Genetic determinants and epigenetic effects of pioneer-factor occupancy

Julie Donaghey[1,2], Sudhir Thakurela [1,2], Jocelyn Charlton[2,3], Jennifer S. Chen[2], Zachary D. Smith[1,2], Hongcang Gu[1], Ramona Pop[2], Kendell Clement[1,2], Elena K. Stamenova[1], Rahul Karnik[1,2], David R. Kelley [2], Casey A. Gifford[1,2,5], Davide Cacchiarelli[1,2,6], John L. Rinn[1,2,7], Andreas Gnirke[1], Michael J. Ziller[4] and Alexander Meissner [1,2,3]*

Transcription factors (TFs) direct developmental transitions by binding to target DNA sequences, influencing gene expression and establishing complex gene-regulatory networks. To systematically determine the molecular components that enable or constrain TF activity, we investigated the genomic occupancy of FOXA2, GATA4 and OCT4 in several cell types. Despite their classification as pioneer factors, all three TFs exhibit cell-type-specific binding, even when supraphysiologically and ectopically expressed. However, FOXA2 and GATA4 can be distinguished by low enrichment at loci that are highly occupied by these factors in alternative cell types. We find that expression of additional cofactors increases enrichment at a subset of these sites. Finally, FOXA2 occupancy and changes to DNA accessibility can occur in G_1-arrested cells, but subsequent loss of DNA methylation requires DNA replication.

Organismal development is orchestrated by the selective use and distinctive interpretation of identical genetic material in each cell. During this process, TFs coordinate protein complexes at associated promoter and distal enhancer elements to modulate gene expression. However, the current understanding of the steps needed for the activation of silent cis-regulatory elements remains incomplete. A generally accepted model assumes that primary access to certain regulatory elements can be restricted by chromatin to ensure spatial and temporal control of gene expression during successive developmental stages[1]. Thus, there is a requirement for a distinct mechanism that transitions repressed cis-regulatory elements toward accessibility and consequently allows for coordinated binding of cell-type-specific TFs. This process may be accomplished through so-called pioneer TFs, which bind nucleosomal target sites and reorganize local chromatin to increase their acessibility[2–7]. Interestingly, despite their supposed universal targeting and remodeling capabilities, pioneer TFs also display a degree of cell-type specificity[8–10], and recent work suggests that cell-type-specific cofactors[11–13], signaling[14] and the underlying chromatin landscape[5,6] can influence their genomic occupancy. However, studying the contributions of individual TFs remains challenging when using native developmental systems, because extrinsic signals may induce rapid transitions downstream of the initial TF binding event without yielding intermediate states that are stable enough to be molecularly characterized. Moreover, cofactors and partially redundant family members may already be present in these systems, further complicating the isolation and interpretation of their individual roles. To overcome these limitations, we compared pioneer-TF occupancy at endogenously bound cis-regulatory elements across multiple cell types to an ectopic cellular environment, thereby allowing us to obtain new insights into the regulatory capabilities of the presumed pioneer TFs FOXA2, GATA4 and OCT4, which are frequently studied in development and used in cellular reprogramming.

