

H3K4me3 Breadth Is Linked to Cell Identity and Transcriptional Consistency

B erence A. Benayoun,^{1,2,7} Elizabeth A. Pollina,^{1,3,7} Duygu Ucar,^{1,6,7} Salah Mahmoudi,¹ Kalpana Karra,¹ Edith D. Wong,¹ Keerthana Devarajan,¹ Aaron C. Daugherty,¹ Anshul B. Kundaje,¹ Elena Mancini,¹ Benjamin C. Hitz,¹ Rakhi Gupta,¹ Thomas A. Rando,^{2,4,5} Julie C. Baker,¹ Michael P. Snyder,¹ J. Michael Cherry,¹ and Anne Brunet^{1,2,3,*}

¹Department of Genetics, Stanford University, Stanford CA 94305, USA

²Paul F. Glenn Laboratories for the Biology of Aging, Stanford University, Stanford CA 94305, USA

³Cancer Biology Program, Stanford University, Stanford CA 94305, USA

⁴Department of Neurology and Neurological Sciences, Stanford University, Stanford CA 94305, USA

⁵RR&D REAP, VA Palo Alto Health Care Systems, Palo Alto, CA 94304, USA

⁶Present address: Jackson Laboratory for Genomic Medicine, Farmington, CT 06030, USA

⁷Co-first author

*Correspondence: anne.brunet@stanford.edu

<http://dx.doi.org/10.1016/j.cell.2014.06.027>

SUMMARY

Trimethylation of histone H3 at lysine 4 (H3K4me3) is a chromatin modification known to mark the transcription start sites of active genes. Here, we show that H3K4me3 domains that spread more broadly over genes in a given cell type preferentially mark genes that are essential for the identity and function of that cell type. Using the broadest H3K4me3 domains as a discovery tool in neural progenitor cells, we identify novel regulators of these cells. Machine learning models reveal that the broadest H3K4me3 domains represent a distinct entity, characterized by increased marks of elongation. The broadest H3K4me3 domains also have more paused polymerase at their promoters, suggesting a unique transcriptional output. Indeed, genes marked by the broadest H3K4me3 domains exhibit enhanced transcriptional consistency rather than increased transcriptional levels, and perturbation of H3K4me3 breadth leads to changes in transcriptional consistency. Thus, H3K4me3 breadth contains information that could ensure transcriptional precision at key cell identity/function genes.

INTRODUCTION

Diverse cell types within multicellular organisms are characterized by specific transcriptional profiles. Chromatin states influence some aspects of transcription, such as expression levels or alternative splicing, and may play a role in the establishment and maintenance of gene expression programs (Bernstein et al., 2005, 2012). For example, subtypes of enhancers direct the high expression of cell identity genes (Parker et al., 2013; Rada-Iglesias et al., 2011; Whyte et al., 2013). Whether other aspects of transcription are linked to cell identity and can be predicted by chromatin states is unknown.

Trimethylation of histone H3 lysine 4 (H3K4me3) is a major chromatin modification in eukaryotes (Santos-Rosa et al., 2002; Strahl et al., 1999). Modifiers of H3K4me3 play roles in fundamental biological processes, including embryonic development (Ingham, 1998) and stem cell biology (Ang et al., 2011; Schmitz et al., 2011). Perturbations in H3K4me3-modifying complexes lead to cancer in mammals (Shilatifard, 2012) and lifespan changes in invertebrates (Greer et al., 2010; Siebold et al., 2010). The H3K4me3 modification is associated with the promoters of actively transcribed genes (Barski et al., 2007; Guenther et al., 2007; Santos-Rosa et al., 2002) and is thought to serve as a transcriptional on/off switch (Dong et al., 2012). However, H3K4me3 can also mark poised genes (Bernstein et al., 2006), and transcription can occur in the absence of H3K4me3 (H odl and Basler, 2012). Thus, how this mark affects specific transcriptional outputs to influence diverse cellular functions is still largely unclear.

Important information regarding specific transcriptional outputs could be contained in the spread of epigenetic modifications over a genomic locus. Repressive chromatin marks, such as H3K9me3, are deposited over broad genomic regions (~megabases) (Shah et al., 2013; Soufi et al., 2012; Zhu et al., 2013). Active chromatin marks are usually restricted to specific genomic loci but have also been observed in broader deposits (~kilobases) (Parker et al., 2013). For example, broad deposits of H3K4me3 have been reported in embryonic stem cells (ESCs), Wilms tumor cells, hematopoietic stem cells, and hair follicle stem cells at some key regulators in these cells (Adli et al., 2010; Aiden et al., 2010; Lien et al., 2011). However, the overall biological significance of H3K4me3 breadth is unexplored.

Here, we performed a meta-analysis of the H3K4me3 mark, which revealed that extremely broad H3K4me3 domains in one cell type mark cell identity/function genes in that cell type across species. Using the broadest H3K4me3 domains, we discovered novel regulators of neural progenitor cells and propose that these domains could be used to identify regulators of a particular cell type. Remarkably, genes marked by the broadest H3K4me3 domains showed increased transcriptional consistency (i.e., low

transcriptional variability), and perturbation of H3K4me3 breadth led to changes in transcriptional consistency. Our study identifies a new chromatin signature linked to transcriptional consistency and cell identity and highlights that breadth is a key component of chromatin states.

RESULTS

Broad H3K4me3 Domains Mark Subsets of Genes in All Organisms but Do Not Predict Expression Levels

To investigate the importance of H3K4me3 breadth, we analyzed the landscape of H3K4me3 domains in >200 data sets of H3K4me3 chromatin immunoprecipitation followed by sequencing (ChIP-seq) or microarray hybridization (ChIP-chip) in stem, differentiated, or cancer cells from nine species (Table S1 available online). Consistent with previous reports, H3K4me3 was mostly present in 1–2 kb regions around transcription start sites (TSSs) (Figures 1A–1C). However, as previously noted in mammalian stem cells (Adli et al., 2010; Aiden et al., 2010; Lien et al., 2011), broader domains of H3K4me3 spanning up to 60 kb were present in all cell types and organisms (Figures 1A–1C and S1A). Broad H3K4me3 domains were mostly found close to genes, extending both 5' and 3' of TSSs (Figure 1C). The genes marked by these regions were different between cell types (Figure S1B). Broad H3K4me3 domains were not associated with higher H3K4me3 ChIP intensities (Figure S1C and S1D) and were observed regardless of sequencing depth, method of chromatin fragmentation, peak-calling algorithm (Figure S1E), or H3K4me3 antibody.

We asked whether broader H3K4me3 domains may be explained by underlying promoter or gene structure or by expression levels. Broader H3K4me3 regions did not mark gene cluster regions (Figure S1F). There was no correlation between H3K4me3 domain breadth and gene length (Figure S1G) nor with the numbers of used TSSs (e.g., alternative promoters) (Figures S1H and S1I). H3K4me3 breadth did not correlate with mRNA levels (Figures 1D and 1E), even when comparing the most extreme examples (the top 5% broadest H3K4me3 domains) to the rest of the breadth distribution (Figure S1J). Thus, broad H3K4me3 domains are present in many cell types across taxa but cannot be explained as simple readouts of promoter complexity or high expression levels. These observations prompted us to investigate the biological relevance of broad H3K4me3 domains.

Broad H3K4me3 Domains Preferentially Mark Cell Identity and Function Genes in a Given Cell Type

To assess whether broad H3K4me3 domains mark specific gene sets, we analyzed H3K4me3 ChIP-seq data sets from >20 different cell and tissue types in mice and humans. When compared to all H3K4me3 domains, the top 5% broadest H3K4me3 domains in a particular cell/tissue type enriched for annotations linked to the “function” (e.g., specialized cytoskeleton for contractile cells) and the “identity” (e.g., factors required to establish that cell lineage) (Figures S2A and S2B). In human embryonic stem cells (hESCs), the set of 5% broadest H3K4me3 domains enriched the most, along the H3K4me3 breadth continuum, for validated embryonic stem cell regulators

(Figures 2A and S2C and Table S2). In contrast, the broadest domains of other histone marks (e.g., H3K27ac) or the top 5% most “intense” H3K4me3 domains (i.e., with the highest ChIP-seq signal normalized to peak breadth) did not strongly enrich for stem cell regulators (Figure 2A). More generally, along the breadth continuum, the top 5% broadest H3K4me3 domains in a given cell/tissue type most enriched for genes with important functions for that particular cell/tissue type across cell types and taxa (Figure 2B).

We next asked whether the broadest H3K4me3 domains could separate cells or tissues by lineage better than the complete set of H3K4me3 domains (Zhu et al., 2013). The top 5% H3K4me3 broadest domains indeed discriminated cells or tissues according to their lineage in human and mouse (Figures 2C and S2D). Clustering quality measures showed that the set of the 5% broadest H3K4me3 domains outperformed all other 5% subsets of H3K4me3 domains (binned by breadth or intensity) as well as the complete set of H3K4me3 domains (Figures 2D and S2E–S2G). The broadest H3K4me3 domains could also distinguish fibroblasts from fully and even partially reprogrammed iPSCs (Figures S2H and S2I). Reciprocally, genes encoding factors known to be critical for cell identity/fate (e.g., *Myod1* in skeletal muscle) had significantly broader H3K4me3 domains in the relevant tissue than expected by chance ($p < 2.22 \times 10^{-308}$ in Kolmogorov-Smirnov test; Figure 2E and Table S3).

A prediction from these observations is that a subset of the broadest H3K4me3 domains should be remodeled as cells differentiate along a lineage. Indeed, some of the top 5% broadest H3K4me3 domains are remodeled between progenitors and derived progeny (e.g., adipocyte progenitors and adipocytes) (Figure 2F). Genes that gained the top 5% broadest H3K4me3 domains during differentiation were enriched for differentiated cell functions (e.g., lipid homeostasis), whereas genes that lost the top 5% broadest H3K4me3 domains were enriched for progenitor cell functions (e.g., cell proliferation) (Figure S2J). A subset of genes was marked by broad H3K4me3 domains in most cells (e.g., *Foxo3*), and those marked genes tended to be shared between mouse and human ($p < 1.4 \times 10^{-75}$ in Fisher's exact test), perhaps representing genes that are key for basic cell function across tissues.

Next, we compared the top 5% broadest H3K4me3 domains to another signature linked to cell identity, “super-enhancers” (Whyte et al., 2013). Genes marked by the top 5% broadest H3K4me3 domains were distinct from genes assigned to super-enhancers (Figure S2K). The top 5% broadest H3K4me3 domains and super-enhancers were similarly effective at predicting validated stem cell regulators in mESCs (Figure S2L), but the broadest H3K4me3 domains discriminated better between lineages (Figure S2M). Additionally, the breadth of H3K4me3 domains provides a ranking system to further enrich for key regulators (Figure S2L). Thus, the top 5% broadest H3K4me3 domains have the potential to identify genes with functions in a given cell type or lineage.

