

**Spring 2021 – Epigenetics and Systems Biology**  
**Discussion Session (Epigenetics and Development)**  
**Michael K. Skinner – Biol 476/576**  
**Week 8 (March 11)**

**Epigenetics of Cell and Developmental Biology**

Primary Papers

1. Hackett, et al., (2013) Science. 339:448. (PMID: 23223451)
2. Colognori, et al. (2020) Dev Cell. 2020 Jul 6;54(1):21-32.e5. (PMID: 32531209)
3. Wu, et al., (2020) Cell Reports. 33(7):108395. (PMID: 33207205)

**Discussion**

Student 19 – Ref #1 above

- What stages of development and cells have reduced DNA methylation?
- What technology was used?
- What role does 5hmC have in the process?

Student 20 – Ref #2 above

- What is Xist?
- What are X inactivation stages and Xist repeats mechanism?
- What if Xist repeats lost?

Student 21 – Ref #3 above

- What histone modifications and chromatin remodeling proteins involved?
- What is Myeloid and Erythroid progenitors?
- What is the epigenetic observation on the regulation of lineage development?

performed with *P*-element–transposon–containing Cg4-neo and DmTNP or hTh9 (Fig. 2, B and C, and fig. S7). These observations imply that the DmTNP and hTh9 proteins can nick DNA, independent of having *P*-element termini, which would lead to elevated gene transfer via DNA linearization. A similar observation was made for the SET domain and mariner transposase fusion gene–containing protein (SETMAR or Metnase protein), but this protein is inactive for transposition of HsMAR transposons (8, 9). However, most important, the presence of *P*-element termini on Cg4-neo enhanced the DNA integration activity of both DmTNP and hTh9 3 to 5 times above background, which suggested transpositional DNA integration. Many G418-resistant colonies were also obtained (fig. S7) when Cg4-neo was transfected into a stable HEK293 cell line induced to express a tetracycline-inducible human THAP9 gene (Fig. 2B).

To analyze the nature of the DNA integration events in the G418-resistant colonies, genomic DNA was isolated from individual colonies obtained from DmTNP or hTh9 cotransfections with Cg4-neo, and the sites of *P*-element insertion were characterized by splinkerette polymerase chain reaction (PCR) (17, 18) followed by DNA sequencing. DNA sequence analysis of PCR integration sites identified distinct integration sites with novel 8–base pair (bp) target-site duplications (TSDs) for individual integration events into the human genome that had occurred with both DmTNP and hTh9 (Table 1 and tables S1 and S2). Taken together, these data indicate that human THAP9 actively integrates genetically marked *Drosophila P*-element vectors into human cells by transposition.

The studies reported here indicate that the human THAP9 gene encodes an active DNA transposase that can mobilize *Drosophila P*-element transposons in *Drosophila* and human cells. It will be interesting to investigate the physiological relevance of THAP9's transposition function and to find out if any THAP9 recombination signal DNA elements can be found in the human genome. This is the first report, beyond the V(D)J recombination system, of an active DNA transposase in the human genome. *P* element–like transposons and THAP9-related genes are not restricted to *Drosophila* or related insect species but are widely distributed in eukaryotic genomes like *Ciona* (sea squirt), zebrafish, chicken, and *Trichomonas vaginalis* (a parasitic protozoan) (7, 19). The THAP9 gene is absent and has apparently been lost from sequenced rodent genomes (6). Although many of the human transposase–related genes are derived from DNA transposons (43 of 47) (2), most have not been characterized, with the exception of the V(D)J recombinase RAG1 and RAG2 (10, 11) and the SETMAR (Metnase) protein (8). It is possible that other human genes of this class, besides THAP9, may also encode active DNA transposases.

#### References and Notes

1. C. Feschotte, E. J. Pritham, *Annu. Rev. Genet.* **41**, 331 (2007).
2. E. S. Lander *et al.*, *Nature* **409**, 860 (2001).
3. S. Schaack, C. Gilbert, C. Feschotte, *Trends Ecol. Evol.* **25**, 537 (2010).
4. H. H. Kazazian Jr., *Science* **303**, 1626 (2004).
5. C. Feschotte, *Nat. Rev. Genet.* **9**, 397 (2008).
6. S. E. Hammer, S. Strehl, S. Hagemann, *Mol. Biol. Evol.* **22**, 833 (2005).
7. H. Quesneville, D. Nouaud, D. Anxolabehere, *Mol. Biol. Evol.* **22**, 741 (2005).

8. M. Shaheen, E. Williamson, J. Nickoloff, S.-H. Lee, R. Hromas, *Genetica* **138**, 559 (2010).
9. S.-H. Lee *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **102**, 18075 (2005).
10. S. D. Fugmann, A. I. Lee, P. E. Shockett, I. J. Villey, D. G. Schatz, *Annu. Rev. Immunol.* **18**, 495 (2000).
11. M. Gellert, *Annu. Rev. Biochem.* **71**, 101 (2002).
12. M. Roussigne *et al.*, *Trends Biochem. Sci.* **28**, 66 (2003).
13. S. Hagemann, W. Pinsker, *Mol. Biol. Evol.* **18**, 1979 (2001).
14. A. Sabogal, A. Y. Lyubimov, J. E. Corn, J. M. Berger, D. C. Rio, *Nat. Struct. Mol. Biol.* **17**, 117 (2010).
15. Y. M. Mul, D. C. Rio, *EMBO J.* **16**, 4441 (1997).
16. P. J. Southern, P. Berg, *J. Mol. Appl. Genet.* **1**, 327 (1982).
17. C. Horn *et al.*, *Nat. Genet.* **39**, 933 (2007).
18. C. J. Potter, L. Luo, *PLoS ONE* **5**, e10168 (2010).
19. J. M. Carlton *et al.*, *Science* **315**, 207 (2007).
20. E. L. Beall, D. C. Rio, *Genes Dev.* **10**, 921 (1996).

**Acknowledgments:** We thank J. M. Taliaferro, J. Aspdren, and M. Francis for comments and discussion; R. Rawat for technical assistance; and C. J. Potter (Johns Hopkins Medical School) for helpful suggestions about the splinkerette PCR experiments. This work was supported by NIH grants R01GM48862, R01GM61987, R01 GM097352, R01GM104385, and R01GM094890. S.M. and D.C.R. conceived the experiments. S.M., A.S., and D.C.R. performed cell culture and molecular biological experiments and analyzed data. S.M. and D.C.R. wrote the paper. The authors declare no competing financial interests. Requests for materials should be addressed to the corresponding author.

#### Supplementary Materials

www.sciencemag.org/cgi/content/full/339/6118/446/DC1  
Materials and Methods  
Figs. S1 to S7  
Tables S1 and S2  
References (21, 22)

19 October 2012; accepted 27 November 2012  
10.1126/science.1231789

## Germline DNA Demethylation Dynamics and Imprint Erasure Through 5-Hydroxymethylcytosine

Jamie A. Hackett,<sup>1,2</sup> Roopsha Sengupta,<sup>1,2\*</sup> Jan J. Zyllicz,<sup>1,2,3\*</sup> Kazuhiro Murakami,<sup>1,2\*</sup> Caroline Lee,<sup>1,2</sup> Thomas A. Down,<sup>1</sup> M. Azim Surani<sup>1,2,3†</sup>

Mouse primordial germ cells (PGCs) undergo sequential epigenetic changes and genome-wide DNA demethylation to reset the epigenome for totipotency. Here, we demonstrate that erasure of CpG methylation (5mC) in PGCs occurs via conversion to 5-hydroxymethylcytosine (5hmC), driven by high levels of TET1 and TET2. Global conversion to 5hmC initiates asynchronously among PGCs at embryonic day (E) 9.5 to E10.5 and accounts for the unique process of imprint erasure. Mechanistically, 5hmC enrichment is followed by its protracted decline thereafter at a rate consistent with replication-coupled dilution. The conversion to 5hmC is an important component of parallel redundant systems that drive comprehensive reprogramming in PGCs. Nonetheless, we identify rare regulatory elements that escape systematic DNA demethylation in PGCs, providing a potential mechanistic basis for transgenerational epigenetic inheritance.

**S**pecification of primordial germ cells (PGCs) from epiblast cells at ~embryonic day (E) 6.25 is linked with extensive epigenetic reprogramming—including global DNA demethylation,

chromatin reorganization, and imprint erasure—that is vital for generating totipotency (1, 2). The erasure of CpG methylation (5mC) is a key component of this program, but the dynamics

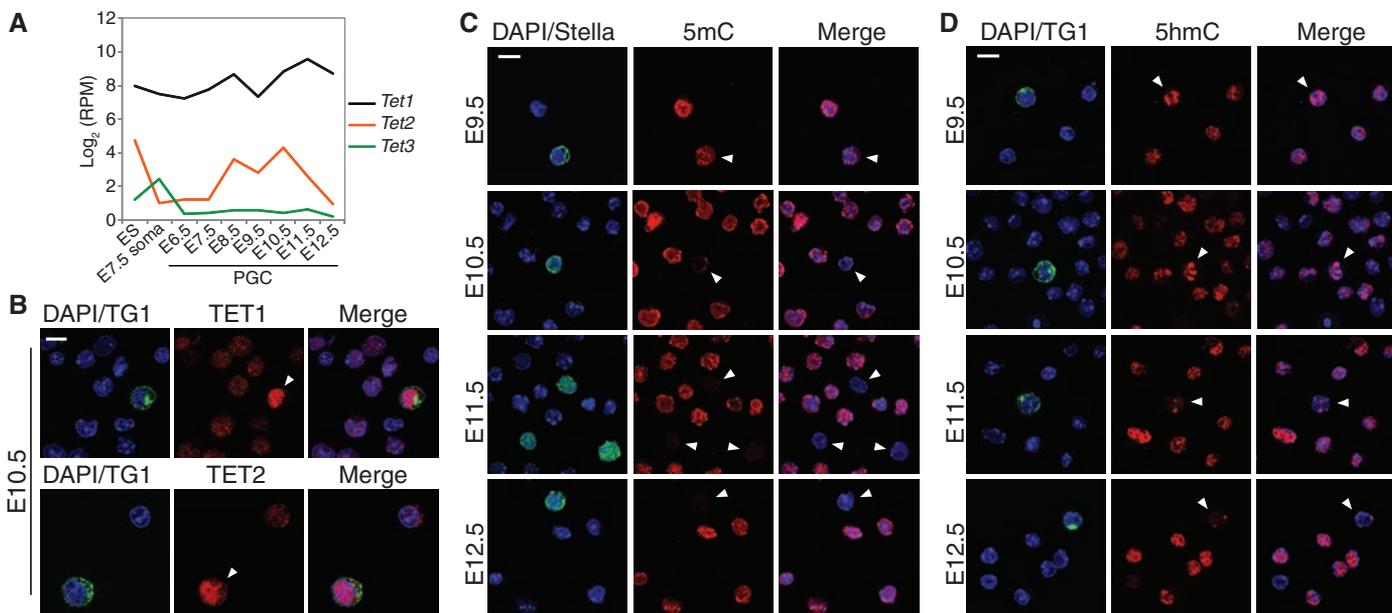
and underlying mechanisms of the process remain unclear (3). Here, we report a comprehensive analysis of PGCs by combining immunofluorescence, genome-wide 5-(hydroxy)methylcytosine DNA immunoprecipitation sequencing [(h)meDIP-seq], single-cell RNA sequencing (RNA-seq), bisulfite sequencing, and functional analyses to address the mechanistic basis of epigenetic reprogramming in PGCs.

We investigated *Tet* expression by using single-cell RNA-seq, which revealed that *Tet1* and *Tet2* are expressed in PGCs and peak between E10.5 and E11.5 but that *Tet3* is undetectable (Fig. 1A). Immunofluorescence (IF) showed that TET1 and TET2 are nuclear and expressed at significantly higher levels in PGCs than in neighboring somatic cells between E9.5 and E11.5 (Fig. 1B and figs.

<sup>1</sup>Wellcome Trust/Cancer Research UK Gurdon Institute, University of Cambridge, Cambridge CB2 1QN, UK. <sup>2</sup>Wellcome Trust/Medical Research Council Stem Cell Institute, University of Cambridge, Cambridge, UK. <sup>3</sup>Department of Physiology, Development and Neuroscience, University of Cambridge, Cambridge, UK.

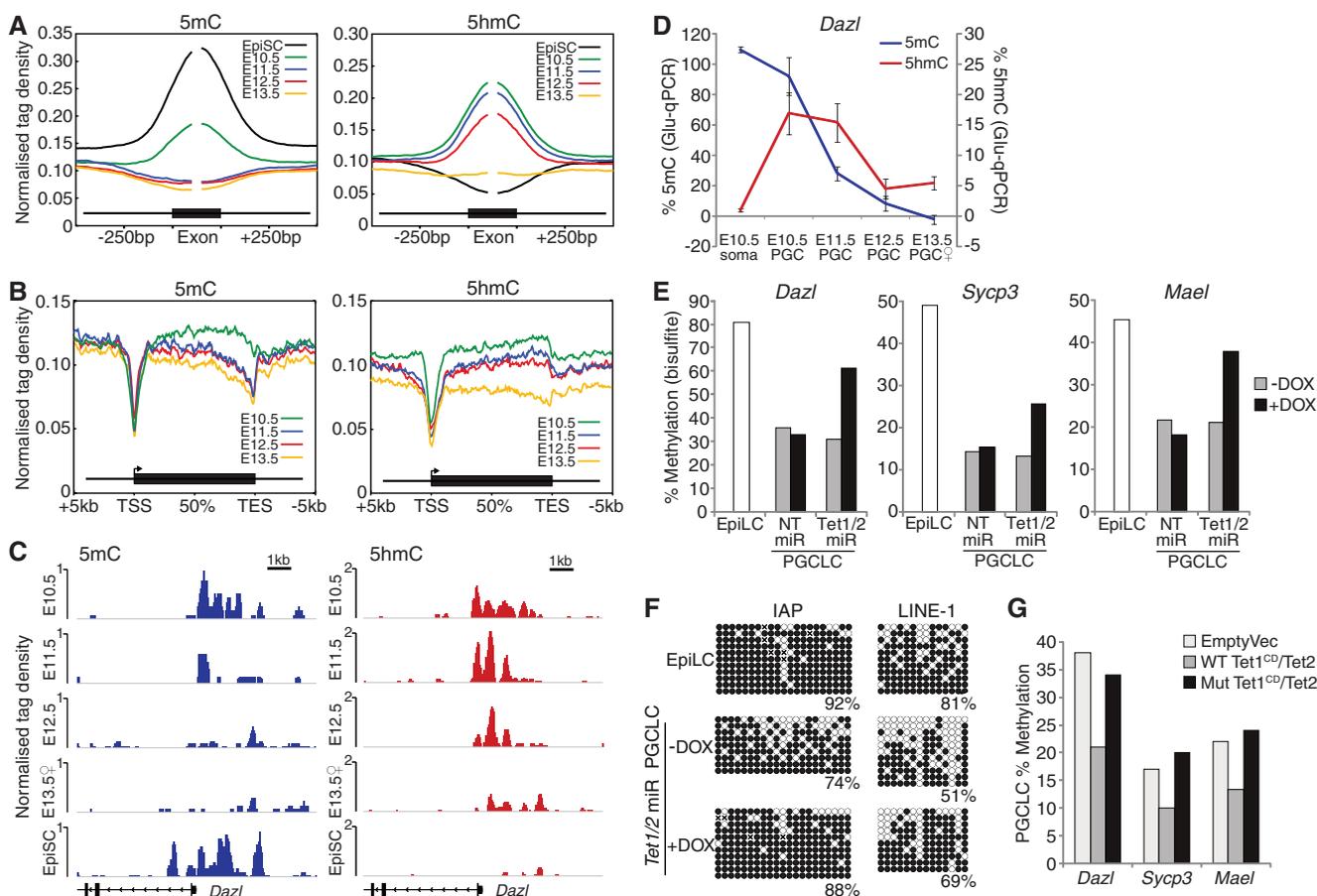
\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: a.surani@gurdon.cam.ac.uk



**Fig. 1.** Global dynamics of 5mC, 5hmC, and TETs in PGCs. **(A)** Single-cell RNA-seq analysis of *Tet1*, *Tet2*, and *Tet3* expression. Shown is log<sub>2</sub> reads per million (RPM). **(B)** Expression of TET1 and TET2 in E10.5 PGCs (arrowheads)

and soma. **(C)** Dynamics of DNA methylation (5mC) in PGCs shows 5mC erasure between E9.5 and E11.5. **(D)** Kinetics of 5hmC in PGCs. TG1/STELLA mark PGCs. Scale bars indicate 10 μm.



**Fig. 2.** Erasure of 5mC is coupled to 5hmC conversion. **(A)** Enrichment of 5mC and 5hmC in E10.5 to E13.5 PGCs and EpiSCs over internal exons. **(B)** Distributions of 5mC and 5hmC relative to a metagene. TSS, transcription start site; TES, transcription end site. **(C)** Profiles of 5mC (blue) and 5hmC (red) at the *Dazl* promoter. **(D)** Glu-qPCR showing quantitative levels of 5mC and 5hmC at a CCGG

site in the *Dazl* promoter. Error bars represent SEM. **(E and F)** DNA methylation (%) by bisulfite sequencing of –DOX or +DOX *Tet1/Tet2* miR or NT miR PGCLCs at (E) gene promoters and (F) repeat elements. Open and solid circles represent unmethylated and methylated CpGs, respectively. **(G)** DNA methylation in PGCLCs stably expressing catalytically active (WT) or mutant (Mut) TET1 and TET2.

Downloaded from <http://science.sciencemag.org/> on February 21, 2017

S1 and S2). This suggests that erasure of 5mC in PGCs could occur through conversion to 5hmC by TET1 and TET2 together (4, 5).

We pursued this possibility by IF and found a progressive reduction of 5mC in PGCs between E9.5 and E10.5, until it became undetectable by E11.5 (Fig. 1C). The loss of 5mC occurs concurrently with a global enrichment of 5-hydroxymethylcytosine (5hmC) in PGCs between E9.5 and E10.5, suggesting a genome-scale conversion of 5mC to 5hmC (Fig. 1D). The global conversion to 5hmC initiates asynchronously among PGCs from E9.5, perhaps reflecting developmental heterogeneity (figs. S3 to S5). Indeed, TET1 up-regulation also initially occurs in a subset of PGCs from E9.5, which apparently also exhibit lower 5mC signal (fig. S6). In contrast to soma and embryonic stem (ES) cells (6), we observed that 5hmC exhibited a distinct localization in PGCs that coincided with 4',6-diamidino-2-phenylindole (DAPI)-dense chromocenters, indicating that the conversion of 5mC to 5hmC includes heterochromatic satellite regions (fig. S7). The enrichment of 5hmC in PGCs at E10.5 is followed by its progressive reduction, suggesting that 5hmC is an intermediate toward demethylation to unmodified cytosine (C) (Fig. 1D). We checked whether 5hmC is subsequently converted

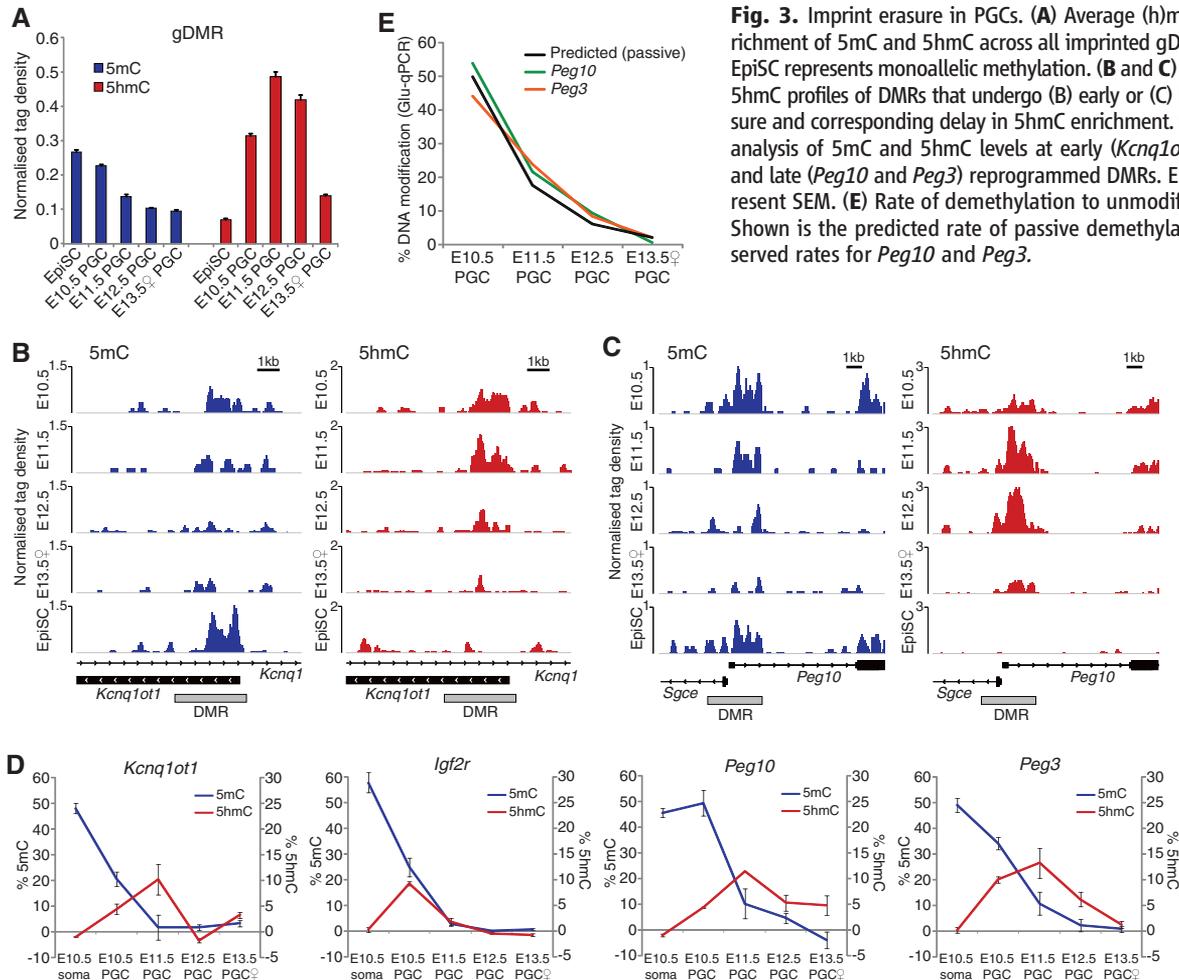
to 5-formylcytosine or 5-carboxycytosine but found no detectable enrichment of these derivatives in PGCs (fig. S8) (7).

To gain further insight into the dynamics of 5mC to 5hmC conversion, we performed meDIP-seq and hmeDIP-seq in E10.5 to E13.5 PGCs (fig. S9). Because before E10.5 PGCs were highly limiting, we also profiled epiblast stem cells (EpiSCs), which are derived from the same post-implantation epiblast as nascent PGCs, and embryonic soma (E10.5) as references (fig. S10). Unlike bisulfite sequencing, our approach distinguishes between 5mC and 5hmC but generates a relative rather than a quantitative measure of modifications (6). We therefore initially examined exonic sequences, which are highly methylated and thus exhibit an informative dynamic range of relative (h)meDIP signal when they become demethylated. We found significantly reduced 5mC in E10.5 PGCs relative to EpiSC and soma and erasure by E11.5 (Fig. 2A and figs. S11 to S13). The loss of 5mC in PGCs is paralleled by a strong exonic enrichment of 5hmC, indicating 5mC-to-5hmC conversion (Fig. 2A and fig. S11). Once 5mC is converted to 5hmC, it is set on a pathway toward demethylation, because there are no 5hmC maintenance mechanisms (6). Consistent with this, 5hmC undergoes

a progressive depletion during PGC development, which is delayed relative to loss of 5mC (Fig. 2, A and B).

Next, we examined methylation-dependent genes such as *Dazl*, which are activated by promoter demethylation in PGCs (8, 9), and observed strong 5hmC enrichment coincident with loss of 5mC at their promoters (Fig. 2C and fig. S14). We confirmed that 5mC erasure is coupled to 5hmC enrichment at the *Dazl* promoter quantitatively, by using the glucosyltransferase-quantitative polymerase chain reaction (Glu-qPCR) assay (Fig. 2D). RNA-seq revealed that transcriptional activation of *Dazl* and other methylation-dependent germline genes initiates at E9.5 and increases progressively until ~E11.5 (fig. S15). This represents an important functional readout of the timing of DNA demethylation in PGCs.

To functionally link 5hmC to DNA demethylation, we used in vitro PGC-like cells (PGCLCs). PGCLCs are specified from epiblast-like cells (EpiLCs) and exhibit the fundamental properties of migratory PGCs in vivo, including global DNA demethylation and chromatin reorganization (fig. S10) (10). TET1 and TET2 are both active in PGCs, so we generated PGCLCs carrying a doxycycline (DOX)-inducible compound microRNA (miR) knockdown of *Tet1* and *Tet2* (T-KD). We



Downloaded from <http://science.sciencemag.org/> on February 21, 2017

found that genes known to be demethylated in PGCs in vivo (8) also underwent DNA demethylation upon specification of control uninduced (–DOX) T-KD PGCLCs and in nontargeting (NT) miR PGCLCs (+/–DOX). In contrast, induction of *Tet1/Tet2* miR (+DOX) resulted in a substantial inhibition of DNA demethylation in PGCLCs but did not reduce the efficiency of their specification (Fig. 2E and fig. S16). Knockdown of *Tet1/Tet2* also inhibited DNA demethylation at long interspersed nuclear element 1 (LINE-1) sites and prevented the limited erasure of 5mC that occurs at intracisternal-A-particles (IAP) (Fig. 2F). These findings are important considering that both the maintenance and de novo DNA methylation systems are repressed in PGCs and PGCLCs (10), which likely accounts for some direct passive demethylation. Moreover, constitutive overexpression of catalytically active, but not catalytic mutant, TET1 and TET2 in PGCLCs promoted 5mC erasure to a greater extent (Fig. 2G). Thus, TET-mediated 5mC conversion is a key event toward DNA demethylation in PGCs.

The reprogramming of gonadal PGCs in vivo uniquely entails the complete erasure of genomic imprints (11). Analysis of imprinted gametic differentially methylated regions (gDMRs) ( $n = 21$ ) in PGCs revealed that erasure of 5mC is coupled to a significant increase of 5hmC enrichment (Fig. 3A). However, the precise timing of 5mC erasure is imprinted locus-specific. For example, the DMRs at *Kcnq1ot1* and *Igf2r* exhibit loss of 5mC by E10.5 relative to EpiSC (which represent ~50% allelic 5mC) and erasure by

E11.5 (Fig. 3B), whereas *Peg10* and *Peg3* remain methylated until E11.5 (Fig. 3C and fig. S17). Moreover, *Kcnq1ot1* and *Igf2r* are enriched in 5hmC by E10.5, whereas 5hmC enrichment at *Peg10* and *Peg3* is delayed until E11.5, suggesting that conversion to 5hmC follows a defined temporal order at imprinted DMRs, which dictates the timing of demethylation in PGCs. Indeed, we observed that other genomic regions also exhibited differential onset of 5mC erasure (compare *Peg10* DMR versus exon, Fig. 3C). Glu-qPCR analysis confirmed that the *Peg10* and *Peg3* DMRs maintained 5mC levels of 50% and 34%, respectively, in E10.5 PGCs, whereas *Kcnq1ot1* and *Igf2r* DMRs were already reduced to 21% and 25%, respectively (Fig. 3D). Glu-qPCR also established the quantitative enrichment of 5hmC at imprinted DMRs in PGCs. The cumulative data suggest that conversion of 5mC to 5hmC by TET1 and TET2 is a general mechanism for the erasure of imprints in PGCs.

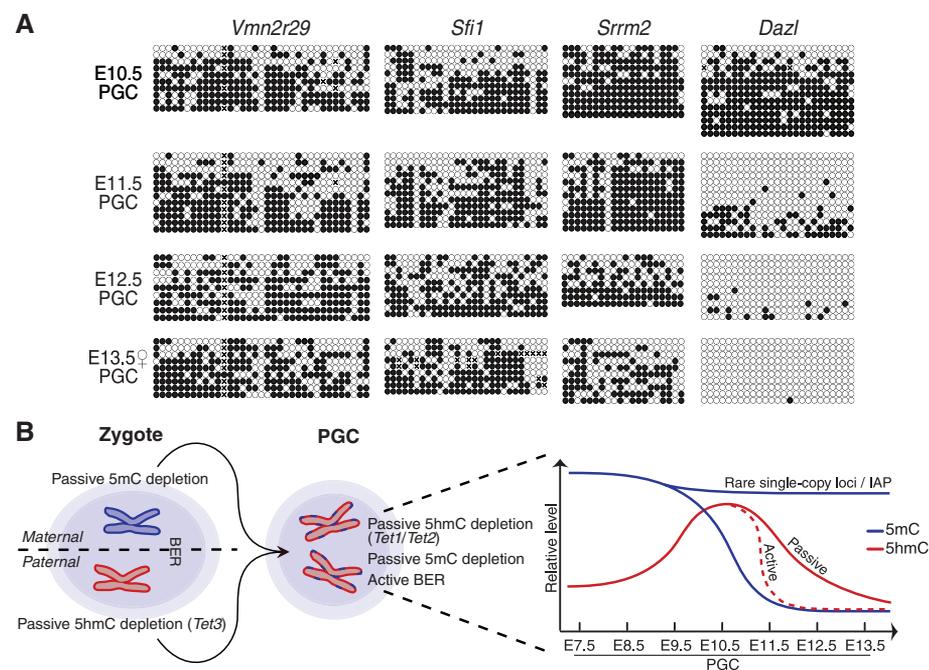
Conversion of 5mC to 5hmC at exons, promoters, and gDMRs in PGCs was followed by a protracted period of progressive 5hmC depletion between E11.5 and E13.5 (Figs. 2, A to D, and 3), suggesting a replication-coupled process (12). This prompted us to examine the rate of DNA demethylation between E10.5 and E13.5 quantitatively by using Glu-qPCR. Because demethylation commences asynchronously in PGCs, it is necessary to examine loci that have not initiated substantial 5mC erasure by E10.5, such as *Peg10* and *Peg3*. Because PGCs have an estimated cell cycle of ~16 hours between E10.5 and E13.5 (13), we would predict a reduction of DNA mod-

ification of ~threefold per 24 hours (1.5 population doublings) if the process is coupled to DNA replication. We observed that the rate of demethylation at *Peg10* ( $P = 0.0022$ ) and *Peg3* ( $P = 0.0019$ ) fits highly significantly with the predicted rate (Fig. 3E), suggesting that 5hmC may be removed from these loci by replication-coupled dilution. We obtained similar results for the *Dazl* promoter ( $P = 0.0014$ ).