Results

FOXA2 occupancy is cell-type specific and restricted to a subset of motif-containing loci. The FOXA motif contains seven core consensus nucleotides with a less distinct flanking sequence and is consequently abundant in the human genome[15] (Supplementary Fig. 1a). Because pioneer factors have the unique ability to access target loci in closed chromatin[16,17], they might be expected to extensively occupy genomic sites containing their core regulatory motif. To investigate this possibility, we determined the proportion of the preferred motif sequence that was occupied across several human cell types with detectable FOXA2 expression by chromatin immunoprecipitation sequencing (ChIP–seq), including HepG2 (hepatocellular liver carcinoma), A549 (lung carcinoma) and embryonic stem cell (ESC)-derived definitive endoderm (dEN)[18]. We used five position-weight matrices for the FOXA2 motif, mapped their positions across the human genome and then considered only those that overlapped with potential genetic regulatory regions, as defined by enrichment of activating histone modifications in at least one of the Encyclopedia of DNA Elements (ENCODE)[19]/Roadmap Epigenomics Project[20] cell types (Supplementary Fig. 1a,b and Methods). Only 6.3–13.7% of identified motifs were significantly bound by FOXA2 in these three cell types (Fig. 1a and Supplementary Fig. 1b; P value >0.01), and enrichment was largely cell-type specific (Fig. 1b), in agreement with results from prior studies[8–10]. However, we did not observe saturation of the binding spectra within the current selected cell types, and data from additional cell types will probably confirm that more motifs are targeted by FOXA2 (Supplementary Fig. 1c).
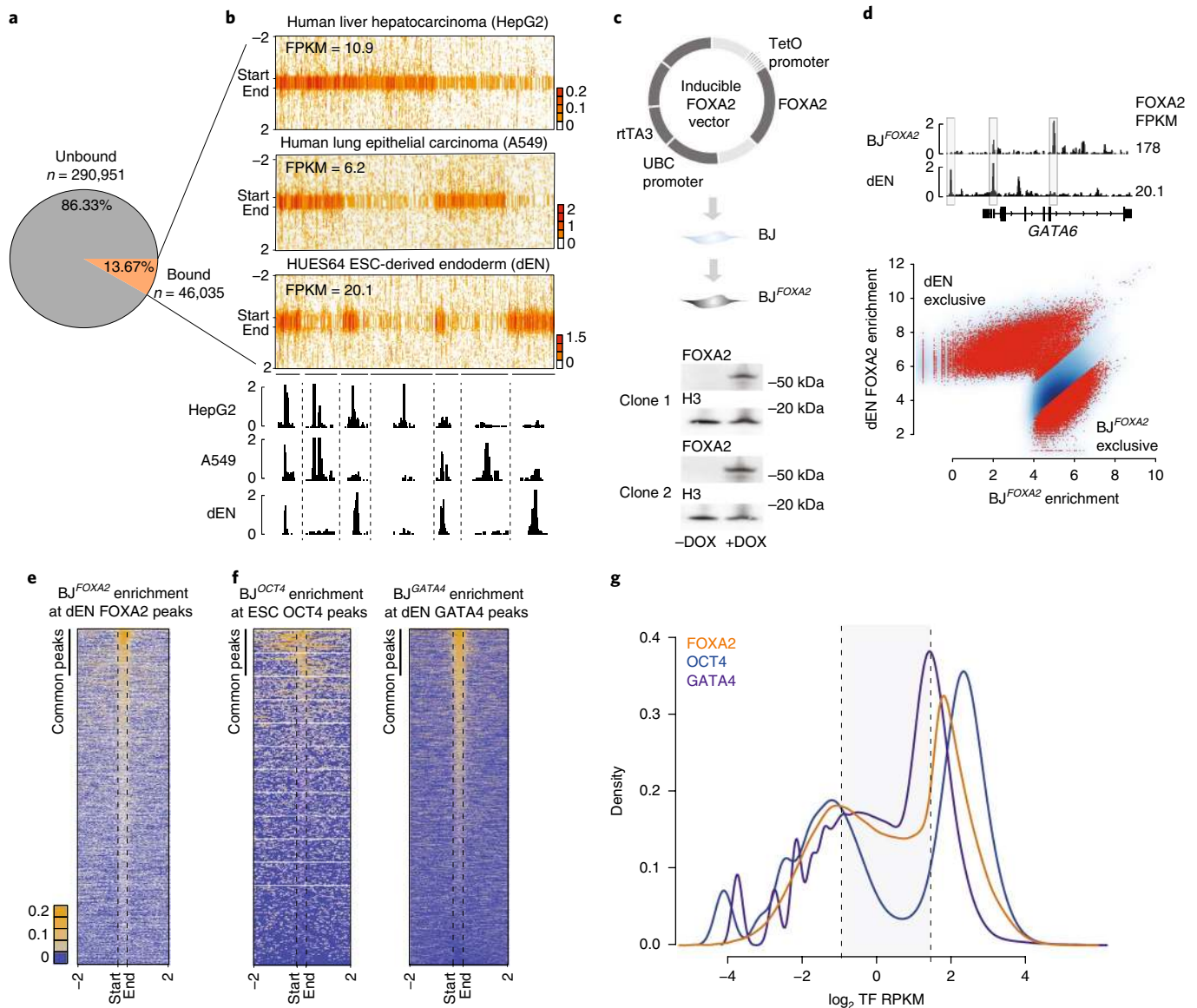
Fig. 1 | Ectopic FOXA2 and GATA4, but not OCT4, display low-level sampling. a, Pie chart displaying the percentage of FOXA motifs (Supplementary Fig. 1a,b) mapped across the genome that were unbound or bound by FOXA2 in HepG2, A549 and dEN cells. **b**, Read-density heat maps (normalized read count) for statistically called FOXA2 peaks in HepG2, A549 and dEN cells that overlapped a motif instance. 