Broad H3K4me3 Domains as a Discovery Tool for Novel Regulators of Neural Progenitor Cells

We tested whether the top 5% broadest H3K4me3 domains could be used to discover novel regulators in a particular cell

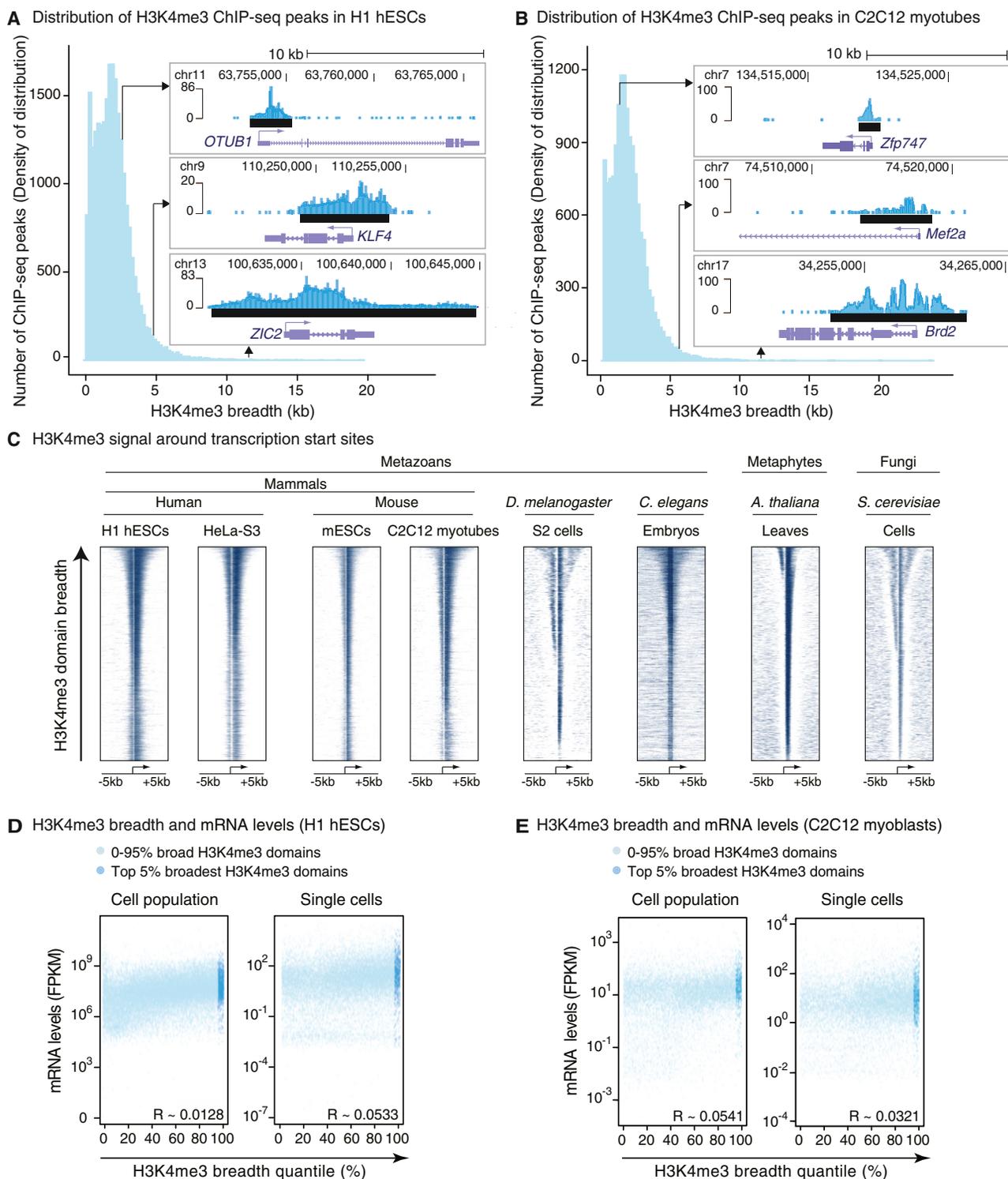


Figure 1. H3K4me3 Breadth Is an Evolutionarily Conserved Feature that Is Not Predictive of Expression Levels

(A and B) Breadth distributions of H3K4me3 ChIP-seq peaks in H1 hESCs (A) and C2C12-derived myotubes (B) display “heavy right tails,” indicative of broader H3K4me3 domains than expected. (Inserts) Example H3K4me3 regions in H1 hESCs or C2C12 myotubes; (black bar) ChIP-seq peaks called by MACS2.

(C) H3K4me3 ChIP signal sorted by breadth at -5 kb, $+5$ kb around transcription start sites (TSSs).

(D and E) mRNA levels are not a function of H3K4me3 breadth quantile at the population (left) or single-cell (right) level by RNA-seq in H1 hESCs (D) or C2C12 myoblasts (E). (Insert) Pearson correlation coefficients.

See also Figure S1J.

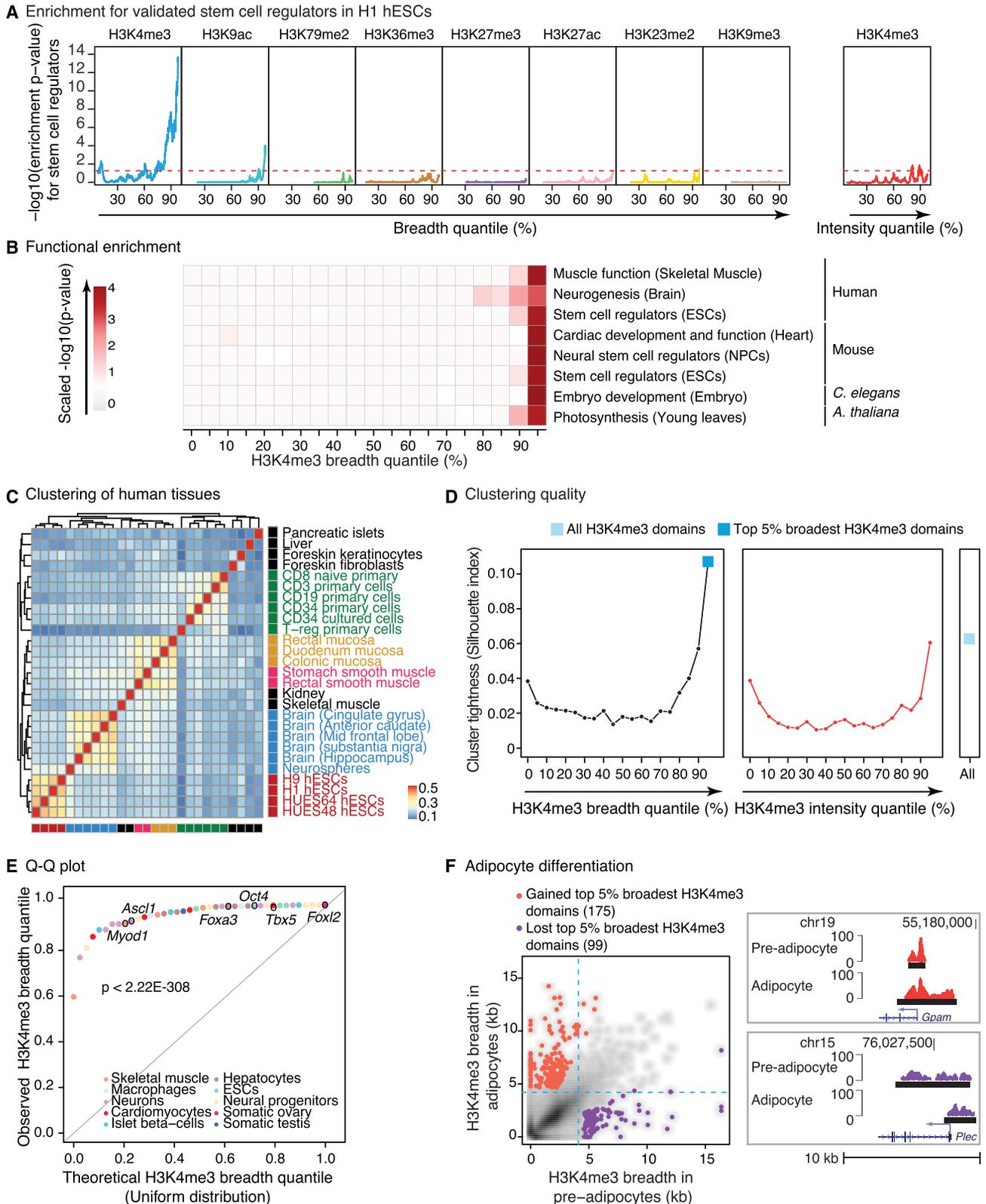


Figure 2. H3K4me3 Breadth Enriches for Genes that Are Important for Cell Identity and Function

(A) The top 5% broadest H3K4me3 domains enrich for stem cell regulators in H1 hESCs. Enrichment expressed as $-\log_{10}$ (p value) in Fisher's exact test. (Red dashed line) $p = 0.05$. See also Figure S2C.

(legend continued on next page)

type. We chose adult neural progenitor cells (NPCs) because these cells give rise to all brain cell types and may be a source of cells for regenerative therapies (Bonaguidi et al., 2011; Doetsch et al., 1999; Lujan et al., 2012). We generated H3K4me3 ChIP-seq data sets from NPC cultures isolated from adult mice and from the microdissected brain region that hosts these cells in vivo (Figures 3A–3C and S3A).

As seen in other cell types, genes marked by the top 5% broadest H3K4me3 domains in adult NPCs were enriched for known regulators based on a literature-curated list (Figures 2B, 3C, and S3B and Table S2). A large fraction of broad H3K4me3 domains in NPCs mark genes encoding transcription factors, noncoding RNAs, or both (Figure 3D). Interestingly, the broadest H3K4me3 domains in NPCs marked genes that had not been previously implicated in the maintenance of the neuronal lineage (e.g., *Bahcc1*, *Fam72a*, *2610017109Rik*; Figure 3D and Table S4) and genes that, while involved in brain development, have not been studied in adult/postnatal NPCs (e.g., *Otx1*) (Sakurai et al., 2010).

