Pioneer factor GAF cooperates with PBAP and NURF to regulate transcription

Julius Judd[1], Fabiana M. Duarte[1,2], John T. Lis[1]*

[1]Department of Molecular Biology and Genetics, Cornell University, Ithaca, New York 14835, USA

[2]Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, MA 02138, USA

*Correspondence to: johnlis@cornell.edu

ORCID: Julius Judd (0000-0002-4602-0205); Fabiana M. Duarte (0000-0002-1261-4929); John T. Lis (0000-0003-1201-9406)

16 **Summary**

17   Transcriptionally silent genes must be activated throughout development. This requires

18   nucleosomes be removed from promoters and enhancers to allow transcription factor binding

19   (TFs) and recruitment of coactivators and RNA Polymerase II (Pol II). Specialized pioneer TFs

20   bind nucleosome-wrapped DNA to perform this chromatin opening by mechanisms that remain

21   incompletely understood[1–3]. Here, we show that GAGA-factor (GAF), a *Drosophila* pioneer

22   factor[4], interacts with both SWI/SNF and ISWI family chromatin remodelers to allow

23   recruitment of Pol II and entry to a promoter-proximal paused state, and also to promote Pol II's

24   transition to productive elongation. We found that GAF functions with PBAP (SWI/SNF) to

25   open chromatin and allow Pol II to be recruited. Importantly this activity is not dependent on

26   NURF as previously proposed[5–7]; however, GAF also functions with NURF downstream of this

27   process to ensure efficient Pol II pause release and transition to productive elongation apparently

28   through its role in precisely positioning the +1 nucleosome. These results demonstrate how a

29   single sequence-specific pioneer TF can synergize with remodelers to activate sets of genes.

30   Furthermore, this behavior of remodelers is consistent with findings in yeast[8–10] and mice[11–13],

31   and likely represents general, conserved mechanisms found throughout Eukarya.

32

33 **Main**

34   Pioneer transcription factors are a class of transcription factors that can bind and open
35 condensed chromatin. They control cell-fate decisions in development by opening chromatin at
36 previously inactive lineage-specific promoters and enhancers via sequence-specific binding[1–3].
37 These factors possess the unique ability to bind nucleosome-wrapped DNA, but the question of
38 how they evict nucleosomes and initiate transcription remains open.
39

40   From yeast to mammals, there is growing evidence that pioneer factors cooperate with
41 multiple ATP-dependent nucleosome remodeling complexes to establish transcription-
42 permissive chromatin architecture[8]. In yeast, the pioneer factor Abf1 synergizes with the RSC
43 complex (SWI/SNF family) to maintain the nucleosome-free region (NFR) of Abf1-bound
44 promoters, while ISW1a and ISW2 are required to properly position the +1 nucleosome and
45 phase downstream nucleosomes[9]. In mouse embryonic stem cells, the pioneer factors OCT4 and
46 NANOG are codependent on BAF complex (SWI/SNF family) subunit BRG1 to bind and open
47 chromatin at target sites[11,13]. Recent structural studies have illuminated how SWI/SNF family
48 remodelers bidirectionally evict nucleosomes from promoter NFRs in yeast[10] and mammals[12].
49

50   GAGA-factor (GAF) is a *Drosophila* transcription factor encoded by the *Trithorax-like*
51 (*trl*) gene[14] that preferentially binds GAGAG repeats, but is capable of binding a single GAG
52 trinucleotide[15]. We have previously demonstrated that, in *Drosophila* cell cultures, GAF is
53 essential for establishing paused Pol II on GAF-bound promoters, and that the NFRs of these
54 promoters fill with nucleosomes upon GAF depletion[16]. Without this activity, the response of a
55 subset of heat shock genes is impaired[17]. In early fly embryos, regions with chromatin signatures
56 similar to those at binding sites of the embryonic pioneer factor Zelda—but lacking Zelda
57 binding—are enriched for GAF binding, suggesting that GAF may also be an additional early
58 embryonic pioneer factor[4]. GAF interacts physically with the NURF complex (<u>N</u>ucleosome
59 <u>R</u>emodeling <u>F</u>actor) and both GAF and NURF are required to remodel nucleosomes on the *hsp70*
60 promoter *in vitro*[5,18]. We have previously speculated that GAF recruits NURF to target
61 promoters, which clears them of nucleosomes and allows Pol II initiation and subsequent pausing
62 to proceed[7]. However, early studies speculated that GAF can also interact with Brahma (Brm)
63 complexes (SWI/SNF family; BAP/PBAP)[6], and recent evidence indicates that GAF physically
64 interacts with PBAP (Polybromo associated Brm) but not BAP[19,20] in addition to NURF.
65

66   To test which of these remodelers is responsible for GAF's ability to establish
67 transcription-permissive chromatin architecture at target genes, we depleted GAF, NURF301,
68 and BAP170 (unique subunits of the NURF and PBAP complexes that are essential for complex
69 functionality)—as well as NURF301 and BAP170 simultaneously—in S2 cells using RNAi (Fig.

70   1a). After confirming knockdown efficiency (Extended Data Fig. 1), we used a combination of

71   PRO-seq, ATAC-seq, and 3'RNA-seq to monitor changes in nascent transcription, chromatin

72   state, and mRNA output.

73

74   **GAF synergizes with PBAP to open chromatin**

75

76   We used a spike-in normalization strategy for PRO-seq and 3'RNA-seq (see methods) to

77   ensure the detection of widespread transcriptional changes that can be hidden by centralizing

78   normalization strategies such as RPKM[21]. A principal component analysis of all genome-wide

79   data sets revealed that GAF-knockdown predominantly clusters with PBAP-knockdown (Fig. 1b,

80   Extended Data Fig. 2). After confirming data quality (Extended Data Fig. 2–4), we defined a set

81   of promoters that have downregulated pause region PRO-seq signal upon GAF knockdown

82   (Extended Data Fig. 5a). Notably, the number of genes with GAF-dependent pausing was far

83   greater than previously reported because our spike-in normalization scheme allowed us to

84   examine the genome-wide effects of GAF depletion with unprecedented sensitivity (n=685 in

85   this study, n=140 reported previously[16]). ATAC-seq hypersensitivity signal (fragments < 120 bp,

86   see methods) revealed that these promoters are substantially less accessible upon GAF, PBAP, or

87   NURF+PBAP knockdown (Fig. 1c–d, Extended Data Fig. 6), and PRO-seq shows that pausing is

88   severely reduced upon GAF, PBAP, or NURF+PBAP knockdown, but not after NURF

89   knockdown (Fig. 1c, e, f, Extended Data Fig. 6). These results clearly demonstrate that GAF

90   coordinates with PBAP–not NURF as previously proposed–to regulate Pol II recruitment by

91   evicting nucleosomes from the NFRs of target promoters. To our knowledge, this is the first

92   report of a pioneer factor synergizing specifically with PBAP (or PBAF, the homologous

93   mammalian complex) to maintain accessible target promoters in metazoans.

94

95   In contrast to PBAP, NURF knockdown increases PRO-seq signal in the pause region

96   and in the gene body region compared to the LACZ-RNAi control, particularly in the early pause

97   region closer to the TSS (Fig. 1f, left panel). We then compared the changes in pause region

98   PRO-seq signal upon GAF knockdown to that observed after PBAP and NURF knockdown on a

99   gene-by-gene basis. This revealed a near-perfect one-to-one correlation between GAF and PBAP

100   knockdowns, but minor anticorrelation between GAF and NURF knockdowns (Fig. 1g, compare

101   left panel to right panel, and Extended Data Fig. 7d for the NURF+PBAP knockdown). When we

102   examined promoters with PBAP-dependent pausing (n=806; Extended Data Fig 5b), we

103   observed similar trends to those seen at GAF-dependent promoters: decreased pausing and

104   promoter accessibility after GAF, PBAP, and NURF+PBAP knockdown, and increased pausing

105   and narrowed promoter accessibility upon NURF knockdown (compare Extended Data Fig. 7a–c

106   to Fig. 1d-f). Taken together, these data indicate that PBAP and GAF act together to free the

107   promoter of nucleosomes, while NURF acts at a downstream step.

108

109   Is GAF's mechanistic role to bind nucleosome-bound DNA and recruit the PBAP

110   remodeling complex where they act synergistically to remove nucleosomes, or does GAF

111   binding have an intrinsic ability to displace nucleosomes? The striking loss of PRO-seq signal

112   and loss of chromatin openness of promoters (Fig. 1c–d) described when either factor is depleted

113   argues for a highly synergistic model, where GAF alone has little intrinsic chromatin opening

114   activity. To investigate further, we compared ATAC-seq signal between the GAF and PBAP

115   knockdown conditions, which revealed significant low-magnitude changes at only a small

116   number of sites (Extended Data Fig. 8a), indicating that GAF does not possess sufficient intrinsic

117   chromatin opening ability to account for the effects of GAF knockdown on chromatin. In further

118   support of this, 88% of promoters with decreased pausing upon GAF knockdown had decreased

119   pausing upon PBAP knockdown (n=603; Extended Data Fig. 8b).

120

121   Small sets of promoters show only GAF- or only PBAP-knockdown effects. We found

122   that GAF-specific promoters (n=82; PBAP knockdown causes no change) had higher levels of

123   GAF ChIP-seq signal and were far less sensitive to the GAF-knockdown than the class of genes

124   dependent on both GAF and PBAP (Extended Data Fig. 8c-d). We speculate that these promoters

125   may be held open by paused Pol II[22] that is generated by mechanisms independent of PBAP, or

126   the level of PBAP remaining after knockdown was sufficient to be recruited by the high level of

127   GAF bound at these promoters. PBAP-specific promoters are mostly not bound by GAF

128   (Extended Data Fig. 8c), and often contain the binding motif for the transcription factor lola

129   (Extended Data Fig. 8f), which might function like GAF in its collaboration with PBAP.

130

131   **M1BP can establish paused Pol II independent of GAF**

132

133   Not all GAF bound promoters have GAF-dependent pausing, and some of these bound

134   but unaffected promoters are bound by M1BP (Motif 1 Binding Protein) and the insulator BEAF-

135   32 (Boundary Element Associated Factor)[16,23]. However, it remains unclear whether M1BP acts

136   redundantly with GAF to open promoters and promote pausing, or if M1BP/BEAF-32 simply

137   insulate promoters from GAF's activity. To investigate this, we divided GAF-bound promoters

138   into two classes based on whether they have GAF-dependent pausing (n=600) or not (n=1,245),

139   and found that the BEAF-32 and M1BP motifs[23,24] were overrepresented in GAF-bound

140   promoters with unchanged pausing (Fig. 2a). We then subdivided the class of GAF-bound, GAF-

141   independent pausing genes by whether they were bound by M1BP[23] (n=152), BEAF-32[25]

142   (n=152), or both (n=159; Fig. 2b). GAF-binding was weaker and more diffuse in GAF-bound

143   genes with GAF-independent pausing (Class II–IV), while these promoters were directly and

144   strongly bound by either M1BP or BEAF-32 or both (Class II–IV). We know from our previous

145   study that genes bound by M1BP have reduced pause region PRO-seq signal upon M1BP

146   knockdown[17], and this reduction in pausing correlates with M1BP binding intensity (Fig. 2b).