We next asked whether any promoters or regulatory elements can escape the comprehensive 5mC reprogramming in PGCs. We screened for CpG islands (CGI) that remain methylated in female PGCs at E13.5, because these cells represent the lowest point of global demethylation (fig. S18) (14). We identified 11 CGIs with significant 5mC enrichment in E13.5 PGCs (figs. S19 and S20). Validation by bisulfite sequencing showed that the promoter CGIs of *Vmn2r29* and *Sfi1* and the exonic CGI of *Srrm2* were all methylated in PGCs at E10.5 and maintained CpG methylation throughout reprogramming (Fig. 4A).

To define the extent of 5mC erasure at single-base resolution, we performed whole-genome bisulfite sequencing (WGBS), which revealed that global CpG methylation is reduced to 2.2% in female E13.5 PGCs (fig. S21). However, we identified 4730 loci that escape demethylation (>40% 5mC) in PGCs, which are predominately repeat associated (>95%). Resistant loci predominantly correspond to IAP elements, but the IAPLTR1 subclass is significantly more methylated than any other (fig. S22). IAPLTR1 is the most active and hence hazardous IAP subclass to genomic integrity, suggesting specific systems are mobilized to maintain 5mC at IAPLTR1 during reprogramming to protect genome stability (15). We were unable to determine any unique sequence characteristics of the 233 single-copy loci with >40% 5mC, suggesting that positional context or chromatin structure may contribute to their escape from reprogramming. Indeed, “escapees” were often adjacent to IAP elements or telomeric regions. Considered with the recent observations that many regulatory elements can evade zygotic 5mC erasure (16, 17), our data suggest that rare but potentially functionally relevant 5mC epialleles could be inherited over multiple generations by evading erasure during both zygotic and PGC reprogramming.

We demonstrate here that comprehensive DNA demethylation in PGCs, including imprint erasure, entails conversion of 5mC to 5hmC, likely redundantly by TET1 and TET2. In vivo 5hmC conversion initiates asynchronously in PGCs between E9.5 and E10.5 and is largely complete by E11.5. The rate of progressive decline of 5hmC thereafter, both globally and at single-copy loci, is consistent with a replication-dependent mechanism of demethylation toward unmodified cytosines (Fig. 4B). In parallel to 5hmC conversion, repression of the de novo (*Dnmt3a/b*) and maintenance (*Uhrf1*) DNA methylation systems in PGCs prevents cyclical remethylation and simultaneously renders PGCs permissive for direct



**Fig. 4.** Inheritance of 5mC through reprogramming. **(A)** The *Vmn2r29*, *Sfi1*, and *Srrm2* CGIs escape reprogramming in PGCs. Open and solid circles represent unmethylated and methylated CpGs, respectively. *Dazl* is representative of demethylation at most loci. **(B)** Model for the mechanisms and dynamics of DNA demethylation in PGCs.

passive 5mC depletion (fig. S23) (18), which may contribute to the partial demethylation observed in *Tet1* and *Tet2* knockdown PGCs. Thus, whereas in zygotes 5mC reprogramming is mechanistically compartmentalized into TET3-mediated 5hmC conversion of the paternal genome and direct passive 5mC depletion on the maternal genome (12, 19–21), both of these mechanisms operate together in PGCs (Fig. 4B). In addition, up-regulation of the base excision repair (BER) pathway in PGCs may both protect against cumulative genetic damage and act as an auxiliary active demethylation mechanism, perhaps for specific loci (22, 23). Reprogramming in PGCs therefore involves multiple redundant mechanisms to reset the epigenome for totipotency, which accounts for the apparent fertility (albeit subfertile) of mice lacking individual components, such as *Tet1* (24). The existence of multiple mechanisms may also underpin the comprehensive nature of DNA demethylation in PGCs (3). Nonetheless, some rare single-copy sites of CpG methylation escape from 5mC erasure (25), which may pro-

vide mechanistic avenues for investigations into transgenerational epigenetic inheritance.

#### References and Notes

1. M. A. Surani, K. Hayashi, P. Hajkova, *Cell* **128**, 747 (2007).
2. P. Hajkova *et al.*, *Nature* **452**, 877 (2008).
3. J. A. Hackett, J. J. Zylitz, M. A. Surani, *Trends Genet.* **28**, 164 (2012).
4. M. Tahiliani *et al.*, *Science* **324**, 930 (2009).
5. S. Ito *et al.*, *Nature* **466**, 1129 (2010).
6. G. Ficz *et al.*, *Nature* **473**, 398 (2011).
7. S. Ito *et al.*, *Science* **333**, 1300 (2011); 10.1126/science.1210597.
8. D. M. Maatouk *et al.*, *Development* **133**, 3411 (2006).
9. J. A. Hackett *et al.*, *Development* **139**, 3623 (2012).
10. K. Hayashi, H. Ohta, K. Kurimoto, S. Aramaki, M. Saitou, *Cell* **146**, 519 (2011).
11. P. Hajkova *et al.*, *Mech. Dev.* **117**, 15 (2002).
12. A. Inoue, Y. Zhang, *Science* **334**, 194 (2011); 10.1126/science.1212483.
13. P. P. Tam, M. H. Snow, *J. Embryol. Exp. Morphol.* **64**, 133 (1981).
14. C. Popp *et al.*, *Nature* **463**, 1101 (2010).
15. C. Qin *et al.*, *Mol. Carcinog.* **49**, 54 (2010).
16. S. A. Smallwood *et al.*, *Nat. Genet.* **43**, 811 (2011).

17. J. Borgel *et al.*, *Nat. Genet.* **42**, 1093 (2010).
18. K. Kurimoto *et al.*, *Genes Dev.* **22**, 1617 (2008).
19. T. P. Gu *et al.*, *Nature* **477**, 606 (2011).
20. M. Wossidlo *et al.*, *Nat. Commun.* **2**, 241 (2011).
21. K. Iqbal, S. G. Jin, G. P. Pfeifer, P. E. Szabó, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 3642 (2011).
22. P. Hajkova *et al.*, *Science* **329**, 78 (2010).
23. S. Cortellino *et al.*, *Cell* **146**, 67 (2011).
24. M. M. Dawlaty *et al.*, *Cell Stem Cell* **9**, 166 (2011).
25. S. Guibert, T. Forné, M. Weber, *Genome Res.* **22**, 633 (2012).

**Acknowledgments:** We thank N. Miller for fluorescence-activated cell sorting analysis, F. Tang and W. Tang for experimental support, and G.-L. Xu for reagents. This work was funded by the Wellcome Trust (RG49135, RG44593, and 083563) and the Human Frontier Science Program. Sequencing data have been deposited in the Sequence Read Archive (SRA060914).

#### Supplementary Materials

www.sciencemag.org/cgi/content/full/science.1229277/DC1  
Materials and Methods  
Figs. S1 to S24  
References (25, 26)

24 August 2012; accepted 16 November 2012  
Published online 6 December 2012;  
10.1126/science.1229277

# Actin, Spectrin, and Associated Proteins Form a Periodic Cytoskeletal Structure in Axons

Ke Xu,<sup>1\*</sup> Guisheng Zhong,<sup>1\*</sup> Xiaowei Zhuang<sup>1,2,†</sup>

Actin and spectrin play important roles in neurons, but their organization in axons and dendrites remains unclear. We used stochastic optical reconstruction microscopy to study the organization of actin, spectrin, and associated proteins in neurons. Actin formed ringlike structures that wrapped around the circumference of axons and were evenly spaced along axonal shafts with a periodicity of ~180 to 190 nanometers. This periodic structure was not observed in dendrites, which instead contained long actin filaments running along dendritic shafts. Adducin, an actin-capping protein, colocalized with the actin rings. Spectrin exhibited periodic structures alternating with those of actin and adducin, and the distance between adjacent actin-adducin rings was comparable to the length of a spectrin tetramer. Sodium channels in axons were distributed in a periodic pattern coordinated with the underlying actin-spectrin-based cytoskeleton.

Actin plays critical roles in shaping and maintaining cell morphology, as well as in supporting various cellular functions, including cell motility, cell division, and intracellular transport (1). In neurons, actin is essential for the establishment of neuronal polarity, cargo transport, neurite growth, and stabilization of synaptic structures (2–4). Despite its importance, our understanding of actin structures in neurons remains incomplete. Electron microscopy has shown detailed actin ultrastructure in growth cones and dendritic spines (5, 6), in which

actin is the dominant cytoskeletal protein, but little is known about the organization of actin in the axonal and dendritic shafts (4). These neurites contain a high density of different types of cytoskeletal filaments, such as microtubules and neurofilaments (6–8). Hence, the challenge of resolving the organization of actin in axons and dendrites requires imaging tools with both high spatial resolution and molecular specificity.

A prototypical actin-spectrin-based cytoskeleton structure is found in red blood cells (erythrocytes) (9, 10), where actin, spectrin, and associated proteins form a two-dimensional (2D) polygonal network (mostly composed of hexagons and pentagons) underneath the erythrocyte membrane (11, 12). Spectrin analogs have been found in many other animal cells (9, 10), including neurons (13, 14). These analogs play important roles, ranging from regulation of the heartbeat

to stabilization of axons, formation of axon initial segments and nodes of Ranvier, and stabilization of synapses in neurons (9, 10, 15). An erythrocyte-like, polygonal lattice structure has been observed for spectrin in the *Drosophila* neuromuscular junction (16), and models similar to the erythrocyte cytoskeleton have also been proposed for other systems (10). However, the ultrastructural organization of spectrin in non-erythrocyte cells is largely unknown due to similar challenges in imaging.

Recent advances in superresolution fluorescence microscopy (17, 18) allow resolutions down to ~10 nm to be achieved with molecular specificity, providing a promising solution to the above challenges. In particular, superresolution studies of neurons have provided valuable structural and dynamic information of actin in dendritic spines (19–22). In this work, we used a superresolution fluorescence imaging method, stochastic optical reconstruction microscopy (STORM) (23–27), to study the 3D ultrastructural organization of actin and spectrin in neurons.

To image actin in neurons, we fixed cultured rat hippocampal neurons at various days in vitro (DIV) and labeled actin filaments with phalloidin conjugated to a photoswitchable dye, Alexa Fluor 647 (Invitrogen, Carlsbad, CA) (Fig. 1) (28). To identify axons and dendrites, we immunolabeled MAP2, a microtubule-associated protein enriched in dendrites, or NrCAM, a cell adhesion molecule found in the initial segments of axons (15), using a dye of a different color (Fig. 1) (28). In the conventional fluorescence images (Fig. 1, A, D, and F), MAP2 specifically stained dendrites, and NrCAM specifically labeled the initial segments of axons, whereas actin was found in both dendrites and axons.

Next, we used 3D STORM (27) with a dual-objective astigmatism-imaging scheme (28, 29)

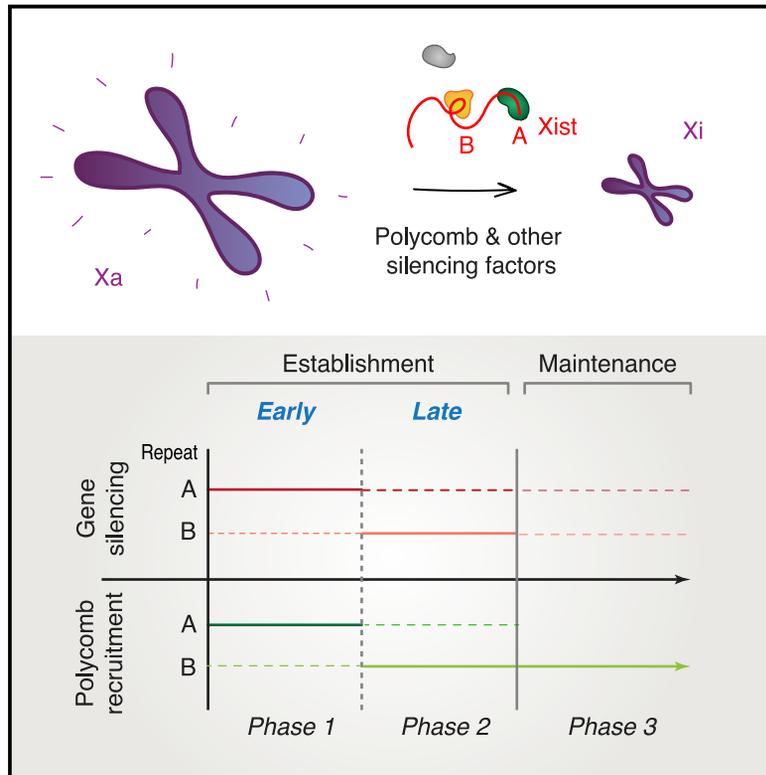
<sup>1</sup>Howard Hughes Medical Institute (HHMI), Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA. <sup>2</sup>Department of Physics, Harvard University, Cambridge, MA 02138, USA.

\*These authors contributed equally to this work.

†To whom correspondence should be addressed. E-mail: zhuang@chemistry.harvard.edu

# Xist Repeats A and B Account for Two Distinct Phases of X Inactivation Establishment

## Graphical Abstract



## Authors

David Cognori, Hongjae Sunwoo,  
Danni Wang, Chen-Yu Wang,  
Jeannie T. Lee

## Correspondence

lee@molbio.mgh.harvard.edu

## In Brief

Cognori et al. show that X inactivation establishment is a biphasic process with distinct genetic requirements. Repeat A initiates Polycomb recruitment and gene silencing, whereas repeat B stabilizes them. Surprisingly, X inactivation can initiate without repeat B. Without repeat A, however, differentiating female cells lose one X chromosome to overcome loss of silencing.

## Highlights

- X inactivation establishment is a biphasic process requiring Xist repeats A and B
- Polycomb complexes can initially be recruited without repeat B
- Repeat A initiates Polycomb recruitment and X silencing, whereas B stabilizes them
- Frequency of X chromosome loss (XO state) is heightened by deleting repeat A



## Short Article

# Xist Repeats A and B Account for Two Distinct Phases of X Inactivation Establishment

David Cognori,<sup>1,2,3</sup> Hongjae Sunwoo,<sup>1,2,3</sup> Danni Wang,<sup>1,2</sup> Chen-Yu Wang,<sup>1,2</sup> and Jeannie T. Lee<sup>1,2,4,\*</sup><sup>1</sup>Department of Molecular Biology, Massachusetts General Hospital, Boston, MA 02114, USA<sup>2</sup>Department of Genetics, Harvard Medical School, Boston, MA 02115, USA<sup>3</sup>These authors contributed equally<sup>4</sup>Lead Contact\*Correspondence: [lee@molbio.mgh.harvard.edu](mailto:lee@molbio.mgh.harvard.edu)<https://doi.org/10.1016/j.devcel.2020.05.021>

## SUMMARY

X chromosome inactivation (XCI) is a global silencing mechanism by which XX and XY mammals equalize X-linked gene dosages. XCI begins with an establishment phase during which Xist RNA spreads and induces *de novo* heterochromatinization across a female X chromosome and is followed by a maintenance phase when multiple epigenetic pathways lock down the inactive X (Xi) state. Involvement of Polycomb repressive complexes 1 and 2 in XCI has been intensively studied but with conflicting conclusions regarding their recruitment and role in Xi silencing. Here, we reveal that establishment of XCI has two phases and reconcile the roles that Xist repeats A and B play in gene silencing and Polycomb recruitment. Repeat A initiates both processes, whereas repeat B bolsters or stabilizes them thereafter. Once established, XCI no longer requires repeat A during maintenance. These findings integrate disparate studies and present a unified view of Xist's role in Polycomb-mediated silencing.

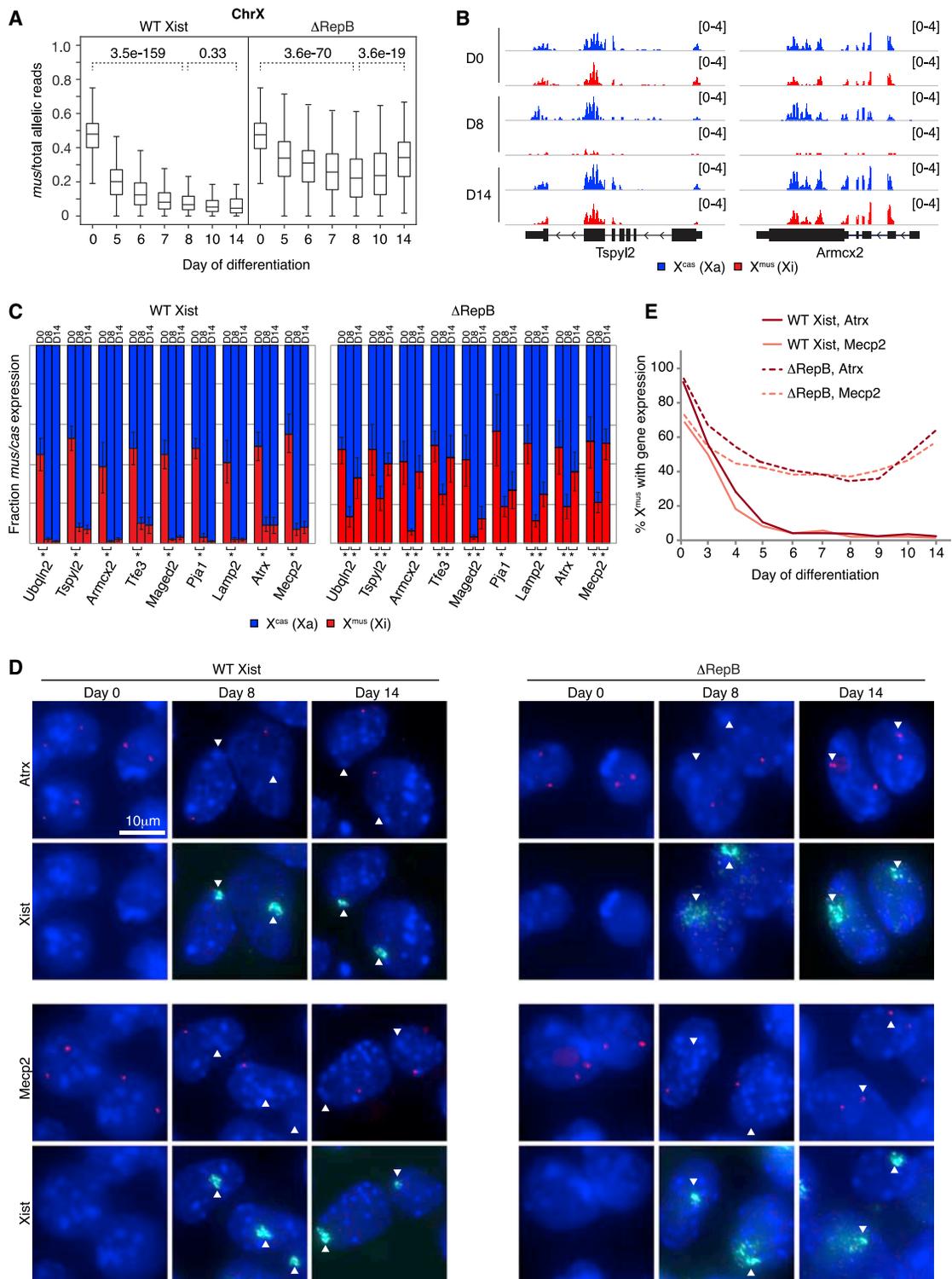
## INTRODUCTION

X chromosome inactivation (XCI) is initiated by a chromosome-counting mechanism that triggers dosage compensation only when there is more than one X chromosome in the early mammalian embryo (Stamer and Magnuson, 2009; Disteché, 2016; Jégu et al., 2017). Once committed to XCI, one X chromosome is selected to undergo a series of distinct epigenetic changes that are separable into establishment and maintenance phases. The long noncoding RNA Xist is instrumental in both phases. During establishment, Xist spreads across the future inactive X (Xi) and carries out several essential functions: (1) eviction of activating factors (Minajigi et al., 2015; Jégu et al., 2019), (2) recruitment of silencing factors, including Polycomb repressive complexes 1 and 2 (PRC1, PRC2) (Wang et al., 2001; Plath et al., 2003; Silva et al., 2003; de Napoles et al., 2004; Kohlmaier et al., 2004; Schoeftner et al., 2006; Zhao et al., 2008; Pintacuda et al., 2017; Cognori et al., 2019), and (3) global transformation of the 3D chromosomal structure (Rao et al., 2014; Minajigi et al., 2015; Giorgetti et al., 2016; Wang et al., 2018). The establishment phase involves building a chromosomal memory that persists through the ensuing maintenance phase and ensures stable retention of repressive heterochromatin (Kohlmaier et al., 2004; Simon et al., 2013). During this early window, the incipient Xi is easily perturbed and reactivated (Wutz and Jaenisch, 2000; Kohlmaier et al., 2004); but once established, the Xi is remarkably stable and becomes more difficult to reactivate (Brown and Willard, 1994; Csankovszki et al., 2001; Minajigi et al., 2015; Carrette

et al., 2017; Adrianse et al., 2018). Even so, the Xi continues to require Xist to fully maintain its silent configuration. In somatic female cells, loss of Xist RNA results in reversal of some 3D chromosomal structures (Minajigi et al., 2015; Wang et al., 2019), loss of repressive Polycomb marks (Plath et al., 2003; Zhang et al., 2007; Nozawa et al., 2013; Pintacuda et al., 2017; Cognori et al., 2019), and partial reactivation of Xi genes (Zhang et al., 2007; Yildirim et al., 2013; Bhatnagar et al., 2014; Carrette et al., 2017; Adrianse et al., 2018). The establishment and maintenance phases are therefore biologically and functionally distinct.

There has been considerable interest in understanding the mechanistic differences between the earlier and more dynamic period of XCI (establishment) versus the later and more stable period of XCI (maintenance). Because Xist RNA and XCI are widely viewed as paradigms for understanding Polycomb-mediated epigenetic regulation (Stamer and Magnuson, 2009; Disteché, 2016; Jégu et al., 2017), similar concepts may extend to autosomal gene regulation as well. Moreover, a deeper understanding could inform growing interest in pharmacological Xi reactivation as a method of treating X-linked neurodevelopmental disorders (Bhatnagar et al., 2014; Carrette et al., 2017, 2018; Sripathy et al., 2017; Adrianse et al., 2018). Notably, for Rett Syndrome, reactivation of the wild-type (WT) *MECP2* allele on the Xi could potentially restore expression of the missing protein for therapeutic impact. More comprehensive knowledge of how the Xi progresses through various stages would enormously benefit design of treatment approaches.





**Figure 1. Xi Gene Silencing Can Be Initiated, but Not Maintained, without Xist Repeat B**

(A) Boxplots of allele-specific RNA-seq time course showing abortive Xi gene silencing in  $\Delta$ RepB versus WT Xist differentiating female ESCs. Two-tailed Student's t test, p values for pairwise comparison as shown.

(B) Zoom in of allele-specific RNA-seq tracks showing reads from individual X-linked genes.

(C) Allele-specific RT-qPCR showing relative expression from each allele for several X-linked genes in WT Xist and  $\Delta$ RepB differentiating female ESCs. Error bars show standard deviation between 3 biological replicates. Two-tailed Student's t test, asterisks indicate  $p < 0.05$ .

(legend continued on next page)

Two motifs within Xist RNA have been linked to the processes of gene silencing and Polycomb recruitment: repeats A and B. Regarding gene silencing, the importance of repeat A is universally accepted (Wutz et al., 2002; Minks et al., 2013), whereas reports on repeat B have argued for varying degrees of silencing defects upon its deletion—though there is agreement that the defects are not as severe as those associated with loss of repeat A (Pintacuda et al., 2017; Bousard et al., 2019; Colognori et al., 2019; Nesterova et al., 2019). Regarding Polycomb recruitment, the relative roles of repeats A and B have yet to be resolved, with some observations supporting repeat A as an important factor (Kohlmaier et al., 2004; Zhao et al., 2008; Davidovich et al., 2013, 2015; Kaneko et al., 2013; Cifuentes-Rojas et al., 2014; da Rocha et al., 2014; Lee et al., 2019) and others arguing that the repeat B region alone is responsible (da Rocha et al., 2014; Pintacuda et al., 2017; Nesterova et al., 2019). Finally, the functional relationship between Polycomb recruitment and gene silencing remains unclear. Some reports show that loss of Polycomb recruitment has only a minor effect on Xi silencing (Kalantry and Magnuson, 2006; Leeb and Wutz, 2007; Bousard et al., 2019), whereas others show a significant effect (Wang et al., 2001; Almeida et al., 2017; Pintacuda et al., 2017; Colognori et al., 2019; Nesterova et al., 2019).

These disparate findings have been difficult to reconcile and urge further investigation to elucidate the functional relationships between repeats A and B in Xi gene silencing and Polycomb recruitment. Here, we reconcile disparate models by demonstrating the existence of two discrete phases during XCI establishment. We ascribe repeats A and B functions in the two establishment phases, with phase 1 being primarily repeat A dependent and phase 2 being primarily repeat B dependent. We provide evidence that XCI can thus best be characterized as a three-part process with discrete genetic requirements and epigenetic outcomes.

## RESULTS

### Xi Gene Silencing Is Initiated but Not Maintained in Female Cells Lacking Xist Repeat B

Recent work has shown that Xist repeat B plays a major role in recruiting PRC1 and PRC2 to the Xi for proper establishment of silencing (Pintacuda et al., 2017; Colognori et al., 2019; Nesterova et al., 2019). In these studies, deleting repeat B significantly impaired Xist-mediated gene silencing and essentially abolished deposition of PRC1/PRC2 histone modifications (H2AK119ub/H3K27me3) across the Xi. A separate study, however, reported that Xist transcripts lacking repeat B do not show significantly impaired silencing and exhibit some residual H2AK119ub/H3K27me3 at Xi regions (Bousard et al., 2019). These seemingly contradictory results warrant closer examination. Indeed, because many of these studies were conducted using an autosomal transgene or inducible Xist system in male embryonic stem cells (ESCs), or on the already-established Xi in

fibroblasts, the effect of deleting repeat B has not been explored fully in a physiological context during ESC differentiation when XCI is established *de novo*.

Here, we examined female ESCs carrying WT Xist or Xist lacking the repeat B region ( $\Delta$ RepB, Figure S1A; Colognori et al., 2019) as they underwent differentiation. Importantly, our parental cell line is a *Mus musculus/Mus castaneus* hybrid, enabling us to distinguish the two X chromosomes by genetic variants between strains. Moreover,  $X^{\text{mus}}$  carries a *Tsix* mutation forcing it to become the future Xi in  $\sim 95\%$  of cells, rather than a random choice between  $X^{\text{mus}}$  and  $X^{\text{cas}}$  (Ogawa et al., 2008). All cell lines were extensively validated by several means to ensure the following: (1) the identity of the deletion, (2) that it occurs on  $X^{\text{mus}}$ , (3) that Xist RNA levels and splicing are not altered, (4) that it does not affect preferential Xist expression from (and thus inactivation of)  $X^{\text{mus}}$  over  $X^{\text{cas}}$ , and (5) that it does not interfere with ESC differentiation. First, Sanger sequencing across the deleted region, while using SNPs to determine allelic identity, ensured deletion of the expected sequence on  $X^{\text{mus}}$  (Table S1). Second, two-color RNA fluorescence *in situ* hybridization (FISH) using one probe specifically targeting the deleted sequence and another targeting a control exon 7 sequence ensured Xist transcripts lacked the deleted sequence (Figure S1B). Third, RT-qPCR for different regions across Xist again confirmed loss of the deleted region without affecting adjacent ones or overall Xist levels (Figure S1C). Fourth, RT-qPCR for Xist using allele-specific (SNP-targeting) primer sets verified  $\sim 95\%$  expression from  $X^{\text{mus}}$  (Figure S1D). Lastly, characteristic expression patterns of common differentiation markers (*Sox2*, *Oct4*, and *Gata4*) were indistinguishable between WT Xist and deletion cells, suggesting normal progression of differentiation across 14 days (Figure S1E).

