


RESEARCH

Open Access



A quantitative metric of pioneer activity reveals that HNF4A has stronger in vivo pioneer activity than FOXA1

Jeffrey L. Hansen^{1,2,3} and Barak A. Cohen^{1,2*} 

*Correspondence:
cohen@wustl.edu

¹ The Edison Family Center for Genome Sciences and Systems Biology, School of Medicine, Washington University in St. Louis, Saint Louis, MO, USA

² Department of Genetics, School of Medicine, Washington University in St. Louis, Saint Louis, MO, USA

³ Medical Scientist Training Program, Washington University in St. Louis, St. Louis, MO, USA

Abstract

Background: We and others have suggested that pioneer activity — a transcription factor's (TF's) ability to bind and open inaccessible loci — is not a qualitative trait limited to a select class of pioneer TFs. We hypothesize that most TFs display pioneering activity that depends on the TF concentration and the motif content at their target loci.

Results: Here, we present a quantitative in vivo measure of pioneer activity that captures the relative difference in a TF's ability to bind accessible versus inaccessible DNA. The metric is based on experiments that use CUT&Tag to measure the binding of doxycycline-inducible TFs. For each location across the genome, we determine the concentration of doxycycline required for a TF to reach half-maximal occupancy; lower concentrations reflect higher affinity. We propose that the relative difference in a TF's affinity between ATAC-seq labeled accessible and inaccessible binding sites is a measure of its pioneer activity. We estimate binding affinities at tens of thousands of genomic loci for the endodermal TFs FOXA1 and HNF4A and show that HNF4A has stronger pioneer activity than FOXA1. We show that both FOXA1 and HNF4A display higher binding affinity at inaccessible sites with more copies of their respective motifs. The quantitative analysis of binding suggests different modes of binding for FOXA1, including an anti-cooperative mode of binding at certain accessible loci.

Conclusions: Our results suggest that relative binding affinities are reasonable measures of pioneer activity and support the model wherein most TFs have some degree of context-dependent pioneer activity.

Keywords: Pioneer factor, Pioneer activity, Transcription factor binding, FOXA1, HNF4A, Genomics

Background

Activating silent genes requires transcription factors (TFs) to bind and open DNA when their motifs are occluded by nucleosomes. Activating silent genes is postulated to involve two qualitatively different classes of TFs, pioneer factors (PFs), and non-pioneer factors (nonPFs) [1, 2]. According to this hypothesis, PFs bind to nucleosome-occluded



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

DNA and make it accessible to nonPFs, which then recruit the cofactors required to activate transcription. However, we recently showed that both a canonical PF, FOXA1, and a nonPF, HNF4A, can independently bind, open, and then activate nearby genes [3], and many TFs possess unique ways of binding and opening nucleosomal DNA [4–8]. From these data, we propose that most TFs have quantifiable pioneer activity that depends on their nuclear concentrations and the motif content at their target loci. Here, we present a metric that can quantify *in vivo* pioneer activity.

A metric for pioneer activity should reflect the relative difference in binding affinity between accessible and inaccessible DNA, where strong pioneer activity is a result of a smaller difference in relative affinity. Normally, affinity is quantified by measurements of *in vitro* equilibrium dissociation constants, or K_d s. We propose a method to estimate *in vivo* affinities in parallel by using our previously established doxycycline-inducible (dox-inducible) TF expression lines [3]. We expressed endodermal TFs FOXA1 and HNF4A across a 1000-fold range, measured binding and accessibility with CUT&Tag [9] and ATAC-seq [10], and then extracted the dox concentration at which each site was half-maximally bound, or the site's "dox₅₀" (Fig. 1A). Each site's dox₅₀ reflected the TF's affinity for that site. The ratio of the average dox₅₀ at accessible versus inaccessible binding sites is the TF's pioneer activity index. We show that HNF4A has stronger overall pioneer activity and that both TFs can compensate for weaker affinity at inaccessible binding sites when there are more copies of their motifs. The distribution of pioneer activities across the genome supports the hypothesis that most TFs have pioneer activity given sufficient TF levels and motif content.

Results

A quantitative metric for pioneer activity

An appropriate measure of pioneer activity should capture the relative difference of TF binding between accessible and inaccessible sites in the genome. In principle, we could compare the dissociation constant (K_d) of a TF at accessible and inaccessible sites as a measure of pioneer activity, since the K_d is the concentration of TF required to reach half-maximal binding. In practice, computing an absolute K_d inside cells is impractical. While it is possible to measure apparent K_d s using fluorescently-labeled TFs, the throughput and resolution required to make accurate genome-wide calculations is impractical. We propose a related measure that uses doxycycline-inducible (dox-inducible) TFs to compute the dox₅₀, the dox concentration required to reach half-maximal binding inside cells. By inducing TF levels over a wide range of dox concentrations and measuring the resulting binding, we determine a dox₅₀ for every location in the genome in parallel. This method assumes linearity of expression as a function of doxycycline concentration, and we have shown previously that the expression of targets of our inducible TFs increases linearly with dox concentrations [3].

