bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Dynamics of transcription-dependent H3K36me3 marking by the SETD2:IWS1:SPT6 ternary complex

Katerina Cermakova[1], Eric A. Smith[1], Vaclav Veverka[2], H. Courtney Hodges[1,3,4,*]

[1] Department of Molecular & Cellular Biology, Center for Precision Environmental Health, and Dan L Duncan Comprehensive Cancer Center, Baylor College of Medicine, Houston, TX, 77030, USA
[2] Institute of Organic Chemistry and Biochemistry, Czech Academy of Sciences, Prague, Czech Republic
[3] Center for Cancer Epigenetics, The University of Texas MD Anderson Cancer Center, Houston, TX, 77030, USA
[4] Department of Bioengineering, Rice University, Houston, TX, 77005, USA
* Lead contact; Correspondence to: *chodges@bcm.edu*

Abstract
SETD2 contributes to gene expression by marking gene bodies with H3K36me3, which is thought to assist in the concentration of transcription machinery at the small portion of the coding genome. Despite extensive genome-wide data revealing the precise localization of H3K36me3 over gene bodies, the physical basis for the accumulation, maintenance, and sharp borders of H3K36me3 over these sites remains rudimentary. Here we propose a model of H3K36me3 marking based on stochastic transcription-dependent placement and transcription-independent spreading. Our analysis of the spatial distributions and dynamic features of these marks indicates that transcription-dependent placement dominates the establishment of H3K36me3 domains compared to transcription-independent spreading processes, and that turnover of H3K36me3 limits its capacity for epigenetic memory. By adding additional terms for asymmetric histone turnover occurring at transcription start sites, our model provides a remarkably accurate representation of H3K36me3 levels and dynamics over gene bodies. Furthermore, we validate our findings by revealing that loss of SPT6 impairs the transcription-coupled activity of the SETD2:IWS1:SPT6 ternary complex, thereby reducing the tight correlation between transcription and H3K36me3 levels at gene bodies.

Keywords: chromatin, epigenetics, methylation, CTD, RNAP2

Introduction

Eukaryotic cells use an ensemble of covalent histone modifications to mark important sites of regulatory activity. One of these marks, trimethylation of lysine 36 on histone 3 (H3K36me3) is an abundant and highly conserved chromatin modification that is enriched at gene bodies of transcriptionally active genes and also at centromeric regions[1–3]. The presence and distribution of H3K36me3 domains at actively transcribed genes is conserved from human to yeast, suggesting that this mark is extremely important for proper cellular function[4]. Important regulatory activities linked to the H3K36me3 mark include transcription elongation[5,6], prevention of cryptic start sites[7], as well as pre-mRNA splicing[8] and processing[9]. In addition to transcription, H3K36me3 also plays important roles in the recruitment of DNA repair machinery to mismatch regions, and for this reason, actively transcribed genes pre-marked with H3K36me3 are especially protected from DNA damage[10].

The genome-wide distribution of H3K36me3 is maintained through various mechanisms. In human cells, H3K36 is mono- and di-methylated by eight distinct histone methyltransferases; however, the predominant writer of the trimethyl mark on H3K36 is SETD2[1,11,12]. Interestingly, SETD2 is a major tumor suppressor in clear cell renal cell carcinoma[13], breast cancer[14], bladder cancer[15], and acute lymphoblastic leukemias[16–18]. In these settings, mutations and deletions of SETD2 often leads to global downregulation of H3K36me3 and are strongly associated with decreased overall survival[13–19]. H3K36me3 is enzymatically removed from chromatin by members of Jumonji domain-containing histone demethylase protein families[20]. However, the mark may be also removed by other mechanisms, including histone turnover. Histone turnover is performed by diverse ATP-dependent chromatin remodeling complexes, such as BAF (SWI/SNF) and related complexes[21,22], as well as histone chaperones (e.g. FACT complex[23] or SPT6[24]), that catalyze histone eviction from chromatin. Interestingly, histone turnover is highest at active promoters and less pronounced within transcribed regions enriched for H3K36me3[25,26]. In addition to the writers and erasers of this mark, H3K36me3 is also read by a range of the proteins harboring the PWWP domain as well as Tudor, chromo and plant homeodomain (PHD) finger domains[27], many of which are known transcription regulators or DNA damage responders (e.g. PHF1[28], Dnmt3a[29] and LEDGF/p75[30,31]).

Transcription-coupled H3K36me3 deposition at coding regions is facilitated through the unstructured yet highly conserved C-terminal domain of RNA polymerase 2 (RNAP2 CTD). During transcription elongation, the RNAP2 CTD recruits many factors to transcribed genes, including the histone methyltransferases MLL (KMT2A; H3K4 methylation), DOT1L (H3K79me2), and SETD2 (H3K36me3) (**Figure 1A**)[32–34]. These methyltransferases are recruited to sites of transcription in part via direct interaction with RNAP2 [6,35–38]. However, this common mode of recruitment curiously does not result in similar patterns of histone modification, as each of these marks has a distinct, characteristic profile across the genome[39–42]. For the H3K36me3 mark, serine-2 phosphorylation on the CTD of the POLR2A (Rpb1) subunit of RNAP2 interacts with the transcription elongation factor and histone chaperone SPT6, which associates with SETD2 through its binding partner IWS1[6,43,44]. Through this association (**Figure 1B**), H3K36me3 is thought to be propagated along gene bodies in a transcription-dependent manner[35]. In yeast,

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

transcription is sufficient for formation of H3K36me3 domain over gene bodies[45]. Here the presence of SPT6 and IWS1 is not only essential for SETD2 recruitment, but also for the catalysis of H3K36me3 placement[6]. However, in mammals, SETD2 associates with a large number of nuclear proteins and complexes that possess their own chromatin recognition domains (e.g. p53[46] or STAT1[47]), which could potentially influence SETD2 methyltransferase activity independently of transcription. Furthermore, it has been proposed that the placement of H3K36me3 involves "spreading" of the mark across the gene body[48–50], and a variety of differing patterns reported for H3K36me3 profiles across gene bodies[51–54]. It has therefore remained uncertain whether SETD2 can be recruited to chromatin to place and spread the H3K36me3 mark through transcription-independent mechanisms in addition to transcription-dependent interaction with the RNAP2 CTD. Given the importance of H3K36me3, we were surprised to find no systematic genome-wide studies describing the spatial distribution, intensity, and dynamics of this mark placed by SETD2 in concert with other factors.

Although great effort has been made to elucidate the participation of H3K36me3 in various cellular processes[1], much less is known about dynamics of the H3K36me3 mark itself in human cells. How are the activities of enzymes that regulate H3K36me3 writing and erasure coordinated to result in formation of these domains? Is the transcription-coupled H3K36me3 placement sufficient to achieve the steady-state of these domains, or are other transcription-independent processes needed to propagate and stabilize these marks? What are the insulation mechanisms that prevent this mark from spreading outside of its boundaries?

Because chromatin and its marks are highly dissipative complex systems[55,56], non-linear and agent-based kinetic models have become invaluable tools to understand the dynamical behavior of the chromatin landscape and to predict the behavior of epigenetic systems[57–62]. We previously developed a model of H3K9me3 dynamics premised on simple nucleation-propagation-turnover Markov processes that provided quantitative predictions of kinetic properties and dynamics of H3K9me3 domains throughout the genome[58]. This model allowed us to elucidate the constraints that lead to the characteristic pattern of enrichment and memory of H3K9me3 domains in embryonic stem cells and fibroblasts[59]. Here we have further developed this approach to model the establishment and dynamic properties of H3K36me3 domains at gene bodies throughout the genome. Our analysis shows that H3K36me3 domains are broad, flat domains with sharp borders at gene bodies despite the lack of discernable insulator features in multiple cell types. We find that the abundance of H3K36me3 at gene bodies scales in proportion to transcription frequency, consistent with transcription-dependent marking by SETD2. In contrast to the direct local spreading of H3K9me3 marks in heterochromatin, our analysis shows that transcription-independent H3K36me3 spreading must be small in proportion to transcription-dependent mechanisms. Furthermore, incorporating measured rates of histone turnover is sufficient to produce remarkably in-vivo-like mark distributions over gene bodies without insulator elements. Finally, we find that SPT6 loss sharply reduces the observed dependence of H3K36me3 levels on transcription frequency, consistent with our model of H3K36me3 placement occurring primarily through transcription-dependent marking. Together our data provide new insights into the placement, persistence, and function of H3K36me3 domains.