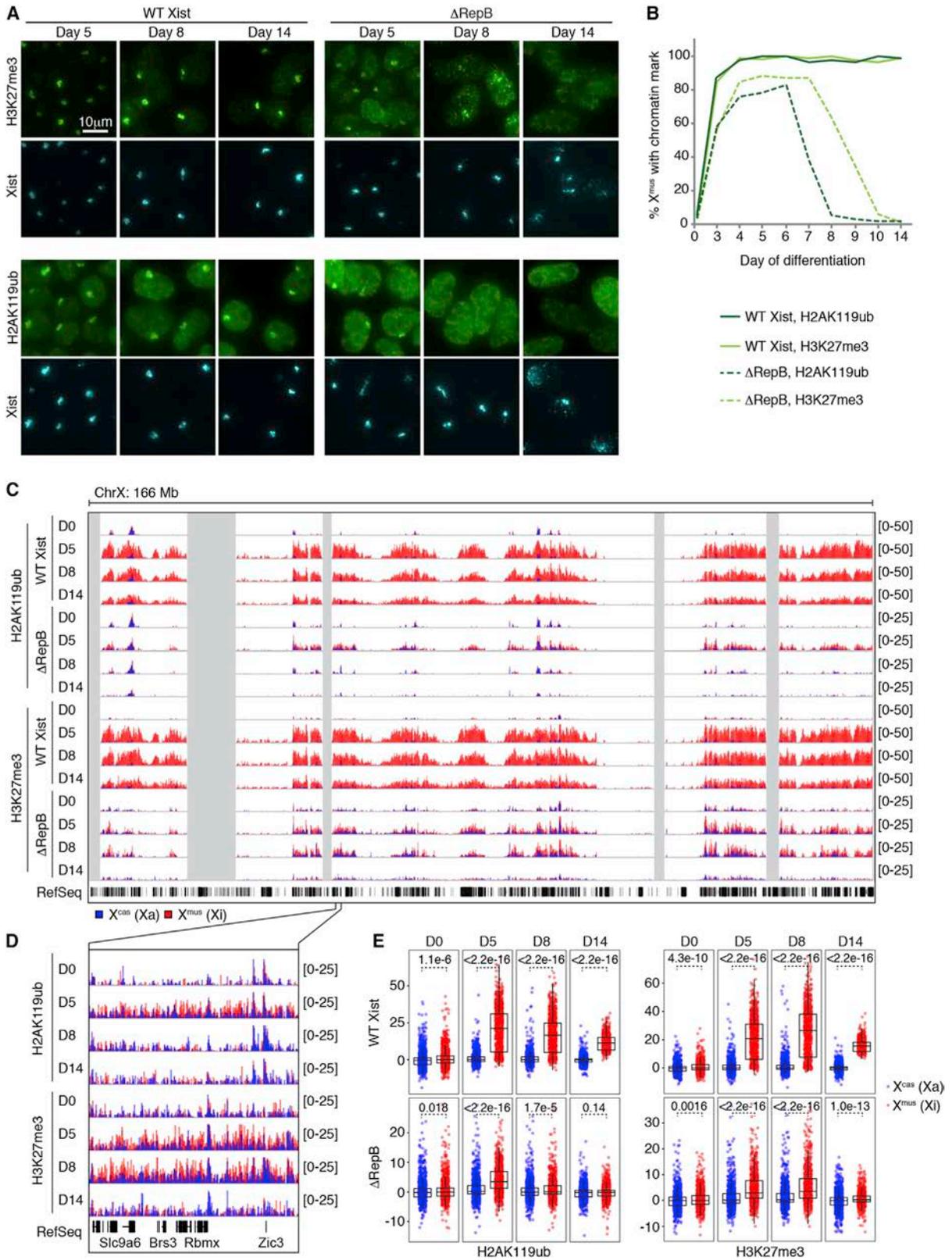
We began by performing a transcriptomic time course for  $X^{\text{mus}}$  (Xi) expression. In WT Xist cells, XCI occurred as expected between days 0–14 of cell differentiation, evidenced by the progressive decrease in activity from  $X^{\text{mus}}$  (Figure 1A). By contrast,  $\Delta$ RepB cells showed a biphasic profile: Between days 0–8, the  $\Delta$ RepB chromosome could initiate noticeable silencing, as demonstrated by the progressive decrease in  $X^{\text{mus}}$  transcripts to  $<20\%$  X-linked gene expression by day 8 (Figure 1A). Beyond day 8, this initial silencing eroded such that, by day 14,  $X^{\text{mus}}$  activity reverted back to  $\sim 40\%$  of X-linked allelic reads. Autosomes, as represented by chromosome 13, did not display any sign of skewed gene expression in  $\Delta$ RepB cells (Figure S2A). Examination of specific X-linked genes (e.g., *Tspyl2* and *Armcx2*) confirmed the overall trends (Figure 1B). These findings suggest a hidden biphasic dynamic during XCI establishment.

We validated these findings using two orthogonal approaches. First, we performed time course allele-specific RT-qPCR to quantify  $X^{\text{mus}}:X^{\text{cas}}$  expression for nine genes located at various positions along the X and that are normally subject to XCI (Figure 1C). Consistent with our transcriptome-wide analysis, all nine genes showed a statistically significant

(D) Atrx and Mecp2 nascent RNA FISH combined with Xist RNA FISH in WT Xist and  $\Delta$ RepB differentiating female ESCs. Diffuse Xist cloud morphology caused by disruption of repeat B/Polycomb was previously described (Colognori et al., 2019). Arrowheads mark positions of Xist cloud.

(E) Quantification of (D). Note that no Xist clouds are present at day 0 to indicate  $X^{\text{mus}}$ , but expression is inferred by the presence of two pinpoint signals.  $n > 100$  per time point.

See also Figures S1 and S2.



(legend on next page)

drop from biallelic expression at day 0 to monoallelic expression from  $X^{\text{mus}}$  at days 8 and 14 in WT Xist cells. By contrast,  $\Delta\text{RepB}$  cells showed a similar drop from day 0 to 8, but a statistically significant reversion back toward biallelic expression at day 14 for eight out of nine genes assayed. Second, we performed nascent RNA FISH for two of the above genes, *Atrx* and *Mecp2*, in combination with Xist RNA FISH to indicate the  $X^{\text{mus}}$  (Figures 1D and 1E). In WT cells, there was a progressive loss of nascent transcription overlapping the Xist cloud between days 0–14, as expected. On the other hand,  $\Delta\text{RepB}$  cells demonstrated an initial silencing of *Atrx* and *Mecp2* between days 0–8 but regained expression between days 8–14. Notably, Xist clouds appeared dispersed on the  $\Delta\text{RepB}$  chromosome, as previously reported (Colognori et al., 2019). Together these data show that XCI does initiate—though transiently and not to a full extent—despite the absence of repeat B, thereby implicating additional Xist motifs in initiating silencing. However, the data also demonstrate that the establishment process cannot be completed without repeat B.

### Polycomb Recruitment to Xi Is Initiated but Cannot Be Maintained without Xist Repeat B

As repeat B has now been associated with Polycomb binding to the Xi (Pintacuda et al., 2017; Bousard et al., 2019; Colognori et al., 2019; Nesterova et al., 2019), we performed immunofluorescence (IF) for the H2AK119ub and H3K27me3 histone marks (deposited by PRC1 and PRC2, respectively) across a time course beginning with the earliest appearance of Xist clouds at day 3. Among nearly all WT Xist-positive cells, strong foci of H2AK119ub and H3K27me3 staining were observed on the  $X^{\text{mus}}$  (Figures 2A and 2B), as expected. Unexpectedly,  $\Delta\text{RepB}$  cells also displayed noticeable enrichment of both repressive marks in Xist-positive cells during early time points (Figures 2A and 2B). These data suggest that, contrary to recent reports, deleting repeat B does not fully abrogate the initiation of Polycomb recruitment.

The PRC2 mark, H3K27me3, was observed overlapping with Xist clouds in ~85% of Xist-positive  $\Delta\text{RepB}$  cells and remained stable between days 4–7, after which time the signal began to fade away to undetectable levels by day 10, suggesting that repeat B is required for maintaining PRC2. Yet at the same time, PRC2 must come independently of Repeat B during the initiation phase. In parallel, the PRC1 mark, H2AK119ub, remained stable between days 4–6 in ~80% of Xist-positive cells, after which point it also faded away and became undetectable by day 8, suggesting repeat B's role in stabilizing PRC1 as well. The differential kinetics for H3K27me3 with respect to H2AK119ub are similar to the previous observation that depleting HNRNPK

(repeat B's direct binding partner) has a more immediate effect on PRC1 than PRC2 (Pintacuda et al., 2017; Colognori et al., 2019; Żylicz et al., 2019). This delay may reflect order of recruitment, difference in turnover rate for each mark, or differential requirements for PRC1/2 recruitment. Importantly, the rise and fall of repressive histone marks coincided with the initiation and erosion of gene silencing (inflection around day 8) (Figure 1) and support the idea of a biphasic dynamic during XCI establishment.

At higher resolution, allele-specific chromatin immunoprecipitation sequencing (ChIP-seq) for H2AK119ub and H3K27me3 agreed with the IF data (Figures 2C–2E). WT Xist cells showed characteristic enrichment of H2AK119ub and H3K27me3 across  $X^{\text{mus}}$  between days 5–14 (Figure 2C). Consistent with IF data,  $\Delta\text{RepB}$  cells also accumulated the two marks on  $X^{\text{mus}}$  (red track) compared with  $X^{\text{cas}}$  (blue track), although overall levels were lower than on the WT Xist  $X^{\text{mus}}$  (Figures 2C and 2D). Quantification of coverages over genes revealed significant enrichment of H2AK119ub at day 5 but less so at days 8–14 (Figure 2E)—consistent with IF data. The H3K27me3 mark was enriched on the  $\Delta\text{RepB}$   $X^{\text{mus}}$  on day 5 and persisted longer to day 8 but less so by day 14 (Figure 2E). When examined individually, there were no obvious differences in coverage dynamics for genic versus intergenic regions, or genes silenced by XCI versus those non-expressed to begin with (at least at this temporal resolution), though genes silenced by XCI exhibited a stronger gain in H2AK119ub and H3K27me3 during the inactivation process (Figures S2B and S2C).

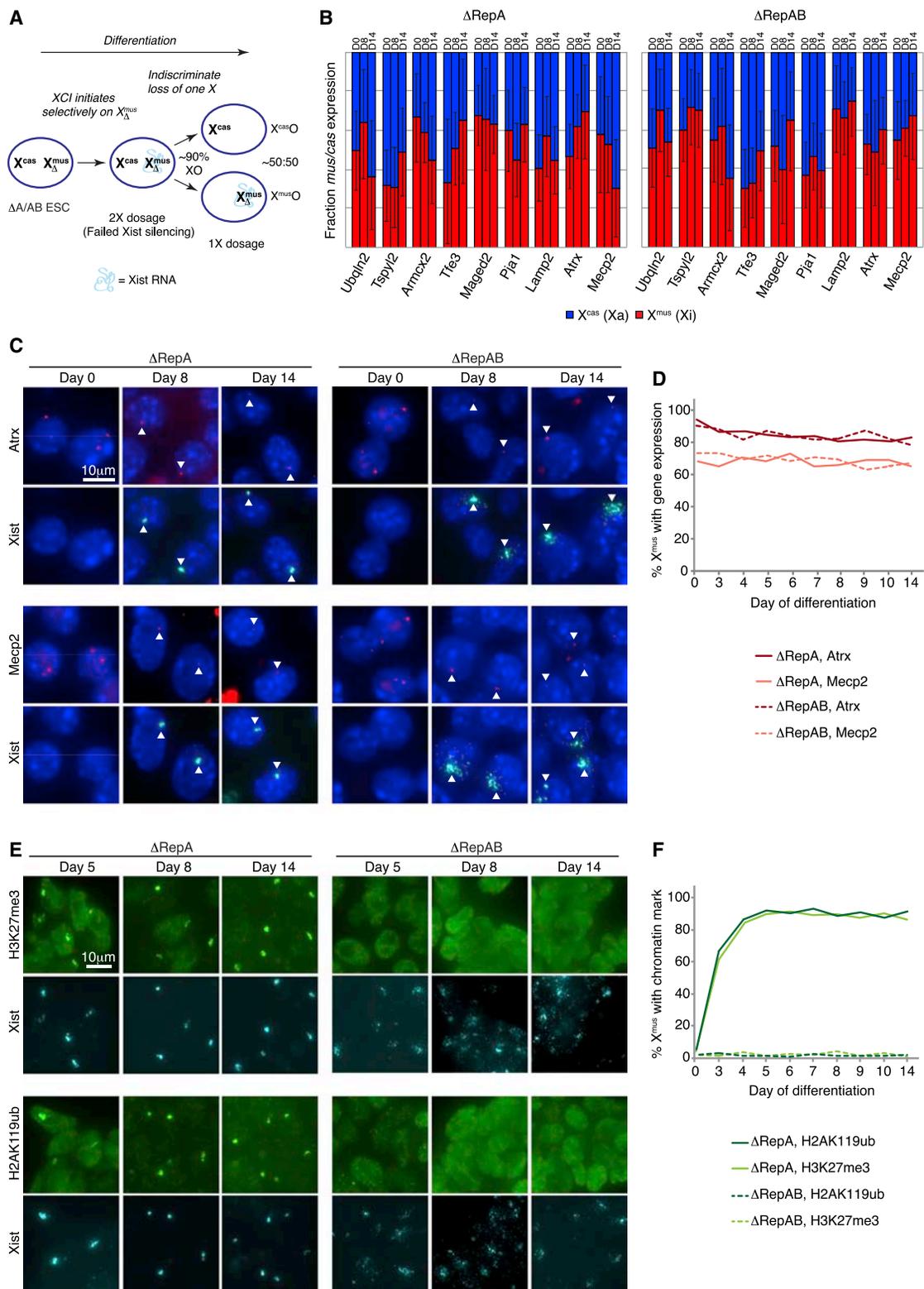
Taken together, these data demonstrate several crucial points: (1) Polycomb complexes can be recruited and XCI can initiate without repeat B; (2) without repeat B, however, Polycomb recruitment and gene silencing cannot be fully established or stabilized; (3) the establishment of gene silencing and Polycomb recruitment can therefore be characterized as “biphasic.” The first phase is primarily repeat B independent, whereas the second is repeat B dependent. This discovery left open the question of what additional Xist motifs are essential for the first phase of establishing silencing.

### Failed Xi Gene Silencing and Selective Pressure toward X Aneuploidy in Cells Lacking Xist Repeat A

The repeat A motif has been reported to play a role in both Xi silencing and Polycomb recruitment (Wutz et al., 2002; Zhao et al., 2008; Minks et al., 2013; Cifuentes-Rojas et al., 2014; Bousard et al., 2019; Lee et al., 2019; Nesterova et al., 2019). However, various deletions of repeat A have produced different findings, resulting in a lack of consensus regarding its *in vivo* role. This is due partly to the fact that existing deletions cover different sequences around repeat A and have been

#### Figure 2. Xi Polycomb Recruitment Can Be Initiated, but Not Maintained, without Xist Repeat B

- (A) H2AK119ub and H3K27me3 IF combined with Xist RNA FISH in WT Xist and  $\Delta\text{RepB}$  differentiating female ESCs.  
 (B) Quantification of (A). Note that no Xist clouds are present at day 0 to indicate  $X^{\text{mus}}$ , but cells accordingly show no focal enrichment of H2AK119ub or H3K27me3.  $n > 100$  per time point.  
 (C) Allele-specific H2AK119ub/H3K27me3 ChIP-seq time course in WT Xist and  $\Delta\text{RepB}$  differentiating female ESCs. Note the 2-fold difference in y axis scaling between WT Xist and  $\Delta\text{RepB}$  tracks.  
 (D) Zoom in of allele-specific ChIP-seq tracks in (C).  
 (E) Boxplots quantifying allele-specific ChIP-seq coverage over X-linked genes. Wilcoxon rank-sum test, p values as shown.  
 See also Figure S2.



**Figure 3. Repeat A Is Required for Gene Silencing and Contributes to Early Polycomb Recruitment on Xi**

(A) Repeat A or AB deletion leads to loss of an X chromosome (either  $X^{cas}$  or  $X^{mus}$ ) in  $\sim 90\%$  of cells over the course of differentiation.

(B) Allele-specific RT-qPCR showing relative expression from each allele for several X-linked genes in  $\Delta RepA$  and  $\Delta RepAB$  differentiating female ESCs. Error bars show standard deviation between 3 biological replicates. Two-tailed Student's t test, asterisks indicate  $p < 0.05$ .

(legend continued on next page)

analyzed in different contexts, such as on an autosomal transgene and/or in male cells. Previous observations also suggested that sequences within or around repeat A can influence Xist expression and/or splicing (Wutz et al., 2002; Zhao et al., 2008; Hoki et al., 2009; Royce-Tolland et al., 2010; Colognori et al., 2019).

*In vivo* investigation of repeat A's role in XCI necessitates a discrete deletion in the endogenous context in female cells. Here, we created a distinct  $\Delta$ RepA clone in female ESCs that removes a minimal region containing the motif without affecting Xist splicing or expression levels (Figures S1A–S1C). As earlier, we carried out extensive validation of our cell line to ensure  $\Delta$ RepA Xist was expressed selectively from  $X^{\text{mus}}$  during differentiation (Table S1; Figures S1B–S1E)—though  $\Delta$ RepA Xist clouds often appeared smaller than WT Xist clouds, as noted previously (Ha et al., 2018). Interestingly, roughly half of cells lacked any Xist cloud over the course of differentiation (Figures S1C and S3A). Follow-up analysis showed that this was due to loss of an X chromosome in most cells, occurring specifically during cell differentiation. While 95% of  $\Delta$ RepA cells carried two X chromosomes on day 0, only 9% retained them both by day 14 (Figure S3B). This contrasted sharply with differentiation of WT Xist or  $\Delta$ RepB cells, which consistently retained their XX status. We suspect that a crucial function for repeat A during differentiation precludes survival (or competitive advantage) of XX cells lacking it. Indeed, because the  $\Delta$ RepA mutation renders Xist unable to properly silence the Xi in cis (see below), there may be strong selective pressure to lose one X during differentiation in order to achieve proper 1 × gene dosage (Figure 3A). Interestingly,  $X^{\text{mus}}$  (selectively expressing  $\Delta$ RepA Xist) and  $X^{\text{cas}}$  (carrying a WT but non-expressed copy of Xist) were lost with roughly equal probability, as determined by allele-specific PCR of genomic DNA from the differentiated cell population (Figure S3B). This lack of discrimination could explain why roughly half of cells were Xist-positive ( $X^{\text{mus}}\text{O}$ ), and half were Xist-negative ( $X^{\text{cas}}\text{O}$ ), and further supports the idea that Xi<sup>ΔA</sup> and Xa are functionally interchangeable in supplying cells with the necessary X-linked gene products (Figure 3A). We also point out that, at least in the context of our *Tsix* mutant background, once an early decision was made to express (mutant) *Xist* from and silence  $X^{\text{mus}}$  (even if unsuccessful), this decision appeared irreversible: Xist expression did not “switch” alleles despite a capable WT *Xist* copy on  $X^{\text{cas}}$  and continued to be expressed from  $X^{\text{mus}}$  even after subsequent loss of  $X^{\text{cas}}$ . Notably, the *Xist* allele expressed from  $X^{\text{mus}}$  is mutated ( $\Delta$ RepA) and cannot initiate silencing on  $X^{\text{mus}}$ . Thus, in these  $X^{\text{mus}}\text{O}$  cells, an overall “count” of one Xa and dosage compensation were preserved (Figures S1B–S1D). The acute tendency to become XO is further testament to the importance of RepA for female cells undergoing dosage compensation.

Due to the stochastic loss of  $X^{\text{cas}}$  or  $X^{\text{mus}}$  and mixture of XO/XX cells during  $\Delta$ RepA ESC differentiation, we could not pursue genomic analyses such as RNA-seq and ChIP-seq. However, in spite of large variation between replicates, we were clearly able to see expression from both alleles for nine X-linked genes using allele-specific RT-qPCR (Figure 3B), indicative of failed Xi silencing. To rule out that apparent biallelic expression could be an artifact of cellular mosaicism, we examined nascent transcription at the single-cell level using RNA FISH. Within the fraction of Xist-positive cells (most being  $X^{\text{mus}}\text{O}$ , some being  $X^{\text{mus}}X^{\text{cas}}$ ), *Atrx* and *Mecp2* demonstrated clear failure to be silenced throughout the entire 14-day differentiation time course, despite an overlying Xist cloud (Figures 3C and 3D). These results reaffirm the consensus in the field that repeat A is critical for gene silencing (Wutz et al., 2002).

### An Early Wave of Xi Polycomb Recruitment Requires Xist Repeat A

To address whether repeat A is required for Polycomb targeting, we inspected H2AK119ub and H3K27me3 modifications in IF experiments. In Xist-positive  $\Delta$ RepA differentiating ESCs, a focus of H2AK119ub and H3K27me3 could still be observed over the Xist domain throughout the time course (Figures 3E and 3F). We conclude that, at the cytological level, loss of repeat A alone is insufficient to abolish bulk enrichment of these marks, consistent with previous transgenic studies on autosomes (Plath et al., 2003; Kohlmaier et al., 2004; da Rocha et al., 2014). Without ChIP-seq analysis, however, we could not rule out finer defects that might be present at the molecular level despite clear foci at the cytological level, as demonstrated previously (Wang et al., 2018, 2019; Żylicz et al., 2019). Indeed, recent data have suggested that repeat A is necessary for spreading both H2AK119ub and H3K27me3 into active genic regions (Bousard et al., 2019; Żylicz et al., 2019).

Additional support for repeat A function came from female ESCs carrying deletions of both repeats A and B on  $X^{\text{mus}}$  (Table S1; Figures S1A–S1E). Similar to  $\Delta$ RepA,  $\Delta$ RepAB differentiating ESCs exhibited a cluster of Xist RNA in roughly half of cells (Figure S3A) and had heightened tendency to become XO during differentiation, with only 8% of cells retaining both X chromosomes by day 14 (Figures 3A and S3B). Furthermore,  $\Delta$ RepAB cells failed to undergo gene silencing, as shown by persistence of biallelic *Mecp2* and *Atrx* expression in allele-specific RT-qPCR and nascent RNA FISH experiments (Figures 3B–3D). Notably, Xist cloud dispersal seen in  $\Delta$ RepB cells (Figures 1D and S1B) (Colognori et al., 2019) became exacerbated by the simultaneous deletion of repeat A, suggesting repeat A may also play a role in localizing and/or spreading Xist on the Xi. However, in contrast to both the  $\Delta$ RepA and  $\Delta$ RepB single deletions, the  $\Delta$ RepAB double

(C) *Atrx* and *Mecp2* nascent RNA FISH combined with Xist RNA FISH in  $\Delta$ RepA and  $\Delta$ RepAB differentiating female ESCs. Only Xist-positive cells ( $X^{\text{mus}}X^{\text{cas}}$  and  $X^{\text{mus}}\text{O}$ ) were examined. Arrowheads mark positions of Xist cloud.

(D) Quantification of (C). Note that no Xist clouds are present at day 0 to indicate  $X^{\text{mus}}$ , but expression is inferred by the presence of two pinpoint signals.  $n > 100$  per time point.

(E) H2AK119ub and H3K27me3 IF combined with Xist RNA FISH in  $\Delta$ RepA and  $\Delta$ RepAB differentiating female ESCs. Note the diffuse Xist cloud morphology caused by disruption of repeat B/Polycomb (Figure 1D) (Colognori et al., 2019) appears exacerbated by additional loss of repeat A/Polycomb (but not by repeat A loss alone). Only Xist-positive cells ( $X^{\text{mus}}X^{\text{cas}}$  and  $X^{\text{mus}}\text{O}$ ) were examined.

(F) Quantification of (E). Note that no Xist clouds are present at day 0 to indicate  $X^{\text{mus}}$ , but cells accordingly show no focal enrichment of H2AK119ub or H3K27me3.  $n > 100$  per time point.

See also Figures S1 and S3.

deletion completely abolished H2AK119ub and H3K27me3 signals on the Xi, as determined by IF (Figures 3E and 3F). This was true throughout the entire differentiation time course. Thus, repeats A and B both contribute to establishing Polycomb recruitment and gene silencing during the early window of XCI.

### Continued Polycomb Recruitment during XCI Maintenance Requires Xist Repeat B but Not A

Given the unexpected relationship between repeats A and B during XCI establishment, we next investigated their roles in XCI maintenance by examining similar  $\Delta$ RepA and  $\Delta$ RepB deletions in female MEFs. Our parental MEF cell line is again a *Mus musculus/Mus castaneus* hybrid, but one that became tetraploid post-XCI and thus carries 2 Xi's ( $X^{\text{mus}}$ ) and 2 Xa's ( $X^{\text{cas}}$ ) (Yildirim et al., 2011). IF for H2AK119ub/H3K27me3 revealed that loss of repeat B alone was sufficient to abolish both marks from Xi (Figure 4A), as previously observed (Colognori et al., 2019). On the other hand, deletion of repeat A alone had no apparent cytological effect (Figure 4A). To obtain higher resolution information, we performed allele-specific ChIP-seq for H2AK119ub and H3K27me3. In agreement with IF data, there was chromosome-wide depletion of both marks from  $X^{\text{mus}}$  (becoming indistinguishable from  $X^{\text{cas}}$ ) in  $\Delta$ RepB cells (Colognori et al., 2019), but no similar effect in  $\Delta$ RepA cells (Figures 4B–4D). Furthermore, we detected no obvious differences in coverage over genic versus intergenic regions, or non-expressed genes versus those subject to XCI, besides again higher coverage over genes subject to XCI (except in  $\Delta$ RepB cells where the marks are both absent) (Figures S4A and S4B). Together, these data indicate that, while repeat A plays a role in establishing Polycomb recruitment during the first phase of XCI, it appears dispensable once recruitment has been established. On the other hand, repeat B is required for Polycomb recruitment during establishment and remains relevant throughout the XCI maintenance phase.

## DISCUSSION

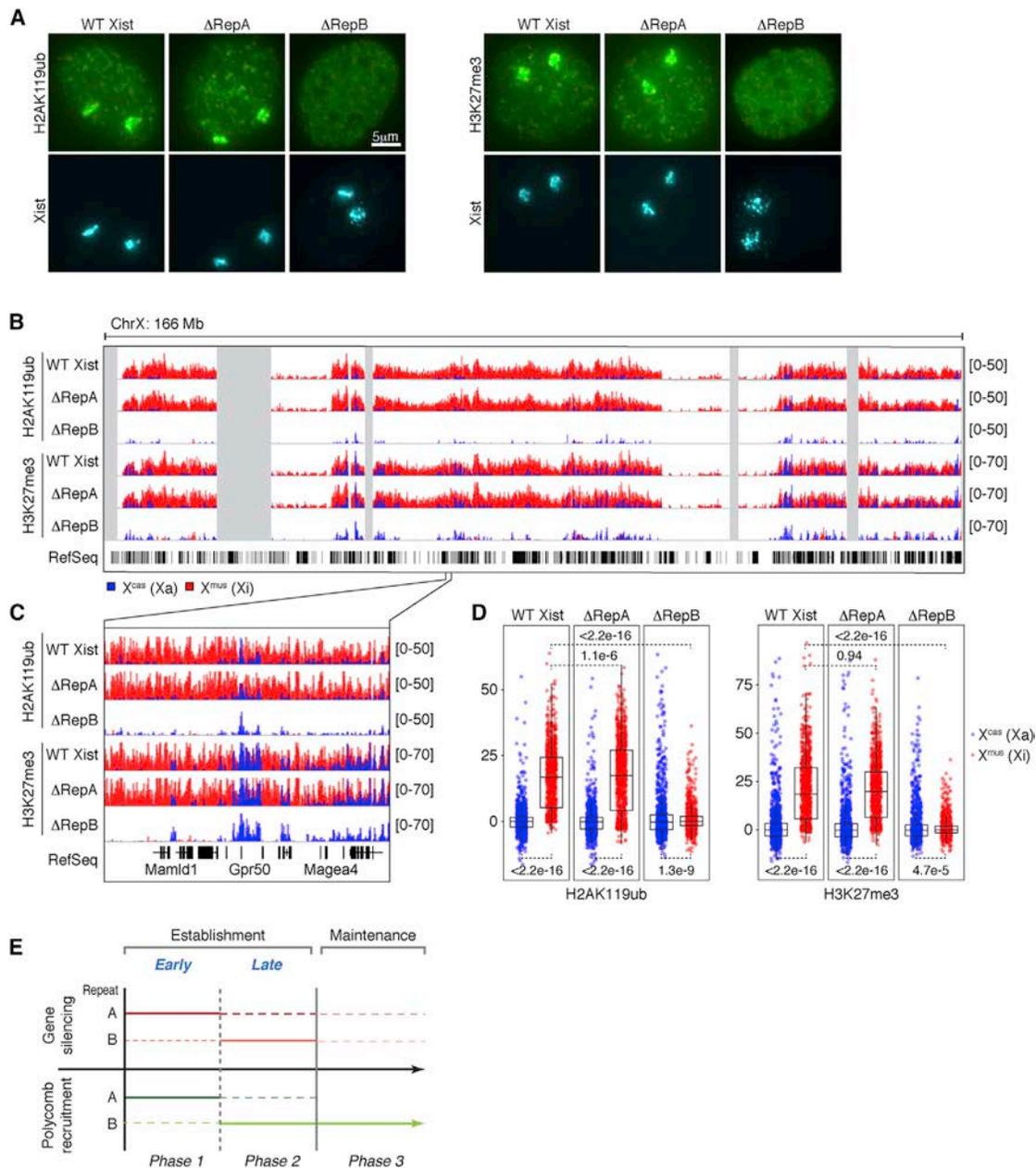
Here, our work has addressed the longstanding confusion over the roles of Xist repeats A and B for Polycomb recruitment and Xi gene silencing. In doing so, we found that XCI can best be characterized as having three distinct phases (Figure 4E): (1) an “early establishment” phase (days 0–8 of ESC differentiation) during which repeat A is required to initiate gene silencing and an early wave of Polycomb recruitment; (2) a “late establishment” phase (days 8–14) during which repeat B is essential for stabilizing Polycomb proteins and gene silencing on the Xi; and (3) a “maintenance” phase (in somatic cells) in which repeat A is no longer required but repeat B continues to play a role in Polycomb maintenance. During the maintenance phase, gene silencing is stabilized but may still depend on continued expression of *Xist* in a context-dependent manner, as post-XCI deletions of *Xist* (in part or in whole) can cause either major physiological perturbations (Yildirim et al., 2013) or minimal reactivation unless combined with other pharmacological agents (Csankovszki et al., 2001; Minajigi et al., 2015; Carrette et al., 2017; Adrianse et al., 2018; Colognori et al., 2019). Of note, our data do not rule out that repeat A might also function during late establishment (dotted lines, Figure 4E) because we did not conditionally remove it during this time frame. Similarly, repeat B may also

contribute somewhat to early establishment (dotted lines, Figure 4E), since initial gene silencing and Polycomb recruitment were less robust upon its removal, and deletion of both repeats A and B was necessary to abolish early Polycomb enrichment. Thus, during the early critical stages, repeats A and B may work together and both be required to establish the typical Polycomb binding patterns, associated enrichment of H3K27me3 and H2AK119ub, and full genic silencing on the Xi.

