

Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription

Julius Judd,¹ Fabiana M. Duarte,^{1,2} and John T. Lis¹

¹Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14853, USA; ²Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA

Transcriptionally silent genes must be activated throughout development. This requires nucleosomes be removed from promoters and enhancers to allow transcription factor (TF) binding and recruitment of coactivators and RNA polymerase II (Pol II). Specialized pioneer TFs bind nucleosome-wrapped DNA to perform this chromatin opening by mechanisms that remain incompletely understood. Here, we show that GAGA factor (GAF), a *Drosophila* pioneer-like factor, functions with both SWI/SNF and ISWI family chromatin remodelers to allow recruitment of Pol II and entry to a promoter-proximal paused state, and also to promote Pol II's transition to productive elongation. We found that GAF interacts with PBAP (SWI/SNF) to open chromatin and allow Pol II to be recruited. Importantly, this activity is not dependent on NURF as previously proposed; however, GAF also synergizes with NURF downstream from this process to ensure efficient Pol II pause release and transition to productive elongation, apparently through its role in precisely positioning the +1 nucleosome. These results demonstrate how a single sequence-specific pioneer TF can synergize with remodelers to activate sets of genes. Furthermore, this behavior of remodelers is consistent with findings in yeast and mice, and likely represents general, conserved mechanisms found throughout eukarya.

[*Keywords:* pioneer factors; chromatin remodelers; transcription regulation; promoter-proximal pausing]

Supplemental material is available for this article.

Received June 21, 2020; revised version accepted October 20, 2020.

Pioneer transcription factors are a class of transcription factors that can bind and open condensed chromatin. They control cell-fate decisions in development by opening chromatin at previously inactive lineage-specific promoters and enhancers via sequence-specific binding (Zaret and Mango 2016; Mayran and Drouin 2018; Vallot and Tachibana 2020). These factors possess the unique ability to bind nucleosome-wrapped DNA, but the question of how they evict nucleosomes and initiate transcription remains open.

From yeast to mammals, there is growing evidence that pioneer factors cooperate with multiple ATP-dependent nucleosome remodeling complexes to establish transcription-permissive chromatin architecture (Kubik et al. 2017). In yeast, the pioneer factor Abf1 synergizes with the RSC complex (SWI/SNF family) to maintain the nucleosome-free region (NFR) of Abf1-bound promoters, while ISW1a and ISW2 are required to properly position the +1 nucleosome and phase downstream nucleosomes (Krietenstein et al. 2016). In mouse embryonic stem cells, the pioneer factors OCT4 and NANOG are codependent

on BAF complex (SWI/SNF family) subunit BRG1 to bind and open chromatin at target sites (King and Klose 2017; Hainer et al. 2019). Recent structural studies have illuminated how SWI/SNF family remodelers bidirectionally evict nucleosomes from promoter NFRs in yeast (Wagner et al. 2020) and mammals (He et al. 2020).

GAGA factor (GAF) is a *Drosophila* transcription factor encoded by the *Trithorax-like* (*trl*) gene (Farkas et al. 1994) that preferentially binds GAGAG repeats, but is capable of binding a single GAG trinucleotide (Wilkins and Lis 1998). We have previously demonstrated that, in *Drosophila* cell cultures, GAF is essential for establishing paused Pol II on GAF-bound promoters, and that the NFRs of these promoters fill with nucleosomes upon GAF depletion (Fuda et al. 2015). Without this activity, the response of a subset of heat-shock genes is impaired (Duarte et al. 2016). In early fly embryos, regions with chromatin signatures similar to those at binding sites of the embryonic pioneer factor Zelda—but lacking Zelda

Corresponding author: johnlis@cornell.edu

Article published online ahead of print. Article and publication date are online at <http://www.genesdev.org/cgi/doi/10.1101/gad.341768.120>.

© 2021 Judd et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see <http://genesdev.cshlp.org/site/misc/terms.xhtml>). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at <http://creativecommons.org/licenses/by-nc/4.0/>.

binding—are enriched for GAF binding, suggesting that GAF may also be an additional early embryonic pioneer factor (Moshe and Kaplan 2017). GAF interacts physically with the NURF complex (nucleosome remodeling factor) and both GAF and NURF are required to remodel nucleosomes on the *hsp70* promoter in vitro (Tsukiyama and Wu 1995; Xiao et al. 2001). We have previously speculated that GAF recruits NURF to target promoters, which clears them of nucleosomes and allows Pol II initiation and subsequent pausing to proceed (Vihervaara et al. 2018). However, early studies speculated that GAF can also interact with Brahma (Brm) complexes (SWI/SNF family; BAP/PBAP) (Tsukiyama and Wu 1995), and recent evidence indicates that GAF physically interacts with PBAP (polybromo-associated Brm) but not BAP (Nakayama et al. 2012; Lomaev et al. 2017) in addition to NURF.

Not all GAF-bound promoters have pausing that is affected by GAF depletion, a phenomenon that we and others have speculated could be a result of the binding of other transcription factors (including M1BP and BEAF-32), which could establish paused Pol II independent of GAF (Li and Gilmour 2013; Fuda et al. 2015); however, it is unclear whether these factors act redundantly with GAF to establish open chromatin and promote Pol II recruitment or whether they simply serve to insulate the promoter from the effects of GAF and thus allow other TFs to orchestrate transcription.

To test which of these remodelers is responsible for GAF's ability to establish transcription-permissive chromatin architecture at target genes, we depleted GAF, NURF301, and BAP170 (unique subunits of the NURF and PBAP complexes that are essential for complex functionality)—as well as NURF301 and BAP170 simultaneously—in S2 cells using RNAi (Fig. 1A). After confirming knockdown efficiency (Supplemental Fig. S1), we used a combination of PRO-seq, ATAC-seq, ChIP-seq, and 3' RNA-seq to monitor changes in nascent transcription, chromatin state, GAF binding, and mRNA output.

Results

GAF synergizes with PBAP to open chromatin

We used a spike-in normalization strategy for PRO-seq and 3' RNA-seq (see Materials and Methods) to ensure the detection of widespread transcriptional changes that can be hidden by centralizing normalization strategies such as RPKM (Mahat et al. 2016). A principal component analysis of all genome-wide data sets revealed that GAF knockdown predominantly clusters with PBAP knockdown (Fig. 1B; Supplemental Fig. S2). After confirming data quality (Supplemental Figs. S2–S4), we defined a set of promoters that have down-regulated pause region PRO-seq signal upon GAF knockdown (Fig. 1D). Notably, the number of genes with GAF-dependent pausing was far greater than previously reported because our spike-in normalization scheme allowed us to examine the genome-wide effects of GAF depletion with unprecedented sensitivity ($n = 685$ in this study; $n = 140$ reported previously (Fuda et al. 2015)). ATAC-seq hypersensitivity signal (frag-

ments < 120 bp) (see the Materials and Methods) revealed that these promoters are substantially less accessible upon GAF, PBAP, or NURF + PBAP knockdown (Fig. 1C, E; Supplemental Fig. S6), and PRO-seq shows that pausing is severely reduced upon GAF, PBAP, or NURF + PBAP knockdown, but not after NURF knockdown (Fig. 1C, F, G; Supplemental Fig. S6). These results clearly demonstrate that GAF coordinates with PBAP—not NURF as previously proposed—to regulate Pol II recruitment by evicting nucleosomes from the NFRs of target promoters. To our knowledge, this is the first report of a pioneer-like factor synergizing specifically with PBAP (or PBAF, the homologous mammalian complex) to maintain accessible target promoters in metazoans.