Using a lentiviral-based RNA interference (shRNA) approach, we knocked down candidate genes marked by the broadest H3K4me3 domains in primary NPCs, and we quantified the ability of NPCs to proliferate and generate new neurons (neurogenesis) (Figures 3E and S3C). Knockdown of known regulators *p53* and *Sox2* had the expected effect on NPC proliferation and neurogenesis (Figures 3F, 3G, and 3I) (Wang et al., 2011). Knocking down two-thirds of genes marked by broad H3K4me3 domains (12/18) significantly decreased NPC proliferation (Figures 3F, 3G, S3D, and S3E), whereas knockdown of control genes (i.e., genes with shorter or no H3K4me3 domains) did not significantly affect NPC proliferation. Knockdown of a subset of genes marked by the broadest H3K4me3 domains also decreased neurogenesis, suggesting that they are necessary for the differentiation ability of NPCs (Figures 3H, 3I, and S3F). These include the genes encoding the transcription factor *SALL1*, the homeobox transcription factor *OTX1*, and *BAHCC1*, a protein with a bromo-adjacent homology domain that can bind acetylated histones (Kuo et al., 2012). In contrast, knockdown of *Fam72a*, a gene misregulated in an Alzheimer's mouse model (Nehar et al., 2009), led to an increase in neurogenesis, suggesting that *FAM72A* may function to restrain differentiation in the progenitor state (Figures 3H, 3I, and S3F). Thus, this targeted screen allowed us to identify previously uncharacterized genes involved in NPC self-renewal and/or neurogenesis and to confirm that the top 5% broadest H3K4me3 domains can be used to discover genes that regulate the biology of a given cell type. To facilitate the use of H3K4me3 breadth as a discovery tool in other systems, we created a searchable data-

base accessible at <http://bddb.stanford.edu> (Figure S3G and Table S5).

The Top 5% Broadest H3K4me3 Domains Represent a Distinct Subclass of H3K4me3 Domains

Considering their association with biological function, we asked whether the top 5% broadest H3K4me3 domains constitute a distinct entity with specific (epi-)genomic characteristics. We used four different classification algorithms—a family of machine-learning algorithms that can learn to separate entities based on discriminating features. We built classification models based on the co-occurrence of protein binding profiles and chromatin modifications with H3K4me3 domains in 13 cell types and organisms (Figures 4A, S4A, Table S1, and Extended Experimental Procedures). All four algorithms were able to discriminate with high accuracy (>75%) the top 5% broadest H3K4me3 domains apart from random sets of the same number of H3K4me3 domains from the rest of the distribution (Figures 4B and S4B–S4D). In contrast, models built from the top 5% most intense H3K4me3 peaks lacked discriminative power (Figures 4B and S4D). The discriminative power of the classification models remained high when the top 5% H3K4me3 broadest domains were compared to domains up to the 85th percentile of the breadth distribution and then sharply dropped (Figure 4C). The presence of this inflexion point supports the notion that a specific set of features distinguishes the broadest H3K4me3 domains from the rest of the distribution. Reciprocally, H3K4me3 domains sharing the discriminating (epi-)genomic characteristics of the broadest H3K4me3 domains were broader than expected by chance (Figure 4D). Together, these results suggest that specific combinations of histone marks and protein binding distinguish the top 5% broadest H3K4me3 domains from the rest of the breadth continuum.

We next asked which (epi-)genomic characteristics most contributed to the ability of models to accurately discriminate the top 5% broadest H3K4me3 domains from the rest of the distribution (Figure S4E). Consistent with the tissue-specific genomic deposition pattern of H3K4me3 breadth, we found that lineage-specific transcription factors (TFs), such as *MYOD*/*MYOG* in myotubes, were recurrent important contributors to the models (Figure 4B). The enrichment of lineage-specific TFs at broad H3K4me3 domains was observed in multiple other cell types (Figure S4G). Other important recurrent contributors included the transcription initiation complex (e.g., TAFs and RNA polymerase II [Pol II]), marks of transcriptional elongation (e.g., H3K79me2), the H3K4me3 reader *CHD1*, which is involved in nucleosome repositioning at transcribed genes (Smolle et al., 2012), and the transcriptional repressor *SIN3A*, which constrains

(B) The top 5% broadest H3K4me3 domains enrich for genes involved in cell/tissue function. Significance as scaled $-\log_{10}$ (p value) in Fisher's exact test (see Extended Experimental Procedures and Table S2). The NPC data set is described in Figures 3A–3C.

(C) Hierarchical clustering of the top 5% broadest H3K4me3 domains from human tissues and cells based on Jaccard Index similarity. See also Figure S2D.

(D) Measure of cluster tightness (Silhouette index) from different sets of H3K4me3 domains in human tissues. See also Figures S2E–S2G.

(E) Quantile-by-quantile (Q-Q) plot of the quantile ranks of H3K4me3 domains marking known cell identity or reprogramming genes in tissue of relevance (Table S3). Significance in Kolmogorov-Smirnov test.

(F) H3K4me3 breadth is remodeled at a subset of loci during differentiation. (Left) Scatterplots of H3K4me3 breadth for adipogenesis (*3T3L1* in pre- versus mature adipocyte). (Right) Remodeled top 5% broadest H3K4me3 domains between pre- and mature adipocytes.

See also Figure S2J.

spurious transcription in yeast (Carrozza et al., 2005) (Figures 4B, 4E, 4F, S4E, and S4F). These contributors point toward a link between extreme H3K4me3 breadth and specific regulation of transcriptional initiation and elongation—a surprising finding given that H3K4me3 breadth is not correlated with expression level (Figures 1E and S1H) and that the models were built to discriminate the top 5% broadest H3K4me3 domains from the rest of H3K4me3 domains, all of which are thought to be transcriptionally active.

The Top 5% Broadest H3K4me3 Domains Exhibit Unique Regulation of Pol II Pausing and Elongation

We examined whether the top 5% broadest H3K4me3 domains had unique features of transcriptional regulation. Increasing H3K4me3 domain breadth was associated with enriched binding of positive regulators of transcription elongation (e.g., super elongation complex, P-TEFb) ($p < 1 \times 10^{-4}$ in permutation test; Figures 5A, 5B, and S5A–S5E) (Luo et al., 2012; Sims et al., 2004). The top 5% broadest H3K4me3 domains were also the most likely to be bound by Pol II (Figures 5A, 5B, and S5B–S5E) despite similar numbers of Pol-II-binding sites (Figure S1H). The top 5% broadest H3K4me3 domains are associated with overall higher levels of Pol II ($p < 4.7 \times 10^{-2}$ in one-sample Wilcoxon tests; Figures 5D and S5G). Serine-2-phosphorylated Pol II (Ser2P), which is characteristic of productive elongation (Sims et al., 2004), was significantly higher at genes marked by the top 5% broadest H3K4me3 domains ($p < 8.0 \times 10^{-3}$ in one-sample Wilcoxon tests; Figures 5E and 5F). Somewhat surprisingly, serine-5-phosphorylated Pol II (Ser5P), which is associated with transcriptional initiation, was also increased at genes marked by top 5% broadest H3K4me3 domains ($p < 8.0 \times 10^{-3}$ in one-sample Wilcoxon tests; Figures S5H and S5I). This observation, coupled with the association of the top 5% broadest H3K4me3 domains to factors that promote Pol II pausing (e.g., NELFA) (Sims et al., 2004) (Figures 5A, 5B, and S5A), suggests that broad H3K4me3 domains are associated with Pol II pausing. Indeed, genes marked by the top 5% broadest H3K4me3 domains had significantly higher traveling ratios, a measure of Pol II pausing ($p < 2.9 \times 10^{-2}$ in one-sample Wilcoxon tests; Figures 5G, 5H, and S5J). Though it may seem contradictory that the same class of domains would associate with increased elongation and Pol II pausing, this phenomenon may represent a foot-

print of steady Pol II release from a heavily preloaded promoter. Increased Pol II pausing at proximal promoters has been suggested to promote chromatin accessibility (Gilchrist et al., 2010). Indeed, promoters of genes coated by the top 5% broadest H3K4me3 domains were more accessible than promoters marked by shorter H3K4me3 domains, as assessed by DNase-seq ($p < 4.2 \times 10^{-20}$ in one-sample Wilcoxon tests; Figures 5I and S5K) or ATAC-seq (Buenrostro et al., 2013) ($p = 4.9 \times 10^{-156}$ in one-sample Wilcoxon tests; Figure S5L). Thus, the top 5% broadest H3K4me3 domains are associated with unique regulation of Pol II initiation and elongation and increased chromatin accessibility, suggesting a specific transcriptional output.

The Top 5% Broadest H3K4me3 Domains Are Associated with Increased Transcriptional Consistency

Increased Pol II pausing and elongation have been linked to transcriptional consistency (e.g., lower transcriptional variability, or “transcriptional noise,” in single cells) (Boettiger et al., 2011; Lagha et al., 2013). In addition, increased chromatin accessibility at promoters may facilitate transcriptional consistency, in part by minimizing transcriptional bursting (Field et al., 2008). We asked whether the top 5% broadest H3K4me3 domains are associated with transcriptional consistency (Figure 6A). To measure transcriptional consistency in single cells, we calculated the variance in expression relative to expression level for each gene in single-cell RNA-seq data sets (Table S6). Remarkably, genes marked by the top 5% broadest H3K4me3 domains had reduced transcriptional variability in single-cell RNA-seq experiments ($p < 4.6 \times 10^{-26}$ in one-sample Wilcoxon tests; Figure 6B). Transcriptional consistency significantly increased with H3K4me3 breadth, but not with peak intensity (Figure S6A).

To expand our analysis, we tested transcriptional consistency in data sets generated from cell populations. A gene with stable transcription levels will tend to have consistent expression values between biological replicates of cell populations despite microvariations in the environment (Figure 6A) (Dong et al., 2011). Similar to our observations at the single-cell level, analysis of 15 different cell population transcriptome data sets (Table S6) showed that genes marked by the top 5% broadest H3K4me3 domains had reduced transcriptional variability ($p < 4 \times 10^{-3}$

Figure 3. The Top 5% Broadest H3K4me3 Domains Can Be Used as a Discovery Tool to Identify New Regulators of Neural Progenitor Cells

- (A) Experimental design for H3K4me3 ChIP-seq data sets in primary cultures of neural progenitors (NPCs) and microdissected niche (subventricular zone).
 (B) H3K4me3 ChIP-seq peaks at a known NPC regulator in independent NPC primary cultures and in the NPC niche. (Black bars) ChIP-seq peaks called by MACS2.
 (C) Distribution of H3K4me3 ChIP-seq peaks as a function of their breadth in NPCs reveals that known NPC regulators are marked by broad H3K4me3 domains.
 (D) Genes associated to the top 30 broadest H3K4me3 domains in NPCs. Domains ranked by decreasing H3K4me3 breadth. Known regulators of NPCs in bold. (RP) rank of rank product.
 (E) Experimental design to test the role of genes marked by the top 5% broadest H3K4me3 domains in NPC proliferation and neurogenesis.
 (F) Proliferation capacity as normalized MTT optical density relative to control. Mean + SD of two independent experiments conducted in triplicate with two or more independent hairpins. (Hashed blue bars) Genes whose role in NPCs was discovered while this study was in preparation (Agoston et al., 2014; Ninkovic et al., 2013). See also Figure S3D.
 (G) Proliferation capacity as percentage of infected cells that formed primary neurospheres relative to control. Mean + SD of at least two independent experiments conducted in triplicate.
 (H) Images of new neurons upon *Fam72a* knockdown. (Green) TUJ1, neurons; (blue) DAPI, nuclei.
 (I) Neurogenesis measured by percentage of DCX+ cells (new neurons) normalized to control.