The ratio of the average dox₅₀ at accessible versus inaccessible sites is a TF's "pioneer activity index." A TF with maximal pioneer activity will bind with equal affinity to accessible and inaccessible DNA and have a pioneer activity index of 1. Decreased affinity (higher dox₅₀) at inaccessible sites will lower the index towards 0 (Fig. 1B). Because the measurements at accessible and inaccessible sites are made at the same time in the same nucleus, the dox concentrations (or TF concentrations) cancel out, allowing us to

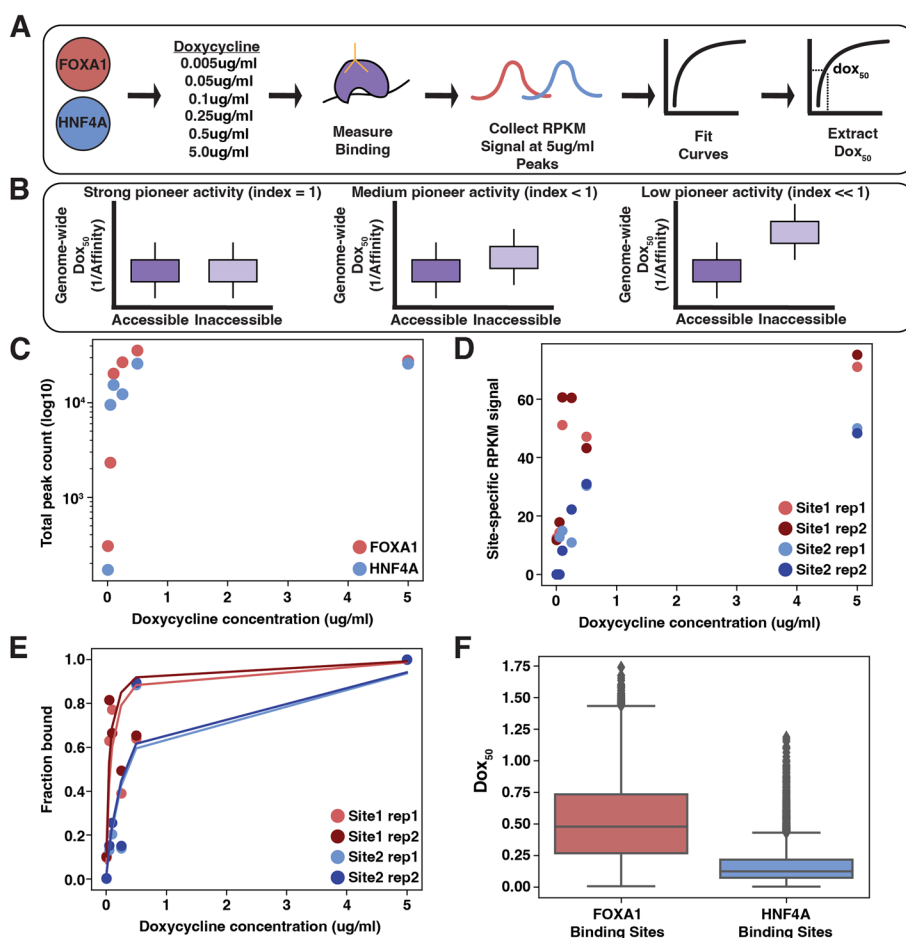


Fig. 1 Experimental design to compute pioneer activity. **A** We induced FOXA1 or HNF4A across a 1000-fold dox range, measured binding, collected binding signals at a reproducible set of binding sites, and then extracted a dox_{50} for each site. **B** Maximal pioneer activity (index = 1) occurs when a TF’s average affinity in accessible regions is similar to that in inaccessible regions. Pioneer activity decreases (index < 1) as the affinity in inaccessible regions decreases, causing an increase in the dox_{50} measurement. **C** The number of total peaks for each TF across each dox induction level. **D** Replicate RPKM binding signal at two example genomic sites. **E** Replicate fitted lines at two example genomic sites. **F** Full distribution of dox_{50} values for each TF

compare pioneer activity indices of different TFs to each other [11]. This strategy allows us to circumvent the challenge of measuring effective nuclear TF concentration while maintaining the physiological relevance of our in vivo pioneer activity measurements.

Measurement of dox_{50} for FOXA1 and HNF4A

FOXA1 and HNF4A are liver TFs that are commonly used to reprogram embryonic fibroblasts to endoderm progenitor cells [12, 13]. FOXA1 is a canonical PF and HNF4A a nonPF, and the two are suggested to work in a collaborative and sequential fashion to activate their target genes [1, 14]. We previously tested FOXA1 and HNF4A’s behavior in an ectopic setting by expressing them within K562 blood cells, a lineage in which neither TF is expressed and that should present the TFs with unique complements of chromatin and cofactors. We created clonal K562 lines that expressed either inducible FOXA1

or HNF4A and showed that both TFs could independently bind and open inaccessible chromatin and activate nearby genes [3].

Based on the ability of FOXA1 and HNF4A to independently bind, open, and activate in an ectopic cell line, we expected both TFs would have similar pioneer activity. To test this prediction we attempted to measure each TF's pioneer activity index using the same dox-inducible FOXA1 or HNF4A K562 lines (Fig. 1A). We first treated each TF line with a 1000-fold range of dox (0.005, 0.05, 0.25, 0.1, 0.5, and 5.0 $\mu\text{g/ml}$) and measured resultant binding. Read normalized binding signal (RPKM) was highly correlated between replicates (Fig. S1). We found that each TF transitioned from binding hundreds of sites to tens of thousands of sites in a similar fashion as we increased the dox concentration and that each TF increased similarly, suggesting that neither TF was operating exclusively at either end of its dynamic range (Fig. 1C). We then collected the overlapping set of binding sites between each TF's replicates in the 5.0 $\mu\text{g/ml}$ sample and at each site plotted the read normalized signal (RPKM) from the other induction levels (Fig. 1D). Generally we observe that at each site, a TF's binding signal saturates in an expected fashion as its expression increases; we refer to this type of binding as the "saturation modality" (Fig. S2). We then fit Eq. 1 to these distributions.

$$\text{Fraction bound} = \frac{1}{1 + \frac{\text{dox}_{50}}{[\text{dox}]}} \quad (1)$$

In order to fit Eq. 1, we normalized each site's RPKM signal to the signal in the 5.0- $\mu\text{g/ml}$ sample to convert our measurements into fractional binding (Fig. 1E). We found that at some sites the binding signal peaked prior to the 5.0- $\mu\text{g/ml}$ sample (fraction bound > 1 in any of the first five induction levels). We removed these sites to prevent poor fitting. This left us with 11,557 FOXA1 binding curves and 5940 HNF4A binding curves with highly similar fitted lines across replicates (Figs. 1E and S3). We extracted dox_{50} values from these lines and found similar results between replicates (Fig. S4) and so we averaged each site's replicate dox_{50} value for the remaining analyses. The resulting distributions of each TF's genome-wide dox_{50} values show that FOXA1 has a much larger variance in dox_{50} values than HNF4A (Fig. 1F), suggesting that FOXA1 binding generally depends more on the genomic environment than HNF4A.