2 kb on each side of the peaks are shown. Heat maps are clustered according to the occurrence of binding across the three cell types. Genome-browser tracks from left to right highlight genomic occupancy across the three cell types: shared (chromosome (chr) 18: 9072728–9075158), unique in HepG2 (chr 18: 9202880–9225100), unique in A549 (chr18: 9008450–9022842), shared in A549 and HepG2 (chr 18: 8725886–8734843), shared in HepG2 and dEN (chr 4: 80986601–81000201), shared in A549 and dEN (chr 4: 75017694–75029960) and unique in dEN (chr 4: 74903404–74905306). **c**, Schematic of the pTripZ vector used for the generation of clonal FOXA2-inducible (BJ^FOXA2) cell lines. Cropped western blots of FOXA2 and H3 protein levels in two distinct BJ^FOXA2 clones. **d**, Browser tracks displaying differential binding across ectopic BJ^FOXA2 and dEN (chr 18: 19745852–19782939). FOXA2 FPKM values are shown at right. Below the scatter plot, output of DiffBind[42] differential peak-set analysis between dEN and BJ^FOXA2 is shown. Red dots, peaks with statistically significant differential enrichment between the two datasets. Axes are reads per kilobase of transcript per million mapped reads (RPKM). **e**, Read-density heat map (normalized read count) of ectopic FOXA2 enrichment in BJ^FOXA2 cells at dEN FOXA2 regions. Vertical side bar, peaks shared between BJ^FOXA2 and dEN. Dashed lines mark the start and end of peaks, extended by 2 kb on either side. Most dEN sites still showed low-level enrichment in BJ^FOXA2 yet were not called as significantly enriched. **f**, Read-density heat maps of ectopic OCT4 signal in BJ^OCT4 at human ESC occupied regions ($n = 22{,}477$) and ectopic GATA4 signal in BJ^GATA4 at dEN occupied regions ($n = 42{,}477$). Vertical side bar, peaks shared between the ectopic and endogenous context for either factor, as shown in **e** for FOXA2. In contrast to FOXA2 and GATA4, fewer ESC OCT4 sites showed any notable enrichment in BJ^OCT4. **g**, Density plot displaying FOXA2, OCT4 and GATA4 ectopic enrichment in BJs at union sets of ectopic and endogenous sites (FOXA2, orange; OCT4, navy; GATA4, purple). Dashed lines demarcate regions within the background distribution, sampled sites (shaded) and peaks.

