Published in final edited form as:

Nat Struct Mol Biol. 2022 July ; 29(7): 665–676. doi:10.1038/s41594-022-00800-z.

Kinetic principles underlying pioneer function of GAGA transcription factor in live cells

Xiaona Tang[1], **Taibo Li**[1,2], **Sheng Liu**[1], **Jan Wisniewski**[3], **Qinsi Zheng**[4], **Yikang Rong**[5], **Luke D. Lavis**[4], **Carl Wu**[1,6,✉]

[1]Department of Biology, Johns Hopkins University, Baltimore, MD, USA.

[2]Department of Biomedical Engineering, Johns Hopkins University, Baltimore, MD, USA.

[3]Experimental Immunology Branch, National Cancer Institute, Bethesda, MD, USA.

[4]Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA.

[5]State Key Laboratory of Bio-Control, Institute of Entomology, School of Life Sciences, Sun Yat-sen University, Guangzhou, China.

[6]Department of Molecular Biology and Genetics, Johns Hopkins School of Medicine, Baltimore, MD, USA.

Abstract

How pioneer factors interface with chromatin to promote accessibility for transcription control is poorly understood in vivo. Here, we directly visualize chromatin association by the prototypical GAGA pioneer factor (GAF) in live *Drosophila* hemocytes. Single-particle tracking reveals that most GAF is chromatin bound, with a stable-binding fraction showing nucleosome-like confinement residing on chromatin for more than 2 min, far longer than the dynamic range of most transcription factors. These kinetic properties require the full complement of GAF's DNA-

✉**Correspondence and requests for materials** should be addressed to Carl Wu. wuc@jhu.edu.

Author contributions

X.T. performed all genetic and imaging experiments with support from T.L., J.W. and Y.R. and all data analysis using R functions created by S.L. and X.T. L.D.L. and Q.Z. synthesized JF552/JFX554 and JF700. X.T. and C.W. designed the study and wrote the paper with input from all authors.

Competing interests

L.D.L. and Q.Z. are listed as inventors on patents and patent applications whose values might be affected by publication. The remaining authors declare no competing interests.

Code availability

Custom R package Sojourner used to process and analyze SPT trajectories is available at https://github.com/sheng-liu/sojourner. Python codes for running Spot-On were adapted from https://gitlab.com/tjian-darzacq-lab/Spot-On-cli to analyze SPT trajectories. MATLAB codes for running vbSPT were adapted from https://gitlab.com/anders.sejr.hansen/anisotropy to classify trajectories.

Extended data are available for this paper at https://doi.org/10.1038/s41594-022-00800-z.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41594-022-00800-z.

Peer review information *Nature Structural & Molecular Biology* thanks Mounia Lagha and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. Peer reviewer reports are available. Primary Handling Editor: Carolina Perdigoto, in collaboration with the *Nature Structural & Molecular Biology* team.

Reporting summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

binding, multimerization and intrinsically disordered domains, and are autonomous from recruited chromatin remodelers NURF and PBAP, whose activities primarily benefit GAF's neighbors such as Heat Shock Factor. Evaluation of GAF kinetics together with its endogenous abundance indicates that, despite on–off dynamics, GAF constitutively and fully occupies major chromatin targets, thereby providing a temporal mechanism that sustains open chromatin for transcriptional responses to homeostatic, environmental and developmental signals.

Drosophila GAGA pioneer factor (GAF), a ubiquitous and essential Zn finger transcription factor (TF) encoded by the Trithorax-like (*Trl*) gene, is a multimeric protein complex that binds specifically to clusters of adjacent GAGAG sequences on numerous genes, including homeotic, steroid- and heat shock-response genes[1–4]. GAF regulates transcription by interactions with the TAF3 and TAF4 components of the TFIID general TF[5–7], the NELF elongation factor[8], and antagonism to histone H1-mediated transcriptional repression[9]. In addition, as a pioneer TF[10,11], GAF is capable of binding to reconstituted nucleosomes[12], directly recruiting chromatin remodelers NURF, PBAP and other factors[6,13–15] to create accessible chromatin for neighboring factors[16] and assembly of the paused RNA Polymerase II (Pol II)[17,18]. Pioneer TFs[19] appear to possess special nucleosome-binding properties defined by their modes of DNA binding that generate chromatin accessibility to benefit the subsequent binding of neighboring TFs (nonpioneers or settlers)[16,20–22]. The beneficiaries of GAF's pioneering activity include *Drosophila* heat shock factor HSF[23], coactivator CBP[23,24], Polycomb repressor PHO (Pleiohomeotic)[25] and the insulator binding Large Boundary Complex[26], whose interactions with genomic targets are all in close proximity to GAGAG elements and are facilitated by GAF.

Genome-wide analysis reveals that GAF is enriched in vivo at promoters as well as distal *cis*-regulatory regions comprising several thousand targets that often include clusters of tandem GA repeats[17,27,28]. GAF-specific RNA or protein depletion experiments have demonstrated the in vivo function of GAF in generating chromatin accessibility in the *Drosophila* embryo and in cultured cells[17,18,28,29]. Despite these advances over decades of research, unifying principles for pioneering activity of TFs such as GAF have remained elusive. How pioneer factors differ from other sequence-specific TFs and the kinetic mechanisms by which pioneers perform their key genomic functions are subjects of continuing debate.

To explain the underlying mechanisms of this pleiotropic TF, we have studied the kinetics of GAF diffusion in the nucleoplasm and on genomic chromatin by single-particle tracking (SPT) in live *Drosophila* hemocytes under different genetic contexts. In comparison, we also measured the kinetic parameters for HSF at normal and heat-stressed conditions. We then determined GAF and HSF protein levels in hemocytes, curated existing databases for numbers of genomic targets and integrated these parameters with the measured kinetics to obtain the target occupancy of each factor in vivo. Our findings uncover crucial quantitative principles for pioneering and maintaining chromatin accessibility over extended time periods, even when the responsible factors continuously bind to and dissociate from their chromatin targets in a dynamic manner.

Results

GAF chromatin-binding fractions and long residence times.

Drosophila Trl encodes two isoforms of GAF that harbor the N-terminal POZ/BTB domain (hereafter called POZ), the central zinc finger (ZF)-containing DNA-binding domain (DBD) and long and short C-terminal Q-rich domains (Fig. 1a)[4,30]. We used CRISPR–Cas9-based gene editing to insert a HaloTag at the N terminus of endogenous *Trl* (Fig. 1a and Supplementary Fig. 1a–d). The Halo knock-in strain is homozygous-viable, expressing long and short Halo-GAF isoforms (GAF^L and GAF^S) as the sole source. Halo-GAF binds to numerous loci on salivary gland polytene chromosomes, consistent with immunostaining studies[12,30,31] (Fig. 1b), and appears in multiple nuclear foci in diploid circulating larval hemocytes (>90% plasmatocytes, counterpart of mammalian macrophages[32]), similar to nuclear foci in the S2 cell line[33] (Fig. 1c). GAF puncta have also been observed in blastoderm nuclei through the nuclear cycles, although the polarized, apical distribution of larger GAF foci does not appear to be recapitulated in hemocytes[34,35]. We also constructed transgenic flies expressing C-terminal tagged isoforms GAF^L-Halo and GAF^S-Halo, which exhibit tissue-specific expression[30] (Supplementary Fig. 1f) and are functionally active, as indicated by rescue of Trl^{13C}/Trl^{R67} lethal alleles[3]. Together, the data demonstrate that fusion of HaloTag does not interfere with the localization and essential functions of GAF.

We investigated the live-cell dynamics of tagged GAF species in live hemocytes (Extended Data Fig. 1a) by SPT, using a 'fast-tracking' regime (10 ms per frame) in dSTORM mode[36–38] to measure slow- and fast-diffusing molecules (Fig. 1d and Supplementary Video 1), and quantified diffusion coefficients with a robust, displacement-based, analytical protocol (Spot-On)[39] (Fig. 1e and Extended Data Fig. 1b). All Halo-tagged GAF versions display similar slow and fast diffusivities that are within a similar dynamic range, as demonstrated by histone H2B (Fig. 1e and Extended Data Fig. 1c), with >75 and >85% slow-diffusing fraction (chromatin bound, F_{bound}), respectively. The diffusion coefficient (D) of the bound fraction ($D_{\text{bound}} = 0.004 - 0.005\ \mu m^2\ s^{-1}$) is two orders of magnitude lower than the free fraction ($D_{\text{free}} = 0.75 - 0.8\ \mu m^2\ s^{-1}$) (Extended Data Fig. 1e,f).

A 'slow-tracking' regime (500 ms per frame) to motion-blur fast particles allowed selective detection of long- and short-lived chromatin-bound populations (Extended Data Fig. 2a–c and Supplementary Video 2). We calculated 1-CDF of dwell times to generate survival curves demonstrating apparent GAF dissociation over time. Halo-GAF and GAF^L-Halo show similar profiles, with a slightly faster decay for the GAF^S-Halo isoform (Extended Data Fig. 2d). Fitting to a double exponential function (Extended Data Fig. 2e) enabled calculation of long- and short-lived (called stable and transient) binding fractions, f_{sb} and f_{tb}, and average stable and transient residence times, τ_{sb} and τ_{tb}, after correction for photobleaching and out-of-focus chromatin motions using Halo-H2B as a control (Extended Data Fig. 2d). The stable-binding fraction f_{sb} multiplied by total binding fraction F_{bound} gives the overall stable-binding fraction $F_{\text{sb}} = 30 - 35\%$ (Fig. 1f and Extended Data Fig. 1d). All stable-binding GAF fractions display protracted residence times τ_{sb}: 130 s for Halo-GAF, 141 s for GAF^L-Halo and 85 s for GAF^S-Halo, which is 20- to 30-fold longer than the transient residence time τ_{tb} of roughly 4 s (Fig. 1g and Extended Data Fig. 2f), and longer

than τ_{sb} values measured for many mammalian TFs[21,38,40–43]. Stable- and transient-binding are assumed to occur at cognate and nonspecific sites, respectively.

Three domains of GAF contribute to stable chromatin binding.

Purified, bacterially expressed GAF associates in oligomeric complexes ranging from monomers to complexes as large as decamers, with a peak at hexamers[44,45], and large GAF complexes are also observed in nuclear extracts of S2 cells[46]. Multimerization is mediated by the POZ and Q-rich domains of GAF[44–46]. To investigate how GAF domains contribute to particle dynamics, we used CRISPR–Cas9 to engineer Halo-GAF deletions in the POZ domain (*ΔPOZ*), the zinc finger (*ZF*^{9}; *ZF*^{10}) and the long and short Q-rich domains (*ΔQ*) (Fig. 2a and Extended Data Fig. 3a–e).

Homozygous *ΔPOZ* and *ZF*^{9} or *ZF*^{10} mutants arrest at early pupal and early third instar larval stages, respectively, while homozygous *ΔQ* is 70% viable, producing infertile adults with impaired longevity (Fig. 2a). GAF has been shown to be important for the maternal to zygotic transition in early embryogenesis[29]. The phenotypes of homozygous GAF mutants manifest later in development because their heterozygous parents harbor the untagged wildtype (WT) gene on the balancer chromosome. Thus, homozygous progeny carrying the GAF mutant alleles would retain WT untagged parental GAF RNA/protein that perdures through early development. Our results indicate that all three GAF domains are essential for *Drosophila* viability, with *ZF*^{9} and *ZF*^{10} showing the most severe phenotype.

SPT using fast and slow-tracking regimes for Halo-GAF *ΔPOZ*, *ZF*^{9}, *ZF*^{10} and *ΔQ* mutants in third instar larval hemocytes found that all three domain mutants display substantial reductions in the stable-binding fraction F_{sb} and residence time τ_{sb} for the slow-diffusing fraction (Fig. 2b,c and Extended Data Fig. 4a–c). Disruption of the POZ domain reduces F_{sb} from 29 to 15%, and τ_{sb} from 130 to 42 s, demonstrating the important contribution of this multimerization domain to stable chromatin association. Deletion of Q-rich domains shows a modest reduction of F_{sb} from 29 to 24%, and reduces the τ_{sb} from 130 to 49 s. Under fast tracking, *ΔPOZ* and *ΔQ* proteins exhibit similar D_{free} but larger D_{bound} compared to WT GAF, suggesting a more diffusive binding mode (Extended Data Fig. 4d,e). These results are consistent with the specific binding patterns of *ΔPOZ* and *ΔQ* on fixed polytene chromosomes (Fig. 2d)

The zinc finger mutants *ZF*^{9} and *ZF*^{10} also show reductions in stable-binding fraction F_{sb} from 29 to 19 and 22%, and residence time τ_{sb} from 130 to roughly 40 s in hemocytes (Fig. 2b,c and Extended Data Fig. 4a–c), similar to *ΔPOZ* and *ΔQ*. However, unlike *ΔPOZ* and *ΔQ*, both *ZF*^{9} and *ZF*^{10} mutants exhibit loss of the specific binding pattern on polytene chromosomes and increase of nucleoplasmic distribution (Fig. 2d). These results confirm the crucial function of the DNA-binding zinc finger for site-specific chromatin binding, and report aberrant, mechanistically unclear diffusive behavior of *ZF*^{9} and *ZF*^{10} mutant proteins. This aberrant behavior is also reflected in the time-averaged mean-squared displacement (MSD) curves from slow tracking of *ZF*^{9} and *ZF*^{10}, which show an initial steep rise followed by a plateau after 10 s, a profile dramatically different from the Brownian behavior of WT GAF, *ΔPOZ, ΔQ* and H2B HaloTag fusions over a 60 s timescale (Extended Data Fig. 4f).

As an orthogonal approach to assess the residence times measured by SPT, we performed fluorescence recovery after photobleaching (FRAP) experiments in live hemocytes (Extended Data Fig. 4g,h). The half-recovery time for WT Halo-GAF is 443 s, and reduces to 91 s for *ΔPOZ*, 381 s for *ΔQ* and 48 s for ZF^{9}. Compared to SPT, residence times measured by FRAP trends similarly between WT and mutant GAFs but generally provide larger values, consistent with previous observations that modeling residence times by FRAP is a more indirect approach and tends to overestimate TF residence times[43,47]. A recent study of GAF dynamics using FRAP and fluorescence correlation spectroscopy in the syncytial embryo that contains a large GAF pool for exchange reported a residence time of roughly 1 min (ref. [35]), in range of our roughly 2 min value determined by SPT.

GAF binding is independent of recruited chromatin remodelers.

At cognate GAGAG sites on *Drosophila* chromatin, GAF recruits NURF and PBAP, ATP-dependent chromatin remodelers of the ISWI and SWI/SNF families, respectively, to drive DNA accessibility for neighboring TFs and establish promoter-proximal paused Pol II (refs. [13,17,23,28,48,49]). This process begins during *Drosophila* embryogenesis, when GAF and the Zelda pioneer factor are individually required to activate and remodel the chromatin accessibility landscape for widespread zygotic transcription[29,33]. However, it was unclear whether the recruitment of NURF and PBAP by GAF is required to assist its own chromatin binding as a pioneer factor. To address this, we performed SPT of GAF^{L}-Halo and GAF^{S}-Halo on third instar larval hemocytes isolated from *bap170* and *nurf301*/*E(bx)* mutants for unique subunits in PBAP[49,50] and NURF[23,49] complexes, respectively (Fig. 3a). The results show little or small changes in the F_{bound}, D_{bound} and τ_{sb} values of the mutants (Fig. 3b,c and Extended Data Fig. 5a–f). GAF^{L}-Halo and GAF^{S}-Halo isoforms also show no qualitative global binding changes on polytene chromosomes in the *bap170* and *nurf301* mutants, although changes in a minority of chromosomal loci might escape detection (Fig. 3d) (by contrast, changes of HSF-Halo binding can be detected in mutants, see below). These findings are generally consistent with ChIP–seq studies showing similar average GAF-binding genome-wide in PBAP-depleted S2 cells (Extended Data Fig. 6), although partial GAF binding is observed in a subset of *Drosophila* promoters (685 promoters displaying reductions in paused RNA Pol II and chromatin accessibility)[28]. Taken together, our live-cell SPT results indicate that GAF chromatin binding and dwell time are largely autonomous from NURF and PBAP, although other remodeling activities are not excluded[51]. We conclude that the main beneficiaries of NURF and PBAF recruitment and nucleosome remodeling are factors such as HSF, CBP, PHO and Large Boundary Complex[23–26] that bind close to GAGAG elements at genomic targets.

Heat shock increases HSF binding but not residence time.

The recruitment of chromatin remodelers by GAF increases accessibility to facilitate binding of HSF to the tripartite heat shock element (HSE) adjacent to GAF-binding sites at heat shock promoters[18,52–55]. Under normal conditions, HSF is predominantly monomeric, with low (submicromolar) affinity of its winged-helix DBD for an NGAAN sequence[56,57]; heat shock induces HSF trimerization and juxtaposition of three DBDs for high-affinity binding to HSEs containing triple NGAAN sequences in alternating orientation[58–63].

The kinetics of HSF monomer and HSF trimer diffusion in the nucleoplasm and association with chromatin has not been investigated in diploid cells. To measure HSF dynamics, we constructed a transgenic HSF-Halo strain under natural expression control, and verified the functionality of HSF-Halo by rescue of *P{PZ} Hsf*^{03091}/*Hsf*^{3} lethal alleles[64,65]. We further validated HSF-Halo functions by confocal imaging of fixed polytene nuclei, which showed that HSF-Halo is mostly nucleoplasmic at room temperature (RT), except for low binding to few sites including *Hsp83* harboring very high-affinity HSEs; heat shock at 37.5 °C for 10 or 30 min induced strong HSF binding to many more chromosomal loci, most prominently reported at *Hsp* genes[58,66] (Extended Data Fig. 7a). This inducible pattern of HSF binding on heat shock is partially reduced in mutants for *Trl*, *Bap170* and *Nurf301* (Extended Data Fig. 7a) (note that there are GAF-independent HSF targets in the genome[18], for which changes of HSF binding at corresponding polytene loci would not be expected). Imaging of fixed hemocytes shows that the heterogeneous distribution of HSF-Halo changes on heat shock to a more punctate pattern including several prominent foci, consistent with previous studies (Fig. 4a)[58,67,68].

We performed live-cell SPT on HSF-Halo in fast- and slow-tracking modes in the *P{PZ}Hsf*^{03091}/*Hsf*^{3} genetic background, using hemocytes cultured at RT or heat shocked at 37.5 °C (Fig. 4b). As expected, the overall binding F_{bound} of HSF from fast tracking increases substantially from 24.9 to 44.3% on heat shock (Fig. 4c and Extended Data Fig. 7b). Two-component exponential decay fitting of the HSF-Halo survival curves derived from slow tracking reveals a substantial increase of stable-binding fraction F_{sb} from 5.9 to 14.2% on heat shock with no measurable change of residence time τ_{sb} (47 s) (Fig. 4c,d and Extended Data Fig. 7c,e). Thus, heat shock elevates the stable chromatin-binding HSF trimer fraction without affecting the dissociation rate (inverse of residence time), suggesting that the limited stable binding at RT ($F_{\text{sb}} = 5.9\%$) is due to low-level trimerization[63]. Distinct from HSF dynamics, GAF shows a small overall reduction in F_{bound} on heat shock (from 77.3 to 69.6%), and similarly for the stable-binding fraction F_{sb} (from 29.4 to 22.1%) (Fig. 4c). The residence time τ_{sb} for GAF remains unchanged after heat shock (Fig. 4d and Extended Data Fig. 7d,e).

We note that the residence time for HSF has been investigated by FRAP on *Drosophila* polytene chromosomes, reporting 15 s and >6 min half-recovery times at non-shock and heat-shock conditions, respectively[66]. The differences between HSF half-recovery times on polytene chromosomes and residence times measured by SPT in hemocytes could be due to the high multiplicity and lateral alignment of binding targets on polytene chromatin.

Chromatin-bound GAF displays H2B-like confinement.

The diffusion coefficient of chromatin-bound HSF measured by fast tracking (D_{bound}, average of both stable- and transient-binding) at RT where HSF monomers predominate is roughly fourfold greater than that of HSF trimers induced by heat shock ($D_{\text{bound}} = 0.075$ versus 0.019 $\mu m^2\ s^{-1}$) (Fig. 4e). HSF monomers also exhibit more than tenfold larger D_{bound} values than GAF ($D_{\text{bound}} = 0.0046\ \mu m^2\ s^{-1}$) (Fig. 4e), indicating that a single DBD is more diffusive on chromatin than multiple DBDs. D_{bound} for GAF approaches the H2B

value (D_{bound} = 0.0020 μm^2 s^{-1}) (Fig. 4e). To strengthen these findings, we analyzed particle trajectories with vbSPT, a variational Bayesian hidden Markov model (HMM) algorithm that assigns bound and free diffusive states to individual particle displacements of each trajectory[69,70]. We classified particle trajectories as either 'bound' or 'free', excluding a small fraction showing two-state diffusivity (Extended Data Fig. 8a–d). The time-averaged MSD plots of the bound particles confirm that bound molecules move in small confined regions (Fig. 4f and Extended Data Fig. 9a,b), while free molecules undergo Brownian motion (Extended Data Fig. 9c,d). The MSD plot of chromatin-bound GAF trajectories reaches a plateau at low values, resembling that of H2B (Extended Data Fig. 9a), while HSF plateaus at higher values (Fig. 4f). The radius of confinement (R_c) gives median values for HSF monomers (0.13 μm), HSF trimers (0.10 μm), GAF multimers (0.07 μm) and H2B (0.06 μm) (Extended Data Fig. 9e). In contrast, the MSD plots of HSF (both RT and 37.5 °C) from slow tracking (reflecting chromatin movements) show similar Brownian motion features as H2B, although on a shorter timescale, possibly owing to differential motion of HSF targets from bulk chromatin (Extended Data Fig. 9f). Together, the results indicate that chromatin-bound GAF is nearly as constrained as nucleosomal histones. Activated HSF trimers are less constrained than GAF, possibly due to fewer DBDs per complex and/or higher local chromatin mobility, but bound HSF monomers are more diffusive, consistent with the presence of only a single DBD.

