L. Mechanisms that regulate localization of a DNA double-strand break to the nuclear periphery. *Genes Dev* **23**, 912-27 (2009).
100. Ryu, T. *et al.* Heterochromatic breaks move to the nuclear periphery to continue recombinational repair. *Nat Cell Biol* **17**, 1401-11 (2015).
101. Lemaitre, C. *et al.* Nuclear position dictates DNA repair pathway choice. *Genes Dev* **28**, 2450-63 (2014).
102. van Sluis, M. & McStay, B. A localized nucleolar DNA damage response facilitates recruitment of the homology-directed repair machinery independent of cell cycle stage. *Genes Dev* **29**, 1151-63 (2015).
103. Caridi, P.C., Delabaere, L., Zapotoczny, G. & Chiolo, I. And yet, it moves: nuclear and chromatin dynamics of a heterochromatic double-strand break. *Philos Trans R Soc Lond B Biol Sci* **372**(2017).
104. Lamm, N., Rogers, S. & Cesare, A.J. Chromatin mobility and relocation in DNA repair. *Trends Cell Biol* **31**, 843-855 (2021).
105. Merigliano, C. & Chiolo, I. Multi-scale dynamics of heterochromatin repair. *Curr Opin Genet Dev* **71**, 206-215 (2021).
106. van Sluis, M. & McStay, B. Nucleolar reorganization in response to rDNA damage. *Curr Opin Cell Biol* **46**, 81-86 (2017).
107. Aten, J.A. *et al.* Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science* **303**, 92-5 (2004).
108. Caron, P. *et al.* Non-redundant functions of ATM and DNA-PKcs in response to DNA double-strand breaks. *Cell Rep* **13**, 1598-609 (2015).
109. Schrank, B.R. *et al.* Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. *Nature* **559**, 61-66 (2018).
110. Belin, B.J., Lee, T. & Mullins, R.D. DNA damage induces nuclear actin filament assembly by Formin -2 and Spire-½ that promotes efficient DNA repair. [corrected]. *Elife* **4**, e07735 (2015).
111. Oshidari, R. *et al.* Nuclear microtubule filaments mediate non-linear directional motion of chromatin and promote DNA repair. *Nat Commun* **9**, 2567 (2018).
112. Caridi, C.P. *et al.* Nuclear F-actin and myosins drive relocalization of heterochromatic breaks. *Nature* **559**, 54-60 (2018).
113. Lottersberger, F., Karssemeijer, R.A., Dimitrova, N. & de Lange, T. 53BP1 and the LINC Complex Promote Microtubule-Dependent DSB Mobility and DNA Repair. *Cell* **163**, 880-93 (2015).
114. Marnef, A. *et al.* A cohesin/HUSH- and LINC-dependent pathway controls ribosomal DNA double-strand break repair. *Genes Dev* **33**, 1175-1190 (2019).
115. Huang, J. *et al.* SLFN5-mediated chromatin dynamics sculpt higher-order DNA repair topology. *Mol Cell* (2023).
116. Klaasen, S.J. *et al.* Nuclear chromosome locations dictate segregation error frequencies. *Nature* **607**, 604-609 (2022).
117. Manjón, A.G. *et al.* Unexpected gene activation following CRISPR-Cas9-mediated genome editing. *EMBO Rep* **23**, e53902 (2022).
118. Xiao, H. *et al.* Insights into the mechanism of microtubule stabilization by Taxol. *Proc Natl Acad Sci U S A* **103**, 10166-10173 (2006).
119. Smith, K.S., Liu, L.L., Ganesan, S., Michor, F. & De, S. Nuclear topology modulates the mutational landscapes of cancer genomes. *Nat Struct Mol Biol* **24**, 1000-1006 (2017).
120. Etourneaud, L. *et al.* Lamin B1 sequesters 53BP1 to control its recruitment to DNA damage. *Sci Adv* **7**(2021).
121. Gibbs-Seymour, I., Markiewicz, E., Bekker-Jensen, S., Mailand, N. & Hutchison, C.J. Lamin A/C-dependent interaction with 53BP1 promotes cellular responses to DNA damage. *Aging Cell* **14**, 162-9 (2015).
122. Murray-Nerger, L.A., Justice, J.L., Rekapalli, P., Hutton, J.E. & Cristea, I.M. Lamin B1 acetylation slows the G1 to S cell cycle transition through inhibition of DNA repair. *Nucleic Acids Res* **49**, 2044-2064 (2021).
123. Wagner, A., Regev, A. & Yosef, N. Revealing

the vectors of cellular identity with single-cell genomics. *Nat Biotechnol* **34**, 1145-1160 (2016).

124. Park, S.R. *et al.* Single-Cell Transcriptome Analysis of Colon Cancer Cell Response to 5-Fluorouracil-Induced DNA Damage. *Cell Rep* **32**, 108077 (2020).
125. Virchow, R. *Die Cellularpathologie in ihrer Begründung auf physiologische und pathologische Gewebelehre: 20 Vorlesungen, gehalten während d. Monate Febr., März u. April 1858 im Patholog. Inst. zu Berlin*, (Hirschwald, 1858).
126. Navin, N. *et al.* Tumour evolution inferred by single-cell sequencing. *Nature* **472**, 90-4 (2011).
127. Zong, C., Lu, S., Chapman, A.R. & Xie, X.S. Genome-wide detection of single-nucleotide and copy-number variations of a single human cell. *Science* **338**, 1622-6 (2012).
128. McConnell, M.J. *et al.* Mosaic copy number variation in human neurons. *Science* **342**, 632-7 (2013).
129. Jerkovic, I. & Cavalli, G. Understanding 3D genome organization by multidisciplinary methods. *Nat Rev Mol Cell Biol* **22**, 511-528 (2021).
130. Cosma, M.P. & Neguembor, M.V. The magic of unraveling genome architecture and function. *Cell Rep* **42**, 112361 (2023).
131. Mateo, L.J. *et al.* Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* **568**, 49-54 (2019).
132. Su, J.H., Zheng, P., Kinrot, S.S., Bintu, B. & Zhuang, X. Genome-Scale Imaging of the 3D Organization and Transcriptional Activity of Chromatin. *Cell* **182**, 1641-1659. e26 (2020).
133. Takei, Y. *et al.* Integrated spatial genomics reveals global architecture of single nuclei. *Nature* **590**, 344-350 (2021).
134. Nguyen, H.Q. *et al.* 3D mapping and accelerated super-resolution imaging of the human genome using in situ sequencing. *Nat Methods* **17**, 822-832 (2020).
135. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nat Methods* **10**, 857-60 (2013).
136. Mota, A. *et al.* FRET-FISH probes chromatin compaction at individual genomic loci in single cells. *Nat Commun* **13**, 6680 (2022).
137. Chen, F., Tillberg, P.W. & Boyden, E.S. Optical imaging. Expansion microscopy. *Science* **347**, 543-8 (2015).
138. Woodworth, M.A. *et al.* Multiplexed single-cell profiling of chromatin states at genomic loci by expansion microscopy. *Nucleic Acids Res* **49**, e82 (2021).
139. Pownall, M.E. *et al.* Chromatin expansion microscopy reveals nanoscale organization of transcription and chromatin. *Science* **381**, 92-100 (2023).
140. Shaban, H.A., Barth, R., Recoules, L. & Bystricky, K. Hi-D: nanoscale mapping of nuclear dynamics in single living cells. *Genome Biol* **21**, 95 (2020).
141. Llères, D., James, J., Swift, S., Norman, D.G. & Lamond, A.I. Quantitative analysis of chromatin compaction in living cells using FLIM-FRET. *J Cell Biol* **187**, 481-96 (2009).
142. Lou, J. *et al.* Phasor histone FLIM-FRET microscopy quantifies spatiotemporal rearrangement of chromatin architecture during the DNA damage response. *Proc Natl Acad Sci U S A* **116**, 7323-7332 (2019).
143. Liang, Z., Solano, A., Lou, J. & Hinde, E. Histone FRET reports the spatial heterogeneity in nanoscale chromatin architecture that is imparted by the epigenetic landscape at the level of single foci in an intact cell nucleus. *Chromosoma* **133**, 5-14 (2024).
144. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-11 (2002).
145. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-54 (2006).
146. Fullwood, M.J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64 (2009).
147. Mumbach, M.R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* **13**, 919-922 (2016).
148. Hughes, J.R. *et al.* Analysis of hundreds of cis-regulatory landscapes at high resolution in a single, high-throughput experiment. *Nat Genet* **46**, 205-12 (2014).
149. Jäger, R. *et al.* Capture Hi-C identifies the chromatin interactome of colorectal cancer risk loci. *Nat Commun* **6**, 6178 (2015).
150. Darrow, E.M. *et al.* Deletion of DXZ4 on the human inactive X chromosome alters higher-order genome architecture. *Proc Natl Acad Sci U S A* **113**, E4504-12 (2016).
151. Oudelaar, A.M. *et al.* Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet* **50**, 1744-1751 (2018).
152. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele topologies. *Nat Genet* **50**, 1151-1160 (2018).
153. Beagrie, R.A. *et al.* Complex multi-enhancer contacts captured by genome architecture

mapping. *Nature* **543**, 519-524 (2017).

154. Quinodoz, S.A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757 e24 (2018).
155. Zheng, M. *et al.* Multiplex chromatin interactions with single-molecule precision. *Nature* **566**, 558-562 (2019).
156. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* **18**, 424-8 (2000).
157. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).
158. Arrastia, M.V. *et al.* Single-cell measurement of higher-order 3D genome organization with scSPRITE. *Nat Biotechnol* **40**, 64-73 (2022).
159. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-47 (2015).
160. Hoffman, M.M. *et al.* Integrative annotation of chromatin elements from ENCODE data. *Nucleic Acids Res* **41**, 827-41 (2013).
161. Rang, F.J. *et al.* Single-cell profiling of transcriptome and histone modifications with EpiDamID. *Mol Cell* **82**, 1956-1970.e14 (2022).
162. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* **33**, 1165-72 (2015).
163. Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* **51**, 1060-1066 (2019).
164. Hainer, S.J., Bošković, A., McCannell, K.N., Rando, O.J. & Fazzio, T.G. Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329.e11 (2019).
165. Ku, W.L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods* **16**, 323-325 (2019).
166. Zeller, P. *et al.* Single-cell sortChIC identifies hierarchical chromatin dynamics during hematopoiesis. *Nat Genet* **55**, 333-345 (2023).
167. Wang, Q. *et al.* CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Mol Cell* **76**, 206-216.e7 (2019).
168. Kaya-Okur, H.S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).
169. Wu, S.J. *et al.* Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat Biotechnol* **39**, 819-824 (2021).
170. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* **39**, 825-835 (2021).
171. Harada, A. *et al.* A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat Cell Biol* **21**, 287-296 (2019).
172. Gopalan, S., Wang, Y., Harper, N.W., Garber, M. & Fazzio, T.G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol Cell* **81**, 4736-4746.e5 (2021).
173. Meers, M.P., Llagas, G., Janssens, D.H., Codomo, C.A. & Henikoff, S. Multifactorial profiling of epigenetic landscapes at single-cell resolution using MulTI-Tag. *Nat Biotechnol* **41**, 708-716 (2023).
174. Bartosovic, M. & Castelo-Branco, G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nat Biotechnol* **41**, 794-805 (2023).
175. Stuart, T. *et al.* Nanobody-tethered transposition enables multifactorial chromatin profiling at single-cell resolution. *Nat Biotechnol* **41**, 806-812 (2023).
176. Xiong, H., Wang, Q., Li, C.C. & He, A. Single-cell joint profiling of multiple epigenetic proteins and gene transcription. *Sci Adv* **10**, eadi3664 (2024).
177. Yeung, J. *et al.* scChIX-seq infers dynamic relationships between histone modifications in single cells. *Nat Biotechnol* **41**, 813-823 (2023).
178. Lochs, S.J.A. *et al.* Combinatorial single-cell profiling of major chromatin types with MAbID. *Nat Methods* **21**, 72-82 (2024).
179. Rooijers, K. *et al.* Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat Biotechnol* **37**, 766-772 (2019).
180. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* **24**, 695-713 (2023).
181. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* **24**, 494-515 (2023).

Chapter 2

Single-cell DamID to capture contacts between DNA and the nuclear lamina in individual mammalian cells

Kim L. de Luca[1,2] and Jop Kind[1,2,3]

[1] Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht; Utrecht, The Netherlands

[2] Oncode Institute; The Netherlands

[3] Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University; Nijmegen, The Netherlands

Published in ***Methods in Molecular Biology*** (2021).
Beatrice Bodega and Chiara Lanzuolo (eds.), *Capturing Chromosome Conformation*

Abstract

The organization of DNA within the eukaryotic nucleus is important for cellular processes such as regulation of gene expression and repair of DNA damage. To comprehend cell-to-cell variation within a complex system, systematic analysis of individual cells is necessary. While many tools exist to capture DNA conformation and chromatin context, these methods generally require large populations of cells for sufficient output. Here we describe single-cell DamID, a technique to capture contacts between DNA and a given protein of interest. By fusing the bacterial methyltransferase Dam to nuclear lamina protein lamin B1, genomic regions in contact with the nuclear periphery can be mapped. Single-cell DamID generates contact maps with sufficient throughput and resolution to reliably identify patterns of similarity as well as variation in nuclear organization of interphase chromosomes.

Key words

DamID, single-cell genomics, chromatin, nuclear lamina, lamin B1

1. Introduction

A long-standing question in biology is how the same genome can give rise to all different cell types and their corresponding functions in the organism. It has become clear that gene transcription occurs in a cell type-specific manner, after which mRNA translation and downstream processes eventually culminate in functional proteins. Still, the regulatory events underlying differences in gene expression and cell fate commitment remain poorly understood. DNA is packaged within chromatin, which, in turn, is coiled into higher-order structures and resides in nuclear compartments with more permissive or restrictive features (reviewed in[1]). Densely packed heterochromatin is typically located at the nuclear lamina (NL), while less compacted, transcriptionally active, euchromatin resides in the nuclear interior (reviewed in[2]). Withal, heterogeneity in cellular function is partly the result of variability in spatial genome organization.

There are multiple ways of experimentally capturing nuclear organization and chromatin context, e.g. chromatin immunoprecipitation followed by next-generation sequencing (ChIP-seq)[3], DNA adenine methyltransferase identification (DamID)[4,5], assay for transposase-accessible chromatin using sequencing (ATAC-seq)[6], chromosome conformation capture (3C)[7] and its derivatives 4C[8], 5C[9], Hi-C[10] and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET)[11]. These methods generally require large populations of cells for sufficient output. If single-cell adaptations exist[12-14] they lack the resolution to conclusively identify variation in chromosomal organization across cells. More recently, genome architecture mapping (GAM)[15] and super-resolution chromatin tracing[16] have interrogated chromatin organization of single cells. However, these techniques are experimentally challenging and largely unavailable to those without designated hardware and prior training. Here we describe the method for performing a modified version of DamID that is suited to single cells[17]. DamID is easily downscaled, with the capacity to process hundreds of cells per day whilst retaining high genomic coverage at a resolution of ~10 kb. As a result, single-cell DamID (scDamID) robustly assesses cellular heterogeneity with easy implementation and common lab equipment.

DamID involves the tethering of *Escherichia coli* DNA adenine methyltransferase (Dam) to a protein of interest (POI), and subsequent identification of Dam-methylated DNA sequences. The NL is a meshwork of proteins lining the inner nuclear membrane (INM), providing structure to the nucleus and serving as a scaffold for the genome. By tethering Dam to NL component lamin B1 (Dam-LMNB1) and expressing the fusion protein *in vivo*, the genome that comes in close proximity to the NL can be recognized, amplified, and sequenced. Untethered Dam is standardly included in the experimental design, as a proxy for accessible chromatin that is methylated by a freely diffusing enzyme. Dam deposits a methyl group on N6 of the adenine

within a GATC motif ($G^{m6}ATC$) on both strands of the DNA. These $G^{m6}ATC$s are cleaved with methylation-sensitive restriction enzyme DpnI, leaving blunt ends to which a universal adapter is ligated. Polymerase chain reaction (PCR) primers hybridize to this adapter in order to specifically amplify methylated genomic fragments. Expanding on the previously published protocol, we include here a multiplexing strategy to pool hundreds of cells in one sequencing library, by adding a cell-specific barcode to the PCR primer.

This chapter describes the single-cell DamID workflow from plasmid choice to sample preparation and data processing. Population-based DamID has been described extensively with regards to the Dam-POI[5,18] and computational analysis[19]; therefore, we devote particular attention to the novelties of the single-cell approach. We reflect on all aspects of the experimental setup and execution in the *Notes* section.

2. Materials

2.1 Hardware

1. Benchtop centrifuge with tube and plate rotors.
2. Nucleic acid spectrophotometer such as NanoDrop (Thermo Scientific).
3. Conventional gel electrophoresis equipment.
4. Real-time thermal cycler with 96-well plate format.
5. Thermal cycler with 96-well plate format.
6. Fluorometer such as Qubit (Invitrogen).
7. Automated electrophoresis system such as Bioanalyzer or TapeStation (Agilent).
8. (Access to a facility providing) fluorescence-activated cell sorter.
9. (Access to a facility providing) Illumina sequencer.
10. Optional: UV PCR workstation.
11. Optional: liquid-handling robot such as Nanodrop II (BioNex).

2.2 Plasmids

See the van Steensel lab website for elaborate explanations on plasmids for DamID in mammalian cells (http://research.nki.nl/vansteensellab/Mammalian_plasmids.htm). For conventional transfection, we have used the pPTuner IRES2 plasmid (Clontech PT4040-5) in which to clone Dam-POI (see [20]). For lentiviral transduction, we have used the plasmid pCCL.sin.cPPT.hPGK.ΔLNGFR.WPRE [21] in which to clone Dam-POI (see [17]).

2.3 Cell culture

12. Cell culture dishes and plasticware
13. Cell type-specific culture medium.
14. Transfection or transduction reagents.
15. Antibiotics for selection after transfection or transduction.
16. Dissociation reagent.
17. FACS tubes (strainer cap is recommended).
18. Cell-permeable fluorescent DNA stain to evaluate DNA content in live cells during FACS, such as Hoechst.
19. Cell-impermeable fluorescent DNA stain to discriminate live and apoptotic cells during FACS, such as DAPI or PI.
20. Optional: chemical induction agent such as Shield1 or indole-3-acetic acid (IAA; auxin-class hormone).

2.4 MboI-qPCR assay

21. qPCR 96-well plates.
22. RT plate sealers.
23. Genomic DNA isolation kit or equivalent separate reagents.
24. 1% agarose gel including DNA stain.
25. Gel loading dye.
26. 1 Kb+ DNA ladder.
27. MboI enzyme plus buffer.
28. Primers flanking GATCs (10 μM; see Table 2.1 at end of document for sequences).
29. Real-time PCR mix including dye.

2.5 Single-cell DamID and next-generation sequencing

30. PCR 96-well plates.
31. Lysis buffer: 10 mM Tris acetate pH 7.5, 10 mM magnesium acetate, 50 mM potassium acetate, 0.67% Tween-20, 0.67% IGEPAL CA-630, freshly added 0.67 mg/mL Proteinase K.
32. DpnI enzyme plus buffer.
33. T4 DNA ligase plus buffer.
34. DamID double-stranded adapter (50 μM). Dissolve Adapter_top and Adapter_bottom to 100 μM in annealing buffer, then mix equal volumes of both oligonucleotides in a tightly closed tube. Place tube in a container with water of ~94 °C and let cool to room temperature to allow slow annealing of adapters.

35. Adapter_top
    5' CTAATACGACTCACTATAGGGCAGCGTGGTCGCGGCCGAGGA 3'
    Adapter_bottom
    5' TCCTCGGCCGCG 3'
36. Annealing buffer: 100 mM potassium acetate, 30 mM HEPES, pH 7.5.
37. Barcoded primers (25 µM). 5' NNNNNNBARCODGTGGTCGCGGCCGAGGATC 3'
38. PCR mix (recommended: including gel loading dye).
39. SPRI beads or spin column purification kit.
40. Illumina library preparation kit or reagents.
41. Qubit DNA HS reagents and tubes.

3. Methods

A proper DamID experimental design includes Dam-LMNB1 as well as untethered Dam. In scDamID, clonal cell lines have to be established (for both constructs separately) to avoid intercellular variation due to differences in e.g. protein expression level and inducibility. Initial experimental steps are performed with multiple clones per construct to select for clones with desired signal-to-noise methylation levels.

3.1 Choosing the DamID plasmid

Dependent on the preferred mode of delivery for your cells, select a vector for the Dam constructs. Including an antibiotic resistance cassette to select for successful plasmid integration is recommended. For temporal regulation of protein expression, consider including a degron system or inducible promoter. We routinely use the ProteoTuner system (degron-tagged Dam construct is degraded by default and stabilized upon addition of Shield1) or the auxin-inducible degron (AID; degron-tagged Dam construct is stable by default and degraded upon addition of auxin).

3.2 Creating a clonal cell line that stably expresses the Dam-fusion constructs

3.2.1 Construct delivery

Introduce the Dam plasmids into cells by your method of choice. We have derived cell lines by conventional liposome transfection and lentiviral transduction. Select successfully transfected cells by antibiotic resistance and/or fluorescence-activated cell sorting (FACS; if the vector includes expression of a fluorescent protein).

3.2.2 Generating clonal cell lines

Once all cells in the culture population are resistant against the antibiotic or contain another selectable marker, establish clones by FACS or limiting dilution followed by colony picking. See Note 1.

3.2.3 Optional: testing clones by MboI-qPCR assay

This is an optional, though recommended, step to evaluate levels of adenine methylation in the different clones before proceeding to scDamID. Purified gDNA is digested with restriction enzyme MboI, which specifically cuts unmethylated (but not hemimethylated or fully methylated) GATC sequences. Quantitative PCR (qPCR) is performed on digested and undigested DNA, with primers flanking GATC sequences in lamina-associated domains (LADs) and inter-LADs (iLADs). Percentage of methylation is calculated by $1/2^{(Ct\ (digested)\ -\ Ct\ (undigested)} * 100\%$. Comparing the percentages of methylation in LADs and iLADs gives an estimation of NL-specific Dam methylation. Typically, the Dam-LMNB1 clones with highest LAD/iLAD ratios are selected for subsequent experiments.

1. Grow cells under appropriate culture conditions. Dependent on Dam-fusion, add or remove chemical induction agent 12 h before cell collection. See Note 2.
2. Isolate genomic DNA from cells using the Wizard Genomic DNA Purification Kit or other method of choice.
3. Measure concentration of purified gDNA using a spectrophotometer.
4. Check integrity of gDNA by running ~500 ng of DNA on a 1% agarose gel. Include a 1 Kb+ DNA ladder. Intact gDNA runs as a single, tight band larger than 10 Kb in size.
5. For each clone, prepare two digestion reactions with 1 ug gDNA each: one containing 5 units (U) MboI, and one undigested control.
6. Incubate reactions at 37 °C for 4 h to digest unmethylated DNA.
7. Incubate reactions at 65 °C for 20 min to heat inactivate MboI.
8. Optional: check MboI digestion by running an aliquot of the digestion reaction on a 1% agarose gel.
9. Dilute digestion reactions 1/5 with nuclease-free water (NFW).
10. Assemble qPCR reactions on ice and protected from light as much as possible. Each reaction of 10 µL final volume contains 4 µL of DNA, 0.5 µL of 10 µM forward primer, 0.5 µL of 10 µM reverse primer, and 5 µL of 2X qPCR mix (including PCR buffer, polymerase and dye). We recommend performing each reaction in triplicates. See Table 2.1 at end of document for primer sequences. Select (at least) two primer sets per region, *i.e.* two for LADs and two for iLADs.

Do include reactions with positive control primers that have performed robustly in your hands, as well as negative control reactions without DNA template.

11. Run the assembled reactions in a qPCR thermocycler using the program described in Table 2.2.
12. Calculate the percentage of methylation per primer set as follows. Take the average of replicate Ct values. Calculate the differences between digested and undigested samples (ΔCt). Then calculate 1 / (2^ΔCt) * 100% for the percentage of methylation.
13. Compare the methylation levels in LADs versus iLADs. A good Dam-LMNB1 clone typically has a high (>3) LAD/iLAD ratio, in addition to a high (>40) LAD methylation percentage. A good Dam clone typically has a low (<1) LAD/iLAD ratio, and a low (<20) LAD methylation percentage. See Figure 2.1 for an example MboI-qPCR result.
14. Proceed to scDamID with selected clones, (at least) two per construct. See Note 3.

Cycle	Denature	Anneal	Extend
1	95 °C for 3 min		
2-40	95 °C for 5 s	60 °C for 10 s	72 °C for 10 s Acquire at end of step
Optional	Melt profile analysis		

Table 2.2. Thermal cycling program for MboI-qPCR.

3.3 Obtaining single cells for DamID

scDamID involves depositing single cells into reaction reservoirs (typically tubes, strips or plates of 200-µL volume capacity), followed by additive reactions without cleaning the sample in between, keeping material loss at a minimum.

All subsequent steps are performed at room temperature, unless otherwise specified. If available, using a multi-step pipette or liquid-handling robot will highly decrease hands-on time and increase throughput. A decontaminated working environment, such as a UV PCR workstation, is advised.

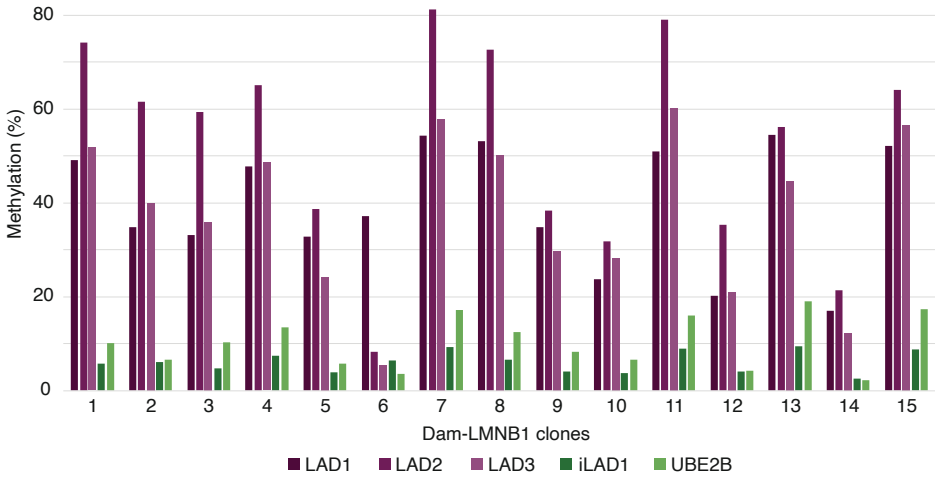


Figure 2.1. Methylation levels across 15 Dam-LMNB1 clones, assayed by MboI-qPCR. Primers flank GATC sequences in LADs (purple) and iLADs (green). UBE2B primers flank a GATC in the *UBE2B* gene promoter.

15. Grow cells under appropriate culture conditions. Dependent on Dam-fusion construct, add or remove chemical induction agent 12 hours before cell collection.
16. Prior to cell collection, prepare an appropriate amount of lysis buffer including freshly added Proteinase K and keep on ice until dispensation. Per sample well, 3 µL of lysis buffer is needed.
17. Dispense 3 µL of lysis buffer per well in an appropriate number of 96-well plates. Keep plates on ice until cell collection.
18. If needed, make a single-cell suspension of the target sample (e.g. by trypsin dissociation).
19. Collect live single cells in the prepared 96-well plates by FACS. See Notes 4-8.
20. As soon as collection per plate is complete, seal plate and centrifuge at 1000 x *g* for 1 min.
21. Continue with step 3.4 or freeze plates at -20 °C.

3.4 DamID: amplification of Dam-marked genomic fragments

Genomic DNA that has been methylated at GATC motifs is specifically digested, leaving blunt ends to which a universal adapter is ligated. Using barcoded primers that hybridize to this adapter, the methylated fragments are specifically enriched for by PCR.

22. Incubate plates at 42 °C for 4 h to lyse cells and digest all proteins.
23. Incubate plates at 80 °C for 20 min to heat inactivate Proteinase K.
24. Assemble DpnI digestion mix and add 7 µL of mix to each well. DpnI digestion mix per well: 6.9 µL of 1X buffer and 0.1 µL of DpnI enzyme.
25. Incubate plates at 37 °C for 8 h to digest methylated DNA.
26. Incubate plates at 80 °C for 20 min to heat inactivate DpnI, then cool on ice.
27. Assemble adapter ligation mix and add 10 µL of mix to each well. Adapter ligation mix per well: 9.55 µL of 1X T4 ligase buffer, 0.2 µL of 50 µM adapter and 0.25 µL of T4 ligase.
28. Incubate plates at 16 °C for 12-16 h.
29. Incubate plates at 65 °C for 10 min to heat inactivate T4 ligase, then cool on ice. See Note 9.
30. Add 2 µL of 25 µM cell-specific, barcoded primer to each well. Take care: each barcoded primer corresponds to a single sample. See Notes 10-11.
31. On ice, assemble PCR mix and add 28 µL of mix to each well. PCR mix per well: 28 µL PCR buffer including polymerase (e.g. 10 µL of 5X PCR buffer, 0.5 µL of polymerase, and 17.5 µL of NFW).
32. Run the assembled reactions in a thermocycler using the program described in Table 2.3. When testing an uncharacterized cell line, evaluate PCR product by agarose gel electrophoresis after 20 cycles, and adjust cycles as necessary. See Note 12.
33. Run 4 µL of PCR product on 1% agarose gel to check control samples, smear intensity and distribution of fragment size. Include a 1 Kb+ DNA ladder. Run more or fewer cycles of PCR if necessary. See Figure 2.2 for an example PCR result.

Cycle	Denature	Anneal	Extend
Heated lid on			
1			68 °C for 10 min
2	94 °C for 3 min	65 °C for 5 min	68 °C for 15 min
3–6	94 °C for 1 min	65 °C for 1 min	68 °C for 10 min
7–32	94 °C for 1 min	65 °C for 1 min	68 °C for 2 min

Table 2.3. Thermal cycling program for DamID PCR.

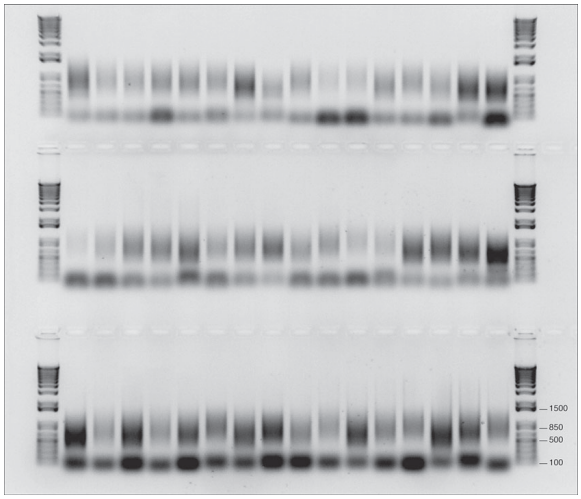


Figure 2.2. DamID PCR product of 48 single cells (with different barcoded primers), loaded on 1% agarose gel.

3.5 Preparation of Illumina sequencing libraries

The amplified product is multiplexed by pooling together all samples with different barcodes. The pools are subsequently cleaned, concentrated and further processed into libraries for deep sequencing. We have used proprietary kits as well as separate reagents to perform library preparation (end repair and phosphorylation, dA-tailing, adapter ligation and PCR).

34. Evaluate PCR product and estimate relative concentration of the different samples.
35. Pool barcoded samples together according to their estimated concentration; the aim is to generate a mixture with equal numbers of molecules across the samples. See Notes 10-11.
36. Purify the multiplexed material with SPRI beads (2:1 volumes beads:samples) or spin columns and elute in an appropriate volume of resuspension buffer; 30 µL of MQ is usual practice in our lab.
37. Prepare libraries for next-generation sequencing with method of choice, preferably starting with 250-500 ng DNA.
38. Measure the concentration of each library with a Qubit fluorometer, per manufacturer's instructions.
39. Evaluate the fragment distribution of each library with an automated electrophoresis system such as Agilent Bioanalyzer or TapeStation.
40. For Dam-LMNB1 in mammalian cells, sequence single-end to a depth of approximately 500K raw reads per single-cell sample.

3.6 Raw data processing and visualization

Raw reads are demultiplexed by their library-specific index and by their sample-specific DamID barcode, after which the DamID primer sequence is removed and sequences are aligned to the reference genome. Reads per GATC are counted, summed across sequencing lanes, aggregated in genomic segments and optionally smoothened for visualization.

4. Notes

1. We recommend seeding an initial number of ~200 clones, as survival rate may be poor. Depending on the cell line, we have experienced survival anywhere between 5-90%. Subsequent testing of clones by MboI-qPCR and conventional DamID has resulted in 10-60% of clones screened as appropriate for scDamID.
2. Depending on the construct and degron or induction system, timing of the protein stabilization may need to be optimized. We have good experiences with inductions between 4 and 24 h.
3. After having selected clones, we recommend performing conventional population-based DamID followed by next-generation sequencing. This information serves as a comparison between bulk DamID and other methods such as ChIP-seq or Hi-C, if available, but most importantly as a comparison for the scDamID data.
4. Collecting cells by FACS is by far the most straightforward and high-throughput method. If FACS is unavailable, pipetting by mouth with a micro-capillary may be an option.
5. If sorting live cells, it is recommended to exclude apoptotic cells by briefly incubating cell suspension with a DNA dye (e.g. DAPI or PI).
6. Dam methylation accumulates in G1 and G2. During DNA replication, the mark remains on only one DNA strand (hemimethylation). We recommend collecting cells at the G1/S or G2/M transition.
7. To sort on cell cycle phase, incubate cells with Hoechst, a cell membrane-permeable DNA dye. Live incubation of the cell suspension with 1 μg/mL Hoechst at 37 °C for 30 min before sorting has generated acceptable DNA content profiles in our hands. Alternatively, the FUCCI reporter can be used to discriminate cell cycle phases.
8. During cell collection, it is recommended to include multiple low-cell samples (between 10-100 cells) and empty wells (0 cells).
9. In conventional DamID, a digestion step with MboI is included to destroy and thereby avoid PCR amplification of fragments with unmethylated GATCs. We do not include this MboI digestion in scDamID, but it is not advised against *per se*.

10. We recommend pooling multiple experimental conditions in the same library to avoid batch effects (or, at the very least, enable batch correction). Therefore, if the number of samples exceeds the number of available barcodes, take care to add barcodes to your samples such that multiple conditions can be pooled together. While one library per condition does facilitate future re-sequencing of particular samples, it is best practice not to pool in that manner until after you have established potential differences between experimental conditions.
11. We recommend pooling and processing 100-cell (or higher) samples separately from the single-cell samples.
12. The number of DamID PCR cycles depends on various aspects, most significantly the expression level of Dam-POI *in vivo*. We recommend starting with 20 cycles. An ultimately time-saving but somewhat laborious approach is to pause the PCR program (at 4 °C) after 20 cycles, remove 4 µL from each PCR reaction and store cold in e.g. a 384-well plate, and repeat this process after every subsequent two cycles. Loading the successive material on 1% agarose gel should give a good indication of when enough product has been generated. Keep in mind that, when pooling many samples in one sequencing run, little material is required per sample. Therefore, avoid over-amplification since it will generate many PCR duplicates and starkly reduce complexity of the sample. See Figure 2.3 for an example gel of DamID PCR products after successive cycling.

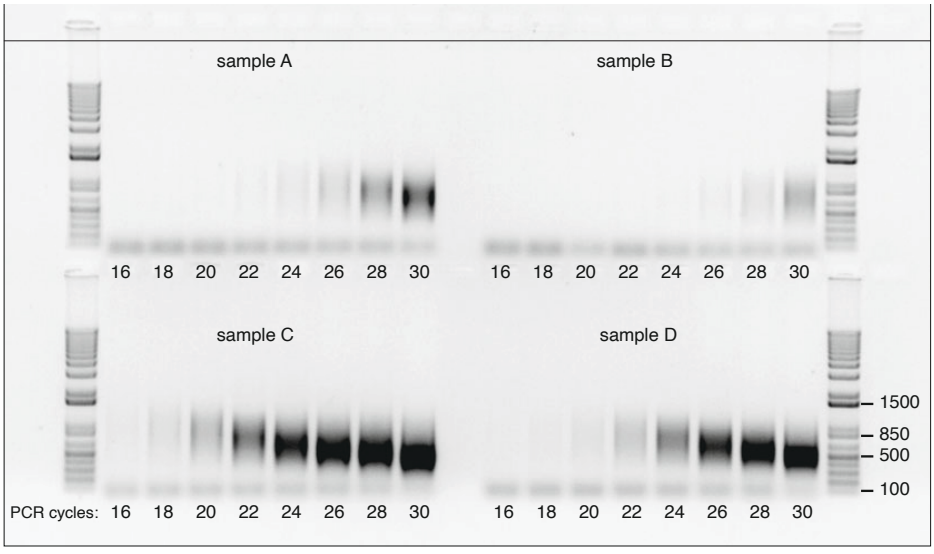


Figure 2.3. DamID PCR products of 4 different samples after successive cycling, loaded on 1% agarose gel.

Name	Sequence
LAD1_for	CATTGGCTTCTTTGGTGCCAGGT
LAD1_rev	ACGGTGGAGGCAGTCAAAAGGC
LAD2_for	ACAGCAGGAAGTACTTGAGATCC
LAD2_rev	ATTAATCTGGCCCGGAGAGT
LAD3_for	AGCTTATATCAAATAAATCCCTGAAA
LAD3_rev	TGTGCATGACAAATATAAAACCAA
iLAD1_for	GAAGGTTCCCCCACAGAAAT
iLAD1_rev	CTGAGGCAAAGACAGGGAAG
iLAD2_for	ACAGCAGGAAGTACTTGAGATCC
iLAD2_rev	ATTAATCTGGCCCGGAGAGT
UBE2B_for	ACTCAGGGGTGGATTGTTGA
UBE2B_rev	GCCAGAGATTTCAGGGAAAG

Table 2.1. Primer sequences for MboI-qPCR.

References

1. Bickmore, W.A. & van Steensel, B. Genome architecture: domain organization of interphase chromosomes. *Cell* **152**, 1270-84 (2013).
2. Zuleger, N. *et al.* The nuclear envelope as a chromatin organizer. *Nucleus* **2**, 339-49 (2011).
3. Johnson, D.S. *et al.* Genome-wide mapping of in vivo protein-DNA interactions. *Science* **316**, 1497-502 (2007).
4. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* **18**, 424-8 (2000).
5. Vogel, M.J *et al.* Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc* **2**, 1467-78 (2007).
6. Buenrostro, J.D. *et al.* Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-8 (2013).
7. Dekker, J. *et al.* Capturing chromosome conformation. *Science* **295**, 1306-11 (2002).
8. Simonis, M. *et al.* Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C). *Nat Genet* **38**, 1348-54 (2006).
9. Dostie, J. *et al.* Chromosome Conformation Capture Carbon Copy (5C): a massively parallel solution for mapping interactions between genomic elements. *Genome Res* **16**, 1299-309 (2006).
10. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
11. Fullwood, M.J. *et al.* An oestrogen-receptor-alpha-bound human chromatin interactome. *Nature* **462**, 58-64 (2009).
12. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).
13. Buenrostro, J.D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-90 (2015).
14. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* **33**, 1165-72 (2015).
15. Beagrie, R.A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519-524 (2017).
16. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**(2018).
17. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-47 (2015).
18. Robson, M.I. & Schirmer, E.C. The Application of DamID to Identify Peripheral Gene Sequences in Differentiated and Primary Cells. *Methods Mol Biol* **1411**, 359-86 (2016).
19. Gomez-Saldivar, G. *et al.* DamID Analysis of Nuclear Organization in Caenorhabditis elegans. *Methods Mol Biol* **1411**, 341-58 (2016).
20. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178-92 (2013).
21. Amendola, M. *et al.* Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. *Nat Biotechnol* **23**, 108-16 (2005).

Chapter 3

Single-cell profiling of transcriptome and histone modifications with EpiDamID

Franka J. Rang[1,2*], Kim L. de Luca[1,2*], Sandra S. de Vries[1,2], Christian Valdes-Quezada[1,2], Ellen Boele[1,2], Phong D. Nguyen[1], Isabel Guerreiro[1,2], Yuko Sato[3], Hiroshi Kimura[3], Jeroen Bakkers[1,4], Jop Kind[1,2,5‡]

[1] Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht; Utrecht, The Netherlands

[2] Oncode Institute; The Netherlands

[3] Cell Biology Center, Institute of Innovative Research, Tokyo Institute of Technology; Yokohama, 226-8503, Japan

[4] Department of Pediatric Cardiology, Division of Pediatrics, University Medical Center Utrecht; Utrecht, The Netherlands

[5] Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University; Nijmegen, The Netherlands

* Co-first authorship

‡ Corresponding authorship

Article published in ***Molecular Cell*** (2022).

Abstract

Recent advances in single-cell sequencing technologies have enabled simultaneous measurement of multiple cellular modalities, but the combined detection of histone post-translational modifications and transcription at single-cell resolution has remained limited. Here, we introduce EpiDamID, an experimental approach to target a diverse set of chromatin types by leveraging the binding specificities of single-chain variable fragment antibodies, engineered chromatin reader domains, and endogenous chromatin-binding proteins. Using these, we render the DamID technology compatible with the genome-wide identification of histone post-translational modifications. Importantly, this includes the possibility to jointly measure chromatin marks and transcriptome at the single-cell level. We use EpiDamID to profile single-cell Polycomb occupancy in mouse embryoid bodies and provide evidence for hierarchical gene regulatory networks. In addition, we map H3K9me3 in early zebrafish embryogenesis, and detect striking heterochromatic regions specific to notochord. Overall, EpiDamID is a new addition to a vast toolbox to study the role of chromatin states during dynamic cellular processes.

Introduction

Histone post-translational modifications (PTMs) contribute to chromatin structure and gene regulation. The addition of PTMs to histone tails can modulate the accessibility of the underlying DNA and form a binding platform for myriad downstream effector proteins. As such, histone PTMs play key roles in a multitude of biological processes, including lineage specification (e.g. refs [1–3]), cell cycle regulation (e.g. refs [4,5]), and response to DNA damage (e.g. refs [6,7]).

Over the past decade, antibody-based DNA-sequencing methods have provided valuable insights into the function of histone PTMs in a variety of biological contexts. Most studies employ ChIP-seq (chromatin immunoprecipitation after formaldehyde fixation[8]), or strategies based on *in situ* enzyme tethering such as chromatin immunocleavage (ChIC)[9], and its derivative Cleavage Under Targets and Release Using Nuclease (CUT&RUN)[10]. However, the requirement of high numbers of input cells consequently provides a population-average view, which disregards the complexity of most biological systems. As a result, several low-input methods have been developed that can assay histone PTMs in individual cells, including but not limited to Drop-ChIP[11], ChIL-seq[12], ACT-seq[13], single-cell ChIP-seq[14], single-cell ChIC-seq[15], single-cell adaptation of CUT&RUN[16], CUT&Tag[17], CoBATCH[18], single-cell itChIP[19], and sortChIC[20]. While these techniques offer an understanding of the epigenetic heterogeneity between cells, they do not provide a direct link to other measurable outputs. Recently, however, three methods have been developed that jointly profile histone modifications and gene expression: Paired-Tag (parallel analysis of individual cells for RNA expression and DNA from targeted tagmentation by sequencing)[21], CoTECH (combined assay of transcriptome and enriched chromatin binding)[22], and SET-seq (same cell epigenome and transcriptome sequencing)[23]. These techniques thus enable linking of gene regulatory mechanisms to transcriptional output and cellular state. Of note, all three methods rely on antibody binding for detection of histone modifications and Tn5-mediated tagmentation for sequencing library preparation. As can be expected from its implementation in ATAC-seq (assay for transposable-accessible chromatin using sequencing)[24], the Tn5 transposase has a high affinity for exposed DNA in open chromatin. While approaches exist to mitigate this bias[25], a recent systematic analysis of Tn5-based studies has provided preliminary indications that accessibility artefacts persist[26].

We recently developed scDam&T-seq, a method that measures DNA-protein contacts and transcription in single cells by combining single-cell DamID and CEL-Seq2[27]. DamID-based techniques attain specificity by tagging a protein of interest (POI) with the *E. coli* Dam methyltransferase, which methylates adenines in a GATC motif in the proximity of the POI[28–30].

The approach is especially suited for single-cell studies, because DNA-protein contacts are recorded directly on the DNA in the living cell, and downstream sample handling is limited. However, Dam cannot be tethered directly to post-translationally modified proteins by genetic engineering, which has precluded the use of DamID for studying histone PTMs.

Here, we present EpiDamID, an extension of existing DamID protocols, based on the fusion of Dam to chromatin-binding modules for the detection of various types of histone PTMs. We validate the specificity of EpiDamID in population (Fig. 3.1) and single-cell samples (Fig. 3.2). Subsequently, we leverage its single-cell resolution to study the Polycomb mark H3K27me3 and its relationship to transcription in mouse embryoid bodies (EBs) (Fig. 3.3), and identify distinct Polycomb-regulated and Polycomb-independent hierarchical TF networks (Fig. 4). Finally, we implement a protocol to assay cell type-specific patterns of the heterochromatic mark H3K9me3 in the zebrafish embryo and discover broad domains of heterochromatin specific to the notochord (Fig. 3.5). Together, these results show that EpiDamID provides a versatile tool that can be implemented in diverse biological settings to obtain single-cell histone PTM profiles.

Design

The conventional DamID approach involves genetically engineering a protein of interest (POI) to the bacterial methyltransferase Dam (Fig. 3.1a). In this study, we adapted the DamID method to detect histone PTMs by fusing Dam to one of the following: 1) full-length chromatin proteins, 2) tuples of well-characterized reader domains[31–33], or 3) single-chain variable fragments (scFv) also known as mintbodies[34–36] (Fig. 3.1a, Methods). Similar strategies have been successfully applied in microscopy, proteomics and ChIP experiments[34–38]. Our approach is henceforth referred to as EpiDamID, and the construct fused to Dam as the targeting domain. Since this approach can be applied to any existing DamID method, EpiDamID makes all these protocols available to the study of chromatin modifications. This includes the possibility to perform (live) imaging of Dam-methylated DNA[39–41], tissue-specific study of model organisms without cell isolation via Targeted DamID (TaDa)[42], DamID-directed proteomics[43], (multi-modal) single-cell[27,39,40,44] and single-molecule[45] sequencing studies, and the processing of samples with little material[40,46].

Results

Targeting domains specific to histone modifications mark distinct chromatin types with EpiDamID

We categorized the various targeting domains into the following chromatin types: accessible, active, heterochromatin, and Polycomb. We generated various expression constructs for each of the different targeting domains, testing promoters (HSP, PGK), orientations (Dam-POI, POI-Dam) and two versions of the Dam protein (DamWT, Dam126) (Table S3.1). The choice of promoter influences the expression level of the Dam-POI, whereas the orientation may affect target binding. In the Dam126 mutant, the N126A substitution diminishes off-target methylation[47,48]. We introduced the Dam constructs by viral transduction in hTERT- immortalized RPE-1 cells and performed DamID2 followed by high-throughput sequencing[49]. To validate our data with an orthogonal method, we generated ChIP-seq samples for various histone modifications.

The DamID samples were filtered on sequencing depth and information content (IC), a metric for determining signal-to-noise levels (Fig. S3.1a-b, Methods). IC additionally showed that tuples of reader domains fused to Dam typically perform better than single domains ($p < 0.05$ for three out of four domains, Fig. S3.1b), in agreement with a recent study employing similar domains for proteomics purposes[38] (Fig. S3.1a-b). Therefore, only data from the triple reader domains were included in further analyses.

Visualization of all filtered samples by uniform manifold approximation and projection (UMAP) shows that EpiDamID mapping identifies distinct chromatin types and that samples consistently group with their corresponding ChIP-seq datasets (Fig. 3.1b). Genome-wide DamID signal also correlates well with ChIP-seq signal (mean Pearson's r correlation coefficients from 0.40-0.64 for active marks, 0.58-0.61 for heterochromatin marks, and 0.56-0.60 for Polycomb marks) (Fig. 3.1c and S3.1c). Importantly, DamID samples do not group based on construct type, promoter, Dam type, sequencing depth, or IC (Fig. S3.1d-e), indicating that those properties do not influence target specificity. All targets display the expected patterns of enrichment along the linear genome (Fig. 3.1d, left), as well as genome-wide on-target signal (Fig. 3.1d, right). To further explore the specificity of constructs that target active chromatin, we compared signal of Dam- H3K9ac and Dam-TAF3 at H3K9ac ChIP-seq peaks with high and low H3K4me3 ChIP-seq levels. Dam-H3K9ac shows enrichment in both categories, while Dam-TAF3 is enriched specifically in the high-H3K4me3 category (Fig. S3.1f). This confirms that, while the untethered Dam protein preferentially marks accessible chromatin, targeting it to active regions of the genome yields specific methylation patterns.

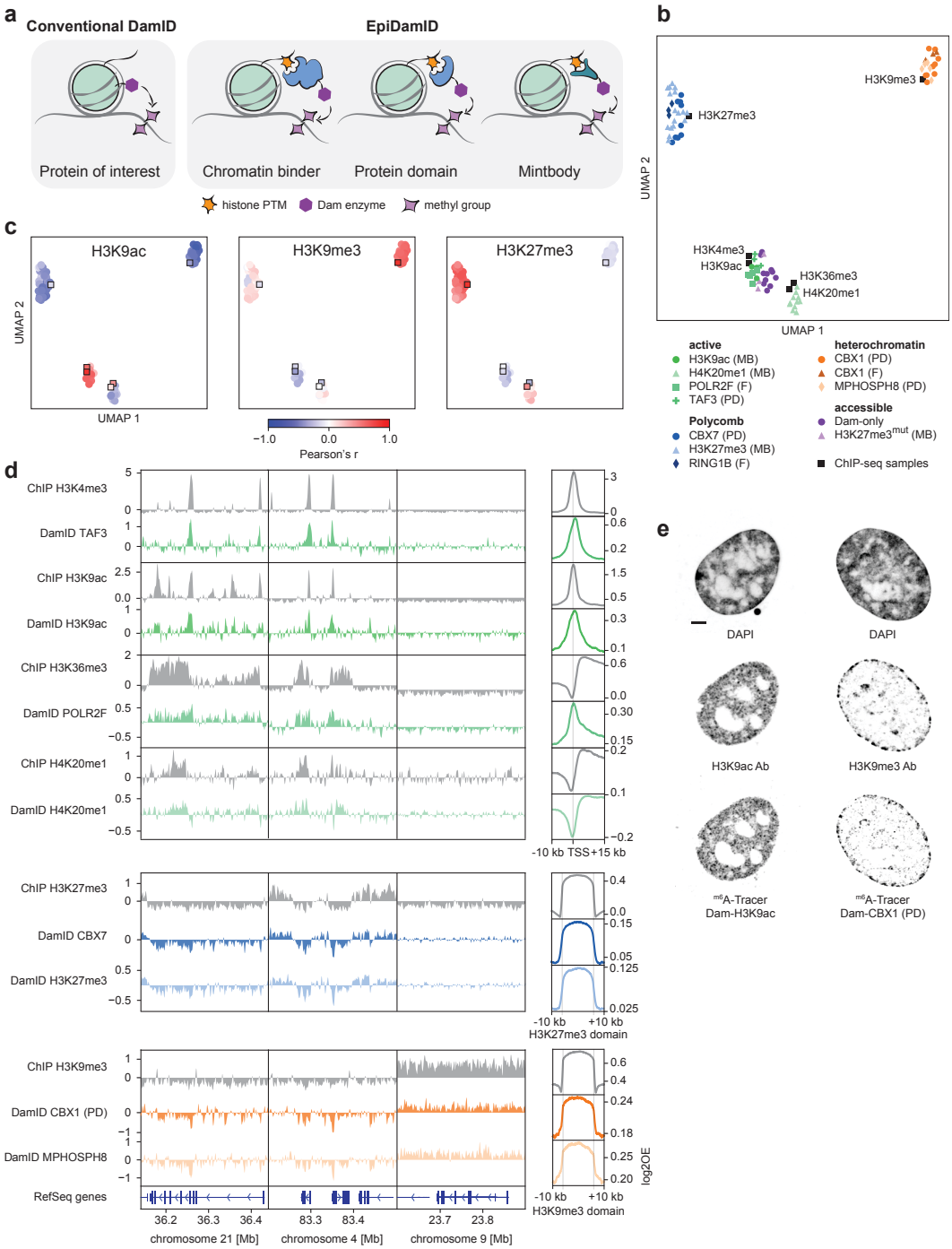


Fig. 3.1. Targeting domains specific to histone modifications mark distinct chromatin types with EpiDamID

a, Schematic overview of EpiDamID concept compared to conventional DamID. **b**, UMAP of DamID samples colored by targeting construct, and ChIP-seq samples of corresponding hPTMS. *[cont]*

MB: mintbody; PD: protein domain; F: full protein. **c**, UMAPs as in **b**, colored by correlation with selected ChIP-seq samples (H3K9ac, H3K9me3, and H3K27me3). Correlation values reflect the Pearson's correlation coefficient of Dam-normalized samples with the indicated ChIP-seq sample. Control constructs (Dam, H3K27me3mut) are excluded from the UMAP. DamID samples are circles; ChIP-seq samples are squares. **d**, Left: genome browser views of ChIP-seq (gray) and DamID (colored) enrichment. Data represent the combined signal of all samples of each targeting domain. Right: average DamID and ChIP-seq enrichment plots over genomic regions of interest. Signal is normalized for untethered Dam or input, respectively. Regions are the TSS (-10/+15 kb) of the top 25% H3K9ac-enriched genes for the active marks (top), and ChIP-seq domains (+/- 10 kb) for H3K27me3 (middle), and H3K9me3 (bottom). **e**, Confocal images of nuclear chromatin showing DAPI (top), immunofluorescent staining against an endogenous histone modifi-cation (middle), and its corresponding EpiDamID construct visualized with ^{m6}A-Tracer (bottom). Left: H3K9ac, right: H3K9me3. Scale bar: 3 µm.

Next, we quantified the spreading of Dam signal from its binding location to determine the resolution for all chromatin types. We found that DamID signal decays to 50% (from 100% at peak center or domain border) across a distance that extends ~1 kb past the ChIP-seq 50% decay point (Fig. S3.1g), implying a resolution of ~1-2 kb, similar to earlier studies with transcription factors[50,51]. It was previously reported that the Dam126 mutant improves signal quality compared to DamWT[47]. Indeed, this mutant markedly improved sensitivity and reduced background methylation (mean IC increase of 0.07-0.21 per construct) (Fig. S3.1h-i).

We further validated the correct nuclear localization of Dam-marked chromatin with microscopy, by immunofluorescent staining of endogenous histone PTMs and DamID visualization using ^{m6}A-Tracer protein[41,52] (Fig. 3.1e). Together, these results show that EpiDamID specifically targets histone PTMs and enables identification of their genomic distributions by next-generation sequencing.

Detection of histone PTMs in single mouse embryonic stem cells with EpiDamID

We next established EpiDamID for single-cell sequencing. To this end, we generated clonal, inducible mESC lines for the following targeting domains fused to Dam: H4K20me1 mintbody, H3K27me3 mintbody, and the H3K27me3-specific CBX7 protein domain (3x tuple). While H4K20me1 is enriched over the gene body of active genes[53], the heterochromatic mark H3K27me3 is enriched over the promoter of developmentally regulated genes[54,55]. As controls, we included an H3K27me3^{mut} mintbody construct whose antigen-binding ability is abrogated by a point mutation in the third complementarity determining region of the heavy chain (Y105F), and a published mESC line expressing untethered Dam[27]. We performed scDam&T-seq to generate 442-1,402 single-cell samples per construct, retaining 283-855 samples after filtering on the number of unique GATCs and IC (10,417-45,067 median unique counts per construct and median IC of 2.0-2.9) (Fig. S3.2a-c, Table S3.2). For subsequent analyses, we also included a published dataset of Dam fused to RING1B[27] as an example of a full-length chromatin reader targeting Polycomb chromatin.

All constructs contained DamWT, as the Dam126 methylation levels were found insufficient to produce high-quality single-cell signal (data not shown).

Dimensionality reduction of the single-cell datasets revealed that the samples primarily separated on chromatin type (Fig. 3.2a). To further confirm the specificity of the constructs, we used mESC H3K27me3 (ENCSR059MBO) and H3K9ac (ENCSR000CGP) ChIP-seq datasets from the ENCODE portal[56] and generated our own for H4K20me1. For all single cells, we computed the enrichment of counts within H3K27me3, H3K9ac and H4K20me1 ChIP-seq domains. These results show a strong enrichment of EpiDamID counts within domains for the corresponding histone PTMs (Fig. 3.2b-d, Fig. S3.2d), indicating that the methylation patterns are specific for their respective chromatin targets, even at the single-cell level.

These results demonstrate that mintbodies and protein domains can be used to map histone PTMs in single cells with EpiDamID.

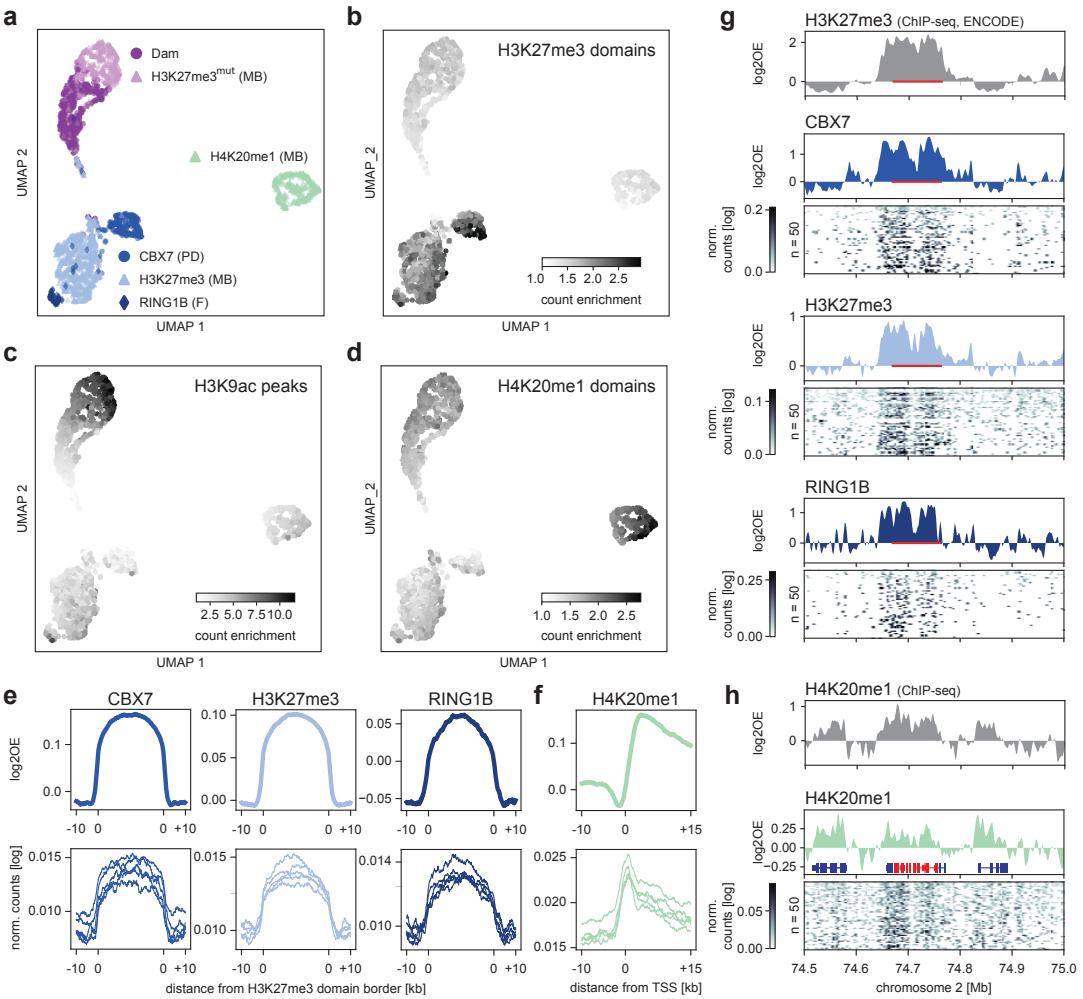


Fig. 3.2. Detection of histone PTMs in single mouse embryonic stem cells with EpiDamID. **a**, UMAP based on the single-cell DamID readout of all single-cell samples. MB: mintbody; PD: protein domain; F: full protein. **b**-**d**, DamID UMAP as in A, colored by the enrichment of counts within H3K27me3 ChIP-seq domains (**b**), H3K9ac ChIP-seq peaks (**c**), and H4K20me1 ChIP-seq domains (**d**). **e**, Average signal over H3K27me3 ChIP-seq domains of CBX7 and H3K27me3 targeting domains and full-length RINGB1B protein. **f**, Average H4K20me1 signal over the TSS of the top 25% active genes (based on H3K9ac ChIP-seq signal). **e**-**f**, Top: in silico populations normalized for Dam; Bottom: five of the best single-cell samples (bottom) normalized only by read depth. **g**-**h**, Signal of various marks over the HoxD cluster and neighboring regions. ChIP-seq data is normalized for input control. DamID tracks show the Dam-normalized in silico populations of the various Dam-fusion proteins, DamID heatmaps show the depth- normalized single-cell data of the fifty richest cells. The HoxD cluster is indicated in red in **g** (bar) and **h** (RefSeq); additional RefSeq genes are shown in **h**. The combined single-cell data also showed the expected enrichment over H3K27me3 ChIP-seq domains (Fig. 3.2e) and active gene bodies (Fig. 3.2f) for the Polycomb-targeting constructs and H4K20me1, respectively. Contrary to the H3K27me3 construct, H3K27me3^{mut} showed little enrichment over H3K27me3 ChIP-seq domains (Fig. S3.2e). The specificity of the signal is also evident at individual loci in both the *in silico* populations and single cells (Fig. 3.2g-h and Fig. S3.2f).

Joint profiling of Polycomb chromatin and gene expression in mouse EBs

To exploit the benefits of simultaneously measuring histone PTMs and transcriptome, we profiled Polycomb chromatin in mouse EBs. We targeted the two main Polycomb repressive complexes (PRC) with EpiDamID using the full-length protein RING1B and H3K27me3-mintbody fused to Dam. RING1B is a core PRC1 protein that mediates H2AK119 ubiquitylation[57,58], and H3K27me3 is the histone PTM deposited by PRC2[59–62]. Both PRC1 and PRC2 have key roles in gene regulation during stem cell differentiation and early embryonic development (see refs [63,64] for recent reviews on this topic).

To assay a diversity of cell types at various stages of differentiation, we harvested EBs for scDam&T-seq at day 7, 10 and 14 post aggregation, next to ESCs grown in 2i/LIF (Fig. 3.3a). We used Hoechst incorporation in combination with fluorescence-activated cell sorting (FACS) to deposit live, single cells into 384-well plates and record their corresponding cell cycle phase (Methods). In addition to RING1B and H3K27me3-mintbody, we included the untethered Dam protein for all time points as a control for chromatin accessibility. Collectively, we obtained 2,943 cells after filtering (Fig. S3.3a-b), in a similar range as CoTECH (~7,000 cells), higher than SET-seq (~500 cells) and lower than Paired-Tag (~65,000 nuclei). The number of unique genomic and transcriptomic counts per cell was similar or higher compared to the other methods (Fig. S3.3a-b). Based on the transcriptional readout, we identified eight distinct clusters across time points (Fig. 3.3b). We integrated the EB transcriptome data with the publicly available mouse embryo atlas[65] to confirm the correspondence of cell types with early mouse development and guide cluster annotations (Fig. S3.3c-d).

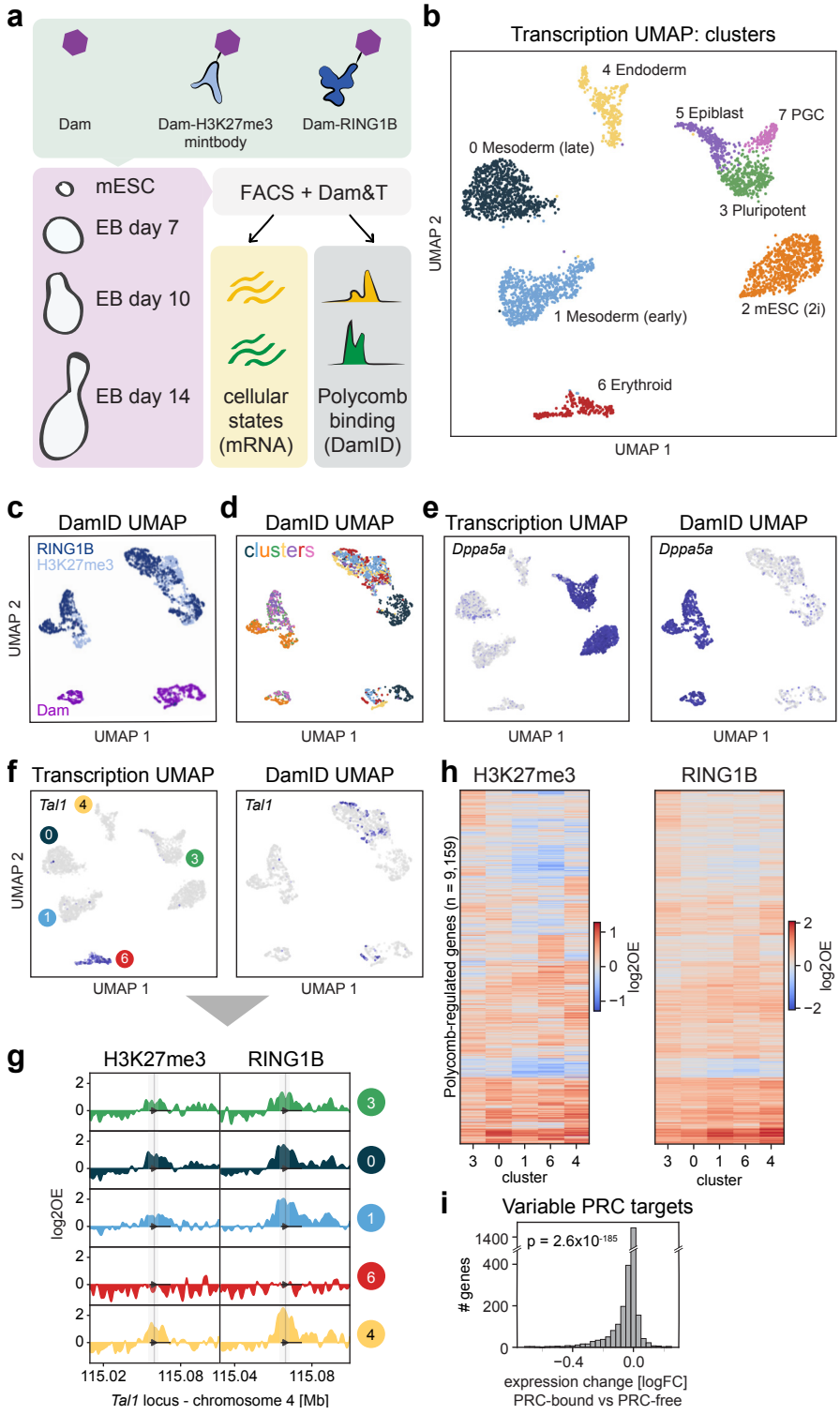

a
Dam
Dam-H3K27me3 mintbody
Dam-RING1B
mESC
EB day 7
EB day 10
EB day 14
FACS + Dam&T
cellular states (mRNA)
Polycomb binding (DamID)
b
Transcription UMAP: clusters
4 Endoderm
5 Epiblast
7 PGC
0 Mesoderm (late)
3 Pluripotent
2 mESC (2i)
1 Mesoderm (early)
6 Erythroid
UMAP 2
UMAP 1
c
DamID UMAP
RING1B
H3K27me3
Dam
UMAP 2
UMAP 1
d
DamID UMAP
clusters
UMAP 1
e
Transcription UMAP
Dppa5a
UMAP 1
DamID UMAP
Dppa5a
UMAP 1
f
Transcription UMAP
Tal1
4
0
3
1
6
UMAP 2
UMAP 1
DamID UMAP
Tal1
UMAP 1
g
H3K27me3
RING1B
log2OE
3
0
1
6
4
115.02
115.08
115.04
115.08
Tal1 locus - chromosome 4 [Mb]
h
H3K27me3
RING1B
Polycomb-regulated genes (n = 9,159)
log2OE
cluster
3 0 1 6 4
i
Variable PRC targets
p = 2.6x10-185
# genes
1400
400
200
0
−0.4
0.0
expression change [logFC]
PRC-bound vs PRC-free

Fig. 3.3. Joint profiling of Polycomb chromatin and gene expression in mouse EBs.
a, Schematic showing the experimental design. **b**, UMAP of samples based on transcriptional readout, colored by cluster. **c**-**d**, UMAP of samples based on DamID readout, colored by construct (**c**) and cluster (**d**). **e**, Transcriptomic UMAP (left) and DamID UMAP (right), colored by expression of pluripotency marker *Dppa5a*. **f**, Transcriptomic UMAP (left) and DamID UMAP (right), colored by expression of hematopoietic regulator Tal1. **g**, Genomic tracks of H3K27me3 and RING1B DamID signal per cluster at the *Tal1* locus. **h**, Heatmaps showing the H3K27me3 (left) and RING1B (right) DamID signal of all identified PRC targets for transcriptional clusters 3, 0, 1, 6, and 4. PRC targets are ordered based on hierarchical clustering. **i**, Fold-change in expression of Polycomb targets between clusters where the gene is PRC-associated and clusters where the gene is PRC-free. The significance was tested with a two-sided Wilcoxon's signed rank test ($p = 2.6 \times 10^{-185}$). This indicated the presence of pluripotent and differentiated cellular states, including epiblast, endoderm, and mesoderm lineages. Notably, the DamID readout alone was sufficient to consistently separate cells on chromatin type (Fig. 3.3c) and to distinguish between the pluripotent and more lineage-committed cells (Fig. 3.3d-e). Thus, the EpiDamID profiles display cell type-specific patterns of chromatin accessibility and Polycomb association. Prompted by this observation, we trained a linear discriminant analysis (LDA) classifier to assign an additional 1,543 cells with poor transcriptional data to cell type clusters, based on their DamID signal (Fig. S3.3e, Table S3.2).

Next, we defined the set of genes that is Polycomb-regulated in the EB system. First, we determined the H3K27me3 and RING1B signal at the promoter region of all genes and compared these two readouts across the clusters. This confirmed good correspondence between H3K27me3 and RING1B profiles (Pearson's r = 0.60-0.82, p = 0 between profiles of the same cluster) (Fig. S3.3f-g), albeit with a slightly higher signal amplitude for RING1B (Fig. S3.3g). This difference between RING1B and H3K27me3 may be biological (e.g., differential binding sites or kinetics) and/or technical (e.g., the use of a full-length protein versus a mintbody to target Dam). Nonetheless, because of the overall similarity, we decided to classify high-confidence Polycomb targets as having both H3K27me3 and RING1B enrichment in at least one of the EB clusters (excluding cluster 7 due to the relatively low number of cells) or in the previous ESC data set. We identified 9,159 Polycomb-regulated targets across the dataset, in good concordance with previous work in mouse development (4,059 overlapping genes out of a total of 5,986; $p = 9.5 \times 10^{-135}$, Chi-square test)[66] (Fig. S3.3h).

Next, we intersected the cluster-specific transcriptome and DamID data to relate gene expression patterns to Polycomb associations. Based on the role of Polycomb in gene silencing, differential binding of PRC1/2 to genes is expected to be associated with changes in expression levels. As exemplified in Fig. 3.3f-g, the cell type-specific expression of *Tal1*, a master regulator in hematopoiesis, is indeed inversely related to Polycomb enrichment. This negative association is apparent for all PRC targets that are upregulated in the hematopoietic cluster (Fig. S3.3i-j).

In addition, unsupervised clustering of H3K27me3 and RING1B promoter occupancy shows variation in signal between target genes as well as between cell types, indicating dynamic regulation of these targets in EBs (Fig. 3.3h). In line with this, Polycomb targets with variable PRC occupancy are typically more highly expressed in those clusters where Polycomb is absent (Wilcoxon's signed-rank test, $p = 2.6 \times 10^{-185}$, Fig. 3.3i). Since the negative relationship between Polycomb occupancy and transcription is not perfect, we were interested to see whether an additional layer of epigenetic regulation could further explain the observed transcriptional changes. To this end, we integrated our data with a publicly available scNMT- seq dataset[67], also generated in EBs (Fig. S3.3k). This resulted in sufficient scNMT-seq samples in four clusters to compare CpG methylation profiles with Polycomb occupancy. The integrated profiles indeed revealed a complementary relationship between the two marks, where genes with either CpG methylation or Polycomb at their promoter tend to be expressed at lower levels (Fig. S3.3l). This was also apparent for CpG methylation and expression of genes with variable Polycomb enrichment between the clusters (Fig. S3.3m). The observed trends are in line with the known repressive effects of both marks and their largely mutually exclusive localizations[68–70].

Collectively, these data illustrate the strength of EpiDamID to jointly capture transcription and chromatin dynamics during differentiation, as well as the potential to integrate the results with datasets derived from different techniques.

Polycomb-regulated transcription factors form separate regulatory networks

We next focused on the Polycomb targets based on their function, and found that TF genes are over-represented within the Polycomb target genes (Fig. S3.4a), in line with previous observations[54]. Nearly half of all TF genes in the genome (761/1689) is bound by Polycomb in at least one cluster. In addition, genes encoding TFs generally accumulate higher levels of H3K27me3 and RING1B compared to other protein-coding genes (Fig. S3.4b). Consistent with an important role in lineage specification, Polycomb-controlled TFs are expressed in a cell type-specific pattern, as opposed to the more constitutive expression across cell types for Polycomb-independent TFs (Fig. S3.4c-d). Accordingly, the Polycomb-controlled TFs are enriched for Gene Ontology (GO) terms associated with animal development (Fig. S3.4e).

The high Polycomb occupancy at developmentally regulated TF genes prompted further investigation into the role of Polycomb in TF network hierarchies. We used SCENIC to systematically identify target genes that are associated with the expression of TFs[71,72]. SCENIC employs co-expression patterns and binding motifs to link TFs to their targets, together henceforth termed "regulons" (per SCENIC nomenclature). We identified 285 "activating" regulons after filtering (Fig. 3.4a, Methods). While regulons and their activity were found independently of RNA-based cluster annotations, regulon activity trends clearly matched the annotated clusters (Fig. 3.4a).

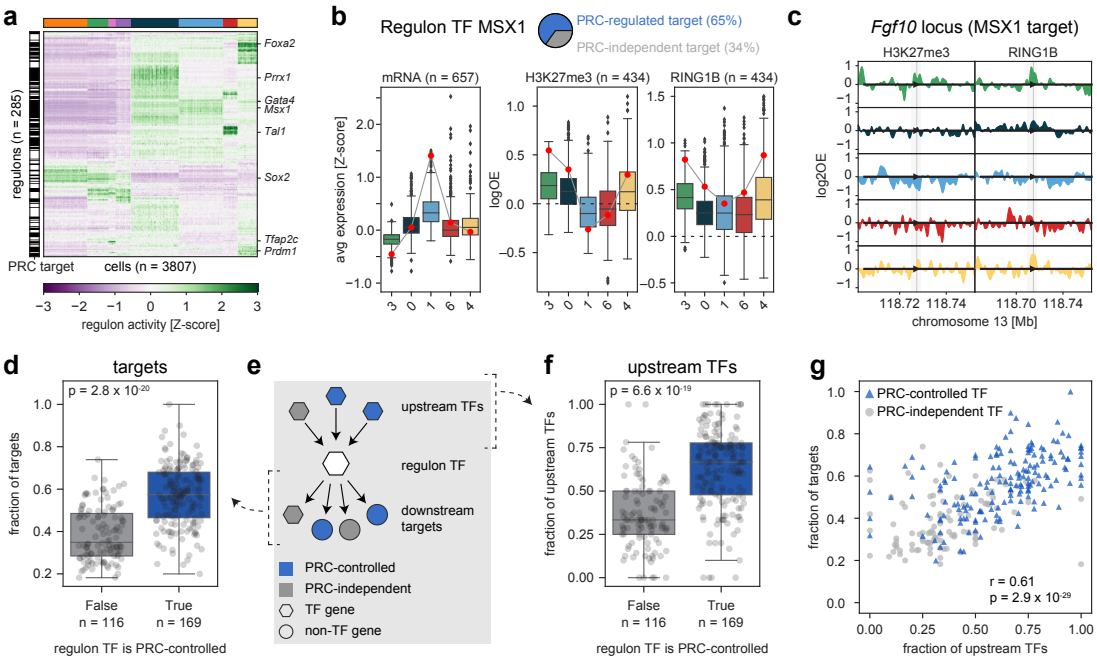


Fig. 3.4. Polycomb-regulated transcription factors form separate regulatory networks.

a, Heatmap showing SCENIC regulon activity per single cell. Cells (columns) are ordered by transcriptional cluster; regulon (rows) are ordered by hierarchical clustering. The black and white bar on the left indicates whether the regulon TF is a PRC target (black) or not (white). **b**, Example of the relationship between expression and Polycomb regulation for the MSX1 regulon. Pie chart indicates the percentages of PRC-controlled (blue) or PRC-independent (grey) target genes. Left: boxplots showing target gene expression per cluster for all target genes. Middle and right: boxplots showing the H3K27me3 and RING1B DamID signal at the TSS per cluster for the PRC-controlled target genes. The expression and DamID signal of Msx1 is indicated with a red circle. **c**, Genomic tracks of H3K27me3 and RING1B DamID signal per cluster at the Fgf10 locus, one of the target genes of MSX1. Arrow head indicates the location of the TSS; shaded area indicates -5kb/+3kb around the TSS. **d**, Boxplots showing the fraction of PRC-controlled target genes, split by whether the TF itself is PRC-controlled. The significance was tested with a two-sided Mann-Whitney U test . **e**, Schematic of the regulatory network, indicating the relationship between a regulon TF (white hexagon), its upstream regulators (colored hexagons), and its downstream targets (colored hexagons/circles). **f**, Boxplots showing the fraction of PRC-controlled upstream regulators, split by whether the regulon TF is PRC-controlled. The significance was tested with a two-sided Mann-Whitney U test. **g**, Scatter plot showing the relationship between the fraction of Polycomb-controlled targets and regulators of a regulon TF. Regulon TFs that are PRC-controlled are indicated in blue; regulon TFs that are PRC-independent are indicated in grey. Correlation was computed using Pearson's correlation.