In contrast to PBAP, NURF knockdown increases PRO-seq signal in the pause region and in the gene body region compared with the LACZ-RNAi control at GAF-dependent promoters ($n = 685$), particularly in the early pause region closer to the TSS (Fig. 1G, left panel). We then compared the changes in pause region PRO-seq signal upon GAF knockdown with that observed after PBAP and NURF knockdown on a gene-by-gene basis. This revealed a near-perfect one-to-one correlation between GAF and PBAP knockdowns, but minor anticorrelation between GAF and NURF knockdowns (Fig. 1H, cf. left panel and right panel; Supplemental Fig. S7D for the NURF + PBAP knockdown). When we examined promoters with PBAP-dependent pausing ($n = 806$) (Supplemental Fig. S5A), we observed trends similar to those seen at GAF-dependent promoters: decreased pausing and promoter accessibility after GAF, PBAP, and NURF + PBAP knockdown, and increased pausing and narrowed promoter accessibility upon NURF knockdown (cf. Supplemental Fig. S7A–C and Fig. 1E–G). Taken together, these data indicate that PBAP and GAF act together to free the promoter of nucleosomes, while NURF acts at a downstream step.

Is GAF's mechanistic role to bind nucleosome-bound DNA and recruit the PBAP remodeling complex where they act synergistically to remove nucleosomes, or does GAF binding have an intrinsic ability to displace nucleosomes? The striking loss of PRO-seq signal and loss of chromatin openness of promoters (Fig. 1C, E, F) described when either factor is depleted argues for a highly synergistic model, where GAF alone has little intrinsic chromatin opening activity. To investigate further, we compared ATAC-seq signal between the GAF and PBAP knockdown conditions, which revealed significant low-magnitude changes at only a small number of sites (Supplemental Fig. S8A), indicating that GAF does not possess sufficient intrinsic chromatin opening ability to account for the effects of GAF knockdown on chromatin. In further support of this, 88% of promoters with decreased pausing upon GAF knockdown had decreased pausing upon PBAP knockdown ($n = 603$) (Supplemental Fig. S8B). In order to test whether GAF is capable of binding target loci prior to synergizing with PBAP to open chromatin, we profiled the genome-wide binding pattern of GAF in the LACZ-RNAi control as well as after GAF or PBAP depletion. First, this revealed strong GAF occupancy of promoters with GAF dependent pausing (Fig. 1C, I; Supplemental

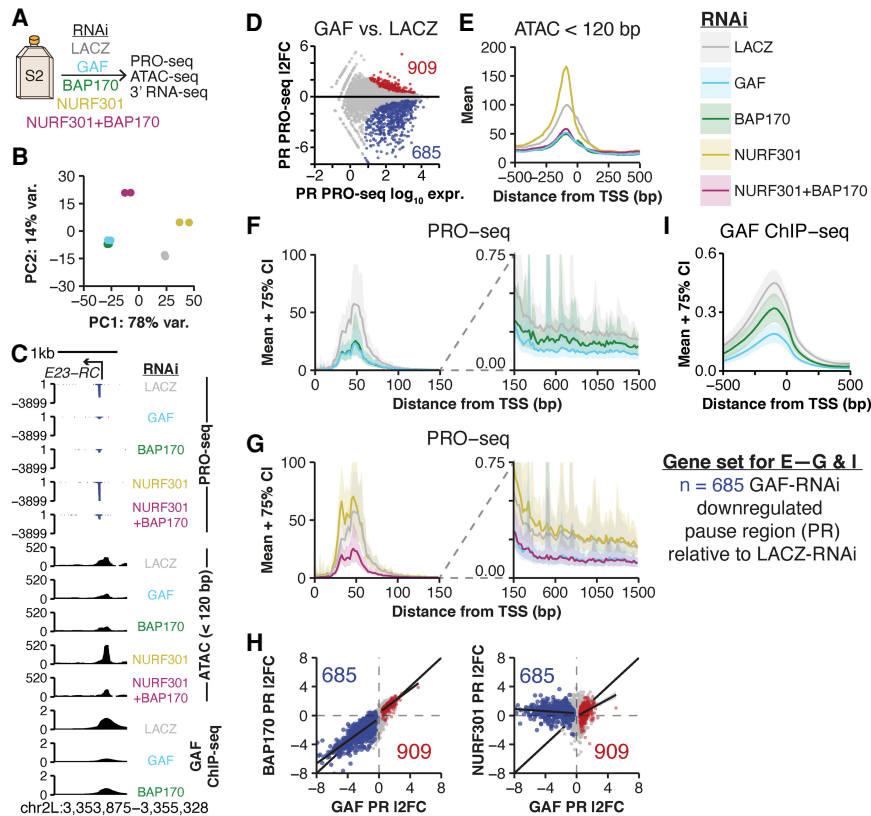


Figure 1. GAGA factor opens chromatin via the PBAP complex. (A) Experimental design. (B) Principal component analysis of spike-in normalized PRO-seq signal in the pause region (TSS -50 to $+100$). (C) Browser shot of *E23-RC*. (D) MA plot for the comparison of spike-in normalized GAF-RNAi versus LACZ-RNAi PRO-seq in the pause region (TSS -50 to $+100$); DESeq2 FDR < 0.01 . (12FC) \log_2 fold change. Number of genes significantly up-regulated or down-regulated is also shown, and we focus on down-regulated genes in subsequent panels as the magnitude of changes is greater and they have properties consistent with being primary targets of GAF. (E) ATAC-seq (< 120 bp) signal in 1-bp bins at promoters with GAF-RNAi down-regulated pausing ($n = 685$; see D; DESeq2 FDR < 0.01). Signal is the mean of 1000 subsamplings of 10% of regions. Confidence interval was calculated but omitted to avoid overplotting. (F) PRO-seq signal for the LACZ, GAF, and BAP170 RNAi treatments. The pause region (left) is in 2-bp bins, and the gene body (right) is in 20-bp bins. Data is shown as mean (line) \pm 75% confidence interval (shaded) from 1000 subsamplings of 10% of regions. Gene set as in E. (G) As in F, but for LACZ, NURF301, and NURF301 + BAP170 RNAi treatments. GAF-RNAi is also shown in the gene body region for comparison (blue line), although it is partially obscured by the NURF + PBAP line (purple) due to similarity of the trace. (H) Pause region (TSS -50 to $+100$) PRO-seq \log_2 fold change (I2FC) versus the LACZ-RNAi control; GAF-RNAi compared with BAP170-RNAi (left) or NURF301-RNAi (right). Red/blue points are significantly changed by GAF-RNAi (DESeq2 FDR < 0.01). Also shown are a GLM and 95% confidence interval for up-regulated and down-regulated promoters. (I) GAF ChIP-seq signal for the LACZ, GAF, and BAP170 RNAi treatments in 10-bp bins. Data is shown as mean (line) \pm 75% confidence interval (shaded) from 1000 subsamplings of 10% of regions. Gene set as in E.

comparison (blue line), although it is partially obscured by the NURF + PBAP line (purple) due to similarity of the trace. (H) Pause region (TSS -50 to $+100$) PRO-seq \log_2 fold change (I2FC) versus the LACZ-RNAi control; GAF-RNAi compared with BAP170-RNAi (left) or NURF301-RNAi (right). Red/blue points are significantly changed by GAF-RNAi (DESeq2 FDR < 0.01). Also shown are a GLM and 95% confidence interval for up-regulated and down-regulated promoters. (I) GAF ChIP-seq signal for the LACZ, GAF, and BAP170 RNAi treatments in 10-bp bins. Data is shown as mean (line) \pm 75% confidence interval (shaded) from 1000 subsamplings of 10% of regions. Gene set as in E.