Because primary TF engagement cannot be adequately dissected by using endogenous systems that already express FOXA2 as part of their regulatory circuitry, we engineered a doxycycline (DOX)-inducible system in immortalized foreskin fibroblasts (BJ), which do not normally express FOXA2 or other FOXA family members (Supplementary Fig. 1d). We derived several clonal cell lines
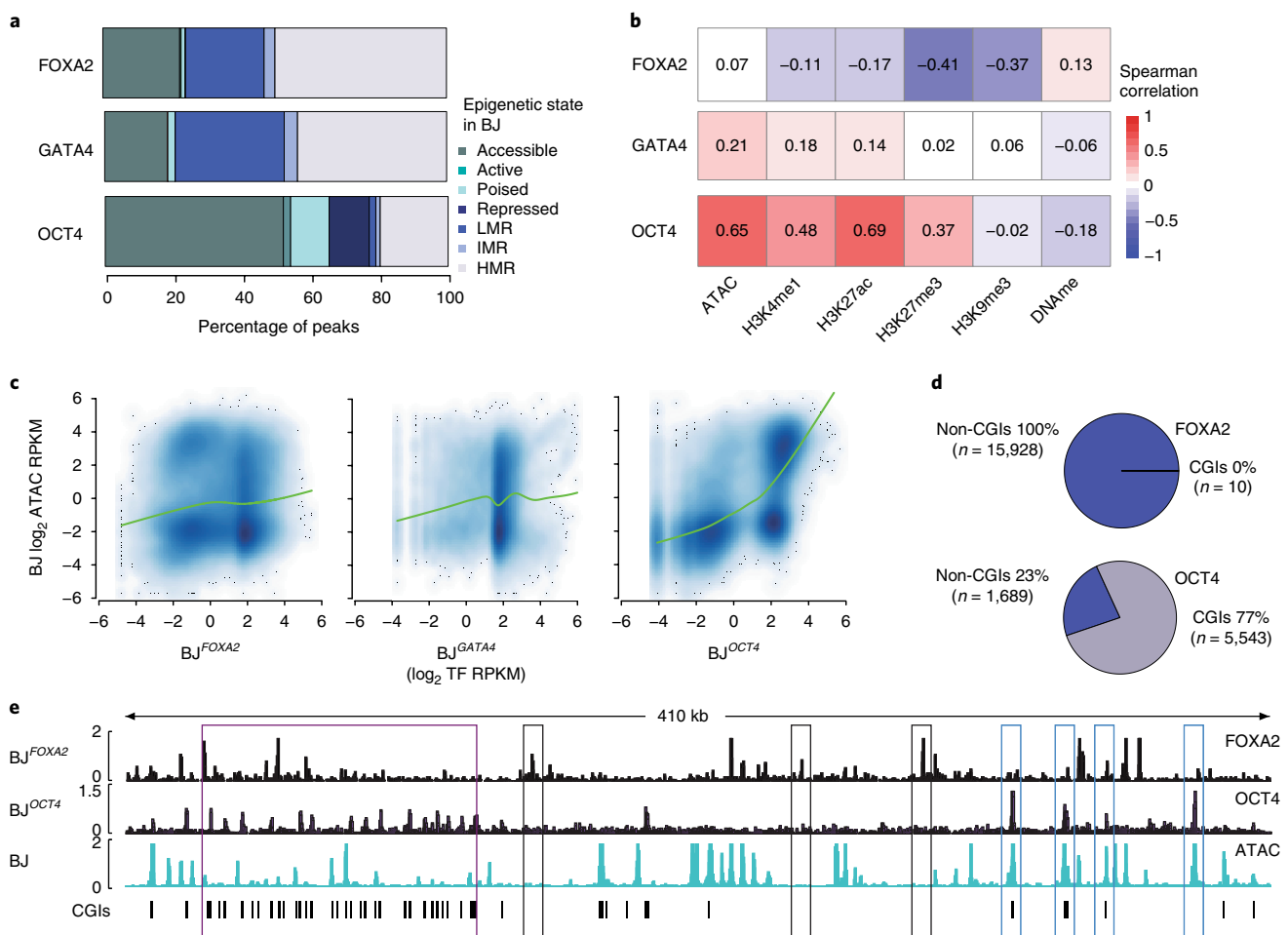
Fig. 2 | Influence of prior epigenetic state on TF occupancy. a, Percentage of TF-bound regions in BJ^{FOXA2}, BJ^{GATA4} and BJ^{OCT4} with assigned chromatin states, which were defined hierarchically in BJ before TF induction. First, 'accessible' regions were categorized by ATAC–seq enrichment. Then regions highly enriched in H3K27ac or H3K4me1 were categorized as 'active' or 'poised', respectively. Regions enriched in H3K27me3 or H3K9me3 were categorized broadly as 'repressed'. Finally, all remaining regions were grouped on the basis of their DNAme levels: highly methylated regions (HMRs, >60% mean methylation), intermediately methylated regions (IMRs, 20–60% mean methylation) and lowly methylated regions (LMRs, <20% mean methylation). LMRs are a 'low signal' state that lacks DNA accessibility as well as enrichment of any assessed histone modifications[4]. **b**, Spearman correlations between TF enrichment for peak sets as described in Fig. 1g and epigenetic features, displayed as a heat map. **c**, Scatter plots and LOWESS fit curves (green line) of FOXA2, GATA4 and OCT4 versus ATAC–seq enrichment in BJ before TF induction. **d**, Pie charts summarizing the percentage of FOXA2 and OCT4 targets that overlapped defined preexisting closed chromatin and were located within or outside of annotated CGIs. **e**, Representative genome-browser tracks displaying FOXA2 and OCT4 enrichment compared with preinduced BJ ATAC–seq data (chr 5: 140657329–141085891). Purple box highlights regions of OCT4 binding in preexisting closed chromatin that overlapped annotated CGIs. Gray boxes highlight FOXA2 binding at preexisting closed chromatin, and blue boxes highlight OCT4 binding in regions of preexisting open chromatin.

