Sequential Catalysis of Histone H3K27 Methylation States in the Cell Cycle and Blood Development

Jacob Edward Greene

A dissertation

submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

University of Washington

2026

Reading Committee

Steven Henikoff, Chair

Andrew B. Stergachis

Kami Ahmad

Program Authorized to Offer Degree:

Molecular Medicine and Mechanisms of Disease

University of Washington

Abstract

Sequential Catalysis of Histone H3K27 Methylation States in the Cell Cycle and Blood Development

Jacob Edward Greene

Chair of the Supervisory Committee:

Steven Henikoff

Department of Genome Sciences

Polycomb domains form at silenced genes and are marked by histone H3 lysine 27 tri-methylation, but it is unclear how this methylation is maintained following DNA replication in proliferating cells. Here, I use CUT&Tag chromatin profiling to track the dynamics of H3K27 methylation in live human cells. In K562 erythroleukemic cells, I find that tri-methylation in Polycomb domains is maintained by stepwise methylation after DNA replication. Outside of Polycomb domains, thousands of inactive genes gain histone H3K27 di-methylation hours after DNA replication. Acute treatment with small-molecule inhibitors of the Polycomb Repressive Complex 2 (PRC2) histone methyltransferase slows H3K27 methylation during S-phase of the cell cycle and increases acetylation of the residue. By applying multi-factor, single-cell CUT&Tag chromatin profiling to developing blood cells, I observe that Polycomb-silencing at genes in the monocyte and B-lymphoid lineages proceed via a di-methyl intermediate to gain H3K27 tri-methylation. Together, My results explain how H3K27 modification patterns are faithfully duplicated in rapidly proliferating cells and indicate that developmental silencing generally initiates from an inactive chromatin state.

Table of Contents

Chapter 1. Introduction

Healthy multicellular development depends on scheduled gene regulation that allows one genome to generate diverse, stable, and responsive cell types. This occurs on chromatin fibers made of DNA packaged into nucleosomes by histone proteins. To influence transcriptional activity, histones are post-translationally modified. Histone H3 lysine 27 methylation is an essential feature of facultative heterochromatin that is dynamically regulated across lineages to maintain silencing of developmental genes. Through the methyltransferase activity of Polycomb repressive complex 2 (PRC2) and coordination by PRC1, H3K27 methylation accumulates to form Polycomb domains enriched for H3K27 di-methylation (H3K27me2) and tri-methylation (H3K27me3).

In this introduction, I review our understanding of PRC2 methyltransferase activity, which is fundamental to how the Polycomb system encodes transcriptional memory. First identified in forward genetic screens of *Drosophila* development, the Polycomb group proteins are repressors of developmental gene expression that are functionally conserved in mammals. Biochemical purification of these proteins reveals they function in complexes and explains PRC2 methyltransferase activity in terms of its structure and interactors. While PRC2 activity on reconstituted nucleosomes reveals the general catalytic principles underlying its rate of H3K27 mono-, di- and tri-methylation, chromatin profiling in live cells indicates that many factors and cellular rates, like proliferation, shape how H3K27 methylation is gained and lost. The current evidence suggests that H3K27 methylation by PRC2 is slower than a ~16-20-hour cell cycle. In line with domain models, slow H3K27 methylation explains how the Polycomb system can both maintain repressive domains and respond to transcription.

1.1 Discovering the Polycomb Group and its Link to Heterochromatin

The Polycomb group proteins were theorized to structurally encode a memory of gene activity following their 1949 discovery in forward genetic screens in *Drosophila* by Pamela Lewis. The namesake member, *Polycomb (Pc)*, was identified from mutant flies with altered segmentation phenotypes [1]. Later screens found additional genes that

produced related developmental defects and functionally defined the Polycomb group as negative regulators of homeotic gene expression, namely the *Hox* loci that control segment identity along the anterior–posterior axis [2]. When Polycomb function was lost, *HOX* genes were expressed outside their normal body position well after they usually turned off, resulting in striking homeotic transformations. Ed Lewis later proposed that Polycomb group proteins are maintenance factors which keep genes repressed across cell divisions, while another group, called Trithorax proteins, reinforce the opposite, active state [2]. This led to the idea that developmental regulation needs a memory system of transcriptional state. This memory function became a widely held hypothesis because it explained stable cell identity in terms of the stabilization of gene expression after the transient signals of development fade [3-5].

In terms of mechanism, two Polycomb group proteins were directly implicated in heterochromatin-mediated gene silencing as dominant modifiers of position effect variegation (PEV). In PEV experiments, a reporter gene that normally resides in euchromatic chromosome arms is translocated near the constitutively compact, dark-staining regions near centromeres, called heterochromatin. This produces mosaic expression of the reporter, and PEV modifiers alter that pattern. Overexpression of *Enhancer of zeste* (*E(z)*) and its mammalian ortholog increase PEV, as judged by silencing of the heterochromatin-associated *white* gene [6], while loss of function of *Suppressor of zeste 12* (*Su(z)12*) activates *white* [7]. *E(z)* and *Su(z)12* were unique in bridging homeotic gene silencing and heterochromatin-mediated silencing. This suggested the two phenomena might use similar molecular mechanisms [8] and motivated studying how *E(z)* and *Su(z)12* interact with the histone proteins that package DNA. Future studies would investigate the 1991 model put forth by Renato Paro [8] where homeotic gene silencing by Polycomb group proteins induced a heterochromatin-like state to prevent unscheduled activation of developmental regulators.

1.2 The Problem with a Histone Memory System

In the early 2000s, the Polycomb group proteins were isolated in multiprotein complexes and shown to covalently modify histones. Biochemical purification confirmed that Polycomb Repressive Complex 2, or PRC2, is an assembly of E(z), Su(z)12, Esc and

Nurf55 [9-11], with a conserved core organization from flies to mammals [12,13] by the EZH2/1, EED, SUZ12 and RBBP4/7 proteins. The E(z) and EZH2 orthologs contain SET domains previously shown to methylate target lysine residues on histones [14], and this motivated study of the purified PRC2 complexes *in vitro* methyltransferase assays to test their enzymatic activity. PRC2 specifically tri-methylates H3K27 [9-12], inspiring the hypothesis that this signal lets PRC2 recruit PRC1 to gene-rich regions to form heterochromatin [15]. In line Paro's model [8], the H3K27me3 mark thus provided a physical way for PRC2 to transmit memory on chromatin and signal to other repressive molecules. This was further confirmed by observation of PRC1 chromodomain proteins binding to H3K27me3 residue [16].

However, this histone-based system challenged contemporary theories of memory maintenance, like the Holliday–Pugh and Riggs model in which faithful inheritance relies on even splitting of a molecule, like DNA methylation into two hemi-methylated duplexes, to transmit memory after replication [17-19]. The problem was that histones offer no such template because H3–H4 tetramers do not split evenly during chromatin replication [20,21]. Therefore, on both strands many newly assembled histones will not carry the mark. Conserving the H3K27me3 mark across cell divisions required a different mechanism than DNA methylation to transfer the mark from the modified parent nucleosome to a new, unmodified one.

1.3 The Structural Basis of PRC2 Catalysis

The solution was a positive-feedback system that could read the H3K27me3 mark from one nucleosome and write it onto a neighbor, [22,23]. Visualizing PRC2 at the replication fork provided the molecular mechanism [24]. In the structure of PRC2 subunit EED, the WD40 β-propeller presents an aromatic cage that binds the H3K27me3 mark and increases catalysis of H3K27 tri-methylation. In methyltransferase assays on reconstituted nucleosomes, addition of H3K27me3 peptide raises PRC2's kinetic rate 7x, and this allosteric activation was abolished by mutating the EED cage [25]. Without the peptide, the tri-methyl step is intrinsically slow and addition of one methyl group to an unmodified, mono-, or di-methylated residue occurs at a frequency of 9:6:1, such that adding the third methyl is about 9-fold slower than the first [26]. A nucleosome already

carrying H3K27me3 can thus activate PRC2 catalysis on its neighbor, which propagates the mark. This dependence on EED also explains why EZH2 is inert alone: its SET domain stays closed until assembly with EED and SUZ12 opens the catalytic site [27,28].

An atomic model of an active PRC2 complex revealed the architecture of its allosteric switch. The complex folds into two lobes bridged by EZH2 [29]: (1) a catalytic lobe holding the EZH2 CXC and SET catalytic domains braced by SUZ12, and (2) a regulatory lobe built around the *EED* β-propeller, whose top face presents an aromatic cage that clamps the tri-methylated H3K27 residue [25,30]. Without allosteric activation, the *EZH2* stimulation-responsive motif (SRM) is disordered. When H3K27me3 occupies the *EED* cage, the SRM folds into an α-helix and packs against the SET-I subdomain, ordering the active site for catalysis [29,31]. At the active site, the substrate lysine must thread into a narrow aromatic channel, and each methyl addition enlarges its ε-amino tip, crowding the active site. The channel size can therefore constrain the tri-methyl step [29]. Indeed, an A677G substitution in EZH2 sufficiently enlarges the channel to facilitate addition of the third methyl group [32], showing how active site structure shapes PRC2 catalysis. This explains the relative rates of mono-, di-, and tri-methylation in terms of active site conformation. These changes at the active site for successive methyl additions are consistent with the distributive catalysis by PRC2 [33,34].

PRC2 allosteric activation explains not only the maintenance of the H3K27me3 mark but also the nucleation and spreading dynamics shown to build Polycomb domains [35]. Nucleation begins with recruitment. In *Drosophila*, the sequence-specific proteins Pleiohomeotic (Pho) and Polycomb-like (Pcl) bind Polycomb response elements and tether PRC2 to these target sites [36,37]. In mammals, an expanded set of accessory subunits both direct PRC2 recruitment and activate the enzyme (reviewed in [38,39]). For example, JARID2 tri-methylated at K116 mimics the H3K27me3 interaction with the EED cage [40-42]. Two assemblies of PRC2 accessory subunits read different features: (1) PRC2.1 through Polycomb-like subunits that bind unmethylated CpG DNA [43], and (2) PRC2.2 through the PRC1-catalyzed H2AK119ub1 mark [44]. Both structurally converge on the same allosteric mechanism to activate catalysis [45]. Likewise, spreading is shown by the structure of PRC2 on chromatin bridging two nucleosomes, simultaneously

reading the tri-methylated tail of one at the EED cage and engaging the unmodified tail of its neighbor via the EZH2 SET domain [46,47]. Whether the same geometry is possible at the replication fork is unresolved. While EZH2 in mouse embryonic stem cells (ESCs) co-precipitates with the p150 subunit of replication-coupled histone chaperone CAF-1 [48], no structure has yet resolved the PRC2 interaction with DNA replication machinery. These structures reveal how allosteric activation is wired into PRC2, not how fast the complex works on chromatin in live cells.

1.4 Mapping H3K27me3 Dynamics to the Cell Cycle

How fast PRC2 catalyzes H3K27me3 after replication can be read from chromatin profiling across the cell cycle, and the answer is slow. The fork-coupled model predicts that PRC2 reads the parental mark and writes it onto adjacent new histones, restoring each domain within S phase [24]. Two similar methods test this on nascent DNA using double-thymidine block to first synchronize cells in G1 phase, which perturbs normal cell-cycle progression, gene regulation, and potentially H3K27me3 levels [49]. To trace recovery of the H3K27me3 mark after replication, Chromatin Occupancy after Replication (ChOR-seq) tags newly replicated DNA with a pulse of the thymidine analogue EdU, immunoprecipitates H3K27me3-marked chromatin, and sequences the EdU-labelled DNA with a chromatin spike-in to expose differences between timepoints [50]. To reveal how the mark segregates on the two daughters, Sister Chromatids After Replication (SCAR-seq) adds a strand-resolution step which shows that H3K27me3-marked parental histones recycle randomly [51]. This symmetric recycling passes the whole domain pattern onto the new DNA as a coherent unit [50]. The PRC1-catalyzed H2AK119ub1 mark reinforces this separately from H3–H4 tetramer via independent recycling of H2A–H2B heterodimers that can be read via JARID2 [52]. The pre-replication level of the H3K27me3 mark is gradually rebuilt during S phase and the subsequent G1 phase. Recovery is fastest at the center of domains with the greatest PRC2 and H3K27me3 occupancy, and slowest at their borders [50]. These rates track with chromatin density and are accelerated by histone H1-mediated compaction [53]. Together these dynamics suggest that domain maintenance is not copied at the fork and is slow and post-replicative.

If PRC2 catalysis is slow, then the H3K27me3 mark is diluted each cell cycle. DNA replication doubles the histone pool with unmodified H3, halving H3K27me3 at every division. Quantitative proteomics using heavy methionine to label new proteins and methyl groups shows that overall levels of H3K27 di- and trimethylation are halved at the fork and re-established slowly, over more than one cell-cycle in HeLa cells [54]. In mouse intestinal cells that conditionally excise the *EZH2* or *EED* gene, chromatin profiling shows that H3K27me3 falls proportionately each division. Based on their amount of histone methylation, genes in domains remain off until successive divisions dilute the residual mark below a threshold [55,56]. Human lymphoma cells require several cell divisions before dilution relieves repression [55]. Restoration of the H3K27me3 mark therefore is slower than the replication that opposes it and naturally extends into the next cell cycle.

Based on dilution of the H3K27me3 mark in S-phase, recovery of it by PRC2 might continue into the subsequent G1 phase. Indeed, single-cell profiling of the mark in exponentially cycling in mouse ESCs shows it recovers mostly during the 2 hours of G1 phase [49]. Lengthening G1 phase by either double-thymidine block to inhibit DNA synthesis or culture conditions that restores the G1 checkpoint in these cells raise global H3K27me3 levels [57]. A similar approach works in H3K27M-mutant glioma cells that fail to normally spread the H3K27me3 mark beyond nucleation sites. In this context, blocking G1-S transition with the CDK4/6 inhibitor palbociclib induces *de novo* domain formation and spreading [57]. Likewise, shortening G1 in human HEK293 cells with a Chk1 inhibitor overrides the G1-S checkpoint and lowers global H3K27me3 levels [57]. In mouse ESCs, PRC2 may be tethered to certain nucleation sites in S and G2 phase and then released during G1 phase to facilitate spreading [58]. This interpretation aligns with the consistent rate of H3K27 tri-methylation observed across cell cycle phases in HeLa cells by ChOR-seq [50]. The mechanism explaining how PRC2 catalysis might increase during G1 phase is currently unknown, and it is possible this simply offers a window when replication no longer dilutes the H3K27me3 mark.

Alternatively, PRC2 may itself control the G1-S transition by regulating the *INK4A-ARF* locus that forms a Polycomb domain [59]. Loss of H3K27me3 mark here triggers cell-cycle

arrest by blocking G1/S transition in two ways: (1) $p16^{INK4A}$ inhibits the CDK4/6 kinases, and (2) $p14^{ARF}$ stabilizes p53 [59]. In rapidly cycling cells, dilution of the H3K27me3 mark at *INK4A-ARF* could activate the G1/S blockade to slow cell cycle progression and promote domain recovery [57,59]. Slow PRC2 tri-methylation may therefore be an integral part of this and possibly other [60]homeostatic loops that regulate cell cycle speed . This raises the question of whether Polycomb domains self-reinforce or rely on nucleation sites to tether PRC2 and limit dilution.

1.5 Maintaining the H3K27me3 Mark Beyond Nucleation

To test whether Polycomb domains rely on sequence to tether PRC2 and limit dilution of the H3K27me3 mark, two research groups designed systems in *Drosophila* with removable Polycomb response elements (PREs). To measure dilution of the H3K27me3 mark after PRE removal, a silenced reporter is positioned next to a removable PRE, and its excision yields a domain without PREs. When cell division is blocked, the domain persists and the reporter is off, suggesting that de-methylation does not oppose domains [61]. However, under normal cell division in the developing wing disc, the reporter activates in step with dilution of the H3K27me3 mark. After PRE removal, Laprell, Finkl and Müller found the H3K27me3 mark dilutes by 50% each cell division and the reporter rapidly activates. This lead to the conclusion that copying the mark requires sequence-specific recruitment of PRC2 [61]. Coleman and Struhl found that dilution is only ~10-12% per cell division such that silencing survives for ~6 divisions after PRE removal. This suggest that the H3K27me3 mark itself can limit dilution and carry memory, but not indefinitely [62].The dilution rates were observed using similar, but distinct 10-12 kb transgene systems that might be explained by different levels of PRC2-tethering at cryptic sites nearby [62]. Both studies agree that sequences-specific tethering of PRC2 is essential for long-term domain maintenance.

Interestingly, reporter reactivation was not spatially uniform across the wing disc [62]. Coleman and Struhl attribute this variegation to the number of divisions since PRE removal. Indeed, removing the PRE earlier after egg laying increases reporter-positive cells and removing it later decreases them. Reactivation across the wing-disc is also higher in more proliferative zones, like the developing wing blade, and lower in less

proliferative zone closer to the body wall. This indicates that the memory-maintenance by the H3K27me3 mark differs across tissues based on proliferation rate.

Based on size alone, nucleation is poorly suited to the job of domain maintenance. The H3K27me3 mark covers broad Polycomb domains spanning hundreds of kilobases, while the elements that recruit PRC2 are 100 base-pairs to 1.6 kilo-bases in size [61,63]. A mark confined to nucleation sites could never blanket a domain, so most of it must be spread and held between nucleation sites, leading to the peaks-and-valleys density distribution characteristic of domains in rapidly cycling cells. For example, mouse ESCs grown in serum cycle every 11-12 hours and lean heavily on re-nucleation each cycle to re-build domains that reach kilobases past those sites [64].

In terms of chromatin on-off rates, the factors responsible for nucleation and PRC2 occupancy differ sharply. The PRC2 core complex engages nucleosomes rather than a defined DNA sequence [46]. Single-molecule tracking shows transcription factors sample their target sequences in brief, repeated visits [65], and this restless occupancy that drives the nucleosome turnover at regulatory elements, including *Drosophila* PREs [66]. After the fork passes, the PRE nucleation factor Pho re-occupies its sites hours later, much slower than other factors [67]. A nucleation factor that operates this slowly makes a poor anchor for fast, faithful domain maintenance after the DNA replication.

1.6 Polycomb Domains are Locally Maintained

Nucleation and spreading might be temporally distinct parts of the Polycomb system. Across heterochromatin systems from fission yeast to human, domain establishment and maintenance are mechanistically separable: sequence-specific factors nucleate the mark at defined sites, it spreads by self-propagation, and spreading is opposed by inhibitory factors that set the domain boundaries [68]. Polycomb domains follow the same logic. For example, accessory factors that mimic the H3K27me3 residue tether PRC2 to domains and removing them initiates spreading beyond domain edges [69]. Within an established domain, maintenance is also local. In mouse ESCs, proximity labelling that marks parental nucleosomes and follows them through S phase shows that Polycomb domains are preserved by re-depositing marked nucleosomes near their original

position [70]. In contrast, nucleosomes in active regions are re-deposited kilobases away [70]. What passes to the daughters is therefore most of the parental domain, broken only by the new, unmodified nucleosomes inserted after replication.

PRC2 fills the gaps in domains locally. Single-molecule imaging of PRC2 on reconstituted chromatin and free DNA shows that the enzyme preferentially binds free DNA, dwells there for seconds at a time, and draws most of its chromatin affinity from the DNA wrapped around histones rather than histone contacts [71]. In this *in* vitro system, PRC2 residence time on chromatin increases by engaging the Polycomblike-related co-factor PHF1 [71]. If this carries over to the nucleus, a Polycomb domain would hold a concentrated, long-dwelling pool of PRC2 and, because dense nucleosome arrays activate catalysis [72], that pool would stay active wherever the domain stays compact. PRC2 can likely traverse the gaps in domains by local diffusion, or “hopping”, between neighboring nucleosomes [73]. Because this depends on the local density of mark and enzyme complex rather than the passing fork, gap-filling can occur anywhere in a domain, at any point in the cycle. What is inherited, then, is a whole self-propagating region rather than the state of any single nucleosome. As formal models suggest, the Polycomb domains are the unit of memory, not individual nucleosomes [22].

1.7 Modelling H3K27 Methylation Dynamics in Domains

Before empirical observations defined molecular workings of the Polycomb system, formal modelling predicted that histone-based memory leverages domains that are more than the sum of induvial nucleosomes. To ask in mathematical terms whether positive-feedback within a domain limits dilution, the Dodd proposal in 2007 modelled a 60-nucleosome domain with three possible nucleosome states: methylated (repressed), unmodified, or acetylated (active) that interconvert in this order, one step at a time. The conversion probability is set by a parameter that models feedback-to-noise as a function of mark frequency in the domain. Replication is modeled by resetting 50% of the nucleosome to the unmodified state. After replication, the region is bi-stable and then settles into a mostly methylated or mostly acetylated state based on two related conditions. First, the feedback must be cooperative because it depends on more than one like-marked nucleosome to bias the outcome. Second, conversion must reach

beyond adjacent nucleosomes because conversion weighted on neighbors alone cannot hold the state. Even though the Dodd model is based on *S. pombe* that lack H3K27 methylation, it suggests that the PRC2 read-write mechanism can carry a heritable state across replicative dilutions, but only at the domain scale.

To incorporate the empirically-derived H3K27 methylation dynamics that form Polycomb domains, Berry, Dean, and Howard used a four-state approach to model each H3 histone as me0 through me3, with me0 and me1 treated as neutral and me2 and me3 as repressive [74]. The conversion rates between states is based on the allosteric activation of PRC2 by the H3K27 di- and tri-methyl marks [25] and time-resolved mass spectrometry data of slow methyl addition on H3K27 over multiple cell cycles [54,75]. To model loss of H3K27 methylation, the model incorporates transcription, where each passage of RNA polymerase II removes a methyl group and turns over some nucleosome to the me0 state. The rate of transcription is limited by the density of the me2/me3 states. The dynamics of a 30-nucleosome domain are simulated by probabilistically drawing each methylation, de-methylation, and transcription based on the current methylation landscape in the context of periodic 'replication' that reset half the nucleosomes to the me0 state. Consistent with the simpler 3-state Dodd model, the 4-state model predicts the domain is bi-stable, but this is contingent on slow methylation kinetics to buffer brief pulses of transcription. Therefore, only sustained transcription converts the domain to an active state. If the methylation rate increases, the domain becomes static. Thus, slow methylation by PRC2 is not a limitation, but likely an evolved mechanism for Polycomb domains to both repressively filter transcriptional noise and respond to sustained transcriptional activation [76].

These earlier models failed to capture the transient, hit-and-run kinetics of PRC2 and separate it from PRE-directed methylation. To add the spatial dimension of Polycomb domains, Lundkvist, Lizana, and Schwartz built a Monte Carlo simulation that inspects a nucleosome each round to decide its next methylation state. If a nucleosome is next to a PRE, or adjacent to a H3K27 di- or tri-methylated nucleosome, the probability of methylation increases. Modelling a 24-hour cell cycle with periodic two-fold dilution and PRC2 interaction every 15 seconds reproduces the H3K27me3 chromatin profiles at

PRE-equipped loci. This shows that the spatial distribution of each H3K27 methylation state depends on (1) nucleation site position, (2) the local concentration of the H3K27me3 mark and (3) the replication frequency. Because H3K27me3-marked nucleosomes allosterically stimulate PRC2, the mark concentrates near PREs rather than spreading across the genome. In the early *Drosophila* embryo, the model finds that rapid syncytial cleavage divisions dilute the inherited mark faster than PRC2 can restore it. This indicates that maternally-deposited H3K27me3 is unlikely to transmit the memory of repression from the germline to the offspring [77]. What survives rapid division is not the H3K27me3 mark, only the capacity to re-nucleate at a PRE and, perhaps, H3K27 mono- and di-methylated nucleosomes.

The Lundkvist model has since been used to test whether the H3K27me2 mark lowers to the barrier to catalyze H3K27me3 [78], as suggested by distributive PRC2 catalysis. Degen and colleagues removed PRC2 hit-and-run catalysis from the original model. In their updated model, PRC2 catalysis increases only near the H3K27me3 mark or PREs marked by E(z) and Pho*,* which is absent until nuclear cycle 10 (NC10) during *Drosophila* embryogenesis [79]. The model is also tuned by live-imaging measurements of nuclear E(z) and Pho. Based on chromatin profiling, Polycomb domains do not appear until ~110 minutes later in NC14 [78], so they predicted each H3K27 methylation state across the first 16 cell cycle dilutions. On the maternal chromosome, which is pre-loaded with H3K27 di- and tri-methylation from the oocyte[77], the model estimates complete dilution of the H3K27me3 mark by NC6, but allostery maintains H3K27 mono- and di-methylation across the 10-minute cleavage divisions that occur until NC8, when the zygotic genome activates. Meanwhile, on the unmarked paternal chromosome, no methylation appears until *Pho* nucleates it. Domains occur on both chromosomes at the same time, and these predictions line up with immunostaining for H3K27 di- and tri-methyl states in staged embryos. This suggests that the H3K27me2 mark is transmitted across generations to potentially seed domains independent of hit-and-run PRC2 methylation, but does not answer if this occurs via specific H3K27me2-reader binding [80].

1.8 Resolving Sequential H3K27 Methylation Dynamics

The role of H3K27 mono- and di-methylation remains unclear in Polycomb domain formation, memory maintenance, and transcriptional regulation. These two lower methylation states are likely the products of hit-and-run catalysis, which explains their broad genomic coverage and high frequency on chromatin. For example, H3K27me2 modification occupies 50-70% of histone H3 tails in chromatin[81,82]. It is possible that the H3K27me2 mark, which covers many transcriptionally inactive genes, may reinforce transcriptional repression by binding to chromodomains in repressors, like PRC1 [16], or by blocking histone acetylation [83,84].

Despite the modelling approaches that predict the conversion dynamics between H3K27 methylation states, we still do not understand the actual dynamics of each mark across the genome. Our understanding of H3K27 methylation dynamics is mostly limited to early embryogenesis in *Drosophila* and mouse ESCs, in part because both provide homogenous systems for time-series measurements. Single-cell chromatin profiling provides an avenue to understand H3K27 methylation dynamics in the complex tissues that arise later in development, and, in cases of Polycomb impairment, can give rise to disease [85,86].

In this thesis, I use CUT&Tag chromatin profiling [87] to track the dynamics of H3K27 methylation states during the cell cycle and in human blood development. In both contexts I find that H3K27 tri-methylation is catalyzed sequentially via a di-methyl intermediate state. During the cell cycle, sequential PRC2 catalysis occurs on the time-scale of hours [54,88]. Outside Polycomb domains, the H3K27me2 mark accumulates on transcriptionally inactive chromatin and does not proceed to tri-methylation. Unlike Polycomb domains, H3K27me2-marked regions do not repress integrated reporter sequences [88]. If PRC2 catalysis is slow, sequential, and samples inactive chromatin, then this supports that H3K27me3 is catalyzed on inactive chromatin [83,89-91].

Chapter 2. Histone H3K27 methylation dynamics during the cell cycle

This chapter is adapted from:

Greene, J. E., Ahmad, K. & Henikoff, S. Histone H3K27 methylation states are sequentially catalyzed in cycling cells. *bioRxiv*, 2026.2004.2030.721988 (2026).

2.1 Introduction

How are H3K27 methylation patterns in the genome propagated through chromatin duplication and cell division? In plants, the ATXR5/6 methyltransferases associate with the DNA replication fork and catalyze H3K27 methylation on newly deposited histones. This restores tri-methylation patterns by the end of S-phase[92]. In mammals, the EZH2 methyltransferase subunit of PRC2 co-precipitates with the replisome subunit PCNA and the chromatin assembly factor 1 (CAF1) that deposits new histones after the DNA replication fork[48]. Studies in mammals suggest that PRC2 catalysis at replication forks[24] maintains the H3K27me3 mark by propagating it onto new, unmodified histones. However, this model has been questioned by the slow catalysis of the mark, which is not recovered during S-phase after DNA replication[50,54,75,93,94].

In this study we use CUT&Tag chromatin profiling to track the dynamics of H3K27 methylation states during the cell cycle in human K562 cells. Comparison of H3K27 methylation dynamics to replication timing, the H3K27ac modification, and active RNA polymerase II (RNAPII) reveals that Polycomb-silenced domains are transiently marked by the H3K27me1 and H3K27me2 modifications as stepwise methylation restores H3K27me3 patterns after DNA replication. Outside of Polycomb domains, thousands of active genes gain the H3K27me1 mark after DNA replication, while genes with weak promoter activity proceed to H3K27me2. We find that sequential catalysis of H3K27 methylation is slow at genes that replicate in early S-phase, and fast later in S-phase. These replicative patterns of H3K27 mono- and di-methylation are catalyzed by PRC2, as they are lost upon either catalytic or allosteric inhibition of the complex. Upon

inhibition, chromatin accumulates H3K27 acetylation. Finally, we show that the H3K27me2 mark does not repress transcription, which suggests a passive process at inactive genes. Together, our results explain how H3K27 modification patterns are faithfully duplicated in proliferating cells.

2.2 Results

2.2.1 The steady-state distributions of histone H3K27 modifications in K562 cells

To determine a baseline for analyzing the cell cycle dynamics of the three H3K27 methylation states, we characterized their steady-state distributions in the human erythroleukemic K562 cell line. We performed profiling of mono-, di-, and tri-methylation via CUT&Tag[87] chromatin profiling using highly specific antibodies validated on synthetic nucleosomes[95] (Supplementary Fig. 1). To identify active genes, we also profiled histone H3K27 acetylation (H3K27ac), and RNA Polymerase II (RNAPII) phosphorylated at serine 2 and/or 5 (active RNAPII). As expected, large >10-kilobase (kb) regions are enriched for the H3K27me3 modification, consistent with the annotation of Polycomb domains in K562 cells[96,97] (Fig. 1A, Supplementary Fig. 2). For example, a genomic region on Chromosome *6* shows an H3K27me3-marked Polycomb domain encompassing four genes next to two active genes (Fig. 1A). In this region, the Polycomb domain lacks H3K27me1 and H3K27me2 modifications, while the two active genes are marked with the H3K27me1 modification and little H3K27me2 modification. In contrast, the intervening regions are enriched with the H3K27me2 mark. Thus, these three methylation states of the histone H3 K27 residue mark largely distinct regions.

To determine the relationship of each H3K27 modification with active RNAPII, we displayed their average signals at protein-coding genes (Fig. 1B). As expected, genes in Polycomb domains exhibited low active RNAPII. Similarly, genes enriched for the H3K27me2 mark are characterized by low RNAPII signal. In contrast, the H3K27me1 and H3K27ac modifications are enriched at active genes[98]. Our finding that di- and tri-

methylated H3K27 vary inversely with active RNAPII at genes is consistent with previous observations in mouse embryonic stem cells[81,84].

2.2.2 Stepwise methylation of the histone H3K27 residue in Polycomb domains

For histone modifications to be stably maintained in the epigenome they must be duplicated every cell cycle after DNA replication. Chromatin duplication occurs via sequential events: (1) old, modified histones are distributively segregated to daughter strands, (2) new, unmodified histones are deposited on the strands to complete chromatin assembly, and (3) enzymatic catalysis modifies the new histones. Studies suggest that H3K27 tri-methylation occurs slowly after DNA replication[24,50,54,93,94], but catalysis of the H3K27me1 and H3K27me2 marks during the cell cycle have not been examined. We therefore set out to resolve the cell cycle dynamics of all three methylation states in K562 cells.

We first asked how the steady-state levels of the H3K27 modifications correlate with transcription and timing in S-phase when a locus replicates. To compare replication timing with H3K27 modifications and active RNAPII, we visualized the density of 10-kb windows enriched for each mark along a continuous axis from early- to late-S-phase (Supplementary Fig. 3A-B). Windows enriched for H3K27me3 mostly replicated in mid-S-phase, although there are both earlier- and later-replicating windows. In contrast, H3K27me1-enriched windows predominantly replicated at the beginning of S-phase, as did windows enriched for H3K27ac and active RNAPII. Finally, H3K27me2-enriched windows replicated throughout S-phase. Thus, regions enriched for the H3K27me3, H3K27me2, and H3K27me1 marks exhibited distinct replication timing (Supplementary Fig. 3B). Our findings are consistent with the association of early replication timing and transcriptional activity[99].

To separate cells during S-phase for CUT&Tag chromatin profiling of H3K27 methylation, we isolated 6 fractions of cells by fluorescence-activated cell sorting (FACS) on DNA content, where each fraction represents about 2.5-3 hours of the cell

cycle (Fig. 1C). We then profiled each H3K27 residue methylation, the H3K27ac modification, and active RNAPII in each fraction. We first analyzed the 2,855 genes in Polycomb domains because new histones deposited there after DNA replication must become tri-methylated. We ranked genes based on their replication timing[100], assigned them to six bins from early to late S-phase, and then displayed their chromatin profiling signals in FACS-isolated fractions (Fig.1D). We expected genes in Polycomb domains to undergo two-fold dilution of the H3K27me3 mark via distributive histone segregation and then recover the mark as PRC2 methylates new histones. Indeed, early-replicating Polycomb domains have their lowest H3K27me3 signals in S1 and S2 fractions, while late-replicating Polycomb domains have their lowest signals in S4 and S/G2 fractions. This low methylation persists for 1-2 fractions after replication, indicating a delay of at least ~5-6 hours before new histones are tri-methylated. This delay appeared consistent regardless of when genes in Polycomb domains replicate, with early-replicating Polycomb domain genes starting to recover tri-methylation by late S-phase and late-replicating domains beginning tri-methylation in the next cell cycle (Supplementary Fig. 4).

Polycomb domains have low mono- and di-methylation in steady-state populations (Fig. 1A). Strikingly, these marks transiently increase during S-phase in FACS-isolated fractions (Fig. 1D). Mono-methylation appears as a transient wave just as genes replicate and then rapidly diminishes (Fig. 1D, Supplementary Fig. 4). Di-methylation appears slightly later by 1-2 fractions after the mono-methylation wave and then diminishes as well. These patterns are simply explained by distributive catalysis by PRC2 in Polycomb domains, where: (1) mono-methylation occurs rapidly on new histones behind the replication fork, (2) mono-methylation is converted to di-methylation, and (3) eventually, di-methylation is converted to tri-methylation.

2.2.3 Stepwise methylation at genes outside Polycomb domains

In steady-state populations, the H3K27me2 modification occurs outside Polycomb domains (Fig. 1A-B), so we next set out to analyze the H3K27 methylation dynamics in these regions. Genes outside Polycomb domains were categorized based on the most

abundant H3K27 modification above a minimum signal threshold (Supplementary Table 1). This classification schema identified 2,268 H3K27me2-, and 3,872 H3K27me1-enriched genes (Supplementary Fig. 5A-B). These H3K27me2- and H3K27me1-enriched genes are not targets of PRC2 or PRC1 subunits SUZ12, CBX2 and CBX8 (Supplementary Fig. 5C), suggesting that they result from untargeted catalysis. The H3K27me1 genes were enriched for MYC, RNAPII, EP300, and other general transcription factors (Sup Fig. 4C). In contrast, H3K27me2-enriched genes were not significantly enriched for any of the 116 factors profiled in K562 cells by the ENCODE consortium, suggesting that they are inactive[96,101].

We hypothesized that stepwise catalysis also occurs outside domains but does not achieve tri-methylation. To test this hypothesis, we displayed the signal of each mark at H3K27me2-enriched genes outside domains and grouped them by replication timing (Fig. 2A). The stepwise gain of mono- and di-methylation was recapitulated at genes enriched for H3K27me2 (Fig. 2A). In general, tri-methylation did not appear to follow from di-methylation at H3K27me2 genes as in Polycomb domains (Fig. 2A, Supplementary Fig. 6). The H3K27me1-enriched genes also gained mono-methylation, but it was not converted to di-methylation in the next fraction (Fig. 2B). Therefore, mono- and di-methylation of the H3K27 residue occurs after DNA replication at thousands of genes outside of Polycomb domains.

2.2.4 Increased catalysis of H3K27 methylation in late S-phase

To compare the dynamics of H3K27 methylation at early- versus late-replicating genes, we grouped Polycomb domain and H3K27me2-enriched genes by replication timing and visualized the complete cycle of H3K27 methylation as the magnitude of change (ΔRPKM) between S-phase fractions (Fig. 3). To resolve conversion of the H3K27 methylation states, we compared the change between marks (Fig. 3A). For example, a clockwise cycle indicates conversion of mark X (x-axis) to mark Y (y-axis). Compression of the x-axis only indicates fast conversion of mark X to mark Y. Compression of the y-axis indicates less catalysis of mark Y, and compression of both axes indicates less change in both marks overall. Lastly, a counterclockwise cycle indicates conversion of

mark Y to mark X. We observed each of these by comparing the H3K27me2 mark to H3K27me3 and H3K27me1 (Fig. 3B-C, top and bottom row, respectively).

At Polycomb domain genes that replicate in early S-phase (Fig. 3B i), we observed well-defined changes in H3K27 methylation catalysis between the G1/S and the S1 fractions (Fig. 3, yellow). During this interval in aggregate, tri-methylation decreased by -0.7 ΔRPKM while mono- and di-methylation increased by +0.5 and +0.15 ΔRPKM, respectively. This was followed after the S1 fraction (Fig. 3, orange) by the loss of mono-methylation (-0.2 ΔRPKM) and gain of large amounts of di-methylation (+1 ΔRPKM) which began to convert to tri-methylation (+0.15 ΔRPKM). Tri-methylation then increased by +0.8 ΔRPKM while di-methylation dropped by 0.5 ΔRPKM from the S2 to the S3 fraction (Fig. 3, peach). Mono-methylation continued to decrease by -0.2 ΔRPKM from S3 to the S4 fraction (Fig. 3, magenta) while di-methylation also decreased (-0.5 ΔRPKM) and tri-methylation increased by +0.2 ΔRPKM. Catalysis of tri-methylation then ceased after the S4 fraction (Fig. 3, light purple). Both di- and tri-methylation rapidly decreased by -0.5 ΔRPKM from the S/G2 to the G1/S fraction (Fig. 3, dark purple), thus forming a complete cycle, consistent with stepwise catalysis.

Notably, early replicating genes in Polycomb domains (Fig. 3B i) exhibited significantly greater loss and gain in H3K27 tri- and in di-methylation than genes that replicated later in S-phase (Supplementary Fig. 7a i-ii). At late-mid replicating genes (Fig. 3B iv) H3K27me1 dynamics were not apparent. These changes suggest that the catalysis of H3K27me3 is slow in early S-phase but rapid in late S-phase. In general, this shift toward H3K27me3 catalysis coincided with S-phase progression as the change in all marks diminished (Fig. 3B, Supplementary Fig. 7A). Our analysis was inconclusive for late-replicating genes (Fig. 3B vi) due to low sample size.

At the H3K27me2-enriched genes outside Polycomb domains, we were able to distinguish the conversion of H3K27 mono- to di-methylation (Fig. 3C). These two marks formed a dynamic cycle as in Polycomb domains but conversion of H3K27 di- to tri-methylation was limited. As in Polycomb domains, we found a drop in the overall change of H3K27me2 and other marks after early S-phase (Supplementary Fig. 7B).

We conclude that overall H3K27 methylation proceeds slowly in early S-phase and speeds up in late S-phase.