Fig. S6), supporting the conclusion that the loss of GAF-dependent pausing and chromatin accessibility are a direct effects of GAF depletion and not secondary effects of long-term GAF depletion. When GAF is depleted, GAF ChIP-seq signal is predictably diminished at these loci, but when PBAP is severely depleted, GAF binding is only partially reduced (Fig. 1C,I; Supplemental Fig. S6). These results demonstrate that GAF still binds to target loci in cells depleted of PBAP, but these interactions are either weaker or more transient. Notably the effects of the depletion of both of these factors on chromatin accessibility and pausing are virtually indistinguishable.

Small sets of promoters show only GAF or only PBAP knockdown effects. We found that GAF-specific promoters ($n = 82$; PBAP knockdown causes no change) had higher levels of GAF ChIP-seq signal and were far less sensitive to the GAF knockdown than the class of genes dependent on both GAF and PBAP (Supplemental Fig. S8C-E). We speculate that these promoters may be held open by paused Pol II (Gilchrist et al. 2010) that is generated by mechanisms independent of PBAP, or the level of PBAP remaining after knockdown was sufficient to be recruited by the high level of GAF bound at these promoters. PBAP-

specific promoters are mostly not bound by GAF (Supplemental Fig. S8C), and often contain the binding motif for the transcription factor *lola* (Supplemental Fig. S8F), which might function like GAF in its collaboration with PBAP.

M1BP can establish paused Pol II independent of GAF

Not all GAF bound promoters have GAF-dependent pausing, and some of these bound but unaffected promoters are bound by M1BP (motif 1-binding protein) and the insulator BEAF-32 (boundary element-associated factor) (Li and Gilmour 2013; Fuda et al. 2015). However, it remains unclear whether M1BP acts redundantly with GAF to open promoters and promote pausing, or whether M1BP/BEAF-32 simply insulate promoters from GAF's activity. To investigate this, we divided GAF-bound promoters into two classes based on whether they have GAF-dependent pausing ($n = 600$) or not ($n = 1245$), and found that the BEAF-32 and M1BP motifs (Yang et al. 2012; Li and Gilmour 2013) were overrepresented in GAF-bound promoters with unchanged pausing (Fig. 2A). We then subdivided the class of GAF-bound, GAF-independent

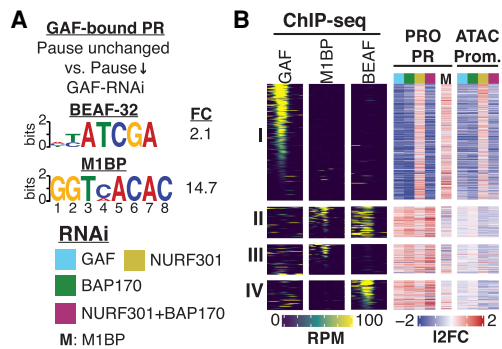


Figure 2. M1BP can establish paused Pol II independent of GAF. (A) Motifs enriched in GAF-bound promoters (GAF ChIP-seq peak within -500 to TSS) with GAF-independent pausing ($n = 1245$) over GAF-bound promoters with GAF-dependent pausing ($n = 600$). (FC) Fold change. DREME E -value < 0.001 . (B) GAF, M1BP, and BEAF-32 ChIP-seq signal in 10-bp bins in the promoter region (left; TSS ± 500), pause region (TSS -50 to $+100$) PRO-seq log₂ fold change (middle), and promoter (-250 to TSS) ATAC-seq (< 120 bp) log₂ fold change (right) at all GAF-bound genes. (Row I) GAF-dependent pausing ($n = 600$). (Rows II–IV) GAF-independent pausing ($n = 1,245$). (Row II) M1BP and BEAF-32 bound ($n = 159$). (Row III) M1BP only ($n = 152$). (Row IV) BEAF-32 only ($n = 152$). Sort order: (Row I) GAF ChIP-seq. (Rows II, III) M1BP ChIP-seq. (Row IV) BEAF-32 ChIP-seq. GAF-bound GAF-independent genes without M1BP or BEAF-32 ChIP-seq signal are not shown ($n = 782$).

pausing genes by whether they were bound by M1BP ($n = 152$) (Li and Gilmour 2013), BEAF-32 ($n = 152$) (Liang et al. 2014), or both ($n = 159$) (Fig. 2B). GAF-binding was weaker and more diffuse in GAF-bound genes with GAF-independent pausing (classes II–IV), while these promoters were directly and strongly bound by either M1BP or BEAF-32 or both (classes II–IV). We know from our previous study that genes bound by M1BP have reduced pause region PRO-seq signal upon M1BP knockdown (Duarte et al. 2016), and this reduction in pausing correlates with M1BP-binding intensity (Fig. 2B). Moreover, all classes of GAF-bound, GAF-independent pause genes had relatively unchanged ATAC-seq hypersensitivity signal in promoters after GAF or BAP knockdown (Fig. 2B, left). This demonstrates that M1BP can create paused Pol II independent of GAF and that it does not simply act redundantly to recruit PBAP, and the weak and diffuse GAF binding at these sites is insufficient to complement depletion of M1BP.

NURF promotes transition to productive elongation

GAF can physically interact with the remodelers PBAP and NURF (Xiao et al. 2001; Nakayama et al. 2012; Lomaev et al. 2017) and appears to function with each remodeler at distinct steps in transcription: GAF and PBAP open chromatin allowing Pol II initiation and entry to the promoter-proximal pause site; while GAF and NURF ensure efficient transition to productive elongation. This role of GAF and PBAP in the first of these two

steps is supported strongly by results described above (Fig. 1). Evidence that NURF's role is downstream from PBAP is provided by the observation that the PBAP + NURF double knockdown primarily mimics PBAP depletion in terms of changes in ATAC-seq and PRO-seq patterns in the pause region (Fig. 1). Support for NURF's role in productive elongation comes in part from the fact that the PBAP knockdown only partially recapitulates the decrease in gene body polymerase density seen after GAF depletion (Fig. 1F, right panel). In contrast, the NURF + PBAP double knockdown mirrors the GAF knockdown (Fig. 1G, right panel). However, only 222/685 GAF-regulated promoters show a statistically significant increase in pause signal upon NURF depletion (in contrast to 603/685 GAF-regulated promoters that show decreased pause signal upon PBAP depletion) (Supplemental Fig. S8B), indicating the interdependence of GAF and NURF is likely to be more limited than that observed for GAF and PBAP. CUT&RUN assays demonstrate co-occupancy of GAF and NURF at promoters genome-wide (Supplemental Fig. S9A), showing that GAF and NURF are likely to act together. These results support the model that GAF coordinates with both remodelers to ensure efficient transcription by first acting with PBAP to open chromatin and allow for the formation of promoter-proximal paused Pol II, and then with NURF at a subset of GAF target promoters to establish chromatin structure at the start of genes, which ensures proper transition to productive elongation by Pol II.