We first determined how overall regulon activity identified by SCENIC correlates to PRC binding. As illustrated for the homeobox TF gene *Msx1*, we found that regulon activity is generally inversely related to PRC association of both the TF gene (red dot) and its PRC-controlled targets (boxplots, 65% of all MSX1 targets) (Fig. 3.4b-c).

We wondered whether there is a general preference for Polycomb-controlled TFs to target genes that themselves are regulated by Polycomb. Indeed: while Polycomb-controlled TFs have a similar number of target genes compared to other TFs (Fig. S3.4f), the expression of the targets is much more frequently controlled by Polycomb than expected by chance (Mann-Whitney-U test $p = 2.8 \times 10^{-20}$, Fig. 3.4d). This effect is even stronger when considering the subset of targets that is exclusively regulated by Polycomb TFs (Chi-square test $p = 0$, Fig. S3.4g). Similarly, upstream TFs controlling the regulon TFs (Fig. 3.4e) also tend to be Polycomb-controlled (Mann-Whitney-U test, $p = 6.6 \times 10^{-19}$, Fig. 3.4f). Moreover, the fractions of Polycomb-controlled upstream regulators and downstream targets are correlated (Pearson's $r = 0.61$, $p = 2.9 \times 10^{-29}$, Fig. 3.4g), indicating consistency in the level of Polycomb regulation across at least three layers of the TF network. This trend is especially strong for the lineage-specific genes (Pearson's $r = 0.48$, $p = 9.2 \times 10^{-8}$), but also holds for other, unspecific, genes (Pearson's $r = 0.41$, $p = 4.0 \times 10^{-4}$) (Fig. S3.4h-i). These results suggest that Polycomb-associated hierarchies exist, forming relatively separate networks isolated from other gene regulatory mechanisms, and that this phenomenon extends beyond lineage-specific genes alone.

Together, the above findings demonstrate that single-cell EpiDamID can be successfully applied in complex developmental systems to gather detailed information on cell type-specific Polycomb regulation and its interaction with transcriptional networks.

Implementation of EpiDamID during zebrafish embryogenesis

Next, we applied EpiDamID in an *in vivo* system to study the heterochromatic mark H3K9me3 during zebrafish development. To bypass the need for genetic engineering, we employed microinjection of mRNA into the zygote (Fig. 3.5a), a strategy successfully applied in the mouse embryo[40]. H3K9me3 is reprogrammed during the early stages of development in several species[73–77] and the deposition of this mark coincides with decreased developmental potential[78]. It was previously shown that H3K9me3 is largely absent before the maternal-to-zygotic transition (MZT)[74], but it remains unclear whether the H3K9me3 distribution undergoes further remodeling after this stage, and whether its establishment differs across cell types during development.

We injected mRNA encoding the H3K9me3-specific construct *Dam-Mphosph8* and untethered *Dam* into the yolk at the one-cell stage and collected embryos at the 15-somite stage (Fig. 3.5a), which comprises a wide diversity of cell types corresponding to all germ layers. We generated 2,127 single-cell samples passing both DamID and CEL-Seq2 thresholds (Fig. S3.5a, Table S3.2). Comparing the DamID data of an *in silico* whole-embryo sample to published H3K9me3 ChIP-seq data of 6-hpf embryos[74] showed good concordance (Pearson's $r = 0.72$, $p = 0$; Fig. S3.5b).

Broad domains of notochord-specific H3K9me3 enrichment revealed by scDam&T

Analysis of the single-cell transcriptome data resulted in 22 clusters of diverse cell types (Fig. 3.5b), which we annotated according to expression of known marker genes (Fig. S3.5c). After dimensionality reduction based on the DamID signal, we observed a clear visual separation of cells in accordance with their Dam construct, and to a lesser extent with their cell type (Fig. 3.5c-d). Cluster-specific DamID profiles allowed us to employ the LDA classifier to assign a further 705 cells with poor transcriptional readout to a cluster (Fig. S3.5d, Table S3.2). Notably, the MPHOSPH8 samples of hatching gland (cluster 1, *he1.1* expression) and notochord (cluster 2, *col9a2* expression) segregated strongly from the other cell types (Fig. 3.5d), implying differences in their single-cell H3K9me3 profiles. In particular, we observed the appearance of large domains of H3K9me3 enrichment in the notochord, and seemingly lower levels of H3K9me3 in the hatching gland (Fig. 3.5e and Fig. S3.5e).

Next, to more systematically identify and characterize regions of differential H3K9me3 enrichment between cell clusters, we performed ChromHMM[79,80]. The approach uses the H3K9me3 signal per cluster to annotate genomic segments as belonging to different H3K9me3 states. We included the 12 cell clusters containing >30 cells per construct and identified five H3K9me3 states across the genome. These represented: A) three states of constitutive H3K9me3 with different enrichment levels [A1-A3], B) notochord-specific H3K9me3 enrichment, and C) constitutive depletion of H3K9me3 (Fig. 3.5f-g). While all 12 clusters had the highest H3K9me3 enrichment in state A1, cells belonging to the hatching gland (cluster 1) tended to have lower signal in these regions compared to other cell types (Fig. S3.5f). Notochord cells (cluster 2), conversely, displayed somewhat higher enrichment in state A1 and dramatically higher enrichment in state B compared to the others. State A (A1-3) chromatin forms broad domains (Fig. S3.5g) that together comprise 27% of the genome (Fig. S3.5h) and, as expected for H3K9me3-associated chromatin regions, are characterized by sparser gene density and lower gene activity compared to the H3K9me3-depleted state C (Fig. 3.5h). Moreover, state A1 is strongly enriched for zinc-finger transcription factors (Fig. S3.5i), which are known to be demarcated by H3K9me3 in other species[81,82]. The notochord-specific state B has similar characteristics to states A1-A3 (Fig. 3.5h, S3.5g-i), yet exhibits broader consecutive regions of H3K9me3 enrichment (Fig. 3.5g and S3.5g) and an even lower active gene density (Fig. 3.5h). However, we did not find a notable increase in H3K9me3 at genes downregulated in notochord (Fig. S3.5j), implying that these domains do not play a role in gene expression regulation.

One of the known functions of H3K9me3 chromatin is the repression of transposable elements[83–85]. Indeed, it was previously observed in zebrafish that nearly all H3K9me3 domains in early embryos are associated with repeats[74].

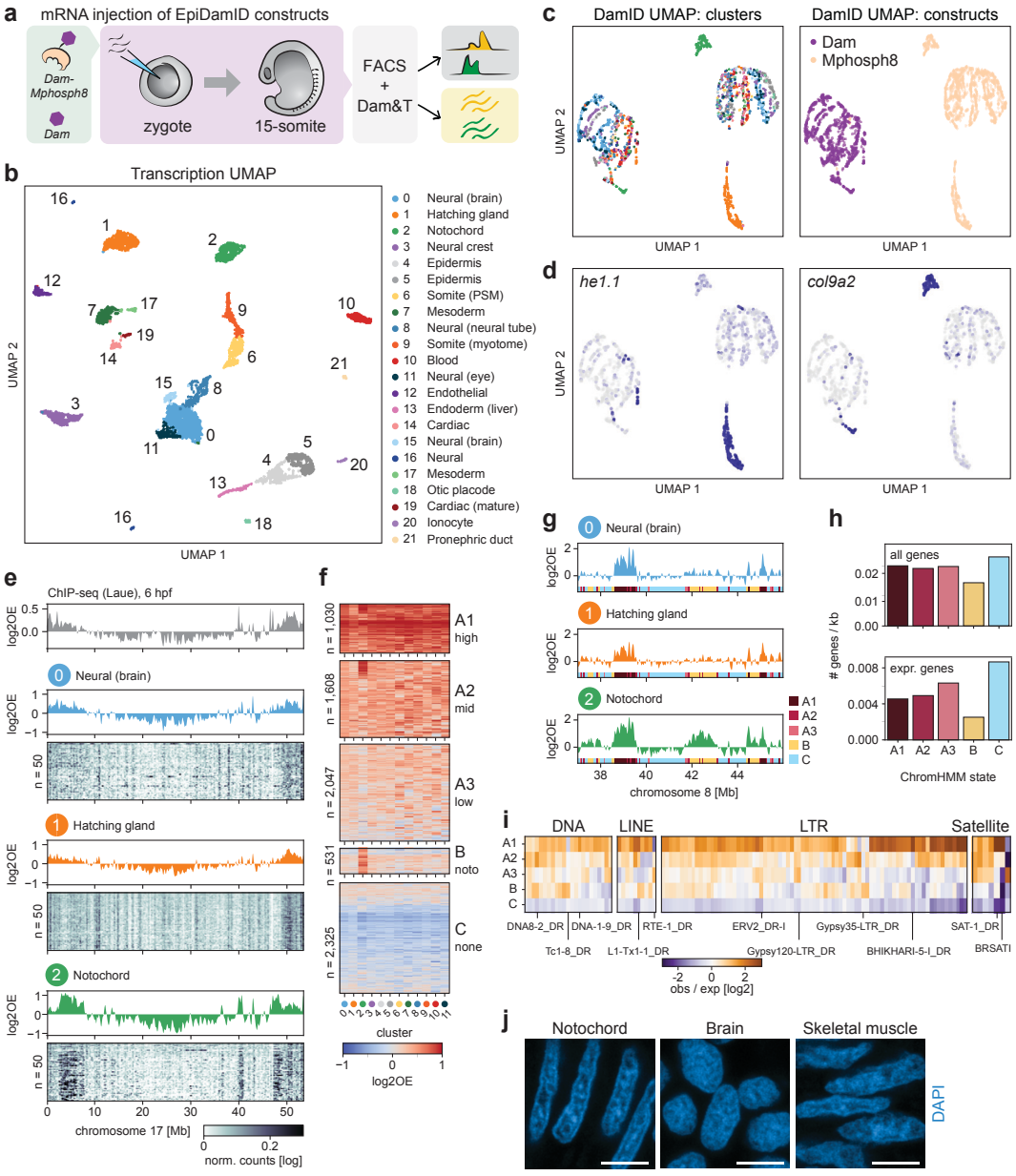


Fig. 3.5. Notochord-specific H3K9me3 enrichment in the zebrafish embryo.

a, Schematic representation of the experimental design and workflow. **b**, UMAP based on the transcriptional readout of all single-cell samples passing CEL-Seq2 thresholds (n = 3902). **c**, UMAP based on the genomic readout of all single-cell samples passing DamID thresholds (n = 2833). Samples are colored by transcriptional cluster (left) and Dam-targeting domain (right). **d**, Expression of the hatching gland marker he1.1 (left) and the notochord marker col9a2 (right) projected onto the DamID UMAP. **e**, Genomic H3K9me3 signal over chromosome 17. Top track: H3K9me3 ChIP-seq signal of 6-hpf embryo. Remaining tracks: combined single-cell Dam-MPHOSPH8 data for clusters 0-2. Heatmaps show the depth-normalized Dam-MPHOSPH8 data of the 50 richest cells. *[cont]*

f, Heatmap showing the cluster-specific average H3K9me3 enrichment over all domains called per ChromHMM state. Per state, domains were clustered using hierarchical clustering. **g**, Genomic H3K9me3 signal over a part of chromosome 8 for clusters 0-2. The colored regions at the bottom of each track indicate the ChromHMM state. **h**, Gene density of all genes (top) and expressed genes (bottom) per state. **i**, Enrichment of repeats among the ChromHMM states. Example repeats are indicated. **j**, Representative images of DAPI staining in cryosections of zebrafish embryos at 15-somite stage. Scale bars: 4 μm.

We determined whether distinct repeat classes were over-represented in each H3K9me3 ChromHMM state (Fig. S3.6a) and found a strong enrichment of several repeat classes in state A1, including LTR and tRNA. Further discrimination within the classes showed a high frequency of pericentromeric satellite repeats SAT-1 and BRSATI in state A1 (Fig. 3.5i), in line with the known occupancy of H3K9me3 at pericentromeric regions. Inspection of the DamID patterns showed a clear increase of signal centered on specific repeat regions in state A1, and to lesser extents in other states (Fig. S3.6b). In addition, we found that state B harbors specific enrichment of certain repeats (Fig. 3.5i and Fig. S3.6c), although further study is required to determine whether H3K9me3 is involved in cell type-specific repression of repetitive genomic regions in the notochord.

Altered expression of chromatin proteins and pronounced nuclear compartmentalization in notochord

Finally, we evaluated cluster-specific expression of known chromatin proteins in relation to the differential H3K9me3 patterns. Expression levels of histone methyltransferases, demethylases and other chromatin factors did not show an upregulation of known H3K9 methyltransferases (setdb2, setdb1a/b, suv39h11a/b, ehmt2) nor demethylases (kdm4aa/ab/b/c, phf8) in notochord (Fig. S3.6d). However, the H3K9- and H3K36-specific demethylase kdm4c was exclusively upregulated in hatching gland, which could explain the low H3K9me3 levels in this cluster. Notably, the notochord cluster showed significant upregulation of lmna, the gene encoding nuclear lamina protein Lamin A/C that associates with heterochromatin[86] and plays an important structural role in the nucleus[86,87]. This could be relevant in relation to the structural role of the notochord and the resulting mechanical forces the cells are subjected to[88]. To more directly investigate chromatin state and nuclear organization in these embryos, we performed confocal imaging of H3K9me3 and DAPI stainings in notochord, brain, and skeletal muscle. H3K9me3-marked chromatin displayed a typical nuclear distribution in all tissues, including heterochromatin foci as previously reported[74] (Fig. S3.6e). DAPI staining showed more structure in the notochord compared to the other tissues (Fig. 3.5j), visible as a clear rim along the nuclear periphery and denser foci within the nuclear interior. This indicates a stronger separation between euchromatin and heterochromatin, although it remains to be elucidated whether these features are related to the notochord-specific H3K9me3 domains in the genome.

The implementation of EpiDamID in zebrafish embryos shows that this strategy provides a flexible and accessible approach to generate high-resolution single-cell information on the epigenetic states that underlie biological processes during organismal development.

Discussion

Advantages of DamID for single-cell multi-modal omics during embryo development

The DamID workflow involves few enzymatic steps and is thus especially suitable for integration with other single-cell protocols to achieve multi-modal measurements[49]. Minimal sample handling prior to molecular processing results in a high recovery rate of collected cells[40]; for example, scDam&T-seq with EpiDamID constructs could be used to individually assay all cells of a single preimplantation mouse embryo and examine epigenetic and transcriptomic differences that may point towards cell fate commitment, while tracking intra- embryonic variability. Further, DamID genomic marks are stable upon deposition, offering the possibility to track ancestral EpiDamID signatures through mitosis to study inheritance and spatial distribution of epigenetic states in daughter cells[41,89].

Comparison to other single-cell transcriptome and chromatin profiling techniques

In the past year, three other techniques have been published that are capable of simultaneously measuring chromatin modifications and transcription: Paired-Tag[21], CoTECH[22], and SET-seq[23]. One major conceptual difference between above methods and DamID-based techniques is the manner of capturing DNA in proximity of the chromatin mark of interest. Strategies leveraging CUT&Tag obtain a readout of chromatin by targeting protein A fused to transposase Tn5 (pA-Tn5) to antibody-bound regions, and integrating barcoded adapters into the surrounding DNA. DamID deposits signal in living cells over time; consequently, it represents a historic record of chromatin state over a period of multiple hours up to a full cell cycle, while antibody-based techniques provide a snapshot view. In DamID, regions that are only transiently bound by the mark of interest will thus be represented more strongly in the signal relative to CUT&Tag-based methods. Another key difference is the extent to which chromatin accessibility affects the data. DamID techniques are known to have an accessibility signature due to extended exposure to free-floating Dam protein (discussed in more detail under *Limitations*), which is controlled for by performing experiments with untethered Dam. While CUT&Tag- and CUT&RUN-based methods have reported less of such an accessibility bias and do not customarily include explicit control experiments, early results[26] suggest that such a bias may indeed be present. The question of data interpretation and normalization in light of this bias should be carefully considered among all existing single-cell genomics techniques.

With regard to the transcriptional readout, the four techniques also employ different approaches: Paired-Tag exclusively amplifies the nuclear fraction of mRNA, SET-seq separates and measures total RNA in the cytoplasm, while CoTECH and scDam&T-seq both amplify the total mRNA. Finally, the Paired-Tag and CoTECH protocols have been adapted for combinatorial indexing and consequently have a higher throughput compared to scDam&T-seq and SET-seq.

Limitations

EpiDamID requires the expression of a construct encoding for the Dam-fusion protein in the system of interest. This may involve a substantial time investment depending on the system of choice and conditions generally need to be optimized for each Dam-fusion protein to optimize signal quality. DamID techniques are also limited in their resolution by the distribution of GATC motifs in the genome (median inter-GATC distance: 263 bp in mouse, 265 bp in human). In addition, we and others[47,50,51] have found that the methylation spreads ~1 kb from the site of binding (Fig. S3.1g), thus yielding an empirical resolution of 1-2 kb. This is sufficient to study the localization of many chromatin factors, but may be restrictive when exact binding sites are required. Finally, due to the *in vivo* expression and consequent roaming of the Dam-POI in the nucleus, spurious methylation gradually accumulates in unspecific, mostly accessible, chromatin regions. The degree of accumulated background signal differs substantially between different Dam-POIs, yet interferes most with proteins that reside within active chromatin. This can be overcome through computational normalization to the untethered Dam protein. In the case of single-cell experiments, this requires the grouping of similar cells into *in silico* populations. While this strategy yields good results, it does not provide a way to eliminate the accessibility component in individual cells, and the signal in single cells should therefore be interpreted as convolution of on-target and accessibility signal. Computational imputation of accessibility signal based on transcriptional similarity between targeted samples and Dam control samples could provide a solution to this problem, similar to current single-cell transcriptional imputation methods (see ref [90] for an overview). We explored one experimental strategy to reduce off-target effects by implementing Dam mutants with decreased affinity for DNA, which yielded promising results in population data but insufficient ^{m6}A-events for single-cell profiling. Further adaptation of the Dam protein to engineer an enzyme with high enzymatic activity and reduced DNA-binding affinity may further improve the quality of EpiDamID profiles in single cells. Alternatively, molecular processing could be extended to facilitate an orthogonal accessibility readout from the same sample.

Acknowledgements

We would like to thank the members of the Kind laboratory for their helpful comments and suggestions. In particular, we thank Koos Rooijers for providing input and support on the computational work. This work was funded by an ERC Starting Grant (ERC-StG 678423-EpiID) to JK. The Oncode Institute is partially funded by the KWF Dutch Cancer Society. IG is supported by an EMBO Long-Term Fellowship ALTF1214-2016, Swiss National Science Fund grant P400PB_186758 and NWO-ENW Veni grant VI.Veni.202.073. PDN is supported by an EMBO Long-Term Fellowship ALTF1129-2015, HFSPO Fellowship (LT001404/2017-L) and an NWO-ZonMW Veni grant (016.186.017-3). The laboratory of JB is supported by the Netherlands Cardiovascular Research Initiative: An initiative with support of the Dutch Heart Foundation and Hartekind, CVON2019-002 OUTREACH. The laboratory of HK is supported by MEXT/JSPS KAKENHI (JP18H05527, JP20K06484, and JP21H04764), and Japan Science and Technology Agency (JPMJCR16G1 and JPMJCR20S6). We additionally thank the Hubrecht Sorting Facility, the Hubrecht Imaging Center, and the Utrecht Sequencing Facility (USEQ) subsidized by the University Medical Center Utrecht.

Author contributions

Conceptualization: FJR, KLdL, SSdV, JK. Data curation & Validation: FJR, KLdL, SSdV. Formal analysis & Software: FJR. Funding acquisition & Project administration: JK. Investigation & Methodology: KLdL and SSdV designed and performed all experiments unless noted otherwise. CVQ and EB designed and generated knock- in mouse ESC lines. PDN performed all zebrafish experiments, with assistance from IG and SSdV. Resources: YS, HK. Supervision: JB, JK. Visualization: FJR, KLdL. Writing – original draft: FJR, KLdL, JK. Writing – review & editing: all authors.

Methods

Cell lines

All cell lines were grown in a humidified chamber at 37°C in 5% CO_2, and were routinely tested for mycoplasma. Human TERT-immortalized RPE-1 cells were cultured in DMEM/F12 (Gibco) containing 10% FBS (Sigma F7524 lot BCBW6329) and 1% Pen/Strep (Gibco). This cell line does not contain a Y chromosome. Human HEK293T cells were cultured in DMEM (Gibco) containing 10% FBS and 1% Pen/Strep (Gibco). This cell line does not contain a Y chromosome. Mouse F1 hybrid Cast/EiJ x 129SvJae embryonic stem cells (mESCs; a kind gift from the Joost Gribnau laboratory) were cultured on irradiated primary mouse embryonic fibroblasts (MEFs), in mESC culture media CM+/+ defined as follows: G-MEM (Gibco) supplemented with 10% FBS (Sigma F7524 lot BCBW6329), 1% Pen/Strep (Gibco), 1x GlutaMAX (Gibco), 1x non-essential amino acids (Gibco), 1x sodium pyruvate (Gibco), 0.1 mM β-mercaptoethanol (Sigma) and 1000 U/mL ESGRO mLIF (EMD Millipore ESG1107). Cells were split every 3 days and medium was changed every other day. Expression of the Dam-POI constructs was suppressed by addition of 0.5 mM indole-3-acetic acid (IAA; Sigma, I5148). This cell line does not contain a Y chromosome.

Zebrafish

All experiments were conducted under the guidelines of the animal welfare committee of the Royal Netherlands Academy of Arts and Sciences (KNAW). Adult Tüpfel long fin (wild type) zebrafish (*Danio rerio*) were maintained and embryos raised and staged as previously described[91,92].

ChIP-seq

ChIP-seq was performed as described previously[93], with the following adaptations. Cells were harvested by trypsinization, and chemically crosslinked with fresh formaldehyde solution (1% in PBS) for 8 minutes while rotating at room temperature. Crosslinking was quenched with glycine on ice and sample was centrifuged at 500 g for 10 min at 4 °C. Pellet was then resuspended in lysis buffer for 5 min on ice and sonicated as follows: 16 cycles of 30 s on / 30 s off at max power (Bioruptor Diagenode), and centrifuged at 14,000 rpm at 4 °C for 10 min. The chromatin in supernatant was treated with RNase A for 30 min at 37 °C, and Proteinase K for 4 hours at 65 °C to reverse crosslinks, then cleared using DNA purification columns and eluted in nuclease-free water. Chromatin was incubated with antibodies (see below), after which Protein G beads (ThermoFisher #88847) were added for antibody binding. After successive washing, samples were cleared using DNA purification columns, eluted in nuclease-free water, and measured using a Qubit fluorometer. Libraries were prepared according to the Illumina TruSeq DNA LT kit and sequenced on the Illumina HiSeq 2500 following manufacturer's protocols. Up to 50 ng of immunoprecipitated chromatin was used as input for library preparation.

Antibodies used were: anti-H3K4me3 Abcam ab8580, anti-H3K9ac Abcam ab4441, anti-H3K9me3 Abcam ab8898, anti-H3K27me3 Merck Millipore 07-449, anti-H3K36me3 Active Motif 61902, anti-H4K20me1 Abcam ab9051.

DamID construct design and lentivirus production

The constructs for mintbodies, chromatin binding domains, and full-length protein constructs were fused to Dam in both possible orientations under the control of the auxin-inducible degron (AID) system[94,95] with either the hPGK or HSP promoter, and cloned into the pCCL.sin.cPPT.ΔLNGFR.Wpre lentiviral construct[96] by standard cloning procedures.

The linkers used for the triple fusion domains are, in order of appearance:
Dam; V5 linker [GKPIPNPLLGLDST]; 1st domain (e.g., chromo); GSAGSAAGSGEF; 2nd domain; linker [KESGSVSSEQLAQFRSLD]; 3rd domain. All other POIs are linked to Dam via a V5 linker, which has been commonly used in DamID constructs[82,97,98]. The Gly- and Ser-rich flexible linker, GSAGSAAGSGEF, was designed to express GFP-fusion proteins for rapid protein-folding assay[99]. The KESGSVSSEQLAQFRSLD flexible linker was previously used for the construction of a bioactive scFv[100]. For context: the Gly and Ser residues in the linker were designed to provide flexibility, whereas Glu and Lys were added to improve the solubility[101].

Bulk DamID2

hTERT-RPE1 cells were grown as described above. At 30% confluence in 6-well plates, cells were transduced with 1500 µL total volume unconcentrated lentivirus, amounts ranging between 20-1500 µL unconcentrated lentivirus (or 0.1-40 µL concentrated) in the presence of 10 µg/mL polybrene. Cells were collected for genomic DNA isolation (Wizard, Promega) 48 h after transduction. Dam methylation levels were checked by ^{m6}A-PCR as previously described[28,102] and sequenced following the DamID2 protocol[49].

Immunofluorescent staining and confocal imaging of RPE-1 cells

Viral transduction was performed as described above for bulk DamID2, with the exception that RPE-1 cells were grown on glass coverslips. Two days after transduction, cells were washed with PBS and chemically crosslinked with fresh formaldehyde solution (2% in PBS) for 10 minutes at RT, then permeabilized (with 0.5% IGEPAL® CA-630 in PBS) for 20 minutes and blocked (with 1% bovine serum albumin (BSA) in PBS) for 30 minutes. All antibody incubations were performed in final 1% BSA in PBS followed by three PBS washes at RT. Incubation with primary antibody against the endogenous histone modification as well as purified ^{m6}A-Tracer protein[52] (recognizing methylated DNA) was performed at 4 °C for 16 hours (overnight), followed by anti-GFP (against ^{m6}A-Tracer protein) incubation at RT for 1 hour, and secondary antibody incubations at RT for 1 hour. The final PBS wash was simultaneously an incubation with DAPI at 0.5 µg/mL for 2 min, followed by a

wash in MilliQ and sample mounting on glass slides using VECTASHIELD Antifade mounting medium (Vector Laboratories). Primary antibodies: anti-H3K9ac abcam ab4441 (rabbit) at 1:1000, anti-H3K9me3 abcam ab8898 (rabbit) at 1:300, anti-GFP Aves GFP-1020 (chicken) at 1:1000. Secondary antibodies: AlexaFluor anti-chicken 488 at 1:500 and anti-rabbit 647 at 1:500. Purified ^{m6}A-Tracer protein (used at 1:1000) was a kind gift from the Bas van Steensel laboratory. Imaging was performed on a Leica TCS SP8 laser scanning confocal microscope with a 63X (NA 1.40) oil-immersion objective. Images were processed in Imaris 9.3 (Bitplane) by baseline subtraction. Additional background correction was done with a 1-µM Gaussian filter for the images of Dam-CBX1 ^{m6}A-Tracer and H3K9me3 stainings.

Generation of mouse embryonic stem cell lines

The various stable clonal F1 hybrid mESC lines for the initial single cell experiments were created by lentiviral co-transduction of pCCL-EF1α-Tir1-IRES-puro and pCCL-hPGK-AID-Dam-POI constructs with a 4:1 ratio in a EF1α-Tir1-IRES-neo mother line[27], after which the cells were selected for 10 days on 0.1% gelatine coated 10- cm dishes in 60% Buffalo Rat Liver (BRL)-conditioned medium containing 0.8 µg/mL puromycin (Sigma P9620), 250 µg/mL G418 (ThermoFisher 11811031) and 0.5 mM IAA. Individual puromycin resistant colonies were handpicked and tested for the presence of the constructs by PCR using Dam-specific primers fw-ttcaacaaaagccaggatcc and rev-gacagcggtgcataaggcgg.

The clonal F1 hybrid knock-in cell lines were CRISPR targeted in a mother line carrying Tir1-Puro in the TIGRE locus[103]. For all CRISPR targeting, cells were cultured on gelatin-coated 6-wells in 60% BRL conditioned medium to 70-90% confluency and transfected with Lipofectamin3000 (Invitrogen L3000008) according to the supplier protocol with 2 µg donor vector and 1 µg Cas9/guide vector. At 24 h after transfection the cells were split to a gelatin-coated 10-cm dish and antibiotic selection of transfected cells is started 48 h after transfection. Cells were selected with 60% BRL conditioned medium containing 0.8 µg/mL puromycin for the Tir1 knock-in and 2.5 µg/mL blasticidin (Invivogen) for the AID-Dam knock-in lines. After 5-10 days of selection, individual colonies were manually picked and screened by PCR for the correct genotype.

All CRISPR knock-in lines were made in a Tir1-TIGRE mother line that was generated by co-transfection of Cas9-gRNA plasmid pX330-EN1201(Addgene plasmid #92144) and donor plasmid pEN396-pCAGGS-Tir1-V5- 2A-PuroR TIGRE (Addgene plasmid #92142)[104]. The Tir1-puro clones were screened by PCR from the CAGG promoter to Tir1 with the primers fw-cctctgctaaccatgttcatg and rev-tccttcacagctgatcagcacc, followed by PCR from the polyA to the TIGRE locus with primers fw-gggaagagaatagcaggcatgct and rev-accagccacttcaaagtggtacc. The Tir1 expression is further confirmed by Western blot using a V5 antibody (Invitrogen R960-25).

A knock-in of AID-Dam in the N-terminus of the RING1B locus was made by co-transfection of a donor vector carrying the blasticidin-p2A-HA-mAID-Dam cassette flanked by 2 500-bp homology arms of the endogenous RING1B locus (pHom-BSD-p2A-HA-mAID-Dam) and p225a-RING1B spCas9-gRNA vector (sgRNA: 5'gctttttattcctagaaatgtctc3') as described above. Picked clones were screened for correct integration by PCR with primers from Dam to the RING1B locus outside the targeting construct; fw-gaacaacaagcgcatctggc and rev-tcctcccctaacctgcttttgg. Presence of the RING1B wildtype allele was checked by PCR with primers fw-tcctcccctaacctgcttttgg and rev-gccttgcctgcttggtttg. The H3K27me3 mintbody coupled to ER-mAID-Dam was knocked into the Rosa26 locus by co-transfection of pHom-ER-mAID-V5-Dam-scFv_H3K27me3-P2A-BSD- Hom donor vector and p225a-Rosa26 spCas9-RNA vector (sgRNA: gtccagtctttctagaagatgggc) as described above.
Picked clones were screened for correct integration by PCR from a sequence adjacent to the Rosa homology arm to the Rosa26 locus with primers fw- gaactccatatatgggctatg and rev-cttggtgcgtttgcgggga. The untethered mAID-Dam was knocked into the Rosa26 locus by co-transfection with the pHom-ER-mAID-V5- Dam-P2A-BSD-Hom donor vector and p225a-Rosa26 spCas9-RNA vector (sgRNA: gtccagtctttctagaagatgggc) as described above. Picked clones were screened for correct integration by PCR with the same primers as for the Dam-H3K27me3 mintbody knock-in line.

All clones with correct integrations were furthermore screened for their level of induction upon IAA removal by ^{m6}A-PCR evaluated by gel electrophoresis[28,102], followed by DamID2 sequencing in bulk[49], to select the clone with a correct karyotype and the best signal-to-noise ratio of enrichment over expected regions or chromatin domains. Finally, the best 3-4 clones were selected for testing of IAA removal timing in single cells by DamID2.

Mouse embryonic stem cell culture and induction of Dam-fusion proteins

When plated for targeting or genomics experiments, cells were passaged at least 2 times in feeder-free conditions, on plates coated with 0.1% gelatin, grown in 60% BRL-conditioned medium, defined as follows and containing 1 mM IAA: 40% CM+/+ medium and 60% of CM+/+ medium conditioned on BRL cells. For timed induction of the constructs the IAA was washed out at different clone-specific times before single-cell sorting.

Embryoid body differentiation and induction of Dam-fusion proteins

For EB differentiation, the stable knock-in F1ES lines were cultured for 2 weeks on plates coated with 0.1% gelatin, grown in 2i+LIF ES cell culture medium defined as follows: 48% DMEM/F12 (Gibco) and 48% Neurobasal medium (Gibco), supplemented with 1x N2 (Gibco), 1x B27 supplement + vitamin A (Gibco), 1x non-essential amino acids, 1% FBS, 1% Pen/Strep, 0.1 mM β-mercaptoethanol, 1 μM PD0325901 (Axon Medchem, PZ0162-5MG), 3 μM CHIR99021 (Tocris, SML1046-5MG), 1000 U/mL

ESGRO mLIF. EB differentiation was performed according to ATCC protocol. On day 1 of differentiation, 2x10^6 cells were grown in suspension on a non-coated bacterial 10-cm dish with 15 mL CM +/- (with β-mercaptoethanol, without LIF) and 0.5 mM IAA. On day 2, half the cell suspension was divided over five non-coated bacterial 10-cm dishes each containing 15mL CM+/- medium and 0.5 mM IAA. Plates were refreshed every other day. EBs were harvested at day 7, 10, and 14. Two days before single-cell sorting, the EBs were grown in CM+/- medium containing 1 mM IAA, and induced as follows: 6 h without IAA (RING1B); 20 h without IAA and 7 h with 1 μM 4OHT (Sigma SML1666) (Dam-H3K27me3-mintbody); 7 h without IAA and 4 h with 1 μM 4OHT (untethered Dam). The EBs were evaluated by brightfield microscopy and hand-picked for further handling (see below).

FACS for single-cell experiments

FACS was performed on BD FACSJazz or BD FACSInflux Cell Sorter systems with BD Sortware. mESCs and EBs were harvested by trypsinization, centrifuged at 300 g, resuspended in medium containing 20 μg/mL Hoechst 34580 (Sigma 63493) per $1x10^6$ cells and incubated for 45 minutes at 37°C. Prior to sorting, cells were passed through a 40-μm cell strainer. Propidium iodide (1 μg/mL) was used as a live/dead discriminant. Single cells were gated on forward and side scatters and Hoechst cell cycle profiles. Index information was recorded for all sorts. One cell per well was sorted into 384-well hard-shell plates (Biorad, HSP3801) containing 5 μL of filtered mineral oil (Sigma #69794) and 50 nL of 0.5 μM barcoded CEL-Seq2 primer[27,49]. In the EB experiment, the knock-in mESC lines were cultured alongside on 2i+LIF medium and included as a reference at each timepoint.

Single-cell Dam&T-seq

The scDam&T-seq protocol was performed as previously described in detail[49], with the adaptation that all volumes were halved to reduce costs. Liquid reagent dispension steps were performed on a Nanodrop II robot (Innovadyne Technologies / BioNex). Addition of barcoded adapters was done with a mosquito LV (SPT Labtech). In short, after FACS, 50 nL per well of lysis mix (0.07% IGEPAL, 1 mM dNTPs, 1:50,000 ERCC RNA spike-in mix (Ambion, 4456740)) was added, followed by incubation at 65 °C for 5 min. 100 nL of reverse transcription mix (1× First Strand Buffer and 10 mM DTT (Invitrogen, 18064-014), 2 U RNaseOUT Recombinant Ribonuclease Inhibitor (Invitrogen, 10777019), 10 U SuperscriptII (Invitrogen, 18064-014)) was added, followed by incubation at 42 °C for 2 h, 4 °C for 5 min and 70 °C for 10 min. Next, 885 nL of second strand synthesis mix (1× second strand buffer (Invitrogen, 10812014), 192 μM dNTPs, 0.006 U *E. coli* DNA ligase (Invitrogen, 18052019), 0.013 U RNase H (Invitrogen, 18021071), 0.26 U *E. coli* DNA polymerase (Invitrogen)) was added, followed by incubation at 16 °C for 2 h. 250 nL of protease mix was added (1× NEB CutSmart buffer, 1.0 mg/mL Proteinase K (Roche, 000000003115836001)), followed by incubation at 50 °C for 10 h and 80 °C for 20 min. Next, 115 nL of DpnI

mix (1× NEB CutSmart buffer, 0.1 U NEB DpnI) was added, followed by incubation at 37 °C for 6 h and 80 °C for 20 min. Finally, 50 nL of 0.5uM DamID2 adapters were dispensed (final concentrations 25 nM), followed by 400 nL of ligation mix (1× T4 Ligase buffer (Roche, 10799009001), 0.13U T4 Ligase (Roche, 10799009001)) and incubation at 16°C for 16 hrand 65°C for 10 min. Contents of all wells were pooled and the aqueous phase was recovered by centrifugation and transfer to clean tubes. Samples were purified by incubation for 10 min with 0.8 volumes magnetic beads (CleanNA, CPCR-0050) diluted 1:7 with bead binding buffer (20% PEG8000, 2.5 M NaCl), washed twice with 80% ethanol and resuspended in 8 µL of nuclease-free water before in vitro transcription at 37 °C for 14 h using the MEGAScript T7 kit (Invitrogen, AM1334). Library preparation was done as described in the CEL-Seq2 protocol with minor adjustments[105]. Amplified RNA (aRNA) was purified with 0.8 volumes beads as described above, and resuspended in 20 µL of nuclease-free water, and fragmented at 94 °C for 90 sec with the addition of 0.25 volumes fragmentation buffer. Fragmentation was stopped by addition of 0.1 volumes of 0.5 M EDTA pH 8 and quenched on ice. Fragmented aRNA was purified with beads as described above, and resuspended in 12 µL of nuclease-free water. Thereafter, library preparation was done as previously described[105] using up to 7 µL or approximately 150 ng of aRNA, and 8-10 PCR cycles depending on input material. Libraries were sequenced on the Illumina NextSeq500 (75-bp reads) or NextSeq2000 (100-bp reads) platform.

Collection of zebrafish samples and FACS

Tüpfel long fin (wild type) pairs were set up and the following morning, approximately 1 nL of 1 ng/µL *Dam- Mphosph8* mRNA or 0.5 ng/µL *Dam-Gfp* mRNA was injected into the yolk at the 1 cell stage. Embryos were slowed down overnight at 23°C and the following morning all embryos were manually dechorionated. At 15- somite stage, embryos were transferred to 2-mL Eppendorf tubes and digested with 0.1% Collagenase type II from Cl. Histolyticum (Gibco) in Hanks Balanced Salt Solution without Mg^{2+}/Ca^{2+} (Thermofisher) for 20-30 mins at 32°C with constant shaking. Once embryos were noticeably digested, cell solution was spun at 2000 g for 5 min at room temperature and the supernatant was removed. Cell pellet was resuspended with TrypLE Express (Thermofisher) and digested for 10 min at 32°C with constant shaking. Cell solution was inactivated with 10% Fetal Bovine Serum (Thermofisher) in Hanks Balanced Salt Solution without Mg^{2+}/Ca^{2+} and filtered through a 70-µm cell strainer (Greiner Bio-One). Cells were pelleted at 2000g 5min room temperature and washed twice with 10% Fetal Bovine Serum (Thermofisher) in Hanks Balanced Salt Solution without Mg^{2+}/Ca^{2+}. Hoechst 34580 at a final concentration of 16.8 µg/mL was added to the cell solution and incubated for 30 mins at 28°C in the dark. Solution was then filtered through a 40-µm cell strainer (Greiner Bio-One), and propidium iodide was added at a final concentration of 5 µL/mL. FACS was performed on BD FACSInflux as described above, retaining only cells in G2/M phase based on Hoechst DNA content. Plates were processed for scDam&T-seq as described above.

Immunofluorescent staining and confocal imaging of zebrafish embryos

Embryos at 15-somite stage were fixed in 4% PFA (Sigma) for 2 h at RT, followed by washes in PBS. Embryos were then washed three times in 4% sucrose/PBS and allowed to equilibrate in 30% sucrose/PBS at 4°C for 3-5 h. Embryos were suspended in Tissue Freezing Medium (Leica) orientated in the sagittal plane and frozen with dry ice. Blocks were sectioned at 8 µm and slides were rehydrated in PBS, treated with -20°C pre-cooled acetone for 7 min at -20°C, washed three times with PBS and digested with Proteinase K (Promega) at a final concentration of 10 µg/mL for 3 min, washed 1x PBS and incubated in blocking buffer (10% Fetal Bovine Serum, 1% DMSO, 0.1% Tween20 in PBS) for 30 min. Primary antibody was diluted in blocking buffer and slides incubated overnight at 4°C. Slides were washed the following day and incubated with the appropriate AlexaFluor secondary antibodies (1:500), DAPI (0.5 µg/mL) and Phalloidin-TRITC (1:200) diluted in blocking buffer for 1 h at RT. Slides were washed, covered with glass coverslips with ProLong Gold Antifade Mountant (Thermofisher) and imaged at 63X with a LSM900 confocal with AiryScan2 (Zeiss). Images were viewed and processed in Imaris 9.3 (Bitplane) and Adobe Creative Cloud (Adobe). Primary antibody: anti-H3K9me3 abcam ab8898 at 1:500[106].

Processing DamID and scDam&T-seq data

Data generated by the DamID and scDam&T-seq protocols was largely processed with the workflow and scripts described in[49]. The procedure is described in short below.

Demultiplexing

All reads are demultiplexed based on the barcode present at the start of R1 using a reference list of barcodes. In the case of scDam&T-seq data, the reference barcodes contain both DamID-specific and CEL-Seq2-specific barcodes and zero mismatches between the observed barcode and reference are allowed. In the case of the population DamID data, the reference barcodes only contain DamID-specific barcodes and one mismatch is allowed. The UMI information, also present at the start of R1, is appended to the read name.

DamID data processing

DamID reads are aligned using bowtie2 (v. 2.3.3.1)[107] with the following parameters: "--seed 42 --very- sensitive -N 1". For human samples, the hg19 reference genome is used; for mouse samples, the mm10 reference genome; and for zebrafish samples the GRCz11 reference genome. The resulting alignments are then converted to UMI-unique GATC counts by matching each alignment to known strand-specific GATC positions in the reference genome. Any reads that do not align to a known GATC position or have a mapping quality smaller than 10 are removed. In the case of bulk DamID samples, up to 64 unique UMIs are allowed per GATC position, while up to 4 unique UMIs are allowed for single-cell samples to account for the maximum number of alleles in G2. Finally, counts are binned at the desired resolution.

CEL-Seq2 data processing

CEL-Seq2 reads are aligned using tophat2 (v. 2.1.1)[108] with the following parameters: "--segment-length 22 --read-mismatches 4 --read-edit-dist 4 --min-anchor 6 --min-intron-length 25 --max-intron-length 25000 --no- novel-juncs --no-novel-indels --no-coverage-search --b2-very-sensitive --b2-N 1 --b2-gbar 200". For mouse samples, the mm10 reference genome and the GRCm38 (v. 89) transcript models are used. For zebrafish samples, the GRCz11 reference genome and the adjusted transcript models published by the Lawson lab[109] are used. Alignments are subsequently converted to transcript counts per gene with custom scripts that assign reads to genes similar to HTSeq's[110] htseq-count with mode "intersection_strict".

Processing of ChIP-seq data

External ChIP-seq datasets were downloaded from the NCBI GEO repository and the ENCODE database[56]. The external ChIP-seq data used in this manuscript consists of: H3K9ac ChIP-seq in mESC (ENCSR000CGP), H3K27me3 ChIP-seq in mESC (ENCSR059MBO), and H3K9me3 ChIP-seq in 6-hpf zebrafish embryos[74] (GSE113086). Internal and external ChIP-seq data were processed in an identical manner. First reads were aligned using bowtie2 (v. 2.3.3.1) with the following parameters: "--seed 42 --very-sensitive -N 1". Indexes for the alignments were then generated using "samtools index" and genome coverage tracks were computed using the "bamCoverage" utility from DeepTools (v. 3.3.2)[111] with the following parameters: "--ignoreDuplicates --minMappingQuality 10". For marks that exist in broad domains in the genome, domains were called using MUSIC[112] according to the suggested workflow. For marks that form narrow peaks in the genome, peaks were called using MACS2 (v. 2.1.1.20160309)[113] using the "macs2 callpeak" utility with the following parameters: "-q 0.05".

Computing the Information Content (IC) of DamID samples

The Information Content (IC) of a DamID sample is a measure of how much structure is in the detected methylation signal. It is essentially an adaptation of the RNA-seq normalization strategy called PoissonSeq[114]. Its goal is to compare the obtained signal to a background signal (the density of mappable GATCs), identify regions where the signal is similar to background, and finally compare the amount of total signal (i.e. total GATC counts) to the total signal in background regions. The IC is the ratio of total signal over background signal and can be used to filter out samples that contain little structure in their data. The code used to compute the IC is available online (https://github.com/KindLab/EpiDamID2022) and the procedure is explained below.

As an input, we use the sample counts binned at 100-kb intervals, smoothened with a 250-kb Gaussian kernel. The large bin size and smoothing are necessary when working with single-cell samples that have very sparse and peaky data and would otherwise be difficult to match to the background signal. As a control, we use the number of

mappable GATCs in the same 100-kb bins, similarly smoothened. We subsequently remove all genomic bins that do not have any observed counts in the sample. Our starting data is then *X*, a matrix with size (*n*, *k*), where *n* is the number of genomic bins and *k* is the number of samples. Since we are comparing one experimental sample with the control, *k* is always 2. *Xij* denotes the number of counts observed in the *i*th bin of the *j*th sample. We first compute the expected number of counts for each *Xij* based on the marginal probabilities of observing counts in each bin and in each sample:

$$d = \sum_{i=1}^{n} \sum_{j=1}^{k} X_{ij}$$

$$p = \sum_{j=1}^{k} Xj/d = (p_1 \ldots p_n)^T$$

$$q = \sum_{i=1}^{n} Xi/d = (q_1, q_2)$$

$$E = d\,(p \cdot q)$$

Where *d* is the total sum of *Xij*; *pi* is the marginal probability of observing counts in bin *i*; *qj* is the marginal probability of observing counts in sample *j*; and *E* is the matrix of size (*n*, *k*) where entry *Eij* is the expected number of counts in bin *i* for sample *j*, computed as *piqjd*.

We subsequently compute the goodness of fit of our predictions compared to the actual counts per bin:

$$g = \sum_{j=1}^{k} \frac{X_j - E_j}{E_j}$$

Where *gi* is the measure of how well the predictions of *Ei* match the observed counts in *Xi* in bin *i*. The better the prediction, the closer *gi* is to zero, indicating that the signal of the experimental sample closely resembles the background in bin *i*. Next, an iterative process is performed where in each step a subset of the original bins is chosen that exclude bins with extreme values of *g*. Specifically, all bins with a goodness of fit in the top and bottom 5th percentiles are excluded to progressively move towards a stable set of bins where the sample resembles the background. After each iteration, the chosen bins are compared to the previous set of bins and when this has stabilized, or when the maximum number of iterations is reached, the procedure stops. In practice, convergence is usually reached after only a couple of iterations. The IC is then computed for the experimental sample as the ratio of its summed total counts to the sum of counts observed in the final subset of bins.

Population DamID data filtering and analyses

The population DamID samples were filtered based on a depth threshold of 300,000 UMI-unique GATC counts and an IC of at least 1.1. Per Dam-construct, the best samples based on the IC were maintained. Samples were normalized for the total number of

counts using reads per kilobase per million (RPKM). Normalization for Dam controls was performed by adding a pseudo count of 1, taking the per bin fold-change with Dam, and performing a log2-transformation, resulting in log2 observed-over-expected (log2OE) values. The UMAP presented in Fig. 3.1B was computed by performing principal component analysis (PCA) on the RPKM-normalized samples (20-kb bins) and using the top components for UMAP computation in python with custom scripts. For the correlations presented in Fig. 3.1c and S3.1c, the RPKM-normalized DamID values were normalized for the density of mappable GATCs and log-transformed. The Spearman's rank correlation was then computed with the input-normalized ChIP-seq values of the various marks.

Resolution analysis on RPE-1 samples

To evaluate the resolution of EpiDamID signal compared to ChIP-seq, we wanted to determine the spread of the signal around regions of known enrichment. To this end, we used ChIP-seq peaks for H3K9ac and H3K4me3, and domains for H3K27me3 and H3K9me3. We computed the average ChIP-seq signal and DamID signal around these regions, using a resolution (i.e. bin size) of 200 bp. The resulting signal was mildly smoothed to get a better representation of the trends. For each sample, we then determined the distance over which the signal measured at the reference point decayed to 50% relative to the background. As a reference point, we chose the center of H3K9ac and H3K4me3 peaks, or the boundary of H3K27me3 and H3K9me3 domains. The spread of the DamID signal can then be determined as the increase in this distance relative to the corresponding ChIP-seq sample.

Single-cell DamID data filtering and analyses

Filtering and normalizing scDamID data

Single-cell DamID samples were filtered based on a depth and an IC threshold. For the mouse samples, these thresholds were 3,000 unique GATCs and an IC within the range of 1.5 to 7 (the upper threshold removes samples with very sparse profiles); for zebrafish, these thresholds were 1,000 unique GATCs and an IC within the range of 1.2 to 7. For the zebrafish samples, chromosome 4 was excluded when determining depth and IC (and in all downstream analyses) since the reference assembly of this chromosome is poor and alignments unreliable. The quality of scDam&T-seq samples is determined separately for the DamID readout and the CEL-Seq2 readout. To preserve as much of the data as possible, we used all samples passing DamID thresholds for analyses that relied exclusively on the DamID readout. Wherever single-cell data was used, samples were normalized for their total number of GATCs, scaled by a factor 10,000, and log-transformed with a pseudo- count of 1, equivalent to the normalizations customarily performed for single-cell RNA-seq samples. To generate in silico populations based on single-cell samples, the binned UMI-unique counts of all single-cells were combined and normalization was performed equivalent to population DamID samples.

scDamID UMAPs

The UMAPs presented in Fig. 3.2a, Fig. 3.3c and Fig. 3.5c were computed by performing PCA on the depth-normalized single-cell samples and using the top components for UMAP computation. Since in EBs inactivation of chromosome X can coincides with a strong enrichment of H3K27me3/RING1B on that chromosome, we depth-normalized these samples using the total number of GATCs on somatic chromosomes.

For the zebrafish samples, chromosome 4 was completely excluded from the analysis. For the mouse UMAPs, the single-cell data were binned at a resolution of 10-kb intervals, while for the zebrafish UMAPs, the resolution was 100 kb. Notably, when the first principal components showed a strong correlation to sample depth, it was excluded.

Single-cell count enrichment

Fig. 3.2b-d show the enrichment of counts in ChIP-seq domains for all single-cell mESC samples; Fig. S3.5f shows the enrichment of counts for all MPHOSPH8 zebrafish samples. The count enrichment is equivalent to the more well-known Fraction Reads in Peaks (FRiP) metric, but has been normalized for the expected fraction of counts within the domains based on the total number of mappable GATCs covered by these domains. In other words, if the domains cover 50% of the mappable GATCs in the genome and we observe that 70% of a sample's counts fall within these domains, the count enrichment is 0.7 / 0.5 = 1.4.

Single-cell CEL-Seq2 data filtering and analyses

Filtering CEL-Seq2 data

Single-cell data sets were evaluated with respect to the number of unique transcripts, percentage mitochondrial reads, percentage ERCC-derived transcripts and the percentage of reads coming from unannoted gene models (starting with "AC" or "Gm") and appropriate thresholds were chosen. For the EB data, the used thresholds were ≥1,000 UMI-unique transcripts, <7.5% mitochondrial transcripts, <1% ERCC-derived transcripts, and <5% transcripts derived from unannotated gene models. In addition, a small group of cells (29/6,554 ≈ 0.4%) from different time points, which formed a cluster that could not be annotated and was characterized by high expression of ribosomal genes, was removed from further analyses. For the zebrafish data, the used thresholds were ≥1,000 UMI-unique transcripts and <5% ERCC-derived transcripts. Only genes observed in at least 5 samples across the entire dataset were maintained in further analyses. The quality of scDam&T-seq samples is determined separately for the DamID readout and the CEL-Seq2 readout. To preserve as much of the data as possible, we used all samples passing CEL-Seq2 thresholds (independent of DamID quality) for transcriptome-based analyses.

Analysis of CEL-Seq2 data with Seurat and Harmony

Single-cell transcription data was processed using Seurat (v3)[115]. First, samples were processed using the "NormalizeData", "FindVariableFeatures", "ScaleData", and "RunPCA" commands with default parameters. Subsequently, batch effects relating to processing batch and plate were removed using Harmony[116] using the "RunHarmony" command, using a theta=2 for the batch variable and theta=1 for the plate variable. Clustering and dimensionality reduction were subsequently performed with the "FindNeighbors", "FindClusters" and "RunUMAP" commands. Differentially expressed genes per cluster were found using the "FindAllMarkers" command.

Integration with external single-cell datasets

The EB data was integrated with part of the single-cell mouse embryo atlas published by[65] and with the transcription data from the scNMT-seq EB dataset published by[67]. In the case of the mouse embryo atlas, the data was loaded directly into R via the provided R package "MouseGastrulationData". One data set per time point was included (datasets 18, 14, 19, 16, 17, corresponding to embryonic stages E6.5, E7.0, E7.5, E8.0, E8.5, respectively). In the case of the scNMT-seq dataset, the transcript count tables were downloaded from the repository provided in the publication. Only cells derived fom wild type embryos were included. The external data and our own data was integrated using the SCTransform[117] and the anchor-based intregration[115] functionalities from Seurat. First, all data was normalized per batch using the "SCTransform" command. Data sets were then integrated using the "SelectIntegrationFeatures", "PrepSCTIntegration", "FindIntegrationAnchors", and "IntegrateData", as per Seurat documentation. To assign scNMT-seq samples to the previously determined EB clusters, we used Seurat's "TransferData" command.

SCENIC

We used SCENIC[72] on the command line according to the documentation provided for the python-based scalable version of the tool (pySCENIC)[71]. Specifically, we ran "pyscenic grn" with the parameters "--method grnboost2"; "pyscenic ctx" with the parameters "--all_modules"; and "pyscenic aucell" with the default parameters. We used the transcription factor annotation and the transcription factor motifs (10 kb +/- of the TSS) provided with SCENIC. This yielded 414 activating regulons. We subsequently filtered regulons based on the expression of the regulon as a whole (at least 50% of cells having an AUCell score > 0 within at least one Seurat cluster) and based on the expression of the regulon transcription factor (detected in at least 5% of cells in at least one cluster) to retain only high confidence regulons. This resulted in 285 remaining activating regulons. However, repeating all analyses with the unfiltered set of regulons yielded the same trends and relationships.

Linear Discriminant Analysis (LDA) classifier to assign samples to transcriptional clusters based on DamID signal

In both the EB results and the zebrafish results, we noticed that there was a substantial number of cells that passed DamID thresholds, but that had a poor CEL-Seq2 readout. Since most of our analyses rely on the separation of cells in transcriptional clusters (i.e. cell types) and cells with a poor CEL-Seq2 readout cannot be included in the clustering, these cells cannot be used in downstream DamID-based analyses. However, we noticed that the separation of different cell types was recapitulated to a considerable extent in low- dimensionality representations of the DamID readout (see the DamID-based UMAPs in Fig. 3.2a and Fig. 3.3d). Since cell type information is captured in the DamID readout, we reasoned that a classifier could be trained based on cells with both good DamID and CEL-Seq2 readouts to assign cells with a poor CEL-Seq2 readout to transcriptional clusters based on their DamID readout.

To this end, we implemented a Linear Discriminant Analysis (LDA) classifier as described below. In addition, the code is available online (https://github.com/KindLab/EpiDamID2022).

Data input and preprocessing

As in input for the classifier, we used the binned DamID data of all samples passing DamID thresholds and the transcriptional cluster labels of these samples (samples with a poor CEL-Seq2 readout had the label "unknown"). The DamID data was depth-normalized (as described above) and genomic bins that contained fewer than 1 mappable GATC motif per kb were excluded, resulting in a matrix of size N x M, where N is the number of samples and M is the number of remaining genomic bins. For the EB data, a bin size of 10 kb was used, while a bin size of 100 kb was used for the zebrafish data. Subsequently, the pairwise correlation was computed between all samples, resulting in a correlation matrix of size N x N. This transformation had two reasons: First, it served as a dimensionality reduction, since $N << M$. Second, it resulted in a data type that effectively describes the similarity of a sample with all other samples, including samples without a cluster label. Consequently, during the training phase, the classifier can indirectly use the information of these unlabeled samples to learn about the overall data structure. We found that using the correlation matrix (N x N) as an input for the classifier yielded much better results than using the original matrix (N x M).

To train the LDA classifier, we used two thirds (~66%) of all samples with cluster labels (i.e. with a good CEL-Seq2 readout). Since the number of cells per cluster varied extensively, we randomly selected two thirds of the samples per cluster and thereby ensured that all clusters were represented in both training and testing. The training data thus consisted of the correlation matrix of size N_{train} x N and a list of sample labels of size N_{train}, where N_{train} is the number of samples used for training.

Consequently, we retained one third (~33%) of labelled samples to test the performance of the LDA classifier, consisting of the correlation matrix of size N_{test} x N and a list of sample labels of size N_{test}, where N_{test} is the number of samples used for testing. In summary, this split the samples into three groups: one group for training, one group for testing, and the group of unlabeled samples.

Training the classifier

For the implementation of the LDA classifier, we used the "LinearDiscriminantAnalysis" function provided in the Python (v. 3.8.10) scikit-learn toolkit (v. 0.24.2). The number of components was set to the number of transcriptional clusters minus one and the LDA classifier was trained using the training samples.

Testing the performance

To test the performance, the trained LDA classifier was used to predict the labels of the training set of samples. Predictions with a probability larger than 0.5 were maintained, while predictions with a lower probability were discarded (and the corresponding cells were thus not labelled). The predicted labels were subsequently compared to the known labels (Fig. S3.3e, S3.5c). In general, we found a very good performance for clusters with many cells, while the performance tended to be lower for clusters with few cells. This is as expected, since the number of samples for these clusters was also very low during training.

Predicting cluster labels for unlabeled samples

After establishing that the performance was satisfactory, the LDA was retrained, this time using all labelled samples. The actual performance on the unlabeled data is likely higher than the performance on the test data, since the number of samples used for the final training is notably higher. Finally, the cluster labels were predicted for the unlabeled samples. Once again, only predictions with a probability higher than 0.5 were maintained. Fig. S3.3d shows the number of EB samples that were attributed to each cluster using the LDA classifier, as well as the number of samples that could not be attributed ("unassigned").

Defining PRC targets

First, we identified for each gene the region of 5 kb upstream and 3 kb downstream of the TSS. Only protein- coding genes and genes for non-coding RNA were considered. When the TSS domains of two genes overlapped, they were merged if the overlap was >4 kb, otherwise the two domains were split in the middle of the overlap. This resulted in 30,356 domains covering a total of 35,814 genes. Subsequently, for all single- cells, the number of observed GATC counts within each domain was determined. *In silico* populations per transcriptional cluster were generated by combining the counts of all cells belonging to each cluster per DamID construct. The *in silico* population counts were subsequently RPKM-normalized, using the total number of GATC counts on

the somatic chromosomes of the combined single-cell samples as the depth (i.e. also counts outside the domains). Normalization for Dam controls was performed for the H3K27me3 and RING1B data per transcriptional cluster by adding a pseudo count of 1, taking the fold-change with Dam, and performing a log2-transformation, resulting in log2 observed-over-expected (log2OE) values. The correlation of the resulting H3K27me3 and RING1B values per cluster is shown in Fig. S3.3f. We subsequently determined PRC targets as those genes that showed H3K27me3 and RING1B log2OE values >0.35 in at least one cluster. PRC targets were defined based on the *in silico* population of the H3K27me3 and RING1B data of the mESCs (Fig. 3.2) and the EB clusters, excluding cluster 7. Cluster 7 was excluded, because it consisted of relatively few cells and the combined data was consequently sparse.

Comparing EpiDamID and scNMT-seq data at transcription start sites

We downloaded the tables of single-cell CpG methylation values at regions +/-2 kb of gene TSS from the repository provided in the scNMT-seq publication[67]. We subsequently averaged the CpG methylation scores across cells per cluster to gain an average CpG methylation for all genes per cluster. This could be done for four out of eight transcriptional clusters to which sufficient scNMT-seq samples were attributed (cluster 3: 31 cells; cluster 5: 21 cells; cluster 1: 37 cells; cluster 4: 43 cells). We subsequently could integrate the CpG methylation scores with our own H3K27me3 and RING1B DamID data for all genes, for which the enrichment scores were computed as described in the previous section. The subsequent analyses were performed on genes that were represented in both datasets.

ChromHMM of zebrafish in silico populations

In order to determine regions that were characterized by H3K9me3-enrichment in specific (sets of) cell types in the zebrafish embryo, we made use of ChromHMM (v. 1.22)[79,80]. As input, we used the *in silico* H3K9me3 signal (log2OE) of all clusters that had at least 30 cells passing DamID thresholds for both Dam and MPHOSPH8 (clusters 0-11). The genome-wide signal at a resolution of 50 kb was used and the values were binarized based on a threshold of log2OE > 0.35. Bins that had fewer than 1 mappable GATC per kb were given a value of 2, indicating that the data was missing. As in all other analysis, chromosome 4 was excluded. The binarized values of clusters 0-11 were provided as input for the ChromHMM and the results were computed using the "LearnModel" function using the following parameters: -b 50000 -s 1 -pseudo. The number of ChromHMM states was varied from 2 to 10 and for each result the differences between the states (based on the emission probabilities) were inspected. We found that a ChromHMM model with 5 states was optimal, since this yielded the most diverse states and increasing the number of states just added redundant states with similar emission probabilities.

Repeat enrichment in ChromHMM states

The RepeatMasker repeat annotations for GRCz11 were downloaded from the UCSC Genome Browser website (https://genome.ucsc.edu/). The enrichment of repeats within each ChromHMM state was computed either for repeat classes as a whole (Fig. S3.6a) or for individual types of repeats (Fig. 3.5i and S3.6c). To compute the enrichment of a repeat class/type in a ChromHMM state, the fraction of repeats belonging to that class/type that fell within the state was computed and normalized for the fraction of the genome covered by that state. In other words, if we observe that 70% of a certain repeat falls within state B and state B covers 7% of the genome, then the repeat enrichment is 0.7 / 0.07 = 10.

GO term and PANTHER protein class enrichment analysis

GO term and PANTHER[118] protein class enrichment analyses were performed via de Gene Ontology Consortium website (http://geneontology.org/). For Fig. S3.4e, the list of PRC-regulated TFs was used as a query and the list of all TFs as a reference to determine enriched Biological Process GO terms. Only the top 10 most significant terms are shown. For Fig. S3.5g, the list of genes in ChromHMM state A1 or B was used as a query and the list of genes in all ChromHMM states as a reference to determine enriched PANTHER protein classes. All hits are shown.

Quantification and statistical analysis

The number of n samples included in analyses is provided within each figure and/or accompanying figure legend. Statistical p values are associated with the significance test as described in the figure legends. The boxes of boxplots indicate the quartiles of the dataset, the middle shows the median, and the error bars of indicate 1.5 times the inter-quartile range.

Data and code availability

All sequencing data generated in this manuscript are deposited on the NCBI Gene Expression Omnibus (GEO) portal and are publicly available as of the data of publication under accession number GSE184036 (see Key Resource Table for further details). Imaging data are publicly available on Mendeley Data
(DOI: 10.17632/sp7hsw68c4.1).

Key scripts are available at https://github.com/KindLab/EpiDamID2022.
Any additional information required to reanalyze the data reported in this paper is available from the Lead Contact upon request.

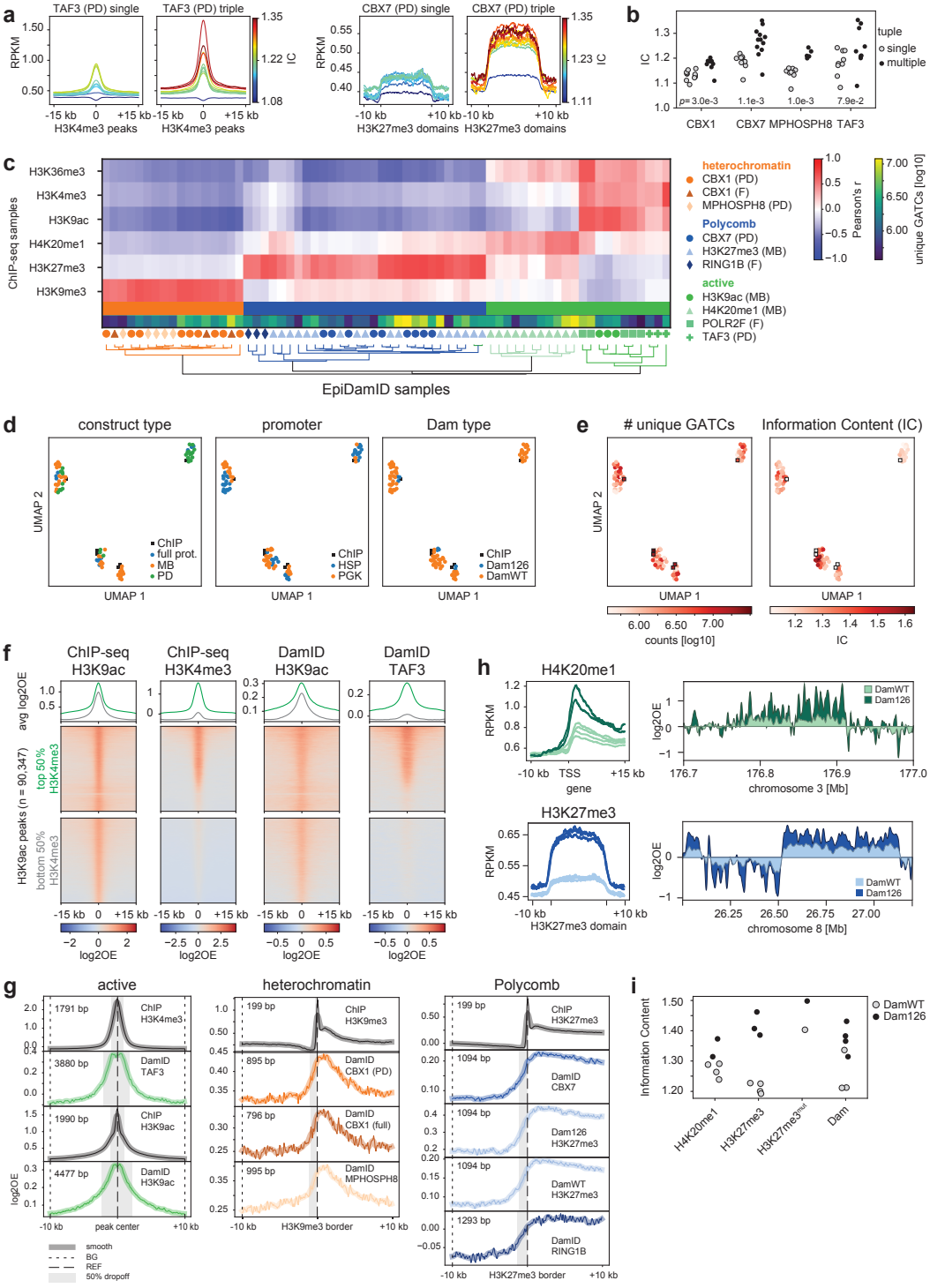

a
TAF3 (PD) single
TAF3 (PD) triple
CBX7 (PD) single
CBX7 (PD) triple
RPKM
IC
H3K4me3 peaks
H3K27me3 domains
b
tuple
single
multiple
p= 3.0e-3
1.1e-3
1.0e-3
7.9e-2
CBX1
CBX7
MPHOSPH8
TAF3
c
ChIP-seq samples
H3K36me3
H3K4me3
H3K9ac
H4K20me1
H3K27me3
H3K9me3
heterochromatin
CBX1 (PD)
CBX1 (F)
MPHOSPH8 (PD)
Polycomb
CBX7 (PD)
H3K27me3 (MB)
RING1B (F)
active
H3K9ac (MB)
H4K20me1 (MB)
POLR2F (F)
TAF3 (PD)
Pearson's r
unique GATCs [log10]
EpiDamID samples
d
construct type
promoter
Dam type
UMAP 1
UMAP 2
ChIP
full prot.
MB
PD
HSP
PGK
Dam126
DamWT
e
unique GATCs
Information Content (IC)
counts [log10]
IC
f
ChIP-seq H3K9ac
ChIP-seq H3K4me3
DamID H3K9ac
DamID TAF3
avg log2OE
top 50% H3K4me3
bottom 50% H3K4me3
H3K9ac peaks (n = 90,347)
log2OE
h
H4K20me1
gene
H3K27me3
H3K27me3 domain
chromosome 3 [Mb]
chromosome 8 [Mb]
DamWT
Dam126
g
active
heterochromatin
Polycomb
ChIP H3K4me3
DamID TAF3
ChIP H3K9ac
DamID H3K9ac
ChIP H3K9me3
DamID CBX1 (PD)
DamID CBX1 (full)
DamID MPHOSPH8
ChIP H3K27me3
DamID CBX7
Dam126 H3K27me3
DamWT H3K27me3
DamID RING1B
peak center
H3K9me3 border
H3K27me3 border
smooth
BG
REF
50% dropoff
i
Information Content
H4K20me1
H3K27me3
H3K27me3rui
Dam

Fig. S3.1. Technical validation of EpiDamID data

a, Average enrichment over genomic regions of interest for TAF3 and CBX7 DamID. Left: data generated by fusing Dam to a single protein domain; Right: data generated by fusing Dam to a trimer of the same protein domain. Sample lines are colored by their Information Content (IC). **b**, Strip plot of samples comparing the IC of single (grey) and multiple (black) targeting domains. Per construct, the significance was tested with a two-sided Mann-Whitney U test. **c**, Clustered heatmap showing the correlation between ChIP-seq and Dam-normalized DamID. Correlations were computed using Pearson's correlation. Samples are labeled by their targeting domain (colored shapes) and number of unique GATC counts. **d**, UMAPs of samples, colored by construct properties. **e**, UMAPs of samples, colored by the number of unique GATC counts and IC. **f**, ChIP-seq and DamID enrichment at H3K9ac peaks (center +/- 15 kb), split into two categories according to ChIP-seq H3K4me3 occupancy (highest and lowest 50%). The heatmaps show the enrichment for each peak region, while the line plots on top show the average enrichment per H3K4me3 category. **g**, Signal resolution analysis. The plots show ChIP-seq and DamID enrichment at genomic regions of interest +/- 10 kb. Left panel shows active marks; signal is centered around ChIP-seq H3K4me3 and H3K9ac peaks for DamID TAF3 and H3K9ac, respectively. Middle panel shows heterochromatin; signal is centered around ChIP-seq H3K9me3 domain borders. Right panel shows Polycomb; signal is centered around ChIP-seq H3K27me3 domain borders. Solid line indicates the mean signal at these regions, shaded line indicates smoothened signal. Large dashed line indicates the location of the highest signal in ChIP (REF); small dashed line indicates the background measuring point (BG). Grey shaded area indicates the region over which the signal at the REF point drops with 50% relative to the BG point. The size of the drop-off distance is indicated in the top left. **h,** Comparison of DamWT and Dam126 signal. Left: average DamID enrichment plots over genomic regions of interest. Regions are the TSS of the top 25% H3K9ac-enriched genes for H4K20me1 (top), and ChIP-seq domains for H3K27me3 (bottom). Right: genome browser views of DamID enrichment corresponding to left panels. The data shown in H represent the combined data of all samples of a particular targeting domain. **i**, Strip plot of samples comparing the IC of DamWT (grey) and Dam126 (black) targeting constructs.

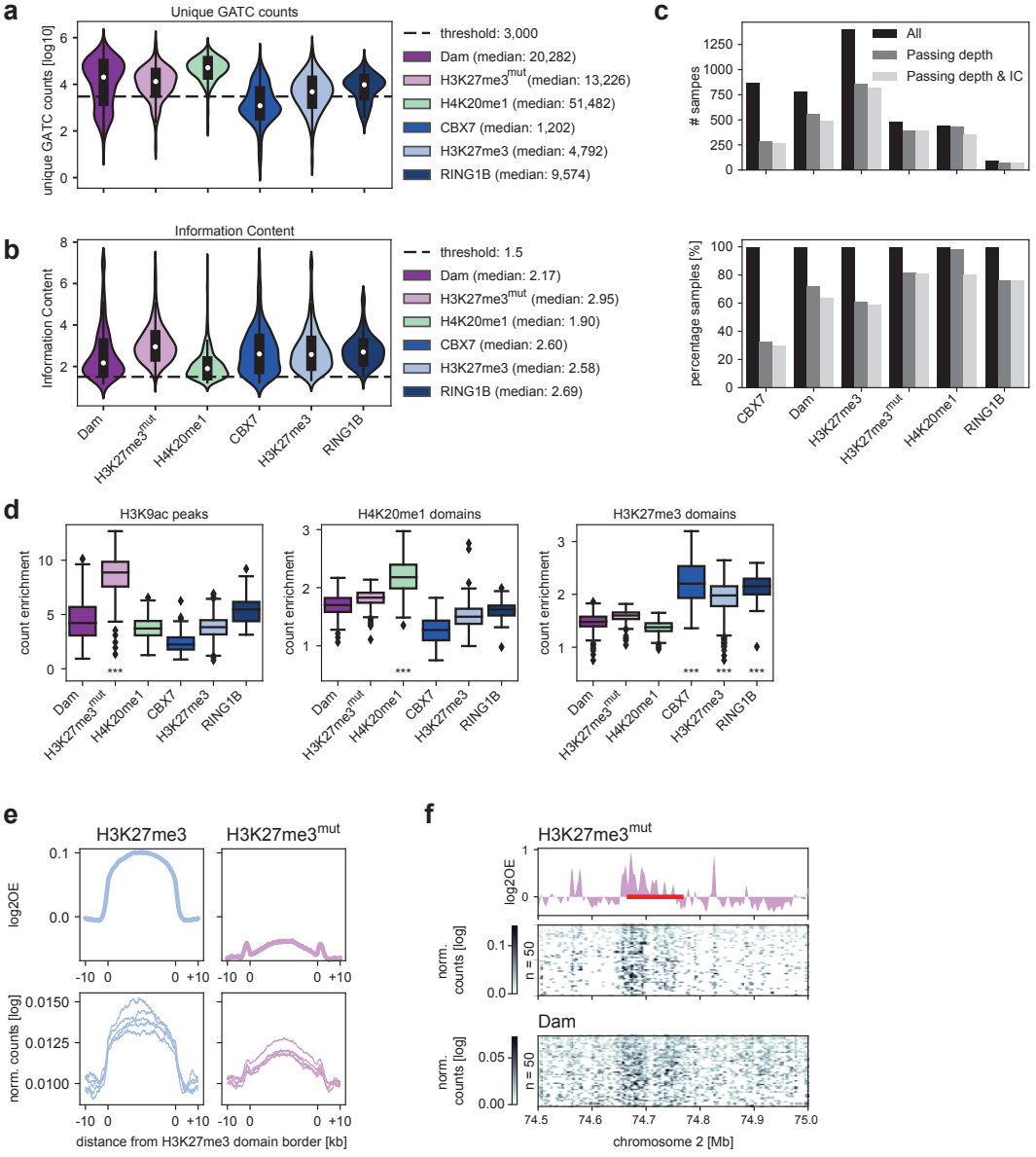

a
Unique GATC counts
unique GATC counts [log10]
threshold: 3,000
Dam (median: 20,282)
H3K27me3mut (median: 13,226)
H4K20me1 (median: 51,482)
CBX7 (median: 1,202)
H3K27me3 (median: 4,792)
RING1B (median: 9,574)
b
Information Content
Information Content
threshold: 1.5
Dam (median: 2.17)
H3K27me3mut (median: 2.95)
H4K20me1 (median: 1.90)
CBX7 (median: 2.60)
H3K27me3 (median: 2.58)
RING1B (median: 2.69)
Dam
H3K27me3mut
H4K20me1
CBX7
H3K27me3
RING1B
c
sampes
All
Passing depth
Passing depth & IC
percentage samples [%]
CBX7
Dam
H3K27me3
H3K27me3mut
H4K20me1
RING1B
d
H3K9ac peaks
H4K20me1 domains
H3K27me3 domains
count enrichment
***
e
H3K27me3
H3K27me3mut
log2OE
norm. counts [log]
distance from H3K27me3 domain border [kb]
f
H3K27me3mut
log2OE
norm. counts [log]
n = 50
Dam
chromosome 2 [Mb]

Fig. S3.2. Detection of histone PTMs in single mouse embryonic stem cells with a single-cell implementation of EpiDamID

a, Violin plots indicating the distribution of the number of unique GATCs detected for each cell line. The dashed line indicates the threshold used for data filtering. **b**, Violin plots indicating the distribution of the Information Content (IC) after filtering on depth for each cell line. The dashed line indicates the threshold used for data filtering. **c**, Overview of the number (top) and percentage (bottom) of samples retained after filtering on depth and IC. **d**, Boxplots showing the count enrichment in H3K9ac ChIP-seq peaks (left), H4K20me1 ChIP-seq domains (middle), and H3K27me3 ChIP-seq domains (right) of all single cells per DamID construct. Count enrichment was computed as the fraction of GATC counts that fell within the regions, relative to the total fraction of genomic GATC positions inside these domains. In each plot, the enrichment of constructs of interest are compared to the enrichment in the Dam control. The significance of the difference was tested with a two-sided Mann-Whitney-U test. *** indicates a p-values smaller than 0.001. Constructs without an indication of significance were not tested. **e**, Average signal over H3K27me3 ChIP-seq domains of H3K27me3 and H3K27me3^{mut} mintbodies. Top: in silico populations normalized for Dam; Bottom: five of the best single-cell samples (bottom) normalized by read depth. **f**, Signal of H3K27me3^{mut} and Dam control over the HoxD cluster and neighboring regions. The DamID track show the Dam-normalized in silico populations of H3K27me3^{mut}, while the heatmaps show the depth-normalized single-cell data of the fifty richest cells for H3K27me3mut and Dam. The red bar around 74.7 Mb indicates the HoxD cluster.