(referred to as BJ^{FOXA2}) with no detectable FOXA2 in the uninduced state but rapid, uniform and consistent mRNA/protein induction after DOX treatment (Fig. 1c and Supplementary Fig. 1e–g) and performed ChIP–seq for FOXA2 after 1, 4 and 10 d of induction. We observed a clear increase in FOXA2 binding between 1 and 4 d but little change afterward (Supplementary Fig. 1h,i and Supplementary Table 1) and identified a total of 49,830 consensus peaks for the combined 4- and 10-d time points, of which 98% contained a FOX family motif[21] (Supplementary Fig. 1j). Despite supraphysiological levels, we still primarily observed cell-type-specific FOXA2 binding: ~ 70% of FOXA2 peaks showed differential enrichment between dEN and BJ^{FOXA2} (Fig. 1d). Therefore, DNA sequence alone is clearly insufficient to direct binding, because many potential FOXA2 targets remained unbound in ectopic conditions.

FOXA2 and GATA4 demonstrate low-level sampling at many of their alternative lineage targets. Although we observed only a

partial overlap between significantly called FOXA2 peaks in endogenous and ectopic contexts, we nevertheless noticed consistent low-level FOXA2 enrichment in BJ^{FOXA2} at most regions that are highly occupied in dEN, HepG2 and A549 cells (Fig. 1e and Supplementary Fig. 2a). Within the union set of previously defined endogenous and ectopic peaks, a notable number of statistically unoccupied regions still displayed low to intermediate FOXA2 enrichment, as compared to the genomic background (Fig. 1g). To determine whether this low-level enrichment was a general feature of ectopic TF expression, we engineered inducible BJ fibroblasts for two other presumed pioneer TFs, OCT4 and GATA4 (BJ^{OCT4} and BJ^{GATA4}; Supplementary Fig. 2b and Supplementary Tables 2 and 3) and found a comparable low-level enrichment for GATA4 but not for OCT4 (Fig. 1f,g). However, ectopic OCT4 has been shown to display low-level enrichment at ESC OCT4 targets when it is coexpressed with SOX2, KLF4 and cMYC in BJ fibroblasts[5], indicating that this ability is context and cofactor dependent (Supplementary Fig. 2c).