How, mechanistically, can NURF contribute to productive elongation? Knockdown of NURF alone leads to increased highly proximal pausing on a set of promoters ($n = 831$) (Fig. 3A) and this is coupled with improper $+1$ nucleosome positioning and phasing of early gene body nucleosomes at these promoters (Fig. 3B). We interpret this decrease in signal at the $+1$ nucleosome as misphasing, because less consistent positioning would lead to a decrease in aggregated signal at the dyad. While ATAC-seq is not the most precise method of mapping nucleosomes, in light of NURF's known activity of sliding $+1$ and sequential nucleosomes away from the TSS and into properly spaced arrays as indicated by the global decrease in MNase-seq signal at the $+1$ nucleosome and early gene-body nucleosomes on most genes in NURF^{-/-} embryos (Kwon et al. 2016), we believe this evidence supports the conclusion that these promoters with increased pause region Pol II density upon NURF knockdown also have misphased $+1$ nucleosomes upon NURF knockdown. While long-duration depletion experiments have the potential to trigger changes that are not directly attributable to depletion of the target factor, the transcriptional changes described here are likely to be the direct result of NURF depletion given that the same promoters also experience changes in nucleosome positioning consistent with the documented function of NURF (Kwon et al. 2016). Therefore, NURF appears to have a role in proper pausing and chromatin architecture in the early gene body, and without the activity of NURF, pause release and the transition to productive elongation are dysregulated (Fig. 1G).

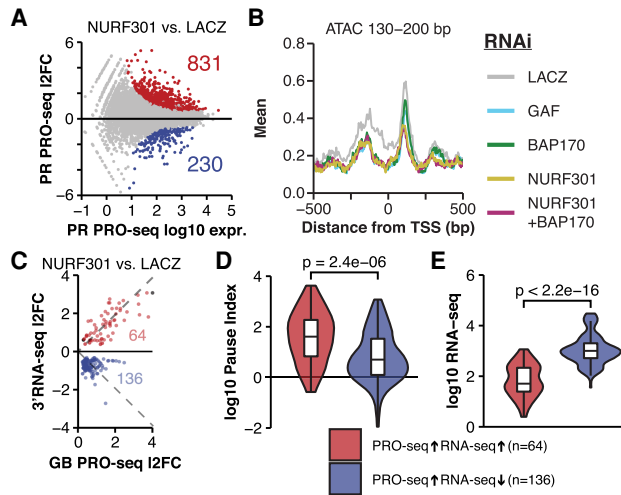


Figure 3. NURF positions nucleosomes, which influences pause release and elongation at highly expressed genes with low pausing. (A) MA plot for the comparison of spike-in normalized NURF301-RNAi versus LACZ-RNAi PRO-seq in the pause region (TSS -50 to $+100$); DESeq2 FDR < 0.01 . Number of genes significantly up-regulated or down-regulated is also shown. (B) Centers of mononucleosome sized ATAC-seq (fragments 130–200 bp) signal in 1-bp bins at all promoters with NURF301-RNAi up-regulated pausing ($n = 831$; see A; DESeq2 FDR < 0.01). Signal is the mean of 1000 subsamplings of 10% of regions. Confidence interval was calculated but omitted to avoid overplotting. (C) PRO-seq gene body (TSS $+200$ to TES -200) log₂ fold change compared with 3' RNA-seq log₂ fold change (in the last 1-kb region) upon NURF301 RNAi treatment. Only genes with DESeq2 FDR < 0.1 for both PRO-seq and RNA-seq are shown. Number of points in each quadrant are also shown. Dashed lines are one-to-one I2FC. (D) Distribution of pause indices (pause region PRO-seq signal/length normalized gene body PRO-seq signal) for genes with increased gene body PRO-seq density and increased RNA-seq signal (DESeq FDR < 0.1 ; $n = 64$) or genes with increased gene body PRO-seq density but decreased RNA-seq signal ($n = 136$). See C for gene classes. (E) Distribution of RNA-seq normalized counts for the two classes of genes described in D. In D and E filled violins represent the distribution and box plots show the median (center line), 25% and 75% quartiles (hinges), and 1.5*IQR (whiskers). Outliers are not plotted, and P-value is from a Mann-Whitney U-test.

Our model that NURF ensures efficient pause release and transition to productive elongation predicts that mRNA output would be decreased upon NURF depletion. Indeed, this is observed quite broadly (Supplemental Fig. S51). GAF interacts physically with NURF (Xiao et al. 2001; Nakayama et al. 2012; Lomaev et al. 2017), and a subset of GAF-dependent promoters have increased gene body Pol II density by PRO-seq (Fig. 1G, right panel); as such we reasoned that genes with increased gene body Pol II density upon NURF knockdown ($n = 831$) might represent primary targets of NURF. Genes with increased GB PRO-seq signal in the NURF knockdown split into two classes with the majority having decreased mRNA-seq signal ($n = 136$) (Fig. 3C). This can be explained by Pol II moving more slowly without the activity of NURF, which leads to decreased mRNA output despite increased Pol II

density (PRO-seq). Further analysis revealed that genes with increased GB PRO-seq and decreased 3' mRNA-seq upon NURF knockdown (Fig. 3C, bottom half), when compared with those that have increased GB PRO-seq and increased mRNA-seq signal (Fig. 3C, top half), are normally (1) less paused, (2) more expressed, (3) characterized by higher promoter ATAC-seq hypersensitivity signal that narrows upstream of the TSS upon NURF knockdown, (4) marked by a well-positioned $+1$ nucleosome that shows decreased signal upon NURF knockdown, and (5) distinguished by greater gene body polymerase density that further increases upon NURF knockdown (Fig. 3D,E; Supplemental Fig. S9B–D). Taken together, we propose that these findings indicate that these moderately expressed, less paused genes depend more strongly upon the activity of NURF to ensure proper nucleosome positioning. Upon NURF depletion, nucleosomes present an energy barrier to productive elongation, which leads to higher gene body polymerase density despite lower mRNA output as a result of slow-moving polymerases.

We speculate that without the activity of NURF, nucleosomes might drift into sequence-determined “energy wells”—tracts of DNA sequence where nucleosome displacement is less energetically favored—that are difficult for Pol II to transit, especially in the early stage of pause release. Under this model, without the assistance of NURF, both pause release and productive elongation would be inefficient due to the increased energy barrier more tightly DNA-associated nucleosomes present to transcribing Pol II. It was previously demonstrated that in the absence of NURF, $+1$ nucleosomes drift toward the TSS, and early gene body nucleosomes are misphased out to ~ 1 kb at NURF-bound promoters using MNase-seq in *Drosophila* embryonic tissue (Kwon et al. 2016). NURF mutant animals have less intense MNase-seq signal associated with the $+1$ nucleosome at NURF-bound promoters, and the signal maxima shifts ~ 12 bp toward the TSS (Kwon et al. 2016). Without NURF, these nucleosomes likely are free to drift into positions that are energetically opposed to Pol II transit, leading to inefficient pause release and therefore increased pause region PRO-seq signal. Taken together, these results indicate that GAF works interdependently with NURF at a subset of GAF target promoters to ensure proper nucleosome positioning in the early gene body for energetically favorable nucleosome transit by Pol II, a process downstream from PBAP’s GAF-directed eviction of nucleosomes from NFRs.