High temporal occupancy defines pioneering activity.

The steady-state open chromatin landscape at promoters and enhancers featuring nucleosome-depleted regions genome-wide[51,71,72] belies highly dynamic interactions with transcription and chromatin factors[73–78]. The establishment and maintenance of chromatin accessibility, that is the sustained opening of chromatin, requires the joint activities of sequence-specific DNA-binding factors and ATP-dependent remodeling enzymes[12,23,28,29,79,80], the latter proteins interacting with chromatin with a lifetime of seconds[73,75,76]. GAF directs pioneering functions not only by virtue of its affinity for nucleosomal DNA targets[12,81] but also recruitment of chromatin remodelers[12,14,15,28,48,82]. Given the highly transient association and variable occupancy levels displayed by remodelers[73,76], we hypothesized that GAF should instead sustain high occupancy along with protracted dwell time to continuously maintain open chromatin at cognate targets.

Temporal occupancy, the percent time of any duration for which a cognate site is factor-bound, depends on the number of GAF molecules ($N_{monomers}$) per cell and the number of target sites (N_{sites}) in the genome. In the context of the facilitated diffusion model[83] in which TFs experience three-dimensional (3D) nucleoplasmic diffusion, nonspecific binding, 1D diffusion, dissociation and rebinding until site-specific chromatin engagement, temporal occupancy is also dependent on the kinetics of target search and dissociation. Integration of our kinetic data from SPT with published genomic data allows calculation of temporal occupancy for GAF.

ChIP–seq identifies 3,622 high-confidence GAF peaks from the hemocyte-like S2 cell line[17,84,85]. Similar numbers of GAF peaks are found by ChIP–seq analysis of larval imaginal tissues and embryos although the peaks from different cell types show partial

overlap[29,86]. We measured GAF abundance by fluorescence flow cytometry using a calibrated CCCTC-binding factor (CTCF) standard[87], estimating $N_{\text{monomers}} = 56{,}683 \pm 6{,}025$ GAF molecules per hemocyte (Extended Data Fig. 10a). Given that circulating hemocytes are largely in the G2 cell-cycle phase (Extended Data Fig. 10b,c), we estimate $N_{\text{sites}} = 3{,}622 \times 4$ genome copies = 14,488. From the overall chromatin-binding fraction (F_{sb}), stable and transient residence times (τ_{sb}, τ_{tb}) and fractions for stable and transient chromatin binding (f_{sb}, f_{tb}), we derived the average search time ($\tau_{\text{search}} = 150$ s, the time from GAF dissociation from one stable target to association with the next) and the sampling interval (71.5 s, the time from the start of one stable-binding event to the next stable event on the same chromatin target; Methods), assuming that GAF binds stably at specific sites and transiently elsewhere (Fig. 5a,b).

We calculated the occupancy (τ_{sb}/sampling interval (SI)) for an average GAF target to be 182% under nonshock conditions and at 154% after heat shock, which indicates full target occupancy by at least one or multiple GAF subunits for any time period, assuming no change in the number of GAF molecules and targets. For the 3,622 GAF peaks, whose intensities are correlated with the number of GAGAG sequences (each nonoverlapping GAGAG sequence counted as one element), roughly 65% harbor more than two GAGAG elements, with median peak intensity rising to a plateau at 6–7 clustered elements[17], for example at *ubx*, *engrailed*, *E74*, *eve* and *Hsp* genes[4] (Fig. 5c). The results indicate that GAF binds with a distribution tending toward large oligomers for peaks showing high ChIP–seq signals, which can be attributed to the cooperative binding of GAF, as demonstrated in vitro[44,45]. At this subset of highly enriched sites, GAF may bind as a multimeric complex of six or more subunits with essentially full temporal occupancy despite factor on–off dynamics for any time period in which GAF levels and the number of GAF targets remain constant (Fig. 6). The average occupancy for *ΔQ* ($N_{\text{monomers}} = 66{,}206 \pm 3{,}064$) remains similar at 177% but is reduced to 89% for *ΔPOZ* ($N_{\text{monomers}} = 44{,}907 \pm 14{,}154$) (Fig. 5b).

We note that many GAF-binding peaks overlap with Pipsqueak, a related POZ-domain TF[88,89] and partially with CLAMP, another GA repeat-binding pioneer factor in *Drosophila*[90,91]. The overall site occupancy at GAF locations on chromatin is therefore likely to be further supplemented should the contributions of Pipsqueak and CLAMP, which are both expressed in hemocytes[92], be quantified.

For HSF, we determined $N_{\text{monomers}} = 9{,}543 \pm 613$ for a sole source, transgenic HSF-Halo under natural expression control in the *P{PZ} Hsf*^{03091}/*Hsf*^{3} background (Extended Data Fig. 10a). A similar calculation for HSF-Halo binding to 442 genomic sites after heat shock[54] gives a search time (τ_{search}) = 119 s and a sampling interval of 31 s, which results in an average HSF occupancy of 153%, or 51% for HSF trimers as the predominant species induced after 10–40 min heat shock (Fig. 5b). At highly enriched HSF locations such as the major *Hsp* genes harboring several HSEs, it follows that one or more HSF trimers may occupy the promoter 100% of the time on full induction to release the paused RNA Pol II and rapidly recruit additional enzymes for a burst of transcription until system attenuation (Fig. 6).

Discussion

Three GAF domains promote stable chromatin association.

Our SPT of GAF in live hemocytes reveals that GAF binds chromatin with an exceptionally long residence time τ_{sb} of roughly 2 min compared to other factors. Systematic mutagenesis showing that the DBD, the POZ multimerization domain and the Q-rich domains are required for viability also found that not just the DNA-binding zinc finger, but all three GAF domains are required for its stable chromatin association and long residence time τ_{sb} (Fig. 2b,c). This includes the Q-rich intrinsically disordered region domains that may promote self-interactions to stabilize the GAF multimer[46,93] or interactions with other basal TFs recruited to GAF targets. Of interest, the zinc finger mutants ZF^{9} and ZF^{10} have a substantial residual slow-diffusing fraction in hemocytes, but show no specific binding on polytene chromosome bands (Fig. 2b,d). This is possibly due to self-association or to nonspecific interactions with other nuclear proteins caused by the altered zinc finger in the context of the remaining basic regions of the DBD (Fig. 2a and Extended Data Fig. 3d), and the intact POZ or Q-rich domains. Overall, our findings are consistent with the contributions to specific and nonspecific DNA binding shown by mutant DBDs of mammalian TFs[38,41,94], and with the observation of intrinsically disordered region-assisted, in vivo chromatin binding for yeast and mammalian TFs[95,96]. The mechanism underlying the aberrant, slow-diffusion of *ZF* mutants is unknown and remains to be further explored.

The POZ domain is found at the N terminus of vertebrate and invertebrate transcriptional regulators implicated in development and disease[97]. Functionally, the POZ domain is involved in protein homo- and hetero-dimerization, as well as multimerization[97]. The POZ domain mediates multimerization of GAF, which facilitates cooperative binding to closely clustered GAGAG elements[44,45], and assists long-distance promoter-enhancer interactions between well-separated GAGAG clusters[98]. As judged by the kinetic behaviors of the GAF POZ mutant, we conclude that multimerization of GAF constitutes a critical element for its ability to pioneer open chromatin. Similarly, a variant glucocorticoid receptor that mimics allosterically induced glucocorticoid receptor tetramerization converts glucocorticoid receptor to a super-receptor that enhances chromatin occupancy at normally inaccessible sites[99].

Autonomy from recruited chromatin remodelers.

The coupling of GAF-mediated pioneering of chromatin accessibility to ATP-dependent chromatin remodeling activities has been reported from the outset of studies on the mechanism underlying DNase hypersensitive sites[12]. Biochemical experiments have since demonstrated that GAF directly recruits remodelers NURF and PBAP via protein-protein interactions, in addition to a number of other chromatin-based factors[82]. Using mutants for NURF and PBAP, we now show that their recruitment is not obligatory for GAF to kinetically engage chromatin targets. This indicates that GAF is largely autonomous from recruited remodelers and that the ensuing chromatin remodeling to antagonize competing processes of nucleosome encroachment primarily benefits the binding or activity of neighboring TFs to chromatin targets. While other factors or the global background of remodeling activities are not excluded from modulating GAF binding, GAF's relative

autonomy from two prominent recruited members of the SWI/SNF and ISWI remodeler families at the initial step of chromatin association may define an important property of TFs that act as pioneering agents.

Constitutively high temporal occupancies by GAF.

By curating genomic databases for the number of genomic targets and integrating these parameters with measured kinetics and abundance, we found that GAF binds to its target sites with temporal occupancy of 182% averaged over 3,622 high-confidence ChIP–seq peaks[17] (Fig. 5a–c), that is with full occupancy by a GAF monomer or oligomer. For genomic sites with greater than average GAF ChIP enrichment and number of GAGAG elements (Fig. 5c), this occupancy is likely to involve higher oligomers, consistent with the native biochemical states of GAF complexes. Such high occupancy, or kinetic persistence, whereby factor dissociation from chromatin and replacement are essentially simultaneous, maintains a constant barrier and magnet for remodeler recruitment at GAF targets (GAF multimers appear as stable biochemical complexes[44,45], but dynamic exchange of GAF monomers within a multimeric complex awaits further study). We envision that a substantial fraction of GAF targets in the genome displays multimeric status and full temporal occupancy (roughly 100%), while the remainder show progressively lower occupancies, consistent with the genome-wide continua of TF-binding levels on metazoan genomes that reflect functional, quasi-functional and nonfunctional transcriptional control[100,101].

Unlike GAF, HSF monomers under RT conditions inducibly trimerize on heat shock[63], which markedly increases the chromatin-bound fraction (F_{bound}) without changing stable residence time (τ_{sb}). For roughly 400 activated HSF-binding sites on the *Drosophila* genome[54], we estimate an average of roughly 50% temporal occupancy by one HSF trimer. At the major *Hsp* loci harboring greater than average ChIP–seq signal intensity and multiple HSE elements, we envision that one or more HSF trimers engage at near full (100%) temporal occupancy. Full occupancy by HSF on heat shock induction may be required to facilitate release of the paused RNA Polymerase II previously established by GAF, and to sustain recruitment-release of new transcription preinitiation complexes for a strong transcriptional burst of HS-responsive genes.

Pioneering of chromatin accessibility.

The pioneer TF concept has been introduced and elaborated for over two decades with a focus on special nucleosome-binding properties of the FoxA1 prototype proposed to initiate establishment of chromatin accessibility for the benefit of consequent binding of neighboring TFs[19]. However, there continues to be debate whether the reported properties of FoxA1 are sufficiently distinct to set it apart from other TFs[16,20–22]. Our early findings on GAF that predate the controversy have documented that nucleosome-binding and ATP-dependent remodeling are functionally coupled in a biochemical assay[12], and additional studies to the present—including the genome-wide effects of remodeler depletion on chromatin accessibility[28] and nucleosome positioning at GAF targets[79]—support the concept that GAF directly recruits remodelers for the site-specific creation of accessible chromatin.

Thus, there is ample evidence to include remodeler recruitment by GAF or other TFs[21] as a fundamental biochemical criterion for pioneering besides affinity for nucleosomes or closed chromatin[41]. Our finding of autonomous remodeler recruitment (chromatin interaction kinetics of GAF being largely unaffected in NURF and PBAP mutants) provides additional insight on the hierarchical nature of pioneering wherein GAF binding to chromatin at the initial stage of pioneering (stage 1) would be followed by remodeler recruitment and ATP-dependent nucleosome mobilization to create DNase hypersensitivity, thereby facilitating assembly of the transcription preinitiation complex, paused Pol II and the inducible binding of HSF (stage 2) (Fig. 6). In addition to remodeler recruitment, constitutively full temporal occupancy by GAF revealed by the single-particle kinetics provides a quantitative criterion for pioneering long-term chromatin accessibility primed for the transcriptional responses to homeostatic, environmental and developmental signals. Notably, TF Zeste has been shown by in vitro reconstitution experiments to be capable of binding the chromatin template independently with subsequent recruitment of the BAP chromatin remodeler[102].

We emphasize that high temporal occupancy is not an obligatory consequence of GAF's long residence time on chromatin, or its multimeric, cooperative binding to GAGAG elements. Occupancy is also dependent on cellular GAF expression, abundance and the number and genomic distribution of GAGAG elements in *Drosophila*. In a similar range as GAF, CTCF has a residence time of roughly 70 s in U2OS cells measured by SPT[43]. A stochastic HaloTag labeling approach revealed roughly 110 and 265 s residence times for Sp1 and CTCF, respectively in U2OS cells[42]. This suggests that long residence times may be shared by other pioneer factors, but are not obligatory, as the pioneer factor Zelda has a short residence time of 3–5 s (refs. [103,104]). For dynamic Zelda, heterogeneous nuclear distributions with locally increased concentrations should also contribute to high occupancy[103,104]. Thus, it may be instructive to consider pioneering as an active process with the multiple inputs we have described: autonomous factor binding to closed chromatin, remodeler recruitment, nucleosome mobilization and for the subset of TFs that maintain accessible chromatin constitutively, correspondingly high temporal occupancy, not excluding additional criteria to be identified. Together with an increased appreciation of overall factor abundance and local concentration in driving transcription[105], we hope that inclusion of these biochemical and kinetic principles guides further investigations on the substantial fraction of computationally identified human TFs (16% of roughly 700 TFs)[16] that may pioneer chromatin accessibility as a basic mechanism of genome regulation in eukaryotic organisms.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41594-022-00800-z.

Methods

Fly strain construction.

CRISPR–Cas9-mediated genome editing.—HaloTag was inserted downstream of the start codon of endogenous *Trl* via homology-directed repair (HDR) and CRISPR–Cas9 to generate the Halo-GAF knock-in strain (Supplementary Fig. 1a–c). The donor repair template was constructed on the pScarlessHD-DsRed plasmid (a gift from K. O'Connor-Giles, Addgene plasmid no. 64703), which contains a DsRed selection marker cassette flanked by PBac transposon ends and TTAA sites[109]. The donor plasmid is designed such that, after HDR, the DsRed cassette is inserted into a nearby genomic TTAA site adjacent to the gRNA target in the coding region that is close to the ATG start codon. Approximately 1 kb downstream of the gRNA site was cloned as the right homology arm, with silent mutations introduced to destroy the gRNA sequence in the donor plasmid. Similarly, 1 kb upstream of the genomic TTAA site was cloned as the left homology arm. The left and right homology arms mediate HDR on Cas9 cleavage, inserting HaloTag along with the DsRed cassette. Flies that underwent HDR were identified with DsRed eye fluorescence. The DsRed cassette was removed with a single cross to a fly strain expressing PBac transposase, as indicated by loss of fluorescence, leaving only one TTAA site, thus allowing scarless HaloTag knock-in with a removable selection marker. A flexible linker GGSGS was added between Halo and GAF. The HaloTag knock-in was verified by fluorescent staining and DNA sequencing.

After constructing the Halo-GAF strain, deletions of the Halo-GAF fusion protein were generated by CRISPR–Cas9 gene editing (Extended Data Fig. 3 and Supplementary Fig. 2). A 90 bp (30 amino acids (aa), Δ90–119) precise deletion was generated in the POZ domain (*ΔPOZ*) by HDR. Small deletions in the zinc finger of the DBD (*ZF*[9], R356Δ, N357Δ; and *ZF*[10], R356Δ) were generated by random indels. For *ΔQ*, two gRNAs targeting the Q-rich domains of the long or short GAF isoforms were introduced at the same time to screen for indels creating frameshifts and truncations of both Q-rich domains. To screen for desired mutants, lethal or reduced viability strains were selected and characterized by PCR and DNA sequencing.

All gRNAs (see Supplementary Table 1 for gRNA sequences) were cloned into pCFD5 (a gift from S. Bullock, Addgene plasmid no. 73914). Donor and gRNA plasmids were mixed to final concentrations of 200 and 500–600 ng μl^{-1}, respectively, and injected into fly strains expressing Cas9 in the germline (yw;nos-Cas9(II-attP40), a gift from NIG-FLY, Japan). To generate Halo-GAF mutants, the Halo-GAF knock-in strain was crossed to yw;nos-Cas9(II-attP40) for injection. All fly embryo injections were performed by BestGene Inc.

Transgenic fly construction via PhiC31 integrase.—*Trl* gene and roughly 1 kb flanking genomic sequence was cloned from a BAC genomic clone (BACR11B23) into pattB (backbone taken from pattB-aubergine-ADH-gf, Addgene plasmid no. 69448, a gift from P. Zamore) via recombineering[110]. HaloTag was inserted upstream of the stop codon for the long and short isoforms, respectively, along with a removable Cam^R selection cassette via recombineering[111]. The CamR cassette was flanked by the 8-bp NotI restriction sites (plus an additional bp to ensure that HaloTag is in frame) and removed by NotI

digestion and religation, leaving a GGSGSAAA linker sequence between GAF and HaloTag. The constructs were incorporated into the attP2 site in the *Drosophila* genome via PhiC31 integrase[112] (Supplementary Fig. 1e), generating the GAF^{L}-Halo and GAF^{S}-Halo transgenic strains, which express a Halo-tagged long or short GAF isoform and the other untagged isoform. The functionality of recombinant fusion proteins was verified by rescue of the lethal alleles Trl^{13C}/Trl^{R67} (ref. [3]). We similarly generated the HSF-Halo transgenic fly strain at the attP2 site from the genomic clone BACR33K09. The functionality of HSF-Halo was verified by rescue of $P\{PZ\}Hsf^{03091}/Hsf^{3}$ lethal alleles[64,65]. The Halo-H2B transgenic strain was similarly constructed at the attP2 site, with a roughly 4.9 kb DNA fragment containing five *Drosophila* histone genes for HaloTag insertion at the N terminus of H2B.

Mutant fly strains.—Mutant alleles were obtained from the Bloomington *Drosophila* Stock center (BDSD, IN): Trl^{13C}(BDSC:58473); Trl^{R67} (BDSC:58475); $P\{PZ\}Hsf^{03091}$ (BDSC:11271); Hsf^{3} (BDSC:5488); $P\{EP\}Bap170^{G5986}$ (BDSC:28471); $Bap170^{\Delta 135}$ (BDSC:63807); $Nurf301^{4}$ (BDSC:9904) and *Df(3L)Exel6084* (BDSC:7563).

Genotypes of fly strains in Fig. 3 and Extended Data Fig. 5 (only the GAF^{L}-Halo isoform is shown, GAF^{S}-Halo strains have the same corresponding genotypes except expressing GAF^{S}-Halo) are as follows:

WT: $p\{GAF^{L}\text{-}Halo\}attP2$

bap170: $Bap170^{\Delta 135}/P\{EP\}Bap170^{G5986}$; $p\{GAF^{L}\text{-}Halo\}attP2$ (generated by crossing $Bap170^{\Delta 135}$; $p\{GAF^{L}\text{-}Halo\}attP2/T(2;3)TSTL, CyO: TM6B, Tb^{1}$ to $P\{EP\}\ Bap170^{G5986}$; $p\{GAF^{L}\text{-}Halo\}attP2/\ T(2;3)TSTL, CyO: TM6B, Tb^{1}$)

nurf301: $Nurf301^{4}$, $p\{GAF^{L}\text{-}Halo\}attP2/Df(3L)Exel6084$, $p\{GAF^{L}\text{-}Halo\}\ attP2$ (generated by crossing $Nurf301^{4}$, $p\{GAF^{L}\text{-}Halo\}attP2/TM6B, Tb^{1}$ to *Df(3L) Exel6084*, $p\{GAF^{L}\text{-}Halo\}attP2/TM6B, Tb^{1}$)

Genotypes of fly strains in Extended Data Fig. 7a are:

WT: *p{HSF-Halo}attP2*

trl: Trl^{13C}, $p\{HSF\text{-}Halo\}attP2/TrlR^{67}$, $p\{HSF\text{-}Halo\}attP2$ (generated by crossing Trl^{13C}, $p\{HSF\text{-}Halo\}attP2/TM6B, Tb^{1}$ to $TrlR^{67}$, $p\{HSF\text{-}Halo\}attP2/TM6B, Tb^{1}$)

bap170: $Bap170^{\Delta 135}/P\{EP\}Bap170^{G5986}$; $p\{HSF\text{-}Halo\}attP2$ (generated by crossing $Bap170^{\Delta 135}$; $p\{HSF\text{-}Halo\}attP2/T(2;3)TSTL, CyO: TM6B, Tb^{1}$ to $P\{EP\}Bap170^{G5986}$; $p\{HSF\text{-}Halo\}attP2/\ T(2;3)TSTL, CyO: TM6B, Tb^{1}$)

nurf301: $Nurf301^{4}$, $p\{HSF\text{-}Halo\}attP2/Df(3\ L)Exel6084$, $p\{HSF\text{-}Halo\}attP2$ (generated by crossing $Nurf301^{4}$, $p\{HSF\text{-}Halo\}attP2/TM6B, Tb^{1}$ to *Df(3L)Exel6084*, $p\{HSF\text{-}Halo\}attP2/TM6B, Tb^{1}$).

Single-particle imaging in live *Drosophila* hemocytes.