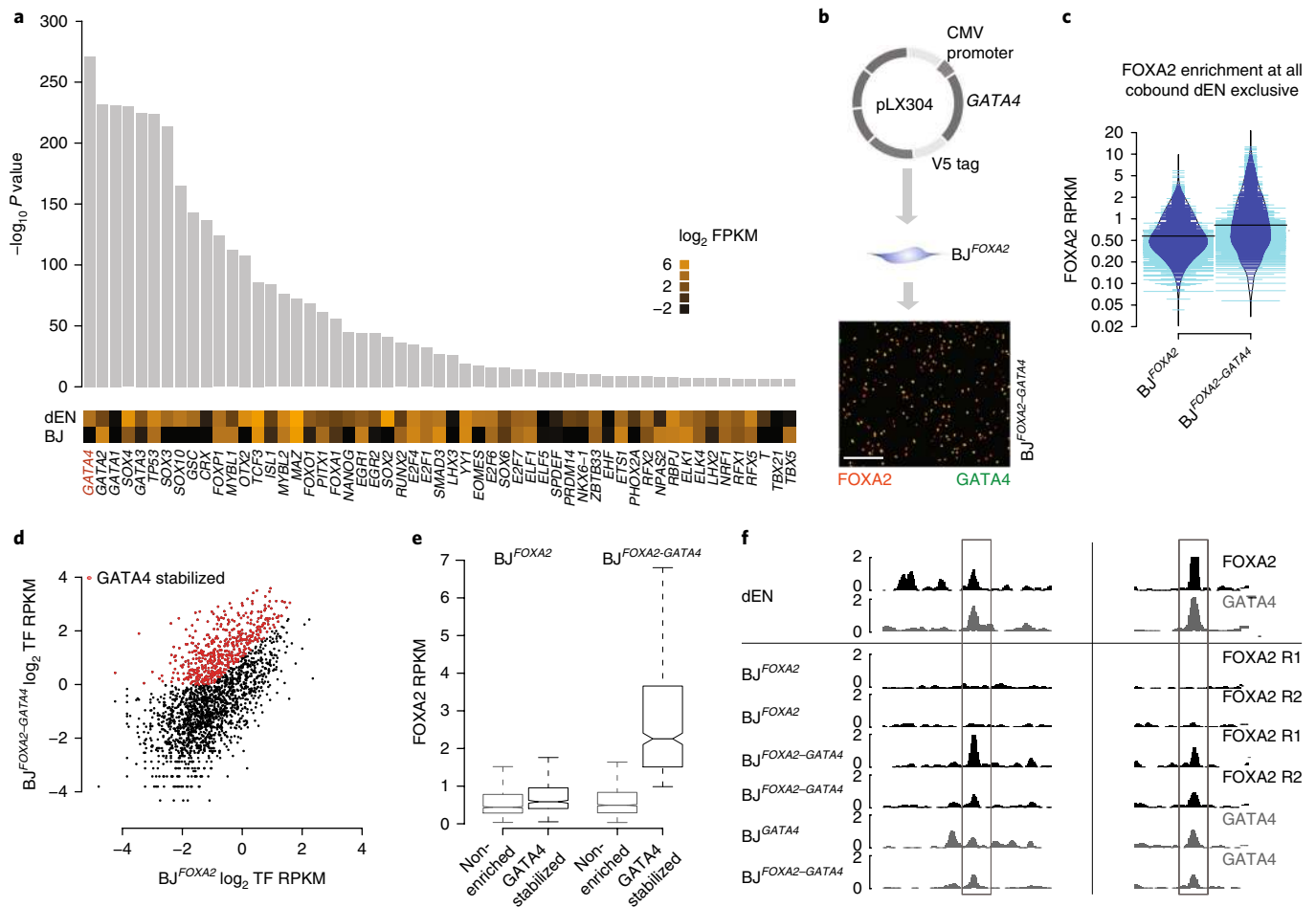
Fig. 3 | Cofactor expression modulates FOXA2 enrichment at a subset of target sites. a, Enrichment of TF motifs in dEN-exclusive sites versus BJ^FOXA2-exclusive sites. Bottom shows expression (log$_2$ FPKM) of the TFs associated with the listed motifs in both dEN and BJ^FOXA2. Of note, although there were many significant differential motifs observed in dEN-exclusive sites, not all motifs were associated with factors that displayed differential expression. **b**, Simplified schematic of the ectopic expression system used to coexpress GATA4 in BJ^FOXA2. Immunostaining of FOXA2 and V5-GATA4 in BJ^FOXA2-GATA4. White scale bar, 345 nm. **c**, Bean plot comparing FOXA2 enrichment at all dEN-exclusive cobound sites ($n = 2,093$) in BJ^FOXA2 and BJ^FOXA2-GATA4. Horizontal black bars, averages; blue lines, data points within the distribution; teal lines, data points outside the distribution. **d**, Scatter plot comparing FOXA2 enrichment of cobound dEN-exclusive sites. Red dots, regions that gained at least twofold enrichment and had RPKM >1 in BJ^FOXA2-GATA4 ($P \leq 2 \times 10^{-16}$). **e**, Box plots displaying the RPKM of FOXA2 enrichment in BJ^FOXA2 and BJ^FOXA2-GATA4 at GATA4 stabilized sites compared with nonenriched sites. 318 of these 504 regions display low FOXA2 enrichment in BJ^FOXA2. Boxes indicate interquartile range, and whiskers show maximum and minimum values. Horizontal line, median. Outliers are removed. **f**, Representative browser tracks of FOXA2 and GATA4 enrichment in dEN and BJ lines, as indicated at left. Gray boxes, regions showing cofactor-mediated recruitment of FOXA2 to its dEN targets at chr 13: 76031782–76039815 (left) and chr 17: 14352627–14360300 (right).