Measurement of pioneer activity indices for FOXA1 and HNF4A

The dox_{50} distributions in Fig. 1 suggest that HNF4A may bind more consistently across the genome but does not explicitly report on the difference in binding affinity between accessible and inaccessible sites. We therefore classified each site as either accessible or inaccessible based on ATAC-seq peaks collected in these cell lines before dox induction [3]. Seventeen percent of FOXA1's binding sites occurred in accessible regions and 83% occurred in inaccessible regions (prior to the above filtering step: 36% accessible and 64% inaccessible). Thirty-six percent of HNF4A's binding sites occurred in accessible regions and 64% occurred in inaccessible sites (prior to filtering, 49% accessible and 51% inaccessible). A larger proportion of FOXA1-accessible sites were filtered out due to the unexpected anti-cooperative binding modality that we identified and that we discuss further below.

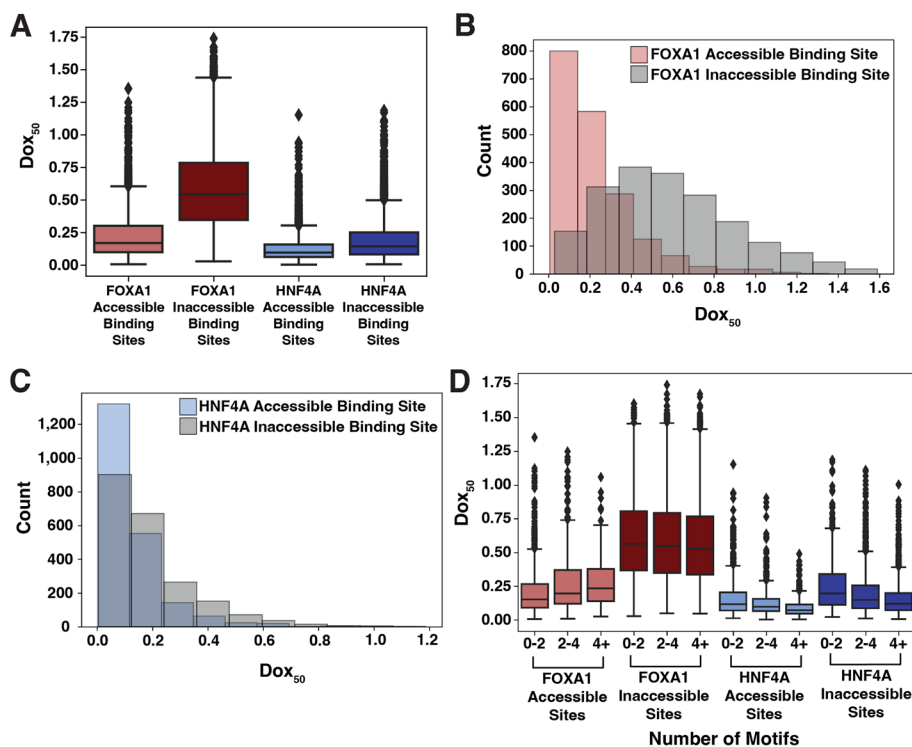


Fig. 2 HNF4A has stronger pioneer activity than FOXA1. **A** Distributions of dox_{50} estimates extracted from binding curves at FOXA1-accessible binding sites ($n = 1930$), FOXA1 inaccessible binding sites ($n = 9627$), HNF4A accessible binding sites ($n = 2135$), and HNF4A inaccessible binding sites ($n = 3805$). **B, C** Distributions from FOXA1 (**B**) and HNF4A (**C**) shown in histogram form. **D** Same plot as **A** but each genomic binding site is binned by whether the site has < 2 , ≥ 2 but < 4 , or ≥ 4 motifs as called by FIMO ($p = 1e-3$)

Comparing the dox_{50} distributions between accessible and inaccessible sites revealed that the binding of HNF4A is less affected by inaccessible DNA than FOXA1 (Figs. 2A–C and S4). We then computed pioneer activity indices for each TF by dividing the average dox_{50} for accessible sites by the average dox_{50} for inaccessible sites. A TF with maximal pioneer activity would have equal affinity across accessible and inaccessible sites and therefore have an index of 1. Any reduction in affinity at inaccessible sites increases the average inaccessible dox_{50} and thus decreases the index. HNF4A’s index was 0.680 and FOXA1’s was 0.391, demonstrating that HNF4A has stronger global pioneer activity than FOXA1. HNF4A has stronger pioneer activity even when considering all sites without filtering (Fig. S5).

We next considered whether the motif content at each binding site affected pioneer activity. We subset each TF’s binding sites into those that had less than 2, between 2 and 4, or more than 4 motifs specific to the respective TF and re-plotted the dox_{50} distributions. For both FOXA1 and HNF4A, higher motif content at inaccessible binding sites correlated with lower dox_{50} distributions (Fig. 2D). Unexpectedly, FOXA1-accessible sites with more motifs had lower affinity. We speculate this may be driven by anti-cooperative binding behavior and explore this phenomenon further below. We conclude that HNF4A has stronger pioneer activity in K562 cells and that

weak affinity binding at inaccessible regions can be compensated for by both TFs by the presence of additional motifs.