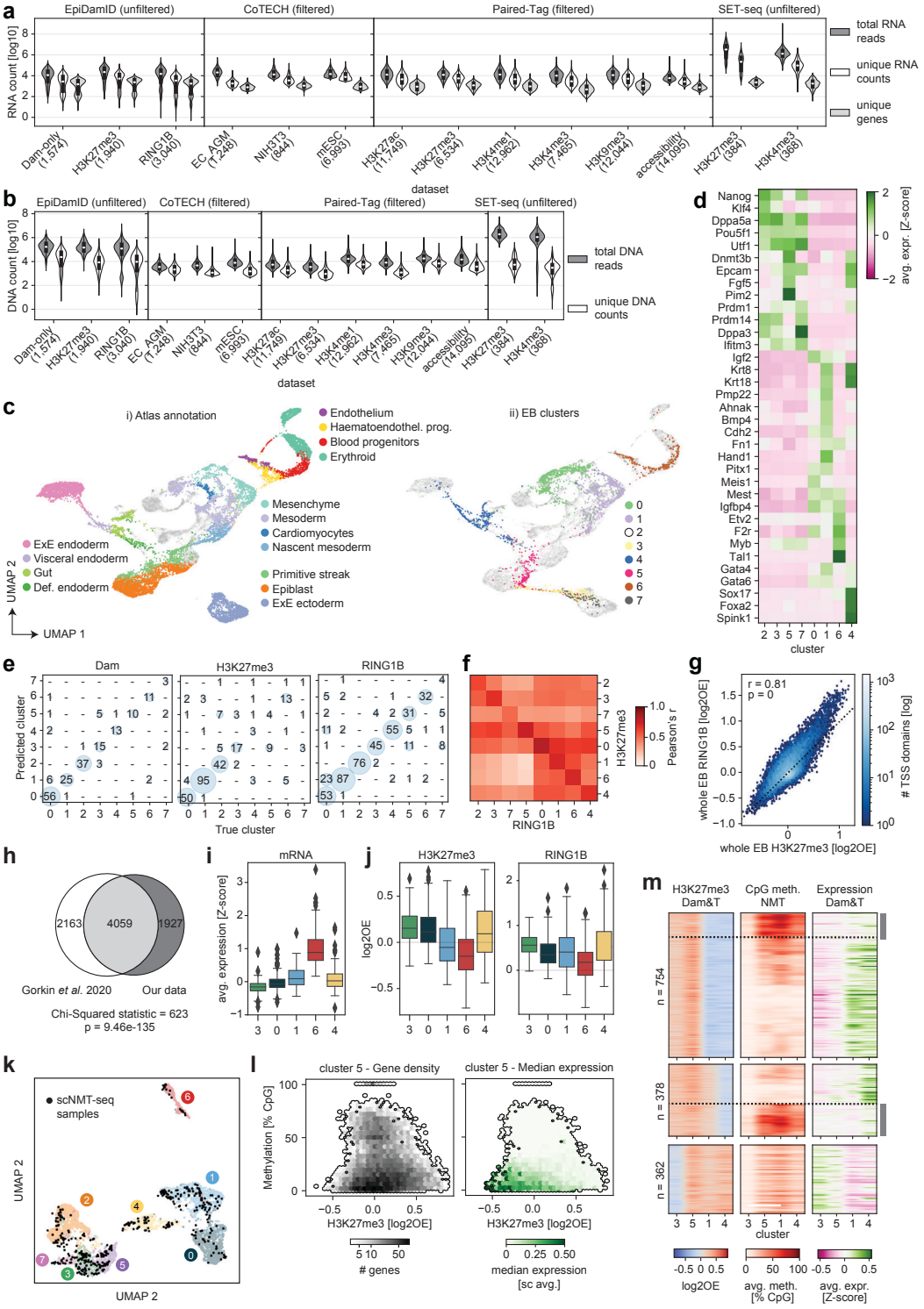

a
EpiDamID (unfiltered)
CoTECH (filtered)
Paired-Tag (filtered)
SET-seq (unfiltered)
RNA count [log10]
total RNA reads
unique RNA counts
unique genes
Dam-only (1,574)
H3K27me3 (1,940)
RING1B (3,040)
EC_AGM (1,248)
NIH3T3 (844)
mESC (6,993)
H3K27ac (11,749)
H3K27me3 (6,534)
H3K4me1 (12,962)
H3K4me3 (7,465)
H3K9me3 (12,044)
accessibility (14,095)
H3K27me3 (384)
H3K4me3 (368)
b
DNA count [log10]
total DNA reads
unique DNA counts
dataset
c
i) Atlas annotation
ii) EB clusters
Endothelium
Haematoendothel. prog.
Blood progenitors
Erythroid
Mesenchyme
Mesoderm
Cardiomyocytes
Nascent mesoderm
Primitive streak
Epiblast
ExE ectoderm
ExE endoderm
Visceral endoderm
Gut
Def. endoderm
UMAP 1
UMAP 2
d
Nanog
Klf4
Dppa5a
Pou5f1
Utf1
Dnmt3b
Epcam
Fgf5
Pim2
Prdm1
Prdm14
Dppa3
Ifitm3
Igf2
Krt8
Krt18
Pmp22
Ahnak
Bmp4
Cdh2
Fn1
Hand1
Pitx1
Meis1
Mest
Igfbp4
Etv2
F2r
Myb
Tal1
Gata4
Gata6
Sox17
Foxa2
Spink1
avg. expr. [Z-score]
cluster
e
Dam
H3K27me3
RING1B
Predicted cluster
True cluster
f
H3K27me3
RING1B
Pearson's r
g
r = 0.81
p = 0
whole EB RING1B [log2OE]
whole EB H3K27me3 [log2OE]
TSS domains [log]
h
2163
4059
1927
Gorkin et al. 2020
Our data
Chi-Squared statistic = 623
p = 9.46e-135
i
mRNA
avg. expression [Z-score]
j
H3K27me3
RING1B
log2OE
k
scNMT-seq samples
UMAP 2
l
cluster 5 - Gene density
cluster 5 - Median expression
Methylation [% CpG]
H3K27me3 [log2OE]
genes
median expression [sc avg.]
m
H3K27me3 Dam&T
CpG meth. NMT
Expression Dam&T
n = 754
n = 378
n = 362
cluster
log2OE
avg. meth. [% CpG]
avg. expr. [Z-score]

Fig. S3.3. Validation and characterization of scDam&T-seq data in mouse embryoid bodies
a-b, Overview of the RNA and DNA outputs for a number of datasets generated by recent single-cell multimodal omics techniques. **a,** The number of raw RNA-derived reads, unique transcripts and unique genes. **b,** The number of raw DNA- derived reads and unique counts. The included techniques are CoTECH (Xiong *et al.*, 2021), Paired-Tag (Zhu *et al.*, 2021) and SET-seq (Sun *et al.*, 2021). The EpiDamID (scDam&T-seq) data shows the statistics of the embryoid body (EB) dataset. Some techniques show only the statistics of cells that passed quality thresholds ("filtered"), while others show the statistics of all obtained cells ("unfiltered"). The labels on the x-axis indicate the name of the various datasets, with the number of samples shown in paratheses. **c,** UMAPs of samples based on the integration of our EB transcription data with single-cell RNA-seq mouse embryonic data (Pijuan-Sala *et al.*, 2019), colored by reference-annotated cell type (i) and EB-annotated cluster (ii). For atlas integration, the day 0 (i.e., mESC) time point was excluded. **d,** Average expression of known marker genes. Expression was standardized over single-cells and the per-cluster average was computed. **e,** Confusion plots showing the performance of the LDA classifier during training, for each construct. **f,** Pearson correlation between the combined H3K27me3 and RING1B DamID signal at the TSS of all genes per transcriptional cluster. **g,** Correlation of combined H3K27me3 and RING1B DamID signal at the TSS of all genes. Data of all single-cell samples passing DamID thresholds was combined for each construct. The correlation was computed using Pearson's correlation. **h,** Overlap between a published set of PRC targets during mouse development66 and our PRC targets. Only genes represented in both datasets could be compared. Significance of the overlap was computed with a Chi-squared test.
i, Boxplots showing the expression (averaged Z-score) of genes identified as significantly upregulated in cluster 6. **j,** Boxplots showing the H3K27me3 (left) and RING1B (right) DamID signal at the TSS of the subset of genes shown in **i** that are PRC targets. **k**, UMAPs of samples based on the integration of our EB transcription data with the transcriptional readout of the EB scNMT-seq data generated by Argelaguet *et al.* EpiDamID samples are colored by the transcriptional clusters determine previously; scNMT-seq samples are indicated in black.
l, Relationship between promoter CpG methylation, promoter H3K27me3 enrichment and gene expression of all genes in cells belonging to cluster 5 (epiblast-like). The left plot shows the relationship between promoter CpG methylation (+/- 2 kb around TSS) and H3K27me3 enrichment (-5 kb/+3 kb around TSS) for all genes. The right plot shows the same relationship, but the color scale indicates the median expression of genes in each region of the plot.
m, Heatmaps indicating the promoter H3K27me3 enrichment, promoter CpG methylation and gene expression for three groups of PRC targets with variable H3K27me3 enrichment. Rows are genes; columns are transcription clusters. Enrichment is shown for the 4 clusters that contained sufficient scNMT-seq samples (cluster 3: 31 cells; cluster 5: 21 cells; cluster 1: 37 cells; cluster 4: 43 cells). Genes are sorted by hierarchical cluster based on their CpG methylation levels. Examples of genes where H3K27me3 and CpG methylation complementary repress genes are indicated with a dotted line and a grey box.

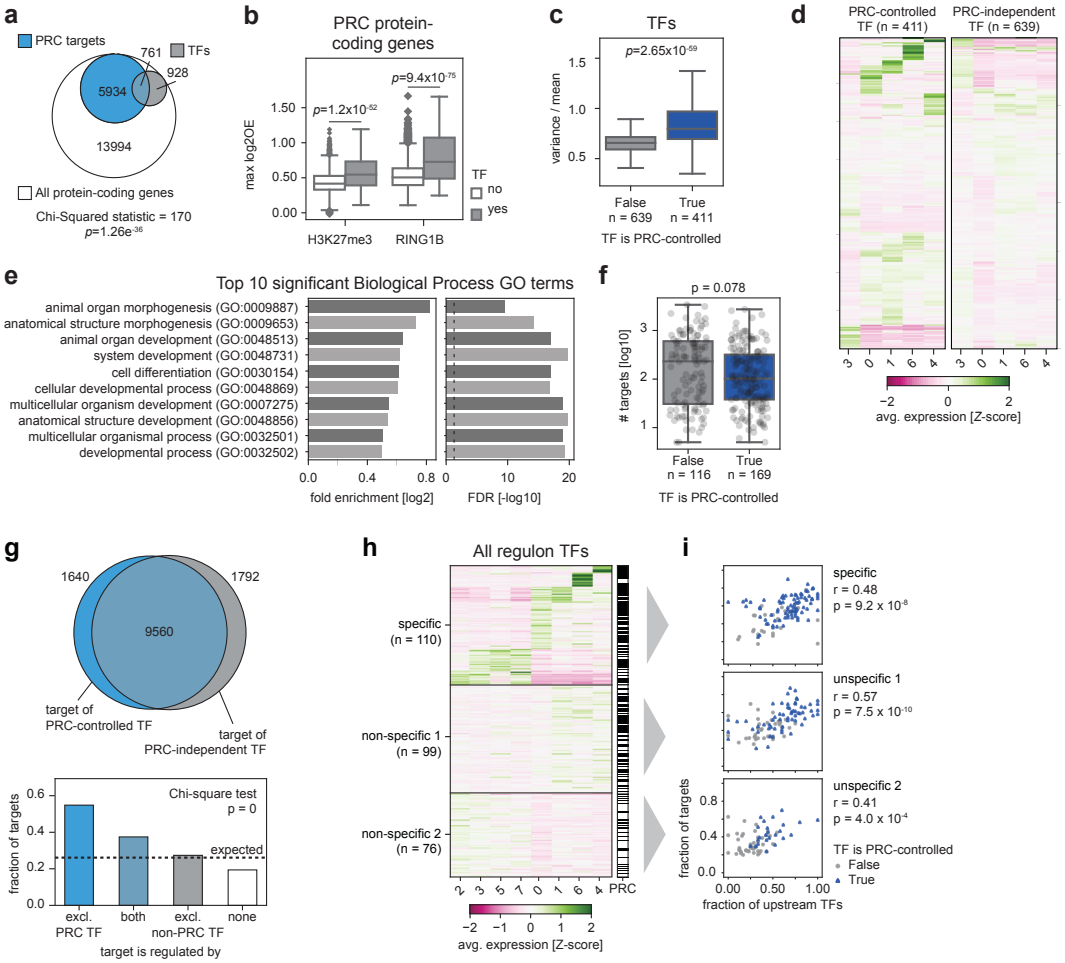

a
PRC targets
TFs
761
928
5934
13994
All protein-coding genes
Chi-Squared statistic = 170
p=1.26e-36
b
PRC protein-coding genes
max log2OE
p=1.2x10-52
p=9.4x10-75
TF
no
yes
H3K27me3
RING1B
c
TFs
variance / mean
p=2.65x10-59
False
n = 639
True
n = 411
TF is PRC-controlled
d
PRC-controlled TF (n = 411)
PRC-independent TF (n = 639)
avg. expression [Z-score]
e
Top 10 significant Biological Process GO terms
animal organ morphogenesis (GO:0009887)
anatomical structure morphogenesis (GO:0009653)
animal organ development (GO:0048513)
system development (GO:0048731)
cell differentiation (GO:0030154)
cellular developmental process (GO:0048869)
multicellular organism development (GO:0007275)
anatomical structure development (GO:0048856)
multicellular organismal process (GO:0032501)
developmental process (GO:0032502)
fold enrichment [log2]
FDR [-log10]
f
p = 0.078
# targets [log10]
False
n = 116
True
n = 169
TF is PRC-controlled
g
1640
9560
1792
target of PRC-controlled TF
target of PRC-independent TF
Chi-square test
p = 0
expected
fraction of targets
excl. PRC TF
both
excl. non-PRC TF
none
target is regulated by
h
All regulon TFs
specific (n = 110)
non-specific 1 (n = 99)
non-specific 2 (n = 76)
PRC
avg. expression [Z-score]
i
specific
r = 0.48
p = 9.2 x 10-8
unspecific 1
r = 0.57
p = 7.5 x 10-10
unspecific 2
r = 0.41
p = 4.0 x 10-4
fraction of targets
fraction of upstream TFs
TF is PRC-controlled
False
True

Fig. S3.4. Characterization of the Polycomb-regulated regulatory network

a, Venn diagram showing the overlap between PRC-controlled protein-coding genes (blue) and transcription factors (TF) (grey) in the context of all protein-coding genes (white). The significance of the overlap between PRC targets and TFs was computed using a Chi-squared test. **b,** Boxplots showing the maximum observed H3K27me3 and RING1B DamID signal across transcriptional clusters for PRC-controlled TFs (grey) and the remaining PRC-controlled protein-coding genes (white). The significance of the difference between TFs and other genes was tested with a two-sided Mann- Whitney-U test. **c**, Quantification of variability in gene expression of PRC-regulated and PRC-independent TFs (only expressed genes are included). Boxplots show variance over mean across all single cells. Significance was computed using a two-sided Mann-Whitney U test. **d**, Clustered heatmaps showing mRNA expression (averaged Z-score) per cluster, of Polycomb-regulated TFs (left) and Polycomb-independent TFs (right). Only expressed genes are included in this plot. **e,** The ten most significant Biological Process GO terms between PRC-controlled and PRC-independent TFs. **f**, Number of targets of each regulon TF, split by whether or not the TF is PRC-regulated. The significance of the difference between the two groups was tested with a two-sided Mann-Whitney U test. **g,** Top: Venn diagram displaying the overlap between genes that are targets of a PRC-controlled TF (blue) and genes that are targets of a PRC-independent TF (grey). Bottom: Bar plot showing the fraction of targets in each category that is PRC-regulated. The dotted line indicates the expected fraction, i.e., the fraction of all genes that is a PRC target. A Chi-square test was performed to evaluate whether the deviation from the expected frequencies is significant. **h,** Clustered heatmap showing mRNA expression (averaged Z-score) per cluster, of all regulon TFs, grouped by lineage-specific or non-specific genes. TFs are annotated as PRC- controlled (black) or PRC-independent (white). **i,** Scatter plot showing the relationship between the fraction of Polycomb- controlled targets and regulators of a regulon TF. Regulon TFs that are PRC controlled are indicated in blue; regulon TFs that are PRC independent are indicated in grey. Regulon TFs are split based on the groups indicated in **i**. Correlation was computed using Pearson's coefficient.

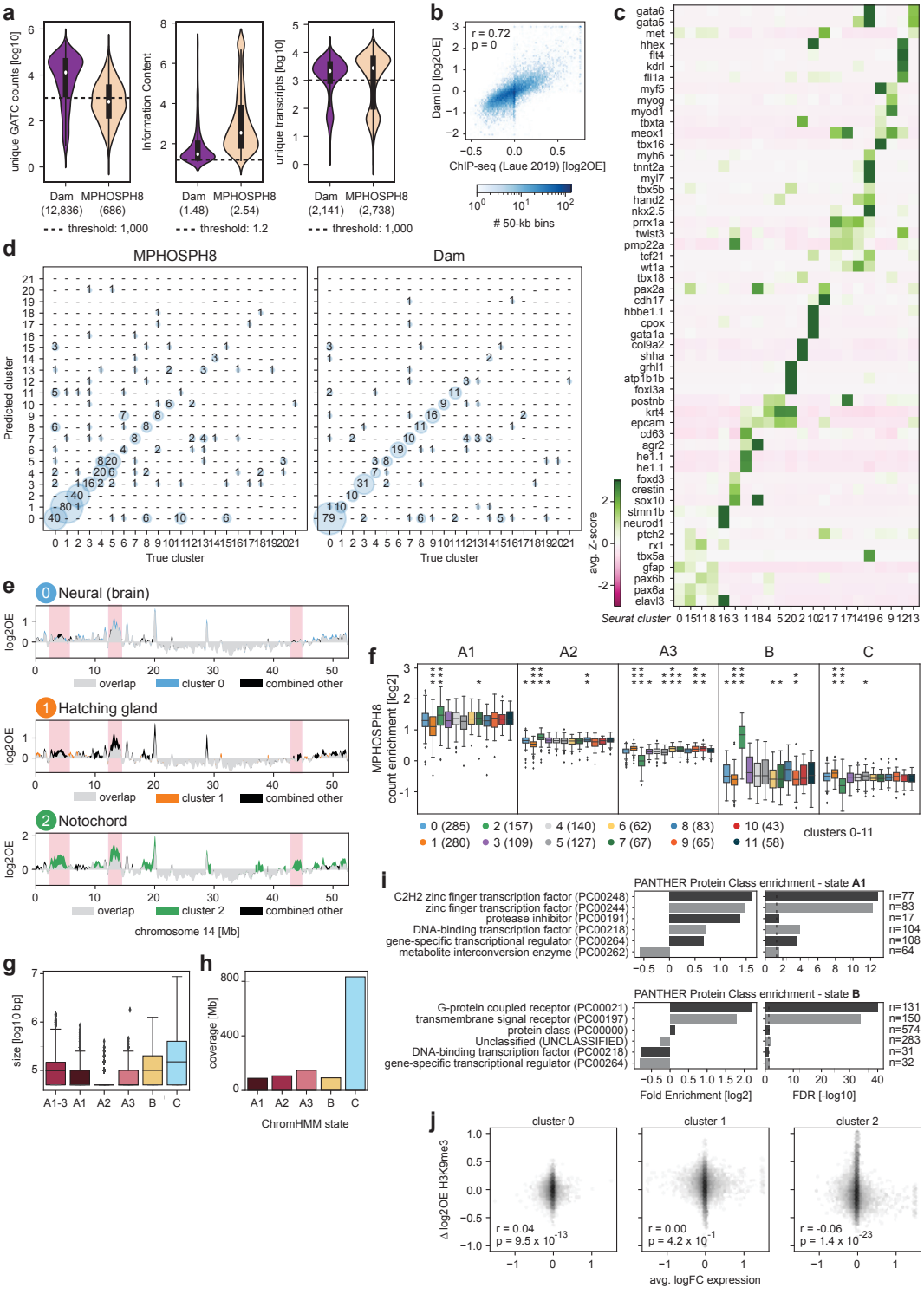

a
unique GATC counts [log10]
Information Content
unique transcripts [log10]
Dam
MPHOSPH8
(12,836)
(686)
(1.48)
(2.54)
(2,141)
(2,738)
threshold: 1,000
threshold: 1.2
threshold: 1,000
b
r = 0.72
p = 0
DamID [log2OE]
ChIP-seq (Laue 2019) [log2OE]
50-kb bins
c
gata6
gata5
met
hhex
flt4
kdrl
fli1a
myf5
myog
myod1
tbxta
meox1
tbx16
myh6
tnnt2a
myl7
tbx5b
hand2
nkx2.5
prrx1a
twist3
pmp22a
tcf21
wt1a
tbx18
pax2a
cdh17
hbbe1.1
cpox
gata1a
col9a2
shha
grhl1
atp1b1b
foxi3a
postnb
krt4
epcam
cd63
agr2
he1.1
he1.1
foxd3
crestin
sox10
stmn1b
neurod1
ptch2
rx1
tbx5a
gfap
pax6b
pax6a
elavl3
avg. Z-score
Seurat cluster
d
MPHOSPH8
Dam
Predicted cluster
True cluster
e
0 Neural (brain)
1 Hatching gland
2 Notochord
log2OE
overlap
cluster 0
cluster 1
cluster 2
combined other
chromosome 14 [Mb]
f
A1
A2
A3
B
C
MPHOSPH8 count enrichment [log2]
0 (285)
1 (280)
2 (157)
3 (109)
4 (140)
5 (127)
6 (62)
7 (67)
8 (83)
9 (65)
10 (43)
11 (58)
clusters 0-11
g
size [log10 bp]
A1-3
A1
A2
A3
B
C
h
coverage [Mb]
ChromHMM state
i
PANTHER Protein Class enrichment - state A1
C2H2 zinc finger transcription factor (PC00248)
zinc finger transcription factor (PC00244)
protease inhibitor (PC00191)
DNA-binding transcription factor (PC00218)
gene-specific transcriptional regulator (PC00264)
metabolite interconversion enzyme (PC00262)
n=77
n=83
n=17
n=104
n=108
n=64
PANTHER Protein Class enrichment - state B
G-protein coupled receptor (PC00021)
transmembrane signal receptor (PC00197)
protein class (PC00000)
Unclassified (UNCLASSIFIED)
DNA-binding transcription factor (PC00218)
gene-specific transcriptional regulator (PC00264)
n=131
n=150
n=574
n=283
n=31
n=32
Fold Enrichment [log2]
FDR [-log10]
j
cluster 0
cluster 1
cluster 2
Δ log2OE H3K9me3
r = 0.04
p = 9.5 x 10^-13
r = 0.00
p = 4.2 x 10^-1
r = -0.06
p = 1.4 x 10^-23
avg. logFC expression

Fig. S3.5. Characterization of transcriptomic clusters and associated genomic H3K9me3 enrichments

a, Violin plots showing the total number of unique GATC counts, the information content (IC) and total number of unique transcripts obtained for all cells in the zebrafish dataset. **b**, Comparison of our data with a published H3K9me3 ChIP-seq dataset of the 6-hpf zebrafish embryo (Laue et al., 2019). All single-cell MPHOSPH8 and Dam samples were combined to generate an in silico whole-embryo data set; DamID data is the log2OE of MPHOSPH8 signal over Dam is shown; ChIP- seq is the log2OE of H3K9me3 over input control. The correlation was computed using Pearson's correlation. **c**, Expression of marker genes over all clusters, ordered by cell type. The average single-cell Z-scores are shown. **d,** Confusion plots showing the performance of the LDA classifier during training, for each construct. **e,** Genomic H3K9me3 signal over chromosome 14. For clusters 0-2, the cluster-specific signal (color) is compared to the combined signal from all other clusters (black). Each set indicates the overlay, where overlapping regions are colored grey. **f**, Boxplots showing the enrichment of counts within genomic regions belonging to each of the five ChromHMM states for all cells belonging to transcriptional clusters 0-11. Count enrichment was computed as the fraction of GATC counts that fell within the regions, relative to the total fraction of genomic GATC positions inside these domains. Per state, the count enrichment of a cluster was compared to the enrichment of cells in all other clusters using a two-sided Mann-Whitney-U test. *** = $p < 0.001$; ** = $p < 0.01$; * = $p < 0.1$; no indication means the result was insignificant. **g**, Distribution of domain sizes per ChromHMM state and for states A1-3 combined. **h,** Total genomic coverage per ChromHMM state. **i,** PANTHER protein- class enrichments (Mi et al., 2013) for genes found in state A1 (top) and B (bottom). **j,** Plot displaying the relationship between differential gene expression and differential H3K9me3 enrichment. The x-axis shows the average log-fold change in gene expression of cells in one cluster relative to all other cells; the y-axis shows the differential log2OE H3K9me3 at these genes of one cluster relative to all other cells. H3K9me3 at a gene was measured as the log2OE value of the 50-kb genomic window containing the TSS of the gene. The relationship between the variables was tested with a Pearson's correlation test.

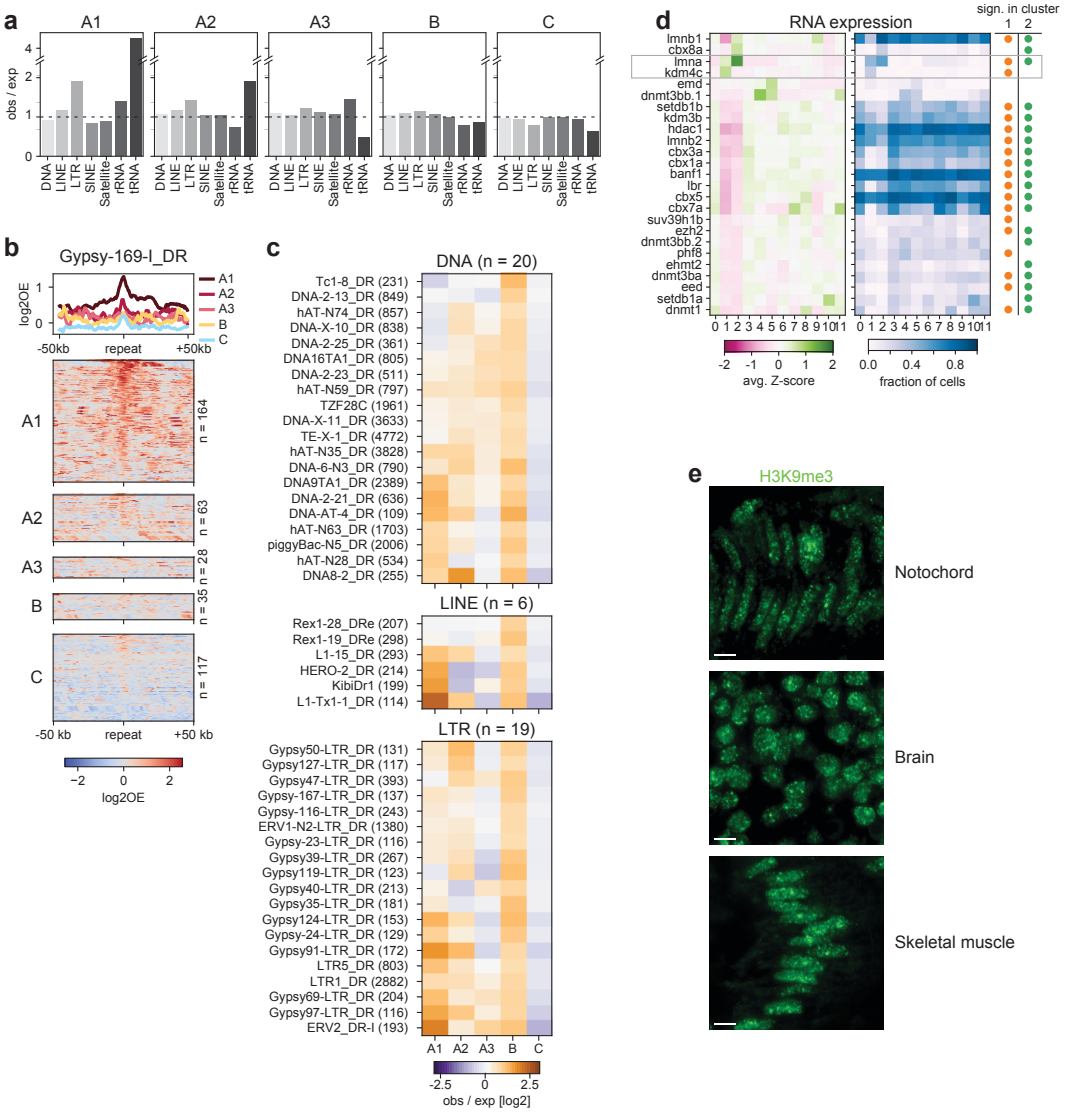

a
A1
A2
A3
B
C
obs / exp
DNA
LINE
LTR
SINE
Satellite
rRNA
tRNA
b
Gypsy-169-I_DR
log2OE
-50kb
repeat
+50kb
n = 164
n = 63
n = 28
n = 35
n = 117
-50 kb
+50 kb
log2OE
c
DNA (n = 20)
Tc1-8_DR (231)
DNA-2-13_DR (849)
hAT-N74_DR (857)
DNA-X-10_DR (838)
DNA-2-25_DR (361)
DNA16TA1_DR (805)
DNA-2-23_DR (511)
hAT-N59_DR (797)
TZF28C (1961)
DNA-X-11_DR (3633)
TE-X-1_DR (4772)
hAT-N35_DR (3828)
DNA-6-N3_DR (790)
DNA9TA1_DR (2389)
DNA-2-21_DR (636)
DNA-AT-4_DR (109)
hAT-N63_DR (1703)
piggyBac-N5_DR (2006)
hAT-N28_DR (534)
DNA8-2_DR (255)
LINE (n = 6)
Rex1-28_DRe (207)
Rex1-19_DRe (298)
L1-15_DR (293)
HERO-2_DR (214)
KibiDr1 (199)
L1-Tx1-1_DR (114)
LTR (n = 19)
Gypsy50-LTR_DR (131)
Gypsy127-LTR_DR (117)
Gypsy47-LTR_DR (393)
Gypsy-167-LTR_DR (137)
Gypsy-116-LTR_DR (243)
ERV1-N2-LTR_DR (1380)
Gypsy-23-LTR_DR (116)
Gypsy39-LTR_DR (267)
Gypsy119-LTR_DR (123)
Gypsy40-LTR_DR (213)
Gypsy35-LTR_DR (181)
Gypsy124-LTR_DR (153)
Gypsy-24-LTR_DR (129)
Gypsy91-LTR_DR (172)
LTR5_DR (803)
LTR1_DR (2882)
Gypsy69-LTR_DR (204)
Gypsy97-LTR_DR (116)
ERV2_DR-I (193)
A1 A2 A3 B C
obs / exp [log2]
d
RNA expression
sign. in cluster
Imnb1
cbx8a
lmna
kdm4c
emd
dnmt3bb.1
setdb1b
kdm3b
hdac1
lmnb2
cbx3a
cbx1a
banf1
lbr
cbx5
cbx7a
suv39h1b
ezh2
dnmt3bb.2
phf8
ehmt2
dnmt3ba
eed
setdb1a
dnmt1
avg. Z-score
fraction of cells
e
H3K9me3
Notochord
Brain
Skeletal muscle

Fig. S3.6. Characterization of repeat content, expression of chromatin factors and nuclear localization of H3K9me3 chromatin

a, Enrichment of repeats per class for all ChromHMM states. Enrichment is computed as the observed number of repeats within a state relative to the expected number based on the genome coverage of each state. **b**, H3K9me3 enrichment at Gypsy-169-I_DR repeats across ChromHMM states. The heatmaps show the enrichment per individual repeat instance, while the line plot shows the average enrichment per state. **c**, Enrichment of repeats in ChromHMM states as in **3.5i**. Only repeats having at least 100 copies throughout the genome and an enrichment ≥1.5 in state B are included. Enrichment is computed as the observed number of repeats in a stated compared to the expected number based on the genome coverage of that state. **d**, RNA expression of various chromatin factors across clusters 0-11. The left heatmap shows the average single-cell expression (Z-score); the right heatmaps shows the fraction of cells in each cluster with at least one transcript of each gene. Only factors that are expressed in at least 10% of cells of at least one cluster are shown. **e**, Representative images of H3K9me3 staining in cryosections of notochord (left), brain (middle), and skeletal muscle (right) in 15-somite embryos. Scale bars represent 4 μm.

Supplementary Tables

Table S3.1. Overview of EpiDamID constructs used in RPE-1 DamID experiments

Table S3.2. Metadata and quality metrics of all single-cell samples of the ESC, EB and zebrafish experiments

Available in online version: https://doi.org/10.1016/j.molcel.2022.03.009

References

1. Juan, A. H. *et al.* Roles of H3K27me2 and H3K27me3 Examined during Fate Specification of Embryonic Stem Cells. *Cell Rep* **17**, 1369–1382 (2016).
2. Nicetto, D. *et al.* H3K9me3-heterochromatin loss at protein-coding genes enables developmental lineage specification. *Science (1979)* **363**, 294–297 (2019).
3. Pengelly, A. R., Copur, Ö., Jäckle, H., Herzig, A. & Müller, J. A histone mutant reproduces the phenotype caused by loss of histone-modifying factor polycomb. *Science (1979)* **339**, 698–699 (2013).
4. Hirota, T., Lipp, J. J., Toh, B.-H. & Peters, J.-M. Histone H3 serine 10 phosphorylation by Aurora B causes HP1 dissociation from heterochromatin. *Nature 2005 438:7071* **438**,1176–1180 (2005).
5. Liu, W. *et al.* PHF8 mediates histone H4 lysine 20 demethylation events involved in cell cycle progression. *Nature* **466**, 508–512 (2010).
6. Rogakou, E. P., Pilch, D. R., Orr, A. H., Ivanova,V. S. & Bonner, W. M. DNA Double-stranded Breaks Induce Histone H2AX Phosphorylation on Serine 139 *. *Journal of Biological Chemistry* **273**, 5858–5868 (1998).
7. Sanders, S. L. *et al.* Methylation of Histone H4 Lysine 20 Controls Recruitment of Crb2 to Sites of DNA Damage. *Cell* **119**, 603–614 (2004).
8. Solomon, M. J. & Varshavsky, A. Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proceedings of the National Academy of Sciences* **82**, 6470–6474 (1985).
9. Schmid, M., Durussel, T. & Laemmli, U. K. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* **16**, 147–57 (2004).
10. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, 1–35 (2017).
11. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nature Biotechnology 2015 33:11* **33**, 1165–1172 (2015).
12. Harada, A. *et al.* A chromatin integration labelling method enables epigenomic profiling with lower input. *Nat Cell Biol* **21**, 287–296 (2019).
13. Carter, B. *et al.* Mapping histone modifications in low cell number and single cells using antibody-guided chromatin tagmentation (ACTseq). *Nat Commun* **10**, (2019).
14. Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* **51**, 1060–1066 (2019).
15. Ku, W. L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods* **16**, 323–325 (2019).
16. Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329.e11 (2019).
17. Kaya-Okur, H. S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1–10 (2019).
18. Wang, Q. *et al.* CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Mol Cell* **76**, 206-216.e7 (2019).
19. Ai, S. *et al.* Profiling chromatin states using single-cell itChIP-seq. *Nat Cell Biol* **21**, 1164–1172 (2019).
20. Zeller, P. *et al.* Hierarchical chromatin regulation during blood formation uncovered by single-cell sortChIC. *bioRxiv* 2021.04.26.440606 (2021) doi:10.1101/2021.04.26.440606.
21. Zhu, C. *et al.* Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat Methods* **18**, 283–292 (2021).
22. Xiong, H., Luo, Y., Wang, Q., Yu, X. & He, A. Single-cell joint detection of chromatin occupancy and transcriptome enables higher-dimensional epigenomic reconstructions. *Nat Methods* **18**, 652–660 (2021).
23. Sun, Z. *et al.* Joint single-cell multiomic analysis in Wnt3a induced asymmetric stem cell division. *Nat Commun* **12**, (2021).
24. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nature Methods 2013 10:12* **10**, 1213–1218 (2013).
25. Kaya-Okur, H. S., Janssens, D. H., Henikoff, J. G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nature Protocols 2020 15:10* **15**, 3264–3283 (2020).
26. Zhang, W. &, Wang, M. & Zhang, Y. Tn5 transposase-based epigenomic profiling methods are prone to open chromatin

bias. *bioRxiv* 2021.07.09.451758 (2021) doi:10.1101/2021.07.09.451758.

27. Rooijers, K. *et al.* Simultaneous quantification of protein–DNA contacts and transcriptomes in single cells. *Nat Biotechnol* (2019) doi:10.1038/s41587-019-0150-y.
28. Vogel, M. J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein DNA interactions using DamID in mammalian cells. *Nat Protoc* **2**, 1467–1478 (2007).
29. Filion, G. J. *et al.* Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143**, 212–224 (2010).
30. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered Dam methyltransferase. *Nature Biotechnology 2000 18:4* **18**, 424–428 (2000).
31. Kungulovski, G., Mauser, R., Reinhardt, R. & Jeltsch, A. Application of recombinant TAF3 PHD domain instead of anti-H3K4me3 antibody. *Epigenetics Chromatin* **9**, (2016).
32. Kungulovski, G. *et al.* Application of histone modification-specific interaction domains as an alternative to antibodies. *Genome Res* **24**, 1842–1853 (2014).
33. Vermeulen, M. *et al.* Selective Anchoring of TFIID to Nucleosomes by Trimethylation of Histone H3 Lysine 4. *Cell* **131**, 58–69 (2007).
34. Sato, Y. *et al.* Genetically encoded system to track histone modification in vivo. *Sci Rep* **3**, (2013).
35. Sato, Y. *et al.* A Genetically Encoded Probe for Live-Cell Imaging of H4K20 Monomethylation. *J Mol Biol* **428**, 3885–3902 (2016).
36. Tjalsma, S. J. D. *et al.* H4K20me1 and H3K27me3 are concurrently loaded onto the inactive X chromosome but dispensable for inducing gene silencing. *EMBO Rep* **22**, 1–17 (2021).
37. Sato, Y., Nakao, M. & Kimura, H. Live-cell imaging probes to track chromatin modification dynamics. *Microscopy* 1–8 (2021) doi:10.1093/ JMICRO/DFAB030.
38. Villaseñor, R. *et al.* ChromID identifies the protein interactome at chromatin marks. *Nat Biotechnol* (2020) doi:10.1038/s41587-020-0434-2.
39. Altemose, N. *et al.* µDamID: A Microfluidic Approach for Joint Imaging and Sequencing of Protein-DNA Interactions in Single Cells. *Cell Syst* **11**, 354-366.e9 (2020).
40. Borsos, M. *et al.* Genome–lamina interactions are established de novo in the early mouse embryo. *Nature* **569**, 729–733 (2019).
41. Kind, J. *et al.* Single-Cell Dynamics of Genome-Nuclear Lamina Interactions. *Cell* **153**, 178–192 (2013).
42. Southall, T. D. *et al.* Cell-type-specific profiling of gene expression and chromatin binding without cell isolation: Assaying RNA pol II occupancy in neural stem cells. *Dev Cell* **26**, 101–112 (2013).
43. Wong, X. *et al.* Mapping the micro-proteome of the nuclear lamina and lamina-associated domains. *Life Sci Alliance* **4**, (2021).
44. Kind, J. *et al.* Genome-wide Maps of Nuclear Lamina Interactions in Single Human Cells. *Cell* **163**, 134–147 (2015).
45. Cheetham, S. W. *et al.* Single-molecule simultaneous profiling of DNA methylation and DNA-protein interactions with NanoporeDamID. *bioRxiv*2021.08.09.455753 (2021) doi:10.1101/2021.08.09.455753.
46. Pal, M., Kind, J. & Torres-Padilla, M. E. DamID to map genome-protein interactions in preimplantation mouse embryos. *Methods in Molecular Biology* **2214**, 265–282 (2021).
47. Szczesnik, T., Ho, J. W. K. & Sherwood, R. Dam mutants provide improved sensitivity and spatial resolution for profiling transcription factor binding. *Epigenetics Chromatin* **12**, 1–11 (2019).
48. Park, M., Patel, N., Keung, A. J. & Khalil, A. S. Construction of a Synthetic, Chromatin-Based Epigenetic System in Human Cells. *SSRN* (2018) doi:10.2139/ssrn.3155804.
49. Markodimitraki, C. M. *et al.* Simultaneous quantification of protein–DNA interactions and transcriptomes in single cells with scDam&T-seq. *Nat Protoc* **15**, 1922–1953 (2020).
50. Tosti, L. *et al.* Mapping transcription factor occupancy using minimal numbers of cells in vitro and in vivo. *Genome Res* **28**, 592–605 (2018).
51. Cheetham, S. W. *et al.* Targeted damid reveals differential binding of mammalian pluripotency factors. *Development (Cambridge)* **145**, (2018).
52. Schaik, T. van, Vos, M., Peric-Hupkes, D., Celie, P. H. & Steensel, B. van. Cell cycle dynamics of lamina-associated DNA. *EMBO Rep* **21**, e50636 (2020).
53. Shoaib, M. *et al.* Histone H4 lysine 20 mono-methylation directly facilitates chromatin openness and promotes transcription of housekeeping genes. *Nat Commun* **12**, 1–16 (2021).

54. Boyer, L. A. *et al.* Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature 2006 441:7091* **441**, 349–353 (2006).

55. Riising, E. M. *et al.* Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Mol Cell* **55**, 347–360 (2014).

56. Davis, C. A. *et al.* The Encyclopedia of DNA elements (ENCODE): data portal update. *Nucleic Acids Res* **46**, D794–D801 (2018).

57. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature 2004 431:7010* **431**, 873–878 (2004).

58. de Napoles, M. *et al.* Polycomb Group Proteins Ring1A/B Link Ubiquitylation of Histone H2A to Heritable Gene Silencing and X Inactivation. *Dev Cell* **7**, 663–676 (2004).

59. Cao, R. *et al.* Role of histone H3 lysine 27 methylation in polycomb-group silencing. *Science (1979)* **298**, 1039–1043 (2002).

60. Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P. & Reinberg, D. Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev* **16**, 2893–2905 (2002).

61. Czermin, B. *et al.* Drosophila Enhancer of Zeste/ESC Complexes Have a Histone H3 Methyltransferase Activity that Marks Chromosomal Polycomb Sites. *Cell* **111**, 185–196 (2002).

62. Müller, J. *et al.* Histone Methyltransferase Activity of a Drosophila Polycomb Group Repressor Complex. *Cell* **111**, 197–208 (2002).

63. Piunti, A. & Shilatifard, A. The roles of Polycomb repressive complexes in mammalian development and cancer. *Nature Reviews Molecular Cell Biology 2021 22:5* **22**, 326–345 (2021).

64. Blackledge, N. P. & Klose, R. J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat Rev Mol Cell Biol* **0123456789**, (2021).

65. Pijuan-Sala, B. *et al.* A single-cell molecular map of mouse gastrulation and early organogenesis. *Nature* **566**, 490–495 (2019).

66. Gorkin, D. U. *et al.* An atlas of dynamic chromatin landscapes in mouse fetal development. *Nature* **583**, 744–751 (2020).

67. Argelaguet, R. *et al.* Multi-omics profiling of mouse gastrulation at single-cell resolution. *Nature* **576**, 487–491 (2019).

68. Li, Y. *et al.* Genome-wide analyses reveal a role of Polycomb in promoting hypomethylation of DNA methylation valleys. *Genome Biol* **19**, 1–16 (2018).

69. Hagarman, J. A., Motley, M. P., Kristjansdottir, K. & Soloway, P. D. Coordinate Regulation of DNA Methylation and H3K27me3 in Mouse Embryonic Stem Cells. *PLoS One* **8**, e53880 (2013).

70. Brinkman, A. B. *et al.* Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res* **22**, 1128–1138 (2012).

71. van de Sande, B. *et al.* A scalable SCENIC workflow for single-cell gene regulatory network analysis. *Nat Protoc* **15**, 2247–2276 (2020).

72. Aibar, S. *et al.* SCENIC: Single-cell regulatory network inference and clustering. *Nat Methods* **14**, 1083–1086 (2017).

73. Wang, C. *et al.* Reprogramming of H3K9me3-dependent heterochromatin during mammalian embryo development. *Nat Cell Biol* **20**, 620–631 (2018).Laue, K., Rajshekar, S., Courtney, A. J., Lewis, Z. A. & Goll, M. G. The maternal to zygotic transition regulates genome-wide heterochromatin establishment in the zebrafish embryo. *Nat Commun* **10**, (2019).

74. Mutlu, B. *et al.* Regulated nuclear accumulation of a histone methyltransferase times the onset of heterochromatin formation in C. elegans embryos. *Sci Adv* **4**, 6224–6246 (2018).

75. Rudolph, T. *et al.* Heterochromatin Formation in Drosophila Is Initiated through Active Removal of H3K4 Methylation by the LSD1 Homolog SU(VAR)3-3. *Mol Cell* **26**, 103–115 (2007).

76. Santos, F., Peters, A. H., Otte, A. P., Reik, W. & Dean, W. Dynamic chromatin modifications characterise the first cell cycle in mouse embryos. *Dev Biol* **280**, 225–236 (2005).

77. Ahmed, K. *et al.* Global Chromatin Architecture Reflects Pluripotency and Lineage Commitment in the Early Mouse Embryo. *PLoS One* **5**, e10531 (2010).

78. Ernst, J. & Kellis, M. ChromHMM: Automating chromatin state discovery and characterization. *Nat Methods* **9**, 215–216 (2012).

79. Ernst, J. & Kellis, M. Chromatin-state discovery and genome annotation with ChromHMM. *Nat Protoc* **12**, 2478–2492 (2017).

80. Hahn, M. A., Wu, X., Li, A. X., Hahn, T. & Pfeifer, G. P. Relationship between Gene Body DNA Methylation and Intragenic H3K9me3 and H3K36me3 Chromatin Marks. *PLoS One* **6**, e18844 (2011).

81. Vogel, M. J. *et al.* Human heterochromatin proteins form large domains containing KRAB-ZNF genes. *Genome Res* **16**, 1493–1504 (2006).
82. Mosch, K., Franz, H., Soeroes, S., Singh, P. B. & Fischle, W. HP1 Recruits Activity-Dependent Neuroprotective Protein to H3K9me3 Marked Pericentromeric Heterochromatin for Silencing of Major Satellite Repeats. *PLoS One* **6**, e15894 (2011).
83. Liu, S. *et al.* Setdb1 is required for germline development and silencing of H3K9me3-marked endogenous retroviruses in primordial germ cells. *Genes Dev* **28**, 2041–2055 (2014).
84. Bulut-Karslioglu, A. *et al.* Suv39h-Dependent H3K9me3 Marks Intact Retrotransposons and Silences LINE Elements in Mouse Embryonic Stem Cells. *Mol Cell* **55**, 277–290 (2014).
85. Gruenbaum, Y. & Foisner, R. Lamins: Nuclear Intermediate Filament Proteins with Fundamental Functions in Nuclear Mechanics and Genome Regulation, **84**, 131–164 (2015).
86. Donnaloja, F., Carnevali, F., Jacchetti, E. & Raimondi, M. T. Lamin A/C Mechanotransduction in Laminopathies. *Cells 2020, Vol. 9, Page 1306* **9**, 1306 (2020).
87. Corallo, D., Trapani, V. & Bonaldo, P. The notochord: structure and functions. *Cellular and Molecular Life Sciences 2015 72:16* **72**, 2989–3008 (2015).
88. Park, M., Patel, N., Keung, A. J. & Khalil, A. S. Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. *Cell* **176**, 227-238. e20 (2019).
89. Hou, W., Ji, Z., Ji, H. & Hicks, S. C. A systematic evaluation of single-cell RNA-sequencing imputation methods. *Genome Biol* **21**, 1–30 (2020).
90. Aleström, P. *et al.* Zebrafish: Housing and husbandry recommendations, **54**, 213–224 (2019).
91. Westerfield, M. *The Zebrafish Book : A Guide for the Laboratory Use of Zebrafish, 4th Edition*. (University of Oregon Press, Eugene, 2000).
92. Collas, P. A Chromatin Immunoprecipitation Protocol for Small Cell Numbers. *Methods in Molecular Biology* **791**, 179–193 (2011).
93. Nishimura, K., Fukagawa, T., Takisawa, H., Kakimoto, T. & Kanemaki, M. An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nature Methods 2009 6:12* **6**, 917–922 (2009).
94. Kubota, T., Nishimura, K., Kanemaki, M. T. & Donaldson, A. D. The Elg1 Replication Factor C-like Complex Functions in PCNA Unloading during DNA Replication. *Mol Cell* **50**, 273–280 (2013).
95. Amendola, M., Venneri, M. A., Biffi, A., Vigna, E. & Naldini, L. Coordinate dual-gene transgenesis by lentiviral vectors carrying synthetic bidirectional promoters. *Nature Biotechnology 2004 23:1* **23**, 108–116 (2005).
96. Meuleman, W. *et al.* Constitutive nuclear lamina–genome interactions are highly conserved and associated with A/T-rich sequence. *Genome Res* **23**, 270–280 (2013).
97. Peric-Hupkes, D. *et al.* Molecular Maps of the Reorganization of Genome-Nuclear Lamina Interactions during Differentiation. *Mol Cell* **38**, 603–613 (2010).
98. Waldo, G. S., Standish, B. M., Berendzen, J. & Terwilliger, T. C. Rapid protein-folding assay using green fluorescent protein. *Nature Biotechnology 1999 17:7* **17**, 691–695 (1999).
99. Bird, R. E. *et al.* Single-Chain Antigen-Binding Proteins. *Science (1979)* **242**, 423–426 (1988).
100. Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: Property, design and functionality. *Adv Drug Deliv Rev* **65**, 1357–1369 (2013).
101. de Luca, K. L. & Kind, J. Single-cell damid to capture contacts between dna and the nuclear lamina in individual mammalian cells. in *Methods in Molecular Biology* vol. 2157 159–172 (Humana, New York, NY, 2021).
102. Zeng, H. *et al.* An Inducible and Reversible Mouse Genetic Rescue System. *PLoS Genet* **4**, e1000069 (2008).
103. Nora, E. P. *et al.* Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. *Cell* **169**, 930-944. e22 (2017).
104. Hashimshony, T. *et al.* CEL-Seq2: Sensitive highly-multiplexed single-cell RNA-Seq. *Genome Biol* **17**, 1–7 (2016).
105. Chandra, T. *et al.* Independence of Repressive Histone Marks and Chromatin Compaction during Senescent Heterochromatic Layer Formation. *Mol Cell* **47**, 203–214 (2012).
106. Langmead, B. & Salzberg, S. L. Fast gappedread alignment with Bowtie 2. *Nature Methods 2012 9:4* **9**, 357–359 (2012).
107. Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome Biology 2013 14:4* **14**, 1–13 (2013).

108. Lawson, N. D. *et al.* An improved zebrafish transcriptome annotation for sensitive and comprehensive detection of cell type-specific genes. *Elife* **9**, 1–76 (2020).

109. Anders, S., Pyl, P. T. & Huber, W. HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166–169 (2015).

110. Ramírez, F. *et al.* deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* **44**, W160–W165 (2016).

111. Harmanci, A., Rozowsky, J. & Gerstein, M. MUSIC: identification of enriched regions in ChIP-Seq experiments using a mappability-corrected multiscale signal processing framework. *Genome Biol* **15**, 474 (2014).

112. Zhang, Y. *et al.* Model-based analysis of ChIPSeq (MACS). *Genome Biol* **9**, 1–9 (2008).

113. Li, J., Witten, D. M., Johnstone, I. M. & Tibshirani, R. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* **13**, 523–538 (2012).

114. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).

115. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat Methods* **16**, 1289–1296 (2019).

116. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* **20**, (2019).

Chapter 4

Retrospective and multi-factorial profiling resolves the order of chromatin dynamics in single cells

Samy Kefalopoulou[1,2]*, P.M.J. Rullens[1,2,3]*, Kim L. de Luca[1,2], Sandra S. de Vries[1,2], Tessy Korthout[1,2], Alexander van Oudenaarden[1,2], Peter Zeller[1,2,4]‡, Jop Kind[1,2,3]‡

[1] Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht; Utrecht, The Netherlands

[2] Oncode Institute; The Netherlands

[3] Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University; Nijmegen, The Netherlands

[4] Department of Molecular Biology and Genetics, Aarhus University; Aarhus, Denmark

* Co-first authorship

‡ Co-corresponding authorship

Revised manuscript accepted in ***Nature Cell Biology***, in press (2025).

Abstract

The regulation of gene expression is governed at multiple levels of chromatin organization. However, how gene regulation is coordinated remains relatively unexplored. Here we develop Dam&ChIC, a method that enables retrospective and multifactorial chromatin profiling in single cells. Dam&ChIC employs *in vivo* chromatin labeling with ^{m6}A to acquire a past chromatin state, coupled with an antibody-mediated readout to capture the present chromatin state. Analyses of diverse chromatin combinations highlight its versatility and superior resolution. By tracking LAD inheritance over the cell cycle, we showcase that Dam&ChIC provides retrospective single-cell chromatin data. Importantly, by applying Dam&ChIC in random X chromosome inactivation, we disentangle the temporal order of chromatin remodeling events. Our data reveals that upon mitotic exit, the inactive X chromosome undergoes extensive genome–lamina detachment, preceding H3K27me3 spreading. We anticipate that Dam&ChIC will be instrumental in unraveling the interconnectivity and order of gene-regulatory events underlying cell-state changes during development.

Introduction

Gene regulation and the establishment of cellular specificity involves the coordinated activity of processes that take place on different levels of chromatin organization. Such processes include, among others, post-translational modification of histone tails (hPTMs)[1], the three-dimensional organization of chromatin[2] and interactions of chromatin with nuclear scaffolds like the nucleolus and the nuclear lamina (NL)[3].

Recent advances in molecular technologies have been instrumental for profiling various features of chromatin organization with single-cell precision[4-11]. These technologies highlighted that the principles of chromatin organization in single cells adhere to findings previously identified by population-based studies, however, for certain chromatin features, cell-to-cell variability and dynamics are more extensive than previously anticipated.

While methods to measure single-cell chromatin organization are very valuable to reveal heterogeneity, they preclude obtaining insights into the interconnectivity between different chromatin features. To address this, recently several methods were developed to measure multiple chromatin features from the same cell. Experimental approaches such as multi-CUT&Tag[12], MulTI-Tag[13], nano-CT[14], NTT-seq[15] and scMAbID[16] all recover multifactorial chromatin information by using mixtures of target-specific antibodies and in situ tagging of the chromatin. The information provided by these methods hold the promise to study cooperativity and the order of gene-regulatory events in the same cell. However, at the same time, the sparsity of the information that most of these technologies provide, makes it challenging to analyze the interplay between chromatin features in single cells. As a result, our understanding of how the different levels of chromatin organization interact and control gene expression in an orchestrated manner remains limited.

Furthermore, the variability and dynamics in chromatin states observed between single cells emphasize the need for tools that can record chromatin state transitions retrospectively in the same cell. Such technologies would provide unique insights into the role of past chromatin states in determining present cellular outcomes. This is not feasible with in situ tagging approaches because these only capture snapshots of chromatin states, right at the time of harvest. Instead, a molecular recording system that maps chromatin features over time in live cells, has the potential to unveil past chromatin states. Molecular recording is inherent to a few technologies that are based on expression of exogenous bacterial methyltransferases in mammalian cells. When fused to endogenous DNA-interacting proteins, these enzymes can label the genome in live cells, by depositing methylation marks in the proximity of the genomic sites of interaction. These labels represent molecular footprints of past

protein DNA-interactions that can be retrieved upon sequencing. This principle was recently illustrated by a technology that is based on the expression of the bacterial cytosine methyltransferase DCM which enabled recordings of gene and enhancer activity during cellular differentiation[17]. However, since this method was not developed at the single-cell level, it is limited to profiling chromatin-state transitions in populations of cells. Therefore, the need to delineate chromatin dynamics that occur within the same cell over time still persists.

A high-resolution single-cell technology that records epigenomic information over a controlled period of time in live cells is (sc)DamID[7,18-22]. DamID involves the fusion of a DNA-binding protein of interest (POI), or an hPTM reader domain, with the *E. coli* methyltransferase Dam. Conditional expression of the fusion protein in live cells enables the deposition of adenine-6-methylation (^{m6}A) in the proximity of the POI or hPTM. Consequently, protein-DNA contacts or hPTM occupancy are directly recorded on the DNA and remain stable until the DNA is replicated. DamID thereby has the promise to directly measure past chromatin states of individual cells, within a defined window of time. However, information by DamID alone is not sufficient to disentangle past and present chromatin states within the same cell.

We reasoned that obtaining both past and present chromatin states would be feasible if DamID[7,20] is applied in single cells together with an antibody-based in situ readout, that provides snapshot information on chromatin states at the time of cell harvest.
An ideal candidate technology for this purpose is sortChIC[11], which recovers high-resolution single-cell chromatin profiles through in situ tethering of the non-specific micrococcal nuclease (MNase) fused to protein A (pA-MNase), to specifically digest and amplify antibody-bound chromatin[23,24]. Building upon these two approaches, here we developed a new versatile method, Dam&ChIC. Our technology combines recording of chromatin states in live cells and antibody-directed chromatin digestion, with a twofold objective: first, to recover multifactorial chromatin information at high resolution to dissect the interplay between chromatin states in single cells; second, to uncover past and present chromatin states in the same individual cell to reveal state transitions along dynamic processes.

We demonstrate both objectives of Dam&ChIC through comprehensive benchmarking and implementation in two distinct biological systems. First, we show that Dam&ChIC recovers high-quality multifactorial chromatin measurements. Additionally, we provide the proof-of-concept of retrospective chromatin measurements by using Dam&ChIC to study the reorganization of lamina-associated domains (LADs) during cell division. With this, we identify that LAD inheritance over mitosis is domain-size dependent. Secondly, we employ Dam&ChIC in X-chromosome inactivation (XCI), an essential part of dosage compensation in mammals that depends on

heterochromatin formation. We uncover an unanticipated, but extensive, loss of genome–lamina interactions across the Xi that occurs prior to accumulation of the Polycomb marks H2AK119ub and H3K27me3, but after expression of the XCI master regulator Xist. Finally, by mapping past and present genome–lamina interactions over differentiation, we show that the release of the Xi from the nuclear lamina occurs upon the exit from mitosis. Collectively, we show that Dam&ChIC provides a new experimental and analytical framework that enables studying the order of chromatin regulatory events associated with cell-state transitions in single cells

Results

Method design

We developed Dam&ChIC by expanding on the methodologies of DamID and sortChIC. While both technologies produce high-resolution single-cell chromatin measurements, their molecular principles and methodologies are fundamentally different. In particular, DamID requires creating a cell line that conditionally expresses an enzymatic fusion of Dam with a protein of interest (POI) to deposit exogenous ^{m6}A over time on proximal genomic GATC motifs. In contrast, sortChIC involves targeting an antibody in fixed cells and subsequent tethering of pA-MNase to capture a static end state.

Based on these principles, we consolidated the optimal series of molecular reactions to yield high-resolution measurements from both modalities in Dam&ChIC. First, Dam expression is induced during a desired time window in living cells (Fig. 4.1a, step 1). Cells are subsequently harvested, permeabilized and stained with an antibody of interest (Fig. 4.1a, step 2). Afterwards, single nuclei are sorted in 384-well plates via fluorescence-activated cell sorting (FACS) (Fig. 4.1a, step 3) and further processed with robotic liquid handling (see Methods). The molecular processing includes activation of pA-MNase (Fig. 4.1a, steps 4-5), blunt-ending of the pA-MNase-produced fragments (Fig. 4.1a, step 6), DpnI digestion to specifically enrich for genomic fragments containing ^{m6}A-marked GATC motifs (Fig. 4.1a, step 7), and ligation of both types of fragments with blunt-end forked adapters (Fig. 4.1a, step 8). The adapter design includes a T7 promoter for *in vitro* transcription (IVT), unique molecular identifiers (UMI), the Illumina P5 sequence and cell-specific barcodes[20], allowing linear amplification of the produced fragments and illumina library preparation (Fig. 4.1a, step 9, Fig. S4.1a). Upon high-throughput sequencing, the unique sequence context of DamID and ChIC reads is leveraged to separate the pool of fragments in silico (Fig. 4.1a, step 10, see Methods). Specifically, DamID reads almost exclusively align to genomic GATC motifs, contrary to ChIC-derived reads that do not have any motif specificity (Fig. S4.1b).

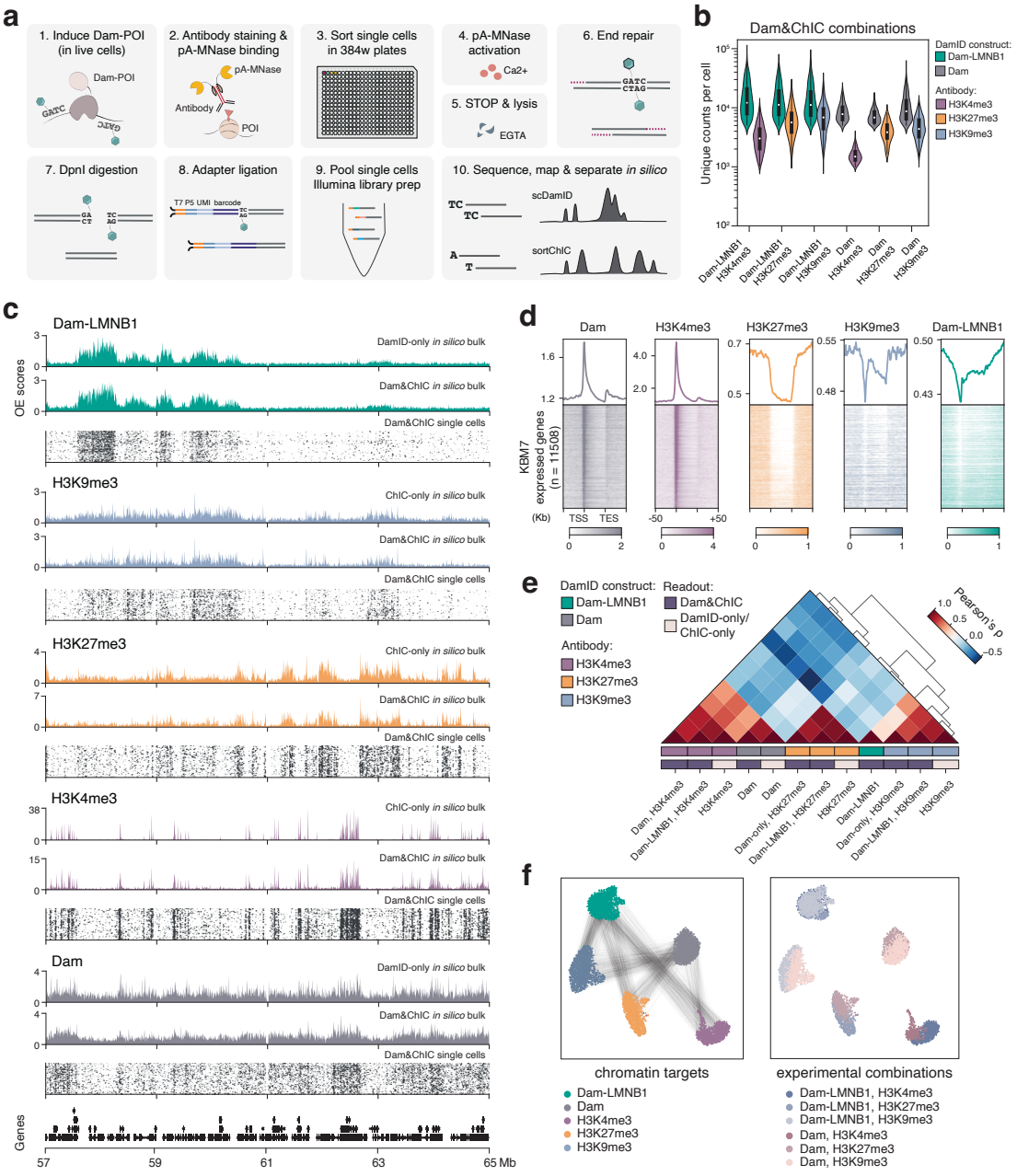


Fig. 4.1. Dam&ChIC enables high resolution joint chromatin profiling in single cells.

a, Graphical overview of the single-cell Dam&ChIC method. **b**, Violin plots depicting the unique number of reads obtained per cell by Dam&ChIC for different combinations of Dam constructs and antibodies. **c**, Single-cell heatmaps (n = 463-1740 cells) of each chromatin feature profiled with Dam&ChIC and their in silico bulk profiles, shown as observed over expected (OE) values. Additionally, corresponding in silico bulk profiles of ChIC-only datasets in KBM7 cells are shown for comparison. *[cont]*

For the untethered Dam, the in silico bulk of a corresponding publicly available DamID-only dataset in KBM7 cells is used for comparison. **d**, Heatmaps showing chromatin features profiled with Dam&ChIC aligned on genes (n = 11508) expressed in KBM7 cells. Genes (rows) are ordered on their relative expression levels, determined by publicly available RNA-seq data. Profiles above show the scaled averages of the heatmaps for each mapped chromatin feature. **e**, Hierarchical clustering depicting genome-wide Pearson correlations relating all pseudobulk Dam&ChIC chromatin data and corresponding ChIC-only datasets, or a publicly available DamID-only dataset. Data was binned in 1Kb bins. Data labelled as "Dam" or "Dam-LMNB1" with the Dam&ChIC color label, contain the respective DamID readouts from all experimental combinations with hPTMs. **f**, UMAP representations of Dam&ChIC data binned in 100-Kb genomic bins. Each cell (dot) is represented twice, once for each measurement. Left; colored by chromatin target, right; colored by experimental combination. Black lines in the left UMAP connect the same cell.

Due to the intrinsic preference of pA-MNase for A/T-rich genomic regions[25], more than 95% of ChIC reads start with either an A or T nucleotide (Fig. S4.1c,d), which we use to achieve more confident read separation (see Methods). Importantly, the number of reads recovered for both modalities in Dam&ChIC is comparable to control datasets, in which either method is performed individually (control datasets are henceforth referred to as DamID-only and ChIC-only) (Fig. S4.1e). In the DamID-only and ChIC-only libraries, the vast majority of reads are correctly assigned to the respective readout, separating target from off-target signal by at least two orders of magnitude (Fig. S4.1e). Therefore, the Dam&ChIC protocol yields sequencing libraries containing both DamID- and ChIC-derived fragments, which can be separated in silico based on sequence context.

Dam&ChIC jointly maps diverse chromatin types with high resolution

To benchmark the quality of the integrated protocol, as well as to test its versatility in profiling different chromatin features, we generated single-cell data of diverse combinations of heterochromatic and euchromatic chromatin types. We used two previously established human KBM7 cell lines; one that conditionally expresses Dam tethered to the core nuclear lamina protein LMNB1 and one that conditionally expresses the untethered Dam enzyme. Dam-LMNB1 has been previously used to characterize LADs in single cells[7], while the untethered Dam enzyme has been reported to accurately detect chromatin accessibility in single cells[20].

Importantly, KBM7 cells have a near complete haploid genome, ensuring that both Dam&ChIC measurements originate from the same chromosome copy. For these experiments, we induced expression of Dam-LMNB1 or untethered Dam in live cells for 15 h, to enable ^{m6}A deposition. Thereafter, we stained nuclei with antibodies specific to the hPTMs H3K4me3, H3K27me3 or H3K9me3. For benchmarking, we additionally performed ChIC-only experiments for the same set of hPTMs and used matching DamID-only data, both derived from KBM7 cells.

For Dam-LMMB1 and the untethered Dam, Dam&ChIC recovers a median of ~12,300 and ~8,500 UMI-flattened (unique) reads per cell respectively, in all combinations with hPTMs (Fig. 4.1b). For the euchromatic H3K4me3 it recovers a median of ~2,600 unique reads per cell and for the heterochromatic H3K27me3 and H3K9me3 modifications a median of ~5,000 and ~5,500 unique reads per cell, respectively (Fig. 4.1b). Dam&ChIC thereby attains higher sensitivity compared to other recently published multifactorial chromatin profiling methods, particularly considering the haploid genome of KBM7 cells (Fig. S4.1f). In order to normalize the two distinct fragment types present in the Dam&ChIC libraries, we computed observed over expected (OE) scores[7]. For the DamID readout, OE scores are calculated over the in silico genomic distribution of GATC motifs (Fig. S4.1g, see Methods). For the ChIC readout, the distribution of maximum expected reads was generated with a bulk ChIC-only experiment against H3 (Fig. S4.1g, see Methods).

We directly compare Dam&ChIC to ChIC-only and DamID-only data, at the resolution of individual genes and lamina-associated domains (LADs). This confirms highly specific single-cell genomic enrichment at the expected regions (Fig. 4.1c,d, Fig. S4.1h), further corroborated by a comparison against matching ENCODE bulk data of the closely related K562 cell line (Fig. S4.1i,j). Moreover, a genome-wide comparison shows strong correlation between Dam&ChIC and the corresponding ChIC-only and DamID-only datasets (Fig. 4.1e). Finally, dimensionality reduction of the entire Dam&ChIC dataset reveals that the different single-cell readouts consistently separate based on chromatin type (Fig. 4.1f, left), regardless of the used experimental combinations (Fig. 4.1f, right). Altogether, these results demonstrate that Dam&ChIC is a versatile multifactorial method for profiling diverse euchromatic and heterochromatic factors in single cells with high sensitivity and specificity.

Dam&ChIC disentangles past and present genome–lamina interactions in the same cell

Using a fluorescent tracker to follow Dam-labeled DNA by microscopy (^{m6}A-Tracer) we previously reported that, during interphase, LADs display constrained localization dynamics within a 1-µm zone underneath the nuclear lamina[26]. These findings indicate that over a 15-h time window, LADs are very dynamic, and that these dynamics can be efficiently detected with DamID[26]. In contrast, the ChIC readout can only detect LADs that are in contact with the nuclear lamina at the time of cell harvest. To explore the possibility to use the Dam&ChIC method to capture temporal dynamics of LADs during interphase, we used the Dam-LMNB1 KBM7 cell line to induce expression of the Dam-LMNB1 fusion protein for a period of 15 h, during which genome–lamina interactions are directly recorded on the DNA. After this period, the cells were stained with an antibody against LMNB1 to detect current-state genome–lamina interactions with the ChIC readout. Provided that genome–lamina

interactions undergo a degree of dynamics in KBM7 cells, we expect that past interactions will be detected exclusively by DamID (Fig. 4.2a; past interactions), whereas very recently established interactions are only measured with ChIC (Fig. 4.2a; *de novo* interactions).

We obtained high quality single-cell LAD profiles for both Dam&ChIC measurements (Fig. S4.2a,b). We started by examining—on a population level—which LADs are identified by both measurements, or by DamID or ChIC alone. For this purpose, we calculated the contact frequency (CF) metric that quantifies the fraction of cells in which a genomic region is in contact with the nuclear lamina (Fig. S4.2b; see Methods and [7]). We find strong overlap between LAD profiles identified by DamID and ChIC, with only a small minority of LADs detected exclusively by one of both measurements (Fig. S4.2c). Moreover, the CF is comparable between both measurements, albeit somewhat higher for several regions in the DamID measurement (Fig. S4.2c).

We next inspected the combined genome–lamina contact profiles in single cells. Interestingly, visual inspection of Dam&ChIC data indicates that although LADs are generally detected by both methods, we regularly observe LADs that are identified exclusively by DamID and to a lesser extent by ChIC alone in single cells (Fig. 4.2b). This was confirmed by systematic analysis of the fraction of 100-Kb genomic bins that are identified only by DamID, ChIC, or both (Fig. 4.2c,d, Fig. S4.2d,e, see Methods). Notably, the genomic bins that display the largest discrepancy between DamID and ChIC signal are the genomic regions that interact with the nuclear lamina with intermediate CFs which indicates that these regions display the highest dynamics during interphase (Fig. 4.2e). In contrast, regions with higher CFs are more comparably detected by both Dam&ChIC readouts (Fig. 4.2e). This suggests more stable genome–lamina interactions for regions with high CF, typically residing within larger domains with high A/T content[7]. These findings discern variable degrees of dynamics between different regions that interact with the nuclear lamina and corroborate previous observations that larger LADs with high CFs are more stably associated with the nuclear lamina through multivalent interactions over hundreds of kilobases[7].

Given that the current Dam&ChIC dataset in KBM7 cells captures genome–lamina dynamics, we reasoned that the ratio between DamID and ChIC signal in a given cell should contain information on the direction of LAD changes. High ChIC and low DamID signal indicate that a LAD is being established. Conversely, low ChIC and high DamID signal implies that a LAD is being released from the nuclear lamina.

Based on this reasoning, we sought to derive structural principles in LAD establishment and release. RNA velocity[27] and the ensuing chromatin velocity[28] estminate the ratio between two interdependent cellular modalities to extract directional change.