Discussion

Here, we demonstrate that *Drosophila* GAF possesses pioneer-like activity that depends on both SWI/SNF (PBAP) and ISWI (NURF) family ATP-dependent nucleosome remodeling complexes to establish optimal chromatin architecture for transcription at target promoters (Fig. 4A). SWI/SNF (PBAP) evicts nucleosomes from promoters, establishing a nucleosome-free region which allows Pol II to

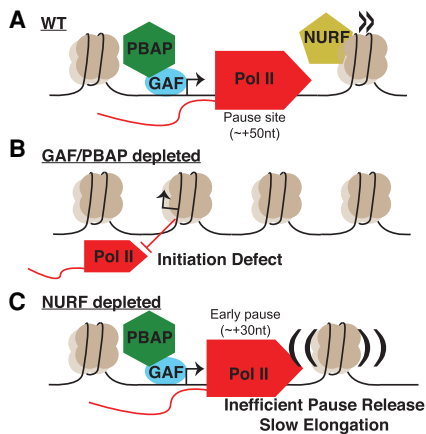


Figure 4. Nucleosome remodelers and pioneer factors coordinate to establish permissive chromatin architecture. (A–C) Cartoon summarizing the findings of this study.

be recruited and initiate transcription (Fig. 4B). This first major step of transcription allows Pol II to begin transcription and progress to the promoter-proximal pause region and is required at most GAF-dependent promoters. ISWI (NURF) then ensures that the nucleosomes along the early gene body are properly phased at a subset of GAF target promoters, thereby facilitating Pol II to transition to pause release and productive elongation in an energetically favorable manner (Fig. 4C). This work solidifies decades of *in vitro* biochemistry findings in *Drosophila* by resolving the roles of these factors *in vivo*, and to our knowledge is the first report of a pioneer-like factor working cooperatively with both ISWI and SWI/SNF remodelers to establish transcription-permissive chromatin at target promoters in metazoans.

These results indicate that a single pioneer-like transcription factor (GAGA factor) is able to orchestrate the activity of multiple nucleosome remodeling complexes that regulate the first three stages of the transcription cycle (recruitment, pausing, and transition to productive elongation). GAF and PBAP knockdowns have virtually identical effects on promoter chromatin accessibility and Pol II pausing, yet GAF still binds to chromatin (weakly) in the absence of PBAP. We speculate that this is because GAF first interacts transiently with nucleosome bound DNA and triggers the recruitment of PBAP and the subsequent removal of nucleosomes, thus allowing GAF to become more stably bound. This data shows that GAF synergizes with PBAP to clear promoters of nucleosomes and allow Pol II to be recruited, where it rapidly initiates transcription and traverses to the pause site. In light of protein interaction data demonstrating that GAF and PBAP interact physically (Nakayama et al. 2012; Lomaev et al. 2017), the most congruent explanation is that GAF directly recruits PBAP to target promoters. We note that we have not tested whether GAF can bind to nucleosome-bound DNA like classic pioneer factors (Zaret and Mango 2016); therefore, we cannot rule out the possibility that GAF and PBAP binding are simply interdepen-

dent events. Further analysis demonstrated that GAF also synergizes with NURF to position the +1 nucleosome at a subset of GAF-dependent promoters, which allows for efficient pause release and transition to productive elongation. The most obvious explanation for this is that GAF is responsible for recruitment of NURF, and indeed this is supported by evidence that GAF interacts physically with NURF and both GAF and NURF are required for *in vitro* nucleosomes remodeling activity on an *hsp70* template (Tsukiyama and Wu 1995; Xiao et al. 2001; Lomaev et al. 2017). However, our data does not rule out the possibility that NURF is not directly recruited by GAF and that they are simply functionally synergistic at a subset of GAF target promoters.

Strikingly, these roles for pioneer factors and specific nucleosome remodeler family members seem to also be consistent with limited recent data in mammals (King and Klose 2017; Hainer et al. 2019), which indicate that this finding might represent a deeply conserved mechanism throughout all of eukarya. To our knowledge, this is the first report of a single transcription factor with these expansive capabilities in metazoans, and the first view of not only how sequence-specific pioneer factors and nucleosome remodelers unite to regulate chromatin, but also how the resulting chromatin structure effects nascent transcription and mRNA production.

Materials and methods

RNAi treatments

Drosophila S2 cells were maintained at 25°C in M3 + BPYE medium with 10% FBS.

Two biological replicates were performed for each RNAi treatment as previously described (Duarte et al. 2016), except dsRNA complementary to *LACZ*, *GAF*, *BAP170*, *NURF301*, or both *BAP170* and *NURF301* were added to cultures. We generated dsRNA by PCR amplifying a dsDNA template from S2 genomic DNA with T7 RNA polymerase promoters on the 5' end of both strands, and then generated dsRNA using laboratory-made T7 RNA polymerase. See Supplemental Table S1 for oligonucleotide primer sequences. All RNAi treatments were done using 10 µg/mL dsRNA, including the *BAP170*+*NURF301* condition (5 µg/mL each). After 5 d, an equal volume of 25°C serum-free M3 + BPYE was added to cultures and they were incubated for 20 min at 25°C (this was to mimic a paired heat stress experiment that was performed alongside these experiments but is not presented in this publication). Cells were then harvested for PRO-seq, ATAC-seq, and 3' RNA-seq, and aliquots were lysed by boiling in 1× Laemmli buffer for Western blot analysis.

Western blots

Western blots were performed using anti-GAF (1:500; laboratory-made) or anti-NURF301 (1:100; Novus Biologicals 40360002), with anti-Chromator (1:2000; laboratory-made) as a loading control. Loading was standardized by cell number and for each RNAi treatment, a serial twofold dilution curve was analyzed compared with the *LACZ*-RNAi condition. Protein was detected using dual-color secondary antibodies and blots were imaged using the LI-COR Odyssey system.

Custom genomes

To facilitate accurate counting of spike-in reads, published PRO-seq data that did not contain spiked-in human cells (Duarte et al. 2016), was aligned to a repeat-masked human genome (hg38 assembly [Lander et al. 2001], retrieved from the UCSC genome browser [Haeussler et al. 2019]) using bowtie2 (Langmead and Salzberg 2012) using default parameters. Unique alignments (mapq > 1) were retained, and any regions with alignments were masked using bedtools maskfasta (Quinlan and Hall 2010). This custom-masked genome was then combined with the *Drosophila* genome (dm6 genome assembly [Hoskins et al. 2015], retrieved from the UCSC genome browser [Haeussler et al. 2019]). This allowed us to align PRO-seq data (containing both human- and fly-derived sequences) to this combined genome and ensured that no *Drosophila*-derived reads aberrantly mapped to the human genome and skewed spike-in normalization factors. We also masked any region in the dm6 genome assembly >100 bp with >80% homology to *Hsp70Aa* in order to uniquely map sequencing data to a single copy of *Hsp70*.

Gene annotations

We started with a list of all unique FlyBase transcripts (Thurmond et al. 2019), and reassigned the TSS based on the site of maximum PRO-cap signal (Kwak et al. 2013) in the window of TSS ± 50 bp. We then filtered out transcripts <500-bp long and removed any duplicate transcripts (occasionally two isoforms with TSSs within 50 bp of each other are corrected to the same PRO-cap maximum site, resulting in a duplicate transcript). We then discarded any transcript for which length-normalized PRO-seq signal in the TSS upstream region (−400 to −100) was more than half that in the pause region (−50 to +100) or more than that in the gene body region (TSS +200 to TES −200). This removed transcripts for which read-through transcription from an upstream gene is a major driver of signal within that gene and removes most transcript isoforms other than the most expressed isoform. This filtering left a list of 9375 genes, which was the starting point for DESeq2 (Love et al. 2014) differential expression testing and PCA analyses.