Sample preparation.—Single-molecule live-cell imaging was performed with third instar larval hemocytes, representing mainly plasmatocytes (>90% of *Drosophila* hemocytes). Hemocytes were released from five to ten thoroughly washed larvae into a sample chamber containing 1 ml of filtered Schneider's *Drosophila* medium (Gibco 21720024) including an EDTA-free protease inhibitor cocktail (Roche 4693159001). The sample chamber is an Attofluor Cell Chamber (Invitrogen, A7816) assembled with round coverglass (Electron Microscopy Sciences, 72290–12) cleaned by flaming. Cells were stained with 0.2–1.5 nM JF552/JFX554 (see below for fast- and slow tracking) for 30 min at RT, during which hemocytes adhered to the coverglass bottom of the imaging chamber. During incubation steps, sample chambers were covered with aluminum foil to minimize evaporation and block light. Cells were then briefly washed twice with Schneider's media without protease inhibitor and imaged on a custom-built wide-field SPT fluorescence microscope[113].

For heat shock experiments, after washing away JF dyes, cells were heat shocked on a 37.5 °C heat block for 10 min, and then transferred to a stage-top incubation chamber (Okolab) preheated to 37.5 °C on the microscope. The objective of the microscope was also heated to 37.5 °C. Hence, for each sample, imaging started 10 min after heat shock and continued over multiple cells for a total of 30 min.

Single-particle imaging.—All single-particle imaging were carried out on an Axio Observer Z1 microscope (Zeiss) equipped with a Plan-Apochromat ×150/1.35 glycerin-immersion objective (Zeiss), and a C9100–13 electron microscopy-CCD camera (Hamamatsu Photonics, Japan) featuring 512 × 512 pixels with 16-μm pixel size. The pixel size of recorded images is 16 μm/150 = 107 nm. JF552/JFX554 were excited with a CL555–100 555 nm laser (CrystaLaser) and the emission light was passed through a filter cube containing a 561 nm BrightLine single-edge beamsplitter, a 612/69 nm BrightLine single-band bandpass emission filter (Semrock), then through a 750 nm blocking edge BrightLine multiphoton short-pass emission filter and a 405/488/561/635 nm StopLine quad-notch filter (Semrock) before entering the camera. The electron microscopy-CCD camera was operated at roughly −80 °C (forced-air cooling) and ×1,200 electron microscopy gain. We used the ZEN (Zeiss) and HCImage (Hamamatsu Photonics) software to operate the microscope and camera, respectively. Imaging was carried out at RT except for heat shock experiments in which samples were imaged in a stage-top incubator with temperature control (H301-MINI chamber with UNO-T-H controller, Okolab)

In the fast-tracking regime, time-lapse videos with a 128 × 128 pixel field of view were acquired with high laser power (roughly 1 kW cm^{-2}) and 10 ms exposure time for 1.5–2 min. Initial laser excitation leads to simultaneous emissions of all labeled molecules, marking locations of individual nuclei in the field of view. Cells with relatively homogenous initial nuclear glow (interphase) were imaged. Emitting molecules quickly enter the 'dark' state and stochastically reemit. Cells were stained with JFX554, with the concentration optimized (1–1.5 nM) to achieve sparse single particles per nucleus per frame after 10–30 s of the initial glow, minimizing mis-tracking.

For slow tracking, we used low laser power (roughly 36 W cm^{-2}) and imaged a 256 × 256 pixel field of view with 500 ms exposure time until no single molecules were observed. The temporal range of most videos varied from 2–5 min, but could be as short as 1.5 min or up to 8 min. We labeled cells with 50 nM of a far-red dye JF700 to block most Halo-tagged proteins and at the same time adjusted a low concentration of JF552 (0.02–0.2 nM) to visualize only 2–10 molecules per frame. The labeling of JF552 in slow tracking was much sparser than the fast-tracking regime. Use of JF700 as a blocker made titration of JF552 concentration easier than using JF552 alone. This sparse labeling approach allows tracking of chromatin-bound molecules with minimal photobleaching.

Single-particle data processing and statistical analysis.

Image preprocessing.—Raw time-lapse data were preprocessed in Fiji[114] to convert to 16-bit TIFF format and extract a substack with sparse single particles (<5 particles per nucleus for fast tracking and <10 particles per nucleus for slow tracking). A maximum-intensity Z projection of each video was generated to outline cell nuclei as the ROI. A corresponding binary mask was created to isolate nuclear trajectories for subsequent analysis.

Single-particle localizing and tracking.—Single particles were localized and tracked using the open-source program DiaTrack v.3.05 (ref. [115]). Tracking was performed with a 6 pixel (roughly 0.65 μm) maximum jump allowance between consecutive frames for fast tracking. All factors we imaged showed displacements within this cutoff as informed by the frequency histogram (Extended Data Figs. 1b, 4a, 5a and 7b). Since slow tracking selectively imaged bound particles, the maximum jump allowance between consecutive frames was set as 3 pixels (roughly 0.32 μm) to minimize misconnection. HSF-Halo showed displacement histograms with a smooth tail within this range, while Halo-GAF and Halo-H2B displacements were mostly within 0.2 μm (not shown). For analysis of slow-tracking experiments to measure residence times, we allowed gaps in trajectories to account for blinking or missed localizations, with three-frame maximum blinking and a more stringent 2-pixel maximum jump than the routine 3-pixel maximum jump (However, both jump criteria yield very similar results).

Analyzing fast-tracking data.—We used a custom R package Sojourner (https://github.com/sheng-liu/sojourner) to extract trajectories from MATLAB files generated by DiaTrack, which contain information on x, y coordinates and frame number (time) that were applied for computation of kinetic parameters. Trajectories found within the nucleus were isolated (‘masked’ trajectories) using Sojourner and binary masks generated during preprocessing. Average length of trajectories was 11–21 displacements (median 5–8).

Spot-On.: We used Spot-On to perform two-state kinetic modeling of displacements from all ‘masked’ trajectories[39], to derive diffusion coefficients (D_{bound}, D_{free}) and the corresponding fractions (F_{bound}, $F_{\text{free}} = 1 - F_{\text{bound}}$). The Spot-On python package was used with the following parameters: bin width 0.01 μm, number of timepoints 6, jumps to consider 4, max jump 1 μm, gapsAllowed 0 and Z correction with $dZ = 0.6$ μm. Mean and s.d. of F_{bound}, D_{bound} and D_{free} were calculated from 3–5 biological replicates.

vbSPT (variational Bayesian SPT) HMM.: A MATLAB program running vbSPT HMM[69,70] (https://gitlab.com/anders.sejr.hansen/anisotropy) was modified to assign each trajectory displacement into two states, 'bound' or 'free'. Then each trajectory was subclassified as 'bound' if all displacements were classified as bound state and 'free' if all displacements were classified as free state; a small fraction of trajectories containing two states were omitted from related analysis. The subclassified 'bound' and 'free' trajectories from biological replicates of the same conditions were pooled together and used to calculate MSD.

To calculate the apparent radius of confinement (R_c) for individual trajectories, we fit each MSD curve with the following confined diffusion model[41,116],

$$\mathrm{MSD} = R_c \times \left(1 - e^{\frac{-4D\Delta t}{R_c^2}}\right)$$

Analyzing slow-tracking data.—Using the Sojourner R package, the apparent dwell times (temporal length of trajectories) were determined for all 'masked' trajectories lasting at least three frames.

1-CDF curves were generated and fit to a double exponential decay model:

$$P(t) = f_{sb}e^{-k_{sb}t} + f_{tb}e^{-k_{tb}t}$$

where k_{sb} and k_{tb} correspond to dissociation rates for stable- and transient-binding events, respectively; f_{sb} and f_{tb} correspond to the fraction of time the molecule spends at stable- and transient-binding sites, respectively, and $f_{sb} + f_{tb} = 1$.

The survival curves reflect not only factor dissociation, but also photobleaching, axial and lateral cell or chromatin movements, fluctuating background and so on. To correct for all these factors, assuming that these processes affect Halo-H2B to the same extent as other proteins, and that bulk H2B dissociation is negligible in the experimental time frame of 2–5 min, we measured the apparent unbinding rate for Halo-H2B in the same way and used it as a correction factor for other proteins' residence time[43]. The corrected average residence times for stable- (τ_{sb}) and transient-binding (τ_{tb}) were calculated as follows:

$$\tau_{sb} = \frac{1}{k_{sb} - k_{sb,\ H2B}}$$

$$\tau_{tb} = \frac{1}{k_{tb} - k_{sb,\ H2B}}$$

Mean and s.d. of k_{sb}, k_{tb}, f_{sb} and f_{tb} were calculated from 100 bootstrap samples, then mean and s.d. of k_{sb} and k_{tb} were used to calculate τ_{sb} and τ_{tb} with error propagation.

H2B decay reports photobleaching kinetics and chromatin mobility, and is therefore expected to be unique depending on fluorophore type, cell type and imaging regime. As a correction standard for dwell time, the 1-CDF survival curve of H2B has been reported with different kinetics by different investigators[40,43,104]. Although not directly comparable, it is of interest that the survival plots of Halo-H2B (JF552) in our study of *Drosophila* hemocytes are of a similar range to H2B-Halo (JF549 or PA-JF646) survival in human U2OS cells.

Calculating search kinetics and target occupancy.—We integrated approaches from previous studies[38,117,118] and calculated temporal occupancy as described[78].

First, search time (τ_{search}) is the average time it takes from a molecule dissociates off a specific site till it find the next specific site, during which the molecule samples nonspecific sites (each lasts for τ_{tb} on average) for a number of times (N_{trials}) before encountering the next specific target (bound for τ_{sb} on average). τ_{free} is the average free time between two nonspecific binding events. Assuming the molecule samples all nonspecific and specific sites at random with equal accessibility, that is, equal probability of binding to all sites, the average search time between two consecutive specific binding events is calculated as:

$$\tau_{search} = N_{trials} \times (N_{trials} + 1) \times \tau_{free}$$

N_{trials} depends on the ratio of number of nonspecific (N_{ns}) to specific sites (N_s), or r_s:

$$N_{trials} = \frac{N_s + N_{ns}}{N_s} = 1 + r_s$$

Thus,

$$\tau_{search} = (1 + r_s) \times \tau_{tb} + (2 + r_s) \times \tau_{free}$$

To determine r_s, we considered two scenarios underlying detection of binding events during slow tracking[78].

1. Blinking-limited ($r_{s,\,bl.}$): f_{sb} obtained by slow tracking is proportional to the fraction of time the molecule spends at specific sites (stable binding) relative to the overall time it is bound to chromatin (stable- and transient-binding):

$$f_{sb} = \frac{N_s \times \tau_{sb}}{N_s \times \tau_{sb} + N_{ns} \times \tau_{tb}} = \frac{\tau_{sb}}{\tau_{sb} + r_s \times \tau_{tb}}$$

$$r_{s,bl.} = \frac{\tau_{sb}}{\tau_{tb}} \times \left(\frac{1}{f_{sb}} - 1\right)$$

2. Diffusion-limited ($r_{s,\,diff.}$): assuming equal probability of binding to all sites, f_{sb} depends on the ratio of number of specific sites to all sites:

$$f_{sb} = \frac{N_s}{N_s + N_{ns}} = \frac{1}{1 + r_s}$$

$$r_{s,\,diff.} = \frac{1}{f_{sb}} - 1$$

These physical processes can happen in the cell coincidentally and likely represent two extremes of a range of behaviors single molecules can exhibit. We reasoned that the relative likelihood of detecting a blinking-limited binding event by slow tracking is proportional to the global fraction of bound molecules (F_{bound} obtained by fast tracking). Therefore, we computed a weighted average value for r_s as follows:

$$r_s = \frac{(F_{bound} \times r_{s,\,bl.}) + (1 - F_{bound}) \times r_{s,\,diff.}}{2}$$

To obtain τ_{free}, we considered that F_{bound} is proportional to the fraction of the time a molecule spends bound to chromatin either stably or transiently:

$$F_{bound} = \frac{\tau_{sb} + N_{trials} \times \tau_{tb}}{\tau_{sb} + N_{trials} \times \tau_{tb} + (N_{trials} + 1) \times \tau_{free}}$$

Thus,

$$\tau_{free} = \frac{\frac{(1 + r_s) \times \tau_{tb} + \tau_{sb}}{F_{bound}} - (1 + r_s) \times \tau_{tb} - \tau_{sb}}{2 + r_s}$$

τ_{search} was calculated with the values derived for r_s and τ_{free}, as shown above. We then estimated the sampling interval (SI), the average time between two consecutive binding events at a specific site[38],

$$SI = \frac{(\tau_{search} + \tau_{sb}) \times N_{sites}}{N_{monomers}}$$

We used N_{sites} values presented by GAF and HSF ChIP–seq studies[17,54]. $N_{monomers}$ was estimated by flow cytometry (below). Finally, the average occupancy is the percentage of time a given specific site is occupied by the protein of interest:

$$Occupancy = \frac{\tau_{sb}}{SI}$$

Confocal microscopy.

Hemocytes were prepared in the same way described above for live-cell imaging, except for staining with 50 nM JFX554 for 30 min at RT or heat shocked on a metal heat block at the indicated temperature (the time for heat shock overlaps with the final stage of staining so

that dye labeling and heat shock were completed at the same time). During all incubation steps the sample chambers were covered with aluminum foil to minimize evaporation and block light. Then cells were briefly washed twice with PBS and immediately fixed in freshly made 4% formaldehyde (diluted from Pierce 16% formaldehyde (w/v), methanol-free, no. 28906) in 1× PBS for 15 min. Fixed samples were washed in PBS for 10 min, then stained with 4,6-diamidino-2-phenylindole (DAPI) for 10 min and washed in PBS for 5 min. Sample chambers were filled with PBS and covered by a coverglass on top. Salivary glands were dissected, then stained and heat shocked similarly as hemocytes except that JF554 staining was performed after fixation. Salivary gland samples were mounted on glass slides in VECTASHIELD Antifade Mounting Medium (Vector Laboratories H-1000–10). We imaged these samples on a LSM 800 Airyscan confocal microscope (Zeiss) with a ×63 objective. *Z*-stacks were taken at optimal resolution recommended by the Zen software (Zeiss) with 0.2-μm step size for hemocyte nuclei and 0.5-μm step size for salivary gland polytene nuclei. The same laser and scanning settings were used between samples in the same experiment.

Estimation of cellular abundances for Halo-tagged proteins.

Cellular abundances of Halo-GAF and HSF-Halo were estimated by flow cytometry using a calibrated C32 CTCF-Halo U2OS human cell line[87,119]. U2OS cell samples were prepared as described[119] with minor modifications. Briefly, we labeled WT and CTCF-Halo U2OS cells with 50 nM JF552 for 30 min at 37 °C/5% CO_2 in a tissue-culture incubator, washed out the dye (removed medium, rinsed with PBS and incubated with fresh media for 5 min in the incubator) and then immediately prepared cells for flow cytometry. Resuspended cells were filtered through a 40-μm filter and placed on ice until their fluorescence was read out by the flow cytometer. To prepare fly hemocytes for flow cytometry, 20–30 third instar larvae were thoroughly washed and dissected on ice in a 1.5-ml Eppendorf tube lid containing Schneider's *Drosophila* medium (Gibco 21720024), EDTA-free protease inhibitor cocktail (Roche 4693159001) and 10% fetal bovine serum (FBS) (HyClone FBS SH30910.03, Cytiva). Hemocytes were collected and stored on ice. After dissection was done for all fly strains, hemocytes were stained with 50 nM JF552 in 1 ml of medium at RT for 30 min. All tubes were prerinsed with FBS to minimize cell stickiness to the tubes. After staining, hemocytes were centrifuged at 200*g* for 5 min, resuspended in 1 ml of fresh medium and washed at RT for 15 min (longer wash time than for U2OS cells to remove nonspecific cytoplasmic signals). The hemocytes were centrifuged again, resuspended in 500 μl of medium and placed on ice until their fluorescence was read out by the flow cytometer.

We used a SH800 (Sony) cell sorter in analyze mode to measure fluorescence intensity in the U2OS cell lines and hemocytes with the same settings. Single live cells were gated using forward and side scattering in SH800 cell sorter software. JF552 fluorescence was excited using a 561 nm laser and emission read out using a 617/30 bandpass filter. The absolute abundance of protein of interest N_{monomers} (mean number of molecules per cell) was obtained according to:

$$N_{\text{monomers}} = \frac{I - I_{\text{Background}}}{I_{\text{CTCF}} - I_{\text{U2 OS Background}}} \times N_{\text{CTCF}}$$

where I is the average measured fluorescence intensity of the cells expressing the protein of interest, $I_{\text{Background}}$ is the average measured fluorescence intensity of hemocytes from a fly strain not expressing HaloTag (*w1118*), I_{CTCF} is the average measured fluorescence intensity of the CTCF-Halo standard U2OS cell line, $I_{\text{U2OSBackground}}$ is the average measured fluorescence intensity of WT U2OS cells not expressing HaloTag and N_{CTCF} is the absolute abundance of CTCF-Halo (roughly 109,800 proteins per cell). For each experiment, hemocytes and U2OS cells were stained with the same aliquot of JF552 stock solution and measured during the same flow cytometry session. We performed three to five biological replicates to get mean and standard deviation values.

FRAP.—Live-hemocyte samples were prepared as described in SPT except that cells were stained with 50 nM JF552 for 30 min to saturate the staining of HaloTag. FRAP experiments were performed on an LSM 800 Airyscan confocal microscope with a Plan-Apochromat ×63/1.40 oil-immersion objective (Zeiss) at RT. For each cell, three frames were acquired before photobleaching. Due to the mobile nature of hemocyte nuclei and nonhomogeneous distribution of GAF, around 25–30% of the nuclear region was bleached by a 561 nm laser at 15–40% power for 25 iterations, then the whole cell nucleus was excited with 0.2% laser power and imaged in a 8.45 × 8.45 μm^2 window at 5 s intervals for at least 6 min. For each cell sample preparation, FRAP was performed on 10–20 cells.

Data were analyzed in imageJ. The StackReg plugin[120] car drift. Mean intensities for a region of unbleached fluorescence (I_{Ref}) and a region slightly smaller than the bleached area (I_{Bl}) were determined and background (I_{Bg}) subtracted. I_{Bl} intensity relative to the I_{Ref}, that is $(I_{\text{Bl}} - I_{\text{Bg}})/(I_{\text{Ref}} - I_{\text{Bg}})$, was normalized against prebleach $(I_{\text{Bl}} - I_{\text{Bg}})/(I_{\text{Ref}} - I_{\text{Bg}})$ to account for the overall photobleaching during time-lapse acquisition. Data from 10 to 20 cells were averaged to generate a single recovery curve for the 'Exponential Recovery' fit. The fit function was $y = a \times (1 - e^{-kt}) + c$, where a represents the fraction of slow recovery and k the corresponding recovery rate, and $1/k$ the estimated half-recovery time.

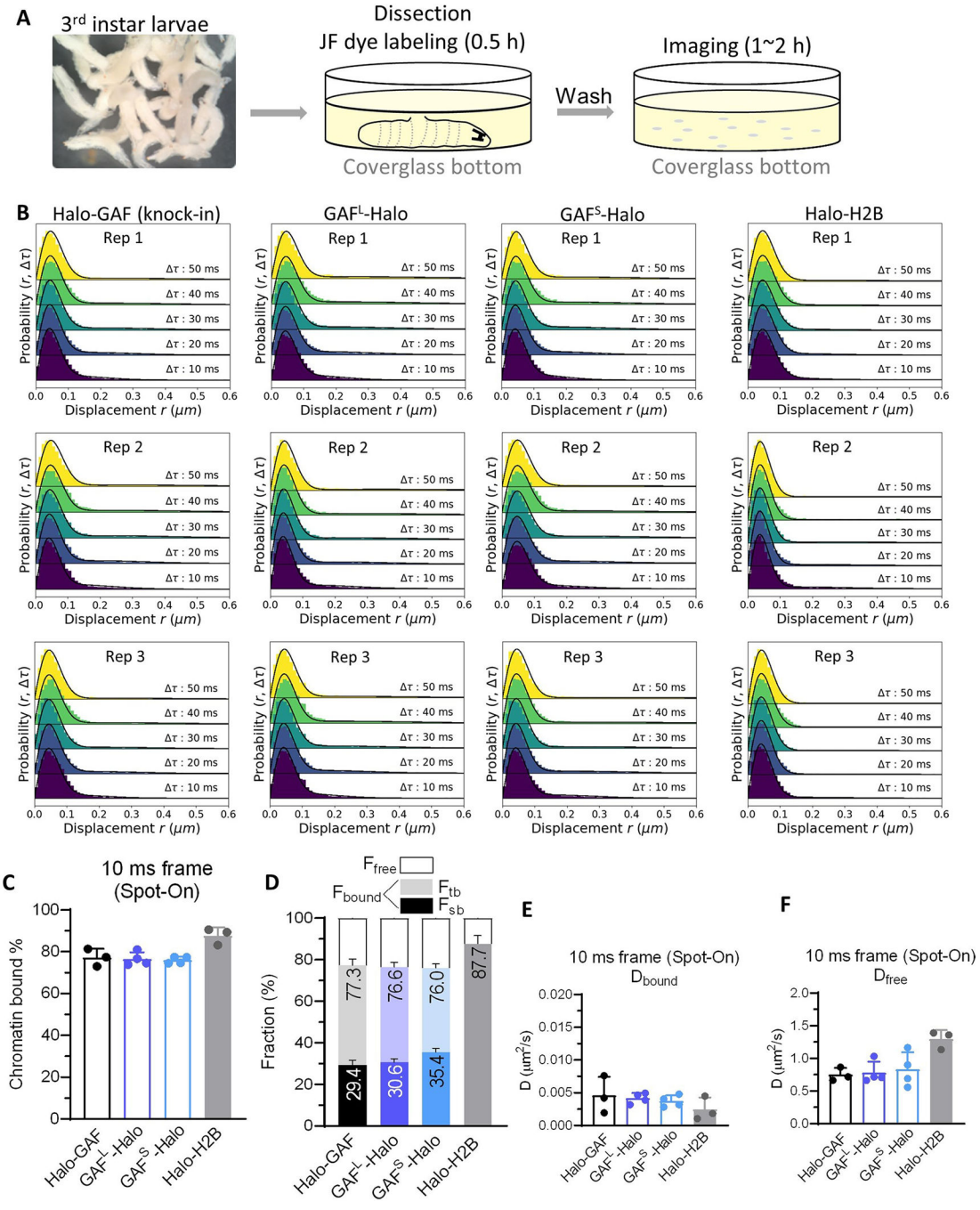
Characterization of cell-cycle stage for larval hemocytes.