Results

The quantitative dependence of H3K36me3 marking on transcription frequency

We began by performing a comprehensive analysis of histone mark ChIP-seq signals at gene bodies based on the degree of transcription. We analyzed the ChIP-seq profiles of H3K4me1, H3K4me3, H3K27ac, H3K27me3, H3K36me3, and H3K79me2 at gene bodies in K562 chronic myeloid leukemia cells. Despite extensive studies of the relationship between transcription and histone modification domains, we found few systematic reports relating the quantitative accumulation of all such marks to the frequency of transcription. We therefore sought to assess the relationship between the levels of each of these marks with transcriptional frequency, using the fragments per kilobase million (FPKM) levels from RNA-seq in the same cells as a proxy. We found that all of these marks highly statistically significant ($p<2.2e-16$) correlation with transcription. All marks positively correlated with transcription except for H3K27me3, which displayed a negative correlation (**Figure 1C and S1**). Of those with positive correlation, all histone marks except for H3K36me3 displayed a strong characteristic peak localized at transcription start sites (TSSs), consistent with regulatory roles during transcription initiation. In contrast, we found that H3K36me3 domains are localized directly over transcriptional units, with sharp, well-defined borders near TSSs and transcription termination sites (TTSs, **Figure 1D**), consistent with many previous studies of H3K36me3[9,19,40]. By comparing the mean levels of H3K36me3 across gene bodies to their FPKM values, we found that H3K36me3 had the highest correlation ($R=0.543$, $p<2.2e-16$) between transcriptional activity and histone mark levels over several powers of 10 (**Figure 1E**). Altogether, the unusually high correlation and presence of the mark throughout the entire transcriptional unit indicated a strong link between H3K36me3 accumulation and transcription elongation.

We next sought to address whether the shape of the H3K36me3 domain varied with transcriptional activity by analyzing these marks at individual loci genome-wide. We restricted our analysis to transcripts with FPKM values >2, and with gene bodies that spanned the 15-kb window of our analysis. The resulting heat map of 9,079 genes (**Figure 1F**) revealed that although the intensity of H3K36me3 marking varies with transcriptional activity, the overall shape of the distribution is largely unaffected, yielding flat plateaus at gene bodies throughout the genome. This interpretation was confirmed by examination of browser tracks of individual genes (**Figure 1G**). Nearly identical genome-wide results were obtained in A549 cells, an unrelated non-small cell lung cancer line, indicating these features are shared across multiple cell types (**Figure S2**).

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

A transcription-dependent model of H3K36me3 dynamics

We next constructed a dynamic Markov state machine model to describe the dynamics of H3K36me3 marking and its relationship to transcription frequency. We constructed a one-dimensional lattice model of chromatin, analogous to the "beads on a string" picture of chromatin. Although higher-order chromatin structure may contribute to certain features of transcription, our relatively simple treatment of chromatin structure is justified, since well-ordered, evenly spaced arrays of nucleosomes are clearly visible at lowly and highly expressed genes in our analysis of MNase-seq data in K562 cells (**Figure S3**). In our simulations, the lattice $\{I_j\}$ spans from j = 1 to j = 512, where each lattice position *j* corresponds to an individual nucleosome. Thus with a nucleosomal spacing of 195 bp, our lattice model corresponds to 99.8 kbp of DNA. To make our model as general as possible, we consider that each nucleosome position along the lattice may occupy one of two states: an unmodified (I_j = 0) state or a modified with H3K36me3 (I_j = 1) state. Briefly, four features define the standard model (**Figure 2**):

(1) *Transcription-dependent marking*: Upon an individual transcription event occurring in an individual time step, nucleosome positions in a transcriptional unit (positions 174-338) are each randomly marked with probability P_n. This process corresponds to the recruitment and stochastic labeling of nucleosomes during transcription events. Transcription events occur within a given time step at probability P_i.

(2) *Transcription-independent spreading*: Our general kinetic scheme integrates any processes that locally spread H3K36me3 into a single net propagation probability (P_+). In our model, P_+ describes the transcription-independent spreading of H3K36me3 to nucleosome positions immediately adjacent to H3K36me3-marked sites at each time step. This scenario represents the hypothetical pathway whereby SETD2 is recruited to sites of H3K36me3, for example directly or indirectly through H3K36me3 reader(s). Mechanistically, this process is likely composed of multiple rate-limiting steps (e.g. recruitment, marking), however for simplicity we treat this spreading as occurring as a single rate-limiting step.

(3) *Background marking*: All nucleosomes across the lattice are subject to weak, low probability of marking at each time step. This probability is P_{bg}, and corresponds to any process, such as non-specific SETD2 recruitment, that lead to non-specific marking.

(4) *Random mark/histone turnover*: At each time step, all positions across the lattice are subject to stochastic removal of the H3K36me3 mark. All nucleosome marks are removed at each time step with probability P_-. This process corresponds to any mechanism that results in the loss of the mark, and includes both H3K36 demethylation and histone/nucleosome turnover.

The implementation of this reaction scheme is performed using Monte Carlo simulation, where each of these processes occurs as a homogeneous Poisson process with exponentially distributed event times. The lattice space of the simulation is implemented with periodic boundary conditions, hence no insulator elements are present. Full details of this simulation are described in the *Methods* section, and MATLAB code of its implementation is provided as supplemental information.

When these processes are permitted to occur simultaneously, a dynamic H3K36me3 domain is established over the simulated gene body (**Figure 2E**). These simulated domains bear several features in common with our observations of H3K36me3 ChIP-seq data. In particular, simulations with low transcription probabilities bear elevated levels of H3K36me3 compared to background. Moreover, H3K36me3 levels of lowly transcribed genes in these simulations are significantly reduced compared to simulations with high transcription frequencies (**Figure 2E-G**).

To assess the contribution of the transcription-independent spreading, we performed 1,014 simulations by spanning a range of P_n, P_i, and P_+, while keeping P_{bg} and P_- constant. The output of each of these simulations spanned a range of outcomes from very little marking throughout the simulation, to full saturation of the simulation space. Between these extremes, a domain of H3K36me3 is established that shows a strong dependency on transcription frequency. The full set of these outcomes is presented in **Figure S4**.

By analyzing the influence of the outcomes of the spreading parameter P_+, we found that elevated mark spreading in a transcriptionally independent manner would result in poor establishment of H3K36me3 domains with sufficient contrast above non-transcribed regions (**Figure 3**). We found that the major effect of increasing the spreading term P+ was to increase background marking intensity without providing a corresponding increase in marking of the gene body. This outcome was true regardless of whether transcriptional frequency was low or high (**Figure 3A**).

To analyze this phenomenon in quantitative terms, we adopted a similar set of metrics to those we used previously to characterize H3K9me3 domains[58]. To quantify the mean intensity of marking, we use the term $\langle N \rangle$, which reports the fraction of nucleosome positions in the gene body with H3K36me3 marks. To quantify the overall specificity of the domain, we used a specificity score *S*:

$$S = \frac{\langle I_{\text{gene body}} \rangle - \langle I \rangle}{\langle I_{\text{gene body}} \rangle}$$

where $\langle I_{\text{gene body}} \rangle$ is the mean mark density at the gene body, and $\langle I \rangle$ is the mean mark density over the entire lattice. Finally, the contrast between gene bodies and background regions is reflected by the efficiency score *E*:

$$E = \langle I_{\text{gene body}} \rangle - \langle I \rangle$$

By plotting the intensity, specificity, and efficiency of the domains across different parameters, it is readily apparent why elevated spreading results in poor domains. While the intensity of marking increases within gene bodies, the specificity of this marking strongly drops due to the presence of additional marks outside of gene bodies (**Figure 3B-C**). The net result of elevated spreading is to reduce the contrast of the marks between the gene bodies and background

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

regions (**Figure 3D**). Based on this result, we conclude that transcription-dependent is the dominant mode of H3K36me3 placement, and therefore anticipate that non-transcriptional propagation of H3K36me3 in cells is likely to produce a small fraction of the overall H3K36me3 marks at gene bodies.

Replication-independent histone turnover induces an asymmetric distribution of H3K36me3 at the TSS

One notable feature of the H3K36me3 ChIP-seq tracks that was not captured by our model is the reduction of mark intensity near the TSS. This reduction is found in many individual tracks and is largely independent of transcription frequency (**Figure 1D**). We hypothesized that increased histone turnover in this region could induce a substantial reduction of H3K36me3 intensity in a dynamic manner by decreasing the effective lifetime of the H3K36me3 mark near the TSS.

We therefore sought a metric for histone turnover across the gene body. Several experimental approaches have been developed to measure the absolute kinetics of histone lifetime, including metabolic labeling and CATCH-IT [63,64]. These methods have revealed a striking correspondence of histone lifetime with the profiles revealed by ChIP of the replication-independent histone H3.3[64]. We developed a variant of our standard model, where the rate of histone turnover was not uniform throughout the gene body, but varied by position (**Figure 4A**). We used the average profile of H3.3[65,66] as a proxy for the relative rates of replication-independent histone turnover across the gene body and compared the results of the standard model (constant turnover) with the variable turnover model.

Compared to the standard model, the resulting H3K36me3 profiles from the variable turnover model are strikingly similar to the genomic profiles observed from ChIP-seq (**Figure 4B-C**). These profiles also show a similar dependence on transcription frequency as the ChIP-seq data. We therefore conclude that a full description of H3K36me3 dynamics requires including the varied rates of histone turnover that occur near transcription start sites. Our findings furthermore suggest that H3K36me3 is not faithfully copied nor rapidly spread onto the replication-independent H3.3 histones, as these processes would otherwise mask this dependency. Consequently our work confirms that the levels and genome-wide profiles of H3K36me3 are regulated in a replication-independent manner.