2.2.5 PRC2 inhibition slows H3K27 methylation in the cell cycle

If stepwise catalysis is responsible for the transient appearance of H3K27me1 and H3K27me2 marks in Polycomb domains, then chemical inhibition of the EZH2/1 methyltransferase subunit of PRC2 should ablate these marks during the cell cycle. To test this, we treated cells for 8 hours with small molecule inhibitors of PRC2 followed by S-phase fractionation and CUT&Tag profiling (Fig. 4A). Since the cell cycle of K562 cells is ~20 hours[102] and fractions are isolated from the same parent population, we can detect temporal drug effects by comparing count-normalized changes between early- and late-replicating genes in fractions. For example, in the S1 fraction only early-replicating genes acquire new histones in the presence of the drug. Only these genes should be affected if a mark is catalyzed in step with DNA replication and those genes should be unaffected in later fractions.

We treated cells with either EPZ-6438, a catalytic inhibitor of the EZH2/1 methyltransferase subunit[103], or with EED-226, which binds the EED subunit responsible for allosteric activation of the complex[104]. We observed a striking delay of ~4 fractions for the H3K27me1 and H3K27me2 marks in Polycomb domains in both treatment conditions compared to the stepwise catalysis in the DMSO control (Fig. 4B). In contrast, H3K27me3 catalysis appeared far less affected as measured by the maximum RPKM per gene and signal trend across the time course. Due to count normalization, this could result from inhibited global catalysis irrespective of when a locus replicates.

We analyzed genes in Polycomb domains and compared the effects of the two PRC2 inhibitors normalized to the DMSO control versus replication timing (Fig. 4C). The change in each S-phase fraction reflects the acute effects of inhibitors on H3K27 methylation for 8 hours preceding sample collection. We therefore hypothesized that H3K27 methylation changes in S-phase fractions will linearly relate to replicating timing if methylation follows in step with DNA replication (Fig. 4C). The magnitude of

association would be revealed in the slope of a simple linear model, where a value of zero indicates no association and a positive slope indicates early-replicating genes are relatively depleted for the mark, while a negative slope indicates late-replicating genes are depleted.

The effects of the two inhibitors were generally consistent across the H3K27 mono- and di-methyl marks, and genes were depleted of the marks during replication in the presence of the drugs (Fig. 4D, Supplementary Fig. 8, Supplementary Table 2). In contrast, the inhibitors had little effect on the tri-methyl mark (Fig. 4B,D). We therefore conclude that the transient H3K27 mono- and di-methylation following replication are acutely affected by both catalytic and allosteric inhibition of PRC2. We attribute the lack of effect on H3K27 tri-methylation to the slow catalysis of this mark, such that it is not achieved in the untreated cells in 8 hours.

We also found consistent effects of both drugs on active RNAPII and the H3K27ac mark at genes in Polycomb domains (Fig. 4B). Our analysis allowed us to temporally link changes in H3K27 methylation with shifts in H3K27ac and active RNAPII signals. The linear relationship between replication timing and the inhibitors' effects on active RNAPII were close-to-zero; these changes appear unrelated to replication timing (Fig. 4D). In contrast, the relationship with the H3K27ac mark was negative for both inhibitors, but only in the S2 and S3 fractions (Fig. 4C, Supplementary Fig. 8). At these time points, many genes will have replicated in early S-phase in the presence of the PRC2 inhibitors

To resolve exactly when the changes in the H3K27ac and H3K27me2 marks occurred at early replicating genes in Polycomb domains, we plotted the change in one mark versus the other in each fraction (Fig. 4E, left). The time when the H3K27ac mark increased was not consistent across the two inhibitor treatments. Under EPZ-6348 treatment, no significant gain of the H3K27ac mark occurred during the interval from the S1 to the S2 fraction (Fig. 4E, upper left, Supplementary Fig. 9A i). Rather, the gain in acetylation occurred one fraction earlier during the interval from the G1/S to the S1 fraction before the H3K27me2 mark was lost. Under EED-226 treatment, an increase of H3K27ac by 1.4x occurred in step with the loss of the H3K27me2 mark (Fig. 4E, bottom

left). To control for any loss of the H3K27me3 mark in Polycomb domains, we analyzed H3K27me2-enriched genes outside Polycomb domains and observed similar timing (Fig. 4E, right column). These patterns were clear at early replicating genes, but not at later replicating genes (Supplementary Table 3). We therefore conclude that catalytic inhibition of PRC2 indirectly allows the acetylation of the H3K27 residue at early replicating genes, and allosteric inhibition does so more slowly.

2.2.6 H3K27me2-marked regions are inactive

We next sought to define the functional relationship between the H3K27me2 mark and gene transcription. At steady-state, this methylation mark is anticorrelated with active RNAPII in genes (Fig. 1B), so we asked whether the mark is repressive *per se* by querying promoter activity in three contexts in K562 cells (Fig. 5A). First, as a measure of promoter activity in the native chromatin context, we used data obtained with the GRO-cap (global run-on sequencing with 5′ cap selection) method[105] (Fig. 5B). Second, to measure autonomous promoter activity when the DNA sequence is moved from native context to a plasmid reporter, we used data obtained with the SuRE (Survey of Regulatory Elements) method[106] (Fig. 5C). Lastly, to determine whether the H3K27me2 mark represses transcription, we asked whether promoter sequences are repressed when integrated into regions marked by H3K27me2 using the TRIP (Thousands of Reporters Integrated in Parallel) method[107,108] (Fig. 5D). In the native context, promoters of H3K27me2-marked genes had weak median activity that was 9-fold higher than those of Polycomb domain genes (Fig. 5B). In contrast, promoters of H3K27me1- and H3K27ac-marked genes had 699-fold and 1437-fold higher median native activity than those of Polycomb domain genes, respectively. The difference between promoters of Polycomb domain genes and H3K27me2-marked genes was erased when these promoters were moved to a plasmid. In contrast, the promoters of H3K27me1- and H3K27ac-marked genes had 16-fold and 21-fold higher median autonomous activity than those of Polycomb domain genes (Fig. 5C). In summary, we found that promoters of genes marked by H3K27me2 are relatively inactive both in their native context and when moved to a plasmid.

We next asked whether seven promoter sequences[107] are repressed when integrated into the regions uniquely marked by H3K27me2 (Fig. 5D, Supplementary Fig. 10). Three of these promoters are repressed in their native chromatin context (*ADAMTS1, ARHGEF9*, and *BRINP1)* three are weakly expressed (*MED30, ZNF300*, and *TMEM106B)*, and one is constitutively expressed (*hPGK*). When integrated into the regions marked only by the H3K27me2 modification, one of these promoters (Fig. 5D iii) was significantly repressed, one was not repressed (Fig. 5D vi), and five were neutrally affected. H3K27me1-marked regions were also found to have a neutral effect on six out of seven promoters, and H3K27me3-marked regions significantly repressed four of the seven promoters. We therefore conclude that the H3K27me2 mark does not repress most gene promoters.

If the H3K27me2 mark is not repressive, then the slow dynamics of PRC2 might create a window for gene activation after DNA replication during the 1-2 fractions before tri-methylation is recovered. To compare cell cycle changes of the H3K27me2 mark versus active RNAPII at Polycomb domain genes, we visualized the magnitude of change (ΔRPKM) between S-phase fractions (Fig. 5E). At early replicating genes, which include many rapid stimulus response genes like *IL1B*, *IL17C*, and *IL10* that regulate inflammatory response (Supplementary Figure 11, Supplementary Table 1), the dynamics of these two marks formed a counter-clockwise cycle where modest gains in active RNAPII (<+0.1RPKM per fraction) preceded catalysis of di-methylation. The increase of active RNAPII initiated after S3 through the S4 fraction. This continued through DNA replication after G1/S, into S1 and until di-methylation increased in the S2 fraction (Fig. 5E i). The cycle compressed along both axes at mid-early replicating genes and was not discernable at genes replicating later in S-phase (Fig. 5E ii-vi). These include the homeotic patterning genes at the *HOXA* and *HOXD* loci (Supplementary Table 1) that are stably repressed in K562 cells. We thus found that early replicating genes in Polycomb domains transiently gain active RNAPII, before di-methylation is catalyzed, and significantly more so than later replicating genes (Supplementary Fig. 7A-B).

At the H3K27me2-enriched genes outside Polycomb domains, transient gains of active RNAPII were also greatest at early replicating genes (Fig. 5F i). These include other stimulus response genes like *IL11, TNF,* and *CCL24* (Supplementary Figure 12, Supplementary Table 1). Active RNAPII at these sites increased only when H3K27me2 decreased or did not change, such that changes in the two marks were inversely related, both at early and mid-early replicating genes (Fig. 5F i-ii). This reveals that loading of RNAPII and catalysis to H3K27 di-methylation are incompatible.

2.3 Discussion

Previous studies dissecting the dynamics of chromatin duplication have shown that the H3K27me3 modification, which is essential for developmental gene silencing, can take as much as 24 hours to regain pre-replication levels[50,54]. The slow recovery kinetics of the H3K27me3 mark challenge the model where it is propagated onto new, unmodified histones by PRC2 allosteric activation at replication forks[24]. Here, we resolve the catalysis of H3K27 methylation after DNA replication in cycling K562 cells and find that tri-methylation at genes in Polycomb domains proceeds via sequential mono- and di-methyl intermediates in the timespan of hours per step. This sequential catalysis in Polycomb domains proceeds to tri-methylation via allosteric activation of PRC2[25,35,109]. In active genes histone methylation stops at H3K27me1. This may be due to transcription of DNA that disengages the PRC2 complex, turns over nucleosomes[89,110], or recruits the UTX H3K27 demethylase[111,112]. In contrast, at inactive genes and other sites of low transcription, sequential catalysis yields the H3K27me2 mark. The H3K27me2 modification is one of the most abundant histone modifications across metazoan genomes, and has been proposed to limit pervasive transcription[83]. However, we find no evidence that the H3K27me2 mark is associated with transcriptional silencing in K562 cells. We conclude that the H3K27me2 modification is not repressive, but rather simply accumulates on inactive chromatin.

PRC2 enzymatic activity is understood based on the biochemical kinetics of the complex on *in vitro* assembled nucleosome substrates[25,41]. *In vitro*, the presence of tri-methylated H3K27 peptides greatly enhances efficient catalysis of H3K27me3 on an

unmodified nucleosome substrate via allosteric activation of PRC2 at the EED subunit[25]. The subsequent observation of PRC2 at replication forks led to the model where PRC2 allosteric activation there reads H3K27me3 on old histones and writes H3K27me3 on new ones[24]. Structural studies have also documented PRC2 complexes with other allosteric activators that are suggested to promote processive methylation of the unmodified H3K27 residue to H3K27me3[40,109]. We find using chromatin profiling in S-phase fractions that H3K27 methylation occurs in sequential steps via mono-, di-, and tri-methylation of the residue *in vivo*. These events can occur over the course of hours, which supports largely distributive methylation by PRC2 where each binding event adds one methyl group.

Chromatin profiling of S-phase fractions provides a rough timing of chromatin dynamics over the course of hours. In K562 cells the doubling time is ~20 hours: G1-phase is 4-6 hours, S-phase is 9-13 hours, and G2-phase is 1.5-2 hours[102]. We find that H3K27me3 is fully regained over the course of 2-4 fractions. Changes in the H3K27 methyl marks are greatest at early replicating genes in and outside Polycomb domains. After DNA replication later in S-phase, stepwise H3K27 methylation is enhanced to promote more efficient catalysis of H3K27me3. That early- versus later-replicating genes experience different rates of PRC2 catalysis may result from increased allosteric activation of the complex later in S-phase. This could result from the limited decompaction of replicating Polycomb domains[113] which maintains a high density of H3K27me3-marked nucleosomes that thereby enhance PRC2 allosteric activation. Many cellular factors change during S-phase progression, and our data suggests that proximity to the onset of DNA replication[114,115] after G1-phase shapes PRC2 kinetics during S-phase.

2.3.1 The H3K27me2 modification accumulates after DNA replication at inactive chromatin

In embryonic development H3K27 methylation is essential for gastrulation[27] and lineage acquisition[116,117]. PRC2 first catalyzes the H3K27me2 mark at sites with histone H2A ubiquitination after zygotic genome activation in flies, zebrafish, and mice[118-120]. These sites later form Polycomb domains marked by H3K27me3. During embryogenesis

nearly all H3K27me2 matures into H3K27me3, but pre-epiblast mouse embryonic stem cells (mESCs) have H3K27me1, H3K27me2, and H3K27me3 that is mutually distinct: H3K27me1 covers active gene bodies, H3K27me2 spans intergenic regions and inactive genes, and H3K27me3 is focused around CpG-rich sequences which occur at ~70% of mammalian promoters[81,84]. Our data find that H3K27me2 results from widespread catalysis in S-phase which may not depend on pre-existing histone H2A ubiquitination.

During H3K27 methylation by PRC2, H3K27me2-modified nucleosomes are the substrate for the H3K27me3 mark. This might suggest that inactive sites marked by H3K27me2 are more likely to gain the H3K27me3 mark. The maintenance of H3K27me2-marked regions devoid of the H3K27me3 mark across cell cycles means this conversion is restricted by the slow kinetics of PRC2 and low local allosteric activation of the complex. This implies that cells with different cell cycle lengths may achieve very different global distributions of the H3K27me3 mark[94].

2.3.2 PRC2 opposes acetyltransferase activity

Whereas H3K27me3 marks Polycomb domains, H3K27 acetylation marks active regulatory elements. Hindrance of H3K27 acetylation is achieved by H3K27me3 and H3K27me2, but not by H3K27me1[121]. Previously di-methylated regulatory elements gain H3K27ac upon PRC2 knockout and are activated in the early mouse embryo, mESCs and *Drosophila* cells, consistent with the idea that di-methylation blocks acetylation[83,84,118]. Our data corroborates that H3K27 acetylation is gained upon PRC2 inhibition and resolves the timing of H3K27 acetylation with respect to changes in H3K27 di-methylation in each cell cycle.

We found that PRC2 catalytic and allosteric inhibitors affect H3K27 methylation in S-phase and allow the gain of H3K27ac. After DNA replication in early S-phase, both drug types result in H3K27ac gains, but there is a slightly faster effect of EPZ-6438, suggesting that catalytic inhibition somehow allows more acetylation. This informs development of PRC2 inhibitors and their application toward treating multiple cancers[103,104,122], because it suggests targeting different parts of PRC2 can shape

cancer response[123]. Catalytic PRC2 inhibitors, including EPZ-6438, mimic S-adenosyl methionine and occupy the catalytic site of EZH2/1 methyltransferase subunit to form a catalytic block. This likely prevents PRC2 from binding the H3K27 residue. In contrast, allosteric PRC2 inhibitors, including EED-226, bind the aromatic cage of the EED subunit and allosterically modulate the EZH2/1 catalytic site. If PRC2 in this context can still bind the H3K27 residue after DNA replication, then PRC2 itself could shield the residue from acetylation.

PRC2 catalytic and allosteric inhibitors are suggested to have distinct effects on catalysis of H3K27me3[94]. In K562 cells, both catalytic and allosteric inhibition slow catalysis of H3K27me1 and H3K27me2 after DNA replication. In this way, PRC2 allosteric activation modulates not only catalysis of H3K27me3, but catalysis of all H3K27 methyl marks.

2.4 Figures

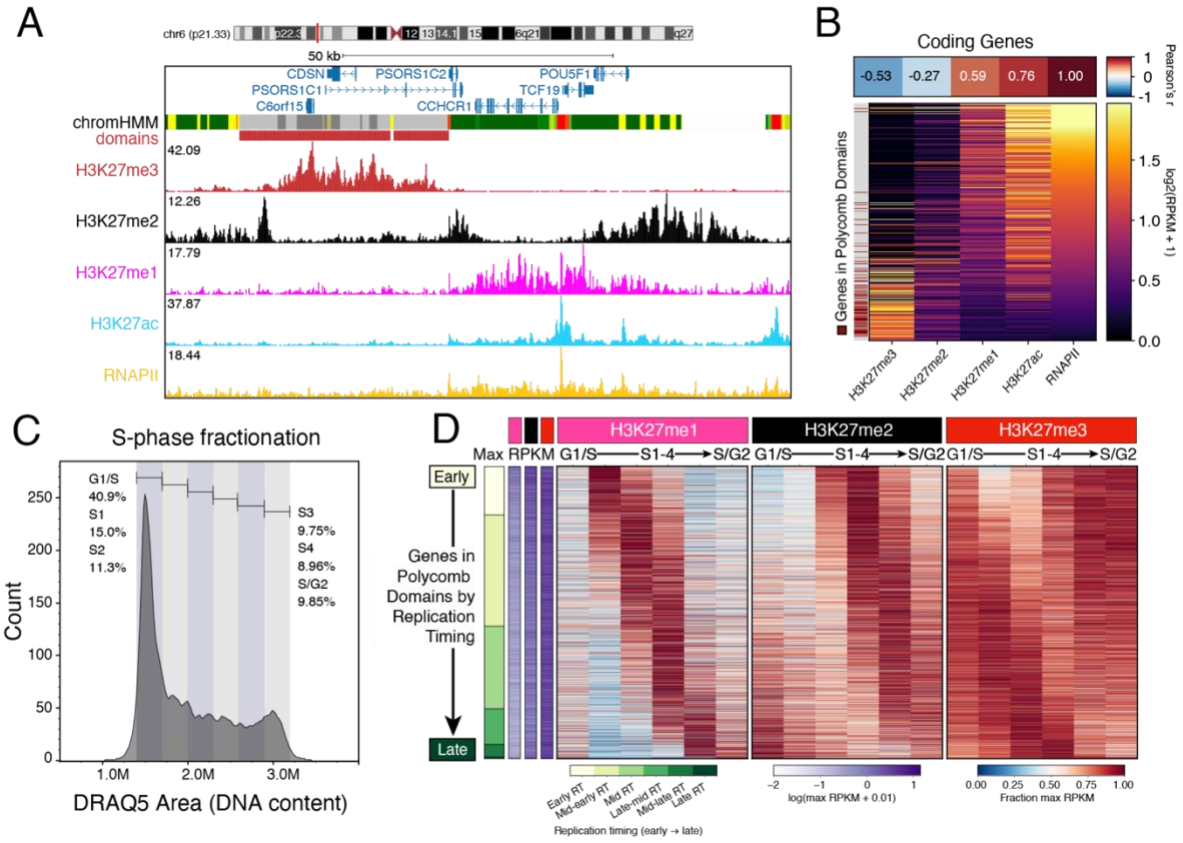


Figure 1. Stepwise methylation of histone H3K27 in Polycomb domains after DNA replication: (A) Browser snapshot of CUT&Tag profiling in K562 cells for the H3K27me3, H3K27me2, H3K27me1, and H3K27ac modifications, and RNAPIIS2/5p at the POU5F1 (OCT4) locus. A Polycomb domain[96] overlaps the PSOR1S1C gene. (B) (top) Pearson correlation heatmap of active RNAPII versus H3K27 modifications at coding genes including a 1 kilobase promoter region upstream (n=19,946). (bottom) Signal heatmap of correlated gene values after log2 transformation. Genes in annotated Polycomb domains are highlighted in red and show high H3K27me3, low active RNAPII signal. (C) Gating schema based on DRAQ5 measurements of DNA-content used to isolate S-phase fractions via fluorescence-activated cell sorting (FACS). (D) CUT&Tag signal at genes in Polycomb domains (n=2,861) ranked by replication timing (y-axis) over the S-phase fractions (x-axis) from (C). Signal is shown as fraction of the maximum averaged across two replicates. Log10(max RPKM) per gene is shown in purple for the H3K27me1, H3K27me2, and H3K27me3 modifications to demonstrate high H3K27me3 signal and relatively less H3K27me2 and H3K27me1 signal at these genes.

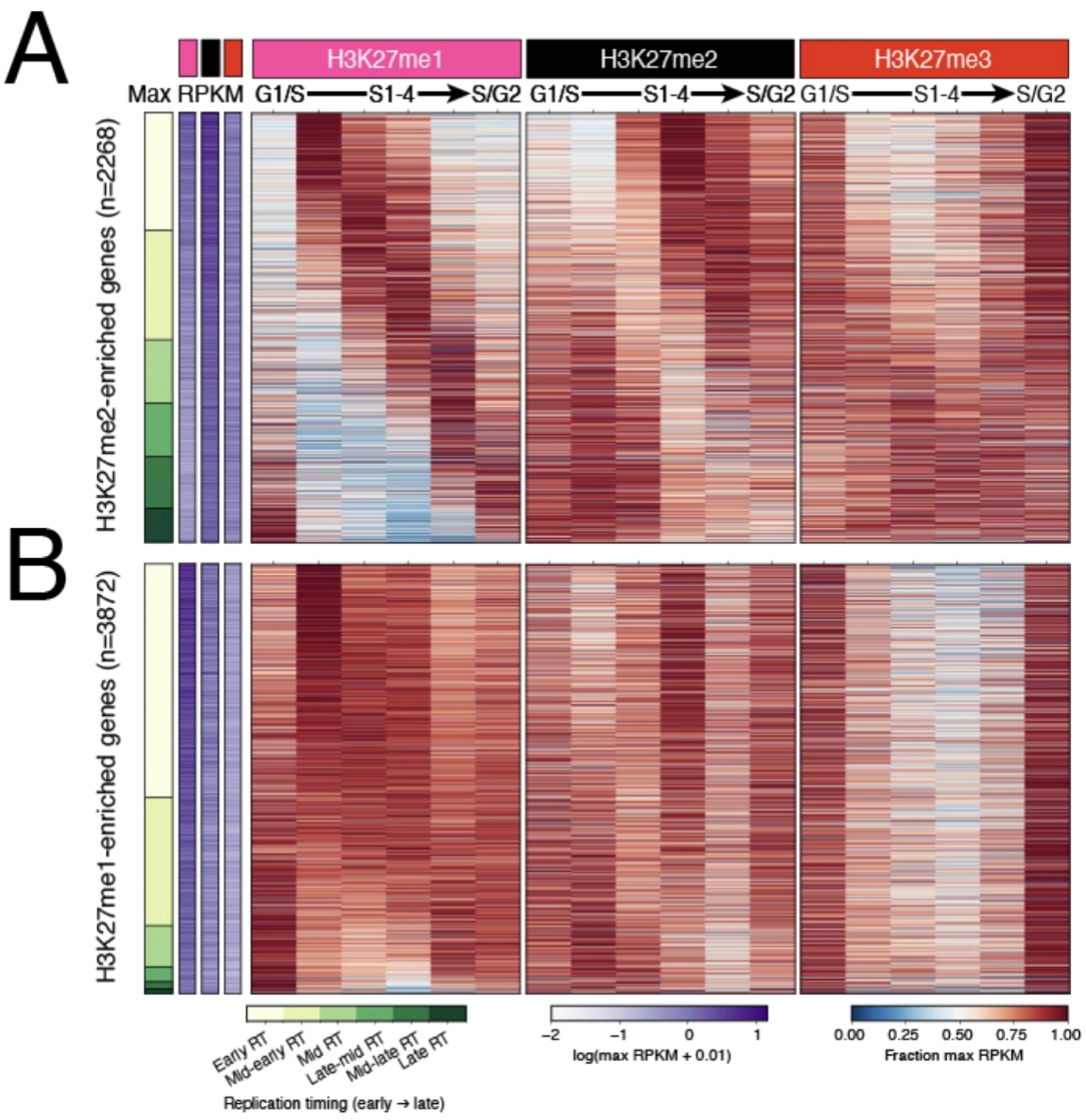


Figure 2. Stepwise methylation of the H3K27 residue outside of Polycomb domains: (A) CUT&Tag signal at genes outside Polycomb domains enriched for H3K27me2 (top, n=2,268) and (B) H3K27me1 (bottom, n=3,872) ranked by replication timing (y-axis) over the S-phase fractions (x-axis). Signal is shown as fraction of the maximum (dilution) averaged across two replicates.

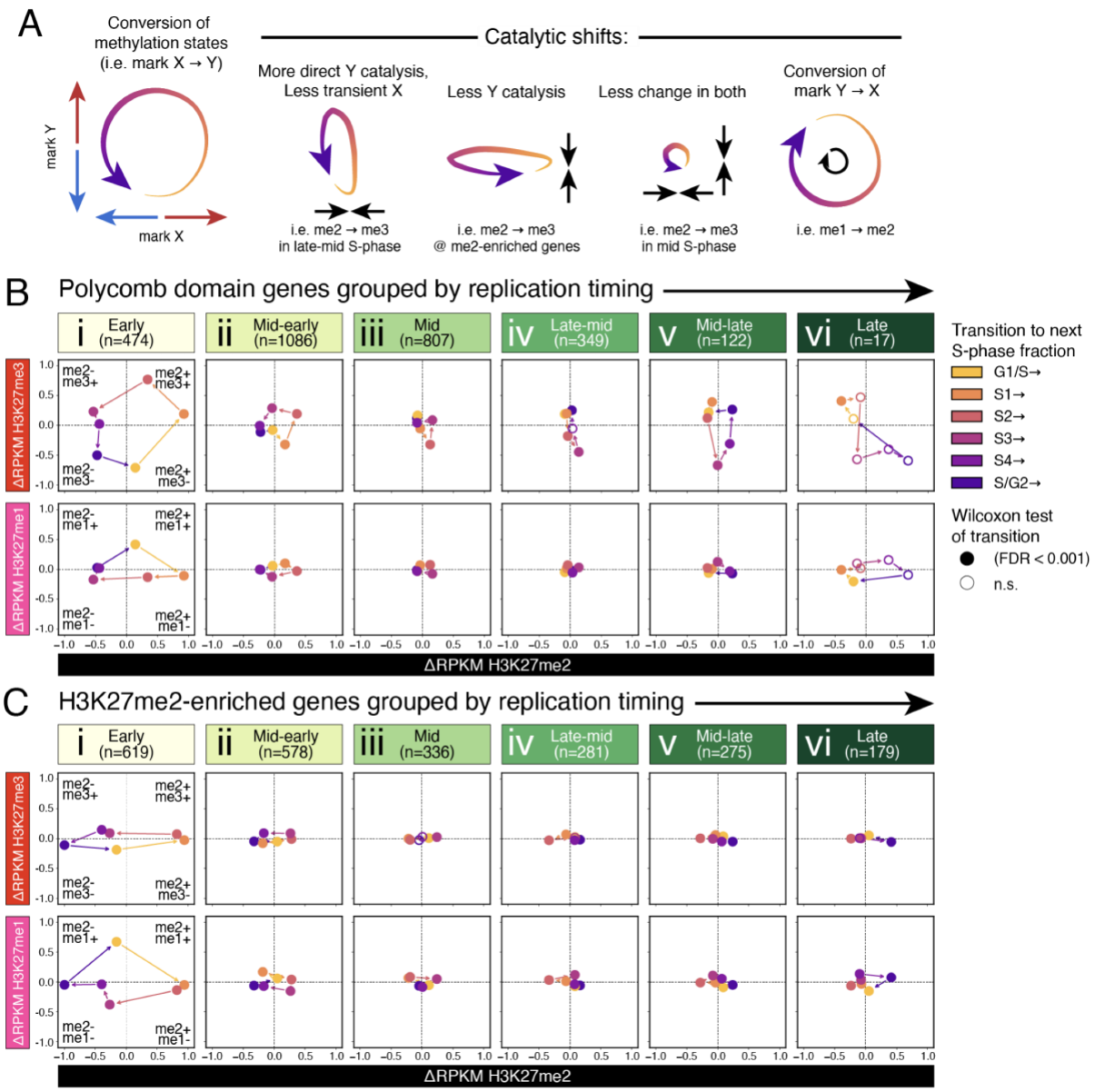


Figure 3. A cycle of H3K27 methylation at genes inside and outside of Polycomb domains: (A) diagram of possible catalytic patterns. (B) magnitude of change shown as DRPKM between S-phase fractions for H3K27me2 versus H3K27me3 (top) and H3K27me1 (bottom) for genes in Polycomb domains grouped by replication timing. Filled dots are medians with a significant change in either mark. Arrows show the direction of the cell cycle time course between S-phase fractions. (C) Same as (B) but for H3K27me2-enriched genes.

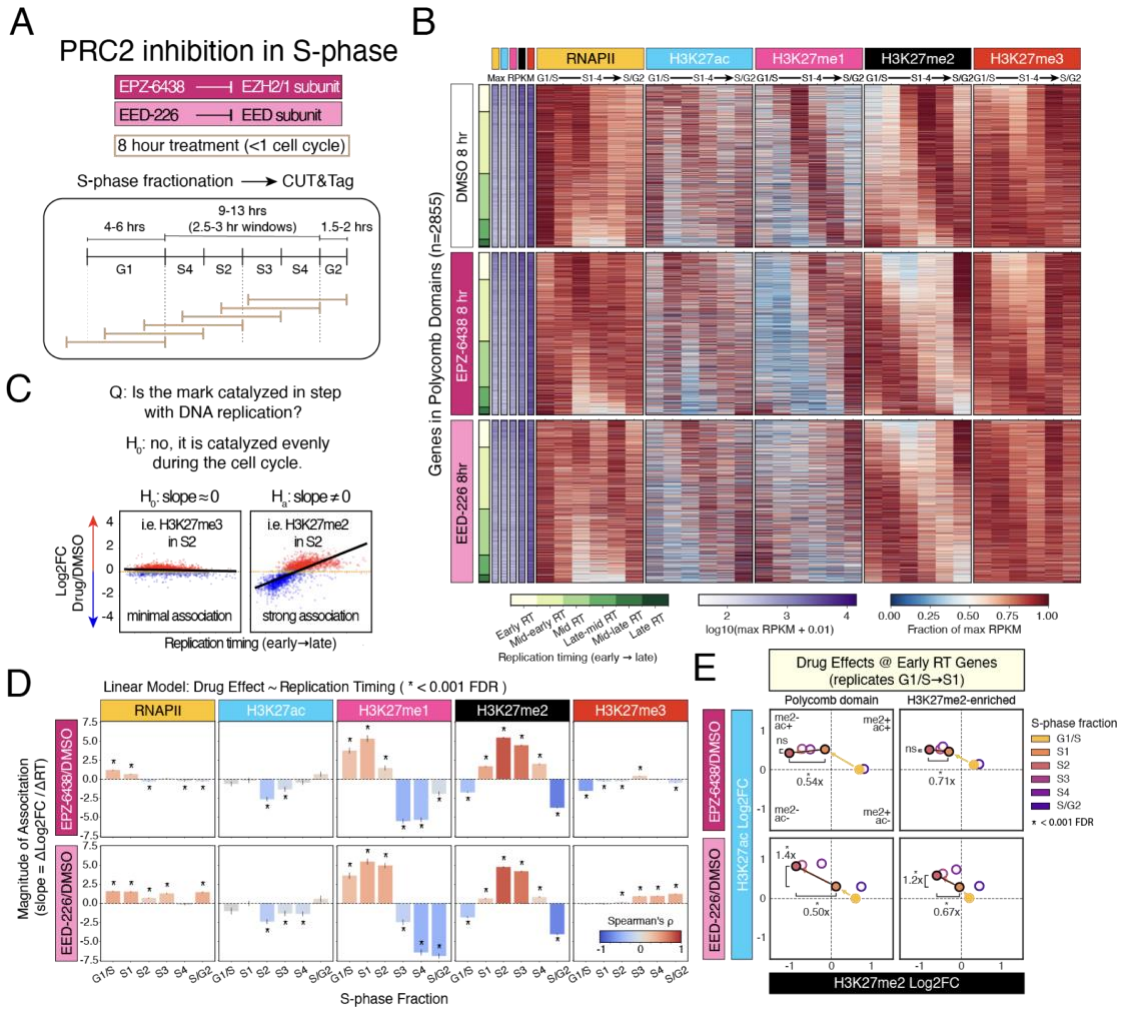


Figure 4. PRC2 inhibition delays H3K27me1 and H3K27me2 modifications in Polycomb domains after DNA replication: (A) Experimental design for 8-hour treatment of cycling K562 cells with small-molecule inhibitors of PRC2. (B) CUT&Tag signal at genes in Polycomb domains (n=2,861) ranked by replication timing (y-axis) over the S-phase fractions (x-axis). Signal is shown as fraction of the maximum averaged across two replicates. Max RPKM is computed per treatment across the S-phase time course. Log10(max RPKM) per gene is show in purple for active RNAPII, H3K27ac, H3K27me1, H3K27me2, and H3K27me3 (left to right). The heatmaps show signal after treatment for 8 hours with DMSO (vehicle control, top), the EZH2 inhibitor EPZ-6438 (middle), and the EED inhibitor EED-226 (bottom). (C) Statistical paradigm for testing the relationship between drug effects versus replication timing at genes in Polycomb domains. Scatter plots are from Supplementary Figure 8. Genes above an absolute of log2 fold-change of 0.2 are colored red for up in treatment or blue for down in treatment, or grey for Log2 fold-change < 0.2. (D) Linear regression analysis (y-axis) of response to EPZ-6438 (top), and EED-226 (bottom) within S-phase fractions (x-axis) versus replication timing at genes in Polycomb domains. Stars indicate a significance <

0.001 after FDR correction. Bar color is Spearman's rho (E) PRC2 inhibition at early replicating genes in and outside Polycomb domains. Median log2 fold-change for EPZ-6438 (top) and EED-226 (bottom) versus DMSO is shown for H3K27me2 (y-axis) and H3K27ac (x-axis) in the S2 fraction. The fold difference between Log2 fold-change in the S1 versus S2 fraction is shown for differences greater than ±0.2. Stars indicate a significant difference in Log2 fold-change values by the Wilcoxon test with a p-value < 0.001 after Bonferroni correction.