Chromatin modifications explain some of the variance in dox_{50} values

We built a linear model (Eq. 2) to try to explain the variance in dox_{50} s for FOXA1 and HNF4A where $C(\text{Accessibility})$ is each binding site’s accessibility prior to TF induction. Accessibility explained 17% of the variance in FOXA1’s dox_{50} s but only 4% of HNF4A’s. While these data further underscore the greater role that accessibility plays on FOXA1 binding than HNF4A, they also reveal that most of the variance in dox_{50} values between genomic loci must be explained by some other variable.

$$Dox_{50} \sim C(\text{Accessibility}) \tag{2}$$

We hypothesized that some of the remaining variance may be explained by the different chromatin modifications present at different target loci and predicted that binding sites with active marks would have lower dox_{50} distributions (easier binding) and binding sites with silent marks would have higher dox_{50} distributions (harder binding). We further subset each TF’s accessible or inaccessible binding sites into those that overlap common K562 marks [15]. H3K4me1 marks enhancers [16], H3K27Ac marks activity [17], and H3K9me3 and H3K27me3 are two modifications shown previously to suppress pioneer activity [18]. The accessible sites overlapped

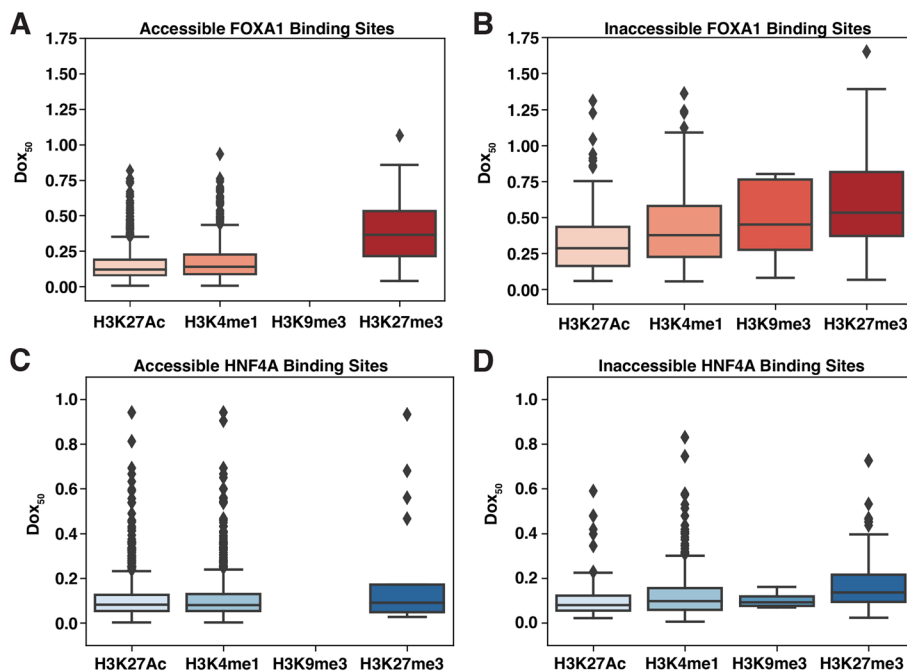


Fig. 3 Dox_{50} distributions across different chromatin modifications. **A** Dox_{50} values for FOXA1-accessible binding sites that overlapped H3K27AC ($n = 1288$), H3K4me1 ($n = 755$), H3K9me3 ($n = 0$), and H3K27me3 ($n = 21$). **B** Dox_{50} values for FOXA1 inaccessible binding sites that overlapped H3K27AC ($n = 203$), H3K4me1 ($n = 352$), H3K9me3 ($n = 12$), and H3K27me3 ($n = 277$). **C** Dox_{50} values for HNF4A accessible binding sites that overlapped H3K27AC ($n = 1147$), H3K4me1 ($n = 1111$), H3K9me3 ($n = 0$), and H3K27me3 ($n = 17$). **D** Dox_{50} values for HNF4A inaccessible binding sites that overlapped H3K27AC ($n = 140$), H3K4me1 ($n = 416$), H3K9me3 ($n = 4$), and H3K27me3 ($n = 135$)

much more often with active marks than silencing marks, and we found that no FOXA1 or HNF4A accessible sites were marked with H3K9me3 (Fig. 3).

As predicted, FOXA1 or HNF4A binding sites that overlapped H3K27Ac or H3K4me1 chromatin modifications had lower dox_{50} distributions than those that overlapped H3K9me3 or H3K27me3 (Fig. 3). These effects were present even after we subset binding sites by accessibility, suggesting that the chromatin modifications can affect binding in ways that are separable from the effects of accessibility. However, when we individually added each chromatin modification (plus an interaction term) to the model in Eq. 1, we found that accounting for these marks did not have large effects on the ability of the model to predict dox_{50} values for either TF. H3K27ac levels explained 2% of FOXA1's dox_{50} variance, H3K4me1 explained 1%, and H3K27me3 explained <1%. For HNF4A, H3K27ac explained 2%, H3K4me1 explained 2%, and H3K27me3 explained <1%. All interaction terms were negligible. We also considered whether a site's DNA methylation level [19] correlated with its dox_{50} value but only identified a minor inhibition of binding exclusively at accessible FOXA1 binding sites (Fig. S6). Together these data suggest that something besides the epigenetic landscape of loci is having a large effect on the pioneering activity of TFs.