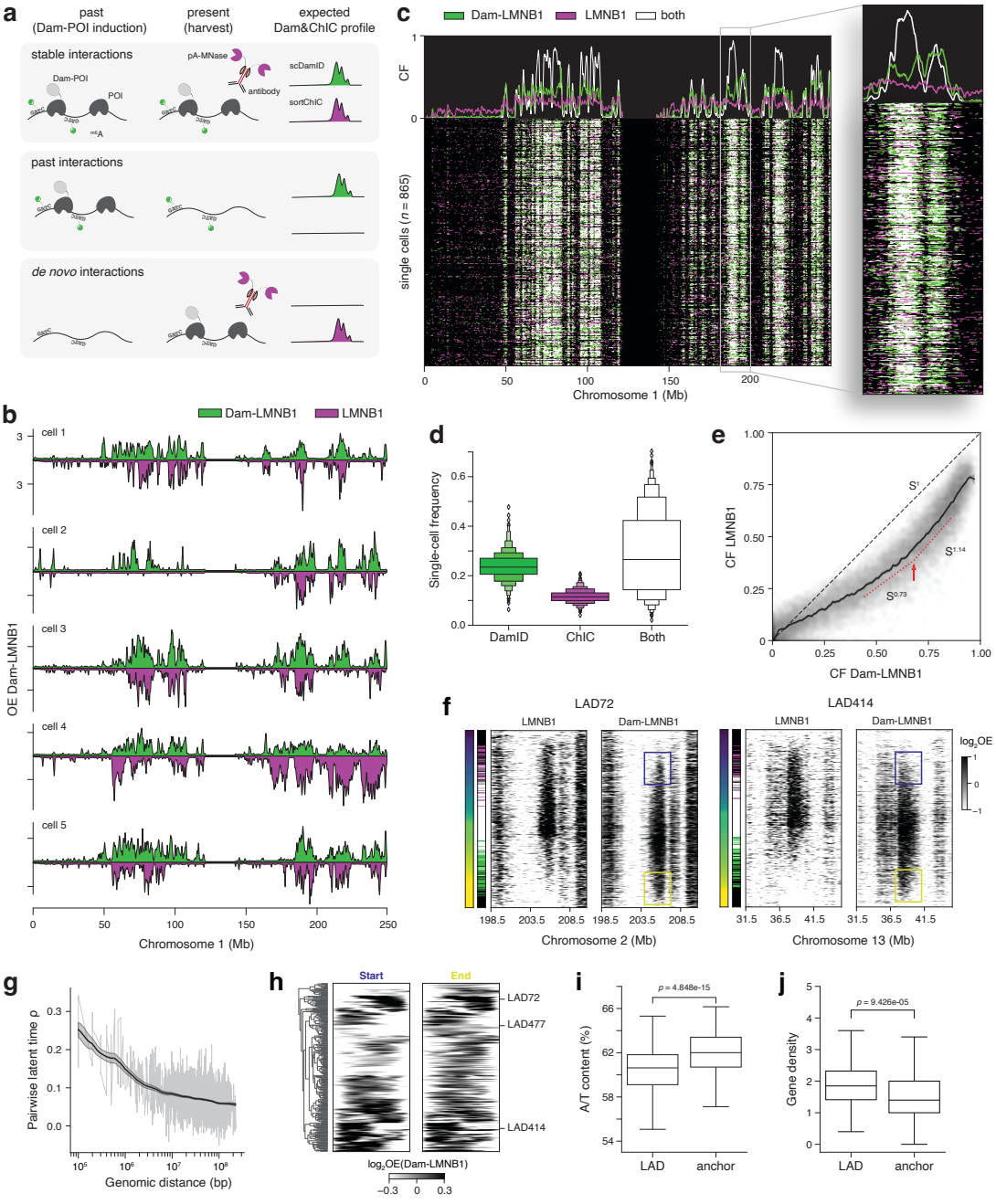


Fig. 4.2. Dam&ChIC disentangles past and present genome–lamina interactions in the same cell.
a, Schematic overview of the chromatin dynamics that Dam&ChIC can resolve between the time of Dam-POI induction (past) and the time of cell harvest for antibody staining (present). Three scenarios are depicted with their corresponding expected Dam&ChIC profiles. **b**, "Mirror plots" of 5 example cells showing quantitative LAD signal, expressed in OE values (see Methods) in 100-Kb genomic bins on chromosome 1, measured by DamID (top, green) and ChIC (bottom, magenta) in the same cell. *[cont]*

c, Combined single-cell heatmap (n = 865) with binarized Dam&ChIC LAD signal (Methods) in 100-Kb genomic bins on chromosome 1. Bins are detected by either DamID (green), ChIC (magenta), or by both measurements (white). Lineplots represent the CF, *i.e.,* the fraction of cells for which a given bin is detected by DamID, ChIC or both, shown with the respective colors. **d**, Boxenplot showing multiple quantiles of quantification of the frequency by which each LAD is measured in a single cell by DamID, ChIC or both. **e**, Scatter plot comparing the CF of the DamID (x-axis) and ChIC (y-axis) measurement. Black uninterrupted line indicates a radius neighbor regression model that was fitted on the data, with its 95% CI. Diagonal black dashed line indicates complete agreement between both measurements. Slope is quantified at different positions. **f**, Single-cell heatmaps for LAD72 (left) and LAD414 (right), with the DamID (right) and ChIC (left) measurements split, and single cells ordered according to latent time (leftmost bar). The middle bar at the left indicates the LAD state (Methods). LAD signal is shown as OE in 100-Kb genomic bins. Blue and yellow boxes indicate the origin and release of each domain respectively, happening from and towards the same genomic location. **g**, Pairwise Pearson's correlation coefficient of the latent time assignment between any pair of LADs (y-axis), plotted against genomic distance (x-axis). **h**, Hierarchical clustering of latent time averaged start (left) and end (right) of LADs at which they emerge or disappear. Latent time averages are from 50 cells up and downstream of start or end, as depicted in blue and yellow boxes, respectively, in **f**. LAD signal is shown as OE in 100 Kb. **i-j,** Boxplot showing **(i)** A/T content and **(j)** gene density quantified in 100-Kb genomic bins on anchor regions (*i.e.,* LAD region where LAD starts attaching and is released; Methods) compared to the rest of the LAD. Two-sided t-test with resulting p-values are designated.

We apply this principle to pseudo-order genome–lamina interactions in Dam&ChIC data, based on the ratio between DamID and ChIC (Fig. S4.2f). This results in a LAD-specific latent time[29], that orders each LAD according to its state of establishment and release over time (Fig. 4.2f, Fig. S4.2f,g). In agreement with previous findings that LADs show coordinated nuclear lamina association along the linear chromosome[7], we find that nearby LADs progress coordinately, with a decay in their coordination according to linear intra-chromosomal distance (Fig. 4.2f,g). Interestingly, the pseudo-ordering based on latent time allows us to observe polarity in the establishment of some LADs; one border of the domain is visible first and defines from where it will start expanding towards the opposing border (Fig. 4.2f, blue boxes), while release of the domain mirrors the pattern of its establishment (Fig. 4.2f, yellow boxes). In a systematic analysis, we examined the genomic position of establishment and release for all LADs, and observed that LADs can establish and release either from the center or the border of the domain (Fig. 4.2h). These findings suggest that LADs may have "anchor" regions for their establishment and release. To test if these anchor regions bear certain (epi)genomic characteristics, we examined the genomic sequence and enrichment of chromatin marks at anchor regions and compared that to the rest of the domain. Interestingly, anchor regions have enriched A/T content (Fig. 4.2i), harbor fewer genes (Fig. 4.2j), and are depleted from hPTMs H3K4me3 (Fig. S4.2h) and H3K27me3 (Fig. S4.2i). However, H3K9me3 levels seem to be equally enriched in anchors compared to the rest of the LAD (Fig. S4.2j). These suggest that anchors have a distinct (epi)genomic character compared to non-anchor regions in LADs.

Collectively, we demonstrate that Dam&ChIC disentangles past and present chromatin states in the same cell, allowing us to reinforce and extend upon previous findings on genome–lamina interaction dynamics[26]. Our analyses reveal that (i) the degree of LAD dynamics in interphase is linked to domain size and (ii) the attachment or detachment of a LAD from the nuclear lamina initiates at anchor sites with a distinct (epi)genomic signature.

Retrospective measurements with Dam&ChIC show that spatial positioning of the genome is partially inherited upon mitosis

Encouraged by the observations that Dam&ChIC disentangles past from present chromatin states, we set out to implement it to study the inheritance of spatial genome organization over a cell division. Previous work described differences in genome–lamina interactions between mother and daughter cells, as a result of stochastic re-shuffling of LADs right after mitosis[26]. These observations were made by microscopy using the ^{m6}A-Tracer, which enabled imaging of LAD inheritance, but left sequence identity and the underlying patterns of genome–lamina reorganization elusive. In order to address this in a genome-wide manner, we sought to leverage Dam&ChIC to measure retrospectively genome–lamina interactions occurring before and after mitosis in the same cell.

To this end, we synchronized KBM7 cells in G1/S using a double thymidine block and expressed Dam-LMNB1 concurrently with the last thymidine incubation (Fig. 4.3a and Methods). We harvested cells in G2 phase and early G1 phase (Fig. S4.3a and Methods) and utilized a multiplexing sorting strategy11 to enable parallel processing of different conditions with Dam&ChIC (Fig. S4.3b and Methods). This experimental setup ensures that Dam-LMNB1 is still present during S and G2 in order to label and record G2 LADs (Fig. S4.3c). Moreover, with our strategy to harvest cells as early as possible in G1 (Fig. S4.3a and Methods), during a time window that is insufficient for accumulation of ^{m6}A after mitosis[26], we prevent measurements of *de novo* G1 LADs in our data. Therefore, in early G1 cells, LADs that are faithfully inherited through mitosis would be detected both by DamID and ChIC (Fig. 4.3b, cell 1, LADs a, b, c). In contrast, LADs that reposition away from the nuclear lamina after mitosis would be exclusively recovered by DamID (Fig. 4.3b, cell 2, LADs a, c), whereas *de novo* established LADs would be exclusively recovered by ChIC (Fig. 4.3b, cell 2, domain d).

We first confirmed that cells in G2 display typical LAD profiles (Fig. S4.3c), verifying that our synchronization and induction approach does not introduce adverse effects on genome–lamina interactions. We then computed the Z-score normalized Jaccard index, to measure the pairwise similarity between DamID and ChIC measurements (see Methods) for three pairs: (i) between DamID and ChIC of the same cell in G2, (ii) between DamID and ChIC of the same cell in early G1 (that is, DamID patterns

of G2 before cell division, and ChIC of early G1 after cell division) or (iii) between DamID and ChIC of random cells in G1. We observed a significant decrease in similarity between DamID and ChIC LAD profiles in G1 compared to G2 cells, yet the similarity in G1 cells is considerably higher than what is observed when comparing different cells (Fig. 4.3c). This suggests that LADs are partially inherited over mitosis, corroborating previous microscopy findings[26].

Interestingly, we can accurately detect three scenarios: inherited LADs in G1 (Fig. 4.3d; white), *de novo* established LADs in G1 (Fig. 4.3d; purple) and non-inherited LADs, unique to G2 cells (Fig. 4.3d; green). The occurrence of these scenarios differs extensively, with some domains showing considerable heterogeneity in inheritance across cells (Fig. 4.3c; left domain), while others are stably preserved across generations (Fig. 4.3d; right domain).

We next wondered if some LADs have a higher propensity to be inherited through mitosis than others. Interestingly, we find that size and A/T content of LADs positively correlate with their likelihood of inheritance (Fig. 4.3e, leftmost panels), which is further corroborated by an increased similarity in the G1 population for large LAD sizes, compared to small LAD sizes (Fig. S4.3d). H3K27me3 enrichment and gene content show a negative correlation with inheritance (Fig. 4.3e, rightmost panels), while H3K9me3, H3K4me3, SINE and LINE content show a weak negative correlation or no correlation with LAD inheritance (Fig. S4.3e). This data indicates that larger, A/T-rich, gene-poor and H3K27me3-poor LADs tend to more faithfully return to the nuclear periphery after their release in mitosis.

In summary, this dataset shows the unique ability of Dam&ChIC to capture chromatin transitions over a cell division, providing an experimental and analytical framework to obtain insights into epigenetic inheritance. Together with our observations above on LAD dynamics during interphase (Figure 2), the results we obtain here underscore that both LAD size and A/T content are predictive factors for genome–lamina interaction stability across the cell cycle.

Dam&ChIC reveals widespread remodeling of genome–lamina interactions during X-chromosome inactivation

We established that Dam&ChIC, as a multifactorial method, can profile diverse chromatin features at high single-cell resolution. This ability makes Dam&ChIC an ideal approach to study the interplay between chromatin features in dynamic cellular settings that undergo multiple chromatin transitions over time. An essential process during mammalian development that entails dynamics of multiple chromatin features, is X-chromosome inactivation (XCI).

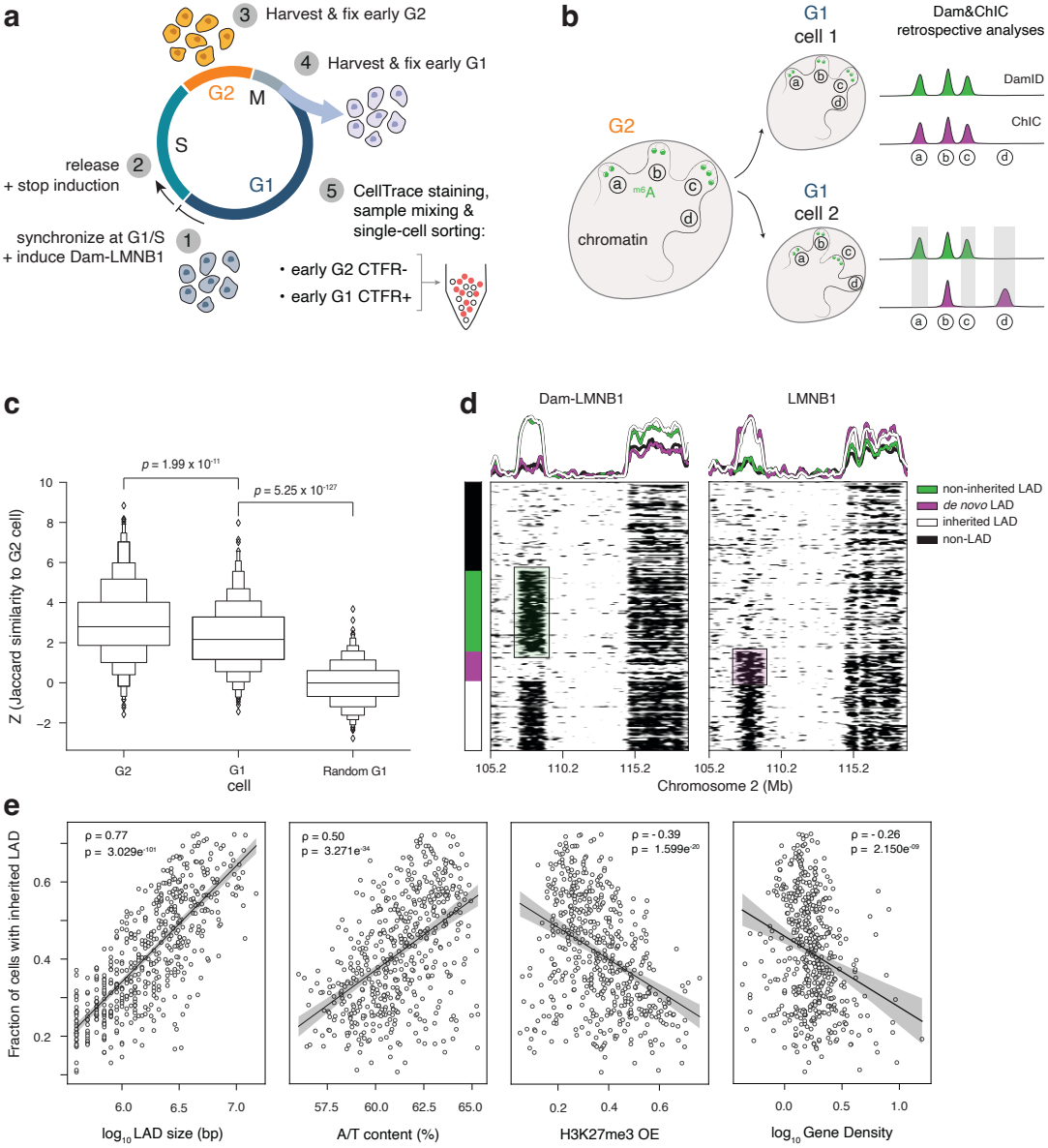


Fig. 4.3. Retrospective measurements with Dam&ChIC show that spatial positioning of the genome is partially inherited upon mitosis.

a, Schematic of experimental setup for the mitotic inheritance experiment (see Methods). **b**, Schematic of the expected retrospective measurements when using Dam&ChIC to follow LADs over mitosis. LAD b is inherited to cell 2 and is expected to be captured in G1-phase by both DamID and ChIC. LADs a and c are non-inherited to cell 2 and are expected to be recovered only by DamID, while the *de novo* LAD d of cell 2 is expected to be recovered only by ChIC in G1 phase cells. *[cont]*

c, Boxenplot of multiple quantiles of Jaccard similarity between DamID of G2 cell and ChIC of the same G2 cell (left), descending G1 cell (middle) or random G1 cell (right). Two-sided t-test with resulting p-values are designated. **d**, Single-cell heatmaps of DamID and ChIC measurements of Dam&ChIC, showing two LADs that are either weakly inherited (left) or faithfully inherited (right). Cells are grouped according to the recovery of the left LAD by DamID and/or ChIC as: non-inherited, *de novo*, inherited, or non-LAD. **e**, Scatterplots for the recovery of frequency of inherited LADs (n = 516) against different (epi)genomic features. Y-axis is the same across all plots. The black lines represent a linear regression fit with its 95% CI in the shaded area. The Pearson Correlation coefficient (ρ) and p-value of two-sided testing for non-correlation are indicated.

XCI is specific to female cells to ensure dosage compensation of X-linked genes between males (XY) and females (XX)[30,31], through the formation of transcriptionally repressive heterochromatin on the inactive X chromosome (Xi)[32]. Heterochromatin formation on the Xi is regulated primarily by the long non-coding RNA *Xist*[33-35] and subsequent deposition and spreading of Polycomb hPTMs H2AK119ub and H3K27me3[36-39]. While remodeling of the hPTM landscape on the Xi has been under continuous investigation, the nuclear positioning of the Xi remains unclear. Previous studies suggest that the Xi associates with different nuclear compartments, such as the nucleolus or the nuclear lamina (NL)[40-45]. Whether this interaction is dynamic and how it is linked to other hPTMs during XCI has remained ambiguous. Moreover, the random nature of XCI has forced population-based genomic studies to work with genetically engineered hybrid systems that predetermine which of the two X alleles will be inactivated[46], making it challenging to decouple observations on XCI from potential biases introduced by the genetic background.

With Dam&ChIC, we aimed to map the interactions of the Xi with the nuclear lamina and dissect the interplay with hPTMs H3K27me3, H2AK119ub and H3K9me3. To this end, we induced random XCI *in vitro*, by differentiating our two previously-derived female hybrid (CAST/Eij x 129/Sv) embryonic stem cell (ESC) lines expressing Dam-LMNB1 and Dam-scFv-H3K27me3[19,47] (Fig. 4.4a). We used an established *in vitro* differentiation protocol that involves Vitamin C treatment on a monolayer of cells and attains asynchronous but robust inactivation within a few days[48,49]. We performed Dam&ChIC at multiple timepoints between day 0 and 6 of differentiation, using the same multiplexing sorting strategy described earlier (see Methods).

We compiled dual measurements of Dam-LMNB1 / H3K27me3 (n = 1656 cells), Dam-LMNB1 / H3K9me3 (n = 573 cells), Dam-LMNB1 / H2AK119ub (n = 1213 cells), Dam-scFv-H3K27me3 / LMNB1 (n = 921 cells), Dam-scFv-H3K27me3 / H3K9me3 (n = 497 cells) and Dam-scFv-H3K27me3 / H2AK119ub (n = 1404 cells) (Fig. 4.4a).

Upon allelically resolving the Dam&ChIC data based on single nucleotide polymorphisms (see Methods), we consistently detect a high number of unique counts per allele for all measurements (Fig. S4.4a), providing ample information to interrogate the interplay between the chromatin features of interest.

First, we examined the Dam-LMNB1 / H3K27me3 dataset. Apart from dissecting its interplay with genome–lamina interactions, the H3K27me3 readout can serve two additional purposes: (i) as a means to estimate the differentiation progression, and (ii) as a proxy for the progression of XCI. We used H3K27me3 enrichment on autosomal genes to visualize cells by UMAP, and observe gradual transitioning of cells over differentiation time (Fig. 4.4b). The gradual accumulation of H3K27me3 on the Hox cluster locus, residing on chromosome 6, is further indicative of chromatin remodelling during the differentiation trajectory (Fig. S4.4b). Additionally, we observe strong H3K27me3 enrichment either on the CAST/Eij or 129/Sv X-chromosome allele over the course of differentiation (Fig. 4.4c), which we subsequently use to categorize cells as CAST/Eij-inactive, 129/Sv-inactive or cells that have yet to undergo XCI (Fig. S4.4c,d). The number of cells that undergo XCI increases with each day of differentiation, with a clear preference for the 129/Sv allele (Fig. 4.4c, Fig. S4.4c).

We next compare the genome–lamina interactions of the X chromosome across the H3K27me3-derived categories (*i.e.*, no XCI, CAST/Eij inactive or 129/Sv inactive). In contrast to the suggested perinuclear localization of the Xi[42], we find widespread loss of interactions between the Xi and the nuclear lamina (Fig. 4.4d, Fig. S4.4e). This is not the case for its active counterpart (Xa), which retains genome–lamina interactions similar to cells that did not undergo XCI (Fig. 4.4d). Reassuringly, we find the exact same patterns in the inverse experiment where H3K27me3 is measured with DamID and LADs with ChIC, confirming that the observed loss in genome–lamina interactions is unlikely a technical artifact (Fig. S4.4f-h). The loss of interactions between the Xi and the nuclear lamina extends through megabase-sized regions across the entire X chromosome and is interspersed with regions that maintain strong interactions (henceforth referred to as “anchor” regions), covering roughly 29% of the chromosome (Fig. 4.4d, bottom bar and Fig. S4.4m). Upon cross-examining the loss of genome–lamina interactions and the increase of H3K27me3 on the Xi across single cells, we find that regions with lost interactions are densely covered with H3K27me3, as opposed to the anchor regions that are largely devoid of this mark (Fig. 4.4e, S4.4i, S4.4m). In addition, we quantitatively corroborate the mutual exclusivity between genome–lamina interactions and H3K27me3 on the Xi, showing that the regions that lose genome–lamina interactions upon XCI specifically gain H3K27me3 (Fig. 4.4f).

We sought to examine the relationship between genome–lamina interactions and H2AK119ub, another hPTM enriched on the Xi. H2AK119ub is deposited by PRC1, with several studies suggesting its deposition is closely followed by that of H3K27me3 by PRC2. We therefore wondered whether H2AK119ub appears mutually exclusive with genome–lamina interactions, similar to H3K27me3. By examining our Dam-LMNB1 / H2AK119ub and Dam-scFv-H3K27me3 / H2AK119ub datasets we find a strong allelic enrichment of H2AK119ub over differentiation, with a similar to H3K27me3 preferential enrichment on the 129/Sv allele (Fig. S4.4j-l). Importantly, single-cell analysis of Dam-LMNB1 profiles with the respective H2AK119ub allelic enrichment show that regions with high H2AK119ub enrichment have lost genome–lamina interactions (Fig. 4.4g), similar to what we observed above for H3K27me3.

Altogether, our data indicate that enrichment of either of the Polycomb-deposited marks H3K27me3 and H2AK119ub coincides with detachment from the nuclear lamina, suggesting that Polycomb heterochromatin is incompatible with genome–lamina interactions during XCI.

Genomic regions with constitutive genome–lamina interactions are generally enriched for the repressive hPTMs H3K9me2/3[18]. We therefore wondered if the resistance of the Xi anchor regions to detachment from the nuclear lamina and H3K27me3 deposition might be related to the presence of H3K9me3. Indeed, Dam&ChIC datasets profiling Dam-LMNB1 / H3K9me3 and Dam-scFv-H3K27me3 / H3K9me3 indicate that H3K9me3 is specifically enriched at the anchor regions of the Xi (Fig. 4.4h), which is consistent with previous observations that H3K27me3 and H3K9me3 occupy distinct non-overlapping domains along the Xi[50]. Further quantification of H3K9me3 levels on all X chromosome regions shows a high enrichment of H3K9me3 in anchors (Fig. S4.4m). H3K9me3 domains preexist in cells that did not yet undergo XCI, where they already demarcate the future anchor regions (Fig. S4.4n). Interestingly, anchors seem to be devoid of Xist spreading, as measured previously by CHART-seq on differentiating mouse ESCs[51] (Fig. 4.4i). Similar to small LADs on autosomes, individual anchor regions are not present in every single cell.

We therefore wondered whether these cell-to-cell differences in anchor usage might be linked to the pre-existing H3K9me3 state. To this end, we quantified the genome–lamina interactions and H3K27me3 deposition at anchors that are either H3K9me3 bound or H3K9me3 unbound across single cells. We find that anchors that are H3K9me3 bound in individual cells are more likely to interact with the nuclear lamina and less likely to be enriched with H3K27me3 (Fig. 4.4j). These findings suggest that lamina-interacting regions on the X chromosome that are devoid of H3K9me3 enrichment are predetermined to detach from the nuclear lamina and accumulate H3K27me3 during XCI.

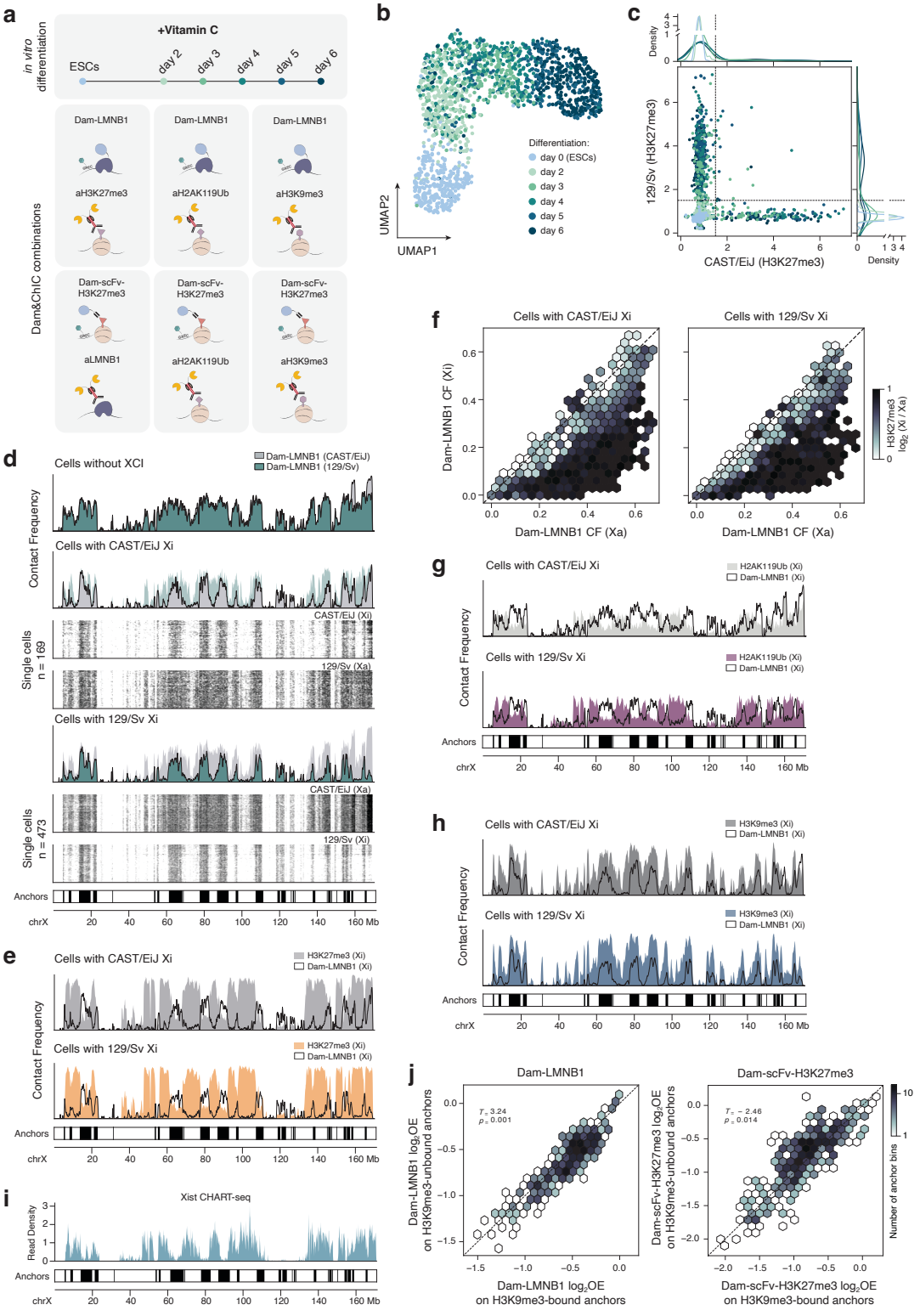

a
in vitro differentiation
+Vitamin C
ESCs
day 2
day 3
day 4
day 5
day 6
Dam&ChIC combinations
Dam-LMNB1
aH3K27me3
aH2AK119Ub
aH3K9me3
Dam-scFv-H3K27me3
aLMNB1
b
UMAP2
UMAP1
Differentiation:
day 0 (ESCs)
day 2
day 3
day 4
day 5
day 6
c
Density
129/Sv (H3K27me3)
CAST/EiJ (H3K27me3)
d
Cells without XCI
Dam-LMNB1 (CAST/EiJ)
Dam-LMNB1 (129/Sv)
Contact Frequency
Cells with CAST/EiJ Xi
Single cells n = 169
CAST/EiJ (Xi)
129/Sv (Xa)
Cells with 129/Sv Xi
CAST/EiJ (Xa)
Single cells n = 473
129/Sv (Xi)
Anchors
chrX
e
Cells with CAST/EiJ Xi
H3K27me3 (Xi)
Dam-LMNB1 (Xi)
Cells with 129/Sv Xi
f
Cells with CAST/EiJ Xi
Cells with 129/Sv Xi
Dam-LMNB1 CF (Xi)
Dam-LMNB1 CF (Xa)
H3K27me3 log2 (Xi / Xa)
g
Cells with CAST/EiJ Xi
H2AK119Ub (Xi)
Dam-LMNB1 (Xi)
Cells with 129/Sv Xi
h
Cells with CAST/EiJ Xi
H3K9me3 (Xi)
Dam-LMNB1 (Xi)
Cells with 129/Sv Xi
i
Xist CHART-seq
Read Density
Anchors
chrX
j
Dam-LMNB1
Dam-scFv-H3K27me3
Dam-LMNB1 log2OE on H3K9me3-unbound anchors
Dam-LMNB1 log2OE on H3K9me3-bound anchors
Dam-scFv-H3K27me3 log2OE on H3K9me3-unbound anchors
Dam-scFv-H3K27me3 log2OE on H3K9me3-bound anchors
Number of anchor bins

Fig. 4.4. Dam&ChIC reveals widespread genome–lamina detachment during X chromosome inactivation.

a, Schematic of the experimental setup with the Vitamin C differentiation trajectory and the various Dam&ChIC combinations. **b**, UMAP (n = 1656 cells) based on H3K27me3 levels on autosomal genes of Dam-LMNB1/H3K27me3-profiled cells, colored by differentiation day. **c**, Scatterplot of allelic X-chromosomal H3K27me3 OE levels per single cell in the Dam-LMNB1 / H3K27me3 dataset, colored by differentiation day. The dashed lines indicate the manually set threshold used to classify cells as "no XCI", "CAST/EiJ inactive", "129S1/Sv inactive" or "Undetermined". **d**, Contact Frequency (CF) values of Dam-LMNB1 on either allele for the cell categories defined in **c**, along the entire X chromosome. The y axes in all CF plots run from 0 to 1. For the categories that undergo XCI, single-cell heatmaps of the active (Xa) and inactive (Xi) X chromosome are shown below, with the respective cell numbers. Plotted cells are spanning all differentiation time points where XCI takes place in the trajectory, based on **c**. Bottom bar shows calls of the anchors, which are defined as LMNB1-retaining and H3K27me3-depleted regions on Xi (Fig. S4.4m, bottom right quadrant). **e**, CF values of Dam-LMNB1 and corresponding H3K27me3 values on either allele in the inactive state for the cell categories defined in **c** along the entire X chromosome. The y axis in all CF plots runs from 0 to 1. Bottom bar shows calls of the anchors. **f**, Hexbin plot of Dam-LMNB1 CF values in cells where the indicated allele is inactivated (left:CAST/EiJ, right:129S1/Sv). For each LAD the average CF is shown for the active (Xa) versus the inactive (Xi) allele Hexbins are colored based on the log-transformed enrichment of H3K27me3 (Xi/Xa). **g**, CF values of Dam-LMNB1 and H2AK119ub on either allele in the inactivated state, as in **e**. **h**, CF values of Dam-LMNB1 and H3K9me3 on either allele in the inactivated state, as in **e** and **g**. **i**, Published Xist CHART-seq data derived from differentiated mouse ESCs (d7)[56] for the X chromosome. The y-axis is expressed as scaled read density values. Bottom bar shows calls of the anchors. **j**, Comparison of Dam-LMNB1 (left) and Dam-scFv-H3K27me3 (right) levels between cells in which a specific nuclear lamina anchor is H3K9me3 bound (x-axis) or unbound (y-axis). Hexbins are colored by anchor abundance. Two-sided t-test with resulting p-values are designated.

In summary, Dam&ChIC provides allelically-resolved single-cell maps of genome–lamina interactions and heterochromatic hPTMs from the same cell. We find that large X-chromosome regions detach from the nuclear lamina during mouse XCI, and instead become strongly enriched for H3K27me3. Regions that are resistant to detachment and H3K27me3 deposition coincide with pre-existing H3K9me3 domains and lack Xist spreading.

Genome-lamina detachment precedes the formation of Polycomb heterochromatin during XCI

The observation that genome–lamina interactions and accumulation of Polycomb marks H3K27me3 and H2AK119ub show a mutually exclusive pattern on the Xi prompted us to examine the temporal order between these seemingly antagonistic events.

To identify the onset of nuclear lamina detachment and H3K27me3 accumulation at single-cell resolution, we first ordered cells along the differentiation trajectory by inferring pseudotime based on the UMAP presented in Fig. 4.4b (Fig. 4.5a). Of note, this UMAP is based solely on H3K27me3 signal on autosomal genes, leaving out the X-linked genes, to avoid skewing the temporal ordering due to H3K27me3 accumulation on the Xi.

The initial loss of genome–lamina interactions appears to coincide with the emergence of H3K27me3 around day 3 of differentiation, further implying their interrelatedness (Fig. 4.5b). We observe a seemingly concurrent onset of both events when visualizing the genomic distribution of genome–lamina interactions and H3K27me3 on the Xi for progressive pseudobulk profiles (Fig. 4.5c) and single cells (Fig. 4.5d). In addition, while H3K27me3 domains progressively expand outwards from sites already present in undifferentiated cells[54], LADs gradually shrink along the Xi (Fig. 4.5c).

To resolve the precise chronology between nuclear lamina detachment and H3K27me3 enrichment, we plot Dam-LMNB1 against H3K27me3 levels on the Xi for each cell in our dataset, and confirm the inverse relationship between H3K27me3 and genome–lamina interactions at the single-cell level (Fig. 4.5e). Interestingly, early time points show a reduction in Dam-LMNB1 levels before H3K27me3 starts to emerge, resulting in a general left-upward mobility of cells over differentiation in the Dam-LMNB1 to H3K27me3 space (Fig. 4.5e, arrow). This observation suggests that the loss of genome–lamina interactions on the Xi initiates before accumulation of H3K27me3. We anticipate that the opposite order of events would result in the temporal appearance of double-positive cells and a general upward-right mobility of cells over differentiation. To corroborate our reasoning, we perform simulations where we control the temporal ordering of the two dynamics (Fig. S4.5a-c), and confirm that the observed left-upward mobility in Fig. 4.5e corresponds to the simulation where loss of genome–lamina interactions occurs prior to H3K27me3 accumulation. Importantly, we identify the same pattern in the inverse Dam&ChIC experiment profiling Dam-scFv-H3K27me3/LMNB1, which detects an even earlier loss of genome–lamina interactions (Fig. 4.5f).

Furthermore, we wondered what the temporal order is between H2AK119ub deposition and loss of genome–lamina interactions of the Xi. We started by examining our Dam-scFv-H3K27me3/H2AK119ub dataset and confirmed that H2AK119ub deposition precedes H3K27me3 (Fig. S4.5d-e), an observation supportive of previous reports that PRC1 activity precedes PRC2 during XCI. However, our Dam-LMNB1 / H2AK119ub dataset shows that H2AK119ub too seems to accumulate after detachment of the Xi from the nuclear lamina (Fig. 4.5g). Taken together, this data suggests that both Polycomb hPTMs H2AK119ub and H3K27me3 start accumulating on the Xi after initiation of the detachment, but these processes continue to consolidate over time.

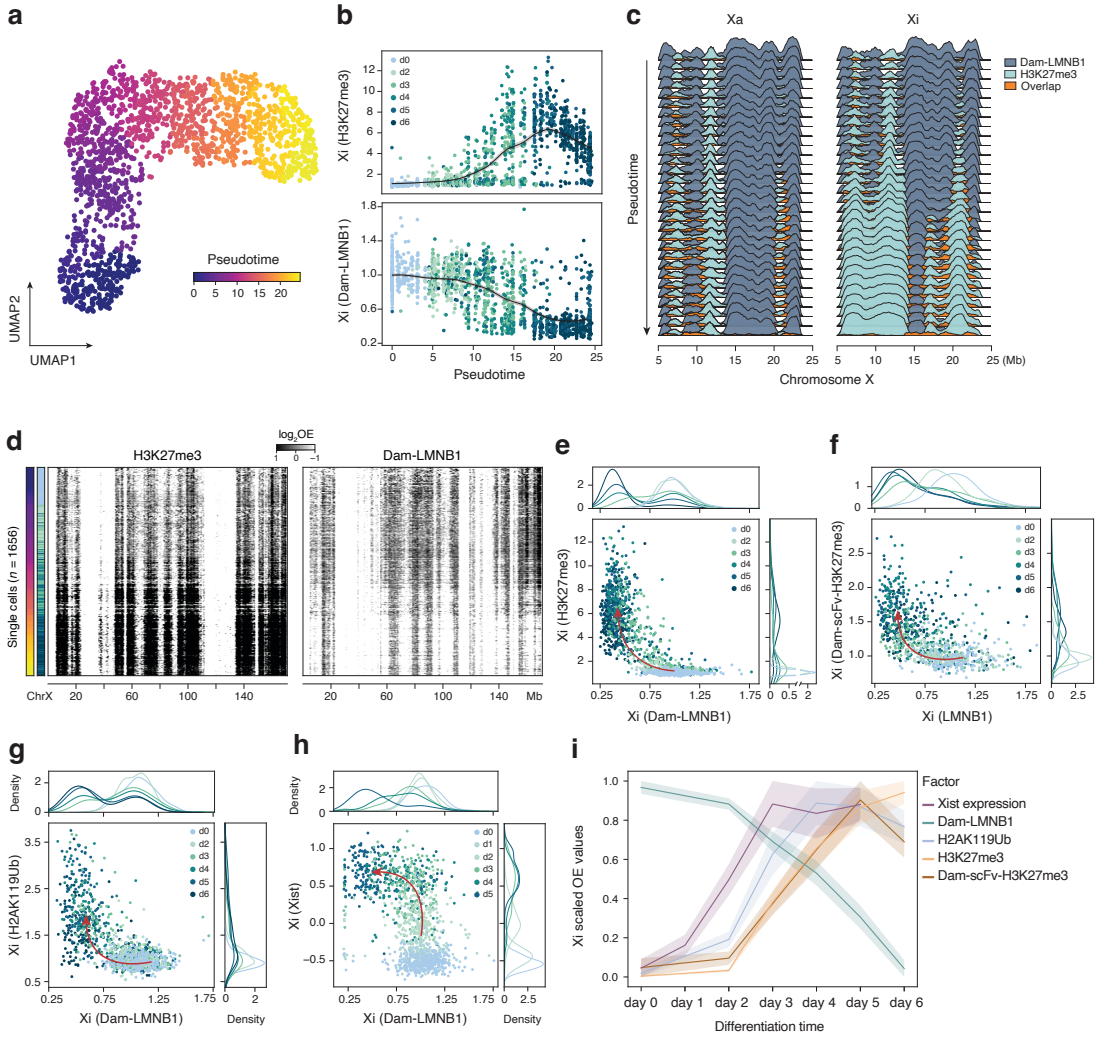


Fig. 4.5. Genome-lamina detachment precedes the formation of Polycomb heterochromatin during XCI. **a**, Same UMAP as in **4b** based on H3K27me3 levels on autosomal genes, but colored by pseudotime[57]. **b**, The Xi average OE H3K27me3 (top) and Dam-LMNB1 (bottom) levels (y-axis) across pseudotime (x-axis) are shown. Cells are colored according to differentiation day. **c**, CF of Dam-LMNB1, H3K27me3 and their overlap for groups of 100 cells progressively slid along pseudotime (y-axis) for a dynamic region of the inactive X chromosome (Xi; right) and its active counterpart (Xa; left). **d**, Single-cell heatmaps (n = 1656) ordered along pseudotime (y-axis) of the entire Xi (x-axis) for H3K27me3 (left) and Dam-LMNB1 (right). The color scale indicates OE values for each modality. **e**, Scatterplot of the average Xi Dam-LMNB1 (x-axis) and H3K27me3 (y-axis) OE levels. Distributions above and on the right show average density per day. Arrow is manually drawn based on directionality of cells over differentiation days. **f**, Same as **e**, but with the inverse readout: Dam-scFc-H3K27me3 (y-axis) / LMNB1 (x-axis). **g**, Same as **e**, for Dam-LMNB1 (x-axis) / H2AK119ub (y-axis). **h**, Same as **e**, for Dam-LMNB1 (x-axis) / Xist expression (y-axis). Xist levels are represented as transcripts per million. **i**, Summary of the order of events profiled during Vitamin C differentiation, with each plotted as average OE values per differentiation day. Shaded area: 95% CI.

As detachment appears to happen early on relative to heterochromatin formation by Polycomb, we sought to characterize its temporal relationship to expression of the master regulator of XCI, Xist. To this end, we performed scDam&T-seq[20,55] during Vitamin C differentiation, to jointly measure genome–lamina contacts and transcriptomes from the same cells (Fig. S4.5f-j).

To examine the temporal relation between Xist expression and nuclear lamina detachment of the Xi, we again projected both modalities on the same plot (Fig. 4.5h). Unlike H2AK119ub and H3K27me3 (Fig. 4.5e-g), we observe clear upward-right mobility of cells, indicating that Xist expression initiates prior to nuclear lamina detachment of the Xi (Fig. 4.5h, Fig. S4.5j).

Collectively, our Dam&ChIC data demonstrates that the loss of genome–lamina interactions is an early event in mouse random XCI, occurring after initiation of Xist expression, but preceding accumulation of the Polycomb marks H2AK119ub and H3K27me3 (Fig. 4.5i). These findings emphasize the progressive nature and temporal relationships of heterochromatin formation events during XCI, which may lead to a better understanding of their interdependencies.

Nuclear lamina detachment of the Xi occurs upon mitotic exit

Our above-described KBM7 experiments demonstrated increased LAD reorganization over mitosis, prompting us to question whether the release of the Xi from the nuclear lamina is also cell cycle-related. To answer this question, we first generated a Dam&ChIC dataset profiling genome–lamina interactions with both DamID and ChIC (Dam-LMNB1 / LMNB1) during Vitamin C differentiation. This allows us to distinguish past from present genome–lamina interactions in the same cells. Using solely inactive X-chromosome data we estimated the ratio between past and present genome–lamina interactions to infer chromatin velocity and latent time over the Vitamin C differentiation trajectory (Fig. 4.6a-b and Methods). This analysis accurately recapitulates the direction of differentiation, as shown by the progressive distribution of cells from different experimental time points across latent time (Fig. S4.6a). Reassuringly, the velocity model fails to deduce any directionality when the ratio between past and present X-chromosome data of randomly coupled DamID and ChIC readouts is used (Fig. S4.6b). Ordering of single-cell Xi profiles based on latent time shows the typical loss of genome–lamina interactions for both DamID and ChIC (Fig. 4.6c-d, Fig. S4.6c). Encouragingly, when comparing the dynamics of genome–lamina interactions measured by DamID and ChIC across latent time, we recover a temporal delay between both measurements, reflecting the rapid detachment of the Xi from the nuclear lamina (Fig. S4.6c).

We next wondered if the loss of genome–lamina interactions of the Xi could be linked to a specific cell-cycle state. We determined the cell-cycle phase for each cell using their Hoechst-based DNA content, that was measured during single-cell sorting (Fig. S4.6e). Interestingly, the precise moment of Xi detachment that we identify through the ratio between past (Dam-LMNB1) and present (LMNB1) genome–lamina interactions (Fig. 4.6e), coincides with a strong enrichment of G1-phase cells (Fig. 4.6f). Additionally, when projecting cell-cycle phase enrichment over the genome–lamina interaction dynamics of single cells (Fig. 4.6g,h, Fig. S4.6f), we observe a transition from G2 to G1 phase right at the moment of Xi release from the nuclear lamina. Taken together, the co-occurrence of nuclear lamina release and entry into the G1 cell-cycle phase, suggests that Xi genome–lamina interactions might be lost over mitosis, analogous to mitotic reorganization of LADs in KBM7 cells.

Discussion

Advantages and limitations of Dam&ChIC

We developed a new chromatin profiling technology, versatile in a twofold manner: 1) for simultaneously profiling two distinct chromatin features at high resolution, and 2) for measuring chromatin-state transitions over time in the same cell. We systematically benchmark Dam&ChIC, by probing diverse combinations of heterochromatic and euchromatic chromatin features. Additionally, when profiling the same chromatin feature (*i.e.*, genome–lamina interactions through NL protein LMNB1) Dam&ChIC detects LADs equally efficiently with either of its two readouts. Overall, we demonstrate that chromatin profiling with Dam&ChIC generates high-quality data both in haploid and diploid cells at allelic resolution.

Unlike other methods, Dam&ChIC does not involve in situ tagging of the genome. This essentially limits Dam&ChIC to measuring two chromatin features at a time. However, the highly distinct means through which both readouts recover chromatin information likely contributes to high per-cell resolution, resulting in unprecedented sensitivity. As a non-Tn5-based approach, Dam&ChIC bypasses the intrinsic bias of other technologies towards recovery of euchromatic features[56,57], and is therefore also suitable to probe the interplay between features of constitutive heterochromatin.

Despite some technologies having circumvented the Tn5-intrinsic bias by using Tn5 fusions targeting constitutive heterochromatin[28] or non-Tn5-dependent genome tagging strategies[16], their sensitivity remains comparatively low. Thus, Dam&ChIC is a suitable approach to examine combinations of chromatin features that range from euchromatic to constitutively heterochromatic, and interrogate their coordinated regulatory function on chromatin, from large domains to individual genes and promoters.

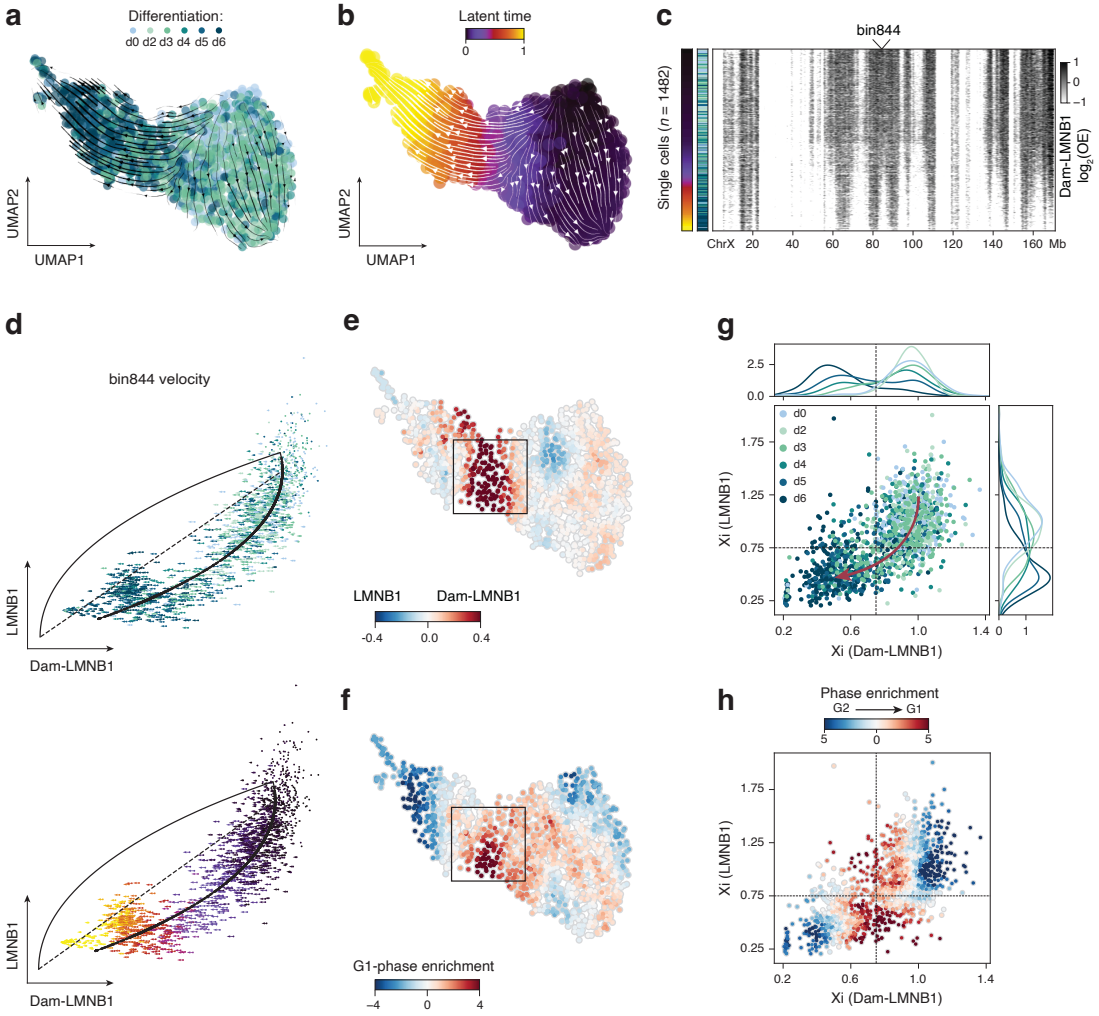


Fig. 4.6. Nuclear lamina detachment of the Xi occurs upon mitotic exit.

a, UMAP (n = 1482 cells) based on Dam-LMNB1 levels in 100-Kb bins of the Xi colored according to differentiation day. Streamlines on top of UMAP represent chromatin velocities inferred by the ratio of Dam-LMNB1 and LMNB1 in 100Kb-bins of the Xi, using scVelo[29]. **b**, Same UMAP as in **a**, but with cells colored according to scVelo-inferred latent time, that is purely based on Dam-LMNB1 / LMNB1 dynamics, not on the underlying UMAP. **c**, Single-cell heatmap showing log-transformed Dam-LMNB1 OE values on the entire inactivating X-chromosome, with cells ordered along the inferred latent time shown in **b**. **d**, Scatter plots showing the ratio of Dam-LMNB1 (x-axis) and LMNB1 (y-axis) in each cell for X-chromosome bin 844 (marked in **c**). Length of arrows reflect the Dam-LMNB1 / LMNB1 velocity of bin844 in that cell. Arrows are colored according to differentiation day (top) or latent time (bottom). **e-f**, Same UMAP as in **a**, colored according to the ratio between Dam-LMNB1 and LMNB1 (**f**) or G1 cell cycle phase enrichment (**g**). The latter is calculated with a kNN approach over the neighborhood of a cell (see Methods). **g**, Scatter plot of average Dam-LMNB1 (x-axis) and LMNB1 (y-axis) OE levels on the Xi. Each dot represents the Xi in one cell, colored according to differentiation day. *[cont]*

Distributions above and on the right show average density of day-specific data points. Arrow is manually drawn based on directionality of cells over differentiation days. **h**, Same scatter plot as in **g**, with points colored according to the ratio between Z-score normalized G1 and G2 cell cycle-phase enrichment (see Methods).

The versatility of Dam&ChIC relies partly on the availability of Dam-POI-expressing cell lines and on high-quality antibodies against targets of interest. A challenge with any DamID-based methodology is the requirement to engineer and integrate the Dam-POI construct for inducible expression in living cells or organisms, which is a significant time investment. However, many cell lines have already been established besides those used in this study, such as for DNA repair proteins[58], and recently the repertoire of DamID has even been extended to single-chain antibodies and engineered chromatin reader domains[19]. Any of these cell lines are directly compatible with Dam&ChIC, expanding the range of chromatin features that can be profiled. Similar to other antibody-based methods, a high-quality sortChIC readout is highly dependent on the sensitivity and specificity of the antibody used to target the chromatin feature of interest. Profiling of certain chromatin features with very sparse occupancy or transient binding may remain a challenge; nevertheless, a multitude of high-quality antibodies have been characterized and commonly used for epigenomic profiling.

Finally, we note that, as a plate-based approach Dam&ChIC is ideal for medium-throughput experiments, providing flexibility at a low cost compared to high-throughput approaches. Additionally, the plate-based nature of Dam&ChIC offers the opportunity to enrich rare cells of interest based on immunodetection or other cellular properties that can be measured by FACS[59]. Upscaling the throughput to sample thousands of cells in parallel is challenging and labor-intensive. Nevertheless, we anticipate that future iterations of Dam&ChIC that render it compatible with combinatorial indexing or microfluidic approaches will substantially increase its throughput.

Recording chromatin transitions over time with Dam&ChIC

Tracking chromatin changes over time in the same cell has been an attractive yet challenging concept. Such technology would enable the study of poorly understood aspects of epigenetic regulation; for example, how past chromatin changes affect present cellular states. With Dam&ChIC we set first steps in these directions, by experimentally integrating the cumulative past chromatin information recorded by scDamID, with snapshot information measured by sortChIC.

Previous work identified the dynamic nature of genome–lamina interactions, using a microscopy-based approach[26]. Consequently, it remains a question whether different types of LADs all undergo a similar degree of dynamics.

Here we used Dam&ChIC to identify extensive variability in genome–lamina dynamics between LADs with different CFs. In addition, we leveraged the different temporal resolutions inherent to Dam&ChIC to pseudo-order single-cells along the life cycle of LADs. We used this pseudo-ordering to study the directional dynamics of LADs and identify diverse LAD behaviors across latent time, such as polarity and focality in the establishment and release of LADs. The regions from where LADs are established and released are enriched in A/T content and are relatively depleted of genes. These observations raise interesting questions on how chromatin is gradually tethered to the nuclear periphery to form LADs[60,61].

Currently, Dam&ChIC is limited to recording short-term chromatin dynamics, given the absence of a maintenance machinery in higher eukaryotes that propagates the ^{m6}A mark during DNA replication. Consequently, the time window of opportunity to study chromatin dynamics with Dam&ChIC currently spans between two replication cycles. The present Dam&ChIC protocol is therefore very suitable to disentangle the role of chromatin dynamics in cell cycle-related phenomena, or other relatively rapid chromatin changes. We anticipate that future iterations will extend its use over many cellular generations. A candidate to aid propagation of the exogenous ^{m6}A mark is the Dam mutant L122A, which has been shown to transfer ^{m6}A only on hemi-methylated DNA[62], similar to DNMT1 for DNA CpG methylation. In the future, improved engineered versions of such exogenous methyltransferases could make tracking of chromatin dynamics over multiple cell generations feasible.

Dam&ChIC unravels the order of heterochromatin formation events during XCI

We used Dam&ChIC to study the large-scale epigenetic rewiring that happens during XCI and explore the interplay of chromatin features, with a main focus on the nuclear localization. We provide single-cell maps of genome–lamina interactions on the inactive X chromosome, combined with heterochromatic hPTMs. Previously, microscopy-based studies have described the nuclear periphery or the nucleolus as the preferred nuclear locations of the Xi[43,45]. More recent findings indicate that the Xi is actively recruited to the nuclear lamina through interactions of Xist with the Lamin B1 receptor (LBR), to enable chromosome-wide silencing[42]. Surprisingly, our data reveals extensive loss of genome–lamina interactions across mega-base regions on the Xi, that coincide with spreading of Polycomb, a known and early event of XCI[36-38]. These Xi regions are interspersed with anchor regions that maintain genome–lamina interactions and coincide with H3K9me3 domains. These anchor regions appear resistant to binding or spreading of Xist.

Although our data seemingly contradict the observation that the Xi is recruited to the nuclear lamina through interactions of LBR with Xist, these processes may be separated in time, and possibly an initial LBR/Xist-mediated recruitment of Xi at the

lamina precedes the dislodgement and Polycomb hPTM spreading we observe here. In addition, our data is consistent with findings showing that active tethering of the X chromosome to the nuclear lamina does not determine XCI initiation[63] and does not exclude that the detaching regions of Xi might instead form interactions with the nucleolus. Interestingly, the H3K9me3 domains that pre-exist already before XCI initiation, demarcate the future Xi regions that resist lamina detachment. This suggests that particularly those genomic regions that are not decorated by H3K9me3, are more vulnerable to detachment.

The observation that precisely those Xi regions that lose genome–lamina interactions strongly accumulate Polycomb-deposited H2AK119ub and H3K27me3, suggests a potentially antagonistic effect between lamina association and Polycomb during XCI. A similar phenomenon has recently been described in two unrelated systems, namely K562 cells and preimplantation embryos, where removal of H3K27me3 led to increased genome–lamina interactions[47,64]. The multifactorial analysis implies that the loss of genome–lamina interactions starts just prior to accumulation of H2AK119ub and H3K27me3. However, both the accumulation of Polycomb marks and the loss of genome–lamina interactions continue to gradually consolidate over XCI. We hypothesize that Polycomb hPTM deposition might further promote exclusion of the Xi from the nuclear lamina, essentially creating a negative feedback loop. Additionally, we find that expression of the key XCI non-coding RNA Xist is initiated before the Xi is released from the nuclear lamina. Despite this clear precedence of Xist expression over Xi nuclear lamina detachment, further studies are necessary to elucidate how the loss of genome–lamina interactions is functionally linked to the spreading of Xist along the X chromosome, transcriptional silencing and recruitment of Polycomb complexes to facilitate the formation of heterochromatin.

Dam&ChIC allowed us to infer chromatin velocity during XCI and resolve the direction of differentiation. More specifically, we aimed to use chromatin velocity during XCI to study the dynamics of genome–lamina interactions in connection to the cell cycle. Our analysis suggests that the loss of genome–lamina interactions occurs directly after mitotic exit, in the G1 phase. Considering our finding that LADs are partially inherited over mitosis, we hypothesize that the mitotic exit itself might offer a window of opportunity for detachment to take place, since the chromosomes need to find their way back to the re-assembled nuclear lamina. Future studies can uncover what triggers release of the Xi from the nuclear lamina, and how this phenomenon coordinates with mitotic exit. Taken together, our comprehensive single-cell analysis highlights the temporal dynamics of spatial positioning of the X chromosome and provides new insights into heterochromatin formation during XCI.

Conclusion

Several aspects of chromatin regulation have been poorly understood due to the limitations posed by conventional technologies. Dam&ChIC introduces unique possibilities to investigate the interplay between diverse chromatin factors, and to track temporal changes within the same cell. Our single-cell multifactorial chromatin data and retrospective analyses elucidate aspects of chromatin regulation that are involved in the inheritance of chromatin upon mitosis and the formation of heterochromatin during XCI. These properties establish Dam&ChIC as a unique tool in the rapidly growing set of single-cell chromatin technologies.

Acknowledgements

We thank members of the Kind and van Oudenaarden labs for useful feedback, discussions throughout the project and critical reading of the manuscript. We thank Michael Robson and Konrad Chudzik for providing materials, as well as Isabel Sierra and Nuno Martins for assistance and efforts that did not make it to publication. We thank Reinier van der Linden, Anita Pfauth and Joost Korver (Hubrecht FACS facility) for performing single-cell sorting. We thank Koos Rooijers for developing data processing methods. We thank the Utrecht Sequencing Facility (USEQ) for providing sequencing service and data. USEQ is subsidized by the University Medical Center Utrecht and The Netherlands X-omics Initiative (NWO project 184.034.019). This work was supported by ERC Consolidator grant (ERC-CoG 1010002885-FateID) and a Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) VIDI grant (016.161.339) to JK. PZ was funded by SNF (P2BSP3-174991), HFSP (LT000209/2018-L) and Marie Skłodowska-Curie Actions (798573). The work in the lab of AvO is supported by (ERC-AdG 101053581-scTranslatomics), and Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NWO) TOP award (NWO-CW714.016.001). The Oncode Institute is partially funded by the KWF Dutch Cancer Society.

Author contributions

SK, PZ and JK conceived the Dam&ChIC method. SK and PZ developed the method. SK, PMJR, PZ and JK designed the project, with input from KLdL. SK designed, optimized and performed all experiments. PMJR performed all computational analyses and developed software. SK and PMJR curated and validated the data. PZ, KLdL and SSdV assisted with experiments. TK developed preliminary data processing methods. TK and PZ performed preliminary analyses. PMJR, SK and PZ visualized the data. JK, PZ and AvO supervised the project. JK and AvO acquired research funding. SK, PMJR, PZ and JK wrote the original manuscript. SK, PZ and PMJR revised the manuscript.

Methods

Experimental methods

Dam-POI cell lines

KBM7 cells expressing Dam-LMNB1 or untethered Dam under the control of the destabilizing domain (DD) system were generated previously using lentiviral transductions[1]. Mouse F1 hybrid Cast/EiJ x 129/Sv embryonic stem cells (ESCs) were a kind gift from the Joost Gribnau laboratory. These mESC lines are derived from female cells and do not contain a Y chromosome. The Dam-LMNB1 or Dam-scFv-H3K27me3 constructs were previously integrated in these cells using CRISPR targeting[2,3]. Both ESC lines express the Dam-POI under the control of the AID degron system. In Dam-scFv-H3K27me3 ESCs, expression of the Dam fusion protein is additionally controlled by tamoxifen-induced nuclear translocation via the estrogen receptor (ER).

Cell culture

All cell lines were cultured in a humidified chamber at 37 °C in 5% CO2 and tested for mycoplasma at a regular basis. Human haploid KBM7 cells were cultured in suspension in IMDM medium (Gibco, 12440053) supplemented with 10% FBS (Sigma, F7514) and 1% Pen/Strep (Gibco, 15140122). Mouse F1 hybrid ESCs were cultured on a monolayer of irradiated primary mouse embryonic fibroblasts (MEFs) in CM+/+, as described previously[3]. CM+/+ consists of G-MEM (Gibco, 11710035) supplemented with 10% FBS, 1% Pen/Strep, 1× GlutaMAX (Gibco, 35050061), 1× non-essential amino acids (Gibco, 11140050), 1× sodium pyruvate (Gibco, 11360070), 0.1 mM β-mercaptoethanol (Sigma, M3148) and 1000 U/mL ESGROmLIF (EMD Millipore, ESG1107). For passaging, mouse ESCs cells were washed once with PBS0, treated with TrypLE (Gibco, 12605010) for 4 min at 37 °C, neutralized with serum-containing medium, and after one wash they were plated at the desired density. Prior to differentiation, mouse ESCs were passaged at least once in feeder-free conditions, on 0.1% gelatin-coated 6-wells, and cultured in medium containing 40% CM +/+ and 60% BRL-conditioned CM +/+ medium. Expression of the Dam-POI in the mouse ESC lines was suppressed by addition of 0.5 mM indole-3-acetic acid (IAA; Sigma, I5148).

Induction of Dam-POI expression

In KBM7 cells, Dam-LMNB1 and the untethered Dam were induced with the addition of 0.5 nM Shield-1 (Glixx Laboratories Inc, GLXC-02939) that enables stabilization of the protein. In hybrid mESCs, 24 h prior to induction the cells were cultured in 1 mM indole-3-acetic acid (IAA; Sigma, I5148), to ensure protein degradation. For induction the cells were washed three times with PBS and fresh medium without IAA was added, followed by the addition of 1 uM 4-Hydroxytamoxifen (4-OHT; Sigma, SML1666), when required.

In detail, Dam-LMNB1 mESCs were induced for 6 h without IAA and Dam-scFv-H3K27me3 mESCs were induced for 20 h without IAA and 6 h with 4-OHT.

In vitro differentiation of mouse hybrid ESCs with Ascorbic Acid (Vitamin C)
Dam-POI expressing mouse hybrid ESCs were passaged at least once in feeder-free conditions and cultured as described above. The culture was depleted from the remaining MEFs and 24-48 h later confluent wells were passaged and plated in AA differentiation medium, defined as IMDM supplemented with 15% FBS, 1% Pen/Strep, 1X GlutaMAX, 1X non-essential amino acids, 50 µg/mL Ascorbic Acid (Sigma, A4544) and 0.1 mM β-mercaptoethanol[4,5]. Expression of the Dam-POI was suppressed by addition of 0.5 mM IAA. The medium was refreshed every day during the course of differentiation.

Cell cycle synchronization with double thymidine block
Confluent cultures of KBM7 cells were passaged 24 h prior to the initiation of synchronizations and 24 h later received 2 mM thymidine (Sigma, T1895) for 15 h, followed by 9 h of release, followed by a second thymidine block for 13 h. Synchronization tests were performed prior to Dam&ChIC experiments, in order to define the cell-cycle progression of these cell lines based on the DNA content at 0, 4, 6, 7, 8, 9, 10, 11, 12, 13 and 14 h after release from the second thymidine block. Cells were harvested at the desired time points, counted on a cytometer, and after one wash with PBS0 0.5 - 1 million cells were fixed with 70% ethanol at -20 °C overnight. Afterwards, the fixed cell suspension was washed once with ice-cold PBS0 and stained in PBS0 containing 2.5 µg/mL Hoechst 34580 (Sigma) for 48 h at 4 °C. To standardize the stainings, equal cell numbers were used from each time point. DNA content measurements were performed on a BD LSR Fortessa FACS, and subsequent analyses were done using FlowJo.

In the single-cell Dam&ChIC experiment, KBM7 cells were synchronized with a double thymidine block as above. During the second thymidine block, 0.5 nM of Shield-1 was added to enable stabilization of Dam-LMNB1. Afterwards, both thymidine and Shield-1 were washed out, cells were released to progress through S-phase, and harvested in G2 phase (8 h post-thymidine release; equivalent to 1 hour in G2) and early G1 phase (13 h post-thymidine release; equivalent to 1 hour in G1). While the Dam-deposited ^{m6}A mark is typically not propagated upon DNA replication in the absence of the Dam-POI, this setup ensures that Dam-LMNB1 is still present during S and G2 at sufficient levels to allow recovery of LADs. Importantly, by harvesting cells as early as possible after mitotic exit, measurements of *de novo* ^{m6}A from any non-degraded protein are prevented, because at that point cells do not have sufficient time to accumulate methylation, as shown before using the ^{m6}A-Tracer[6].

Preparation of samples for Dam&ChIC

Cell harvesting and permeabilization

Nuclei isolation was done like described before by Zeller 2023. In short, cells were collected in 15-mL falcon tubes and washed three times with room temperature PBS0. After counting, 0.5-1 million cells per staining were used for permeabilization in ice-cold Wash Buffer 1 (20 mM HEPES pH 7.5, 150 mM NaCl, 66.6 µg/mL Spermidine (Sigma, S2626-1G), 1x cOmplete protease inhibitor cocktail (Roche, 11836170001), 0.05% Saponin (Sigma, 47036-50G-F), 2 mM EDTA) and washed once at 300g for 4 minutes at 4 °C, before addition of the primary antibody. Subsequent washes were done with the same centrifuge settings.

Cell harvesting, ethanol fixation and permeabilization

For ethanol fixation, cells were collected in 15-mL falcon tubes and washed three times with room temperature PBS0. After counting, 0.5-1 million cells were fixed with 70% ice-cold ethanol for 1 hour at -20 °C. Fixed cells were washed once and permeabilized in ice-cold Wash Buffer 1F (20 mM HEPES pH 7.5, 150 mM NaCl, 1x cOmplete protease inhibitor cocktail, 0.05% Tween). Afterwards, samples were stained with CellTrace dyes (see below) or frozen at -80 °C for long-term storage in Wash buffer 1F that contains additionally 66.6 µg/mL Spermidine, 2 mM EDTA and 10% DMSO.

Staining with CellTrace dyes

Upon fixation and washing, 0.5-1 million cells were stained in 1 mL Wash buffer 1F containing 0.25 µL of a CellTrace dye (CFSE; CellTrace Far Red, CTFR; CellTrace Yellow, CTY) at 4 °C for 20 minutes protected from light, as described before[7]. When multiple dyes were used per sample, each was used equimolarly. The stainings were quenched with the addition of 50 µL Rat Serum (Sigma, R9759) and incubation at 4 °C for 10 minutes protected from light. Samples were washed once with Wash buffer 1F that contains additionally 66.6 µg/mL Spermidine and 2 mM EDTA, and resuspended in the same buffer for subsequent antibody stainings. Part of the sample was aliquoted and frozen at -80 °C for long term storage, as described above.

Multiplexing of samples based on CellTrace stainings

Samples stained with different combinations of CellTrace dyes were mixed in one "super-sample" to enable parallel processing and thereby minimize batch effects, as described before[7]. Specifically, samples from each day of the Vitamin C differentiation time-course were stained with a unique combination of CellTrace dyes, including CFSE, CTFR and CTY or left unstained. This enabled the distinction of up to eight populations in the "super-sample" based on fluorescence. In the synchronization and mitosis experiment, the G2 sample was stained with CTY and CFSE, and the early G1 sample was stained with CFSE, CTFR and CTY. In the latter experiment,

more CellTrace-stained control samples were mixed in the same "super-sample", but only the relevant stainings are presented here for simplicity (Fig. S4.3b). For all experiments, equal cell numbers from each sample were combined together to a final total amount of 0.5-1 M cells for subsequent antibody staining.

Antibody staining and pA-MNase tethering

Nuclei or fixed cells were stained with the primary antibody overnight at 4 °C on a roller. The concentrations used varied per antibody and lot number, and were defined by titrations done in bulk samples prior to single-cell experiments. Specifically, anti-LMNB1 (Abcam, ab16048) was used at 1:200 or 1:400; anti-H3K4me1 (Abcam, ab8895) at 1:400; anti-H3K27me3 (Cell Signaling Technologies, 9733S) at 1:200; anti-H3K9me3 RM389 (Thermofisher, MA5-33395) at 1:200; anti-H3K4me3 (Thermofisher, MA5-11199) at 1:400; anti-H2AK119Ub (Cell Signaling, D27C4) at 1:400. After one wash with Wash Buffer 2 (for nuclei; 20 mM HEPES pH 7.5, 150 mM NaCl, 66.6 µg/mL Spermidine, 1x cOmplete protease inhibitor cocktail, 0.05% Saponin) or Wash Buffer 2F (for fixed/permeabilized cells; 20 mM HEPES pH 7.5, 150 mM NaCl, 66.6 µg/mL Spermidine, 1x cOmplete protease inhibitor cocktail, 0.05% Tween), pA-MNase was added at a final concentration of 3 ng/µL for KBM7 samples and 0.6 ng/µL for ESC and Vitamin C samples, in the respective buffers. Hoechst 34580 was added in the same mix at a final concentration of 2.5 µg/mL. The samples were incubated in a roller at 4 °C for 1 hour, followed by two washes in the same wash buffer, and finally passed through a cell-strainer for sorting.

Single-cell sorting and FACS gating strategies

FACS sorting was performed on the BD FACS Influx Cell Sorter System or BD FACS Jazz Cell Sorter System, with sample and plate cooling. Gating was based on size (forward-side scatter), exclusion of doublets, and Hoechst-defined DNA content (Fig. S4.3b). For experiments in which different samples were multiplexed, for example in the mitosis and XCI experiments, an additional gating strategy based on fluorescence levels of the CellTrace dyes ensured distinction of the different populations (Fig. S4.3b). In the mitosis experiment, G1 and G2 cells were additionally selected as positive for CellTrace CFSE and CellTrace Yellow (CTY), and were sorted on a combination of Hoechst and CellTrace Far Red (CTFR) to ensure selection of pure synchronized cells in G2 and G1 (Fig. S4.3b). One cell per well was sorted in hard-shell 384 well-plates that were pre-filled with 5 uL/well mineral oil using the Freedom EVO liquid handling platform (Tecan), and 100 nL/well Wash buffer 3 (for nuclei; 20 mM HEPES pH 7.5, 150 mM NaCl, 66.6 µg/mL Spermidine, 0.05% Saponin) or Wash buffer 3F (for fixed/permeabilized cells; 20 mM HEPES pH 7.5, 150 mM NaCl, 66.6 µg/mL Spermidine, 0.05% Tween). After sorting, plates were sealed with aluminum seals (Greinier, 676090), spun at 2000g for 1-2 minutes at 4 °C, and kept at 4 °C until further processing.

Single-cell Dam&ChIC

Liquid dispensions were performed using the Nanodrop II liquid handler (Innovadyne) and adapter dispensions were performed using the Mosquito LV (STP Labtech), as described before[3,8,9]. In short, pA-MNase in single nuclei was activated after dispension of 100 nL/well of Activation Solution (Wash buffer 3 or Wash buffer 3F, containing 4 mM CaCl2) to a total volume of 200 nL/well. The plates were incubated at 4 °C for exactly 30 minutes, followed by dispension of 200 nL/well of Stop Solution (containing 0.04M EGTA, 1.5% NP40, 2 µg/µL Proteinase K) to a total volume of 400 nL/well. For lysis and proteinase K treatment, plates were incubated at 65 °C for 6 h, followed by heat inactivation at 80 °C for 10 minutes. Plates were stored at -20 °C until further processing. 200 nL/well of Blunt-ending mix (containing 4 nL (0.02 U) Klenow, 4 nL (0.04 U) T4 PNK, 10 nL 10 mM dNTPs, 60 nL 10 mM ATP, 20 nL 25 mM MgCl2, 10 nL 50% PEG8000, 6 nL 20 µg/µL BSA, 60 nL 10x PNK buffer, 26 nL ultra-pure water) were dispensed to a total volume of 600 nL/well, followed by incubation at 37 °C for 30 minutes and heat inactivation at 75 °C for 20 minutes. Next, digestion with 400 nL/well DpnI mix (containing 20 nL (0.4U) DpnI, 40 nL 10x PNK buffer, 10 nL 20 µg/µL BSA, 330 nL ultra-pure water) was done to a total volume of 1000 nL/well, followed by incubation at 37 °C for 8 h and heat inactivation at 80 °C for 20 minutes. Finally, the content of each well was barcoded through dispension of 50 nL/well of DamID2 adapters at a final concentration of 25 nM and dispension of 950 nL/well Ligation mix (containing 75 nL (0.375U) T4 Ligase, 200 nL 10x Ligase buffer, 675 nL ultra-pure water) to a cumulative final volume of 2000 nL/well, followed by incubation at 4 °C for 20 min, 16 °C for 16 h and heat inactivation at 65 °C for 10 minutes. After each dispension step, the plates were sealed with aluminum plate seals and spun at 2000g for 1-2 minutes at 4 °C.

Single-cell Dam&T

Single-cell Dam&T-seq was performed as described before[8] with major adaptations in reaction volumes and adapter concentration, to facilitate cost efficiency. One cell per well was sorted in hard-shell 384 well-plates that were pre-filled with 5 µL/well mineral oil using the Freedom EVO liquid handling platform (Tecan) and 50 nL/well of 1.5 µM barcoded CELseq2 primers[10] using the Mosquito LV liquid handler (STP Labtech). After sorting, plates were sealed with aluminum seals (Greinier, 676090), spun at 2000g for 2 minutes at 4 °C, and frozen at -80 °C until processing. For processing with single-cell Dam&T, liquid dispensions were performed using the Nanodrop II liquid handler (Innovadyne) and adapter dispensions were performed using the Mosquito LV (STP Labtech), as described before[3,8]. First, 50 nL/well of Lysis mix (containing 10 nL 10 mM dNTPs, 10 nL 1:50000 RNA spike-ins, 7.5 nL 1% NP-40, 22.5 nL ultra-pure water) were added to a total volume of 100 nL/well and plates were incubated at 65 °C for 5 minutes. Directly after that, reverse transcription (RT) of the RNA was performed with dispension of 75 nL/well of RT mix (containing 35 nL 5x

First Strand buffer, 17.5 nL 0.1 M DTT, 8.75 nL (0.35U) RNase OUT, 8.75 nL (1.75U) Superscript II, 5 nL ultra-pure water) to a total volume of 175 nL/well, followed by incubation at 42 °C for 1h, 4 °C for 5 minutes and 70 °C for 10 minutes. Second-strand synthesis (SSS) followed, with dispension of 925 nL/well of SSS mix (containing 220 nL 5x Second Strand buffer, 25 nL 10-mM dNTPs, 7.5 nL (0.075 U) E. coli DNA Ligase, 25 nL (0.25U) DNA polymerase I, 7.5 nL (0.015 U) RNase H, 640 nL ultra-pure water) to a total volume of 1100 nL/well, and incubation at 16 °C for 2 h. Next, the samples were treated with 250 nL/well of Proteinase K mix (containing 65 nL 20 mg/mL Proteinase K, 135 nL 10x CutSmart buffer, 50 nL ultra-pure water) to a cumulative volume of 1350 nL/well, and incubation at 50 °C for 10h and heat inactivation at 80 °C for 20 minutes. Then, 150 nL/well of DpnI mix (15 nL (0.3 U) DpnI, 15 nL 10x CutSmart buffer, 120 nL ultra-pure water) were added to a cumulative volume of 1500 nL/well, and incubation at 37 °C for 6 h and heat inactivation at 80 °C for 20 minutes. Finally, ligation of barcoded DamID2 adapters[8] was performed with dispension of 100 nL/well of adapters to a final concentration of 25 nM, followed directly by dispension of 400 µL/well of Ligation mix (200 nL 10x Ligase buffer, 50 nL (0.25 U) T4 Ligase, 150 nL ultra-pure water) to a cumulative final volume of 2000 nL/well, incubation at 16 °C for 16 h and heat inactivation at 65 °C for 10 minutes. After each dispension step, the plates were sealed with aluminum plate seals and spun at 2000 g for 1-2 minutes at 4 °C.