Mean + SEM of at least two independent experiments conducted in triplicate with two or more independent hairpins. See also Figure S3F. In all relevant panels: n.s., not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.005$ in a Wilcoxon test against control with Bonferroni correction for multiple testing.

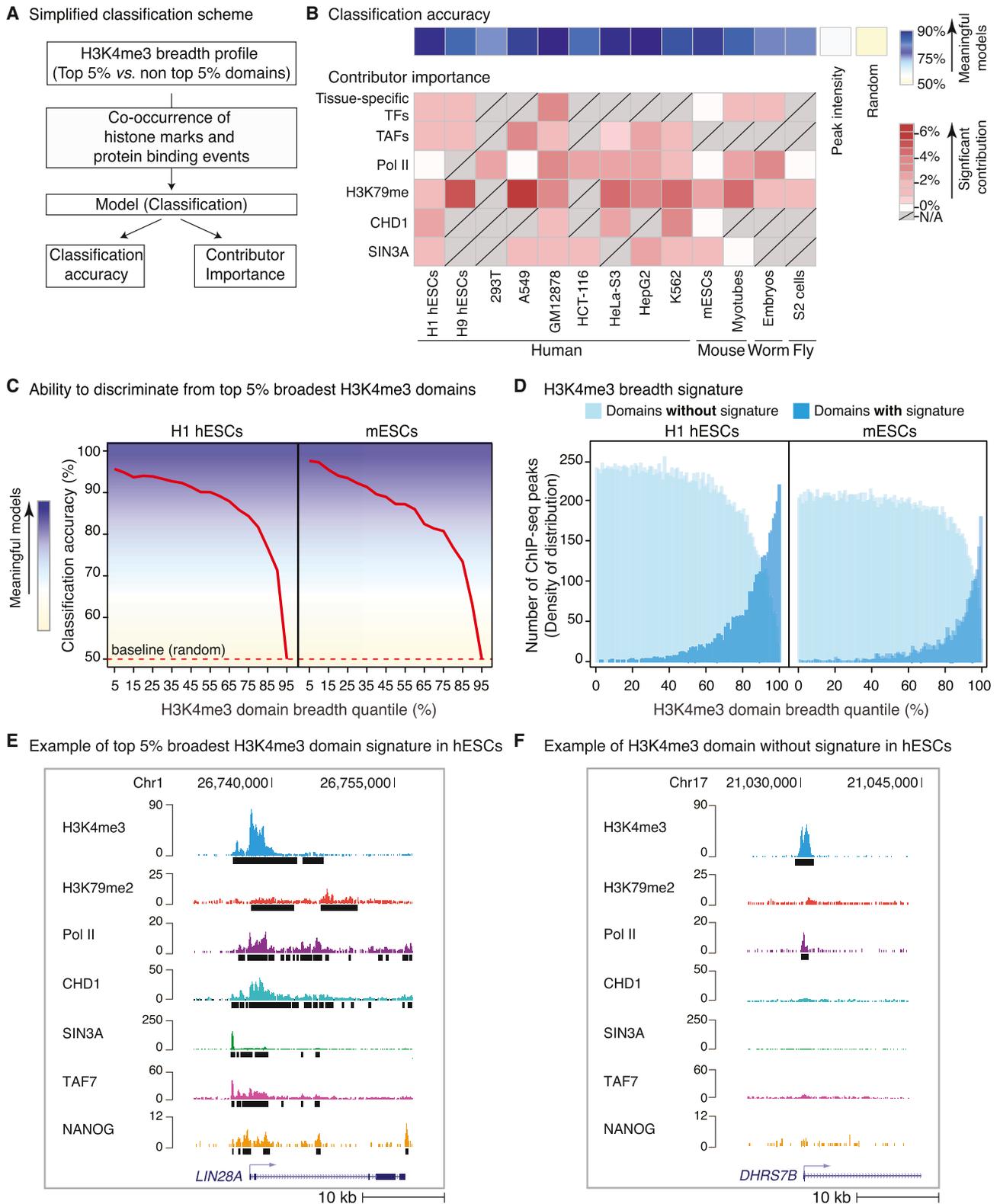


Figure 4. The Broadest H3K4me3 Domains Are Characterized by a Specific Epigenomic Signature

(A) Simplified scheme of computational modeling. See also Figure S4A.

(B) Average classification accuracy and most important contributors associated to the top 5% broadest H3K4me3 domains identified by Random Forest models in 13 cell types and organisms. Contributors for which no data was available are shown in gray with diagonal lines. Tissue-specific transcription factors (TFs) refer (legend continued on next page)

in one-sample Wilcoxon tests; [Figures 6C](#) and [S6B–S6D](#)). The broadest H3K4me3 domains were also associated with increased transcriptional consistency when considering only nascent RNA levels in cell populations by GRO-seq (general run-on sequencing) ($p < 6.6 \times 10^{-94}$ in one-sample Wilcoxon tests; [Figures 6D](#), [S6E](#) and [Table S6](#)), indicating that transcriptional consistency at the broadest H3K4me3 domains does not result from increased mRNA stability. By performing RNA-seq, we confirmed that genes marked by the top 5% broadest H3K4me3 domains in primary NPC cultures were more transcriptionally consistent in these cells ($p = 1.2 \times 10^{-130}$ in a one-sample Wilcoxon test; [Figures 6E](#), [6F](#) and [7E](#)). Transcriptional consistency increased as a function of H3K4me3 domain breadth in NPCs ([Figure 6G](#)). Together, our results indicate that the broadest H3K4me3 domains are linked to transcriptional consistency.

Perturbation of H3K4me3 Breadth Leads to Changes in Transcriptional Consistency

We next tested whether perturbation of H3K4me3 breadth could result in changes in transcriptional consistency. H3K4me3 is deposited by members of the COMPASS/Trithorax/Trithorax-related family of methyltransferase complexes and is removed by the JARID1 family of demethylases ([Black et al., 2012](#)). Components of the H3K4me3 regulatory machinery were enriched to bind loci marked by the broadest H3K4me3 domains ([Figure S7A](#)), suggesting that these regions may require increased presence of regulators for their deposition/maintenance.

We used the adult NPC model to test whether decreasing H3K4me3 breadth could affect transcriptional consistency ([Figure 7A](#)). Because WDR5 is an essential scaffolding subunit shared by all COMPASS/Trithorax-like complexes ([Trievel and Shilatifard, 2009](#)), we reasoned that its knockdown might be more efficient at reducing H3K4me3 breadth than that of individual methyltransferases. We tested the effect of short-term (24 hr) *Wdr5* knockdown on H3K4me3 levels and breadth in NPCs ([Figure 7A](#)). By 24 hr of knockdown, *Wdr5* mRNA and WDR5 protein levels were reduced ([Figures 7B](#), [S7B](#) and [S7C](#)) without major consequences on cell viability ([Figure S7D](#)). Short-term WDR5 depletion led to a global reduction of H3K4me3 levels in NPCs but did not significantly affect methylation of other histone residues ([Figure 7B](#)). In line with the global decrease of H3K4me3, there was a decrease in the signal-to-noise ratio of H3K4me3 ChIP-seq from *Wdr5* shRNA-treated cells compared to control cells ([Figures S7E](#) and [S7F](#)). After accounting for the overall loss of H3K4me3 ChIP-seq intensity ([Extended Experimental Procedures](#), [Figure S7E](#), and [S7F](#)), we observed that H3K4me3 breadth decreased genome wide in response to WDR5 depletion ([Figures 7C](#) and [S7G](#)). The mean proportion of H3K4me3 domains maintaining or losing breadth was fairly con-

stant above the 30th percentile of breadth ([Figure S7G](#)). Thus, a short-term *Wdr5* knockdown provides a way to compare genes at which H3K4me3 breadth is reduced or maintained ([Figure 7C](#)). To evaluate the impact of H3K4me3 breadth reduction on transcriptional consistency, we generated RNA-seq data sets from replicates of NPC cultures infected with control or *Wdr5* shRNA ([Figures 7A](#) and [S7I](#)). In line with our results in uninfected NPCs ([Figure 6G](#)), H3K4me3 breadth enriched for increased transcriptional consistency in control-infected NPCs ([Figure 7E](#), top). Upon short-term *Wdr5* knockdown, only a few genes were differentially expressed ([Figure S7H](#)). In contrast, genes with reduced H3K4me3 breadth (<50% of original breadth) upon *Wdr5* knockdown exhibited increased transcriptional variability compared to genes with maintained breadth ([Figure 7D](#)). Reduction of H3K4me3 breadth was most significantly associated with loss of transcriptional consistency at the top 5% broadest H3K4me3 domains ([Figures 7D](#) and [7E](#), bottom). Thus, reduction of H3K4me3 breadth is associated with a decrease in transcriptional consistency.

To test the effect of H3K4me3 breadth extension on transcriptional consistency, we used H3K4me3 ChIP-seq data sets and expression microarrays obtained in mESCs following short-term (48 hr) knockdown of the H3K4me3 demethylase *Jarid1b/Kdm5b* ([Schmitz et al., 2011](#)). Depletion of JARID1B induces a substantial increase in H3K4me3 levels in mESCs ([Schmitz et al., 2011](#)). After accounting for changes in H3K4me3 ChIP-seq intensity, we observed that H3K4me3 breadth tended to increase genome wide in response to *Jarid1b* knockdown ([Figures 7F](#), [S7K](#), and [S7L](#)). Few genes were differentially expressed after short-term knockdown of *Jarid1b* ([Figures S7K](#) and [S7M](#)). However, genes that gained H3K4me3 breadth (>2× their original breadth) upon *Jarid1b* knockdown had a significant decrease in transcriptional variability compared against those that maintained H3K4me3 breadth ([Figures 7G](#) and [S7L](#)). There was no further gain of transcriptional consistency upon *Jarid1b* knockdown for genes marked by the top 5% broadest domains ([Figure 7G](#)), perhaps because genes marked by these domains have already reached their lowest level of variability. The decreased transcriptional variability associated with the extension of H3K4me3 breadth in mESCs mirrors our findings that reduced H3K4me3 breadth leads to increased transcriptional variability in NPCs. Together, these results are consistent with the idea that the broadest H3K4me3 domains may promote transcriptional consistency at key cell identity/function genes ([Figure 7H](#)).