FOXA1 behaves anti-cooperatively at a subset of accessible binding sites

While examining individual binding sites and their fitted curves, we observed a repeating pattern at a subset of genomic locations where the binding signal increased to a peak at the third (0.1 μ g/ml) or fourth (0.25 μ g/ml) induction level and then decreased, suggesting anti-cooperative behavior (Figs. 4A and S7). To quantify the prevalence of this "anti-cooperative modality" we sampled 10,000 peaks from the original set of unfiltered FOXA1 or HNF4A accessible or inaccessible binding sites and then counted how many displayed saturation behavior (peak at 5 μ g/ml, Fig. S2) and how many displayed anti-cooperative behavior (peak at 0.1 μ g/ml or 0.25 μ g/ml, Figs. 4A and S7). We drew strict criteria for the saturation behavior (binding signal must increase sequentially until maximum at 5 μ g/ml) and for the anti-cooperative behavior (binding signal must increase sequentially until peak and then decrease sequentially) such that not every one of the 10,000 sampled binding sites were included within one of these two modalities. We found that the anti-cooperative behavior occurs most often at accessible FOXA1 binding sites (Fig. 4B–C). Anti-cooperative behavior does not appear to depend on the number of motifs at each peak (Fig. 4D) or the length of each peak (Fig. 4E).

We considered whether another TF might be contributing to anti-cooperative behavior by searching for enriched motifs in either saturation-type accessible FOXA1 binding sites or anti-cooperative-type sites. While FOXA1 motifs were enriched in both types of loci (Fig. 4F), the AP1 motif was only enriched at anti-cooperative sites (Fig. 4G). AP1 is an important K562 TF that exhibits some pioneer activity [20]. This finding suggests that a protein-protein interaction between FOXA1 and AP1 underlies anti-cooperative behavior at accessible loci.

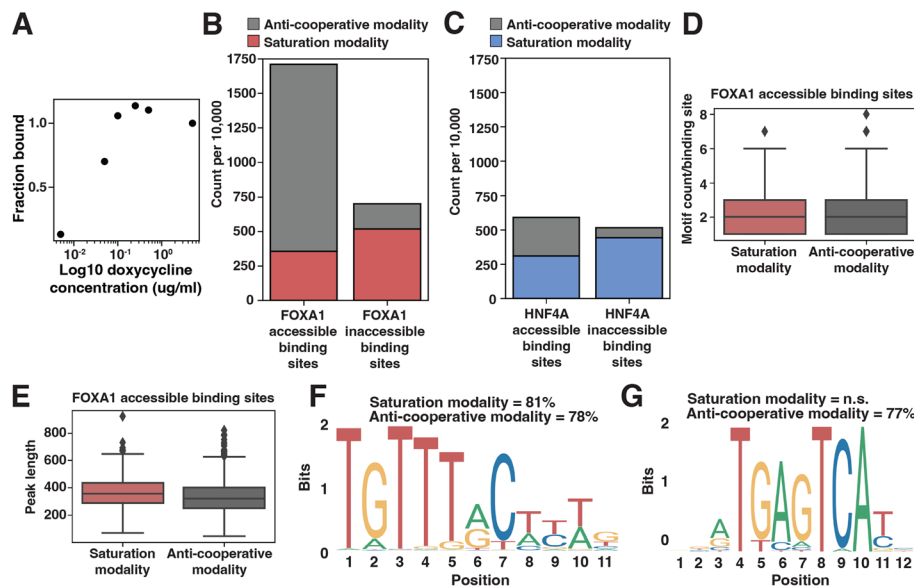


Fig. 4 Characterization of anti-cooperative binding behavior. **A** Example binding curve at a single genomic site that exhibits anti-cooperative behavior. **B, C** A sample of 10,000 FOXA1 (**B**) or HNF4A (**C**) accessible (left bar) or inaccessible (right bar) binding sites colored by if they display saturation binding behavior (red or blue) or anti-cooperative binding behavior (gray). **D** FOXA1 motif count between the accessible binding sites from **B** that display either saturation or anti-cooperative binding behavior. Motifs were called from FIMO with a p -value threshold of $1e-3$. **E** Binding peak length between the accessible binding sites from **B** that display either saturation or anti-cooperative binding behavior. **F** The most enriched motif discovered in FOXA1-accessible saturation and anti-cooperative peaks was FOXA1 (JASPAR MA0148.1). It is significantly enriched for both the saturation behavior ($p = 2.19e-001$) and anti-cooperative behavior ($p = 1.23e-01$). **G** The second most enriched motif discovered in FOXA1-accessible anti-cooperative peaks was AP1 (JASPAR MA1141.1). It was not discovered in the saturation behavior peaks. It is significantly enriched for only the anti-cooperative behavior ($p = 1e-008$)

Discussion

Given a definition of pioneer activity that is a TF's ability to bind at inaccessible genomic locations, we suggest that the difference in the average binding affinity between accessible and inaccessible DNA is a quantitative measure of this activity. We measured pioneer activity indices for FOXA1 and HNF4A in K562 cells and showed that HNF4A has stronger pioneer activity in this cell type than FOXA1. However, both TFs showed a range of dox_{50} values across the genome, which demonstrates that a TF's pioneer activity may depend on accessibility, native chromatin marks, and other factors. Some of these differences are explained by the motif content at different locations, suggesting that low affinity interactions at inaccessible binding sites can be overcome by strong motif content. While our work shows that the pioneering activity of a TF can vary across the genome, what accounts for this variation across sites remains mostly unexplained. DNA accessibility had the largest effect on pioneer activity but only explained 17% of the variance in dox_{50} values. We speculate that much of the remaining variance in dox_{50} values might be explained by interactions with other specific TFs, chromatin remodelers, or with the general transcription machinery that can differ across the genome.

Our work supports the hypothesis that pioneer activity is not a qualitative trait limited to a few TFs, but rather a quantitative property of TFs that manifests differently depending on the TF and the genomic environment [1, 3, 4, 7, 8, 21, 22]. Pioneer activity as a quantitative trait fits with data that show that expression level dictates whether a TF adopts a pioneering or collaborating role [23]. We and others have shown that pioneer activity can be enhanced by increasing TF expression [24] or dampened by decreasing TF expression [3].