SPT6 loss impairs the scaling relationship between H3K36me3 and transcriptional frequency

To validate our model and test whether transcription-coupled activities are responsible for most H3K36me3 placement, we sought experimental confirmation that disruption to the SETD2:IWS1:SPT6 ternary complex would affect the scaling relationship between H3K36me3 abundance and transcription frequency. We therefore compared mononucleosomal H3K36me3 signals obtained following siRNA transfection against either luciferase negative control (siLuc) or SPT6 (siSPT6), previously obtained elsewhere[67]. To measure transcription frequency, we used 3' RNA-seq data from the same cells. Counts obtained using 3' RNA-seq directly report on the frequency of transcription events at a given locus.

By plotting the H3K36me3 profiles based on transcription frequency in the siLuc control, we observed the scaling relationship between transcriptional frequency and H3K36me3 levels as described in Figure 1 above, confirming that the control does not affect this relationship (**Figure 5A-B,** R=0.462). In contrast, transfection of cells with siSPT6 causes a profound but incomplete reduction in the levels of H3K36me3 at gene bodies (**Figure 5C**). Moreover, this treatment alters the relationship between transcription frequency and H3K36me3 levels: Instead of observing a strong scaling relationship, the H3K36me3 levels show a reduced dependence on transcription frequency (**Figure 5C-D**, R=0.05). Pooling of transcripts into quintiles reveals that this effect persists despite pooling and across genomic replicates (**Figure 5E**) [$F(1,14) = 13.4$, $p = 2.6e\text{-}3$]. Together, our analysis shows that disruption of the SETD2:IWS1:SPT6 ternary complex alters the relationship between H3K36me3 accumulation and transcription frequency, providing an important validation of our kinetic model.

Discussion

Here we have employed quantitative approaches to accurately model the properties of H3K36me3 domains. Our comprehensive analysis of histone marks indicates that H3K36me3 domains are uniquely located throughout the transcribed regions of genes in different cell types. In contrast to some reports suggesting that H3K36me3 have strongly biased distributions over the gene bodies, our analysis shows that H3K36me3 domains are characterized as being broad, flat plateaus that span the entirety of gene bodies, with sharply defined boundaries and only modest deviations in intensity from the TSS to the TTS. We find a direct correspondence between transcription frequency and the intensity of H3K36me3 marking, indicating a tight relationship between transcription and H3K36me3 placement. Of the histone modifications we analyzed, the plateau-like distribution is characteristic only of H3K36me3, unlike other marks which show a high degree of enrichment near the TSS. The even marking throughout the gene body suggests that placement of the H3K36me3 mark occurs continuously throughout the entire process of transcription elongation.

Many questions have remained regarding the mechanisms that explain the accumulation of these marks across gene bodies. We used lattice-based simulation approaches to assess the contribution of transcription-dependent marking, transcription-independent maintenance (spreading) and histone turnover to the H3K36me3 domain shape. Our results indicate that transcription-independent spreading at even low rates would increase the H3K36me3 background levels outside of the coding regions. We find that elevated rates of spreading would eventually lead to collapse of the sharp domain boundaries that are observed in vivo. Therefore, our model suggests that transcription-dependent marking is the major pathway for H3K36me3 domain formation. This is in agreement with observation that H3K36me3 levels are proportional to the transcription

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

frequency of individual genes. Importantly, SETD2-dependent H3K36me3 marking has essential roles at sites of double-stranded DNA breaks[68]. Although we find that spreading mechanisms do not contribute substantially to normal sites of transcription, we cannot currently exclude the involvement of transcription-independent spreading mechanisms during double-strand break repair.

Incorporation of replication-independent histone turnover into our model reduces the mark intensity near the TSS, yielding a strikingly similar domain shape to profiles measured in cells. Interestingly, genome-wide studies show that many active genes have a nucleosome-free region at the transcription start site. Our analysis (**Figure S5**) demonstrate that unlike H3K36me3 intensity, the size of this region does not correlate with transcription frequency, suggesting that the steady-state exclusion of histones from promoter regions is independent of H3K36me3 accumulation.

Transcription-dependent H3K36me3 marking relies on RNAP2:SETD2:IWS1:SPT6 complex formation[6]. Because our analysis suggested that co-transcriptional marking was the major pathway of H3K36me3 placement, we sought to test and validate this model by quantifying the effects following disruption of the SETD2:IWS1:SPT6 ternary complex. We analyzed ChIP-seq data obtained in the presence and absence of SPT6. Despite the loss of transcription scaling in SPT6 knock-down cells, a modest accumulation of H3K36me3 is preserved over gene bodies, suggesting that some kind of non-transcriptional mechanisms may still be operating in the absence of SPT6. As such, it is tempting to speculate that residual marking might reflect transcription-independent processes that may contribute to bookmarking of gene bodies independently of complexes involving SPT6. Such processes might be essential to prevent the complete silencing of genes by Polycomb complexes, since H3K36me3-containing nucleosomes are poor PRC2 substrates and H3K36me3 antagonizes PRC2-mediated H3K27 methylation[69,70]. In general, we propose that significant deviations of the transcription dependence of H3K36me3 levels may reflect biologically important sources of regulation, for example sites where polymerases move unusually quickly or slowly through a gene[71], or where other regulatory factors such as H3K36 demethylases are in close proximity (**Figure 5F**)

Genome-wide analysis of H3K36me3 shows that it does not spread outside of the gene body boundaries, despite the lack of any obvious insulator elements. Furthermore, others have shown using an inducible system that these marks also do not spread onto adjacent loci[45]. In the same study, the authors found that the H3K36me3 mark persisted in the locus for 60 minutes after transcription inhibition, suggesting a short epigenetic memory for recently occurred transcriptional activity. This finding is consistent with our model, which suggests a short persistence of H3K36me3 in the absence of transcription. We therefore conclude that the strong dependence on co-transcriptional processes limits the capacity of the H3K36me3 mark to serve as a stable, heritable epigenetic mark at gene bodies.

Many of the processes dependent on H3K36me3 are ubiquitous and highly conserved from yeast to humans. Additionally, H3K36 oncohistone mutations[72] and loss of the H3K36me3 methyltransferase SETD2[73] both play important roles during cancer development. In particular, lysine-to-methionine mutations of H3K36 (H3K36M) contribute to distinct subtypes of sarcomas and pediatric chondroblastoma[74,75]. In these patients, expression of H3K36M mutants leads to global reduction of H3K36 methylation, including H3K36me3 throughout gene bodies, and shapes the oncogenic transcriptome of these cancer cells. Additionally, SETD2 deletions and loss-of-function mutations have been identified in patients with clear cell renal cell carcinoma (ccRCC), and alterations of SETD2 have been uncovered in pediatric high-grade gliomas, colorectal cancer and several hematologic malignancies. In ccRCC, progressive deregulation of H3K36me3 is linked to metastasis and is associated with overall lower survival [19,76]. The H3K36me3 demethylases (JMJD2A and B) have also been implicated in colorectal cancer[77] and ER positive breast cancer[78] development. Overall, dysregulation of H3K36me3 homeostasis represents a common pathway that contributes to several malignancies.

Drugs targeting chromatin regulators are currently viewed as promising treatment avenues for various diseases, and the number of such drugs in clinical use is anticipated to grow in the foreseeable future[79–83]. Unfortunately, the cross-talk between chromatin modifying enzymes and other factors responsible for the maintenance of their modifications remains enigmatic. Insight into this interplay is an essential first step towards rational targeting of the chromatin landscape. Additionally, understanding of marking kinetics is invaluable for drug dosage and sustainability of the desired effect. Therefore, investigation of dynamic mechanisms and kinetic properties that underlie steady-state activities of each mark is crucial for the development of future treatments.

The chromatin landscape is a highly dissipative complex system that constantly integrates hundreds of signals and coordinates appropriate responses. Simulation approaches premised on simple assumptions like those that we present here enable modeling of different scenarios and are therefore ideal for quantitative profiling of the dynamical behavior and kinetics of chromatin marking. Additionally, a number of modern chemical biology tools and chemical probes enable rapid perturbation and characterization of highly dynamic and complex systems like the chromatin landscape for experimental validation of these processes[83]. Future models incorporating additional processes, such as co-transcriptional regulation, may further illuminate the functions of H3K36me3 and its role in development and disease.

Acknowledgments

This work was supported by NIH grant R00CA187565 (H.C.H.), the Cancer Prevention & Research Institute of Texas Grant RR170036 (H.C.H.), the V Foundation grant V2018-003 (H.C.H.), and Gabrielle's Angel Foundation for Cancer Research (H.C.H.), as well as GACR grants 16-06357S and 19-14360S (V.V.)