A fly strain with the Fly-Fucci markers under UAS control (UAS-CFP.E2f1.1–230 and UAS-Venus.NLS. CycB.1–266, BDSC:55122) was crossed to a hemocyte-specific driver Cg-Gal4 (BDSC:7011). Third instar larvae were washed then dissected to release hemocytes in a drop of PBS on a slide, then the fluorescence of hemocytes were imaged immediately on a Zeiss Axioplan 2 compound microscope.

Statistics and reproducibility.

Statistical tests and P values are reported in figure legends. All experiments were independently repeated at least three times with consistent results. Representative microscopic images from at least five independent acquisitions are shown in the figures.

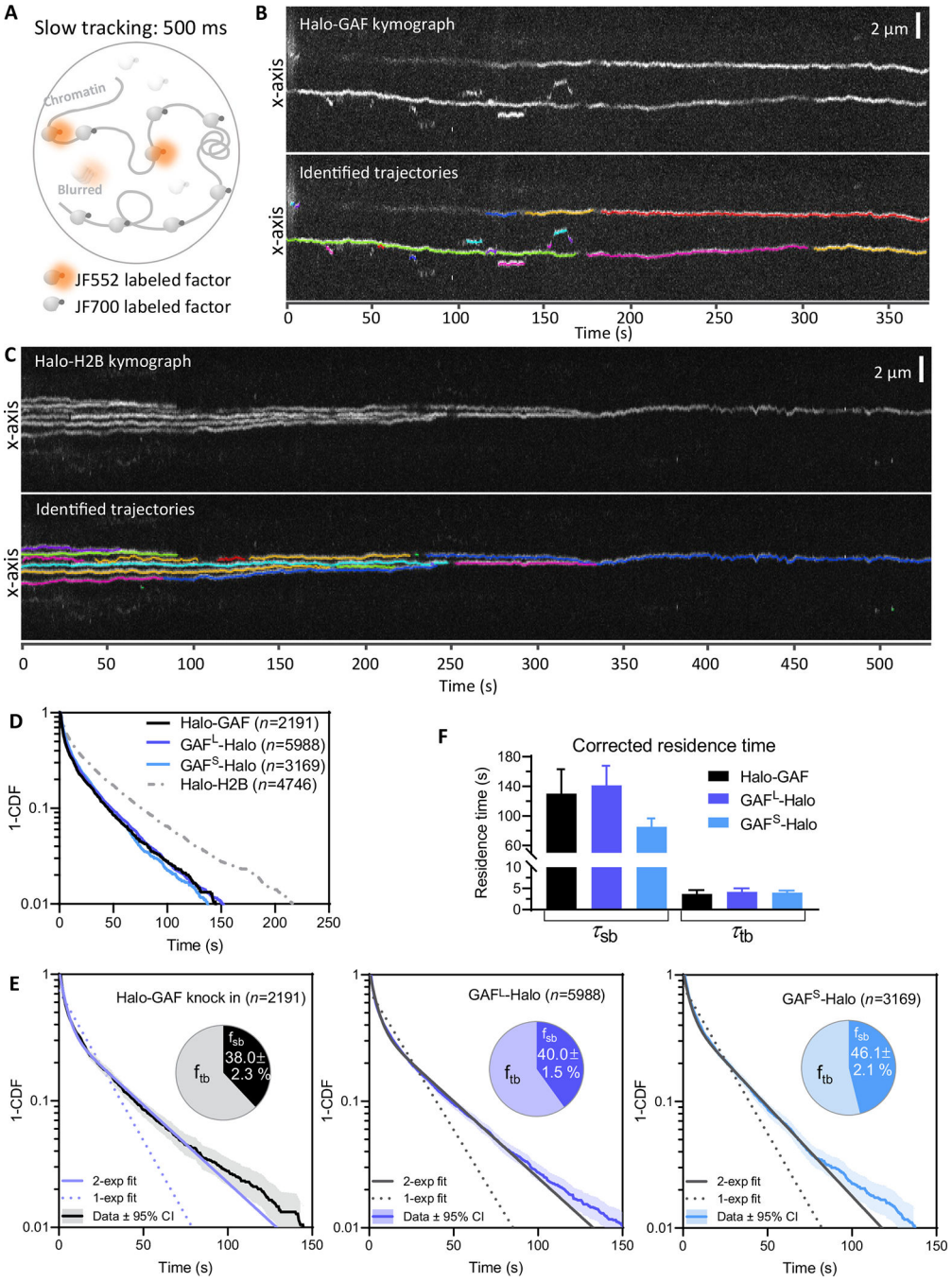
Extended Data



Extended Data Fig. 1 |. Hemocyte imaging and fast-tracking diffusive parameters for Halo-GAF, GAF^{L}-Halo, GAF^{S}-Halo and Halo-H2B.

(A) Experimental timeline of single-particle imaging with 3rd instar larval hemocytes. 3rd instar larvae are washed with DI H2O (left) and dissected in a coverglass bottom dish containing Schneider's medium and JF dye at room temperature. Upon dissection hemocytes are released into the medium and labeled for 30 min, the rest of the larval tissues are discarded (middle). Cells are briefly washed twice with fresh media and imaged for 1–2

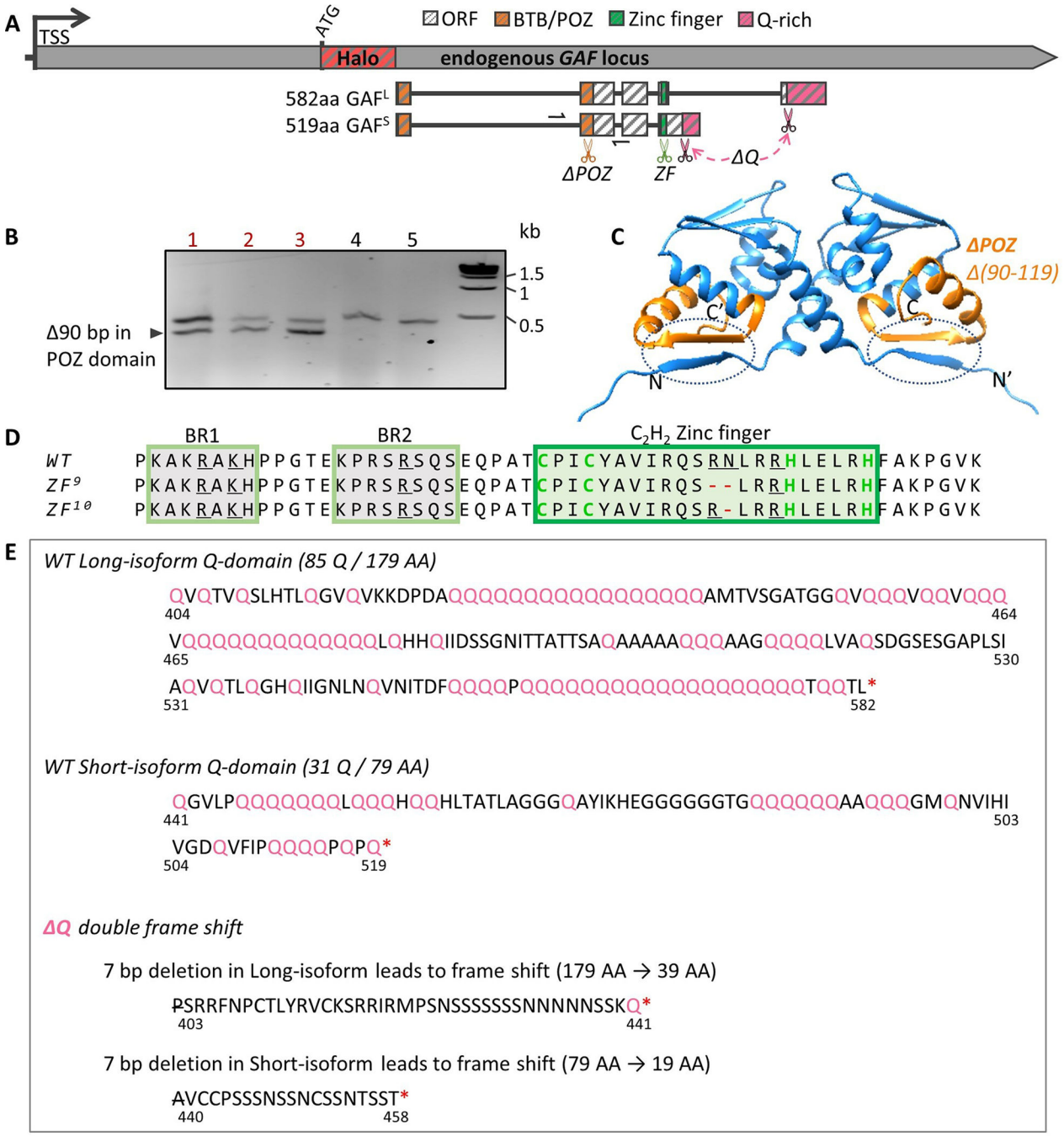
h. (B) Spot-On fits of Halo-GAF, GAF^{L}-Halo, GAF^{S}-Halo, Halo-H2B fast-tracking data. (C) Spot-On kinetic modeling of fast-tracking data shows 77% of Halo-GAF is chromatin bound. Similar values are obtained for isoforms GAF^{L} and GAF^{S} individually tagged in the presence of untagged GAF isoforms. Results are mean ± SD from three biological replicates. (D) Chromatin-free fraction (F_{free}), long- and short-lived chromatin-binding fractions (F_{sb} and F_{tb}) of HaloTagged GAF fusions extracted from fast-tracking in (C, $n = 3$ biological replicates for Halo-GAF and Halo-H2B, $n = 4$ biological replicates for the rest conditions) and slow-tracking data (Extended Data Fig. 2e, $n = 100$ resamplings), respectively, with error propagation. Results are mean ± SD. (E) Diffusion coefficients of bound fraction (D_{bound}) for Halo-GAF, GAF^{L}-Halo, GAF^{S}-Halo, Halo-H2B derived by Spot-On. Results are mean ± SD ($n = 3$ biological replicates for Halo-GAF and Halo-H2B, $n = 4$ biological replicates for the rest conditions). (F) Diffusion coefficients of free fraction (D_{free}) for Halo-GAF, GAF^{L}-Halo, GAF^{S}-Halo, Halo-H2B derived by Spot-On. Results are mean ± SD ($n = 3$ biological replicates for Halo-GAF and Halo-H2B, $n = 4$ biological replicates for the rest conditions).



Extended Data Fig. 2 |. Slow tracking results for Halo-GAF, GAF^{L}-Halo, GAF^{S}-Halo and Halo-H2B.

(A) Fast and slow tracking regimes. Fast tracking with 10-ms frame rate and high laser power allows single molecule imaging to distinguish slow (chromatin-bound) and fast (chromatin-free) diffusing subpopulations. Slow tracking uses low-intensity excitation and 500 ms exposure time to motion blur diffusing molecules and selectively observe the dwell times of chromatin-bound molecules. A higher concentration of JF700 is added to block labeling of most HaloTag protein fusions, while a much lower concentration of JF552 is used to sparsely label a small fraction of HaloTag so that each nucleus shows only 2~10

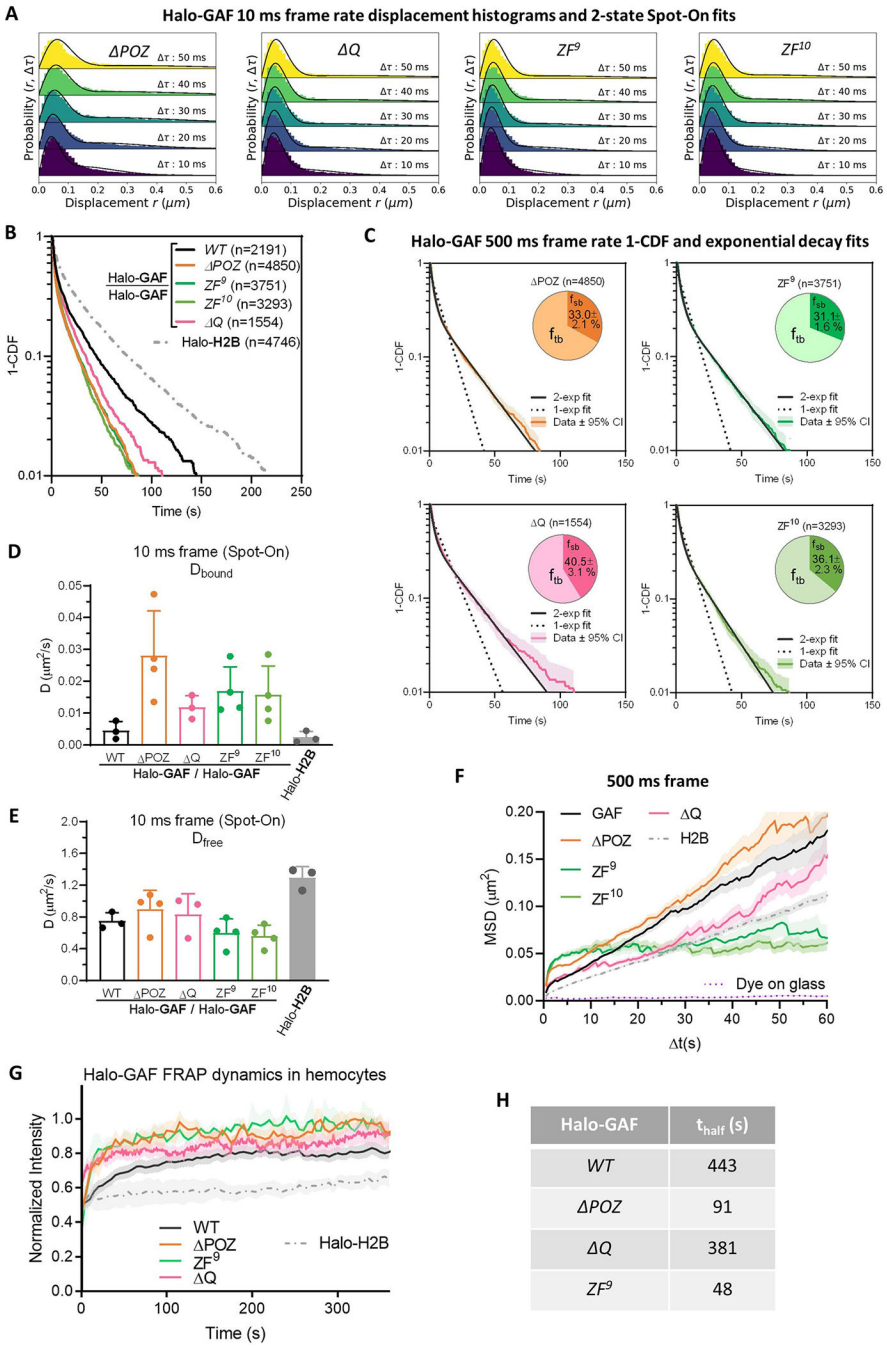
molecules per frame during image acquisition. (B) Kymograph of a Halo-GAF slow tracking movie shows traces of bound GAF molecules over time (upper). Trajectories identified from the raw movie are plotted on the kymograph using separate colors (lower). (C) Kymograph of a Halo-H2B slow tracking movie shows traces of bound H2B molecules over time (upper). Trajectories identified from the raw movie are plotted on the kymograph (lower). (D) Survival probability curves (1-CDF) plotted from apparent dwell times of thousands (n) of single-molecule chromatin-binding events for Halo-GAF, GAFL-Halo and GAFS-Halo. (E) One-component and two-component exponential fit of survival probabilities (1-CDF) from slow tracking data (with 95% CI, confidence interval) of Halo-GAF, GAF^{L}-Halo and GAF^{S}-Halo. Pie charts show the stable-binding (f_{sb}) and transient-binding (f_{sb}) fractions derived from two-component fits, and errors represent bootstrapped SD ($n = 100$ resamplings). (F) Corrected average residence times for stable- (τ_{sb}) and transient- (τ_{tb}) binding by transgenic GAF^{L}-Halo and GAF^{S}-Halo. Error bars represent bootstrapped SD after resampling 100 times ($n = 100$).



Extended Data Fig. 3 |. Generation of mutations in functional domains of Halo-GAF by CRiSPR/ Cas9 gene editing.

(A) In the Halo-GAF fly strain, Cas9 and gRNA were introduced to target the BTB/POZ domain, zinc finger and Q-rich domains, respectively. The BTB/POZ domain is separated by a large intron. A gRNA target site in the second exon (orange scissors) was selected and a donor plasmid containing a 90 bp deletion (*ΔPOZ*) was constructed for homology-directed repair (HDR). For zinc finger mutations, we selected a gRNA target site in the zinc finger coding region (green scissors) and screened for in-frame small deletions generated by non-homologous end joining (NHEJ). To generate deletions of both Q-rich domains in long

and short isoforms (*ΔQ*), two gRNAs targeting the upstream ends of two Q-rich domains (pink scissors) were introduced at the same time, and we screened for double frame-shift deletions induced by NHEJ. Half arrows indicate positions of the PCR primers used in (B). TSS, transcription start site. (B) PCR validation of *ΔPOZ*. Lanes 1–3 show two PCR bands indicating precise deletion in one allele; lanes 4–5 are two lines without the precise deletion. Sanger sequencing verified a precise 90 bp deletion in one allele. (C) AlphaFold[121] predicted structure for a homodimer of GAF POZ domains (residues 1–120) is highly similar to published crystal structures of PLZF POZ domains[122,123]. A 90-bp deletion in the second exon generates a 30-AA deletion (*Δ90–119*, *ΔPOZ*, orange), which includes a β sheet that mediates two of three principal contacts stabilizing the dimer (dashed circles). This functionally lethal mutation is likely to impair GAF multimerization, although the degree to which the multimerization is reduced is unclear. N and C indicate two termini of one monomer; N' and C' for the other monomer. (D) Amino acid sequence of GAF DNA binding domain, which contains a single C2H2 zinc finger (green rectangle) and two upstream basic regions (BR1 and BR2, yellow rectangle). Amino acids involved in recognizing the GAGAG consensus sequence are underlined[56]. Two zinc finger mutations were isolated and verified by sanger sequencing, ZF^{9} and ZF^{10}, with R356 and N357, or R356 deleted, respectively. (E) Amino acid sequence of GAF Q-rich domains for long and short isoforms. In *ΔQ*, a 7 bp deletion was identified by sanger sequencing in both isoforms at the upstream ends of Q-rich domains, resulting in frameshifts and truncations of the Q-rich domains from both isoforms. P403 in the long isoform and A440 in the short isoform are deleted and the subsequent amino acids are newly introduced by the frame shifts.