147   Moreover, all classes of GAF-bound, GAF-independent pause genes had relatively unchanged

148   ATAC-seq hypersensitivity signal in promoters after GAF or BAP knockdown (Fig 2b, left).

149   This demonstrates that M1BP can open chromatin at promoters and create paused Pol II

150   independent of GAF/PBAP, and the weak and diffuse GAF binding at these sites is insufficient

151   to complement depletion of M1BP.

152

153   **NURF promotes transition to productive elongation**

154

155   GAF can physically interact with the remodelers PBAP and NURF[5,19,20] and appears to

156   function with each remodeler at distinct steps in transcription: GAF and PBAP open chromatin

157   allowing Pol II initiation and entry to the promoter-proximal pause site; while GAF and NURF

158   ensure efficient transition to productive elongation. This role of GAF and PBAP in the first of

159   these two steps is supported strongly by results described above (Fig 1). Evidence that NURF's

160   role is downstream of PBAP is provided by the observation that the PBAP+NURF double-

161   knockdown primarily mimics PBAP depletion in terms of changes in ATAC-seq and PRO-seq

162   patterns in the pause region (Fig 1). Support for NURFs role in productive elongation comes in

163   part from the fact that the PBAP knockdown only partially recapitulates the decrease in gene

164   body polymerase density seen after GAF depletion (Fig. 1e, right panel). In contrast, the

165   NURF+PBAP double-knockdown mirrors the GAF knockdown (Fig. 1f, right panel).

166   Furthermore, CUT&RUN assays demonstrate co-occupancy of GAF and NURF at promoters

167   genome-wide (Extended Data Fig. 9a), indicating GAF and NURF are likely to act together.

168   These results support the model that GAF coordinates with both remodelers to ensure efficient

169   transcription by first acting with PBAP to open chromatin and allow for the formation of

170   promoter-proximal paused Pol II, and then with NURF to establish chromatin structure at the

171   start of genes which ensures proper transition to productive elongation by Pol II.

172

173   How mechanistically can NURF contribute to productive elongation? Knockdown of

174   NURF alone leads to increased highly proximal pausing on a set of promoters (n=831; Extended

175   Data Fig. 5c) and this is coupled with improper +1 nucleosome positioning and phasing of early

176   gene body nucleosomes at these promoters (Fig. 3a). We interpret this decrease in signal at the

177   +1 nucleosome as misphasing, because less consistent positioning would lead to a decrease in

178   aggregated signal at the dyad. While ATAC-seq is not the most precise method of mapping

179   nucleosomes, in light of NURF's known activity of sliding +1 and sequential nucleosomes away

180   from the TSS and into properly spaced arrays[26], we believe this evidence supports the conclusion

181   that these promoters with increased pause region Pol II density upon NURF knockdown also

182   have misphased +1 nucleosomes upon NURF knockdown. Therefore, NURF has a role in proper

183   pausing and chromatin architecture in the early gene body, and without the activity of NURF,

184   pause release and the transition to productive elongation are dysregulated (Fig. 1f).

185

186   Our model that GAF recruits and functionally synergizes with NURF to ensure efficient

187   pause release and transition to productive elongation predicts that mRNA output would be

188   decreased upon NURF depletion. Indeed, this is observed quite broadly (Extended Data Fig. 5k).

189   GAF interacts physically with NURF[5,19,20], and GAF-dependent promoters have increased gene

190   body Pol II density by PRO-seq (Fig. 1f, right panel); as such we reasoned that genes with

191   increased gene body Pol II density upon NURF knockdown (n=831) might represent primary

192   targets of NURF. Genes with increased GB PRO-seq signal in the NURF knockdown split into

193   two classes with the majority having decreased mRNA-seq signal (Fig. 3b). This can be

194   explained by Pol II moving more slowly without the activity of NURF, which leads to decreased

195   mRNA output despite increased Pol II density (PRO-seq). Further analysis revealed that genes

196   which increased GB PRO-seq and decreased 3'mRNA-seq upon NURF knockdown (Fig. 3b,

197   bottom half), when compared to those that have increased GB PRO-seq and increased mRNA-

198   seq signal (Fig. 3b, upper half), are normally: (i) less paused; (ii) more expressed; (iii)

199   characterized by higher promoter ATAC-seq hypersensitivity signal that narrows upstream of the

200   TSS upon NURF knockdown; (iv) marked by a well-positioned +1 nucleosome that shows

201   decreased signal upon NURF knockdown; and (v) distinguished by greater gene body

202   polymerase density that further increases upon NURF knockdown (Extended Data Fig. 9b–f,

203   respectively). Taken together, we propose that these findings indicate that these moderately

204   expressed, less paused genes depend more strongly upon the activity of NURF to ensure proper

205   nucleosome positioning. Upon NURF depletion, nucleosomes present an energy barrier to

206   productive elongation, which leads to higher gene body polymerase density despite lower mRNA

207   output as a result of slow-moving polymerases.

208

209   We speculate that without the activity of NURF, nucleosomes might drift into sequence-

210   determined "energy wells"—tracts of DNA sequence where nucleosome eviction is less

211   energetically favored—that are difficult for Pol II to transit, especially in the early stage of pause

212   release. Under this model, without the assistance of NURF, both pause release and productive

213   elongation would be inefficient due to the increased energy barrier more tightly DNA-associated

214   nucleosomes present to transcribing Pol II. It was previously demonstrated that in the absence of

215   NURF, +1 nucleosomes drift toward the TSS, and early gene body nucleosomes are mis-phased

216   out to ~1kb at NURF-bound promoters using MNase-seq in *Drosophila* embryonic tissue[26].

217 NURF mutant animals have less intense MNase-seq signal associated with the +1 nucleosome at
218 NURF-bound promoters, and the signal maxima shifts ~12 bp towards the TSS[26]. Without
219 NURF, these nucleosomes likely are free to drift into positions that are energetically opposed to
220 Pol II transit, leading to inefficient pause release and therefore increased pause region PRO-seq
221 signal. Taken together, these results indicate that GAF recruits NURF to promoters where it
222 ensures proper nucleosome positioning in the early gene body for energetically favorable
223 nucleosome transit by Pol II, a process downstream of PBAP's GAF-directed eviction of
224 nucleosomes from NFRs.
225
226 **Summary**
227

228   We demonstrate that GAF is a sequence-specific pioneer factor in *Drosophila*, that
229 depends on the activity of both SWI/SNF (PBAP) and ISWI (NURF) family ATP-dependent
230 nucleosome remodeling complexes to establish optimal chromatin architecture for transcription
231 at target promoters (Fig. 4a). SWI/SNF (PBAP) evicts nucleosomes from promoters, establishing
232 a nucleosome-free region which allows Pol II to be recruited and initiate transcription (Fig. 4b).
233 This first major step of transcription allows Pol II to begin transcription and progress to the
234 promoter-proximal pause region. ISWI (NURF) then ensures that the nucleosomes along the
235 early gene body are properly phased, thereby facilitating Pol II to transition to pause release and
236 productive elongation in an energetically favorable manner (Fig. 4c). This work solidifies
237 decades of *in vitro* biochemistry findings in *Drosophila* by resolving the roles of these factors *in*
238 *vivo,* and to our knowledge is the first report of a pioneer factor working cooperatively with both
239 ISWI and SWI/SNF remodelers to establish transcription-permissive chromatin at target
240 promoters in metazoans.
241

242   These results indicate that a single pioneer transcription factor (GAGA-factor) is able to
243 orchestrate the activity of multiple nucleosome remodeling complexes that regulate the first three
244 stages of the transcription cycle (recruitment, pausing, and transition to productive elongation).
245 GAF and PBAP knockdowns have virtually identical effects on promoter chromatin accessibility
246 and Pol II pausing, showing that GAF synergizes with PBAP to clear promoters of nucleosomes
247 and allow Pol II to be recruited, where it rapidly initiates transcription and traverses to the pause
248 site. However, further analysis demonstrated that GAF also recruits NURF to position the +1
249 nucleosome, which allows for efficient pause release and transition to productive elongation. The
250 effects of the NURF knockdown are masked by the effects of the GAF/PBAP knockdowns,
251 because without Pol II recruitment, nucleosome positioning along the gene body appears to be
252 mostly irrelevant.
253

254 Strikingly, these roles for pioneer factors and specific nucleosome remodeler family
255 members seem to also be consistent with limited recent data in mammals[11,13], which indicate that
256 this finding might represent a deeply conserved mechanism throughout all of Eukarya. To our
257 knowledge, this is the first report of a single transcription factor with these expansive capabilities
258 in metazoans, and the first view of not only how sequence specific pioneer factors and
259 nucleosome remodelers unite to regulate chromatin, but also how the resulting chromatin
260 structure effects nascent transcription and mRNA production.
261
262 **Data availability.**
263 All sequencing data has been deposited in GEO (GSE149339). All DESeq2 results tables, raw
264 signal and normalized bigWig files, gene lists, and ATAC-seq peaks can be accessed at
265 https://www.github.com/jaj256/GAF. For ease of viewing, we have also created a custom UCSC
266 track hub with pooled normalized data that can be imported to the UCSC genome browser using
267 this link: https://github.com/JAJ256/GAF/raw/master/hub.txt.
268
269 **Code availability.**
270 All code used to analyze data and create figures is available at:
271 https://www.github.com/jaj256/GAF.
272