Additionally, we found that this low-level enrichment was also present in cells endogenously expressing FOXA2 and therefore is not just a product of ectopic or supraphysiological expression (Supplementary Fig. 2d,e).

Influence of prior epigenetic state on FOXA2, GATA4 and OCT4 binding. To determine how a cell's preexisting epigenome might affect pioneer-factor binding, we performed ChIP–seq for select histone modifications associated with active (acetylated histone H3 K27 (H3K27ac) and monomethylated H3 K4 (H3K4me1)) and repressive (trimethylated H3 K27 (H3K27me3)) states, assay for transposon-accessible chromatin (ATAC–seq)[22] for DNA accessibility and whole-genome bisulfite sequencing (WGBS) for DNA methylation (DNAme) levels (Supplementary Table 4). We then defined chromatin states by using simple hierarchical rules reflecting prior knowledge of these modifications and how they interact (Supplementary Fig. 3a). For this analysis, we focused on the most highly enriched targets in a given cell

type and found that ectopic FOXA2 and GATA4 predominantly engaged sites that are devoid of the selected histone modifications and contain variable DNAme levels (Fig. 2a). Endogenous FOXA2 displayed a similar behavior in undifferentiated ESCs at sites that are bound by FOXA2 in dEN (Supplementary Fig. 3b). There was little correlation between FOXA2 or GATA4 enrichment and selected epigenetic features, yet OCT4 binding was positively correlated with preexisting accessible chromatin (Fig. 2a–c and Supplementary Fig. 3b) and frequently overlapped with CpG islands (CGIs) (Fig. 2d,e and Supplementary Fig. 3c). To compare these behaviors to the ectopic binding of a presumed nonpioneer factor, we generated another BJ line expressing hepatocyte nuclear factor 1A (HNF1A). We observed no significant HNF1A binding when expressed alone, but enrichment became readily detectable in combination with FOXA2 (Supplementary Fig. 3d–f). Finally, FOXA2 enrichment was generally depleted in H3K9me3 heterochromatin domains[5] (Supplementary Fig. 3g–i). However, few endogenously occupied FOXA2 regions reside
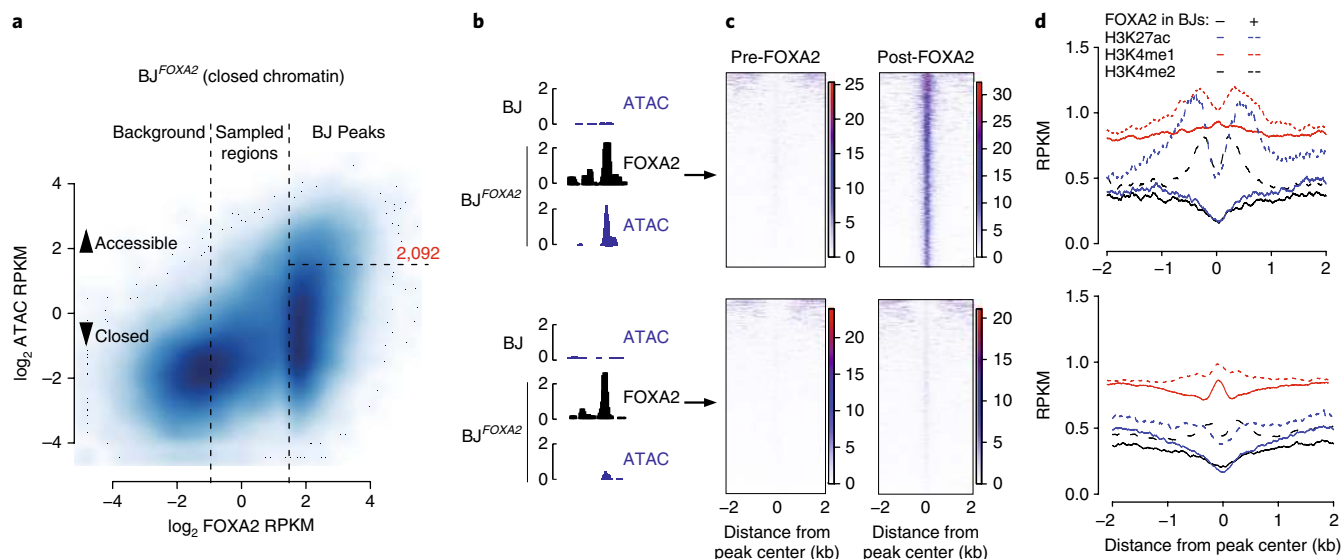
Fig. 4 | Epigenetic effects of ectopic FOXA2 binding. a, Scatter plot displaying FOXA2 enrichment in BJ^FOXA2 compared with ATAC–seq signal at the union set of all FOXA2-binding sites that are located within preexisting closed-chromatin regions (BJ ATAC RPKM <1). Vertical lines separate the union set of FOXA2 peaks into background, sampled and called peaks within BJ^FOXA2, as defined in Fig. 1. Horizontal line demarcates regions in BJ^FOXA2 with ATAC RPKM ≥3. **b**, Representative genome browser tracks of regions that become accessible after FOXA2 binding (top; chr 20: 36008193–36009335) versus regions that remain inaccessible (bottom; chr 19: 1867722–1868322). **c**, Read-density heat maps (normalized read counts) of ATAC-seq signal prior to and after FOXA2 induction for regions as classified in **a**. **d**, Composite plots displaying mean H3K27ac, H3K4me1 and H3K4me2 enrichment over regions in **c**, prior to (solid lines) and after (dashed lines) FOXA2 induction.