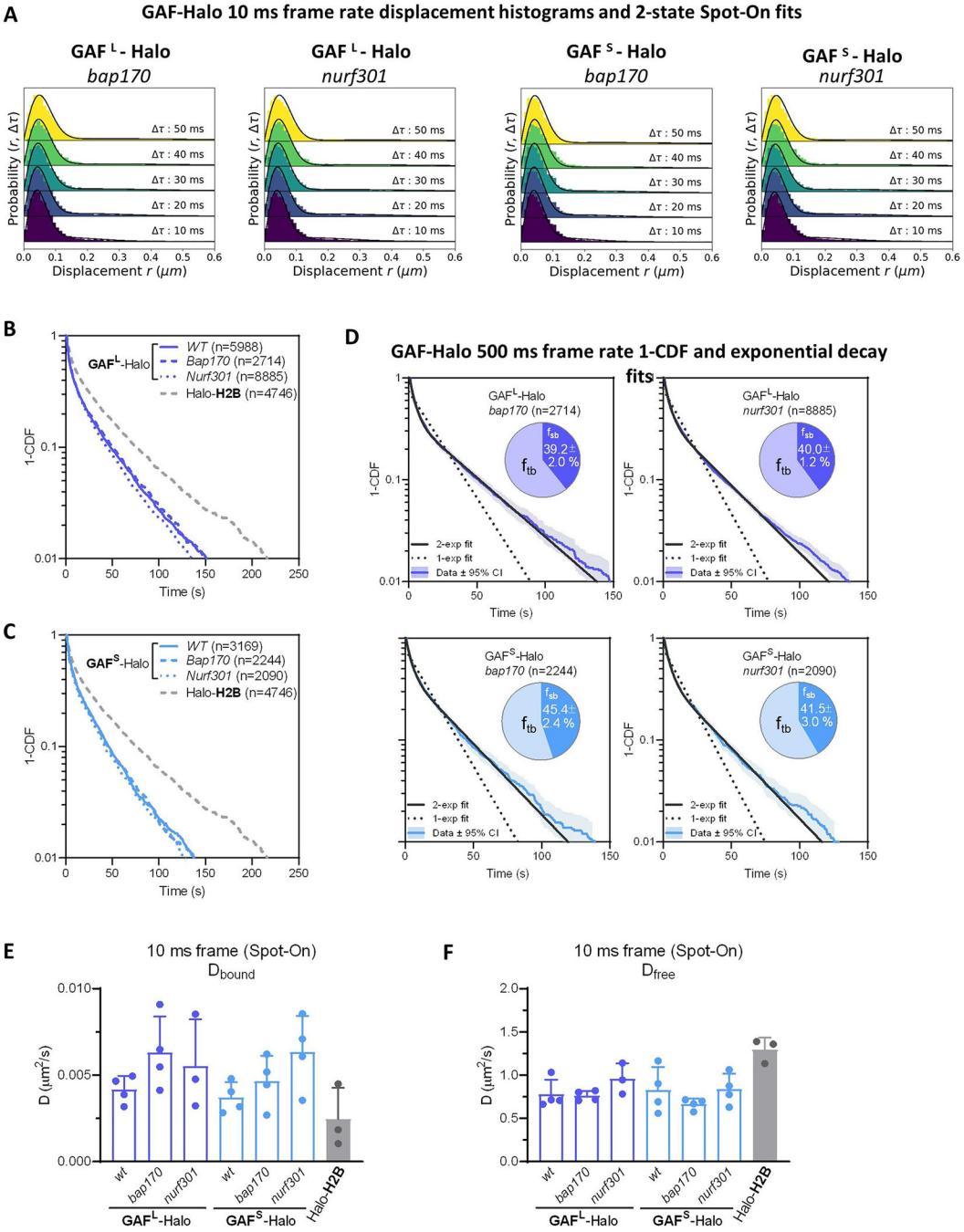


H

Halo-GAF	t_{half} (s)
WT	443
ΔPOZ	91
ΔQ	381
ZF^{9}	48

Extended Data Fig. 4 |. Live-cell SPT and FRAP diffusive parameters for Halo-GAF mutants.
(A) Spot-On fits of fast-tracking data for Halo-GAF mutants (see Extended Data Fig. 1b for *WT*). (B) Survival probability curves (1-CDF) from apparent dwell times of >1,000 single-molecule chromatin-binding events, for *WT* and mutant Halo-GAF. (C) One-component and two-component exponential fit of survival probabilities (1-CDF) from slow tracking data (with 95% CI, confidence interval) of Halo-GAF mutants (see Extended Data Fig. 2e for *WT*). Pie charts show the stable-binding (f_{sb}) and transient-binding (f_{sb}) fractions derived from two-component fits. (D) Diffusion coefficients of bound fraction (D_{bound}) for Halo-GAF

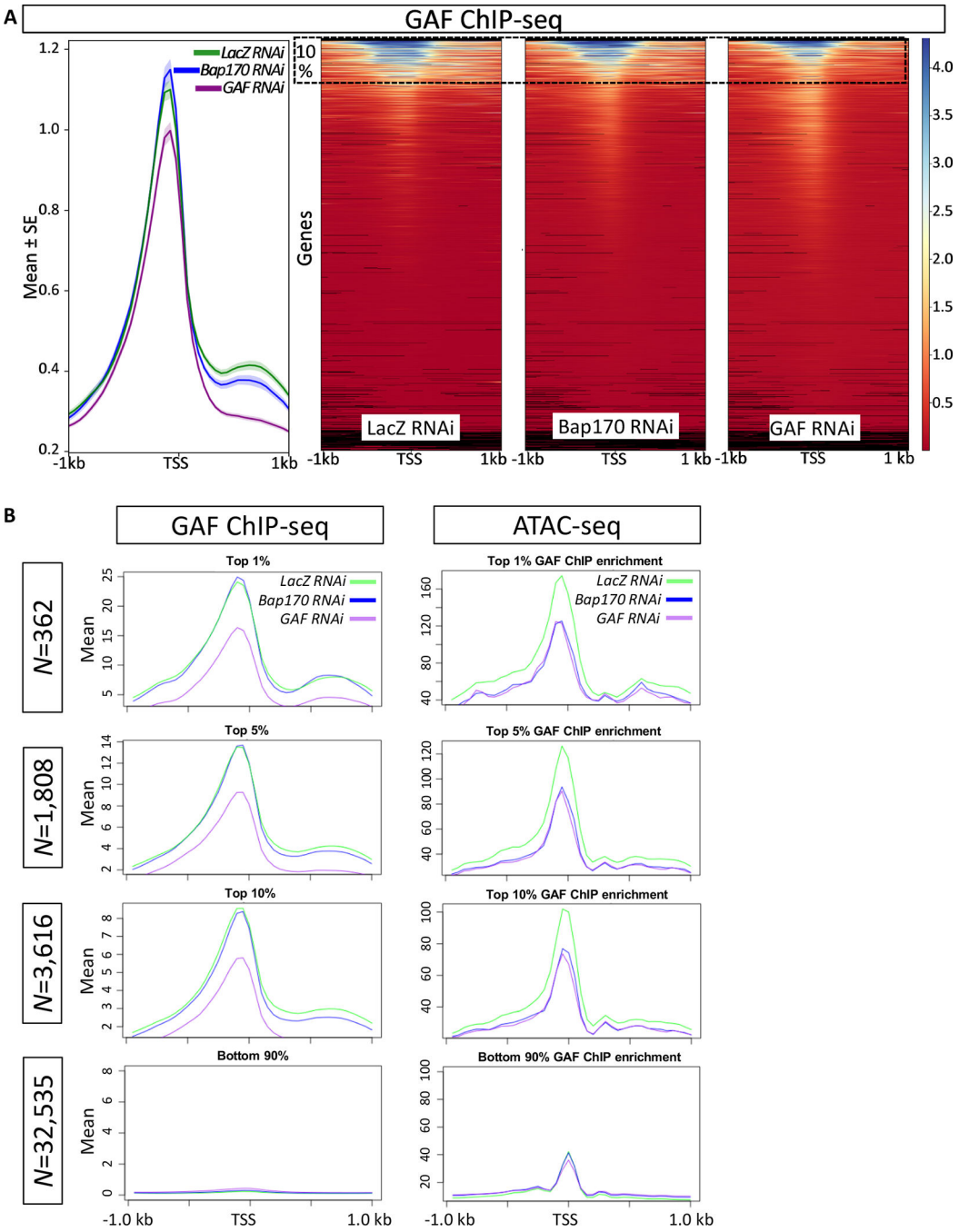
and Halo-H2B derived by Spot-On. Results are mean ± SD ($n = 3$ biological replicates for Halo-GAF WT, ΔQ and Halo-H2B, $n = 4$ biological replicates for the remaining conditions). (E) Diffusion coefficients of free fraction (D_{free}) for Halo-GAF and Halo-H2B derived by Spot-On. Results are mean ± SD ($n = 3$ biological replicates for Halo-GAF WT, ΔQ and Halo-H2B, $n = 4$ biological replicates for the remaining conditions). (F) Average MSD versus lag time for *WT* and Halo-GAF mutants at 500-ms frame rate. Mean and SE (shaded) are shown. System noise is shown by the MSD of dye molecules stuck on coverglass. (G) Mean fluorescence recovery curves from FRAP experiments for Halo-GAF WT and mutants in hemocytes labeled with 50 nM JF552. Shaded areas represent SE. (H) Half recovery times (t_{half}) of FRAP experiments.



Extended Data Fig. 5 |. Live-cell SPT diffusive parameters for GAFL-Halo and GAFS-Halo in *bap170* and *nurf301* mutants.

(A) Spot-On fits of fast-tracking data for GAF^{L}-Halo and GAF^{S}-Halo in *bap170* and *nurf301* mutants (see Extended Data Fig. 1b for *WT*). See methods for genotypes of *WT*, *bap170* and *nurf301*. (B) Survival probability curves (1-CDF) from apparent dwell times of more >1,000 single-molecule chromatin-binding events for GAF^{L}-Halo in *WT*, *bap170* and *nurf301* mutants. (C) Survival probability curves (1-CDF) from apparent dwell times of more >1,000 single-molecule chromatin-binding events for GAF^{S}-Halo in *WT*, *bap170* and *nurf301* mutants. (D) One-component and two-component exponential fit of

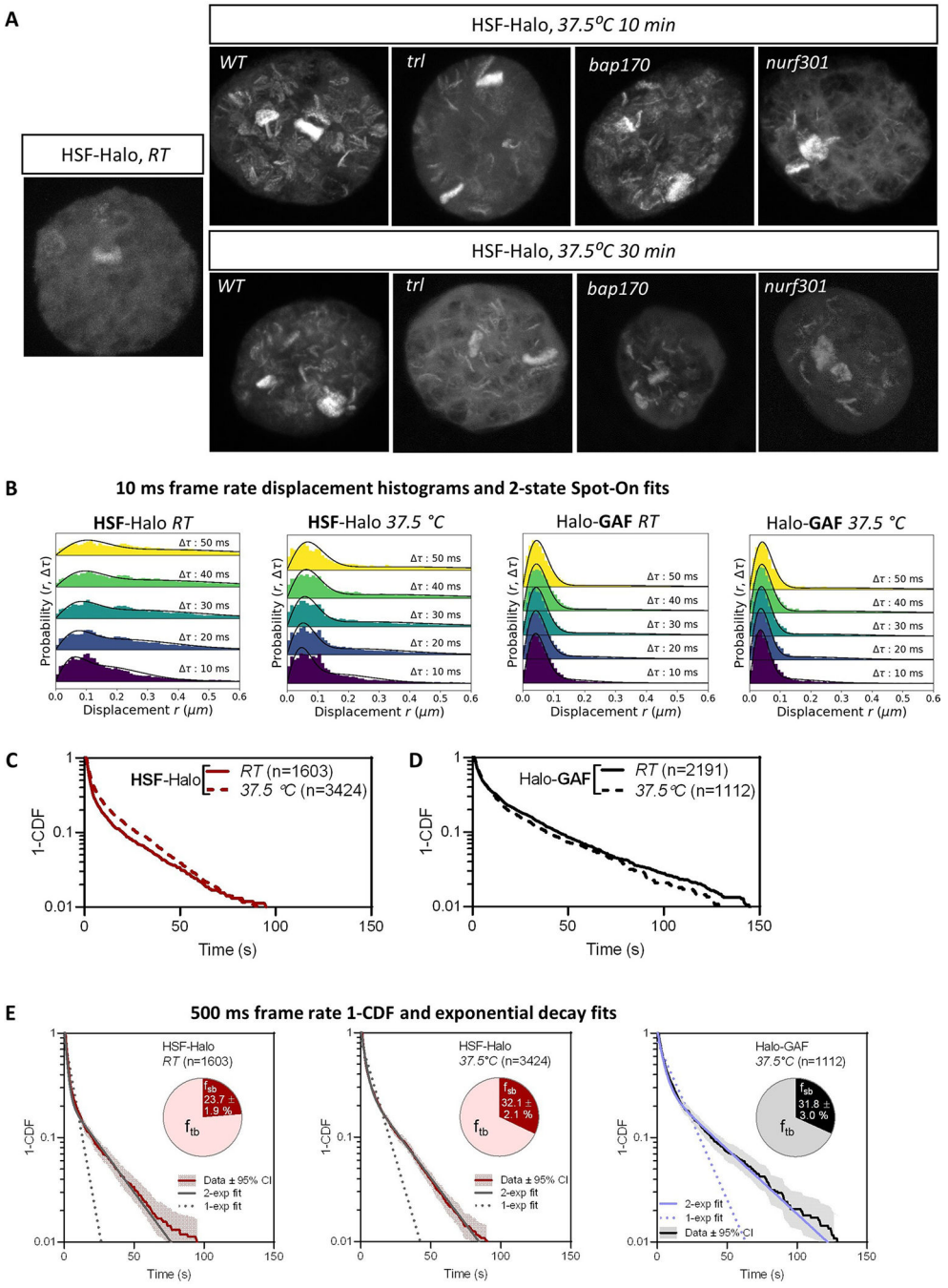
survival probabilities (1-CDF) from slow tracking data (with 95% CI, confidence interval) for GAF^{L}-Halo and GAF^{S}-Halo in *bap170* and *nurf301* mutants (see Extended Data Fig. 2d for *WT*). Pie charts show the stable-binding (f_{sb}) and transient-binding (f_{sb}) fractions derived from two-component fits. (E) Diffusion coefficients of bound fraction (D_{bound}) for GAF^{L}-Halo, GAF^{S}-Halo and Halo-H2B derived by Spot-On. Results are mean ± SD ($n = 3$ biological replicates for GAFL-Halo in nurf301 mutant and Halo-H2B, and $n = 4$ biological replicates for the remaining conditions). (F) Diffusion coefficients of free fraction (D_{free}) for GAF^{L}-Halo, GAF^{S}-Halo and Halo-H2B derived by Spot-On. Results are mean ± SD ($n = 3$ biological replicates for GAFL-Halo in nurf301 mutant and Halo-H2B, and $n = 4$ biological replicates for the remaining conditions).



Extended Data Fig. 6 |. Comparison of GAF ChiP-seq signals in LacZ RNAi (control) and Bap170 RNAi experiments of Judd et al. 2021.

(A) Comparison of GAF ChIP-seq signals in LacZ RNAi (control), Bap170 RNAi and GAF RNAi experiments (bw files from GSE157235) derived from Judd et al. 2021[28]. The left graph shows mean ChIP enrichment (mean ± SE) for all regions ±1 kb centered around transcription start sites (TSS); The right graphs show heat maps of all genes (generated by computeMatrix/plotHeatmap of deepTools[124]). Dashed rectangle indicates the top 10% regions with the highest GAF ChIP enrichment in control (for which the mean enrichment is plotted in (B)). (B) Comparison of GAF ChIP-seq, and ATAC-seq in LacZ RNAi (control),

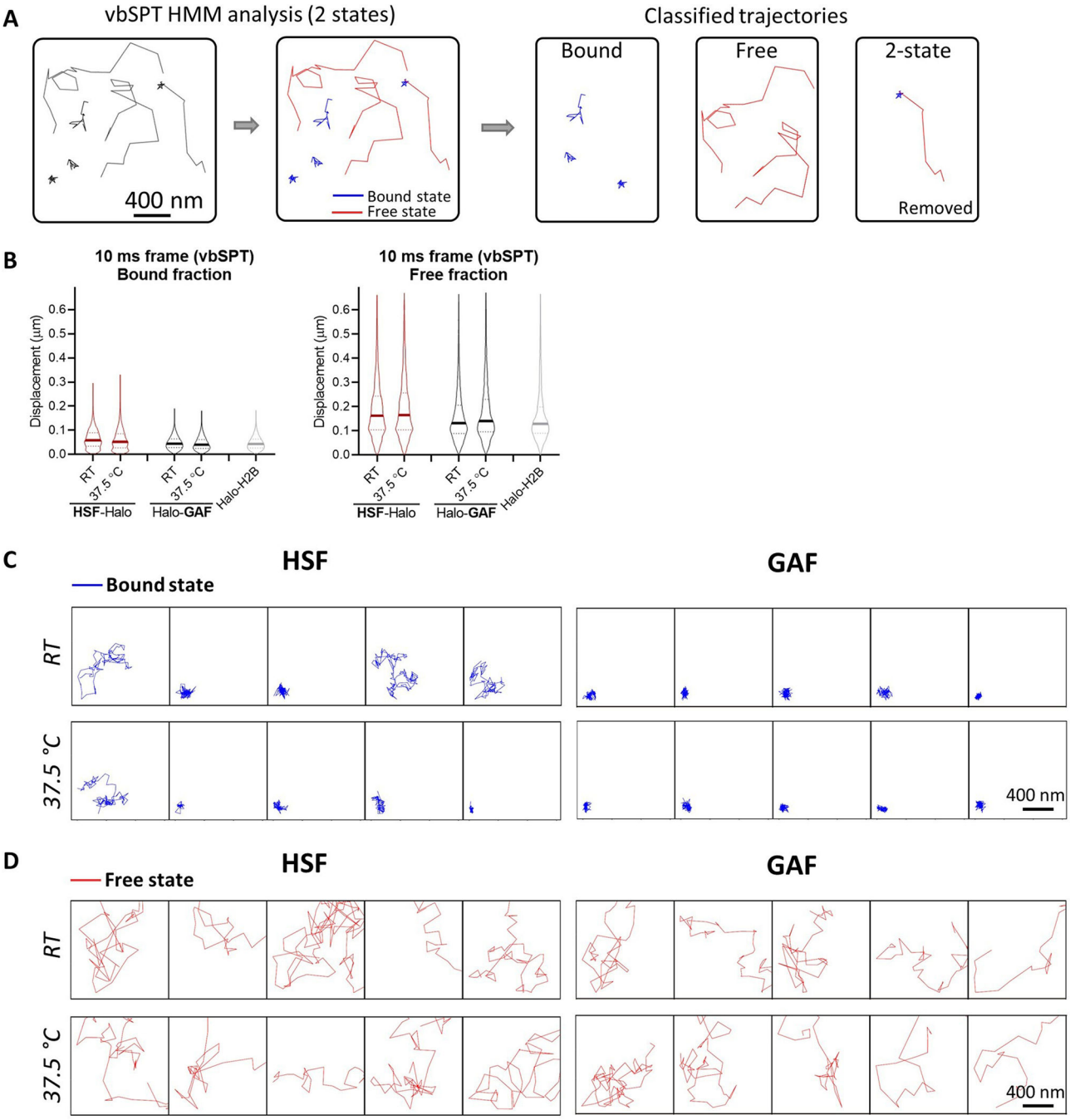
Bap170 RNAi and GAF RNAi experiments (bw files from GSE157235, and GSE149336, respectively) derived from Judd et al.[28]. Regions ±1 kb flanking TSS were sorted according to mean GAF ChIP enrichment in LacZ RNAi from high to low as shown in (A). Mean values of GAF ChIP enrichment (left column) and ATAC-seq (right column) enrichment are plotted for the top 1%, 5%, 10% of regions with the highest GAF ChIP signal and for the remaining 90% regions. For 3616 TSS-flanking regions with highest GAF ChIP enrichment, on average, chromatin accessibilities (ATAC-seq) are reduced in both *Bap170 RNAi* and *GAF RNAi* conditions, while the mean enrichment for GAF ChIP-seq shows no change in *Bap170 RNAi*. These analyses indicate that although there are differential effects at specific sites, the overall genome-wide GAF chromatin binding is not affected in PBAP-depleted condition.



Extended Data Fig. 7 |. HSF-Halo binding on polytene chromosomes and live-cell SPT at RT and HS conditions.

(A) Confocal images of HSF-Halo in fixed salivary glands. HSF-Halo is mostly nucleoplasmic at room temperature (*RT*) and binds to many loci after heat shock (HS) at 37.5 °C for 10 and 30 min. Maximum projections of confocal z-stacks are shown because major HSF bands are located in different focal planes. The pattern of HSF binding on heat shock is substantially reduced in *trl* and *nurf301* mutants and partially affected in the *bap170* mutant. Polytene loci showing little or no change of HSF binding in the *trl*

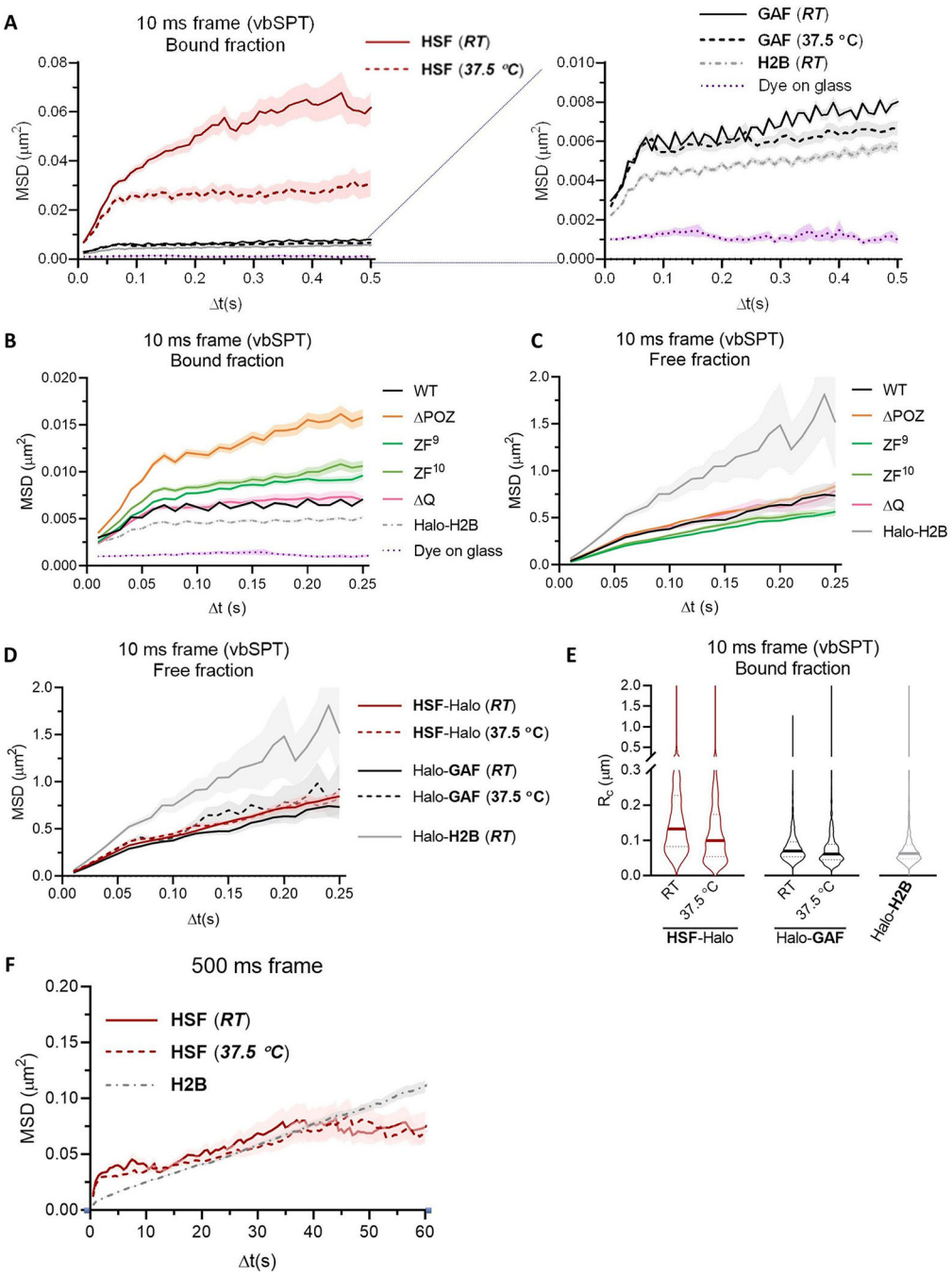
mutant is consistent with findings that not all HS genes are GAF-dependent[18]. (These genes presumably require an analogous pioneer factor and attendant remodelers). See methods for genotypes of *WT*, *trl*, *bap170* and *nurf301*. (B) Spot-On fits of fast-tracking data for HSF-Halo (RT, 37.5 °C) and Halo-GAF (37.5 °C, see Extended Data Fig. 1b for RT). (C) Survival-probability curves (1-CDF) from apparent dwell times of >1,000 single-molecule chromatin-binding events for HSF-Halo at RT and 37.5 °C. (D) Survival-probability curves (1-CDF) from apparent dwell times of >1,000 single-molecule chromatin-binding events for Halo-GAF at RT and 37.5 °C. (E) One-component and two-component exponential fit of survival probabilities (1-CDF) from slow tracking data (with 95% CI, confidence interval) for HSF-Halo (RT, 37.5 °C) and Halo-GAF (37.5 °C, see Extended Data Fig. 2d for *RT*). Pie charts show the stable-binding (f_{sb}) and transient-binding (f_{sb}) fractions derived from two-component fits.