In vitro, FOXA1 has higher affinity (a lower K_d) for naked DNA than HNF4A [21, 25, 26], and yet HNF4A has stronger pioneer activity than FOXA1 in K562 cells. These results demonstrate that even though we may parsimoniously consider pioneer activity as a product of binding energies, it may not be sufficient to solely consider a single TF's DNA-binding domain and its cognate motif. Inside cells, pioneer activity likely depends on the interactions a TF makes with other TFs and with cofactors. Because these interactions will differ in different cell types, a TF's pioneering activity is also likely to depend on the cell type in which it is expressed and the co-bound TFs present at certain locations.

At some locations in the genome, an interaction between FOXA1 and AP1 appears to have a dramatic effect on FOXA1 activity. In the presence of AP1 sites, FOXA1 displays anti-cooperative binding dynamics where occupancy decreases at the highest levels of FOXA1 expression. We speculate that at these sites monomers of FOXA1 interact with AP1 to potentiate binding, whereas dimers of FOXA1 cannot cobind with AP1. In this model, high concentrations of FOXA1 favor its dimeric form which accounts for the loss of binding at these sites when FOXA1 is expressed at high levels.

Conclusion

Regardless of the mechanism underlying anti-cooperative behavior, our results show that pioneer activity can be modified by the interactions a TF makes inside cells. Thus, pioneer activity is contingent on many properties of a TF including its levels, its intrinsic affinity for its motif, the motif content at its targets, and the different interactions it makes with other proteins when bound at different locations. Given these contingencies, we suggest that most TFs will display some degree of pioneer activity and that our pioneer activity index will be a useful metric to quantify it.

Methods

Cell lines

We grew K562 cells (ATCC CCL-243, Manassas, VA) in Iscove's modified Dulbecco serum supplemented with 10% fetal bovine serum, 1% penicillin-streptomycin and 1% non-essential amino acids. For each of our functional assays, we split each line into replicate flasks, treated with doxycycline (dox) (Sigma #D9891-1G), and then waited 24 hours to extract nuclei. We used doses of 0.005 μ g/ml, 0.05 μ g/ml, 0.1 μ g/ml, 0.25 μ g/ml, 0.5 μ g/ml, and 5 μ g/ml for our dox_{50} experiments.

Cloning, production, and infection of viral vectors

We used FOXA1 and HNF4A K562 clonal lines and lentiviral vectors carrying inducible FOXA1 and HNF4A ORFs as described previously [3].

Sequencing library preparations and analysis

We prepared sequencing libraries and analyzed the two replicates of CUT&Tag as described previously [3]. In our previous work, we already used ATAC-seq to measure the uninduced (-dox) accessibility in the FOXA1 and HNF4A K562 lines [3]. Because we used the same clones to perform these experiments, we re-used these data as uninduced accessibility. We also had already sequenced CUT&Tag libraries for the 0.5- $\mu\text{g}/\text{ml}$ and 0.05- $\mu\text{g}/\text{ml}$ doxycycline induction levels and re-used these data as well.

Binding curve analysis

We first established a set of all possible binding sites for each TF by creating a list of binding sites in the sample with the highest dox induction concentration (5 $\mu\text{g}/\text{ml}$). We subset this list into those accessible binding sites (called accessible peak in the -dox uninduced condition) and inaccessible binding sites (absence of called accessible peak). Then we used the multiBigwigSummary from the deepTools suite [27] to count the normalized read intensity at each peak from each induction level. We normalized each induction level to the read intensity at the highest induction level in order to convert read intensity into fraction bound.

With these data we fit a binding curve using SciPy curvefit [28] to the equation (Eq. 1) where dox_{50} is unknown and represents a binding affinity parameter similar to K_d and where [dox] is the concentration of dox used to induce TF expression. When we plotted examples of randomly selected genomic sites and examined the binding curves, we noticed that at some sites, binding peaked (fraction bound ≥ 1) prior to the highest concentration. In these cases, the fit line estimated a negative dox_{50} . For this reason, we filtered out any site that peaked prior to the sample with the highest dox concentration. We also estimated dox_{50} distributions without this filtering step and found similar distributions (Fig. S2). We have listed all of our filtered FOXA1 and HNF4A accessible and inaccessible binding sites, their coordinates, and their dox_{50} values in a Additional file 2.

In order to quantify the early peak, or “anti-cooperative” behavior that we observed, we classified a binding site as exhibiting a “saturation binding” modality if only the highest dox concentration had a fraction bound of 1, and then each subsequent lower concentration had a lower fraction bound. We classified a binding site as exhibiting an “anti-cooperative” modality if the site peaked at either the third (0.1 $\mu\text{g}/\text{ml}$) or fourth (0.25 $\mu\text{g}/\text{ml}$) dox concentrations and then declined in each direction.

We calculated reproducibility in three ways. We first showed that the binding signal was reproducible by plotting the RPKM signal from each replicate for each of the concentrations at all of the binding sites collected as described above. We then showed that the lines fit similarly between replicates by both replicates’ binding signal and fit binding curves at many different randomly chosen genomic sites and showing that the lines look similar. And finally, we showed that the distributions of dox_{50} s from each replicate were

highly overlapping. After showing these, we averaged the dox_{50} from each replicate at each site and used the average value moving forward.

Motif analysis

To discover or count motifs in binding sites, we extracted the sequence from each CUT&Tag binding peak and then used XSTREME [29] for de novo motif discovery and FIMO [30] for specific motif occurrence counting. We used $1e-3$ as a p -value threshold and JASPAR [31] PWMs for FOXA1 (MA0148.1), HNF4A (MA0114.2), and AP-1 (MA1141.1). We used these motif counts to subset the FOXA1/HNF4A accessible/inaccessible peaks into those with less than 2 motifs, more than 2 but less than 4, or 4 or more, and then re-ran the analysis (Fig. 2D).