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Author contributions
K.C., V.V., and H.C.H. conceived of work, K.C. and H.C.H. analyzed data and wrote the manuscript. E.A.S., V.V., and H.C.H. provided critical feedback.

Methods
Processing of ChIP-seq data. Data set files for H3K4me1, H3K4me3, H3K27me3, H3K27ac, H3K79me2, and H3K36me3 were obtained as bigWig files from the ENCODE Consortium for both K562 and A549 cells. These files were provided by the ENCODE Consortium based on uniquely mapped reads to the hg19 genome build. H3.3 ChIP seq profiles from HeLa cells were obtained from GEO accession GSE45023. All data was processed and analyzed from at least 2 independent biological replicates.

Heat maps and metagene plots. ChIP-seq bigWig files were processed using bwtool[84] to extract the regions 34,600 bp upstream of TSSs, and 17,300 bp downstream of TSSs, and the resulting values were averaged at 200 bp windows. The resulting matrix was used to generate metagene profiles and heatmaps using R. Similar matrices were obtained using the same approach for the region 17,300 bp upstream of TTSs and 34,600 bp downstream of TTSs. Scaled whole-gene body metagene plots were not used because they were found to introduce biases that did not exist in unscaled plots.

RNA-seq analysis and integration with ChIP-seq data sets. FPKM values were obtained directly from the ENCODE Consortium following mapping to the hg19 genome. To harmonize the relationship between RNA-seq and ChIP-seq data, the UCSC table browser was used to map annotations reported using Ensembl transcript identifiers to gene symbols. For RNA-seq, the longest transcript was chosen as the canonical transcript, and its TSS was used for ChIP-seq analysis. All analyses were performed by matching the gene symbols between ChIP-seq and RNA-seq data sets. All data was processed and analyzed from at least 2 independent biological replicates.

MNase analysis and measurement of nucleosomal spacing. MNase-seq data was obtained from the ENCODE Consortium for K562 cells as bigWig files. For each canonical TSS defined above, we began by Z scaling the MNase-seq signal in the window ± 4 kbp from the TSS. The +1 nucleosome was then identified by the position where the Z-scaled MNase-seq signal reached 0.5 downstream of the TSS, and heat maps were created beginning 200 bp upstream of this position. Phase scores were assigned for each locus, and the top-scoring TSSs containing strong phases were used to plot the matrix of MNase signals. The resulting signal was averaged across all TSSs with sufficiently high phase scores and showed a readily apparent 195-bp periodicity. The precise value of this periodicity was measured with greater precision by examining the strongest frequency component following fast Fourier transform (FFT). The highest frequency signal in the FFT corresponded to 1/195 bp, confirming the canonical 195-bp periodicity reported elsewhere[85].

Protein-protein interaction network graphing. Protein-protein graphs of functional interactions between proteins were plotted using STRING[86].

Browser tracks. Genome browser tracks were generated using the WashU Epigenome Browser, with the resulting tracks exported as SVG files.

Monte Carlo simulations. Implementation of Monte Carlo simulations was performed using custom code written in MATLAB (Mathworks, Natick, MA). All Monte Carlo steps were implemented using classic Monte Carlo methods and will be provided as supplemental data upon acceptance of the manuscript. Briefly, The stochastic trajectory of a given genetic locus is carried out by iteration of the following processes at each time step (see Figure 2):

(1) Transcription-coupled marking: if transcription is occurring in a given time step, each gene body site at position j that is unmarked ($l_j = 0$) is converted to a marked state ($l_j = 1$) with probability P_n (where $P_n = k_n \Delta t$). Transcription occurs within each given time step randomly with probability P_i.

(2) Transcription-independent spreading: for each marked nucleosome j on the lattice, if the j − 1 position is unmarked ($l_{j-1} = 0$), it is converted to a marked state ($l_{j-1} = 1$) with probability P_+. Similarly, if the j + 1 position is unmarked ($l_{j+1} = 0$) , it is converted to a marked state ($l_{j+1} = 1$) with probability P_+ . Mark propagation thus proceeds outward in both directions from each marked site.

(3) Background marking: each position j that is unmarked ($l_j = 0$) is converted to a marked state ($l_j = 1$) with probability P_{bg} (where $P_{bg} = k_{bg} \Delta t$).

(4) Turnover: for each marked nucleosome j on the lattice, conversion to an unmarked state ($l_j = 0$) occurs with probability P_- (where $P_- = k_- \Delta t$).

(5) Time evolution: Simulation time t is incremented by Δt.

The above procedure treats transcription, marking, and turnover as homogeneous Poisson processes with exponentially distributed lifetimes. We also allow all simulations to evolve using periodic boundary conditions; therefore, there is no boundary implicit in the simulations. For all simulations, the value of Δt was chosen to be sufficiently small such that $P_- = 0.05$. Source code for Monte Carlo simulations will be provided before or upon acceptance.

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Statistical analyses. All statistical procedures were performed using R. All reported P values were obtained using two-sided tests. When calculated p-values are smaller than the 64-bit double precision machine epsilon (2^{-52} = 2.2e-16), p-values are reported as p<2.2e-16.

Figure Legends

Figure 1. H3K36me3 placement by the SETD2:IWS1:SPT6 complex and genome-wide dependence on transcription. **(A)** STRING analysis of validated interactions between RNA Polymerase 2 (RNAP2) and histone methyltransferases MLL (KMT2A), DOT1L, and SETD2. The SETD2:IWS1:SPT6 ternary complex has extensive interactions with RNAP2 subunits. **(B)** IWS1 and SPT6 interact with the RNAP2 unstructured C-terminal domain to recruit SETD2, leading to H3K36me3 placement. **(C)** Analysis of histone ChIP-seq signals at gene bodies in K562 cells, organized by the degree of transcription measured via RNA-seq. H3K36me3 is the histone mark with the highest correlation to transcription. Measurement of correlation values is provided in **Figure S1**. **(D)** H3K36me3 profiles are flat plateaus over transcriptional units starting with the transcription start site (TSS) through the transcription termination site (TTS) in K562 cells. **(E)** The mean ChIP-seq enrichment of H3K36me3 scales with the transcript abundance measured with RNA-seq. **(F)** Heatmap of H3K36me3 plateaus at individual transcription units with their associated RNA-seq output measured by fragments per kilobase million (FPKM). **(G)** Individual genome browser tracks in K562 cells.

Figure 2. General components of the kinetic model of H3K36me3 at transcriptional units. **(A)** During a transcription event, nucleosomes within a transcriptional unit (gene body) are marked with constant probability. **(B)** H3K36me3 marks may be maintained by transcription-independent processes that lead to spreading of the marks to neighboring sites with constant probability. **(C)** All sites are marked with a very low background activity with constant probability. **(D)** Sites marked with HK36me3 are randomly turned over to unmarked sites with constant probability. **(E)** Visualization of an individual simulation over time reveals transient accumulation of H3K36me3 marks that are dissipated when transcription is infrequent. **(F)** Visualization of an individual simulation over time reveals steady-state levels of H3K36me3 marks are maintained when transcription is frequent. **(G)** The profile of simulated H3K36me3 levels in simulations scales in intensity with transcription frequency.

Figure 3. The effects of transcription-independent spreading of H3K36me3 in simulations. **(A)** For both low- and high-transcription simulations, as local spreading probability increases, the entire simulation experiences increased levels of marking, reducing the contrast of the mark between gene body and background. **(B)** Plot of mean H3K36me3 mark intensity at gene bodies as a function of transcription probability and spreading probability. **(C)** Plot of H3K36me3 specificity at gene bodies as a function of transcription probability and spreading probability. **(D)** Plot of H3K36me3 efficiency at gene bodies as a function of transcription probability and spreading probability.

Figure 4. Variable histone turnover reproduces in vivo-like distributions of H3K36me3. **(A)** In addition to the standard constant turnover model, a variable turnover model where histone turnover is incorporated is presented. The rate of replication-independent histone turnover is obtained by measuring the profiles of H3.3 over gene bodies from ChIP-seq. **(B)** The constant turnover model reproduces essential features of H3K36me3 as observed from ChIP-seq data, including the observed scaling with transcription frequency. **(C)** The variable turnover model reproduces several in vivo-like features of H3K36me3 as observed from ChIP-seq data, including the characteristic asymmetric depletion and scalloped edges at the TSS, as well as the observed scaling with transcription frequency.

Figure 5. Validation of transcription-dependent kinetic model of H3K36me3 marking. **(A)** Analysis of the profile of H3K36me3 at gene bodies as a function of transcription frequency following control RNAi silencing of luciferase (siLuc). **(B)** Mean H3K36me3 level scales in the siLuc control with transcription frequency based on analysis of 3' RNA-seq counts. **(C)** Analysis of the profile of H3K36me3 at gene bodies as a function of transcription frequency following RNAi silencing of SPT6 (siSPT6). **(D)** Mean H3K36me3 level scales in the siSPT6 with transcription frequency based on analysis of 3' RNA-seq counts. **(E)** Analysis of transcription dependent H3K36me3 marking following segmenting transcripts into quintiles. Analysis from n=2 independent genome-wide replicates reveals a reduced scaling relation between transcription frequency and H3K36me3 levels. **(F)** Conceptual model describing the scaling relationship between transcription frequency and H3K36me3 levels. Deviations from this relation can arise from several sources of site-specific biological regulation.