PRO-seq library preparation

PRO-seq library preparation was performed as previously described (Kwak et al. 2013; Mahat et al. 2016) using 2×10^7 cells per condition. We spiked in 2.7×10^5 human K562 cells immediately after harvesting cells to facilitate robust normalization of PRO-seq data. We substituted MyOne C1 Streptavidin beads for the M280 beads recommended by the published protocol, as their negatively charged surface is thought to reduce nonspecific nucleic acid binding, and we used 5' and 3' adapters that each had a 6N unique molecular identifier at the ligation junction to facilitate computational PCR deduplication of reads. PRO-seq libraries were all amplified for 11 PCR cycles and sequenced on an Illumina NextSeq in 37×37 paired end mode.

PRO-seq data analysis

Data quality was assessed with fastqc (<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>). Adapters were trimmed and UMIs were extracted using fastp (Chen et al. 2018), and rRNA reads were removed using bowtie2 (Langmead and Salzberg 2012). Reads were then aligned to the combined dm6/hg38 genome assembly described above and reads aligning uniquely (mapq > 10) to the human genome were counted for spike-in normalization. Reads were then mapped to the dm6 genome using bowtie2

(Langmead and Salzberg 2012), and only uniquely mapping reads (mapq > 10) were retained. Alignments were PCR-deduplicated using UMI-tools (spike-in alignments were also deduplicated) (Smith et al. 2017). BigWig coverage tracks of alignment 3' end positions in single-base-pair bins were then generated using deepTools (Ramírez et al. 2014). Normalization factors were derived by taking the minimum number of reads mapped to the spike-in genome across all samples and dividing that by the number of mapped spike-in reads for each sample (Supplemental Eq. 1). The alignment pipeline used can be found at http://github.com/jaj256/PROseq_alignment.sh, commit 55a08db. See Supplemental Table S2 for PRO-seq alignment metrics and normalization factors.

ATAC-seq library preparation

ATAC-seq was performed as previously described (Buenrostro et al. 2015), with some modifications for *Drosophila* cells. Briefly, 10^5 cells were washed with ice-cold PBS, and then resuspended in ice-cold lysis buffer (10 mM Tris-Cl at pH 7.4, 10 mM NaCl, 3 mM MgCl₂, 0.1% NP-40, 1× Pierce protease inhibitors [Thermo Scientific]) and incubated for 3 min on ice. Nuclei were then pelleted and resuspended in transposition buffer (10 mM Tris-Cl at pH 7.4, 10% DMF, 5 mM MgCl₂), and 1.5 μ L of laboratory-made Tn5 transposase was added. After a 30-min incubation in a thermomixer at 37°C, DNA was extracted using phenol:chloroform, PCR amplified for 11 cycles, and sequenced on an Illumina NextSeq in 37×37 paired end mode.

ATAC-seq data analysis

Reads were aligned to the dm6 genome assembly using bowtie2 (Langmead and Salzberg 2012) in local mode, and only unique alignments were retained (mapq > 10). Signal was then divided into two classes: hypersensitivity (paired end alignments with fragment size <120 bp, which represents hypersensitive chromatin and generates fragments smaller than mononucleosomes), and mononucleosome (paired end alignments with fragment size 130–200 bp, which represents two transposition events that roughly flank a mononucleosome-sized region). Coverage tracks were generated using deepTools (Ramírez et al. 2014). For hypersensitivity signal, entire alignments were “piled up” to generate coverage tracks, and for mononucleosome data only the central 3 bp of each alignment were considered. ATAC-seq peaks were called using macs2 (Zhang et al. 2008). See Supplemental Table S3 for alignment metrics.

3' RNA-seq

3' RNA-seq libraries were prepared using the QuantSeq 3' mRNA-seq library prep kit (Lexogen) with the UMI add-on kit. For each condition, 10^6 cells were added to a fixed amount of ERCC spike-in RNA mix (Invitrogen), and RNA was extracted using TRIzol reagent (Invitrogen). RNA treated with RNase free DNase I (Thermo Scientific), and the absence of DNA was confirmed using the Qubit dsDNA-HS assay (Thermo Scientific). RNA quality was confirmed using denaturing agarose gel electrophoresis. 3' RNA-seq libraries were prepared using 325 ng of total RNA per condition according to manufacturer's instructions and sequenced on an Illumina NextSeq in 75-bp single-end mode. Reads were trimmed of adapter and poly(A) sequences and UMIs were extracted using fastp (Chen et al. 2018). Reads were then aligned to a combined dm6/ERCC reference genome using STAR (Dobin et al. 2013), and reads mapped to the ERCC standards were counted for spike-in normalization. Alignments were PCR-deduplicated using UMI-tools (Smith et al. 2017) and only unique reads

were retained (mapq = 255). The 5' ends of reads were used to generate signal tracks (so that transcripts were scored in a read-length independent manner) using deepTools (Ramírez et al. 2014). Spike-in normalization factors were calculated as described above for PRO-seq. See Supplemental Table S4 for alignment metrics and normalization factors.

CUT&RUN

CUT&RUN was performed as described (Skene and Henikoff 2017; Skene et al. 2018). We used both anti-GAF (laboratory-made) or anti-NURF301 (Novus Biologicals 40360002) at a 1:10 dilution for the antibody binding step. ProteinA-MNase was incubated with calcium on ice for 30 min, and cleaved fragments were recovered by phenol:chloroform extraction. Library prep was performed using the following steps: (1) Ends of digested fragments were repaired by incubation for 30 min at 25°C with 0.5 U/μL T4 PNK, 0.12 U/μL T4 DNA polymerase, and 0.05 U/μL Klenow fragment in 1× T4 DNA ligase buffer (with ATP) and 0.5 mM dNTPs. (2) Fragments were A-tailed by incubation for 30 min at 37°C with 0.25 U/μL Klenow exo- and 0.5 mM dATP in 1× NEBuffer 2. (3) Adapters were added by incubation on the lab bench for 2 h with 4.38 nM laboratory-made Illumina TruSeq forked adapters and 24 U/μL T4 DNA ligase in 1× T4 DNA ligase buffer (with ATP). (4) Library DNA was recovered using AMPure XP beads (1.8× concentration) and PCR-amplified for 15 cycles (all enzymes from New England Biolabs). Libraries were sequenced on an Illumina NextSeq in 37×37 paired-end mode. Adapter sequences were removed using fastp (Chen et al. 2018), and reads were aligned to the dm6 reference genome using bowtie2 (Langmead and Salzberg 2012). Only uniquely mapped reads (mapq > 10) with fragment size smaller than 120 bp were retained, and signal coverage tracks were generated using deepTools (Ramírez et al. 2014). Signal was normalized per million mapped reads. See Supplemental Table S5 for alignment metrics.