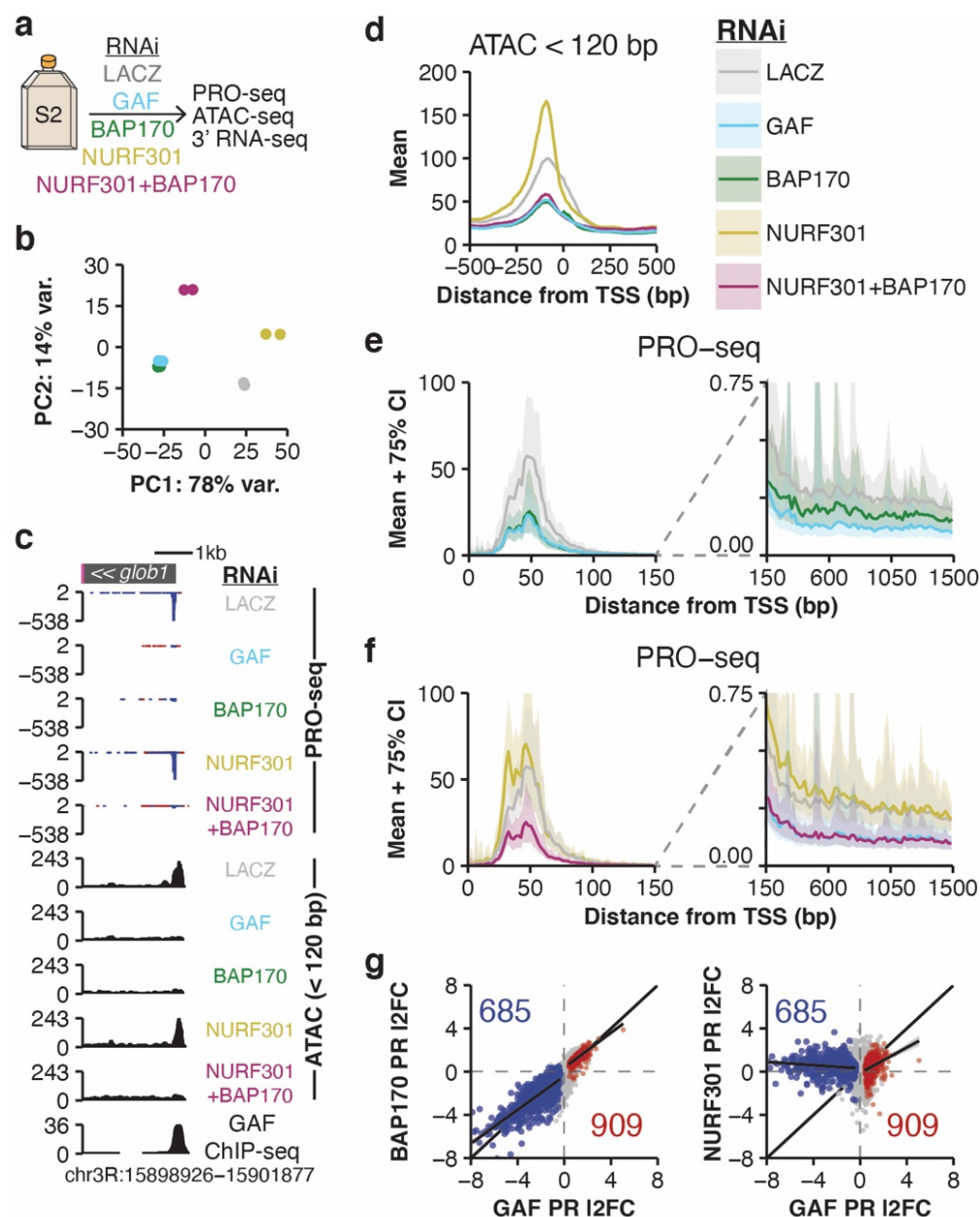273 **Acknowledgements**

282

283 **Author contributions**
284 J.J., F.M.D., and J.T.L. conceptualized the study and designed the experimental plan. J.J.
285 performed all experiments and data analysis and wrote the first draft of the manuscript. J.J.,
286 F.M.D., and J.T.L. revised the manuscript.
287

288 **Competing interests**
289 The authors declare no competing interests.
290

291

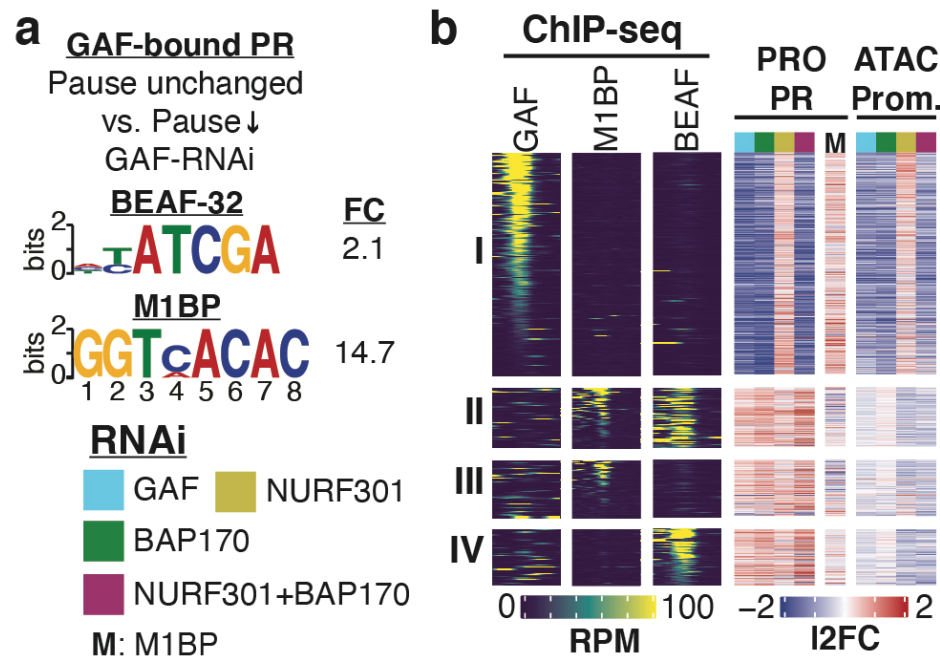**Figure 1: GAGA-factor opens chromatin via the PBAP complex.**

(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 *glob1-RB*.

(d) ATAC-seq (< 120 bp) signal in 1 bp bins at promoters with GAF-RNAi downregulated pausing (n=685; see Extended Data Figure 5a; DESeq2 FDR < 0.01). Signal is the mean of 1,000 sub-samplings of 10% of regions.
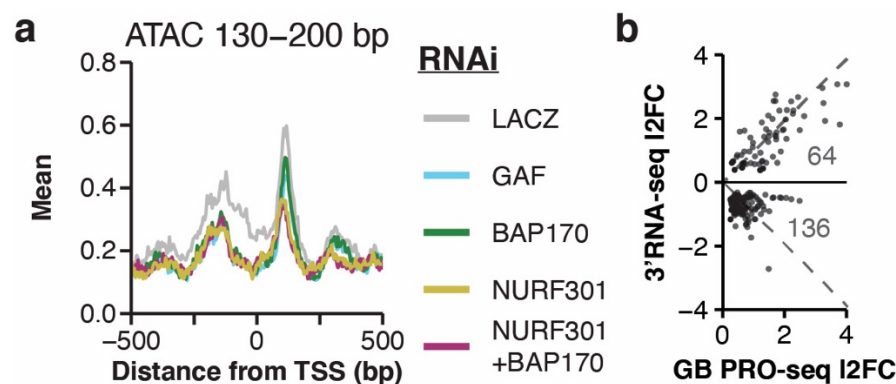
300   **(e)** PRO-seq signal for the LACZ, GAF, and BAP170 RNAi treatments. The pause region
301   (left) is in 2 bp bins, and the gene body (right) is in 20 bp bins. Data is shown as mean
302   (line) ± 75% confidence interval (shaded) from 1,000 sub-samplings of 10% of regions.
303   Gene set as in **(d)**.

304   **(f)** As in **(e)**, but for LACZ, NURF301, and NURF301+BAP170 RNAi treatments. GAF-
305   RNAi is also shown in the gene body region for comparison (blue line), though it is
306   partially obscured by the NURF+PBAP line (purple) due to similarity of the trace.

307   **(g)** Pause region (TSS -50 to +100) PRO-seq log2 fold change (l2FC) vs. the LACZ-RNAi
308   control; GAF-RNAi compared to BAP170-RNAi (left) or NURF301-RNAi (right).
309   Red/blue points are significantly changed by GAF-RNAi (DESeq2 FDR < 0.01). Also
310   shown: a GLM and 95% confidence interval for up- and down-regulated promoters.
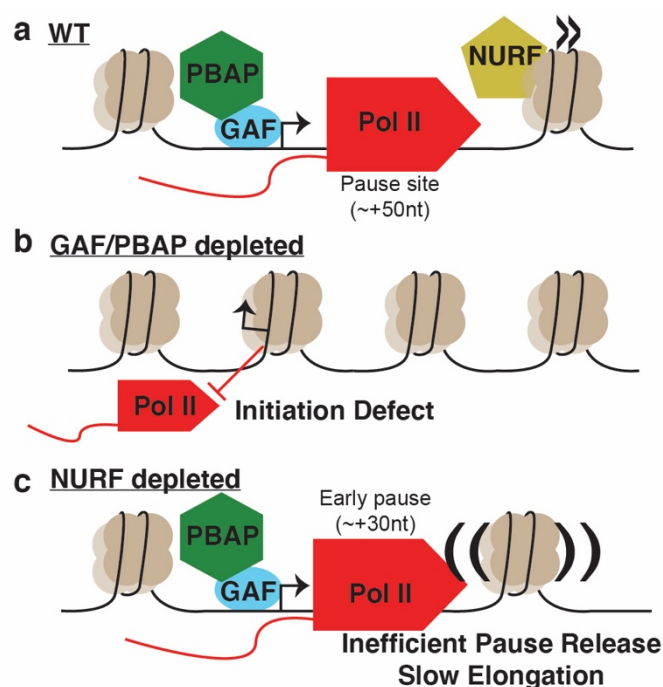
311

Figure 2: M1BP and BEAF-32 block GAFs activity

(a) Motifs enriched in GAF-bound promoters (GAF ChIP-seq peak within -500–TSS) with GAF-independent pausing (n=1,245) 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. (I) GAF-dependent pausing (n=600); (II–IV) GAF-independent pausing (n=1,245); (II) M1BP and BEAF-32 bound (n=159); (III) M1BP only (n=152); (IV) BEAF-32 only (n=152). Sort order: (I) GAF ChIP-seq; (II–III) M1BP ChIP-seq; (IV) BEAF-32 ChIP-seq. GAF-bound GAF-independent genes without M1BP or BEAF-32 ChIP-seq signal are not shown (n=782).

Figure 3: NURF positions nucleosomes which influences pause release and elongation.

(a) Centers of mononucleosome sized ATAC-seq (fragments 130–200 bp) signal in 1 bp bins at all promoters with NURF301-RNAi upregulated pausing (n=831; see Extended Data Figure 5C; DESeq2 FDR < 0.01). Signal is the mean of 1,000 sub-samplings of 10% of regions.

(b) PRO-seq gene body (TSS+200–TES-200) log2 fold change compared to 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 1-to-1 l2FC.

Figure 4: Nucleosome remodelers and pioneer factors coordinate to establish permissive chromatin architecture.

(a–c) Cartoon summarizing the findings of this article.

References

1. Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016).

2. Mayran, A. & Drouin, J. Pioneer transcription factors shape the epigenetic landscape. *J. Biol. Chem.* **293**, 13795–13804 (2018).

3. Vallot, A. & Tachibana, K. The emergence of genome architecture and zygotic genome activation. *Curr. Opin. Cell Biol.* **64**, 50–57 (2020).

4. Moshe, A. & Kaplan, T. Genome-wide search for Zelda-like chromatin signatures identifies GAF as a pioneer factor in early fly development. *Epigenetics Chromatin* **10**, 33 (2017).