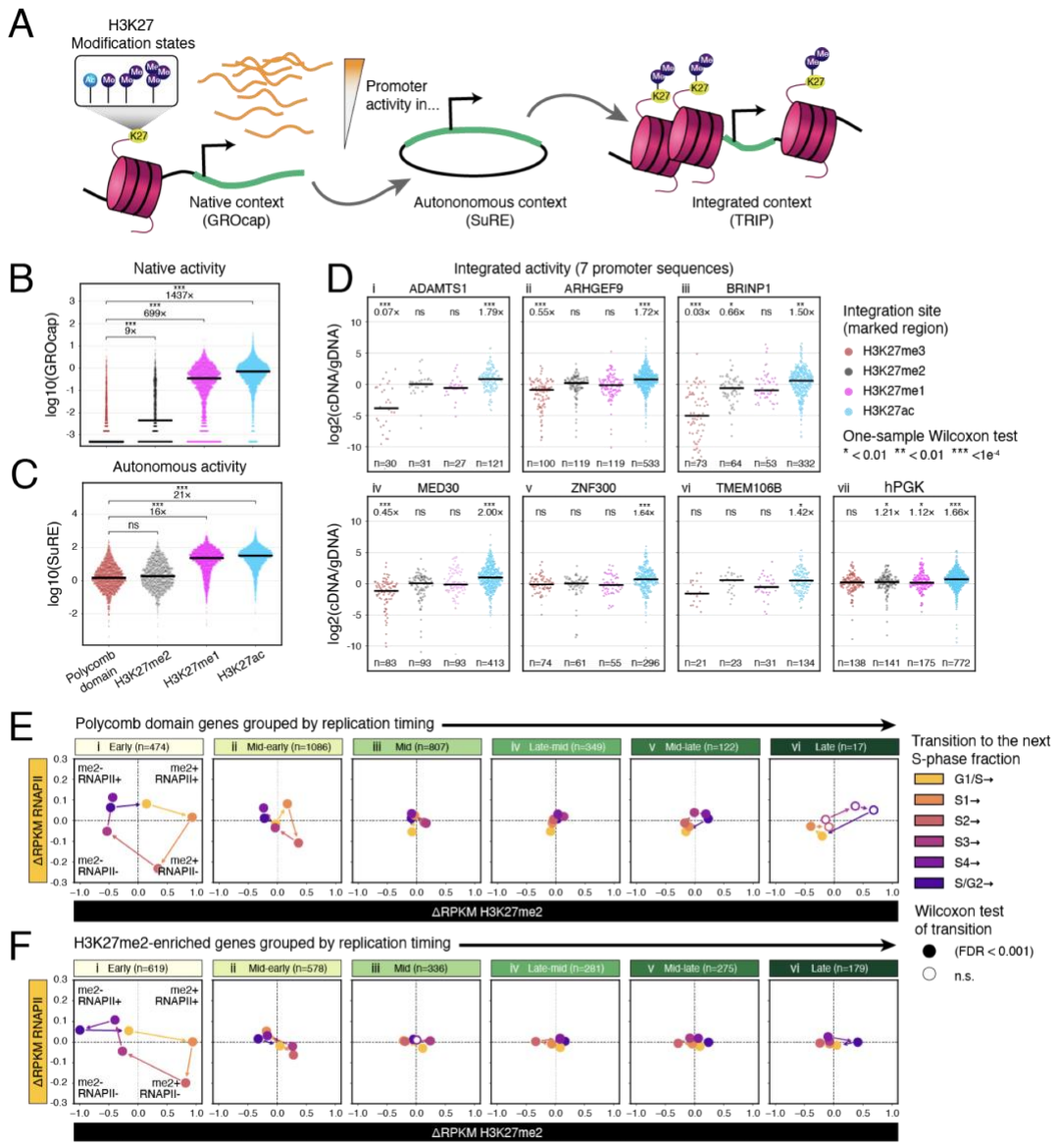
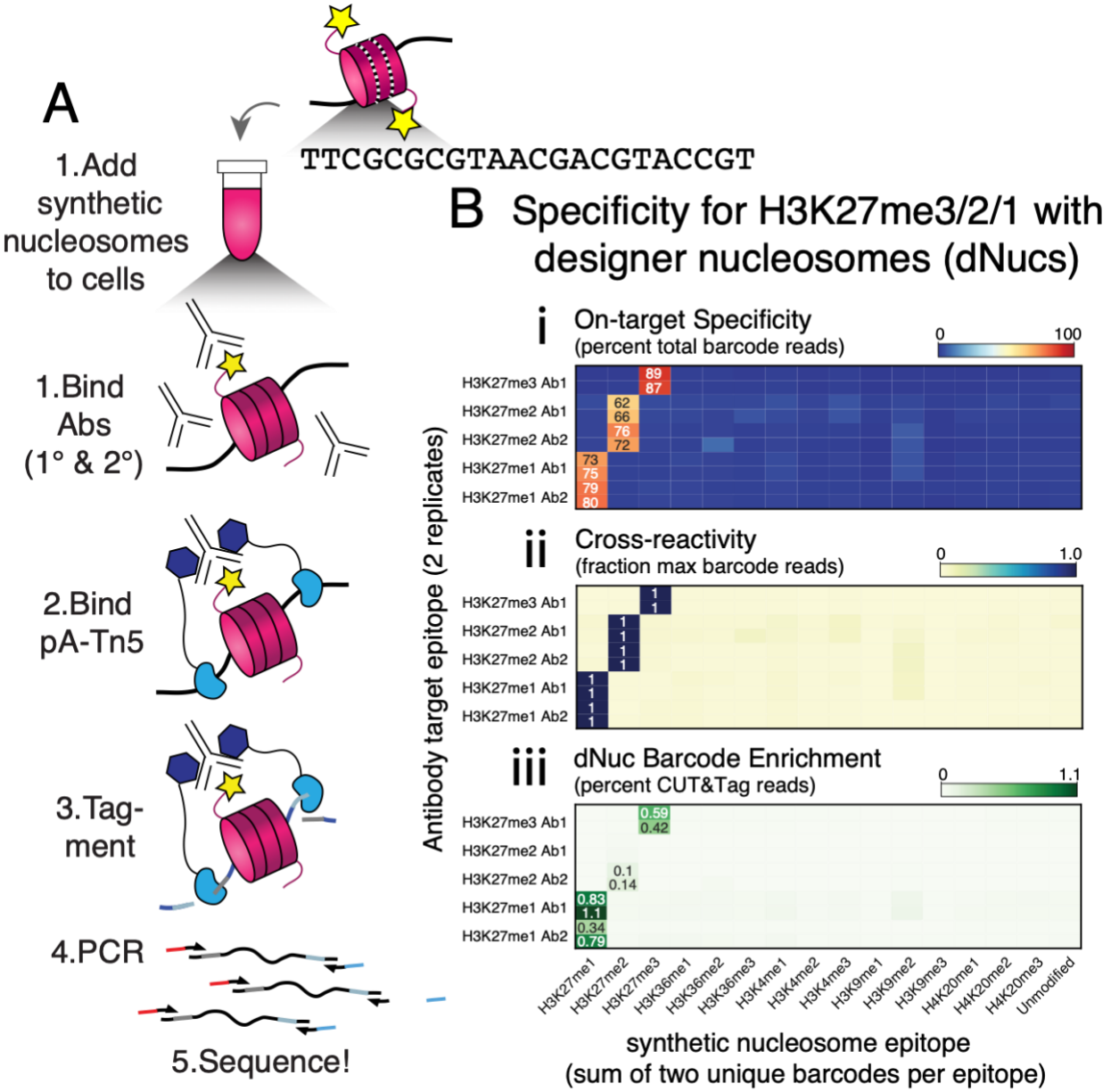


Figure 5. Histone H3K27me2 modification marks inactive chromatin: (A) Schematic of experimental design to measure promoter activity in three different contexts. (B) Native promoter activity measured by GROcap[105] signal at gene promoters grouped by most abundant H3K27 modification (defined in Supplementary Fig. 5A). (C) Same as (B) but for SuRE[106] signal. (D) Expression of promoters integrated into H3K27-modified chromatin. TRIP data are from Ref. [107]. Dot plots are expression values per integration normalized to genomic DNA with median per group. H3K27 classifications are the top 5% of regions defined in Supplementary Figure 9E. FDR-corrected one-sample Wilcoxon test (null: log2(cDNA/gDNA = 0) and fold change is computed against that expectation (E) magnitude of change shown as delta-RPKM between S-phase fractions for active RNAPII versus H3K27me2 at genes in Polycomb domains grouped by

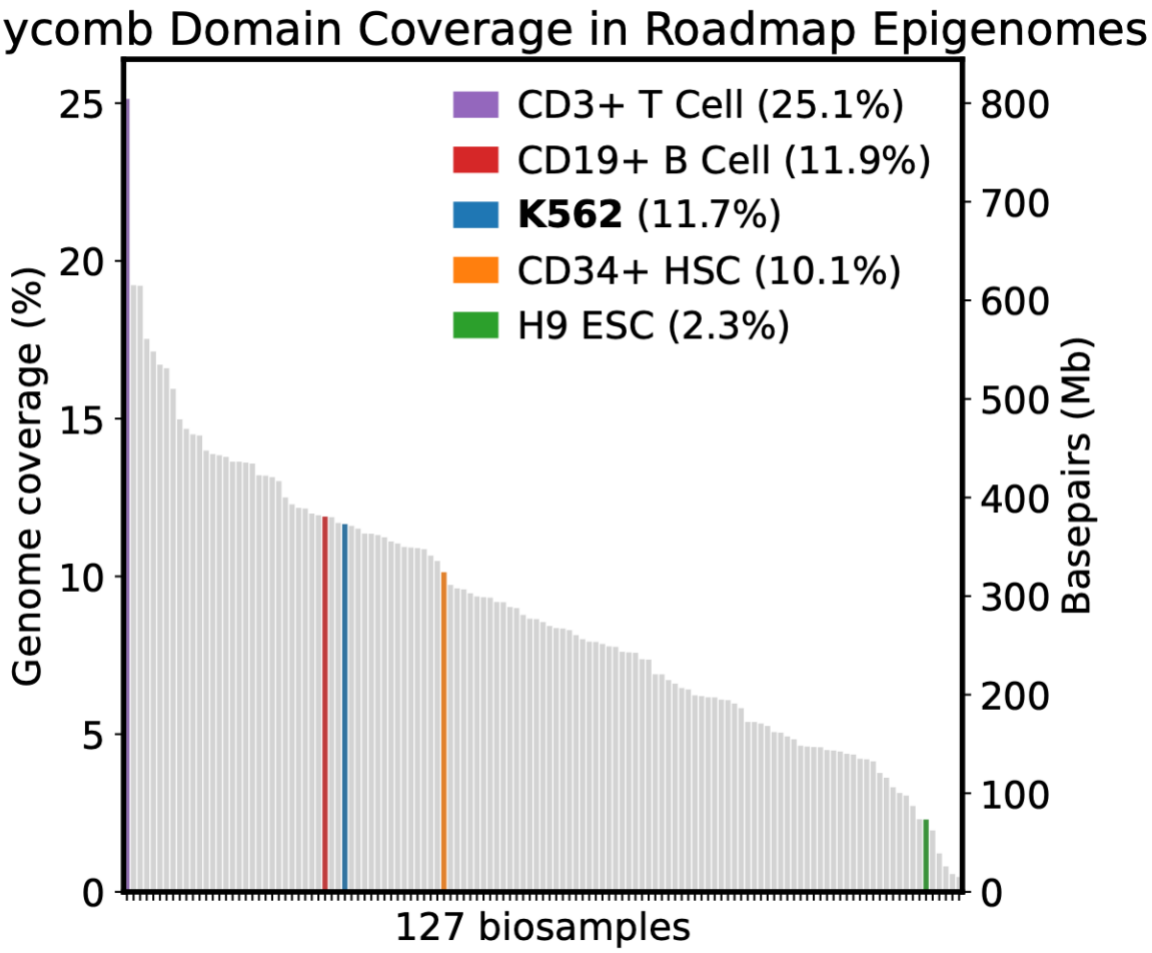
replication timing. Filled dots are medians with a significant change in either mark. Insignificant change versus the null of no change is denoted by no fill. Arrows show the direction of the cell cycle time course between S-phase fractions. (F) Same as (E) but for H3K27me2-enriched genes.

2.5 Supplementary Figures

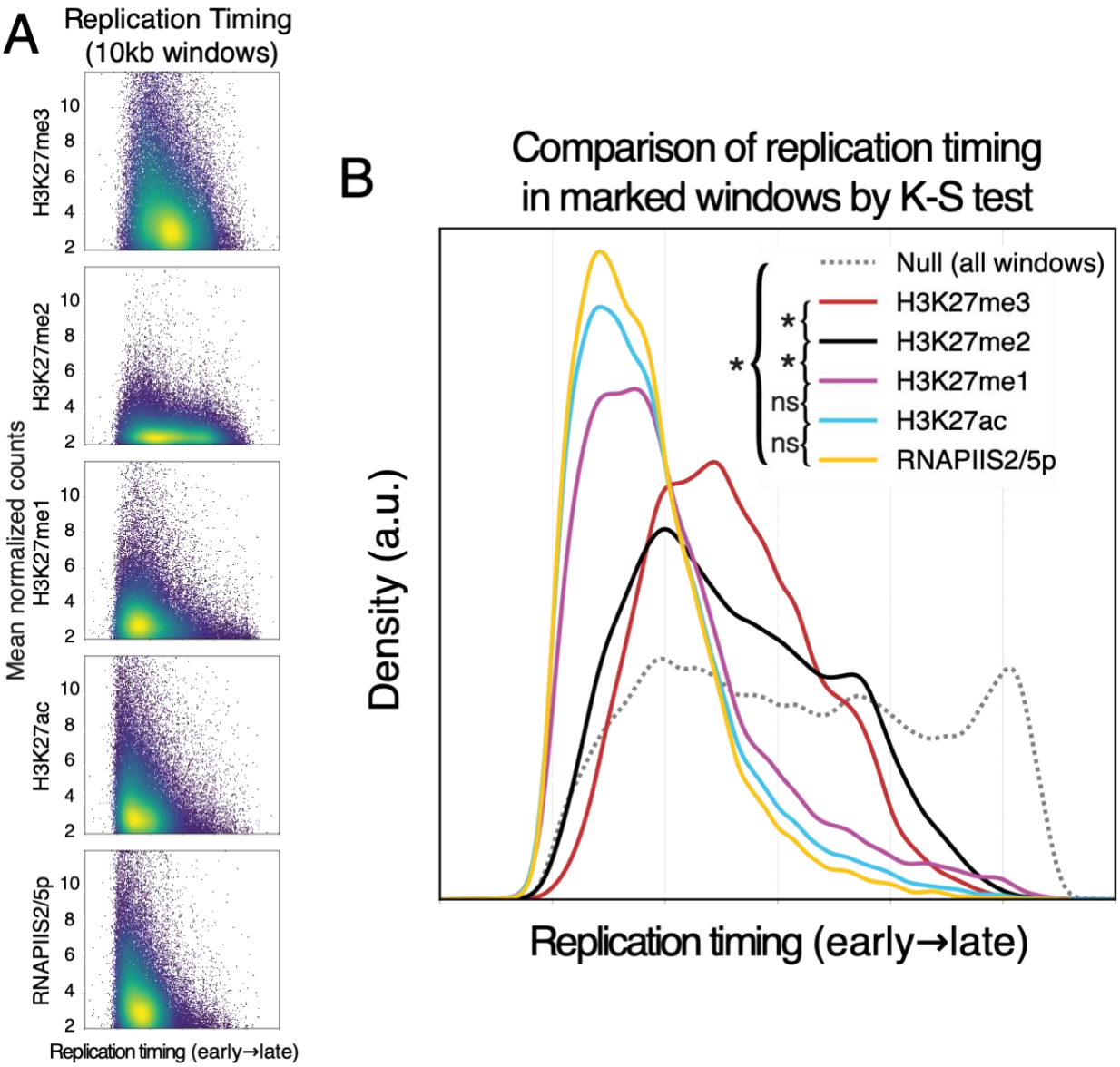


Supplementary Figure 1. Confirmation of H3K27 methylation profiling with designer barcoded nucleosomes: (A) Cleavage Under Targets & Tagmentation (CUT&Tag) chromatin profiling paradigm with a synthetic nucleosome spike-in panel (EpiCypher, Cat# 19-1002). The synthetic nucleosomes are barcoded according to their lysine methylation status using a non-endogenous sequence that can quantified prior to reference genome alignment. The star denotes one of many possible lysine methylations also present in cells input to CUT&Tag. (Bi) Heatmap of barcodes collated

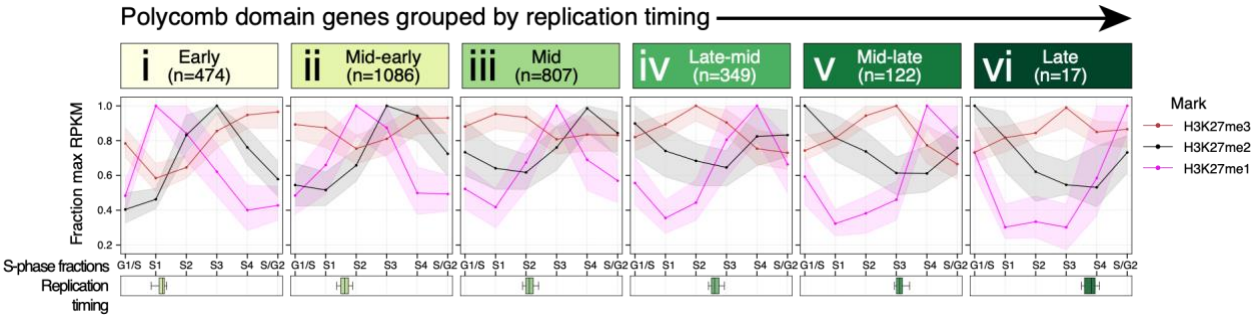
from FASTQ files of CUT&Tag libraries. Y-axis are individual libraries generated with a commercial antibody specific for H3K27me3, H3K27me2 (2 antibodies tested), or H3K27m1 (2 antibodies tested). Profiling was performed in duplicate. Values are the percent of total barcodes in the spike-in panel. (ii & iii) Same as (i) but values are (ii) fraction of the maximum barcode detected and (iii) percent total reads in the FASTQ file.



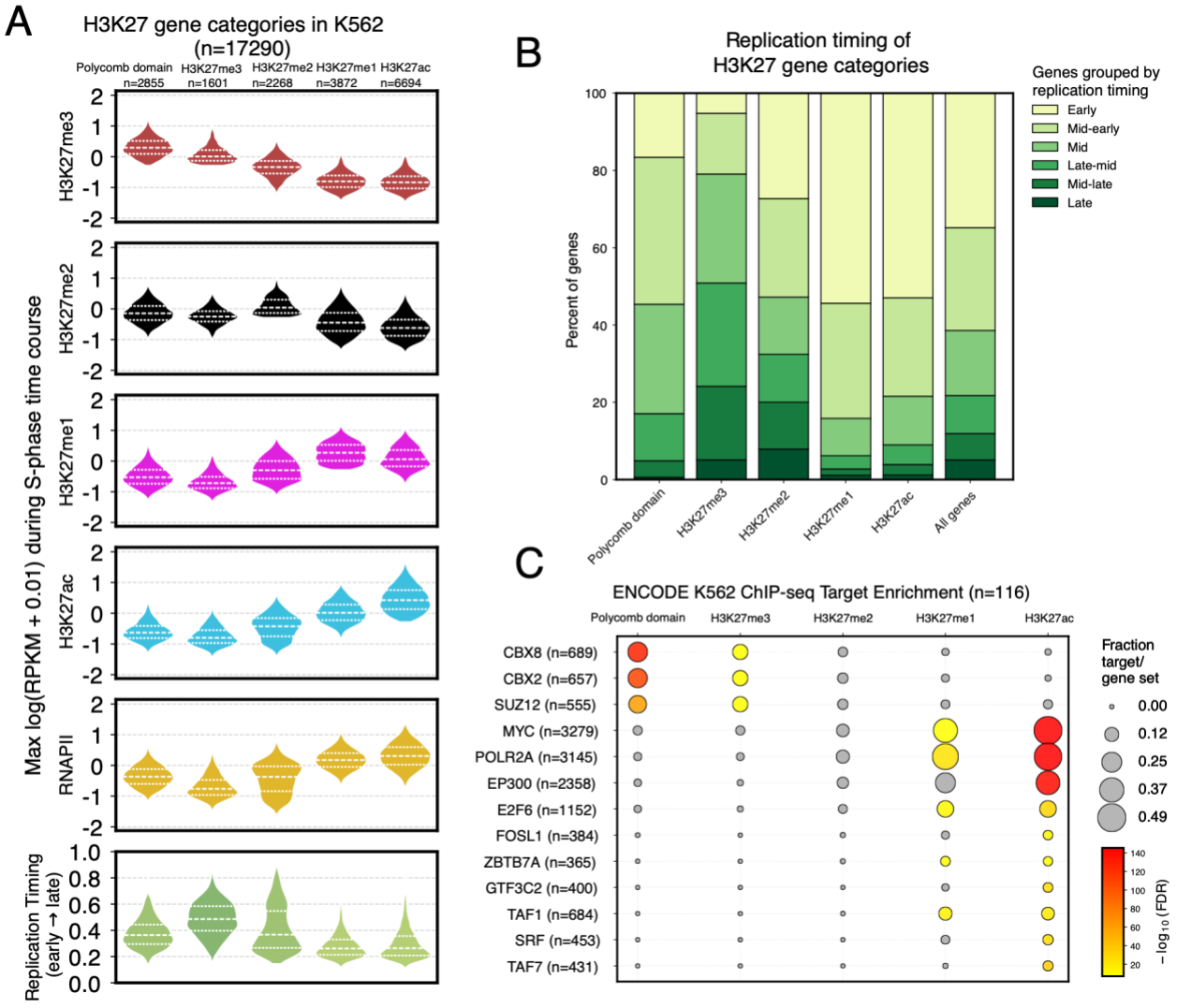
Supplementary Figure 2. Polycomb domains defined by the Roadmap Epigenomes Consortium in 127 biosamples: Polycomb domains defined by the Roadmap Epigenomes Consortium[96] in 127 biosamples ranked by genome coverage (left) and base pair coverage (right). K562 and four representative biosamples are highlighted with their percent genomic coverage. Polycomb domains occupy 11.7% of the genome in K562, which is near the median of the 127 Roadmap epigenomes.



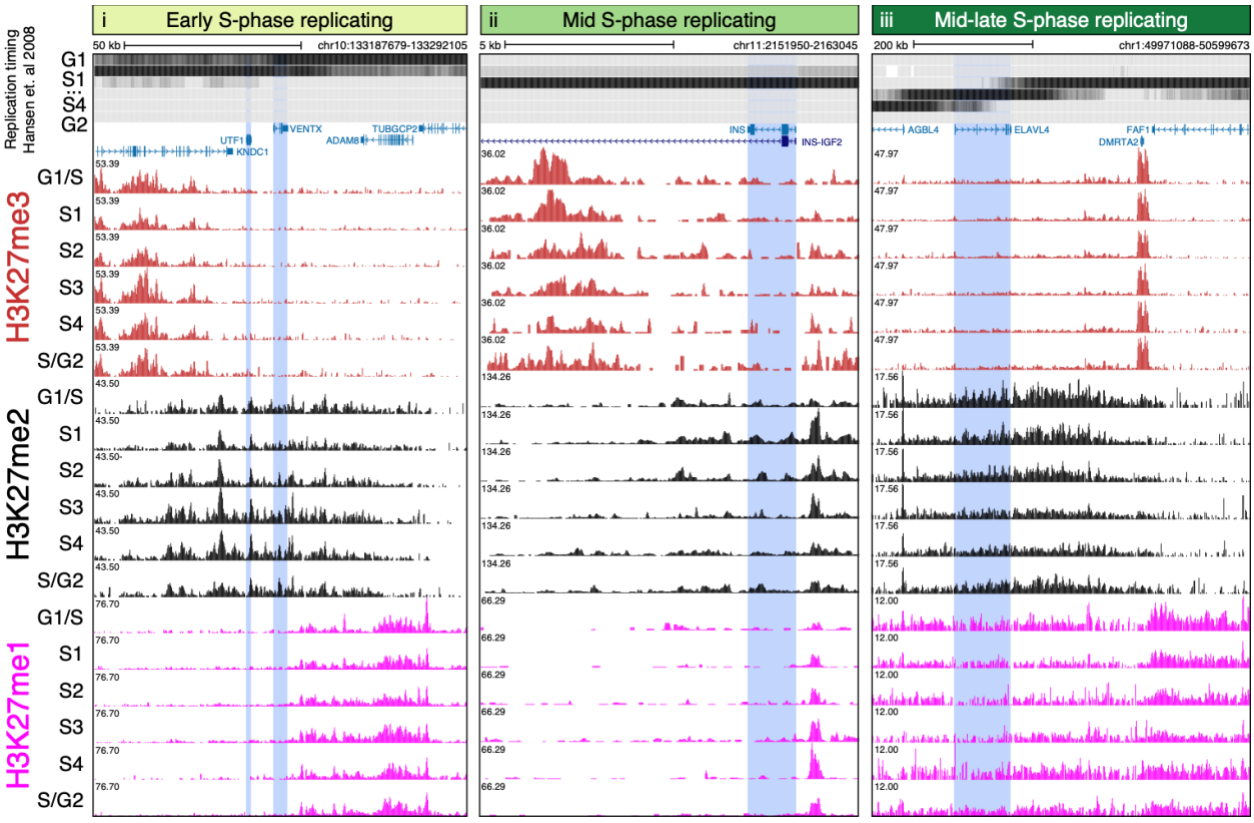
Supplementary Figure 3. Replication timing of 10-kilobase (kb) windows with H3K27 modifications: (A) Mean normalized counts for H3K27me3, H3K27me2, H3K27me1, H3K27ac, and RNAPIIS2/5p (y-axis) in 10-kb windows versus mean replication timing (early-to-late) colored by 2d density (blue is low, yellow is high). A minimum threshold of 2 mean RPKM is applied to select for marked windows. (B) Dsitribution of replication timing in marked 10-kb windows compared across marks via the Kolmogorov-Smirnov test. The null distribution is replication timing in all 10kb windows. Comparisons are pairwise between marks and subject to Bonferroni correction. Comparisons between each mark and the null distribution were also performed (all significant). * = $p < 0.01$.



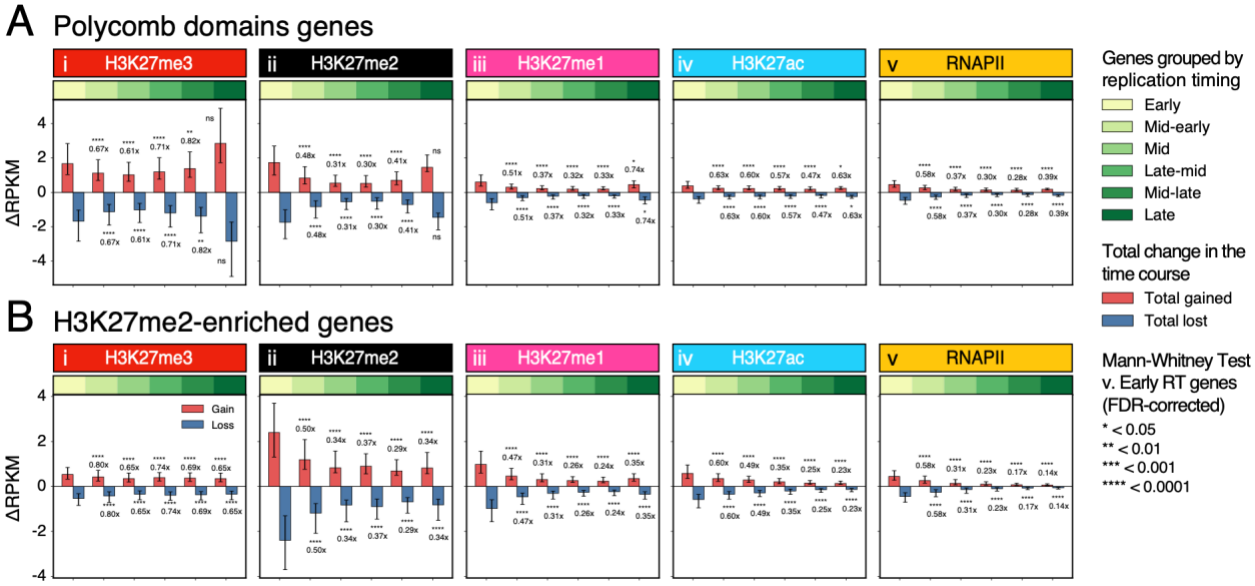
Supplementary Figure 4. Trends of H3K27 methylation at Polycomb domain genes grouped by replication timing: Trends are median with interquartile range per S-phase fraction for H3K27me1 (magenta), H3K27me2 (black), and H3K27me3 (red) in Polycomb domain genes grouped by replication timing from left to right (i-vi). Replication timing is shown per group as one boxplot where each fraction is evenly spaced such that G1/S = 0, S1 = 0.2, S2 = 0.4, S3 = 0.6, S4 = 0.8, and S/G2 = 1.



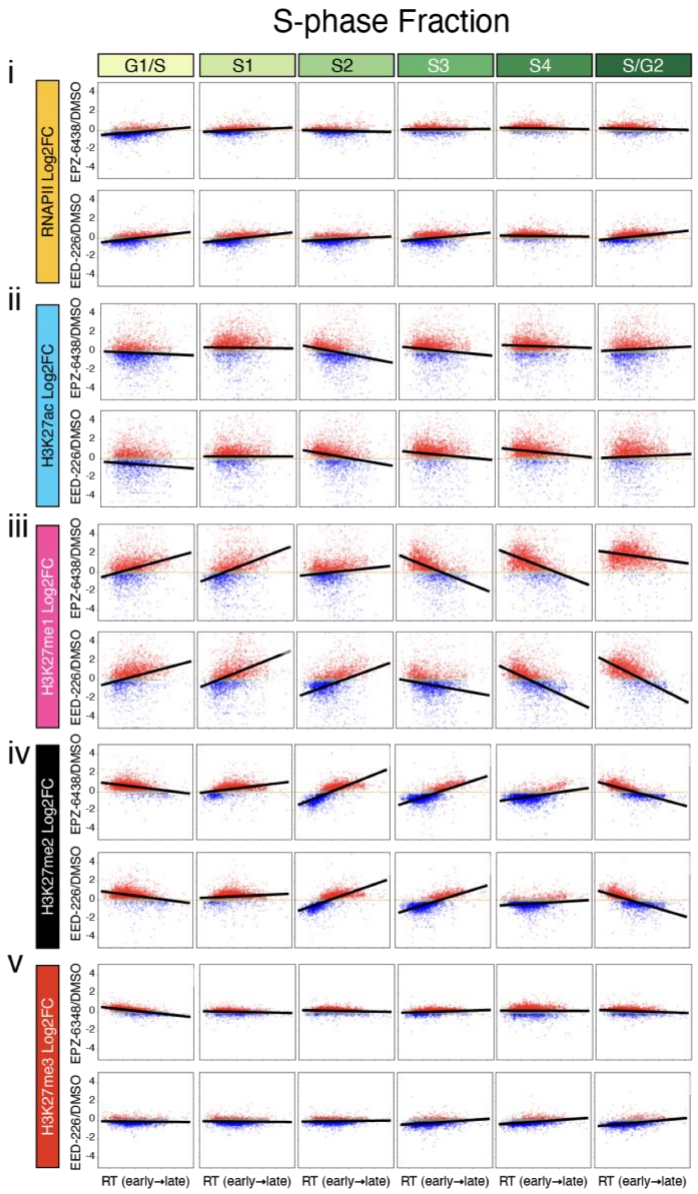
Supplementary Figure 5. Assignment of H3K27 gene categories used in Figure 2: (A) Genes in Polycomb domain (n = 2855) were assigned based on at least 50% base pair overlap with regions defined by the Roadmaps Epigenome Consortium[96] using chromHMM[97] in K562 cells. The remaining genes were then categorized based on the highest maximum H3K27 modification (RPKM) during the time course above a minimum of 0.5 RPKM. Signal is shown as log10(RPKM + 0.01) per gene, and replication timing is show from 0-1 (early-to-late) per category. (B) Replication timing groups per H3K27 category. (C) Overlap enrichment per H3K27 category versus ENCODE_TF_ChIP-seq_2015 target genes[124] in K562. Top 3 gene sets (y-axis) are shown per H3K27 category.



Supplementary Figure 6. Examples of H3K27me2-enriched genes: (i) UCSC browser snapshot of coverage profiles for H3K27me3, H3K27me2, and H3K27me1 in S-phase fractions around the VENTX and UTF1 genes, highlighted in blue. Each mark is group auto-scaled. Replication timing in K562 cells[100] is show in S-phase fractions at the top. Gene annotation is Gencode v49 for hg38. (ii-ii) same as (i) but (ii) at the INS gene and (iii) ELVAL4 gene.

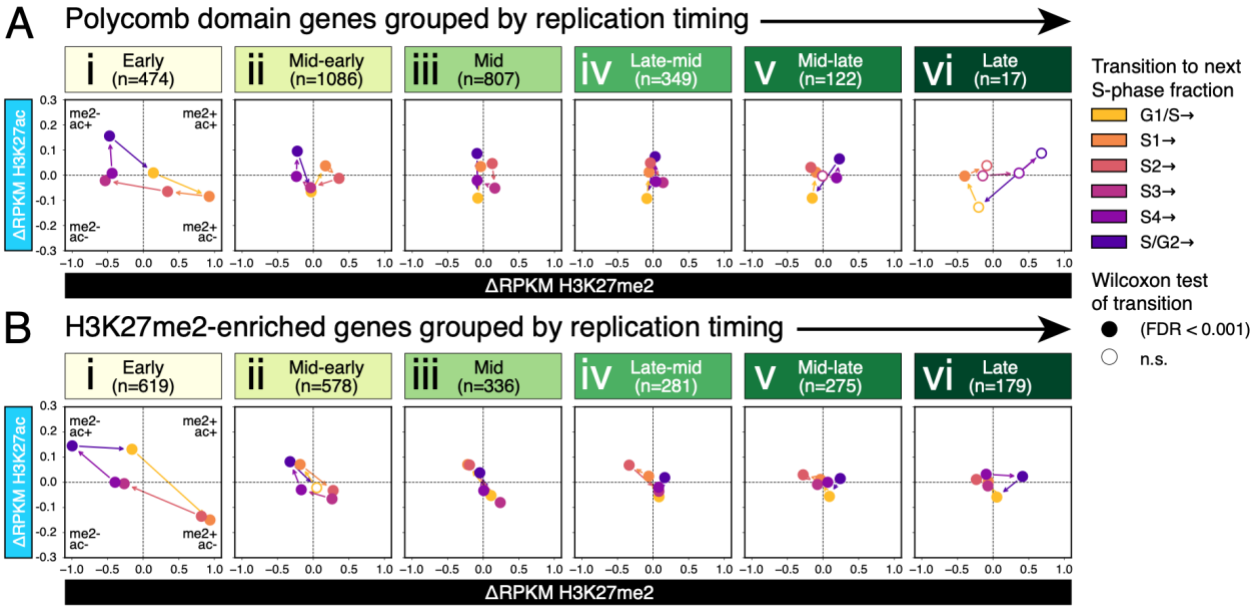


Supplementary Figure 7. High change of H3K27 modification and active RNAPII at earl replicating genes in and outside Polycomb domains: (A) total change shown as delta-RPKM for (i) H3K27me3, (ii) H3K27me2, (iii) H3K27me1, (iv) H3K27ac, and (v) active RNAPII during the time course at genes in Polycomb domains grouped by replication timing (x-axis). Early replicating genes are compared to the other groups using the Mann-Whitney test and fold change is shown between medians. Error bars are interquartile range. (B) same as (A) but for H3K27me2-enriched genes.

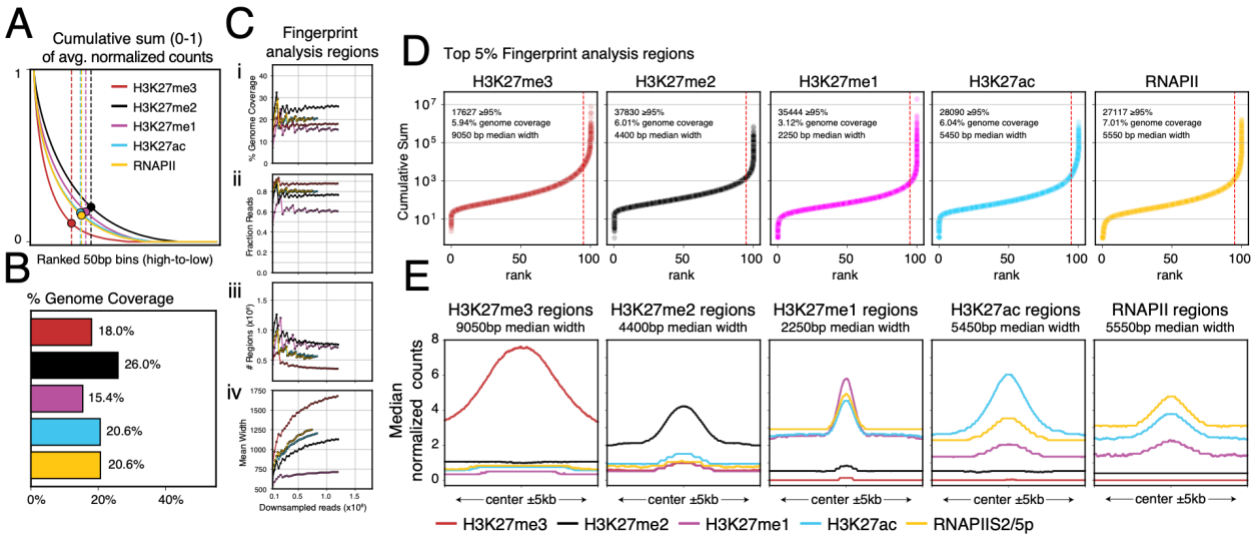


Supplementary Figure 8. Linear regression of active RNAPII, H3K27ac, and H3K27 methylation in response to PRC2 inhibition versus replication timing at genes in Polycomb domains: (i) Log2 fold-change (y-axis) for EPZ-6438 (TAZ, top) and EED-226 (EED, bottom) versus DMSO is shown for RNAPII versus RT (x-axis, early-to-late) at genes in Polycomb domains (n=2861) per S-phase fraction (columns, left to right). Genes above an absolute of log2 fold-change of 0.2 are colored red for up in treatment or blue for down in treatment, or grey for Log2 fold-change < 0.2. The best fit line is

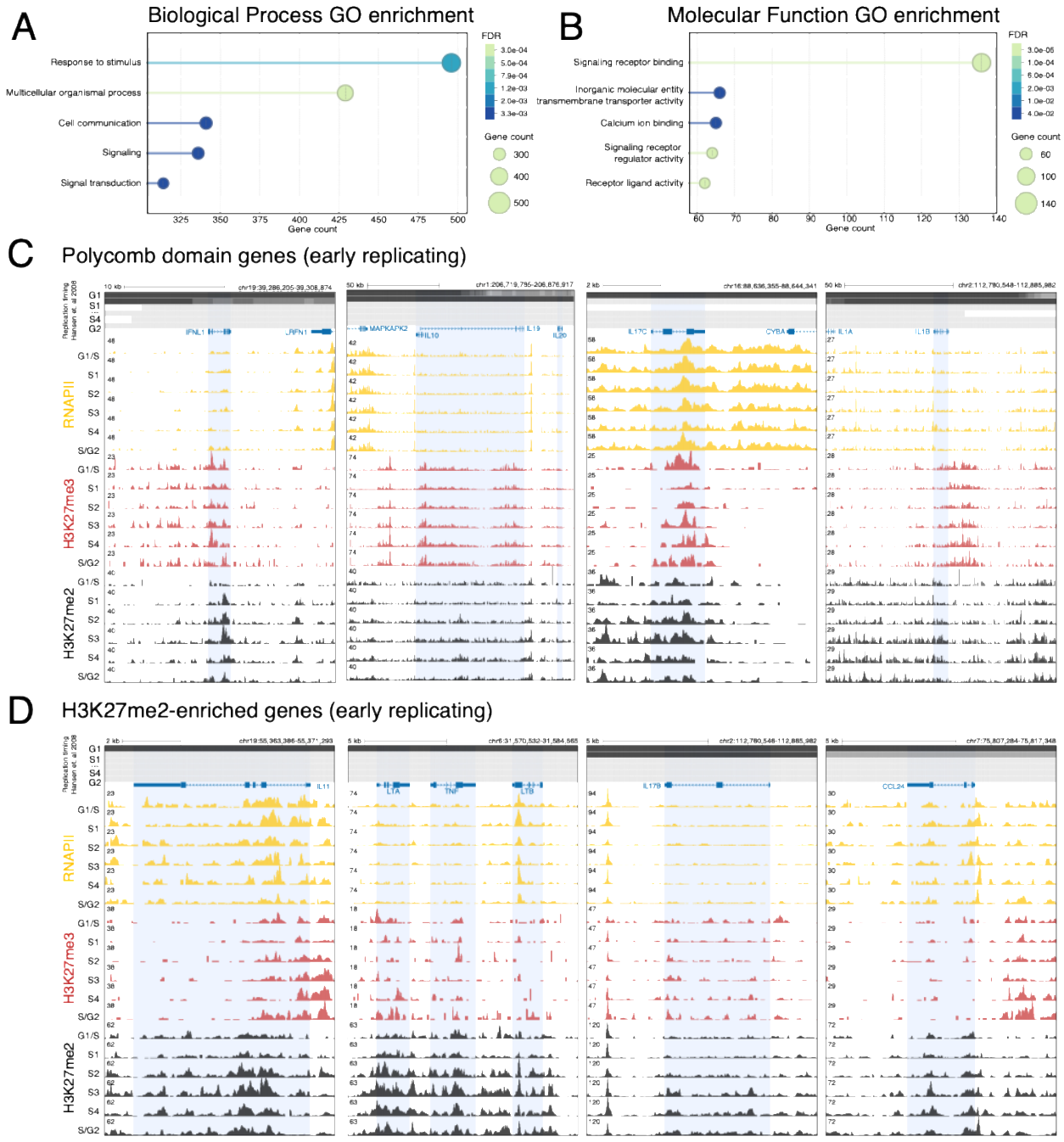
shown per S-phase fraction. (ii-v) Same as (i) but for H3K27ac, H3K27me1, H3K27me2, and H3K27me3.



Supplementary Figure 9. Changes in H3K27 acetylation versus di-methylation at genes in and outside Polycomb domains without drug treatment: (E) magnitude of change shown as delta-RPKM between S-phase fractions for active H3K27ac versus H3K27me2 at genes in Polycomb domains grouped by replication timing. Filled dots are medians with a significant change in either mark. Insignificant change versus the null of no change is denoted by no fill. Arrows show the direction of the cell cycle time course between S-phase fractions. (F) Same as (E) but for H3K27me2-enriched genes.



Supplementary Figure 10. Definition of H3K27-marked regions for TRIP analysis in Figure 5: Fingerprint analysis[125] of H3K27 modifications and active RNAPII steady-state profiles performed in 50-basepair (bp) intervals. Adjacent bins above a slope threshold of 1 were merged, and intervals <300bp were culled. This approach yielded consistent results down to ~10 million reads per mark. (A) Cumulative sum in ranked 50bp bins. Slope threshold of 1 is shown as a dotted line. (B) Percent genome covered by intervals > 300bp per mark. (C) Benchmarking of fingerprint analysis regions via down sampling the reads used to make the coverage file. Metrics (y-axis) versus down sampled reads (x-axis) are shown from top-to-bottom: (i) % genome coverage, (ii) fraction of reads in fingerprint regions, (iii) # of regions, and (iv) mean region width. (D) Regions ranked by the sum of normalized counts. Statistics are shown for the top 5% of intervals. (E) Median normalized coverage of H3K27me3, H3K27me2, H3K27me1, H3K27ac, and RNAPIIS2/5p centered over the top 5% of regions per mark. Regions that did not overlap those of other marks were used for TRIP analysis, except for regions common to H3K27me1 and H3K27ac.