DISCUSSION

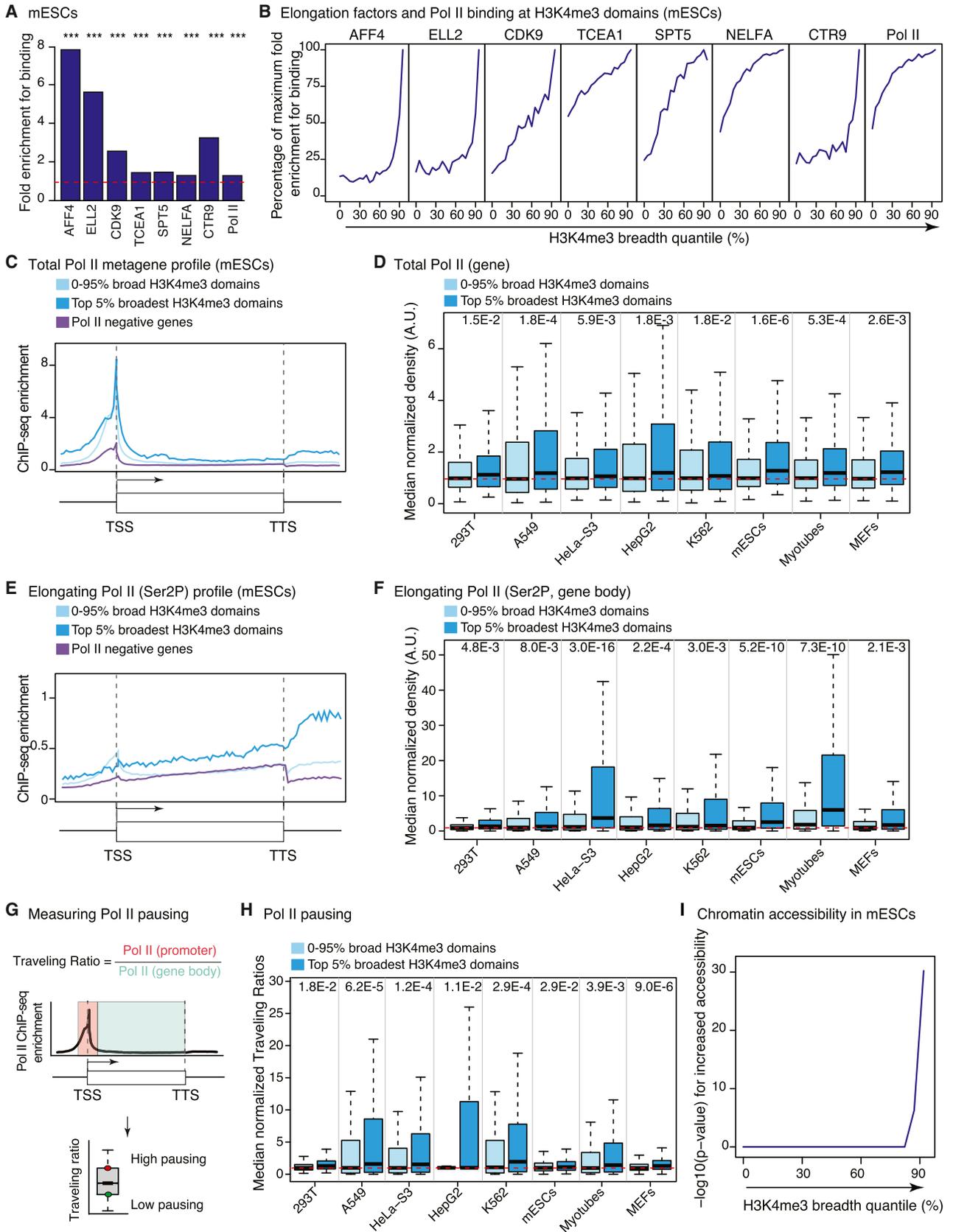
Through meta-analysis of high-throughput genomics data, construction of machine-learning models, and experimental

to: NANOG (H1 hESCs), SMAD2/3 (H9 hESCs), STAT5 (GM12878), NANOG (mESCs), MYOG/MYOD (myotubes), and LIN-13 (*C. elegans* embryos). See also [Figures S4B–S4E](#).

(C) Accuracy of progressive classifications in H1 hESCs and mESCs. Classifications performed between the top 5% broadest H3K4me3 domains and other 5% quantile subsets along the breadth continuum. The accuracy of progressive classifications reflects the ability to discriminate domains of that quantile from the top 5% broadest H3K4me3 domains.

(D) Breadth of H3K4me3 domains “with/without the top 5% broadest H3K4me3 domain signature” in H1 hESCs and in mESCs.

(E and F) Example domains with/without signature in H1 hESCs. (Black bars) Peaks called by MACS2.



(legend on next page)

validation of predictions, we uncover the existence of a subclass of H3K4me3 domains that preferentially marks genes that are important for cell identity and function. Our study also identifies for the first time a genome-wide link between a specific chromatin landscape, H3K4me3 breadth, and low transcriptional variability. We propose to refer to this subclass of broad H3K4me3 domains as “buffer domains,” as they may help to prevent spurious bursts of transcription.

Buffer Domains as a Cell Identity Signature and Discovery Tool

Several epigenetic signatures have been linked to cell identity or developmental genes: active enhancers (regions of enrichment for H3K4me1, H3K27ac, and p300 binding) (Heintzman et al., 2009; Rada-Iglesias et al., 2011), super-enhancers (regions intensely bound by mediator and cell-specific transcription factors) (Whyte et al., 2013), stretch enhancers (extended enhancer regions) (Parker et al., 2013), and DNA methylation canyons or valleys (large hypomethylated regions) (Jeong et al., 2014; Xie et al., 2013). Although there is some overlap in the genes marked by buffer domains and other signatures, many genes are uniquely captured by each of these signatures (Figure S2K; B.A.B., D.U., and A.B., unpublished data). Thus, complementary strategies of transcriptional regulation may control cell identity, with buffer domains tuning transcriptional consistency and super/stretch enhancers promoting high expression (Parker et al., 2013; Whyte et al., 2013).

Here, we test the potential of the buffer domain signature to assist in the discovery of new cell identity/function genes in NPCs. We identify new regulators of NPC proliferation and neurogenesis, such as the putative chromatin reader *Bahcc1* or the noncoding RNA *2610017109Rik*. Although we have not assessed the function of all genes marked by buffer domains in NPCs, our study illustrates the potential of this signature to identify new regulators and annotate poorly characterized genes. As reprogramming factors tend to be marked by buffer domains, this signature may also help to find candidates for reprogramming cells into a cell type of interest, a task facilitated by our web-accessible database (<http://bddb.stanford.edu>). The predictive value of buffer domains could also be used to identify genes in contexts for which few functional genes

have been identified so far, such as aging or neurological disorders.

A Specific Transcriptional Output Linked to H3K4me3 Breadth

There is a general consensus that H3K4me3 plays a role in transcriptional initiation (Laubert et al., 2013). However, recent work in *Drosophila* suggests that H3K4me3 may be dispensable for the expression of some genes (Hödl and Basler, 2012). Here, we find that short-term manipulation of H3K4me3 breadth impacts transcriptional consistency, which suggests that H3K4me3 may ensure the robustness of transcriptional outputs. Interestingly, deficiencies in histone acetyltransferases or chromatin remodelers lead to increased transcriptional noise in yeast (Hansen and O’Shea, 2013; Raser and O’Shea, 2004; Weinberger et al., 2012), supporting the idea that chromatin states are critical for transcriptional precision.

The mechanisms leading to the deposition/maintenance of broad H3K4me3 domains and their connection with Pol II regulation and transcriptional consistency are still unknown. Subunits of H3K4me3-depositing complexes are required for Pol II priming (Pérez-Lluch et al., 2011) and may promote Pol II release into productive elongation (Ardehali et al., 2011). Conversely, PAF1, which associates with elongating Pol II, can recruit methyltransferase complexes responsible for the deposition of both H3K4me3 and H3K79me2 (Krogan et al., 2003), an elongation mark enriched at the broadest H3K4me3 domains. Based on these previous findings and our own, we propose a model in which, at some key regulatory genes, a positive feedback mechanism may link robust initiation and elongation of Pol II with sustained H3K4me3 deposition over broad regions (Figure 7H). Sustained elongation may first allow increased recruitment of H3K4me3-depositing complexes to specific genes. In turn, this recruitment would promote the broadening of H3K4me3 deposits and subsequent recruitment of the H3K4me3 reader CHD1, which facilitates the passage of elongating Pol II (Smolle et al., 2012). This loop could work toward ensuring transcriptional consistency at specific loci. How these loci get selected is unclear, though tissue-specific transcription factors, which are enriched at the broadest H3K4me3 domains, might be involved. As transcription factor

Figure 5. The Top 5% Broadest H3K4me3 Domains Are Associated with Marks of Transcriptional Elongation and Pol II Pausing

(A) Differential binding of components of the elongation machinery to top 5% versus non-top 5% broadest H3K4me3 domains in mESCs. p values from permutation test.

(B) Enrichments for components of the elongation machinery in mESCs expressed as a percentage of the maximal binding enrichment that can be observed along the H3K4me3 breadth continuum (see A for enrichment at the top 5% broadest H3K4me3 domains). See also Figures S5A–S5E.

(C) Mean ChIP-seq enrichment of total Pol II in mESCs. TSS, transcription start site; TTS, transcription termination site.

(D) Normalized Pol II ChIP-seq density over the proximal promoter and gene body. Comparisons of the top 5% broadest H3K4me3 domains against the rest of the distribution are also significant in one-sided Wilcoxon tests ($9.6 \times 10^{-10} < p < 5.4 \times 10^{-3}$) (continued in Figure S5G).

(E) Mean ChIP-seq enrichment of elongating Pol II (Ser2P) in mESCs. TSS, transcription start site; TTS, transcription termination site.

(F) Normalized elongating Pol II (Ser2P) ChIP-seq density over gene bodies. Comparisons of the top 5% broadest H3K4me3 domains against the rest of the distribution are also significant in one-sided Wilcoxon tests ($7.3 \times 10^{-23} < p < 2.6 \times 10^{-5}$) (continued in Figure S5J).

(G) Measure of Pol II pausing. Traveling ratio is defined as background-subtracted ChIP-seq density value of Pol II at the promoter versus gene body.

(H) Normalized traveling ratios. Comparisons of top 5% broadest H3K4me3 domains against the rest of the distribution that is also significant in one-sided Wilcoxon tests ($1.8 \times 10^{-9} < p < 4.7 \times 10^{-2}$) (continued in Figure S5K).

(I) Significance for increased chromatin accessibility in mESCs against expected genome-wide value shown as $-\log_{10}$ (p value) in one-sided Wilcoxon tests. In all relevant panels: p values for the top 5% broadest H3K4me3-domain-associated genes calculated in one-sided, one-sample Wilcoxon tests against expected genome-wide value from 10,000 random samplings (red dashed line).