Library preparation

The content of all wells in a plate was pooled by centrifugation of the 384-well plates on collection plates, transferred to microtubes, and the aqueous phase was separated from the oil phase by a few rounds of centrifugation and transfers to clean tubes. Sample purification with beads followed, like described before[11], by incubation of the sample with 0.8 volume 1:10 diluted beads (CleanNGS; GC Biotech, CNGS-0050) in bead-binding buffer, three washes with 80% Ethanol and elusion with 7 µL ultra-pure nuclease-free water. The eluted sample was then *in vitro* transcribed at 37 °C for 14 h, using the MEGAScript T7 transcription kit (Invitrogen, AMB13345). For single-cell Dam&ChIC, the amplified RNA (aRNA) was purified using 0.8 volume undiluted beads, washed three times with 80% Ethanol, and eluted in 13 µL ultra-pure nuclease-free water. For single-cell Dam&ChIC, the quality of the eluate was evaluated using the Agilent RNA 6000 Pico Assay, and, if necessary, it was bead-purified again. For single-cell Dam&T, the aRNA was fragmented with the addition of 0.2 volume fragmentation buffer (500 mM potassium acetate, 150 mM magnesium acetate, 200 mM Tris-acetate) at 94 °C for 90 seconds, the reaction was stopped with 0.1 volume 0.5M EDTA and quenched on ice, followed by another 0.8 volume bead cleanup. Library preparation followed, like described previously[11], with a few adjustments. In particular, a maximum of 6 µL of 100-200 ng aRNA was used for reverse transcription, followed by 8-11 cycles of library PCR, depending on the amount of input aRNA, and

the amplified material was purified twice, as described above, before final elution in 13 µL of ultra-pure nuclease-free water. Libraries were quantified using the Qubit dsDNA High Sensitivity Assay and the Agilent High Sensitivity DNA Assay, and subsequently sequenced with single-end or paired-end sequencing on the Illumina NextSeq500 (75-bp reads) or the Illumina NextSeq2000 (100-bp reads).

Computational methods

Data processing

Processing Dam&ChIC data

Dam&ChIC data processing is largely based on the workflow and scripts described in Rooijers 2019 and Markodimitraki 2020, but adapted to allow for computational separation of scDamID and sortChIC derived reads and further processing of sortChIC reads. The key steps of the procedure are described below (detailed explanation and code are available at https://github.com/KindLab/DamChIC).

Raw data pre-processing

All reads are demultiplexed by comparing their sample barcode (e.g. single-cell barcode) located at the start of each read (R1 in case of paired-end sequenced libraries) conform to the read layout of 5′-[3 nt UMI][8 nt barcode]TC[gDNA]-3′ for scDamID and 5′-[3 nt UMI][8 nt barcode]WN[gDNA]-3′ for sortChIC reads, to a reference list of barcodes and zero mismatches between the observed barcode and reference are allowed. Sample barcodes and UMIs are trimmed off the reads to only retain gDNA sequences. The UMI sequences were appended to the read name for downstream processing. Per sample barcode, scDamID and sortChIC reads remain mixed at this point.

Sequence alignments

After preprocessing, the reads were aligned to a reference genome using bowtie2 (v. 2.4.1) with parameters: "--seed 42 --very-sensitive -N 1". Before alignment a 'GA' dinucleotide was prepended to the reads that has effectively been digested off scDamID fragments during DpnI gDNA digestion. For human samples, reads were aligned to reference genome hg19 (GRCh37) and mouse samples to mm10 (GRCm38). For allele-specific processing for X-chromosome inactivation experiments, data were aligned to reference genomes generated by imputing 129/Sv and CAST/EiJ single nucleotide polymorphisms obtained from the Sanger Mouse Genomes project[12] onto the mm10 reference genome. The reads were assigned to either genotype by aligning reads to both references. Reads that aligned with lower edit distance (SAM tag 'NM') or higher alignment score (SAM tag 'AS') in case of equal edit distance to one of the genotypes were assigned to that genotype. Reads aligning with equal edit distance and alignment score to both genotypes were considered of 'ambiguous' genotype.

Reads that yielded an alignment with mapping quality (BAM field 'MAPQ') lower than 10 were discarded.

In silico separation of scDamID and sortChIC reads

Following read alignment with 'GA' prepended dinucleotide, reads were considered scDamID in case of perfect alignment on genomic GATC motifs. Residual reads were considered sortChIC and their original sequences (*i.e.*, without prepended GA dinucleotide) were realigned using hisat2 (v. 2.1.0) with parameters: "--seed 42 --no-spliced-alignment --mp '2,0' --sp '4,0'". These sortChIC-considered reads were further pruned given MNase's strong adenine or thymidine-nucleotide cut preference (Fig. S4.1c); only sortChIC reads were kept that start with an 'A' or 'T' nucleotide. In addition, we observed that a larger fraction of sortChIC reads (*i.e.*, that do not map on a genomic GATC motif) are 'TC'-dinucleotide starting in Dam&ChIC libraries than in sortChIC libraries (Fig. S4.1d) and hence decided to exclude TC-starting reads from Dam&ChIC-derived sortChIC libraries.

PCR duplicate filtering

To exclude PCR read-duplicates, both scDamID and sortChIC read counts were collapsed based on genomic position, strand and unique molecular identifier (UMI) sequence. In case of haploid KBM7 cells or allele-specific data processing, multiple reads with the same UMI count as 1, for diploid data processing as 2. The number of observed unique UMIs was taken as the number of unique methylation events for scDamID or unique MNase cut sites for sortChIC.

Filtering of samples

To exclude single-cell samples from the analysis that failed we applied construct-specific cutoffs on UMI-unique reads for scDamID: 1,000 for Dam and Dam-scFv-H3K27me3 (Rang *et al.*, 2022), and 5,000 for Dam-LMNB1. A general cutoff of 1,000 UMI-unique reads per sample was applied for ChIC. Samples in allele-specific analysis were included if ≤200 UMI-unique reads could be assigned to both parental alleles for scDamID as well as sortChIC-derived reads.

ChIP-seq data processing

External K562 ChIP-seq datasets were downloaded from the ENCODE database[13]. ChIP-seq datasets from K562 cells used for this manuscript include H3K4me3 (ENCSR668LDD), H3K36me3 (ENCSR000DWB), H3K27me3 (ENCSR000EWB) and H3K9me3 (ENCSR000APE). ChIP-seq reads were aligned using bowtie2 (v. 2.4.1) with parameters: "--seed 42 --very-sensitive -N 1". ChIP-seq data were binned using the "bamCoverage" function from DeepTools (v. 3.3.2) with parameters: "--ignoreDuplicates --minMappingQuality 10". NarrowPeak calls were likewise downloaded from the ENCODE database from the corresponding ChIP-seq datasets.

Downstream analyses

Binning and calculation of OE values

For most analyses, scDamID and sortChIC data were binned using consecutive non-overlapping 100-Kb bins. For gene-based analysis (Fig. 4.1c,d) data were binned at a higher resolution of 1 Kb. To be able to calculate OE values, an "expected" vector for scDamID data was generated in silico by generating sequences of 65 nt (in both orientations) from the reference genome and aligning and processing them identically to the data. By binning the in silico generated reads, the maximum amount of mappable unique events per bin was determined (Fig. S4.1g). The expected for sortChIC was generated experimentally using MNase-H3 (Fig. S4.1g), unless stated otherwise. OE values were calculated as described in Rooijers 2019.

Data binarization, contact frequency and combined single-cell heatmap

The contact frequency (CF) metric is used throughout the manuscript and is defined by calculating the fraction of samples (single cells) that meet the binarization threshold for a given 100-Kb bin[1]. The combined single-cell heatmap presented in Fig. 4.3c is determined by first binarizing the 100-Kb binned data of both Dam&ChIC constituents separately, resulting in two presence-absence M x N matrices; 1) scDamID defined as $M_d = (a_{ij})$ and 2) sortChIC defined as $M_c = (a_{ij})$, where i is a cell and j a 100-Kb bin. Both matrices are combined in one by simple summation as $(M_d + M_c) + M_c$. In the resulting combined matrix 0 denotes absence of both, 1 presence only of scDamID, 2 presence only of sortChIC and 3 presence of both.

Dimensionality reduction on KBM7 Dam&ChIC data

The UMAP presented in Fig. 4.1f was computed by first calculating the first 50 principal components (Python; sklearn v. 1.0.2) on the RPKM normalized single-cell data binned at 100-Kb resolution, on which UMAP (Python; umap v. 0.5.2) was then run.

(Single-cell) LAD calling

Bulk LADs presented in Fig. S4.1c are called by setting a threshold of log2(≥1) based on the bimodal distribution of OE values of 100-Kb genomic bins. (Fig. S4.1a) [1]. Notably, LAD calls <300 Kb were excluded and LAD calls intervened by ≤100 Kb were combined into one. For single-cell LAD calling used for Fig. 4.2, LADs we first separately called on pseudobulk profiles of both Dam&ChIC readouts as described above. Only LADs that were identified by both readouts were kept for single-cell analysis (Fig. S4.2c). The resulting LAD coordinates were tested in each cell for having an average OE log2(≥1), if this requirement was met, the LAD was called in the given cell. This resulted in two LAD-based binary presence-absence matrices; 1) scDamID defined as $M_d = (a_{ij})$ and 2) sortChIC defined as $M_c = (a_{ij})$, where i is a cell and j a LAD.

Jaccard Z-score normalization by matrix permutation

To compare the single-cell co-ocurrence of scDamID and sortChIC identified LADs between different Dam&ChIC experiments and across LADs with different CFs, the Jaccard similarity was index was computed, defined as $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$.

A represents the scDamID and *B* the sortChIC measurement from the same cell (Fig. S4.2f, Fig. 4.3b, Fig. S4.3e). Two inherent problems are posed on this analysis that might introduce undesirable bias and ultimately unreliable comparison between different experiments or LADs: 1) typical sparsity of single-cell data results in non-uniformly distributed signal dropout and 2) binary similarity metrics can be sensitive to site prominence that varies between different LADs, but which does not reflect true co-occurrence of scDamID and sortChIC signal. To solve both problems, we applied a previously described matrix randomization algorithm [14], to permute the aforementioned presence-absence LAD matrix M_c *n*-times ($n = 100$), without altering row and column totals. The resulting randomized matrices were used to compute Jaccard similarity with LAD matrix M_d to serve for Z-score normalization, which can be written as $Z = \frac{x-\mu}{\sigma}$, where x is the similarity score of the observed, μ the mean similarity and σ the standard deviation of the random controls.

Chromatin velocity

Chromatin velocity analysis presented in Fig. 4.2 and 6 were performed using the scVelo Python package v.0.2.5[15]. An average OE value was calculated in the abovementioned LAD calls for each cell for both the Dam-LMNB1 and LMNB1 Dam&ChIC readout. The resulting Dam-LMNB1 matrix was provided to scVelo as the "spliced" layer and the LMNB1 matrix as "unspliced".

XCI dimensionality reduction and trajectory inference

Dimensionality reduction of the XCI analysis was performed with Seurat v.4.1.0[16] and Signac v.1.3.0[17]. The UMAP introduced in Fig. 4.5b is based on H3K27me3 signal on autosomal genes (*i.e.,* 5 Kb upstream of TSS + gene body). Different batches (*i.e.,* Dam&ChIC experiments) were integrated using Harmony v.0.1.0[18]. Pseudotime inference was performed using Monole3 v.1.0.0[19].

Z-score normalized cell cycle analysis on XCI data

The cell cycle analysis on the XCI dataset presented in Fig. 4.6 is based on Hoechst staining of sorted cells, which reflects DNA content and therefore enables their categorization to the different cell cycle phases (*i.e.,* G1, S or G2). However, these phases are not equally divided across single cells, as the majority of the cells in culture are in G1-phase (Fig. S4.6e). This poses a problem on the cell cycle analysis presented in Fig. 4.6h, that may lead to overestimation of one stage over another.

To solve this, we performed Z-score normalization of the observed cell cycle distributions within k-nearest neighbors *kNN*, $k = 100$. For each neighborhood, the observed cell cycle distribution was compared to n times randomly sampled controls ($n = 100$). Z-score normalization was defined as follows, $Z = \frac{x-\mu}{\sigma}$, where x is the cell cycle distribution of the observed, μ the mean similarity and σ the standard deviation of the randomly sampled controls.

Definition of start and end point anchors in latent time ordering

For the analysis in Fig. 4.2h, the start and end point of LADs through latent time were determined as follows. All bins of each LAD were averaged over latent time, resulting in a 1-dimensional vector containing LAD intensity as a function of latent time. These vectors were first smoothed using a Gaussian filter (sigma=5) before the start increase and end decay were determined using a binary segmentation algorithm ruptures.Binseg. From these single start and end points 75 cells upstream and 10 cells downstream in latent time were averaged.

Simulating in silico data

To corroborate our interpretation of the temporal relationship between chromatin features in Fig. 4.5, in silico data was simulated and presented in Fig. S4.5. The in silico data is generated by an adapted function from scvelo.datasets.simulation[15]. Briefly, the parameters for each reaction are randomly sampled from a log-normal distribution and time events follow the Poisson law. Both the switch moment and noise levels are controlled.

Inactive X chromosome selection

In the mESC vitamin C differentiation datasets the Xi chromosome (i.e, Cast/EiJ or 129/Sv allele) was always determined based on one of the two measured features. In case of Dam-LMNB1/H3K27me3 Dam&ChIC data the allele with the highest H3K27me3 was defined as the Xi, for Dam-LMNB1/H2AK119Ub Dam&ChIC the allele with the highest H2AK119Ub, for both Dam-LMNB1/H3K9me3 and Dam-LMNB1/LMNB1 Dam&ChIC data the allele with the lowest Dam-LMNB1, for all Dam&ChIC datasets that involved the Dam-scFv-H3K27me3 DamID construct the allele with the highest levels of that was defined as the Xi, and for the scDam&T-seq dataset that involved Dam-LMNB1 the RNA expression levels of Xist were used the define the Xi.

Data and code availability

All genomic and transcriptomic data generated in this study have been deposited on GEO under accession number GSE247458. Previously published K562 ChIP-seq data that was used to benchmark Dam&ChIC data are available in the ENCODE database with numbers ENCSR668LDD (H3K4me3), ENCSR000EWB (H3K27me3) and ENCSR000APE (H3K9me3). K562 ATAC-seq data is available in the ENCODE database (ENCSR956DNB). Previously published single-cell multifactorial profiling data is available under GSE171554 (multi-CUT&Tag[20]), GSE179756 (MulTI-Tag[21]), GSE212588 (NTT-seq[22]) and GSE198467 (nano-CT[23]). CHART-seq data[24] is available under GSE48649.

All custom code generated as part of this study has been deposited on GitHub.

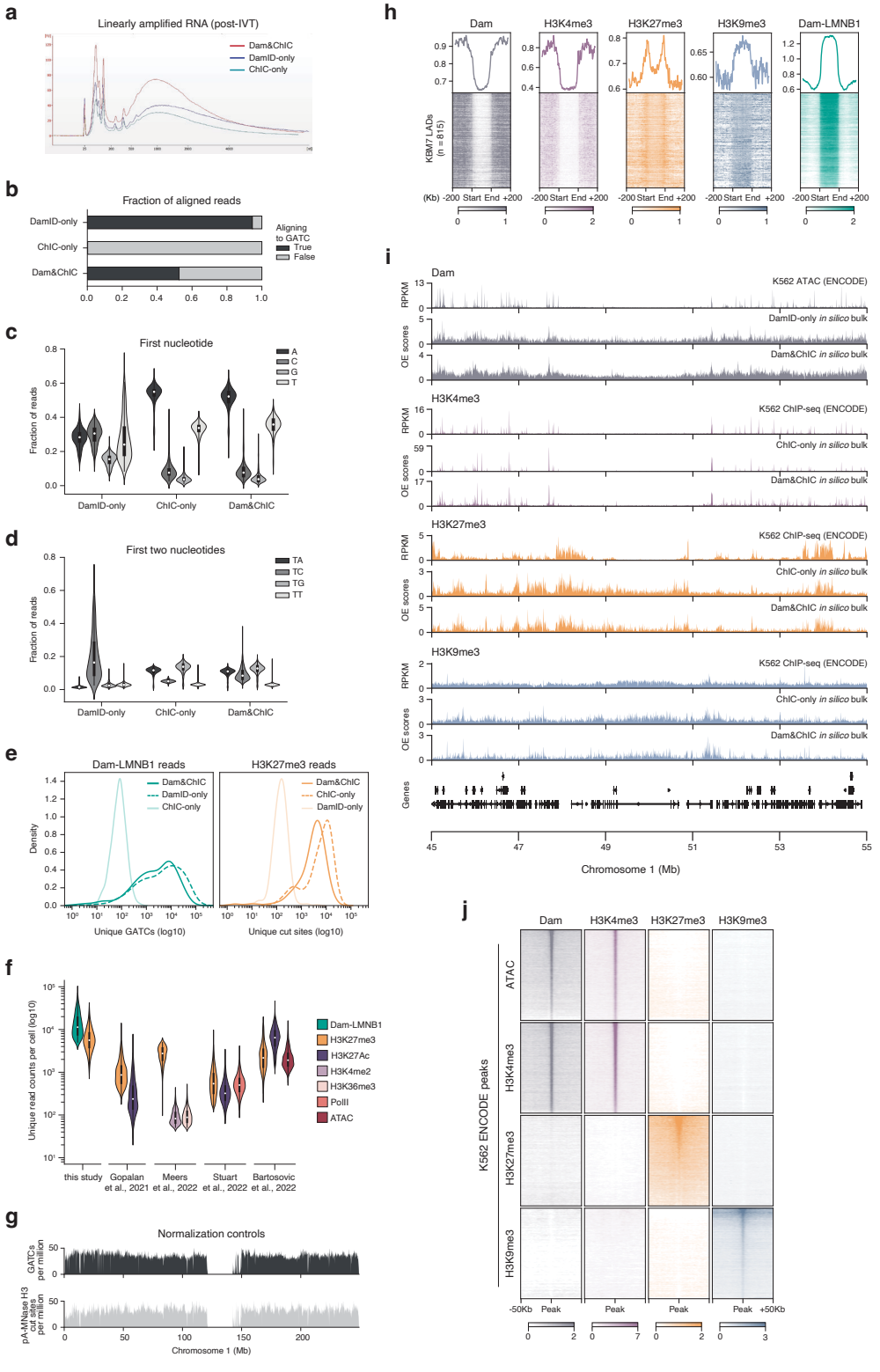

a
Linearly amplified RNA (post-IVT)
Dam&ChIC
DamID-only
ChIC-only
b
Fraction of aligned reads
DamID-only
ChIC-only
Dam&ChIC
Aligning to GATC
True
False
0.0 0.2 0.4 0.6 0.8 1.0
c
First nucleotide
Fraction of reads
A
C
G
T
DamID-only
ChIC-only
Dam&ChIC
d
First two nucleotides
Fraction of reads
TA
TC
TG
TT
DamID-only
ChIC-only
Dam&ChIC
e
Dam-LMNB1 reads
H3K27me3 reads
Density
Dam&ChIC
DamID-only
ChIC-only
Unique GATCs (log10)
Unique cut sites (log10)
f
Unique read counts per cell (log10)
Dam-LMNB1
H3K27me3
H3K27Ac
H3K4me2
H3K36me3
PolII
ATAC
this study
Gopalan et al., 2021
Meers et al., 2022
Stuart et al., 2022
Bartosovic et al., 2022
g
Normalization controls
GATCs per million
pA-MNase H3 cut sites per million
Chromosome 1 (Mb)
h
Dam
H3K4me3
H3K27me3
H3K9me3
Dam-LMNB1
KBM7 LADs (n = 815)
(Kb) -200 Start End +200
i
Dam
K562 ATAC (ENCODE)
DamID-only in silico bulk
Dam&ChIC in silico bulk
H3K4me3
K562 ChIP-seq (ENCODE)
ChIC-only in silico bulk
H3K27me3
H3K9me3
RPKM
OE scores
Genes
Chromosome 1 (Mb)
j
Dam
H3K4me3
H3K27me3
H3K9me3
K562 ENCODE peaks
ATAC
-50Kb Peak
Peak
Peak +50Kb

Figure S4.1.

a, Bioanalyzer plot showing the amount and size distribution of linearly-amplified RNA (aRNA), derived after in vitro transcription with the Dam&ChIC protocol (see Methods). Dam&ChIC was performed to profile Dam-LMNB1 and H3K27me3, alongside respective control DamID-only and ChIC-only experiments. Each aRNA product was derived from one 384-well plate containing 380 cells, barcoded and pooled together prior to amplification. The samples were processed in parallel. **b,** Comparison of the fraction of reads mapping on a genomic GATC motif (black) between Dam&ChIC, DamID-only and ChIC-only. **c,** Violin plots showing a quantification of the first nucleotide of invalid DamID reads (i.e., not mapping on genomic GATC motifs) for Dam&ChIC, DamID-only and ChIC-only. **d,** Violin plot depicting a quantification of the first two nucleotides of invalid DamID reads for Dam&ChIC, DamID-only and ChIC-only. **e,** Distribution of the unique number of Dam-LMNB1 (left) or H3K27me3 (right) reads recovered per cell by DamID-only, ChIC-only or Dam&ChIC. **f,** Comparison of unique number of reads obtained per cell between Dam&ChIC and other recently published multifactorial chromatin profiling methods. **g,** Genomic distribution of in silico GATC mappability (see Methods) (top, black) and experimental pA-MNase H3 mappability (bottom, gray). **h,** Heatmaps showing chromatin features profiled with Dam&ChIC aligned on LADs (n = 815) defined in KBM7 cells. Profiles above show the averages of the heatmaps for each mapped chromatin feature. **i,** Genomic tracks of average (in silico bulk) Dam&ChIC profiles compared to corresponding ChIC-only datasets in KBM7 cells and corresponding ENCODE ChIP-seq or ATAC-seq datasets in K562 cells. **j,** Tornado plots depicting pseudobulk Dam&ChIC data aligned on ENCODE ChIP-seq or ATAC-seq peak coordinates in K562 cells. Peaks (rows) are ordered on the peak signal of corresponding Dam&ChIC dataset (*i.e.*, ATAC-seq peaks are ordered on Dam DamID data, H3K4me3 peaks are ordered on H3K4me3 ChIC data and so on.

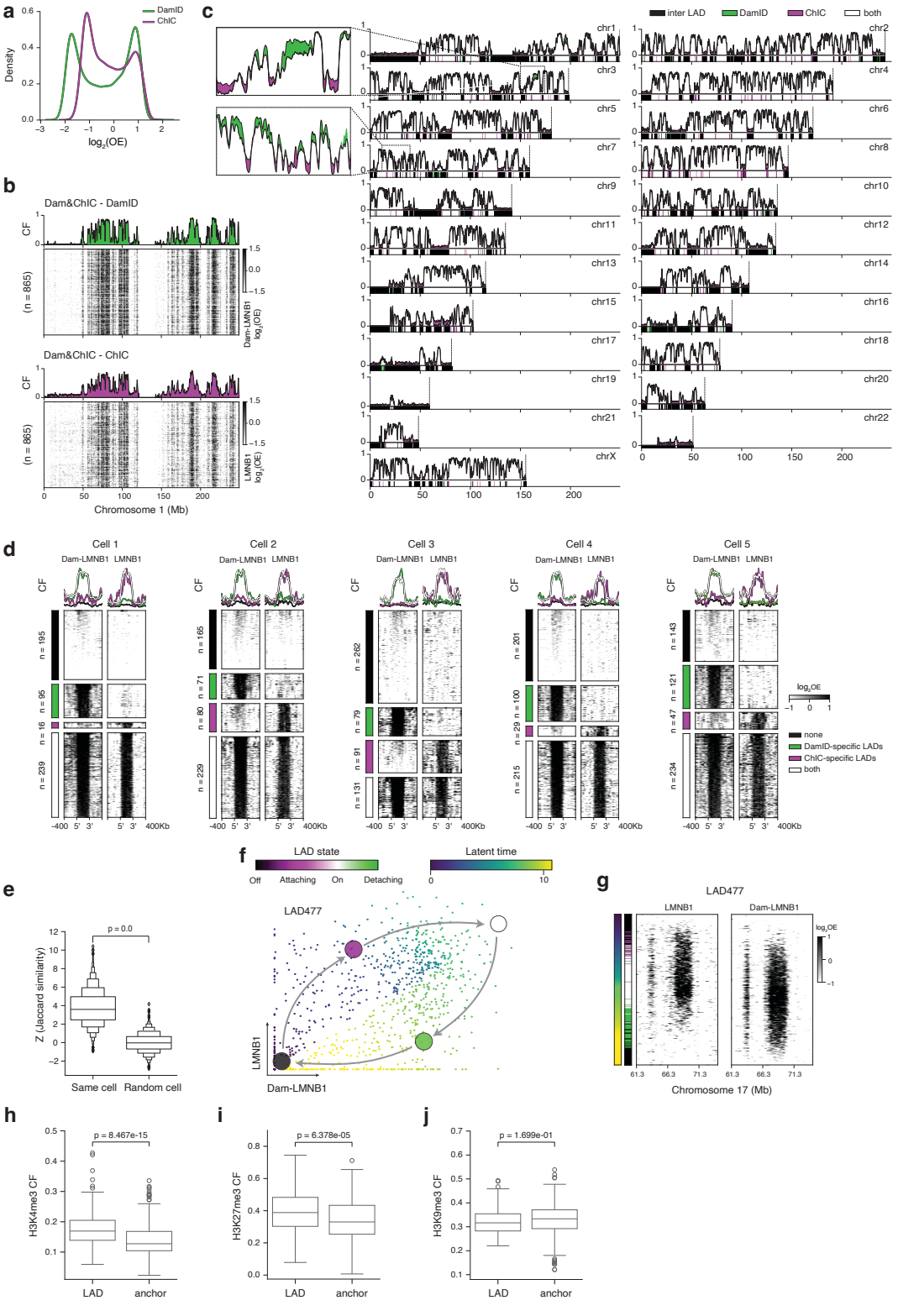

a
DamID
ChIC
Density
log2(OE)
b
Dam&ChIC - DamID
Dam&ChIC - ChIC
CF
(n = 865)
Dam-LMNB1 log2(OE)
LMNB1 log2(OE)
Chromosome 1 (Mb)
c
inter LAD
DamID
ChIC
both
d
Cell 1
Cell 2
Cell 3
Cell 4
Cell 5
Dam-LMNB1
LMNB1
none
DamID-specific LADs
ChIC-specific LADs
both
e
Z (Jaccard similarity)
p = 0.0
Same cell
Random cell
f
LAD state
Off
Attaching
On
Detaching
Latent time
LAD477
LMNB1
Dam-LMNB1
g
LAD477
Chromosome 17 (Mb)
h
H3K4me3 CF
p = 8.467e-15
LAD
anchor
i
H3K27me3 CF
p = 6.378e-05
j
H3K9me3 CF
p = 1.699e-01

Figure S4.2.

a, Distribution of OE values of 100-Kb binned pseudobulk Dam&ChIC data.**b,** Single-cell heatmaps (n = 865) showing OE LAD signal in 100-Kb genomic bins of either the DamID (top) or the ChIC (bottom) readout. Lineplots represent CF values. **c,** Comparison of LAD CF on depicted chromosomes between the DamID (green) and ChIC (magenta) measurement of Dam&ChIC. Overlapping signal is shown in white. LAD calls are shown below each track. **d,** Five example single cells aligned and grouped for all bulk-based LADs according to their detection in the given cell by neither measurement (black), DamID (green), ChIC (magenta) or both (white). DamID (left) and ChIC (right) measurements are split, and respective CF values for each category are shown as line plots at the top, with y axis running from 0 to 1. LAD signal is shown as OE in 100-Kb genomic bins. **e,** Boxenplot of multiple quantiles of z-score normalized Jaccard similarity between the DamID and ChIC readouts from the same cell and any two random cells. Two-sided t-test with resulting p-values is designated. **f,** Phase portrait of LAD477 colored by latent time inferred by scVelo[29], describing dynamics of the LAD based on the ratio of Dam&ChIC LMNB1 data (DamID: x and ChIC: y axes). LAD states are depicted with larger dots according to the likelihood of being off (black), attaching to the lamina (magenta), on (white) or detaching from the lamina (green). **g,** Single-cell heatmaps for LAD477, with the DamID (right) and ChIC (left) measurements split and single cells sorted according to LAD477-based latent time (leftmost bar). The middle bar at the left indicates the LAD state (see Methods). **h-j,** Boxplot showing (**h**) H3K4me3, (**i**) H3K27me3, and (**j**) H3K9me3 CF values, quantified in 100-Kb genomic bins on anchor regions (*i.e.* LAD region where LAD starts attaching and is released; Methods) compared to the rest of the LAD. Two-sided t-test with resulting p-values are designated.

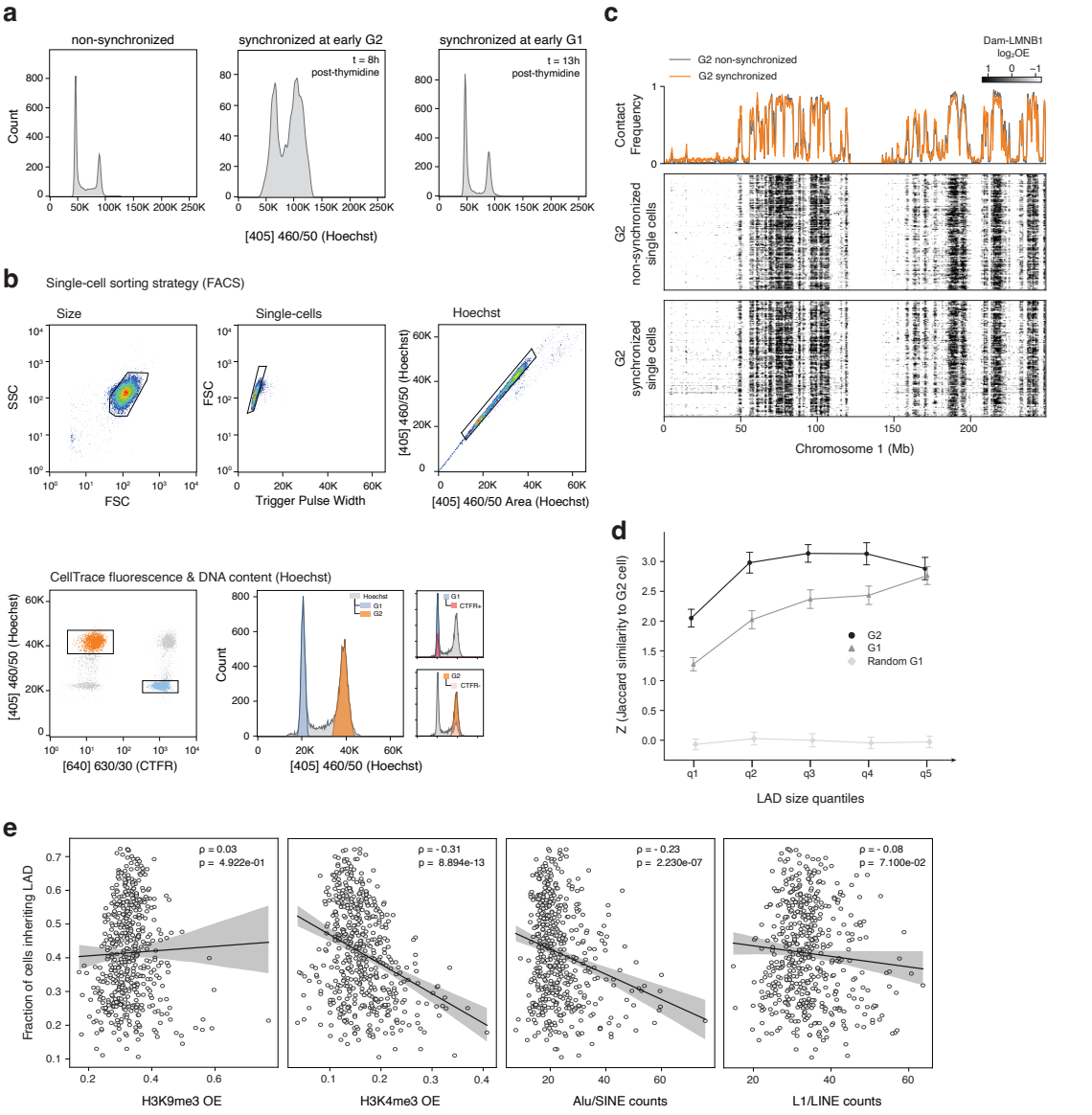

a
non-synchronized
synchronized at early G2
synchronized at early G1
t = 8h post-thymidine
t = 13h post-thymidine
Count
[405] 460/50 (Hoechst)
b
Single-cell sorting strategy (FACS)
Size
Single-cells
Hoechst
SSC
FSC
Trigger Pulse Width
[405] 460/50 (Hoechst)
[405] 460/50 Area (Hoechst)
CellTrace fluorescence & DNA content (Hoechst)
[640] 630/30 (CTFR)
Hoechst
G1
G2
CTFR+
CTFR-
c
G2 non-synchronized
G2 synchronized
Dam-LMNB1 log2OE
Contact Frequency
G2 non-synchronized single cells
G2 synchronized single cells
Chromosome 1 (Mb)
d
Z (Jaccard similarity to G2 cell)
G2
G1
Random G1
LAD size quantiles
e
Fraction of cells inheriting LAD
ρ = 0.03
p = 4.922e-01
ρ = - 0.31
p = 8.894e-13
ρ = - 0.23
p = 2.230e-07
ρ = - 0.08
p = 7.100e-02
H3K9me3 OE
H3K4me3 OE
Alu/SINE counts
L1/LINE counts

Figure S4.3.

a, Hoechst-based DNA content distributions of the non-synchronized control and the synchronized in G2 and synchronized in G1 samples prior to mixing and single-cell sorting for Dam&ChIC. **b,** Single-cell sorting strategy (see Methods), including the used gates based on size, singlets, Hoechst staining and CellTrace stainings. CellTrace Far Red is noted as CTFR. **c,** Single-cell heatmaps and corresponding Contact Frequency quantifications, comparing G2 synchronized with G2 non-synchronized cells. **d,** Z-score normalized Jaccard similarity between DamID and ChIC readouts within G2 cells, within G1 cells, and between random G1 cells, across different LAD size quantiles. **e,** Scatterplots for the recovery of frequency of inherited LADs (n = 516) against different (epi)genomic features. Y axis is the same across all plots. The black lines represent a linear regression fit with its 95% CI in the shaded area. The Pearson Correlation coefficient (ρ) and p-value of two-sided testing for non-correlation are indicated.

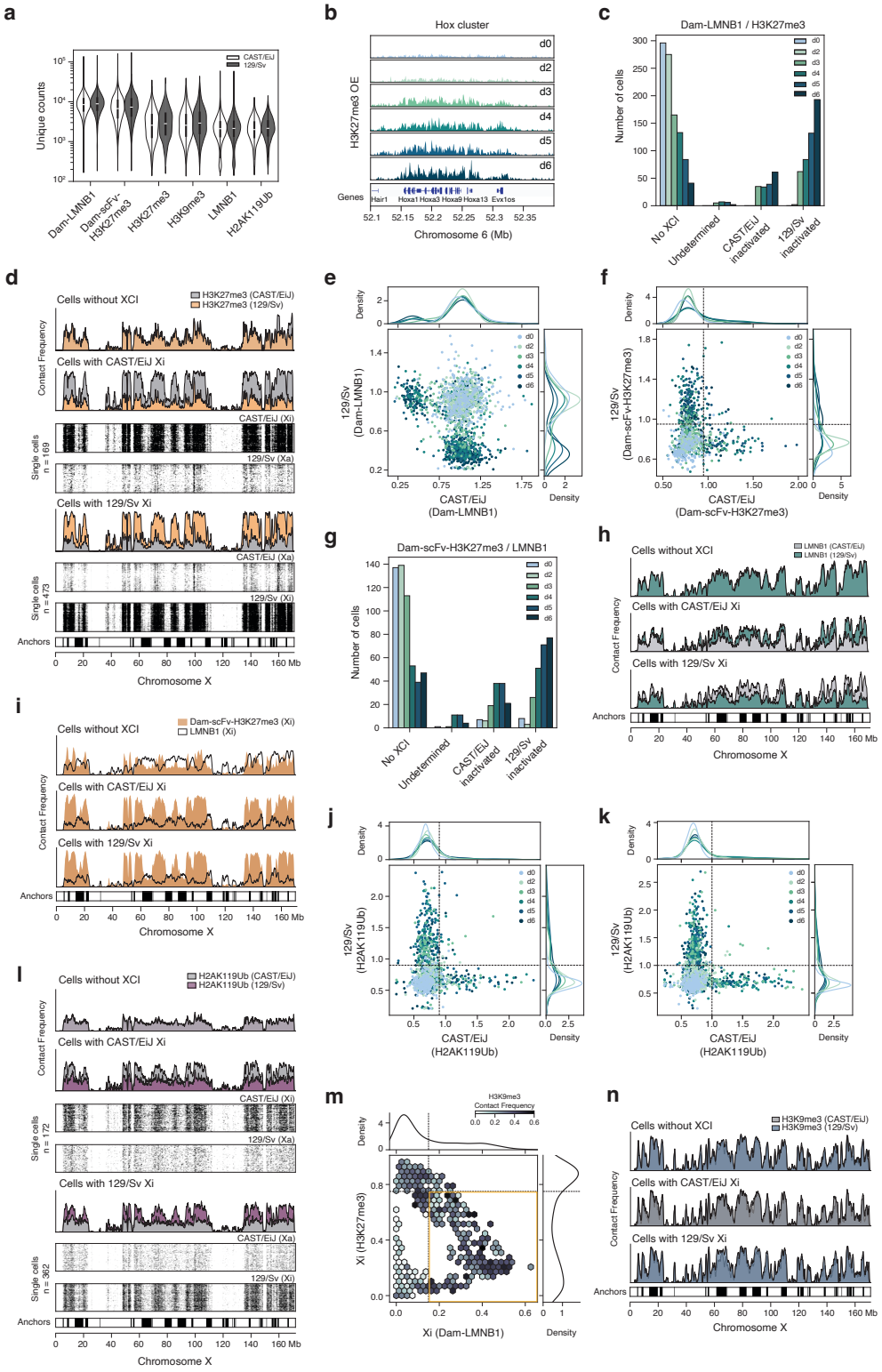

a
Unique counts
CAST/EiJ
129/Sv
Dam-LMNB1
Dam-scFv-H3K27me3
H3K27me3
H3K9me3
LMNB1
H2AK119Ub
b
Hox cluster
H3K27me3 OE
d0
d2
d3
d4
d5
d6
Genes
Hair1
Hoxa1 Hoxa3 Hoxa9 Hoxa13
Evx1os
52.1 52.15 52.2 52.25 52.3 52.35
Chromosome 6 (Mb)
c
Dam-LMNB1 / H3K27me3
Number of cells
No XCI
Undetermined
CAST/EiJ inactivated
129/Sv inactivated
d
Cells without XCI
H3K27me3 (CAST/EiJ)
H3K27me3 (129/Sv)
Contact Frequency
Cells with CAST/EiJ Xi
CAST/EiJ (Xi)
Single cells n = 169
129/Sv (Xa)
Cells with 129/Sv Xi
CAST/EiJ (Xa)
Single cells n = 473
129/Sv (Xi)
Anchors
Chromosome X
e
Density
129/Sv (Dam-LMNB1)
CAST/EiJ (Dam-LMNB1)
f
129/Sv (Dam-scFv-H3K27me3)
CAST/EiJ (Dam-scFv-H3K27me3)
g
Dam-scFv-H3K27me3 / LMNB1
Number of cells
h
LMNB1 (CAST/EiJ)
LMNB1 (129/Sv)
i
Dam-scFv-H3K27me3 (Xi)
LMNB1 (Xi)
j
129/Sv (H2AK119Ub)
CAST/EiJ (H2AK119Ub)
k
l
H2AK119Ub (CAST/EiJ)
H2AK119Ub (129/Sv)
Single cells n = 172
Single cells n = 362
m
H3K9me3 Contact Frequency
Xi (H3K27me3)
Xi (Dam-LMNB1)
n
H3K9me3 (CAST/EiJ)
H3K9me3 (129/Sv)

Figure S4.4.

a, Violin plots depicting the number of unique counts per cell and allele for the various readouts obtained with Dam&ChIC in the Vitamin C differentiation trajectory. **b,** OE levels of H3K27me3 over the Hoxa gene cluster on Chromosome 6, divided by experimental time point. **c,** Number of cells in each Vitamin C differentiation day for all categories of cells based on allelic H3K27me3 levels in **4.4c**. Undetermined cells are defined as those with high levels of H3K27me3 on both alleles (*i.e.* top right quadrant of **4.4c**). **d,** CF values of H3K27me3 on either allele for the cell categories defined in **4.4c**, along the entire X chromosome. The y axis in all CF plots run from 0 to 1. Opacity in the "no XCI" CF plot is decreased to allow visualization of both alleles. For the categories that undergo XCI, single-cell heatmaps of the active (Xa) and inactive (Xi) X chromosome are shown below, with the respective cell numbers, independent of differentiation time point. Bottom bar shows calls of the anchors. **e,** Scatterplot of allelic Dam-LMNB1 levels per single cell in the Dam-LMNB1/H3K27me3 dataset, colored by differentiation day. **f,** Scatterplot of allelic Dam-scFv-H3K27me3 levels per single cell in the Dam-scFv-H3K27me3/LMNB1 dataset, colored by differentiation day. The dashed lines indicate the manually set threshold used to classify cells as no XCI, CAST/EiJ inactive or 129S1/Sv inactive. **g,** Number of cells in each Vitamin C differentiation day for all categories of cells based on Dam-scFc-H3K27me3 in (S4.4f). **h,** Contact Frequency values of LMNB1 on either allele for the cell categories defined in **S4.4f** along the entire X chromosome. The y axes run from 0 to 1. Bottom bar shows calls of the anchors. **i,** CF values of LMNB1 and corresponding Dam-scFv-H3K27me3 values on either allele in the inactive state for the cell categories defined in **S4.4f** along the entire X chromosome. The y axes run from 0 to 1. Bottom bar shows calls of the anchors. **j,** Scatterplot of allelic H2AK119ub levels per single cell in the Dam-LMNB1/H2AK119ub dataset, colored by differentiation day. The dashed lines indicate the manually set threshold used to classify cells as no XCI, CAST/EiJ inactive or 129S1/Sv inactive. **k,** Scatterplot of allelic H2AK119ub levels per single cell in the Dam-scFv-H3K27me3/H2AK119ub dataset, colored by differentiation day. The dashed lines indicate the manually set threshold used to classify cells as no XCI, CAST/EiJ inactive or 129S1/Sv inactive. **l,** CF values of H2AK119ub on either allele for the cell categories defined in (S4.4j), along the entire X chromosome, plotted as in **S4.4d**. The H2AK119ub readout derives from the Dam-LMNB1/H2AK119ub dataset. **m,** Hexbin plots of Dam-LMNB1 (x-axis) and H3K27me3 (y-axis) CF for all X chromosome bins. Each hexbin is a collection of X chromosome bins. Dashed lines indicate the manually-set thresholds used to classify regions as anchors (yellow box; bottom right quadrant). Color projection on hexbins is average H3K9me3 CF values, showing an enrichment at anchor regions. **n,** Contact Frequency values of H3K9me3 on either allele for the cell categories defined as in **S4.4e**, along the entire X chromosome. Opacities are decreased to allow visualization of both allelic categories. The H3K9me3 readout derives from the Dam-LMNB1/H3K9me3 dataset. The y axes run from 0 to 1.

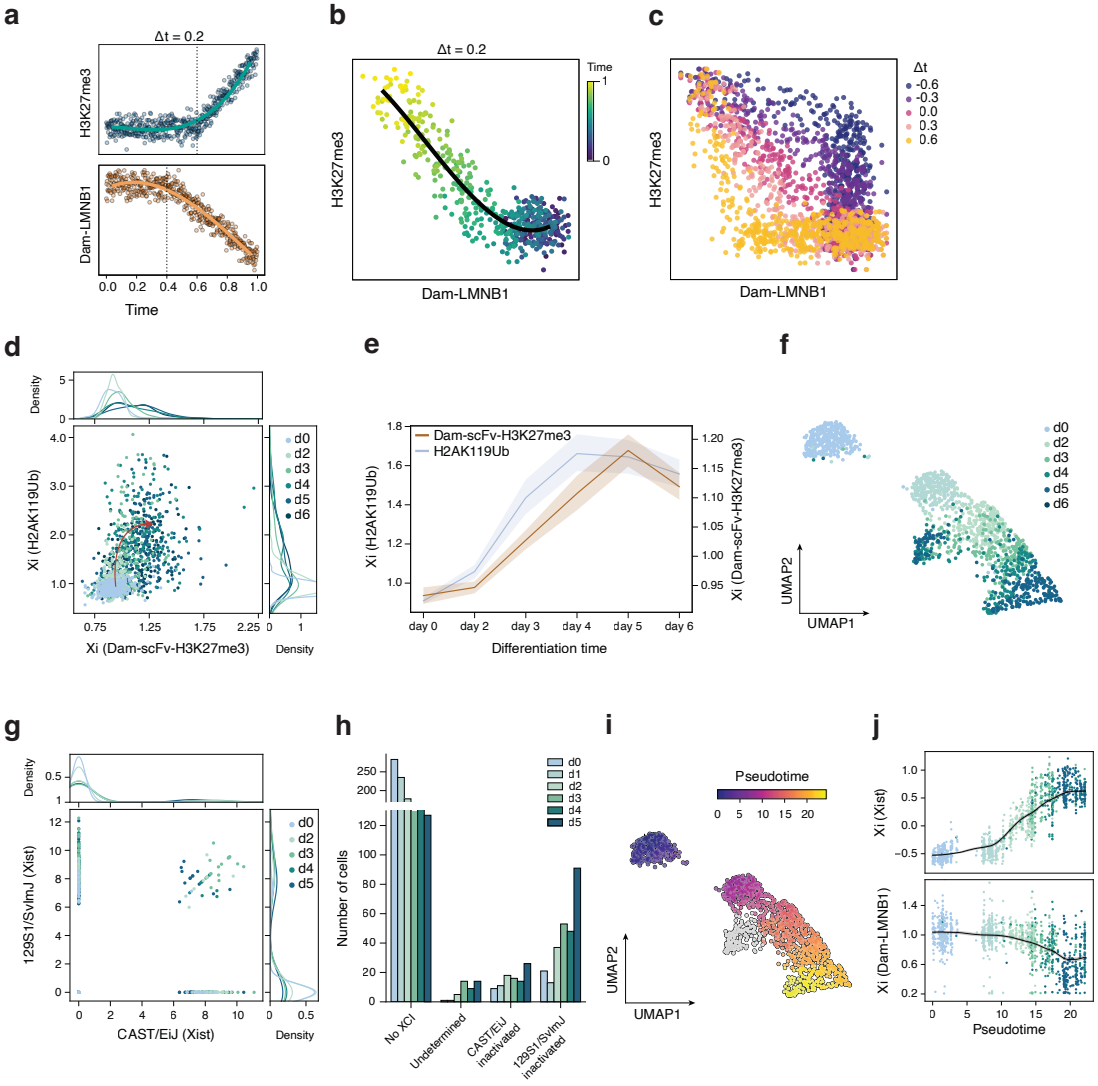

a
Δt = 0.2
H3K27me3
Dam-LMNB1
0.0 0.2 0.4 0.6 0.8 1.0
Time
b
Δt = 0.2
Time
1
0
H3K27me3
Dam-LMNB1
c
Δt
-0.6
-0.3
0.0
0.3
0.6
H3K27me3
Dam-LMNB1
d
Density
Xi (H2AK119Ub)
d0
d2
d3
d4
d5
d6
Xi (Dam-scFv-H3K27me3)
Density
e
Dam-scFv-H3K27me3
H2AK119Ub
Xi (H2AK119Ub)
Xi (Dam-scFv-H3K27me3)
day 0
day 2
day 3
day 4
day 5
day 6
Differentiation time
f
UMAP2
UMAP1
d0
d2
d3
d4
d5
d6
g
Density
129S1/SvImJ (Xist)
CAST/EiJ (Xist)
d0
d2
d3
d4
d5
Density
h
Number of cells
d0
d1
d2
d3
d4
d5
No XCI
Undetermined
CAST/EiJ inactivated
129S1/SvImJ inactivated
i
Pseudotime
0 5 10 15 20
UMAP2
UMAP1
j
Xi (Xist)
Xi (Dam-LMNB1)
Pseudotime

Figure S4.5.

a, Simulated single-cell H3K27me3 (top) and Dam-LMNB1 (bottom) values, where the onset of both dynamics is controlled to enable comparison of their relative timing. Values are generated by randomly sampling from a log-normal distribution and the onset of dynamics is simulated by randomly sampling from a Poisson distribution. Simulated time difference (Δt) is 0.2. **b,** Scatterplot of simulated Dam-LMNB1 (x-axis) and H3K27me3 (y-axis) values from (S4.5a). Cells are colored according to simulated time. **c,** Same as (S4.5b), but with cells following and colored according to progressive Δt values. **d,** Scatterplot of the average inactive X (Xi) Dam-scFv-H3K27me3 (x-axis) and H2AK119ub (y-axis) observed over expected (OE) levels. Cells are colored according to differentiation day. Distributions above and on the right show average density per day. Arrow is manually drawn based on directionality of cells over differentiation days. **e,** Lineplot showing Dam-scFv-H3K27me3 & H2AK119ub dynamics on the Xi over differentiation. Values plotted as average observed over expected (OE) with the 95% confidence interval in the shaded area. **f,** UMAP (n = 1427 cells) based on transcriptomic data of scDam&T-seq, colored by differentiation day. **g,** Scatterplot of allelic *Xist* expression per single cell, colored by differentiation day. **h,** Number of cells in each Vitamin C differentiation day for all categories of cells based on allelic Xist expression in **S4.5g**. **i,** Same UMAP as in **S4.5f** colored by Monocle3-inferred pseudotime[57]. **j,** *Xist* expression (top) and Dam-LMNB1 (bottom) levels (y-axis) across pseudotime (x-axis) are shown over the Xi. *Xist* levels are represented as transcripts per million and Dam-LMNB1 levels as average OE. Cells are colored according to differentiation day.

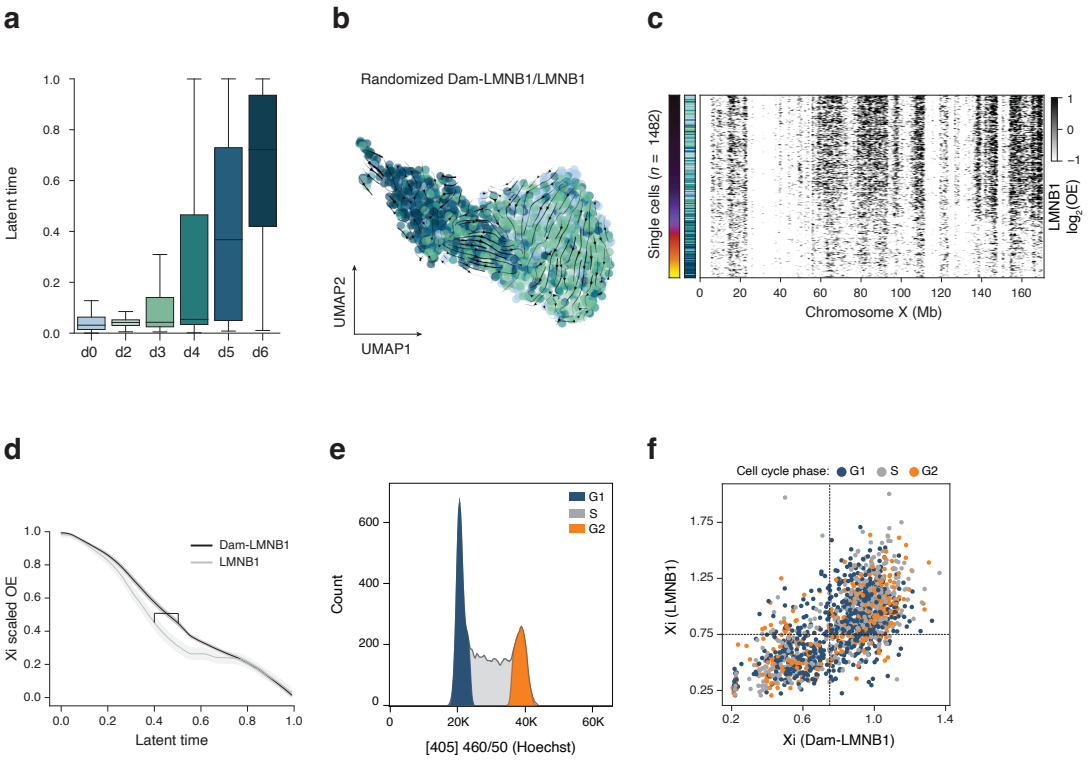


Figure S4.6.

a, Distribution of cells from Vitamin C differentiation days across latent time defined in **6b**. **b,** Same UMAP as in **6a**, based on Dam-LMNB1 levels in 100Kb-bins of the Xi, with streamlines representing chromatin velocities of randomly coupled Dam-LMNB1 and LMNB1 readouts per cell. **c,** Single-cell heatmap, showing log-transformed LMNB1 OE values on the entire inactivating X-chromosome, with cells ordered along the inferred latent time shown in **6b**. **d,** Average Dam-LMNB1 (black) and LMNB1 (grey) OE levels of the inactive X chromosome along latent time. Gap marks the moment of highest genome–lamina dynamics on the Xi. **e,** Example Hoechst profile of 380 sorted single cells from the Dam-LMNB1/LMNB1 Vitamin C experiment, showing their DNA content and respective cell-cycle stage. **f,** Same scatter plot as in **6g** with raw cell cycle-phase annotations determined based on Hoechst values.

References

1. Millán-Zambrano, G., Burton, A., Bannister, A.J. & Schneider, R. Histone post-translational modifications - cause and consequence of genome function. *Nat Rev Genet* **23**, 563-580 (2022).
2. van Steensel, B. & Furlong, E.E.M. The role of transcription in shaping the spatial organization of the genome. *Nat Rev Mol Cell Biol* **20**, 327-337 (2019).
3. van Steensel, B. & Belmont, A.S. Lamina-Associated Domains: Links with Chromosome Architecture, Heterochromatin, and Gene Repression. *Cell* **169**, 780-791 (2017).
4. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* **39**, 825-835 (2021).
5. Buenrostro, J.D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-90 (2015).
6. Kaya-Okur, H.S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).
7. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-47 (2015).
8. Ku, W.L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods* **16**, 323-325 (2019).
9. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).
10. Wu, S.J. *et al.* Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat Biotechnol* **39**, 819-824 (2021).
11. Zeller, P. *et al.* Single-cell sortChIC identifies hierarchical chromatin dynamics during hematopoiesis. *Nat Genet* **55**, 333-345 (2023).
12. Gopalan, S., Wang, Y., Harper, N.W., Garber, M. & Fazzio, T.G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol Cell* **81**, 4736-4746.e5 (2021).
13. Meers, M.P., Llagas, G., Janssens, D.H., Codomo, C.A. & Henikoff, S. Multifactorial profiling of epigenetic landscapes at single-cell resolution using MulTI-Tag. *Nat Biotechnol* **41**, 708-716 (2023).
14. Bartosovic, M. & Castelo-Branco, G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nat Biotechnol* **41**, 794-805 (2023).
15. Stuart, T. *et al.* Nanobody-tethered transposition enables multifactorial chromatin profiling at single-cell resolution. *Nat Biotechnol* **41**, 806-812 (2023).
16. Lochs, S.J.A. *et al.* Combinatorial single-cell profiling of major chromatin types with MAbID. *Nat Methods* **21**, 72-82 (2024).
17. Boers, R. *et al.* Retrospective analysis of enhancer activity and transcriptome history. *Nat Biotechnol* **41**, 1582-1592 (2023).
18. Guelen, L. *et al.* Domain organization of human chromosomes revealed by mapping of nuclear lamina interactions. *Nature* **453**, 948-951 (2008).
19. Rang, F.J. *et al.* Single-cell profiling of transcriptome and histone modifications with EpiDamID. *Mol Cell* **82**, 1956-1970.e14 (2022).
20. Rooijers, K. *et al.* Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat Biotechnol* **37**, 766-772 (2019).
21. van Steensel, B. & Henikoff, S. Identification of *in vivo* DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* **18**, 424-8 (2000).
22. Vogel, M.J., Peric-Hupkes, D. & van Steensel, B. Detection of *in vivo* protein-DNA interactions using DamID in mammalian cells. *Nat Protoc* **2**, 1467-78 (2007).
23. Schmid, M., Durussel, T. & Laemmli, U.K. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* **16**, 147-57 (2004).
24. Skene, P.J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *elife* **6**, e21856 (2017).
25. Dingwall, C., Lomonossoff, G.P. & Laskey, R.A. High sequence specificity of micrococcal nuclease. *Nucleic Acids Res* **9**, 2659-73 (1981).
26. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178-92 (2013).
27. La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494-498 (2018).
28. Tedesco, M. *et al.* Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. *Nat Biotechnol* **40**, 235-244 (2022).
29. Bergen, V. *et al.* Generalizing RNA velocity

to transient cell states through dynamical modeling. *Nat Biotechnol* **38**, 1408-1414 (2020).

30. Lyon, M.F. Gene action in the X-chromosome of the mouse (Mus musculus L.). *Nature* **190**, 372-3 (1961).
31. Lyon, M.F. Sex chromatin and gene action in the mammalian X-chromosome. *Am J Hum Genet* **14**, 135-48 (1962).
32. Heard, E., Chaumeil, J., Masui, O. & Okamoto, I. Mammalian X-chromosome inactivation: an epigenetics paradigm. *Cold Spring Harb Symp Quant Biol* **69**, 89-102 (2004).
33. Brown, C.J. *et al.* The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527-42 (1992).
34. Panning, B., Dausman, J. & Jaenisch, R. X chromosome inactivation is mediated by Xist RNA stabilization. *Cell* **90**, 907-16 (1997).
35. Penny, G.D., Kay, G.F., Sheardown, S.A., Rastan, S. & Brockdorff, N. Requirement for Xist in X chromosome inactivation. *Nature* **379**, 131-7 (1996).
36. Plath, K. *et al.* Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131-5 (2003).
37. Silva, J. *et al.* Establishment of histone h3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev Cell* **4**, 481-95 (2003).
38. Żylicz, J.J. *et al.* The implication of early chromatin changes in X chromosome inactivation. *Cell* **176**, 182-197. e23 (2019).
39. Żylicz, J.J. & Heard, E. Molecular mechanisms of facultative heterochromatin formation: an X-chromosome perspective. *Annual Review of Biochemistry* **89**, 255-282 (2020).
40. Belmont, A.S., Bignone, F. & Ts'o, P.O. The relative intranuclear positions of Barr bodies in XXX non-transformed human fibroblasts. *Exp Cell Res* **165**, 165-79 (1986).
41. Bourgeois, C.A., Laquerriere, F., Hemon, D., Hubert, J. & Bouteille, M. New data on the in-situ position of the inactive X chromosome in the interphase nucleus of human fibroblasts. *Hum Genet* **69**, 122-9 (1985).
42. Chen, C.K. *et al.* Xist recruits the X chromosome to the nuclear lamina to enable chromosome-wide silencing. *Science* **354**, 468-472 (2016).
43. Rego, A., Sinclair, P.B., Tao, W., Kireev, I. & Belmont, A.S. The facultative heterochromatin of the inactive X chromosome has a distinctive condensed ultrastructure. *J Cell Sci* **121**, 1119-27 (2008).
44. Teller, K. *et al.* A top-down analysis of Xa- and Xi-territories reveals differences of higher order structure at >/= 20 Mb genomic length scales. *Nucleus* **2**, 465-77 (2011).
45. Zhang, L.F., Huynh, K.D. & Lee, J.T. Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell* **129**, 693-706 (2007).
46. Augui, S., Nora, E.P. & Heard, E. Regulation of X-chromosome inactivation by the X-inactivation centre. *Nat Rev Genet* **12**, 429-42 (2011).
47. Guerreiro, I. *et al.* Antagonism between H3K27me3 and genome–lamina association drives atypical spatial genome organization in the totipotent embryo. *Nat Genet* **56**, 2228-2237 (2024).
48. Loda, A. *et al.* Genetic and epigenetic features direct differential efficiency of Xist-mediated silencing at X-chromosomal and autosomal locations. *Nat Commun* **8**, 690 (2017).
49. Robert-Finestra, T. *et al.* SPEN is required for Xist upregulation during initiation of X chromosome inactivation. *Nat Commun* **12**, 7000 (2021).
50. Vallot, C. *et al.* Erosion of X Chromosome Inactivation in Human Pluripotent Cells Initiates with XACT Coating and Depends on a Specific Heterochromatin Landscape. *Cell Stem Cell* **16**, 533-46 (2015).
51. Simon, M.D. *et al.* High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465-469 (2013).
52. Cao, J. *et al.* The single-cell transcriptional landscape of mammalian organogenesis. *Nature* **566**, 496-502 (2019).
53. Trapnell, C. *et al.* The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* **32**, 381-386 (2014).
54. Oksuz, O. *et al.* Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol Cell* **70**, 1149-1162 e5 (2018).
55. Markodimitraki, C.M. *et al.* Simultaneous quantification of protein-DNA interactions and transcriptomes in single cells with scDam&T-seq. *Nat Protoc* **15**, 1922-1953 (2020).
56. Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y. & Greenleaf, W.J. Transposition

of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat Methods* **10**, 1213-8 (2013).

57. Kaya-Okur, H.S., Janssens, D.H., Henikoff, J.G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nat Protoc* **15**, 3264-3283 (2020).
58. de Luca, K.L. *et al.* Genome-wide profiling of DNA repair proteins in single cells. *Nat Commun* **15**, 9918 (2024).
59. Baron, C.S. *et al.* Cell Type Purification by Single-Cell Transcriptome-Trained Sorting. *Cell* **179**, 527-542 e19 (2019).
60. Manzo, S.G., Dauban, L. & van Steensel, B. Lamina-associated domains: Tethers and looseners. *Curr Opin Cell Biol* **74**, 80-87 (2022).
61. Rullens, P.M.J. & Kind, J. Attach and stretch: Emerging roles for genome–lamina contacts in shaping the 3D genome. *Curr Opin Cell Biol* **70**, 51-57 (2021).
62. Horton, J.R., Liebert, K., Hattman, S., Jeltsch, A. & Cheng, X. Transition from nonspecific to specific DNA interactions along the substrate-recognition pathway of dam methyltransferase. *Cell* **121**, 349-61 (2005).
63. Pollex, T. & Heard, E. Nuclear positioning and pairing of X-chromosome inactivation centers are not primary determinants during initiation of random X-inactivation. *Nat Genet* **51**, 285-295 (2019).
64. Siegenfeld, A.P. *et al.* Polycomb-lamina antagonism partitions heterochromatin at the nuclear periphery. *Nat Commun* **13**, 4199 (2022).

Chapter 5

Genome-wide profiling of DNA repair proteins in single cells

Kim L. de Luca[1,2*‡], Pim M. J. Rullens[1,2,3*], Magdalena A. Karpinska[4], Sandra S. de Vries[1,2], Agnieszka Gacek-Matthews[5], Lőrinc S. Pongor[6], Gaëlle Legube[7], Joanna W. Jachowicz[5], A. Marieke Oudelaar[4], Jop Kind[1,2,3‡]

[1] Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht; Utrecht, The Netherlands

[2] Oncode Institute; The Netherlands

[3] Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University; Nijmegen, The Netherlands

[4] Genome Organization and Regulation, Max Planck Institute for Multidisciplinary Sciences; Göttingen, Germany

[5] Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna Biocenter (VBC); Vienna, Austria

[6] Cancer Genomics and Epigenetics Core Group, Hungarian Center of Excellence for Molecular Medicine (HCEMM); Szeged, Hungary

[7] MCD, Centre de Biologie Intégrative (CBI), CNRS, Université de Toulouse, Université Toulouse III – Paul Sabatier; Toulouse, France

* Co-first authorship

‡ Co-corresponding authorship

Adapted from article published in ***Nature Communications*** (2024).
Author list reproduced as published.

Abstract

Accurate repair of DNA damage is critical for maintenance of genomic integrity and cellular viability. Because damage occurs non-uniformly across the genome, single-cell resolution is required for proper interrogation, but sensitive detection has remained challenging. Here, we present a comprehensive analysis of repair protein localization in single human cells using DamID and ChIC sequencing techniques. We validate this method by benchmarking against ChIP-seq and imaging-based approaches, demonstrating its high sensitivity and ability to detect repair domains at individual damage sites.

This study reports genome-wide binding profiles in response to DNA double-strand breaks induced by AsiSI, and explores variability in genomic damage locations and associated repair features in the context of spatial genome organization. By unbiasedly detecting repair factor localization, we find that repair proteins often occupy entire topologically associating domains, mimicking variability in chromatin loop anchoring. This work provides a scalable and multimodal framework for studying DNA repair dynamics at single-cell resolution, with applications in spatial genome organization and chromatin biology.

Introduction

The eukaryotic nucleus is constantly exposed to endogenous and exogenous sources of damage to the genome. Among these, double-stranded breaks (DSBs) in the DNA are particularly hazardous lesions because they completely sever the DNA fiber, leaving the genome at risk of small nucleotide changes and larger structural aberrations such as translocations and deletions. These types of genomic instability are associated with tumorigenesis as well as aging-related diseases. To ensure genome integrity, the cell is dependent on the DNA damage response (DDR), an intricate signaling cascade that includes recognition, processing, and restoration of the lesion[1]. The two main groups of DSB repair pathways are end joining (EJ) and homology-directed repair (HDR). EJ involves re-ligation of minimally processed DNA ends and can occur throughout interphase, while HDR requires a homologous template sequence (usually its sister chromatid) and is therefore generally restricted to S and G2 phase[2]. Processing and repair outcome are thus influenced by the phase of the cell cycle, complexity of the DSB, as well as its transcriptional status, chromatin environment, and location within the nucleus[3,4]. In response to damage, DDR proteins accumulate into DNA repair foci, which are formed by local rearrangement of the chromatin at the level of topologically associating domains (TADs)[5].

Further, damaged chromatin exhibits more large-scale mobility, forming clusters of DSBs that reside in specific sub-compartments[6,7].

Given the variability in the occurrence and repair of DNA damage across cells, it becomes imperative to collect information from individual cells to accurately profile their distribution. Single-cell detection of spontaneous damage was described using whole-genome amplification[8], but insight into DNA repair at this level has not been reported from a genomic perspective. In this study, we address this technological gap by presenting a detailed analysis of DSB repair factor localization in single cells. We use two methods to map DNA-binding proteins, namely DamID[9] and ChIC[10], combined with a new computational framework for signal detection, and compare our approach to the state of the art. Additionally, we report simultaneous measurement of repair protein signal (with DamID) and chromatin features such as histone modifications or structural proteins (with ChIC) in the same cell[11], enabling direct analysis of the interplay between repair proteins and the damaged chromatin substrate.

A distinctive property of single-cell data lies in the ability to measure signal across the entire genome on a per-cell basis. With a sufficient number of cells, and diversity in signal among those cells, patterns can start to emerge that reflect underlying processes of interest.

To illustrate this concept, we induce damage at many (~100) known locations in the human genome with the DIvA system[12], and quantify repair protein binding at all sites within individual cells. Specifically, we investigate whether sites are simultaneously occupied by the repair machinery, referred to as coordination. We explore such coordination in the context of damage-specific genome reorganization. Overall, our data reveal heterogeneity in repair protein localization that was previously unappreciated. We demonstrate the utility of (multifactorial) protein profiling in single cells, setting the stage for future investigations into DNA repair and genome stability.

Detecting double-strand break repair proteins genome-wide with DamID and ChIC

We established an experimental and computational workflow to unbiasedly identify DSB repair profiles in single cells (Fig. 5.1a). Using DamID[9,13-15] and ChIC[10,16,17], we measured genomic contacts of proteins involved in EJ and HDR, choosing different Dam-fusion proteins (for DamID) or antibodies (for ChIC). After filtering on quality criteria (see Methods), we obtained a collective total of ~15,000 single-cell profiles of DSB repair proteins 53BP1, MDC1, and γH2AX, and HDR-specific proteins RAD51 and

BRCA1 (Fig. 5.1b, Supplementary Fig. 5.1a-b). We used the previously established human DIvA cell line, which generates DSBs at sequence-specific positions in the genome by the endonuclease AsiSI-ER under control of 4-hydroxytamoxifen (4OHT)[12,18].

To determine significant enrichments of repair occupancy on the genome, we developed a computational method for single-cell domain calling. We modified a multi-scale representation of genomic signals (MSR) approach[19] that unbiasedly identifies enrichments of variable sizes (Sup. Fig. 5.2, Methods). Upon damage induction, specific repair signal accumulates at DSB sites (Fig. 5.1b, Sup. Fig. 5.1c-e), which compares well with publicly available population-based ChIP-seq datasets (Sup. Fig. 5.1e-g). Measured across all cells and DSB sites, we calculated the proportion of cells in which each DSB was bound by the different target proteins (referred to as repair protein frequency). Our collective dataset indicates superior sensitivity compared to previous reports (Fig. 5.1c), allowing for the study of a broader range of infrequently captured DSB sites. In accordance with the DDR signaling cascade, nearly all DSB sites are most frequently occupied by mediator of DNA damage checkpoint protein 1 (MDC1) and ubiquitous DSB marker γH2AX (the direct chromatin substrate for MDC1 binding), followed by EJ-promoting 53BP1, and HDR-specific RAD51 and BRCA1 (Fig. 5.1c). Nonetheless, repair signatures clearly vary across sites, distinguishing between EJ and HDR preference. To further quantify the repair protein domains per cell, we defined a set of single-cell profiles in which we could most accurately determine the number of MSR segments, also referred to as repair protein domains (see Methods). At the median sequencing depth (3×10^3 UMIs), DSB-treated and control cells respectively contained 28 and 3 repair-enriched segments, reaching a plateau at 50 and 7 segments (Sup. Fig. 5.1h), in agreement with the number of AsiSI-induced 53BP1 foci detected by imaging[20].

DSB repair has been widely studied with imaging-based methods that detect proteins, while DamID relies on detection of genomic DNA that has been marked by the Dam-fusion protein. We previously developed the ^{m6}A-Tracer system for visualization and tracking of Dam-methylated contacts[21], and apply it here to validate that HDR-specific Dam-RAD51 generates foci that visually overlap with γH2AX, a universal DSB marker (Fig. 5.1d). Of the DamID RAD51 signal, ~65% colocalizes with γH2AX; vice versa, a smaller ~30% of γH2AX signal corresponds to Dam-RAD51 (Fig. 5.1e). The latter is anticipated to be lower since not all γH2AX-marked DSBs are repaired by RAD51, as can also be observed from the smaller number of Dam-RAD51 foci.

Together, these results show that implementation of DamID and ChIC can be used for sensitive and specific single-cell genomics as well as quantitative imaging analyses of DSB repair.

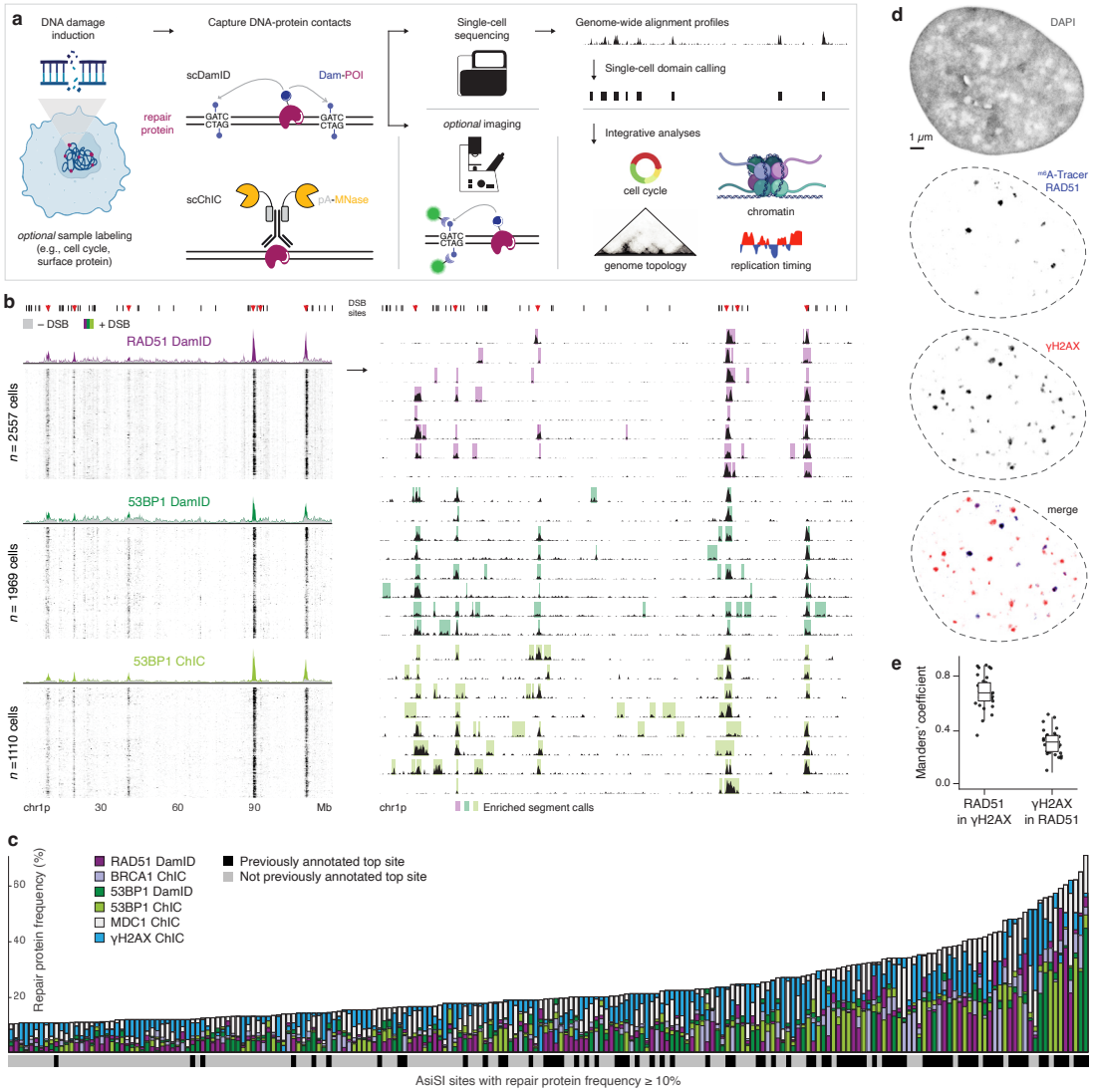


Fig. 5.1. Detecting double-strand break repair proteins with DamID and ChIC.

a, Overview of experimental setup and integrations, partly adapted from ref. [63]. **b,** Signal on chromosome 1p. Left: single-cell heatmaps (RPKM) in +DSB condition, with single-cell aggregates of +DSB (colored) and –DSB (grey) conditions on top. Right: single-cell line plots with overlaid MSR calls. AsiSI motifs, annotated with black lines, and red triangles indicating frequently cleaved top sites. **c,** Bar plot of all AsiSI sites with repair protein frequency ≥ 10%, colored by repair protein frequencies per dataset (target protein & method) per site. Sites are ordered (on the *x*-axis) by increasing absolute repair protein frequency (*i.e.*, highest frequency in any dataset). Per site, bars are ordered by increasing repair protein frequency per dataset (front to back, *i.e.*, bars are not stacked). Bottom horizontal bar indicates previous (lack of) annotation as top site. **d,** Confocal images of one representative nucleus showing DAPI, RAD51 DamID ^{m6}A-Tracer, and endogenous γH2AX immunofluorescent staining. **e,** Quantification of signal colocalization (Manders' A and B per nucleus), $n = 32$ nuclei. The box represents the median, first and third quartiles of the data.