ChIP-seq

ChIP-seq was performed as previously described (Fuda et al. 2015). Briefly, 2×10^7 cells from each of two replicates of LACZ, GAF, and BAP170 RNAi-treated S2 cells were washed with ice-cold PBS, then cross-linked in 1% formaldehyde for 10 min at 25°C. Formaldehyde was quenched for 5 min at 25°C with 0.125 M glycine, then fixed cells were washed twice with ice-cold PBS. Affinity purified anti-GAF (laboratory-made) was used at 1:200 dilution and was coupled to prewashed Sheep anti-Rabbit IgG M280 Dynabeads in IP Buffer (150 mM NaCl, 20 mM Tris-HCl at pH 8.0, 1% Triton X-100, 5 mg/mL BSA) for 12 h at 4°C. Fixed cells were resuspended in lysis buffer (1% SDS, 10 mM EDTA, 10 mM Tris-HCl at pH 8.0) and incubated for 5 min on ice, then the lysate was sonicated in a Diagenode Bioruptor on high for 10 cycles of 30 sec on and 30 sec off. Immunoprecipitation was performed overnight at 4°C. Beads were washed three times with wash buffer 1 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 150 mM NaCl, 20 mM Tris-HCl at pH 8.0), twice with wash buffer 2 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM NaCl, 20 mM Tris-HCl at pH 8.0), once with wash buffer 3 (0.1% SDS, 1% Triton X-100, 2 mM EDTA, 500 mM LiCl, 20 mM Tris-HCl at pH 8.0), and three times with wash buffer 4 (2 mM EDTA, 10% Glycerol, 20 mM Tris-HCl pH 8.0). Cross-links were reversed using 0.5% SDS and Proteinase K overnight at 65°C in a thermomixer, and DNA was purified by performing two sequential phenol:chloroform extractions followed by a chloroform extraction and concentration by ethanol precipitation. Af-

ter degrading RNA using a cocktail of RNase A and RNase T1 (Thermo), DNA was prepared for sequencing using the NEB Ultra II DNA library preparation kit for Illumina according to manufacturer instructions, and libraries were sequenced on a NextSeq500 instrument in paired-end mode. Data was analyzed as described for published ChIP-seq experiments (see below), except alignment was performed in paired end mode and instead of extending reads, actual fragments (determined using paired end reads) were used to generate coverage tracks.

Reanalysis of published data

GAF ChIP-seq (Fuda et al. 2015) raw reads were downloaded and mapped to the dm6 genome assembly using bowtie2 (Langmead and Salzberg 2012), and only uniquely mapping reads were retained (mapq > 10). Single end reads were extended 200 bp and reads-per-million normalized coverage tracks were generated using deepTools (Ramírez et al. 2014). Peaks were called using macs2 (Zhang et al. 2008). A M1BP ChIP-seq (Li and Gilmour 2013) signal track was downloaded and converted for the dm6 assembly using liftOver (Haeussler et al. 2019), and signal was normalized on a per-million basis. M1BP knockdown PRO-seq (Duarte et al. 2016) normalized signal tracks were accessed and converted to the dm6 genome assembly as above. BEAF-32 ChIP-seq (Liang et al. 2014) raw reads were downloaded, aligned using bowtie2 (Langmead and Salzberg 2012), and only uniquely mapping reads were retained (mapq > 10). Single end reads were extended 200 bp and reads-per-million normalized coverage tracks were generated using deepTools (Ramírez et al. 2014). See Supplemental Table S6 for accession numbers for all published data used in this manuscript.

DE testing

Signal counting in each set of regions for each data type was performed using functions from the BRGenomics package (<https://bioconductor.org/packages/BRGenomics>). Differential expression testing and principal component analysis was performed using DESeq2 (Love et al. 2014). Genes with adjusted *P*-value < 0.01 were considered differentially expressed.

Browser shots

Browser shots were generated using a custom R function, which can be found at https://github.com/IAJ256/browser_plot.R (commit 1352d5c).

Metaprofiles

Metaprofiles were generated using the BRGenomics package (<https://bioconductor.org/packages/BRGenomics>) by calculating a signal matrix across all regions in a set using the bin size specified, then sampling 10% of regions 1000 times to calculate the mean and 75% confidence interval. In some cases, confidence intervals were removed to avoid overplotting. Visualization was performed using ggplot2 (Wickham 2016).

Motif analysis

To search for motifs overrepresented in one set of promoters compared with another, we used DREME (Bailey 2011) with an *e*-value threshold of 0.001.

Classification of GAF-bound promoters

We considered a promoter GAF-bound if the promoter region (−500 to TSS) overlapped with a GAF ChIP-seq peak (see above). We then considered these GAF-bound promoters as having GAF-dependent pausing or GAF-independent pausing on the basis of whether or not they had significantly decreased PRO-seq in the pause region compared with the LACZ-RNAi control (DESeq2 FDR < 0.01, log₂ Fold Change < 0). We further subdivided the GAF-bound promoters with GAF-independent pausing by whether they were bound by M1BP, BEAF-32, or both, with “bound” defined as falling within the top 25% of promoters in our total set of GAF-bound promoters with GAF-independent pausing when rank-ordered by total ChIP-seq signal within the promoter (−500 to TSS) for a given factor. Heat maps were created using the ComplexHeatmap R package (Gu et al. 2016).

Data and code availability

All sequencing data has been deposited in GEO (GSE149339). All DESeq2 results tables, raw signal, and normalized bigWig files, gene lists, and ATAC-seq peaks can be accessed at <https://www.github.com/jaj256/GAF>. For ease of viewing, we have also created a custom UCSC track hub with pooled normalized data that can be imported to the UCSC genome browser using the following link: <https://github.com/JAJ256/GAF/raw/master/hub.txt>. All code used to analyze data and create figures is available at <https://www.github.com/jaj256/GAF>.

Acknowledgments

We thank members of the Lis laboratory and Feschotte laboratories for critical reading of the manuscript and helpful discussions regarding data analysis. We thank Paul Badenhorst for open discussion and data sharing regarding his work on NURF in *Drosophila* embryos. We are grateful to Dr. John Tamkun for providing us with an anti-BRM antibody. We thank the Cornell Biotechnology Resource Center Genomics facility and Peter Schweitzer for assistance with Illumina sequencing. We are grateful to Abdullah Ozer for assistance with PRO-cap TSS correction. This work was supported by National Institutes of Health grants GM025232 and HG009393 to J.T.L. J.J. was supported by National Human Genome Research Institute fellowship F31HG010820. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author contributions: J.J., F.M.D., and J.T.L. conceptualized the study and designed the experimental plan. J.J. performed all experiments and data analysis and wrote the first draft of the manuscript. J.J., F.M.D., and J.T.L. revised the manuscript.

References

Bailey TL. 2011. DREME: motif discovery in transcription factor ChIP-seq data. *Bioinformatics* **27**: 1653–1659. doi:10.1093/bioinformatics/btr261

Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr Protoc Mol Biol* **109**: 21.29.1–21.29.9. doi:10.1002/0471142727.mb2129s109

Chen S, Zhou Y, Chen Y, Gu J. 2018. Fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**: i884–i890. doi:10.1093/bioinformatics/bty560

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**: 15–21. doi:10.1093/bioinformatics/bts635

Duarte FM, Fuda NJ, Mahat DB, Core LJ, Guertin MJ, Lis JT. 2016. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. *Genes Dev* **30**: 1731–1746. doi:10.1101/gad.284430.116

Farkas G, Gausz J, Galloni M, Reuter G, Gyurkovics H, Karch F. 1994. The Trithorax-like gene encodes the *Drosophila* GAGA factor. *Nature* **371**: 806–808. doi:10.1038/371806a0

Fuda NJ, Guertin MJ, Sharma S, Danko CG, Martins AL, Siepel A, Lis JT. 2015. GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. *PLoS Genet* **11**: e1005108. doi:10.1371/journal.pgen.1005108

Gilchrist DA, Dos Santos G, Fargo DC, Xie B, Gao Y, Li L, Adelman K. 2010. Pausing of RNA polymerase II disrupts DNA-specified nucleosome organization to enable precise gene regulation. *Cell* **143**: 540–551. doi:10.1016/j.cell.2010.10.004

Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**: 2847–2849. doi:10.1093/bioinformatics/btw313