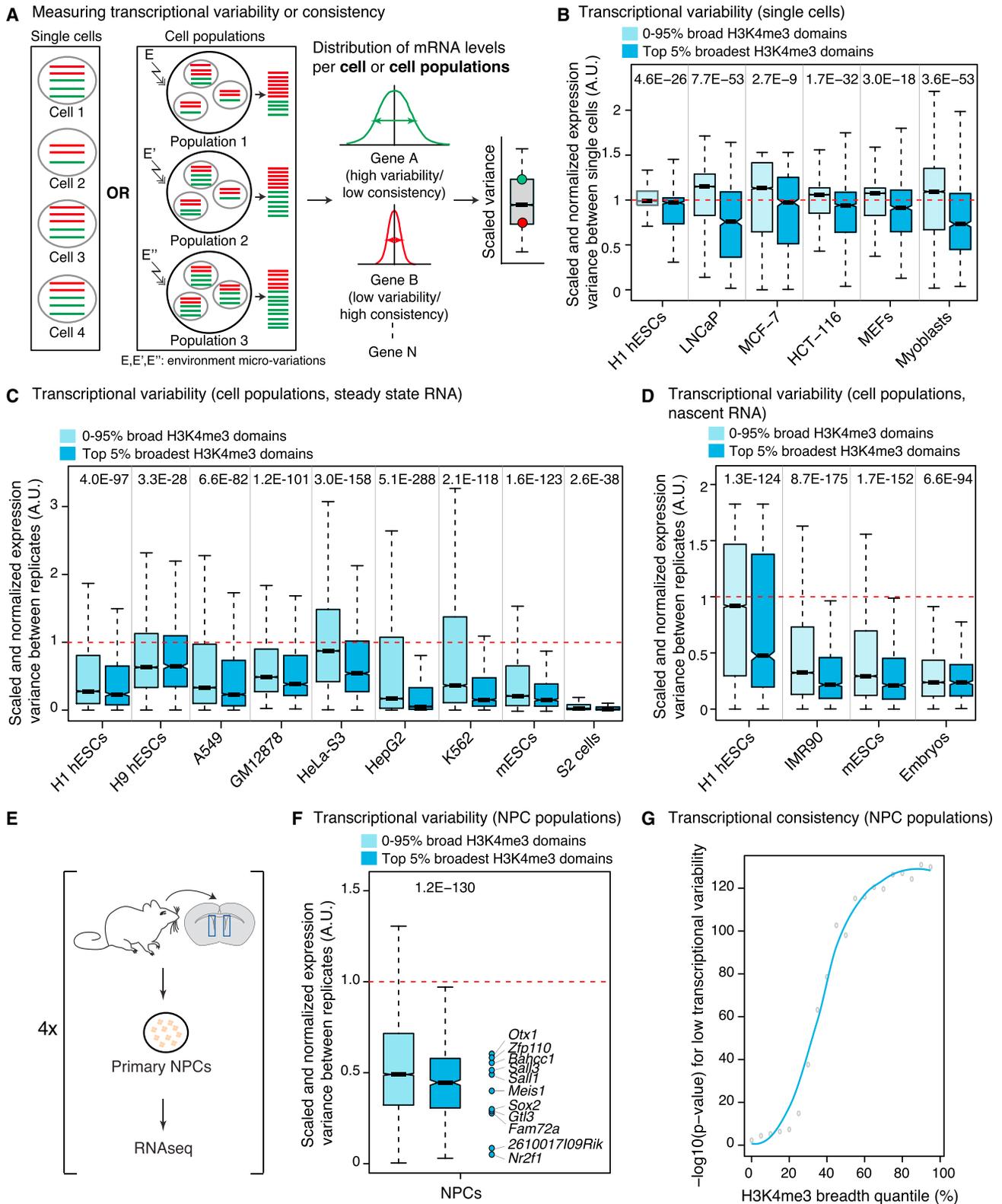


Figure 6. H3K4me3 Breadth Is Associated with Transcriptional Consistency

(A) Transcriptional consistency/variability at the level of single cells or cell populations is defined as variance of expression levels scaled to expression levels (i.e., scaled variance).

(legend continued on next page)

dynamics has been associated with transcriptional noise (Hansen and O'Shea, 2013), there may also exist an interplay between transcription factor binding and H3K4me3 breadth in tuning transcriptional consistency.

Biological Implications of Regulating Transcriptional Consistency

Control of transcriptional noise, through Pol II pausing or nucleosome positioning, has emerged as a new layer of complexity in biological processes (Levine, 2011; Raser and O'Shea, 2004). In metazoans, a dual requirement for transcriptional consistency has been observed. Reduction of transcription stochasticity is required for embryo patterning (Lagha et al., 2013; Raj et al., 2010), yet increased transcriptional variability may be permissive for cell-fate decisions (Balázsi et al., 2011). Although the importance of transcriptional consistency postdevelopment is not fully understood, transcriptional noise increases with age in cardiomyocytes (Bahar et al., 2006), but not in hematopoietic stem cells (Warren et al., 2007). The prevalence of buffer domains suggests that control of transcriptional consistency is critical in most cells and organisms. This conserved signature may ensure consistent expression of key genes involved in identity maintenance against variations of a fluctuating environment, a feature that may be lost during aging or disease.

EXPERIMENTAL PROCEDURES

H3K4me3 ChIP-Seq Analysis

Publicly available data sets were obtained from ENCODE (Bernstein et al., 2012), Roadmap Epigenomics (Zhu et al., 2013), GEO data sets, ArrayExpress/EBI, or Sequence Read Archive (SRA) (Table S1). ChIP-seq experiments in NPCs were performed as previously described (Webb et al., 2013), using H3K4me3 antibody (Active Motif, antibody 39159). Libraries were generated according to Illumina instructions and were sequenced on an Illumina GAII sequencer. Reads were mapped to reference genomes using bowtie0.12.7 (Langmead et al., 2009). ChIP-seq peaks were called using MACS2.08 (Feng et al., 2012) with the “—broad” option for histone marks. Peaks were assigned to the gene with the closest TSS.

Proliferation and Neurogenesis in Primary NPC Culture

Adult/postnatal mouse NPCs were isolated by microdissection of the subventricular zone and were maintained as nonadherent neurospheres, as described previously (Webb et al., 2013). For proliferation or neurogenesis assays, adherent adult (proliferation) or postnatal NPCs (neurogenesis) were transduced with a 30% lentiviral dilution and were selected using 0.5 $\mu\text{g/ml}$ of puromycin (Invivogen). For proliferation, cell growth was quantified by MTT (Molecular Probes) 4 days following infection. For neurogenesis, cells were

plated to account for proliferation differences and were then switched to NPC differentiation media for 4 days. The number of new neurons was assessed by Doublecortin (DCX) staining (Santa Cruz, sc-8066).

Transcriptional Variability

Transcriptional variability was assessed using microarray or RNA-seq data sets with ≥ 3 replicates generated in conditions matching H3K4me3 ChIP data sets. The expression variance per gene across replicates was scaled to the expression level of the gene (i.e., normalized to maximum level observed for that gene). Details are in the [Extended Experimental Procedures](#).

ACCESSION NUMBERS

Previously published data sets are reported in [Tables S1](#) and [S6](#). Sequencing data have been deposited to the Sequence Read Archive (SRA) under accession number SRP024294.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven tables, and seven figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2014.06.027>.

AUTHOR CONTRIBUTIONS

B.A.B., E.A.P., D.U., and A.B. designed the study. B.A.B. and D.U. performed the computational analysis with help from E.A.P. E.A.P. initiated the project and performed experiments with help from B.A.B. S.M. constructed the RNA-seq libraries. B.A.B. created the MySQL database, and E.D.W. and K.K. wrote the web interface with supervision from B.C.H. and J.M.C. A.C.D. did chromatin accessibility analyses. A.B.K. advised on the pipeline for breadth change analysis. E.M. and T.A.R. provided intellectual contributions. K.D. assisted in experiments. R.G. and J.C.B. gave advanced access to frog H3K4me3 data sets and M.P.S. to unpublished ENCODE data. B.A.B., D.U., E.A.P., and A.B. wrote the paper. All authors commented on the manuscript.

ACKNOWLEDGMENTS

We especially thank Joanna Wysocka and Michael Bassik for stimulating scientific discussion on this research. We thank Sivan Bercovici, Andreas Beyer, Will Greenleaf, Eric Greer, Courtney Hodges, Jonathan Pritchard, Julien Sage, Dario Valenzano, Anton Valouev, Reiner Veitia, and members of the Brunet lab for critical discussion and feedback on the manuscript. This work was supported by NIH DP1 AG044848 (A.B.), P01 GM099130 (A.B. and M.P.S.), P01 AG036695 (A.B. and T.A.R.), R01 GM103787 (J.C.B.), the Glenn Laboratories for the Biology of Aging (A.B. and T.A.R.), the Dean's fellowship at Stanford University (B.A.B.), NSF GRFP and NIH F31 AG043232 graduate fellowship (E.A.P.), the Department of Genetics at Stanford University (D.U.), and EMBO long-term postdoctoral fellowship and the Wenner-Gren Fellowship (S.M.).

(B) Transcriptional variability at the single-cell level (steady-state mRNA). Comparisons of the top 5% broadest H3K4me3 domains against the rest of the distribution are also significant in Wilcoxon tests ($2.9 \times 10^{-93} < p < 2.4 \times 10^{-16}$). See also [Figure S6A](#).

(C) Transcriptional variability at the cell population level (steady-state mRNA). Comparisons of the top 5% broadest H3K4me3 domains against the rest of the distribution are also significant in Wilcoxon tests ($1.5 \times 10^{-170} < p < 3.9 \times 10^{-3}$) (continued in [Figure S6B](#)).

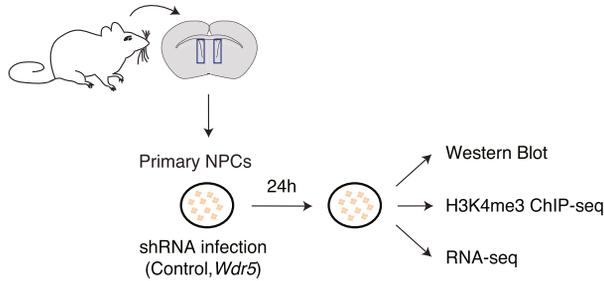
(D) Transcriptional variability at the cell population level (nascent mRNA by GRO-seq). Comparisons of the top 5% broadest H3K4me3 domains against the rest of the distribution are also significant in Wilcoxon tests ($1.8 \times 10^{-51} < p < 4.4 \times 10^{-20}$). See also [Figure S6C](#).

(E) Experimental design for RNA-seq data sets in primary NPCs cultures.

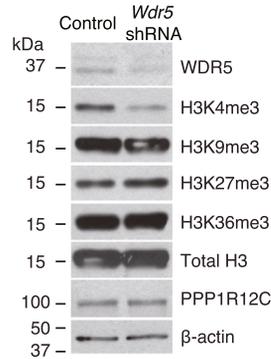
(F) Transcriptional variability at the cell population level in adult NPCs (steady-state mRNA). Comparison of the top 5% broadest H3K4me3 domains against the rest of the distribution are also significant in a Wilcoxon test ($p = 3.7 \times 10^{-12}$).