Supplementary Figure 11. Description and examples of early replicating genes in and outside Polycomb domains: (A-B) Gene ontology (GO) enrichment analysis of early replicating genes in and outside Polycomb domains for (A) Biological process and (B) Molecular function generated using the String database[126]. Categories were ranked by number of overlapping genes. (C-D) Examples of (C) Polycomb domain genes and (D) H3K27me2-enriched genes in the 'Response to stimulus' Biological process GO category that encode cytokine signaling molecules. UCSC browser snapshots show of

coverage profiles for Active RNAPII (yellow), H3K27me3 (red), and H3K27me2 (black) in S-phase fractions around the gene loci, highlighted in blue. Each mark is group auto-scaled. Replication timing in K562 cells[100] is show in S-phase fractions at the top. Gene annotation is Gencode v49 for hg38. Genes in (C) are *IFNL1, IL19, IL10, IL17C, IL20, and IL10*. Genes in (D) are *ILII, LTA, TNF, LTB, IL17B, and CCL24.*

2.6 Methods

Cell culture

Human K562 cells (ATCC, Manassas, VA, Catalog #CCL-243) were cultured following the supplier's protocol. For inhibitor treatments, stock solutions in DMSO were diluted in cell culture medium and added directly to the cultures at the specified final concentrations. The PRC2 inhibitors used were EPZ-6438 at 10 μM (Selleckchem, Catalog # S7128), EED-226 at 10 μM (Selleckchem, Catalog # S8496). DMSO at 1:1,000 v/v was used as control. Cells were harvested after 8 hours of treatment and processed for CUT&Tag profiling.

Drosophila S2 cells (RRID: CVCL_Z232) were grown to log phase in HYQ-SFX Insect medium (Invitrogen) supplemented with 18 mM L-glutamine prior to harvest. Cell counts and cell sizes were measured using the Vi-CELL XR Cell Viability Analyzer (www.beckman.com).

Antibodies

We used the following antibodies: Guinea Pig anti-Rabbit IgG (Heavy & Light Chain) (Antibodies Online, Cat# ABIN101961), H3K27me3 (Rabbit monoclonal anti-H3K27me3, Cell Signaling Technology, Cat# 53207), H3K27me2 (Rabbit polyclonal anti-H3K27me2, Abcam, Cat# ab24684), H3K27me1 (H3K27me1 Recombinant Polyclonal Antibody, Thermo Fisher Scientific, Cat# 712817), H3K27ac (Rabbit monoclonal anti-H3K27me3, Cell Signaling Technology, Cat# 53207), RNAPIIS2/5P (Rabbit monoclonal Phospho-Rpb1 CTD (Ser2/Ser5), Cell Signaling Technology, Cat#

13546), H3K4me3 (Rabbit monoclonal anti-H3K4me3, EpiCypher, Cat# 13-0060), H3K4me2 (Rabbit monoclonal anti-H3K4me2, EpiCypher, Cat# 13-0027), and H3K4me1 (H3K4me1 Recombinant superclonal Antibody, Thermo Fisher Scientific, Cat# 710785). The final concentrations of the antibodies used in CUT&Tag are specified in the following sections. Antibodies were validated using the SNAP-CUTANA™ K-MetStat Panel according manufacturers guidelines (EpiCypher, Cat# 19-1002).

Fluorescence-Activated Cell Sorting

To purify S-phase fractions of K562 cells, 10 million cells were resuspended in ice-cold harvest buffer (1x PBS with 2 mM EDTA, 2% FBS, 1 mM Spermidine, 0.05% Triton-X-100) and briefly fixed with 0.1% PFA for 2 minutes. Fixation was quenched with 2.5 M glycine for 5 minutes, then cells were spun down and resuspended in harvest buffer with 2.5 µM DRAQ5 (Thermo Scientific, 62251) to stain DNA. After at least 10 minutes of incubation, 6 cell cycle fractions were isolated on a BD Discover S8 imaging flow cytometer based on DRAQ5 intensity.

CUT&Tag-direct for whole cells

CUT&Tag reactions were performed according to the CUT&Tag-direct protocol[127] with minor modifications. 25,000 cells (no treatment) or 8,000 cells (DMSO, EPZ-6438, and EED22-6 treatment) were aliquoted in harvest buffer. *Drosophila* S2 cells were added to each sample at a ratio of 1.6:1. 2 µl of Bio-Mag Plus Concanavalin A (ConA)-coated magnetic beads (Bangs Laboratories Catalog #BP531) per sample were activated and added to the cells, then incubated for 10 min at room temperature. ConA-bound cells were suspended in wash buffer (20 mM HEPES pH 7.5, 150 mM NaCl, 0.5 mM spermidine, 0.05% Triton-X100, Roche Complete Protease Inhibitor EDTA-Free tablet, 2 mM EDTA) and split into individual 0.5 ml tubes for overnight incubation with the primary antibody at a 1:50 dilution at 4°C. After removing unbound primary antibody by washing, the samples were resuspended in wash buffer containing the secondary antibody (guinea pig anti-rabbit IgG) at a 1:100 dilution and incubated at room temperature for 1 h. Following another wash, the samples were resuspended in 300-

wash buffer (wash buffer with an additional 150 mM NaCl) containing Protein AG-Tn5 (pAG-Tn5 at a 1:20 dilution, EpiCypher, Catalog #15-1117) and incubated at room temperature for 1 h. The samples were then washed in 300-wash buffer and resuspended in tagmentation buffer (300-wash buffer with 10 mM $MgCl_2$), followed by incubation at 37°C for 1 h to complete the Tn5 tagmentation reaction. The samples were washed with TAPS wash buffer (10 mM TAPS with 0.2 mM EDTA) and resuspended in 5 µl of release solution (10 mM TAPS, 1% SDS, and 1:10 Thermolabile Proteinase K (New England Biolabs, Catalog #P8111S)). They were then incubated in a thermocycler with a heated lid at 37°C for 1 h and 58°C for 1 h to release fragments. To quench SDS, 15 µl of 6% Triton-X100 was added. PCR was performed by adding 2 µl each of barcoded 10 mM i5 and i7 primer solutions and 25 µl of premixed KAPA PCR Master Mix (10 µl of HiFi buffer, 1.5 µl of 10 mM dNTPs, 1 µl of KAPA HiFi polymerase, and 29.5 µl of H_2O) (Roche, Catalog #07958846001). The following cycling conditions were used: Cycle 1: 58 °C for 5 min; Cycle 2: 72 °C for 5 min; Cycle 3: 98 °C for 30 s; Cycle 4: 98 °C for 10 s; Cycle 5: 60 °C for 10 s; Repeat Cycles 4-5 11 times; 72 °C for 1 min; Hold at 12 °C. CUT&Tag libraries were cleaned with HighPrep Paramagnetic bead-based post PCR clean-up reagent (MagBio, Catalog # AC-60500) at a 1.3:1 (vol/vol) ratio of beads to sample, quantified on the Agilent 4200 D1000 TapeStation (Agilent, Catalog #5067-5584) and pooled for sequencing.

DNA sequencing data processing

The size distributions and molar concentration of libraries were determined using an Agilent 4200 TapeStation and libraries were mixed to achieve equal representation as desired aiming for a final concentration as recommended by the manufacturer. Paired-end 50x50 bp sequencing on the Illumina NextSeq 2000 platform was performed by the Fred Hutchinson Cancer Center Genomics Shared Resources and data were analyzed as described (https://www.protocols.io/view/cut-amp-tag-data-processing-and-analysis-tutorial-e6nvw93x7gmk/v1).

For processing sequencing data, we used cutadapt 4.4[128] to trim adapters from 50 bp paired-end reads with parameters “-j 8 --nextseq-trim 20 -m 20 -a

AGATCGGAAGAGCACACGTCTGAACTCCAGTCA -A
AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -Z”

We used Bowtie2 2.5.1[129] to map the paired-end 50 bp reads to the hg38 human genome reference sequence from UCSC with parameters "--very-sensitive-local --soft-clipped-unmapped-tlen --dovetail --no-mixed --no-discordant -q --phred33 -I 10 -X 1000".

To remove spike-in reads, hg38 positions mapped to by reads from S2 profiles were blacklisted and reads that completely overlapped with of these intervals were removed using bedtools intersects. The spike-in reads were not used for data analysis.

Data analysis and visualization

We used bedtools[130] genomecov to generate bedGraph files, which were then converted to bigWig format using bedGraphToBigWig. Coverage-normalized bigWig files represent the fraction of counts at each base pair, scaled by the size of the reference hg38 genome (3,298,912,062), ensuring that if the counts were uniformly distributed, each position would have a value of 1. BigWigs were visualized in the UCSC genome browsers.

Reads overlapping the longest transcript of coding genes defined by gencode v48 plus one kilobase upstream were quantified using the Featurecounts function of the subread package. RPKM values per gene were calculated by dividing by total counts in millions and interval length in kilobases and then averaged across replicates.

Replicating timing assignment

Processed replication timing data in K562 cells[100] was obtained from the UCSC table browser for hg19 and lifted over to hg38 using CrossMap. Replication timing per gene was calculated as the average wavelet-transformed signal and inverted such that 0 is early and 1 is late. Six replication timing bins were fit to the distribution at even intervals between the 0.2 and 0.98 percentiles.

Promoter activity quantification

Gene promoter regions were defined in[107] and processed promoter values for GroCap and SuRE were obtained from the supplement. Processed TRIP data[107] was also obtained from the supplement.

H3K27 gene category assignment

Polycomb domains for 127 roadmap epigenomes[96] were obtained from the WUSTL portal and defined as continuous ChromHMM intervals labelled “Repressed Polycomb”, “Weak Repressed PolyComb”, “Bivalent Enhancer”, “Flanking Bivalent TSS/Enh”, or “Bivalent/Poised TSS” greater than 1 kilobase. Genes were defined to be in Polycomb domains if they overlapped by at least 50%. The remaining genes were assigned to a H3K27 modification state based on the most abundant mark in S-phase fractions above a minimum threshold of 0.5 RPKM.

Quantification and statistical analysis

For each CUT&Tag experiment, at least two biological replicates were included. Statistical comparisons between groups were performed using Pearson correlation and linear regression, as well as the Kolmogorov-Smirnov, Mann-Whitney, and Wilcoxon tests implemented in SciPy 1.17.0[131]. Bonferroni correction was applied to adjust for multiple comparisons and only adjusted p-values below the corrected threshold of 0.001 were considered statistically significant unless otherwise noted.

Data and code availability

Raw data is available at GEO under accession XXX. Processed data is available at Zenodo under https://doi.org/10.5281/zenodo.19489947. Data processing and plotting scripts are publicly available on GitHub @ https://github.com/jacob-greene/H3K27me2. An interactive UCSC browser session with the manuscript data is available @ https://genome.ucsc.edu/s/jegreene/H3K27me2_manuscript.

Chapter 3. Scalable single-cell chromatin profiling with sciCUT&Tag

This chapter is adapted from:

Janssens, D. H. & Greene, J. E. *et al.* Scalable single-cell profiling of chromatin modifications with sciCUT&Tag. *Nature Protocols* **19**, 83-112 (2024).

3.1 Introduction

Cleavage Under Targets & Tagmentation (CUT&Tag) is an antibody-directed *in situ* chromatin profiling strategy that is rapidly replacing immune precipitation-based methods, such as ChIP-seq. The efficiency of the method enables chromatin profiling in single cells but is limited by the numbers of cells that can be profiled. Here, we describe a combinatorial barcoding strategy for CUT&Tag that harnesses a nanowell dispenser for simple, high-resolution, high-throughput, single-cell chromatin profiling. In this single-cell indexed CUT&Tag (sciCUT&Tag) protocol, lightly cross-linked nuclei are bound to magnetic beads and incubated with primary and secondary antibodies in bulk and then arrayed in a 96-well plate for a first round of cellular indexing by antibody-directed-Tn5 tagmentation. The sample is then repooled, mixed and arrayed across 5184 nanowells at a density of 12-24 nuclei per well for a second round of cellular indexing during PCR amplification of the sequencing-ready library. This protocol can be completed in two days by a research technician, and we illustrate the optimized protocol by profiling histone modifications associated with developmental gene repression (H3K27me3) as well as transcriptional activation (H3K4me1-2-3) in human peripheral blood mononuclear cells and use SNPs to facilitate collision removal. We have also used sciCUT&Tag for simultaneous profiling of multiple chromatin epitopes in single cells. The reduced cost, improved resolution and scalability of sciCUT&Tag make it an attractive platform to profile chromatin features in single cells.

3.1.1 Development of the Protocol

Within the nucleus, genomic DNA is tightly packaged by histone octamers and DNA-binding proteins into chromatin, such that only a small minority of the genome is accessible to transcriptional machinery. Chromatin proteins and their modified states are the molecular determinants of protein-DNA interactions, and therefore scalable, genome-wide profiling of chromatin modifications is a powerful method to map the molecular mechanisms that govern not only transcription[132], but also cell-fate specification[133].

To quantify chromatin features genome-wide, we previously introduced Cleavage Under Targets and Tagmentation (CUT&Tag)[87], a precipitation-free method that scales to single cells[85,134]. CUT&Tag uses a Tn5 cut-and-paste transposase fused to a protein A moiety (pA-Tn5), which tagments antibody-targeted genomic loci with PCR-ready adapters. As the Tn5 fusion protein remains bound to DNA following magnesium-catalyzed tagmentation, fragments are retained within intact cells, which makes CUT&Tag compatible with single-cell profiling. Single-cell resolution can then be achieved by partitioning the sample with either a droplet-based approach or nanowell dispenser and then introducing a cell-specific barcode by PCR that is read out during next-generation sequencing.

Since the advent of CUT&Tag, derivative methods have combined single-cell CUT&Tag with strategies for profiling cell features other than chromatin, such as mRNAs[135] and cell-surface proteins[136]. This complementary information can then be used *a priori* to ascribe cell identity and generate a representative chromatin landscape of each cell type by aggregating the scCUT&Tag profiles. However, the genome-wide distributions of numerous chromatin modifications are tightly correlated with both tissue and cell type[96]. Thus, by obtaining sufficient data quality, unbiased high-resolution clustering and cell type identification is possible using the scCUT&Tag chromatin profiles alone. Here, we expand on the scCUT&Tag technology using a sample partitioning method called combinatorial indexing to dramatically increase the number of individual cells that can be profiled per experiment while maintaining precise control of the CUT&Tag reaction chemistry at each step. We provide a detailed description of the single-cell indexed CUT&Tag (sciCUT&Tag) protocol as performed on a nanowell dispenser. This sciCUT&Tag protocol is simple, highly reproducible, and is suitable for parallel profiling of multiple samples and epitope combinations in a single experiment. By multiplexing samples, sciCUT&Tag not only streamlines the data acquisition, but also facilitates data integration by limiting batch-to-batch variation and provides an opportunity to acquire biological replicates that

power downstream statistical analysis. This sciCUT&Tag protocol was first used for the Multiple Target Identification by Tagmentation (MulTI-Tag)[137] method that we recently introduced to profile multiple chromatin epitopes in individual cells. The high recovery of both cellular input and individual fragments in MulTI-Tag is enabled by an underlying single-cell partitioning method called combinatorial indexing, which facilitates the precise control of CUT&Tag reaction chemistry at each step.

Relative to droplet-based approaches, combinatorial indexing exponentially increases the number of single cells possible to profile per experiment for RNA-seq[138], ATAC-seq[139] and CUT&Tag[135,140,141]. Furthermore, combinatorial indexing presents the opportunity to optimize chemistry for fragment recovery without needing to also optimize chemistry for droplet formation. Through a single round of indexing, pooling, splitting and indexing, reads are labeled and then grouped by unique combinations of barcodes that statistically correspond to an individual cell of origin.

Two key variables can be scaled to increase the cellular yield from a combinatorial indexing experiment: number of split-pooling rounds and number of indices per round. However, scaling either or both increases the probability of technical error. Step-by-step protocols[142,143] have therefore been introduced to streamline and standardize combinatorial indexing experiments. In the sciCUT&Tag protocol, we minimize the number of split-pooling rounds by increasing the number of indices per round from 96 to 5184 with the TaKaRa ICELL8 nanowell system[144]. We find that employing a robotic system for combinatorial indexing achieves a high degree of cell-per-well accuracy and tunability.

sciCUT&Tag builds on the bulk CUT&Tag protocol[145] in which nuclei are first isolated by a detergent wash and then immobilized on paramagnetic beads to facilitate supernatant removal while preserving yield (**Fig. 1a**). Incubation with primary and then secondary antibodies is then performed in bulk for each sample and epitope under investigation. The sciCUT&Tag workflow can accommodate simultaneous profiling of multiple samples, as individual sample/epitope combinations are then arrayed into a 96-well plate. Uniquely barcoded pA-Tn5 transposomes are then added to each well and tagmentation is performed in a 300 mM NaCl-containing buffer to add the first index *in situ*. In a streamlined combinatorial indexing paradigm, we perform only one subsequent round of indexing after pooling all nuclei from the 96-well plate and then splitting into a 5184-nanowell chip via a robotic device that facilitates reproducibility by reducing technical variation across experiments. Lastly, lysis is followed by PCR amplification to add the second index via barcoded primers in 5184-well format prior to sequencing.

To leverage informatics (**Fig. 1b**) for the removal of barcodes that correspond to 'collisions' where two nuclei stick together, we recommend designing sciCUT&Tag experiments to include samples with distinct genotypes that can be read out per cell to identify cross-genotype collisions. This incorporates a built-in quality control check, as certain tissues or cells may be 'stickier' than others. For sciCUT&Tag, we increase the sequencing read length relative to bulk CUT&Tag, as the increased genomic information facilitates deconvolution of multiple genotypes *de novo*. Barcoded read groups that contain multiple overlapping single-nucleotide polymorphisms (SNPs) are then removed prior to downstream analysis with dimensionality reduction, graph-based clustering, and cell type annotation.

3.1.2 Applications of the Method

In the development of this protocol, we first optimized sciCUT&Tag single-cell acquisition parameters by profiling a mixture of human and mouse cell lines. As an example of the protocol, we applied sciCUT&Tag to profile a mixture of human peripheral blood mononuclear cells (PBMCs) from two healthy donors with (1) an antibody cocktail that recognizes histone H3 lysine 4 mono-, di-, and tri- methylation (H3K4me1-2-3) and labels transcriptionally active promoters and enhancers, and with (2) a single monoclonal antibody that recognizes histone H3 lysine 27 trimethylation (H3K27me3), which labels Polycomb-associated, developmentally repressed chromatin. We show that single cell profiles of either chromatin modification are sufficient to perform high-resolution clustering and *de novo* identification of the cell types found in human PBMCs. In addition, we previously used the multifactorial variation of sciCUT&Tag, MultiTag, to resolve cells along a trilineage differentiation from human embryonic stem cells to definitive endoderm, mesoderm or neuroectoderm[137].

In principle, sciCUT&Tag can be applied to any biological sample that is compatible with the isolation of lightly cross-linked nuclei. For example, this sciCUT&Tag platform is well suited for profiling patient samples to identify rare cell types or cell states associated with disparate clinical outcomes. The sciCUT&Tag method is comparable to single-cell RNA-seq and ATAC-seq methods that have facilitated comparative genomic analysis of human blood samples to identify gene regulatory changes in response to inflammatory stresses, such as those induced by COVID-19 infection, as well as inter- and intra-tumoral heterogeneity[146-148]. This sciCUT&Tag platform will enable investigators to

expand these comparative genomic analyses to include other critical aspects of the epigenome, including regions maintained in developmentally repressed states.

3.1.3 Comparison with other methods

Improvements have been made to sciCUT&Tag since it was first introduced[137] that increase both reads/cell as well as cell recovery. Compared to other single-cell CUT&Tag approaches[85,134-136,140,141,149], sciCUT&Tag offers at least a four-fold increase in throughput. Importantly, the sciCUT&Tag method drastically reduces the cost per sample via scalable multiplexing. With optimal loading conditions across an entire experiment (12-24 cells/nanowell), we estimate that sciCUT&Tag achieves a consistent yield of ~40,000 cells/chip, which corresponds to ~0.11 USD per cell in library preparation and sequencing costs. In comparison, standard droplet-based kits currently cost ~0.85 USD per cell, while obtaining fewer unique reads per cell than this sciCUT&Tag platform. Because each ICELL8 nanowell chip costs only 600 USD, it is cost-efficient to dispense the same pool of tagmented nuclei multiple times to generate additional sequencing libraries and capture more unique single-cell profiles per experiment. The sciCUT&Tag method also represents a near doubling in reads/cell versus our initial scCUT&Tag[85] method (2116 versus 1110 median reads/cell, respectively, for H3K27me3 in human PBMCs) and is comparable to methods that use linear amplification-based library preparation to bolster reads/cell[134,140].

3.1.4 Overview of the procedure

The sciCUT&Tag method for single-cell chromatin profiling can be performed on the benchtop and completed in 1.5 days using the TaKaRa ICELL8 robotic nanowell system. In brief, lightly cross-linked nuclei are prepared and immobilized on wheat germ agglutinin (WGA)-coated paramagnetic beads (**Steps 1-19; Fig. 1a**)[150]. Bead-bound nuclei are incubated with a primary antibody followed by incubation with a secondary antibody to increase the number of IgG molecules at each epitope bound by the primary antibody (**Steps 20-30; Fig. 1a**). Bead-bound nuclei are washed and arrayed in a 96 well plate for incubation with protein A-Tn5 loaded with differentially barcoded mosaic-end adapters and washed under stringent conditions (**Steps 31-40; Fig. 1a**). Tn5 is

activated by addition of Mg^{2+}, and tagmentation is carried out at 37 ^{o}C (**Steps 41-42; Fig. 1a**). Tagmented bead-bound nuclei are repooled, washed and passed through a 20 µm Mini-Strainer to remove cellular aggregates, and are then dispensed on the ICELL8 into a 5184 350v chip, imaged and prepared for PCR in 0.19% SDS Release buffer (**Steps 43-83; Fig. 1a**). Triton-X neutralizing solution and PCR reagents are added through a series of dispenses on the ICELL8 and sciCUT&Tag libraries are enriched by PCR amplification (**Steps 84-107; Fig. 1a**) and two Solid Phase Reversible Immobilization (SPRI) magnetic bead cleanup steps (**Steps 108-125; Fig. 1a**). The sciCUT&Tag library is quantified on an Agilent TapeStation capillary gel analyzer and diluted for sequencing on an appropriate flow cell as determined by the initial library concentration (**Steps 126-141; Fig. 1a**). After sequencing, data processing steps142-148 demultiplex samples according to the four-barcode combinatorial indexing scheme using the sciEXtract custom code, align fastq files to the genome of interest and remove doublets using SNP-based detection or using a nearest neighbor approach with synthetic doublets. Dimensionality reduction and clustering is performed and annotated based on distribution of chromatin features surrounding marker genes.

3.1.5 Experimental Design

For sciCUT&Tag, we recommend designing experiments to include multiple samples with distinct genotypes, as this helps detect ‘collisions’ where a barcode combination corresponds to more than one cell. As few as two samples can be mixed prior to immobilization on paramagnetic beads for this approach. **To keep nuclei intact and to reduce the formation of multi-nuclear aggregates,** light cross-linking is critical.

While multiple epitopes can be profiled in parallel with sciCUT&Tag[137], the most abundant epitopes should be prioritized to increase reads/cell yield, which is essential for dimensionality reduction during data analysis. The number of nuclei per antibody can vary greatly depending on the experimental design. For example, for an experiment in which two different histone modifications will be profiled in an equal number of nuclei, the sample should be split evenly in two tubes and processed in parallel and allocated a corresponding proportion of the 96-well plate for sci-pA-Tn5 binding and tagmentation. Pooling antibodies may also be used to garner sufficient analyzable signal for less

frequent epitopes, and we expect that antibodies validated by the manufacturer for CUT&RUN or immunofluorescence are most likely to work for this application. Throughout the protocol, we recommend incubating the samples with a relatively high concentration of the primary antibody (1:10 dilution by volume) and secondary antibody (1:20 dilution by volume) as well as the sci-pA-Tn5 complexes (1:10 dilution by volume) in order to saturate the binding of the chromatin protein of interest and increase the number of unique reads obtained per cell. We have found the secondary antibody step is required for CUT&Tag to increase the number of Protein A binding sites for each chromatin target. Without the secondary antibody, the efficiency of tagmentation is very low. The immunoglobin-binding moiety of pA-Tn5 binds tightly only to certain IgG isotypes of some host species, for example rabbit and guinea pig IgG, but poorly to mouse IgG_1 and goat IgG. Thus, the secondary amplifying antibody must be chosen to provide high affinity for protein-A. For example, for mouse primary antibodies we recommend a rabbit anti-mouse antibody be used.

In order to increase the number of single cell profiles captured in each experiment, the second round of indexing is performed using the ICELL8 nanowell dispenser. Because the bead-bound nuclei slowly sink, which can cause the ICELL8 chip to be loaded unevenly, we use the “Pause and Dispense protocol”, available upon request from TaKaRa technical support. This custom program will load the negative control and then pause three times during the cell dispense process to allow for sample mixing immediately prior to aspiration of the sample solution by the ICELL8.

During sciCUT&Tag library sequencing, the length of the barcode reads is fixed and read out using 43 cycles for the second read and 37 cycles for the third read. The length of the genomic reads can be variable depending on the application. When mixing samples from multiple donors for SNP-based doublet detection, it is important to obtain sufficient genomic information to detect donor specific SNPs. For this we have used either the Next-Seq P2-200 flow cell, which yields 79 bp by 79 bp paired ends (79 x 43 x 37 x 79) or the Next-Seq P1-300 flow cell, which yields 129 bp by 129 bp paired ends (129 x 43 x 37 x 129).

3.1.6 Example validation of the method

To optimize our sciCUT&TAG protocol, we identified split-pooling and sample loading conditions that minimize the number of cellular collisions arising from two or more cells receiving the same combination of barcodes, and also empirically determined the H3K4me1-2-3 and H3K27me3 modification specific parameters for *in silico* dimensionality reduction that robustly separate the major cell types found in human peripheral blood. These steps were necessary for development of the protocol, but most users can follow the guidelines here and do not need to repeat optimization experiments.

3.1.7 Loading conditions for the ICELL8 5184-well chip

We first estimated the rate of 'collisions' by overloading the ICELL8 5184-well chip in a 'barnyard' experiment that mixed human and mouse cell-lines at a 1:1 ratio and dispensed them across a serial dilution of 40 to 10 cells/nanowell in duplicate (Supplementary Fig. 1a and b). Collisions were called as barcodes with fewer than 90% of reads uniquely aligned to either hg38 (human) or mm10 (mouse). We found that collisions dominated the low-content barcodes with fewer than 500 reads (Supplementary Fig. 1c; grey box). This allowed us to set a low threshold of at least 500 reads (Supplementary Fig. 1c; black dashed line) for a barcode to be called as a single cell. We excluded these low-content barcodes when calculating the rate of collision across the serial dilution of dispense targets (Supplementary Fig. 1d). For the dispense target of 40 nuclei/nanowell, we observed the highest collision rate of 21.2%; and this condition was also estimated to have the highest yield at just under 40,000 cells/chip. In line with the serial dilution, cells/chip yield decreased ratiometrically with the dispense target and the rate of collisions also decreased in accordance with the number of nuclei dispensed per nanowell. Specifically, at 20 nuclei/nanowell the rate of collisions was 17.7%, at 10 nuclei/nanowell the rate of collisions was 16.0%, and at 5 nuclei/nanowell the rate of collisions fell below 12%. We conclude that the number of nuclei loaded per nanowell must be carefully controlled to limit collisions and obtain an optimum number of single-cell profiles per experiment.

As expected from overloading the barnyard experiment, we observed that barcodes/well across both the 96-well plate and 5184-well chip highly correlated with low-content barcodes and collisions (Supplementary Fig. 1c and d). Given that nuclei are barcoded by well in combinatorial indexing, a well that is overloaded in the 96 well plate during the first round of indexing will give rise to a higher frequency of collisions during the second round of indexing. In addition, overloading may cause the tagmentation and/or PCR chemistry to become rate-limiting, which can potentially reduce reads/barcode and weaken analytical power during dimensionality reduction. Indeed, we observed that that one well in the 96-well plate gave rise to 3126 of the barcodes recovered by sequencing, the majority of which had either low content or were a collision (Supplementary Fig. 1a). Furthermore, we observed that barcodes/well strongly correlated with low-content cells/well across both the 96-well plate (Pearson's $r = 0.99$) and 5184-well chip (Pearson's $r = 0.93$) (Supplementary Fig. 1e and f). Concordantly, collisions/well also exhibited a high correlation with barcodes/well across both the 96-well plate (Pearson's $r = 0.90$) and 5184-well chip (Pearson's $r = 0.63$).

While this pattern was maintained when correlating barcodes/well and collisions with high-content cells/well (e.g., barcodes with >500 reads) for the 96-well plate (Supplementary Fig. 1g; Pearson's $r = 0.91$ and 0.88, respectively), it was not well-maintained for the 5184-well chip (Supplementary Fig. 1h; Pearson's $r = 0.43$ and 0.49). This suggests two important conclusions: (1) that events which produce low-content barcodes and collisions originate in the 96-well plate, and (2) that the nanowells which yield the most cells above threshold are distinct from those which yield low-content barcodes and collisions. While overloaded chemistry in the 5148-cell chip can account for the second observation, possible explanations for the first observation are that clumping and/or lysis in the 96-well plate can increase total barcode count at the expense of low-content barcodes and collisions. To limit clumping and lysis, we emphasize careful pipetting with wide-bore tips and filtering the pooled tagmentation product with a 20 µm Mini-Strainer prior to dispense on the ICELL8.

Given the possibility of high collision frequency in combinatorial indexing experiments, we implemented a collision removal strategy based on *de novo* genotyping that can be

widely applied to any tissue or cell line for which biological replicates with distinct genotypes can be obtained. As little as two samples can be mixed prior to immobilization on paramagnetic beads for this strategy and multiple epitopes can be simultaneously profiled in parallel in a single sciCUT&Tag experiment. We demonstrated this scalable multiplexing approach by mixing PBMCs from two donors and profiling both H3K27me3 and H3K4me1-2-3 across the mixed sample as well as the single donor samples representing ground truth. Both H3K27me3 and H3Kme1-2-3 profiles exhibited classic 146 base-pair nucleosomal laddering (Fig. 2a), and we recovered a greater number of reads/cell above the low-content threshold of at least 500 reads/cell (Fig. 2b; black dashed line) for H3K27me3 (μ = 2649 reads/cell) than H3K4me1-2-3 (μ = 1420 reads/cell).

We next built on the Souporcell[151] tool for identifying cross-genotype collisions based on single nucleotide polymorphisms (SNPs) in mixed donor samples, integrating it into an easy-to-use and reproducible Nextflow pipeline[152,153]. The pipeline combines sciCUT&Tag genome alignments for multiple epitopes into a single Souporcell run to maximize sequence space for *de novo* genotyping, and then outputs barcoded bed files, principal component analysis (PCA) of Souporcell cluster assignment probabilities, and tabulated cell metadata (Fig. 2c). This collision removal strategy may be insufficient to distinguish two unique genotypes from sciCUT&Tag profiles generated by paired-end 25 x 25 or 50 x 50 base sequencing, so we increased the read length to paired-end 79 x 79.

We benchmarked the pipeline in the mixed-donor PBMC experiment by first counting the cells assigned to each genotype and mapping them to their donor of origin. Indeed, for both the H3K27me3 and H3K4me1-2-3 profiles, Donor A associated with genotype 1 and Donor B associated with genotype 0 (Fig. 2d). As expected, the mixed donor population was evenly distributed across both genotype assignments with only a small intermediate population. Because only two donors were mixed, we were able to plot the cluster assignment probabilities against each other instead of by PCA, which revealed that genotypes 0 and 1 assignments neatly separated, and collisions aligned with the small intermediate population (Fig. 2e). The collision population also scored significantly

higher than single-genotype cells for DoubletEnrichment in the artificial nearest neighbor (ANN) approach implemented in ArchR[154], which corroborated their mixed-phenotype character (Supplementary Fig. 2).

Including the single-donor ground truth samples in addition to the mixed donor sample provided us with a comprehensive basis to assess the accuracy of Souporcell[151] for identifying cross-genotype collisions in sciCUT&Tag (Fig. 2f). Indeed, we observed nearly zero collisions in the single-donor ground truth samples across dispense targets. Compared to H3K27me3 (μ = 7.0% collision rate; max = 8.5%), H3K4me1-2-3 exhibited a higher percentage of collisions across the serial dilution (μ = 13.0%; max = 17.2%). Together, efficient combinatorial indexing on ICELL8 system augmented by *de novo* genotyping and SNP-based collision removal comprise a robust approach to single-cell chromatin profiling in sciCUT&Tag. Importantly, while we were able to obtain a ground truth of cellular collisions using Souporcell, the accuracy of SNP-based doublet callers is highly dependent on the data quality, including the number of reads obtained per cell, the genetic diversity of the donor samples to be mixed, the length of the genomic reads and the fraction of reads that overlap one another to allow SNP comparisons. So, while Souporcell is a viable approach for calling doublets in H3K4me1-2-3 and H3K27me3 sciCUT&Tag data, whether this extends to sciCUT&Tag profiles of other epitopes must be determined empirically.

3.1.8 Tailoring dimensionality reduction to the biology of different chromatin epitopes.

The advantage of single-cell approaches like sciCUT&Tag over bulk CUT&Tag are that they enable identification of cell populations within heterogenous samples. By parsing sparse chromatin profiles with unsupervised dimensionality reduction, latent representations are revealed that drive clustering of like cells for downstream analysis. Across single-cell analyses, the input to dimensionality reduction algorithms is a cell-by-feature matrix, and the output is a latent representation with a condensed number of dimensions/cell that can be further reduced to two for visualization with algorithms such as uniform manifold approximation and projection (UMAP)[155], 2018 #157}. To identify the optimal dimensionality reduction parameters for different types of chromatin, such

as variably phased heterochromatic epitopes that form broad domains and tightly positioned nucleosomes at promoters, we tested a range of cell-by-genomic-bin feature representations for H3K27me3 and H3K4me1-2-3 profiling in PBMCs, which we expected to separate into distinct immune cell types based on previous analysis[85].

To systematically optimize dimensionality reduction of sciCUT&Tag data, we used the iterative latent semantic indexing (LSI) approach implemented in ArchR[154] to reduce matrices with 500 bp, 5 kb, and 50 kb genomic bins. Information content and noise for the first 45 dimensions of LSI was estimated for each binning paradigm as proportional variance and correlation to reads/cell, respectively. As a realistic next step for data processing, we then computed UMAPs for the first 15, 30, and 45 LSI dimensions and shaded them by reads/cell to check for embedding bias. For the repressive epitope H3K27me3, we found that increasing genomic bins to 50 kb pushed nearly all variance into early LSI dimensions, while correlation to reads/cell remained distributed into late dimensions (**Fig. 3a**). Concordant with the low variance in late dimensions, inputting more than 15 dimensions to UMAP did not improve separation across genomic bin widths (**Fig. 3b**). For H3K27me3, these findings advocate for representation of broad domains by larger genomic intervals.

While increasing bin width may improve dimensionality reduction for broad H3K27me3 domains, it can reduce resolution by merging distinctly regulated loci in larger features. For H3K4 methylation that adorns well-phased nucleosomes in active domains like transcriptional start sites (TSSs), smaller bins can maximize feature resolution. Indeed, for H3K4me1-2-3, unlike for H3K27me3, we found that increasing genomic bins to 50 kb did not increase early dimension variance, but rather increased correlation to reads/cell in early LSI dimensions (**Fig. 3c**), and polarized UMAP embeddings by reads/cell (**Fig. 3d**). In contrast, we observed the best UMAP separation for H3K4me1-2-3 with 500 bp bins and 15 LSI dimensions. Not only does this confirm that fewer dimensions are sufficient for UMAP embedding of sparse chromatin data, but also that smaller genomic intervals are better suited to represent narrow chromatin domains, such as those marked by H3K4 methylation.

3.1.9 Limitations

The high throughput that we obtain requires the 72x72 nanowell chip used in the ICELL8 robotic system, which is not as widely available as droplet systems. However, the strategy described here can be carried out in microtiter plates. For example, using a 384-well plate for the split in place of the 5184-nanowell chip will provide up to ~4000 single cells.

As in bulk CUT&Tag, tagmentation during sciCUT&Tag is performed under 300mM NaCl to limit off-target localization by pA-Tn5 to accessible DNA. While high-salt conditions ensure accurate tagmentation at the targeted epitope, they also competitively reduce occupancy of transcription factors (TFs), making sciCUT&Tag best for higher-affinity targets, such as histone modifications. Of note, histone modifications are also the preferred epitopes for single-cell profiling because of their high abundance in most cell types, and it is rare for single-cell methods to be applied to TFs because of their low abundance, which limits the power of dimensionality reduction. Bulk CUT&RUN remains the preferable method for both base-pair resolution and high signal-to-noise profiling of TFs.

3.2 Materials

Biological materials

- Cell lines. We have used human K562 cells (ATCC; CCL-243; https://scicrunch.org/resolver/RRID:CVCL_0004), H1 cells (WiCell; WA01-lot#WB35186; https://scicrunch.org/resolver/RRID:CVCL_9771), mouse 3T3 cells (https://scicrunch.org/resolver/RRID:CVCL_0594) and other mammalian cell lines, as well as healthy human peripheral blood mononuclear cells (PBMCs). A description of the PBMC acquisition and processing is provided in the Supplementary methods.
  CAUTION: The cell lines used in your research should be regularly checked to ensure they are authentic and are not infected with mycoplasma. When working with primary human samples ensure the study complies with all relevant

governmental and institutional review boards, and that the informed consent is obtained from all study participants.

Reagents

- 3XFlag-pA-Tn5-Fl (pA-Tn5) plasmid (Addgene, cat. no. 124601)
- 5 mL of purified Protein A–Tn5 (pA-Tn5) fusion (https://www.protocols.io/view/3xflag-patn5-protein-purification-and-meds-loading-j8nlke4e5l5r/v1). This is sufficient to prepare a loaded 96-well sci-pA-Tn5 for up to 30 experiments.
- Tn5 Mosaic-End Adapters: 1 universal MED-REV oligo as well as 8 s5 adapter oligos and 12 s7 adapter oligos with unique barcodes are listed in Supplementary Table 1. We ordered s5 and s7 oligos from IDT prepared by PAGE purification on a 1 µmole scale.
- PCR primers: 72 i5 PCR primers and 72 i7 PCR primers with unique barcodes adapted from Buenrostro et al[156] are listed in Supplementary Table 1. We order i5 and i7 PCR primers from IDT, in salt-free format on a 25 nmole scale; higher purification formats are optional. Our PCR conditions are optimized for these custom primers.