Extended Data Fig. 8 |. vbSPT analysis of fast-tracking data for HSF-Halo and Halo-GAF at RT and HS conditions.

(A) Overview of fast-tracking trajectory classification with displacement-based HMM classification (vbSPT). After assigning each displacement as either in bound or free state, each trajectory is sub-classified as 'bound' or 'free', a small fraction of trajectories containing 2 states are excluded from the following analysis in (B–G) and (Fig. 5). (B) Violin plots of displacements show distinct distributions for bound and free trajectories classified by vbSPT. (C) Examples of bound trajectories at 10-ms frame rate classified by vbSPT for HSF-Halo, Halo-GAF at RT and 37.5 °C and Halo-H2B at RT. (D) Examples of

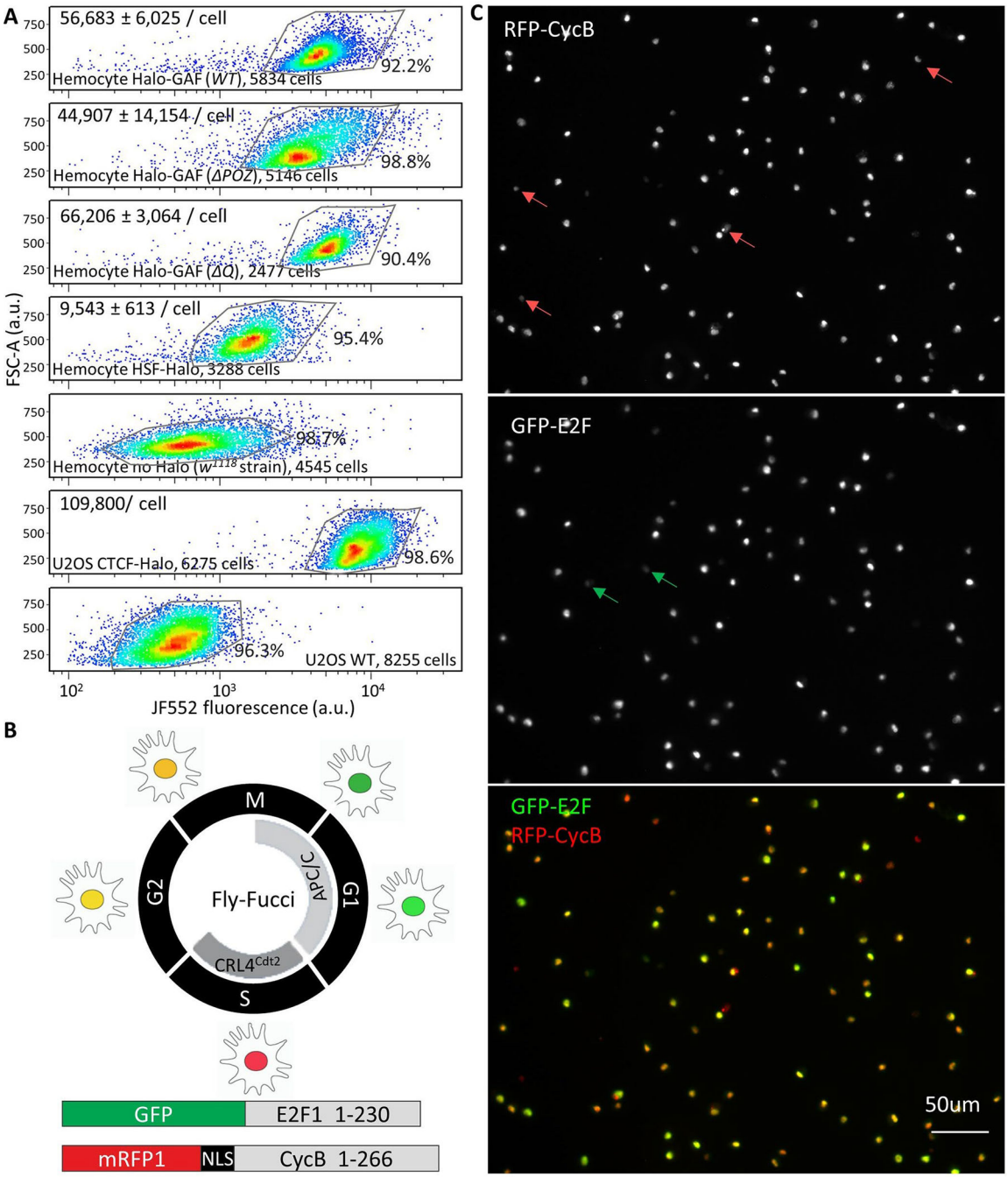
free trajectories at 10-ms frame rate classified by vbSPT for HSF-Halo, Halo-GAF at RT and 37.5 °C and Halo-H2B at RT.



Extended Data Fig. 9 |. MSD analysis of vbSPT-classified HSF-Halo and Halo-GAF fast-tracking trajectories.

(A) Plot of average MSD as a function of lag time Δt of bound trajectories classified by vbSPT for HSF-Halo, Halo-GAF at RT and 37.5 °C and Halo-H2B at RT. The right panel shows a zoomed-in section of the same plot. System noise is shown by MSD of dye molecules stuck on the coverglass. Mean and SE (shaded) are shown. (B) Average MSD

over Δt of bound trajectories at 10-ms frame rate classified by vbSPT for Halo-GAF *WT* and mutants, and Halo-H2B. Mean and SE (shaded) are shown. (C) Average MSD over Δ*t* of free trajectories at 10-ms frame rate classified by vbSPT for Halo-GAF *WT* and mutants, and Halo-H2B. Mean and SE (shaded) are shown. (D) Average MSD over Δ*t* of free trajectories at 10-ms frame rate classified by vbSPT for HSF-Halo, Halo-GAF at RT and 37.5 °C and Halo-H2B at RT. Mean and SE (shaded) are shown. (E) Radius of confinement (Rc) is derived by fitting individual MSD curves with a confined diffusion model, for bound trajectories at 10-ms frame rate classified by vbSPT, for HSF-Halo, Halo-GAF at RT and 37.5 °C, and Halo-H2B at RT. (F) Average MSD over Δ*t* for *WT* and Halo-GAF mutants at 500-ms frame rate. Mean and SE (shaded) are shown.



Extended Data Fig. 10 |. FACS quantitation of Halo-GAF and HSF-Halo in *Drosophila* hemocytes and cell cycle phase identification.

(A) Total Halo-GAF (knock-in *WT*, *ΔPOZ* and *ΔQ*) and HSF-Halo (transgenic in the *P{PZ}Hsf*^{03091}/*Hsf*^{3} background) fluorescence per cell for JF552-stained late 3rd instar larval hemocytes and CTCT-Halo in U2OS cells quantified by flow cytometry. Cellular abundance of Halo-GAF and HSF-Halo molecules are estimated using CTCF-Halo in U2OS cells as a standard (see methods)[87,119]. Hemocytes (*w*^{1118} strain) or U2OS cells not expressing HaloTag were used as controls for background subtraction. One of three representative flow cytometry experiments is shown. Mean ± SD of estimated protein

abundance is shown at the upper left corner of each plot. A much larger number of molecules (in the order of one million) for GAF was reported earlier in the S2 cell line[125]; the reason for the discrepancy is unclear. FSC-A, forward scatter area. (B) Conceptual diagram of the Fly-FUCCI system[126]. Both GFP-E2F1$_{1–230}$ and mRFP1-CycB$_{1–266}$ are expressed with the GAL4/UAS system. In early M phase, both GFP-E2F1$_{1–230}$ and mRFP1-CycB$_{1–266}$ are present and thus display yellow. In mid-mitosis, the APC/C marks mRFP1-CycB$_{1–266}$ for proteasomal degradation, leaving the cells fluorescing green. As cells progress from G1 to S phase, CRL4^{Cdt2} degrades GFP-E2F1$_{1–230}$, and cells are labeled red due to the presence of mRFP1-CycB$_{1–266}$ only. After cells enter G2 phase, GFP-E2F1$_{1–230}$ protein levels reaccumulate, marking the cells yellow due to the presence of mRFP1-CycB$_{1–266}$. (C) Characterization of cell-cycle stage for late 3rd instar larval hemocytes. Only 4 out of 96 cells in the field of view show ‘red only’ fluorescence (S phase), and 2 cells have ‘green only’ fluorescence (M to G1 phase). The majority of hemocytes have both red and green fluorescence, indicating G2 phase or early M phase. Given that a previous study shows only 0.32% of larval hemocytes stain positive with the mitotic phosH3 antibody[127], we conclude that most larval hemocytes are in the G2 phase.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank A. Hansen and M. Woringer for discussion on Spot-On analyses, A. Hansen, C. Cattoglio, X. Darzacq and R. Tjian for the CTCF-Halo U2OS cell line and discussions on measuring factor abundance, T. Lionnet, J.Z. Liu, P. Dong and B. Mehl for advice on SPT, Y.H. Ling and S.J. Yoo for assistance with data analysis, P. Vallotton for customization of DiaTrack software, P. Badenhorst for advice on larval hemocyte culture, S. Deluca for advice on bioinformatics, W. Dai for imaging assistance, E. Pryce and the Integrated Imaging Center for LSM 800 training, Y. Liu for assembly and maintenance of a high performance computational platform, J. Lis, M. Levine and G. Hager for discussion, and Wu Laboratory members for comments on the manuscript. This study was supported by HHMI funding to the TIC (C.W., Q.Z. and L.L.), Johns Hopkins Bloomberg Distinguished Professorship funds (C.W.), and National Institutes of Health grant nos. GM132290–01 (C.W.) and DK127432 (C.W.).

Data availability

All raw SPT trajectory coordinates have been deposited at the 4D Nucleome Data Portal: https://data.4dnucleome.org/tang_et_al_2021. Due to the large number and size of files, original videos will be available upon request. GSE157235 and GSE149336 are the published ChIP–seq and ATAC-seq data sets, respectively, used for analysis in this study and both are available at https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE149339. Source data are provided with this paper.

References

1. Biggin MD & Tjian R Transcription factors that activate the Ultrabithorax promoter in developmentally staged extracts. Cell 53, 699–711 (1988). [PubMed: 2897243]
2. Gilmour DS, Thomas GH & Elgin SC *Drosophila* nuclear proteins bind to regions of alternating C and T residues in gene promoters. Science 245, 1487–1490 (1989). [PubMed: 2781290]
3. Farkas G et al. The Trithorax-like gene encodes the *Drosophila* GAGA factor. Nature 371, 806–808 (1994). [PubMed: 7935842]

4. Soeller WC, Oh CE & Kornberg TB Isolation of cDNAs encoding the *Drosophila* GAGA transcription factor. Mol. Cell. Biol. 13, 7961–7970 (1993). [PubMed: 7504178]
5. Chopra VS et al. Transcriptional activation by GAGA factor is through its direct interaction with dmTAF3. Dev. Biol. 317, 660–670 (2008). [PubMed: 18367161]
6. Lomaev D et al. The GAGA factor regulatory network: identification of GAGA factor associated proteins. PLoS ONE 12, e0173602 (2017). [PubMed: 28296955]
7. Giot L et al. A protein interaction map of *Drosophila melanogaster*. Science 302, 1727–1736 (2003). [PubMed: 14605208]
8. Li J et al. Kinetic competition between elongation rate and binding of NELF controls promoter-proximal pausing. Mol. Cell 50, 711–722 (2013). [PubMed: 23746353]
9. Croston GE, Kerrigan LA, Lira LM, Marshak DR & Kadonaga JT Sequence-specific antirepression of histone H1-mediated inhibition of basal RNA polymerase II transcription. Science 251, 643–649 (1991). [PubMed: 1899487]
10. Cirillo LA & Zaret KS An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. Mol. Cell 4, 961–969 (1999). [PubMed: 10635321]
11. Soufi A et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. Cell 161, 555–568 (2015). [PubMed: 25892221]
12. Tsukiyama T, Becker PB & Wu C ATP-dependent nucleosome disruption at a heat-shock promoter mediated by binding of GAGA transcription factor. Nature 367, 525–532 (1994). [PubMed: 8107823]
13. Okada M & Hirose S Chromatin remodeling mediated by *Drosophila* GAGA factor and ISWI activates fushi tarazu gene transcription in vitro. Mol. Cell. Biol. 18, 2455–2461 (1998). [PubMed: 9566866]
14. Nakayama T, Shimojima T & Hirose S The PBAP remodeling complex is required for histone H3.3 replacement at chromatin boundaries and for boundary functions. Development 139, 4582–4590 (2012). [PubMed: 23136390]
15. Xiao H et al. Dual functions of largest NURF subunit NURF301 in nucleosome sliding and transcription factor interactions. Mol. Cell 8, 531–543 (2001). [PubMed: 11583616]
16. Sherwood RI et al. Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. Nat. Biotechnol. 32, 171–178 (2014). [PubMed: 24441470]
17. Fuda NJ et al. GAGA factor maintains nucleosome-free regions and has a role in RNA polymerase II recruitment to promoters. PLoS Genet. 11, e1005108 (2015). [PubMed: 25815464]
18. Duarte FM et al. Transcription factors GAF and HSF act at distinct regulatory steps to modulate stress-induced gene activation. Genes Dev. 30, 1731–1746 (2016). [PubMed: 27492368]
19. Zaret KS Pioneer transcription factors initiating gene network changes. Annu. Rev. Genet. 54, 367–385 (2020). [PubMed: 32886547]
20. Hansen JL, Loell KJ, Cohen BA A test of the pioneer factor hypothesis using ectopic liver gene activation. eLife 10.7554/eLife.73358 (2022).
21. Swinstead EE et al. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. Cell 165, 593–605 (2016). [PubMed: 27062924]
22. Johnson TA et al. Conventional and pioneer modes of glucocorticoid receptor interaction with enhancer chromatin in vivo. Nucleic Acids Res. 46, 203–214 (2018). [PubMed: 29126175]
23. Badenhorst P, Voas M, Rebay I & Wu C Biological functions of the ISWI chromatin remodeling complex NURF. Genes Dev. 16, 3186–3198 (2002). [PubMed: 12502740]
24. Boija A et al. CBP regulates recruitment and release of promoter-proximal RNA Polymerase II. Mol. Cell 68, 491–503.e5 (2017). [PubMed: 29056321]
25. Kuroda MI, Kang H, De S & Kassis JA Dynamic competition of polycomb and trithorax in transcriptional programming. Annu. Rev. Biochem. 89, 235–253 (2020). [PubMed: 31928411]
26. Wolle D et al. Functional requirements for Fab-7 boundary activity in the bithorax complex. Mol. Cell. Biol. 35, 3739–3752 (2015). [PubMed: 26303531]

27. van Steensel B, Delrow J & Bussemaker HJ Genomewide analysis of *Drosophila* GAGA factor target genes reveals context-dependent DNA binding. Proc. Natl Acad. Sci. USA 100, 2580–2585 (2003). [PubMed: 12601174]
28. Judd J, Duarte FM & Lis JT Pioneer-like factor GAF cooperates with PBAP (SWI/SNF) and NURF (ISWI) to regulate transcription. Genes Dev. 35, 147–156 (2021). [PubMed: 33303640]
29. Gaskill MM, Gibson TJ, Larson ED & Harrison MM GAF is essential for zygotic genome activation and chromatin accessibility in the early embryo. eLife 10.7554/eLife.66668 (2021).
30. Benyajati C et al. Multiple isoforms of GAGA factor, a critical component of chromatin structure. Nucleic Acids Res. 25, 3345–3353 (1997). [PubMed: 9241251]
31. Greenberg AJ & Schedl P GAGA factor isoforms have distinct but overlapping functions in vivo. Mol. Cell. Biol. 21, 8565–8574 (2001). [PubMed: 11713290]
32. Tepass U, Fessler LI, Aziz A & Hartenstein V Embryonic origin of hemocytes and their relationship to cell death in *Drosophila*. Development 120, 1829–1837 (1994). [PubMed: 7924990]
33. Read D, Butte MJ, Dernburg AF, Frasch M & Kornberg TB Functional studies of the BTB domain in the Drosophila GAGA and Mod(mdg4) proteins. Nucleic Acids Res. 28, 3864–3870 (2000). [PubMed: 11024164]
34. Raff JW, Kellum R & Alberts B The *Drosophila* GAGA transcription factor is associated with specific regions of heterochromatin throughout the cell cycle. EMBO J. 13, 5977–5983 (1994). [PubMed: 7813435]
35. Bellec M et al. The control of transcriptional memory by stable mitotic bookmarking. Nat. Commun. 13, 1176 (2022). [PubMed: 35246556]
36. Rust MJ, Bates M & Zhuang X Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (STORM). Nat. Methods 3, 793–795 (2006). [PubMed: 16896339]
37. Heilemann M et al. Subdiffraction-resolution fluorescence imaging with conventional fluorescent probes. Angew. Chem. Int. Ed. Engl. 47, 6172–6176 (2008). [PubMed: 18646237]
38. Chen J et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. Cell 156, 1274–1285 (2014). [PubMed: 24630727]
39. Hansen AS et al. Robust model-based analysis of single-particle tracking experiments with Spot-On. eLife 10.7554/eLife.33125 (2018).
40. Chong S et al. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science 361, eaar2555 (2018). [PubMed: 29930090]
41. Lerner J et al. Two-parameter mobility assessments discriminate diverse regulatory factor behaviors in chromatin. Mol. Cell 79, 677–688.e6 (2020). [PubMed: 32574554]
42. Liu H et al. Visualizing long-term single-molecule dynamics in vivo by stochastic protein labeling. Proc. Natl Acad. Sci. USA 115, 343–348 (2018). [PubMed: 29284749]
43. Hansen AS, Pustova I, Cattoglio C, Tjian R & Darzacq X CTCF and cohesin regulate chromatin loop stability with distinct dynamics. eLife 10.7554/eLife.25776 (2017).
44. Espinás ML et al. The N-terminal POZ domain of GAGA mediates the formation of oligomers that bind DNA with high affinity and specificity. J. Biol. Chem. 274, 16461–16469 (1999). [PubMed: 10347208]
45. Katsani KR, Hajibagheri MA & Verrijzer CP Co-operative DNA binding by GAGA transcription factor requires the conserved BTB/POZ domain and reorganizes promoter topology. EMBO J. 18, 698–708 (1999). [PubMed: 9927429]
46. Wilkins RC & Lis JT DNA distortion and multimerization: novel functions of the glutamine-rich domain of GAGA factor. J. Mol. Biol. 285, 515–525 (1999). [PubMed: 9878426]
47. Mazza D, Abernathy A, Golob N, Morisaki T & McNally JG A benchmark for chromatin binding measurements in live cells. Nucleic Acids Res. 40, e119 (2012). [PubMed: 22844090]
48. Tsukiyama T & Wu C Purification and properties of an ATP-dependent nucleosome remodeling factor. Cell 83, 1011–1020 (1995). [PubMed: 8521501]
49. Badenhorst P et al. The *Drosophila* nucleosome remodeling factor NURF is required for Ecdysteroid signaling and metamorphosis. Genes Dev. 19, 2540–2545 (2005). [PubMed: 16264191]