Supplemental Figure Legends

Figure S1: Relationship between transcription frequency and ChIP-seq enrichment of chromatin marks at gene bodies. The mean ChIP-seq enrichment of **(A)** H3K4me1, **(B)** H3K4me3, **(C)** H3K27ac, **(D)** H3K27me3, **(E)** H3K36me3 and **(F)** H3K72me2 marks in K562 cells is plotted against the transcript abundance measured by RNA-seq.

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S2: Analysis of histone marks at gene bodies in A549 cells. (A) H3K36me3 profiles over transcriptionally active genes; transcription start sites (TSS) and transcription termination sites (TTS) are highlighted. The individual genes clustered by transcription abundance determined by RNA-seq. **(B)** The mean ChIP-seq enrichment of H3K36me3 scales with transcript abundance measured by RNA-seq. **(C)** Heatmap of H3K36me3 plateaus at individual transcription units with their associated RNA-seq output measured by fragments per kilobase million (FPKM). **(G)** Individual genome browser tracks for different chromatin marks in A549 cells.

Figure S3: Analysis of nucleosome occupancy downstream of transcription start sites. (A) Heatmap of MNase-seq fragments in K562 cells at individual genes aligned by their transcription start sites (TSSs) and arranged by transcription frequency. Arrays of nucleosomes positioned downstream of TSSs are highlighted. **(B)** Mean nucleosome density determined by MNase-seq aggregated across individual genes ordered by their transcription frequency. **(C)** Phasing of the MNase-seq signal is constant.

Figure S4: Model output spanning all parameters. Dependence of H3K36me3 domain shape and boundaries on frequency of transcription-independent spreading, background marking and transcription-dependent marking. The probabilities of transcription event (Pi), H3K36me3 propagation by transcription-independent spreading (P+) and by transcription-dependent placement (Pn) are underlying variables.

Figure S5: Nucleosome-free regions do not scale with transcription frequency. (A) Heatmap of MNase-seq fragments at individual genes aligned by their TSS and ordered by the size of nucleosome free region. The abundance of transcripts of individual genes measured by RNA-seq are highlighted. **(B)** The transcript abundance measured by RNA-seq plotted against size of nucleosome-depleted region (NDR). **(C)** H3K36me3 levels show no strong relationship with the size of the NDR.

References

1. Wagner, E. J. & Carpenter, P. B. Understanding the language of Lys36 methylation at histone H3. *Nature Reviews Molecular Cell Biology* (2012). doi:10.1038/nrm3274
2. Ernst, J. *et al.* Mapping and analysis of chromatin state dynamics in nine human cell types. *Nature* (2011). doi:10.1038/nature09906
3. Chen, E. S. *et al.* Cell cycle control of centromeric repeat transcription and heterochromatin assembly. *Nature* (2008). doi:10.1038/nature06561
4. Woo, H., Ha, S. D., Lee, S. B., Buratowski, S. & Kim, T. S. Modulation of gene expression dynamics by co-Transcriptional histone methylations. *Experimental and Molecular Medicine* (2017). doi:10.1038/emm.2017.19
5. Krogan, N. J. *et al.* Methylation of Histone H3 by Set2 in Saccharomyces cerevisiae Is Linked to Transcriptional Elongation by RNA Polymerase II. *Mol. Cell. Biol.* (2003). doi:10.1128/mcb.23.12.4207-4218.2003
6. Yoh, S. M., Lucas, J. S. & Jones, K. A. The Iws1:Spt6:CTD complex controls cotranscriptional mRNA biosynthesis and HYPB/Setd2-mediated histone H3K36 methylation. *Genes Dev.* (2008). doi:10.1101/gad.1720008
7. Neri, F. *et al.* Intragenic DNA methylation prevents spurious transcription initiation. *Nature* (2017). doi:10.1038/nature21373
8. Kim, H., Fong, N., Kim, S., Erickson, B. & Bentley, D. L. Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci.* (2011). doi:10.1073/pnas.1109475108
9. Kolasinska-Zwierz, P. *et al.* Differential chromatin marking of introns and expressed exons by H3K36me3. *Nat. Genet.* (2009). doi:10.1038/ng.322
10. Li, F. *et al.* The histone mark H3K36me3 regulates human DNA mismatch repair through its interaction with MutSα. *Cell* (2013). doi:10.1016/j.cell.2013.03.025
11. Sun, X. J. *et al.* Identification and characterization of a novel human histone H3 lysine 36-specific methyltransferase. *J. Biol. Chem.* (2005). doi:10.1074/jbc.M504012200
12. Edmunds, J. W., Mahadevan, L. C. & Clayton, A. L. Dynamic histone H3 methylation during gene induction: HYPB/Setd2 mediates all H3K36 trimethylation. *EMBO J.* (2008). doi:10.1038/sj.emboj.7601967
13. Duns, G. *et al.* Histone methyltransferase gene SETD2 is a novel tumor suppressor gene in clear cell renal cell carcinoma. *Cancer Res.* (2010). doi:10.1158/0008-5472.CAN-10-0120
14. Sarakbi, W. Al, Sasi, W., Jiang, W. & Roberts, T. The mRNA expression of SETD 2 in human breast cancer: correlation with clinico-pathological parameters. *BMC* (2009).
15. Weinstein, J. N. *et al.* Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* (2014). doi:10.1038/nature12965
16. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* (2012). doi:10.1038/nature10725
17. Zhu, X. *et al.* Identification of functional cooperative mutations of SETD2 in human acute leukemia. *Nat. Genet.* (2014). doi:10.1038/ng.2894
18. Mar, B. G. *et al.* Mutations in epigenetic regulators including SETD2 are gained during relapse in paediatric acute lymphoblastic leukaemia. *Nat. Commun.* (2014). doi:10.1038/ncomms4469