  CAUTION: Do not use Nextera primers, which will not work efficiently.
- Custom sequencing primers: 4 Custom oligos for sequencing the sciCUT&Tag library on an Illumina NextSeq 2000 instrument are listed in Supplementary Table 1. We order these primers from IDT using standard desalting conditions on a 25 nmole scale.
- Nuclease-free H_2O (Promega, cat. no. P1197)
- HEPES Sodium Salt (Sigma-Aldrich, cat. no. H3375)
- 1 M Potassium Chloride (KCl; Sigma-Aldrich, cat. no. P3911)
- 10% Triton X-100
- 2 M Spermidine (Sigma-Aldrich, cat. no. S2501)
- Glycerol (Sigma-Aldrich, cat. no. G5516)
- Roche Complete Protease Inhibitor EDTA-Free tablets (Sigma-Aldrich, cat. no. 5056489001)

- 10X Phosphate Buffer (PBS; Fisher Scientific cat. no. BP3991)
- Tween 20 (Sigma-Aldrich, cat. no. 274348-500mL)
- Pierce 16% Formaldehyde (w/v), Methanol-free (Fisher Scientific cat no. 28906)
  CAUTION: Formaldehyde is a cross-linking reagent that is hazardous if swallowed, inhaled, or absorbed through the skin. Proper PPE should be worn, and the workspace should be adequately ventilated. Preferably keep the formaldehyde solution in a fume hood to avoid inhalation.
- 0.5 M Ethylenediaminetetraacetic acid (EDTA; Research Organics, cat. no. 3002E)
- 5 M Sodium chloride (NaCl; Sigma-Aldrich, cat. no. S5150-1L)
- 5 mg/mL 4', 6-Diamidino-2-phenylindole (1000X DAPI aliquots; Sigma-Aldrich, cat. no. D9542-50 mg)
- Dimethyl sulfoxide (DMSO; Sigma-Aldrich cat. no. D4540)
- Dynabeads™ MyOne™ Streptavidin C1 (Invitrogen, cat. no. 65001). Alternatively, Magnefy™ Concanavalin A ready-to-use beads (Bangs Laboratories, Inc. cat no. MFY531) can be used in the same way.
- Wheat Germ Agglutinin (WGA), Biotinylated (Vector Laboratories, Inc., cat. no. B-1025-5)
- 1 M Manganese Chloride ($MnCl_2$; Sigma-Aldrich, cat. no. 203734)
- 1 M Calcium Chloride ($CaCl_2$; Fisher, cat. no. BP510)
- 1 M Potassium Chloride (KCl; Sigma-Aldrich, cat. no. P3911)
- 100 mM Magnesium Chloride ($MgCl_2$; Sigma-Aldrich, cat. no. M8266-100G) 3002E)
- TAPS (Millipore Sigma, cat. no. T5130-25G)
- 10% Sodium dodecyl sulfate (SDS; Sigma-Aldrich, cat. no. L4509)
- KAPA HiFi PCR Kit with dNTPs (Roche, cat no. 07958838001)
- Solid Phase Reversible Immobilization (SPRI) magnetic beads (e.g. Agencourt AMPure XP, Beckman Coulter, cat. no. A63880)
- 1 M Tris-HCl pH 8.0 (Fisher cat no. BP1521)
- Ethanol (Decon Labs, cat. no. 2716)

- Qubit dsDNA HS kit (Life Technologies, cat. no. Q32851)
- Illumina NextSeq Sequencing Reagents (e.g. P2 Reagents 200 cycles)

Antibodies

- Primary antibody to chromatin protein

  CRITICAL: High abundance epitopes such as histone modifications are optimal to obtain sufficient reads per cell and produce meaningful separation of cell types and cell states during downstream analysis. For example validation of this protocol, we used Anti-H3K27me3 rabbit monoclonal antibody (Cell Signaling Technology, cat. no. 9733); https://scicrunch.org/resolver/RRID:AB_2616029), anti-H3K4me1 rabbit oligoclonal antibody (Thermo, cat. no. 710795; https://scicrunch.org/resolver/RRID:AB_2532764), and anti-H3K4me2 rabbit polyclonal antibody (Millipore, cat. no. 07-030; https://scicrunch.org/resolver/RRID:AB_310342) primary antibodies.

- Secondary antibody to primary antibody

  CRITICAL: The secondary antibody must recognize the host type of the primary antibody and is necessary to increase the number of IgG molecules per targeted epitope. This increases the nucleation of pA-Tn5 at antibody targeted regions of the genome to ensure robust tagmentation. For rabbit primary antibodies we recommend a guinea pig anti-rabbit (Antibodies-Online ABIN101961; https://scicrunch.org/resolver/RRID:AB_10775589). For mouse primary antibodies we recommend a rabbit anti-mouse secondary antibody (e.g. Abcam ab46540; https://scicrunch.org/resolver/RRID:AB_2614925)

Cell Culture Reagents

- Matrigel (Corning cat. no. CLS354234)
- mTeSR1 Basal Media (STEMCELL Technologies cat. no. 85851)
- mTeSR1 Supplement (STEMCELL Technologies cat. no. 85852)
- RPMI 1640 Medium (ThermoFisher cat. no. 11875093)

- Glutamine and HEPES supplement (ThermoFisher cat. no. 72400047)
- FBS (ThermoFisher cat. no. 16141079)
- DMEM + Glutamax (ThermoFisher cat. no. 10566016)
- Gibco Antibiotic-Antimyocotic (Fisher Scientific cat. no. 15-240-062)

EQUIPMENT

- ICELL8 single-cell systems (TaKaRa)
- ICELL8 350v chips (TaKaRa, cat. no. 640019)
- SMARTer ICELL8 Blank Chip Reagent Kit (TaKaRa Bio USA, cat. no. 640196)
- SMARTer ICELL8 Loading Kit (TaKaRa Bio USA, cat. no. 640109)
- SMARTer ICELL8 Collection Kit (TaKaRa Bio USA, cat. no. 640048)
- Countess 3 Automated Cell Counter (ThermoFisher Scientific)
- T100 thermal cycler (BIORAD cat. no. 1861096)
- 96S Super Magnet Plate (Alpaqua cat no. A001322)
- Centrifuge Eppendorf 5810, swinging bucket
- MicroAmp Support Bases (ThermoFisher, cat. no. N801-0531)
- Centrifuge Eppendorf 5424, fixed angle rotor
- Centrifuge Eppendorf 5415R, refrigerated fixed angle rotor
- Macsimag magnetic separator (Miltenyi, cat. no. 130-092-168).
- Vortex mixer (e.g., VWR Vortex Genie)
- Micro-centrifuge (e.g., VWR Model V)
- Nutator (e.g., Corning cat. no. 6720)
- pH Meter (e.g., Thermo Scientific cat. no. VSTAR83)
- Cell culture incubator (Sanyo cat no. MCO-19AIC)
- Conical centrifuge tubes (15 mL or 50 mL)
- Cryogenic screw-cap vials (Corning cat. no. 430658)
- Reagent Reservoirs (e.g. Thermo Scientific cat. no. 8095)
- AlumaSeal 96 film (e.g., Sigma Aldrich cat. no. Z721549-100EA)
- Mr. Frosty containers (Thermo cat. no. 5100-0001)
- 96 well LoBind PCR plates Semi-skirted (Eppendorf, cat. no. 0030129504)

- 0.65 mL Prelubricated microcentrifuge tube, DNase/RNAse free (Costar, cat. no. 3206)
- 1.5-mL microcentrifuge tubes (Genesee, cat. no. 22-282)
- 1.7-mL Microtube, Low Retention, DNase/RNAse free (Genesee Scientific, cat. no. 24-282LR)
- Mini-Strainer 20 µm (PluriSelect USA, cat. no. 43-10020-40)
- 2-mL microcentrifuge tubes (Axygen, cat. no. MCT-200-C)
- Capillary electrophoresis instrument (e.g. Agilent Tapestation 4200)
- 2-20 µL 8-Channel Multi Pepette (e.g. Rainin cat. no. 17013803)
- 20-200 µL 8-Channel Multi Pepette (e.g. Rainin cat. no. 17013805)
- 200 µL Wide Bore Tip Boxes (e.g., Rainin cat. no. 30389241)
- 5-50 µL 16-Channel Multi Pipette (e.g. CappAero cat. no. C50-16)
- Illumina NextSeq2000

Software

- Souporcell (https://github.com/wheaton5/souporcell)
- Souper-star (https://github.com/FredHutch/souper-star)
- ArchR (https://github.com/GreenleafLab/ArchR)
- sciEXtract (https://github.com/mfitzgib/sciCTextract)

REAGENT SETUP

Cell culture

Cells of choice should be cultured or isolated before beginning the experiment. We have used human K562 cells, H1 human embryonic stem cells, ML-2 and RS4;11 cells, mouse 3T3 cells and PBMCs; culture and isolation conditions for these cells can be found in the Supplemental methods.

Sci-pA-Tn5 Plate:

1. Anneal each of the 8 Barcoded Mosaic end - s5 adapters (ME-s5) and the 12 Barcoded Mosaic end - s7 adapters (ME- s7) oligonucleotides with Mosaic end –

reverse oligonucleotides (**Supplementary Table 1**)[25]. To anneal, first dilute oligonucleotides to 200 µM in annealing buffer (10mM Tris pH8, 50mM NaCl, 1 mM EDTA). Mix 200 µM ME-s5 oligo with 200 µM ME-reverse oligo resulting in 100 µM ME-s5 adapter. Mix 200 µM ME-s7 oligo with 200 µM ME-reverse resulting in 100 µM ME-s7 adapter.

2. Place the tubes in a Thermocycler and run an annealing program with a 105ºC heated lid that starts with a 95°C 2-minute incubation and decreases the temperature in 5°C increments incubating 5 minutes each, ending with a 5 minute incubation at 25°C and the hold at 8°C. Keep on ice for immediate use or store at -20°C. Annealed adapter stocks can be kept at -20 ºC for at least a year.
3. Place 96 well LoBind PCR plate on ice and mix 3.2 µl of the appropriate preannealed ME-s5 adapter (100 µM) with 3.2 µl of the appropriate preannealed ME-s7 adapter (100 µM). Immediately add 40 µl of 5.5 µM pA - Tn5 fusion protein in 50% Glycerol, pipette up and down ~5 times to mix. An example of the Sci-pA-Tn5 Plate Layout is provided in **Supplementary Table 2**.
4. Incubate the mixture for 1 hour at room temperature (23-27 ºC) and then store at -20 °C until use. The loaded sci-pA-Tn5 plate is stable at -20 ºC for at least a year.

CRITICAL: The volumes of ME-s5 adapter, ME-s7 adapter and the pA-Tn5 fusion can be scaled in proportion to one another. In addition, before loading an entire 96-well plate, it is advisable to perform a serial dilution experiment adjusting the ratio of the ME-s5 and ME-s7 adapter mixture to the pA-Tn5 to ensure the loading conditions are optimized for a given pA-Tn5 prep. The optimum loading condition can then be determined empirically by running parallel samples using the various pA-Tn5 preps. The relative efficiency of pA-Tn5 loading can be determined by the CUT&Tag library yield as read out on a Tapestation or Bioanalyzer.

Sci-i5 and i7 Primer Plate:

1. Resuspend 72 barcoded i5 and 72 barcoded i7 Primers at 100 µM in 10 mM Tris pH 8 (**Supplementary Table 1**)[25].

2. Prepare a master primer plate by individually pipetting 200 µL of 100 µM of barcoded i5 and i7 primers into the appropriate wells of a 384-deep well plate (Fisher Sci cat. No. 269390) and store at -20 °C for up to one year.
3. Prepare the sciCUT&Tag PCR Primer Plates by filing the appropriate wells of a 384-well ICELL8 source plate with 90 µl of 10 mM Tris pH 8 and, using a 16-channel pipette, add 10 µl of 100 µM barcoded i5 and i7 primers from the 384-deep well plate master plate to each well.
4. Using 16-channel pipette, stamp 23 µL each of the diluted 10 µM barcoded i5 and i7 primers into four iCELL8 source plates focusing on one column at a time. For sci-i5 and i7 primer plate layout see **Supplementary Table 3**. Store sciCUT&Tag PCR Primer Plates at 4 °C for up to one month or at -20 °C long term.

WGA-coated MyOne-C1 magnetic beads[150]:

1. Prepare a 50 mL Stock of 1xPBS (pH 6.8) by diluting 10X PBS in Nuclease-Free Water then adjust pH by adding 1M HCl dropwise while monitoring with a pH Meter. This stock can be kept at room temperature for up to 1 year.
2. Prepare 5 mL of 1xPBS (pH 6.8) + 0.01% Tween-20 (v/v). 50 µL of 1% Tween20 (v/v) in 4950 µL of PBS (pH 6.8).
3. Resuspend 500 µl Dynabeads™ MyOne™ Streptavidin C1 in 1 mL 1xPBS (pH6.8) + 0.01% Tween-20 in standard 1.5mL Microfuge tube.
4. Place on magnet stand and allow to clear. Remove supernatant and discard without disturbing the magnetic beads. Resuspend in 1 mL 1xPBS (pH6.8) + 0.01% Tween-20 to Wash.
5. Repeat Step 3 to wash Dynabeads™ MyOne™ Streptavidin C1 a third time and then resuspend in 500 µl 1xPBS (pH6.8) + 0.01% Tween-20 in standard 1.5 mL tube.
6. Add 125 µl of 5 mg/mL Biotin-WGA to the beads with gentle vortexing to ensure even mixing.
7. Incubate on a nutator for 30 min at RT to allow Biotin-WGA to bind Dynabeads™ MyOne™ Streptavidin C1.

8. Quick spin (100xg for 2-3 seconds), and place on magnet stand for ~ 2 min to clear, then remove and discard supernatant.
9. Resuspend in 1 mL 1xPBS + 0.01% Tween.
10. Store at 4°C until use for up to six months.

CRITICAL: Magnefy™ Concanavalin A ready-to-use beads can be used as an alternative to the WGA-coated MyOne-C1 magnetic beads, but may differ in their propensity to bind specific cell types.

1M HEPES-KOH pH 7.9: Add 23.83g HEPES to 80 mL of deionized water. Adjust to pH 7.0 by adding 1M KOH to solution dropwise while monitoring with a pH meter. Add deionized water to almost 100 mL then do a final pH adjustment to pH 7.9 by adding 1M KOH to solution dropwise while monitoring with a pH meter. Bring up to 100 mL with deionized water, filter sterilize and store at room temperature for up to 1 year.

1 M HEPES pH 7.5: Add 23.83g HEPES to 70 mL of deionized water. Adjust to pH 7.0 by adding 1M NaOH to solution dropwise while monitoring with a pH meter. Add deionized water to almost 100 mL then do a final pH adjustment to pH 7.5 by adding 1M NaOH to solution dropwise while monitoring with a pH meter. Bring up to 100 mL with deionized water, filter sterilize and store at room temperature for up to 1 year.

NE1 buffer: Prepare fresh by mixing 1 mL 1M HEPES-KOH pH 7.9, 500 µl 1M KCl, 12.5 µl 2 M spermidine, 500 µl 10% Triton X-100 (v/v), and 10 mL glycerol in 38 mL dH_2O, and add 1 Roche Complete Protease Inhibitor EDTA-Free tablet. Chill on ice before use.

Wash buffer: Prepare fresh by mixing 1 mL 1 M HEPES pH 7.5, 1.5 mL 5 M NaCl, 12.5 µl 2 M spermidine, 250 µl of 10% (v/v) Triton-X solution and bring the final volume to 50 mL with dH_2O, and add 1 Roche Complete Protease Inhibitor EDTA-Free tablet. Wash Buffer can be stored at 4 °C for up to 1-2 days prior to using, but a Roche Complete Protease Inhibitor EDTA-Free tablet should be added fresh before use.

Binding buffer: Mix 200 µl 1M HEPES-KOH pH 7.9, 100 µl 1M KCl, 10 µl 1M $CaCl_2$ and 10 µl 1M $MnCl_2$, and bring the final volume to 10 mL with dH2O. Store the buffer at 4 °C for up to 6 months.

Antibody buffer: Prepare fresh by mixing 8 µl 0.5 M EDTA with 2 mL Wash buffer and chill on ice.

300-wash buffer: Mix 1 mL 1 M HEPES pH 7.5, 3 mL 5 M NaCl and 12.5 µl 2 M spermidine, 250 µl of 10% (v/v) Triton-X solution, bring the final volume to 50 mL with dH_2O and add 1 Roche Complete Protease Inhibitor EDTA-Free tablet. 300-wash buffer can be stored at 4 °C for 1-2 days prior to using, and in this case an additional Roche Complete Protease Inhibitor EDTA-Free tablet should be added fresh before use.

Tagmentation buffer: Prepare fresh by mixing 10 mL 300-wash buffer and 100 µl 1 M $MgCl_2$ (to 10 mM).

1 M TAPS pH 8.5 buffer: dissolve 24.3 g TAPS ([tris(hydroxymethyl)methylamino]propanesulfonic acid) in 80 mL dH_2O. Adjust to pH 8.5 by addition of 1 M NaOH. Bring to final 1 L volume with dH_2O and filter sterilize.

10 mM TAPS pH 8.5 buffer: Mix 500 µl 1 M TAPS pH 8.5 in 50 mL dH_2O (to 10 mM). Can be kept at Room Temperature for up to 6 months.

0.19% SDS Release buffer: Prepare fresh the day of the experiment by diluting 10% SDS (w/v) 1:10 in 1 mL dH_2O. Then mix 152 µl 1% SDS and 648 µl 10 mM TAPS pH 8.5 buffer.

2.5% Triton-X neutralization solution: Prepare fresh the day of the experiment by mixing 250 µl 10% Triton-X100 (v/v) in 750 dH_20

3.4 Procedure

Prepare WGA-coated MyOne-C1 magnetic beads TIMING 15 min

1) Gently resuspend and withdraw 200 µl of the WGA-coated MyOne-C1 magnetic bead slurry.

CRITICAL STEP: As compared to the BioMag-Plus Concanavalin A (Bangs Laboratories, Inc.) beads used in bulk CUT&RUN and CUT&Tag protocols, the WGA-Coated MyOne-C1 beads and Magnefy™ Concanavalin A ready-to-use beads are smaller and more uniformly sized, reducing the formation of multi-nuclei aggregates and allowing bead-bound nuclei to pass through a 20-micron filter during split-pooling.

2) At room temperature, transfer 200 µl WGA-coated MyOne-C1 bead slurry into 1.5 mL Binding buffer in a 2 mL tube and mix by pipetting. Place the tube on a magnet stand to clear for 2 minutes.
3) Withdraw the liquid completely and remove the 2 mL tube from the magnet stand. Add 2 mL Binding buffer and mix by pipetting.
4) Place tube on magnet stand to clear, withdraw liquid, resuspend beads in 200 µl Binding buffer and leave at room temperature until nuclei are ready.

Prepare lightly cross-linked nuclei and cryopreserve TIMING 1 hr

5) Harvest cell culture(s) in a 15 mL conical centrifuge tube at room temperature or thaw frozen cells and count cells using an automated cell counting devise (e.g. Countess 3) or by hand using a hemocytometer. For efficient recovery of lightly cross-linked nuclei, this protocol requires starting with ~10 million mammalian cells per sample.
6) Centrifuge 3 min 600 x g in a swing-bucket rotor at room temperature and drain liquid.
7) Resuspend in 1.5 mL PBS at room temperature while pipetting and transfer to a 1.5 mL standard microfuge tube (not low bind) to increase visibility.
8) Centrifuge 3 min 600 x g in a benchtop centrifuge at room temperature and carefully pipette off the liquid. Do not disturb the pelleted cells.
9) Resuspend in 1 mL ice-cold NE1 buffer with gentle vortexing. Let sit on ice 10 min.
10) Centrifuge 8 min 1300 x g at 4°C in a temperature-controlled benchtop centrifuge. Carefully pipette off the liquid. Do not disturb the pelleted nuclei.
11) Resuspend in 1 mL of room temperature PBS.

12) To cross-link the nuclei, add 6.4 µl of 16% formaldehyde to the sample (final concentration is 0.1%) and immediately mix by inverting the tube 5 times, then incubate at room temperature for 2 min.
13) Stop cross-linking by adding 2.5 M glycine to twice the molar concentration of the formaldehyde (e.g., 30 µl to 1 mL).
14) Centrifuge 8 min 1300 x g at 4°C in a temperature-controlled benchtop centrifuge. Carefully pipette off the liquid. Do not disturb the pelleted nuclei.

CRITICAL STEP: Nuclei become difficult to see after cross-linking and can be quite slippery in the tube. To avoid losing nuclei, aspirate the supernatant slowly at the surface of the liquid and leave a small volume of PBS (~5 µl) at the bottom of the tube.

15) Resuspend lightly cross-linked nuclei in Wash buffer and dilute to quantify nuclei using the Countess 3 or cell counter slide.

? TROUBLESHOOTING

16) Dilute lightly cross-linked nuclei to a concentration of ~1 million nuclei/mL in Wash buffer.

PAUS|E POINT: If not using immediately, nuclei may be slow frozen. To do so, aliquot 900 µl into cryogenic vials containing 100 µl DMSO, mix by pipetting, then add to a Mr. Frosty container filled to the line with isopropanol and place in a -80°C freezer overnight for storage at -80°C.

Bind nuclei to WGA-coated MyOne-C1 beads TIMING 15 min

CRITICAL: For each sciCUT&Tag experiment, use 2 million lightly cross-linked nuclei as input. The protocol can also be run using fewer than 2 million lightly cross-linked nuclei, but this may result in suboptimal recovery. Keep freshly prepared lightly cross-linked nuclei on ice or thaw slow-frozen lightly cross-linked nuclei aliquots at room temperature. If thawing nuclei, we recommend counting nuclei prior to bead-binding on a Countess 3 instrument to account for potential lysis during freeze-thaw. For SNP-based detection of

collisions and *In Silico* doublet removal during downstream analysis, it may be desirable to mix samples from multiple different donors.

17) Resuspend the 200 µl of WGA-coated MyOne-C1 beads from step 4 by inverting the tube 2-3 times or by pipetting up and down with a wide-bore tip. Then, in a dropwise manner, mix 200 µl of WGA-coated MyOne-C1 beads with the nuclei suspension. To ensure even binding, close and invert samples to mix in-between additions of WGA-coated MyOne-C1 beads. Once the full 200 µl WGA-coated MyOne-C1 bead slurry has been added, close the tube and incubate on a nutator at room temperature for 10 min.

CRITICAL STEP: If numerous distinct samples are to be profiled in the same experiment, the proportion of nuclei and the WGA-coated MyOne-C1 bead slurry should correspond to the proportion of the 96-well sci-pA-Tn5 plate that will be dedicated to each sample (for example, for an experiment with a total of 2 million nuclei split evenly between K562 cells, H1 cells and PBMCs, ~670K nuclei of each cell type should be bound to 67 µl WGA-coated MyOne-C1 beads).

18) Divide the sample into 0.65 mL pre-lubricated microcentrifuge tubes at the appropriate ratio for each antibody to be profiled.

CRITICAL STEP: The number of nuclei per antibody can vary greatly depending on the experimental design (For example, for an experiment in which two different histone modifications will be profiled in an equal number of nuclei, the sample should be split evenly between two 0.65 mL pre-lubricated microcentrifuge tubes. In an experiment in which four different histone marks will be profiled and two of the histone marks will be profiled in twice as many cells as the other two histone marks, the sample will be split into four tubes: 1/3 the volume of bead-bound nuclei will be allocated to the two “high-cell-number histone marks” and 1/6 the volume of bead-bound nuclei being allocated to the two “low-cell-number histone marks”. Each of the four samples will then be allocated a corresponding proportion of the 96-well plate for sci-pA-Tn5 binding and tagmentation).

19) Place the tubes on a magnet stand to clear and withdraw the liquid. If desired, the unbound supernatants may be counted on a Countess 3 instrument to check efficiency of nuclei binding to beads.

? TROUBLESHOOTING

Bind primary antibody TIMING 1 hr 15 min

20) Resuspend samples in Antibody Buffer at a concentration of ≤ 20,000 bead bound nuclei per µl (for example, bring up 1 million bead bound nuclei to 50 µl in Antibody Buffer). Then add the relevant antibodies to each sample at a 1:10 dilution (v/v) of Antibody solution to sample (for example, add 5 µL of anti-H3K27me3 antibody to 50 µL of sample). If multiple antibodies are to be added to a single sample (e.g., anti-H3K4me1, anti-H3K4me2 and anti-H3K4me3 antibodies), the volume of each antibody solution should correspond to 1/10 the initial volume of the sample.

21) Mix by gentle vortexing, or flicking the tube, then quick spin (100xg for 2-3 seconds) to collect the solution at the bottom of the tube.

22) Place in a tube rack at room temperature and incubate for 30 minutes, then mix by gentle vortexing or flicking the tube, quick spin to collect solution at the bottom of the tube (100xg for 2-3 seconds) and incubate for an additional 30 minutes at room temperature or at 4˚C overnight.

PAUSE POINT Antibody incubation may proceed overnight at 4 ºC; we have not noticed any difference between the efficiency of a 1 hr room temperature incubation and an overnight 4 °C incubation.

Bind secondary antibody TIMING 45 min

23) Quick spin at 100xg for 2-3 seconds, place each tube on the magnet stand to clear for 2 min and remove the liquid from each tube by pipetting.

CRITICAL: The quick spin on a micro-centrifuge minimizes the carry-over of reagents that could result in increased background tagmentation or an overall reduction in pA-Tn5 tethering to targeted chromatin.

24) Remove tubes from the magnet stand and wash each with 150 µl of Wash buffer. Place each tube back on the magnet stand to clear for 2 minutes and remove the liquid.
25) Repeat step 24 for a second wash.
26) Resuspend samples in 30 µl of Wash buffer and add 1.5 µl secondary antibody (1:20). Mix by gentle vortexing, or flicking the tubes, then quick spin (100xg for 2-3 seconds) to collect the solution at the bottom of the tube.
27) Place the tubes in a tube rack at room temperature for 30 min or 4 ºC for 1 hr. Halfway through the incubation, again mix by gentle vortexing, or flicking the tube, then quick spin to collect the solution at the bottom of the tube.

PAUSE POINT Secondary antibody incubation may proceed overnight at 4 ºC.

28) After a quick spin (100xg for 2-3 seconds), place each tube on the magnet stand to clear for 2 min and remove the liquid from each tube by pipetting.
29) Wash with 150 µl of Wash buffer. Place each tube on the magnet stand to clear for 2 min and remove the liquid from each tube by pipetting.
30) Repeat step 29 for a second wash.

Bind pA-Tn5 adapter complex TIMING 1 hr 15 min

31) Resuspend WGA-coated MyOne-C1 bead bound samples in 300-wash buffer to a final volume of 20 µl per well of the 96-well sci-pA-Tn5 plate (for example, if the experiment is split evenly between two samples, 48 wells each, then both samples should be brought up in 960 µl 300-wash buffer. Include an additional 20 µl of 300-wash buffer to avoid running out of solution while loading the plate).
32) Transfer 20 µL of the appropriate sample into each well of a fresh 96 well LoBind Semi-skirted PCR plate.

CRITICAL STEP: Be sure to track the distribution of each sample across the plate for demultiplexing after sequencing.

33) Centrifuge the sealed Sci-pA-Tn5 Plate for 1 minute at 1200 g before opening.

34) Using an 8-channel 20 µl multi pipette, transfer 2 µl of adapter-loaded pA-Tn5 from the Sci-pA-Tn5 Plate into the corresponding wells of the 96 well sample plate. Mix by pipetting up and down ~5 times. Seal the Sci-pA-Tn5 Plate and store at -20 ºC.

35) Allow pA-Tn5 binding to proceed for 1hr at 4 ºC while keeping the 96 well sample plate stationary.

36) Add 50 µl of 300-wash buffer to each well of the 96 well sample plate. Then place on the 96 well plate ring magnet. Allow samples to collect on the side of the wells (~1 min).

37) Using an 8-channel 200 µl multi pipette, remove unbound pA-Tn5 solution and discard in waste.

CRITICAL STEP: During wash steps, avoid cross-contamination of pA-Tn5 solution between wells.

38) Decant 300-wash buffer into a reservoir and add 150 µl to each well of the 96 well sample plate. Incubate for 2-5 min while keeping the sample plate on the magnet.

39) Using an 8-channel 200 µl multi pipette, remove wash solution and transfer to waste.

40) Repeat steps 38-39 (second wash).

***Tagmentation* TIMING 1 hr 15 min**

41) Decant Tagmentation buffer into a reservoir and add 75 µL to each well of the 96 well sample plate.

42) Seal sample plate with an AlumaSeal 96 cover, and place in a ThermoCycler set to 37ºC with heated lid for 1 hr.

Repool, Filter and Count TIMING 40 min

43) Place 96 well sample plate on the 96 well plate ring magnet. Allow samples to collect on the side of the wells (~1 min). Remove AlumaSeal and using an 8-channel 200 µl multi pipette, remove Tagmentation Buffer and discard in waste.

44) Remove the 96 well sample plate from the magnet. Decant Wash buffer into a reservoir and using an 8-channel 200 µl multi pipette with wide-bore tips transfer 200 µl Wash buffer into the 8 wells of column 1 of the sample plate.
45) Gently pipette the Wash buffer up and down, dispensing the solution along the side of the well in a circular motion to dislodge all the bead-bound nuclei from the sides of the wells and resuspend them in solution.
46) Transfer the 200 µl Wash buffer with bead-bound tagmented nuclei from column 1 into the 8 wells of column 2.
47) Repeat Step 45 to resuspend the bead-bound nuclei from column 2.
48) Transfer the 200 µl Wash buffer with bead-bound tagmented nuclei from columns 2 and 3 into the 8 wells of column 3. Repeat Steps 45 and 46 for each successive column to collect all the nuclei from each row in 200 µl of Wash buffer in column 12.
49) Using a single-channel 200 µl pipette with wide-bore tips, transfer Wash buffer and nuclei from each well of column 12 to a 0.65 mL pre-lubricated microcentrifuge tube on a magnet stand. As the tube fills up, wait for the beads to clear for 2 min, and transfer clear Wash buffer to the waste.
50) Repeat Step 49 until all the bead-bound nuclei have been collected from the 96 well sample plate into the 0.65 mL pre-lubricated microcentrifuge tube.
51) Remove tube from the magnet stand and wash bead-bound nuclei with 500 µl Wash buffer, pH 8.5.
52) Return sample to the magnet stand and allow to clear for 2 min, then transfer Wash buffer to waste.
53) Remove sample from the magnet stand and resuspend bead-bound nuclei in 540 µl Wash buffer.
54) Place two 20 µm Mini-Strainers in two open 1.7-mL Low Retention Microtubes.
55) Using a single-channel 200 µl pipette with wide-bore tips, transfer 270 µl of sample to each of the 20 µm Mini-Strainers
56) Carefully move the loaded 20 µm Mini-Strainers and Low Retention Microtubes to a bench-top centrifuge.

57) Centrifuge for 3 min at 100 g at room temperature.

CRITICAL STEP: Passing the sample solution through a 20 µm Mini-Strainer is necessary to remove any large multi-nuclear aggregates prior to loading the ICELL8. Following centrifugation, the surface of the 20 µm Mini-Strainer should appear brown from the visible WGA-coated beads that are filtered out along with the multicellular aggregates.

58) If liquid remains on top of the 20 µm Mini-Strainers, twist the strainer to break any seal with the lip of the Low Retention Microtubes and repeat Step 57.
59) Using a single channel 200 µl pipette with wide-bore tips, recombine the two filtered samples into a single 1.7-mL Low Retention Microtube.
60) Dilute a 10,000X DAPI stock 1:100 in 10 mM TAPS buffer pH 8.5, and then add 5.4 µl of the 100X DAPI solution to the sample solution for a final concentration of 0.5 µg/mL DAPI.

    CRITICAL STEP: DAPI is used for detecting and counting nuclei on the Countess Cell Counter as well as the ICELL8. To obtain an accurate count of the concentration of nuclei, DAPI staining is necessary to distinguish bead-bound nuclei from free-floating beads.
61) Load 10 µl of Sample solution onto a plastic slide for the Countess.
62) Insert the slide into the Countess then turn on the bright field and DAPI Channels and image for quantification.

    CRITICAL STEP: To obtain an accurate quantification of DAPI-positive nuclei, it may be necessary to (1) adjust the intensity of the DAPI laser or (2) adjust the detection parameters for calling nuclei during the image processing. Typically, an accurate count ranges from 1 x 10^5 DAPI-positive nuclei/mL to 4 x10^5 DAPI-positive nuclei/mL.
63) Repeat Steps 61 and 62 three more times to get four replicate counts of the DAPI-positive nuclei/mL.

? TROUBLESHOOTING

Dispense on the ICELL8, Image and Perform SDS Release. TIMING 2 hr 30 min

64) Calculate the average of the four replicate counts of the DAPI-positive nuclei/mL.

65) Thaw one aliquot of the 100X Diluent and the Fiducial reagents from the SMARTer ICELL8 Blank Chip Reagent Kit at room temperature.

66) Prepare the ICELL8 by running (1) a system prime and (2) a daily warmup. Turn on the microscope system for imaging the chip.

67) Place a 350v chip on the target platform within the humidity-controlled chamber of the ICELL8.

68) Load 25 µl of 10 mM TAPS pH 8.5 into the positive and negative control wells of a 384 well ICELL8 loading plate (A24 and P24).

69) Load 25 µl of the fiducial reagent into the fiducial well of the 384 well ICELL8 loading plate (P1).

70) Place sample on a magnet stand and allow to clear for 2 minutes. Discard Wash Buffer plus DAPI and resuspend the bead-bound sample in 10 mM TAPS pH 8.5 at a concentration of 2.85 - 5.71 x 10^5 DAPI-positive nuclei. Then add the appropriate volume of 100X DAPI solution and the 100X Diluent so that the final concentration in the sample is 1X DAPI and 1X Diluent. Proceed immediately to sample dispense.

CRITICAL STEP: The minimum volume necessary to load the ICELL8 is 540 µL and the maximum concentration of nuclei that should be dispensed on the ICELL8 is 5.71 x 10^5 DAPI-positive nuclei/mL, which corresponds to 20 nuclei per 35 nL dispense. If a sufficient number of tagmented nuclei are obtained, it may be desirable to increase the volume of the sample to 800 µL which is enough to dispense two 350v ICELL8 chips; this strategy captures more single cells, but also requires an additional flow cell for sequencing.

? TROUBLESHOOTING

71) Using a wide bore 200 µl tip, homogenize the sample by pipetting up and down 3-4 times then load 65 µl of sample solution into each of the 8 source wells for cellular dispense on the 384 well ICELL8 loading plate (rows A-D, columns 1&2).

CRITICAL STEP: During protocol optimization, it may be desirable to perform a serial dilution experiment. For example, we performed a serial dilution to determine the collision frequency at different loading concentrations. For this, load different concentrations of sample into the 8 wells of the source plate.

72) Secure the source plate in the humidified chamber of the ICELL8
Make sure to load the Pause and Dispense protocol prior to dispensing cells. Proceed with an unfiltered, 35 nL cell dispense cycle. When prompted, mix the sample in the source wells prior to each dispense by pipetting up and down 5 times using a wide bore tip, and then resume dispensing.

73) Once the cell dispense is complete, remove the 350v chip from the humidity-controlled chamber of the ICELL8. Dry the 350v chip with the circular paper from the SMARTer ICELL8 Loading Kit, then cover the chip with the translucent three leaf imaging film provided in the SMARTer kit and seal.

74) Using the ICELL8 chip adapter and balance provided with the ICELL8, spin at 1200 x g for 1 min at room temp.

75) Image the ICELL8 chip to examine the distribution of nuclei at different regions of the plate. Here, we would like to obtain 10-20 nuclei per well, and the imaging is used only as a quality control to examine the distribution of nuclei across the chip.

CRITICAL STEP: This combinatorial indexing approach does not require identifying wells that contain a single nucleus, or the generation of a filter file for downstream dispenses.

? TROUBLESHOOTING

76) Remove the imaging film and place the 350v chip on the target platform within the humidity-controlled chamber of the ICELL8.

77) Load a fresh 384 well plate with the 0.19% SDS Release buffer. Dispense 25 µl of the 0.19% SDS Release buffer into the positive control, negative control and fiducial wells (A24, P24 and P1). Dispense 80 µl of the 0.19% SDS Release buffer into the 8 source wells for cellular dispense (rows A-D, columns 1&2).

CRITICAL STEP: During the 58 ºC Tn5-release step, the concentration of SDS is 0.095% (w/v). This concentration has been optimized to maximize Tn5 release while allowing for consistent quenching of the SDS with Triton-X prior to the addition of PCR reagents[157].

78) Secure the source plate in the humidified chamber of the ICELL8.

79) Proceed with an unfiltered, 35 nL cell dispense cycle.

CRITICAL STEP: Because all other solutions remain homogenous throughout dispense cycles, the rest of the protocol does not require the Pause and Dispense protocol.

80) Once the SDS Release buffer dispense is complete, remove the 350v chip from the humidity-controlled chamber of the ICELL8. Dry the 350v chip with the circular paper from the SMARTer ICELL8 Loading Kit, then cover with the thermal film (white backing) and seal.

81) Using the ICELL8 chip adapter and balance, spin at 1200 x g for 1 min at Room Temperature.

82) Place the 350v chip in a thermocycler modified with the ICELL8 chip adapter. Incubate for 1hr at 58ºC, then hold at 12ºC.

PAUSE POINT: Following the 1hr incubation at 58ºC the chip can be wrapped in parafilm and stored at 4ºC indefinitely.

Triton-X Quench and PCR amplification. TIMING 1 hr 30 min

83) Using the ICELL8 chip adapter and balance, spin the sealed 350v chip at 1200 x g for 1 min at room temp.

84) Remove the thermal film and place the 350v chip on the target platform within the humidity-controlled chamber of the ICELL8.

85) Load 80 µl of the 2.5% Triton-X Neutralizing solution into the 8 source wells of the 384 well sciCUT&Tag PCR Primer Plate (rows A-D, columns 1&2). Add 25 µL 10 mM TAPS to fiducial, negative and positive control wells (P1, A24, P24).

86) Secure the sciCUT&Tag PCR Primer source plate in the humidified chamber of the ICELL8.

87) Proceed with an unfiltered, 35 nL cell dispense cycle.

CRITICAL STEP: Similar to the 0.19% SDS Release buffer dispense**,** here, the cell dispense cycle is used for dispensing 35 nL of the Triton-X Neutralizing solution into all the wells of the 350v chip.

88) Once the dispense is complete, remove the 350v chip from the humidity-controlled chamber of the ICELL8. Dry the 350v chip with the circular paper from the SMARTer ICELL8 Loading Kit, then cover with either an imaging film or thermal film and seal.

89) Using the ICELL8 chip adapter and balance, spin at 1200 x g for 1 min at room temp.

90) Remove the film and return the 350v chip to the target platform within the humidity-controlled chamber of the ICELL8.

91) Proceed with the 35 nL unfiltered dispense of the first set of sciCUT&Tag primers (Index1, sci i5s).

92) Repeat Steps 89-91 to dry and spin down the contents of the chip between dispenses.

93) Proceed with the 35 nL unfiltered dispense of the second set of sciCUT&Tag primers (Index2, sci i7s).

94) Repeat Steps 89-91 to dry and spin down the contents of the chip between dispenses.

95) Prepare the KAPA HiFi PCR Mix. In a 1.5-mL microcentrifuge tube, mix 553.7 µl 5X HiFi Buffer, 306.6 µl ddH_2O, 84.8 µl dNTPs (10mM) and 54.9 µl KAPA HiFi Polymerase.

96) Load 100 µl of KAPA HiFi PCR Mix into the 8 source wells for cellular dispense on the 384 well ICELL8 loading plate (rows A-D, columns 1&2) and save remaining PCR mix. Add 25 µL 10 mM TAPS to fiducial, negative and positive control wells (P1, A24, P24).

97) Secure the source plate in the humidified chamber of the ICELL8.

98) Proceed with a 50 nL unfiltered cell dispense of the KAPA HiFi PCR Mix.