Homology-directed repair mediated by RAD51 correlates with replication timing and transcriptional activity

Repair pathway choice between EJ and HDR is highly regulated at multiple levels, including nuclear structure, global spatial genome organization, local chromatin context, and sequence specificity[22]. In addition to such regulatory processes, the cell cycle state of individual cells has been linked to heterogeneity in repair pathway usage[2,23]. Our DamID experimental setup includes recording of live-cell DNA content during fluorescence-activated cell sorting (FACS), thereby establishing a procedure to address DSB repair protein occupancy in relation to the cell cycle at unprecedented resolution (Fig. 5.2a). We observed that some DSB sites exhibit differences in repair protein frequency according to cell cycle phase, particularly when bound by HDR factors (Sup. Fig. 5.3a). To further explore this relationship, we ordered all Dam-RAD51 cells on their cell cycle stage, and noticed differences in repair enrichment at DSB sites during S phase (Fig. 5.2b-c), a genome-wide trend that is not present for 53BP1 (Sup. Fig. 5.3b). Because HDR requires a sister chromatid as its repair template, we annotated the genomic regions by their replication timing (RT) using publicly available Repli-Seq data[24,25]. This showed considerable concordance between RT of a site and the relative frequency with which it is bound by Dam-RAD51 along S and G2 (Fig. 5.2d), in agreement with imaging data suggesting that active replication influences HDR employment[23].

RT is significantly correlated with chromatin state and genome organization[26,27]; moreover, RT-driven remodeling of the epigenome has been linked to cancer-specific chromosomal rearrangements[28]. Several studies have further highlighted the role of chromatin in DNA repair[29-38]; notably, RAD51 is preferentially recruited to transcribed loci enriched in active histone modification H3K36me3[29,39]. We sought to directly measure the relationship between DNA repair and chromatin, by jointly profiling RAD51 occupancy (with DamID) and either H3K36me3 or repressive mark H3K9me3 (with ChIC, using an antibody) in the same cell (see Methods and ref. [11]). Our combined single-cell Dam&ChIC profiles were of similar quality to DamID as the only genomic readout (Sup. Fig. 5.3c-d), and ChIC histone modification signal was in excellent concordance with published ChIP-seq (Sup. Fig. 5.3e-f). We show that H3K36me3 enrichment is significantly higher in cells that have RAD51 binding than in cells without, for a given AsiSI site (Fig. 5.2e) and across all sites (Sup. Fig. 5.3g). In contrast, H3K9me3 is expectedly low, and similar between bound states (Fig. 5.2e, Sup. Fig. 5.3g). This corroborates the previously established link between HDR and transcriptionally active chromatin at the single-cell level.

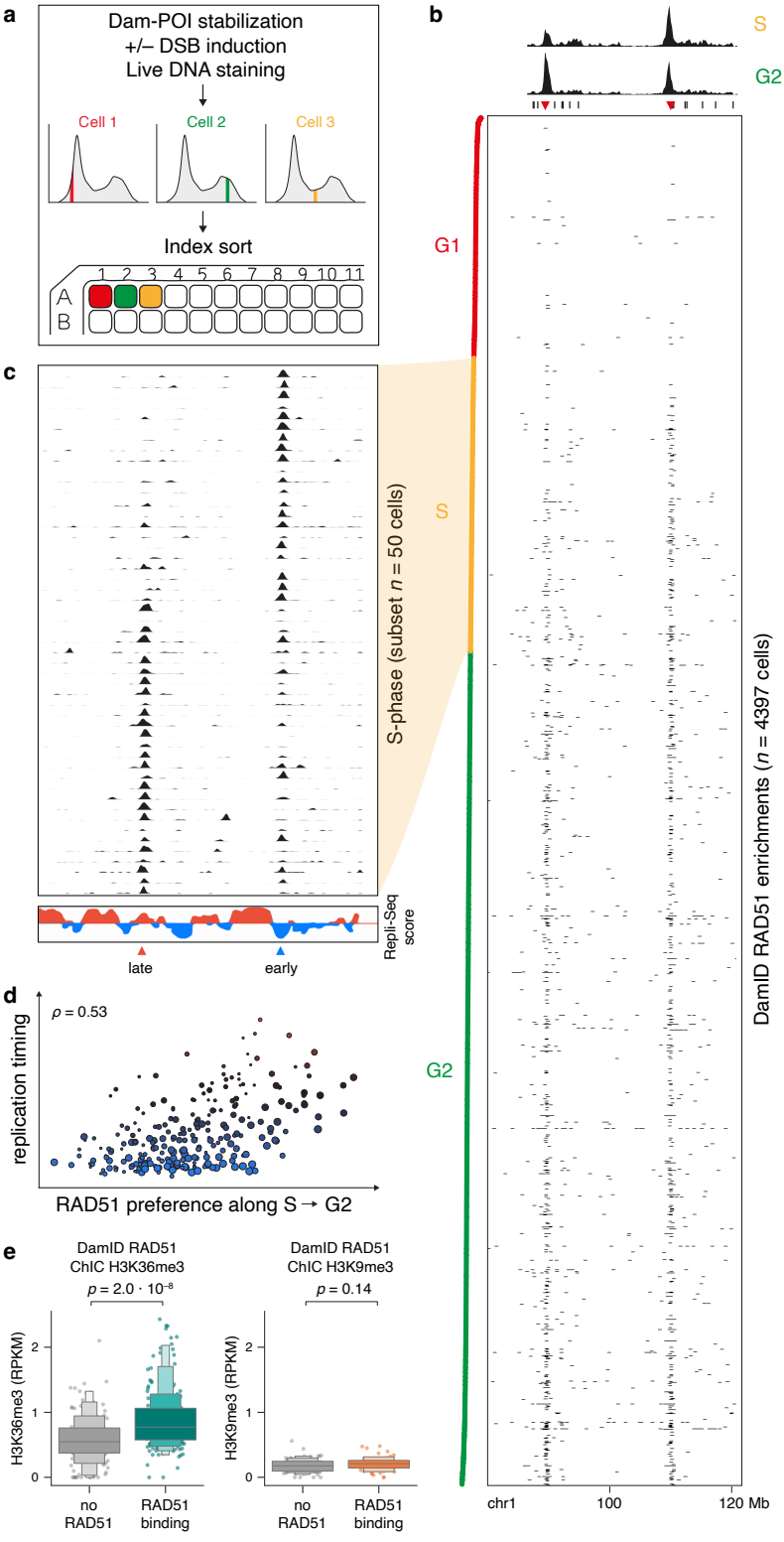

a
Dam-POI stabilization
+/– DSB induction
Live DNA staining
Cell 1
Cell 2
Cell 3
Index sort
1 2 3 4 5 6 7 8 9 10 11
A
B
b
S
G2
G1
S
G2
c
S-phase (subset n = 50 cells)
Repli-Seq score
late
early
DamID RAD51 enrichments (n = 4397 cells)
chr1
100
120 Mb
d
p = 0.53
replication timing
RAD51 preference along S → G2
e
DamID RAD51
ChIC H3K36me3
p = 2.0 · 10⁻⁸
H3K36me3 (RPKM)
no RAD51
RAD51 binding
DamID RAD51
ChIC H3K9me3
p = 0.14
H3K9me3 (RPKM)
no RAD51
RAD51 binding

Fig. 5.2. RAD51 correlates with cell cycle phase, replication timing, and transcriptionally active chromatin.

a, Overview of DamID experimental setup up to and including FACS. **b,** Heatmap showing binarized RAD51 DamID enrichment (MSR calls) of single cells (n = 4397), at a selected region on chromosome 1. DNA content as measured by live Hoechst staining is annotated along the vertical axis. Single-cell aggregates of +DSB condition for S and G2 phases on top, with AsiSI motifs annotated as in **5.1b**. **c,** Line plots showing RAD51 signal (RPKM) for a subset of cells in S phase, region as in **b**. **d,** Quantification (Pearson's r) between cell cycle preference and Repli-Seq scores for AsiSI clusters with repair protein frequency ≥ 5%. Circle size indicates repair protein frequency, circle color corresponds to the y-axis. **e,** H3K36me3 (left) and H3K9me3 (right) ChIC signal at AsiSI site 505, for cells with and without RAD51 enrichment. Boxen plots show data median and multiple quantiles. Number of samples per plot, from left to right: n = 163, 203, 75, 51 cells. Statistic was calculated with the two-sample Kolmogorov-Smirnov test.

Repair protein occupancy in single cells follows pre-existing genome topology

Next, we further explored the binding of repair proteins at DSBs, and the potential spreading of proteins along the genome. In particular, we set out to address chromatin conformation of DSB repair sites at the resolution of single cells. Recent population-based work has reported local chromatin reorganization upon DNA damage[5,7], and proposed cohesin-mediated loop extrusion as the mechanism by which DNA repair domains are established[5]. This further substantiates the thought that chromatin is compacted during the DDR, forming incontiguous domains according to topological features[40-44].

Repair factors are constrained within topologically associating domains in multiple modalities

To examine the interplay between genome topology and repair factor binding in single cells, we projected the DamID 53BP1- and RAD51-enriched segments onto Hi-C maps (Sup. Fig. 5.4a, Fig. 5.3a). Repair domain enrichment occurs in different scenarios, varying mostly according to extent, directionality, and modality of spreading. Modality refers to whether one (*i.e.*, unimodal) or different (*i.e.*, multimodal) enrichment patterns are found, and is thus a measure of intercellular variability. Domains of repair occupancy seem generally constrained by borders of chromosomal domains.

In highly insulated chromatin, repair protein-enriched segments clearly overlap with the local genome structure, while DSB sites devoid of detectable structure show repair protein occupancy in a scattered fashion, without topological constraints. Of note, cell-to-cell heterogeneity is strongly reflected in the repair protein segments,

since a spectrum of domain sizes can be observed (Sup. Fig. 5.4a, Fig. 5.3a). We note that these heterogeneous measurements are a direct result of detecting enrichment without any a priori information of segment size, enabled by our newly implemented multi-scale domain calling approach.

To further illustrate multimodal enrichment, we grouped the repair protein segments according to their three most frequent edge positions (Fig. 5.3a-b). We noticed that segment edges correspond to relative dips in the insulation score, which is a measure of TAD border strength. To investigate if the variability in repair protein spreading is linked to topological variability, we generated split-pool recognition of interactions by tag extension (SPRITE) data. SPRITE is a method that, unlike Hi-C, detects higher-order chromatin structures[45], which can be interpreted as single-molecule topologies. Indeed, we find various SPRITE clusters that overlap the three most frequent repair protein segment edges (Sup. Fig. 5.4b), mimicking the multiple modes of repair spreading.

Next, we sought to generalize this apparent relationship between repair protein spreading in single cells and topological domains. Accordingly, taken across all DSB sites genome-wide, there is a strong minimum in the insulation score at repair protein segment boundaries, indicating overlap with TAD borders (Fig. 5.3c). Repair protein segments also more frequently end at stronger borders (*i.e.*, with lower insulation scores), as quantified by the negative overall correlation (Fig. 5.3d).

Structural protein CTCF demarcates individual repair domains

TAD-like partitioning of repair domains is thought to stabilize chromatin topology and thereby safeguard genome integrity[41]. Further, insulator protein CTCF was found to be in close spatial proximity to radiation-induced γH2AX repair foci (as measured by super-resolution microscopy[40]). CTCF-bound loop anchors are also particularly fragile sites (as measured by mapping DSBs[46]). Yet, a genome-centered view of repair factors and 3D organization is lacking. We used the combined Dam&ChIC method to gather genome-wide information of repair protein RAD51 (with DamID) and structural protein CTCF (with ChIC) within the same cell, akin to two-color imaging. As quality control, alignment of the aggregated CTCF signal on CTCF binding motifs indicates high specificity (Sup. Fig. 5.4b), while retaining sensitivity of median ~10^3 reads per cell (Sup. Fig. 5.4c).

In the combined RAD51&CTCF dataset, we again grouped the repair protein segments (that is, cells) into modes based on the RAD51 edge positions. This was done by calling peaks on the distribution of segment edge frequencies (as seen in Fig. 5.3a) and selecting segments within a 100-kb window surrounding each peak (shaded areas).

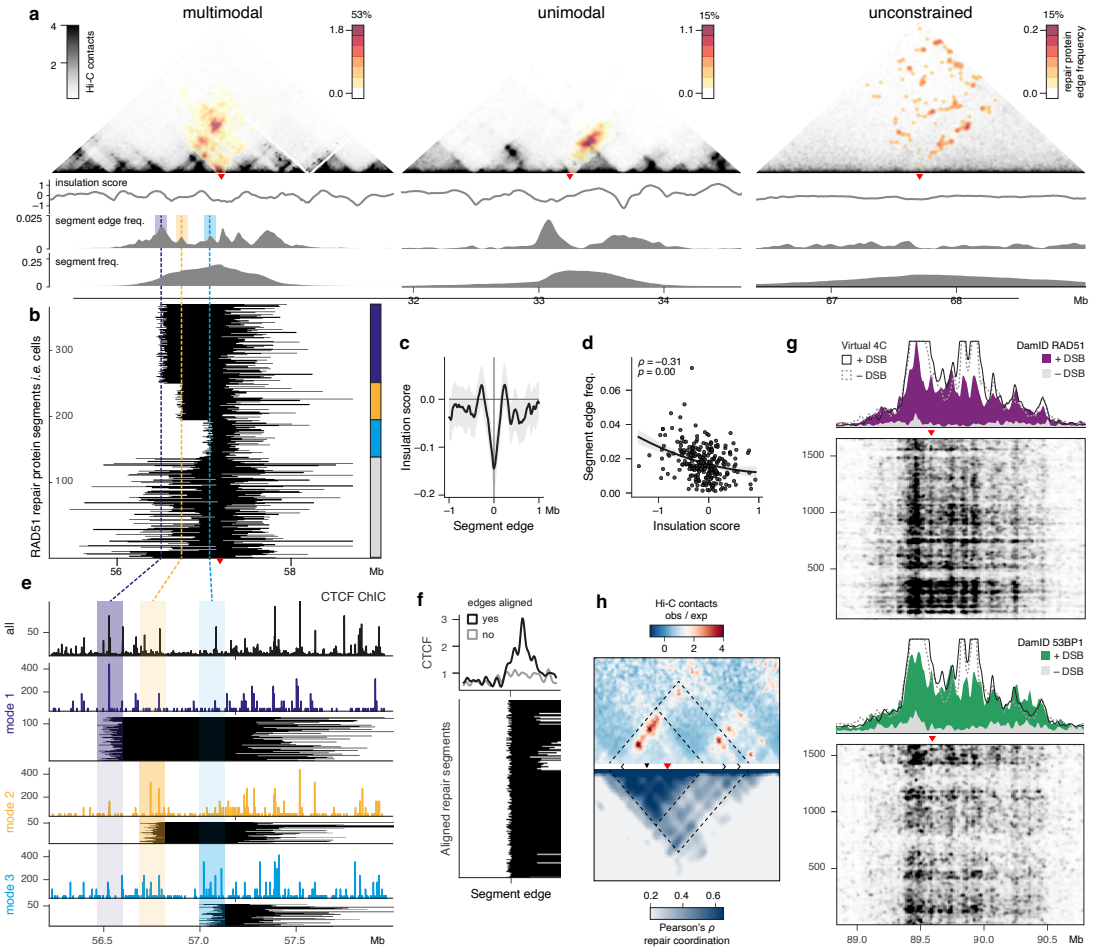


Fig. 5.3. Repair protein spreading in single cells mimics underlying genome topology.
a, Overlay of RAD51 (left) and 53BP1 (middle, right) DamID repair protein segment edges onto the Hi-C contact matrix for regions harboring a DSB site (red triangle). Repair protein binding signal is colored by the frequencies with which those edges are observed across single cells. The total frequency of each locus (% of cells bound by repair protein) is shown above the color bar. One example region is displayed for different scenarios of repair protein spreading: multimodal (left), unimodal (middle), and unconstrained (right). Dashed lines indicate local maxima of edge frequency: outermost (purple), middle (orange), and nearest (blue). The shaded area marks the window (150 kb) used to group segment edges into modes. **b,** Individual repair protein segments per single cell, grouped into modes on the left border, based on windows as in **a**. Colored blocks represent modes, with unassigned segments in grey. **c,** Quantification of insulation score across segment edges within a 2-Mb window, presented as mean ± 95% CI. **d,** Quantification (Pearson's *r*) of the relationship between repair protein-based segment edge frequency and topology-based insulation score at that edge. Line and shaded area represent best-fit linear regression ± 95% CI. *P*-value corresponds to the result of a two-sided null hypothesis test. **e,** Dam&ChIC experiment co-measuring Dam-RAD51 and CTCF in the same cells. *[cont]*

Upper track shows RPKM-normalized CTCF signal of all cells. Below, three groups of spreading modes from **b** are plotted. Corresponding CTCF signals for the same cells are plotted above. Shaded areas indicate spreading modes. **f,** Genome-wide version of **e**, aligning segment edges across DSB sites (colored blocks). The average CTCF signal is shown of aligned cells (black) and for cells without segment edges (grey) on top. **g,** Comparison of virtual 4C (Hi-C contacts along the linear genome) with Dam-RAD51 (top) and Dam-53BP1 (bottom). Aggregates and heatmaps show single-cell data for both +/–DSB conditions. **h,** Comparison of the normalized Hi-C contact matrix (top) with single-cell repair coordination (bottom) for a 5-Mb region. Repair coordination across cells is measured per pair of genomic bins.

For each spreading mode, we then plotted the CTCF signal corresponding to the cells in that mode (Fig. 5.3e, colored lines). By visual inspection, binding of CTCF is consistently highest in the area surrounding its corresponding segment edges. Loci with seemingly high CTCF occupancy (e.g., blue shaded area) may be near-devoid in most cells (modes 1 and 2), while particularly enriched in others (mode 3).

We quantify this observation genome-wide, by aligning all repair protein segment edges found within peak mode windows (*i.e.*, all colored blocks across the whole genome), and comparing them with all other segments not found within those windows, (*i.e.*, all grey blocks). CTCF signal is strongly enriched on the aligned segment edges, while no such pattern can be found for unaligned segments (Fig. 5.3f). With this, our multifactorial data show that proteins such as CTCF demarcate TAD-like structures in various conformations. The observation of heterogeneity within the sizes of repair domains—and their relationship to genome organization—thereby highlights the importance of sequencing-based profiling at single-cell resolution.

Repair protein signal is highly coordinated within loop extrusion borders

In the comparison with Hi-C above, we defined repair protein occupancy as the enrichment from start to end of the MSR segments. To further strengthen the notion that such occupancy occurs in 3D, we compared repair spreading at a given DSB site to the chromosomal conformation of that locus (virtual 4C from Hi-C data). The quantitative repair counts—*i.e.*, sequencing read output, not domain calls—captured by both DamID and ChIC strongly resemble topological contacts along the linear genome (Fig. 5.3g, Sup. Fig. 5.4e). Hence, the quantitative *in silico* population signal, single-cell profiles, and binarized MSR segments all show that DNA repair occurs in the context of genome organization.

Importantly, single-cell Hi-C and super-resolution chromatin tracing methods have indicated that boundaries of TAD-like domains vary across cells, despite preferential anchoring at population-based boundaries[47-50]. We reasoned that, if repair proteins spread according to these topology-driven rules of variability and boundary anchoring, loci with stronger Hi-C contacts should more frequently show coordinated

occupancy of repair proteins across single cells. Conversely, repair signal should be independent for loci with weak (or few) Hi-C contacts. Coordination is thus a measure of how frequently a repair protein occupies a DSB locus, measured across all cells: it should reflect the diversity of topological configurations commonly observable in a given Hi-C window. To best interpret chromatin contacts, we calculated the normalized Hi-C matrix (observed/expected), where distinct architectural stripes and dots can be observed (Fig. 5.3h), which are features of cohesin-mediated loop extrusion. Indeed, we find that repair protein coordination surrounding a DSB corresponds to those features, suggesting that the repair machinery spreads within an ample variety of preferentially anchored but dynamic single-cell loops (Fig. 5.3h).

Altogether, we interpret these results to mean that the spreading of repair proteins on the genome follows underlying topology, explaining differences in repaired genome segment size as well as quantitative repair signal across DSB loci.

Conclusion

Our single-cell sequencing approaches provides a powerful framework to study other DNA damage and repair mechanisms, including stochastic systems and in vivo models. Stratification of single-cell DNA repair profiles according to an additional layer of information is amenable to many different questions of interest. We demonstrated the applicability by using live DNA content staining prior to index sorting to explore the cell cycle. A similar approach can be envisioned for any cellular observation that is compatible with FACS, e.g., antibody staining for cell type annotation, classification of apoptosis or other stress responses, and mitochondrial labeling. We foresee that (multimodal) single-cell measurements will disentangle the role of chromatin, transcription, and other factors involved in DNA repair at high resolution, with implications for our understanding of cellular response and fate after damage.

Acknowledgements

We thank members of the Kind and Legube groups and Eva Brinkman for feedback throughout the project. Specifically, we acknowledge Koos Rooijers for developing computational methods and performing bioinformatic analyses. Purified pA-MNase and ^{m6}A-Tracer were respective gifts from Alexander van Oudenaarden and Bas van Steensel. We thank Reinier van der Linden at the Hubrecht FACS Facility and Anko de Graaff at the Hubrecht Imaging Center for technical support and training. High-throughput sequencing and computational resources were provided by the Utrecht Sequencing Facility (USEQ) and High-Performance Computing Facility. USEQ is subsidized by the University Medical Center Utrecht and the Netherlands X-omics Initiative (NWO project 184.034.019). SPRITE experiments performed in the Jachowicz lab were supported by OEAW core funds. The Oncode Institute is supported by the KWF Dutch Cancer Society. This work was further funded by: Dutch Research Council (NWO-ALW) Vidi grant 161.339 (JK), Fulbright Commission the Netherlands promovendus grant (KLdL), European Molecular Biology Organization grant 9178 (KLdL).

Author contributions

Conceptualization: KLdL and JK. Methodology, Investigation, Data curation & Validation: KLdL. SSdV performed revision experiments with KLdL. AGM and LSP respectively contributed SPRITE experiments and data processing, advised by JWJ. Formal analysis & Visualization: KLdL and PMJR. Software: PMJR. Project administration: JK. Resources: GL, JWJ, and JK. Funding acquisition: KLdL, JWJ, and JK. Writing – original draft: KLdL with contributions from PMJR and JK. Writing – review & editing: all authors.

Methods

Experimental methods

DamID construct design

Shield1-inducible Dam lines were created by cloning into the multiple cloning site (MCS) downstream of the double degron (DD) of ProteoTuner vector pPTuner IRES3 (PT4040-5). This generates an in-frame construct of DD-POI-V5-Dam. POI 53BP1 contains the minimal focus-forming region (amino acids 1221-1711). POI RAD51 contains the full endogenous protein (amino acids 1-339).

Generation of cell lines

Stable, clonal cell lines containing Dam-POI constructs were established by transfection and antibiotic selection. DIvA cells were grown in 24-well plates and transfected with 500 ng DamID plasmid and 1.5 µL Lipofectamine 2000 per well. Each well was passaged to a 15-cm dish and subjected to antibiotic resistance selection with 500 µg/mL G418 (Gibco) for 10 days (at complete death of untransfected control dishes). Monoclonal cell populations were hand-picked, expanded, and characterized by performing bulk DamID. Dam methylation levels were checked by evaluating methylation-specific amplification on agarose gel (as previously described in refs. [14,51]) and on-target methylation signal was evaluated with high-throughput sequencing. One clone per construct with the highest signal-to-noise ratio (SNR) was chosen for single-cell experiments.

Cell culture and experimental treatment conditions

Cell lines were grown in DMEM containing high glucose, GlutaMAX supplement, and sodium pyruvate (Gibco) supplemented with 10% fetal bovine serum (FBS; Sigma-Aldrich) and 1X Penicillin Streptomycin (Gibco), at 37 °C with 5% CO_2. For maintenance, cells were split 1/10 every 3 days and routinely tested for *Mycoplasma*. For experimental procedures, cells were plated the day prior, to achieve 60-70% confluency during induction treatment. Dam-POIs were stabilized by addition of 500 nM Shield1 ligand (AOBIOUS, dissolved in DMSO to 500 mM) for 4 h (Dam-53BP1) or 8 h (Dam-RAD51). Nuclear translocation of AsiSI-ER was simultaneously induced by addition of 300 nM 4-hydroxytamoxifen (4OHT; Sigma, dissolved in ethanol to 13 mM). Cells were harvested by washing twice with PBS0 followed by trypsinization with 1X TrypLE (Gibco), inactivation with DMEM, pipetting to yield a single-cell suspension, and centrifugation at 300 *g*. For FACS, cell pellets were resuspended in growth medium containing 10 µg/mL Hoechst 34580 per $1x10^6$ cells and incubated for 45-60 minutes at 37 °C, for live DNA content staining. Prior to sorting, cell suspension was passed through a 20-µm mesh. For bulk DamID, genomic DNA was isolated from cell pellets using commercial reagents (e.g., Promega Wizard). During each

genomics experiment, cells were concurrently plated on glass and treated to verify proper induction and cleavage activity of AsiSI-ER by immunofluorescent staining and imaging.

Immunofluorescent staining

Cells were grown as described above, with the exception of plating on glass coverslips the day prior to experimental treatment. At the end of induction, cells were washed twice with PBS and chemically crosslinked with fresh formaldehyde solution (2% in PBS) for 10 min at RT, then permeabilized (with 0.5% IGEPAL® CA-630 in PBS) for 20 min and blocked (with 1% bovine serum albumin (BSA) in PBS) for 30 min. All antibody incubations were performed in final 1% BSA in PBS followed by three PBS washes at RT. Incubation with primary antibody against the endogenous protein as well as purified ^{m6}A-Tracer protein[52] (recognizing methylated DNA) was performed at 4 °C for 16 h (overnight), followed by anti-GFP (against ^{m6}A-Tracer protein) incubation at RT for 1 h, and secondary antibody incubations at RT for 1 h. The final PBS wash was simultaneously an incubation with DAPI at 0.5 µg/mL for 2 min, followed by a wash in MilliQ and sample mounting on glass slides using VECTASHIELD Antifade mounting medium (Vector Laboratories).

Primary antibodies: anti-53BP1 Santa Cruz [H-300] sc-22760 (rabbit) at 1/500, anti-γH2AX BioLegend [2F3] 61340x (mouse) at 1/1000, anti-MDC1 Bethyl Laboratories A300-053A (rabbit) at 1/500, anti-GFP Aves GFP-1020 (chicken) at 1/1000.
Secondary antibodies: all Alexa Fluor (Invitrogen) at 1/500. Anti-chicken 488 (goat) cat#A-11039, anti-mouse 555 (goat) cat#A32727, anti-mouse 647 (donkey) cat#A-21245, anti-rabbit 555 (goat) cat#A-21428, anti-rabbit 647 (goat) cat#A-21245.

Confocal imaging

Imaging (12-bit) was performed on an inverted scanning confocal microscope (Leica TCS SP8) with a HC PL APO CS3 63X (NA 1.40) oil-immersion objective and HyD detectors. Pinhole was set to 1 Airy Unit. Scanning zoom was set to 4-5X at a speed of 400 Hz. Full-nucleus images were acquired as *Z*-stacks at 0.2-µm intervals. Multi-color images were acquired sequentially (by frame).

Image analysis

Images were processed in Imaris 9.3 (Bitplane) by baseline subtraction and background correction with a 3x3(x3) median filter. Colocalization was calculated by Manders' coefficients M1 and M2 between channel 2 (DamID ^{m6}A-Tracer) and channel 3 (endogenous DSB repair marker). Pixels were retained that contained signal (intensity >10) in channel 1 (DAPI). Pixel intensity thresholds for the colocalization analysis were determined using Costes' method with default settings.

FACS

FACS was performed on BD FACSJazz or BD FACSInflux Cell Sorter instruments with BD Sortware. Index information was recorded for all sorts. Single cells were gated on forward and side scatters, trigger pulse width, and Hoechst cell cycle profiles. One cell per well was sorted into 384-well hard-shell plates containing 5 µL of filtered mineral oil and protocol-specific reagent.

DamID and derivative methods

High-throughput sequencing

Libraries were sequenced on the Illumina NextSeq 500 (75-bp single-end reads) or NextSeq 2000 (100-bp paired-end reads) platform.

Bulk DamID

DamID on populations was performed as described in ref. [51], briefly as follows. Genomic DNA (gDNA) was isolated from cell pellets, digested with DpnI to enrich for Dam-methylated GATCs, and ligated to universal (not barcoded) double-stranded DamID adapter molecules. Methylation-specific PCR was performed with barcoded primer (unique per sample). Samples were pooled per clone, further processed to construct Illumina-compatible libraries, and sequenced to approximately 10 M raw reads per sample.

Automated liquid handling

Liquid reagent dispensing steps for single-cell protocols in microwell plates were performed on a Nanodrop II robot (Innovadyne Technologies / BioNex). Addition of barcoded adapters was done with a mosquito LV (SPT Labtech).

Single-cell DamID2

DamID on single cells was performed as previously described in detail[53], briefly as follows. After FACS, cells were lysed and treated with proteinase K, after which methylated GATCs resulting from Dam enzyme activity were specifically digested with DpnI. Double-stranded adapters containing cell-specific barcodes and a T7 promoter were ligated to the blunt (DpnI-cleaved) DNA ends. Cells with non-overlapping barcodes were pooled together to undergo in vitro transcription (IVT), amplifying the genomic DamID-specific product in a linear manner. Library preparation was then performed on the amplified RNA, to generate molecules compatible with Illumina sequencing.

Single-cell Dam&T-seq

One single-cell DamID experiment was performed using the combinatorial scDam&T-seq approach capturing DamID and transcriptome[15], as previously

described in detail[53], with the exception that all volumes were halved to reduce costs. Briefly, after FACS, cells were lysed, followed by reverse transcription and second-strand synthesis in order to convert cellular mRNA into cDNA. Subsequent steps were followed according to the scDamID2 protocol.

(Single-cell) ChIC

ChIC was performed as described in detail in ref. [10], with adaptations as follows. After experimental treatment of cell cultures as described above, nuclei were isolated and permeabilized, incubated with primary antibody, then incubated with pA-MNase (Protein A IgG-binding domain fused to micrococcal nuclease, for antibody-specific binding) and Hoechst (for DNA content staining). If the primary antibody was raised in mouse, nuclei were incubated with secondary antibody (rabbit anti-mouse) before incubation with pA-MNase. After FACS, proximity-based cleaving by pA-MNase was activated (for exactly 30 min on ice), followed by inactivation and proteinase K treatment. MNase-cleaved ends were then blunted and phosphorylated, and double-stranded adapters were ligated. For one experiment, A-tailing was performed after end repair of MNase-cleaved ends, followed by ligation of T-tailed adapters. DNA molecules were then further processed for sequencing as in scDamID2.

Antibodies: anti-53BP1 Santa Cruz [H-300] sc-22760 (rabbit) at 1/500, anti-γH2AX BioLegend [2F3] 61340x (mouse) at 1/500, anti-MDC1 Bethyl Laboratories A300-053A (rabbit) at 1/500, anti-BRCA1 Santa Cruz [D-9] sc-6954 (mouse) at 1/500, anti-H3K9me3 Abcam ab8898 (rabbit) at 1/1000, anti-H3K36me3 Active Motif 6110x (rabbit) at 1/500, anti-mouse IgG Abcam ab6709 (rabbit) at 1/500, anti-CTCF Merck 07-729 (rabbit) at 1/200.

Single-cell Dam&ChIC

Combinatorial profiling of DamID and ChIC was performed as described in ref. [11], with adaptations as follows. The same procedures of nuclei isolation, antibody treatment, FACS, and molecular preparation were followed as for scChIC. After end repair of MNase-cleaved ends, the scDamID procedure was followed, namely DpnI digestion to enrich for Dam-methylated GATCs, adapter ligation, and subsequent library preparation steps.

SPRITE

SPRITE was performed as described in ref. [54] with minor modifications. U2OS cells were crosslinked with DSG/1%PFA, followed by permeabilization and nuclear extraction according to the protocol. Next, nuclei were digested with a mix of restriction enzymes HpyCH4V and AluI for 16 h, washed in PBS buffer and sonicated using a Covaris E220. From this step onwards, the standard SPRITE protocol[54] was

followed. Sequencing was performed using the Illumina NovaSeq X platform (10B chemistry), with 500 million reads obtained.

Computational methods

Raw data processing

Data generated by the scDamID and scDam&T-seq protocols was largely processed with the workflow and scripts described in [53] (see also www.github.com/KindLab/scDamAndTools). The procedure is described below, in brief. For detailed parameters and exact software versions, see www.github.com/KindLab/scRepair or https://doi.org/10.5281/zenodo.13927895 (ref. [55]).

DamID and Dam&T samples

DamID data contains reads that result from DpnI-restriction activity. Briefly, raw sequencing data was processed by removing reads that contained contaminants resulting from library preparation, using cutadapt (v2.0). Samples were then demultiplexed (using an in-house Python script), and for every single sample one file was obtained. Demultiplexed data were aligned using HISAT2 (v2.1.0) to human reference genome hg19/GRCh37, supplemented with the ERCC spike-in sequences. Since DpnI cleaves GAˇTC, we prefixed each read in silico with a 'GA' dimer to improve alignment rates. Aligned reads were counted per genomic GATC position and a vector of counts per chromosome was stored. For DamID, reads not aligning at GATC positions were discarded. Reads were then binned in either 1-kb or 100-kb bins for further analysis and plotting.

ChIC and Dam&ChIC samples

Contaminant removal, demultiplexing and alignment was done as for DamID samples. Reads were counted per GATC position (for both ChIC-only and Dam&ChIC samples), but reads not aligning on GATC position, as well as unaligned reads were stored in an auxiliary BAM file. These non-GATC reads were processed by removal of the (in silico prefixed) 'GA' dimer, and realignment to the human reference genome. Realigned non-GATC reads were then counted UMI-unique per genomic position and counts were binned in either 1-kb or 100-kb bins for further analysis and plotting.

SPRITE samples

Raw sequencing data were analyzed based on ref. [45] using the available pipeline (https://github.com/GuttmanLab/sprite-pipeline) and reference genome hg38. The pipeline was modified to enable local alignment using the "--local" flag of Bowtie 2. The barcode sequence for creating the SPRITE clusters was `[Y|SPACER|ODD|SPACER|EVEN|SPACER|ODD]`.

RPKMs and scaling factors

RPKMs were only used for inspection of raw read densities and were calculated using the common definition. The nature of DSB binding patterns make the signal generally inherently peaky (high variance in read density along linear chromosomes). However, the number of DSBs and the ratio of signal (signal from DSB loci) and background (signal from inter-DSB loci) varies substantially between samples (i.e., SNR, signal-to-noise ratios). Naively scaling the read density as with RPKMs led to false enrichments/depletions, e.g., comparing a high SNR sample to an uninduced (no DSBs) background would register depletions across most inter-DSB regions. To overcome the inherent variation in SNR between the single-cell samples we employed the scaling factor normalization from PoissonSeq[56], section 3.2, which weakens the assumptions implicit of RPKM scaling by assuming only that a fraction (here: 50%) of bins are not differentially enriched between two conditions.

Enriched segment calling

To identify regions of significant enrichment and depletion we adapted the workflow of multi-scale representation of genomic signals as described in ref. [19]. A diagram of the workflow can be found in Sup. Fig. 5.2. Each induced sample is compared to a suitable background. For DamID, this is an average of single-cell samples that are not induced with 4OHT (one background per clone). For ChIC, this is an average of single-cell samples that are not induced with 4OHT (one background per antibody target). Starting from 1-kb bins, both foreground and background signals are convolved with a Gaussian kernel whose standard deviation increases with $\sqrt{2}$ at each level, up to a standard deviation of 10 Mb. This creates the scale-space.

Each level of the scale-space is segmented along the linear chromosome by comparing where background signal is either above or below the foreground signal (in practice a small confidence interval is built around the foreground signal to deal with zero data in the foreground signal and numerical rounding issues after Gaussian convolution, raising the possibility that background signal is within the confidence interval of the foreground signal; those segments are ignored in further steps). Each segment, where background signal is either below or above the foreground signal, at each level in the scale space is then tested for significant enrichment or depletion, respectively. We use the Gamma distribution to create a confidence interval around the background, using the observed signal density and the scaling factor (see above) of the foreground versus the background signal, and a *p*-value of 10^{-9}.

A segment is considered significantly enriched or depleted when the foreground signal is either above or below this confidence interval. Similar to the original MSR, we calculate the significant fold change (SFC) as the observed foreground signal

over the confidence interval boundary. In addition, we also record the true observed fold change of observed versus expected densities. Finally, we prune the enriched/depleted segments across all levels of the scale-space by selecting the level that yields the highest (absolute) SFC score.

Post-hoc filtering

We focus on high-fidelity enrichments in downstream analyses, satisfying the following conditions: observed $\log_2$FC ≥ 1.25, size ≤ 10 Mb, segments encompass at least 10 UMI-unique reads in the foreground sample.

Normalization of Hoechst measurements across batches

A Gaussian mixture model with 2 or 3 components, depending on fit, was applied to Hoechst intensity values, after which G1 and G2 peaks were assigned.

Defining top AsiSI sites

Top AsiSI site annotations were taken from ref. [33].

External data processing

ChIP-seq

Raw ChIP-seq data was obtained from GSE48423, GSE97589 and E-MTAB-5817. Reads were aligned using HISAT2 (same parameters as our DamID and ChIC samples). Aligned reads with MAPQ ≥ 10 were counted in genomic bins of either 1 kb or 100 kb and used for downstream analyses.

Repli-seq

Repli-seq data was obtained from the 4DN project, SRP126407 and SRP197558. Samples used: 4DNFI6KIPWXQ; 4DNFID41JKT6; 4DNFIFEZTAI1; 4DNFIZDPE9T6; SRR6363337; SRR6363338; SRR6363341; SRR6363342; SRR6363345; SRR6363348; SRR6363350; SRR9040713; SRR9040714.

Sequencing data was aligned using HISAT2 using the same parameters as DamID and ChIC samples. Reads were counted in bins of 1 kb. Using sample annotations of early and late replication (and for one sample set, mid-replication) we used glmPCA[57] to obtain a principal component per chromosome indicative of replication timing in U2OS cells.

Hi-C

Processed Hi-C files were obtained from E-MTAB-8851. Samples used: HiC_mOHT_rep1; HiC_mOHT_rep2; HiC_pOHT_rep1; HiC_pOHT_rep2.

The observed over expected Hi-C matrix presented in Fig. 5.3f was calculated by dividing the normalized Hi-C matrix (binned at 25-kb resolution) by the average intrachromosomal contact probability between increasingly distant pairs of loci (i.e., distance-dependent contact decay) using cooltools.lib.numutils.observed_over_expected.

For Sup. Fig. 5.4, the normalized Hi-C matrices (binned at 25-kb resolution) were smoothed using a Gaussian kernel (σ = 200 kb). The fold change was calculated (by dividing the smoothed induced matrix over the uninduced) and subsequently $\log_2$-transformed, after which insignificant $\log_2$FCs were masked from the interaction maps.

hPTMs in cells with and without RAD51 repaired sites

The analysis presented in Fig. 5.2e and Sup. Fig. 5.3f, that demonstrates enriched H3K36me3 on AsiSI sites in cells where sites are repaired by RAD51, was performed as follows. Per site, cells were stratified by the presence or absence of an MSR segment call. Only sites that overlap with an MSR segment in at least 5 cells were included. In the resulting two groups of cells (*i.e.*, with and without repair protein bound), the mean enrichment of either H3K36me3 or H3K9me3 ChIC signal was calculated in the consecutive non-overlapping 100-kb bin wherein the geometric mean of the AsiSI cluster was located.

MSR segment edge peak calling

For systematic comparison between MSR segment edge frequency and Hi-C insulation score (computed using cooltools.insulation on 25-kb binned normalized Hi-C matrix) presented in Fig. 5.3c-d, segment edge peaks were identified as follows. The segment edge frequency was calculated by averaging the binary MSR calls over the single cells in 25-kb bins. The edge frequency vector was smoothed by fitting a Gaussian kernel (25-kb standard deviation). Segment edge frequency peaks were subsequently detected by scipy.signal.find_peaks (with prominence = percentile 97.5). Hi-C insulation score and segment edge frequency at called peak positions were used.

SPRITE analysis

We extracted SPRITE clusters from the aligned data using a modified version of get_sprite_contacts.py (https://github.com/GuttmanLab/sprite-pipeline). Only SPRITE clusters containing between 2-10000 fragments were included for visualization. In Sup. Fig. 5.4b all SPRITE clusters overlapping the locus are plotted and clustered according to overlap with repair protein segment edge peaks highlighted in Fig. 5.3a.

Identification of pairs of contacting repair sites

For the identification of contacting pairs of repair sites, a total of 79 AsiSI clusters that contain a previously annotated top site was included. The theoretical number of pairs, defined as $P = \frac{n!}{k!(n-k)!}$, where n is the number of AsiSI clusters (79) and $k = 2$, was tested for the presence of DSB-induced (+4OHT) 3D contacts in the logFC Hi-C matrix described above. The mean logFC was calculated within a window of 50-kb surrounding the pixel where both cluster geometric means of the pair meet in the Hi-C matrix. Based on the distribution of mean logFC values across the 3081 pairs, a threshold was set at ≥ 0.2 to discriminate contacting from not-contacting pairs (Sup. Fig. 5.4a). Pairs of sites less than 2 Mb apart in linear genomic distance were excluded from the analysis.

Analysis of repair coordination

Analysis of repair coordination across single cells is largely categorized in two: 1) based on the binary MSR segment calls, that were used for both pairwise and multi-way similarity metrics (Fig. 5.4e-g, Sup. Fig. 5.4d-e) and 2) based on quantitative depth-normalized repair signal, the Pearson's coefficient was calculated to evaluate pairwise coordination. The details of both approaches are described below.

Coordination of binary MSR repair-enriched segments

To measure coordination of binary repair signal, MSR segments overlapping the AsiSI cluster geometric mean were used to determine whether a cluster is repaired in a given cell. A total of 79 AsiSI clusters that contain a previously annotated top site was included. The resulting $m \times n$ matrix is defined as $M = (a_{ij})$, where i is a cell and j is an AsiSI cluster. Pairwise distances were calculated between all pairs of clusters using the Jaccard similarity index, defined as the size of the intersection over the size of the union: $J(A,B) = \frac{|A \cap B|}{|A \cup B|}$ where A and B represent any pair of clusters in M.

For multi-way coordination analysis of repair hubs (of ≥ 3 contacting AsiSI clusters), we applied a multiple-site extension of the Sørensen similarity index on binary matrix M[58]. For a detailed justification and explanation of the Sørensen-Dice multi-way similarity metric, see refs. [58-60]. We briefly describe the way the metric was used here, below. The multiple-site similarity index for any number of T AsiSI clusters can be formulated as Equation (1)

$$C_S^T = \frac{T}{T-1}\left(\frac{\sum_{i<j} a_{ij} - \sum_{i<j<k} a_{ijk} + \sum_{i<j<k<l} a_{ijkl} - \ldots}{\sum_i a_i}\right)$$

where a_i is the number of cells in which cluster A_i is repaired, a_{ij} the number of cells that share repair of cluster A_i and A_j, and a_{ijk} the number of cells that share repair of cluster A_i, A_j, and A_k and so on.

In case of $T = 2$ the outcome would reflect the definition of the original pairwise Sørensen-Dice similarity index.

We wrote a Python implementation of the `betapart` R package[60] to compute the multiple-site Sørensen similarity index on a subset of matrix M defined as: $N = (M_{ij})_{\substack{1 \le i < a \\ j \in c}}$ where c is a vector of AsiSI cluster indices of any size in M.

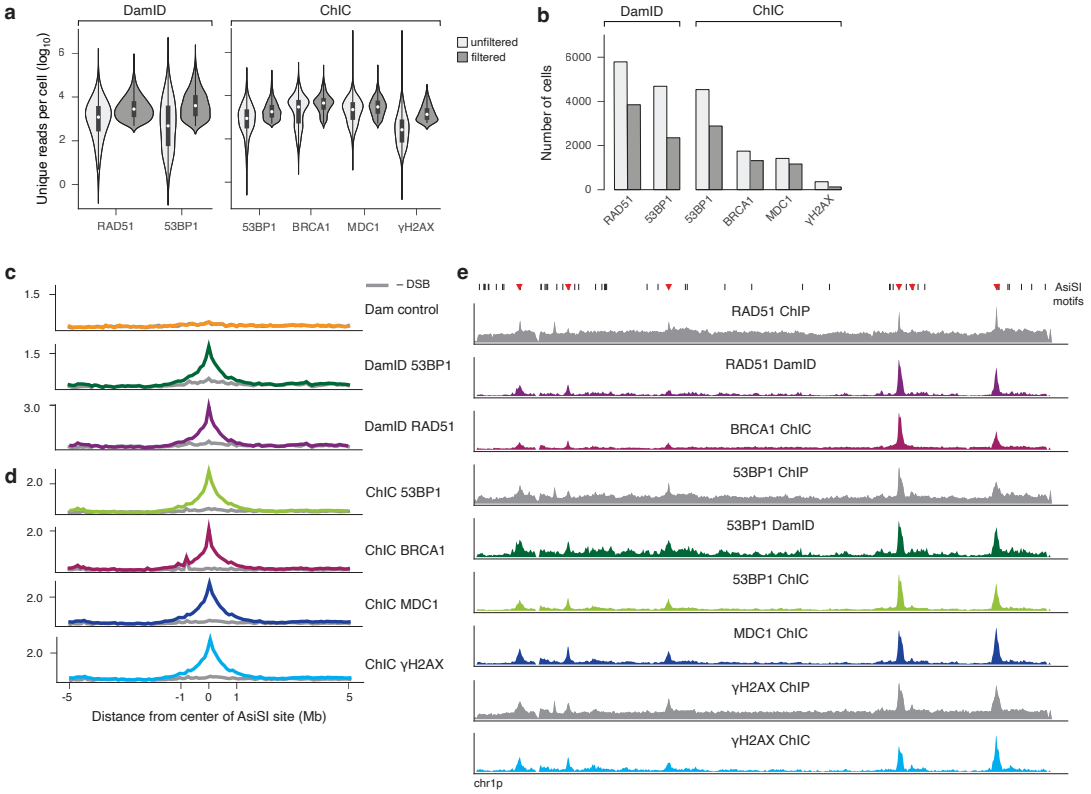
Our `pyBetapart` function is available on GitHub.

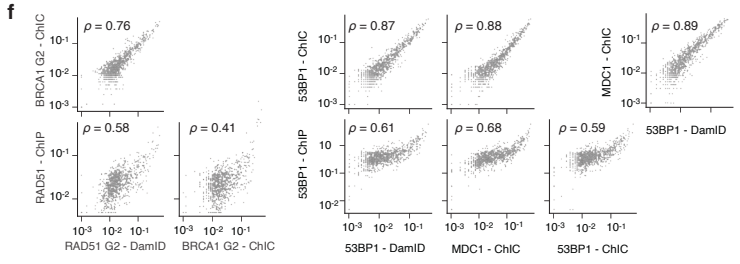
Coordination of quantitative repair signal

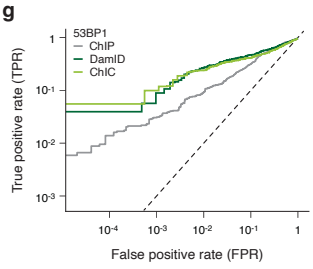
Coordinated behavior of quantitative repair signal for the analysis presented in Fig. 5.3 and Sup. Fig. 5.4 was computed as follows. For intra-chromosomal coordination maps, the single-cell data was first binned at 20-kb resolution, RPKM-normalized and smoothed using a Gaussian kernel (40-kb standard deviation). For inter-chromosomal coordination maps, data was binned at 40 kb, RPKM-normalized and smoothed. For all possible pairs of bins surrounding the AsiSI cluster geometric mean i and j, repair coordination was calculated as the Pearson's correlation coefficient between the vectors b_i and b_j that contain the RPKM values of bin i and j across the single cells.

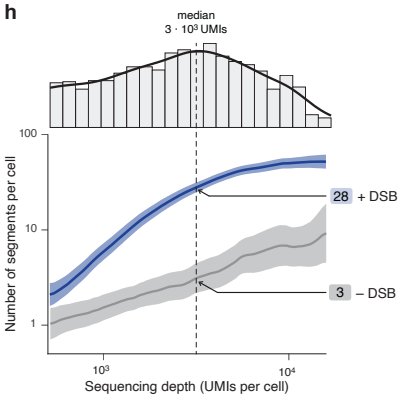
Data and code availability

Raw sequencing data, metadata, and count tables generated in this study are publicly available on the NCBI Gene Expression Omnibus under Series accession code GSE229874 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE229874].
Imaging data generated in this study are publicly available on Mendeley Data [https://data.mendeley.com/datasets/84x5yfhdw4/1][62].
Previously published sequencing data are available under GEO accession codes GSE48423 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE48423], GSE97589 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE97589], GSE131018 [https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE131018], and SRA BioProject numbers PRJEB21297 [https://www.ncbi.nlm.nih.gov/sra/PRJEB21297], PRJEB42378 [https://www.ncbi.nlm.nih.gov/bioproject/PRJEB42378], PRJNA397123 [https://www.ncbi.nlm.nih.gov/bioproject/PRJNA397123].
Key scripts are available on GitHub [www.github.com/KindLab/scRepair].
DOI-versioned code is available on Zenodo [https://doi.org/10.5281/zenodo.13927895][55].


a
DamID
ChIC
unfiltered
filtered
Unique reads per cell (log10)
RAD51
53BP1
53BP1
BRCA1
MDC1
γH2AX
b
DamID
ChIC
Number of cells
RAD51
53BP1
53BP1
BRCA1
MDC1
γH2AX
c
– DSB
Dam control
DamID 53BP1
DamID RAD51
d
ChIC 53BP1
ChIC BRCA1
ChIC MDC1
ChIC γH2AX
Distance from center of AsiSI site (Mb)
e
AsiSI motifs
RAD51 ChIP
RAD51 DamID
BRCA1 ChIC
53BP1 ChIP
53BP1 DamID
53BP1 ChIC
MDC1 ChIC
γH2AX ChIP
γH2AX ChIC
chr1p

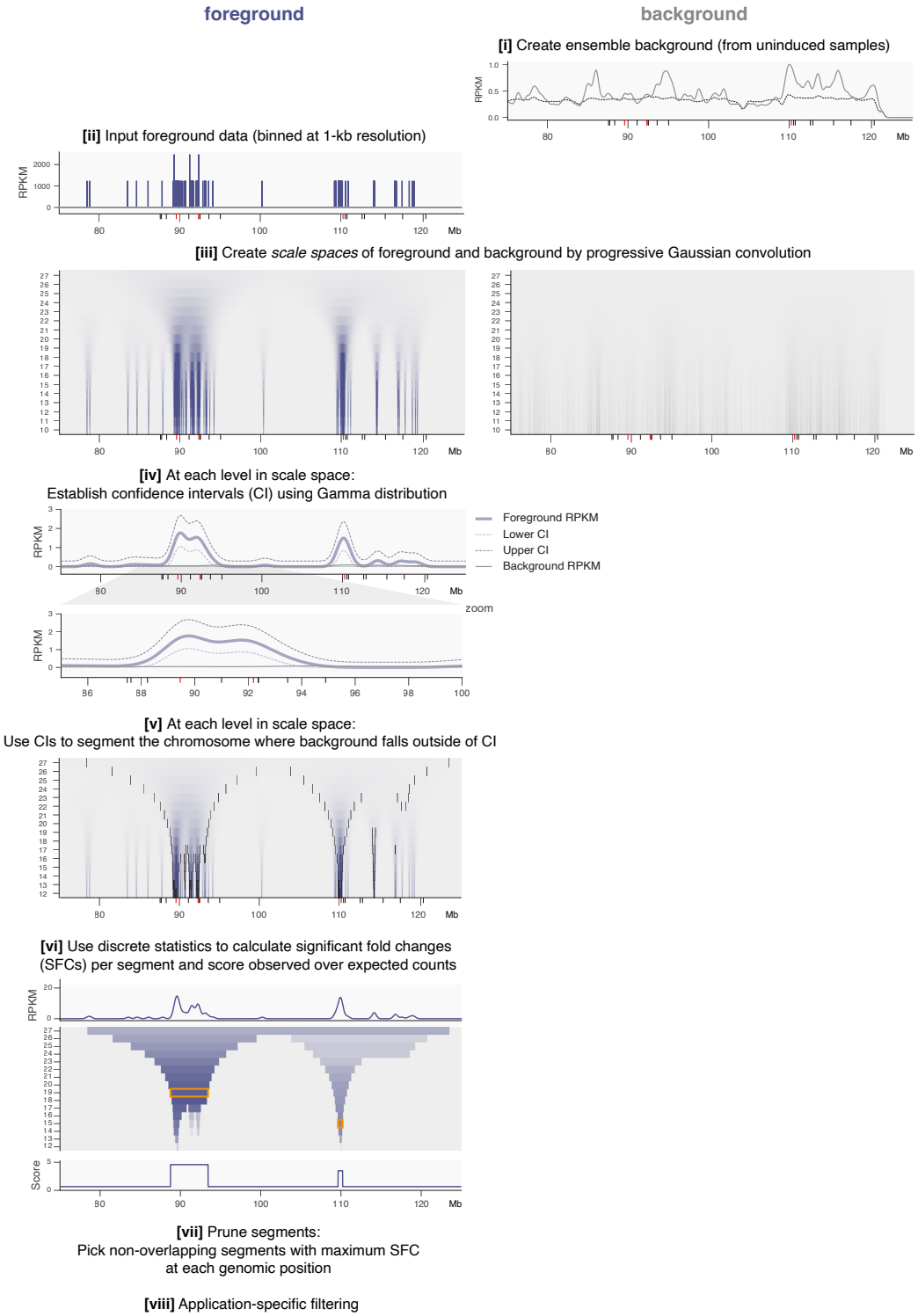

f
ρ = 0.76
ρ = 0.58
ρ = 0.41
BRCA1 G2 - ChIC
RAD51 - ChIP
RAD51 G2 - DamID
BRCA1 G2 - ChIC
ρ = 0.87
ρ = 0.88
ρ = 0.89
ρ = 0.61
ρ = 0.68
ρ = 0.59
53BP1 - ChIC
53BP1 - ChIP
MDC1 - ChIC
53BP1 - DamID
53BP1 - DamID
MDC1 - ChIC
53BP1 - ChIC


g
53BP1
ChIP
DamID
ChIC
True positive rate (TPR)
False positive rate (FPR)

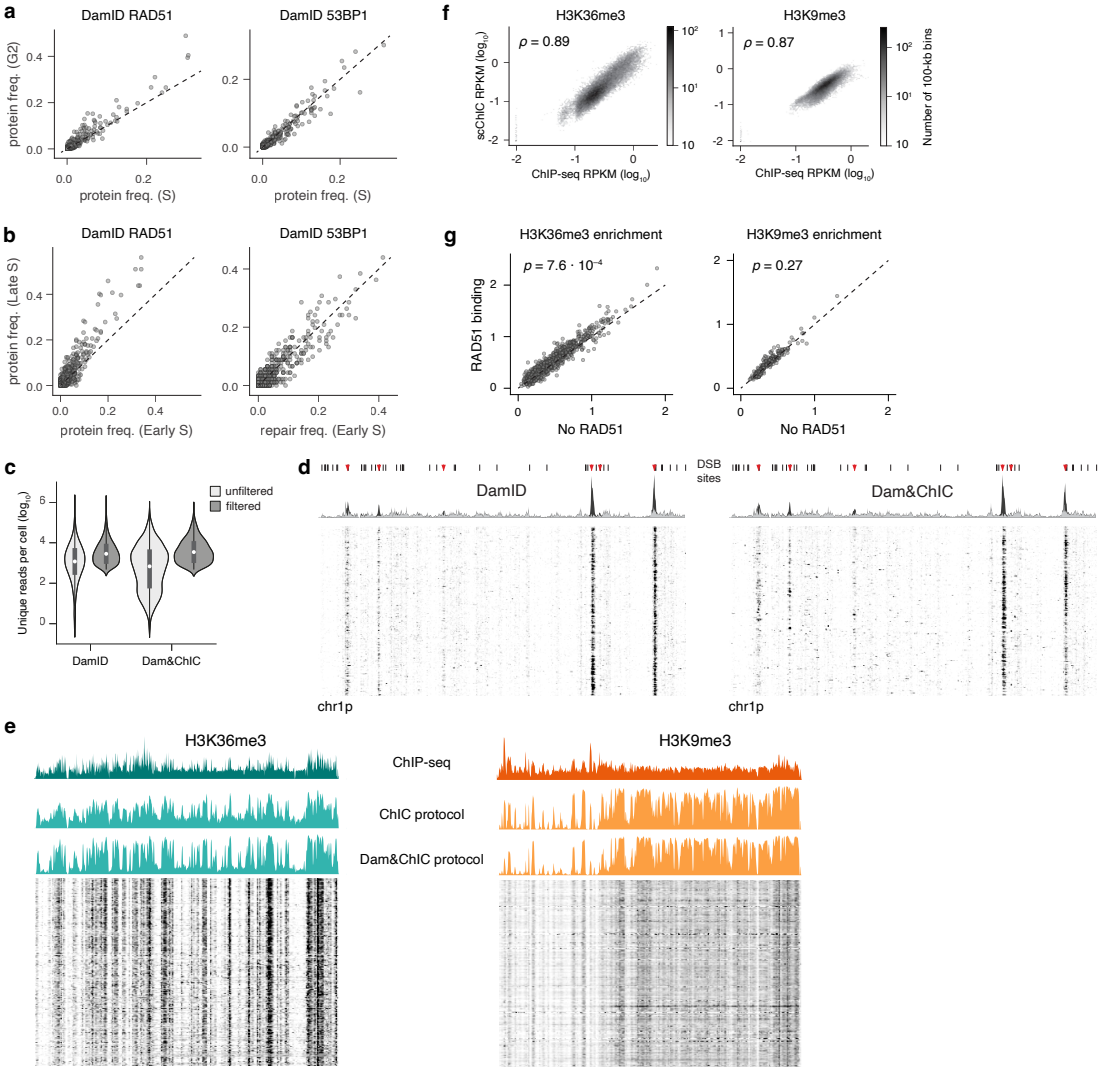

h
median
3 · 10³ UMIs
Number of segments per cell
28 + DSB
3 – DSB
Sequencing depth (UMIs per cell)

Sup. Fig. 5.1. Technical validation of single-cell DamID and ChIC repair protein data.

a, Distribution of sequencing depth per protein target per method, unfiltered and filtered (retaining samples passing threshold). **b,** Number of cells per protein target per method, unfiltered and filtered. **c,** DamID alignment plots (averaged across single cells) over top AsiSI sites, in +DSB (colored) and –DSB (grey) conditions. **d,** ChIC alignment plots (averaged across 100-cell samples) as in **c**. **e,** Line plots of chromosome 1p as in **5.1b**, including external ChIP-seq data. **f,** Scatter plots of ChIP RPKM versus DamID and ChIC repair protein frequency, per protein and relevant cell cycle phase. Each dot is an AsiSI site. Correlation values indicate Spearman's *rho*. **g,** Receiver-Operator Characteristic (ROC) curves showing TPR versus FPR for 53BP1 ChIP, ChIC and DamID, testing presence of a called repair protein segment and whether it falls on an AsiSI motif. Dashed line: $x = y$. **h,** Line graphs showing the number of enriched segments as a function of sequencing depth in both DSB induction conditions. Bold line indicates mean, shaded area indicates 95% CI. Histogram of sequencing depth is annotated on top, dashed vertical line indicating the median.

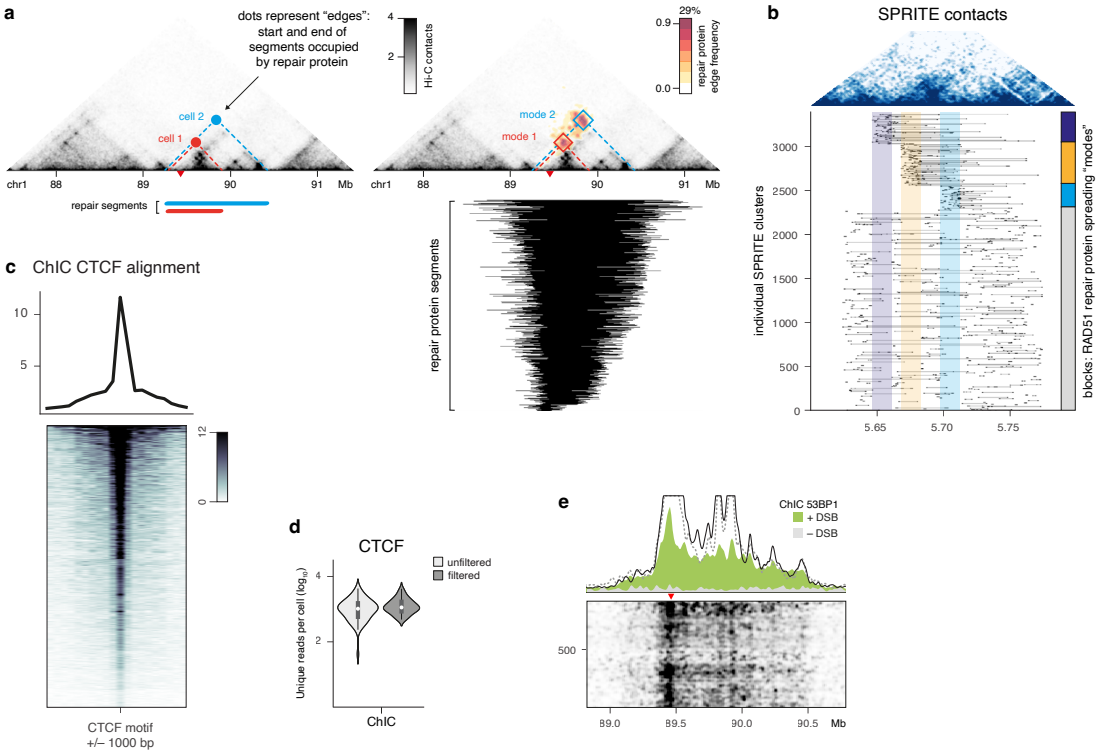


Sup. Fig. 5.2. Computational model for unbiased detection of enriched genomic segments.



Sup. Fig. 5.3. Validation and characterization of pathway-specific repair protein frequencies.

a, Scatter plots of RAD51 and 53BP1 DamID repair protein frequency in G2 vs S phase. Each dot is an AsiSI cluster. Dashed line: x = y. **b,** Scatter plots as in **a**, showing Late vs Early S phase. **c,** Sequencing depth (UMIs per cell) for samples prepared with ChIC or Dam&ChIC protocols. **d,** DamID RAD51 signal on chromosome 1p, generated with DamID (left) or Dam&ChIC (right) protocols. **e,** Signal on chromosome 1, for H3K36me3 (top) and H3K9me3 (bottom). Line plots show ChIP-seq (RPKM) and single-cell aggregate ChIC (contact frequency). Heatmaps show scChIC (RPKM) of 1000 cells. Aggregate ChIC signal is split on sample preparation by ChIC or Dam&ChIC. **f,** Hexbin density plots showing the bin-based correlations between ChIP-seq and scChIC sequencing depth. Hexbins are colored by the number of 100-kb genomic bins. **g,** Genome-wide quantification of H3K36me3 (left) or H3K9me3 (right) signal based on presence or absence of RAD51 enrichment. For each AsiSI site, scatter plots show the mean hPTM ChIC signal in cells that are repaired by RAD51 (y) or not (x). Statistical significance was calculated by the one-sample Kolmogorov-Smirnov test.



Sup. Fig. 5.4. Intercellular variability visualized by projection of repair protein segment edges onto Hi-C contact profiles.

a, Overlay of repair protein segment edges onto the Hi-C contact matrix for regions harboring a DSB site. Left: cartoon indicates repair protein segments of different sizes, originating from two individual cells. Right: Each repair data point projected onto the Hi-C map represents the genomic coordinates of a segment start and end. Repair protein signal is colored by the frequencies with which those edges are observed across single cells. Repair protein segments (as called on the linear genome) are annotated underneath, sorted on segment size (large to small, top to bottom). **b,** Pairwise SPRITE contact map (top) of the site shown in **5.3a**. Individual SPRITE clusters (rows) are plotted underneath, aggregated by the three RAD51 DamID repair protein spreading scenarios as in **5.3a-b**. **c,** ChIC CTCF alignment plots (averaged across single cells) over CTCF motifs +/– 1000 bp. **d,** Distribution of ChIC CTCF sequencing depth, unfiltered and filtered (retaining samples passing threshold). **e,** Comparison of virtual 4C (representation of Hi-C contacts formed along the linear genome) and ChIC 53BP1 (bottom) signal, as in **5.3g**. Single-cell aggregates for both conditions are plotted above single-cell heatmaps +/–DSB.

References

1. Ciccia, A. & Elledge, S.J. The DNA damage response: making it safe to play with knives. *Mol Cell* **40**, 179-204 (2010).
2. Hustedt, N. & Durocher, D. The control of DNA repair by the cell cycle. *Nat Cell Biol* **19**, 1-9 (2016).
3. Mitrentsi, I., Yilmaz, D. & Soutoglou, E. How to maintain the genome in nuclear space. *Curr Opin Cell Biol* **64**, 58-66 (2020).
4. Carre-Simon, A. & Fabre, E. 3D Genome Organization: Causes and Consequences for DNA Damage and Repair. *Genes (Basel)* **13**(2021).
5. Arnould, C. *et al.* Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660-665 (2021).
6. Aymard, F. *et al.* Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nat Struct Mol Biol* **24**, 353-361 (2017).
7. Zagelbaum, J. *et al.* Multiscale reorganization of the genome following DNA damage facilitates chromosome translocations via nuclear actin polymerization. *Nat Struct Mol Biol* **30**, 99-106 (2023).
8. Zhu, Q., Niu, Y., Gundry, M. & Zong, C. Single-cell damagenome profiling unveils vulnerable genes and functional pathways in human genome toward DNA damage. *Sci Adv* **7**(2021).
9. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-47 (2015).
10. Zeller, P. *et al.* Single-cell sortChIC identifies hierarchical chromatin dynamics during hematopoiesis. *Nat Genet* **55**, 333-345 (2023).
11. Kefalopoulou, S., Rullens, P. M. J., de Luca, K. L., de Vries, S. S., Korthout T., van Oudenaarden, A., Zeller, P., Kind, J. Time-resolved and multifactorial profiling in single cells resolves the order of heterochromatin formation events during X-chromosome inactivation. (bioRxiv, 2023).
12. Iacovoni, J.S. *et al.* High-resolution profiling of gammaH2AX around DNA double strand breaks in the mammalian genome. *EMBO J* **29**, 1446-57 (2010).
13. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* **18**, 424-8 (2000).
14. Vogel, M.J., Peric-Hupkes, D. & van Steensel, B. Detection of in vivo protein-DNA interactions using DamID in mammalian cells. *Nat Protoc* **2**, 1467-78 (2007).
15. Rooijers, K. *et al.* Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat Biotechnol* **37**, 766-772 (2019).
16. Schmid, M., Durussel, T. & Laemmli, U.K. ChIC and ChEC; genomic mapping of chromatin proteins. *Mol Cell* **16**, 147-57 (2004).
17. Ku, W.L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods* **16**, 323-325 (2019).
18. Massip, L., Caron, P., Iacovoni, J.S., Trouche, D. & Legube, G. Deciphering the chromatin landscape induced around DNA double strand breaks. *Cell Cycle* **9**, 2963-72 (2010).
19. Knijnenburg, T.A. *et al.* Multiscale representation of genomic signals. *Nat Methods* **11**, 689-94 (2014).
20. Caron, P. *et al.* Non-redundant functions of ATM and DNA-PKcs in response to DNA double-strand breaks. *Cell Rep* **13**, 1598-609 (2015).
21. Kind, J. *et al.* Single-cell dynamics of genome-nuclear lamina interactions. *Cell* **153**, 178-92 (2013).
22. Scully, R., Panday, A., Elango, R. & Willis, N.A. DNA double-strand break repair-pathway choice in somatic mammalian cells. *Nat Rev Mol Cell Biol* **20**, 698-714 (2019).
23. Karanam, K., Kafri, R., Loewer, A. & Lahav, G. Quantitative live cell imaging reveals a gradual shift between DNA repair mechanisms and a maximal use of HR in mid S phase. *Mol Cell* **47**, 320-9 (2012).
24. Macheret, M. & Halazonetis, T.D. Intragenic origins due to short G1 phases underlie oncogene-induced DNA replication stress. *Nature* **555**, 112-116 (2018).
25. Moiseeva, T.N. *et al.* An ATR and CHK1 kinase signaling mechanism that limits origin firing during unperturbed DNA replication. *Proc Natl Acad Sci U S A* **116**, 13374-13383 (2019).
26. Klein, K.N. *et al.* Replication timing maintains the global epigenetic state in human cells. *Science* **372**, 371-378 (2021).
27. Marchal, C., Sima, J. & Gilbert, D.M. Control of DNA replication timing in the 3D genome. *Nat Rev Mol Cell Biol* **20**, 721-737 (2019).

28. Du, Q. *et al.* Replication timing and epigenome remodelling are associated with the nature of chromosomal rearrangements in cancer. *Nat Commun* **10**, 416 (2019).
29. Aymard, F. *et al.* Transcriptionally active chromatin recruits homologous recombination at DNA double-strand breaks. *Nat Struct Mol Biol* **21**, 366-74 (2014).
30. Janssen, A. *et al.* A single double-strand break system reveals repair dynamics and mechanisms in heterochromatin and euchromatin. *Genes Dev* **30**, 1645-57 (2016).
31. Tsouroula, K. *et al.* Temporal and Spatial Uncoupling of DNA Double Strand Break Repair Pathways within Mammalian Heterochromatin. *Mol Cell* **63**, 293-305 (2016).
32. Pellegrino, S., Michelena, J., Teloni, F., Imhof, R. & Altmeyer, M. Replication-Coupled Dilution of H4K20me2 Guides 53BP1 to Pre-replicative Chromatin. *Cell Rep* **19**, 1819-1831 (2017).
33. Clouaire, T. *et al.* Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. *Mol Cell* **72**, 250-262 e6 (2018).
34. Nakamura, K. *et al.* H4K20me0 recognition by BRCA1-BARD1 directs homologous recombination to sister chromatids. *Nat Cell Biol* **21**, 311-318 (2019).
35. Becker, J.R. *et al.* BARD1 reads H2A lysine 15 ubiquitination to direct homologous recombination. *Nature* **596**, 433-437 (2021).
36. Schep, R. *et al.* Impact of chromatin context on Cas9-induced DNA double-strand break repair pathway balance. *Mol Cell* **81**, 2216-2230 e10 (2021).
37. Clouaire, T. & Legube, G. A Snapshot on the Cis Chromatin Response to DNA Double-Strand Breaks. *Trends Genet* **35**, 330-345 (2019).
38. Ferrand, J., Plessier, A. & Polo, S.E. Control of the chromatin response to DNA damage: Histone proteins pull the strings. *Semin Cell Dev Biol* **113**, 75-87 (2021).
39. Pfister, S.X. *et al.* SETD2-dependent histone H3K36 trimethylation is required for homologous recombination repair and genome stability. *Cell Rep* **7**, 2006-18 (2014).
40. Natale, F. *et al.* Identification of the elementary structural units of the DNA damage response. *Nat Commun* **8**, 15760 (2017).
41. Ochs, F. *et al.* Stabilization of chromatin topology safeguards genome integrity. *Nature* **574**, 571-574 (2019).
42. Sanders, J.T. *et al.* Radiation-induced DNA damage and repair effects on 3D genome organization. *Nat Commun* **11**, 6178 (2020).
43. Collins, P.L. *et al.* DNA double-strand breaks induce H2Ax phosphorylation domains in a contact-dependent manner. *Nat Commun* **11**, 3158 (2020).
44. Huang, J. *et al.* SLFN5-mediated chromatin dynamics sculpt higher-order DNA repair topology. *Mol Cell* (2023).
45. Quinodoz, S.A. *et al.* Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. *Cell* **174**, 744-757 e24 (2018).
46. Canela, A. *et al.* Genome Organization Drives Chromosome Fragility. *Cell* **170**, 507-521 e18 (2017).
47. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).
48. Nagano, T. *et al.* Cell-cycle dynamics of chromosomal organization at single-cell resolution. *Nature* **547**, 61-67 (2017).
49. Stevens, T.J. *et al.* 3D structures of individual mammalian genomes studied by single-cell Hi-C. *Nature* **544**, 59-64 (2017).
50. Bintu, B. *et al.* Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* **362**(2018).
51. de Luca, K.L. & Kind, J. Single-cell DamID to capture contacts between DNA and the nuclear lamina in individual mammalian cells. *Methods Mol Biol* **2157**, 159-172 (2021).
52. van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P. & van Steensel, B. Cell cycle dynamics of lamina-associated DNA. *EMBO Rep* **21**, e50636 (2020).
53. Markodimitraki, C.M. *et al.* Simultaneous quantification of protein-DNA interactions and transcriptomes in single cells with scDam&T-seq. *Nat Protoc* **15**, 1922-1953 (2020).
54. Quinodoz, S.A. *et al.* SPRITE: a genome-wide method for mapping higher-order 3D interactions in the nucleus using combinatorial split-and-pool barcoding. *Nat Protoc* **17**, 36-75 (2022).
55. prullens. github.com/KindLab/scRepair [10.5281/zenodo.13927895] (2024)
56. Li, J., Witten, D.M., Johnstone, I.M. & Tibshirani, R. Normalization, testing, and false discovery rate estimation for RNA-sequencing data. *Biostatistics* **13**, 523-38 (2012).
57. Townes, F.W., Hicks, S.C., Aryee, M.J. & Irizarry, R.A. Feature selection and dimension reduction for single-cell RNA-Seq based on a multinomial model. *Genome Biol* **20**, 295 (2019).

58. Diserud, O.H. & Odegaard, F. A multiple-site similarity measure. *Biol Lett* **3**, 20-2 (2007).
59. Baselga, A., Jimenez-Valverde, A. & Niccolini, G. A multiple-site similarity measure independent of richness. *Biol Lett* **3**, 642-5 (2007).
60. Baselga, A. Separating the two components of abundance-based dissimilarity: balanced changes in abundance vs. abundance gradients. *Methods in Ecology and Evolution* **4**, 552-557 (2013).
61. Baselga, A. & Orme, C.D.L. Betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution* **3**, 808-812 (2012).
62. de Luca, K.L. "scRepair2024 — imaging files belonging to de Luca & Rullens et al. 2024", Mendeley Data, V1, doi: 10.17632/84x5yfhdw4.1 (2024)
63. Created in BioRender. de Luca, K.L. (2024) BioRender.com/c99h067

Chapter 6

Higher-order clustering and coordinated repair in response to DNA double-strand breaks

Kim L. de Luca[1,2*‡], Pim M. J. Rullens[1,2,3*], Magdalena A. Karpinska[4], Sandra S. de Vries[1,2], Agnieszka Gacek-Matthews[5], Lőrinc S. Pongor[6], Gaëlle Legube[7], Joanna W. Jachowicz[5], A. Marieke Oudelaar[4], Jop Kind[1,2,3‡]

[1] Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht; Utrecht, The Netherlands

[2] Oncode Institute; The Netherlands

[3] Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University; Nijmegen, The Netherlands

[4] Genome Organization and Regulation, Max Planck Institute for Multidisciplinary Sciences; Göttingen, Germany

[5] Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna Biocenter (VBC); Vienna, Austria

[6] Cancer Genomics and Epigenetics Core Group, Hungarian Center of Excellence for Molecular Medicine (HCEMM); Szeged, Hungary

[7] MCD, Centre de Biologie Intégrative (CB), CNRS, Université de Toulouse, Université Toulouse III – Paul Sabatier; Toulouse, France

* Co-first authorship

‡ Co-corresponding authorship

Adapted from article published in ***Nature Communications*** (2024).
Author list reproduced as published.