5. Xiao, H. *et al.* Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. *Mol. Cell* **8**, 531–543 (2001).

6. Tsukiyama, T. & Wu, C. Purification and properties of an ATP-dependent nucleosome remodeling factor. *Cell* **83**, 1011–1020 (1995).

7. Vihervaara, A., Duarte, F. M. & Lis, J. T. Molecular mechanisms driving transcriptional stress responses. *Nat. Rev. Genet.* **19**, 385–397 (2018).

8. Kubik, S., Bruzzone, M. J. & Shore, D. Establishing nucleosome architecture and stability at promoters: Roles of pioneer transcription factors and the RSC chromatin remodeler. *Bioessays* **39**, (2017).

9. Krietenstein, N. *et al.* Genomic Nucleosome Organization Reconstituted with Pure Proteins. *Cell* **167**, 709-721.E12 (2016).

10. Wagner, F. R. *et al.* Structure of SWI/SNF chromatin remodeller RSC bound to a nucleosome. *Nature* **579**, 448–451 (2020).

11. Hainer, S. J., Bošković, A., McCannell, K. N., Rando, O. J. & Fazzio, T. G. Profiling of Pluripotency Factors in Single Cells and Early Embryos. *Cell* **177**, 1319-1329.E11 (2019).

12. He, S. *et al.* Structure of nucleosome-bound human BAF complex. *Science* **367**, 875–881 (2020).

13. King, H. W. & Klose, R. J. The pioneer factor OCT4 requires the chromatin remodeller BRG1 to support gene regulatory element function in mouse embryonic stem cells. *Elife* **6**, (2017).

14. Farkas, G. *et al.* The Trithorax-like gene encodes the Drosophila GAGA factor. *Nature* **371**, 806–808 (1994).

15. Wilkins, R. C. & Lis, J. T. GAGA factor binding to DNA via a single trinucleotide sequence element. *Nucleic Acids Res.* **26**, 2672–2678 (1998).

16. Fuda, N. J. *et al.* GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. *PLoS Genet.* **11**, e1005108 (2015).

17. Duarte, F. M. *et al.* Transcription factors GAF and HSF act at distinct regulatory steps to

379   modulate stress-induced gene activation. *Genes Dev.* **30**, 1731–1746 (2016).

380 18. Tsukiyama, T. & Wu, C. Purification and properties of an ATP-dependent nucleosome
381   remodeling factor. *Cell* **83**, 1011–1020 (1995).

382 19. Lomaev, D. *et al.* The GAGA factor regulatory network: Identification of GAGA factor
383   associated proteins. *PLoS One* **12**, e0173602 (2017).

384 20. Nakayama, T., Shimojima, T. & Hirose, S. The PBAP remodeling complex is required for
385   histone H3.3 replacement at chromatin boundaries and for boundary functions.
386   *Development* **139**, 4582–4590 (2012).

387 21. Mahat, D. B. *et al.* Base-pair-resolution genome-wide mapping of active RNA
388   polymerases using precision nuclear run-on (PRO-seq). *Nat. Protoc.* **11**, 1455–1476
389   (2016).

390 22. Gilchrist, D. A. *et al.* Pausing of RNA polymerase II disrupts DNA-specified nucleosome
391   organization to enable precise gene regulation. *Cell* **143**, 540–551 (2010).

392 23. Li, J. & Gilmour, D. S. Distinct mechanisms of transcriptional pausing orchestrated by
393   GAGA factor and M1BP, a novel transcription factor. *EMBO J.* **32**, 1829–1841 (2013).

394 24. Yang, J., Ramos, E. & Corces, V. G. The BEAF-32 insulator coordinates genome
395   organization and function during the evolution of Drosophila species. *Genome Res.* **22**,
396   2199–2207 (2012).

397 25. Liang, J. *et al.* Chromatin immunoprecipitation indirect peaks highlight long-range
398   interactions of insulator proteins and Pol II pausing. *Mol. Cell* **53**, 672–681 (2014).

399 26. Kwon, S. Y., Grisan, V., Jang, B., Herbert, J. & Badenhorst, P. Genome-Wide Mapping
400   Targets of the Metazoan Chromatin Remodeling Factor NURF Reveals Nucleosome
401   Remodeling at Enhancers, Core Promoters and Gene Insulators. *PLoS Genet.* **12**,
402   e1005969 (2016).

403

404 **Methods**

405 **Cell culture and RNAi treatments.** *Drosophila* S2 cells were maintained at 25 °C in M3 +

406 BPYE medium with 10% FBS. Two biological replicates were performed for each RNAi

407 treatment as previously described[17], except dsRNA complementary to LACZ, GAF, BAP170,

408 NURF301, or both BAP170 and NURF301 was added to cultures. We generated dsRNA by PCR

409 amplifying a dsDNA template from S2 genomic DNA with T7 RNA Polymerase promoters on

410 the 5' end of both strands, and then generated dsRNA using lab-made T7 RNA Polymerase. See

411 Supplementary Table 1 for oligonucleotide primer sequences. All RNAi treatments were done

412 using 10 μg/mL dsRNA, including the BAP170+NURF301 condition (5 μg/mL each). After 5

413 days, an equal volume of 25 °C serum-free M3 + BYPE was added to cultures and they were

414 incubated at 25 °C for 20 min (this was to mimic a paired heat-stress experiment that was

415 performed alongside these experiments but is not presented in this publication). Cells were then

416 harvested for PRO-seq, ATAC-seq, and 3'RNA-seq, and aliquots were lysed by boiling in 1x

417 Laemmli buffer for western blot analysis.

418

419 **Western blots.** Western blots were performed using anti-GAF (lab-made, 1:500) or anti-

420 NURF301(Novus Biologicals Cat. No. 40360002; 1:100), with anti-Chromator (lab-made,

421 1:2000) as a loading control. Loading was standardized by cell number and for each RNAi

422 treatment, a serial 2-fold dilution curve was analyzed compared to the LACZ-RNAi condition.

423 Protein was detected using dual-color secondary antibodies and blots were imaged using the LI-

424 COR Odyssey system.

425

426 **Custom genomes.** To facilitate accurate counting of spike-in reads, published PRO-seq data that

427 did not contain spiked-in human cells[17], was aligned to a repeat-masked human genome (hg38

428 assembly[27], retrieved from the UCSC genome browser[28]) using bowtie2[29] using default

429 parameters. Unique alignments (mapq > 1) were retained, and any regions with alignments were

430 masked using bedtools maskfasta[30]. This custom-masked genome was then combined with the

431 *Drosophila* genome (dm6 genome assembly[31], retrieved from the UCSC genome browser[28]. This

432 allowed us to align PRO-seq data (containing both human- and fly- derived sequences) to this

433 combined genome and ensured that no drosophila-derived reads aberrantly mapped to the human

434 genome and skewed spike-in normalization factors. We also masked any region in the dm6

435 genome assembly larger than 100 bp with greater than 80% homology to *Hsp70Aa* in order to

436 uniquely map sequencing data to a single copy of *Hsp70*.

437

438 **Gene annotations.** We started with a list of all unique FlyBase transcripts[32], and reassigned the

439 TSS based on the site of maximum PRO-cap signal[33] in the window of TSS ± 50 bp. We then

440 filtered out transcripts less than 500 bp long and removed any duplicate transcripts (occasionally

441   two isoforms with TSSs within 50 bp of each other are corrected to the same PRO-cap maximum

442   site, resulting in a duplicate transcript). We then discarded any transcript for which length-

443   normalized PRO-seq signal in the TSS-upstream region (-400 to -100) was more than half that in

444   the pause region (-50 to +100) or more than that in the gene body region (TSS+200 to TES-200).

445   This removed transcripts for which read-through transcription from an upstream gene is a major

446   driver of signal within that gene and removes most transcript isoforms other than the most

447   expressed isoform. This filtering left a list of 9,375 genes, which was the starting point for

448   DESeq2[34] differential expression testing and PCA analyses.

449

450   **PRO-seq library preparation.** PRO-seq library preparation was performed as previously

451   described[21,33] using 2×10^7 cells per condition. We spiked in 2.7×10^5 human K562 cells

452   immediately after harvesting cells to facilitate robust normalization of PRO-seq data. We

453   substituted MyOne C1 Streptavidin beads for the M280 beads recommended by the published

454   protocol, as their negatively charged surface is thought to reduce non-specific nucleic acid

455   binding, and we used 5' and 3' adapters that each had a 6N unique molecular identifier at the

456   ligation junction to facilitate computational PCR deduplication of reads. PRO-seq libraries were

457   all amplified for 11 PCR cycles and sequenced on an Illumina NextSeq in 37x37 paired end

458   mode.

459

460   **PRO-seq data analysis.** Data quality was assessed with fastqc[35]. Adapters were trimmed and

461   UMIs were extracted using fastp[36], and rRNA reads were removed using bowtie2[29]. Reads were

462   then aligned to the combined dm6/hg38 genome assembly described above and reads aligning

463   uniquely (mapq > 10) to the human genome were counted for spike-in normalization. Reads

464   were then mapped to the dm6 genome using bowtie2[29], and only uniquely mapping reads (mapq

465   > 10) were retained. Alignments were PCR-deduplicated using UMI-tools[37] (spike-in alignments

466   were also de-duplicated). BigWig coverage tracks of alignment 3' end positions in single base

467   pair bins were then generated using deepTools[38]. Normalization factors were derived by taking

468   the minimum number of reads mapped to the spike-in genome across all samples and dividing

469   that by the number of mapped spike-in reads for each sample (Supplementary Equation 1). The

470   alignment pipeline used can be found at http://github.com/jaj256/PROseq_alignment.sh, commit

471   55a08db). See Supplementary Table 2 for PRO-seq alignment metrics and normalization factors.

472

473   **ATAC-seq library preparation.** ATAC-seq was performed as previously described[39], with

474   some modifications for *Drosophila* cells. Briefly, 10^5 cells were washed with ice-cold PBS, and

475   then resuspended in ice-cold lysis buffer (10 mM Tris-Cl pH 7.4, 10 mM NaCl, 3 mM $MgCl_2$,

476   0.1% NP-40, and 1x Pierce Protease Inhibitors [Thermo Scientific]) and incubated on ice for 3

477   min. Nuclei were then pelleted and resuspended in Transposition buffer (10 mM Tris-Cl pH 7.4,

478   10% DMF, and 5 mM MgCl$_2$), and 1.5 μL of lab-made Tn5 transposase was added. After a 30

479   min incubation in a thermomixer at 37 °C, DNA was extracted using phenol:chloroform, PCR

480   amplified for 11 cycles, and sequenced on an Illumina NextSeq in 37x37 paired end mode.