99) Repeat Steps 89-91 to dry and spin down the contents of the chip between dispenses.
100) Load an additional 24 µl of KAPA HiFi PCR Mix into the 8 source wells for cellular dispense on the 384 well ICELL8 loading plate (rows A-D, columns 1&2).
101) Secure the source plate in the humidified chamber of the ICELL8.
102) Proceed with a second 50 nL unfiltered cell dispense of the KAPA HiFi PCR Mix.
103) Once the dispense is complete, remove the 350v chip from the humidity-controlled chamber of the ICELL8. Dry the 350v chip with the circular paper from the SMARTer ICELL8 Loading Kit. Then cover with a thermal film and seal well to avoid evaporation.
104) Using the ICELL8 chip adapter and balance, spin at 1200 x g for 1 min at room temp.
105) Place the 350v chip in a thermocycler modified with the ICELL8 chip adapter.
106) Proceed with PCR reaction according to the following PCR cycling conditions).

Cycle number	**Denature**	**Anneal**	**Extend**	**Final**
1			58° C, 5 min	
2			72° C, 10 min	
3	98° C, 45 s			
4–19	98° C, 15 s	60° C, 10 s	72° C, 5 s	
20			72° C, 1 min	

21				8° C, hold

Sample Collection TIMING 15 min

107) Open the ICELL8 Collection Kit and assemble the collection module by attaching the 0.7 mL microfuge tube to the chip collection funnel. Remove the thermal film from the 350v chip and invert the chip, with the nanowells facing down, onto the chip collection funnel. Insert the collection module into the ICELL8 collection adapter. Seal the chip to the collection module with the provided sealing tape.

108) Using the ICELL8 collection adapter balance, centrifuge the collection assembly at 1200 x g for 8 sec at room temp and then stop. Transfer the sample to a 1.7 mL tube, re-assemble the collection module, repeat the 8 sec centrifugation step, and transfer the sample. Centrifuge the chip for 5 min and collect the remaining sample.

CRITICAL STEP: The volume of PCR product exceeds the capacity of a single 0.7 mL microfuge collection tube. To avoid overflow, the sample must be collected in intervals.

***Post-PCR Cleanup* TIMING 1 hr 15 min**

109) Determine the volume of the PCR amplified sciCUT&Tag library by adjusting the volume of a 1 mL Pipette until it matches the sample volume upon aspiration. Split evenly between two 1.7 mL low retention tubes and add 1.1 x the volume of Ampure XP beads, mixing by pipetting up and down.

110) Quick spin (100xg for 1-2 seconds) and let sit at room temperature 5-10 min.

111) Place the tubes on a magnet stand and allow the solutions to clear for 3 min before carefully withdrawing liquid. While the empty tubes are still on the magnet stand, add 1.5 mL 80% ethanol to each tube.

112) Allow solutions to clear before carefully withdrawing liquid. While still on the magnet, add 1.5 mL 80% ethanol.

113) Remove liquid and discard.

CRITICAL STEP: Overdrying can reduce recovery of DNA, so continue to the elution step within 5 min.

114) Remove from the magnet stand, add 100 µL 10 mM Tris-HCl pH 8 to each of the tubes and vortex on full for 30 seconds. Then collect sample at the bottom of the tube using a quick spin (100xg for 2-3 sec).

115) Incubate tubes for 5 min at room temperature and then place the tubes on the magnet stand and allow to clear for 2 min.

116) Combine the two 100 µL 10 mM Tris-HCl pH 8 solutions containing the eluted sciCUT&Tag library into a single fresh 1.5 mL tube with a pipette.

117) Add 0.9 x the volume of Ampure XP beads (180 µl), mixing by pipetting up and down.

118) Quick spin (100xg for 2-3 seconds) and let sit at room temperature 5-10 min.

119) Place tube on the magnet stand and allow to clear before carefully withdrawing liquid. While the tube is still on the magnet, add 500 µl 80% ethanol.

120) Allow to clear before carefully withdrawing liquid. While the tube is still on the magnet stand, add 500 µl 80% ethanol.

121) Withdraw the liquid and discard in liquid waste.

CRITICAL STEP: Over-drying can reduce recovery of DNA, so continue to the elution step within < 5 min.

122) Remove from magnet stand, add 25 µl 10 mM Tris-HCl pH 8 and vortex on full for 30 seconds.

123) After a 5 minute incubation at room temperature place on the magnet stand and allow to clear for 2 minutes.

124) Transfer the 25 µl 10 mM Tris-HCl pH 8 solution containing the eluted sciCUT&Tag library to a fresh 1.5 mL tube with a pipette.

PAUSE POINT: Libraries can be stored at 4˚C for up to two weeks or at -20˚C for up to 1 year.

Library analysis and DNA sequencing TIMING: 1 d

125) Determine the size distribution of libraries by capillary electrophoresis (*e.g.* using the Agilent 4200 TapeStation) using 2 µl of the library, following the manufacturer's instructions (https://www.agilent.com/cs/library/usermanuals/public/4200-TapeStation_SystemManual.pdf). For mixing single-tube samples with load buffer, use a low-volume (*e.g.* 2 µl) pipettor to avoid injecting bubbles.

CRITICAL STEP: Molarity estimates are based on the capillary electrophoresis profile by choosing the region between 160 bp and 1000 bp and using the value reported as "Region Molarity [pmol/L]".

? TROUBLESHOOTING

126) Allow the Illumina NextSeq 2000 Flow Cell and cartridge to thaw at room temperature for 15 minutes.

CRITICAL STEP: The Next-Seq P2-200 is expected to produce 400 Million reads and is preferable to obtain sufficient library saturation for samples with an initial concentration > 10 nM. The Next-Seq P1-300 is expected to produce 100 Million reads and is preferable to reduce the cost while obtaining sufficient library saturation for samples with an initial concentration < 10 nM.

127) Remove cartridge from the bag and invert it 10 times to mix reagents.

128) For sequencing on a NextSeq2000 instrument, dilute the sciCUT&Tag library to 2nM using the RSB with Tween supplied as part of the Illumina flow cell kit. Then dilute to the final loading concentration specified in the Illumina Onboard Denature and Dilute Guide. We do not include the PhIX spike in.

129) Resuspend NextSeq_Read1_seq_primer, NextSeq_Read2_seq_primer, NextSeq_Index1_seq_primer and NextSeq_Index2_seq_primer at 100 µM in 10 mM Tris HCl pH 8.0 (For sequences see Supplementary Table 1).

130) For loading, mix 1.8 µL of NextSeq_Read1_seq_primer and 1.8 µL of NextSeq_Read2_seq_primer and dilute to 0.3 µM, by adding 596.4 µL HT1 Buffer.

131) Mix 3.6 µL of NextSeq_Index1_seq_primer and 3.6 µL of NextSeq_Index2_seq_primer and dilute to 0.6 µM, by adding 592.8 µL HT1 Buffer.
132) Using a new P1000 tip, pierce the Library Reservoir and push the foil to the edges.
133) Add 20 µL of the diluted sciCUT&Tag library to the bottom of the Library Reservoir. Avoid touching the foil.
134) Mix the 0.3 µM Read1&2 primer solution by pipetting up and down 4 times and then transfer 550 µL to the well “Custom 1” on the Illumina instrument.
135) Mix the 0.6 µM Index1&2 primer solution by pipetting up and down 4 times and then transfer 550 µL to the well “Custom 2” on the Illumina instrument.
136) Pull the flow cell out of the package and hold by the gray tab with the label facing up.
137) Push to insert the flow cell into the slot on the front of the cartridge.
138) Pull back the gray tab and remove to expose the flow cell.
139) Click Load to load the cartridge.
140) Wait 2-3 minutes while the sequencer is scanning consumables. Then double check the Sequencing configuration and press Sequence.

Data Processing and Analysis **TIMING: Variable compute timing**

141) Perform demultiplexing using the sciCTextract custom software. This requires six input files: the four Fastq files corresponding to genomic read 1, genomic read 2, index read 1 and index read 2, as well as two barcode tables in comma-separated value format (.csv). The Fastq files should follow Illumina naming conventions, including read headers (e.g. @VH00319:342:AACKYJMM5:1:1101:31410:1000 1:N:0:0). The first barcode.csv file defines the prefixes of the sample names based on the s5 and s7 adapter sequences (Supplementary Table 4). The second barcode.csv file defines the suffixes of the sample names based on the i5 and i7 primer sequences (Supplementary Table 5). The output of sciCTextract consists of a pair of compressed Fastq files with rewritten headers that

include the error corrected barcode sequences, as well as enough information to unambiguously identify each source read (e.g. @AACKYJMM5:1:1101:31410:1000_GCGTTAAA_GTGTATCG_AGCGATAG_CAGGACGT 1:N:0:0).

142) Clip adapter sequences using cutadapt 2.9 with parameters: -j 8 -m 20 -a CTGTCTCTTATACACATCT -A CTGTCTCTTATACACATCT -Z

143) Perform alignments using Bowtie2 version 2.4.2 to UCSC HG38 with parameters: --very-sensitive-local --soft-clipped-unmapped-tlen --no-mixed --no-discordant --dovetail --phred33 -I 10 -X 1000

144) Convert the sam file created by bowtie2 to a bam file and run "bedtools bamtobed -bedpe" on the resulting bam file.

CRITICAL STEP: If the experiment includes multiple donor samples mixed together, perform SNP-based identification of doublets using the aligned reads as input to Souporcell[151]. To streamline this process, we assembled an easy-to-use Nextflow pipeline called Souper-star that removes duplicate reads, merges files for Souporcell, generates quality control output, and splits out bed files for each epitope profiled (https://github.com/FredHutch/souper-star).

145) From the bed file, extract columns 1, 2, 6, 7 which correspond to the chr, start, end and the rewritten headers including the barcode.

146) Prepare the bedfile for input into single-cell analysis software such as ArchR by reformatting the fourth column to include just the barcode sequence and removing duplicate reads with the same chromosome, start, stop, and barcode. This can be done using the bash script: cat $1 | awk '{gsub(":","\t",$4);gsub("_","\t",$4); print $1"\t"$2"\t"$3"\t"$4"\t"$5}4' | awk '{print $1"\t"$2"\t"$3"\t"$9"_"$10"_"$11"_"$12}' | sort -k1,1 -k2,2n -k4,4 | uniq -c | awk '{print $2"\t"$3"\t"$4"\t"$5"\t"$1}' > $2

147) Use the resulting bed file as the input to a single-cell analysis package for dimensionality reduction and graph-based clustering (*e.g.*, using ArchR, Signac, chromVAR[158]). Aggregate bed files can also be used for peak-calling

(SEACR[159]). Example analysis can be found here: https://github.com/FredHutch/sciCnT_manuscript_2023.

3.5 Timing

Steps 1-4, Prepare WGA-coated MyOne-C1 magnetic beads: 15 min

Steps 5-16, Prepare lightly cross-linked nuclei and cryopreserve: 1 hr

Steps 17-19, Bind nuclei to WGA-coated MyOne-C1 beads: 15 min

Steps 20-21, Bind primary antibody: 1 hr 15 min

Steps 23-30, Bind secondary antibody: 45 min

Steps 31-40, Bind pA-Tn5 adapter complex: 1 hr 15 min

Steps 41-42, Tagmentation: 1 hr 15 min

Steps 43-63, Repool, filter and count: 40 min

Steps 64-83, Dispense on the ICELL8, image and perform SDS Release: 2 hr 30 min

Steps 84-107, Triton-X Quench and PCR amplification: 1 hr 30 min

Steps 108-109, Sample collection: 15 min

Steps 110-125, Post-PCR cleanup: 1 hr 15 min

Steps 126-141, Library analysis and DNA sequencing: 1 day

Steps 142-148, Data processing and analysis: variable

Troubleshooting

Troubleshooting advice can be found in Table 2.

Table 2: Troubleshooting table.

Steps	Problem	Possible reasons	Solution

15	An insufficient number of nuclei are recovered	The starting number of cells is too low for the nuclei-prep protocol.	Increase the number of cells used as input.
			Use a low input nuclei-prep method.
19	> 10% Nuclei remain in solution and are not bound by the beads	WGA-coated paramagnetic beads are suboptimal	Remake the WGA-coated beads
			Switch to ConA-coated MyOne-C1 beads
63	Very few tagmented nuclei pass through the 20 micron strainer	Nuclear lysis and aggregation	Limit work-times and harsh pipetting of the cells.
			Increase the cross-linking time from 2 to 3 min to prevent lysis
70	The number of nuclei is insufficient for 540 µL of 2.85 x 10^5 nuclei / mL	Started protocol with fewer than 2 million nuclei, or nuclei were lost throughout the protocol.	Depending on the number of nuclei present, proceed with the whole sample in 540 µL.
76	Upon imaging, ICELL8 chip wells	There was an error in the quantification or dilution of	Redo the sample quantification and

	contain far more or less than 10 nuclei	tagmented nuclei prior to dispense	dilution, and dispense on a new ICELL8 chip.
126	The library yield is bellow 5 nM	Inefficient Tagmentation or Inefficient PCR.	Remake the sci-PA-Tn5 plate with a high efficiency prep of pA-Tn5
			Perform an SDS vs Triton-X titration to ensure concentrations are optimum for PCR.

3.6 Anticipated results

3.6.1 Sequencing results and distribution of reads across barcodes

We demonstrated the sciCUT&Tag protocol by mixing human PBMCs from two donors and profiling H3K27me3 and H3K4me1-2-3. H3K27me3 labels regions of Polycomb-mediated developmental gene repression. Of note, we pooled antibodies to target all methylated forms of H3K4, which yielded biologically interpretable signal as H3K4 methylation is controlled by highly conserved enzymatic complexes that label nucleosomes at regions of transcriptionally active chromatin (i.e., active promoters and enhancers)[160].

The results of the quality control step described in step 126 are shown in Fig. 4. The undiluted library that contains barcoded reads for both H3K27me3 and H3K4me1-2-3 profiles yielded a peak at around 400 bp (**Fig. 4a**), which is anticipated given the sciCUT&Tag barcoding schema (**Fig. 4b**). Paired-end sequencing at 79 X 43 X 37 X 79 cycles was then be performed with a Next-Seq P2-200 kit.

For preliminary analysis of demultiplexed reads described in step 142, we first calculated the number of unique barcode combinations sequenced and map them to the 96-well plate using the s5 and s7 indices. These indices distinguish the H3K27me3 and H3K4me1-2-3 wells in the same 96-well plate (**Fig. 4b**), and we recovered 363 barcodes/well for H3K27me3 (*%yield = 1.7; stdev. = 73.8, n = 48*) and 240 barcodes/well for H3K4me1-2-3 (*%yield = 1.2; stdev. = 57.4, n = 48*). There should be little variation in mean recovery across the eight s5 and twelve s7 indices, which was observed (**Fig. 4c** and marginal bar plots). Also, the relatively low percent recovery is expected and results from (1) passing the sample through a 20 µm Mini-Strainer to remove multicellular aggregates and (2) loading only a fraction of the sample to obtain between 3-24 nuclei / nanowell on the ICELL8 depending on the experiment (see below).

During library preparation, tagmented nuclei from the 96-well plate are pooled and then dispensed in the 5184-well chip in blocks of 96 that include control wells (**Fig. 4c**, green rectangle). We then performed a similar analysis described in step 142-145 to map the number of unique barcode combinations sequenced to the 5184-well chip using the i5 and i7 indices. A key metric that affects the rate of collisions is the number of cells/nanowell, so we programmed the ICELL8 to perform a serial dilution in duplicate at 24, 12, 6, and 3 cells/nanowell (**Fig. 4d**). Concordantly, the serial dilution was reflected in mean recovery per nanowell, as indicated by the stepwise reduction in barcodes/nanowell along the y axis of the chip (**Fig. 4d**). Combinatorial indexing with the ICELL8 system therefore affords sciCUT&Tag a high degree of tunable cell-per-well accuracy.

3.6.2 Measuring signal-to-noise and antibody specificity with genomic coverage and peak calling analysis

An important step in data validation for chromatin profiling is measuring signal-to-noise to confirm that profiling is targeted. The signal for antibodies that target distinct epitopes should be likewise distinct, otherwise the given antibodies are non-specific. For chromatin profiling experiments, signal-to-noise can be visually inspected with tools like integrated genomics viewer (IGV)[161] or the UCSC genome browser[162], as well as with peak-calling algorithms that distinguish enrichment in genomic coverage over

background. While peak-calling algorithms have yet to be developed for single-cell chromatin profiling data, it is possible to apply bulk peak-calling algorithms to aggregated single-cell data, and we recommend using algorithms for sciCUT&Tag that are designed for bulk CUT&RUN and CUT&Tag data, such as SEACR[159].

For example, we used SEACR to call enriched regions in aggregated sciCUT&Tag profiling of H3K27me3 and H3K4me1-2-3 in PBMCs by selecting the top 1% of regions by area under curve (AUC). Peaks can be plotted with genomic coverage and visualized in the UCSC genome browser[162]. We observed that genomic coverage at the *HOXA* locus, which is differentially regulated during hematopoiesis[163], varied substantially between H3K27me3 and H3K4me1-2-3 bulk profiles (**Fig. 5a**). Notably, both profiles show a dynamic signal range exemplified by high-coverage regions sharply juxtaposed against largely unoccupied regions. At a large genomic window spanning multiple megabases, which represents numerous 500-50kb genomic bins input as features for dimensionality reduction, H3K27me3 occupies separate regions from H3K4me1-2-3. Distinct coverage patterns are maintained even at smaller regions occupied by both H3K27me3 and H3K4me1-2-3, such as directly over the *HOXA* genes, where H3K27me3 forms broader domains than H3K4me1-2-3. In genome-wide analysis, we found that H3K27me3 peaks were broader than H3K4me1-2-3 peaks (**Fig. 5b**), which likely facilitates SNP-based collision identification for H3K27me3, as broad peaks contain overlapping reads which cover more sequence space than narrow peaks. Accordingly, H3K27me3 peaks contained 45.75% of the 192,668 SNPs detected across the pooled chromatin profiles, whereas H3K4me1-2-3 peaks contained only 36.14% of SNPs. Raw reads were also enriched within their respective peak set (**Fig. 5c**), which suggested a general trend of mutual exclusivity across the two bulk profiles. Indeed, only 4117 peaks had at least 10% reciprocal intersection (Jaccard similarity = 0.06) between the 19,853 H3K27me3 peaks and 22,785 H3K4me1-2-3 peaks (**Fig. 5d**). Together, visual inspection of genomic coverage coupled with peak calling on the aggregated data demonstrates high signal-to-noise and confirms the specificity of these two antibody sets.

3.6.3 Cell-type annotation and confirmation with bulk projection

Analysis of single-cell genomics data as described in step 148 generally includes cell-type annotation that is rooted in graph-based clustering. For graph-based clustering of sciCUT&Tag data, we used the approach implemented in Seurat[164], and then compared common metrics across clusters, including information content (reads/cell) and gene coverage (**Fig. 6**). Computing coverage at marker genes this way provides a strong basis for cell type annotation because sciCUT&Tag maps the chemical impression of chromatin modifying enzymes within individual cells.

sciCUT&Tag profiling of H3K27me3 in PBMCs revealed six different cell populations with different similarities to one another and varying information content (**Fig. 6a and b**). We identified two B cell populations that clustered together and had higher reads/cell than the dataset median (**Fig. 6b**; black dashed line; median = 2116 reads/cell), as well as three more lymphoid populations with fewer reads/cell. We also identified one myeloid population with information content far below the median, which is consistent with previous H3K27me3 profiling of PBMCs[85] and suggests that H3K27me3 may provide less signal in myeloid cells because they are less defined by the Polycomb group (PcG) chromatin modifiers that deposit H3K27me3[165].

To annotate cell types in the H3K27me3 PBMC dataset, we computed gene coverage to identify genes with high and low differential coverage across clusters. For H3K27me3, high coverage indicates a ‘repressed’ status (**Fig. 6c**), whereas low coverage indicates the opposite (e.g., ‘unrepressed’; **Fig. 6d**). We found that differentially H3K27me3-enriched genes included common immune cell biomarkers (i.e., *LEF1, CD4, CD14, EBF1, NCAM2, etc.*).

We performed a similar analysis for H3K4me1-2-3 in PBMCs, which revealed seven different cell populations, some of which were not distinguished by H3K27me3. For example, H3K4me1-2-3 provided less granular resolution of B cells than H3K27me3, but more granular resolution of T cells, which separated into three distinct clusters (**Fig. 6e**) versus only two identified by H3K27me3. H3K4me1-2-3 also distinguished two myeloid populations, where H3K27me3 only distinguished one. Of note, the two myeloid

populations identified by H3K4me1-2-3 exhibited different information content, with CD14 monocytes having lower reads/cell than CD16 monocytes. This suggests that H3K4 methylation distinguishes two myeloid populations by activity at only a few loci (Fig. 6f). Together, side-by-side analysis of H3K27me3 and H3K4me1-2-3 demonstrates that certain cell types are better distinguished by certain epitopes.

Importantly, gene coverage analysis validated our graph-based clustering by confirming that each cluster has a distinct coverage profile (**Fig. 6c,d,g,h**). Cell type annotation with H3K4me1-2-3 is straightforward because H3K4 methylation is mechanistically linked with chromatin remodeling[166] and correlated with gene activation[167]. High H3K4me1-2-3 coverage can therefore be understood to mark 'active' genes (**Fig. 6g**), whereas low coverage means that a gene remains inactive (**Fig. 6h**). We applied the same list of gene biomarkers used for H3K27me3 to H3K4me1-2-3, which revealed a surprising overlap of crucial immune genes regulated by both PcG and Trithorax group (TrxG) chromatin modifiers[160].

We used bulk chromatin profiling data from the ENCODE project[96] for orthogonal validation of CUT&Tag cell type annotation. Bulk chromatin profiling (*e.g.*, ChIP-seq, CUT&RUN, CUT&Tag) of any epitope can be randomly split into parts to simulate single-cell sparsity and projected into the same latent representation used for UMAP. To exemplify this validation technique, we projected ENCODE ChIP-seq data from immune populations purified by flow-assisted cell sorting (FACS) into our H3K27me3 and H3K4me1-2-3 UMAP embeddings (**Fig. 7**). For H3K27me3, this approach cleanly corroborated our cell type annotation based on gene coverages for all populations. For H3K4me1-2-3, however, bulk projection was less precise, as we profiled all forms of H3K4 methylation with sciCUT&Tag and only obtained bulk data for individual epitopes. Nonetheless, data for H3K4me1 and H3K4me3 projected into nearly all H3K4me1-2-3 sciCUT&Tag clusters and generally substantiated our gene coverage-based annotation.

In conclusion, the ability to obtain abundant and high-quality single-cell chromatin data at low cost with our sciCUT&Tag protocol enables the simultaneous profiling of multiple epitopes across multiple samples. Not only does combinatorial indexing on the ICELL8 make sciCUT&Tag tunable and transparent, but the easy-to-use pipeline for collision

removal via *de novo* genotyping yields high-confidence data for tens of thousands of single cells per experiment.

3.7 Figures

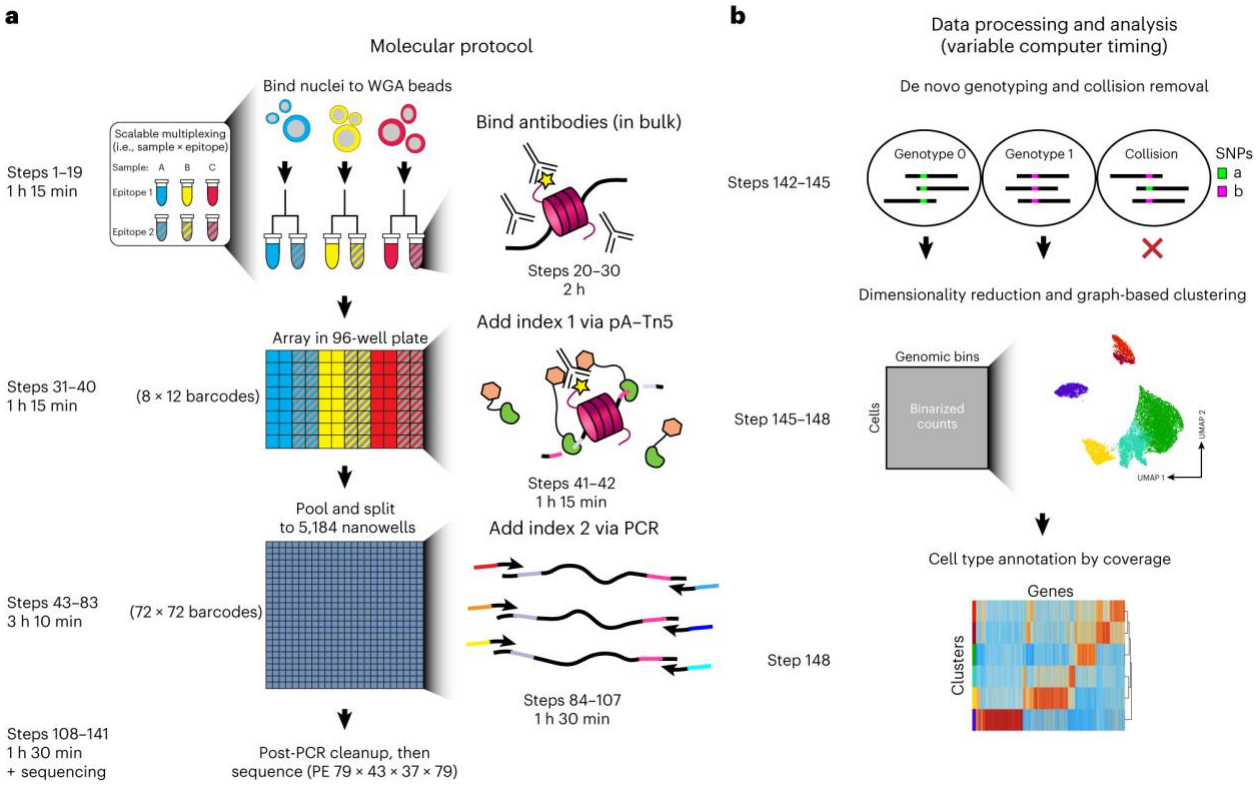


Figure 1 | Overview of the sciCUT&Tag experimental workflow. a) Schematic of the sciCUT&Tag sample processing procedure. Cells are first incubated with primary and secondary antibodies in bulk, then arrayed in 96-well format for tagmentation with barcoded pA-Tn5. The 96-well plate is then pooled and split with a TaKaRa ICELL8 robot to the 5184-well chip for PCR amplification that adds a second barcode prior to sequencing. Cells from different samples are indicated in blue, yellow and red; chromatin epitope conditions are hashed and unhashed; nucleosomes are colored fuschia with yellow star-shaped epitopes; pA-Tn5 is colored peach and green; barcoded indices are multicolored line segments next to black genomic template. **b**) Schematic of the informatics approach for sciCUT&Tag data processing. For mixed-donor samples, collisions are first detected by *de novo* genotyping to identify and remove barcodes with SNPs from multiple genotypes. Dimensionality reduction is then performed on matrix of

cells by genomic bins prior to graph-based clustering to identify subpopulations within a sample. These subpopulations than can then be assigned a celltype annotation based on gene coverage. Distinct SNPs at the same locus are green and pink within black cells. See text for details of the procedure.

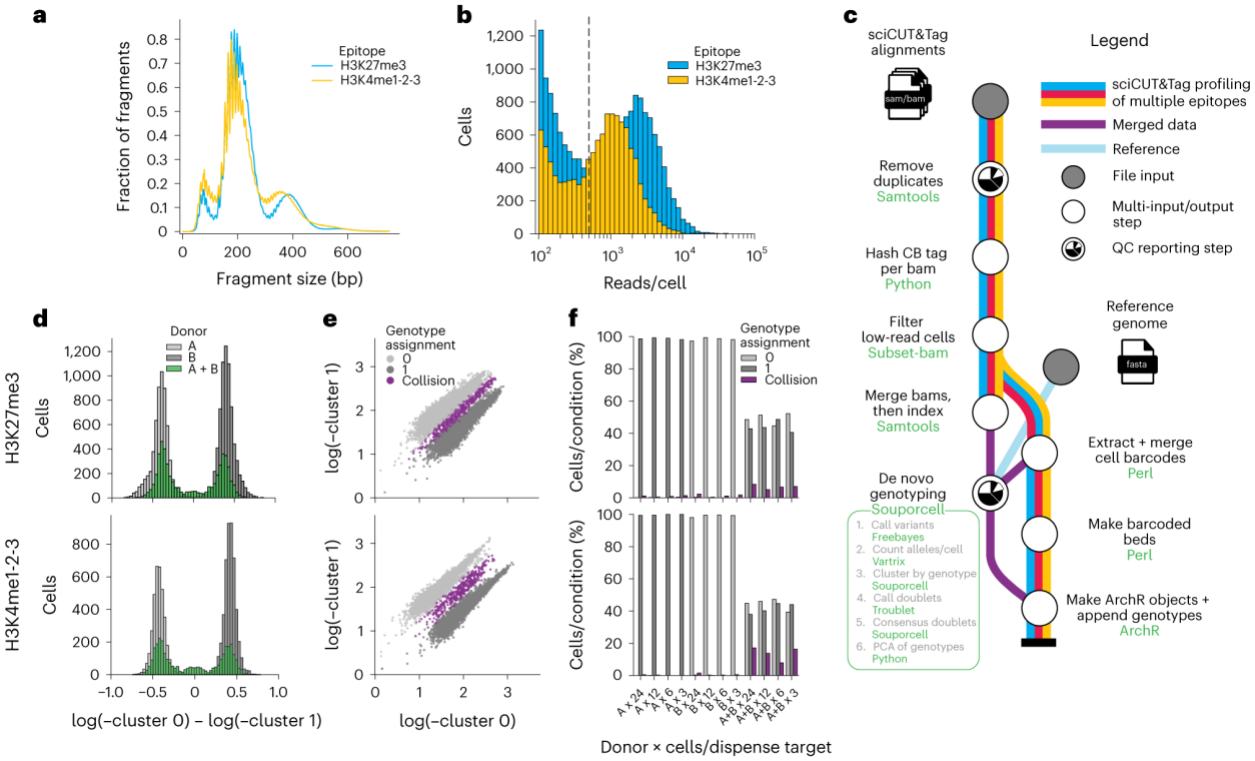


Figure 2 | Validation of a de novo genotyping strategy for SNP-based collision removal in sciCUT&Tag of mixed-donor PBMCs a) Fragment size distributions for H3K27me3 (blue) and H3K4me1-2-3 (yellow) sciCUT&Tag profiles. **b)** Reads/cell distributions for H3K27me3 (blue) and H3K4me1-2-3 (yellow) sciCUT&Tag profiles. A black dashed line marks 500 reads/cell and barcodes with less than 500 reads are excluded from further analyses. **c)** Nextflow[153] pipeline schematic built on Souporcell[151] for *de novo* genotyping and SNP-based collision removal. sciCUT&Tag profiling of multiple epitopes are the blue, red, and yellow paths that are merged into one purple path prior to *de novo* genotyping with Souporcell. Input files are genomic alignments and the reference genome used for alignment. Outputs are barcoded bed files, principal component analysis (PCA) of Souporcell cluster assignment probabilities, an ArchR[154] object annotated with genotype assignments, and tabulated cell metadata. Software

used in the pipeline are shown in green and the pipeline terminus is a black line. **d)** Cell counts versus the log differential of the genotype assignment cluster probabilities colored by donor for H3K27me3 (top) and H3K4me1-2-3 (bottom). **e)** The two genotype assignment cluster probabilities plotted against each other and colored by genotype assignment for H3K27me3 (top) and H3K4me1-2-3 (bottom). **f)** The percentage of cells per condition (donor x dispense target) for each genotype assignment for H3K27me3 (top) and H3K4me1-2-3 (bottom).

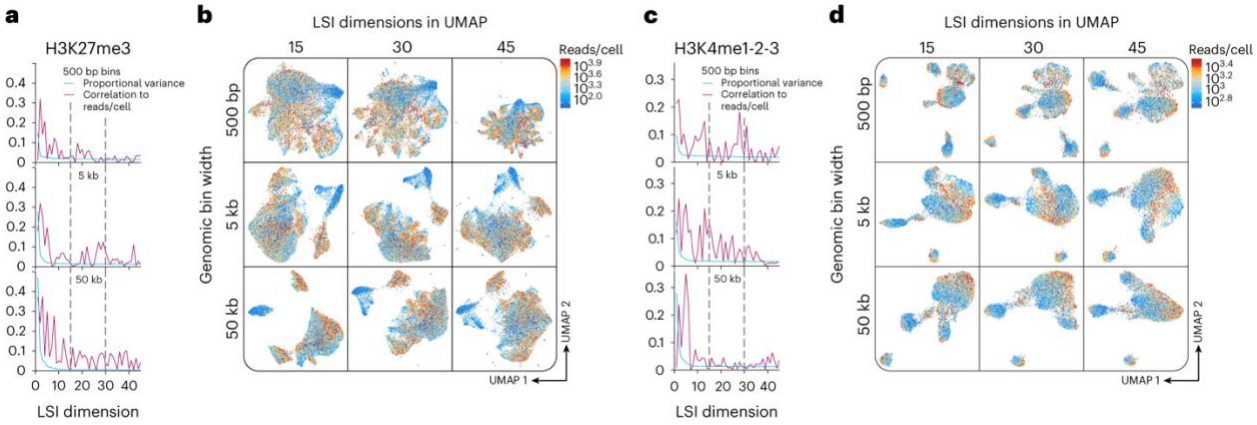


Figure 3 | Systematic investigation of dimensionality reduction parameters for sciCUT&Tag a) H3K27me3 profiling in PBMCs was binned at 500 bp, 5kb and 50kb genomic intervals, then input to iterative LSI[154] for dimensionality reduction. Proportional variance (blue) and correlation to reads/cell (red; Pearson's r) was calculated across the first 45 LSI dimensions. Grey lines mark dimensions 15, 30, and 45. **b)** For each bin width, UMAPs were computed from the first 15, 30, and 45 LSI dimensions and colored by reads/cell. **c and d)** same as **a** and **b** for H3K4me1-2-3. LSI, latent semantic indexing; UMAP, uniform manifold approximation and projection.

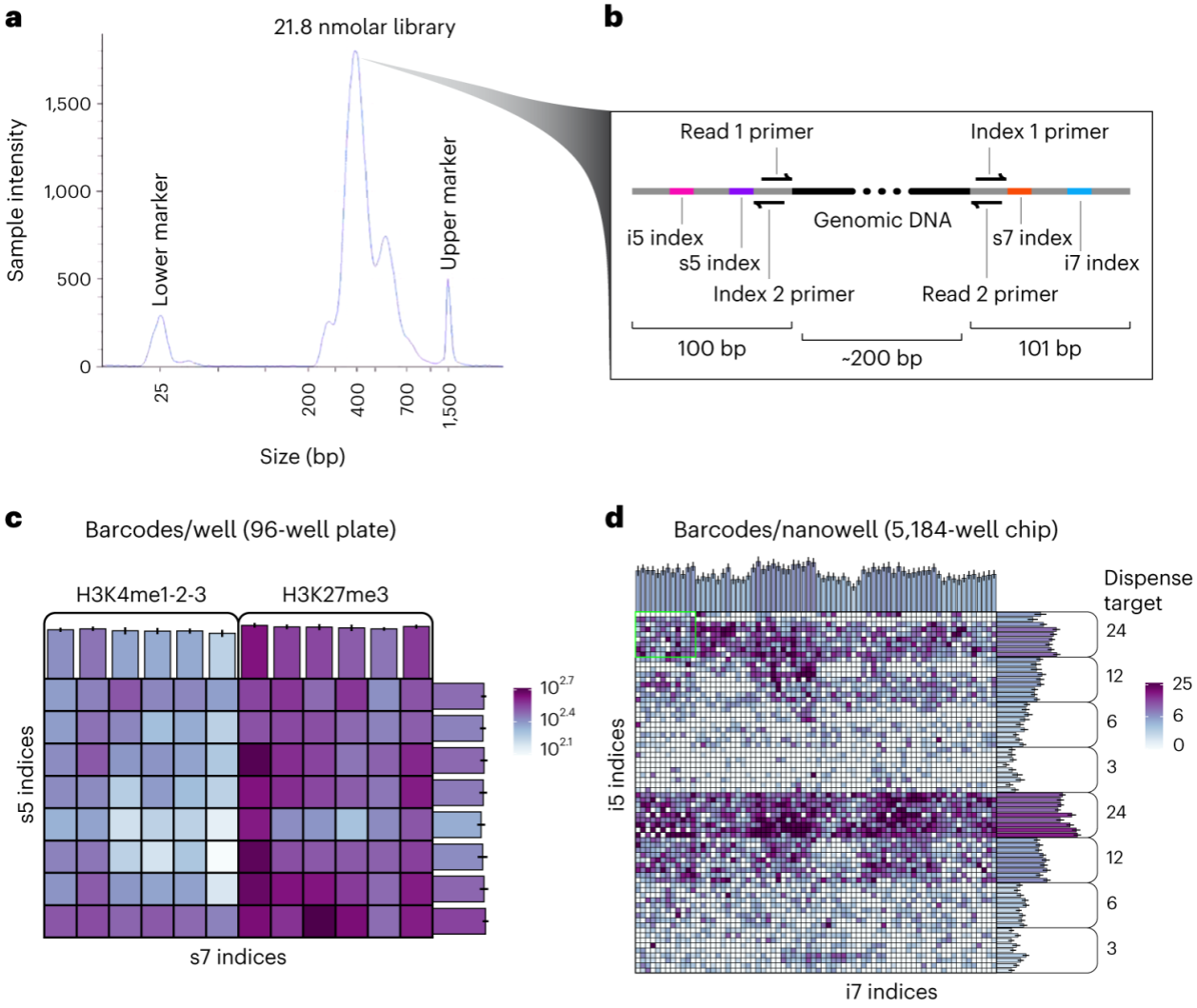


Figure 4 | Tunable combinatorial indexing in sciCUT&Tag with the ICELL8 system

a) Readout from the TapeStation for sciCUT&Tag profiling of H3K27me3 and H3K4me1-2-3 in PBMCs showing fragment size distribution. **b**) sciCUT&Tag barcoding schema. **c**) 96-well plate heatmap shaded per well by unique barcode combinations with greater than 100 reads recovered by sequencing. Marginal bar plots show the mean number of unique combinations recovered per s5 and s7 index and are shaded accordingly. Marginal bar plots for the s7 indices are labeled by profiled epitope. Error bars are standard deviation. **d)** 5184-well chip shaded by unique barcode combinations with greater than 100 reads recovered by sequencing per well. Marginal bar plots show the mean number of unique combinations recovered per column and row and are

shaded accordingly. Marginal bar plots along the y-axis are labeled by the cell/well dispense target. Error bars are standard error.