Haussler M, Zweig AS, Tyner C, Speir ML, Rosenbloom KR, Raney BJ, Lee CM, Lee BT, Hinrichs AS, Gonzalez JN, et al. 2019. The UCSC genome browser database: 2019 update. *Nucleic Acids Res* **47**: D853–D858. doi:10.1093/nar/gky1095

Hainer SJ, Bošković A, McCannell KN, Rando OJ, Fazio TG. 2019. Profiling of pluripotency factors in single cells and early embryos. *Cell* **177**: 1319–1329.e11. doi:10.1016/j.cell.2019.03.014

He S, Wu Z, Tian Y, Yu Z, Yu J, Wang X, Li J, Liu B, Xu Y. 2020. Structure of nucleosome-bound human BAF complex. *Science* **367**: 875–881. doi:10.1126/science.aaz9761

Hoskins RA, Carlson JW, Wan KH, Park S, Mendez I, Galle SE, Booth BW, Pfeiffer BD, George RA, Svirskas R, et al. 2015. The release 6 reference sequence of the *Drosophila melanogaster* genome. *Genome Res* **25**: 445–458. doi:10.1101/gr.185579.114

King HW, Klose RJ. 2017. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife* **6**: e22631. doi:10.7554/eLife.22631

Krietenstein N, Wal M, Watanabe S, Park B, Peterson CL, Pugh BF, Korber P. 2016. Genomic nucleosome organization reconstituted with pure proteins. *Cell* **167**: 709–721.e12. doi:10.1016/j.cell.2016.09.045

Kubik S, Bruzzone MJ, Shore D. 2017. Establishing nucleosome architecture and stability at promoters: roles of pioneer transcription factors and the RSC chromatin remodeler. *Bioessays* **39**: 1600237. doi:10.1002/bies.201600237

Kwak H, Fuda NJ, Core LJ, Lis JT. 2013. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science* **339**: 950–953. doi:10.1126/science.1229386

Kwon SY, Grisan V, Jang B, Herbert J, Badenhorst P. 2016. Genome-wide mapping targets of the metazoan chromatin remodeling factor NURF reveals nucleosome remodeling at enhancers, core promoters and gene insulators. *PLoS Genet* **12**: e1005969. doi:10.1371/journal.pgen.1005969

Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K, Dewar K, Doyle M, Fitzhugh W, et al. 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860–921. doi:10.1038/35057062

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**: 357–359. doi:10.1038/nmeth.1923

- Li J, Gilmour DS. 2013. Distinct mechanisms of transcriptional pausing orchestrated by GAGA factor and M1BP, a novel transcription factor. *EMBO J* **32**: 1829–1841. doi:10.1038/emboj.2013.111
- Liang J, Lacroix L, Gamot A, Cuddapah S, Queille S, Lhoumaud P, Lepetit P, Martin PGP, Vogelmann J, Court F, et al. 2014. Chromatin immunoprecipitation indirect peaks highlight long-range interactions of insulator proteins and Pol II pausing. *Mol Cell* **53**: 672–681. doi:10.1016/j.molcel.2013.12.029
- Lomaev D, Mikhailova A, Erokhin M, Shaposhnikov A V, Moresco JJ, Blokhina T, Wolle D, Aoki T, Ryabykh V, Yates JR, et al. 2017. The GAGA factor regulatory network: identification of GAGA factor associated proteins. *PLoS One* **12**: e0173602. doi:10.1371/journal.pone.0173602
- Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**: 550. doi:10.1186/s13059-014-0550-8
- Mahat DB, Kwak H, Booth GT, Jonkers IH, Danko CG, Patel RK, Waters CT, Munson K, Core LJ, Lis JT. 2016. Base-pair-resolution genome-wide mapping of active RNA polymerases using precision nuclear run-on (PRO-seq). *Nat Protoc* **11**: 1455–1476. doi:10.1038/nprot.2016.086
- Mayran A, Drouin J. 2018. Pioneer transcription factors shape the epigenetic landscape. *J Biol Chem* **293**: 13795–13804. doi:10.1074/jbc.R117.001232
- Moshe A, Kaplan T. 2017. Genome-wide search for Zelda-like chromatin signatures identifies GAF as a pioneer factor in early fly development. *Epigenetics Chromatin* **10**: 33. doi:10.1186/s13072-017-0141-5
- Nakayama T, Shimojima T, Hirose S. 2012. The PBAP remodeling complex is required for histone H3.3 replacement at chromatin boundaries and for boundary functions. *Development* **139**: 4582–4590. doi:10.1242/dev.083246
- Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**: 841–842. doi:10.1093/bioinformatics/btq033
- Ramírez F, Dündar F, Diehl S, Grüning BA, Manke T. 2014. DeepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res* **42**: W187–W191. doi:10.1093/nar/gku365
- Skene PJ, Henikoff S. 2017. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**: e21856. doi:10.7554/eLife.21856
- Skene PJ, Henikoff JG, Henikoff S. 2018. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* **13**: 1006–1019. doi:10.1038/nprot.2018.015
- Smith T, Heger A, Sudbery I. 2017. UMI-tools: modeling sequencing errors in unique molecular identifiers to improve quantification accuracy. *Genome Res* **27**: 491–499. doi:10.1101/gr.209601.116
- Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, Matthews BB, Millburn G, Antonazzo G, Trovisco V, et al. 2019. Flybase 2.0: the next generation. *Nucleic Acids Res* **47**: D759–D765. doi:10.1093/nar/gky1003
- Tsukiyama T, Wu C. 1995. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**: 1011–1020. doi:10.1016/0092-8674(95)90216-3
- Vallot A, Tachibana K. 2020. The emergence of genome architecture and zygotic genome activation. *Curr Opin Cell Biol* **64**: 50–57. doi:10.1016/j.ceb.2020.02.002
- Vihervaara A, Duarte FM, Lis JT. 2018. Molecular mechanisms driving transcriptional stress responses. *Nat Rev Genet* **19**: 385–397. doi:10.1038/s41576-018-0001-6
- Wagner FR, Dienemann C, Wang H, Stützer A, Tegunov D, Urlaub H, Cramer P. 2020. Structure of SWI/SNF chromatin remodeler RSC bound to a nucleosome. *Nature* **579**: 448–451. doi:10.1038/s41586-020-2088-0
- Wickham H. 2016. *Ggplot2: elegant graphics for data analysis*. Springer-Verlag, New York.
- Wilkins RC, Lis JT. 1998. GAGA factor binding to DNA via a single trinucleotide sequence element. *Nucleic Acids Res* **26**: 2672–2678. doi:10.1093/nar/26.11.2672
- Xiao H, Sandaltzopoulos R, Wang HM, Hamiche A, Ranallo R, Lee KM, Fu D, Wu C. 2001. Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol Cell* **8**: 531–543. doi:10.1016/S1097-2765(01)00345-8
- Yang J, Ramos E, Corces VG. 2012. The BEAF-32 insulator coordinates genome organization and function during the evolution of *Drosophila* species. *Genome Res* **22**: 2199–2207. doi:10.1101/gr.142125.112
- Zaret KS, Mango SE. 2016. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr Opin Genet Dev* **37**: 76–81. doi:10.1016/j.gde.2015.12.003
- Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, Nussbaum C, Myers RM, Brown M, Li W, et al. 2008. Model-based analysis of ChIP-seq (MACS). *Genome Biol* **9**: R137. doi:10.1186/gb-2008-9-9-r137