Our data show that deleting repeat B alone does not preclude initiation of Polycomb recruitment and gene silencing during early XCI between days 0–8 (Figures 1 and 2). Thus, in contrast with recent proposals (Pintacuda et al., 2017; Nesterova et al., 2019), repeat B is not the only motif involved in these activities. However, without repeat B, gene silencing and Polycomb recruitment cannot proceed to completion and are also unstable, exhibiting a reversion to biallelic expression and an inability to retain the Polycomb marks beyond day ~8. Thus, repeat B functions as a parallel pathway to stabilize and/or bolster gene silencing and Polycomb on Xi. In this regard, our study reconciles disparate conclusions of prior studies in which deletions of repeat B have been shown to both significantly affect Xi silencing (Pintacuda et al., 2017; Colognori et al., 2019; Nesterova et al., 2019) or oppositely, to have little effect (Bousard et al., 2019). Our work explains this disparity in that repeat B's impact on Xi silencing changes over the time course of XCI, being more pronounced at later versus earlier timepoints. It also explains the residual amounts of H2AK119ub and H3K27me3 detected in one study after deleting repeat B (Bousard et al., 2019).

Moreover, while all reports agree that repeat A is required for Xi silencing (Wutz et al., 2002; Bousard et al., 2019; Nesterova et al., 2019), there has been a lack of consensus regarding its role in targeting Polycomb complexes (Plath et al., 2003; Kohlmaier et al., 2004; Zhao et al., 2008; Hoki et al., 2009; Davidovich et al., 2013, 2015; Kaneko et al., 2013; Cifuentes-Rojas et al., 2014; da Rocha et al., 2014). Although we were unable to pursue a more detailed epigenomic analysis of the  $\Delta$ RepA mutant due to its instability and propensity to become XO during differentiation, our cytological comparison of  $\Delta$ RepB versus  $\Delta$ RepAB cells promote the idea that repeat A contributes to early targeting of Polycomb complexes. Deleting either repeat A or B alone does not fully abolish initiation of Polycomb recruitment, but simultaneous deletion of both does. Therefore, by inference, repeat A must collaborate with repeat B during XCI establishment for full recruitment of Polycomb complexes, with repeat A being more critical for the early phase and repeat B being more critical for subsequent phases. This was most likely missed in the past because the requirement is revealed only in conjunction with a repeat B deletion (Figure 3).

Although on its own, the  $\Delta$ RepA mutant shows Xi enrichment of PRC1 and PRC2 marks in most Xist-positive cells, there could be underlying local defects that cannot be discerned by cytological assays. Indeed, SMCHD1-depleted cells also demonstrate an apparently normal Xist cloud and enrichment of H2AK119ub/H3K27me3 by IF, but regional defects become clear in higher-resolution molecular assays (Wang et al., 2018, 2019). Together, our data affirm a role for repeat A in the initiation of Polycomb recruitment, consistent with previous studies (Kohlmaier et al., 2004; Zhao et al., 2008; Cifuentes-Rojas et al., 2014; Davidovich et al., 2015) and are also conceptually consistent



**Figure 4. Repeat B, but Not A, Continues to Play a Role in Maintaining Polycomb across Xi**

(A) H2AK119ub and H3K27me3 IF combined with Xist RNA FISH in WT Xist,  $\Delta$ RepA, and  $\Delta$ RepB female MEFs.  $n > 100$  per deletion cell line, with nearly all cells showing the indicated pattern of enrichment for each mark.

(B) Allele-specific H2AK119ub/H3K27me3 ChIP-seq in WT Xist,  $\Delta$ RepA, and  $\Delta$ RepB MEFs.

(C) Zoom in of allele-specific ChIP-seq tracks in (B).

(D) Boxplots quantifying allele-specific ChIP-seq coverage over X-linked genes. Wilcoxon rank-sum test, p values as shown.

(E) Diagram summarizing roles of repeats A and B in gene silencing and Polycomb recruitment throughout XCI. Repeat A is required for initial gene silencing and an early wave of Polycomb recruitment; repeat B is required for sustained gene silencing and Polycomb recruitment. Dashed line indicates uncertain or minimal contribution.

See also Figures S1 and S4.

with work attributing to repeat A the recruitment of Polycomb to initially active genes in ESCs (Simon et al., 2013; Bousard et al., 2019; Żylicz et al., 2019). Mechanistically, how this recruitment occurs is still under debate. One possibility is through direct RNA-mediated recruitment of Polycomb complexes (Zhao

et al., 2008; Cifuentes-Rojas et al., 2014). An alternative is through indirect recruitment as a consequence of *de novo* gene silencing by other repeat A-interacting proteins such as SPEN (Nesterova et al., 2019). It is also possible that both types of mechanisms are at play. Our findings also have implications

for the role of Polycomb complexes during Xi gene silencing. Previous literature suggests Polycomb may be dispensable for initial silencing (Kalantry and Magnuson, 2006; Leeb and Wutz, 2007) but is required for its stabilization (Kalantry and Magnuson, 2006; Wang et al., 2001) until additional mechanisms such as DNA methylation solidify the silenced state in maintenance phase (Csankovszki et al., 2001). Our observation that deleting repeat B leads to reversal of gene silencing coincident with loss of Polycomb marks at differentiation day 8 supports the latter half of this hypothesis (Figures 1 and 2). As for a role in initial silencing, Polycomb mark enrichment was observed on Xi in a repeat A-dependent manner coincident with initial transient silencing in  $\Delta$ RepB cells (Figure 2). However, the marks were also enriched on Xi in  $\Delta$ RepA cells despite failure to initiate silencing (Figure 3). Thus, it is less clear whether Polycomb recruitment in this case is a cause or consequence of initial silencing.

The functional importance of repeat A is further underscored by our inability to derive stable XX female differentiated ESCs lacking it. The instability occurred only during differentiation and not in undifferentiated cells, suggesting that the propensity toward aneuploidy is caused by selective pressure to make up for failed dosage compensation. Intriguingly, loss of repeat B does not similarly lead to aneuploidy, potentially because of the less drastic effect on Xi gene silencing (still some silencing at day 14 [Figure 1A]) or compensation by partial downregulation of Xa (Colognori et al., 2019). It is also tempting to speculate that proper X dosage may be more critical during early differentiation (when repeat A is critical) than late (when repeat B is critical).

In a broader context, our study reveals establishment of gene silencing and Polycomb domains to be more complex than previously thought, entailing distinct molecular requirements compared with their maintenance. Xi gene silencing and Polycomb domains occurring in the presence of Xist repeat A eventually disappear without repeat B. This is consistent with burgeoning evidence that pre-existing Polycomb marks alone are insufficient to recruit the complexes and maintain Polycomb domains on autosomes; *de novo* recruitment occurs via different means (Kahn et al., 2016; Højfeldt et al., 2018; Oksuz et al., 2018). Notably, the division of labor between repeats A and B in recruiting Polycomb during XCI establishment does not persist into the maintenance phase, when repeat B but not A continues to be required (Figure 4). Perhaps repeat A is required during *de novo* XCI establishment to silence initially active genes but becomes dispensable in the maintenance phase when Xi genes are already silenced. Whether similar multi-phasic and/or context-dependent mechanisms are required to establish and maintain epigenetic silencing over autosomal Polycomb targets would be of high interest to future investigations.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability

- Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Xist Deletion Cell Lines
  - Cell Culture
  - ESC Differentiation
- METHOD DETAILS
  - Oligo FISH Probes
  - RNA FISH
  - X-Chromosome Painting
  - IF/RNA FISH
  - Microscopy
  - Antibodies
  - (Allele-specific) qPCR and RT-qPCR
  - ChIP-seq
  - ChIP-seq Analysis
  - RNA-seq
  - RNA-seq Analysis
- QUANTIFICATION AND STATISTICAL ANALYSIS

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.devcel.2020.05.021>.

## ACKNOWLEDGMENTS

We thank all members of the Lee lab for critical comments and stimulating discussions. H.S. was supported by NIH 5T32HD007396-24 and J.T.L. by NIH R01-HD097665 and the Howard Hughes Medical Institute.

## AUTHOR CONTRIBUTIONS

D.C., H.S., and J.T.L. conceived the project, analyzed data, and wrote the paper. D.W. performed allele-specific RT-qPCR experiments. C.Y.W. performed bioinformatic analysis of ChIP-seq datasets. D.C. and H.S. performed all other experiments and analyses together.

## DECLARATION OF INTERESTS

J.T.L. is a co-founder of Translate Bio and Fulcrum Therapeutics and serves as Advisor to Skyhawk Therapeutics.

Received: September 4, 2019

Revised: February 16, 2020

Accepted: May 14, 2020

Published: June 11, 2020

## REFERENCES

- Adrianse, R.L., Smith, K., Gattbonton-Schwager, T., Sripathy, S.P., Lao, U., Foss, E.J., Boers, R.G., Boers, J.B., Gribnau, J., and Bedalov, A. (2018). Perturbed maintenance of transcriptional repression on the inactive X-chromosome in the mouse brain after Xist deletion. *Epigenet. Chromatin* **11**, 50.
- Almeida, M., Pintacuda, G., Masui, O., Koseki, Y., Gdula, M., Cerase, A., Brown, D., Mould, A., Innocent, C., Nakayama, M., et al. (2017). PCGF3/5-PRC1 initiates Polycomb recruitment in X chromosome inactivation. *Science* **356**, 1081–1084.
- Bhatnagar, S., Zhu, X., Ou, J., Lin, L., Chamberlain, L., Zhu, L.J., Wajapeyee, N., and Green, M.R. (2014). Genetic and pharmacological reactivation of the mammalian inactive X chromosome. *Proc. Natl. Acad. Sci. USA* **111**, 12591–12598.
- Bousard, A., Raposo, A.C., Żylicz, J.J., Picard, C., Pires, V.B., Qi, Y., Gil, C., Syx, L., Chang, H.Y., Heard, E., and da Rocha, S.T. (2019). The role of Xist-mediated Polycomb recruitment in the initiation of X-chromosome inactivation. *EMBO Rep.* **20**, e48019.

- Brown, C.J., and Willard, H.F. (1994). The human X-inactivation centre is not required for maintenance of X-chromosome inactivation. *Nature* 368, 154–156.
- Carrette, L.L.G., Blum, R., Ma, W., Kelleher, R.J., 3rd, and Lee, J.T. (2018). Tsix-Mecp2 female mouse model for Rett syndrome reveals that low-level MECP2 expression extends life and improves neuromotor function. *Proc. Natl. Acad. Sci. USA* 115, 8185–8190.
- Carrette, L.L.G., Wang, C.Y., Wei, C., Press, W., Ma, W., Kelleher, R.J., 3rd, and Lee, J.T. (2017). A mixed modality approach towards Xi reactivation for Rett syndrome and other X-linked disorders. *Proc. Natl. Acad. Sci. USA* 115, E668–E675.
- Cifuentes-Rojas, C., Hernandez, A.J., Sarma, K., and Lee, J.T. (2014). Regulatory interactions between RNA and polycomb repressive complex 2. *Mol. Cell* 55, 171–185.
- Colognori, D., Sunwoo, H., Kriz, A.J., Wang, C.Y., and Lee, J.T. (2019). Xist deletion analysis reveals an interdependency between Xist RNA and polycomb complexes for spreading along the inactive X. *Mol. Cell* 74, 101–117.e10.
- Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* 153, 773–784.
- da Rocha, S.T., Boeva, V., Escamilla-Del-Arenal, M., Ancelin, K., Granier, C., Matias, N.R., Sanulli, S., Chow, J., Schulz, E., Picard, C., et al. (2014). Jarid2 is implicated in the initial Xist-induced targeting of PRC2 to the inactive X chromosome. *Mol. Cell* 53, 301–316.
- Davidovich, C., Wang, X., Cifuentes-Rojas, C., Goodrich, K.J., Gooding, A.R., Lee, J.T., and Cech, T.R. (2015). Toward a consensus on the binding specificity and promiscuity of PRC2 for RNA. *Mol. Cell* 57, 552–558.
- Davidovich, C., Zheng, L., Goodrich, K.J., and Cech, T.R. (2013). Promiscuous RNA binding by Polycomb repressive complex 2. *Nat. Struct. Mol. Biol.* 20, 1250–1257.
- de Napolés, M., Mermoud, J.E., Wakao, R., Tang, Y.A., Endoh, M., Appanah, R., Nesterova, T.B., Silva, J., Otte, A.P., Vidal, M., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* 7, 663–676.
- Disteche, C.M. (2016). Dosage compensation of the sex chromosomes and autosomes. *Semin. Cell Dev. Biol.* 56, 9–18.
- Giorgetti, L., Lajoie, B.R., Carter, A.C., Attia, M., Zhan, Y., Xu, J., Chen, C.J., Kaplan, N., Chang, H.Y., Heard, E., and Dekker, J. (2016). Structural organization of the inactive X chromosome in the mouse. *Nature* 535, 575–579.
- Glaab, W.E., and Skopek, T.R. (1999). A novel assay for allelic discrimination that combines the fluorogenic 5' nuclease polymerase chain reaction (TaqMan) and mismatch amplification mutation assay. *Mutat. Res.* 430, 1–12.
- Ha, N., Lai, L.T., Chelliah, R., Zhen, Y., Yi Vanessa, S.P., Lai, S.K., Li, H.Y., Ludwig, A., Sandin, S., Chen, L., and Zhang, L.-F. (2018). Live-cell imaging and functional dissection of Xist RNA reveal mechanisms of X chromosome inactivation and reactivation. *iScience* 8, 1–14.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* 38, 576–589.
- Højfeldt, J.W., Laugesen, A., Willumsen, B.M., Damhofer, H., Hedehus, L., Tvardovskiy, A., Mohammad, F., Jensen, O.N., and Helin, K. (2018). Accurate H3K27 methylation can be established de novo by SUZ12-directed PRC2. *Nat. Struct. Mol. Biol.* 25, 225–232.
- Hoki, Y., Kimura, N., Kanbayashi, M., Amakawa, Y., Ohhata, T., Sasaki, H., and Sado, T. (2009). A proximal conserved repeat in the Xist gene is essential as a genomic element for X-inactivation in mouse. *Development* 136, 139–146.
- Jégu, T., Aeby, E., and Lee, J.T. (2017). The X chromosome in space. *Nat. Rev. Genet.* 18, 377–389.
- Jégu, T., Blum, R., Cochrane, J.C., Yang, L., Wang, C.Y., Gilles, M.E., Colognori, D., Szanto, A., Marr, S.K., Kingston, R.E., and Lee, J.T. (2019). Xist RNA antagonizes the SWI/SNF chromatin remodeler BRG1 on the inactive X chromosome. *Nat. Struct. Mol. Biol.* 26, 96–109.
- Kahn, T.G., Dorafshan, E., Schultheis, D., Zare, A., Stenberg, P., Reim, I., Pirrotta, V., and Schwartz, Y.B. (2016). Interdependence of PRC1 and PRC2 for recruitment to polycomb response elements. *Nucleic Acids Res.* 44, 10132–10149.
- Kalantry, S., and Magnuson, T. (2006). The polycomb group protein EED is dispensable for the initiation of random X-chromosome inactivation. *PLoS Genet.* 2, e66.
- Kaneko, S., Son, J., Shen, S.S., Reinberg, D., and Bonasio, R. (2013). PRC2 binds active promoters and contacts nascent RNAs in embryonic stem cells. *Nat. Struct. Mol. Biol.* 20, 1258–1264.
- Kharchenko, P.V., Tolstorukov, M.Y., and Park, P.J. (2008). Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nat. Biotechnol.* 26, 1351–1359.
- Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: accurate alignment of transcriptsomes in the presence of insertions, deletions and gene fusions. *Genome Biol.* 14, R36.
- Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol.* 2, E171.
- Lee, D.M., Trotman, J.B., Cherney, R.E., Inoue, K., Schertzer, M.D., Bischoff, S.R., Cowley, D.O., and Calabrese, J.M. (2019). A 5' fragment of Xist can sequester RNA produced from adjacent genes on chromatin. *Nucleic Acids Res.* 47, 7049–7062.
- Leeb, M., and Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J. Cell Biol.* 178, 219–229.
- Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., and Durbin, R.; 1000 Genome Project Data Processing Subgroup (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079.
- Li, B., Kadura, I., Fu, D.J., and Watson, D.E. (2004). Genotyping with TaqMAMA. *Genomics* 83, 311–320.
- Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* 30, 923–930.
- Minajigi, A., Froberg, J., Wei, C., Sunwoo, H., Kesner, B., Colognori, D., Lessing, D., Payer, B., Boukhali, M., Haas, W., and Lee, J.T. (2015). Chromosomes. A comprehensive Xist interactome reveals cohesin repulsion and an RNA-directed chromosome conformation. *Science* 349.
- Minks, J., Baldry, S.E., Yang, C., Cotton, A.M., and Brown, C.J. (2013). XIST-induced silencing of flanking genes is achieved by additive action of repeat monomers in human somatic cells. *Epigenet. Chromatin* 6, 23.
- Nesterova, T.B., Wei, G., Coker, H., Pintacuda, G., Bowness, J.S., Zhang, T., Almeida, M., Bloechl, B., Moindrot, B., Carter, E.J., et al. (2019). Systematic allelic analysis defines the interplay of key pathways in X chromosome inactivation. *Nat. Commun.* 10, 3129.
- Nozawa, R.S., Nagao, K., Igami, K.T., Shibata, S., Shirai, N., Nozaki, N., Sado, T., Kimura, H., and Obuse, C. (2013). Human inactive X chromosome is compacted through a PRC2-independent SMCHD1-HBIX1 pathway. *Nat. Struct. Mol. Biol.* 20, 566–573.
- Ogawa, Y., Sun, B.K., and Lee, J.T. (2008). Intersection of the RNA interference and X-inactivation pathways. *Science* 320, 1336–1341.
- Oksuz, O., Narendra, V., Lee, C.H., Descostes, N., LeRoy, G., Raviram, R., Blumenberg, L., Karch, K., Rocha, P.P., Garcia, B.A., et al. (2018). Capturing the onset of PRC2-mediated repressive domain formation. *Mol. Cell* 70, 1149–1162.e5.
- Pintacuda, G., Wei, G., Roustan, C., Kirmizitas, B.A., Solcan, N., Cerase, A., Castello, A., Mohammed, S., Moindrot, B., Nesterova, T.B., and Brockdorff, N. (2017). hnRNPK recruits PCGF3/5-PRC1 to the Xist RNA B-repeat to establish polycomb-mediated chromosomal silencing. *Mol. Cell* 68, 955–969.e10.
- Pinter, S.F., Colognori, D., Beliveau, B.J., Sadreyev, R.I., Payer, B., Yildirim, E., Wu, C.T., and Lee, J.T. (2015). Allelic imbalance is a prevalent and tissue-specific feature of the mouse transcriptome. *Genetics* 200, 537–549.

- Pinter, S.F., Sadreyev, R.I., Yildirim, E., Jeon, Y., Ohsumi, T.K., Borowsky, M., and Lee, J.T. (2012). Spreading of X chromosome inactivation via a hierarchy of defined Polycomb stations. *Genome Res.* **22**, 1864–1876.
- Plath, K., Fang, J., Mlynarczyk-Evans, S.K., Cao, R., Worringer, K.A., Wang, H., de la Cruz, C.C., Otte, A.P., Panning, B., and Zhang, Y. (2003). Role of histone H3 lysine 27 methylation in X inactivation. *Science* **300**, 131–135.
- Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. *Nat Protoc.* **8**, 2281–2308.
- Rao, S.S., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680.
- Royce-Tolland, M.E., Andersen, A.A., Koyfman, H.R., Talbot, D.J., Wutz, A., Tonks, I.D., Kay, G.F., and Panning, B. (2010). The A-repeat links ASF/SF2-dependent Xist RNA processing with random choice during X inactivation. *Nat. Struct. Mol. Biol.* **17**, 948–954.
- Schoeftner, S., Sengupta, A.K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., Jenuwein, T., and Wutz, A. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* **25**, 3110–3122.
- Silva, J., Mak, W., Zvetkova, I., Appanah, R., Nesterova, T.B., Webster, Z., Peters, A.H., Jenuwein, T., Otte, A.P., and Brockdorff, N. (2003). Establishment of histone H3 methylation on the inactive X chromosome requires transient recruitment of Eed-Enx1 polycomb group complexes. *Dev. Cell* **4**, 481–495.
- Simon, M.D., Pinter, S.F., Fang, R., Sarma, K., Rutenberg-Schoenberg, M., Bowman, S.K., Kesner, B.A., Maier, V.K., Kingston, R.E., and Lee, J.T. (2013). High-resolution Xist binding maps reveal two-step spreading during X-chromosome inactivation. *Nature* **504**, 465–469.
- Sripathy, S., Leko, V., Adrianse, R.L., Loe, T., Foss, E.J., Dalrymple, E., Lao, U., Gatbonton-Schwager, T., Carter, K.T., Payer, B., et al. (2017). Screen for reactivation of MeCP2 on the inactive X chromosome identifies the BMP/TGF-beta superfamily as a regulator of XIST expression. *Proc. Natl. Acad. Sci. USA* **114**, 1619–1624.
- Starmer, J., and Magnuson, T. (2009). A new model for random X chromosome inactivation. *Development* **136**, 1–10.
- Wang, C.Y., Colognori, D., Sunwoo, H., Wang, D., and Lee, J.T. (2019). PRC1 collaborates with SMCHD1 to fold the X-chromosome and spread Xist RNA between chromosome compartments. *Nat. Commun.* **10**, 2950.
- Wang, C.Y., Jégu, T., Chu, H.P., Oh, H.J., and Lee, J.T. (2018). SMCHD1 merges chromosome compartments and assists formation of super-structures on the inactive X. *Cell* **174**, 406–421.e25.
- Wang, J., Mager, J., Chen, Y., Schneider, E., Cross, J.C., Nagy, A., and Magnuson, T. (2001). Imprinted X inactivation maintained by a mouse polycomb group gene. *Nat. Genet.* **28**, 371–375.
- Wutz, A., and Jaenisch, R. (2000). A shift from reversible to irreversible X inactivation is triggered during ES cell differentiation. *Mol. Cell* **5**, 695–705.
- Wutz, A., Rasmussen, T.P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist RNA. *Nat. Genet.* **30**, 167–174.
- Yildirim, E., Kirby, J.E., Brown, D.E., Mercier, F.E., Sadreyev, R.I., Scadden, D.T., and Lee, J.T. (2013). Xist RNA is a potent suppressor of hematologic cancer in mice. *Cell* **152**, 727–742.
- Yildirim, E., Sadreyev, R.I., Pinter, S.F., and Lee, J.T. (2011). X-chromosome hyperactivation in mammals via nonlinear relationships between chromatin states and transcription. *Nat Struct Mol Biol.* **19**, 56–61.
- Zhang, L.F., Huynh, K.D., and Lee, J.T. (2007). Perinucleolar targeting of the inactive X during S phase: evidence for a role in the maintenance of silencing. *Cell* **129**, 693–706.
- Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* **322**, 750–756.
- Żylicz, J.J., Bousard, A., Žumer, K., Dossin, F., Mohammad, E., da Rocha, S.T., Schwalb, B., Syx, L., Dingli, F., Loew, D., et al. (2019). The implication of early chromatin changes in X chromosome inactivation. *Cell* **176**, 182–197.e23.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Rabbit monoclonal anti-H3K27me3	GeneTex	Cat#GTX60892
Rabbit monoclonal anti-H2AK119ub	Cell Signaling	Cat#CST8240
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
Recombinant mouse LIF	Sigma	Cat#ESG1107
<b>Critical Commercial Assays</b>		
NEBNext Poly(A) mRNA Magnetic Isolation Module	New England BioLabs	Cat#E7490S
Agencourt AMPure XP Beads	Beckman Coulter	Cat#A63881
NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina	New England BioLabs	Cat#E6240S
NEBNext Multiplex Oligos for Illumina (Index Primers Set 1)	New England BioLabs	Cat#E7335S
Quant-iT PicoGreen dsDNA Reagent	Thermo Fisher Scientific	Cat#P7581
NEBNext Ultra Directional RNA Library Prep Kit for Illumina	New England BioLabs	Cat#E7420S
<b>Deposited Data</b>		
RNA-seq in WT Xist female mouse differentiating ESC (days 0,5,6,7,8,10)	This study	GEO: GSE135389
RNA-seq in WT Xist female mouse differentiating ESC (day 14)	<a href="#">Colognori et al., 2019</a>	GEO: GSE107217
RNA-seq in ΔRepB Xist female mouse differentiating ESC (days, 0,5,6,7,8,10)	This study	GEO: GSE135389
RNA-seq in ΔRepB Xist female mouse differentiating ESC (day 14)	<a href="#">Colognori et al., 2019</a>	GEO: GSE107217
H3K27me3 ChIP-seq in WT Xist female mouse differentiating ESC (days 0,5,8,14)	This study	GEO: GSE135389
H3K27me3 ChIP-seq in ΔRepB Xist female mouse differentiating ESC (days 0,5,8,14)	This study	GEO: GSE135389
H2AK119ub ChIP-seq in WT Xist female mouse differentiating ESC (days 0,5,8,14)	This study	GEO: GSE135389
H2AK119ub ChIP-seq in ΔRepB Xist female mouse differentiating ESC (days 0,5,8,14)	This study	GEO: GSE135389
H3K27me3 ChIP-seq in WT Xist female MEF	<a href="#">Colognori et al., 2019</a>	GEO: GSE107217
H3K27me3 ChIP-seq in ΔRepA Xist female MEF	This study	GEO: GSE135389
H3K27me3 ChIP-seq in ΔRepB Xist female MEF	<a href="#">Colognori et al., 2019</a>	GEO: GSE107217
H2AK119ub ChIP-seq in WT Xist female MEF	<a href="#">Colognori et al., 2019</a>	GEO: GSE107217
H2AK119ub ChIP-seq in ΔRepA Xist female MEF	This study	GEO: GSE135389
H2AK119ub ChIP-seq in ΔRepB Xist female MEF	<a href="#">Colognori et al., 2019</a>	GEO: GSE107217
<b>Mendeley data</b>	This study	<a href="https://doi.org/10.17632/77fjk9p346.1">https://doi.org/10.17632/77fjk9p346.1</a>
<b>Experimental Models: Cell Lines</b>		
WT Xist female ESC ( <i>Tsix</i> <sup>TST/+</sup> )	<a href="#">Ogawa et al., 2008</a>	N/A

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
$\Delta$ RepA Xist female ESC (clone 8)	This study	N/A
$\Delta$ RepB Xist female ESC (clone D2)	<a href="#">Colognori et al., 2019</a>	N/A
$\Delta$ RepAB Xist female ESC (clone 9)	This study	N/A
WT Xist female MEF (EY.T4)	<a href="#">Yildirim et al., 2011</a>	N/A
“Old” $\Delta$ RepA Xist female MEF (clone X9)	<a href="#">Colognori et al., 2019</a>	N/A
“New” $\Delta$ RepA Xist female MEF (clone ds4)	This study	N/A
$\Delta$ RepB Xist female MEF (clone 22)	<a href="#">Colognori et al., 2019</a>	N/A
Oligonucleotides		
Oligo FISH probes used for Xist RNA FISH	<a href="#">Colognori et al., 2019</a>	N/A
Stellaris FISH probes used for Atrx and Mecp2 nascent RNA FISH	LGC Biosearch Technologies	Custom
XMP X Green Mouse Chromosome Paint	MetaSystems	Cat#D-1420-050-FI
gRNAs used to generate Xist deletions (see <a href="#">Table S1</a> )	Integrated DNA Technologies	N/A
PCR primers (see <a href="#">Table S2</a> )	Integrated DNA Technologies	N/A
Recombinant DNA		
pSpCas9(BB)-2A-GFP (PX461)	<a href="#">Ran et al., 2013</a>	Addgene Cat#48140
pSpCas9(BB)-2A-Puro (PX459) v2.0	<a href="#">Ran et al., 2013</a>	Addgene Cat#62988
Software and Algorithms		
HOMER v4.8	<a href="#">Heinz et al., 2010</a>	<a href="http://homer.ucsd.edu/homer/index.html">http://homer.ucsd.edu/homer/index.html</a>
NovoAlign v3.02	Novocraft	<a href="http://www.novocraft.com/products/novoalign/">http://www.novocraft.com/products/novoalign/</a>
TopHat2 v2.0.10	<a href="#">Kim et al., 2013</a>	<a href="https://ccb.jhu.edu/software/tophat/index.shtml">https://ccb.jhu.edu/software/tophat/index.shtml</a>
SPP	<a href="#">Kharchenko et al., 2008</a>	<a href="http://compbio.med.harvard.edu/Supplements/ChIP-seq/">http://compbio.med.harvard.edu/Supplements/ChIP-seq/</a>
featureCounts v1.5.0-p1	<a href="#">Liao et al., 2014</a>	<a href="http://subread.sourceforge.net">http://subread.sourceforge.net</a>
SAMtools v1.4.1	<a href="#">Li et al., 2009</a>	<a href="http://samtools.sourceforge.net/">http://samtools.sourceforge.net/</a>

**RESOURCE AVAILABILITY****Lead Contact**

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Jeannie T. Lee ([lee@molbio.mgh.harvard.edu](mailto:lee@molbio.mgh.harvard.edu)).