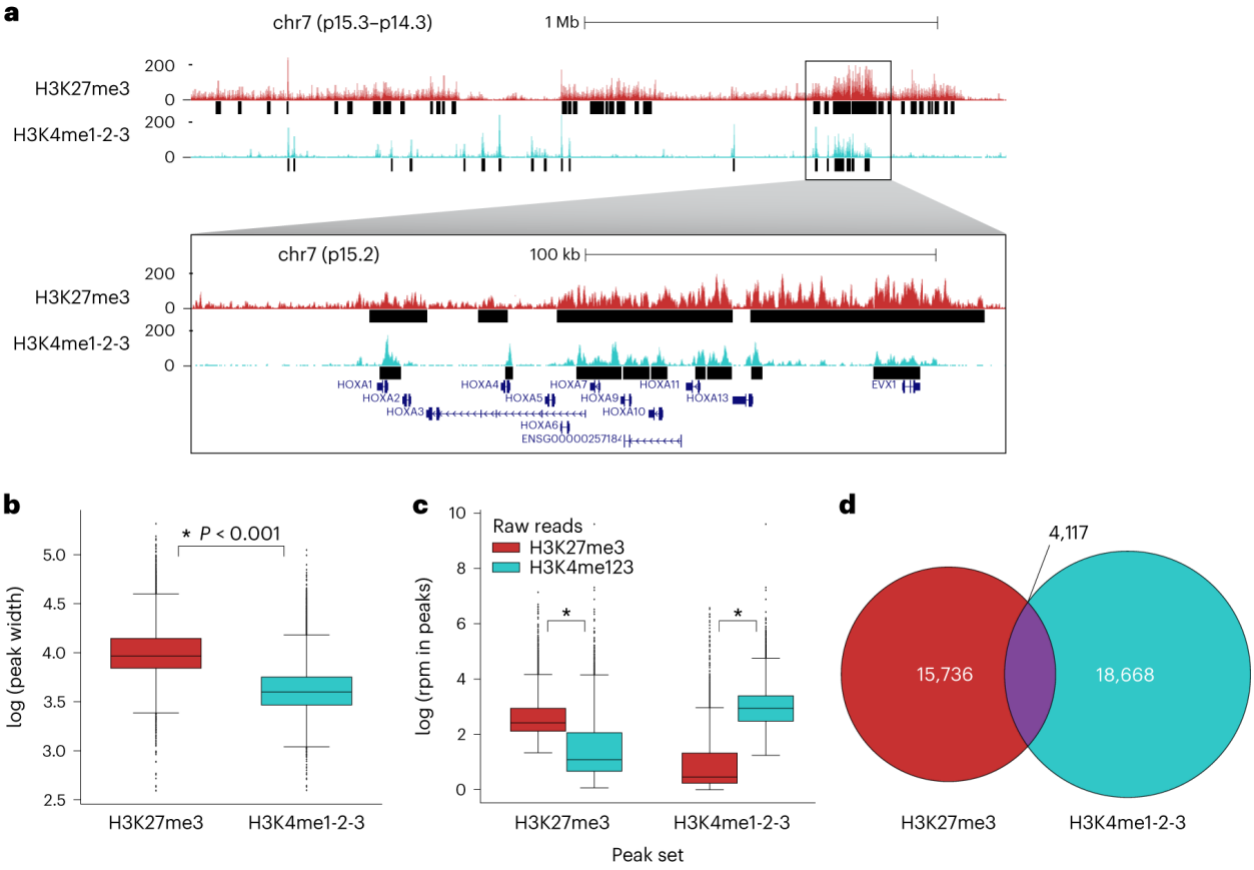


Figure 5 | Visual inspection of genomic coverage and peak-calling analysis a) Genomic coverage at the *HOXA* locus on chromosome 7 for H3K27me3 (red) and H3K4me1-2-3 (teal) aggregated profiles. Peaks detected by SEACR[159] (black) are plotted below their corresponding track. **b**) Peak width comparison between H3K27me3 and H3K4me1-2-3 peak sets (independent t-test). Boxplots show the median with 1st and 3rd quartiles and outliers. **c)** Comparison of H3K27me3 and H3K4me1-2-3 reads-per-million in the two respective peak sets. Raw reads-per-peak were normalized by total fragment counts per million. Boxplots show the median with 1st and 3rd quartiles and outliers. **d)** Quantification of overlapping peaks across the H3K27me3 and H3K4me1-2-3 peak sets.

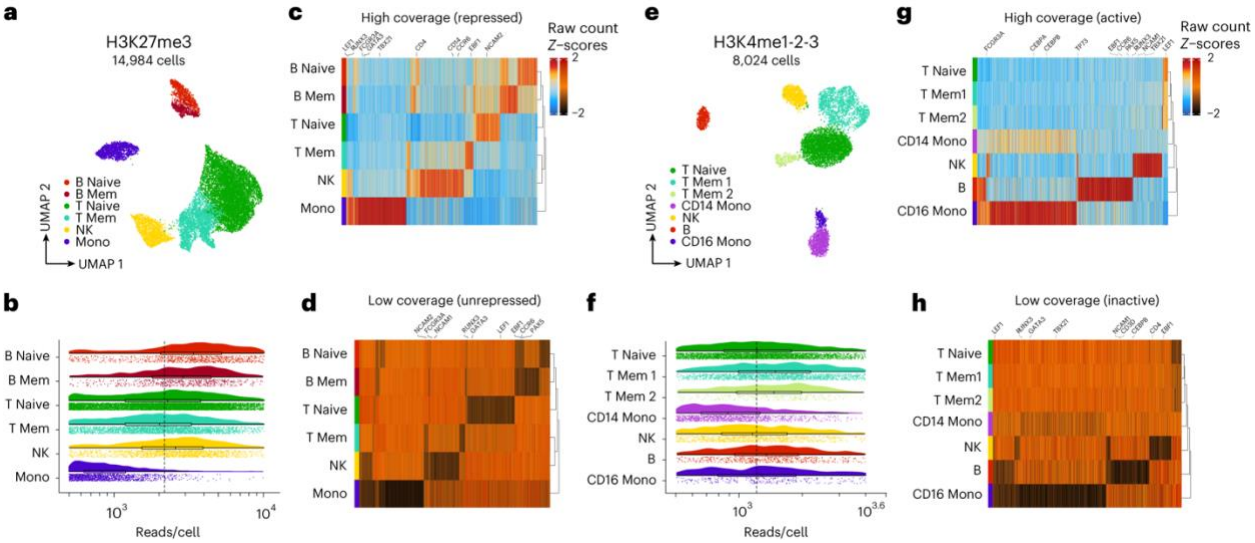


Figure 6 | sciCUT&Tag supports graph-based clustering and cell type annotation of human peripheral blood mononuclear cells

a) UMAP embedding for H3K27me3 profiling in PBMCs colored by sciCUT&Tag cell type annotation based on gene coverage. Each point represents one cell. **b)** Reads/cell distribution per population in H3K27me3 dataset. Black dashed line is the dataset median. Density plot heights are normalized, boxplots show the median with 1st and 3rd quartiles. Each point represents one cell. **c)** Heatmap of gene coverages per population for H3K27me3 with a log fold-change greater than 0.5 under an FDR of 0.01. **d)** Heatmap of gene coverages per population for H3K27me3 with a log fold-change less than 0.5 under an FDR of 0.0. Heatmap data is row-normalized, binary sorted by row, and clustered by column. **e-h)** same as **a-d** for H3K4me1-2-3. PBMCs, peripheral blood mononuclear cells; Mem, memory; NK, natural killer; Mono, monocyte; UMAP, uniform manifold approximation and projection.

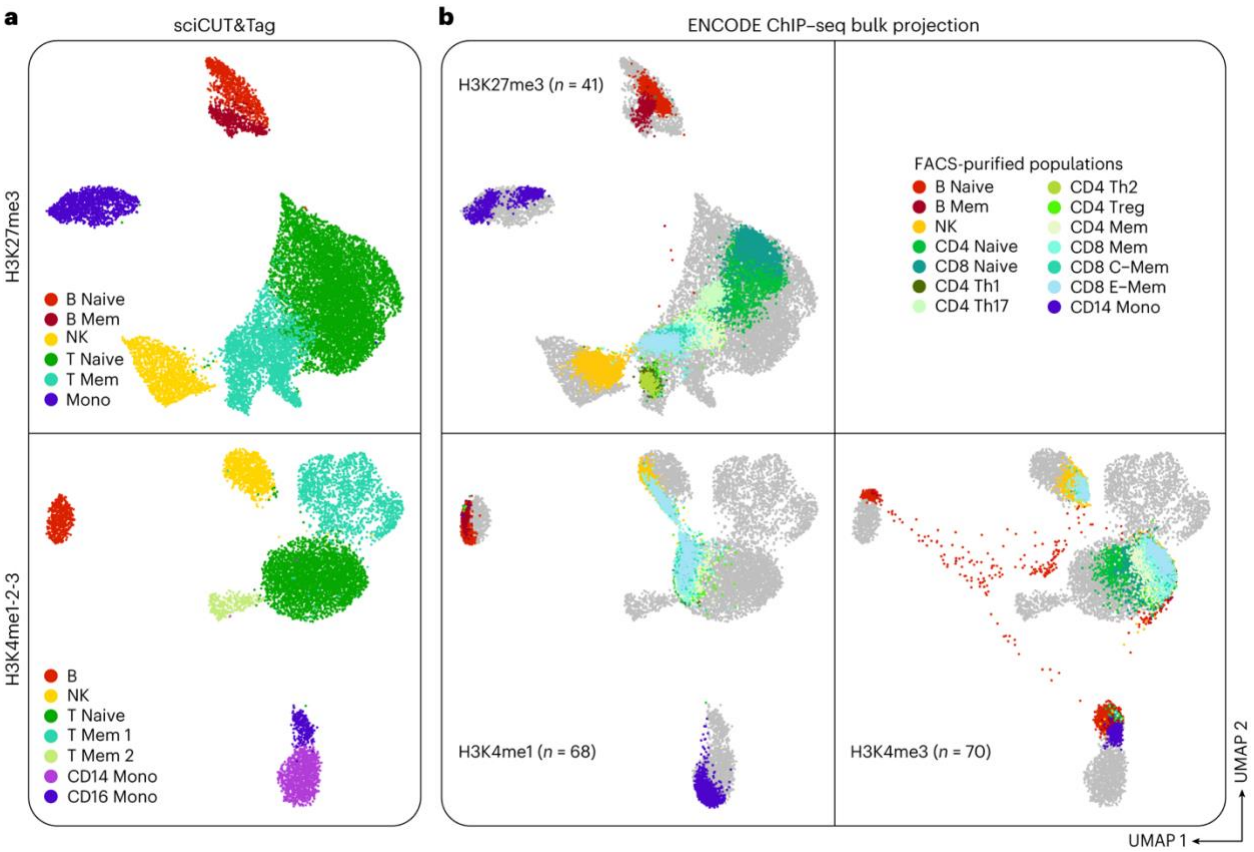
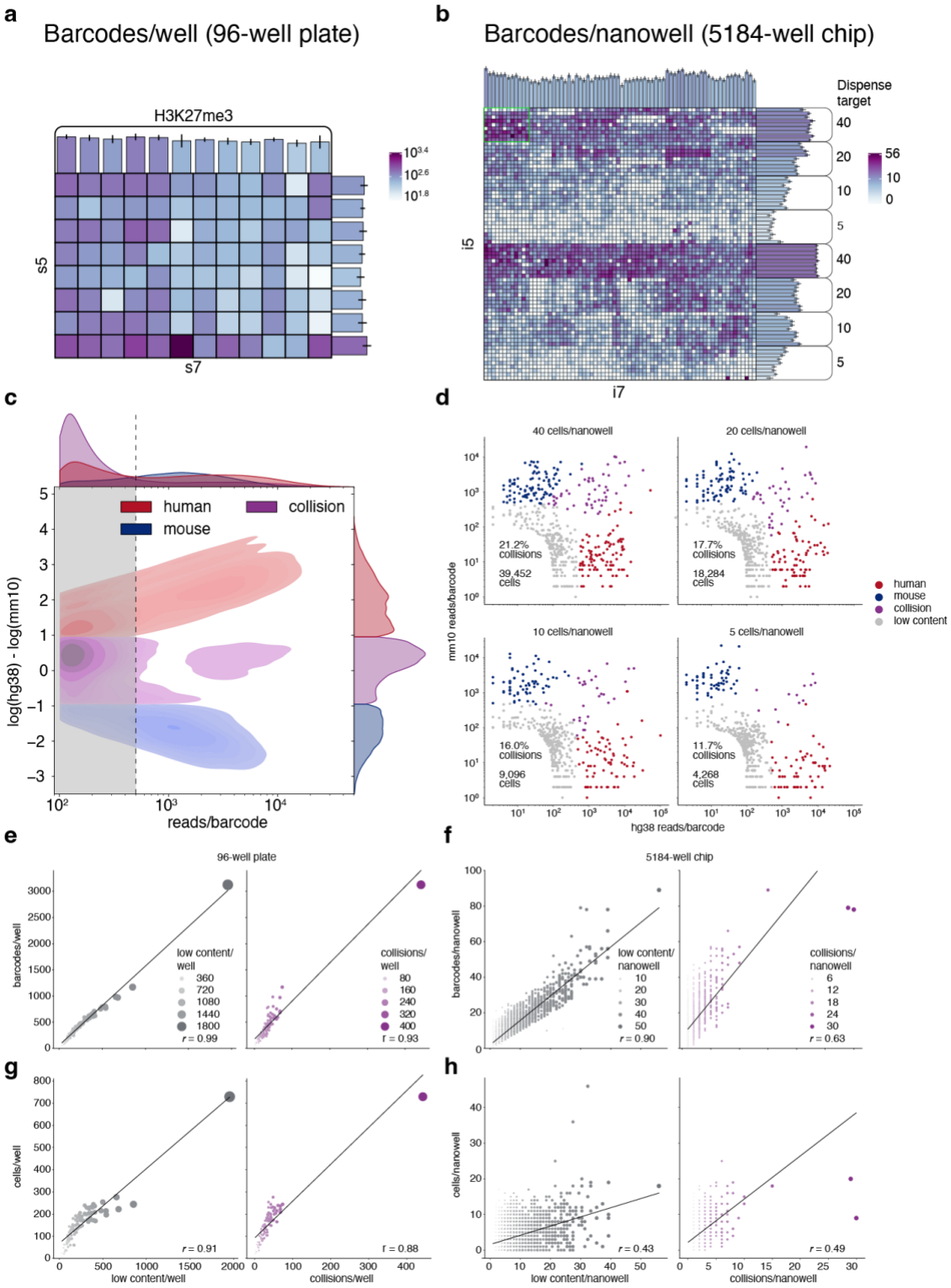


Figure 7 | Orthogonal validation of sciCUT&Tag cell type annotation by bulk projection of public data

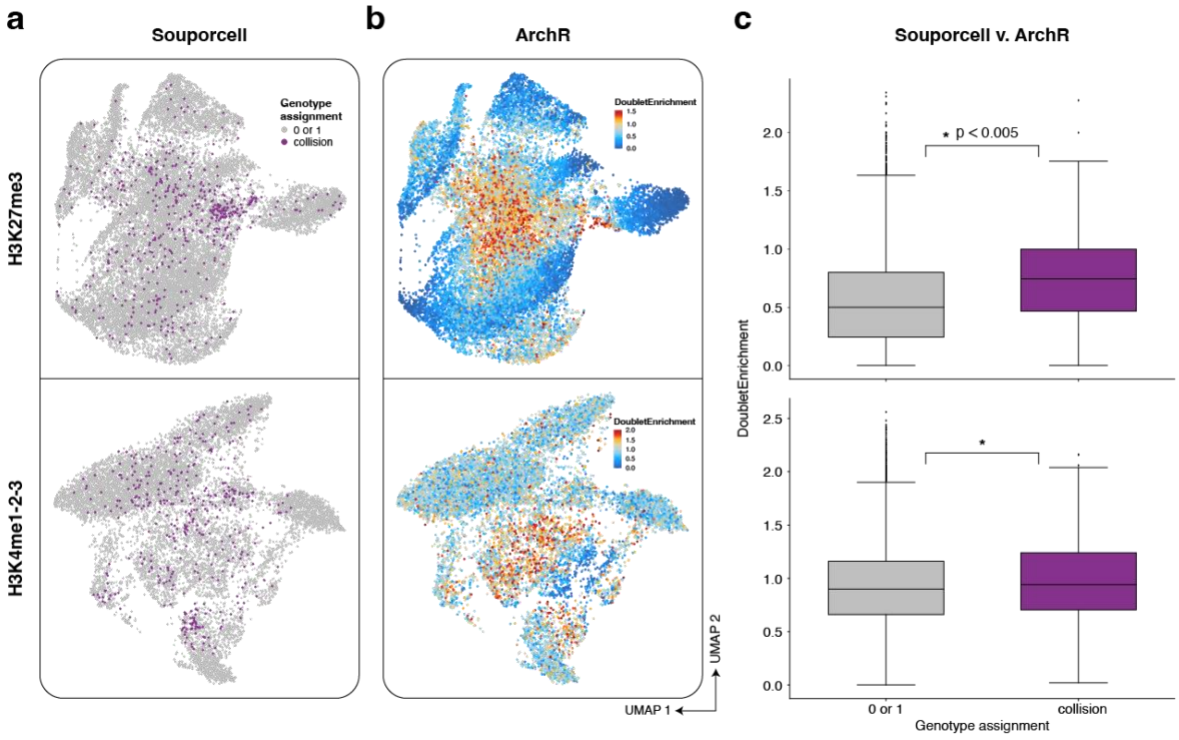
a) UMAP embeddings for H3K27me3 (top) and H3K4me1-2-3 (bottom) profiling in PBMCs colored by cell type annotation based on gene coverage in sciCUT&Tag. Each point represents one cell. **b)** Bulk projections of ENCODE ChIP-seq into sciCUT&Tag UMAP embeddings. Counts per genomic bin (50 kb for H3K27me3, 500 bp for H3K4me1 and H3K4me3) were randomly split into 500 pseudo-cells per bulk dataset, then projected via sciCUT&Tag LSI representations. H3K27me3 profiles for FACS-purified populations were projected into H3K27me3 sciCUT&Tag (top) and H3K4me1 (bottom left) and H3K4me3 (bottom right) profiles were projected into H3K4me1-2-3 sciCUT&Tag. Mem, memory; NK, natural killer; Th, T-helper; Treg, T-regulatory; C-Mem, central memory; E-mem, effector memory; Mono, monocyte.

3.8 Supplementary Figures



Supplementary Figure 1 | Mapping the determinants of collisions and low-content barcodes in sciCUT&Tag a) 96-well plate heatmap shaded per well by unique barcode combinations with greater than 100 reads recovered by sequencing in the barnyard experiment that mixed human and mouse cell lines at a 1:1 ratio. Marginal bar plots show the mean number of unique combinations recovered per s5 and s7 index and are shaded accordingly. Marginal bar plots for the s7 indices are labeled by profiled epitope. Error bars are standard deviation. **b)** 5184-well chip shaded by unique barcode combinations with greater than 100 reads recovered by sequencing per well in the barnyard experiment. Marginal bar plots show the mean number of unique combinations recovered per column and row and are shaded accordingly. Marginal bar plots along the y-axis are labeled by the cell/well dispense target. Error bars are standard error. **c)** Bivariate density plot of the log differential of reads/barcode that uniquely align to hg38 or mm10 versus total reads/barcode. Barcodes with >90% of reads uniquely aligned to hg38 are called as human (red), likewise for mm10 (mouse, blue), and barcodes with <90% of reads uniquely aligned to either hg38 or mm10 are called as a collision (purple). Marginal plots show univariate density. A black dashed line marks 500 reads/barcode and barcodes with less than 500 reads are shaded gray. **d)** Scatter plots of 500 randomly sampled barcodes for each dispense target. Each dot represents one unique barcode combination plotted according to number of reads that uniquely map to hg38 or mm10. Barcodes with >90% of reads uniquely aligned to hg38 are called as human (red), likewise for mm10 (mouse, blue), and otherwise are called as a collision (purple). Low-content barcodes, which are excluded from downstream analysis, are colored gray. Rate of collision is percentage of collisions per total cells. Cells/chip is estimated as four times the yield from a quarter of one 5184-well chip. **e)** Barcodes/well linearly correlated against low-content barcodes/well (gray) and collisions/well (purple) for the 96-well plate. Each dot represents one well of the 96-well plate and is sized and shaded according to low-content barcodes/well and collisions/well, respectively. **f**) same as **c** for the 5184-well chip. **e)** High-content (e.g., >500 reads/barcode) cells/well linearly correlated against low-content barcodes/well (gray) and collisions/well (purple) for the 96-well plate. Each dot represents one well of the 96-well plate and is sized and shaded according to low-content barcodes/well and collisions/well, respectively. **f**) same as **c** for

the 5184-well chip. Correlations are Pearson's R and the best-fit linear regression is shown in black for **e-h**.



Supplementary Figure 2 | Souporcell identifies collisions also detected by artificial nearest neighbor testing in ArchR a) UMAPs for H3K27me3 (top) and H3K4me1-2-3 (bottom) profiling in PBMCs that include the collisions identified by Souporcell[151] (purple). **b)** Same as **a** but UMAPs are colored by DoubletEnrichment score from addDoubletScores() in ArchR[154]. **c)** Comparison of the ArchR DoubletEnrichment score in Souporcell collisions v. non-collisions (0 or 1 genotype assignment).

Chapter 4. Histone H3K27 methylation dynamics during blood development

4.1. Introduction

How do H3K27 methylation patterns in the genome arise during development? These patterns are first established during embryogenesis, maintained in committed progenitors, and built upon during lineage specification. In blood development hematopoietic stem and progenitor cells (HSPCs) in the bone marrow give rise to the white and red blood cells of the lymphoid, myeloid, and erythroid lineages through coordinated changes in gene expression that are orchestrated at the level of chromatin. In each lineage, many genes gain the H3K27me3 mark. In this chapter, I investigate the dynamics of H3K27 methylation gained during blood cell development using single-cell CUT&Tag profiling of human bone marrow.

4.2. Results

4.2.1 Establishing single-cell profiling of H3K27 methylation states in human blood cells

To separately validate single-cell measurement of each H3K27 methylation state, H3K27ac, and active RNA polymerase II (RNAPII) phosphorylates at serine 2 or 5, we applied sciCUT&Tag chromatin profiling [86] to human peripheral blood mononuclear cells from two healthy donors. After genotypic doublet removal, we assessed cell type resolution via iterative latent semantic indexing followed by uniform manifold approximation [154] and projection (UMAP) [155]. For the H3K27 tri- and di-methyl modifications enriched at repressed and inactive genes, we expected depletion of the marks. For the H3K27 mono-methyl and acetyl modifications, as well as active RNAPII, we expected enrichment at actively transcribed genes. Like our previous application of the method [86], the H3K27me3 mark resolved 3-4 major immune cell types based on marker gene coverage (Figure 1). For example, H3K27me3 coverage over B-lymphoid

transcription factor *PAX5* was depleted for one region, marking putative B-cells. The T-lymphoid transcription factor *LEF1* and natural killer gene *NCAM1* that encodes the cell surface marker CD56 were depleted over a separate region of the embedding, marking T and natural killer cells, respectively. This confirmed resolution of the lymphoid subtypes by the H3K27me3 mark (Figure 1 i).

Consistent with mutually exclusive coverage of H3K27 methylation states [88], genes exhibited different levels of coverage for each H3K27 mark. The *SPI1* gene, which encodes the key myeloid transcription factor PU.1, did not show variable enrichment for the H3K27me3 mark. In contrast, we did distinguish depletion of the H3K27me2 mark over the *SPI1* gene in a cluster of cells defined by H3K27me2 modification (Figure 1 ii). Likewise, the H3K27me2 modification separated out multiple clusters of cells in UMAP space, some which also showed H3K27me2 depletion at the lymphoid markers *PAX5* and *LEF1*. *NCAM1* was poorly resolved by the di-methyl mark. Of the five marks, the H3K27me1 modification, which corresponds with gene activity [88], exhibited the lowest median number of fragments and had low cell-type specificity (Figure 1 iii). In contrast, the H3K27ac modification as well as active RNAPII resolved four separate clusters in the embedding distinguished by all four marker genes (Figure 1 iv-v). This indicates that chromatin profiling of H3K27 modifications can distinguish immune cell types to varying degrees.

4.2.2 Single-cell profiling of H3K27 methylation states in human bone-marrow

We next set out to capture the dynamics of H3K27 modification during human blood development. We applied sciCUT&Tag profiling to bone-marrow mononuclear cells depleted for CD3-positive T-cells, which mature outside the bone marrow in thymus. To anchor our analysis in hematopoietic stem cells (HSCs), we enriched for the CD34-positive population. In 4 donors, we profiled each H3K27 modification alongside active RNAPII using the 2-in-1 variation of the sciCUT&Tag protocol (Figure 2a) [168]. In this experiment, the first round of tagmentation barcodes sites occupied by active RNAPII. After blocking with a single-chain variable fragment (scFv), the second round of tagmentation barcodes sites occupied by a H3K27 modification. We next assigned cell

types based on the active RNAPII profiles paired with each H3K27 modification state in single cells (Figure 2b). The two BMMC donors were enriched in lineage-committed monocyte, B-lymphoid, and erythroid cells, and the two CD34+ donors were enriched in the progenitor populations and HSCs (Figure 2c). Across cell types, we observed a consistent distribution of the paired profiles for each H3K27 modification (Figure 2d). This enabled comparison of the four H3K27 modifications based on a shared, RNAPII-defined reference embedding and cell type annotation.

4.2.3 Comparative dynamics of H3K27 methylation in three hematopoietic lineages

To compare H3K27 methylation in the monocyte, B-lymphoid, and erythroid lineages, we first defined which genes gained and lost the H3K27me3 mark versus HSCs (Figure 3). For statistical analysis, we used donors as replicates and found between 2.5-4 thousand differential genes per lineage. Consistent with its repression in B-lymphocytes, the *INPP4B* gene that regulates T-lymphocyte fate [169] significantly gained the H3K27me3 mark (Figure 3a). In contrast the *PAX5* gene encoding the B-lymphoid transcription factor significantly lost the mark. In erythrocytes, the gene encoding the *FLI1* megakaryocyte transcription factor [170] significantly gained the mark (Figure 3b). Consistent with its repression in monocytes, the *GATA2* gene encoding a key HSC transcription factor [171] acquired the H3K27me3 mark (Figure 3c). In contrast the myeloid gene *IRF8* lost the mark in the monocytes. Our data confirmed that stemness genes and regulators of alternate lineages acquire the H3K27me3 mark during hematopoietic development.

It is unknown how H3K27 methylation dynamics give rise to the H3K27me3 mark in hematopoietic lineages. This may occur via processive catalysis of the tri-methyl mark [24,75,172] or via an intermediate di-methyl state, like in metazoan embryogenesis [78,118,119] and cycling human cells [88]. To chart H3K27me2 dynamics preceding H3K27 tri-methylation, we constructed a pseudotime axis based on the RNAPII embedding in BMMCs. After assigning cells to B-lymphoid, monocyte, and erythroid lineages, we plotted pseudotime trends for active RNAPII, H3K27me2, and H3K27me3 at the genes which significantly gain the H3K27me3 mark (Figure 4a). In general, we observed an

early RNAPII maxima at these genes in the HSCs and progenitor populations that diminished over pseudotime, albeit less rapidly in the erythroid lineage. After loss of RNAPII in B-lymphoid and monocyte development, the H3K27me2 mark increased at most genes, followed by the H3K27me3 mark. In the erythroid lineage, the H3K27 di- to tri-methyl conversion appeared concomitant at the pseudotime terminus. To quantify the dynamics for these three marks, we charted the differences between trend maxima per gene across the lineages. In the B-lymphoid and monocyte lineages, we found a significant difference between the H3K27 di- to tri-methyl maximum occupancy that did not hold in the erythroid lineage (Figure 4b i). At the vast majority of genes gaining the H3K27me3 mark in all three lineages, the H3K27me2 and H3K27me3 marks followed after RNAPII max occupancy (Figure 4b ii-iii). This suggests that the H3K27me3 mark is gained sequentially in B-lymphoid and monocyte cells, but not erythroid development. In all three lineages, the H3K27me3 mark accumulated only after RNAPII levels thoroughly diminished.

4.3. Discussion

Our analysis of H3K27 methylation dynamics reveals that thousands of genes which gain the H3K27me3 mark in B-lymphoid and monocyte development proceed via a di-methyl intermediate state. We have previously shown that H3K27me2 marks inactive chromatin that does not repress transcription [88]. Like in embryogenesis, this would supports that H3K27 tri-methylation initiates Polycomb repression at many inactive genes after transient activation signals fade [39]. Sequential H3K27 methylation supports a distributive catalytic mechanism by PRC2 that is consistent across early embryogenesis, where the mark is established *de novo,* and lineage specification of committed progenitors, where some genes have previously gained the mark. While the H3K27me3 mark promotes allosteric activation of PRC2 to propagate itself onto neighboring nucleosomes [25], this process is likely slow and step-wise before nucleosomes compact. Indeed, chromatin condensation promotes PRC2 catalytic activity [53]. In erythroid development, it is possible that the chromatin condensation which precedes enucleation facilitates rapid H3K27 tri-methylation [173]. Further analysis of

single-cell chromatin profiling that captures H3K27 methylation dynamics promises to resolve the function of the Polycomb repressive system across developmental contexts.

4.4 Figures

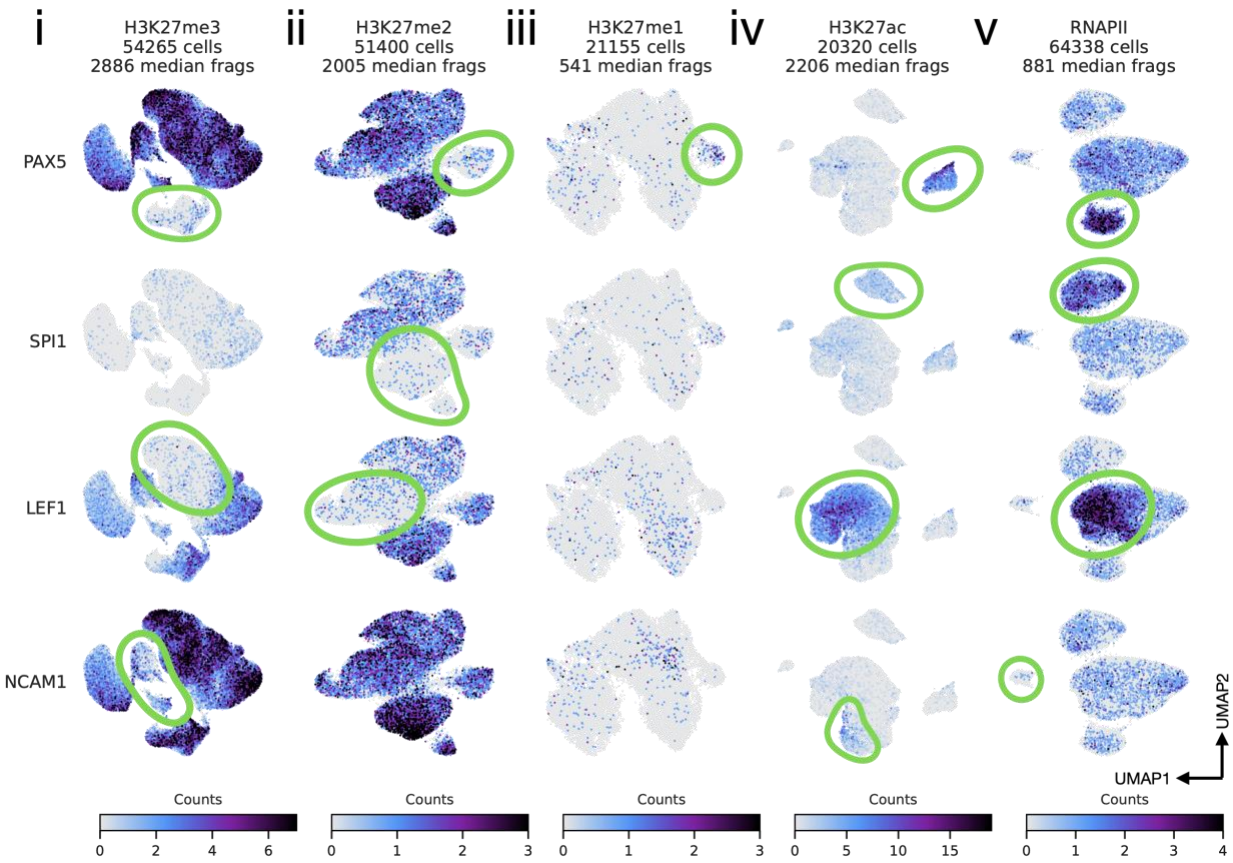


Figure 1. Marker gene coverage of H3K27 modifications and active RNAPII in healthy peripheral blood mononuclear cells (PBMCs): Read counts in select marker genes are visualized in UMAP space for the (i) H3K27me3, (ii) H3K27me2, (iii) H3K27me1, (iv) H3K27ac, and (v) RNAPII marks. Regions of each embedding that correspond to putative cell types are outlined in green. For the H3K27me3 and H3K27me2 marks, the outlined regions are depleted for the mark. For the H3K27me1, H3K27ac, and RNAPII marks the outlined regions are enriched for the mark. *PAX5* coverage marks B cells, *SPI1* coverage marks monocytes, *LEF1* coverage marks T cells, and *NCAM1* coverage marks natural killer cells. The number of cells and median fragment counts are denoted per dataset.

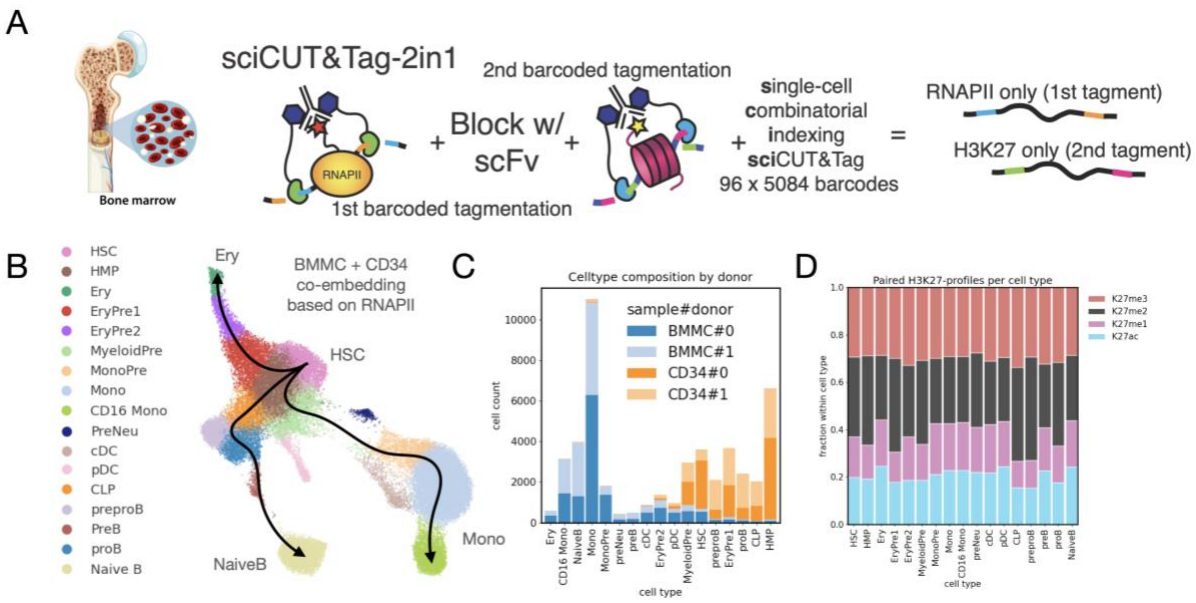


Figure 2. sciCUT&Tag-2-in1 profiling of active RNAPII and H3K27 modifications in healthy bone-marrow mononuclear cells (BMMCs): (A) Diagram of the sciCUT&Tag-2-in1 paradigm applied to BMMCs. (B) Cell type assignment based on the RNAPII co-embedding of BMMCs and CD34+ positive cells. Black arrows denote the B-lymphoid, monocyte, and erythroid lineages stemming from the HSCs. (C) Cell type composition by donor for the two BMMC and two CD34+ samples profiled in parallel in one sciCUT&Tag-2-in1 experiment. (D) Cell type distriubtion of the H3K27 modification profiles paired with the RNAPII in each single cell.

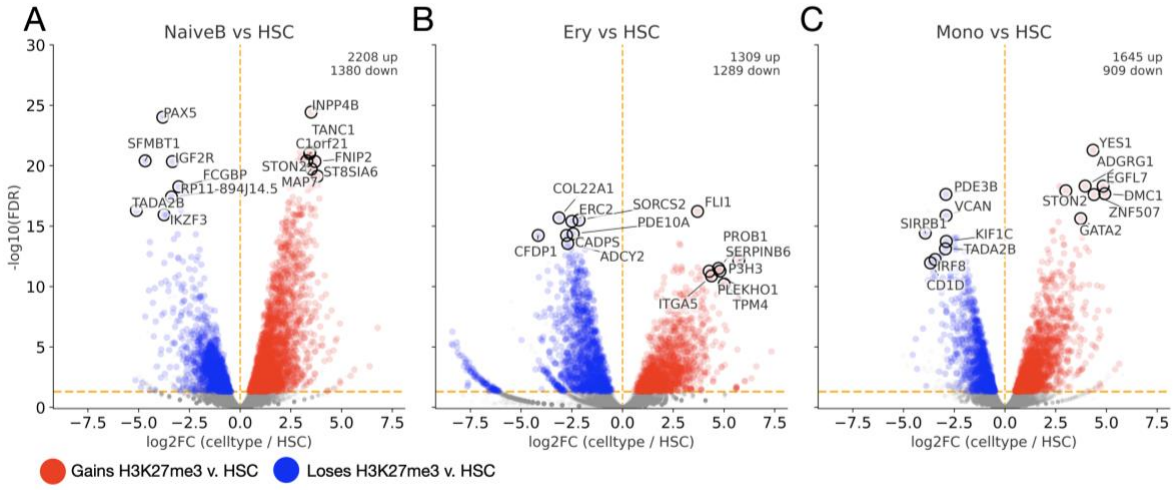


Figure 3. Differential analysis of the H3K27me3 mark in three hematopoietic lineages: Volcano plots of the H3K27me3 mark in the (A) B-lymphoid, (B) erythroid, and (C) monocyte lineages versus HSCs. Above 0.5 RPKM coverage, genes which significantly gain the mark versus HSCs are shown in red, and genes which significantly lose the mark versus HSCs are shown in blue.