Abstract

DNA double-strand breaks (DSBs) are among the most severe forms of genomic damage, posing a major threat to genome stability and cell survival. Failure to repair DSBs accurately can lead to mutations, chromosomal rearrangements, and disease, including cancer. The repair process unfolds within a complex nuclear environment, where chromatin organization plays a key role in damage recognition and processing. Repair outcomes are influenced not only by local chromatin state but also by the spatial positioning of DSBs within the 3D genome. While individual repair factor recruitment has been extensively studied, it remains unclear how multiple DSBs coordinate their repair across the nuclear space.

Here, we investigate repair protein clustering and multi-way coordination at damaged loci using single-cell DamID and ChIC sequencing, in combination with Hi-C and Tri-C topological maps. Using Tri-C, we provide direct evidence of multi-way chromatin contacts between DSB sites, demonstrating long-range clustering in response to damage. Single-cell repair binding profiles further show that these loci are coordinately bound by repair proteins. Notably, larger hubs show increased coordination, suggesting cooperative repair mechanisms. These findings reveal a hierarchical organization of repair protein binding in single cells, highlighting the role of 3D genome architecture in coordinating DNA repair.

Introduction

Besides local reorganization of the genome in response to damage as described in *Chapter 5*, DSBs exhibit large-scale intra-nuclear motion[1-5]. The phenomenon of repair foci clustering was first observed by microscopy[6]: over time, a reduction in the number of γH2AX foci and an increase in their size supported the breakage-first theory of chromosomal translocations. Live-cell imaging further showed fusion of separate repair foci by fluorescent tagging of repair proteins 53BP1[2,7,8] and Rad52[5].

At the genomic (rather than the protein) level, high-throughput sequencing experiments comparing chromosomal contacts in damaged and undamaged conditions showed that DSB loci physically interact[3,8,9]. This provides indirect evidence of DSB clustering; however, these approaches rely on bulk population averaging, making it impossible to determine whether clustering occurs within individual cells, or represents an ensemble of distinct interactions. To date, no experimental evidence has been presented that directly couples repair foci clustering and genomic identity of DNA breaks.

To bridge this gap, we use single-cell sequencing of DNA repair factor binding, which allows for the direct quantification of repair protein occupancy at individual DSB sites, and the identification of coordinated repair events across damaged loci. In parallel, we perform single-molecule sequencing of chromatin conformation to directly assess DSB clustering. By quantifying repair protein occupancy within damaged clusters, we find that repair coordination is predominantly cooperative but can also involve competitive interactions. Together, our findings provide a high-resolution, single-cell perspective on how genome topology contributes to the DNA damage response.

Given prior evidence that DSBs can engage in large-scale nuclear reorganization, we set out to determine whether repair proteins exhibit coordinated binding across damaged loci. By integrating single-cell repair binding data with Hi-C chromatin conformation maps, we assessed whether spatially contacting DSBs may share repair machinery.

Coordination of single-cell repair protein binding corresponds to Hi-C contact frequency

First, we re-analyzed recent Hi-C data[10] for downstream visualization and statistical purposes, confirming that some DSB sites form long-range and often inter-chromosomal contacts upon damage (Fig. 6.1a, Supplementary Fig. 6.1a). Damage-specific contacts were identified by computing a differential Hi-C matrix, which quantifies the fold-change between control and damage-induced conditions (Fig. 6.1bi, different *x*-axis region in Fig. 6.1bii). We defined pairs of sites as contacting by setting a threshold on the differential matrix (Sup. Fig. 6.1b), and validate Hi-C signal up to +/- 0.5 Mb surrounding those DSB sites, while non-contacting sites are fully devoid (Fig. 6.1c, top).

We reasoned that, in order for DSB pairs to physically interact upon DNA damage in a given cell, both sites should be simultaneously occupied by the repair machinery to gain affinity for one another. Genome-wide single-cell data offer a unique possibility to examine such coordinated protein binding, also between distant and inter-chromosomal loci. We applied the same bin-based coordination metric as in Figure 3 of *Chapter 5*: if two loci are both coordinately bound by repair protein across single cells, these loci show high correlation. Indeed, we find highly coordinated 53BP1 binding events between pairs of individual break sites, which strongly overlaps the enriched Hi-C damage-induced contacts (Fig. 6.1d). We systematically distinguish all contacting DSB pairs from those that do not contact and find coordinated 53BP1 binding exclusively at contacting DSB pairs (Fig. 6.1c, bottom).

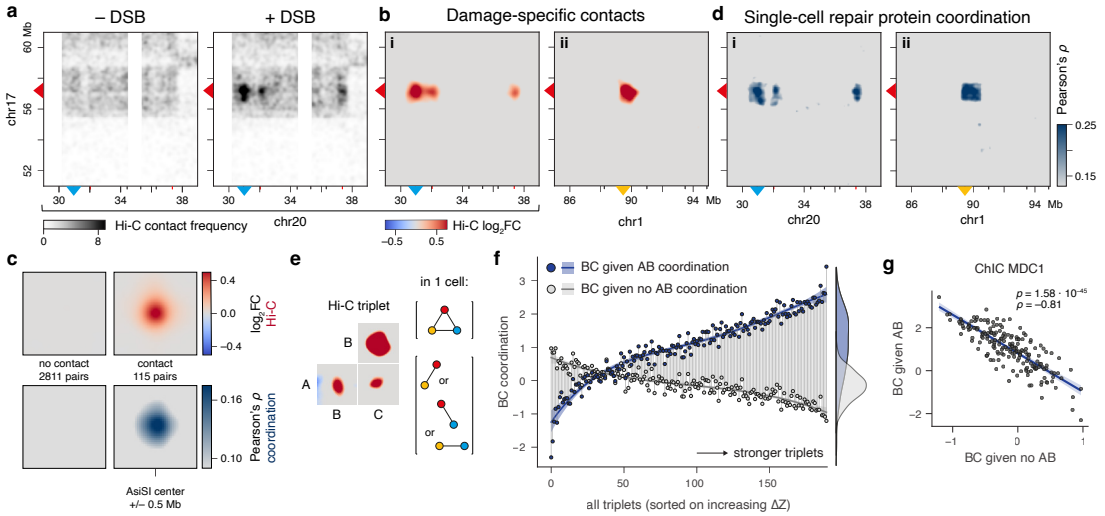


Fig. 6.1. Pairwise repair coordination of long-range contacts in single cells.
a, Hi-C signal in 10-Mb regions harboring DSB sites. The left and right panel show Hi-C contacts of chr17 (*y*) and chr20 (*x*) in –DSB and +DSB conditions, respectively. **b,** Left panel shows differential Hi-C contacts ($\log_2$FC (+DSB / –DSB)) of chr17 with chr20 as presented in **a**. Right panel shows differential Hi-C contacts of a different region between chr17 and chr1. **c,** Comparison of Hi-C contact score (top) and single-cell repair coordination (bottom) across all pairs of DSB sites, categorized by contact (right) or no contact (left) based on log2FC threshold. A region of 1 Mb surrounding the AsiSI motif is shown. Coordination is measured as Pearson's correlation between 40-kb bins across cells. **d,** Single-cell repair protein coordination (53BP1) of all pairwise combinations of bins in selected regions as in **b**. Coordination is measured as Pearson's correlation between 40-kb bins across cells. **e,** Left panel shows Hi-C based definition of a triplet where all three sites A, B and C have pairwise contacts. Right panel shows a schematic representation of higher-order (top) and mutually exclusive (bottom) contacts. **f,** Scatter plot measures three-way coordination of triplets. Each triplet is represented twice: BC co-repair is quantified as the Z-score normalized Jaccard index between B and C in 1) cells that have AB co-repair (blue) and 2) in cells without AB co-repair (grey). Triplets are sorted on the *x*-axis based on the ΔZ(Jaccard). Lines and shaded areas represent polynomial regressions ± 95% CI. **g,** Scatter plot combining both groups of **f**, so that each triplet is represented only once. Line and shaded area represent best-fit linear regression ± 95% CI. Pearson's correlation coefficient is plotted together with the *p*-value derived from a two-sided null hypothesis test.

While this correlation analysis captures quantitative signals in individual genomic bins, it does not consider whether signal-containing bins are part of a consecutive sequence (as is the case in a repair domain). To know whether entire repair domains also behave in a coordinated manner, we turned to our MSR segment calls, which represent presence or absence of repair protein along the entire DSB region.

Because these segment calls are binarized, we quantified coordination with the Jaccard index, which measures the similarity between two sets of observations. In this context, that is the overlap in presence of repair protein segment A and B among all cells. This analysis confirmed that contacting pairs show highly increased repair protein coordination, at the level of whole domains, in all DDR protein single-cell datasets (Sup. Fig. 6.1c).

Coordination of single-cell repair protein binding is variable across cells

In the analyses described above for Figure 6.1, both the Hi-C contacts and repair coordination measure pairwise (two-way) events, meaning they are limited to two DSB sites at a time. Notably, the population-based Hi-C data indicates that various DSBs are in close physical proximity to more than one other DSB—some DSBs contact >15 other sites (Sup. Fig. 6.1d). Consequently, it is currently unknown to what extent contacting sites form multi-way, or higher-order, DSB repair hubs. The simplest form of such a multi-way hub is a triplet of given DSB sites A, B, and C (Fig. 6.1e). In a triplet, Hi-C contacts are found between all three sites, forming pairwise combinations AB, AC, and BC. However, the question remains whether triplet ABC is formed within one cell, or represents mutually exclusive contacts that occur independently across cells.

Although scDamID and scChIC profiles do not measure spatial localization of DSBs, pairwise damage-specific contacts are prominently recapitulated. Hence, we reasoned we could test the premise of intercellular heterogeneity by quantifying repair coordination of DSB triplets. A triplet is considered cooperative if AB and BC are coordinated, and competitive if AB hinders BC.

Our analysis reveals that repair coordination is predominantly cooperative: if two DSBs share a repair factor, they are more likely to recruit the same factor at a third site. This is indicated by higher coordination scores of BC given AB (blue) compared to BC given no AB (grey) (Fig. 6.1f). However, in a subset of cases, repair factor occupancy is mutually exclusive, suggesting competition for repair resources. These competitive triplets exhibit lower coordination scores overall, indicating variability in repair hub formation across cells (Fig. 6.1f). Notably, cooperative and competitive behavior are highly anti-correlated, for all measured repair factors (Fig. 6.1g, Sup. Fig. 6.1c). This suggests that, in most cells, DSBs are coordinately bound by repair protein within the triplet; only in very few cells will triplet sites be bound coordinately with another DSB (or remain unbound by repair protein altogether).

In sum, by distinguishing multi-way repair coordination within one cell from different combinations of pairwise coordination across cells, we find increased cooperative repair, in support of multi-way DSB repair hub formation.

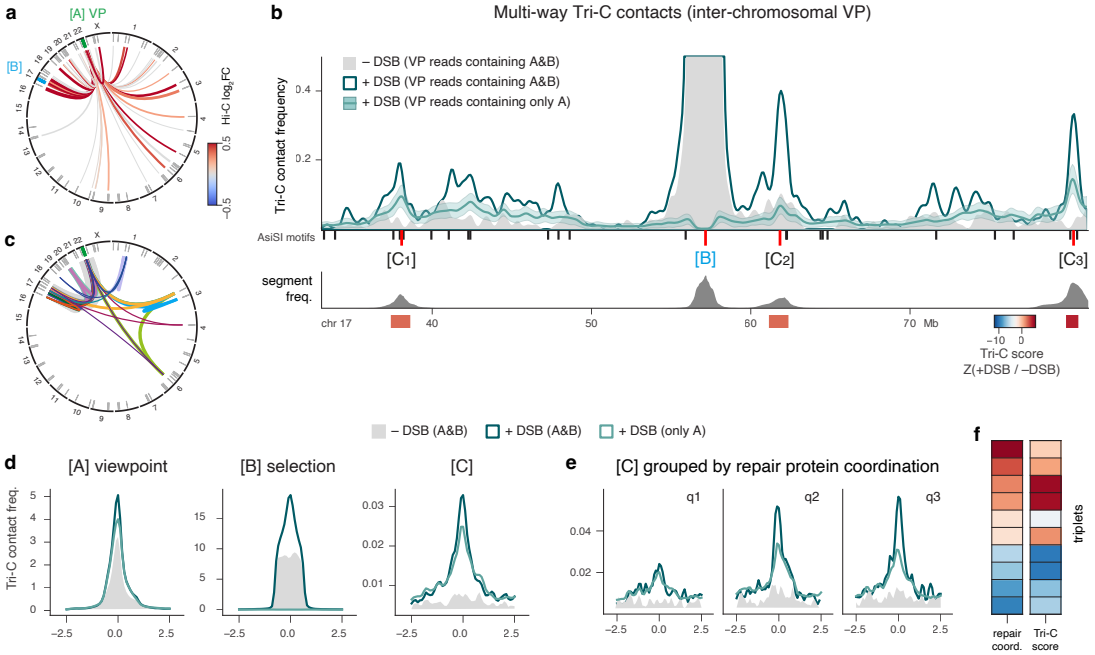


Fig. 6.2. Damage-specific three-way contacts identified by Tri-C.
a, Circos plot showing pairwise Tri-C contacts (lines) between the given VP and all other DSB top sites across the genome. The width of the lines represents the Tri-C contact frequency and the color represents the Hi-C fold-change between –DSB and +DSB conditions. Ticks indicate the DSB site locations. **b,** (top) Contact profile of positively selected reads that, aside from the VP (defined as site A), also have a fragment overlapping a second DSB site B. Conditions are –DSB (grey) and +DSB (teal). Conversely, contact profile of negatively selected reads, specifically lacking a fragment overlapping site B (light teal, mean ± SD). (bottom) Repair protein segment frequency of 53BP1. Below, repair domains (boxes) are used to quantify the Z-score at DSB sites [C] between positively and negatively selected profiles. **c,** Circos plot showing all three-way Tri-C contacts formed by the given VP, with a minimum Tri-C score of 0.5. Line width represents Tri-C score. Each color represents one three-way contact. **d,** Alignment plot of +DSB contact profiles as in **b** for all triplets of all 13 Tri-C viewpoints, where each time the VP was defined as site A, positive (teal) or negative (light teal) selection was done on site B, and the resulting contact profile is examined on site C. The grey contact profile is that of positive selection for site B in the –DSB condition. **e,** Same as the right panel of site C in **c**, but here stratified over 3 quantiles (q) that are based on increasing single-cell repair protein coordination. **f,** Heatmaps comparing single-cell repair protein coordination (left) to three-way Tri-C contacts (right). Coordination is measured as BC co-repair given AB co-repair. Tri-C contact is measured as Z-score of triplet site C. For this VP, all triplets are shown that contain a minimum of 30 reads at site C.

Repair protein foci cluster in multi-way chromatin hubs

To solidify our hypothesis that higher-order coordination of repair protein binding events suggests the presence of DSB repair contact hubs, we sought to prove that damaged sites form multi-way contacts on the genomic level. We addressed this gap by performing Tri-C experiments[11,12] in damaged and undamaged conditions. Tri-C is a multi-way 3C approach, which enables identification of multiple ligation junctions in 3C concatemers. Since the order of fragments in these concatemers represents the 3D conformation of individual alleles, Tri-C gives insight into higher-order chromatin structures formed in single cells. By using capture oligonucleotide-mediated enrichment of regions of interest, Tri-C allows for analysis at high resolution and sensitivity. We designed 13 unique viewpoints (VPs) that each cover a single DSB site. The final DNA-sequencing reads are thus expected to contain a viewpoint fragment, and one or more proximal fragments, originating from single-allele chromatin conformations.

First, we used Tri-C to validate the presence of DSB-induced pairwise contacts between a given VP and other DSB sites. Both long-range and inter-chromosomal contacts are specifically formed in the damaged condition (Fig. 6.2a, Sup. Fig. 6.2a-b). As intended, many of the Tri-C reads are composed of more than two contact fragments (Sup. Fig. 6.2c). To analyze multi-way hub formation between DSB sites, we turned to the minimal multi-way hub case: a triplet of given DSB sites A, B and C (Fig. 6.1e). We defined the Tri-C experimental VP as site A, and performed in silico selection at site B (see Methods). Briefly: all reads containing site A are divided into two sets: one that contains only site A (negative selection), and one that contains both A and B (positive selection). We then visualized the contact profiles of the remaining proximal fragments. These fragments are specifically enriched at other DSBs (*i.e.*, sites C), thus forming three-way interactions on single alleles (Fig. 6.2b, Sup. Fig. 6.2d). To quantify this observation, we calculated damage-specific contacts at site C, termed Tri-C score, for each possible triplet ABC (see Methods and Fig. 6.2b, bottom).

Using this measure, we could now visualize triplets formed with this VP across the genome (Fig. 6.2c). Our finding of three-way topologies is supported by all 13 Tri-C datasets (Fig. 6.2d). Finally, we aimed to integrate the multi-way Tri-C analysis with single-cell coordination of DSB repair protein. Triplets were grouped into three quantiles based on increasing repair coordination. Indeed, higher repair protein coordination correlated with increased Tri-C contact enrichment (Fig. 6.2e), even at the level of individual triplets (Fig. 6.2f). In sum, we present direct evidence of multi-way clustering in response to damage, at single-molecule resolution.

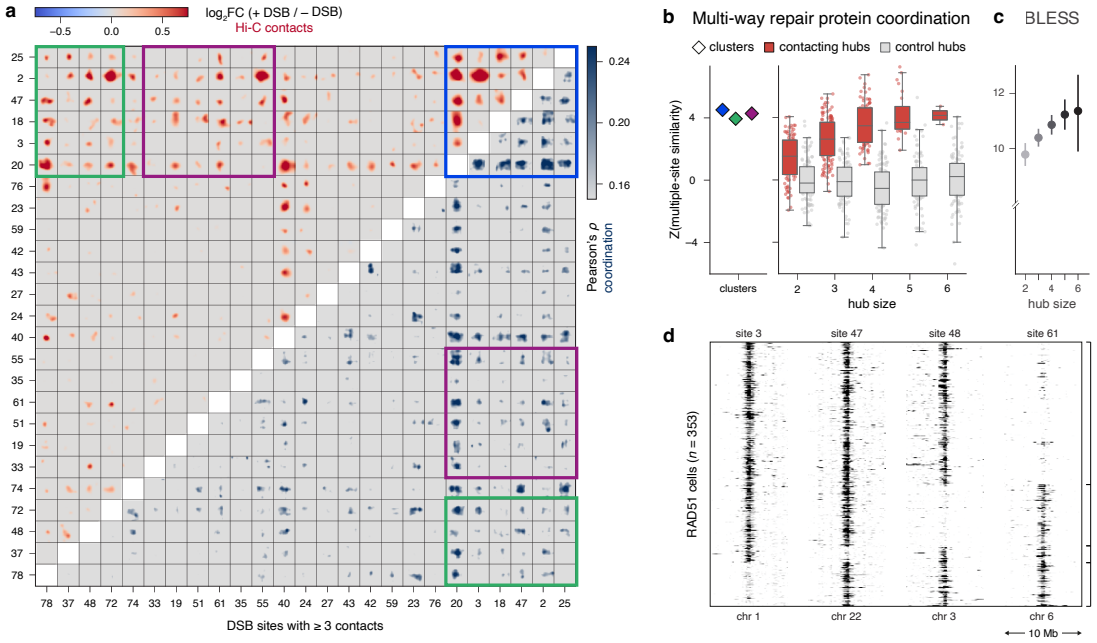


Fig. 6.3. Higher-order coordination of repair protein binding in single cells. **a,** Clustered heatmap and hub annotation as in **Sup. Fig. 6.3a**, displaying pairwise damage-specific Hi-C contacts (red, top left) and single-cell 53BP1 repair coordination (marine blue, bottom right), in a region of 1 Mb surrounding each AsiSI site. **b,** Higher-order repair coordination. Left: clusters corresponding to those identically colored in **a**. Right: Across all hub sizes, coordination is measured for hubs in contact (red) and control hubs of the same size (grey). Controls are of randomly drawn top sites (100 times). The boxes represent the median, first and third quartiles of the data. **c,** BLESS score (ref. [3]) measured across hub sizes, as mean ± 95% CI. **d,** Heatmap showing single-cell Dam-RAD51 signal (RPKM) at 4 DSB sites in 10-Mb windows. Rows (*i.e.*, cells) are ordered according to different combinations of simultaneous RAD51 binding in each cell.

Higher-order repair coordination in single cells increases with hub size

As described above, multi-way DSB repair hub formation is prominently observed within single cells and on single molecules. While Tri-C contacts are currently limited to predominantly three-way structures, genome-wide single-cell profiles allow for theoretically infinite combinations. Thus, we set out to evaluate higher-order repair coordination, using Hi-C as an independent and orthogonal measurement. Based on pairwise Hi-C data, we selected DSB sites that each form 3 or more proximal contacts (Sup. Fig. 6.1d). Hierarchical clustering identified subsets of sites that frequently interact (Fig. 6.3a, Sup. Fig. 6.3a; clusters indicating colored boxes). The clusters are remarkably mirrored by pairwise repair coordination across single cells (Fig. 6.3a).

From this DSB Hi-C matrix, we identified a few hundred hubs of sites that have pairwise contacts between all participating loci. These hubs vary in size (*i.e.*, number of contacts), with some extremes consisting of up to 6 DSBs (Sup. Fig. 6.3b).

To directly test higher-order coordination in hubs, we applied a multiple-site similarity measure (see Methods and refs. [13,14]). First, we measured multi-way coordination in the large clusters identified from the Hi-C matrix. Indeed, these clusters showed very high coordination of repair protein binding (Fig. 6.3b, left). Next, in a systematic analysis, contacting hubs (red) showed considerably more multi-way repair coordination than control hubs of the same size (grey) that were randomly selected from all sites (Fig. 6.3b, right). Moreover, coordination consistently increased according to hub size, implying that cooperative repair (rather than multiple separate hubs) preferentially occurs within one cell. In support of this hypothesis, larger hubs show more frequent repair protein binding (Sup. Fig. 6.3c). That is, a given DSB is bound in (many) more cells when that site is part of a larger hub. This is in line with our finding of cooperative versus competitive repair within triplets (Figure 6.1). In addition, larger hubs contain sites that are more frequently damaged, as shown by BLESS[15], a method that detects genome-wide DSB distribution. Damage propensity can thus be interpreted as a source of variability in hub formation that is mimicked by repair protein recruitment and subsequent cooperative motion. Finally, we sought to explore variability in repair protein binding within hubs. A given hub of 4 DSBs (with pairwise Hi-C contacts between all) shows different combinations of simultaneous repair protein binding (Fig. 6.3d).

Our analysis thus shows evidence of cooperative and mutually exclusive higher-order DSB repair contacts, which are indistinguishable in Hi-C population data where all cases appear as pairwise interactions. Collectively, these data illustrate a range of repair coordination, with multi-way architectural repair hubs that may heterogeneously exist across single cells in various conformations.

Conclusion

We established the framework for sensitive, genome-wide, single-cell DNA repair factor profiling, enabling detection of repair protein enrichment at kilobase-scale resolution across thousands of individual cells. The sequence-specific DSB induction system allowed us to examine coordinated repair protein occupancy at known genomic positions. We provide the first single-cell genomic evidence of multi-way repair hubs—directly linking chromatin conformation to coordinated repair factor recruitment—a phenomenon not captured by population-based spatial contact approaches.

Our findings support emerging models in which the physical properties of chromatin, compartmentalization, and phase separation of repair proteins drive the self-organization of damaged loci[8,16-19]. This suggests that spatial genome architecture and mechanochemical signaling act in concert to regulate genome stability, adding a new layer of complexity to the DNA damage response.

Acknowledgements

We thank members of the Kind and Legube groups and Eva Brinkman for feedback throughout the project. Specifically, we acknowledge Koos Rooijers for developing computational methods and performing bioinformatic analyses. Purified pA-MNase and ^{m6}A-Tracer were respective gifts from Alexander van Oudenaarden and Bas van Steensel. We thank Reinier van der Linden at the Hubrecht FACS Facility and Anko de Graaff at the Hubrecht Imaging Center for technical support and training. High-throughput sequencing and computational resources were provided by the Utrecht Sequencing Facility (USEQ) and High-Performance Computing Facility. USEQ is subsidized by the University Medical Center Utrecht and the Netherlands X-omics Initiative (NWO project 184.034.019). AMO acknowledges funding from the Max Planck Society. The Oncode Institute is supported by the KWF Dutch Cancer Society. This work was further funded by: Dutch Research Council (NWO-ALW) Vidi grant 161.339 (JK), Fulbright Commission the Netherlands promovendus grant (KLdL), European Molecular Biology Organization grant 9178 (KLdL).

Author contributions

Conceptualization: KLdL and JK. Methodology, Investigation, Data curation & Validation: KLdL. MAK contributed Tri-C experiments, advised by AMO. SSdV performed revision experiments with KLdL. Formal analysis & Visualization: KLdL and PMJR. Software: PMJR. Project administration: JK. Resources: GL, AMO, and JK. Funding acquisition: KLdL, AMO, and JK. Writing – original draft: KLdL with contributions from PMJR and JK. Writing – review & editing: all authors.

Methods

This section describes experimental and computational methods that relate only to *Chapter 6*, namely Tri-C data and multi-way coordination analyses. See *Chapter 5* for all other methods.

Experimental methods

Tri-C

Measurement of multi-way contacts with Tri-C was performed following previously published protocols[12], in 2 biological replicates per experimental condition and 4 technical replicates per biological replicate. Briefly, cells were collected in culture medium in batches of 15 M cells per technical replicate and cross-linked for 10 minutes with 2% formaldehyde (ThermoFischer, 28908). To prepare 3C libraries, aliquots of cells were split equally into 3 reactions and digested with *Nla*III enzyme (NEB, R0125L). Then, a proximity ligation was performed and ligated chromatin was extracted with Phenol-Chloroform method. The separate digest reactions were combined. 8 µg of 3C library per technical replicate was sheared with Covaris S220 Focused-Ultrasonicator to the mean size of 450 bp (time: 55 s, duty factor: 10%, peak incident power: 140 W, cycles per burst: 200). In order to exclude fragments shorter than 300 bp, which are unlikely to contain more than one ligation junction, the samples were size-selected with 0.7x of Mag-Bind TotalPure NGS beads (Omega Bio-Tek, M1378-01). Samples were indexed in duplicate in order to increase sample complexity using NEBNext Ultra II DNA Library Prep Kit for Illumina (NEB, E7645S), and Herculase II polymerase (Agilent, 600677) for sample amplification.

Indexed libraries were enriched for the viewpoints of interests in a double-capture procedure. Probes used for capture were designed with the Python-based oligo tool (https://oligo.readthedocs.io/en/latest/). The 120 nt long, 5'-biotinylated, ssDNA probes were ordered as a multiplexed panel of oligos (IDT, xGen™ Custom Hybridization Capture Panels), and used at 2.9 nM concentration. The enrichment was performed using the KAPA Hyper Capture Reagent Kit (Roche, 9075828001). A total of 12 µg of indexed sample per biological replicate was used as input for the first capture. Captured DNA was pulled down with M-270 Streptavidin Dynabeads (Invitrogen, 65305), washed, and PCR amplified. All recovered material was used as input for the second capture. The quality of final samples was assessed by fragment analyzer and samples were sequenced on an Illumina platform with 300 cycles paired-end reads.

Computational methods

Raw data processing

Tri-C samples

Tri-C data were processed using the CapCruncher pipeline[12] (v.0.3.11) in tiled mode.

Identifying contacting hubs

The identification of hubs of contacting AsiSI clusters builds on the pair identification described above, but is elaborated by walking through Hi-C matrix M^2 (Sup. Figure 6.1g), testing all theoretical combinations of hubs with defined sizes for Hi-C contacts. The hub identification function for any number of k AsiSI clusters included in theThe hub identification function for any number of k AsiSI clusters included in the hub can be written as $H = \frac{n!}{k!(n-k)!}$ where n is the total number of AsiSI clusters included in M^2. All theoretical combinations of hubs were called as contacting hub if all AsiSI clusters within H contact each other (Hi-C logFC ≥ 0.2). Our hub identification function written in Python is available on GitHub.

Triplet categorization

For the analysis of (in)complete triplets (Figure 6.1), we built on our pairwise identification described above. We first determined all combinations of three among a total of 79 AsiSI clusters that contain a previously annotated top site. The theoretical number of triplets can be defined as above, where n is the number AsiSI clusters (79) and k = 3. Theoretical triplets T were tested for the presence of Hi-C based contacts (logFC ≥ 0.2) and called "complete" in case all three sites contact each other (AB, AC, BC), or "incomplete" if one pair within the triplet does not contact while both other pairs do (AB, AC, BC).

Conditioning of cells for triplet analysis

For triplet analysis presented in Fig. 6.1f-g, cells were stratified by the presence of a binary MSR segment on both A and B (*i.e.*, AB co-repair). Cells with an MSR segment on either A or B were used to classify no AB co-repair. In these two conditions of cells the pairwise coordination of BC was measured using the Jaccard similarity index as described above.

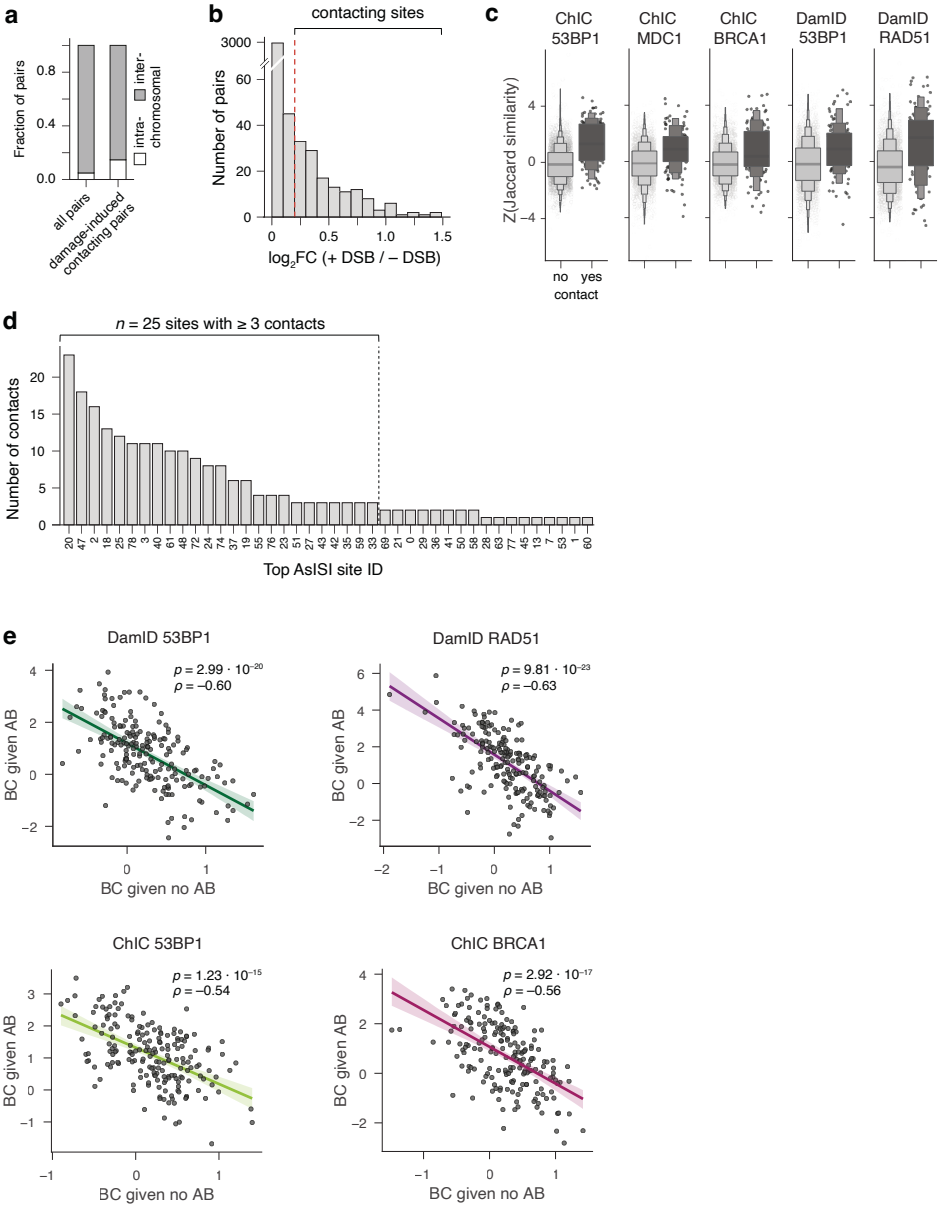
Coordination Z-score normalization by matrix permutation

Two potential problems are posed on our coordination analysis that might introduce undesirable bias: 1) typical sparsity of single-cell data results in non-uniformly distributed signal dropout and 2) binary similarity metrics can be sensitive to site prominence that differs between sites, but that does not reflect coordination.

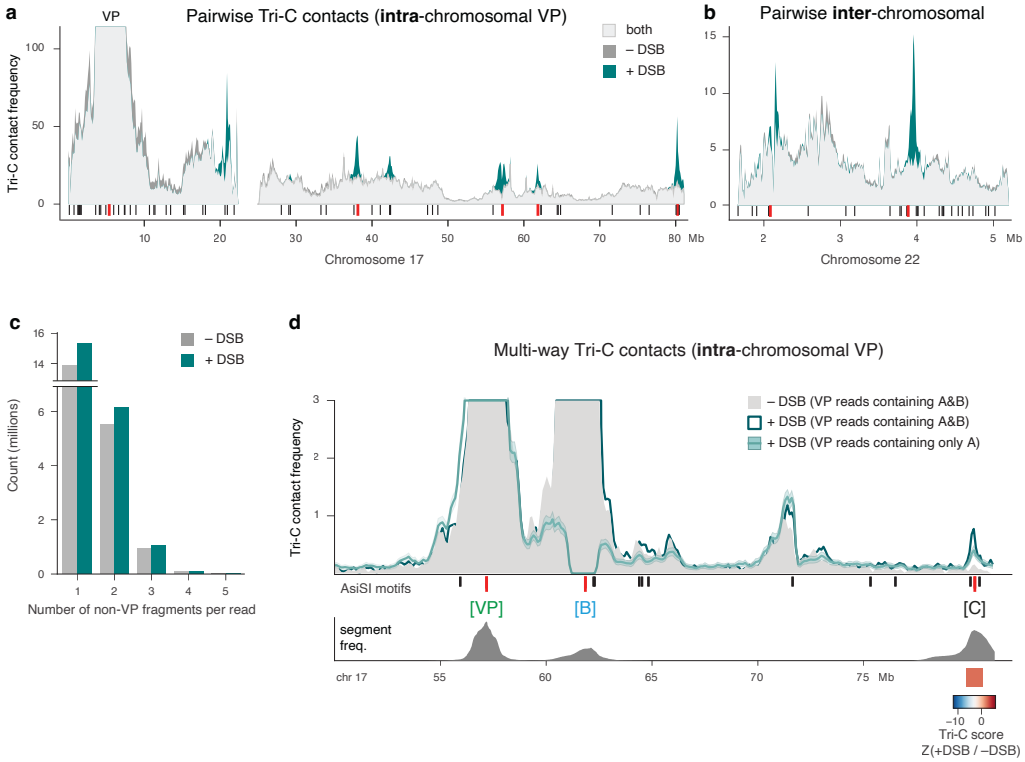
To solve both problems, we applied a previously described algorithm[20], to randomize the abovementioned presence-absence matrix M n-times (n = 100), without altering row and column totals. The resulting randomized matrices were used to Z-score normalize binary coordination metrics, which can be written as $Z = \frac{x - \mu}{\sigma}$, where x is the similarity score of the observed, μ the mean similarity, and σ the standard deviation of the random controls.

Multi-way contact triplet analysis in Tri-C data

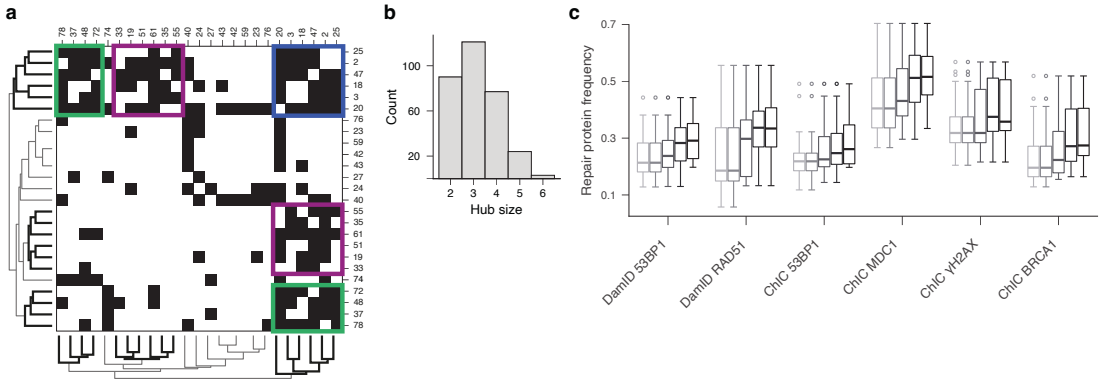
To measure the presence or absence of multi-way contacts of triplets, we devised an association analysis inspired by previous work[21]. As described above, a triplet is defined as three AsiSI clusters (A, B, and C) that are all observed to have contact in the pairwise Hi-C data. In our design, the experimental Tri-C viewpoint (VP) captures AsiSI cluster A, ensuring that each observed read represents a contact that includes A. To discern if A, B, and C all contact within a single hub or rather form mutually exclusive contacts, we computationally positively selected reads containing B and quantify the presence of third interaction partner C. In parallel we did the inverse, negatively selecting reads without B and quantified the presence of C. Negative selection was repeated 100 times to create a statistical background profile (mean ± standard deviation) to which the positively selected profile could be compared. Importantly, the positively selected reads are required to contain B, while in negative selection no such constraints are imposed. The positive profile is therefore effectively generated by smaller reads, where one fewer fragment contributes to each profile. To compensate for this technical artifact, we randomly remove one fragment from each read prior to negative selection. The presence of C is statistically quantified by computing the Z-score between the positively and n-times negatively selected profile as $Z = \frac{x - \mu}{\sigma}$, where x is the mean of the positively selected Tri-C profile in a defined window surrounding C, μ the mean and σ the standard deviation in the same window surrounding C, across the n-times negatively selected profile. The defined window surrounding C is based on the width of our repair segment enrichment. A positive Z-score indicates that A, B, and C preferentially coalesce in a single hub, while a negative Z-score indicates that, although A, B, and C all demonstrate pairwise interactions with each other, they preferentially do so in a mutually exclusive manner.



Sup. Fig. 6.1. Characterization of pairwise contacts and associated repair features. **a,** Quantification of (damage-specific) contacting pairs as intra- or inter-chromosomal. **b,** Histogram showing the number of DSB pairs across Hi-C contact scores (log_2FC (+DSB / –DSB)). **c,** Pairwise single-cell coordination of binarized repair enrichment (MSR calls). Coordination is measured with the normalized Jaccard index, for pairs categorized as in **6.1d**. **d,** Bar plot showing number of contacts formed per site. **e,** Scatter plot indicating pairwise similarities of BC for both cases of AB. Same as in **6.1g** for the indicated repair proteins.



Sup. Fig. 6.2. Characterization of Tri-C contacts.
a-b, Pairwise Tri-C contact profile in –DSB (grey) and +DSB (teal) conditions. Ticks underneath indicate location of DSB sites with previously annotated top sites in red. Intra-chromosomal contacts on chr 17 in **a,** and inter-chromosomal contacts on chr 22 of the same VP in **b**. **c,** Histogram showing the distribution of reads according to the number of non-viewpoint fragments per read. **d,** Multi-way Tri-C contact profile as in **6.2b**, but showing an intra-chromosomal VP.



Sup. Fig. 6.3. Characterization of higher-order contacts and associated repair features. **a,** Heatmap of all sites forming ≥ 3 contacts, clustered based on binarized contact score. Colored rectangles indicate hubs of sites that display increased multi-way contacts between them. **b,** Histogram showing the number of formed hubs containing 2-6 sites. **c,** Repair protein frequency measured across hub sizes, for all datasets.

References

1. Neumaier, T. *et al.* Evidence for formation of DNA repair centers and dose-response nonlinearity in human cells. *Proc Natl Acad Sci U S A* **109**, 443-8 (2012).
2. Caron, P. *et al.* Non-redundant functions of ATM and DNA-PKcs in response to DNA double-strand breaks. *Cell Rep* **13**, 1598-609 (2015).
3. ymard, F. *et al.* Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. *Nat Struct Mol Biol* **24**, 353-361 (2017).
4. Caridi, C.P. *et al.* Nuclear F-actin and myosins drive relocalization of heterochromatic breaks. *Nature* **559**, 54-60 (2018).
5. Schrank, B.R. *et al.* Nuclear ARP2/3 drives DNA break clustering for homology-directed repair. *Nature* **559**, 61-66 (2018).
6. Aten, J.A. *et al.* Dynamics of DNA double-strand breaks revealed by clustering of damaged chromosome domains. *Science* **303**, 92-5 (2004).
7. Krawczyk, P.M. *et al.* Chromatin mobility is increased at sites of DNA double-strand breaks. *J Cell Sci* **125**, 2127-33 (2012).
8. Arnould, C. *et al.* Chromatin compartmentalization regulates the response to DNA damage. *Nature* **623**, 183-192 (2023).
9. Zagelbaum, J. *et al.* Multiscale reorganization of the genome following DNA damage facilitates chromosome translocations via nuclear actin polymerization. *Nat Struct Mol Biol* **30**, 99-106 (2023).
10. Arnould, C. *et al.* Loop extrusion as a mechanism for formation of DNA damage repair foci. *Nature* **590**, 660-665 (2021).
11. Oudelaar, A.M. *et al.* Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet* **50**, 1744-1751 (2018).
12. Downes, D.J. *et al.* Capture-C: a modular and flexible approach for high-resolution chromosome conformation capture. *Nat Protoc* **17**, 445-475 (2022).
13. Diserud, O.H. & Odegaard, F. A multiple-site similarity measure. *Biol Lett* **3**, 20-2 (2007).
14. Baselga, A. & Orme, C.D.L. Betapart: An R package for the study of beta diversity. *Methods in Ecology and Evolution* **3**, 808-812 (2012).
15. Crosetto, N. *et al.* Nucleotide-resolution DNA double-strand break mapping by next-generation sequencing. *Nat Methods* **10**, 361-5 (2013).
16. Kilic, S. *et al.* Phase separation of 53BP1 determines liquid-like behavior of DNA repair compartments. *EMBO J* **38**, e101379 (2019).
17. essina, F. *et al.* Functional transcription promoters at DNA double-strand breaks mediate RNA-driven phase separation of damage-response factors. *Nat Cell Biol* **21**, 1286-1299 (2019).
18. Zhang, L. *et al.* 53BP1 regulates heterochromatin through liquid phase separation. *Nat Commun* **13**, 360 (2022).
19. Levone, B.R. *et al.* FUS-dependent liquid-liquid phase separation is important for DNA repair initiation. *J Cell Biol* **220**(2021).
20. Strona, G., Nappo, D., Boccacci, F., Fattorini, S. & San-Miguel-Ayanz, J. A fast and unbiased procedure to randomize ecological binary matrices with fixed row and column totals. *Nat Commun* **5**, 4114 (2014).
21. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele topologies. *Nat Genet* **50**, 1151-1160 (2018).

Chapter 7

Chromatin profiling with single-cell sequencing technologies

Kim L. de Luca[1,2] and Jop Kind[1,2,3]

[1] Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW) and University Medical Center Utrecht; Utrecht, The Netherlands

[2] Oncode Institute; The Netherlands

[3] Department of Molecular Biology, Faculty of Science, Radboud Institute for Molecular Life Sciences, Radboud University; Nijmegen, The Netherlands

Review article in preparation[i].

[i] This chapter serves as the foundation for a forthcoming review article. Certain sections from the Discussion (Chapter 8) will be incorporated and expanded for publication. To avoid redundancy, this thesis does not include those overlapping elements.

Abstract

Over the past decade, vast progress in single-cell sequencing has revolutionized our understanding of complex biological processes. Here, we provide a cutting-edge overview of recent progress in the development of sequencing technologies to study chromatin and the epigenome at single-cell resolution. Further, we briefly highlight current approaches using long-read single-molecule sequencing. We discuss advantages and limitations, and consider how these may influence experimental design, altogether aiming to facilitate and expedite the implementation of these new methods for the advancement of biological discoveries.

Introduction

Biological processes within the nucleus are often dynamic, creating a heterogeneity of states across cells. The same genetic information in each cell undergoes complex regulation to give rise to many different cell states and types, as reflected in their transcriptional profiles. Gene expression is thus controlled by an interplay between many factors including chromatin packaging, epigenetic modifications, interacting proteins, and 3D genome configuration. Collectively, these regulatory events ultimately determine the genomic output of a cell.

High-throughput sequencing forms the basis of several powerful methods to probe genome organization and regulation. Over the past decade, "omics" research has witnessed remarkable technological advancements, particularly in the realm of single-cell (sc) approaches. Since its introduction in 2009[1], scRNA-seq has become widely implemented, providing unparalleled insights into cell type diversity and tissue compositions. The single-cell omics field now stands at the forefront of DNA-centric methods aimed at measuring chromatin states and dynamics that precede or may be modulated by transcriptional output. For instance, assay for transposase-accessible chromatin (ATAC)-seq[2,3] has offered detailed insights into regulatory chromatin landscapes associated with variable transcriptional states.

However, thorough understanding of the establishment, maintenance, and changes in chromatin states requires more direct measurement of regulatory factors. To that end, methods are emerging to study epigenetic modifications, protein-DNA interactions, and genome organization at the single-cell level. Moreover, improvements in long-read sequencing allow for measurement of regulatory chromatin events on individual DNA molecules. In this review, we outline recent methodological developments with a focus on chromatin, including single-cell sequencing-based techniques to assay spatial genome organization, epigenetic histone modifications, and other chromatin-binding proteins. Where relevant, we include our perspective and future outlook.

Single-cell (epi)genomics

To start grasping the regulatory logic that governs the processes that altogether control gene expression programs in complex biological settings, many technologies have been developed to measure chromatin features by high-throughput sequencing. We distinguish between chromatin state (profiling DNA accessibility, epigenetic modifications, or DNA-binding proteins) and chromatin conformation (mapping genomic proximities). Where available, we present multi-omic assays that combine chromatin conformation or accessibility with another modality.

Isolation and barcoding of single cells

When embarking on single-cell genomics experiments, it is important to consider assay throughput and cell confinement. In single-cell genomics, molecules derived from an individual cell require unique labeling and amplification in an isolated reaction chamber. For most methods we discuss, three isolation configurations exist: a (nano)well in a plate (plate-based), a droplet (microfluidics), or the intact cell or nucleus itself (combinatorial indexing). They differ fundamentally, as reviewed extensively elsewhere[4-6] and we recommend picking the approach that best matches the research question and desired throughput.

In plate-based approaches, cells are processed individually in wells. Although this limits throughput to a few thousand cells per week and results in higher reagent costs, plate-based approaches are preferred when using fluorescence-activated cell sorting (FACS) to enrich for desired cell types, for rare cells, and after manual dissection or laser capture. To reduce costs as well as technical variation, implementing dispensing robotics is advisable when frequent usage is anticipated. In microfluidic encapsulation, barcoding and enzymatic processing of cells occurs in small droplets, allowing higher throughput of up to tens of thousands of cells per week. An attractive aspect of droplet-based platforms is their commercial availability, facilitating technological access, basic computational pipelines, and quality control. Lastly, combinatorial indexing relies on split-pool barcoding of intact cells or nuclei to create unique sample barcode combinations, followed by processing according to assay choice. Depending on the number of barcoding rounds (typically three or four), the throughput ranges from tens of thousands to millions of cells per experiment. Combinatorial indexing requires optimization but can be performed in any lab with standard molecular biology equipment. Once established, it is highly cost-effective and customizable, which is why we consider it the preferred choice, especially when frequent use is envisioned.

Chromatin accessibility

Chromatin accessibility is often considered a proxy for transcriptional regulation. Measuring accessibility is suitable, for example, when a priori knowledge of specific epigenetic regulation is (largely) absent. Various strategies map the DNA packaging based on the principle of nucleosome occupancy, with scATAC-seq representing the most common technology at present (Fig. 7.1). This approach leverages the hyperactive Tn5 transposase, which inserts pre-loaded NGS adapters into nucleosome-free, open chromatin sites. scATAC-seq has been adapted to achieve high throughput using combinatorial indexing[7,8] nanowells[9], and droplet-based microfluidics[10,11]. Availability of an extensive computational toolbox greatly facilitates data analysis (see the Single-Cell Best Practices handbook).

ATAC-seq can be combined with other modalities, most notably the transcriptome and DNA methylation (extensively reviewed in[6]), whole-genome sequencing[12], and genotyping[13].

Histone post-translational modifications

To study chromatin states with more detail in individual cells, several approaches are available to map the genomic locations of HPTMs (Fig. 7.1). These chromatin marks are abundant, relatively stable, and their distribution ranges from narrow peaks (<1 kb) to broad domains (>1 Mb). The availability of high-quality antibodies is key to the successful profiling of HPTMs, and most techniques thus use primary antibodies to confer specificity to chromatin epitopes, as in the gold standard in epigenetic profiling: population-based chromatin immunoprecipitation followed by sequencing (ChIP-seq). Although droplet-based scChIP has been developed[14,15], detailed analysis is currently limited to clusters of cells, due to signal sparsity.

For low-input HPTM profiling, immunocleavage (IC) is especially useful, since it does not rely on pulldown of material, thereby minimizing sample loss. Several single-cell IC variations exist that fuse Protein A (pA) to MNase (in CUT&RUN[16] and ChIC[17,18]) or Tn5 (in CUT&Tag[19]) (Fig. 7.1). CUT&Tag can be implemented with combinatorial indexing (sciCUT&Tag[20]) or with nanowell (ICELL8 [21]) and droplet-based (10x Chromium[22]) platforms. To improve the sensitivity of CUT&Tag, its recent adaptations fused a secondary nanobody directly to Tn5 (nano-CT[23] and NTT-seq[24]), which also facilitates multiplexing of target features.

In case the research question involves rare cell types or would benefit from retaining cellular metadata per cell (e.g., cell cycle or protein expression), sortChIC[18] is recommended. This plate-based method relies on FACS and processing in 384-well plates, enabling cell selection and index sorting. A second low-input method to map epigenetic modifications is EpiDamID[25]. DamID involves *in vivo* expression of a fusion protein, consisting of a protein of interest (POI) tethered to the bacterial DNA adenine methyltransferase (Dam)[26,27]. To profile HPTMs, Dam is fused to either chromatin reader domains or single-chain antibodies with affinities for different chromatin modifications[25,28]. Because EpiDamID involves the *in vivo* "stamping" of DNA, sample handling is minimal, and individual cells can be directly sorted or manually selected. Thus, unlike sortChIC that requires an input of ~500.000 cells, EpiDamID profiles can be obtained starting with a single cell[28,29]. A downside is that EpiDamID depends on genetic engineering, and its resolution is limited by the distribution of the GATC sequence recognized by Dam.

Chromatin proteins and transcription factors

In comparison to HPTMs, mapping of TFs with ChIP-seq or IC methods is typically difficult due to a lack of high-quality antibodies, a low POI count at genomic targets, and the short residence time of TFs on chromatin. Additionally, accessible chromatin presents a general challenge, because the enzymes used to record binding (e.g., MNase, Tn5, Dam) exert intrinsic activity there, resulting in non-specific targeting[16]. To assess specificity, we advise stratifying POI binding by consensus sequence motifs and validated enrichment calls. To mitigate accessibility-related artifacts in CUT&Tag, it is further recommended to use stringent binding and washing conditions that reduce pA-Tn5 affinity for exposed DNA[19]. However, this may also minimize binding of other proteins. Hence, for mapping TFs or proteins with lower DNA-binding affinity, CUT&RUN (employing MNase) is expected to generate better results compared to CUT&Tag, as the target protein is less likely to be inadvertently removed during washing.

In Table 7.1 (at end of document), we provide an overview of proteins and HPTMs that have been mapped in single cells (current at time of writing). Before embarking on mapping a new POI in single cells, it is good practice to perform antibody titrations in iteratively small populations (e.g., 1000, 100, 10 cells) to determine sensitivity and specificity. We also refer the reader to the following resources on antibody validation {EpiCypher; ImmGen; Harvard CompBio}, and extensive experimental comparisons and considerations for targeted IC[19,30].

An alternative method to map TF binding sites (TFBSs) uses a TF-transposase fusion to trace genome-wide POI occupancy as "calling cards"[31]. Notably, its single-cell implementation recovers TFBSs from mRNA of the same material, enabling joint profiling of TF occupancy and transcriptome (scCC[32]). This could be a powerful alternative to antibody-based multi-omics; however, the method requires genetic engineering and the insertional mutagenesis may affect cellular physiology.

Finally, single-cell profiling of heterochromatic regions is more feasible, largely owing to broad domain sizes and antagonism with DNA accessibility. SortChIC and scDamID have reported profiling of proteins associated with constitutive as well as facultative heterochromatin, such as nuclear lamina components and Polycomb group proteins[18,25,28,33,34]. Simultaneous profiling of heterochromatin and DNA accessibility is possible with genome and epigenome by transposases (GET)-seq, using an H3K9me3-specific chromatin reader fused to Tn5[35]. With its single-cell implementation on 10x Chromium, the dual readout was leveraged to infer genome-wide CNV patterns and chromatin velocity (*i.e.*, trajectories of epigenetic modifications).

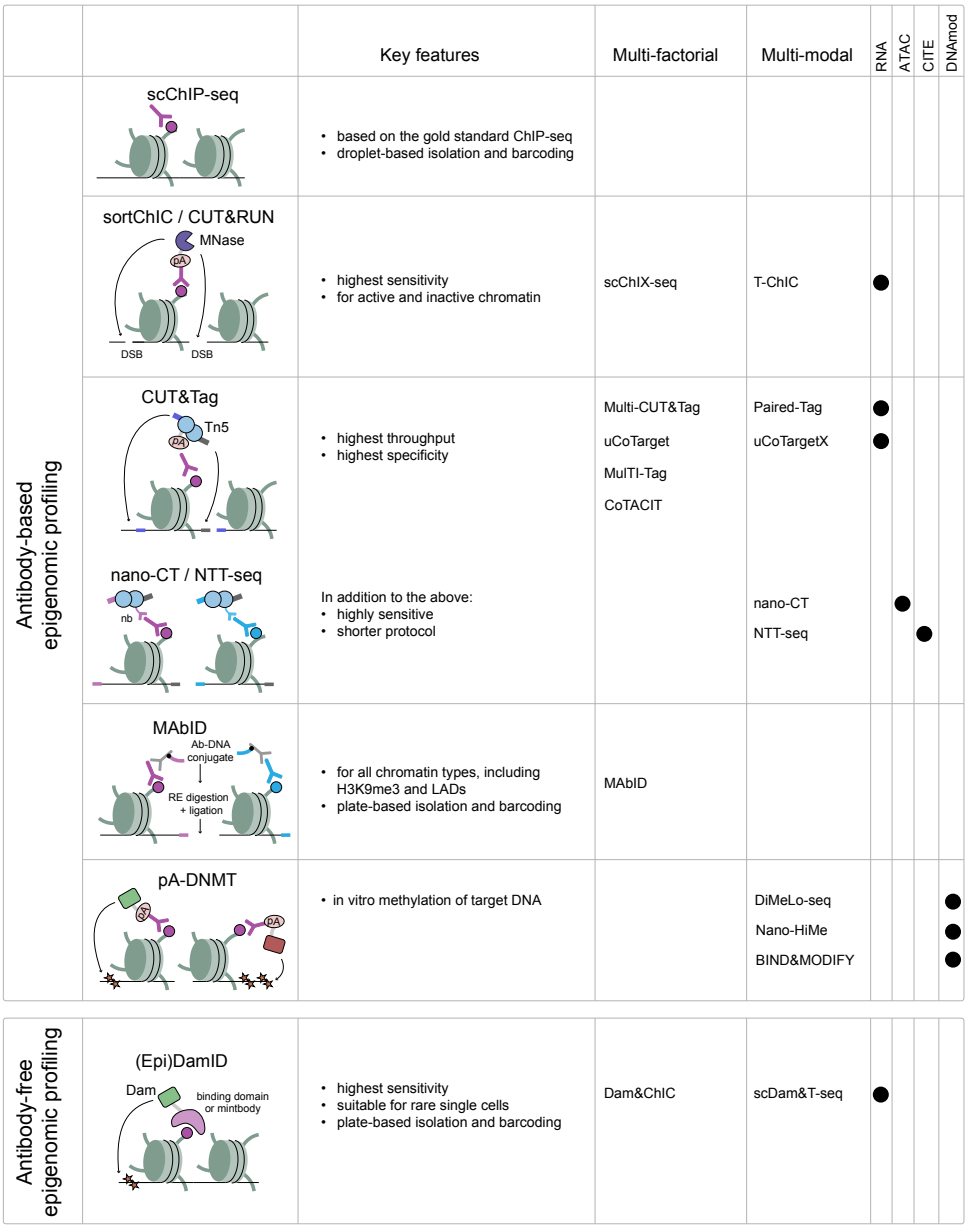


Figure 7.1. Schematic comparison of (sc) epigenomics methods.

Overview of different approaches for profiling of chromatin types in single cells. Note that we distinguish between multi-factorial methods (multiple readouts, same modality) and multi-modal methods (readouts of different modalities) and we indicate which methods adhere to these criteria. We categorize the different types of methods that exist and distinguish between antibody-based profiling and antibody-free alternatives.

Multi-modal profiling of HPTMs and chromatin proteins

Several of the above single-cell methods have been multiplexed to simultaneously map multiple HPTMs and chromatin proteins, with or without additional modalities. While the terms have been used interchangeably, here we distinguish between multi-factorial methods (multiple readouts of the same modality, such as different HPTMs) and multi-modal methods (readouts of different modalities, such as HPTMs plus transcriptome). The integration of spatial information as a separate modality in sequencing-based assays has been reviewed in[5,6].

Multi-factorial profiling: various HPTMs or chromatin proteins

Simultaneous profiling of multiple chromatin marks presents an exciting opportunity to investigate the coordination of epigenetic regulation in single cells. Most of these multi-factorial methods are based on tagmentation (Fig. 7.1), including the pioneering methods scMulti-CUT&Tag[36] and MulTI-tag[37], which yielded comparatively sparser data than their unifactorial equivalents. Sensitivity and specificity were improved in nano-CT[23] and NTT-seq[24] which are both implemented on 10x Chromium. Their multiplexing requires the availability of antibodies of different origins to distinguish between the different chromatin marks. More recently, uCoTarget[38] reported five HPTM profiles in the same cell, using pA-Tn5 fusions and barcoded adaptors for combinatorial indexing-based multiplexing. Combined target chromatin indexing and tagmentation (Co)TACIT) reported three HPTMs per cell with a similar barcoding strategy, but applicable to low-input samples[39]. Thus far, the only reported multi-factorial method not based on tagmentation is MAbID, a plate-based approach that employs antibody-adapter conjugates for epitope-specific ligation of barcodes into the genome[40]. MAbID reports six simultaneous measurements with similar sensitivity as uCoTarget, notably including heterochromatin-associated H3K9me3 and lamina-associated domains (LADs).

Multi-modal profiling: HPTMs or chromatin proteins with another modality

Using 10x Chromium, nano-CT can be combined with an ATAC-seq readout, while NTT-seq can include cell surface protein profiling with CITE-seq[41]. The integration of transcriptome with two antibody readouts makes uCoTargetX[38] the most versatile multi-modal method currently available. Several other technologies combine single HPTM profiling and transcriptomics. Among others, Paired-Tag (sci-based[42] or 10x Chromium implementation[43]), SET-seq[44], and CoTECH[45] achieve similar quality as their unimodal components. An alternative, plate-based strategy combines scDamID and transcriptome in scDam&T-seq[34] for mapping chromatin proteins or HPTMs[25,28]. Lastly, sortChIC and VASA-seq[46] are combined in T-ChIC, to detect full-length transcripts and HPTMs[47].

Multi-omic measurements of transcription and chromatin-associated proteins enable stratification on one modality of interest, for instance cell type- or gene-specific epigenetic maps. The transcriptional component may also elucidate order of events. However, we note that general data sparsity of single-cell genomics methods often precludes inference of "true" single-cell relationships between chromatin states and transcriptional changes. This is especially relevant when combining (many) epigenetic readouts that may not be equally distributed across the genome, or where the more prominent mark dominates the signal. Therefore, continued development of sensitive and versatile methods is essential, in addition to the further advancement of powerful data imputation methods[48]. Although analysis of scHPTM profiles has not been standardized yet, some popular approaches have been compared[49]. See Table 7.1 at end of document for a list of proteins targeted with multi-modal and multi-factorial sc genomics methods.

Chromatin conformation

The last regulatory layer we discuss in the context of single-cell genomics is chromatin conformation. From larger to smaller scale, DNA is folded into chromosome territories, compartments, topologically associating domains (TADs), and loops. While TADs can facilitate functional interactions such as enhancer-promoter looping, the actual contribution of loop extrusion to transcriptional control is nuanced and context-dependent (reviewed in [50]). As a result, the last decade has seen considerable advances to probe chromatin conformation in single cells. We focus on single-cell adaptations of Hi-C, GAM, and SPRITE. An in-depth analysis of methods to probe chromatin contacts genome-wide has been performed in [51], showing that they all faithfully capture 3D organization in bulk, with large variations among individual cells.

Single-cell Hi-C and extensions

Genome folding is most commonly studied with so-called C-based techniques, originating from chromosome conformation capture (3C)[52]. Current implementations use *in situ* chromatin crosslinking, digestion and ligation to detect genomic regions that are close in 3D space. While derivatives exist, only the genome-wide (all-vs-all) approach Hi-C[53,54] has been adapted to single cells (scHi-C[55]). Because scHi-C typically yields a very small fraction of possible contacts per cell, the data sparsity poses a challenge for downstream computational analyses (reviewed in [56]). Obtaining high numbers of scHi-C profiles is therefore desired to more confidently extract general principles of 3D genome organization, for example by applying combinatorial indexing as in sciHi-C[57], or with Droplet Hi-C[58]. By implementing a high-coverage whole genome amplification strategy, Dip-C achieved >1 million contacts per cell and a resolution of ~20 kb[59].

Initial throughput of ~20 cells has been scaled up to ~400 cells[60], which remains low in comparison to combinatorial indexing or droplet-based barcoding. Recently, incorporation of MNase has further improved resolution to ~5 kb[61]. Other recent efforts focused on joint profiling of chromatin conformation with other modalities, such as transcription (in HiRES[62], scCARE-seq[63], LiMCA[60], and GAGE-seq[64]), and accessibility and methylation (in ChAIR[65], sn-m3C-seq[66] and scMethyl-HiC[67]). Higher-order, multi-contact proximities can be directly measured with scNanoHi-C, which implements long-read (third-generation) sequencing to detect contacts in large concatemers[68]. Similar approaches that retain single-molecule information from population-based sequencing are discussed below under *Single-molecule chromatin genomics.*

GAM and multiplex-GAM

While not strictly a genome-wide single-cell method, GAM involves DNA sequencing of hundreds of cryosections in random orientations, from which a co-segregation map can be constructed, representing the frequency with which genomic loci appear together across many slices[69]. Like scHi-C, GAM measures all-vs-all contacts, but independently of proximity ligation. Complex (multi-way) interactions are captured more readily by GAM compared to scHi-C[70]. Other advantages and limitations of GAM are mostly inherent to the approach to capture single nuclear profiles, which keeps cellular structure intact and retains spatial information within a tissue. Extensions of GAM include immunoGAM, in which immunostaining of cryosections enables selection of (rare) cells of interes[71], and the faster, more affordable, automation-compatible multiplex-GAM[70]. Lastly, Sequential-GAM[72] was developed to infer genome topology and radial position from the same cell, by sequencing sequential sections throughout the nucleus.

SPRITE and extensions

Single-cell SPRITE[73] relies on uniquely split-pool barcoding and sequencing DNA fragments within physically associated chromatin complexes. The method records multi-way genomic contacts, also over long distances and in trans, and yields ~10 times more interactions than scHi-C, at lower cost. Bulk SPRITE has been combined with measurement of RNA associations (RD-SPRITE[74]). Due to the modularity and relative user-friendliness of scSPRITE, we anticipate that other readouts will soon be integrated. Nevertheless, population-based SPRITE generates a single-molecule readout allowing for exploration of locus-specific variability in bulk SPRITE data (also see *Higher-order chromatin conformation*). A similar approach of barcoding DNA- and/or RNA-associated complexes has been implemented in MUSIC, using the 10x Chromium platform[75].

While scHi-C is more readily implemented than (multiplex-)GAM and scSPRITE, the latter two provide more detailed views of chromatin conformation. Overall, these three approaches give insights into the heterogeneous nature of genome topologies, with main differences in throughput, sparsity and related resolution, and capture of pairwise versus higher-order contacts. Computational methods for scHi-C are most established, but processing and analysis pipelines are available for both GAM and SPRITE.

Single-molecule and long-read genomics

Chromatin structure and function can also be analyzed on individual DNA molecules extracted from a population of cells. This section briefly highlights the characteristics of single-molecule chromatin genomics, focusing on long-read sequencing methods to study DNA-protein contacts and higher-order chromatin organization. For a more extensive review of long-read epigenomic profiling, see [39].

Although single-molecule assays are performed in bulk (thus discarding information about the cell of origin), information from many individual DNA molecules provides a snapshot of regulatory heterogeneity. Because all DNA molecules are sequenced regardless of status (e.g., bound or unbound by a TF), it is possible to quantify frequencies in a population[76]. Depending on sequence read length, these methods can record co-occurring regulatory events on the same DNA molecule, revealing functional dependencies such as cooperativity or antagonism. While single-cell methods can measure co-occurring features of interest genome-wide, this is typically challenging to extract from due to signal drop-out and sparsity. In contrast, single-molecule methods provide higher resolution within a given region, but genomic distance on the same molecule is inherently limited by read length.

(Multi-modal) profiling of chromatin proteins and HPTMs

Based on the general principle of proximal methylation, a handful of methods has recently been developed to map DNA-protein contacts with long-read sequencing. Akin to pA-DamID, this is achieved by tethering a methyltransferase to pA, conferring antibody-based specificity, followed by modification-sensitive long-read sequencing to identify POI-DNA interactions. Directed methylation with long-read sequencing (DiMeLo-seq)[77], nanopore-sequencing-based histone-modification and methylome (nanoHiMe-seq)[78], and BIND&MODIFY[79] all employ this approach. Instead of the GATC-specific Dam, these methods use the non-specific adenine methyltransferases Hia5 or EcoGII. DamID has been implemented with LRS in Nanopore-DamID. However, as in regular DamID, this uses methylation-specific endonucleases to cleave the modified regions of interest, limiting the read length at present. Retention of uncleaved fragments for LRS is possible with direct modification calling.

DiMeLo-seq was used to efficiently detect the presence of heterochromatin-associated HPTM H3K9me3, Lamin B1, centromere-specific CENP-A, and the insulator protein CTCF. NanoHiMe-seq targeted H3K4me3- and H3K27me3-marked regions, and BIND&MODIFY targeted H3K9me3, H3K27me3, and CTCF. Importantly, LRS allows for mapping to low-complexity regions of the genome, opening up the possibility of studying repetitive regions, such as telomeres or centromeres, which are typically inaccessible for interpretation with short-read technologies. A current downside of such approaches is that interpretation of the single-molecule data requires a priori evaluation of the efficiency of each new antibody—and of the selected methyltransferase—in saturating its target epitopes. Lastly, native single-molecule sequencing is amplification-free, meaning that all reads shall be distributed (approximately) evenly across the genome, and coverage may thus be limited. For instance, a standard ONT MinION or PacBio SMRT might generate ~20 Gb of LRS output—the equivalent of ~7x coverage of the human genome.

Note that above methods can generate dual readouts of endogenous and exogenous DNA methylation measured simultaneously. This is facilitated by the activity of the methyltransferase in the experimental setting (resulting in ^{m}A or ^{m}GpC), which is different than endogenously occurring methylation in the genome (^{m}CpG). Upon sequencing, signal deconvolution of the two types of methylated bases occurs computationally. This inherent multi-modal capacity should greatly facilitate the next frontier of studies to disentangle genome regulation at several layers.

Higher-order chromatin conformation

Single-molecule approaches can also be harnessed to decipher 3D genome conformation. This involves capturing proximal contacts formed by DNA in its native folded state. To do so, the basic notion of C-based techniques is combined with high-throughput SRS or LRS. In brief, C-based techniques generate *in situ* ligation junctions between regions residing in spatial proximity in the nucleus, serving as a proxy for 3D chromatin configuration. While typical C-based analyses (with some exceptions) focus on pairwise 3D contacts only, the intermediate products generated during classical 3C protocols consist of long concatemers of proximity ligated DNA fragments. As these concatemers hold valuable insights into higher-order, multi-way, and potentially synergistic, chromatin conformation, they can be extracted and treated as single DNA molecules and read out using long(er)-read sequencing. Note that ligation-free methods that allow measuring multi-way chromatin contacts (SPRITE and GAM) are not concatemer-based and will therefore not benefit much from long-read sequencing.

The initial 3C step in most multi-way methods is performed in bulk. However, as the concatemers are ligated *in situ*, ligation junctions are delimited by the boundaries of each single nucleus. Because no inter-nuclear or inter-allelic ligation events are expected, this allows for derivation of single-allele multi-way topologies from population-based assays. Examples of this are the methods C-walks[80] and Tri-C[81,82], which employ longer-read Illumina sequencing, and PacBio and regular Illumina sequencing, respectively. In contrast, multi-contact (MC)-4C[83] and MC-3C[84] involve ONT and PacBio sequencing, respectively. C-walks and MC-4C both employ targeted analyses by either PCR amplifying concatemers containing genomic viewpoints of interest or incorporating a Cas9 digestion to direct all measurement towards one locus of interest, respectively. In contrast, MC-3C was performed without any target enrichment and can be considered genome-wide. Similarly, Pore-C[85] is an 'all-versus-all' method but relies on ONT, and achieved much higher sequence depth in comparison to MC-3C. This not only allowed for uncovering of multi-way interactions, but also investigation of CpG methylation on haplotyped concatemers, a possibility that is now further advanced in HiPore-C[86], enabling improved throughput and reduced costs in comparison to its predecessor.

Despite the single-allele resolution of these setups, their population-based nature prevents exploration of cell-to-cell heterogeneity in multi-way topologies. By incorporating a FACS step to sort single cells after proximity ligation, the recent scNano-HiC[68] addressed this limitation. scNano-HiC relies on low-density Tn5 tagmentation to sparsely insert Tn5 PCR barcodes, prior to amplification and ONT sequencing. scNanoHi-C indeed profiled 3D chromatin structures and distinguished subpopulations with distinct configurations, in addition to its ability to detect genomic variants and scaffold assembly of sc genomes.

Conclusions and future perspectives

Future directions inevitably include optimizing detection sensitivity and specificity, measuring more modalities at once, and developing hybrid technologies combining single-cell genomics, single-molecule and long-read sequencing, and advanced microscopy into integrated frameworks. Of particular interest are technologies that bridge the strengths of the different fields discussed. To link cellular traits to molecular states, we envision combining optics and omics through live-cell imaging and single-cell (long-read) sequencing to quantifiably relate cellular traits and molecular features of interest.

Acknowledgement

We thank Britta Bouwman and Linxuan Zhao for their contributions to this work.

Method	Target
CoTACIT	H3K27ac
CoTACIT	H3K27me3
CoTACIT	H3K9me3
CoTECH	H3K27me3
CoTECH	H3K4me3
CUT&Tag2for1	H3K27me3
CUT&Tag2for1	RNAPIISer5
DamID	53BP1
DamID	H3K27me3
DamID	H3K9ac
DamID	H3K9me3
DamID	H4K20me1
DamID	Lamin B1
DamID	RAD51
MAbID	H3K27ac
MAbID	H3K27me3
MAbID	H3K36me3
MAbID	H3K4me1
MAbID	H3K9me3
MAbID	Lamin B1
MulTI-Tag	H3K27me3
MulTI-Tag	H3K36me3
MulTI-Tag	H3K4me1
MulTI-Tag	H3K4me2
nano-CT	H3K27ac
nano-CT	H3K27me3
NTT-seq	H3K27ac
NTT-seq	H3K27me3
NTT-seq	RNAPIISer2/5
Paired-Tag	H3K27ac
Paired-Tag	H3K27me3
Paired-Tag	H3K4me1
Paired-Tag	H3K4me3
Paired-Tag	H3K9me3
scChIX-seq	H3K27me3
scChIX-seq	H3K36me3
scChIX-seq	H3K4me1
scChIX-seq	H3K9me3
scGET-seq	H3K9me3
scMulti-CUT&Tag	H3K27ac
scMulti-CUT&Tag	H3K27me3
scSET-seq	H3K27me3
scSET-seq	H3K4me3
sortChIC	53BP1
sortChIC	BRCA1
sortChIC	CTCF
sortChIC	H2AK119ub
sortChIC	H3K27me3
sortChIC	H3K36me3
sortChIC	H3K4me1
sortChIC	H3K4me3
sortChIC	H3K9me3
sortChIC	Lamin B1
sortChIC	MDC1
sortChIC	γH2AX
uCoTarget	H3K27ac
uCoTarget	H3K27me3
uCoTarget	H3K36me3
uCoTarget	H3K4me1
uCoTarget	H3K4me3
uCoTarget	RUNX1
uCoTargetX	H3K27ac
uCoTargetX	H3K27me3

Table 7.1. DNA-binding proteins and HPTMs mapped in single cells with multi-modal or multi-factorial methods.

References

1. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods* **6**, 377-82 (2009).
2. Buenrostro, J.D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-90 (2015).
3. Cusanovich, D.A. *et al.* Multiplex single cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-4 (2015).
4. Adey, A.C. Tagmentation-based single-cell genomics. *Genome Res* **31**, 1693-1705 (2021).
5. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. *Nat Rev Mol Cell Biol* **24**, 695-713 (2023).
6. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. *Nat Rev Genet* **24**, 494-515 (2023).
7. Domcke, S. *et al.* A human cell atlas of fetal chromatin accessibility. *Science* **370**(2020).
8. Cusanovich, D.A. *et al.* A Single-Cell Atlas of In Vivo Mammalian Chromatin Accessibility. *Cell* **174**, 1309-1324 e18 (2018).
9. Mezger, A. *et al.* High-throughput chromatin accessibility profiling at single-cell resolution. *Nat Commun* **9**, 3647 (2018).
10. Satpathy, A.T. *et al.* Massively parallel single-cell chromatin landscapes of human immune cell development and intratumoral T cell exhaustion. *Nat Biotechnol* **37**, 925-936 (2019).
11. Zhang, X., Marand, A.P., Yan, H. & Schmitz, R.J. scifi-ATAC-seq: massive-scale single-cell chromatin accessibility sequencing using combinatorial fluidic indexing. *Genome Biol* **25**, 90 (2024).
12. Queitsch, K. *et al.* Accessible high-throughput single-cell whole-genome sequencing with paired chromatin accessibility. *Cell Rep Methods* **3**, 100625 (2023).
13. Izzo, F. *et al.* Mapping genotypes to chromatin accessibility profiles in single cells. *Nature* **629**, 1149-1157 (2024).
14. Rotem, A. *et al.* Single-cell ChIP-seq reveals cell subpopulations defined by chromatin state. *Nat Biotechnol* **33**, 1165-72 (2015).
15. Grosselin, K. *et al.* High-throughput single-cell ChIP-seq identifies heterogeneity of chromatin states in breast cancer. *Nat Genet* **51**, 1060-1066 (2019).
16. Hainer, S.J., Bošković, A., McCannell, K.N., Rando, O.J. & Fazzio, T.G. Profiling of pluripotency factors in single cells and early embryos. *Cell* **177**, 1319-1329.e11 (2019).
17. Ku, W.L. *et al.* Single-cell chromatin immunocleavage sequencing (scChIC-seq) to profile histone modification. *Nat Methods* **16**, 323-325 (2019).
18. Zeller, P. *et al.* Single-cell sortChIC identifies hierarchical chromatin dynamics during hematopoiesis. *Nat Genet* **55**, 333-345 (2023).
19. Kaya-Okur, H.S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun* **10**, 1930 (2019).
20. Janssens, D.H. *et al.* Scalable single-cell profiling of chromatin modifications with sciCUT&Tag. *Nat Protoc* **19**, 83-112 (2024).
21. Wu, S.J. *et al.* Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nat Biotechnol* **39**, 819-824 (2021).
22. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nat Biotechnol* **39**, 825-835 (2021).
23. Bartosovic, M. & Castelo-Branco, G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. *Nat Biotechnol* **41**, 794-805 (2023).
24. Stuart, T. *et al.* Nanobody-tethered transposition enables multifactorial chromatin profiling at single-cell resolution. *Nat Biotechnol* **41**, 806-812 (2023).
25. Rang, F.J. *et al.* Single-cell profiling of transcriptome and histone modifications with EpiDamID. *Mol Cell* **82**, 1956-1970.e14 (2022).
26. van Steensel, B. & Henikoff, S. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat Biotechnol* **18**, 424-8 (2000).
27. van Steensel, B., Delrow, J. & Henikoff, S. Chromatin profiling using targeted DNA adenine methyltransferase. *Nat Genet* **27**, 304-8 (2001).
28. Guerreiro, I. *et al.* Antagonism between H3K27me3 and genome-lamina association drives atypical spatial genome organization in the totipotent embryo. *Nat Genet* **56**, 2228-2237 (2024).
29. Borsos, M. *et al.* Genome-lamina interactions are established de novo in the early mouse embryo. *Nature* **569**, 729-733

(2019).