**Materials Availability**

Cell lines generated in this study will be available upon request following completion of an MTA.

**Data and Code Availability**

Original unprocessed microscope images in this manuscript have been deposited at Mendeley Data and are available at: <https://doi.org/10.17632/77fjk9p346.1>

Raw high-throughput sequencing data and processed files for RNA-seq and ChIP-seq reported in this paper have been deposited at GEO under accession number: GSE135389

**EXPERIMENTAL MODEL AND SUBJECT DETAILS****Xist Deletion Cell Lines**

Xist deletions were generated by CRISPR/Cas9 using a pair of gRNAs flanking the target region. gRNA sequences ([Table S1](#)) were designed using tools available online (<http://crispr.mit.edu>) and cloned into pSpCas9(BB)-2A-GFP or pSpCas9(BB)-2A-Puro vectors ([Ran et al., 2013](#)). gRNA/Cas9 plasmid was delivered into ESCs by electroporation (Bio-Rad Gene Pulser Xcell) or MEFs by nucleofection (Lonza Nucleofector II) as per manufacturer’s instructions. Following plasmid delivery, cells were cultured for one week to allow enough time for DNA cutting and repair. Single cells were then sorted into 96-well plates by FACS, expanded, and screened by genomic PCR, Sanger sequencing ([Table S1](#)), and two-color Xist RNA FISH.

*Xist* deletion ESCs were generated in the parental (“WT *Xist*”) *M. musculus/M. castaneus* F2 hybrid female ESC line carrying a mutated *Tsix* allele previously described as “*Tsix*<sup>TST/+</sup>” (Ogawa et al., 2008). This mutation drives selective inactivation of *X*<sup>mus</sup>. *Xist* deletion MEFs were generated in the parental (“WT *Xist*”) *M. musculus/M. castaneus* F1 hybrid (tetraploid) female MEF line previously described as “EY.T4”. All deletion cell lines used in this study are listed in the [Key Resources Table](#), and characterized in detail in [Table S1](#); [Figure S1](#).

### Cell Culture

MEFs were grown in medium containing DMEM, high glucose, GlutaMAX supplement, pyruvate (Thermo Fisher Scientific), 10% FBS (Sigma), 25 mM HEPES pH 7.2-7.5 (Thermo Fisher Scientific), 1x MEM non-essential amino acids (Thermo Fisher Scientific), 1x Pen/Strep (Thermo Fisher Scientific), and 0.1 mM  $\beta$ ME (Thermo Fisher Scientific) at 37°C with 5% CO<sub>2</sub>. ESCs were grown on  $\gamma$ -irradiated MEF feeders in medium containing DMEM, high glucose, GlutaMAX supplement, pyruvate, 15% Hyclone FBS (Sigma), 25 mM HEPES pH 7.2-7.5, 1x MEM non-essential amino acids, 1x Pen/Strep, 0.1 mM  $\beta$ ME, and 500 U/mL ESGRO recombinant mouse Leukemia Inhibitory Factor (LIF) protein (Sigma, ESG1107) at 37°C with 5% CO<sub>2</sub>. LIF was excluded from the medium during ESC differentiation procedures (see below).

### ESC Differentiation

Undifferentiated ESCs were grown on  $\gamma$ -irradiated MEF feeders for 3 days, after which ESC colonies were trypsinized and feeders removed (day 0). ESCs were then switched to medium lacking LIF and grown in suspension for 4 days, forming embryoid bodies (EBs). EBs were cytopspun onto glass slides or settled down onto gelatin-coated coverslips at day 4 and allowed to further differentiate until the indicated timepoints.

### METHOD DETAILS

#### Oligo FISH Probes

Oligo FISH probes for *Xist* RepA, RepB, or exon 7 were previously described (Colognori et al., 2019). Briefly, probe sequences were designed using tools available online (<https://www.idtdna.com/calc/analyzer>). Amino-ddUTP (Kerafast) was added to the 3'-ends of pooled oligos by Terminal Transferase (New England BioLabs) as per manufacturer's instructions. Oligos were purified by phenol/chloroform extraction, concentrated by ethanol precipitation, resuspended in 0.1 M sodium borate, and labeled with Cy3B (GE Healthcare) or Alexa647 NHS-ester (Life Technologies). After another ethanol precipitation, labeled oligos were resuspended in water and labeling efficiency was evaluated by absorbance using NanoDrop (Thermo Fisher Scientific). Custom Stellaris FISH probes (LGC Biosearch Technologies) were designed against the first intron of *Atrx* or *Mecp2* using the Stellaris RNA FISH probe designer available online ([www.biosearchtech.com/stellarisdesigner](http://www.biosearchtech.com/stellarisdesigner)) and labeled with Quasar570 dye.

#### RNA FISH

Cells were cytopspun onto glass slides or settled down onto gelatin-coated coverslips and rinsed with PBS. They were permeabilized with cold CSKT buffer (100 mM NaCl, 300 mM sucrose, 10 mM PIPES, 3 mM MgCl<sub>2</sub>, 0.5% Triton X-100, pH 6.8) for 10 min and then fixed with 4% paraformaldehyde for 10 min at room temp. Cells were rinsed with PBS and dehydrated in a series of increasing ethanol concentrations. Labeled oligo probe pool (1-5 nM for *Xist* RNA FISH, 100 nM for *Atrx* or *Mecp2* nascent RNA FISH) was added to hybridization buffer containing 25% formamide, 2x SSC, 10% dextran sulfate, and nonspecific competitor (0.1 mg/mL mouse Cot-1 DNA [Thermo Fisher Scientific]). Hybridization was performed in a humidified chamber at 37°C overnight. After being washed once in 25% formamide/2x SSC at 37°C for 20 min and three times in 2x SSC at 37°C for 5 min each, cells were mounted for wide-field fluorescent imaging. Nuclei were counter-stained with Hoechst 33342 (Life Technologies).

#### X-Chromosome Painting

Cells were cytopspun onto glass slides or settled down onto gelatin-coated coverslips, rinsed with PBS, treated with RNase A (0.5 mg/mL in PBS) at 37°C for 40 min to remove RNA signal, and denatured for DNA FISH in 70% formamide/2x SSC at 80°C for 15 min. Slides were quenched in ice cold 70% ethanol and dehydrated in a series of increasing ethanol concentrations. 1:10 (v/v) XMP X Green mouse chromosome paint (MetaSystems, D-1420-050-FI) was added to hybridization buffer containing 50% formamide, 2x SSC, 10% dextran sulfate, and 0.2 mg/mL mouse Cot-1 DNA (Thermo Fisher Scientific) and denatured at 95°C for 10 min. Hybridization was performed in a humidified chamber at 37°C overnight. After being washed once in 0.2x SSC at 65°C for 10 min and three times in 2x SSC at room temp for 5 min each, cells were mounted for wide-field fluorescent imaging. Nuclei were counter-stained with Hoechst 33342 (Life Technologies).

#### IF/RNA FISH

Cells were cytopspun onto glass slides or settled down onto gelatin-coated coverslips and rinsed with PBS. They were permeabilized with cold CSKT buffer (100 mM NaCl, 300 mM sucrose, 10 mM PIPES, 3 mM MgCl<sub>2</sub>, 0.5% Triton X-100, pH 6.8) for 10 min and then fixed with 4% paraformaldehyde for 10 min at room temp. After being blocked for 30 min in PBS/1% BSA supplemented with 10 mM ribonucleoside vanadyl complex (New England BioLabs), primary antibodies were added and allowed to incubate at room temp for 1 hr. Cells were washed three times with PBS/0.05% Tween-20 at room temp for 5 min each. After incubating with dye-conjugated

secondary antibody for 30 min at room temp, cells were washed again with PBS/0.05% Tween-20 at room temp for 5 min each. Cells were post-fixed in 4% paraformaldehyde and dehydrated in a series of increasing ethanol concentrations. Xist RNA FISH was then performed as described above.

### Microscopy

For Wide-Field Fluorescent Imaging, Cells Were Observed on a Nikon 90i Microscope Equipped with 60×/1.4 N.A. VC Objective Lens, Orca ER CCD Camera (Hamamatsu), and Volocity Software (PerkinElmer).

### Antibodies

The following primary antibodies were used for ChIP-seq and IF: H3K27me3 (GeneTex, GTX60892) and H2AK119ub (Cell Signaling, CST8240). Dye-conjugated secondary antibodies were purchased from Life Technologies.

### (Allele-specific) qPCR and RT-qPCR

For RT-qPCR, RNA was isolated from cells using TRIzol Reagent (Thermo Fisher Scientific) as per manufacturer's instructions. Genomic DNA was removed using TURBO DNase from the TURBO DNA-free Kit (Thermo Fisher Scientific). After inactivating TURBO DNase with DNase Inactivation Reagent (also enclosed in TURBO DNA-free Kit), RNA was reverse transcribed into cDNA using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific) with random primers (Promega, C118A) at 25°C for 5 min, 50°C for 1 hr, and enzyme inactivated at 85°C for 15 min. Depending on the experiment, qPCR was performed on cDNA or genomic DNA using iTaq Universal SYBR Green Supermix (Bio-Rad) in a CFX96 Real-Time PCR Detection System (Bio-Rad). For allele-specific detection, primers were designed to target genetic variants within each gene, as previously described (Glaab and Skopek, 1999; Li et al., 2004). The relative abundance of alleles was calculated using the formula:  $cas/mus$  fold difference =  $2^{-(Ct^{mus} - Ct^{cas})}$ , and corrected for primer bias/efficiency by comparing to standard curves using pure *cas*, *mus*, or hybrid *cas/mus* genomic DNA as previously described (Pinter et al., 2015). Primer sequences are listed in Table S2.

### ChIP-seq

Cells were cross-linked in PBS with 1% formaldehyde at room temp for 10 min with rotation at 1 million cells/mL and quenched with 0.125 M glycine for 5 min. Cross-linked cells were washed twice with cold PBS, pelleted, and snap-frozen in liquid nitrogen. 10 million cross-linked cells per ChIP were thawed on ice and resuspended in 1 mL buffer 1 (50 nM HEPES pH 7.5, 140 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.25% Triton X-100, 10% glycerol, 1x cOmplete EDTA-free protease inhibitor cocktail [Roche]), and rotated for 10 min at 4°C. Nuclei were pelleted by centrifugation at 1,000 g for 5 min at 4°C, resuspended in 1 mL buffer 2 (10 mM HEPES pH 7.5, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1x cOmplete EDTA-free protease inhibitor cocktail) supplemented with 0.2 mg/mL RNase A (Thermo Fisher Scientific), and rotated for 10 min at 4°C. Nuclei were pelleted by centrifugation at 1,000 g for 5 min at 4°C and resuspended in 1.3 mL buffer 3 (10 mM HEPES pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 1x cOmplete EDTA-free protease inhibitor cocktail, 0.1% sodium deoxycholate, 0.5% sarkosyl). Nuclei were sonicated (Qsonica Q800 Sonicator) in polystyrene tubes at 45% power reading, 30 sec on/30 sec off for a total sonication time of 4 min at 4°C. Triton X-100 was added to the lysate to 1%, which was then centrifuged for 10 min at 16,000 g to remove debris. The lysate was pre-cleared for 2 hr at 4°C with rotation using 20  $\mu$ L Dynabeads Protein G (Thermo Fisher Scientific) pre-washed with PBS/0.5% BSA. After saving 10% as "input" sample, the pre-cleared lysate was combined with 20  $\mu$ L Dynabeads Protein G pre-bound to 2  $\mu$ g antibody (H3K27me3, GeneTex GTX60892; H2AK119ub, Cell Signaling CST8240), and rotated overnight at 4°C. Afterwards, beads were washed five times with wash buffer (50 mM HEPES pH 7.5, 0.5 M LiCl, 1 mM EDTA, 1% NP-40, 0.7% sodium deoxycholate), once with TEN buffer (10 mM Tris pH 8.0, 1 mM EDTA, 50 mM NaCl), and once with TE buffer (10 mM Tris pH 8.0, 1 mM EDTA). Input sample and beads containing ChIP material were resuspended in 400  $\mu$ L TES buffer (50 mM Tris pH 8.0, 10 mM EDTA, 1% SDS) supplemented with 0.5 mg/mL Proteinase K (Sigma) and incubated for 1 hr at 55°C and then for >3 hr at 65°C to reverse cross-links. DNA was purified by phenol-chloroform extraction and quantified with Quant-iT PicoGreen dsDNA Reagent (Thermo Fisher Scientific). Input and ChIP-seq libraries were prepared using NEBNext ChIP-seq Library Prep Master Mix Set for Illumina (New England BioLabs) as per manufacturer's instructions. Libraries were sequenced on Illumina HiSeq2000 (high-throughput run) or HiSeq2500 (rapid run), generating ~50 million 50-nt paired-end reads per sample.

### ChIP-seq Analysis

To account for the hybrid character of our cell lines, adaptor-trimmed reads were separately aligned to custom *mus*/129 and *cas* genomes using NovoAlign (Novocraft), then mapped back to reference mm9 genome using SNPs (Pinter et al., 2012). This generated three tracks: composite (comp) of all reads, and two allele-specific tracks using only allele-specifically mappable reads. After allele-specific alignment, input-subtracted allele-specific ChIP-seq tracks were generated using SPP (Kharchenko et al., 2008), with smoothing using 1-kb windows recorded every 500 bp, as previously described (Wang et al., 2018). To account for different sequencing depths for ChIP-seq, samples differing by >10% were compensated by random downsampling with SAMtools (Li et al., 2009). The densities of H2AK119ub and H3K27me3 over gene bodies versus intergenic regions, and genes subject to XCI versus non-expressed on the X chromosome were computed by Homer (Heinz et al., 2010). For the timecourse ChIP-seq analysis in differentiating ESCs, genes subject to XCI were defined as having non-zero FPKM in RNA-seq for both undifferentiated WT and  $\Delta$ RepB ESCs, and non-expressed genes as having zero FPKM in any of the two datasets. For MEF ChIP-seq analysis, genes subject

to XCI were defined as having non-zero FPKM in RNA-seq for WT MEFs (expressed on Xa but not Xi), and non-expressed genes as having zero FPKM (not expressed on Xa or Xi). *Xist* and *Tsix*, escapee genes, and regions too short (<200 bp) or unmappable were excluded from the analysis. H2AK119ub and H3K27me3 densities were displayed as boxplots produced using R and ggplot2. *p* values determined by Wilcoxon ranked sum test (two-sided).

### RNA-seq

Total cell RNA was extracted using TRIzol Reagent (Thermo Fisher Scientific), from which mRNA was isolated using NEBNext Poly(A) mRNA magnetic isolation module (New England BioLabs) as per manufacturer's instructions. RNA-seq libraries were prepared using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England BioLabs) as per manufacturer's instructions. Libraries were sequenced on Illumina HiSeq2000 (high-throughput run) or HiSeq2500 (rapid run), generating ~50 million 50-nt paired-end reads per sample.

### RNA-seq Analysis

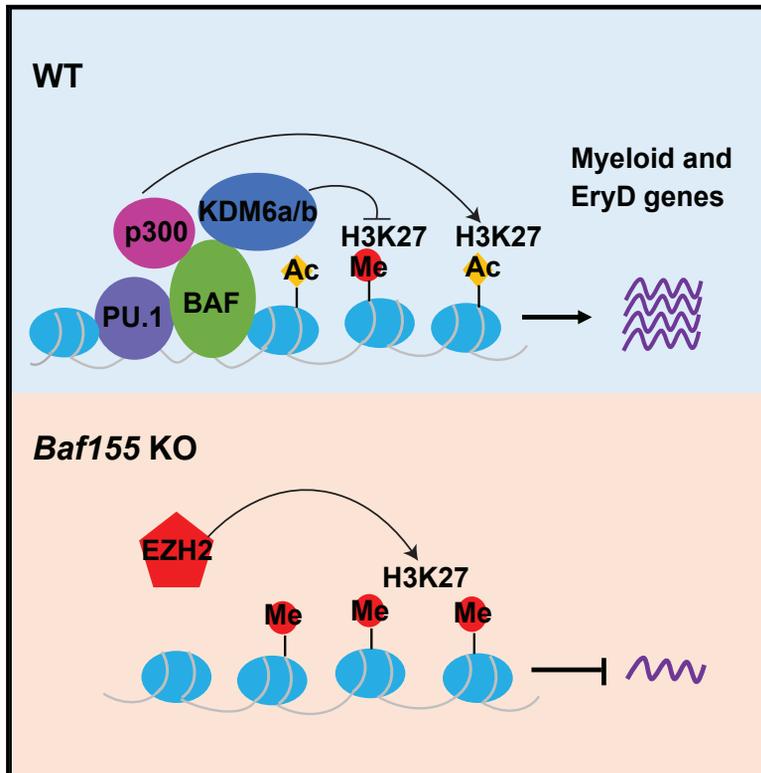
PCR duplicates were removed and reads separately aligned to custom *mus*/129 and *cas* genomes using TopHat2 (Kim et al., 2013). Final allele-specific mapping to reference mm9 genome was generated based on SNPs (Pinter et al., 2012). Only uniquely aligned concordantly mapped sequences were used in downstream analysis. Counts per gene were calculated using featureCounts (Liao et al., 2014). Using MatLab (MathWorks), library sizes were normalized and genes with insufficient allelic information (<13 allele-specific reads) were removed. We also removed potentially miscalled genes from our alignment pipeline, defined as genes incorrectly assigned to *mus* from a pure *cas* RNA-seq library. Allele-specific RPKM was calculated using allelic ratio (allele-specific counts) applied to comp RPKM (total counts). Genes with comp RPKM<1 or overlapping unmappable regions were excluded from the analysis, along with *Xist* and *Tsix*.

### QUANTIFICATION AND STATISTICAL ANALYSIS

Statistical parameters including the statistical tests used and the values of *n*, *p*, and *R* are reported in the figures, figure legends, or associated main texts. Statistical significance is determined by the value of *p* < 0.05 by the indicated tests. For microscope images, *n* generally refers to the total number of counted cells or *Xist* clouds.

# Requisite Chromatin Remodeling for Myeloid and Erythroid Lineage Differentiation from Erythromyeloid Progenitors

## Graphical Abstract



## Authors

Jun Wu, Karen Krchma, Hyung Joo Lee, ..., Maxim N. Artyomov, Ting Wang, Kyunghye Choi

## Correspondence

kchoi@wustl.edu

## In Brief

The mammalian chromatin-remodeling BAF (BRG1/BRM-associated factor) complex has an essential role in developmental and pathological processes. Wu et al. show that BAF-mediated chromatin remodeling and activation of the myeloid and definitive erythroid transcriptional program at the EMP stage is critical for myeloid and definitive erythroid lineage development.

## Highlights

- *Baf155* is required for yolk sac myeloid and definitive erythroid lineage development
- BAF-mediated chromatin remodeling of myeloid gene loci occurs at the EMP stage
- Inaccessible chromatin in *Baf155*-deficient EMPs is enriched by the ETS binding motif
- BAF155 interacts with PU.1 and is recruited to PU.1 target gene loci



## Article

# Requisite Chromatin Remodeling for Myeloid and Erythroid Lineage Differentiation from Erythromyeloid Progenitors

Jun Wu,<sup>1,7</sup> Karen Krchma,<sup>1,7</sup> Hyung Joo Lee,<sup>2,3</sup> Sairam Prabhakar,<sup>1</sup> Xiaoli Wang,<sup>1</sup> Haiyong Zhao,<sup>1</sup> Xiaoyun Xing,<sup>2,3</sup> Rho H. Seong,<sup>4</sup> Daved H. Fremont,<sup>1</sup> Maxim N. Artyomov,<sup>1</sup> Ting Wang,<sup>2,3,6</sup> and Kyunghee Choi<sup>1,5,8,\*</sup>

<sup>1</sup>Department of Pathology and Immunology, Washington University School of Medicine, St. Louis, MO, USA

<sup>2</sup>Department of Genetics, Washington University School of Medicine, St. Louis, MO, USA

<sup>3</sup>Edison Family Center for Genome Sciences and Systems Biology, Washington University School of Medicine, St. Louis, MO, USA

<sup>4</sup>Department of Biological Sciences, Institute of Molecular Biology and Genetics, Seoul National University, Seoul, Korea

<sup>5</sup>Graduate School of Biotechnology, Kyung Hee University, Yong In, Korea

<sup>6</sup>McDonnell Genome Institute, Washington University School of Medicine, St Louis, MO, USA

<sup>7</sup>These authors contributed equally

<sup>8</sup>Lead Contact

\*Correspondence: [kchoi@wustl.edu](mailto:kchoi@wustl.edu)

<https://doi.org/10.1016/j.celrep.2020.108395>

## SUMMARY

The mammalian SWItch/Sucrose Non-Fermentable (SWI/SNF) chromatin-remodeling BAF (BRG1/BRM-associated factor) complex plays an essential role in developmental and pathological processes. We show that the deletion of *Baf155*, which encodes a subunit of the BAF complex, in the *Tie2(+)* lineage (*Baf155* (CKO) leads to defects in yolk sac myeloid and definitive erythroid (EryD) lineage differentiation from erythromyeloid progenitors (EMPs). The chromatin of myeloid gene loci in *Baf155* CKO EMPs is mostly inaccessible and enriched mainly by the ETS binding motif. BAF155 interacts with PU.1 and is recruited to PU.1 target gene loci together with p300 and KDM6a. Treatment of *Baf155* CKO embryos with GSK126, an H3K27me2/3 methyltransferase EZH2 inhibitor, rescues myeloid lineage gene expression. This study uncovers indispensable BAF-mediated chromatin remodeling of myeloid gene loci at the EMP stage. Future studies exploiting epigenetics in the generation and application of EMP derivatives for tissue repair, regeneration, and disease are warranted.

## INTRODUCTION

The mammalian hematopoietic system is established from multiple embryonic origins. The first tissue to produce blood cells is the yolk sac, which generates primitive erythroid (EryP) cells and megakaryocytes (MegPs) presumably from bipotential MegP erythroid progenitors (pMEPs) (Tober et al., 2007). Although EryPs, which express embryonic globin genes, can be detected as early as embryonic day 7.25 (E7.25) in the blood islands of the yolk sac, mature MegP cells are not detected until later, around E9.5, in the yolk sac (Tober et al., 2007). The primary function of EryP cells is to provide developing embryos with oxygen and nutrients to accommodate the rapid growth of the embryo. Until recently, hematopoietic stem cells (HSCs) originating from the hemogenic endothelium of the aorta-gonad-mesonephros (AGM) have been thought to initiate definitive hematopoiesis. However, this paradigm has been challenged by the identification of erythromyeloid progenitors (EMPs), which generate definitive erythroid (EryD) and myeloid cell lineages before the establishment of HSCs (Hoeffel and Ginhoux, 2018; McGrath et al., 2015; Ginhoux et al., 2010; Schulz et al., 2012;

Hashimoto et al., 2013; Gomez Perdiguero et al., 2015). EMPs develop transiently from the *Tie2*-expressing hemogenic endothelium of the yolk sac around E8.5–E10.5, migrate to the fetal liver, expand, and provide embryos with EryD and myeloid cells in fetal life (Chen et al., 2011; McGrath et al., 2015; Gomez Perdiguero et al., 2015; Palis, 2016). EMPs are ultimately replaced by HSCs, which generate a full spectrum of blood cells, including lymphoid cell lineages. The importance of EMPs was realized by the finding that tissue-resident macrophages originate from EMPs through monocyte intermediates, although microglia are believed to originate from yolk sac macrophages (Ginhoux et al., 2010). Despite the critical establishment of the cellular origin of tissue-resident macrophages from EMPs, few studies have examined the molecular mechanisms that regulate myeloid and EryD differentiation from EMPs.

In eukaryotes, DNA is packaged into nucleosomes and subsequent higher-order chromatin structures. As a result, DNA is not easily accessible for transcriptional machinery. Two fundamental mechanisms that allow cells to respond to signals and trigger gene expression are chromatin remodeling and histone modification. Transcription factors and other proteins are



believed to gain access to nucleosome-bound DNA using ATP-dependent chromatin-modifying and remodeling enzymes, such as the SWI/SNF complex family (Hargreaves and Crabtree, 2011; Kadoch and Crabtree, 2015). SWI/SNF complexes, by utilizing energy derived from ATP hydrolysis, destabilize histone-DNA interactions and create an open chromatin state. SWI/SNF complexes are composed of one core ATPase (Brg1 or Brm) and distinct paralogous BRG1/BRM-associated factor (BAF) subunit family members that can interact with cofactors, including transcription factors. Combinatorial assembly of alternative families of subunits confers functional specificity to BAF complexes in different tissues and cell types (Wu et al., 2009). A recent study showing that BAF60c, a subunit of BAF critical for heart development (Lickert et al., 2004), together with GATA4 and TBX5, can reprogram mesoderm to cardiomyocytes in the mouse embryo (Takeuchi and Bruneau, 2009) highlights the fact that SWI/SNF-mediated chromatin remodeling is integral to the lineage-specific transcriptional network. Notably, enforced expression of just one component of the multi-protein BAF complex, BAF60c, in that study was sufficient to confer lineage-specific gene regulation.

Although chromatin-remodeling proteins are critical for establishing cell lineage specification in mammalian development, functional details of how these proteins affect specific cell lineage development are still lacking. Gene knockout studies have demonstrated that BRG1 ATPase and BAF155 are required for blood and vascular development. Particularly, *Brg1*- or *Baf155*-deficient embryos die around implantation (Bultman et al., 2000; Kim et al., 2001). The peri-implantation lethality of *Baf155*-deficient embryos can be extended to mid-gestation by *Baf155* transgene expression (*Baf155*<sup>-/-</sup>; *Tg*<sup>+</sup>). However, these animals display severe blood vessel formation defects in the yolk sac around E10.5 (Han et al., 2008). Moreover, *Tie2-Cre*; *Brg1*<sup>fl/fl</sup> mice are embryonic lethal because of apoptosis of EryP and lack of embryonic globin gene expression (Griffin et al., 2008). Angiogenesis is also defective in these animals. While these studies demonstrate the critical role of BAF-mediated chromatin remodeling in EryP and angiogenesis, it remains unclear whether BAF-mediated chromatin remodeling is subsequently required for the yolk sac hematopoiesis. In this study, we determined BAF-mediated chromatin remodeling requirements from hemogenic endothelium to EMPs and myeloid and EryD lineage differentiation by deleting *Baf155* in the *Tie2* lineage. Our data demonstrate that BAF155-mediated chromatin remodeling of myeloid and EryD gene loci at the EMP stage is critical for activation of the myeloid and EryD transcriptional program as well as myeloid and EryD lineage differentiation.