481

482   **ATAC-seq data analysis.** Reads were aligned to the dm6 genome assembly using bowtie2[29] in

483   local mode, and only unique alignments were retained (mapq > 10). Signal was then divided into

484   two classes: hypersensitivity (paired end alignments with fragment size < 120 bp, which

485   represents hypersensitive chromatin and generates fragments smaller than mononucleosomes),

486   and mononucleosome (paired end alignments with fragment size 130–200 bp, which represents

487   two transposition events that roughly flank a mononucleosome sized region). Coverage tracks

488   were generated using deepTools[38]. For hypersensitivity signal, entire alignments were "piled up"

489   to generate coverage tracks, and for mononucleosome data only the central 3 bp of each

490   alignment were considered. ATAC-seq peaks were called using macs2[40]. See Supplementary

491   Table 3 for alignment metrics.

492

493   **3'RNA-seq.** 3'RNA-seq libraries were prepared using the QuantSeq 3' mRNA-seq Library Prep

494   Kit (Lexogen) with the UMI add-on kit. For each condition, 10^6 cells were added to a fixed

495   amount of ERCC Spike-In RNA Mix (Invitrogen), and RNA was extracted using TRIzol reagent

496   (Invitrogen). RNA treated with RNase free DNase I (Thermo Scientific), and the absence of

497   DNA was confirmed using the Qubit dsDNA-HS assay (Thermo Scientific). RNA quality was

498   confirmed using denaturing agarose gel electrophoresis. 3'RNA-seq libraries were prepared

499   using 325 ng of total RNA per condition according to manufacturer instructions and sequenced

500   on an Illumina NextSeq in 75 bp single end mode. Reads were trimmed of adapter and polyA

501   sequences and UMIs were extracted using fastp[36]. Reads were then aligned to a combined

502   dm6/ERCC reference genome using STAR[41], and reads mapped to the ERCC standards were

503   counted for spike-in normalization. Alignments were PCR-deduplicated using UMI-tools[37], and

504   only unique reads were retained (mapq = 255). The 5' ends of reads were used to generate signal

505   tracks (so that transcripts were scored in a read-length independent manner) using deepTools[38].

506   Spike-in normalization factors were calculated as described above for PRO-seq. See

507   Supplemental Table 4 for alignment metrics and normalization factors.

508

509   **CUT&RUN.** CUT&RUN was performed as described[42,43]. We used both anti-GAF (lab-made)

510   or anti-NURF301(Novus Biologicals Cat. No. 40360002) at a 1:10 dilution for the antibody

511   binding step. ProteinA-MNase was incubated with calcium on ice for 30 minutes, and cleaved

512   fragments were recovered by phenol:chloroform extraction. Library prep was performed using

513   the following steps: (i) Ends of digested fragments were repaired by incubation for 30 min at 25

514   °C with 0.5 U/μL T4 PNK, 0.12 U/μL T4 DNA Polymerase, and 0.05 U/μL Klenow Fragment in

515   1X T4 DNA ligase buffer (with ATP) and 0.5 mM dNTPs; (ii) Fragments were A-tailed by

516   incubation for 30 min at 37 °C with 0.25 U/μL Klenow exo- and 0.5 mM dATP in 1X NEBuffer

517   2; (iii) Adapters were added by incubation on the lab bench for 2 h with 4.38 nM lab-made

518   Illumina TruSeq forked adapters and 24 U/μL T4 DNA ligase in 1X T4 DNA Ligase buffer

519   (with ATP); (iv) library DNA was recovered using AMPure XP beads (1.8X concentration) and

520   PCR amplified for 15 cycles (all enzymes from New England Biolabs). Libraries were sequenced

521   on an Illumina NextSeq in 37x37 paired end mode. Adapters sequences were removed using

522   fastp[36], and reads were aligned to the dm6 reference genome using bowtie2[29]. Only uniquely

523   mapped reads (mapq > 10) with fragment size smaller than 120 bp were retained, and signal

524   coverage tracks were generated using deepTools[38]. Signal was normalized per million mapped

525   reads. See Supplementary Table 5 for alignment metrics.

526

527   **Reanalysis of Published Data.** GAF ChIP-seq[16] raw reads were downloaded and mapped to the

528   dm6 genome assembly using bowtie2[29], and only uniquely mapping reads were retained (mapq >

529   10). Single end reads were extended 200 bp and reads-per-million normalized coverage tracks

530   were generated using deepTools[38]. Peaks were called using macs2[40]. A M1BP ChIP-seq[23] signal

531   track was downloaded and converted for the dm6 assembly using liftOver[28], and signal was

532   normalized on a per-million basis. M1BP-knockdown PRO-seq[17] normalized signal tracks were

533   accessed and converted to the dm6 genome assembly as above. BEAF-32 ChIP-seq[25] raw reads

534   were downloaded, aligned using bowtie2[29], and only uniquely mapping reads were retained

535   (mapq > 10). Single end reads were extended 200 bp and reads-per-million normalized coverage

536   tracks were generated using deepTools[38]. See Supplementary Table 6 for accession numbers for

537   all published data used in this manuscript.

538

539   **DE testing**. Signal counting in each set of regions for each data type was performed using

540   functions from the BRGenomics package[44]. Differential expression testing and principal

541   component analysis was performed using DESeq2[34]. Genes with adjusted p-value < 0.01 were

542   considered differentially expressed. See Supplemental Code 6 for details.

543

544   **Browser shots.** Browser shots were generated using a custom R function, which can be found at

545   https://github.com/JAJ256/browser_plot.R (commit 1352d5c).

546

547   **Metaprofiles.** Metaprofiles were generated using the BRGenomics package[44] by calculating a

548   signal matrix across all regions in a set using the bin size specified, then sampling 10% of

549   regions 1000 times to calculate the mean and 75% confidence interval. In some cases, confidence

550   intervals were removed to avoid over-plotting. Visualization was performed using ggplot2[45].

551

552   **Motif analysis.** To search for motifs overrepresented in one set of promoters compared to
553   another, we used DREME[46] with an e-value threshold of 0.001.
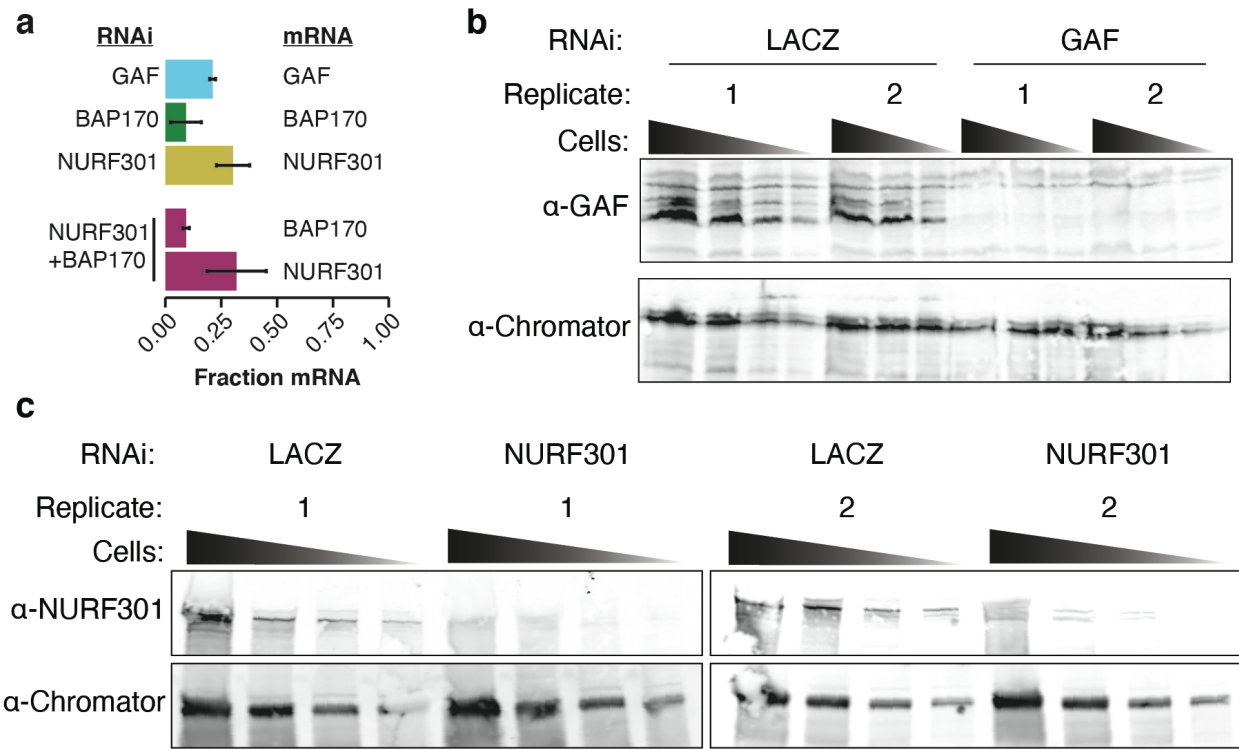554
555   **Classification of GAF-bound promoters.** We considered a promoter GAF-bound if the
556   promoter region (-500 to TSS) overlapped with a GAF ChIP-seq peak (see above). We then
557   considered these GAF-bound promoters as having GAF-dependent pausing or GAF-independent
558   pausing on the basis of whether or not they had significantly decreased PRO-seq in the pause
559   region compared to the LACZ-RNAi control (DESeq2 FDR < 0.01, log2 Fold Change < 0). We
560   further subdivided the GAF-bound promoters with GAF-independent pausing by whether they
561   were bound by M1BP, BEAF-32, or both, with "bound" defined as falling within the top 25% of
562   promoters in our total set of GAF-bound promoters with GAF-independent pausing when rank-
563   ordered by total ChIP-seq signal within the promoter (-500 to TSS) for a given factor. Heatmaps
564   were created using the ComplexHeatmap R package[47].
565

References (Methods)

27. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).

28. Haeussler, M. *et al.* The UCSC Genome Browser database: 2019 update. *Nucleic Acids Res.* **47**, D853–D858 (2019).

29. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).

30. Quinlan, A. R. & Hall, I. M. BEDTools: A flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**, 841–842 (2010).

31. Hoskins, R. A. *et al.* The Release 6 reference sequence of the Drosophila melanogaster genome. *Genome Res.* **25**, 445–458 (2015).

32. Thurmond, J. *et al.* FlyBase 2.0: the next generation. *Nucleic Acids Res.* **47**, D759–D765 (2019).

33. Kwak, H., Fuda, N. J., Core, L. J. & Lis, J. T. Precise maps of RNA polymerase reveal how promoters direct initiation and pausing. *Science.* **339**, 950–953 (2013).

34. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

35. Andrews, S. FastQC: A quality control tool for high throughput sequence data. (2010). Available at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

36. Chen, S., Zhou, Y., Chen, Y. & Gu, J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).

37. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* **27**, 491–499 (2017).

38. Ramírez, F., Dündar, F., Diehl, S., Grüning, B. A. & Manke, T. deepTools: a flexible platform for exploring deep-sequencing data. *Nucleic Acids Res.* **42**, W187–W191 (2014).

39. Buenrostro, J. D., Wu, B., Chang, H. Y. & Greenleaf, W. J. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* **109**, 21.29.1-21.29.9 (2015).

40. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, (2008).

41. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

42. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, (2017).

43. Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat. Protoc.* **13**, 1006–1019 (2018).

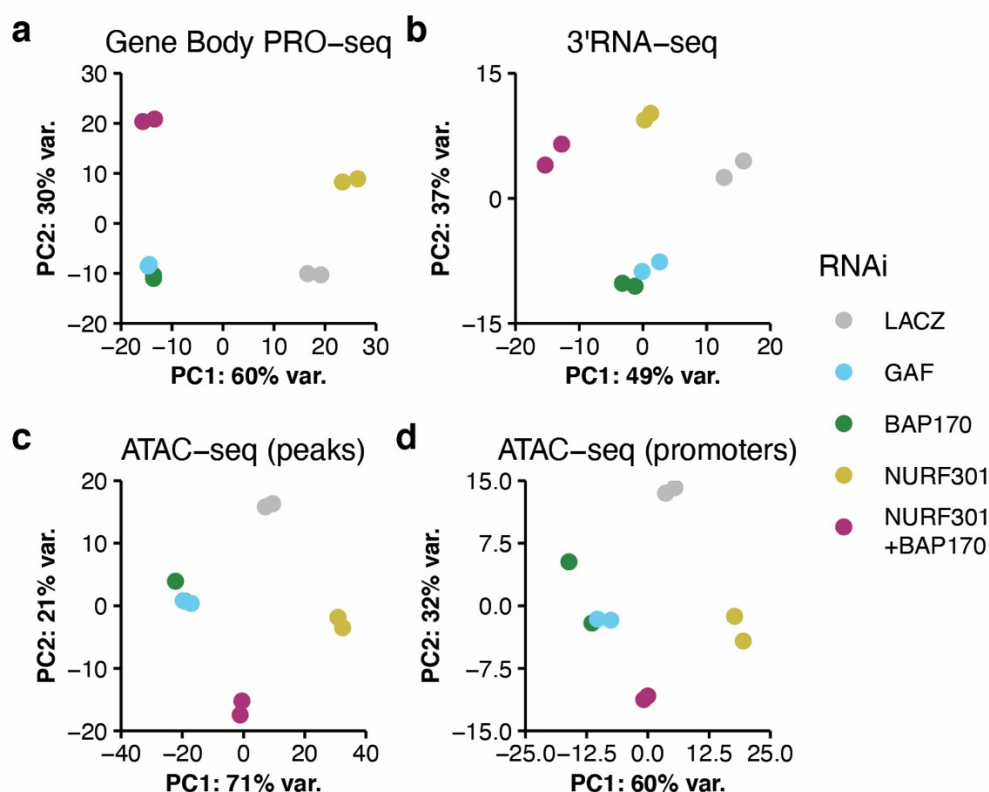44. DeBerardine, M. BRGenomics: Tools for the efficient analysis of high-resolution

603   genomics data. R package version 0.99.31. (2020). doi:10.18129/B9.bioc.BRGenomics

604   45.   Wickham, H. *ggplot2: Elegant Graphics for Data Analysis*. (Springer-Verlag, New York,

605   2016).

606   46.   Bailey, T. L. DREME: Motif discovery in transcription factor ChIP-seq data.

607   *Bioinformatics* **27**, 1653–1659 (2011).

608   47.   Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in

609   multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).

610

611 **Extended Data Figures**

612

613 **Extended Data Figure 1: Knockdown efficiency of GAF-RNAi and NURF301-RNAi by**
614 **western blot.**

615 **(a)** Fraction mRNA remaining compared to the LACZ control for each knockdown target. *BAP170*
616 and *NURF301* transcript levels are displayed independently for the double knockdown. Bar
617 height is mean; error bars are SD.

618 **(b)** Western blot showing the amount of GAF protein remaining after 5 days of RNAi treatment
619 compared to the LACZ-RNAi control. For each condition a serial 2-fold dilution series of cells
620 was loaded. Chromator is included as a loading control.

621 **(c)** Western blot showing the amount of NURF301 protein remaining after 5 days of RNAi
622 treatment compared to the LACZ-RNAi control. For each condition a serial 2-fold dilution series
623 of cells was loaded. Chromator is included as a loading control.

624

Extended Data Figure 2: Principal component analysis of gene body PRO-seq, 3'RNA-seq, and ATAC-seq.
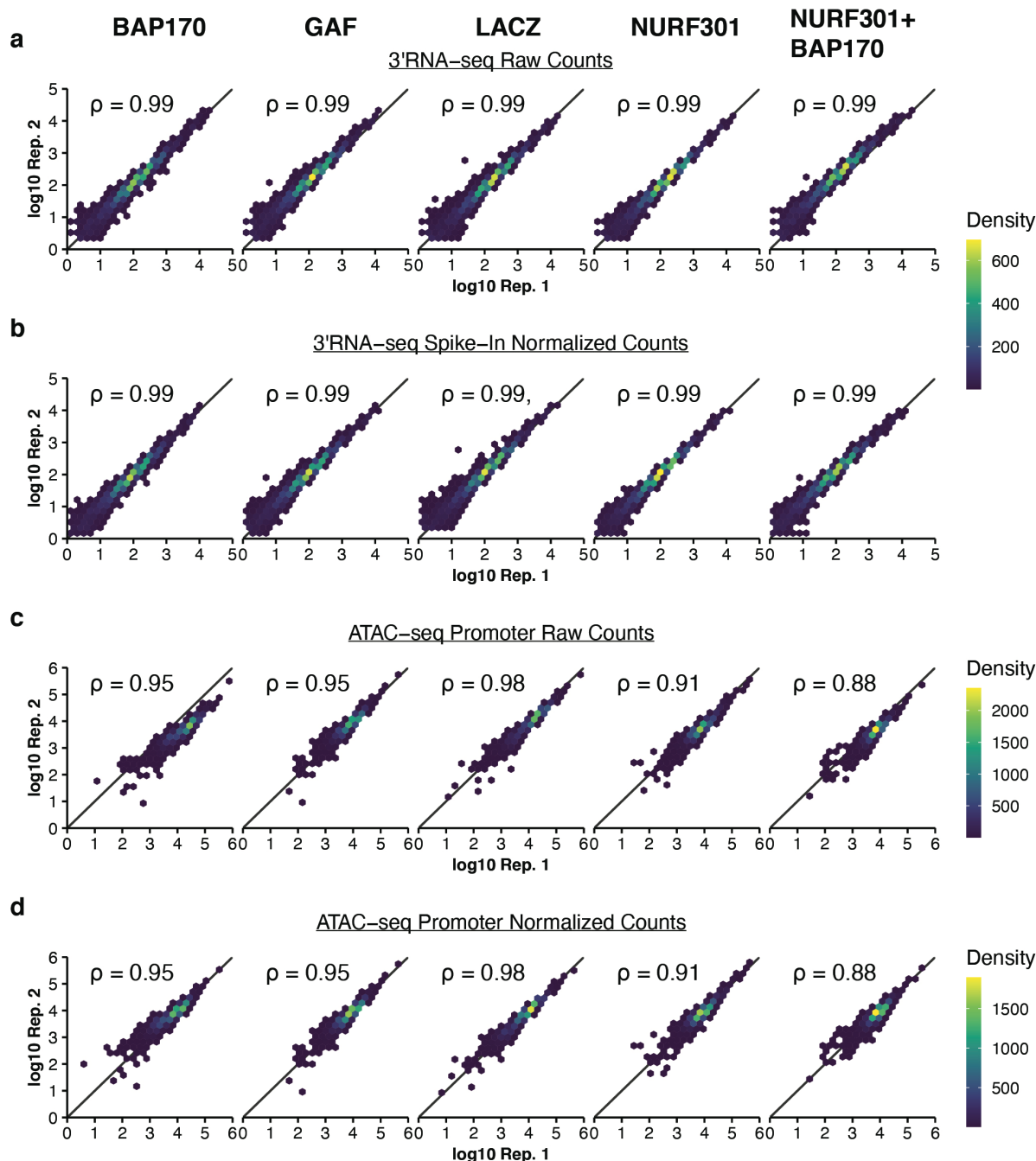
(a) Principal component analysis of spike-in normalized PRO-seq signal in the gene body (TSS+200 to TES-200) of genes in a filtered list (see methods, n=9,375).

(b) Principal component analysis of spike-in normalized 3'RNA-seq signal of genes in a filtered list (n=9,375), signal was counted in the last 1 kb region of each gene.

(c) Principal component analysis of library size-normalized ATAC-seq signal in ATAC-seq peak summits ± 100 bp (n=39,806).

(d) Principal component analysis of library size-normalized ATAC-seq signal in promoter regions (-1000–TSS) of genes in a filtered list (n=9,375).

637

Extended Data Figure 3: Correlation and normalization of PRO-seq data

(a) Raw number of PRO-seq reads mapped to the gene body (TSS+200 to TES-200) of each gene in a filtered list (n=9,375, see methods) for all conditions, with replicate 1 on the x-axis and replicate 2 on the y-axis. ρ is Spearman's rho. To accommodate overplotting, the plot space was divided into hexbins and color-mapped by the number of genes in each bin.

(b) As in **(a)**, but counts were normalized using spike-in scaling factors.

644   **(c)** As in **(a)**, but raw number of PRO-seq reads mapped to the pause region (-50 to +100) of each

645   gene.