Chromatin modifications and methylation analysis and modeling

We used previously published datasets of histone ChIP-seq and whole-genome bisulfite sequencing [15] to identify patterns of H3K27Ac, H3K4me1, H3K9me3, and H3K27me3 marks, as well as DNA methylation [19]. We used BEDTools [32] to overlap FOXA1 or HNF4A's binding sites with chromatin marks. We then used python's statsmodels to run ANOVA analyses on ordinary least squares linear regressions. Each reported variance is the parameter's sum of squares contribution divided by the total sum of squares.

The methylation dataset that we used reported the percent of reads at each CpG that were methylated. We converted the hg19 coordinates of our filtered FOXA1 or HNF4A accessible and inaccessible binding sites to GrCh38, overlapped them with methylation data, retained sites that had at least 10 reads of coverage, and then averaged the percent methylation of each CpG in each binding site. This analysis resulted in a list of binding sites that each had both a dox_{50} value and a value representing the average methylation level at each CpG. We then binned binding sites by whether the sites were <33%, >33% but less than 66%, or >66% methylated and plotted distributions of dox_{50} s.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13059-022-02792-x>.

Additional file 1: Fig. S1. Reproducibility of binding signal. RPKM signal from each replicate of CUT&Tag data across each TF across each dox induction concentration. Pearson's R correlation displayed on each graph. Fig. S2. Common saturation behavior binding pattern. 16 examples from different genomic sites showing saturating binding signal as dox induction increases. Signal is first read normalized (RPKM) and then normalized to the signal at the highest concentration. These sites were sampled from FOXA1 accessible binding sites, but are common across accessible and inaccessible HNF4A binding sites as well. Fig. S3. Sample of replicate fit binding curves. RPKM binding signal and fitted lines for each CUT&Tag replicate at 16 representative genomic loci. Fig. S4. Replicate dox_{50} distributions. Dox_{50} distributions extracted from fitted lines from each CUT&Tag replicate for each TF for each accessibility state. Fig. S5. Dox_{50} distributions without filtering out early saturation peaks. Dox_{50} distributions from all of the FOXA1 accessible binding sites ($n = 10,118$), FOXA1 inaccessible binding sites ($n = 17,644$), HNF4A accessible binding sites ($n = 16,137$), and HNF4A inaccessible binding sites ($n = 16,507$), without filtering out those peaks where binding signal peaked prior to the 5ug/ml dox sample. Fig. S6. Effect of DNA methylation on dox_{50} distributions. The average CpG methylation (% methylated reads at CpG) per sequence versus the sequence's dox_{50} at FOXA1 accessible ($R = 0.217$) (A), FOXA1 inaccessible ($R = -0.107$) (C), HNF4A accessible ($R = -0.002$) (E), or HNF4A inaccessible ($R = -0.126$) (G) sites. The dox_{50} distributions at FOXA1 accessible (B), FOXA1 inaccessible (D), HNF4A accessible (F), or HNF4A inaccessible (H) sites binned by whether the site's CpGs were <33% methylation, between 33% and 66% methylated, or >66% methylated. Fig. S7. Common "anti-cooperative" binding pattern. 16 examples from different genomic sites showing a pattern of increasing and then decreasing binding signal as dox induction increases. Signal is first read normalized (RPKM) and then normalized to the signal at the highest concentration. These sites were sampled from FOXA1 accessible binding sites.

Additional file 2: Table S1. Dox₅₀ values at TFBSs with hg19 coordinates.

Additional file 3. Review history

Acknowledgements

We thank members of the Cohen Lab for reading and critiquing the manuscript and for helpful discussion; Jessica Hoisington-Lopez and Marialynn Crosby in the DNA Sequencing Innovation Lab for assistance with high-throughput sequencing; the Genome Engineering and iPSC Center for allowing us to use their Sony Flow Cytometer for cell sorting; and Mingjie Li in the Hope Center Viral Vectors Core for assistance with producing lentiviral expression vectors.

Review history

The review history is available as Additional file 3.

Peer review information

Barbara Cheifet and Andrew Cosgrove were the primary editors of this article and managed its editorial process and peer review in collaboration with the rest of the editorial team.

Authors' contributions

J.L.H. and B.A.C. designed the overall project. J.L.H. conducted the experiments. J.L.H. conducted the analysis. J.L.H. and B.A.C. wrote the manuscript. The authors read and approved the final manuscript.

Funding

This work was supported by grants from the National Institutes of Health: R01GM092910 (Dr. Barak Cohen), T32HG000045 (Dr. Michael Brent, Washington University in St. Louis Genome Analysis Training Program), and T32GM007200 (Dr. Wayne Yokoyama, Washington University in St. Louis Medical Scientist Training Program).

Availability of data and materials

All genomic sequencing data have been deposited on Gene Expression Omnibus (GEO) under accession number GSE204726 [33]. We used previously published datasets of histone ChIP-seq and whole-genome bisulfite sequencing [15] to identify patterns of H3K27Ac, H3K4me1, H3K9me3, and H3K27me3 marks, as well as DNA methylation [19].