## RESULTS

### **Baf155 CKO Mice Are Embryonic Lethal, Showing Defects in Myeloid and EryD Lineage Development**

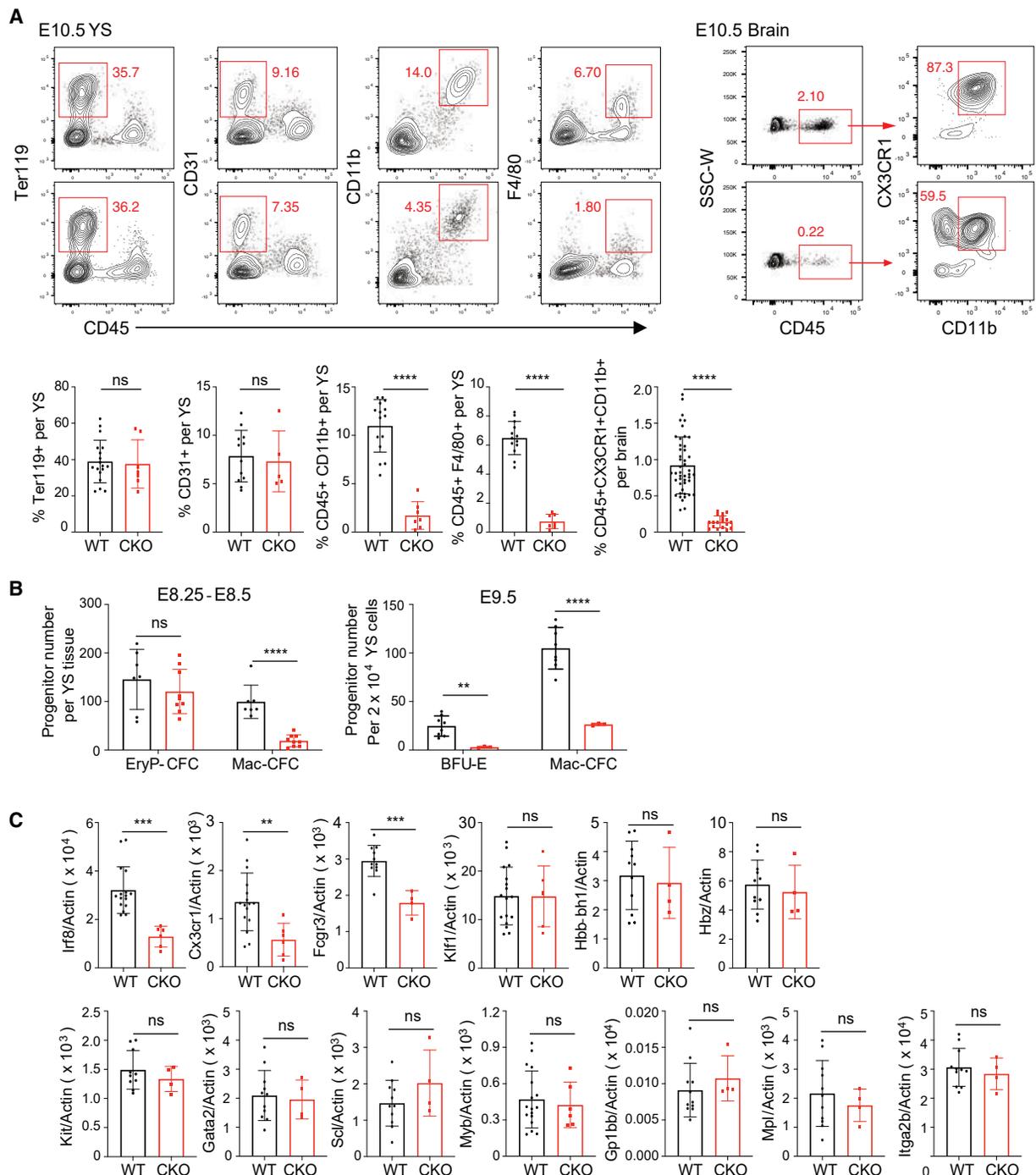
To assess the role of chromatin remodeling in hematopoietic lineage development downstream of the yolk sac hemogenic endothelium, we first generated *Tie2-Cre*; *Baf155*<sup>fl/fl</sup> mice from matings between *Baf155*<sup>fl/fl</sup> (Choi et al., 2012) and *Tie2-Cre* mice, which have been shown to target the yolk sac hemogenic endothelium (Chen et al., 2011; Gomez Perdiguerro et al., 2015).

We also observed effective tdTomato expression in EryP cells, endothelial cells, myeloid cells, and microglia in *Tie2-Cre*; *Rosa-loxp-stop-loxp-tdTomato* yolk sac and brain (Figure S1A). *Tie2-Cre*; *Baf155*<sup>fl/fl</sup> mice were then crossed with *Baf155*<sup>fl/fl</sup> mice to generate *Tie2-Cre*; *Baf155*<sup>fl/fl</sup> mice (hereafter called *Baf155* CKO). We found no live *Baf155* CKO mice at weaning (Table S1). We found no *Baf155* CKO embryos with abnormal morphology up to E9.5. However, at E10.5, some *Baf155* CKO embryos were smaller compared with littermate controls and displayed occasional hemorrhage around the distal end of the tail and yolk sac (8 of 49), suggesting abnormal vessel formation. This is consistent with previous findings that *Brg1* or *Baf155* deficiency leads to angiogenesis defects (Griffin et al., 2008; Han et al., 2008). All *Baf155* CKO embryos showed severe growth retardation and seemingly abnormal gross morphology at E13.5 (data not shown).

We assessed whether the embryonic lethality in *Baf155* CKO mice was due to hematopoietic defects. Ter119<sup>+</sup> erythroid cells from E9.5 and E10.5 yolk sacs, predominantly EryP cells at this stage, were present at similar levels in wild-type and *Baf155* CKO yolk sacs (Figures 1A and S1B). CD45<sup>+</sup>CD31<sup>+</sup> endothelial cells were also present similarly in wild-type and CKO yolk sacs (Figures 1A and S1B). Notably, there was a significant reduction in myeloid cells, CD45<sup>+</sup>CD11b<sup>+</sup>, in *Baf155* CKO yolk sacs (Figures 1A and S1B). Yolk sac macrophages, CD45<sup>+</sup>F4/80<sup>+</sup>, as well as microglia of the brain, CD45<sup>+</sup>CX3CR1<sup>+</sup>CD11b<sup>+</sup>, were also reduced significantly in *Baf155* CKO embryos (Figures 1A, S1B, and S1C). Although EryP progenitors from wild-type littermate control and *Baf155* CKO yolk sacs were present similarly, myeloid progenitors and BFU-Es were decreased significantly in *Baf155* CKO yolk sacs (Figure 1B). Corroborating these data, expression of myeloid lineage genes, including *Irf8*, *Cx3cr1*, and *Fcgr3* (CD16), was reduced significantly in *Baf155* CKO yolk sacs (Figures 1C and S1D). However, genes expressed in the EryP lineage, *Klf1*, *Hbb-bh1*, and *Hbz*, were detected similarly in wild-type and CKO yolk sacs (Figure 1C). Genes expressed in EMPs, *cKit*, *Gata2*, *Tal1/Scf*, and *Myb*, were also detected similarly in E9.5 wild-type (WT) and *Baf155* CKO yolk sacs (Figure 1C). *GP1bb* (CD42c), *Mpl*, and *Itga2b* (CD41), genes expressed in megakaryocytes, were also expressed similarly in WT and CKO yolk sacs (Figure 1C). These results suggest that *Baf155* deficiency leads to selective defects in myeloid and EryD lineage development.

### **EMPs Develop in the Absence of Baf155**

The selective myeloid and EryD lineage defects seen in *Baf155* CKO yolk sacs might be due to defects in EMP generation. Alternatively, EMPs are generated, but their differentiation might be blocked. To differentiate these two possibilities, we first analyzed EMPs based on phenotypic markers in *Baf155* CKO yolk sacs. EMPs are enriched in cKIT<sup>+</sup>CD41<sup>+</sup>CD16/32<sup>+</sup> cells (McGrath et al., 2015). Because expression of *Fcgr3* encoding CD16 was clearly decreased in *Baf155* CKO yolk sacs (Figure 1C), we reasoned that inclusion of CD16/32 in the EMP analysis might be inadequate for measuring EMPs from *Baf155* CKO yolk sacs. Indeed, the cKIT<sup>+</sup>CD16/32<sup>+</sup>CD41<sup>+</sup> or CD16/32<sup>+</sup>CD41<sup>+</sup> cell population was reduced greatly in *Baf155* CKO yolk sacs (Figure S2). Because cKIT is critical for EMP



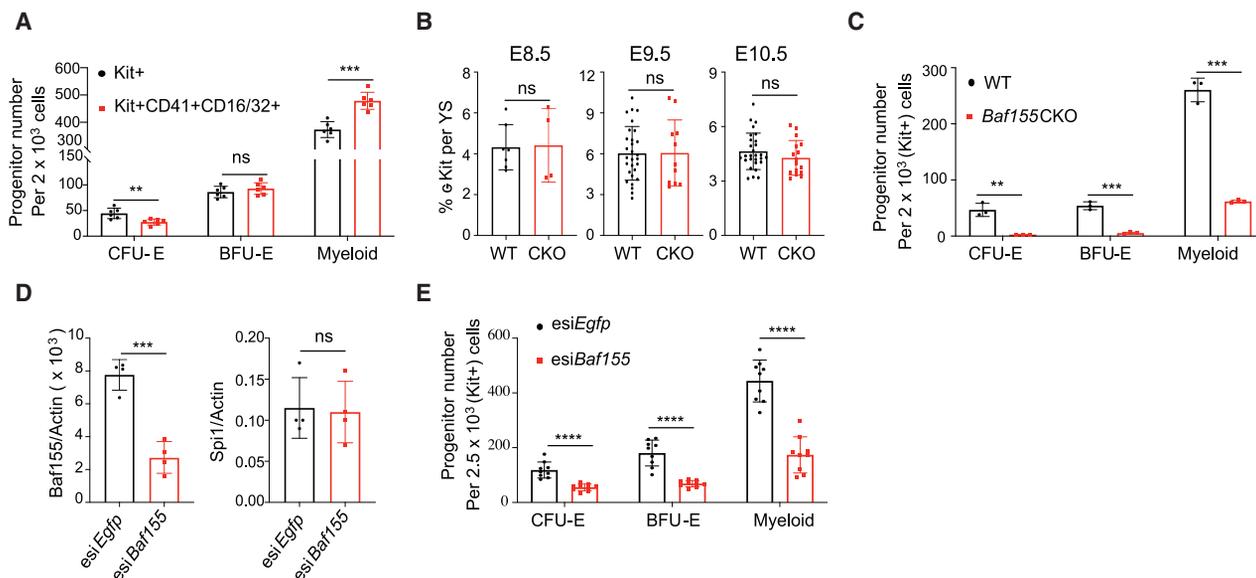
**Figure 1. *Baf155* KO Mice Show Defects in Myeloid and EryD Lineage Development**

(A) A representative flow cytometry analysis of E10.5 yolk sacs (YS) EryP cells (CD45<sup>+</sup>Ter119<sup>+</sup>), ECs (CD45<sup>+</sup>CD31<sup>+</sup>), myeloid cells (CD45<sup>+</sup>CD11b<sup>+</sup>), macrophages (CD45<sup>+</sup>F4/80<sup>+</sup>), and brain microglia (CD45<sup>+</sup>CX3CR1<sup>+</sup>CD11b<sup>+</sup>) in wild-type (WT) and *Baf155* KO mice is shown in the top panel. The percentage of each population is shown in the bottom panel. At least 5 biological replicates in 4 independent experiments for either genotype were analyzed, each representing an individual YS. Data are presented as mean  $\pm$  SD. Student's t test; ns, not significant; \*\*\*\*p < 0.0001.

(B) Distribution of EryP (EryP-CFC), EryD (BFU-E), and macrophage (Mac-CFC) progenitors from E8.25–E8.5/E9.5 WT and *Baf155* KO YS cells. E8.25–E8.5 data from 7 WT and 9 *Baf155* KO biological replicates and E9.5 data from 8 WT and 3 *Baf155* KO biological replicates representing two independent experiments, with each replicate consisting of a single YS, are shown. Data are presented as mean  $\pm$  SD. Student's t test; \*\*p < 0.005, \*\*\*\*p < 0.0001.

(C) qRT-PCR analysis of the indicated gene expression in E9.5 WT and *Baf155* KO YS cells is shown. Data are presented as mean  $\pm$  SD. Student's t test; \*\*p < 0.005, \*\*\*p < 0.001.

See also Figure S1.



**Figure 2. *Baf155* Is Required for Myeloid and EryD Lineage Differentiation from EMPs**

(A) Distribution of CFU-E, BFU-E, and myeloid colonies developing from  $\text{cKIT}^+$  and  $\text{cKIT}^+\text{CD41}^+\text{CD16/32}^+$  population from WT E10.5 YSs. Data are presented as mean  $\pm$  SD. Student's t test; \*\* $p < 0.005$ , \*\*\* $p < 0.001$ ; 6 biological replicates.

(B) Flow cytometry analysis of the  $\text{cKIT}^+$  population per YS from WT and *Baf155* CKO embryos on the indicated embryonic day. E8.5 data (mean  $\pm$  SD) are from 2 independent experiments. E9.5 and E10.5 data (mean  $\pm$  SD) are from 6 independent experiments, each representing a single YS. Data are presented as mean  $\pm$  SD. Student's t test.

(C) CFU-E, BFU-E, and myeloid colonies from  $\text{cKIT}^+$  cells from WT ( $n = 3$ ) and *Baf155* CKO ( $n = 3$ ) YSs. Data are from 2 independent experiments, with each replicate consisting of a single or 2 pooled YSs of the same genotype. Data are presented as mean  $\pm$  SD. Student's t test; \*\* $p < 0.01$ , \*\*\* $p < 0.001$ .

(D) qRT-PCR analysis of the indicated gene expression in  $\text{cKIT}^+$  cells from E10.5 WT YSs transfected with esiRNA against *Baf155* or *Egfp*. Data are presented as mean  $\pm$  SD. Student's t test; \*\*\* $p < 0.001$ ; four biological replicates from 3 independent experiments.

(E) CFU-E, BFU-E, and myeloid colonies from  $\text{cKIT}^+$  cells from E10.5 WT YSs transfected with esiRNA against *Baf155* or *Egfp*. Data are presented as mean  $\pm$  SD. Student's t test; \*\*\* $p < 0.001$ . Nine biological replicates from 3 independent experiments.

See also Figure S2.

development (Azzoni et al., 2018), and *cKit* expression was similar in WT and CKO yolk sacs (Figure 1C), we sorted  $\text{cKIT}^+$  cells and evaluated their myeloid and EryD potential compared with  $\text{cKIT}^+\text{CD41}^+\text{CD16/32}^+$  cells. There was an  $\sim 80\%$  overlap between  $\text{cKIT}^+$  and  $\text{cKIT}^+\text{CD16/32}^+\text{CD41}^+$  cell populations (data not shown). Although  $\text{cKIT}^+$  and  $\text{cKIT}^+\text{CD41}^+\text{CD16/32}^+$  cells generated similar levels of BFU-Es,  $\text{cKIT}^+$  cells generated slightly more CFU-Es compared with  $\text{cKIT}^+\text{CD41}^+\text{CD16/32}^+$  cells, suggesting that  $\text{cKIT}^+$  cells also contain more committed erythroid progenitors (Figure 2A). Slightly more myeloid colonies were generated from  $\text{cKIT}^+\text{CD41}^+\text{CD16/32}^+$  cells compared with  $\text{cKIT}^+$  cells, suggesting that  $\text{cKIT}^+\text{CD41}^+\text{CD16/32}^+$  cells also enrich committed myeloid progenitors (Figure 2A). Importantly,  $\text{cKIT}^+$  cells were present at similar levels in WT and *Baf155* CKO yolk sacs at E8.5–E10.5 (Figure 2B). However,  $\text{cKIT}^+$  cells from E10.5 *Baf155* CKO yolk sacs generated significantly fewer myeloid or BFU-E colonies compared with controls (Figure 2C), suggesting that EMPs were generated in *Baf155* CKO embryos but that *Baf155* CKO EMPs have a block in myeloid and EryD differentiation. We assessed whether *Baf155* inhibition in EMPs was sufficient to block myeloid and EryD differentiation. To this end, we knocked down *Baf155* in EMPs by transfecting *Baf155* esiRNA into sorted  $\text{cKIT}^+$  cells. We achieved more than 60% KD efficiency of *Baf155* expression, whereas expression of irrel-

evant genes, such as *Pu.1*, was unaffected (Figure 2D). Notably, myeloid and EryD output from EMPs was reduced greatly by *Baf155* knockdown (KD) (Figure 2E), indicating that acute deletion of *Baf155* in EMPs is sufficient to inhibit myeloid and EryD lineage differentiation. These data collectively suggest that *Baf155*-mediated chromatin remodeling at the EMP stage is critical for efficient downstream myeloid and EryD lineage differentiation.

### Single-Cell RNA Sequencing Reveals Myeloid Lineage Differentiation Defects of *Baf155*-Deficient EMPs

To better understand the myeloid lineage differentiation block in *Baf155* CKO embryos, we subjected yolk sacs from WT and *Baf155* CKO mice to single-cell RNA sequencing (scRNA-seq). After filtering out low-quality cells, 722 WT and 791 *Baf155* CKO yolk sac cells were chosen for further analysis. t-stochastic neighbor embedding (t-SNE) was used to visualize the populations. WT yolk sac cells were clustered into 7 populations based on similarities of the transcriptome (Figure S3A). High expression of *Kdr*, *Cdh5*, and *Pecam1* highlights cluster 5 to be an endothelial cell population (Figure S3B). Two distinct erythroid cell populations were visible based on the erythroid lineage markers *Gata1*, *Klf1*, and *EpoR* (Figure S3C). Expression of *Hbb-y*, *Hbb-x*, and *Hbb-bh1* (embryonic  $\beta$ -globin genes) separated

cluster 2 as primitive and cluster 4 as an EryD cell population (Figure S3D). We also identified *Gm15915*, *Ccl17*, *Muc13*, and *Gdf3* to be expressed in an EryD-specific manner (Figures S3E and S3I). Expression of the mature myeloid lineage genes *Trem2*, *Emr1*, *Cx3cr1*, *Csf1r*, *Irf8*, and *Cd68* identified cluster 6 as a myeloid cell population (Figures S3F and S3J). Enriched expression of *cKit* and *Itga2b*, encoding CD41, and *Cd34* identified cluster 0 as EMPs (Figure S3G). Expression of *Pu.1* (Spi1), a critical factor for myeloid lineage development, expression was high in EMPs, and its expression was detected continuously in the myeloid lineage arm (Figures S3G and S3K). *Bcl11a*, *Myb*, *Scf* (Tal1), and *Gata2* expression was detected in EMPs, and their expression was detected continuously in the EryD cell cluster (Figure S3K). It is also notable that *Baf155* and *Cd34* expression was high in the EMP population (Figure S3M). As we reported recently (Zhao and Choi, 2019), clusters 1 and 3 represent smooth muscle cells and pericytes, based on *Hand1*, *Hand2*, *Acta2*, *Tbx20*, and *Cd248* expression (Figures S3H and S3L). *Desmin* expression separated pericytes (cluster 3) from smooth muscle cells (cluster 1; Figure S3L).

When we overlaid WT and *Baf155* CKO scRNA-seq data, we observed that the transcriptomes of endothelial cells (ECs; Figure 3A, cluster 7), EryP cells (Figure 3A, clusters 1 and 6), and smooth muscle cells and pericytes (Figure 3A, clusters 2, 4, 5, and 8) overlapped each other, indicating that *Baf155* deficiency did not affect their overall transcriptome (Figures 3A–3D). However, clusters 0 and 3, both expressing EMP genes (Figures 3E and 3F), showed clear separation between WT and *Baf155* CKO yolk sac cells (Figures 3B–3D). *Baf155* expression was clearly absent in cluster 3, indicating that *Baf155* was deleted effectively in this population (Figure 3B). Strikingly, a population with the mature myeloid gene signature (cluster 9) was readily visible in WT yolk sacs but absent in *Baf155* CKO yolk sacs (Figures 3E and 3G). A population with the EryD gene signature was also reduced greatly in *Baf155* CKO yolk sacs (Figures 3E and 3H). Intriguingly, although EryD cells were reduced greatly, the megakaryocytic lineage gene signature was high in the presumptive *Baf155* CKO EryD cell population (Figure 3I). This suggests that the megakaryocytic lineage may be the default pathway in EryD and megakaryocytic lineage choice and that chromatin remodeling is also critical for EryD and megakaryocyte lineage bifurcation. Additionally, endothelial genes were still expressed in the *Baf155* CKO EMP cell population (Figure 3J). These data collectively suggest that chromatin remodeling is needed at the EMP stage for further differentiation into myeloid and EryD lineages to occur. The endothelial gene program is sustained in EMPs when subsequent differentiation is blocked. Alternatively, termination of the endothelial gene program might require chromatin remodeling.

### Chromatin Accessibility of Myeloid and EryD Gene Loci Is Reduced Greatly in *Baf155* CKO EMPs

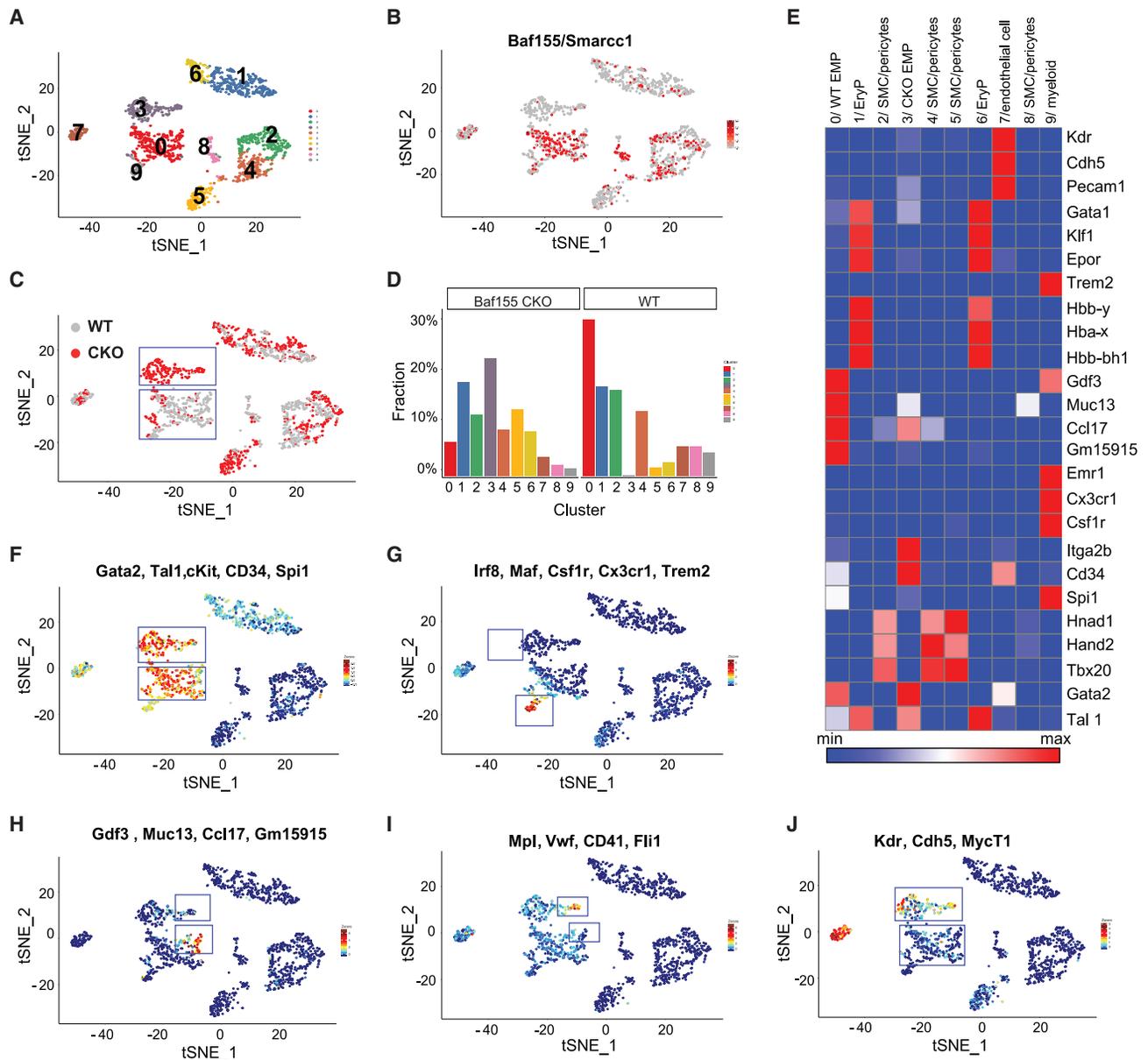
So far, the data suggested that *Baf155* deficiency leads to myeloid and EryD differentiation block from EMPs. To better understand the myeloid lineage differentiation defect in *Baf155*-deficient EMPs, we sorted cKIT<sup>+</sup> cells, EMP enriched, from WT and *Baf155* CKO yolk sacs (Figure S4A) and assessed genome-wide chromatin accessibility by assay for transpo-

sase-accessible chromatin using sequencing (ATAC-seq; Buenroostro et al., 2013). We identified 103,043 and 143,021 accessible chromatin regions in *Baf155* CKO and WT cKIT<sup>+</sup> cells, respectively. Among these ATAC-seq peaks, 7,422 regions were more accessible in WT than in *Baf155* CKO cells, whereas only 56 regions were more accessible in *Baf155* CKO cells than in WT cells (Figures 4A and 4B). The differentially accessible genomic regions (DARs) were enriched for the binding motifs of transcription factors (TFs) such as PU.1, IRF8, AP-1, and CEBP, whereas 13,213 unaffected accessible genomic regions, commonly open in CKO and WT EMPs (fold change < 1.1 and *p* > 0.05), were enriched for the different TF binding motifs, such as CTCF (Figures 4B, 4C, and S4B). Of the 7,422 DARs specific to WT EMPs, 3,679 peaks contained PU.1 motifs, whereas 1,472 of 13,213 unaffected peaks contained PU.1 motifs. This suggests that loss of *Baf155* leads to a closed chromatin structure at selective genomic regions. The data also suggest potential interplay between ETS factors and BAF-mediated chromatin remodeling in activating the myeloid lineage program.

The genes associated with selective genomic regions with loss of chromatin accessibility in *Baf155* CKO EMPs were enriched for biological functions related to immune responses, including inflammatory response and myeloid leukocyte activation, whereas genes with unaffected accessible regions were enriched for biological functions different from those (Figure 4D). Genes under the inflammatory response and myeloid leukocyte activation categories mainly include myeloid genes such as *Cx3cr1*, *Trem2*, *Csf1r*, *Irf8*, *Emr1*, and *Fcgr3* (CD16). Importantly, chromatin of these gene loci was largely inaccessible in *Baf155* CKO EMPs (Figures 4E and S4C), explaining the block of myeloid lineage differentiation. In contrast, EC gene loci, *Kdr*, *Cdh5*, and *Esam*, were similarly accessible in WT and *Baf155* CKO EMPs (Figure S4D). EMP gene loci, *cKit*, *Itga2b* (*Cd41*), and *Cd34*, were also readily accessible in WT and *Baf155* CKO EMPs (Figure S4E). Importantly, although chromatin of erythroid lineage genes that were commonly expressed in EryP cells was readily accessible, chromatin of the erythroid lineage genes that were specifically upregulated in EryD cells, *Hbb-bt*, *Gm15915*, and *Gdf3*, were not (Figure 4F). These data suggest that BAF155-mediated chromatin remodeling of myeloid and EryD gene loci at the EMP stage is necessary for subsequent myeloid and EryD lineage differentiation.