646   **(d)** As in **(c)**, but counts were normalized using spike-in scaling factors.
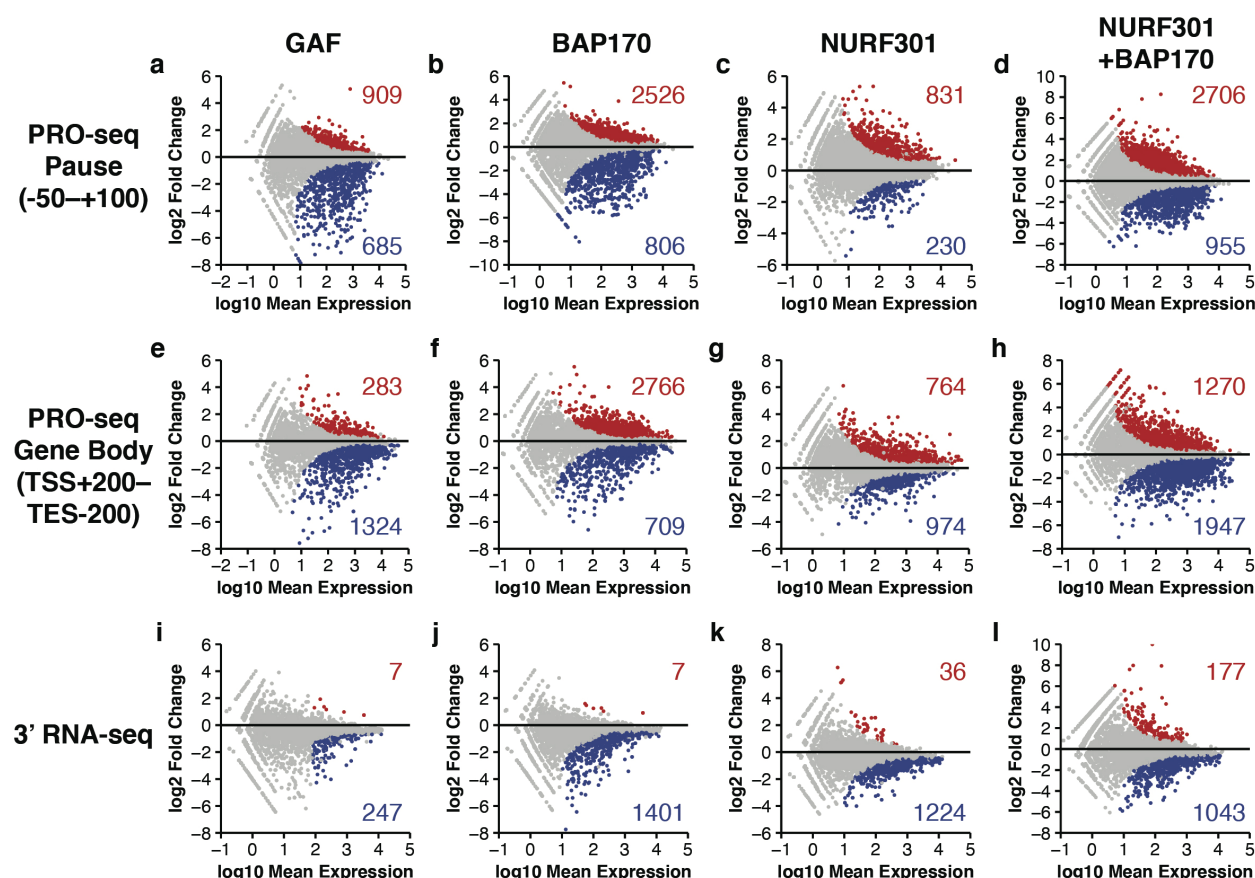
647

648

Extended Data Figure 4: Correlation and normalization of RNA-seq and ATAC-seq data

(a) Raw number of 3'RNA-seq reads mapped to the last 1 kb region of each gene in a filtered list (n=9,375, see methods) for all conditions, with replicate 1 on the x-axis and replicate 2 on the y-axis. ρ is Spearman's rho. To accommodate overplotting, the plot space was divided into hexbins and color-mapped by the number of genes in each bin.

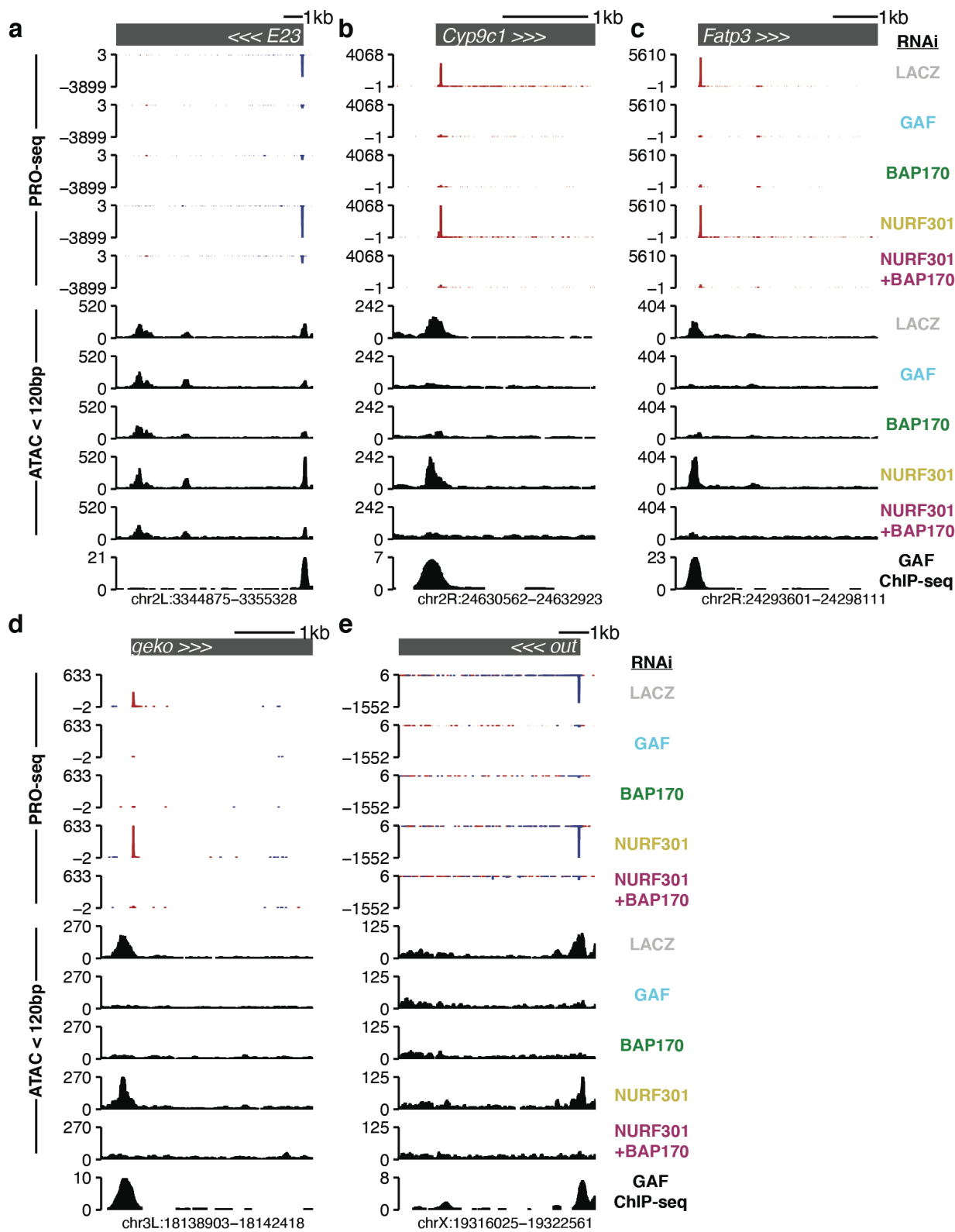**(b)** As in **(a)**, but counts were normalized using spike-in scaling factors.

655   **(c)** As in **(a)**, but raw number of ATAC-seq reads (< 120bp) mapped to the promoter region (-1000–
656   TSS) of each gene.
657   **(d)** As in **(c)**, but counts were normalized using library-size scaling factors (DESeq2).
658

659

Extended Data Figure 5: MA plots for all conditions and data types.

(a) MA plot for the comparison of spike-in normalized GAF-RNAi vs. LACZ-RNAi PRO-seq in the pause region (TSS -50 to +100); DESeq2 FDR < 0.01. Number of genes significantly up- or down-regulated is also shown.

(b) As in (a) but BAP170 RNAi vs. LACZ-RNAi.

(c) As in (a) but NURF301 RNAi vs. LACZ-RNAi.

(d) As in (a) but NURF301+BAP170 RNAi vs. LACZ-RNAi.

(e) MA plot for the comparison of spike-in normalized GAF-RNAi vs. LACZ-RNAi PRO-seq in the gene body (TSS+200–TES-200); DESeq2 FDR < 0.01. Number of genes significantly up- or down-regulated is also shown.

(f) As in (e) but BAP170 RNAi vs. LACZ-RNAi.

(g) As in (e) but NURF301 RNAi vs. LACZ-RNAi.

(h) As in (e) but NURF301+BAP170 RNAi vs. LACZ-RNAi.

(i) MA plot for the comparison of spike-in normalized GAF-RNAi vs. LACZ-RNAi 3'RNA-seq signal in the last 1 kb region of each gene; DESeq2 FDR < 0.01. Number of genes significantly up- or down-regulated is also shown.

(j) As in (i) but BAP170 RNAi vs. LACZ-RNAi.

677   **(k)** As in **(i)** but NURF301 RNAi vs. LACZ-RNAi.

678   **(l)**   As in **(i)** but NURF301+BAP170 RNAi vs. LACZ-RNAi.

679

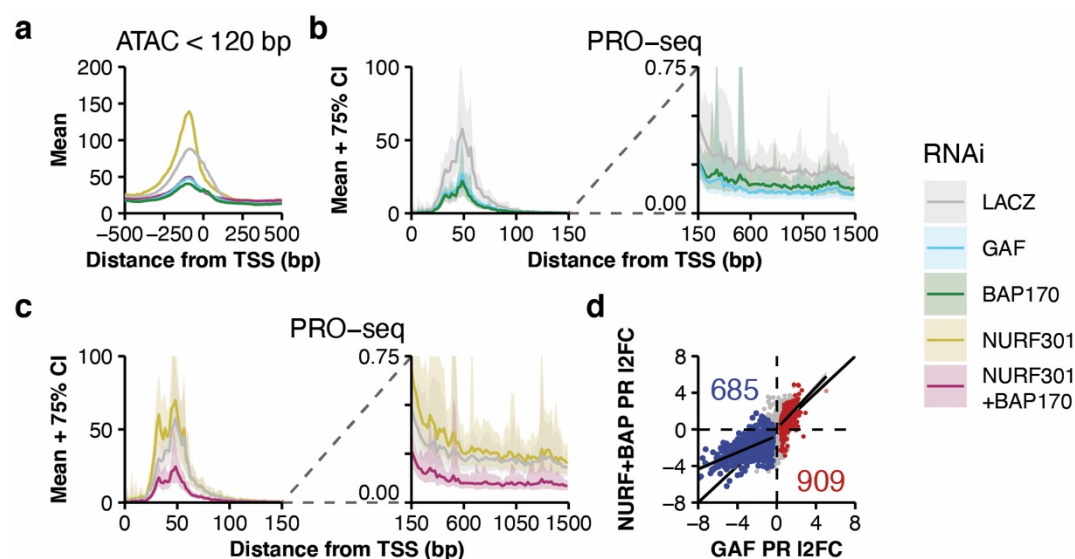681   **Extended Data Figure 6: Examples of GAF-dependent promoters.**

682   **(a)** Browser shot of a gene with GAF-dependent pausing (*E23-RC*). Tracks are each shown
683   as the mean of two replicates. ATAC-seq was filtered to retain only paired-end
684   alignments with insert size < 120 bp.