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

19. Ho, T. H. *et al.* High-resolution profiling of histone h3 lysine 36 trimethylation in metastatic renal cell carcinoma. *Oncogene* (2016). doi:10.1038/onc.2015.221
20. Klose, R. J., Kallin, E. M. & Zhang, Y. JmjC-domain-containing proteins and histone demethylation. *Nat. Rev. Genet.* (2006). doi:10.1038/nrg1945
21. Hodges, C., Kirkland, J. G. & Crabtree, G. R. The many roles of BAF (mSWI/SNF) and PBAF complexes in cancer. *Cold Spring Harb. Perspect. Med.* (2016). doi:10.1101/cshperspect.a026930
22. Hodges, H. C. *et al.* Dominant-negative SMARCA4 mutants alter the accessibility landscape of tissue-unrestricted enhancers. *Nat. Struct. Mol. Biol.* **25**, 61–72 (2018).
23. Lin, L. -j., Minard, L. V., Johnston, G. C., Singer, R. A. & Schultz, M. C. Asf1 Can Promote Trimethylation of H3 K36 by Set2. *Mol. Cell. Biol.* (2010). doi:10.1128/MCB.01229-09
24. Youdell, M. L. *et al.* Roles for Ctk1 and Spt6 in regulating the different methylation states of histone H3 lysine 36. *Mol. Cell. Biol.* (2008). doi:10.1128/MCB.00001-08
25. Dion, M. F. *et al.* Dynamics of replication-independent histone turnover in budding yeast. *Science (80-.).* (2007). doi:10.1126/science.1134053
26. Kraushaar, D. C. *et al.* Genome-wide incorporation dynamics reveal distinct categories of turnover for the histone variant H3.3. *Genome Biol.* (2013). doi:10.1186/gb-2013-14-10-r121
27. Hyun, K., Jeon, J., Park, K. & Kim, J. Writing, erasing and reading histone lysine methylations. *Experimental and Molecular Medicine* (2017). doi:10.1038/emm.2017.11
28. Cai, L. *et al.* An H3K36 Methylation-Engaging Tudor Motif of Polycomb-like Proteins Mediates PRC2 Complex Targeting. *Mol. Cell* (2013). doi:10.1016/j.molcel.2012.11.026
29. Dhayalan, A. *et al.* The Dnmt3a PWWP domain reads histone 3 lysine 36 trimethylation and guides DNA methylation. *J. Biol. Chem.* (2010). doi:10.1074/jbc.M109.089433
30. Eidahl, J. O. *et al.* Structural basis for high-affinity binding of LEDGF PWWP to mononucleosomes. *Nucleic Acids Res.* (2013). doi:10.1093/nar/gkt074
31. Sharma, S. *et al.* Affinity switching of the LEDGF/p75 IBD interactome is governed by kinase-dependent phosphorylation. *Proc. Natl. Acad. Sci.* (2018). doi:10.1073/pnas.1803909115
32. Hsin, J. P. & Manley, J. L. The RNA polymerase II CTD coordinates transcription and RNA processing. *Genes and Development* (2012). doi:10.1101/gad.200303.112
33. Harlen, K. M. & Churchman, L. S. The code and beyond: Transcription regulation by the RNA polymerase II carboxy-terminal domain. *Nature Reviews Molecular Cell Biology* (2017). doi:10.1038/nrm.2017.10
34. Hampsey, M. & Reinberg, D. Tails of intrigue: Phosphorylation of RNA polymerase II mediates histone methylation. *Cell* (2003). doi:10.1016/S0092-8674(03)00360-X
35. Kizer, K. *et al.* A novel domain in Set2 mediates RNA polymerase II interaction and couples histone H3 K36 methylation with transcript elongation. *Mol. Cell. Biol.* (2005). doi:10.1128/MCB.25.8.3305-3316.2005
36. Kim, S. K. *et al.* Human histone H3K79 methyltransferase DOT1L methyltransferase binds actively transcribing RNA polymerase II to regulate gene expression. *J. Biol. Chem.* (2012). doi:10.1074/jbc.M112.384057
37. Fuchs, S. M., Kizer, K. O., Braberg, H., Krogan, N. J. & Strahl, B. D. RNA polymerase II carboxyl-terminal domain phosphorylation regulates protein stability of the set2 methyltransferase and histone H3 di- and trimethylation at lysine 36. *J. Biol. Chem.* (2012). doi:10.1074/jbc.M111.273953
38. Milne, T. A. *et al.* From The Cover: MLL associates specifically with a subset of transcriptionally active target genes. *Proc. Natl. Acad. Sci.* (2005). doi:10.1073/pnas.0503630102
39. Chantalat, S. *et al.* Histone H3 trimethylation at lysine 36 is associated with constitutive and facultative heterochromatin. *Genome Res.* (2011). doi:10.1101/gr.118091.110
40. Kimura, H. Histone modifications for human epigenome analysis. *Journal of Human Genetics* (2013). doi:10.1038/jhg.2013.66
41. Vakoc, C. R., Sachdeva, M. M., Wang, H. & Blobel, G. A. Profile of Histone Lysine Methylation across Transcribed Mammalian Chromatin. *Mol. Cell. Biol.* (2006). doi:10.1128/mcb.01529-06
42. Barski, A. *et al.* High-resolution profiling of histone methylations in the human genome. *Cell* (2007). doi:10.1016/j.cell.2007.05.009
43. Komarnitsky, P., Cho, E. J. & Buratowski, S. Different phosphorylated forms of RNA polymerase II and associated mRNA processing factors during transcription. *Genes Dev.* (2000). doi:10.1101/gad.824700
44. Yoh, S. M., Cho, H., Pickle, L., Evans, R. M. & Jones, K. A. The Spt6 SH2 domain binds Ser2-P RNAPII to direct Iws1-dependent mRNA splicing and export. *Genes Dev.* (2007). doi:10.1101/gad.1503107
45. Sein, H., Värv, S. & Kristjuhan, A. Distribution and maintenance of histone H3 Lysine 36 trimethylation in transcribed locus Lysine 36 trimethylation in transcribed locus. *PLoS One* (2015). doi:10.1371/journal.pone.0120200
46. Xie, P. *et al.* Histone methyltransferase protein SETD2 interacts with p53 and selectively regulates its downstream genes. *Cell. Signal.* (2008). doi:10.1016/j.cellsig.2008.05.012
47. Chen, K. *et al.* Methyltransferase SETD2-Mediated Methylation of STAT1 Is Critical for Interferon Antiviral Activity. *Cell* (2017). doi:10.1016/j.cell.2017.06.042
48. Hawkins, R. D., Hon, G. C. & Ren, B. Next-generation genomics: An integrative approach. *Nature Reviews*