(G) Significance for lower transcriptional variability in adult NPCs against expected transcriptome-wide value expressed as $-\log_{10}$ (p value) in one-sided Wilcoxon tests. In all relevant panels: p values for the top 5% broadest H3K4me3-domain-associated genes calculated in one-sided, one-sample Wilcoxon tests against expected transcriptome-wide value from 10,000 random samplings (red dashed line).

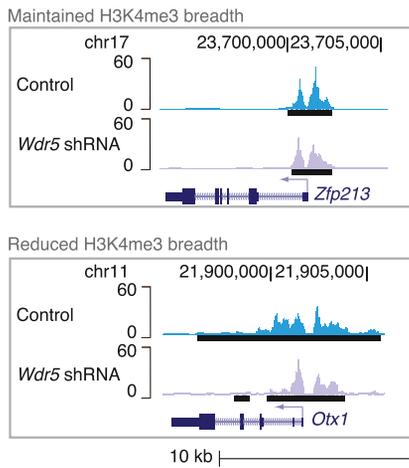
A Effect of *Wdr5* knock-down on H3K4me3 breadth in NPCs



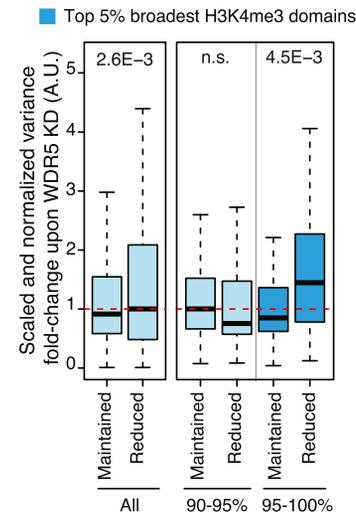
B Western blots



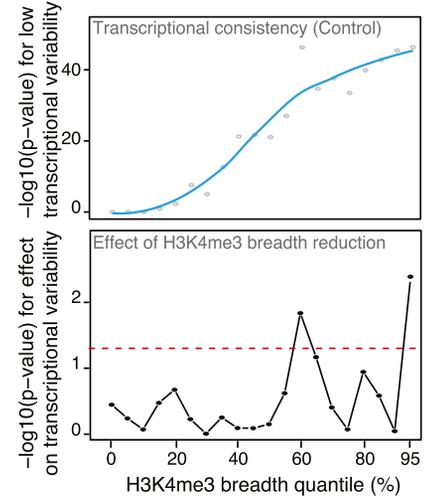
C Example effect of *Wdr5* knock-down (NPCs)



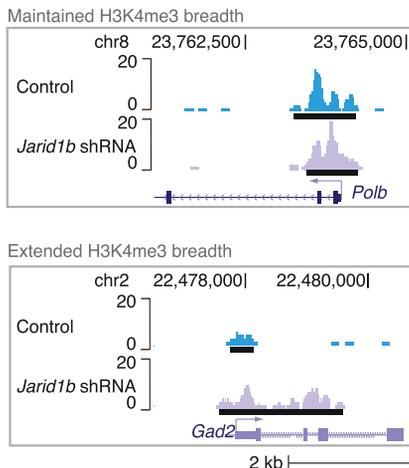
D Effect of H3K4me3 breadth reduction



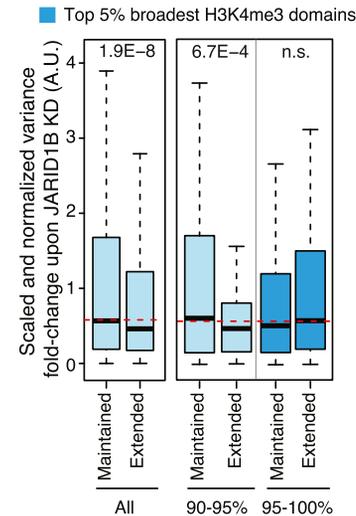
E H3K4me3 breadth changes and loss of transcriptional consistency



F Example effect of *Jarid1b* knock-down (mESCs)



G Effect of H3K4me3 breadth extension



H Summary Model

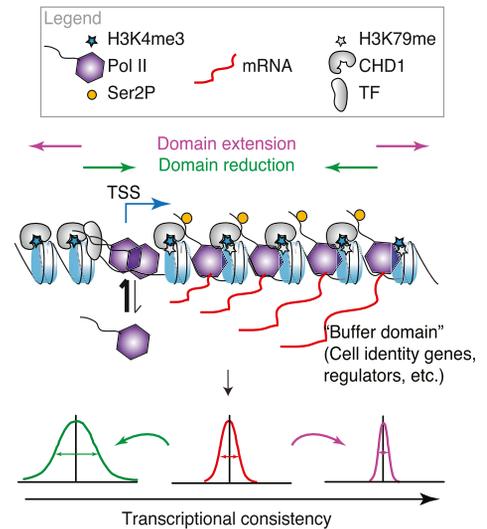


Figure 7. Experimental Perturbation of H3K4me3 Breadth Results in Changes to Transcriptional Consistency

(A) Experimental design to study the effect of knocking down *Wdr5* in primary NPC cultures.

(B) Western blot analysis of NPCs treated in control (empty vector) or *Wdr5* knockdown after 24 hr of infection. See also Figure S7B.

(legend continued on next page)

Received: June 11, 2013
 Revised: April 3, 2014
 Accepted: June 10, 2014
 Published: July 31, 2014

REFERENCES

- Adli, M., Zhu, J., and Bernstein, B.E. (2010). Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. *Nat. Methods* **7**, 615–618.
- Agoston, Z., Heine, P., Brill, M.S., Grebbin, B.M., Hau, A.C., Kallenborn-Gerhardt, W., Schramm, J., Götz, M., and Schulte, D. (2014). Meis2 is a Pax6 co-factor in neurogenesis and dopaminergic periglomerular fate specification in the adult olfactory bulb. *Development* **141**, 28–38.
- Aiden, A.P., Rivera, M.N., Rheinbay, E., Ku, M., Coffman, E.J., Truong, T.T., Vargas, S.O., Lander, E.S., Haber, D.A., and Bernstein, B.E. (2010). Wilms tumor chromatin profiles highlight stem cell properties and a renal developmental network. *Cell Stem Cell* **6**, 591–602.
- Ang, Y.S., Tsai, S.Y., Lee, D.F., Monk, J., Su, J., Ratnakumar, K., Ding, J., Ge, Y., Darr, H., Chang, B., et al. (2011). *Wdr5* mediates self-renewal and reprogramming via the embryonic stem cell core transcriptional network. *Cell* **145**, 183–197.
- Ardehali, M.B., Mei, A., Zobeck, K.L., Caron, M., Lis, J.T., and Kusch, T. (2011). *Drosophila* Set1 is the major histone H3 lysine 4 trimethyltransferase with role in transcription. *EMBO J.* **30**, 2817–2828.
- Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttill, R.A., Dollé, M.E., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A., and Vijg, J. (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. *Nature* **441**, 1011–1014.
- Balázs, G., van Oudenaarden, A., and Collins, J.J. (2011). Cellular decision making and biological noise: from microbes to mammals. *Cell* **144**, 910–925.
- Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. *Cell* **129**, 823–837.
- Bernstein, B.E., Kamal, M., Lindblad-Toh, K., Bekiranov, S., Bailey, D.K., Huebert, D.J., McMahon, S., Karlsson, E.K., Kulbokas, E.J., 3rd, Gingeras, T.R., et al. (2005). Genomic maps and comparative analysis of histone modifications in human and mouse. *Cell* **120**, 169–181.
- Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A bivalent chromatin structure marks key developmental genes in embryonic stem cells. *Cell* **125**, 315–326.
- Bernstein, B.E., Birney, E., Dunham, I., Green, E.D., Gunter, C., Snyder, M., Epstein, C.B., Fritze, S., Harrow, J., Kaul, R., et al.; ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74.
- Black, J.C., Van Rechem, C., and Whetstine, J.R. (2012). Histone lysine methylation dynamics: establishment, regulation, and biological impact. *Mol. Cell* **48**, 491–507.
- Boettiger, A.N., Ralph, P.L., and Evans, S.N. (2011). Transcriptional regulation: effects of promoter proximal pausing on speed, synchrony and reliability. *PLoS Comput. Biol.* **7**, e1001136.
- Bonaguidi, M.A., Wheeler, M.A., Shapiro, J.S., Stadel, R.P., Sun, G.J., Ming, G.L., and Song, H. (2011). In vivo clonal analysis reveals self-renewing and multipotent adult neural stem cell characteristics. *Cell* **145**, 1142–1155.
- Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218.
- Carrozza, M.J., Li, B., Florens, L., Suganuma, T., Swanson, S.K., Lee, K.K., Shia, W.J., Anderson, S., Yates, J., Washburn, M.P., and Workman, J.L. (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. *Cell* **123**, 581–592.
- Doetsch, F., Caillé, I., Lim, D.A., Garcia-Verdugo, J.M., and Alvarez-Buylla, A. (1999). Subventricular zone astrocytes are neural stem cells in the adult mammalian brain. *Cell* **97**, 703–716.
- Dong, D., Shao, X., Deng, N., and Zhang, Z. (2011). Gene expression variations are predictive for stochastic noise. *Nucleic Acids Res.* **39**, 403–413.
- Dong, X., Greven, M.C., Kundaje, A., Djebali, S., Brown, J.B., Cheng, C., Gingeras, T.R., Gerstein, M., Guigó, R., Birney, E., and Weng, Z. (2012). Modeling gene expression using chromatin features in various cellular contexts. *Genome Biol.* **13**, R53.
- Feng, J., Liu, T., Qin, B., Zhang, Y., and Liu, X.S. (2012). Identifying ChIP-seq enrichment using MACS. *Nat. Protoc.* **7**, 1728–1740.
- Field, Y., Kaplan, N., Fondufe-Mittendorf, Y., Moore, I.K., Sharon, E., Lubling, Y., Widom, J., and Segal, E. (2008). Distinct modes of regulation by chromatin encoded through nucleosome positioning signals. *PLoS Comput. Biol.* **4**, e1000216.
- Gilchrist, D.A., Dos Santos, G., Fargo, D.C., Xie, B., Gao, Y., Li, L., and Adelman, K. (2010). Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**, 540–551.
- Greer, E.L., Maures, T.J., Hauswirth, A.G., Green, E.M., Leeman, D.S., Maro, G.S., Han, S., Banko, M.R., Gozani, O., and Brunet, A. (2010). Members of the H3K4 trimethylation complex regulate lifespan in a germline-dependent manner in *C. elegans*. *Nature* **466**, 383–387.
- Guenther, M.G., Levine, S.S., Boyer, L.A., Jaenisch, R., and Young, R.A. (2007). A chromatin landmark and transcription initiation at most promoters in human cells. *Cell* **130**, 77–88.
- Hansen, A.S., and O’Shea, E.K. (2013). Promoter decoding of transcription factor dynamics involves a trade-off between noise and control of gene expression. *Mol. Syst. Biol.* **9**, 704.