### BAF155 Is Recruited to PU.1 Target Gene Loci

PU.1 is a master ETS factor critical for myeloid lineage development. PU.1 regulates its own expression (Chen et al., 1995), although the mechanisms of this autoregulation have not been elucidated clearly. Because reduced accessible regions in *Baf155* CKO EMPs are represented predominantly by the ETS binding motif, we assessed whether PU.1 requires the BAF complex for activating the myeloid lineage program. Intriguingly, *Pu.1* expression itself was diminished in CKO yolk sacs (Figure 5A). Chromatin of the *Pu.1* locus was less accessible in CKO EMPs (Figure 5A). Because *Baf155* KD in EMPs could block myeloid and EryD differentiation without affecting *Pu.1* expression (Figures 2D and 2E), we reasoned that diminished expression of *Pu.1* and its target genes in *Baf155* CKO yolk sacs could be due to deficiency of BAF-mediated chromatin remodeling at

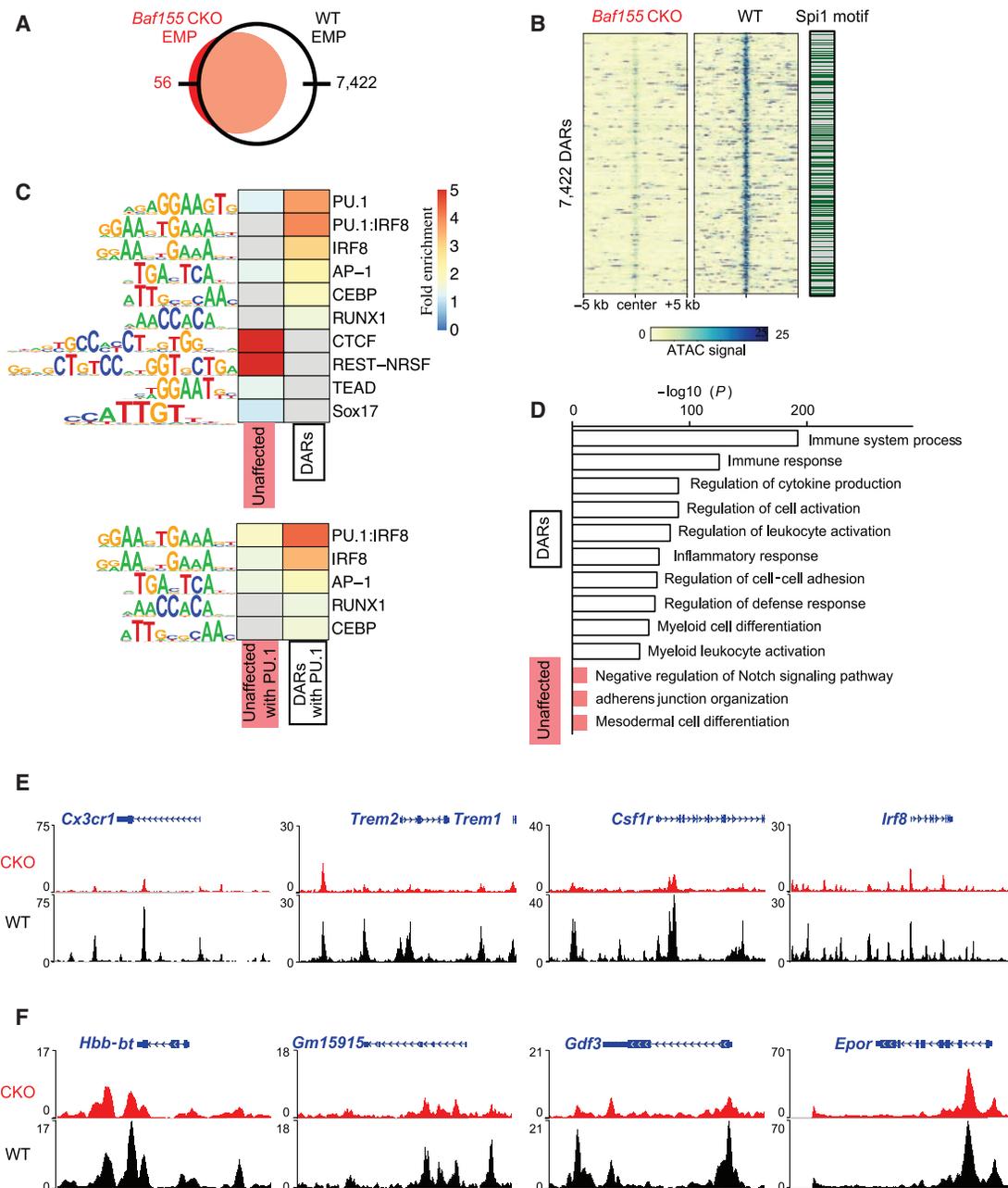


**Figure 3. scRNA-Seq Data Reveal Myeloid and EryD Differentiation Defects from *Baf155*-Deficient EMPs**

- (A) t-SNE projection of all cells, showing 10 different clusters.  
 (B) *Baf155/Smarcc1* expression in all YSs, showing the absence of *Baf155* expression in the CKO EMP cell population.  
 (C) An overlay of scRNA-seq data between WT and *Baf155* CKO YSs.  
 (D) Percentage of cells in each cluster from WT versus *Baf155* CKO YSs.  
 (E) Heatmap showing differentially expressed genes in each cluster.  
 (F) The EMP signature genes *Gata2*, *Tal1*, *cKit*, *Cd34*, and *Pu.1/Spi1* are similarly expressed in WT and *Baf155* CKO EMPs.  
 (G) A cell population with myeloid lineage signature gene expression (*Irf8*, *Maf*, *Csf1r*, *Cx3cr1*, and *Trem2*) is absent in *Baf155* CKO YSs.  
 (H) A cell population with EryD lineage signature genes expression (*Gdf3*, *Muc13*, *Ccl17*, and *Gm15915*) is significantly lower in *Baf155* CKO YSs.  
 (I) A population with elevated megakaryocyte lineage signature gene expression is increased in *Baf155* CKO YS cells.  
 (J) Endothelial lineage signature genes are still expressed in the *Baf155*-deficient EMP cell population.  
 See also [Figure S3](#).

*Pu.1* and its target gene loci. We first assessed whether BAF155 is recruited to PU.1 target genes. We selected 7 genomic regions that contain the ETS motif and are differentially accessible in WT and *Baf155* CKO yolk sacs. These include genomic regions that

are associated with the *Cx3cr1*, *Trem2*, *Csf1r*, *Irf8*, and *Cd68* genes (Figure 5B; Table S3). We also selected 2 unaffected ETS regions, UER1 and UER2, from the 1,472 peaks that contain the PU.1 motif but whose chromatin accessibility is unaffected



**Figure 4. *Baf155* CKO EMPs Have Reduced Chromatin Accessibility at the Myeloid and EryD Gene Loci**

(A) A Venn diagram of the numbers of ATAC-seq peaks found in *Baf155* CKO and WT EMPs.

(B) ATAC-seq signals over 10-kb regions centered on the differentially accessible regions (DARs) with reduced signals in *Baf155* CKO EMPs compared with the WT (left) and the presence of the *Spi1/Pu.1* motif in the DARs (right).

(C) Heatmaps of HOMER known TF motif fold enrichment in the DARs and unaffected accessible regions. Gray cells indicate no enrichment found ( $p > 0.05$ ).

(D) Enriched Gene Ontology (GO) terms and their binomial p values from analyzing the DARs with reduced signals in *Baf155* CKO EMPs (white) and the unaffected peaks (red) using GREAT.

(E) Epigenome browser views of representative myeloid gene loci.

(F) Epigenome browser views of representative erythroid gene loci.

See also Figure S4.

by *Baf155* deficiency (Figure 5B; Table S3). Although a significant mean enrichment was observed for PU.1 and BAF155 binding at these PU.1 target gene loci, only BAF155 binding, not

PU.1, was enriched at UER1 and UER2 (Figure 5B). We next determined whether BAF155 can form a complex with PU.1. Specifically, we generated a mouse embryonic fibroblast (MEF)



line that expresses *Flag-Baf155* and *Pu.1* or *HA-Pu.1*. Cells were then subjected to immunoprecipitation using an antibody against the FLAG tag, followed by PU.1 or hemagglutinin (HA) immunoblot. PU.1 was co-immunoprecipitated with BAF155 (Figure 5C). Conversely, when the anti-HA antibody was used for immunoprecipitation, BAF155 was co-immunoprecipitated with PU.1 (Figure S5A). As reported previously (Alver et al., 2017; Narayanan et al., 2015), we additionally found BRG1, p300, and KDM6a (UTX) to be co-immunoprecipitated with BAF155, suggesting that PU.1 activates its target genes by forming a transcriptional complex with BAF, p300, and KDM6a. Because p300 and KDM6a mainly target H3K27, we postulated that PU.1 target gene loci remain methylated at H3K27 sites in the absence of BAF155, leading to repression of PU.1 target gene expression. Indeed, although inaccessible regions in *Baf155* CKO yolk sacs showed higher H3K27me3 levels compared with WT yolk sacs, H3K27me3 levels were similar at UER1 and UER2 in WT and CKO yolk sacs (Figure S5B). If this were truly the case, then we would expect that inhibition of EZH2, the catalytic subunit of Polycomb repressive complex 2 (PRC2), which methylates H3K27 (Laugesen et al., 2019), might rescue PU.1 target gene expression in *Baf155* CKO yolk sacs. Thus, we set up matings between *Tie2-Cre;Baf155<sup>f/+</sup>* (father) and *Baf155<sup>f/f</sup>* (mother) mice and injected GSK-126 (an EZH2 inhibitor) intraperitoneally into the mother at E8. E9.5 yolk sacs were collected and subjected to qRT-PCR. Although vehicle treatment did not affect myeloid lineage gene expression, indicating that the GSK-126 effect would be specific (Figure S5C), expression of many PU.1 target genes was rescued in *Baf155* CKO yolk sacs when EZH2 was inhibited (Figure 5D). *Pu.1/Spi1* expression was also rescued, suggesting the BAF-mediated remodeling mechanism of PU.1 autoregulation. It is worth noting that GSK-126 treatment led to baseline elevation of expression of some genes, including *Irf8*, *Itgam*, *Gata2*, *Scl*, and *cKit* (Figures 5D and S5D), suggesting that these genes are normally repressed by the EZH2-mediated mechanism.

## DISCUSSION

Chromatin remodeling by the mammalian BAF complex is required for the development of multiple lineages during embryogenesis. *Brg1* deletion within the *Tie2(+)* lineage leads to defective yolk sac angiogenesis and primitive erythropoiesis. EryP cell defects in these mice are due to increased apoptosis and lack of embryonic  $\alpha$ - and  $\beta$ -globin gene expression (Griffin et al., 2008). We show that *Baf155* deletion within the *Tie2(+)* lineage also causes angiogenesis defects, as evidenced by hemorrhage in some *Baf155* CKO mice. However, we found that embryonic globin gene expression was similar, and EryPs and their progenitors were present at similar levels in WT and

*Baf155* CKO yolk sacs. Unexpectedly, *Baf155* CKO mice display defective yolk sac myelopoiesis and definitive erythropoiesis. We attribute the phenotype difference between the two mice, *Tie2-Cre; Brg1<sup>f/f</sup>* and *Tie2-Cre; Baf155<sup>f/f</sup>*, to the nature of the deleted gene. *Brg1* encodes for an ATPase that is the core of the BAF complex, whereas *Baf155* encodes a BAF structural component. Presumably, *Baf155* deletion might have delayed the phenotype's manifestation to reflect the hematopoietic lineage hierarchy; i.e., the EC and EryP lineages arise before EMPs, which generate myeloid and EryD lineages. Intriguingly, we observed that EC genes were still expressed in *Baf155*-deficient EMPs, suggesting that the extinction of previous lineage genes is necessary for new lineage establishment. Alternatively, the previous lineage gene loci are still accessible in the absence of the next lineage gene loci's active chromatin remodeling. Collectively, these data establish that the BAF complex has a critical role in myeloid and EryD lineage differentiation from EMPs by remodeling the chromatin of the myeloid and EryD lineage gene loci.

Although chromatin remodeling is critical for developing many different lineages, it is still unclear whether and how the specificity of the BAF chromatin remodeling complex of the target genes is achieved. Our data demonstrate that BAF-mediated chromatin remodeling of myeloid and EryD lineage genes at the EMP stage is necessary for downstream myeloid and EryD lineage development. We found that DARs in WT and *Baf155*-deficient EMPs are enriched predominantly for the ETS motif. Moreover, BAF155 interacted with PU.1 and was recruited to PU.1 target gene loci. This strongly argues that the BAF complex's target gene specificity is achieved by ETS TFs. Consistent with this idea, recent studies have shown that AP-1 and ETS motifs are enriched in enhancer regions sensitive to *Smadcb1* loss (Alver et al., 2017). Moreover, TMPRSS2-ERG, a fusion gene product from a chromosomal translocation in prostate cancer, interacts with the BAF complex in an ETS-dependent manner (Sandoval et al., 2018), indicating that the BAF complex is required for ERG-mediated prostate oncogenesis. The BAF complex has been shown to interact with p300 and acetylates H3K27 (Alver et al., 2017). The BAF complex also interacts with KDM6a/6b and demethylates H3K27 (Narayanan et al., 2015). We also observed that BAF155 forms a complex with p300 and KDM6a. Our data show that DARs in *Baf155* CKO EMPs included mostly PU.1 target genes and displayed higher H3K27me3 levels. EZH2 inhibitor treatment could rescue some of the PU.1 target gene expression in *Baf155* CKO yolk sacs. These data suggest that PU.1 activates its target genes by forming a complex with BAF, p300, and KDM6a/6b and triggering H3K27 acetylation/demethylation of the target genes (Figure 5E). In the absence of the BAF complex, PU.1 target loci are occupied by EZH2, suppressing PU.1 target gene expression (Figure 5E).

(D) qRT-PCR analysis of *Cx3cr1*, *Irf8*, *Csf-1r*, *Pu.1/Spi1*, *Itgam*, and *Emr1* gene expression in E9.5 WT and *Baf155* CKO YSs with or without GSK126 treatment. Gene expression was normalized to the untreated WT mean value. Data are from at least four biological replicates for either genotype, with each replicate consisting of an individual YS. Data are presented as mean  $\pm$  SD. Student's t test; \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001.

(E) A model showing BAF-mediated chromatin remodeling in PU.1 transcriptional gene activation. See also Figure S5.

Reduced *Pu.1* expression leads to acute myeloid leukemia (Will et al., 2015; Steidl et al., 2007). Moreover, *Pu.1* activation in hematopoietic stem and progenitor cells can lead to myeloid lineage skewing and deregulated hematopoiesis in chronic inflammatory conditions (Pietras et al., 2016; Etzrodt et al., 2019). Methylation of BAF155 at the R1064 residue by coactivator-associated arginine methyltransferase 1 (CARM1; also known as PRMT4) is critical for tumor progression and metastasis (Wang et al., 2014). Although CARM1 is essential for myeloid leukemogenesis, it is dispensable for normal hematopoiesis (Greenblatt et al., 2018). UTX (KDM6a) suppresses myeloid leukemogenesis partially by repressing an ETS-mediated transcriptional program (Gozdecka et al., 2018). *Kdm6b* is required for fetus-derived T-ALL and adult-derived AML (Mallaney et al., 2019). These studies collectively suggest that controlling *Pu.1* expression and its activity might be critical for managing cancer and chronic inflammatory diseases. Intriguingly, ETS factors can regulate *Baf155* expression (Ahn et al., 2005), supporting an interplay between ETS TFs and BAF expression and function. Future studies delineating the crosstalk between ETS factors and BAF and interaction among BAF, p300, PU.1, and KDM6a/6b and PU.1 target gene expression will be critical for further understanding myeloid lineage and leukemia development. Future applications of the epigenetics involving PU.1 and BAF155 expression and function for tissue repair, regeneration, and diseases are warranted.

## STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead Contact
  - Materials Availability
  - Data and Code Availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
  - Mice
  - Cell culture and transduction
- METHOD DETAILS
  - Genotyping
  - Lentiviral and retroviral particle production
  - esiRNA transfection
  - Hematopoietic progenitor assays
  - *In vivo* GSK-126 treatment
  - Nuclear extract preparation
  - Immunoprecipitation
  - Western blotting
  - Chromatin immunoprecipitation (ChIP) coupled with quantitative real-time PCR (ChIP-qPCR)
  - Tissue processing for flow cytometry
  - Flow cytometry and cell sorting
  - Quantitative real-time reverse transcription PCR (qRT-PCR)
  - Single-cell RNA sequencing
  - ATAC-seq library generation, sequencing, and mapping

## ● QUANTIFICATION AND STATISTICAL ANALYSIS

- scRNA-seq bioinformatics analyses
- Identification of ATAC peaks
- Identification and analyses of DARs
- Statistics

## SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at <https://doi.org/10.1016/j.celrep.2020.108395>.

## ACKNOWLEDGMENTS

We thank the Choi lab members for constructive criticism and support. We also thank the Washington University Pathology FACS core. This work was supported by NIH grants R01HL55337 and R01HL149954 (to K.C.) and the Edward Mallinckrodt Foundation (to K.C. and D.H.F.).

## AUTHOR CONTRIBUTIONS

J.W., K.K., and K.C. conceived and designed experiments and wrote the paper. J.W. and K.K. performed the experiments and analyzed the data. H.J.L. and X.X. performed and analyzed the ATAC-seq data. H.Z., S.P., and M.N.A. performed and analyzed the scRNA-seq data. R.H.S. provided the *Baf155*<sup>fl/fl</sup> mice. X.W. and D.H.F. performed BAF155 and PU.1 immunoprecipitation and western blot analyses. S.P., H.J.L., M.N.A., and T.W. helped write the manuscript. K.C. provided overall supervision and coordinated all experimental activities.

## DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: February 24, 2020  
 Revised: September 25, 2020  
 Accepted: October 26, 2020  
 Published: November 17, 2020

## REFERENCES

- Ahn, J., Ko, M., Lee, K., Oh, J., Jeon, S.H., and Seong, R.H. (2005). Expression of SRG3, a core component of mouse SWI/SNF chromatin-remodeling complex, is regulated by cooperative interactions between Sp1/Sp3 and Ets transcription factors. *Biochem. Biophys. Res. Commun.* 338, 1435–1446.
- Alver, B.H., Kim, K.H., Lu, P., Wang, X., Manchester, H.E., Wang, W., Haswell, J.R., Park, P.J., and Roberts, C.W. (2017). The SWI/SNF chromatin remodeling complex is required for maintenance of lineage specific enhancers. *Nat. Commun.* 8, 14648.
- Azzoni, E., Frontera, V., McGrath, K.E., Harman, J., Carrelha, J., Nerlov, C., Palis, J., Jacobsen, S.E.W., and de Bruijn, M.F. (2018). Kit ligand has a critical role in mouse yolk sac and aorta-gonad-mesonephros hematopoiesis. *EMBO Rep.* 19, e45477.
- 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.
- Bultman, S., Gebuhr, T., Yee, D., La Mantia, C., Nicholson, J., Gilliam, A., Ranzazzo, F., Metzger, D., Chambon, P., Crabtree, G., and Magnuson, T. (2000). A *Brg1* null mutation in the mouse reveals functional differences among mammalian SWI/SNF complexes. *Mol. Cell* 6, 1287–1295.
- Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* 36, 411–420.

- Chen, H., Ray-Gallet, D., Zhang, P., Hetherington, C.J., Gonzalez, D.A., Zhang, D.E., Moreau-Gachelin, F., and Tenen, D.G. (1995). PU.1 (Spi-1) autoregulates its expression in myeloid cells. *Oncogene* *11*, 1549–1560.
- Chen, M.J., Li, Y., De Obaldia, M.E., Yang, Q., Yzaguirre, A.D., Yamada-Inagawa, T., Vink, C.S., Bhandoola, A., Dzierzak, E., and Speck, N.A. (2011). Erythroid/myeloid progenitors and hematopoietic stem cells originate from distinct populations of endothelial cells. *Cell Stem Cell* *9*, 541–552.
- Choi, J., Ko, M., Jeon, S., Jeon, Y., Park, K., Lee, C., Lee, H., and Seong, R.H. (2012). The SWI/SNF-like BAF complex is essential for early B cell development. *J. Immunol.* *188*, 3791–3803.
- Corces, M.R., Trevino, A.E., Hamilton, E.G., Greenside, P.G., Sinnott-Armstrong, N.A., Vesuna, S., Satpathy, A.T., Rubin, A.J., Montine, K.S., Wu, B., et al. (2017). An improved ATAC-seq protocol reduces background and enables interrogation of frozen tissues. *Nat. Methods* *14*, 959–962.
- Etzrodt, M., Ahmed, N., Hoppe, P.S., Loeffler, D., Skylaki, S., Hilsenbeck, O., Kokkaliaris, K.D., Kaltenbach, H.M., Stelling, J., Nerlov, C., and Schroeder, T. (2019). Inflammatory signals directly instruct PU.1 in HSCs via TNF. *Blood* *133*, 816–819.
- Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., Mehler, M.F., Conway, S.J., Ng, L.G., Stanley, E.R., et al. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science* *330*, 841–845.
- Gomez Perdiguero, E., Klapproth, K., Schulz, C., Busch, K., Azzone, E., Crozet, L., Garner, H., Trouillet, C., de Bruijn, M.F., Geissmann, F., and Rodewald, H.R. (2015). Tissue-resident macrophages originate from yolk-sac-derived erythroid progenitors. *Nature* *518*, 547–551.
- Gozdecka, M., Meduri, E., Mazan, M., Tzelepis, K., Dudek, M., Knights, A.J., Pardo, M., Yu, L., Choudhary, J.S., Metzakopian, E., et al. (2018). UTX-mediated enhancer and chromatin remodeling suppresses myeloid leukemogenesis through noncatalytic inverse regulation of ETS and GATA programs. *Nat. Genet.* *50*, 883–894.
- Greenblatt, S.M., Man, N., Hamard, P.J., Asai, T., Karl, D., Martinez, C., Bilbao, D., Stathias, V., Jermakowicz, A.M., Duffort, S., et al. (2018). CARM1 Is Essential for Myeloid Leukemogenesis but Dispensable for Normal Hematopoiesis. *Cancer Cell* *33*, 1111–1127.e5.
- Griffin, C.T., Brennan, J., and Magnuson, T. (2008). The chromatin-remodeling enzyme BRG1 plays an essential role in primitive erythropoiesis and vascular development. *Development* *135*, 493–500.
- Han, D., Jeon, S., Sohn, D.H., Lee, C., Ahn, S., Kim, W.K., Chung, H., and Seong, R.H. (2008). SRG3, a core component of mouse SWI/SNF complex, is essential for extra-embryonic vascular development. *Dev. Biol.* *315*, 136–146.
- Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* *21*, 396–420.
- Hashimoto, D., Chow, A., Noizat, C., Teo, P., Beasley, M.B., Leboeuf, M., Becker, C.D., See, P., Price, J., Lucas, D., et al. (2013). Tissue-resident macrophages self-maintain locally throughout adult life with minimal contribution from circulating monocytes. *Immunity* *38*, 792–804.
- Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol. Cell* *38*, 576–589.
- Hoeffel, G., and Ginhoux, F. (2018). Fetal monocytes and the origins of tissue-resident macrophages. *Cell. Immunol.* *330*, 5–15.
- Kadoch, C., and Crabtree, G.R. (2015). Mammalian SWI/SNF chromatin remodeling complexes and cancer: Mechanistic insights gained from human genomics. *Sci. Adv.* *1*, e1500447.
- Kim, J.K., Huh, S.O., Choi, H., Lee, K.S., Shin, D., Lee, C., Nam, J.S., Kim, H., Chung, H., Lee, H.W., et al. (2001). Srg3, a mouse homolog of yeast SWI3, is essential for early embryogenesis and involved in brain development. *Mol. Cell. Biol.* *21*, 7787–7795.
- Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* *9*, 357–359.
- Laugesen, A., Højfeldt, J.W., and Helin, K. (2019). Molecular Mechanisms Directing PRC2 Recruitment and H3K27 Methylation. *Mol. Cell* *74*, 8–18.
- Lickert, H., Takeuchi, J.K., Von Both, I., Walls, J.R., McAuliffe, F., Adamson, S.L., Henkelman, R.M., Wrana, J.L., Rossant, J., and Bruneau, B.G. (2004). Baf60c is essential for function of BAF chromatin remodeling complexes in heart development. *Nature* *432*, 107–112.
- Lybarger, L., Wang, X., Harris, M.R., Virgin, H.W., 4th, and Hansen, T.H. (2003). Virus subversion of the MHC class I peptide-loading complex. *Immunity* *18*, 121–130.
- Mallaney, C., Ostrander, E.L., Celik, H., Kramer, A.C., Martens, A., Kothari, A., Koh, W.K., Haussler, E., Iwamori, N., Gontarz, P., et al. (2019). Kdm6b regulates context-dependent hematopoietic stem cell self-renewal and leukemogenesis. *Leukemia* *33*, 2506–2521.
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. J.* *17*, 10–12.
- McGrath, K.E., Frame, J.M., Fegan, K.H., Bowen, J.R., Conway, S.J., Catherman, S.C., Kingsley, P.D., Koniski, A.D., and Palis, J. (2015). Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Rep.* *11*, 1892–1904.
- McLean, C.Y., Bristor, D., Hiller, M., Clarke, S.L., Schaar, B.T., Lowe, C.B., Wenger, A.M., and Bejerano, G. (2010). GREAT improves functional interpretation of cis-regulatory regions. *Nat. Biotechnol.* *28*, 495–501.
- Narayanan, R., Pirouz, M., Kerimoglu, C., Pham, L., Wagener, R.J., Kiszka, K.A., Rosenbusch, J., Seong, R.H., Kessel, M., Fischer, A., et al. (2015). Loss of BAF (mSWI/SNF) Complexes Causes Global Transcriptional and Chromatin State Changes in Forebrain Development. *Cell Rep.* *13*, 1842–1854.
- Palis, J. (2016). Hematopoietic stem cell-independent hematopoiesis: emergence of erythroid, megakaryocyte, and myeloid potential in the mammalian embryo. *FEBS Lett.* *590*, 3965–3974.
- Pietras, E.M., Mirantes-Barbeito, C., Fong, S., Loeffler, D., Kovtonyuk, L.V., Zhang, S., Lakshminarasimhan, R., Chin, C.P., Techner, J.M., Will, B., et al. (2016). Chronic interleukin-1 exposure drives haematopoietic stem cells towards precocious myeloid differentiation at the expense of self-renewal. *Nat. Cell Biol.* *18*, 607–618.
- Ramírez, F., Ryan, D.P., Grüning, B., Bhardwaj, V., Kilpert, F., Richter, A.S., Heyne, S., Dündar, F., and Manke, T. (2016). deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res.* *44* (W1), W160–5.
- Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R., et al. (2012). Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* *481*, 389–393.
- Sandoval, G.J., Pulice, J.L., Pakula, H., Schenone, M., Takeda, D.Y., Pop, M., Boulay, G., Williamson, K.E., McBride, M.J., Pan, J., et al. (2018). Binding of TMPRSS2-ERG to BAF Chromatin Remodeling Complexes Mediates Prostate Oncogenesis. *Mol. Cell* *71*, 554–566.e7.
- Schulz, C., Gomez Perdiguero, E., Chorro, L., Szabo-Rogers, H., Cagnard, N., Kierdorf, K., Prinz, M., Wu, B., Jacobsen, S.E., Pollard, J.W., et al. (2012). A lineage of myeloid cells independent of Myb and hematopoietic stem cells. *Science* *336*, 86–90.
- Steidl, U., Steidl, C., Ebralidze, A., Chapuy, B., Han, H.J., Will, B., Rosenbauer, F., Becker, A., Wagner, K., Koschmieder, S., et al. (2007). A distal single nucleotide polymorphism alters long-range regulation of the PU.1 gene in acute myeloid leukemia. *J. Clin. Invest.* *117*, 2611–2620.
- Takeuchi, J.K., and Bruneau, B.G. (2009). Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* *459*, 708–711.
- Tober, J., Koniski, A., McGrath, K.E., Vemishetti, R., Emerson, R., de Mesy-Bentley, K.K., Waugh, R., and Palis, J. (2007). The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood* *109*, 1433–1441.
- Wang, L., Zhao, Z., Meyer, M.B., Saha, S., Yu, M., Guo, A., Wisinski, K.B., Huang, W., Cai, W., Pike, J.W., et al. (2014). CARM1 methylates chromatin

remodeling factor BAF155 to enhance tumor progression and metastasis. *Cancer Cell* 25, 21–36.

Will, B., Vogler, T.O., Narayanagari, S., Bartholdy, B., Todorova, T.I., da Silva Ferreira, M., Chen, J., Yu, Y., Mayer, J., Barreyro, L., et al. (2015). Minimal PU.1 reduction induces a preleukemic state and promotes development of acute myeloid leukemia. *Nat. Med.* 21, 1172–1181.

Wu, J.I., Lessard, J., and Crabtree, G.R. (2009). Understanding the words of chromatin regulation. *Cell* 136, 200–206.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.

Zhao, H., and Choi, K. (2019). Single cell transcriptome dynamics from pluripotency to FLK1<sup>+</sup> mesoderm. *Development* 146, dev182097.