30. Nordin, A., Pagella, P., Zambanini, G. & Cantù, C. Exhaustive identification of genome-wide binding events of transcriptional regulators. *Nucleic Acids Res* **52**, e40 (2024).
31. Wang, H., Mayhew, D., Chen, X., Johnston, M. & Mitra, R.D. "Calling cards" for DNA-binding proteins in mammalian cells. *Genetics* **190**, 941-9 (2012).
32. Moudgil, A. *et al.* Self-Reporting Transposons Enable Simultaneous Readout of Gene Expression and Transcription Factor Binding in Single Cells. *Cell* **182**, 992-1008 e21 (2020).
33. Kind, J. *et al.* Genome-wide maps of nuclear lamina interactions in single human cells. *Cell* **163**, 134-47 (2015).
34. 34. Rooijers, K. *et al.* Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells. *Nat Biotechnol* **37**, 766-772 (2019).
35. Tedesco, M. *et al.* Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. *Nat Biotechnol* **40**, 235-244 (2022).
36. Gopalan, S., Wang, Y., Harper, N.W., Garber, M. & Fazzio, T.G. Simultaneous profiling of multiple chromatin proteins in the same cells. *Mol Cell* **81**, 4736-4746.e5 (2021).
37. Meers, M.P., Llagas, G., Janssens, D.H., Codomo, C.A. & Henikoff, S. Multifactorial profiling of epigenetic landscapes at single-cell resolution using MulTI-Tag. *Nat Biotechnol* **41**, 708-716 (2023).
38. Xiong, H., Wang, Q., Li, C.C. & He, A. Single-cell joint profiling of multiple epigenetic proteins and gene transcription. *Sci Adv* **10**, eadi3664 (2024).
39. Liu, M. *et al.* Genome-coverage single-cell histone modifications for embryo lineage tracing. *Nature* (2025).
40. Lochs, S.J.A. *et al.* Combinatorial single-cell profiling of major chromatin types with MAbID. *Nat Methods* **21**, 72-82 (2024).
41. Stoeckius, M. *et al.* Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods* **14**, 865-868 (2017).
42. Zhu, C. *et al.* Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nat Methods* **18**, 283-292 (2021).
43. Xie, Y. *et al.* Droplet-based single-cell joint profiling of histone modifications and transcriptomes. *Nat Struct Mol Biol* **30**, 1428-1433 (2023).
44. Sun, Z. *et al.* Joint single-cell multiomic analysis in Wnt3a induced asymmetric stem cell division. *Nat Commun* **12**, 5941 (2021).
45. Xiong, H., Luo, Y., Wang, Q., Yu, X. & He, A. Single-cell joint detection of chromatin occupancy and transcriptome enables higher-dimensional epigenomic reconstructions. *Nat Methods* **18**, 652-660 (2021).
46. Salmen, F. *et al.* High-throughput total RNA sequencing in single cells using VASA-seq. *Nat Biotechnol* **40**, 1780-1793 (2022).
47. Zeller, P. *et al.* T-ChIC: multi-omic detection of histone modifications and full-length transcriptomes in the same single cell. *bioRxiv*, 2024.05.09.593364 (2024).
48. Heumos, L. *et al.* Best practices for single-cell analysis across modalities. *Nat Rev Genet* **24**, 550-572 (2023).
49. Raimundo, F., Prompsy, P., Vert, J.P. & Vallot, C. A benchmark of computational pipelines for single-cell histone modification data. *Genome Biol* **24**, 143 (2023).
50. Karpinska, M.A. & Oudelaar, A.M. The role of loop extrusion in enhancer-mediated gene activation. *Curr Opin Genet Dev* **79**, 102022 (2023).
51. Fiorillo, L. *et al.* Comparison of the Hi-C, GAM and SPRITE methods using polymer models of chromatin. *Nat Methods* **18**, 482-490 (2021).
52. Dekker, J., Rippe, K., Dekker, M. & Kleckner, N. Capturing chromosome conformation. *Science* **295**, 1306-11 (2002).
53. Lieberman-Aiden, E. *et al.* Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**, 289-93 (2009).
54. Dixon, J.R. *et al.* Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* **485**, 376-80 (2012).
55. Nagano, T. *et al.* Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* **502**, 59-64 (2013).
56. Galitsyna, A.A. & Gelfand, M.S. Single-cell Hi-C data analysis: safety in numbers. *Brief Bioinform* **22**(2021).
57. Ramani, V. *et al.* Massively multiplex single-cell Hi-C. *Nat Methods* **14**, 263-266 (2017).
58. Chang, L. *et al.* Droplet Hi-C enables scalable, single-cell profiling of chromatin architecture in heterogeneous tissues. *Nat Biotechnol* (2024).
59. Tan, L., Xing, D., Chang, C.H., Li, H. & Xie, X.S. Three-dimensional genome structures of single diploid human cells. *Science* **361**,

924-928 (2018).

60. Wu, H. *et al.* Simultaneous single-cell three-dimensional genome and gene expression profiling uncovers dynamic enhancer connectivity underlying olfactory receptor choice. *Nat Methods* **21**, 974-982 (2024).
61. Wu, H., Zhang, J., Tan, L. & Xie, X.S. Extruding transcription elongation loops observed in high-resolution single-cell 3D genomes. *bioRxiv*, 2023.02.18.529096 (2023).
62. Liu, Z. *et al.* Linking genome structures to functions by simultaneous single-cell Hi-C and RNA-seq. *Science* **380**, 1070-1076 (2023).
63. Qu, J. *et al.* Simultaneous profiling of chromatin architecture and transcription in single cells. *Nat Struct Mol Biol* **30**, 1393-1402 (2023).
64. Zhou, T. *et al.* GAGE-seq concurrently profiles multiscale 3D genome organization and gene expression in single cells. *Nat Genet* **56**, 1701-1711 (2024).
65. Chai, H. *et al.* Tri-omic mapping revealed concerted dynamics of 3D epigenome and transcriptome in brain cells. *bioRxiv*, 2024.05.03.592322 (2024).
66. Lee, D.S. *et al.* Simultaneous profiling of 3D genome structure and DNA methylation in single human cells. *Nat Methods* **16**, 999-1006 (2019).
67. Li, G. *et al.* Joint profiling of DNA methylation and chromatin architecture in single cells. *Nat Methods* **16**, 991-993 (2019).
68. Li, W. *et al.* scNanoHi-C: a single-cell long-read concatemer sequencing method to reveal high-order chromatin structures within individual cells. *Nat Methods* **20**, 1493-1505 (2023).
69. Beagrie, R.A. *et al.* Complex multi-enhancer contacts captured by genome architecture mapping. *Nature* **543**, 519-524 (2017).
70. Beagrie, R.A. *et al.* Multiplex-GAM: genome-wide identification of chromatin contacts yields insights overlooked by Hi-C. *Nat Methods* **20**, 1037-1047 (2023).
71. Winick-Ng, W. *et al.* Cell-type specialization is encoded by specific chromatin topologies. *Nature* **599**, 684-691 (2021).
72. Li, Y., Wang, K., Shi, M., Zhang, M.Q. & Gao, J. Sequential-GAM constructs the single-cell geometric 3D genome structure. *bioRxiv*, 2024.03. 27.586762 (2024).
73. Arrastia, M.V. *et al.* Single-cell measurement of higher-order 3D genome organization with scSPRITE. *Nat Biotechnol* **40**, 64-73 (2022).
74. Quinodoz, S.A. *et al.* RNA promotes the formation of spatial compartments in the nucleus. *Cell* **184**, 5775-5790 e30 (2021).
75. Wen, X. *et al.* Single-cell multiplex chromatin and RNA interactions in ageing human brain. *Nature* **628**, 648-656 (2024).
76. Krebs, A.R. Studying transcription factor function in the genome at molecular resolution. *Trends Genet* **37**, 798-806 (2021).
77. Altemose, N. *et al.* DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome wide. *Nat Methods* **19**, 711-723 (2022).
78. Yue, X. *et al.* Simultaneous profiling of histone modifications and DNA methylation via nanopore sequencing. *Nat Commun* **13**, 7939 (2022).
79. Weng, Z. *et al.* BIND&MODIFY: a long-range method for single-molecule mapping of chromatin modifications in eukaryotes. *Genome Biol* **24**, 61 (2023).
80. Olivares-Chauvet, P. *et al.* Capturing pairwise and multi-way chromosomal conformations using chromosomal walks. *Nature* **540**, 296-300 (2016).
81. Oudelaar, A.M. *et al.* Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. *Nat Genet* **50**, 1744-1751 (2018).
82. Oudelaar, A.M., Downes, D.J. & Hughes, J.R. Assessment of Multiway Interactions with Tri-C. *Methods Mol Biol* **2532**, 95-112 (2022).
83. Allahyar, A. *et al.* Enhancer hubs and loop collisions identified from single-allele topologies. *Nat Genet* **50**, 1151-1160 (2018).
84. Tavares-Cadete, F., Norouzi, D., Dekker, B., Liu, Y. & Dekker, J. Multi-contact 3C reveals that the human genome during interphase is largely not entangled. *Nat Struct Mol Biol* **27**, 1105-1114 (2020).
85. Deshpande, A.S. *et al.* Identifying synergistic high-order 3D chromatin conformations from genome-scale nanopore concatemer sequencing. *Nat Biotechnol* **40**, 1488-1499 (2022).
86. Zhong, J.Y. *et al.* High-throughput Pore-C reveals the single-allele topology and cell type-specificity of 3D genome folding. *Nat Commun* **14**, 1250 (2023).

Chapter 8

Discussion

Discussion

8.1 Technical features of single-cell omics
- 8.1.1 Resolution
- 8.1.2 DNA accessibility bias
- 8.1.3 Temporal dynamics
- 8.1.4 Contactomics

8.2 Polycomb in cellular differentiation and pathology
- 8.2.1 Hierarchical regulatory networks in lineage specification
- 8.2.2 Epigenetic oncogenesis

8.3 Genome mobility upon damage
- 8.3.1 Repair protein accumulation within topological constraints
- 8.3.2 Coordinated repair protein binding at clustered genomic loci

8.4 Innovations for DamID-based omics and beyond
- 8.4.1 Approaches to study DNA repair and chromatin
- 8.4.2 Broader perspectives for technical advances with DamID

8.5 Concluding remarks

First, I will discuss technical features related to the single-cell methodologies presented in this thesis, as well as orthogonal approaches where relevant. Then I discuss the considerations and implications of the biological questions we set out to address, and frame the findings in their respective contexts.

8.1 Technical features of single-cell omics

This section delves into the technical considerations that shape single-cell sequencing experiments, focusing on genomic resolution, DNA accessibility bias, temporal dynamics, and spatial contact mapping. In doing so, it highlights challenges such as sparse coverage and molecular proxy biases alongside methodological strategies designed to mitigate them.

8.1.1 Resolution

In the context of single-cell sequencing, genomic resolution describes the granularity at which the features of interest can be mapped. The higher this (genomic) resolution, the more detail we can assign to the binding patterns of chromatin-associated proteins, and gain insights into their functions. Single-cell genomics studies, including those described in this thesis, face several challenges in achieving high resolution, primarily due to technical factors. Because of these challenges, coverage is generally poor (compared to bulk sequencing): signal detection across the whole genome in single cells inherently requires that all possible targeted regions be recovered and analyzed. This is the major experimental bottleneck. Even with in-well approaches and limited sample handling, loss of genomic material prior to amplification has remained inevitable. Sequencing throughput is the next source of signal dropout—in practice, only a fraction of the amplified DNA molecules will be sequenced. Increasing the sequencing depth should improve the resolution but is highly dependent on the complexity of the molecular pool; that is, if the pool contains many similar molecules (amplified from the same piece of original DNA), it will require a disproportionate amount of sequencing to collect newly informative reads. Moreover, larger sequencing space reduces experimental scalability and raises costs, which may be prohibitive. Together, these factors highlight that achieving a highly complex pool of amplified molecules is of critical importance.

For single-cell DamID, two technological advances are included in this thesis. First, in *Chapter 2*, I present a multiplexing approach that is now considered standard practice: incorporating a cell-specific barcode and a unique molecular identifier (UMI) during amplification. The cell-specific barcode allows for pooling of cells during preparation of sequencing libraries, thereby decreasing sample handling by

two orders of magnitude (from 1 cell per library to ~10^2 cells). The use of UMIs diminishes the notorious "jackpot effect" associated with PCR, where certain DNA fragments are disproportionately amplified, leading to an overrepresentation of particular sequences. This issue is exacerbated when amplifying cells individually. Second, from *Chapter 3* onwards, the experiments use scDam&T-seq or its derivative scDamID2, in which the PCR-step is entirely eliminated and replaced by *in vitro* transcription (IVT) for the amplification of single-cell material. Because IVT amplification is linear instead of exponential, the resulting pool is more complex. In addition, IVT is one-sided (whereas PCR is double-sided), thereby retaining molecules that only contain an amplification initiator on one end of the genomic fragment, and increasing the complexity of the amplified pool. Thus, although genome-wide single-cell profiles remain sparse (depending on the POI distribution, sequencing output may reach a median of 1-100K reads per cell), the experimental adaptations described above have yielded improved complexity, throughput, laboriousness, and cost-effectiveness.

Prospects

Future refinements to improve resolution and ensure a more comprehensive capture of target regions include locus-specific coverage strategies. This could be implemented via nested PCR on the single-cell material, as done in retrospective lineage tracing approaches such as GESTALT[1], LINNAEUS[2], and ScarTrace[3], which use CRISPR/Cas9-induced barcode editing to track cellular lineage and mutation accumulation over time. Alternatively, pulldown strategies on amplified material, such as probe-based hybridization or CRISPR-based enrichment, can increase coverage. This would be particularly insightful in the context of DNA damage and repair, where high coverage of the damage site would enable study of mutational signatures that currently go undetected due to data sparsity. Note that these strategies require *a priori* information on the loci of interest, and are thus not an unbiased approach to improving resolution genome-wide. Nevertheless, in combination with ongoing improvements in amplification strategies and sequencing throughput, such targeted adaptations hold promise for achieving higher resolution and enhancing the robustness of single-cell genomic analyses.

8.1.2 DNA accessibility bias

As extensively described in this thesis, current single-cell methods rely on a "proxy" to infer binding of DNA-associated proteins, most commonly via Dam- or Tn5- or MNase-mediated cleavage. Because these proxy proteins favor open chromatin, one of the most important considerations when interpreting such patterns is the bias towards genomic regions that are easily accessible. This bias depends on several

factors, including the inherent sensitivity of the proxy protein for DNA conformation, the concentration of the proxy, the concentration of the target protein, the affinity of the proxy for its target, and the effective capture time. As a result, this generally means that the final readout is a mixture of protein-specific signal and accessibility background.

It should be standard practice to include controls of freely diffusing proxy proteins, *i.e.*, untethered to the DNA-binding protein of interest or instead tethered to a binding-deficient version of the protein. Proxy protein mutants can be (computationally designed and) tested to exhibit less affinity for DNA and/or tunable enzymatic activity, resulting in reduced background and more on-target signal. Similarly, proxy proteins could be "caged"[4] to temporally limit their activity—this is especially applicable to Dam, which does not have a separate "activation" step to induce methylation. Lastly, depending on the genomic binding patterns, molecular processing steps (e.g., DNaseI, untethered MNase, or ATAC-like cleavage) could be included to generate an orthogonal accessibility readout of the same cell. This strategy has the added advantage that it could increase the complexity and effective resolution of the single-cell assay. By integrating these strategies—molecular tuning of proxy proteins and a dual readout of accessibility and binding—it is possible to minimize accessibility-related and better distinguish on-target from background signal.

8.1.3 Temporal dynamics

To understand (the order of) changes in cellular states, particularly in the context of epigenetic regulation, temporal resolution is essential. How can temporal dynamics be measured experimentally, and how would such a temporal axis be inferred from the data? These questions are especially pertinent for multi-modal techniques, where the different cellular readouts may have distinct temporal resolutions.

Traditionally, changes are measured directly over a defined period of time, often through time-course experiments following a given developmental or cellular differentiation trajectory, disease progression, therapeutic response, etc. When applied in bulk, such approaches are useful for revealing significant changes over time, but have important limitations. First, the observed changes may not be an accurate representation of the underlying population of cells. Second, time-courses rely on the assumption that all cells are progressing along the measured axis at the same rate, which may or may not be true. Third, dynamic changes that occur within individual cells cannot be accurately tracked. To address those limitations, single-cell omics have integrated temporal axes in various ways, utilizing both snapshot and cumulative measurements.

In this context, snapshot or static refers to data obtained from a single timepoint rather than continuous tracking over time. These measurements provide a high-resolution view of molecular states at a specific moment but do not inherently capture temporal progression. From static data, a "pseudotime" axis can be inferred after identifying gradual changes in gene expression or chromatin states. In brief, the approach quantifies similarity in the dataset, and orders the cells accordingly along a linear approximation of time. The assumption is thus that cells closer together on the inferred similarity axis are also closer on the real time trajectory. Velocity-based approaches infer temporal dynamics by using different cellular readouts within the same timepoint[5]. The general concept of velocity entails that both the rate and direction of cellular state can be inferred from two measurements that have a temporal relationship, such as spliced and unspliced transcripts. In single-cell genomics, chromatin velocity has been calculated by using the ratio between two different chromatin states, such as accessible chromatin and H3K27ac. Among others, chromatin velocity has served as an indicator for relaxation versus compaction[6], and to predict directionality of cellular differentiation[7,8].

Rather than inferring time or velocity from a single timepoint, cumulative measurements follow a pulse-chase experimental design to quantify changes. Molecules of interest are metabolically labeled during a specific period, such scEU-seq for mRNA turnover[9], or scEdU-seq for DNA replication[10]. By inferring the rate (and location) of incorporation, these methods thus reveal the dynamics of the molecular process.

Dam&ChIC: time-resolved measurement of the same chromatin protein

Chapter 4 of this thesis presents a unique approach to temporal chromatin dynamics. In essence, Dam&ChIC combines features of velocity-based methods and metabolic labeling. Velocity is inferred by measuring the ratio between two different modalities (DamID and ChIC) that capture the genomic binding of the same protein at different time resolutions. DamID represents the cumulative signal of methylated DNA (akin to metabolic labeling), while ChIC provides a snapshot of the current state. In the use case of Lamin B1, velocity thus infers a latent space that captures the directionality and kinetics of genome-lamina association. A high DamID/ChIC ratio would suggest the loss or release of DNA from the NL; conversely, a low DamID/ChIC ratio indicates a recently established contact and potential establishment of a LAD.

As such, Dam&ChIC is the only single-cell sequencing technology currently available to measure patterns of DNA-binding proteins from a temporal viewpoint. Inherent to the DamID technology, the main limitation is the loss of signal once DNA replication occurs. Temporal resolution is thus limited to the cell cycle between two S phases, prohibiting longer-term studies. On the other hand, since DamID signal generally requires induction times of several hours—which, additionally, are highly protein- and clone-dependent—rapid processes are poorly captured. Similar to the mitigation of accessibility bias, I envision Dam mutants and other experimental adjustments to enable DamID-based tracking at higher temporal resolution as well as across multiple cell divisions.

Overall, the development of methods that integrate temporal dynamics at the single-cell level marks a significant advancement, enabling detailed exploration of chromatin state transitions over time.

8.1.4 Contactomics

Several parts of this thesis examine spatial genome organization and its influence on nuclear processes and cellular functions. The overarching field can be termed "contactomics"[11], because it commonly relies on measurement of 3D genomic contacts. Here, I discuss the technical features that should be considered when interpreting contactomics data, with a particular focus on scDamID.

Proximity-based mapping of genomic loci

Proximity mapping captures the relative closeness of genomic loci within nuclear space. As described in *Chapter 7*, most DNA-based contact mapping techniques, such as chromosome conformation capture, GAM[12], and SPRITE[13], infer spatial proximity based on the frequency at which two (or more) genomic loci are detected together. These methods generate genome contact maps, where the measured interactions reflect chromatin folding and organization. While scDamID is primarily designed to infer the proximity of DNA to proteins, the resulting data can also be leveraged to infer aspects of genome conformation. This inference operates under the assumption that if two or more genomic regions consistently display coordinated protein binding across single cells, they are likely in physical proximity within the nuclear space. However, coordinated binding may also result from independent but functionally related processes, rather than spatial proximity. Also see [14] for the development of DamC, an adaptation of DamID to directly capture chromosome structure.

In *Chapter 6*, we leveraged the single-cell nature of our DNA-protein binding patterns to show that co-occurrence can reflect DNA-DNA proximity. This raises a key question: how is co-occurrence quantified, and to what extent do the available data points affect its interpretation?

Quantifying co-occurrence and higher-order contacts

In practice, interpretation of genomic contacts depends on the number of loci that can be simultaneously measured. Most studies, particularly those using Hi-C, quantify pairwise contacts; that is, they capture contacts between two (and only two) loci at a time. However, genome regulation often involves complex chromatin hubs, underscoring the need to quantify higher-order, multi-way, contacts. Many population-based attempt to address this by assuming transitivity. In the context of contactomics, the transitivity problem assumes the following: if locus A contacts locus B, and B contacts C, then A must also directly contact C, in all cases. This assumption is problematic because chromatin interactions do not always follow simple transitive rules. To better resolve higher-order contacts, experimental adaptations have been developed that enable capture of multiple DNA loci on the same molecule, such as Tri-C[15], MC-4C[16], Multiplex-GAM[17], and SPRITE[13]. These methods, discussed in *Chapters 1 and 7*, help disentangle the formation of complex chromatin configurations, but the information is still limited to the set of loci captured on each individual molecule. The relationship between one chromatin hub and another remains unknown.

In contrast, single-cell approaches have the distinct advantage that all measured contacts originate from the same genome. Co-occurrence can thus be measured between more than two loci simultaneously, in theory even for all genomic loci that show a minimum binding frequency to the protein of interest. However, note that co-occurrence is only inferred statistically across a population of cells, rather than directly quantified within a single cell. Lastly, many different chromatin configurations—and the relationship between them—can be evaluated in detail. This capability not only enhances our understanding of the spatial dynamics of chromatin, but also provides insights into the regulatory diversity that underpins cellular function.

In sum, it is the combination of methodological choices and the analytical approach that allows for a more complete picture of spatial positioning and genome organization.

8.2 Polycomb in cellular differentiation and pathology

This section explores how Polycomb repressive complexes and their associated chromatin modifications affect both normal developmental processes and the emergence of cancer. First, I discuss my work on Polycomb-mediated control of gene expression and describe recent links to genome topology. Then, I consider how disruptions in epigenetic programs can lead to oncogenesis, highlighting the multiple roles of Polycomb in cellular identity and genomic stability.

8.2.1 Hierarchical regulatory networks in lineage specification

The developmental trajectory of a cell is largely governed by regulatory networks that impose specific gene expression programs in a lineage-specific manner. Chromatin modifications, particularly histone marks, are at the core of this regulation, as they modulate both the accessibility of regulatory sequences and the recruitment of additional regulatory complexes. The work presented in this thesis showcases how novel single-cell mapping techniques can illuminate these chromatin-based mechanisms in unprecedented detail, including their influence on both normal development and disease.

During early embryonic development, cells undergo progressive lineage restriction that is driven by both transcriptional and epigenetic cues. In that context, Polycomb group proteins and associated histone modifications (such as H3K27me3) have emerged as crucial regulators, helping to establish "poised" or repressed states for lineage-specific genes to ensure activation in the correct developmental window. In *Chapter 3*, we employed EpiDamID to map Polycomb-associated histone modifications and gene expression in single cells of mouse embryoid bodies.

Polycomb chromatin is particularly associated with transcription factor (TF) genes, many of which control fundamental developmental processes and display cell type-specific expression patterns. With SCENIC[18], we used co-expression patterns and binding motifs to identify "regulons"—sets of TFs and their target genes—and observed that regulon activity is inversely correlated with Polycomb occupancy. This finding indicates that Polycomb-mediated repression of TF networks plays a key role in regulating transcriptional dynamics. Furthermore, we uncovered Polycomb-controlled regulatory network hierarchies, where both upstream TFs and their downstream targets are frequently co-regulated by Polycomb, forming tightly interconnected subnetworks.

Spatial proximity of Polycomb target genes

In parallel to these transcriptional insights, extensive research over the years has highlighted the role of 3D chromatin organization in regulating Polycomb target genes (reviewed in e.g., [19]). For example, recent studies in Drosophila have begun to reveal how the spatial proximity of these targets offers an additional layer of regulatory control. Notably, long-range contacts between Polycomb targets can lead to stable transgenerational epigenetic inheritance, demonstrating how spatial organization can drive long-lasting gene silencing[20]. Likewise of interest, spatial proximity between Polycomb-bound loci appears to be relatively rare and influenced by gene expression: co-repressed or co-expressed genes tend to interact within discrete nuclear compartments[21]. Moreover, that study showed that such interactions are typically limited to pairwise contacts rather than large clusters.

The single-cell profiles generated in *Chapter 3* offer a powerful lens for examining spatial proximity among Polycomb target genes on a genome-wide scale. While it fell outside the scope of the presented research, it would be possible to measure which loci are co-occupied by Polycomb within the same nucleus and thus infer whether those loci likely converge into the same repressive compartment. Such an approach would provide an orthogonal way of studying pairwise versus higher-order long-range coordination of Polycomb genes, and in a mammalian system.

Altogether, these insights emphasize that Polycomb-mediated repression not only regulates TF hierarchies at the sequence level but also leverages chromatin architecture to sculpt distinct developmental trajectories, reinforcing robust cell type-specific gene expression programs.

8.2.2 Epigenetic oncogenesis

Whereas lineage specification relies on finely orchestrated epigenetic programs, disruptions in these programs often underlie pathological states such as cancer. Mutations in chromatin regulators, aberrant activity of histone-modifying complexes (such as Polycomb), or deregulated deposition of specific histone marks can promote uncontrolled cell proliferation and block normal differentiation pathways. See [22,23] for extensive reviews on Polycomb repressive complexes in cancer.

“Epigenetic oncogenesis” refers to cancers that are epigenetically induced, where epigenome dysregulation alone can initiate malignancy (in the absence of driver mutations). Recent research in *Drosophila* larvae provides compelling evidence for such a role for Polycomb. One study showed that even a transient loss of Polycomb group protein function is sufficient to induce an irreversible tumorigenic state[24].

Disruption of Polycomb-mediated transcriptional silencing led to the de-repression of specific oncogenic pathways and established heritable alterations in cell fate. The resulting epigenetic cancers exhibited stable transcriptional changes and increasingly aggressive phenotypes over time, underscoring how a temporary epigenetic insult can have lasting consequences. More recent work in the same system has further revealed that sustained inactivation of the Polycomb repressive complex can drive these epigenetically initiated tumors toward genomic instability, ultimately resulting in DNA breaks and repair defects[25]. This transition from a primarily epigenetic to a genetic mode of tumor evolution highlights the importance of Polycomb repressive complexes not only as a regulator of cell identity but also a guardian of genomic stability.

In *Chapter 4* of this thesis, the Dam&ChIC method provides a way to track two chromatin proteins in parallel, exemplified by examining the interplay between Polycomb and the nuclear lamina during random X inactivation. Although the primary focus was on understanding the fundamental biology of X-chromosome silencing, it also highlights that nuclear organization and repressive domains can coordinate to maintain stable gene expression states. We can speculate that if these mechanisms are compromised—much like in the *Drosophila* model of Polycomb disruption—cells may acquire atypical epigenetic landscapes that potentiate or sustain oncogenesis. By combining a Polycomb and a DNA repair readout, Dam&ChIC technology can more directly investigate the relationship between epigenetic repression and genomic instability, potentially shedding light on the origins and evolution of epigenetic malignancies.

8.3 Genome mobility upon damage

Mobility of damaged DNA is an evolutionarily conserved feature—in yeast, fly, and mammalian cells, DNA breaks exhibit efficient motion through the chromatin network. As described in *Chapter 1*, such movement occurs across multiple scales, from nucleosomal remodeling to chromatin compartment switching.

8.3.1 Repair protein accumulation within topological constraints

A recent study (Arnould 2021) proposed that repair domain formation occurs via the mechanism of cohesin-mediated loop extrusion[26]. Upon damage, cohesin accumulates at the DSB site and unidirectionally extrudes the DNA at either end, forming loops until reaching a TAD border. This would enable DSB repair signaling across the entire TAD, including the accumulation of repair proteins such as 53BP1

and RAD51. In *Chapter 6*, we applied scDamID and sortChIC (separately) to study the accumulation of repair proteins in single cells. Because we used a cellular system of break induction at specified positions, it was possible to study the heterogeneity in protein accumulation across thousands of cells.

Where population-based studies present one (ideally replicated) pattern of protein accumulation or "spreading" per biological sample, we show that spreading patterns are actually highly variable. Prior to single-cell insights, the view on repair protein spreading was that the speed of loop extrusion might vary along the DNA fiber, thereby leading to a stronger accumulation of protein at certain regions. Our work suggests that the irregularity seen in population-based samples can (partly) be explained by the aggregation of many different cellular patterns. Certain patterns may occur more frequently, which we describe as "modes", so that we can distinguish between unimodal or multimodal spreading scenarios. These modes largely follow underlying genome topology, meaning that pre-damage TAD structures guide the extent of protein spreading.

The loop extrusion model also poses that repair domain formation should stop at TAD borders, typically marked by convergent CTCF sites. We tested this hypothesis using our single-cell data in two different ways: 1) quantification of coordinated repair protein signal, and 2) simultaneous measurement of repair protein signal and CTCF localization. First, we assessed the relationship between repair protein binding at each pair of genomic bins across all cells. This generates a "coordination map", indicating which genomic regions are coordinately bound by a repair protein. We show that, indeed, such coordination is constrained within TAD-like structures. Moreover, coordination is enhanced within actively extruding loops, as evidenced by increased normalized Hi-C contact frequencies.

Second, we sought to measure whether repair protein spreading is halted by the presence of a boundary element. We took advantage of the recent development of Dam&ChIC (presented in *Chapter 4*) to jointly profile RAD51 and CTCF, and show that variability in spreading correlates with presence of CTCF. That is, protein continues to accumulate beyond a potential TAD border lacking CTCF, whereas accumulation stops when CTCF is present. While this does not exclude the possibility that loop extrusion may merely decelerate at potential borders, we additionally attribute accumulation of signal at those regions to the full stalling of the extrusion process.

Cohesin as a DNA repair factor — loops need not apply

As briefly described in *Chapter 1*, upon DSBs, ATM phosphorylates H2AX, resulting in γH2AX, which is then bound by MDC1 (Figure 8.1 inset). According to the model described above, γH2AX domains form as chromatin passes by ATM at the DSB, facilitated by cohesin-mediated loop extrusion. How do cohesins accumulate at the break site? One possibility involves *de novo* loading of cohesin at either end of the DSB, thereby anchoring one-sided loops. Alternatively, ongoing loop extrusion across the genome may be disrupted by DNA damage, trapping cohesins at those locations.

A recent study (Fedkenheuer 2025) proposes that cohesin has two independent functions: 1) a purely architectural role, maintaining spatial separation between DSBs, and 2) as a repair factor, where phosphorylated cohesin accelerates DNA end-joining for more efficient repair[27]. In this work, cohesin phosphorylation was shown to have a limited effect on local chromatin topology, indicating a non-architectural function. Further, cohesin accumulation within TADs argues against an anchoring function (where cohesin would be concentrated at the break site and TAD borders). More evidence for separate roles of cohesin comes from time-lapse microscopy[28] and recent preprints[29,30]. First, cohesin accumulated at DSBs after MDC1 foci formation[28]. Similarly, γH2AX and 53BP1 domains formed after detection of the extrusion-associated "stripes" seen in Hi-C[29]. This is not in line with a role for loop-extruding cohesin in establishing γH2AX repair domains (the platform for MDC1 binding). Further, mathematical modeling of MDC1 and ATM kinetics found that a minimal model of ATM diffusion was both sufficient and necessary to explain MDC1 spreading[28]. Finally, (in Teloni 2025), loop-forming cohesin appears to control the homology search range during HDR, while "cohesive" cohesin tethers broken ends to the sister chromatid. I attempt to consolidate these findings in Figure 8.1.

Thus, cohesin likely plays distinct roles in both genome topology and DNA repair, and loop extrusion may not be a major driver of γH2AX domain formation.

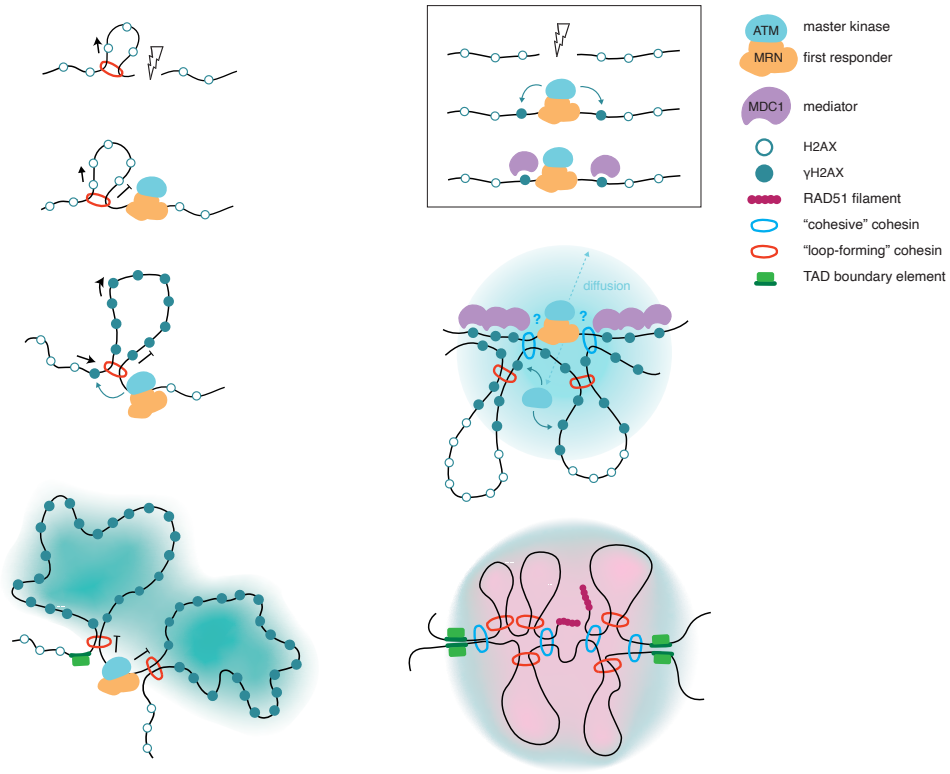


Figure 8.1. Models for DNA repair domain formation.

Inset: the initial DNA damage response. MRN senses the DSB and recruits and activates ATM, which then phosphorylates H2AX, which serves as a binding platform for MDC1.

Left: loop extrusion model. Upon a DSB, loop extrusion is halted and further blocked by DNA repair factors including ATM, leading to an accumulation of cohesin at the DSB. Extrusion resumes unidirectionally, phosphorylating nearby H2AX-containing nucleosomes as the chromatin passes by ATM. The entire TAD is thereby phosphorylated, establishing the DNA repair domain.

Right: diffusion and dual cohesin model. After the initial response, activated ATM diffuses from the DSB outwards, phosphorylating proximal H2AX. This establishes the repair domain. Cohesins are recruited *de novo* to the DSB, but the exact timing and function remain unclear. For DSBs undergoing HDR, RAD51 filaments on either side of the break scan the entire TAD in search of homology. Cohesive cohesins ensure scanning is limited to the damaged, while extruding cohesins facilitate linear (1D) long-range scanning. Figure elements adapted and incorporated from [26,30].

Spreading of HDR factors across large genomic regions

Previous work showed that proteins such as MDC1 and 53BP1 form broad domains, while proteins specific to homology-directed repair (HDR) were shown to have a more localized distribution[31]. We observe similarly broad domains for MDC1 and 53BP1, but also for HDR factor RAD51. The broad RAD51 domains detected by DamID may be explained by the cumulative nature of the signal. Unlike ChIP-based methods used in earlier studies, DamID is highly sensitive to rapid, iterative DNA-binding events, especially within the topological constraints of the chromatin.

Relatedly, in yeast, Rad51 has been shown to bind double-stranded (ds) DNA during the homology search process[32,33], rather than exclusively binding to single-stranded (ss) DNA during end resection. These binding events related to homology search were postulated to be transient interactions occurring within the immediate 3D environment of the break[32,34]. Recent preprints support those descriptions of RAD51 binding (earlier described as controversial)[29,35]. First (in Marin-Gonzalez 2025), RAD51 exhibits two-tiered accumulation around DSBs, forming both a narrow peak (~5 kb, corresponding to ssDNA interaction) and a broad domain (~1 Mb, representing dsDNA interaction during homology search). Second (in Yeh 2025), pre-existing chromatin structure influences the landscape of homology search: proximity—rather than sequence homology—determines which DNA templates are searched. Together, these results align with our observations of broad RAD51 occupancy and its constraints by genome topology.

Limitations and potential prospects

Given the (initially) surprisingly broad distribution of RAD51 described above, the first limitation of our study is the lack of validation with an antibody-based approach in single cells. While we were able to successfully capture signal in 100-cell populations, sortChIC with RAD51 yielded very sparse data. Instead of a pA-MNase approach, a fusion of pA-Dam[36] may improve the signal; however, it remains a snapshot and does not benefit from the classic *in vivo* accumulation of methylation.

Repair domain formation was studied by evaluating protein accumulation at one locus at a time, across all cells. Therefore, we could have implemented approaches to specifically amplify those loci in order to improve coverage. Alternatively, single-molecule approaches could be appropriate. In particular, directed methylation with long-read sequencing (DiMeLo-seq[37]) can map DNA-protein contacts on reads >1 Mb in length. RAD51 and 53BP1 have been successfully mapped with DiMeLo-seq as well as long-read DamID (unpublished data). I envision multiplexed approaches obtain dual readout of e.g., RAD51 and CTCF to disentangle repair protein spreading and genome organization. In section *8.4.2*

Broader perspectives for technical advances with DamID I further expand on single-molecule adaptations.

Data sparsity is also a limitation of the Dam&ChIC experiments to jointly profile RAD51 and CTCF. By aggregating cells based on patterns in the RAD51 signal, we were able to gain insights from the CTCF signal in those "clusters" of cells. While this is similar to transcription-based clustering that should capture different cell types (or a similar feature), the resulting clusters no longer represent true single-cell resolution.

Lastly, we propose that the heterogeneity in repair protein spreading is related to underlying genome organization, which we support with single-molecule SPRITE experiments. Still, the lack of validation using a non-sequencing-based approach could be seen as a limitation, to which DNA-FISH experiments may provide an answer.

8.3.2 Coordinated repair protein binding at spatially clustered genomic loci

The clustering of damaged DNA is a poorly understood phenomenon. Studies have shown spatial coalescence of repair protein foci by (live-cell) microscopy, and contact of two proximal genomic loci by sequencing, but the orthogonal methods remained separate. That is, it was not possible to infer the identity of the DNA within the repair foci by imaging, nor to study clustering of repair foci by sequencing.

In *Chapter 6*, we address this gap with two separate approaches. Single-cell repair protein profiles (DamID or ChIC) were used to assess coordinated binding at >2 DNA loci simultaneously. Single-molecule contact profiles (Tri-C) were used to assess spatial proximity of those loci. Our goal was to tackle the transitivity problem by determining whether contacts represent true multi-way interactions within single cells or simply pairwise interactions observed across different cells. Transitivity implies a direct, sequential relationship: if A is related to B, and B is related to C, then A is *necessarily* related to C. Our findings revealed a predominantly cooperative relationship between clusters of damaged loci, indicating that DSB triplets often form genuine multi-way interactions in a single cell. Notably, however, we also uncovered competitive behavior: if loci A and B were jointly bound by a repair protein, then locus C was not bound. This observation argues against a strictly transitive relationship. In other words, while we confirmed multi-way hubs, the subset of non-transitive patterns reflects the inherent cell-to-cell heterogeneity in how loci cluster for repair.

This mixed outcome further underscores the complexity of multi-locus damage responses, leaving open the question of what forces ultimately bring these broken sites together in nuclear space.

Further open questions

Mechanistically, the function of DSB clustering remains elusive. Notably, clustering mainly involves breaks in genic regions that are set to undergo HDR[38,39]. Since HDR is inhibited in G1, this could serve as a temporary pausing of DNA repair, and clustered DSBs would then be spatially segregated from the rest of the genome. However, how would a damaged region, clustered in G1, be correctly replicated to generate a sister chromatid that serves as a template for HDR in G2? In addition, sequestering broken DNA in clusters increases the risk of chromosomal translocations and oncogenic fusions, which can have deleterious consequences.

Limitations and potential prospects

Our study gave the first description of multiple DSBs simultaneously bound by repair protein in an unbiased manner. However, we did not explore the position of DSB clusters inside the nucleus. This could be achieved by combining DNA-FISH (visualizing specific genomic loci) with IF (visualizing repair protein), even in combination with the ^{m6}A-Tracer (visualizing protein-bound DNA). Aside from proving information on nuclear position, such an experiment would also validate the spatial contact approaches and coordinated repair protein accumulation.

We used the human U2OS DIvA cell system in our proof-of-concept study, because it is well characterized and extensive public data availability. However, U2OS is a cancer cell line with a highly altered karyotype. The single-cell work in this system cannot distinguish between alleles, which is particularly relevant for events on the same molecule, such as DSB clustering. Therefore, we relied on Tri-C for allelic resolution to validate multi-way DSB clustering. Future work might further address this limitation by using a haploid cell line, or a hybrid cross from known genotypes with allele-specific SNPs. Still, studies in G2 phase of the cell cycle remain challenging in this regard.

Altogether, the findings in this section demonstrate that genome mobility upon damage is an active, regulated process shaped by chromatin context, nuclear organization, and likely some mechanochemical forces. In the next section, I discuss approaches to further study DNA repair in the context of chromatin, as well as broader innovations in sequencing and imaging technologies.

8.4 Innovations for DamID-based omics and beyond

8.4.1 Approaches to study DNA repair and chromatin

Throughout this thesis, the aim has been to show how single-cell measurements provide valuable insights that may not be gleaned from population-level studies. This is most poignant in systems or biological processes where observations are stochastic—such as DNA-binding proteins that are unequally distributed across the genome.

Stochastic DNA damage and the role of epigenetics in therapeutic interventions

Exogenous DNA damage, such as that induced by ionizing radiation or chemotherapeutic agents, occurs stochastically and can be a double-edged sword. On one hand, damage may lead to mutations, cancer, or other pathologies. On the other, this very property of overwhelming cells with harmful lesions is harnessed therapeutically to drive cell death. A key limitation is that the precise genomic locations of damage are often unknown, making large-scale population averages uninformative for understanding which chromatin contexts incur breaks or how individual cells respond.

In *Chapters 5 and 6* I present the first report of genome-wide repair factor distributions and the extensive heterogeneity between cells. We used an endonuclease that induces damage at known genomic sequences, because it provided the best opportunity for negative controls. In this way, we could show that 1) our experimental approach of using DNA repair proteins as a proxy for damage is feasible and valid and 2) our computational approach can unbiasedly detect signal without a priori knowledge of position or breadth. Of course, the true power of this approach lies in the ability to identify DSB sites in response to damaging agents used therapeutically. This thesis introduces the toolbox necessary to perform such experiments, underscoring the value of discovery-driven science.

By jointly measuring transcriptional output with scDam&T-seq, we also gain the ability to correlate distinct repair patterns with specific cellular outcomes, such as changes in gene expression or differences in proliferative capacity. Importantly, using Dam&ChIC, we can now pinpoint where damage occurs, and whether certain epigenetic or chromatin states predispose cells to more—or fewer—DNA breaks. This opens the door to testing epigenetic drugs that modulate chromatin organization. Already, compounds targeting histone deacetylases or methyltransferases have shown promise in affecting DNA repair outcomes[40], potentially enabling more targeted or less toxic treatments by lowering the dose of damaging agents. Likewise, CRISPR-based therapies are increasingly sensitive to epigenetic influences on DSB

formation and repair. As alluded to under *8.2.2 Epigenetic oncogenesis*, epigenetic cancers exhibit a signature of altered chromatin states, suggesting that therapies aimed at correcting such states could be a powerful treatment strategy.

Altogether, these insights provide a foundation for optimizing therapy—balancing the beneficial effects of inducing DNA damage against the risk of deleterious outcomes, while leveraging epigenetic modulation to refine treatment efficacy.

A targeting system for chromatin-specific DNA damage

In *Chapter 5*, we demonstrated how Dam&ChIC can map DNA repair factors in specific chromatin contexts, initially validating that RAD51 localizes to actively transcribed euchromatin marked by H3K36me3 rather than to H3K9me3-marked heterochromatin. However, a known limitation of the DIvA cell system is that the vast majority of breaks occurs in open chromatin, because the endonuclease AsiSI is sensitive to CpG methylation, leaving other regions unstudied.

To overcome this, I envision that DamID can be repurposed into an approach to induce DNA damage wherever a particular protein resides. For example, regions near the nuclear periphery can be targeted by Lamin B1; facultative heterochromatin and euchromatin can be targeted by H3K27me3 and H3K9ac mintbodies, respectively. Because Dam is fused to the protein of interest, it selectively methylates binding sites, effectively "tagging" those loci for *in vivo* cleavage. The introduction of the methylation-specific restriction enzyme DpnI into cell cultures expressing Dam fusion proteins enables selective double-stranded DNA damage at those locations. As a proof of concept, I performed a pilot experiment using Dam-Lamin B1, which resulted in targeted DNA damage at the nuclear periphery (Figure 8.2). While this approach is broadly sequence-unspecific compared to CRISPR/Cas9-based damage induction, it allows for genome-wide cleavage at regions bound by particular proteins. Future implementations integrating sequencing-based approaches or proteomics could further explore chromatin-specific DNA repair pathways.

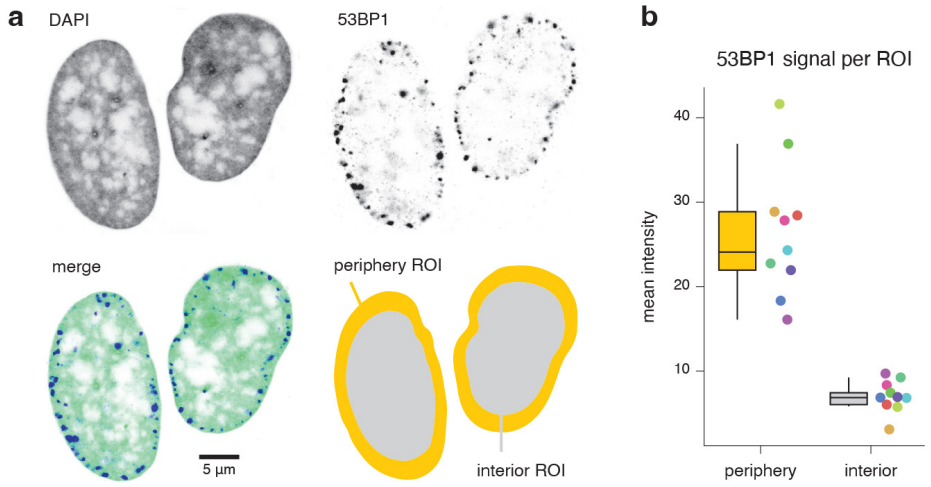


Figure 8.2. Targeted DNA damage at regions methylated by Dam-Lamin B1.
a, Top: confocal images of two representative nuclei showing DAPI and 53BP1 immunofluorescent staining. Bottom left: colored overlay. Bottom right: representation of ROIs selected for quantification, based on expected Lamin B1 localization (nuclear periphery and interior). **b**, Quantification of 53BP1 signal (normalized for area) per ROI per nucleus. Each nucleus is colored separately ($n = 10$). Boxes represent the median, first and third quartiles.

8.4.2 Broader perspectives for technical advances with DamID

As described in *Chapter 7*, single-cell epigenomic sequencing technologies continue to develop at unprecedented rate. Here, I propose future technical advances that build upon the work in this thesis.

Native long-read sequencing

Thanks to high-throughput short-read sequencing, our collective understanding of genome regulation has advanced considerably. Yet the complete annotation of the human genome was only achieved in 2022, and this feat required long-read sequencing (LRS) to assemble repetitive regions. With LRS named *Nature*'s Method of the Year 2022, the field is rapidly expanding into epigenomics (see [41] for a comprehensive review).

A distinct advantage of LRS is its single-molecule resolution, as illustrated by DiMeLo-seq[37], which maps protein-DNA interactions along individual chromatin fibers. However, the single-molecule advantage is inherently coupled to the loss of

single-cell resolution. Because DNA modifications cannot currently be amplified *in vitro*, isolated (modified) DNA must be prepared for sequencing as-is, and LRS platforms still require many orders of magnitude more input material than is present in a single cell.

Still, the DamID approach of *in vivo* proximal methylation can be modified for LRS platforms to capture protein-DNA contacts over large genomic distances, especially in heterochromatic regions. In addition, akin to Dam&ChIC, a dual protein readout can be achieved by combining long-read DamID with DiMeLo-seq. Because LRS platforms capture DNA modifications directly, such methods also yield information on endogenous CpG methylation, essentially providing multimodal insights.

Imaging and multimodal sequencing with DamID

The research in this thesis was introduced with the notion that imaging and sequencing have historically been divided. Only in the last five years have these fields begun to merge, simultaneously capturing morphological and molecular features. Spatially resolved transcriptomics and proteomics were respectively named *Nature*'s Method of the Year 2020 and 2024, but spatial (epi)genomics remains an ongoing challenge. See [42,43] for comprehensive reviews on spatial omics.

Spatial resolution generally refers to cellular location and associated phenotypes within a tissue. Here, rather than the multicellular context, I focus on the potential to capture genomic sequence identity as well as nuclear position from the same cell. Until now, even with the advent of single-cell genomics, a comparison of imaging and sequencing data has been limited to describing the distributions of their respective measurements across cells. DamID-based imaging is made possible by the ^{m6}A-Tracer, a protein module containing a domain that recognizes Dam-methylated DNA, fused to a fluorescent protein[44]. Paired imaging and sequencing of LADs has been performed using an integrated microfluidic device (microDamID[45]). Because cells are dissociated and individually processed on the device, this approach can achieve very high optical resolution. A downside of this design is the limited throughput at 10 cells per device. Throughput could be increased by using microfluidics only to screen and image cells, subsequently depositing them into individual wells for further off-device molecular processing. Many applications of this approach can be envisioned, from nuclear organization to epigenetic regulation to DNA repair. In light of recent findings on atypical genome organization during early development[46], microDamID could be especially useful for ultra-low-input samples such as preimplantation embryos. It would be of interest to explore the nuclear lamina and the associated genome from a microscopic point of view.

Moreover, as shown in *Chapter 5*, DamID Tracer imaging visualizes DNA bound by repair proteins, thereby revealing its spatial position inside the nucleus. In *Chapter 6*, I showed that multiple damaged regions cluster together into repair protein hubs that may serve as repair factories. Building upon this framework, spatiotemporal tracking would provide critical insights into the persistence, mobility, and fate of these repair foci over time. The ability to visualize the Tracer bound to methylated DNA even after the repair protein has dissociated from the site allows for extended temporal tracking, either through time series or live-cell imaging (manuscript in preparation). This would be particularly relevant in understanding how damage incurred in the mother cell propagates through mitosis and influences repair dynamics in daughter cells. Coupled measurements thus enable direct comparisons of the imaged foci and sequencing signal for that same cell. The multimodal readout described here is powerful because it better leverages the information on location-specific repair mechanisms: a more complete view is greater than the sum of its parts.

Of course, the ultimate aim would be to consider sub-nuclear spatial position within the multicellular context of a heterogeneous sample.

8.5 Concluding remarks

> *I will ask you to mark again that rather typical feature of the development of our subject; how so much progress depends on the interplay of techniques, discoveries and new ideas, probably in that order of decreasing importance.*
>
> *—Sydney Brenner, 1980*

This thesis reflects that idea—that scientific progress is driven not only by answering existing questions but by reshaping the very way we ask them. The methods developed and applied here aim to expand the limits of what can be measured and, in doing so, refine our understanding of genome biology.

A unifying lesson is that cellular processes are neither uniform nor static. The power of single-cell data lies precisely in capturing these fluctuations. This thesis explores how the chromatin environment changes dynamically with cell state, from epigenetic control during lineage specification to genome topology during DNA repair.

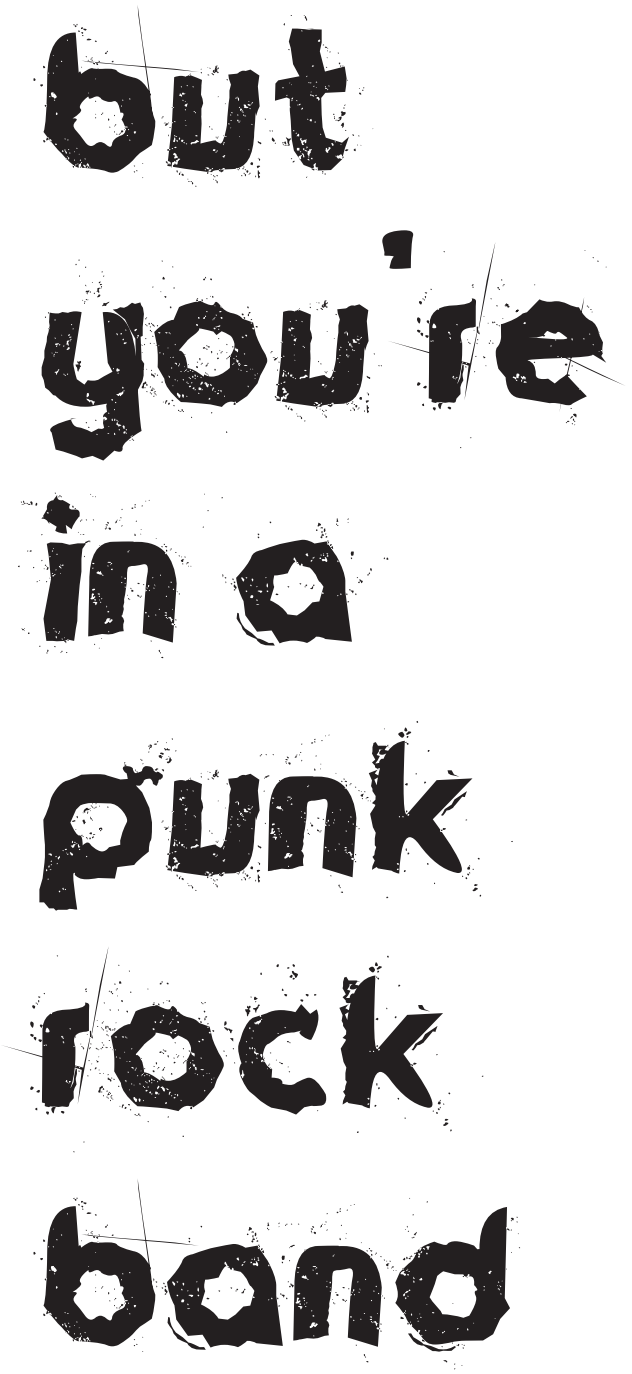
As single-cell methodologies and their computational counterparts continue to advance, they will not only deepen our understanding of genome regulation, but may also lead to practical applications, from improving disease treatment to informing ecological conservation strategies. While the specific applications of these approaches will evolve, the underlying philosophy remains the same—technology shapes knowledge, and knowledge, in turn, reshapes the way we see the world.

References

1. McKenna, A. et al. Whole-organism lineage tracing by combinatorial and cumulative genome editing. Science 353, aaf7907 (2016).
2. Spanjaard, B. et al. Simultaneous lineage tracing and cell-type identification using CRISPR-Cas9-induced genetic scars. Nat Biotechnol 36, 469-473 (2018).
3. Alemany, A., Florescu, M., Baron, C.S., Peterson-Maduro, J. & van Oudenaarden, A. Whole-organism clone tracing using single-cell sequencing. Nature 556, 108-112 (2018).
4. Zorn, J.A. & Wells, J.A. Turning enzymes ON with small molecules. Nat Chem Biol 6, 179-188 (2010).
5. La Manno, G. et al. RNA velocity of single cells. Nature 560, 494-498 (2018).
6. Tedesco, M. et al. Chromatin Velocity reveals epigenetic dynamics by single-cell profiling of heterochromatin and euchromatin. Nat Biotechnol 40, 235-244 (2022).
7. Bartosovic, M. & Castelo-Branco, G. Multimodal chromatin profiling using nanobody-based single-cell CUT&Tag. Nat Biotechnol 41, 794-805 (2023).
8. Yeung, J. et al. scChIX-seq infers dynamic relationships between histone modifications in single cells. Nat Biotechnol 41, 813-823 (2023).
9. Battich, N. et al. Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. Science 367, 1151-1156 (2020).
10. van den Berg, J. et al. Quantifying DNA replication speeds in single cells by scEdU-seq. Nat Methods 21, 1175-1184 (2024).
11. Bouwman, B. Contactomics - exploring new dimensions of 3D genome architecture. (2016).
12. Beagrie, R.A. et al. Complex multi-enhancer contacts captured by genome architecture mapping. Nature 543, 519-524 (2017).
13. Quinodoz, S.A. et al. Higher-Order Inter-chromosomal Hubs Shape 3D Genome Organization in the Nucleus. Cell 174, 744-757 e24 (2018).
14. Redolfi, J. et al. DamC reveals principles of chromatin folding in vivo without crosslinking and ligation. Nat Struct Mol Biol 26, 471-480 (2019).
15. Oudelaar, A.M. et al. Single-allele chromatin interactions identify regulatory hubs in dynamic compartmentalized domains. Nat Genet 50, 1744-1751 (2018).
16. Allahyar, A. et al. Enhancer hubs and loop collisions identified from single-allele topologies. Nat Genet 50, 1151-1160 (2018).
17. Beagrie, R.A. et al. Multiplex-GAM: genome-wide identification of chromatin contacts yields insights overlooked by Hi-C. Nat Methods 20, 1037-1047 (2023).
18. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods 14, 1083-1086 (2017).
19. Cheutin, T. & Cavalli, G. The multiscale effects of polycomb mechanisms on 3D chromatin folding. Crit Rev Biochem Mol Biol 54, 399-417 (2019).
20. Fitz-James, M.H., Sabarís, G., Sarkies, P., Bantignies, F. & Cavalli, G. Interchromosomal contacts between regulatory regions trigger stable transgenerational epigenetic inheritance in Drosophila. Mol Cell 85, 677-691.e6 (2025).
21. Gurgo, J. et al. Multiplexed chromatin imaging reveals predominantly pairwise long-range coordination between Drosophila Polycomb genes. Cell Rep 43, 114167 (2024).
22. Piunti, A. & Shilatifard, A. The roles of Polycomb repressive complexes in mammalian development and cancer. Nat Rev Mol Cell Biol 22, 326-345 (2021).
23. Parreno, V., Martinez, A.M. & Cavalli, G. Mechanisms of Polycomb group protein function in cancer. Cell Res 32, 231-253 (2022).
24. Parreno, V. et al. Transient loss of Polycomb components induces an epigenetic cancer fate. Nature 629, 688-696 (2024).
25. Rawal, C.C. et al. Sustained inactivation of the Polycomb PRC1 complex induces DNA repair defects and genomic instability in epigenetic tumors. Histochem Cell Biol 162, 133-147 (2024).
26. Arnould, C. et al. Loop extrusion as a mechanism for formation of DNA damage repair foci. Nature 590, 660-665 (2021).
27. Fedkenheuer, M. et al. A dual role of Cohesin in DNA DSB repair. Nature Communications 16, 843 (2025).
28. Danovski, G. et al. Diffusion of activated ATM explains γH2AX and MDC1 spread beyond the DNA damage site. Iscience 27(2024).

29. Marin-Gonzalez, A. et al. Cohesin drives chromatin scanning during the RAD51-mediated homology search. bioRxiv, 2025.02. 10.637451 (2025).
30. Teloni, F. et al. Cohesin guides homology search during DNA repair via loops and sister chromatid linkages. bioRxiv, 2025.02. 10.637359 (2025).
31. Clouaire, T. et al. Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks Deciphers Repair Pathway Chromatin Signatures. Mol Cell 72, 250-262 e6 (2018).
32. Renkawitz, J., Lademann, C.A., Kalocsay, M. & Jentsch, S. Monitoring homology search during DNA double-strand break repair in vivo. Molecular cell 50, 261-272 (2013).
33. Peritore, M., Reusswig, K.-U., Bantele, S.C., Straub, T. & Pfander, B. Strand-specific ChIP-seq at DNA breaks distinguishes ssDNA versus dsDNA binding and refutes single-stranded nucleosomes. Molecular cell 81, 1841-1853. e4 (2021).
34. Kalocsay, M., Hiller, N.J. & Jentsch, S. Chromosome-wide Rad51 spreading and SUMO-H2A. Z-dependent chromosome fixation in response to a persistent DNA double-strand break. Molecular cell 33, 335-343 (2009).
35. Yeh, C.D. et al. Proximity determines donor candidacy during DNA double-stranded break homology directed repair. bioRxiv, 2025.02. 10.637161 (2025).
36. van Schaik, T., Vos, M., Peric-Hupkes, D., Hn Celie, P. & van Steensel, B. Cell cycle dynamics of lamina-associated DNA. EMBO Rep 21, e50636 (2020).
37. Altemose, N. et al. DiMeLo-seq: a long-read, single-molecule method for mapping protein-DNA interactions genome wide. Nat Methods 19, 711-723 (2022).
38. Aymard, F. et al. Genome-wide mapping of long-range contacts unveils clustering of DNA double-strand breaks at damaged active genes. Nat Struct Mol Biol 24, 353-361 (2017).
39. Arnould, C. et al. Chromatin compartmentalization regulates the response to DNA damage. Nature 623, 183-192 (2023).
40. Schep, R. et al. Chromatin context-dependent effects of epigenetic drugs on CRISPR-Cas9 editing. Nucleic Acids Research 52, 8815-8832 (2024).
41. Liu, M. et al. Genome-coverage single-cell histone modifications for embryo lineage tracing. Nature (2025).
42. Baysoy, A., Bai, Z., Satija, R. & Fan, R. The technological landscape and applications of single-cell multi-omics. Nat Rev Mol Cell Biol 24, 695-713 (2023).
43. Vandereyken, K., Sifrim, A., Thienpont, B. & Voet, T. Methods and applications for single-cell and spatial multi-omics. Nat Rev Genet 24, 494-515 (2023).
44. Kind, J. et al. Single-cell dynamics of genome-nuclear lamina interactions. Cell 153, 178-92 (2013).
45. Altemose, N. et al. muDamID: A Microfluidic Approach for Joint Imaging and Sequencing of Protein-DNA Interactions in Single Cells. Cell Syst 11, 354-366 e9 (2020).
46. Guerreiro, I. et al. Antagonism between H3K27me3 and genome-lamina association drives atypical spatial genome organization in the totipotent embryo. Nat Genet 56, 2228-2237 (2024).

you say
you wanna
play
country


but
you're
in a
punk
rock
band

Addendum

Samenvatting in het Nederlands

Resúmen kòrtiku na Papiamentu

Curriculum vitae

List of publications

Samenvatting in het Nederlands

Al het leven op aarde wordt gedreven door het goed functioneren van haar basiseenheid: de cel. Hoe wordt een cel gebouwd? De complexe informatie die nodig is om een cel (en daarmee een organisme) te bouwen is gecodeerd in het genoom. In menselijke cellen is 2 meter DNA verpakt in de celkern, wat van grote invloed is op de processen die plaatsvinden in die kern. Bovendien wordt het DNA gelezen en onderhouden door gespecialiseerde eiwitten die DNA kunnen reguleren, repliceren, recombineren en repareren. Hoewel de ruimtelijke organisatie van het genoom al sinds de 19e eeuw wordt bestudeerd, is er weinig bekend over de precieze wisselwerking tussen gecodeerde genetische informatie, de geassocieerde eiwitten en, uiteindelijk, cellulaire functie.

De afgelopen decennia hebben een sterke ontwikkeling doorgemaakt dankzij nieuwe technieken in microscopie (beeldvorming) en zogeheten *sequencing* (uitlezen) van cellen en DNA. Op basis van deze technieken kunnen we verkennen waar en hoe DNA interacteert met eiwitten en of dit functionele effecten heeft. Processen in de celkern zijn vaak zeer dynamisch en heterogeen. Dezelfde genetische informatie ondergaat bijvoorbeeld "epigenetische" regulatie waardoor verschillende celtypen kunnen ontstaan. Epigenetica betekent hier epi (bovenop) genetica (DNA), wat dus verwijst naar veranderingen buiten de genetische code. Dergelijke regulatie kan worden bereikt door ruimtelijke herpositionering van het DNA in de kern, of door aanvullende mechanismen die de uitleesbaarheid van het DNA veranderen. Daarnaast kan reorganisatie of modificatie van het genoom optreden als reactie op breuken en andere vormen van schade aan het DNA. Zowel het optreden van DNA-schade alsook het herstel daarvan zijn variabele processen. Samen benadrukt dit de noodzaak om cellen op individueel niveau te kunnen bestuderen.

Deze dissertatie presenteert nieuwe technieken voor sequencing van individuele cellen. Ik richt me daarbij op het bestuderen van genoomorganisatie, epigenetische regulatie, en DNA-herstel. Nogmaals: de cellen verschillen aanzienlijk in DNA-gerelateerde processen, en het is belangrijk om die variatie in kaart te brengen voor accurate analyses en conclusies.

In **Hoofdstuk 1** (Introductie), geef ik een kort perspectief van de historische ontwikkelingen die relevant zijn geweest voor de studie van het genoom en de celkern. Daarnaast leg ik de biologische en technologische basis om de rest van de thesis in de juiste context te kunnen plaatsen.

Hoofdstuk 2 presenteert een gedetailleerd experimenteel protocol om single-cell DamID sequencing uit te voeren. De ruimtelijke organisatie van het DNA kan in kaart worden gebracht door specifiek uit te lezen welke delen van het DNA in contact

waren met de rand van de celkern (in plaats van met het midden). Dit is mogelijk omdat specifieke eiwitten specifieke contacten maken met het DNA. Zo kan deze techniek dus ook gebruikt worden om andere eiwitten te bestuderen.

Hoofdstuk 3 introduceert EpiDamID, een nieuwe aanpak voor het detecteren van bepaalde eiwit-modificaties die belangrijk zijn voor het reguleren van het genoom. Door tegelijkertijd zulke eiwit-modificaties en genexpressie te meten, kunnen we genoomregulatie koppelen aan cellulaire functie. Dit geeft bijvoorbeeld nieuwe inzichten in differentiatie van celtypen tijdens embryonale ontwikkeling.

Hoofdstuk 4 beschrijft een technische innovatie om met twee methodes (DamID en ChIC) in dezelfde cel eiwitten te meten. Op deze manier kunnen twee verschillende eiwitten tegelijkertijd gemeten worden. Ook is de timing van de twee methodes verschilt: DamID geeft een cumulatieve blik van eiwitbinding in de afgelopen uren, terwijl ChIC een momentopname is van de huidige staat van eiwitbinding. Dit geeft een zeer unieke kijk op de dynamiek van eiwitbinding, waar we gebruik van maken om bijvoorbeeld genoomorganisatie te bestuderen tijdens celdeling.

Met behulp van deze methoden genereerde ik de eerste *single-cell* profielen van eiwitten die betrokken zijn bij DNA-herstel. **Hoofdstuk 5** beschrijft de aanzienlijke heterogeniteit in eiwitprofielen die beïnvloed worden door genoomorganisatie. **Hoofdstuk 6** bouwt voort op deze data, waarbij ik laat zien dat beschadigde stukken DNA ruimtelijk samenkomen. Die contacten worden gecoördineerd gebonden door reparatie-eiwitten, wat duidt op een coöperatief proces. Vergelijk dit enigszins met een repair café waarbij men samenwerkt, in plaats van ieder alleen in hun eigen werkplaats.

In **Hoofdstuk 7** geef ik een overzicht van recente vooruitgang in *single-cell* en *single-molecule* methodes die toepasbaar zijn op studies omtrent epigenetica en genoomorganisatie. **Hoofdstuk 8** (Discussie) plaatst bovenstaande bevindingen in bredere context. Eerst ga ik in op de technische kenmerken van single-cell technieken. Daarna beschouw ik mijn werk aan DNA-herstel en genoomorganisatie, waarbij ik recente inzichten uit ander werk belicht en ook toekomstige richtingen voorstel.

Samenvattend levert dit proefschrift zowel nieuwe technologieën als biologische inzichten op, waarmee ons begrip van genoomregulatie en stabiliteit wordt vergroot.

Resúmen kòrtiku na Papiamentu

Tur sélula den nos kurpa tin mesun material genétiko (yamá “ADN” pa e deskripshon molekular di ásido desoksiribonukléiko), pero diferente sélula por tin diferente funshon. Por ehèmpel, sélulanan di wesu tin forma i funshon distinto for di sélulanan di kueru. E diferensha aki ta surgi pasobra kada sélula ta organisá i “lesa” su ADN di manera úniko. Métodonan tradishonal ta studia miones di sélula huntu i konsekuentemente no ta logra detektá e diferenshanan importante ku ta eksistí entre sélula individual. Ta importante pa komprondé kon e prosesonan nuklear ta sosodé na nivel individual, spesialmente ora algu bai robes, manera sa pasa den kanser.

Mi a desaroyá métodonan nobo pa investigá e proseso di regulashon di ADN den sélulanan individual. Ku e métodonan aki, mi por a sigui kua parti di ADN ta forma interakshon ku sierto proteina, inkluso esnan ku ta repará daño na ADN. Aunke ADN ta sufri daño diariamente sin konsekuensia grave, sierto daño por kousa kanser i otro malesa.

Usando e téknikanan aki, mi a mustra ku reparashon di ADN no ta un proseso uniforme. Kada sélula ta trata e daño di manera diferente, kual ta dependé di nan organisashon interno (di e núkleo). Ademas, mi a deskubrí ku diferente pida ADN dañá ta bini huntu i ta forma grupo, kual ta permití reparashon di varios daño simultáneamente.

Huntu, mi resultadonan ta duna informashon nobo tokante e prosesonan fundamental ku ta mantené nos salú genétiko. Den futuro, e métodonan aki por determiná kon kada kanser ta reakshoná despues di sierto tratamentu, i asina habri kaminda pa medisina personalisá basá riba analisis individual.

Curriculum vitae

Kim de Luca (1990) was born and raised on Curaçao, former Netherlands Antilles, Southern Caribbean. After graduating *cum laude* from Peter Stuyvesant College, she moved to the Netherlands in 2008 to pursue a Bachelor's degree at University College Utrecht, focusing on molecular biology and environmental sciences, graduating *cum laude*. Her undergraduate studies included an exchange to UCLA and a thesis project in the laboratory of Edwin Cuppen at the Hubrecht Institute in Utrecht, under supervision of Sebastiaan van Heesch. They studied the effects of structural genetic variation on enhancer function, which continued as part of a Master's degree in Molecular and Cellular Life Sciences at Utrecht University. A subsequent move to the weird and wonderful Austin, Texas led to further research on histone variant exchange and transcriptional regulation in the laboratory of Vishy Iyer at the UT Austin Center for Systems and Synthetic Biology. Kim returned to the Hubrecht Institute as a PhD candidate in the group of Jop Kind, together establishing the lab and initiating several lines of research. This work led to the development of single-cell technologies for various applications in genome biology. During this time, Kim served a term on the staff association (PV), and on the Works Council (OC) for six years. To pursue further training in engineering and microfluidics, Kim was a visiting scholar in the lab of Aaron Streets (UC Berkeley), supported by competitive fellowships from EMBO and the Fulbright Commission. While living in Oakland, Kim was engaged with the grassroots social justice movement and continues to contribute remotely. Extensive volunteering at 113 Suicide Prevention, the Dutch national crisis hotline, has informed a personal ambition to meet others with compassion. In recognition of their contributions to single-cell genomics, particularly the work on DNA repair profiling, Kim was selected as one of MIT Technology Review's 2025 Innovators Under 35. As an avid diver and driven by concern for coral reef decline, Kim shifted research focus towards marine conservation. Kim currently holds a postdoctoral position in the lab of Iliana Baums at the Helmholtz Institute for Functional Marine Biodiversity, working at the intersection of advanced omics technologies, ecology, and coral reef resilience.

List of publications

* Co-first author
‡ Corresponding author

Manuscripts related to PhD trajectory, included in PhD thesis

Accepted

Chapter 4
Retrospective and multi-factorial single-cell profiling reveals sequential chromatin reorganization during X inactivation.
Kefalopoulou S*, Rullens PMJ*, **de Luca KL**, de Vries SS, Korthout T, van Oudenaarden A, Zeller P‡, Kind J‡.
Nature Cell Biology (2025).

Published

Chapters 5 and 6
Genome-wide profiling of DNA repair proteins in single cells.
de Luca KL*‡, Rullens PMJ*, Karpinska MA, Gacek-Matthews A, Pongor LS, Legube G, Jachowicz JW, Oudelaar AM, Kind J‡.
Nature Communications (2024).

Chapter 3
Single-cell profiling of transcriptome and histone modifications with EpiDamID.
Rang FJ*, **de Luca KL***, de Vries SS, Valdes-Quezada C, Boele E, Nguyen PD, Guerreiro I, Sato Y, Kimura H, Bakkers J, Kind J‡.
Molecular Cell (2022).

Chapter 2
Single-cell DamID to capture contacts between DNA and the nuclear lamina in individual mammalian cells.
de Luca KL, Kind J‡.
Methods in Molecular Biology (2021).

In preparation

Chapter 7
Epigenetic chromatin profiling with single-cell sequencing technologies.
de Luca KL, Kind J.

Manuscripts related to PhD trajectory, not included in PhD thesis

Published

Combinatorial single-cell profiling of major chromatin types with MAbID.
Lochs SJA*, van der Weide RH*, **de Luca KL**, Korthout T, van Beek RE, Kimura H, Kind J‡. *Nature Methods* (2023).

Nuclear chromosome locations dictate segregation error frequencies.
Klaasen SJ, Truong MA°, van Jaarsveld RH°, Koprivec I, Štimac V, de Vries SG, Risteski P, Kodba S, Vukušić K, **de Luca KL**, Marques JF, Gerrits EM, Bakker B, Foijer F, Kind J, Tolić IM, Lens SMA, Kops GJPL‡. *Nature* (2022).

Simultaneous quantification of protein-DNA interactions and transcriptomes in single cells with scDam&T-seq.
Markodimitraki CM*, Rang FJ*, Rooijers K, de Vries SS, Chialastri A, **de Luca KL**, Lochs SJA, Mooijman D, Dey SS‡, Kind J‡. *Nature Protocols* (2020).

Simultaneous quantification of protein-DNA contacts and transcriptomes in single cells.
Rooijers K*, Markodimitraki CM*, Rang FJ°, de Vries SS°, Chialastri A, **de Luca KL**, Mooijman D, Dey SS‡, Kind J‡. *Nature Biotechnology* (2019).

Genome-lamina interactions are established de novo in the early mouse embryo.
Borsos M*, Perricone SM*, Schauer T, Pontabry J, **de Luca KL**, de Vries SS, Ruiz-Morales ER, Torres-Padilla ME‡, Kind J‡. *Nature* (2019).

In preparation

Spatiotemporal tracing of DNA repair foci with DamID imaging.
de Luca KL, Bouke Bentvelsen, Kind J.

Other

Genomic and functional overlap between somatic and germline chromosomal rearrangements.
van Heesch S*, Simonis M*, van Roosmalen MJ, Pillalamarri V, Brand H, Kuijk EW, **de Luca KL**, Lansu N, Braat AK, Menelaou A, Hao W, Korving J, Snijder S, van der Veken LT, Hochstenbach R, Knegt AC, Duran K, Renkens I, Alekozai N, Jager M, Vergult S, Menten B, de Bruijn E, Boymans S, Ippel E, van Binsbergen E, Talkowski ME, Lichtenbelt K, Cuppen E‡, Kloosterman WP‡. *Cell Reports* (2014).