50. Chalkley GE et al. The transcriptional coactivator SAYP is a trithorax group signature subunit of the PBAP chromatin remodeling complex. Mol. Cell. Biol. 28, 2920–2929 (2008). [PubMed: 18299390]
51. Yen K, Vinayachandran V, Batta K, Koerber RT & Pugh BF Genome-wide nucleosome specificity and directionality of chromatin remodelers. Cell 149, 1461–1473 (2012). [PubMed: 22726434]
52. Shopland LS, Hirayoshi K, Fernandes M & Lis JT HSF access to heat shock elements in vivo depends critically on promoter architecture defined by GAGA factor, TFIID, and RNA polymerase II binding sites. Genes Dev. 9, 2756–2769 (1995). [PubMed: 7590251]
53. Lis JT, Mason P, Peng J, Price DH & Werner J P-TEFb kinase recruitment and function at heat shock loci. Genes Dev. 14, 792–803 (2000). [PubMed: 10766736]
54. Guertin MJ & Lis JT Chromatin landscape dictates HSF binding to target DNA elements. PLoS Genet. 6, e1001114 (2010). [PubMed: 20844575]
55. Lis J & Wu C Protein traffic on the heat shock promoter: parking, stalling, and trucking along. Cell 74, 1–4 (1993). [PubMed: 8334697]
56. Vuister GW et al. Solution structure of the DNA-binding domain of *Drosophila* heat shock transcription factor. Nat. Struct. Biol. 1, 605–614 (1994). [PubMed: 7634100]
57. Kim SJ, Tsukiyama T, Lewis MS & Wu C Interaction of the DNA-binding domain of *Drosophila* heat shock factor with its cognate DNA site: a thermodynamic analysis using analytical ultracentrifugation. Protein Sci. 3, 1040–1051 (1994). [PubMed: 7920249]
58. Westwood JT, Clos J & Wu C Stress-induced oligomerization and chromosomal relocalization of heat-shock factor. Nature 353, 822–827 (1991). [PubMed: 1944557]
59. Baler R, Dahl G & Voellmy R Activation of human heat shock genes is accompanied by oligomerization, modification, and rapid translocation of heat shock transcription factor HSF1. Mol. Cell. Biol. 13, 2486–2496 (1993). [PubMed: 8455624]
60. Sarge KD, Murphy SP & Morimoto RI Activation of heat shock gene transcription by heat shock factor 1 involves oligomerization, acquisition of DNA-binding activity, and nuclear localization and can occur in the absence of stress. Mol. Cell. Biol. 13, 1392–1407 (1993). [PubMed: 8441385]
61. Perisic O, Xiao H & Lis JT Stable binding of *Drosophila* heat shock factor to head-to-head and tail-to-tail repeats of a conserved 5 bp recognition unit. Cell 59, 797–806 (1989). [PubMed: 2590940]
62. Neudegger T, Verghese J, Hayer-Hartl M, Hartl FU & Bracher A Structure of human heat-shock transcription factor 1 in complex with DNA. Nat. Struct. Mol. Biol. 23, 140–146 (2016). [PubMed: 26727489]
63. Westwood JT & Wu C Activation of *Drosophila* heat shock factor: conformational change associated with a monomer-to-trimer transition. Mol. Cell. Biol. 13, 3481–3486 (1993). [PubMed: 8497263]
64. Spradling AC et al. The Berkeley Drosophila Genome Project gene disruption project: single P-element insertions mutating 25% of vital *Drosophila* genes. Genetics 153, 135–177 (1999). [PubMed: 10471706]
65. Jedlicka P, Mortin MA & Wu C Multiple functions of *Drosophila* heat shock transcription factor in vivo. EMBO J. 16, 2452–2462 (1997). [PubMed: 9171358]
66. Yao J, Munson KM, Webb WW & Lis JT Dynamics of heat shock factor association with native gene loci in living cells. Nature 442, 1050–1053 (2006). [PubMed: 16929308]
67. Chowdhary S, Kainth AS, Pincus D & Gross DS Heat Shock Factor 1 drives intergenic association of its target gene loci upon heat shock. Cell Rep. 26, 18–28.e5 (2019). [PubMed: 30605674]
68. Gaglia G et al. HSF1 phase transition mediates stress adaptation and cell fate decisions. Nat. Cell Biol. 22, 151–158 (2020). [PubMed: 32015439]
69. Persson F, Lindén M, Unoson C & Elf J Extracting intracellular diffusive states and transition rates from single-molecule tracking data. Nat. Methods 10, 265–269 (2013). [PubMed: 23396281]
70. Hansen AS, Amitai A, Cattoglio C, Tjian R & Darzacq X Guided nuclear exploration increases CTCF target search efficiency. Nat. Chem. Biol. 16, 257–266 (2020). [PubMed: 31792445]
71. Zentner GE & Henikoff S High-resolution digital profiling of the epigenome. Nat. Rev. Genet. 15, 814–827 (2014). [PubMed: 25297728]

72. Kasinathan S, Orsi GA, Zentner GE, Ahmad K & Henikoff S High-resolution mapping of transcription factor binding sites on native chromatin. Nat. Methods 11, 203–209 (2014). [PubMed: 24336359]
73. Tilly BC et al. In vivo analysis reveals that ATP-hydrolysis couples remodeling to SWI/SNF release from chromatin. eLife 10.7554/eLife.69424 (2021).
74. Iurlaro M et al. Mammalian SWI/SNF continuously restores local accessibility to chromatin. Nat. Genet. 53, 279–287 (2021). [PubMed: 33558757]
75. Schick S et al. Acute BAF perturbation causes immediate changes in chromatin accessibility. Nat. Genet. 53, 269–278 (2021). [PubMed: 33558760]
76. Kim JM et al. Single-molecule imaging of chromatin remodelers reveals role of ATPase in promoting fast kinetics of target search and dissociation from chromatin. eLife 10.7554/eLife.69387 (2021).
77. Jeronimo C et al. FACT is recruited to the +1 nucleosome of transcribed genes and spreads in a Chd1-dependent manner. Mol. Cell 10.1016/j.molcel.2021.07.010 (2021).
78. Nguyen VQ et al. Spatiotemporal coordination of transcription preinitiation complex assembly in live cells. Mol. Cell 10.1016/j.molcel.2021.07.022 (2021).
79. Kwon SY, Grisan V, Jang B, Herbert J & Badenhorst P 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 (2016). [PubMed: 27046080]
80. Liu Z & Tjian R Visualizing transcription factor dynamics in living cells. J. Cell Biol. 217, 1181–1191 (2018). [PubMed: 29378780]
81. Wall G, Varga-Weisz PD, Sandaltzopoulos R & Becker PB Chromatin remodeling by GAGA factor and heat shock factor at the hypersensitive *Drosophila* hsp26 promoter in vitro. EMBO J. 14, 1727–1736 (1995). [PubMed: 7737124]
82. Chetverina D, Erokhin M & Schedl P GAGA factor: a multifunctional pioneering chromatin protein. Cell. Mol. Life Sci. 78, 4125–4141 (2021). [PubMed: 33528710]
83. von Hippel PH & Berg OG Facilitated target location in biological systems. J. Biol. Chem. 264, 675–678 (1989). [PubMed: 2642903]
84. Schneider I Cell lines derived from late embryonic stages of *Drosophila melanogaster*. J. Embryol. Exp. Morphol. 27, 353–365 (1972). [PubMed: 4625067]
85. Cherbas L et al. The transcriptional diversity of 25 *Drosophila* cell lines. Genome Res. 21, 301–314 (2011). [PubMed: 21177962]
86. Tsai S-Y, Chang Y-L, Swamy KBS, Chiang R-L & Huang D-H GAGA factor, a positive regulator of global gene expression, modulates transcriptional pausing and organization of upstream nucleosomes. Epigenetics Chromatin 9, 32 (2016). [PubMed: 27468311]
87. Cattoglio C et al. Determining cellular CTCF and cohesin abundances to constrain 3D genome models. eLife 10.7554/eLife.40164 (2019).
88. Schwendemann A & Lehmann M Pipsqueak and GAGA factor act in concert as partners at homeotic and many other loci. Proc. Natl Acad. Sci. USA 99, 12883–12888 (2002). [PubMed: 12271134]
89. Gutierrez-Perez I et al. Ecdysone-induced 3D chromatin reorganization involves active enhancers bound by Pipsqueak and Polycomb. Cell Rep. 28, 2715–2727.e5 (2019). [PubMed: 31484080]
90. Duan J et al. CLAMP and Zelda function together to promote zygotic genome activation. eLife 10.7554/eLife.69937 (2021).
91. Kaye EG et al. Differential occupancy of two GA-binding proteins promotes targeting of the *Drosophila* dosage compensation complex to the male X chromosome. Cell Rep. 22, 3227–3239 (2018). [PubMed: 29562179]
92. Tattikota SG et al. A single-cell survey of blood. eLife 10.7554/eLife.54818 (2020).
93. Agianian B et al. The glutamine-rich domain of the *Drosophila* GAGA factor is necessary for amyloid fibre formation in vitro, but not for chromatin remodelling. J. Mol. Biol. 285, 527–544 (1999). [PubMed: 9878427]

94. Sekiya T, Muthurajan UM, Luger K, Tulin AV & Zaret KS Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. Genes Dev. 23, 804–809 (2009). [PubMed: 19339686]

95. Brodsky S et al. Intrinsically disordered regions direct transcription factor in vivo binding specificity. Mol. Cell 79, 459–471.e4 (2020). [PubMed: 32553192]

96. Chen Y et al. Mechanisms governing target search and binding dynamics of hypoxia-inducible factors. Preprint at bioRxiv 10.1101/2021.10.27.466110 (2021).

97. Chaharbakhshi E & Jemc JC Broad-complex, tramtrack, and bric-à-brac (BTB) proteins: critical regulators of development. Genesis 54, 505–518 (2016). [PubMed: 27521773]

98. Mahmoudi T, Katsani KR & Verrijzer CP GAGA can mediate enhancer function in trans by linking two separate DNA molecules. EMBO J. 21, 1775–1781 (2002). [PubMed: 11927561]

99. Paakinaho V, Johnson TA, Presman DM & Hager GL Glucocorticoid receptor quaternary structure drives chromatin occupancy and transcriptional outcome. Genome Res. 29, 1223–1234 (2019). [PubMed: 31337711]

100. Biggin MD Animal transcription networks as highly connected, quantitative continua. Dev. Cell 21, 611–626 (2011). [PubMed: 22014521]

101. Fisher WW et al. DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. Proc. Natl Acad. Sci. USA 109, 21330–21335 (2012). [PubMed: 23236164]

102. Kal AJ, Mahmoudi T, Zak NB & Verrijzer CP The *Drosophila* brahma complex is an essential coactivator for the trithorax group protein zeste. Genes Dev. 14, 1058–1071 (2000). [PubMed: 10809665]

103. Dufourt J et al. Temporal control of gene expression by the pioneer factor Zelda through transient interactions in hubs. Nat. Commun. 9, 5194 (2018). [PubMed: 30518940]

104. Mir M et al. Dynamic multifactor hubs interact transiently with sites of active transcription in embryos. eLife 10.7554/eLife.40497 (2018).

105. McSwiggen DT, Mir M, Darzacq X & Tjian R Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences. Genes Dev. 33, 1619–1634 (2019). [PubMed: 31594803]

106. Omichinski JG, Pedone PV, Felsenfeld G, Gronenborn AM & Clore GM The solution structure of a specific GAGA factor-DNA complex reveals a modular binding mode. Nat. Struct. Biol. 4, 122–132 (1997). [PubMed: 9033593]

107. Carrera I, Zavadil J & Treisman JE Two subunits specific to the PBAP chromatin remodeling complex have distinct and redundant functions during drosophila development. Mol. Cell. Biol. 28, 5238–5250 (2008). [PubMed: 18573871]

108. Heinz S et al. Simple combinations of lineage-determining transcription factors prime *cis*-regulatory elements required for macrophage and B cell identities. Mol. Cell 38, 576–589 (2010). [PubMed: 20513432]

References

109. Bier E, Harrison MM, O'Connor-Giles KM & Wildonger J Advances in engineering the fly genome with the CRISPR-Cas system. Genetics 208, 1–18 (2018). [PubMed: 29301946]

110. Sharan SK, Thomason LC, Kuznetsov SG & Court DL Recombineering: a homologous recombination-based method of genetic engineering. Nat. Protoc. 4, 206–223 (2009). [PubMed: 19180090]

111. Zhang Y, Schreiner W & Rong YS Genome manipulations with bacterial recombineering and site-specific integration in *Drosophila*. Methods Mol. Biol. 1114, 11–24 (2014). [PubMed: 24557894]

112. Groth AC, Fish M, Nusse R & Calos MP Construction of transgenic *Drosophila* by using the site-specific integrase from phage phiC31. Genetics 166, 1775–1782 (2004). [PubMed: 15126397]

113. Ranjan A et al. Live-cell single particle imaging reveals the role of RNA polymerase II in histone H2A.Z eviction. eLife 10.7554/eLife.55667 (2020).

114. Schindelin J et al. Fiji: an open-source platform for biological-image analysis. Nat. Methods 9, 676–682 (2012). [PubMed: 22743772]
115. Vallotton P & Olivier S Tri-track: free software for large-scale particle tracking. Microsc. Microanal. 19, 451–460 (2013). [PubMed: 23448973]
116. Wieser S & Schütz GJ Tracking single molecules in the live cell plasma membrane-do's and don't's. Methods 46, 131–140 (2008). [PubMed: 18634880]
117. Loffreda A et al. Live-cell p53 single-molecule binding is modulated by C-terminal acetylation and correlates with transcriptional activity. Nat. Commun. 8, 313 (2017). [PubMed: 28827596]
118. Tatavosian R et al. Live-cell single-molecule dynamics of PcG proteins imposed by the DIPG H3.3K27M mutation. Nat. Commun. 9, 2080 (2018). [PubMed: 29802243]
119. Cattoglio C, Darzacq X, Tjian R & Hansen AS Estimating cellular abundances of Halo-tagged proteins in live mammalian cells by flow. Cytometry. Bio. Protoc. 10, e3527 (2020).
120. Thévenaz P, Ruttimann UE & Unser M A pyramid approach to subpixel registration based on intensity. IEEE Trans. Image Process. 7, 27–41 (1998). [PubMed: 18267377]
121. Jumper J et al. Highly accurate protein structure prediction with AlphaFold. Nature 596, 583–589 (2021). [PubMed: 34265844]
122. Ahmad KF, Engel CK & Privé GG Crystal structure of the BTB domain from PLZF. Proc. Natl Acad. Sci. USA 95, 12123–12128 (1998). [PubMed: 9770450]
123. Li X et al. Structure-function studies of the BTB/POZ transcriptional repression domain from the promyelocytic leukemia zinc finger oncoprotein. Cancer Res. 59, 5275–5282 (1999). [PubMed: 10537309]
124. Ramírez F et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 44, W160–W165 (2016). [PubMed: 27079975]
125. Wilkins RC & Lis JT GAGA factor binding to DNA via a single trinucleotide sequence element. Nucleic Acids Res. 26, 2672–2678 (1998). [PubMed: 9592153]
126. Zielke N et al. Fly-FUCCI: a versatile tool for studying cell proliferation in complex tissues. Cell Rep. 7, 588–598 (2014). [PubMed: 24726363]
127. Asha H et al. Analysis of Ras-induced overproliferation in *Drosophila* hemocytes. Genetics 163, 203–215 (2003). [PubMed: 12586708]

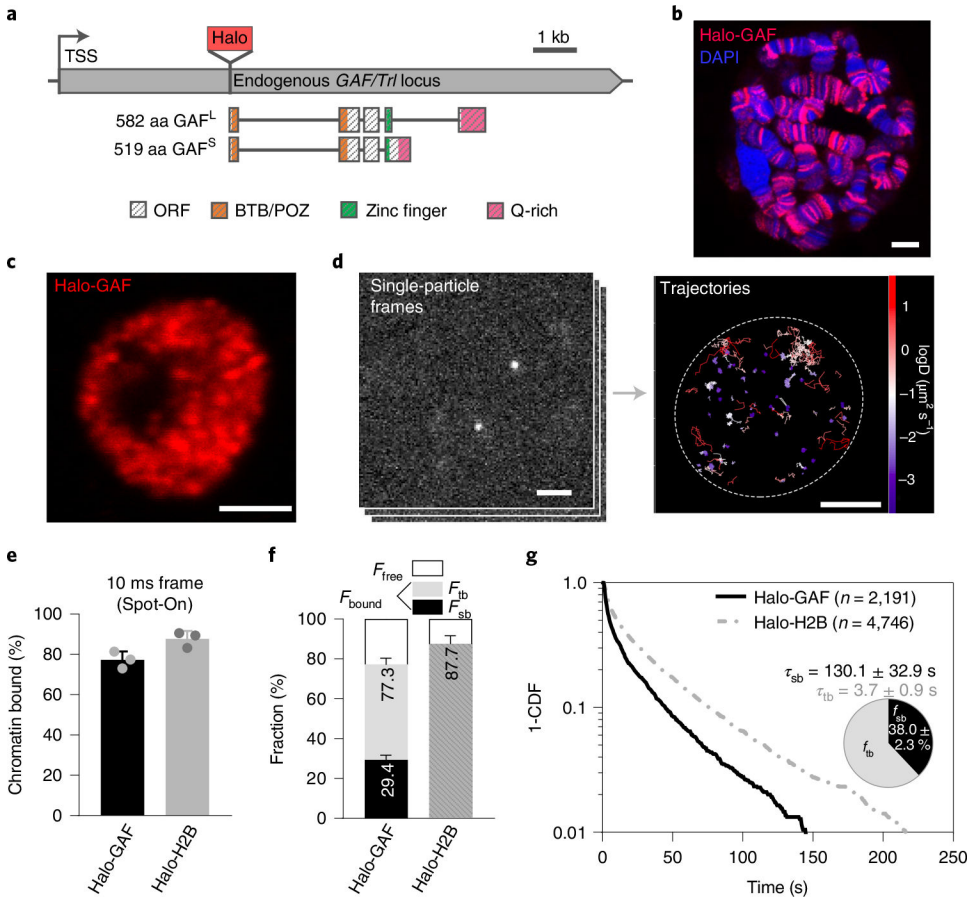


Fig. 1 |. Chromatin-binding dynamics of GAF shown by SPT in live *Drosophila* hemocytes.

a, Diagram of HaloTag (Halo) knock-in at the N terminus of endogenous *GAF/Trl* locus. Protein domains in color. TSS, transcription start site. C-terminal GAF-HaloTag fusions for two splicing isoforms GAF^S (short) and GAF^L (long) are shown in Supplementary Fig. 1. b, Halo-GAF binds to specific polytene chromosome loci in fixed third instar larval salivary gland nuclei, revealed by JF552 fluorescence (red); DNA counterstained with DAPI (blue). Scale bar, 5 μm. **c**, Confocal distribution of Halo-GAF (JF552) foci in fixed third instar larval hemocyte nuclei (predominantly diploid plasmatocytes in G2 phase of the cell cycle). Scale bar, 2 μm. **d**, SPT frames and superimposed fast-tracking trajectories of Halo-GAF in live hemocytes, color-coded according to diffusion coefficients. Dashed oval marks the nucleus. Scale bar, 2 μm. **e**, Spot-On kinetic modeling of fast-tracking data shows chromatin bound and free fractions for Halo-GAF and Halo-H2B. Results are mean ± s.d from three biological replicates. Scale bar, 2 μm. **f**, Chromatin-free fraction (F_{free}), global stable- and transient-binding fractions (F_{sb} and F_{tb}) of Halo-GAF extracted from fast- and slow-tracking data ($n = 3$ biological replicates, **e**) and ($n = 100$ resamplings, **g**), Results are mean ± s.d. with error propagation for F_{sb} from the error bars in **e** and **g**. **g**, Survival-probability curves (1-CDF) plotted from apparent dwell times of thousands (n) of single-particle chromatin-binding events for Halo-GAF. Average residence times for stable- (τ_{sb}) and transient- (τ_{tb}) binding by Halo-GAF are corrected by using Halo-H2B as

a standard. Pie charts show stable- (f_{sb}) and transient-binding (f_{tb}) fractions. Mean and s.d. from bootstrap after resampling 100 times ($n = 100$) are provided.