(C) Examples of H3K4me3 peaks in control (empty vector) and *Wdr5* knockdown in NPCs after 24 hr of infection.

(D) Reduction of H3K4me3 breadth upon 24 hr *Wdr5* knockdown is linked to increased transcriptional variability in NPCs. Variability was measured between three biological replicates at genes whose H3K4me3 domains were maintained or reduced upon *Wdr5* knockdown. (Red dashed line) Expected transcriptome-wide value. p values between genes with maintained versus reduced breadth in Wilcoxon tests.

(E) H3K4me3 breadth remodeling upon *Wdr5* knockdown and loss of transcriptional consistency in NPCs. (Top) $-\log_{10}(p \text{ value})$ in one-sided Wilcoxon test for lower transcriptional variability in control-infected NPCs than expected transcriptome-wide value. (Bottom) $-\log_{10}(p \text{ value})$ in one-sided Wilcoxon tests for increased variability of genes losing H3K4me3 breadth versus genes of the same original H3K4me3 quantile with maintained breadth. (Red dashed line) $p = 0.05$.

(F) Examples of H3K4me3 peaks in control (scramble) and *Jarid1b* knockdown in mESCs after 48 hr of infection.

(G) Gain of H3K4me3 breadth upon *Jarid1b* knockdown is linked to increased transcriptional consistency in mESCs after 48 hr of infection. Variability between three biological replicates at genes whose H3K4me3 domains were maintained versus extended upon *Jarid1b* knockdown. (Red dashed line) Expected transcriptome-wide value. p values between genes with maintained or extended breadth using Wilcoxon tests.

(H) Summary model. Broad H3K4me3 domains extend 5' and 3' of TSSs and mark genes that are important for cell identity/function and genes with increased transcriptional consistency. Broad H3K4me3 domains may promote chromatin accessibility, thereby allowing efficient Pol II loading and elongation. The mechanism responsible for the deposition of broad H3K4me3 domains is unknown but may involve tissue-specific transcription factors and the elongation machinery in a positive feedback loop. These domains may help to “buffer” important cell lineage/function genes against environmental fluctuation and can serve as discovery tool for such genes.

- Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112.
- Hödl, M., and Basler, K. (2012). Transcription in the absence of histone H3.2 and H3K4 methylation. *Curr. Biol.* 22, 2253–2257.
- Ingham, P.W. (1998). *trithorax* and the regulation of homeotic gene expression in *Drosophila*: a historical perspective. *Int. J. Dev. Biol.* 42, 423–429.
- Jeong, M., Sun, D., Luo, M., Huang, Y., Challen, G.A., Rodriguez, B., Zhang, X., Chavez, L., Wang, H., Hannah, R., et al. (2014). Large conserved domains of low DNA methylation maintained by Dnmt3a. *Nat. Genet.* 46, 17–23.
- Krogan, N.J., Dover, J., Wood, A., Schneider, J., Heidt, J., Boateng, M.A., Dean, K., Ryan, O.W., Golshani, A., Johnston, M., et al. (2003). The Paf1 complex is required for histone H3 methylation by COMPASS and Dot1p: linking transcriptional elongation to histone methylation. *Mol. Cell* 11, 721–729.
- Kuo, A.J., Song, J., Cheung, P., Ishibe-Murakami, S., Yamazoe, S., Chen, J.K., Patel, D.J., and Gozani, O. (2012). The BAH domain of ORC1 links H4K20me2 to DNA replication licensing and Meier-Gorlin syndrome. *Nature* 484, 115–119.
- Lagha, M., Bothma, J.P., Esposito, E., Ng, S., Stefaniak, L., Tsui, C., Johnston, J., Chen, K., Gilmour, D.S., Zeitlinger, J., and Levine, M.S. (2013). Paused Pol II coordinates tissue morphogenesis in the *Drosophila* embryo. *Cell* 153, 976–987.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* 10, R25.
- Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 152, 1021–1036.
- Levine, M. (2011). Paused RNA polymerase II as a developmental checkpoint. *Cell* 145, 502–511.
- Lien, W.H., Guo, X., Polak, L., Lawton, L.N., Young, R.A., Zheng, D., and Fuchs, E. (2011). Genome-wide maps of histone modifications unwind *in vivo* chromatin states of the hair follicle lineage. *Cell Stem Cell* 9, 219–232.
- Lujan, E., Chanda, S., Ahlenius, H., Südhof, T.C., and Wernig, M. (2012). Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells. *Proc. Natl. Acad. Sci. USA* 109, 2527–2532.
- Luo, Z., Lin, C., and Shilatifard, A. (2012). The super elongation complex (SEC) family in transcriptional control. *Nat. Rev. Mol. Cell Biol.* 13, 543–547.
- Nehar, S., Mishra, M., and Heese, K. (2009). Identification and characterisation of the novel amyloid-beta peptide-induced protein p17. *FEBS Lett.* 583, 3247–3253.
- Ninkovic, J., Steiner-Mezzadri, A., Jawerka, M., Akinci, U., Masserdotti, G., Petricca, S., Fischer, J., von Holst, A., Beckers, J., Lie, C.D., et al. (2013). The BAF complex interacts with Pax6 in adult neural progenitors to establish a neurogenic cross-regulatory transcriptional network. *Cell Stem Cell* 13, 403–418.
- Parker, S.C., Stitzel, M.L., Taylor, D.L., Orozco, J.M., Erdos, M.R., Akiyama, J.A., van Bueren, K.L., Chines, P.S., Narisu, N., Black, B.L., et al.; NISC Comparative Sequencing Program; National Institutes of Health Intramural Sequencing Center Comparative Sequencing Program Authors; NISC Comparative Sequencing Program Authors (2013). Chromatin stretch enhancer states drive cell-specific gene regulation and harbor human disease risk variants. *Proc. Natl. Acad. Sci. USA* 110, 17921–17926.
- Pérez-Lluch, S., Blanco, E., Carbonell, A., Raha, D., Snyder, M., Serras, F., and Corominas, M. (2011). Genome-wide chromatin occupancy analysis reveals a role for ASH2 in transcriptional pausing. *Nucleic Acids Res.* 39, 4628–4639.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Bruggmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–283.
- Raj, A., Rifkin, S.A., Andersen, E., and van Oudenaarden, A. (2010). Variability in gene expression underlies incomplete penetrance. *Nature* 463, 913–918.
- Raser, J.M., and O’Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. *Science* 304, 1811–1814.
- Sakurai, Y., Kurokawa, D., Kiyonari, H., Kajikawa, E., Suda, Y., and Aizawa, S. (2010). Otx2 and Otx1 protect diencephalon and mesencephalon from caudalization into metencephalon during early brain regionalization. *Dev. Biol.* 347, 392–403.
- Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407–411.
- Schmitz, S.U., Albert, M., Malatesta, M., Morey, L., Johansen, J.V., Bak, M., Tommerup, N., Abarrategui, I., and Helin, K. (2011). Jarid1b targets genes regulating development and is involved in neural differentiation. *EMBO J.* 30, 4586–4600.
- Shah, P.P., Donahue, G., Otte, G.L., Capell, B.C., Nelson, D.M., Cao, K., Aggarwala, V., Cruickshanks, H.A., Rai, T.S., McBryan, T., et al. (2013). Lamin B1 depletion in senescent cells triggers large-scale changes in gene expression and the chromatin landscape. *Genes Dev.* 27, 1787–1799.
- Shilatifard, A. (2012). The COMPASS family of histone H3K4 methylases: mechanisms of regulation in development and disease pathogenesis. *Annu. Rev. Biochem.* 81, 65–95.
- Siebold, A.P., Banerjee, R., Tie, F., Kiss, D.L., Moskowicz, J., and Harte, P.J. (2010). Polycomb Repressive Complex 2 and Trithorax modulate *Drosophila* longevity and stress resistance. *Proc. Natl. Acad. Sci. USA* 107, 169–174.
- Sims, R.J., 3rd, Belotserkovskaya, R., and Reinberg, D. (2004). Elongation by RNA polymerase II: the short and long of it. *Genes Dev.* 18, 2437–2468.
- Smolle, M., Venkatesh, S., Gogol, M.M., Li, H., Zhang, Y., Florens, L., Washburn, M.P., and Workman, J.L. (2012). Chromatin remodelers Isw1 and Chd1 maintain chromatin structure during transcription by preventing histone exchange. *Nat. Struct. Mol. Biol.* 19, 884–892.
- Soufi, A., Donahue, G., and Zaret, K.S. (2012). Facilitators and impediments of the pluripotency reprogramming factors’ initial engagement with the genome. *Cell* 151, 994–1004.
- Strahl, B.D., Ohba, R., Cook, R.G., and Allis, C.D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in Tetrahymena. *Proc. Natl. Acad. Sci. USA* 96, 14967–14972.
- Triebel, R.C., and Shilatifard, A. (2009). WDR5, a complexed protein. *Nat. Struct. Mol. Biol.* 16, 678–680.
- Wang, Y.Z., Plane, J.M., Jiang, P., Zhou, C.J., and Deng, W. (2011). Concise review: Quiescent and active states of endogenous adult neural stem cells: identification and characterization. *Stem Cells* 29, 907–912.
- Warren, L.A., Rossi, D.J., Schiebinger, G.R., Weissman, I.L., Kim, S.K., and Quake, S.R. (2007). Transcriptional instability is not a universal attribute of aging. *Aging Cell* 6, 775–782.
- Webb, A.E., Pollina, E.A., Vierbuchen, T., Urbán, N., Ucar, D., Leeman, D.S., Martynoga, B., Sewak, M., Rando, T.A., Guillemot, F., et al. (2013). FOXO3 shares common targets with ASCL1 genome-wide and inhibits ASCL1-dependent neurogenesis. *Cell Rep.* 4, 477–491.
- Weinberger, L., Voickek, Y., Tirosh, I., Hornung, G., Amit, I., and Barkai, N. (2012). Expression noise and acetylation profiles distinguish HDAC functions. *Mol. Cell* 47, 193–202.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.
- Xie, W., Schultz, M.D., Lister, R., Hou, Z., Rajagopal, N., Ray, P., Whitaker, J.W., Tian, S., Hawkins, R.D., Leung, D., et al. (2013). Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* 153, 1134–1148.
- Zhu, J., Adli, M., Zou, J.Y., Verstappen, G., Coyne, M., Zhang, X., Durham, T., Miri, M., Deshpande, V., De Jager, P.L., et al. (2013). Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* 152, 642–654.