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

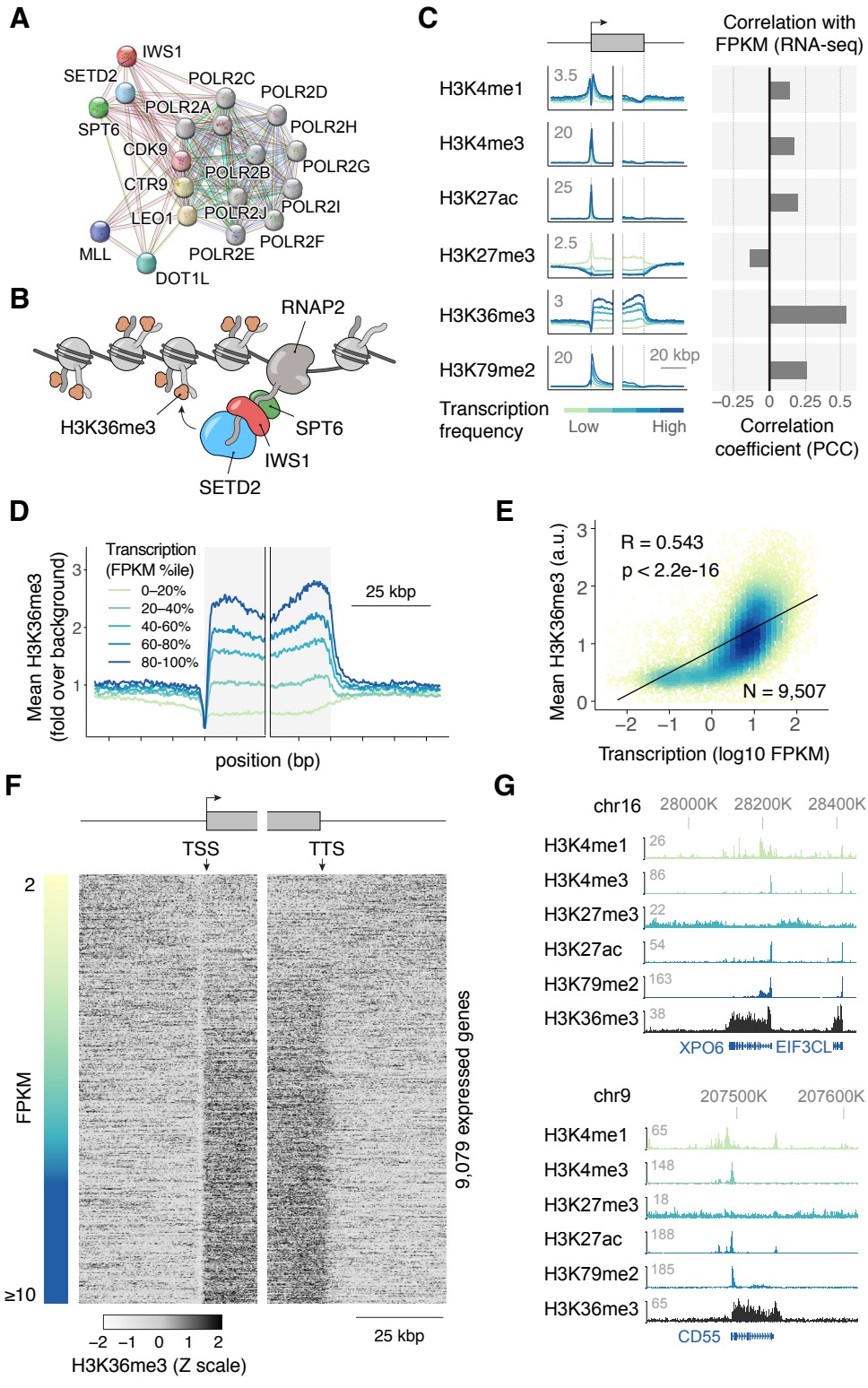
Genetics (2010). doi:10.1038/nrg2795
49. Pack, L. R., Yamamoto, K. R. & Fujimori, D. G. Opposing chromatin signals direct and regulate the activity of lysine demethylase 4C (KDM4C). *J. Biol. Chem.* (2016). doi:10.1074/jbc.M115.696864
50. Lim, P. S., Li, J., Holloway, A. F. & Rao, S. Epigenetic regulation of inducible gene expression in the immune system. *Immunology* (2013). doi:10.1111/imm.12100
51. Li, B., Carey, M. & Workman, J. L. The role of chromatin during transcription. *Cell* (2007). doi:10.1016/j.cell.2007.01.015
52. Barth, T. K. & Imhof, A. Fast signals and slow marks: the dynamics of histone modifications. *Trends in Biochemical Sciences* (2010). doi:10.1016/j.tibs.2010.05.006
53. Gates, L. A., Foulds, C. E. & O'Malley, B. W. Histone Marks in the 'Driver's Seat': Functional Roles in Steering the Transcription Cycle. *Trends in Biochemical Sciences* (2017). doi:10.1016/j.tibs.2017.10.004
54. Bicocca, V. T., Ormsby, T., Adhvaryu, K. K., Honda, S. & Selker, E. U. ASH1-catalyzed H3K36 methylation drives gene repression and marks H3K27me2/3-competent chromatin. *Elife* (2018). doi:10.7554/elife.41497
55. Li, G. & Reinberg, D. Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics and Development* (2011). doi:10.1016/j.gde.2011.01.022
56. Nicol-Benoit, F., Le Goff, P. & Michel, D. Drawing a Waddington landscape to capture dynamic epigenetics. *Biol. Cell* (2013). doi:10.1111/boc.201300029
57. Dodd, I. B., Micheelsen, M. A., Sneppen, K. & Thon, G. Theoretical Analysis of Epigenetic Cell Memory by Nucleosome Modification. *Cell* (2007). doi:10.1016/j.cell.2007.02.053
58. Hodges, C. & Crabtree, G. R. Dynamics of inherently bounded histone modification domains. *Proc. Natl. Acad. Sci.* (2012). doi:10.1073/pnas.1211172109
59. Hathaway, N. A. *et al.* Dynamics and memory of heterochromatin in living cells. *Cell* (2012). doi:10.1016/j.cell.2012.03.052
60. Chory, E. J. *et al.* Nucleosome Turnover Regulates Histone Methylation Patterns over the Genome. *Mol. Cell* (2019). doi:10.1016/j.molcel.2018.10.028
61. MacPherson, Q., Beltran, B. & Spakowitz, A. J. Bottom–up modeling of chromatin segregation due to epigenetic modifications. *Proc. Natl. Acad. Sci.* (2018). doi:10.1073/pnas.1812268115
62. Park, M., Patel, N., Keung, A. J. & Khalil, A. S. Engineering Epigenetic Regulation Using Synthetic Read-Write Modules. *Cell* (2019). doi:10.1016/j.cell.2018.11.002
63. Teves, S. S., Deal, R. B. & Henikoff, S. Measuring genome-wide nucleosome turnover using CATCH-IT. *Methods Enzymol.* (2012). doi:10.1016/B978-0-12-391938-0.00007-0
64. Deal, R. B., Henikoff, J. G. & Henikoff, S. Genome-wide kinetics of nucleosome turnover determined by metabolic labelinq of histones. *Science (80-.).* (2010). doi:10.1126/science.1186777
65. Jin, C. *et al.* H3.3/H2A.Z double variant-containing nucleosomes mark 'nucleosome-free regions' of active promoters and other regulatory regions. *Nat. Genet.* (2009). doi:10.1038/ng.409
66. Pchelintsev, N. A. *et al.* Placing the HIRA Histone Chaperone Complex in the Chromatin Landscape. *Cell Rep.* (2013). doi:10.1016/j.celrep.2013.03.026
67. Nojima, T. *et al.* Deregulated Expression of Mammalian lncRNA through Loss of SPT6 Induces R-Loop Formation, Replication Stress, and Cellular Senescence. *Mol. Cell* (2018). doi:10.1016/j.molcel.2018.10.011
68. Pfister, S. X. *et al.* SETD2-Dependent Histone H3K36 Trimethylation Is Required for Homologous Recombination Repair and Genome Stability. *Cell Rep.* (2014). doi:10.1016/j.celrep.2014.05.026
69. Yuan, W. *et al.* H3K36 methylation antagonizes PRC2-mediated H3K27 methylation. *J. Biol. Chem.* (2011). doi:10.1074/jbc.M110.194027
70. Schmitges, F. W. *et al.* Histone Methylation by PRC2 Is Inhibited by Active Chromatin Marks. *Mol. Cell* (2011). doi:10.1016/j.molcel.2011.03.025
71. Fong, N., Saldi, T., Sheridan, R. M., Cortazar, M. A. & Bentley, D. L. RNA Pol II Dynamics Modulate Co-transcriptional Chromatin Modification, CTD Phosphorylation, and Transcriptional Direction. *Mol. Cell* (2017). doi:10.1016/j.molcel.2017.04.016
72. Nacev, B. A. *et al.* The expanding landscape of 'oncohistone' mutations in human cancers. *Nature* (2019). doi:10.1038/s41586-019-1038-1
73. Gerlinger, M. *et al.* Intratumor Heterogeneity and Branched Evolution Revealed by Multiregion Sequencing. *N. Engl. J. Med.* (2012). doi:10.1056/NEJMoa1113205
74. Dong Fang, 1 *et al.* The histone H3.3K36M mutation reprograms the epigenome of chondroblastomas. *Science (80-.).* (2016). doi:10.1126/science.aaf6419
75. Lu, C. *et al.* Cancer: Histone H3K36 mutations promote sarcomagenesis through altered histone methylation landscape. *Science (80-.).* (2016). doi:10.1126/science.aac7272
76. Ho, T. H. *et al.* Loss of histone H3 lysine 36 trimethylation is associated with an increased risk of renal cell carcinoma-specific death. *Mod. Pathol.* (2016). doi:10.1038/modpathol.2015.123
77. Fu, L. *et al.* HIF-1α-induced histone demethylase JMJD2B contributes to the malignant phenotype of colorectal cancer cells via an epigenetic mechanism. *Carcinogenesis* (2012). doi:10.1093/carcin/bgs217
78. Berry, W. L., Shin, S., Lightfoot, S. A. & Janknecht, R. Oncogenic features of the JMJD2A histone demethylase in

bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

breast cancer. *Int. J. Oncol.* (2012). doi:10.3892/ijo.2012.1618
79. Buckstein, R., Yee, K. & Wells, R. A. 5-Azacytidine in myelodysplastic syndromes: A clinical practice guideline. *Cancer Treatment Reviews* (2011). doi:10.1016/j.ctrv.2010.05.006
80. Greenberg, P. L. *et al.* A randomized controlled trial of romiplostim in patients with low-or intermediate-risk myelodysplastic syndrome receiving decitabine. *Leuk. Lymphoma* (2013). doi:10.3109/10428194.2012.713477
81. Braiteh, F. *et al.* Phase I study of epigenetic modulation with 5-azacytidine and valproic acid in patients with advanced cancers. *Clin. Cancer Res.* (2008). doi:10.1158/1078-0432.CCR-08-1247
82. Thomas, X. G. *et al.* Results from a randomized phase III trial of decitabine versus supportive care or low-dose cytarabine for the treatment of older patients with newly diagnosed AML. *J. Clin. Oncol.* (2017). doi:10.1200/jco.2011.29.15_suppl.6504
83. Cermakova, K. & Courtney Hodges, H. Next-generation drugs and probes for chromatin biology: From targeted protein degradation to phase separation. *Molecules* (2018). doi:10.3390/molecules23081958
84. Pohl, A. & Beato, M. bwtool: A tool for bigWig files. *Bioinformatics* (2014). doi:10.1093/bioinformatics/btu056
85. Valouev, A. *et al.* Determinants of nucleosome organization in primary human cells. *Nature* (2011). doi:10.1038/nature10002
86. Szklarczyk, D. *et al.* STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* (2015). doi:10.1093/nar/gku1003

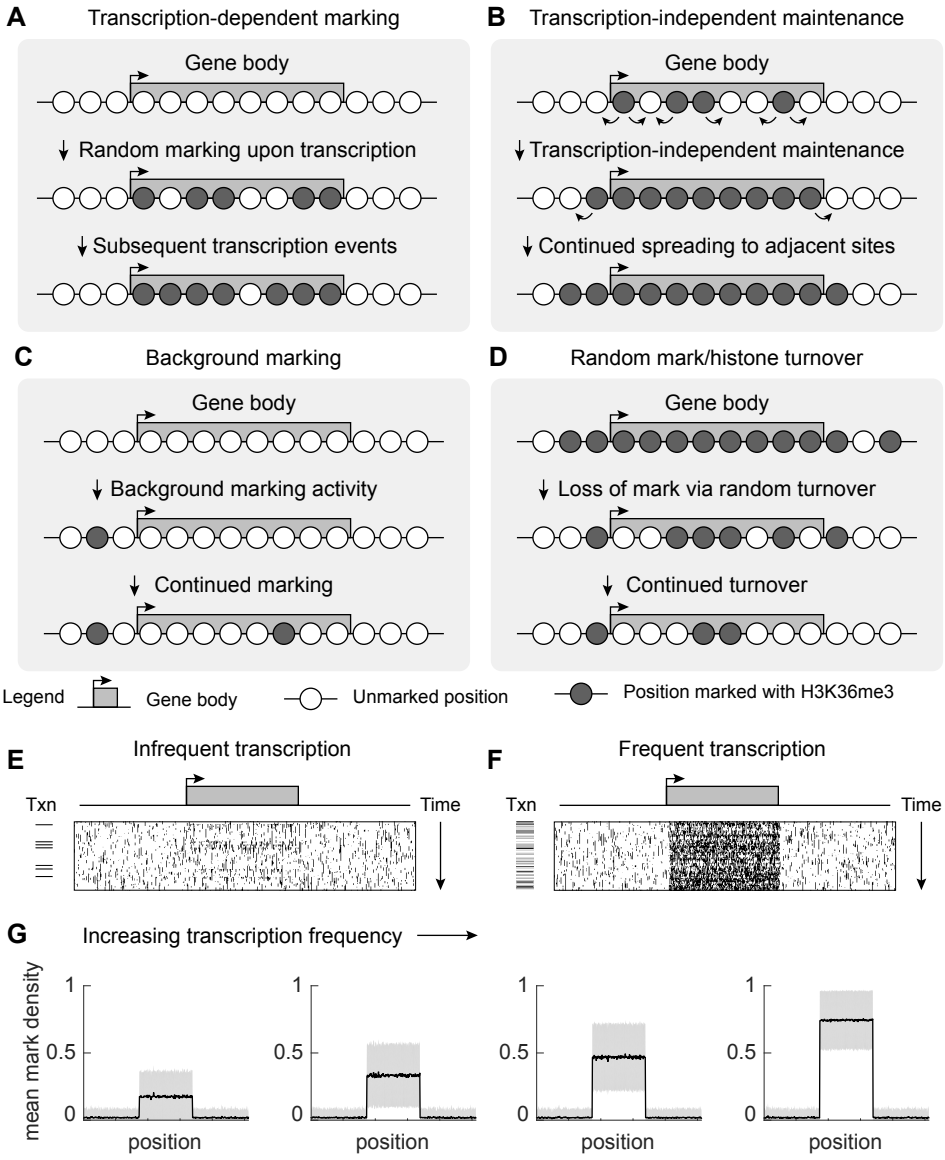
bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1