685   **(b)** As in **(a)** but *Cyp9c1-RA.*

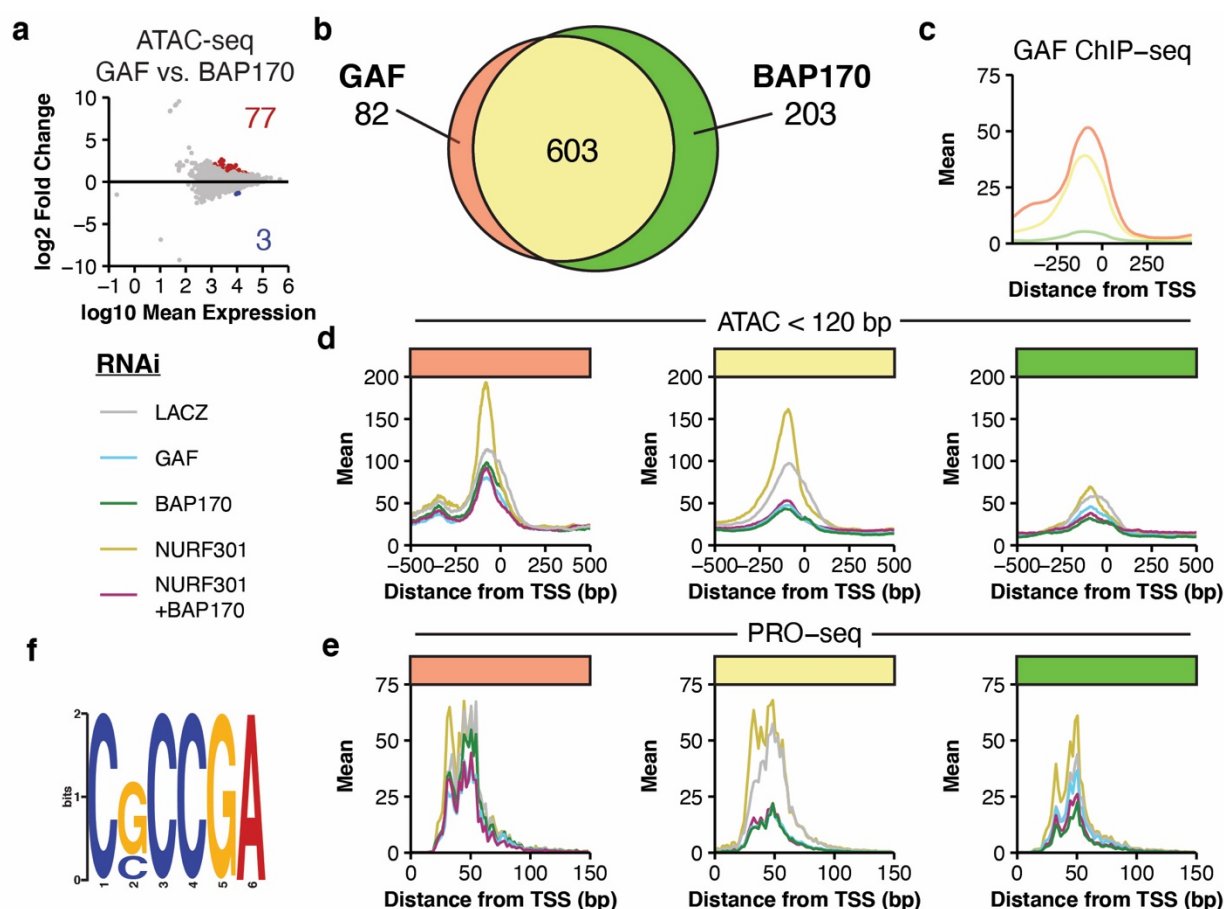686   **(c)** As in **(a)** but *Fatp3-RA.*

687   **(d)** As in **(a)** but *geko-RB.*
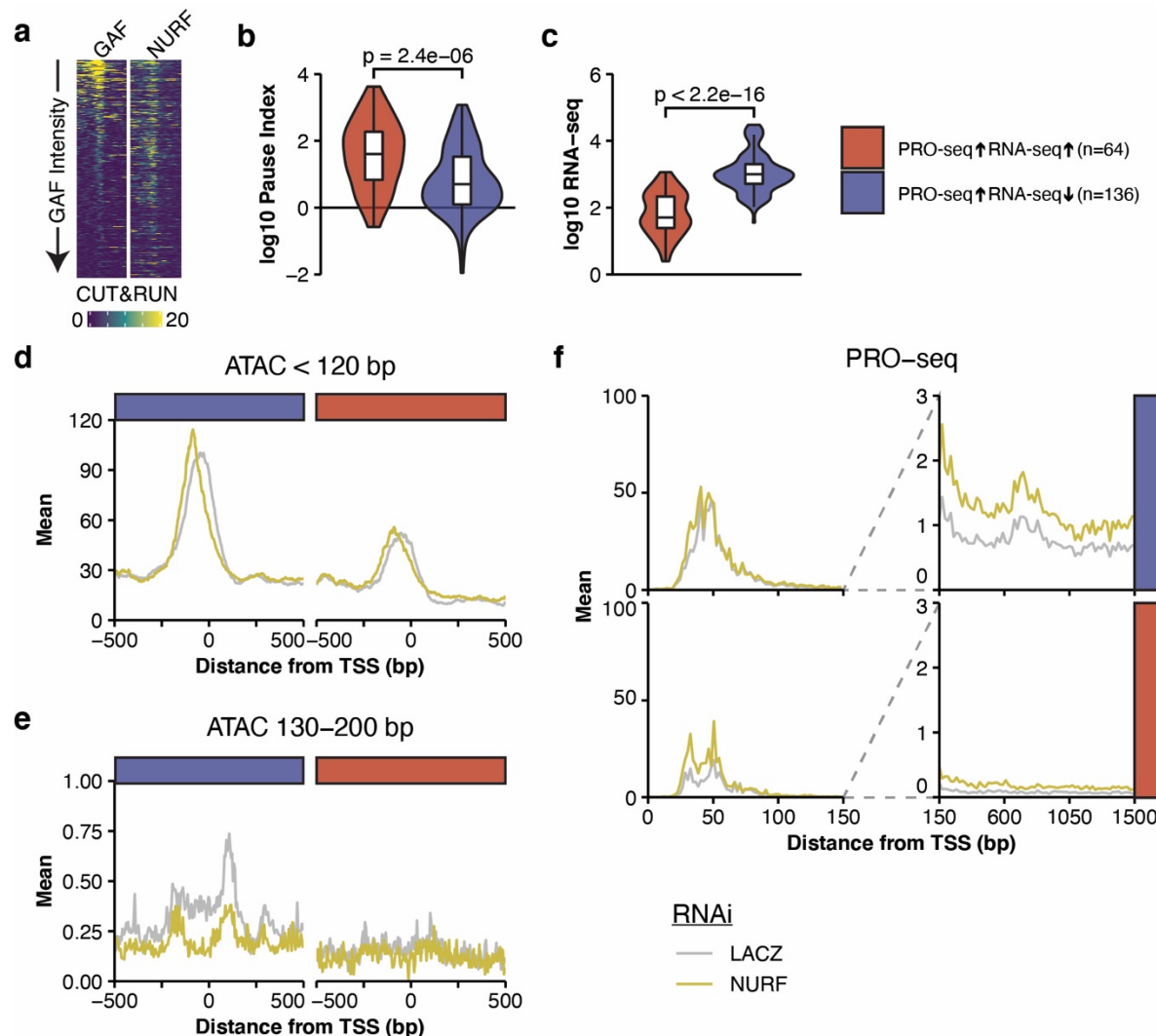
688   **(e)** As in **(a)** but *out-RA.*

689

Extended Data Figure 7: PBAP opens promoter chromatin in coordination with GAF.

(a) ATAC-seq (< 120 bp) signal for each RNAi treatment in 1 bp bins across all promoters with significantly decreased PRO-seq signal in the pause region (-50–+100) upon BAP170-RNAi treatment (n=806; see Extended Data Figure 5b; DESeq2 FDR < 0.01). Signal is the mean of 1,000 sub-samplings of 10% of regions.

(b) PRO-seq signal at genes with BAP170-dependent pausing (n=806) for the LACZ, GAF, and BAP170 RNAi treatments. The pause region (left panel) is in 2 bp bins, and the first 1.5 kb of the gene body (right panel) is in 20 bp bins. Data is shown as mean (solid line) and 75% confidence interval (shaded area), derived from 1,000 sub-samplings of 10% of regions.

(c) As in (b), but for LACZ, NURF301, and NURF301+BAP170 RNAi treatments.

(d) Pause region (TSS -50 to +100) PRO-seq log2 fold change after GAF-RNAi treatment compared to NURF301+BAP170-RNAi. Red and blue points are significantly up- or down-regulated upon GAF-RNAi treatment (DESeq2 FDR < 0.01). The line and shaded area are a GLM and 95% confidence interval fit to significantly up- or down-regulated genes.

708

Extended Data Figure 8: Effects of GAF and BAP-170 RNAi are indistinguishable.

(a) MA plot for the comparison of GAF-RNAi vs. BAP170-RNAi ATAC-seq (fragments < 120 bp) in ATAC-seq peak summits ± 100 bp (n=39,806); DESeq2 FDR < 0.01. Number of regions significantly up- or down-regulated is also shown.

(b) Intersection of promoters significantly downregulated by either GAF-RNAi (n=685) or BAP170-RNAi (n=806).

(c) Promoter GAF ChIP-seq signal in 5 bp bins at gene sets color coded as in (b).

(d) Promoter ATAC-seq (fragments < 120 bp) signal for each RNAi treatment in 1 bp bins at gene sets defined in (b). Signal is the mean of 1,000 sub-samplings of 10% of regions.

(e) Pause region PRO-seq signal for each RNAi treatment in 2 bp bins at gene sets defined in (b). Signal is the mean of 1,000 sub-samplings of 10% of regions.

(f) Motif enriched in BAP170 exclusive promoters (-500 to TSS; n=203; green in (b)–(e)) over GAF/BAP170 dependent promoters (-500 to TSS; n=603; yellow in (b)–(e)). DREME E-value < 0.001.

Extended Data Figure 9: NURF-RNAi causes bifurcated changes in nascent transcription and mRNA levels.

(a) GAF and NURF301 CUT&RUN signal at the top quartile of genes by GAF CUT&RUN signal (TSS ± 500). Rows are sorted by GAF signal and sort order is maintained.

(b) 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 Figure 3a for gene classes.

(c) Distribution of RNA-seq normalized counts for the two classes of genes described in (b).

(d) ATAC-seq (fragments < 120 bp) signal for NURF301 and LACZ RNAi treatments in 1 bp bins across promoters of the genes described in (b). Signal is the mean of 1,000 sub-samplings of 10% of regions.

737   **(e)** As in **(e)**, but ATAC-seq fragments 130–200 bp. Only the central 3 bp of each alignment
738   were considered when generating signal tracks.

739   **(f)** PRO-seq signal for NURF301 and LACZ RNAi treatments at genes described in **(b)**. The
740   pause region (left panel) is in 2 bp bins, and the first 1.5 kb of the gene body (right panel)
741   is in 20 bp bins. Signal is the mean of 1,000 sub-samplings of 10% of regions.

742   In **(b)** and **(c)** filled violins represent the distribution and boxplots show the median (center
743   line), 25% and 75% quartiles (hinges), and 1.5*IQR (whiskers). Outliers are not plotted, and
744   p-value is from a Mann-Whitney U test.