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Received: 2 June 2022 Accepted: 11 October 2022

Published online: 17 October 2022

References

1. Cirillo LA, Lin FR, Cuesta I, Friedman D, Jarnik M, Zaret KS. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol Cell*. 2002;9(2):279–89.
2. Iwafuchi-Doi M, Zaret KS. Pioneer transcription factors in cell reprogramming. *Genes Dev*. 2014;28(24):2679–92.
3. Hansen JL, Loell KJ, Cohen BA. The pioneer factor hypothesis is not necessary to explain ectopic liver gene activation. *Elife*. 2022;11. <https://doi.org/10.7554/eLife.73358>.
4. Zhu F, Farnung L, Kaasinen E, Sahu B, Yin Y, Wei B, et al. The interaction landscape between transcription factors and the nucleosome. *Nature*. 2018;562(7725):76–81.
5. Swinstead EE, Miranda TB, Paakinaho V, Baek S, Goldstein I, Hawkins M, et al. Steroid Receptors Reprogram FoxA1 Occupancy through Dynamic Chromatin Transitions. *Cell*. 2016;165(3):593–605.
6. Miller JA, Widom J. Collaborative competition mechanism for gene activation in vivo. *Mol Cell Biol*. 2003;23(5):1623–32.
7. Soufi A, Garcia MF, Jaroszewicz A, Osman N, Pellegrini M, Zaret KS. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell*. 2015;161(3):555–68.
8. Yu X, Buck MJ. Defining TP53 pioneering capabilities with competitive nucleosome binding assays. *Genome Res*. 2019;29(1):107–15.
9. Kaya-Okur HS, Wu SJ, Codomo CA, Pledger ES, Bryson TD, Henikoff JG, et al. CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*. 2019;10(1):1930.
10. Buenrostro JD, Wu B, Chang HY, Greenleaf WJ. ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr Protoc Mol Biol*. 2015;109:21.29.1–9.
11. Man TK, Stormo GD. Non-independence of Mnt repressor-operator interaction determined by a new quantitative multiple fluorescence relative affinity (QuMFRA) assay. *Nucleic Acids Res*. 2001;29(12):2471–8.
12. Biddy BA, Kong W, Kamimoto K, Guo C, Wayne SE, Sun T, et al. Single-cell mapping of lineage and identity in direct reprogramming. *Nature*. 2018;564(7735):219–24.

13. Sekiya S, Suzuki A. Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. *Nature*. 2011;475(7356):390–3.
14. Horisawa K, Udono M, Ueno K, Ohkawa Y, Nagasaki M, Sekiya S, et al. The Dynamics of Transcriptional Activation by Hepatic Reprogramming Factors. *Mol Cell*. 2020;79(4):660–76.e8.
15. Zhang J, Lee D, Dhiman V, Jiang P, Xu J, McGillivray P, et al. An integrative ENCODE resource for cancer genomics. *Nat Commun*. 2020;11(1):3696.
16. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet*. 2007;39(3):311–8.
17. Creighton MP, Cheng AW, Welstead GG, Kooistra T, Carey BW, Steine EJ, et al. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc Natl Acad Sci U S A*. 2010;107(50):21931–6.
18. Mayran A, Khetchoumian K, Hariri F, Pastinen T, Gauthier Y, Balsalobre A, et al. Pioneer factor Pax7 deploys a stable enhancer repertoire for specification of cell fate. *Nat Genet*. 2018;50(2):259–69.
19. ENCODE Project Consortium, Moore JE, Purcaro MJ, Pratt HE, Epstein CB, Shores N, et al. Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*. 2020;583(7818):699–710.
20. Biddie SC, John S, Sabo PJ, Thurman RE, Johnson TA, Schiltz RL, et al. Transcription factor AP1 potentiates chromatin accessibility and glucocorticoid receptor binding. *Mol Cell*. 2011;43(1):145–55.
21. Garcia MF, Moore CD, Schulz KN, Alberto O, Donague G, Harrison MM, et al. Structural Features of Transcription Factors Associating with Nucleosome Binding. *Mol Cell*. 2019. <https://doi.org/10.1016/j.molcel.2019.06.009>.
22. Rogerson C, Britton E, Withey S, Hanley N, Ang YS, Sharrocks AD. Identification of a primitive intestinal transcription factor network shared between esophageal adenocarcinoma and its precancerous precursor state. *Genome Res*. 2019;29(5):723–36.
23. Blassberg R, Patel H, Watson T, Gouti M, Metzys V, Delás MJ, et al. Sox2 levels regulate the chromatin occupancy of WNT mediators in epiblast progenitors responsible for vertebrate body formation. *Nat Cell Biol*. 2022;24(5):633–44.
24. Yan C, Chen H, Bai L. Systematic Study of Nucleosome-Displacing Factors in Budding Yeast. *Mol Cell*. 2018;71(2):294–305.e4.
25. Jiang G, Lee U, Sladek FM. Proposed mechanism for the stabilization of nuclear receptor DNA binding via protein dimerization. *Mol Cell Biol*. 1997;17(11):6546–54.
26. Rufibach LE, Duncan SA, Battle M, Deeb SS. Transcriptional regulation of the human hepatic lipase (LIPC) gene promoter. *J Lipid Res*. 2006;47(7):1463–77.
27. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res*. 2016;44(W1):W160–5.
28. Virtanen P, Gommers R, Oliphant TE, Haberland M, Reddy T, Cournapeau D, et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261–72.
29. Grant CE, Bailey TL. XSTREME: Comprehensive motif analysis of biological sequence datasets. *bioRxiv*; 2021. p. 2021.09.02.458722. Available from: <https://www.biorxiv.org/content/10.1101/2021.09.02.458722v1> [cited 22 Apr 2022]
30. Grant CE, Bailey TL, Noble WS. FIMO: scanning for occurrences of a given motif. *Bioinformatics*. 2011;27(7):1017–8.
31. Fornes O, Castro-Mondragon JA, Khan A, van der Lee R, Zhang X, Richmond PA, et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res*. 2020;48(D1):D87–92.
32. Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics*. 2010;26(6):841–2.
33. Hansen JL, Cohen BA. A quantitative metric of pioneer activity reveals that HNF4A has stronger in vivo pioneer activity than FOXA1. *Datasets*. *Gene Expression Omnibus*. 2022. <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE204726>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