Zhou, X., Maricque, B., Xie, M., Li, D., Sundaram, V., Martin, E.A., Koebbe, B.C., Nielsen, C., Hirst, M., Farnham, P., et al. (2011). The Human Epigenome Browser at Washington University. *Nat. Methods* 8, 989–990.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Anti-mouse CDD45-BV421 (Clone 30-F11)	BioLegend	Cat#103134; RRID:AB_2562559
Anti-mouse CDD45-APC (Clone 30-F11)	BioLegend	Cat#103112; RRID:AB_312977
Anti-mouse TER119-PE (Clone TER-119)	BioLegend	Cat#116208; RRID:AB_313709
Anti-mouse TER119-APC-Cy7 (Clone TER-119)	BioLegend	Cat#116223; RRID:AB_2137788
Anti-mouse CD31-Biotin (Clone MEC13.3)	BioLegend	Cat#102504; RRID:AB_312911
Anti-mouse CD31-FITC (Clone MEC13.3)	BD Biosciences	Cat#553372; RRID:AB_394818
Anti-mouse/human CD11b-BV421 (Clone M1/70)	BioLegend	Cat#101235; RRID:AB_10897942
Anti-mouse/human CD11b-Biotin (Clone M1/70)	BioLegend	Cat#101204; RRID:AB_312787
Anti-mouse F4/80-PE (Clone BM8)	BioLegend	Cat#123110; RRID:AB_893486
Anti-mouse F4/80-APC (Clone BM8)	BioLegend	Cat#123116; RRID:AB_893481
Anti-mouse CX3CR1-PE (Clone SA011F11)	BioLegend	Cat#149006; RRID:AB_2564315
Anti-mouse CD117(c-Kit)-PE (Clone 2B8)	BioLegend	Cat#105807; RRID:AB_313216
Anti-mouse CD117(c-Kit)-APC (Clone 2B8)	BioLegend	Cat#105812; RRID:AB_313221
Anti-mouse CD41-APC (Clone MWRReg30)	BioLegend	Cat#133914; RRID:AB_11125581
Anti-mouse CD16/32-BV421 (Clone 93)	BioLegend	Cat#101331; RRID:AB_2562188
Anti-mouse CD16/32-PE (Clone 2.4G2)	BD Biosciences	Cat#553145; RRID:AB_394660
Purified anti-mouse CD16/32 (Clone 93)	BioLegend	Cat# 101302; RRID:AB_312801
BV421-Streptavidin	BioLegend	Cat#405225 (No RRID number available)
BV605-Streptavidin	BD Biosciences	Cat#563260 (No RRID number available)
SMARCC1/BAF155 (D7F8S) Rabbit mAb antibody	Cell Signaling Tech.	Cat#11956; RRID:AB_2797776
PU.1 (9G7) Rabbit mAb antibody	Cell Signaling Tech.	Cat# 2258; RRID:AB_2186909
p300 (D8Z4E) Rabbit mAb antibody	Cell Signaling Tech.	Cat# 86377; RRID:AB_2800077
Brg1 (D1Q7F) Rabbit mAb antibody	Cell Signaling Tech.	Cat# 49360; RRID:AB_2728743
UTX (D3Q1I) Rabbit mAb antibody	Cell Signaling Tech.	Cat# 33510; RRID:AB_2721244
HA-Tag(C29F4) Rabbit mAb antibody	Cell Signaling Tech.	Cat# 3724; RRID:AB_1549585
Rabbit IgG	Cell Signaling Tech.	Cat#2729; RRID:AB_1031062
Rabbit Anti-Histone H3, trimethyl (Lys27) Polyclonal antibody	Millipore	Cat# 07-449; RRID:AB_310624
Monoclonal ANTI-FLAG® M2 antibody	Sigma-Aldrich	Cat# F3165; RRID:AB_259529
EZview Red ANTI-FLAG® M2 Affinity Gel	Sigma-Aldrich	Cat# F2426; RRID:AB_2616449
EZview Red Anti-HA Affinity Gel	Sigma-Aldrich	Cat# E6779; RRID:AB_10109562
<b>Bacterial and Virus Strains</b>		
One Shot Stbl3 Chemically Competent E.Coli	Thermo Fisher Scientific	Cat#C737303
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
DMEM	GIBCO	Cat#11965092
FBS	Atlanta Biologicals	Cat#S12450
L-Glutamine	GIBCO	Cat#35050061
MEM Non-Essential Amino Acids	Corning	Cat#25-025-CI
MEM Sodium Pyruvate	Corning	Cat#25-000-CI
Geneticin	GIBCO	Cat#10131-035

(Continued on next page)

**Continued**

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Puromycin	Sigma	Cat#P8833
Blasticidin S Hydrochloride	Research Products International Corp	Cat#B12200
penicillin-streptomycin	GIBCO	Cat# 15140122
IMDM	GIBCO	Cat#2440046
Hexadimethrine bromide	Sigma	Cat#H9268
Interleukin-3 (IL-3) supernatant	This paper	N/A
Murine stem cell factor	PeprTech	Cat#250-03
M-CSF	PeprTech	Cat#315-02
GM-CSF	PeprTech	Cat#315-03
IL-6	PeprTech	Cat#216-16
IL-11	R&D Systems	Cat#418-ML
Erythropoietin (EPO)	PeprTech	Cat#100-64
Lipofectamine 3000	Thermo Fisher Scientific	Cat#L3000-001
esiBaf155	Sigma	Cat#EMU012611-50UG
esiEGFP	Sigma	Cat#EHUEGFP-50UG
MethoCult3434	Stem Cell Technologies	Cat#M3434
GSK-126	MedChem	Cat#HY-13470
SBE- $\beta$ -CD	MedChem	Cat# HY-17031
DSP (Dithiobis (succinimidyl propionate))	Thermo Fisher Scientific	Cat#22585
0.25% trypsin-EDTA	GIBCO	Cat#25200-056
Protease inhibitor cocktail	Sigma	Cat#11836170001
Tween 20	Sigma	Cat#P9416
Iodoacetamide	Sigma	Cat#I6125
Sepharose 4B	Sigma	Cat#4B200
LDS buffer	Invitrogen	Cat#NP0007
Beta - mercaptoethanol	Sigma	Cat#444203
ECL chemiluminescence substrate	Thermo Fisher Scientific	Cat#32106
Collagenase type IV	Worthington	Cat#LS004188
Deoxyribonuclease I	Worthington	Cat#LS002139
0.25% Collagenase	Stem Cell Technologies	Cat#07902
ProteinA-Sepharose® 4	Sigma-Aldrich	Cat#P9424
<b>Critical Commercial Assays</b>		
QuickTiter Lentivirus Associated HIV p24 Titer Kit	Cell Biolabs, INC	Cat#VPK-107
SimpleChIP® Plus Enzymatic Chromatin IP Kit	Cell Signaling Tech.	Cat#9005
RNeasy Micro/Mini Kit	QIAGEN	Cat#74004/74106
qScript cDNA SuperMix	Quanta	Cat#101414-106
DNA Clean and Concentrator 5	Zymo Research	Cat#D4014
AMPure XP beads	Beckman Coulter	N/A
<b>Deposited Data</b>		
ATAC-seq data	This paper	GEO: GSE144243
GRCm10/mm10	UCSC genome browser	<a href="http://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/">http://hgdownload.soe.ucsc.edu/goldenPath/mm10/bigZips/</a>
scRNA-seq data	This paper	GEO: GSE159381
scRNA-seq data of WT cells	<a href="#">Zhao and Choi, 2019</a>	GEO: GSE130146
<b>Experimental Models: Cell Lines</b>		
293FT cell line	ThermoFisher	Cat#R70007

(Continued on next page)

<b>Continued</b>		
REAGENT or RESOURCE	SOURCE	IDENTIFIER
Platinum-E (Plat-E) retroviral packaging cell line	Cell Biolabs, INC	Cat#RV-101
Mouse Embryonic Fibroblast (MEF)	Hansen, T.H. Washington University in St. Louis; <a href="#">Lybarger et al., 2003</a>	N/A
MEF-Flag-Baf155-IRES-GFP+Pu.1-IRES-mCherry	This paper	N/A
MEF-Flag-Baf155+HA-PU.1	This paper	N/A
Experimental Models: Organisms/Strains		
C57Bl6/J Wild Type	Jackson Laboratories	Stock No:000664
Tie2-Cre	Jackson Laboratories	Stock No:004128
<i>Baf155<sup>ff</sup></i>	Rho Hyun Seong; <a href="#">Choi et al., 2012</a>	N/A
Oligonucleotides		
See <a href="#">Tables S2</a> and <a href="#">S3</a> for a list of oligonucleotide sequences		N/A
Recombinant DNA		
pRRL_CAGpN-Flag-Baf155-IRES-GFP	Addgene	Cat#24561; RRID:Addgene_24561
CSII-EF-MCS-IRES2-bsr-PU.1-HA	This paper	N/A
psPAX2	Addgene	Cat#12260; RRID:Addgene_12260
pMD2.G	Addgene	Cat#12259; RRID:Addgene_12259
Pu.1-IRES-mCherry retroviral plasmid DNA	Addgene	Cat#80140; RRID:Addgene_80140
pLKO.1-puro-Ubc-TurboGFP	Sigma	Cat#SHC014; (No RRID number available)
Software and Algorithms		
FlowJo software version 10.5.3	TreeStar Inc.	<a href="https://www.flowjo.com">https://www.flowjo.com</a>
Graphpad Prism version 8.4.3	Graphpad Software, LLC.	<a href="https://www.graphpad.com/scientific-software/prism/">https://www.graphpad.com/scientific-software/prism/</a>
Cutadapt version 1.11	<a href="#">Martin, 2011</a>	<a href="https://github.com/marcelm/cutadapt/">https://github.com/marcelm/cutadapt/</a>
Bowtie 2 version 2.3.4.1	<a href="#">Langmead and Salzberg, 2012</a>	<a href="http://bowtie-bio.sourceforge.net/bowtie2/index.shtml">http://bowtie-bio.sourceforge.net/bowtie2/index.shtml</a>
MACS2 version 2.1.1	<a href="#">Zhang et al., 2008</a>	<a href="https://github.com/macs3-project/MACS">https://github.com/macs3-project/MACS</a>
WashU Epigenome Browser	<a href="#">Zhou et al., 2011</a>	<a href="https://epigenomegateway.wustl.edu/">https://epigenomegateway.wustl.edu/</a>
DiffBind version 2.6.6	<a href="#">Ross-Innes et al., 2012</a>	<a href="https://bioconductor.org/packages/release/bioc/html/DiffBind.html">https://bioconductor.org/packages/release/bioc/html/DiffBind.html</a>
deepTools	<a href="#">Ramírez et al., 2016</a>	<a href="https://github.com/deeptools/deepTools">https://github.com/deeptools/deepTools</a>
GREAT version 4.0.4	<a href="#">McLean et al., 2010</a>	<a href="http://great.stanford.edu/public/html/index.php">http://great.stanford.edu/public/html/index.php</a>
HOMER version 4.8	<a href="#">Heinz et al., 2010</a>	<a href="http://homer.ucsd.edu/homer/index.html">http://homer.ucsd.edu/homer/index.html</a>
Cell Ranger Single Cell Software Suite (v2.0.1)	Cell Ranger	<a href="https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger">https://support.10xgenomics.com/single-cell-gene-expression/software/pipelines/latest/what-is-cell-ranger</a>
Seurat version 2.3.4	Seurat	<a href="https://satijalab.org/seurat/">https://satijalab.org/seurat/</a>
Other		
Fisherbrand Model 120 Sonic Dismembrator	Fisher scientific	N/A

## RESOURCE AVAILABILITY

### Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the corresponding author and Lead Contact, Dr. Kyunghee Choi ([kchoi@wustl.edu](mailto:kchoi@wustl.edu)).

### Materials Availability

This study did not generate any unique reagents.

### Data and Code Availability

The ATAC-seq data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO Series accession number GSE144243 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE144243>).

The accession number for the scRNA-Seq data reported in this paper is GEO: GSE159381.

## EXPERIMENTAL MODEL AND SUBJECT DETAILS

### Mice

*Tie2-Cre;Baf155* CKO mice were obtained by first crossing *Tie2-Cre* (Stock No: 4128, Jackson Labs) males (2-3 months old) to *Baf155<sup>fl/fl</sup>* (Choi et al., 2012) females (2-3 months old) to generate *Tie2-cre; Baf155<sup>fl/+</sup>* mice. Next, timed matings using *Tie2-Cre; Baf155<sup>fl/+</sup>* males (2-8 months old) and *Baf155<sup>fl/fl</sup>* females (2-3 months old) were set up in the evening and females checked for vaginal plugs the following morning (12pm = E0.5). Females were separated from males and housed in the animal barrier until the desired time point. Females were euthanized using CO<sub>2</sub> asphyxiation and uteri removed for embryo collection. Embryos and collected tissue were kept on ice in PBS with 10% FBS until processed for analysis. Wild-type (WT) littermates were used as controls. Animal husbandry, generation, and handling were performed in accordance with protocols approved by the Institutional Animal Care and Use Committee of Washington University School of Medicine in St. Louis.

### Cell culture and transduction

The Mouse Embryonic Fibroblast (MEF) cell line has been previously described (Lybarger et al., 2003). MEF and MEF-derived stable cell lines were cultured in Iscove's Modified Dulbecco's Medium (IMDM) (2440046, GIBCO) supplemented with 10% (v/v) Fetal bovine serum (FBS) (S12450, Atlanta Biologicals), and 100U/ml penicillin-streptomycin (15140122, GIBCO). MEF cells were transduced with Flag-*Baf155*-IRES-GFP lentiviral and *Pu.1*-IRES-mCherry retroviral particles. Hexadimethrine bromide (8 μg/ml) (H9268, Sigma) was added during transductions to increase viral particle uptake. Transduced cells were sorted twice to ensure greater than 90% purity. When HA-*Pu.1* lentivirus was used, MEF-Flag-*Baf155* cells were transduced with HA-*Pu.1* lentiviral particles and selected with 1 μg/ml Blastidicin S Hydrochloride for 2 weeks. The overexpression efficiency of target proteins was confirmed by western blot.

## METHOD DETAILS

### Genotyping

The following primers were used to obtain genotype information for breeders and embryos:

*Baf155* - TGTCATCCATGAGGAGTGGTC3' (F); 5'GGTAGCTCACAAATGCCTGT3' (R); WT = 400 bp; Floxed = 450 bp. *Cre* - 5'ACCAGAGACGGAATCCATCG3' (F); 5'CCACGACCA AGTGACAGCAATG3' (R); *Cre* = 390 bp.

### Lentiviral and retroviral particle production

*Baf155* lentiviral plasmid DNA, pRR\_L\_CAGpN-Flag-*Baf155*-IRES-GFP, was a gift from Jerry Crabtree (Addgene, plasmid# 2456). MISSION pLKO.1-puro-Ubc-TurboGFP (Sigma, SHC014) was used as a transduction efficiency control. The *Pu.1* lentiviral plasmid was constructed by adding a HA tag at the *Pu.1* N-terminal and inserting it into the CSII-EF-MCS-IRES2-bsr lentiviral backbone. Lentiviral packaging plasmid psPAX2 (Addgene, plasmid# 12260) and VSV-G envelope expressing plasmid pMD2.G (Addgene, plasmid# 12259) were gifts from Didier Trono. For cloning purposes, viral plasmid DNA was transformed using One Shot Stbl3 Chemically Competent E.Coli (C737303, ThermoFisher). Lentiviral particles were produced using the 293FT cell line (R70007, ThermoFisher), which was maintained in high glucose Dulbecco's Modified Eagle Medium (DMEM) (11965092, GIBCO), 10% FBS, 200 mM L-Glutamine (35050061, GIBCO), 10 mM MEM Non-Essential Amino Acids (25-025-CI, Corning), 100 mM MEM Sodium Pyruvate (25-000-CI, Corning), and 500 μg/ml Geneticin (10131-035, GIBCO). Cells were transfected with lentiviral DNA using the calcium phosphate method. Sixteen hours after transfection, media was replaced and cells were incubated at 37°C in 5% CO<sub>2</sub> for an additional 48 hours. Virus titer was determined by QuickTiter Lentivirus Associated HIV p24 Titer Kit (Cell Biolabs, INC). *Pu.1*-IRES-mCherry retroviral plasmid DNA was a gift from Ellen Rothenberg (Addgene, plasmid# 80140). Platinum-E (Plat-E) retroviral packaging cell line was used to generate the *Pu.1* retrovirus using the calcium phosphate method. Cells were maintained in high glucose DMEM supplemented with 10% FBS, 1 μg/ml puromycin (P8833, Sigma), 10 μg/ml Blastidicin S Hydrochloride (B12200, Research Products International Corp), and 100 U/ml penicillin-streptomycin. Media was replaced the following morning and virus harvested 48 hours after transfection.

### esiRNA transfection

E10.5 WT yolk sac (YS) cKIT<sup>+</sup> cells (1x10<sup>4</sup>) were plated in 100 μL maturation media (IMDM, 20% FBS, 1% interleukin-3 (IL-3) supernatant, 10 ng/ml murine stem cell factor (PeproTech), 10 ng/ml M-CSF (PeproTech), 10 ng/ml GM-CSF (PeproTech), 10 ng/ml IL-6,

10 ng/ml IL-11 (R&D Systems) and 2 U/ml erythropoietin (PeproTech) in a 96-well plate and transfected with 300 ng esiRNA against either *Baf155* (Sigma, EMU012611) or *Egfp* (Sigma, EHUEGFP) with 2  $\mu$ L lipofectamine 3000 (ThermoFisher). Cells were cultured in a 37°C incubator with 5% CO<sub>2</sub> for 36-48 hours and then subjected to either RNA extraction or re-plating in methylcellulose (MethoCult 3434, Stem Cell Technologies).

### Hematopoietic progenitor assays

Methylcellulose colony-forming assays were performed using MethoCult 3434 (Stem Cell Technologies). E10.5 WT YS were pooled and sorted for either cKIT<sup>+</sup> or cKIT<sup>+</sup>CD41<sup>+</sup>CD16/32<sup>+</sup> populations. Sorted cells were mixed in methylcellulose (2,000/ml) and plated in triplicate using 35mm Petri-dishes. Cultures were maintained in a humidified incubator at 37°C, 5% CO<sub>2</sub>. CFU-E colonies were counted after 2-3 days of culture. Primitive erythroid, definitive erythroid (BFU-E), macrophage, and granulocyte/macrophage colonies were counted following 5-7 days of culture.

### In vivo GSK-126 treatment

GSK-126 (HY-13470, MedChem) was dissolved in SBE- $\beta$ -CD (HY-17031, MedChem) at a final concentration of 20 mg/ml. Pregnant mice (E8) were injected intraperitoneally with equal volumes of either SBE- $\beta$ -CD (vehicle) or GSK126 (100mg/kg). At E9.5, pregnant females were euthanized using CO<sub>2</sub> asphyxiation and the uteri removed for embryo collection and dissection.

### Nuclear extract preparation

Transduced MEF cells expressing *Baf155* and *Pu.1* were treated with DSP (Dithiobis [succinimidyl propionate], Thermo Scientific), a membrane permeable cleavable crosslinker, before subjected to nuclear extraction. Cells were detached using 0.25% trypsin-EDTA, washed twice with phosphate-buffered saline (PBS), resuspended in PBS containing 1mM DSP at approximately 1 $\times$ 10<sup>7</sup>/ml, and incubated at room temperature (RT) for 30 min. The crosslink reaction was stopped by adding 1M Tris-HCl (pH7.5) at a final concentration of 20mM for 10min. DSP treated cells after wash were then incubated in hypotonic buffer (25mM HEPES (pH7.6), 25mM KCl, 5mM MgCl<sub>2</sub>, 0.05mM EDTA, 0.1% NP40, 5% Glycerol, 1mM PMSF) on ice for 10min, vigorously vortexed for 10 s, and centrifuged at 13,000 xg for 1 min. Supernatant, primarily containing soluble cytosolic protein, was collected for immunoblot while the pellet containing nuclei was resuspended in nuclear extraction buffer (10mM HEPES (pH7.6), 100mM KCl, 3mM MgCl<sub>2</sub>, 0.1mM EDTA, 5% Glycerol supplemented with 1mM PMSF and 1X protease inhibitor cocktail (complete mini Roche, 11836170001, Sigma)), incubated on ice for 15min, and followed by 3 $\times$ 10s of sonication at 50% amplitude. Insoluble proteins and debris were removed from the nuclear extract by high-speed centrifugation (10 min at 18,000 g).

### Immunoprecipitation

Nuclear extract was mixed 1:1 with IP buffer (10 mM Tris-HCl (pH 7.5), 200 mM NaCl, 1 mM EDTA, 0.2% Tween 20, 1X protease inhibitor cocktail, 20mM Iodoacetamide) and precleared by Sepharose 4B (4B200, Sigma). Immunoprecipitation was performed by incubating precleared nuclear extract with antibody-bound beads (anti-Flag M2 affinity gel, anti-HA (clone HA-7) affinity gel, or Protein A as a negative control (Sigma)) at 4°C overnight. Precipitates were washed 4 times with washing buffer (10 mM Tris-HCl (pH 7.5), 150 mM NaCl, 0.1% Tween 20, 20 mM Iodoacetamide) and eluted in 1X LDS buffer (Invitrogen). DSP was cleaved by adding 5% beta- mercaptoethanol in LDS loading buffer at 100°C for 10 min.

### Western blotting

Western blotting was conducted following standard protocols. Primary antibodies used for western blotting are listed in key resources table. Secondary antibodies were horseradish peroxidase-conjugated mouse anti-rabbit IgG light chain and mouse IgG kappa chain binding protein (Santa Cruz Biotechnology, 1:10,000 dilution). Membranes were developed with ECL chemiluminescence substrate (ThermoFisher) and visualized using photographic film.

### Chromatin immunoprecipitation (ChIP) coupled with quantitative real-time PCR (ChIP-qPCR)

Chromatin immunoprecipitation (ChIP) was performed according to the manufacture's protocol (ChIP kit, 9005, Cell Signaling Technologies) with the following modification: 25 mg cross-linked YS tissue was used per preparation. After cell lysis, nuclei extracts were digested by adding 0.5  $\mu$ L Micrococcal Nuclease per IP prep and incubating for 20 min at 37°C with frequent mixing to digest DNA to a size of approximately 150-900 bp. Digestion was stopped by adding 10  $\mu$ L 0.5 M EDTA and samples placed on ice for 2 min. Nuclei was pelleted and resuspend in 100  $\mu$ L ChIP buffer. Nuclear lysates were further subjected to sonication to break nuclear membrane using a 120 Sonic Dismembrator (Fisher Scientific) at 4°C for 3 cycles, cycling ON for 10 s and OFF for 30 s at 50% amplitude. Approximately 10  $\mu$ g of digested, cross-linked chromatin and 5  $\mu$ g of antibody (BAF155, Cell Signaling Technologies, 11956; PU.1, Cell Signaling Technologies, 2258; H3K27me3, Millipore,07-449) were used per immunoprecipitation. IP samples were incubated overnight at 4°C with rotation, followed by 30  $\mu$ L of protein G Magnetic Beads per IP reaction, and incubated for an additional 2 hr at 4°C with rotation. After elution of chromatin from the antibody/protein G magnetic beads, reverse cross-link performed by adding 6  $\mu$ L 5 M NaCl and 2  $\mu$ L proteinase K per IP, and incubating for 6 hr at 65°C. Immunoprecipitated DNA fragments were isolated using spin

columns provided by the kit and subjected to qPCR with appropriate primers indicated in Table S2. Rabbit IgG (Cell Signaling Technologies, 2729) was used as a negative control. Quantitative PCR was performed in triplicate from 3 independent experiments, and data were normalized to input values.

### Tissue processing for flow cytometry

YS were collected between E8.25–10.5, and brain rudiments collected between E9.5–10.5. To obtain single-cell suspension, tissues were incubated in Hank's balanced salt solution (HBSS) containing 0.2 mg/ml collagenase type IV (Worthington), 100 U/ml deoxyribonuclease I (Worthington) and 5% FBS at 37°C for 1 h with tubes inverted every 5 to 10 min. Tissues were further dissociated by gently passing through a 20G needle 5 to 10 times. Cells were pelleted and resuspend in 0.3–1 mL IMDM media with 10% FBS. Cells were then passed through a 70  $\mu$ m cell strainer and counted for viability.

### Flow cytometry and cell sorting

Single cell suspensions were centrifuged at 400 g for 5 min, resuspended in 200  $\mu$ L staining buffer (1X PBS, 1% BAS, 2 mM EDTA), placed in 5ml round-bottom tubes, and immunolabeled for FACS analysis. Before immunostaining, cell suspensions were pre-incubated with diluted (1:50) purified anti-CD16/32 (clone: 93, Biolegend, 101302) for 10 min on ice to block non-specific binding to Fc-receptors. Next, antibodies were added and incubated for 40 min on ice. Where appropriate, cells were further incubated with streptavidin conjugates for 20 min. All antibodies used can be found in key resources table. All FACS analyses were carried out on LSR Fortessa or Fortessa X-20 (BD Biosciences). Cell sorting was performed on FACS Aria II (BD Biosciences) sorter using 85  $\mu$ m nozzle. All data were analyzed using FlowJo10 software (Tree Star).

### Quantitative real-time reverse transcription PCR (qRT-PCR)

Total RNA from YS was prepared with RNeasy Micro/Mini Kit (QIAGEN), and reverse-transcribed into cDNA with qScript cDNA SuperMix (101414-106, Quanta) according to the manufacturer's protocol. Gene expression was measured by quantitative real-time PCR with primers indicated in Table S3. Gene expression levels were normalized to  $\beta$ -actin.

### Single-cell RNA sequencing

An equal number of E9.5 and E10.5 WT YS were combined and dissociated with 0.25% collagenase at 37°C for 30 minutes. Cells were briefly stored at  $-80^{\circ}\text{C}$  in 90% FBS and 10% DMSO. Cells were thawed, washed with PBS, and stained with TER-119 antibody. Dead cells and TER-119+ cells were excluded by sorting to enrich live non-erythroid cells. Single cell suspension at 300 cells/ $\mu$ L in PBS were subjected to Chromium 10x Genomics library construction and HiSeq2500 sequencing (The Genome Technology Access Center, Washington University in St. Louis).

### ATAC-seq library generation, sequencing, and mapping

For ATAC-seq library generation, approximately 50,000 cKIT<sup>+</sup> cells were isolated from WT and *Baf155* CKO YS using FACS sorter as described above. ATAC-seq libraries were generated following the Omni-ATAC protocol (Corces et al., 2017) with the following modification: Cells were harvested by centrifuging at 500 g for 5 minutes at 4°C. Supernatant was carefully aspirated and cells were washed once with cold PBS. Cell pellets were lysed in 100  $\mu$ L of ATAC-seq RSB (10 mM Tris pH 7.4, 10 mM NaCl, 3 mM MgCl<sub>2</sub>) containing 0.1% NP40, 0.1% Tween-20, and 0.01% Digitonin by pipetting up and down and incubating on ice for 3 minutes. Next, 1 mL of ATAC-seq RSB containing 0.1% Tween-20 was added and mixed with the lysis reaction. Nuclei were pelleted by centrifuging at 800 g for 5 minutes at 4°C. Supernatant was carefully removed, and nuclear pellets were resuspended in 20  $\mu$ L 2x TD buffer (20 mM Tris pH 7.6, 10 mM MgCl<sub>2</sub>, 20% Dimethyl Formamide). Nuclei were counted using trypan blue. Approximately 50,000 nuclei were transferred to 25  $\mu$ L of 2x TD buffer. 25  $\mu$ L of transposition mix (2.5  $\mu$ L transposase (100 nM final), 16.5  $\mu$ L PBS, 0.5  $\mu$ L 1% digitonin, 0.5  $\mu$ L 10% Tween-20, and 5  $\mu$ L H<sub>2</sub>O) was then added to the nuclei. Transposition reactions were mixed and incubated at 37°C for 30 min gently tapping every 10 min to mix. Reactions were cleaned up with Zymo DNA Clean and Concentrator 5 columns. ATAC-seq library was prepared by amplifying the DNA for 9 cycles on a thermal cycler. The PCR reaction was purified with AMPure XP beads using double size selection following the manufacture's protocol, in which 27.5  $\mu$ L beads (0.55x sample volume) and 50  $\mu$ L beads (1.5x sample volume) were used based on 50  $\mu$ L PCR reaction. ATAC-seq libraries were quantitated by Qubit assays. Paired-end ATAC-seq libraries were sequenced on an Illumina NextSeq 500 machine. The reads were de-multiplexed by using sample-specific index sequences. Nextera adaptor sequences were trimmed by using cutadapt (Martin, 2011) version 1.11. The trimmed reads were mapped to the mouse genome sequence by using bowtie2 (Langmead and Salzberg, 2012) version 2.3.4.1 with the following parameters:  $-\text{local} -\text{k} 4 -\text{X} 2000 -\text{mm}$ . Secondary alignment, multiply mapped reads, and PCR duplicated reads were removed from the total aligned reads.

## QUANTIFICATION AND STATISTICAL ANALYSIS

### scRNA-seq bioinformatics analyses

The sequenced reads were mapped to the GRCm38 assembly using Cell Range 2.0.1 (10x Genomics). Sample demultiplexing, barcode processing, and single-cell 3' counting was performed using the Cell Ranger Single-Cell Software Suite (10x Genomics).

Cellranger count was used to align samples to the mm10 reference genome, quantify reads, and filter reads with a quality score below 30. The resultant files were input into Seurat for normalization across all samples and merging. The Seurat package in R was used for subsequent analysis (Butler et al., 2018). Cells with Mitochondrial content greater than 5 percent were removed for downstream analysis. Data were normalized using a scaling factor of 10,000, and nUMI was regressed with a negative binomial model. Principal component analysis was performed using the top 3000 most variable genes and t-SNE analysis was performed with the top 15 PCAs. Clustering was performed using the FindClusters function which works on K-nearest neighbor (KNN) graph model with the granularity ranging from 0.1-0.9 and selected 0.6 for the downstream clustering. For identifying the markers for each cluster, we performed differential expression of each cluster against all other clusters identifying negative and positive markers for that cluster

### Identification of ATAC peaks

Filtered aligned ATAC-seq reads were used to map to the transposon insertion sites, and ATAC peaks were called from those insertion sites. First, ATAC-seq reads mapped to mitochondrial DNA were removed from the aligned reads. Both ends of the paired-end reads were then treated as two Tn5 insertion sites. Tn5 insertion sites were adjusted to reflect the actual binding center of transposons as follows. All reads mapped to the + strand were offset by +4 bp, and all reads mapped to the – strand were offset by –5 bp. The ATAC peaks were identified from these insertion sites by using the MACS2 (Zhang et al., 2008) version 2.1.1 callpeak function with the following parameters: -g mm --keep-dup all -B --SPMR --nomodel --extsize 73 --shift -37 -p 0.01 --call-summits. The ATAC-seq signals were visualized on the WashU Epigenome Browser (Zhou et al., 2011) as fold change over background using bedGraph tracks generated by using the MACS2 bdgcmp function with the following parameter: -m FE.

### Identification and analyses of DARs

To identify DARs, DiffBind (Ross-Innes et al., 2012) version 2.6.6 was used on the union set of ATAC peaks with the following parameters: minOverlap = 1, fragmentSize = 1, summits = 0. DARs were defined as the ATAC peaks with fold change > 2 and *P*-value < 0.05. Unaffected accessible regions were defined as the ATAC peaks that are present both wild-type and *Baf155* CKO cells and that are also with fold change < 1.1 and *P*-value > 0.05 from DiffBind. Heatmaps of ATAC-seq signal levels of DARs along with their neighboring regions were plotted by using deepTools (Ramírez et al., 2016). Gene Ontology enrichment analysis on DARs and unaffected regions were performed using GREAT (McLean et al., 2010) version 4.0.4. Motif enrichment analysis on DARs was performed using HOMER (Heinz et al., 2010) version 4.8. HOMER scanned the sequences of DARs for known motifs, and calculated enrichment score *P*-values using a binomial test. The heatmap of the selected known motifs were plotted using fold enrichment against the background. HOMER also discovered *de novo* motifs with their best matches to a known motif in DARs.

### Statistics

GraphPad Prism 8 software was used for performing statistical analysis and generating graphs/plots. Data are presented as mean with standard deviation for all the measurements. All experimental data were reliably reproduced in two or more individual biological replicates. Details of the statistical tests performed are given in the respective figure legends. *p* < 0.05 was considered statistically significant.