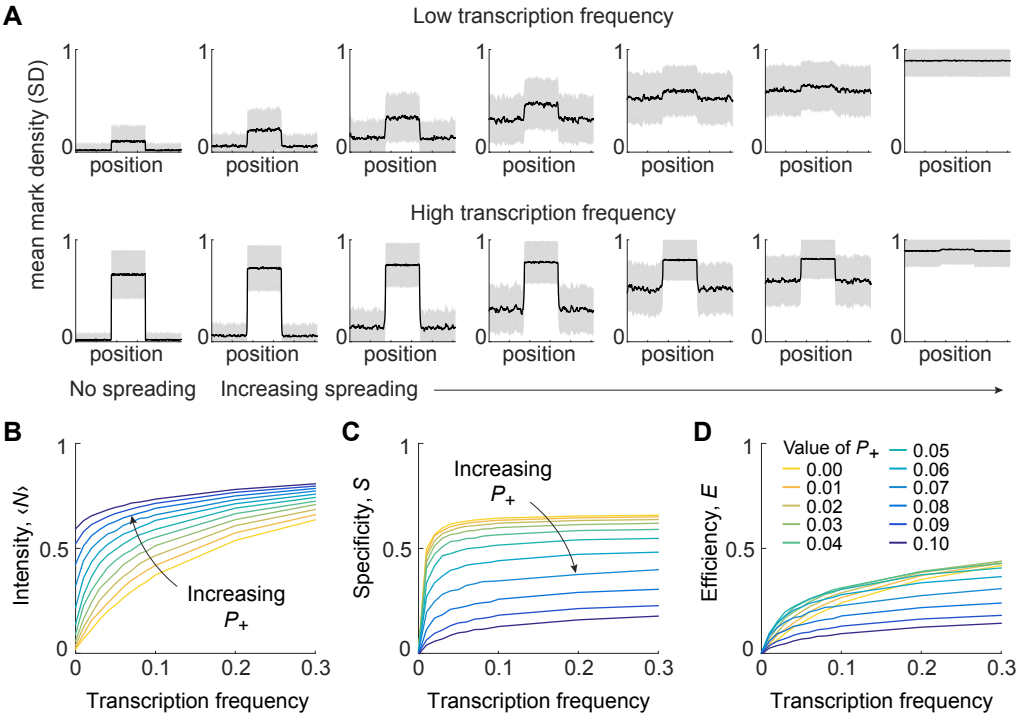
bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2



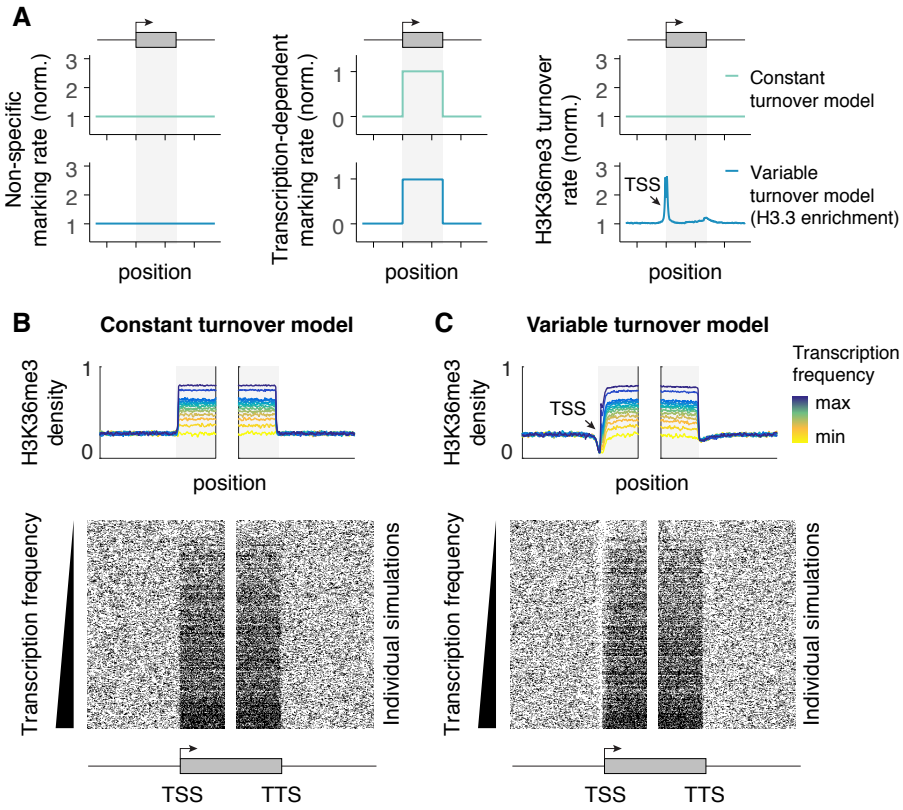
bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 3



bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 4



bioRxiv preprint doi: https://doi.org/10.1101/636084; this version posted May 14, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

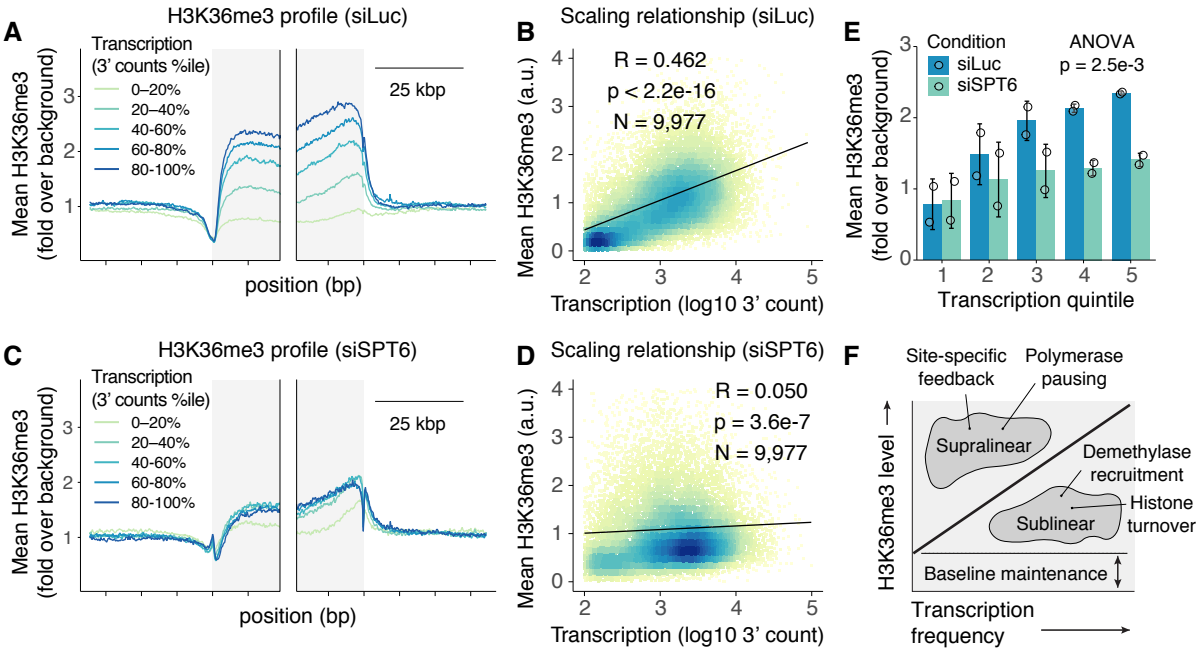



Figure 5