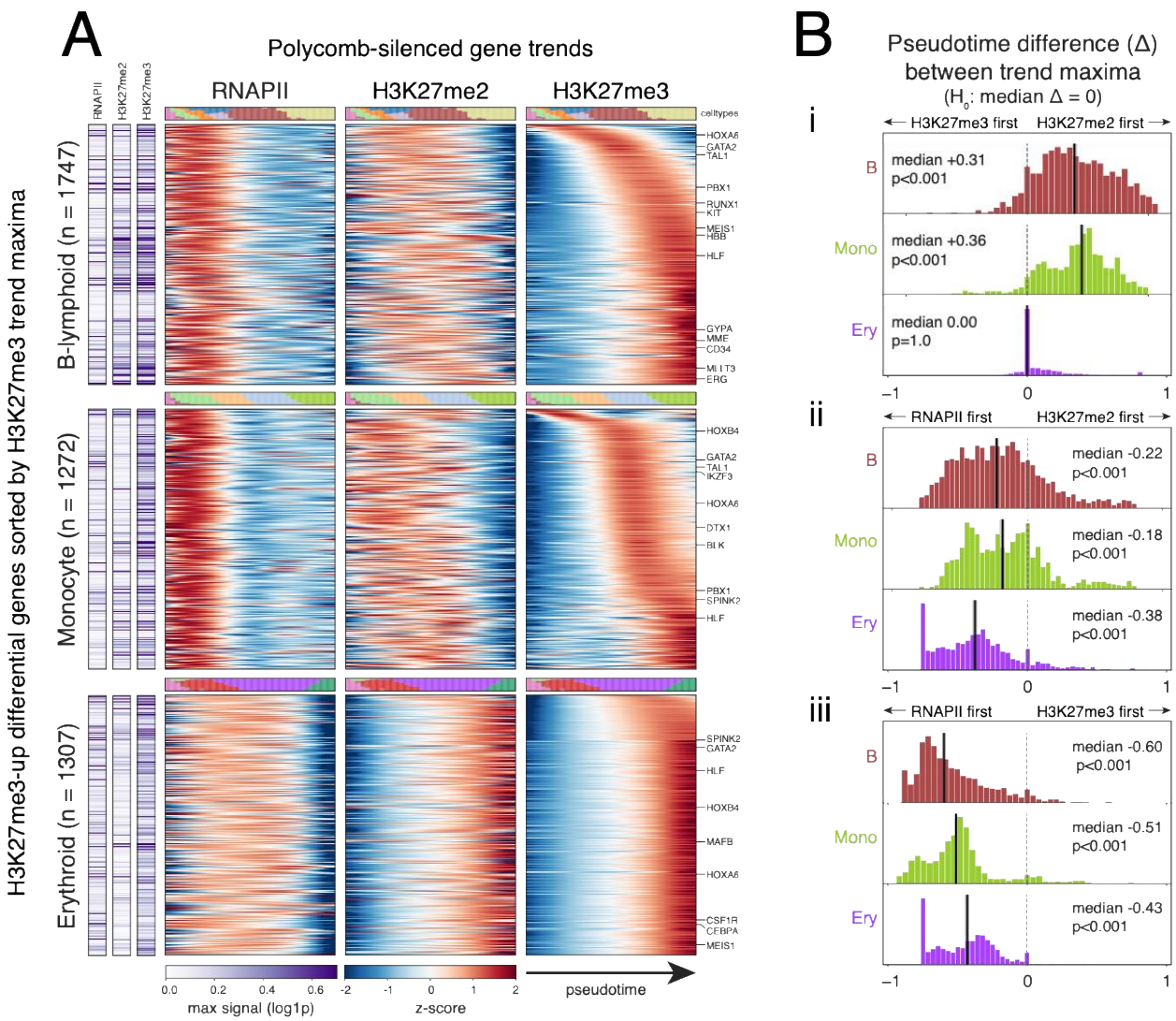


Figure 4. Trend analysis of genes that gain the H3K27me3 mark in three hematopoietic lineages: (A) z-scored trend heatmaps for RNAPII, H3K27me2, and H3K27me3 at genes that gain the H3K27me3 mark in the (top) B-lymphoid, (middle) monocyte, and (bottom) erythroid lineages. Max signal for each mark is shown per gene as log1p-normalized counts. Genes with less than 0.15 max H3K27me3 signal were excluded. Cell type colors (from figure 2b) are shown along pseudotime on the x-axis. Genes are ordered along the y-axis by pseudotime position of H3K27me3 trend maxima. Select hematopoietic marker genes are highlighted. (B) Pseudotime difference between trend maxima for (i) H3K27me3-H3K27me2 (ii) H3K27me2-RNAPII and (iii) H3K27me3-RNAPII in B-lymphoid, monocyte, and erythroid lineages. Median difference is shown with the empirically derived p-value based on bootstrapping versus a null hypothesis of no difference.

4.5 Methods

Peripheral blood mononuclear cell isolation

Peripheral blood mononuclear cells were isolated as in [86].

CD3 depletion of human bone marrow

CD3-targeted magnetic separation was used to deplete the T cells from bone marrow mononuclear cells (BMMCs). 2.5 × 10^7 BMMCs were thawed in a 37°C water bath and diluted dropwise with 15 mL IMDM (Life Technologies, #12440061) supplemented with 10% FBS (Thermo Fisher Scientific, #A5670801) in a 50 mL conical tube. Cells were passed through a 100 µm MACS SmartStrainer (Miltenyi Biotec, #130-098-463) and centrifuged at 400 × *g* for 5 minutes. The cell pellet was resuspended in 200 µL of wash buffer (1X PBS, 2 mM EDTA, 0.5% BSA) and transferred to a 1.5 mL tube. Mixed CD3 MicroBeads (50 µL; Miltenyi Biotec, #130-050-101) were added, and samples were incubated at 4°C for 15 minutes. Following incubation, samples were washed with 1.25 mL wash buffer, centrifuged at 300 × *g* for 5 minutes, and resuspended in 500 µL wash buffer.

For magnetic separation, an LS column (Miltenyi Biotec, #130-042-401) was equilibrated with 3 mL wash buffer. Samples were applied to the column, and the CD3-negative fraction (T cell-depleted BMMCs) was collected in a 15 mL conical tube. The column was rinsed twice with wash buffer, and the flow-through was collected in the same tube. Cell viability and concentration were determined using a Countess cell counter.

Antibodies

We used the following antibodies: Guinea Pig anti-Rabbit IgG (Heavy & Light Chain) (Antibodies Online, Cat# ABIN101961), H3K27me3 (Rabbit monoclonal anti-H3K27me3, Cell Signaling Technology, Cat# 53207), H3K27me2 (Rabbit polyclonal anti-H3K27me2, Abcam, Cat# ab24684), H3K27me1 (H3K27me1 Recombinant Polyclonal Antibody, Thermo Fisher Scientific, Cat# 712817), H3K27ac (Rabbit monoclonal anti-H3K27me3, Cell Signaling Technology, Cat# 53207), RNAPIIS2/5P

(Rabbit monoclonal Phospho-Rpb1 CTD (Ser2/Ser5), Cell Signaling Technology, Cat# 13546), RNAPIIS2P (Rabbit monoclonal anti-Phospho-Ser2-Rpb1 CTD, Cell Signaling Technologies, Cat#13499), and RNAPIIS5P (Rabbit monoclonal anti-Phospho-Ser2/Ser5-Rpb1 CTD, Cell Signaling Technologies, Cat#13546). The final concentrations of the antibodies used in sciCUT&Tag are specified in the following sections. Antibodies were validated using the SNAP-CUTANA™ K-MetStat Panel according manufacturers guidelines (EpiCypher, Cat# 19-1002).

sciCUT&Tag assay

sciCUT&Tag was performed as in [86], and the 2-in-1 variation performed as in [168]

sciCUT&Tag data analysis

Data preprocessing was performed in [86] to generate paired-end bed files with cell barcodes. 50-kilobase (kb) tile matrices were constructed using ArchR [154] and input to iterative LSI for dimensionality reduction using default parameters. UMAPs [155] are based on 30 k-nearest neighbors, 0.5 minimum dist, and cosine distance. Gene coverage was computed at coding genes based on the longest transcript (GENCODE v49) + 1 kb upstream and + 5 kb downstream sequence, unless this overlapped a neighboring gene. BMMC cell type annotations are based on RNAPII gene coverage in clusters based on this modality. Differential genes analysis of the H3K27me3 mark was performed for the naïve B, erythroid, and monocyte clusters separately versus the HSC cluster using EdgeR [174] with donors as replicates. Differential genes were defined as having an FDR <0.05 and RPKM <0.5. For trend analysis, differential genes that gained the H3K27me3 mark were used above a log1p max signal threshold of 0.15. Pseudotime and lineages were computed using Palantir [175], and gene trends were smoothed using Kompot.smooth_expression [176]. Trend maxima were defined using np.max() and significance testing versus the null of no difference was performed via gene-level bootstrapping.

Chapter 5. Future Directions

This thesis establishes that the Polycomb repressive system functions by slow, step-wise H3K27 methylation that can accumulate on inactive chromatin. This suggests that H3K27me3 levels might increase with time and be limited by fast cell cycles. Upon dilution of the H3K27me3 mark, genes in Polycomb domains activate at different rates [55]. This warrants study of gene activation dynamics in rapidly cycling cell state transitions, like in myeloid development and lymphocyte activation [177,178]. Furthermore, the EZH2 methyltransferase subunit of PRC2 is not expressed in many post-mitotic cells, where it is replaced by the less-catalytically active EZH1 subunit [179]. It remains unknown how this affects Polycomb domain maintenance, and whether this ultimately leaves domains susceptible to degradation in aging [180]. Polycomb domain dynamics appear to be temporally linked to the cell cycle and studying them in this context will further reveal their role metazoan genome regulation.

Acknowledgements

I thank members of the Henikoff, Setty, Furlan, Reh and Stergachis labs for their support, collaboration, and helpful feedback during my graduate training. I thank Olivia Waltner for being my best friend and research companion during grad school. I thank my parents, sister, pugdog and grandparents for their constant support, patience, and love. Lastly, I thank my thesis committee for their faith and guidance.

References

1. Lewis, P. Pc: Polycomb. *Drosophila Information Service* **21**, 69 (1949).
2. Lewis, E.B. A gene complex controlling segmentation in Drosophila. *Nature* **276**, 565–570 (1978).
3. Struhl, G. A gene product required for correct initiation of segmental determination in Drosophila. *Nature* **293**, 36–41 (1981).
4. Jürgens, G. A group of genes controlling the spatial expression of the bithorax complex in Drosophila. *Nature* **316**, 153–155 (1985).
5. Struhl, G. & Akam, M. Altered distributions of Ultrabithorax transcripts in extra sex combs mutant embryos of Drosophila. *The EMBO Journal* **4**, 3259–3264 (1985).
6. Laible, G. *et al.* Mammalian homologues of the Polycomb-group gene Enhancer of zeste mediate gene silencing in Drosophila heterochromatin and at S. *cerevisiae telomeres. The EMBO Journal* **16**, 3219–3232 (1997).
7. Birve, A. *et al.* Su(z)12, a novel Drosophila Polycomb group gene that is conserved in vertebrates and plants. *Development* **128**, 3371–9 (2001).
8. Paro, R. & Hogness, D.S. The Polycomb protein shares a homologous domain with a heterochromatin-associated protein of Drosophila. *Proc Natl Acad Sci U S A* **88**, 263–7 (1991).
9. 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 & Development* **16**, 2893–905 (2002).
10. Müller, J. *et al.* Histone methyltransferase activity of a Drosophila Polycomb group repressor complex. *Cell* **111**, 197–208 (2002).
11. Czermin, B. *et al.* Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. *Cell* **111**, 185–96 (2002).
12. Cao, R. *et al.* Role of Histone H3 Lysine 27 Methylation in Polycomb-Group Silencing. *Science* **298**, 1039–1043 (2002).
13. Levine, S.S. *et al.* The Core of the Polycomb Repressive Complex Is Compositionally and Functionally Conserved in Flies and Humans. *Molecular and Cellular Biology* **22**, 6070–6078 (2002).
14. Rea, S. *et al.* Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **406**, 593–599 (2000).

15. Bracken, A.P., Dietrich, N., Pasini, D., Hansen, K.H. & Helin, K. Genome-wide mapping of Polycomb target genes unravels their roles in cell fate transitions. *Genes Dev* **20**, 1123–36 (2006).

16. Bernstein, E. *et al.* Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* **26**, 2560-9 (2006).

17. Holliday, R. & Pugh, J.E. DNA Modification Mechanisms and Gene Activity During Development. *Science* **187**, 226–232 (1975).

18. Riggs, A.D. X inactivation, differentiation, and DNA methylation. *Cytogenet Cell Genet* **14**, 9–25 (1975).

19. Vermaak, D., Ahmad, K. & Henikoff, S. Maintenance of chromatin states: an open-and-shut case. *Current Opinion in Cell Biology* **15**, 266–274 (2003).

20. Yamasu, K. & Senshu, T. Conservative Segregation of Tetrameric Units of H3 and H4 Histones during Nucleosome Replication. *The Journal of Biochemistry* **107**, 15–20 (1990).

21. Annunziato, A.T. Split Decision: What Happens to Nucleosomes during DNA Replication? Journal of Biological Chemistry 280, 12065–12068. (2005).

22. Dodd, I.B., Micheelsen, M.A., Sneppen, K. & Thon, G. Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification. *Cell* **129**, 813–822 (2007).

23. Grunstein, M. Yeast Heterochromatin: Regulation of Its Assembly and Inheritance by Histones. *Cell* **93**, 325–328 (1998).

24. Hansen, K.H. *et al.* A model for transmission of the H3K27me3 epigenetic mark. *Nature Cell Biology* **10**, 1291–1300 (2008).

25. Margueron, R. *et al.* Role of the polycomb protein EED in the propagation of repressive histone marks. *Nature* **461**, 762-767 (2009).

26. Sneeringer, C.J. *et al.* Coordinated activities of wild-type plus mutant EZH2 drive tumor-associated hypertrimethylation of lysine 27 on histone H3 in human B-cell lymphomas. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 20980-20985 (2010).

27. Pasini, D., Bracken, A.P., Jensen, M.R., Lazzerini Denchi, E. & Helin, K. Suz12 is essential for mouse development and for EZH2 histone methyltransferase activity. *EMBO Journal* **23**, 4061–4071 (2004).

28. Cao, R. & Zhang, Y. SUZ12 is required for both the histone methyltransferase activity and the silencing function of the EED-EZH2 complex. *Molecular Cell* **15**, 57-67 (2004).

29. Jiao, L. & Liu, X. Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science 350, aac4383* (2015).

30. Xu, C. *et al.* Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc Natl Acad Sci U S A* **107**, 19266–71 (2010).

31. Justin, N. *et al.* Structural basis of oncogenic histone H3K27M inhibition of human polycomb repressive complex 2. *Nature Communications* **7**, 11316 (2016).

32. McCabe, M.T. *et al.* Mutation of A677 in histone methyltransferase EZH2 in human B-cell lymphoma promotes hypertrimethylation of histone H3 on lysine 27 (H3K27). *Proc. Natl. Acad. Sci. U.S.A.* **109**, 2989–2994 (2012).

33. Swalm, B.M. *et al.* Reaction Coupling between Wild-Type and Disease-Associated Mutant EZH2. *ACS Chemical Biology* **9**, 2459-2464 (2014).

34. Schnee, P., Pleiss, J. & Jeltsch, A. Approaching the catalytic mechanism of protein lysine methyltransferases by biochemical and simulation techniques. *Critical Reviews in Biochemistry and Molecular Biology* **59**, 20 - 68 (2024).

35. Oksuz, O. *et al.* Capturing the Onset of PRC2-Mediated Repressive Domain Formation. *Mol Cell* **70**, 1149-1162.e5 (2018).

36. Brown, J.L., Mucci, D., Whiteley, M., Dirksen, M.L. & Kassis, J.A. The Drosophila Polycomb group gene pleiohomeotic encodes a DNA binding protein with homology to the transcription factor YY1. *Mol. Cell* **1**, 1057–1064 (1998).

37. Wang, L. *et al.* Hierarchical recruitment of polycomb group silencing complexes. *Mol. Cell* **14**, 637–646 (2004).

38. Laugesen, A., Højfeldt, J.W. & Helin, K. Molecular mechanisms directing PRC2 recruitment and H3K27 methylation. *Mol. Cell* **74**, 8–18 (2019).

39. Blackledge, N.P. & Klose, R.J. The molecular principles of gene regulation by Polycomb repressive complexes. *Nat Rev Mol Cell Biol* **22**, 815-833 (2021).

40. Kasinath, V. *et al.* Structures of human PRC2 with its cofactors AEBP2 and JARID2. *Science* **359**, 940–944 (2018).

41. Lee, C.H. *et al.* Distinct stimulatory mechanisms regulate the catalytic activity of Polycomb repressive complex 2. *Molecular Cell* **70**, 435–448 (2018).

42. Sanulli, S. *et al.* Jarid2 methylation via the PRC2 complex regulates H3K27me3 deposition during cell differentiation. *Mol. Cell* **57**, 769–783 (2015).

43. Li, H. *et al.* Polycomb-like proteins link the PRC2 complex to CpG islands. *Nature* **549**, 287 - 291 (2017).

44. Cooper, S. *et al.* Jarid2 binds mono-ubiquitylated H2A lysine 119 to mediate crosstalk between Polycomb complexes PRC1 and PRC2. *Nature Communications* **7**, 13661 (2016).

45. Zhang, Q. *et al.* PALI1 facilitates DNA and nucleosome binding by PRC2 and triggers an allosteric activation of catalysis. *Nature Communications* **12**, 4592 (2021).

46. Poepsel, S., Kasinath, V. & Nogales, E. Cryo-EM structures of PRC2 simultaneously engaged with two functionally distinct nucleosomes. *Nature Structural & Molecular Biology* **25**, 154–162 (2018).

47. Kasinath, V. *et al.* JARID2 and AEBP2 regulate PRC2 in the presence of H2AK119ub1 and other histone modifications. *Science* **371**, eabc3393 (2021).

48. Cheng, L. *et al.* Chromatin Assembly Factor 1 (CAF-1) facilitates the establishment of facultative heterochromatin during pluripotency exit. *Nucleic Acids Res.* **47**, 11114–11131 (2019).

49. Whitfield, M.L. *et al.* Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13**, 1977-2000 (2002).

50. Reverón-Gómez, N. *et al.* Accurate recycling of parental histones reproduces the histone modification landscape during DNA replication. *Mol. Cell* **72**, 239-49.e5 (2018).

51. Wenger, A. *et al.* Symmetric inheritance of parental histones governs epigenome maintenance and embryonic stem cell identity. *Nat. Genet.* **55**, 1567-78 (2023).

52. Flury, V. *et al.* Recycling of modified H2A-H2B provides short-term memory of chromatin states. *Cell* **186**, 1050-65 (2023).

53. Liu, C. *et al.* Histone H1 facilitates restoration of H3K27me3 during DNA replication by chromatin compaction. *Nat. Commun.* **14**, 4081 (2023).

54. Alabert, C. *et al.* Two distinct modes for propagation of histone PTMs across the cell cycle. *Genes and Development* **29**, 585-590 (2015).

55. Jadhav, U. *et al.* Replicational dilution of H3K27me3 in mammalian cells and the role of poised promoters. *Mol. Cell* **78**, 141-51 (2020).

56. Jadhav, U. *et al.* Acquired Tissue-Specific Promoter Bivalency Is a Basis for PRC2 Necessity in Adult Cells. *Cell* **165**, 1389-1400 (2016).

57. Trouth, A. *et al.* The length of the G1 phase is an essential determinant of H3K27me3 landscapes across diverse cell types. *PLoS Biology* **23**, e3003119 (2025).

58. Asenjo, H.G. *et al.* Polycomb regulation is coupled to cell cycle transition in pluripotent stem cells. *Sci. Adv.* **6**, eaay4768 (2020).

59. Bracken, A.P. *et al.* The Polycomb group proteins bind throughout the INK4A-ARF locus and are disassociated in senescent cells. *Genes & Development* **21**, 525–530 (2007).

60. Longhurst, A.D. *et al.* The PRC2.1 subcomplex opposes G1 progression through regulation of CCND1 and CCND2. *eLife* **13**, RP97577 (2025).

61. Laprell, F., Finkl, K. & Müller, J. Propagation of Polycomb-repressed chromatin requires sequence-specific recruitment to DNA. *Science* **356**, 85–88 (2017).

62. Coleman, R.T. & Struhl, G. Causal role for inheritance of H3K27me3 in maintaining the OFF state of a Drosophila HOX gene. *Science* **356**, eaai8236 (2017).

63. Klymenko, T. *et al.* A Polycomb group protein complex with sequence-specific DNA-binding and selective methyl-lysine-binding activities. *Genes Dev* **20**, 1110-22 (2006).

64. Veronezi, G.M.B. & Ramachandran, S. Nucleation and spreading maintain Polycomb domains every cell cycle. *Cell Rep.* **43**, 114090 (2024).

65. Izeddin, I. *et al.* Single-molecule tracking in live cells reveals distinct target-search strategies of transcription factors in the nucleus. *eLife* **3**, e02230 (2014).

66. Deal, R.B., Henikoff, J.G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labeling of histones. *Science* **328**, 1161–1164 (2010).

67. Wooten, M., Nguyen, K., Takushi, B.N., Ahmad, K. & Henikoff, S. Nascent CUT&Tag captures transcription factor binding after chromatin duplication. *bioRxiv* (2025).

68. Allshire, R.C. & Madhani, H.D. Ten principles of heterochromatin formation and function. *Nature Reviews Molecular Cell Biology* **19**, 229–244 (2018).

69. Agius, S.C. *et al.* Accessory subunits of PRC2 mimic H3K27me3 to restrict the spread of Polycomb domains. *Molecular Cell* **86**, 1032-1045.e10 (2026).

70. Escobar, T.M. *et al.* Active and repressed chromatin domains exhibit distinct nucleosome segregation during DNA replication. *Cell* **179**, 953–963.e11 (2019).

71. Choi, J. *et al.* DNA binding by PHF1 prolongs PRC2 residence time on chromatin and thereby promotes H3K27 methylation. *Nature Structural & Molecular Biology* **24**, 1039–1047 (2017).

72. Yuan, W. *et al.* Dense chromatin activates Polycomb repressive complex 2 to regulate H3 lysine 27 methylation. *Science* **337**, 971–975 (2012).

73. Talbert, P.B. & Henikoff, S. Spreading of silent chromatin: inaction at a distance. *Nature Reviews Genetics* **7**, 793–803 (2006).

74. Berry, S., Dean, C. & Howard, M. Slow chromatin dynamics allow Polycomb target genes to filter fluctuations in transcription factor activity. *Cell Systems* **4**, 445–457.e8 (2017).

75. Zee, B.M., Britton, L.M., Wolle, D., Haberman, D.M. & Garcia, B.A. Origins and formation of histone methylation across the human cell cycle. *Mol. Cell. Biol.* **32**, 2503-14 (2012).

76. Marasca, F., Bodega, B. & Orlando, V. How Polycomb-Mediated Cell Memory Deals With a Changing Environment. *BioEssays* **40**, 1700137 (2018).

77. Zenk, F. *et al.* Germ line-inherited H3K27me3 restricts enhancer function during maternal-to-zygotic transition. *Science* **357**, 212-216 (2017).

78. Degen, E.A., Gonzaga-Saavedra, N. & Blythe, S.A. A maternal Polycomb imprint is maintained in early Drosophila development through lower-order methylation states. *Cell Rep* **45**, 117679 (2026).

79. Gonzaga-Saavedra, N., Degen, E.A., Soluri, I.V., Croslyn, C. & Blythe, S.A. Nucleation-dependent propagation of Polycomb modifications emerges during the Drosophila maternal to zygotic transition. *bioRxiv* (2025).

80. Lukauskas, S. *et al.* Decoding chromatin states by proteomic profiling of nucleosome readers. *Nature* **627**, 671-679 (2024).

81. Ferrari, Karin J. *et al.* Polycomb-Dependent H3K27me1 and H3K27me2 Regulate Active Transcription and Enhancer Fidelity. *Molecular Cell* **53**, 49-62 (2014).

82. Loyola, A., Bonaldi, T., Roche, D., Imhof, A. & Almouzni, G. PTMs on H3 variants before chromatin assembly potentiate their final epigenetic state. *Mol Cell* **24**, 309-16 (2006).

83. Lee, H.G., Kahn, T.G., Simcox, A., Schwartz, Y.B. & Pirrotta, V. Genome-wide activities of Polycomb complexes control pervasive transcription. *Genome Res* **25**, 1170-81 (2015).

84. Lavarone, E., Barbieri, C.M. & Pasini, D. Dissecting the role of H3K27 acetylation and methylation in PRC2 mediated control of cellular identity. *Nature Communications* **10**, 1679 (2019).

85. Wu, S.J. *et al.* Single-cell CUT&Tag analysis of chromatin modifications in differentiation and tumor progression. *Nature Biotechnology* **39**, 819-824 (2021).

86. Janssens, D.H. *et al.* Scalable single-cell profiling of chromatin modifications with sciCUT&Tag. *Nature Protocols* **19**, 83-112 (2024).

87. Kaya-Okur, H.S. *et al.* CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nature Communications* **10**, 1930 (2019).

88. Greene, J.E., Ahmad, K. & Henikoff, S. Histone H3K27 methylation states are sequentially catalyzed in cycling cells. *bioRxiv*, 2026.04.30.721988 (2026).

89. Riising, Eva M. *et al.* Gene Silencing Triggers Polycomb Repressive Complex 2 Recruitment to CpG Islands Genome Wide. *Molecular Cell* **55**, 347-360 (2014).

90. Jermann, P., Hoerner, L., Burger, L. & Schübeler, D. Short sequences can efficiently recruit histone H3 lysine 27 trimethylation in the absence of enhancer activity and DNA methylation. *Proceedings of the National Academy of Sciences* **111**, E3415-E3421 (2014).

91. Kraushaar, D.C. *et al.* Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biology* **14**, R121 (2013).

92. Jiang, D. & Berger, F. DNA replication–coupled histone modification maintains Polycomb gene silencing in plants. *Science* **357**, 1146-1149 (2017).

93. Beck, D.B. *et al.* Chromatin in the nuclear landscape. *Cold Spring Harb Symp Quant Biol* **75**, 11-22 (2010).

94. Trouth, A. *et al.* G1 length dictates H3K27me3 landscapes. *bioRxiv* (2024).

95. Shah, R.N. *et al.* Examining the Roles of H3K4 Methylation States with Systematically Characterized Antibodies. *Molecular Cell* **72**, 162-177.e7 (2018).

96. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330 (2015).

97. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nature Methods* **9**, 215-216 (2012).

98. Steiner, L.A., Schulz, V.P., Maksimova, Y., Wong, C. & Gallagher, P.G. Patterns of histone H3 lysine 27 monomethylation and erythroid cell type-specific gene expression. *J Biol Chem* **286**, 39457-65 (2011).

99. Klein, K.N. *et al.* Replication timing maintains the global epigenetic state in human cells. *Science* **372**, 371-378 (2021).

100. Hansen, R.S. *et al.* Sequencing newly replicated DNA reveals widespread plasticity in human replication timing. *Proc Natl Acad Sci U S A* **107**, 139-44 (2010).

101. Filion, G.J. *et al.* Systematic Protein Location Mapping Reveals Five Principal Chromatin Types in Drosophila Cells. *Cell* **143**, 212-224 (2010).

102. Karagiannis, T.C. *et al.* Characterization of K562 cells: uncovering novel chromosomes, assessing transferrin receptor expression, and probing pharmacological therapies. *Cell Mol Life Sci* **80**, 248 (2023).

103. Italiano, A. *et al.* Tazemetostat, an EZH2 inhibitor, in relapsed or refractory B-cell non-Hodgkin lymphoma and advanced solid tumours: a first-in-human, open-label, phase 1 study. *Lancet Oncol* **19**, 649-659 (2018).

104. Qi, W. *et al.* An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. *Nat Chem Biol* **13**, 381-388 (2017).

105. Core, L.J. *et al.* Analysis of nascent RNA identifies a unified architecture of initiation regions at mammalian promoters and enhancers. *Nature Genetics* **46**, 1311-1320 (2014).

106. van Arensbergen, J. *et al.* Genome-wide mapping of autonomous promoter activity in human cells. *Nature Biotechnology* **35**, 145-153 (2017).

107. Leemans, C. *et al.* Promoter-Intrinsic and Local Chromatin Features Determine Gene Repression in LADs. *Cell* **177**, 852-864.e14 (2019).

108. Akhtar, W. *et al.* Chromatin Position Effects Assayed by Thousands of Reporters Integrated in Parallel. *Cell* **154**, 914-927 (2013).

109. Sauer, P.V. *et al.* Activation of automethylated PRC2 by dimerization on chromatin. *Molecular Cell* **84**, 3885-3898.e8 (2024).

110. Chory, E.J. *et al.* Nucleosome Turnover Regulates Histone Methylation Patterns over the Genome. *Mol Cell* **73**, 61-72.e3 (2019).

111. Agger, K. *et al.* UTX and JMJD3 are histone H3K27 demethylases involved in HOX gene regulation and development. *Nature* **449**, 731-734 (2007).

112. Smith, E.R. *et al.* Drosophila UTX is a histone H3 Lys27 demethylase that colocalizes with the elongating form of RNA polymerase II. *Mol Cell Biol* **28**, 1041-6 (2008).

113. Shermoen, A.W., McCleland, M.L. & O'Farrell, P.H. Developmental control of late replication and S phase length. *Curr Biol* **20**, 2067-77 (2010).

114. Limas, J.C. & Cook, J.G. Preparation for DNA replication: the key to a successful S phase. *FEBS Lett* **593**, 2853-2867 (2019).

115. Stewart-Morgan, K.R., Petryk, N. & Groth, A. Chromatin replication and epigenetic cell memory. *Nature Cell Biology* **22**, 361-371 (2020).

116. Yagi, M. *et al.* Bivalent chromatin instructs lineage specification during hematopoiesis. *Cell* **188**, 4314-4331.e29 (2025).

117. Juan, Aster H. *et al.* Roles of H3K27me2 and H3K27me3 Examined during Fate Specification of Embryonic Stem Cells. *Cell Reports* **17**, 1369-1382 (2016).

118. Matsuwaka, M., Kumon, M. & Inoue, A. H3K27 dimethylation dynamics reveal stepwise establishment of facultative heterochromatin in early mouse embryos. *Nature Cell Biology* **27**, 28-38 (2025).

119. Hickey, G.J.M. *et al.* Establishment of developmental gene silencing by ordered polycomb complex recruitment in early zebrafish embryos. *eLife* **11**, e67738 (2022).

120. Kalb, R. *et al.* Histone H2A monoubiquitination promotes histone H3 methylation in Polycomb repression. *Nature Structural & Molecular Biology* **21**, 569-571 (2014).

121. Lu-Culligan, W.J. *et al.* Acetyl-methyllysine marks chromatin at active transcription start sites. *Nature* **622**, 173-179 (2023).

122. Yamagishi, M. *et al.* Targeting Excessive EZH1 and EZH2 Activities for Abnormal Histone Methylation and Transcription Network in Malignant Lymphomas. *Cell Reports* **29**, 2321-2337.e7 (2019).

123. Gambacorta, V. *et al.* Integrated Multiomic Profiling Identifies the Epigenetic Regulator PRC2 as a Therapeutic Target to Counteract Leukemia Immune Escape and Relapse. *Cancer Discov* **12**, 1449-1461 (2022).

124. Xie, Z. *et al.* Gene Set Knowledge Discovery with Enrichr. *Current Protocols* **1**, e90 (2021).

125. Diaz, A., Nellore, A. & Song, J.S. CHANCE: comprehensive software for quality control and validation of ChIP-seq data. *Genome Biology* **13**, R98 (2012).

126. Szklarczyk, D. *et al.* The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res* **51**, D638-d646 (2023).

127. Henikoff, S. CUT&Tag-direct for whole cells with CUTAC V.5 (2024).

128. Martin, M. CUTADAPT removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**(2011).

129. Langmead, B. & Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357-9 (2012).

130. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841-2 (2010).

131. Virtanen, P. *et al.* SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nature Methods* **17**, 261-272 (2020).

132. Talbert, P.B. & Henikoff, S. The Yin and Yang of Histone Marks in Transcription. *Annual Review of Genomics and Human Genetics* **22**, 147-170 (2021).

133. Blanco, M.A. *et al.* Chromatin-state barriers enforce an irreversible mammalian cell fate decision. *Cell Reports* **37**, 109967 (2021).

134. Bartosovic, M., Kabbe, M. & Castelo-Branco, G. Single-cell CUT&Tag profiles histone modifications and transcription factors in complex tissues. *Nature Biotechnology* **39**, 825-835 (2021).

135. Zhu, C. *et al.* Joint profiling of histone modifications and transcriptome in single cells from mouse brain. *Nature Methods* **18**, 283-292 (2021).

136. Zhang, B. *et al.* Characterizing cellular heterogeneity in chromatin state with scCUT&Tag-pro. *Nature Biotechnology* **40**, 1220-1230 (2022).

137. 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. *Nature Biotechnology* **41**, 708-716 (2023).

138. Cao, J. *et al.* Comprehensive single-cell transcriptional profiling of a multicellular organism. *Science* **357**, 661-667 (2017).

139. Cusanovich, D.A. *et al.* Multiplex single-cell profiling of chromatin accessibility by combinatorial cellular indexing. *Science* **348**, 910-914 (2015).

140. Bartlett, D.A. *et al.* High-throughput single-cell epigenomic profiling by targeted insertion of promoters (TIP-seq). *Journal of Cell Biology* **220**, e202103078 (2021).

141. Wang, Q. *et al.* CoBATCH for High-Throughput Single-Cell Epigenomic Profiling. *Molecular Cell* **76**, 206-216.e7 (2019).

142. Martin, B.K. *et al.* Optimized single-nucleus transcriptional profiling by combinatorial indexing. *Nature Protocols* (2022).

143. Del Priore, I. *et al.* Protocol for single-cell ATAC sequencing using combinatorial indexing in mouse lung adenocarcinoma. *STAR Protocols* **2**, 100583 (2021).

144. Mezger, A. *et al.* High-throughput chromatin accessibility profiling at single-cell resolution. *Nature Communications* **9**, 3647 (2018).

145. Kaya-Okur, H.S., Janssens, D.H., Henikoff, J.G., Ahmad, K. & Henikoff, S. Efficient low-cost chromatin profiling with CUT&Tag. *Nature Protocols* **15**, 3264-3283 (2020).

146. Ren, X. *et al.* COVID-19 immune features revealed by a large-scale single-cell transcriptome atlas. *Cell* **184**, 1895-1913 e19 (2021).

147. Granja, J.M. *et al.* Single-cell multiomic analysis identifies regulatory programs in mixed-phenotype acute leukemia. *Nat Biotechnol* **37**, 1458-1465 (2019).

148. Chen, C. *et al.* Single-cell multiomics reveals increased plasticity, resistant populations, and stem-cell-like blasts in KMT2A-rearranged leukemia. *Blood* **139**, 2198-2211 (2022).

149. Stuart, T. *et al.* Nanobody-tethered transposition enables multifactorial chromatin profiling at single-cell resolution. *Nature Biotechnology* (2022).

150. Fujiwara, Y. & Okada, Y. CUT&Tag Using "Stress-Free" Con A-Conjugated Dynabeads((R)). *Methods Mol Biol* **2519**, 141-153 (2023).

151. Heaton, H. *et al.* Souporcell: robust clustering of single-cell RNA-seq data by genotype without reference genotypes. *Nature Methods* **17**, 615-620 (2020).

152. Ewels, P.A. *et al.* The nf-core framework for community-curated bioinformatics pipelines. *Nature Biotechnology* **38**, 276-278 (2020).

153. Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nat Biotechnol* **35**, 316-319 (2017).

154. Granja, J.M. *et al.* ArchR is a scalable software package for integrative single-cell chromatin accessibility analysis. *Nature Genetics* **53**, 403-411 (2021).

155. Leland McInnes and John Healy and James Melville}. UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. *arXiv* (2018).

156. Buenrostro, J.D. *et al.* Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature* **523**, 486-90 (2015).

157. Janssens, D.H. *et al.* Automated CUT&Tag profiling of chromatin heterogeneity in mixed-lineage leukemia. *Nature Genetics* **53**, 1586-1596 (2021).

158. Schep, A.N., Wu, B., Buenrostro, J.D. & Greenleaf, W.J. chromVAR: inferring transcription-factor-associated accessibility from single-cell epigenomic data. *Nat Methods* **14**, 975-978 (2017).

159. Meers, M.P., Tenenbaum, D. & Henikoff, S. Peak calling by Sparse Enrichment Analysis for CUT&RUN chromatin profiling. *Epigenetics Chromatin* **12**, 42 (2019).

160. Dou, Y. *et al.* Regulation of MLL1 H3K4 methyltransferase activity by its core components. *Nature Structural & Molecular Biology* **13**, 713-719 (2006).

161. Robinson, J.T. *et al.* Integrative genomics viewer. *Nat Biotechnol* **29**, 24-6 (2011).

162. Kent, W.J. *et al.* The human genome browser at UCSC. *Genome Res* **12**, 996-1006 (2002).

163. Lebert-Ghali, C.-É. *et al.* Hoxa cluster genes determine the proliferative activity of adult mouse hematopoietic stem and progenitor cells. *Blood* **127**, 87-90 (2016).

164. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nature Biotechnology* **36**, 411-420 (2018).

165. Simon, J.A. & Kingston, R.E. Occupying chromatin: Polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. *Mol Cell* **49**, 808-24 (2013).

166. Wysocka, J. *et al.* A PHD finger of NURF couples histone H3 lysine 4 trimethylation with chromatin remodelling. *Nature* **442**, 86-90 (2006).

167. Cruz, C. *et al.* Tri-methylation of histone H3 lysine 4 facilitates gene expression in ageing cells. *eLife* **7**, e34081 (2018).

168. Janssens, D.H. *et al.* Sequential RNA Polymerase II Activation Drives Human Hematopoiesis. *Cell reports* **45**, 116802-116802 (2026).

169. Peng, T. *et al.* Multi-omics integration analysis identifies INPP4B as a T-cell-specific activation suppressor. *Clin Transl Med* **15**, e70430 (2025).

170. Dalby, A. *et al.* Transcription Factor Levels after Forward Programming of Human Pluripotent Stem Cells with GATA1, FLI1, and TAL1 Determine Megakaryocyte versus Erythroid Cell Fate Decision. *Stem Cell Reports* **11**, 1462-1478 (2018).

171. Rodrigues, N.P., Tipping, A.J., Wang, Z. & Enver, T. GATA-2 mediated regulation of normal hematopoietic stem/progenitor cell function, myelodysplasia and myeloid leukemia. *Int J Biochem Cell Biol* **44**, 457-60 (2012).

172. Margueron, R. *et al.* Ezh1 and Ezh2 maintain repressive chromatin through different mechanisms. *Molecular Cell* **32**, 503-518 (2008).

173. Xu, J. *et al.* Developmental control of polycomb subunit composition by GATA factors mediates a switch to non-canonical functions. *Mol Cell* **57**, 304-316 (2015).

174. Robinson, M.D., McCarthy, D.J. & Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139-40 (2010).

175. Setty, M. *et al.* Characterization of cell fate probabilities in single-cell data with Palantir. *Nature Biotechnology* **37**, 451-460 (2019).

176. Otto, D.J. *et al.* Comparing phenotypic manifolds with Kompot: Detecting differential abundance and gene expression at single-cell resolution. *bioRxiv* (2025).

177. Dalton, S. Cell cycle dynamics in the reprogramming of cellular identity. *FEBS Letters* (2019).

178. Laurenti, E. Linking cell cycle to hematopoietic stem cell fate decisions. *Frontiers in Cell and Developmental Biology* (2023).

179. McCole, R. *et al.* A conserved switch to less catalytically active Polycomb repressive complexes in non-dividing cells. *Cell reports* **44**, 115192-115192 (2025).

180. Leichter, S.M., Ahmad, K. & Henikoff, S. Polycomb misregulation in enterocytes drives tissue decline in the aging Drosophila intestine. *Genome Research* **36**, 102-114 (2026).