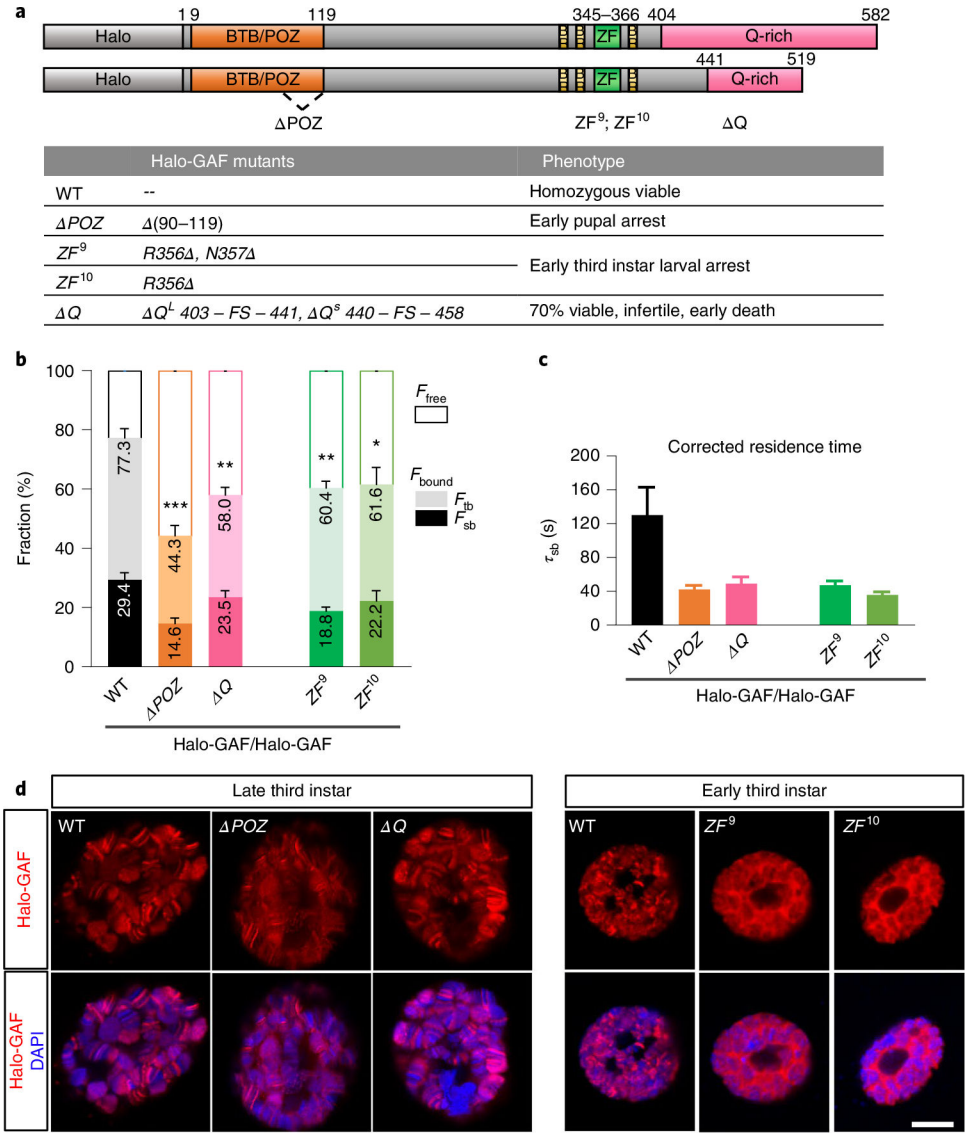


Fig. 2 |. PoZ, Q-rich and DBD domains of GAF all contribute to stable chromatin binding and long residence times.

a, Schematics of Halo-GAF^{L} (long) and Halo-GAF^{S} (short) isoforms with functional domains to scale. BTB/POZ, Broad-complex, Tramtrack and Bric-à-brac/Poxvirus and zinc finger; BR, basic region; ZF, zinc finger and Q-rich, glutamine-rich. Halo-GAF deletions were generated by CRISPR–Cas9. *ΔPOZ* contains a 90 bp deletion in the second exon, generating a 30-aa deletion (*Δ*90–119) of the POZ domain, which includes G93 and L112 that are essential for transcription activation[33]. The deleted Arg and Asn amino acids in the zinc finger (*ZF^{9}*, *ZF^{10}*) make contact with 'GAG' of the consensus GAF-binding site (Omichinski et al.[106]). *ΔQ* contains small deletions at the beginning of Q-rich domains of both long and short isoforms, resulting in frameshift and truncation of Q-rich domains. Table reports mutant phenotypes. FS, frameshift. **b**, Global chromatin bound and free fractions (%) for WT and mutant Halo-GAF from fast-tracking ($n = 3$ biological replicates for Halo-GAF WT and *ΔQ*, $n = 4$ biological replicates for the remaining conditions) and

slow-tracking ($n = 100$ resamplings). Results are mean ± s.d. with error propagation for F_{sb}. $\star P = 0.028$; $\star\star P = 0.0039$ for ΔQ and 0.0065 for ZF9; $\star\star\star P = 0.0002$, two-sided, unpaired *t*-test for fast tracking. **c**, Average residence times for WT and mutant Halo-GAF, corrected as in Fig. 1g. Error bars represent bootstrapped s.d. after resampling 100 times ($n = 100$). **d**, Halo-GAF distribution on fixed salivary gland polytene nuclei for WT and GAF mutants. ΔPOZ and ΔQ nuclei are from late third instar larvae with WT nuclei from the same stage as control. ZF^{9} and ZF^{10} polytene nuclei are from early instar larvae with WT nuclei from the same stage as control. Red, Halo-GAF; blue, DAPI. One representative confocal z-section is shown. Scale bar, 10 μm.

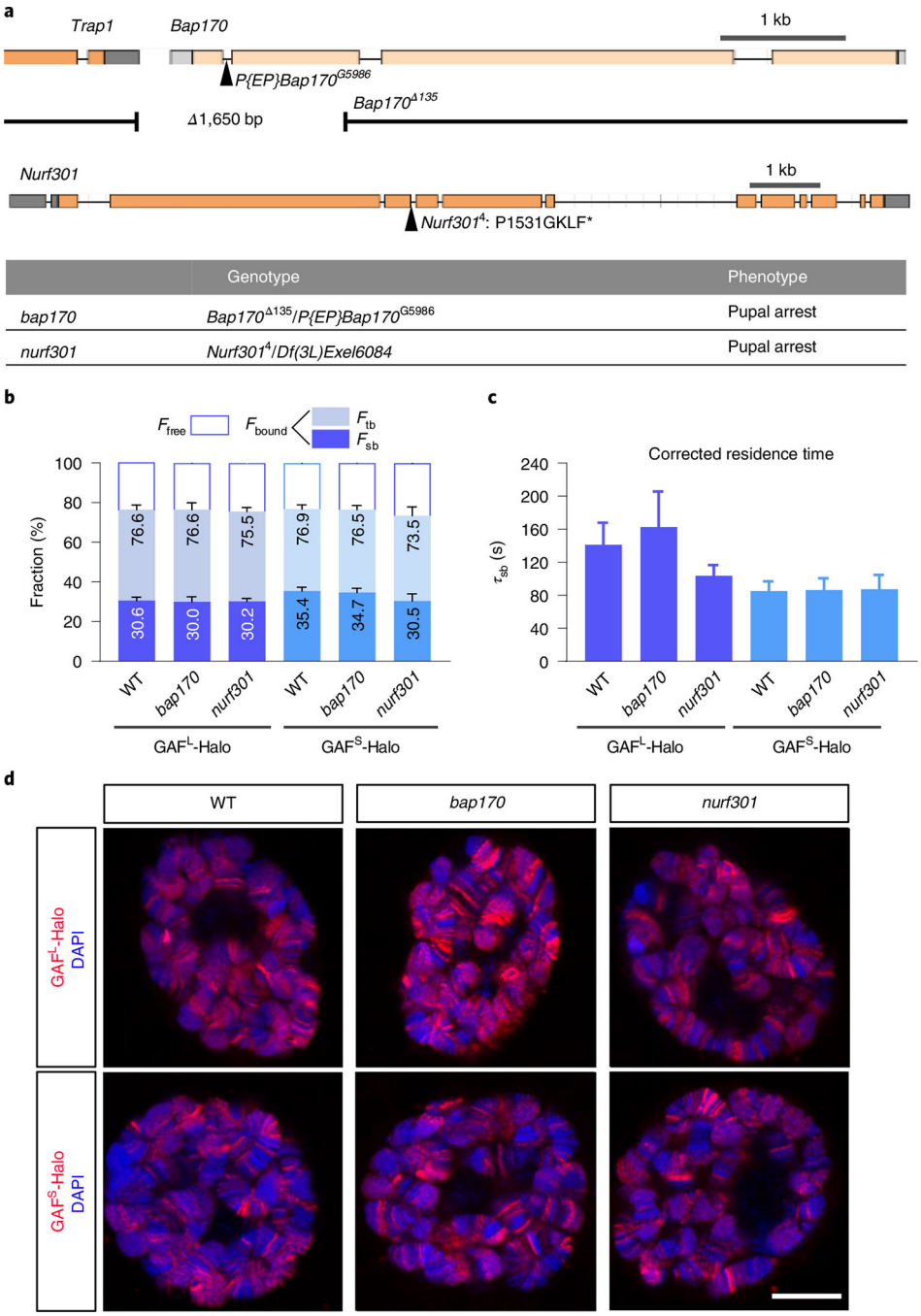


Fig. 3 |. Chromatin binding by GAF is largely independent of remodelers PBAP and NuRF.

a, *Bap170* and *Nurf301*/*E(bx)* mutants and phenotypes. *Bap170Δ*^{135} harbors a 1,650 bp deletion spanning the promoter, 5′ untranslated region and first two exons of *Bap170*, *P{EP}Bap170*^{G5986} has a *P{EP}* element insertion at 447 bp downstream from the transcription start site in the first intron of the gene. *Bap170*$^{\Delta 135}$ abolishes *Bap170* expression and strongly reduces polybromo protein level at larval stage, causing pupal lethality when homozygous[107], while flies homozygous for the *P{EP}Bap170*^{G5986} die primarily as pharate adults, with only 3% male homozygous adults[50]. Transheterozygous

Bap170Δ^{135}/*P{EP}Bap170*^{G5986} (denoted as *bap170*) also exhibits pupal arrest. *Nurf301*^{4} contains a splice-donor site mutation that blocks splicing of the fourth intron. The aberrant transcript introduces four additional amino acids (GKLF) and an in-frame stop codon after P1531, truncating C-terminal 609–1,230 aa including the essential PHD finger and Bromodomain (Badenhorst et al.[23]). *Nurf301*^{4} is pupal-lethal with <60% pupariation rate[49]. When transheterozygous with a deletion allele spanning 31 genes including *Nurf301*, *Nurf301*^{4}/*Df(3L)Exel6084* (denoted as *nurf301*) also shows pupal arrest. **b**, Global chromatin-bound fractions for GAF^{L}-Halo and GAF^{S}-Halo in WT, *bap170* and *nurf301* mutants from fast tracking ($n = 3$ biological replicates for GAF^{L}-Halo in nurf301 mutant, and $n = 4$ biological replicates for the remaining conditions) and slow tracking ($n = 100$ resamplings). Results are mean ± s.d. with error propagation for F_{sb}. All three strains express transgenic GAF^{L}-Halo or GAF^{S}-Halo under natural regulation, in the presence of endogenous, untagged GAF. See Methods for genotypes of WT, *bap170* and *nurf301*. **c**, Corrected average residence times for GAF^{L}-Halo and GAF^{S}-Halo in WT, *bap170* and *nurf301* mutants. Error bars represent bootstrapped s.d. after resampling 100 times ($n = 100$). **d**, GAF^{L}-Halo and GAF^{S}-Halo distributions on fixed salivary gland polytene nuclei for WT, *bap170* and *nurf301* mutants. Red, GAF-Halo; blue, DAPI. One representative confocal *z*-section is shown. Scale bar, 10 μm.

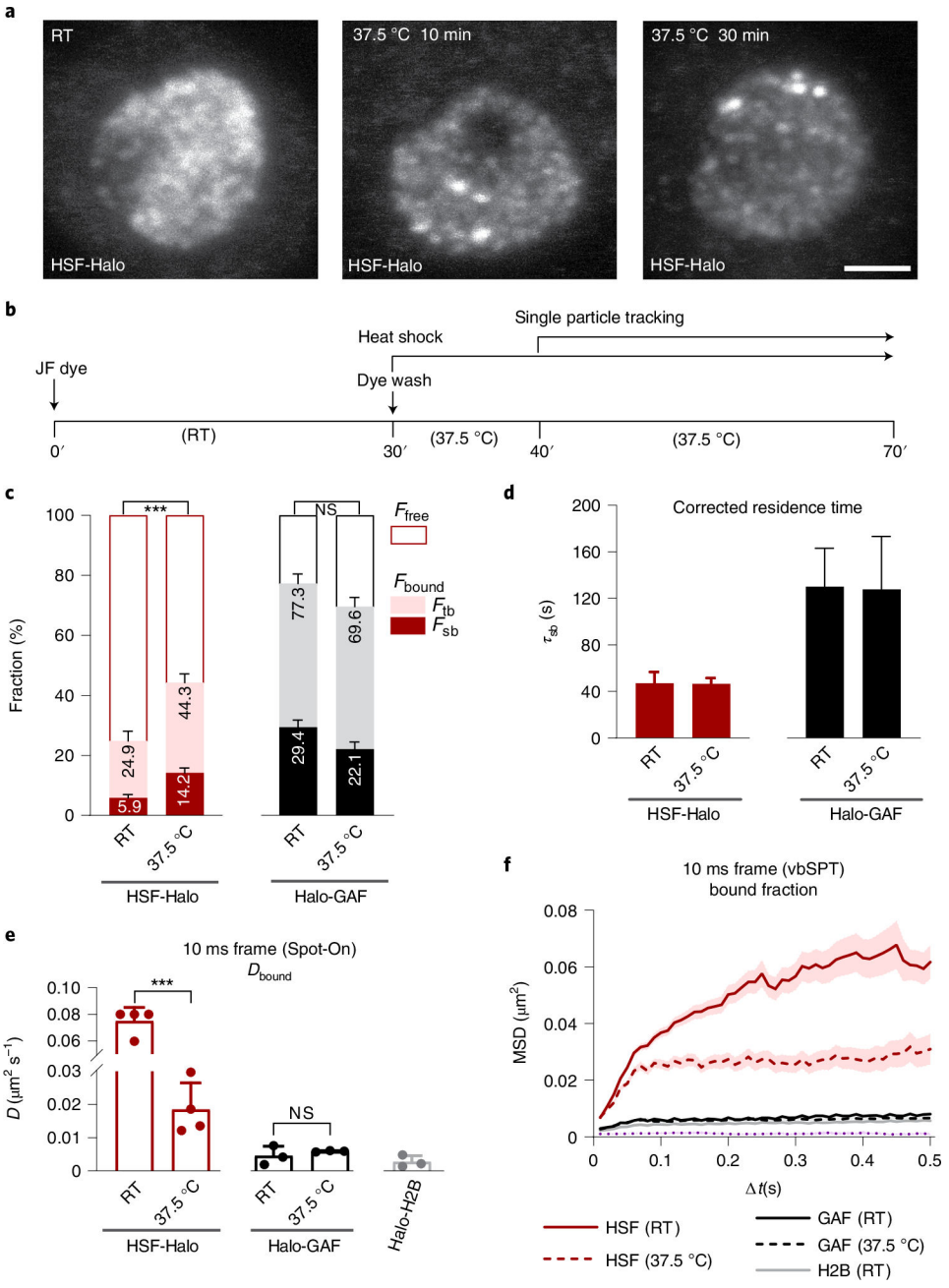
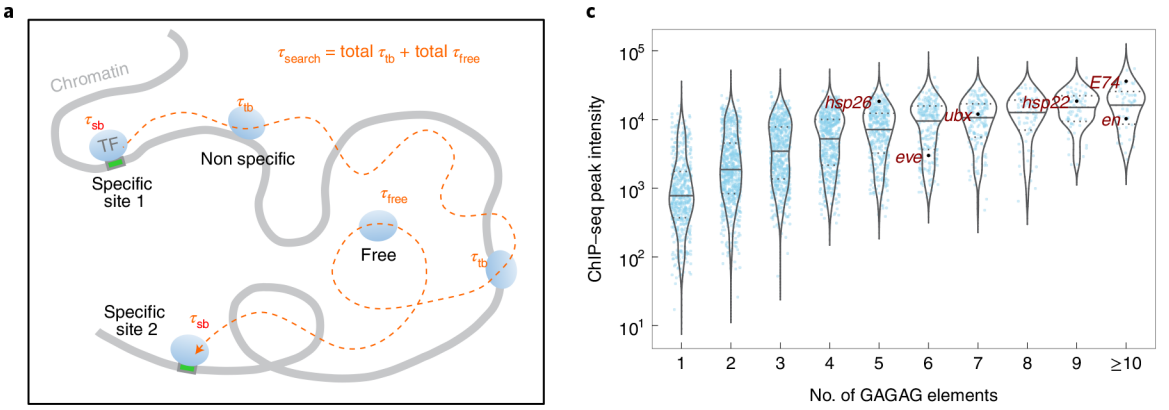


Fig. 4 |. Heat shock increases chromatin-binding fraction of HSF without affecting dwell time.

a, Maximum-intensity *z*-stack projection of HSF-Halo in fixed hemocytes. HSF-Halo forms several prominent foci on heat shock at 37.5 °C. Maximum projections of confocal *z*-stacks are shown. Scale bar, 2 μm. **b**, Flow chart of the heat shock and live-hemocyte imaging procedure. SPT starts 10 min after heat shock and continues over multiple cells (1–2 min per cell) for a total of 30 min with each sample. **c**, Global chromatin-bound fractions for HSF-Halo and Halo-GAF at RT and 37.5 °C, from fast tracking ($n = 4$ biological replicates for HSF-Halo, $n = 3$ biological replicates for Halo-GAF) and slow tracking ($n = 100$ resamplings). Results are mean ± s.d. with error propagation for F_{sb}. ⋆ ⋆ ⋆ $P = 0.00051$,

two-sided, unpaired *t*-test for fast tracking. **d**, Corrected average residence times for HSF-Halo and Halo-GAF at RT and 37.5 °C. Error bars represent bootstrapped s.d. after resampling 100 times ($n = 100$). **e**, Diffusion coefficients for bound fractions of HSF-Halo and Halo-GAF at RT and 37.5 °C, and Halo-H2B at RT derived using Spot-On. Results are mean ± s.d. ($n = 4$ biological replicates for HSF-Halo, $n = 3$ biological replicates for Halo-GAF and Halo-H2B). ⋆ ⋆ ⋆ $P = 0.00016$, two-sided, unpaired *t*-test. **f**, Average MSD (mean ± s.e.) versus lag time of bound trajectories classified by vbSPT for HSF-Halo, Halo-GAF and Halo-H2B at RT and 37.5 °C and Halo-H2B at RT. See Extended Data Fig. 9a for a zoomed-in section for GAF and H2B.



b

Factor	F_{bound}	f_{sb}	τ_{sb} (s)	τ_{tb} (s)	$N_{monomers}$	N_{sites}	N_{trials}	τ_{search} (s)	SI (s)	O
GAF WT	0.77	0.38	130.0	3.70	56,683 ± 6,025	3,622 × 4	23 ± 8	149.9 ± 50.1	71.5 ± 7.6	182 ± 50%
GAF *ΔPOZ*	0.44	0.33	42.3	2.46	44,907 ± 14,154	3,622 × 4	9 ± 2	104.9 ± 17.3	47.5 ± 15.0	89 ± 30%
GAF *ΔQ*	0.58	0.41	49.2	2.93	66,206 ± 3,064	3,622 × 4	8 ± 2	77.6 ± 13.0	27.7 ± 1.3	177 ± 29%
GAF 37.5 °C	0.70	0.32	127.7	2.46	56,683 ± 6,025	3,622 × 4	40 ± 16	197.6 ± 66.5	83.6 ± 8.8	154 ± 57%
HSF 37.5 °C	0.44	0.32	46.7	2.87	9,543 ± 613	442 × 4	9 ± 2	118.5 ± 20.7	30.6 ± 2.0	153 ± 21%

Fig. 5 |. High site occupancy and remodeler autonomy quantifies pioneering criteria.

a, Schematic of a TF trajectory between two specific chromatin targets showing the search time τ_{search}, stable τ_{sb} and transient τ_{tb} dwell times, and τ_{free}. The TF molecule dissociates from a specific target, and samples nonspecific sites for a number of trials before encountering the next specific target. The τ_{search} equation is indicated (Methods). **b**, Key SPT and $N_{monomers}$ parameters measured in this study and N_{sites} from the literature[17,54] are used to calculate occupancy levels for GAF and HSF. **c**, Violin plots of GAF ChIP-seq peak intensities (analysis of table s3 from Fuda et al.)[17] in hemocyte-like S2 cells plotted by the number of nonoverlapping GAGAG elements between the start and end coordinates of each peak identified with HOMER[108].

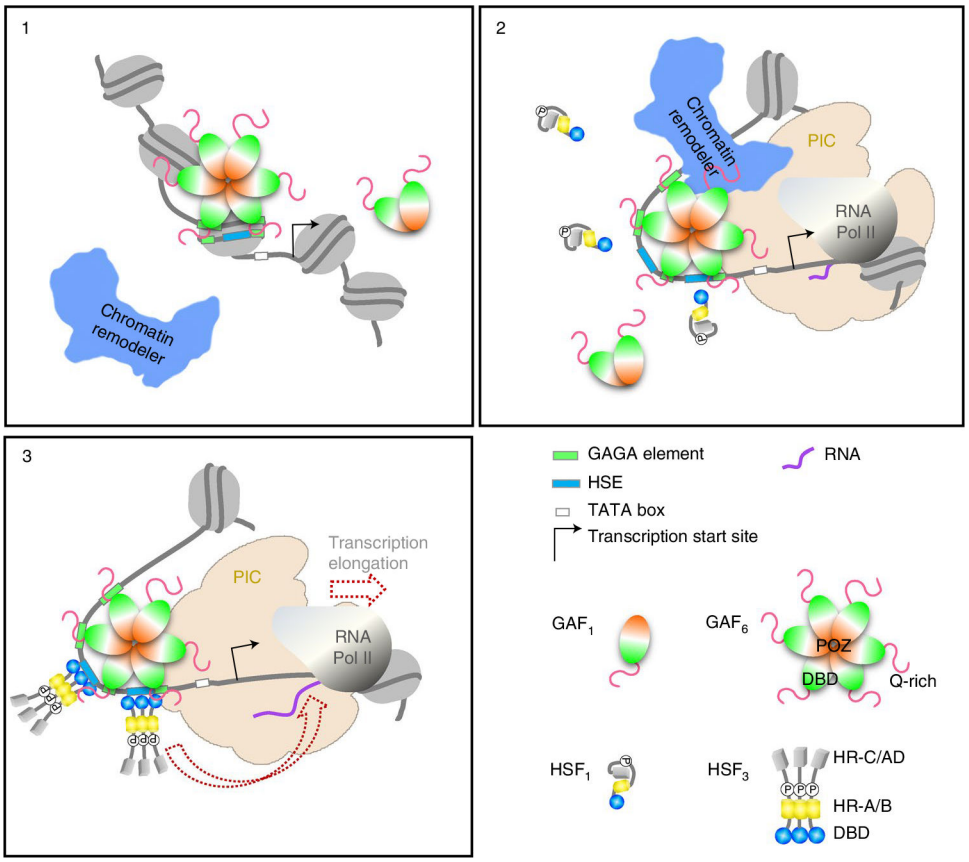


Fig. 6 |. Pioneering of chromatin accessibility is a process involving multiple inputs.

Model, GAF binds autonomously to nucleosomal sites at the first stage of pioneering (box 1). High GAF occupancy at a *Hsp* promoter with clustered GAGA elements maintains chromatin accessibility for neighboring factor HSF and assembly of the preinitiation complex and paused RNA Pol II (box 2). The substantially constrained diffusivity (D_{bound}) of stably bound GAF may reflect multisite interactions of a GAF multimer with clustered GAGA elements[45], locking down GAF with prolonged residence time. HSF trimers bind to accessible chromatin DNA with high affinity on heat shock to trigger RNA Pol II elongation (box 3). See Discussion for details.