within these domains, which are therefore unlikely to be the major cause of the cell-type-specific occupancy (Supplementary Fig. 3j).

GATA4 coexpression increases FOXA2 enrichment at a subset of previously sampled targets. Given the limited ability of the epigenome to determine FOXA2 binding, we speculated that occupancy might instead be directed primarily through cooperativity with cell-type-specific cofactors, as is common among nonpioneer TFs[1,23] and has recently also been suggested for some pioneer TFs[12,13]. To identify potential cofactors, we searched for differentially enriched motifs between regions bound by FOXA2 exclusively in dEN or BJ^FOXA2 and cross-referenced those motifs against RNA-seq data for the corresponding expressed TFs (Fig. 3a and Supplementary Fig. 4a). Motif sequences for several known endodermal regulators were enriched at dEN-exclusive sites, including GATA4, which is known to bind to the *ALB* enhancer locus with FOXA2 in early gut endoderm cells prior to *ALB* expression[24–26]. Thus, we selected GATA4 as a candidate cofactor that might influence FOXA2 binding in the ectopic system. Using our previously published data[27] for FOXA2 and GATA4 binding in dEN, we found that the two factors colocalize at 2,364 genomic sites, 2,093 of which overlap with FOXA2 dEN-exclusive targets (Supplementary Fig. 4b). We infected BJ^FOXA2 with a second lentiviral construct containing constitutively expressed V5-tagged *GATA4*, induced simultaneous expression of both factors for 4 d and performed ChIP–seq for FOXA2 (Fig. 3b; BJ^FOXA2–GATA4). We found that a specific subset of targets (504 of 2,093) displayed a substantial increase upon GATA4 coexpression (referred to as 'GATA4 stabilized'; Fig. 3c–f). Intriguingly, the majority of the GATA4-stabilized sites show evidence of low enrichment sampling when either FOXA2 or GATA4 are induced alone (Fig. 3e and Supplementary Fig. 4c). Ultimately, coexpression could explain only a subset of the dEN-exclusive cobound FOXA2 targets. When we searched for additional confounding factors, we found that the GATA motif was differentially enriched in the GATA4-stabilized subset compared with the nonenriched subset ($P = 1.0 \times 10^{-5}$; motif occurring at 76% of regions). In addition, we observed weak differential enrichment of other endodermal-TF motifs at regions where

GATA4 did not stabilize FOXA2 binding (T box, $P = 1.0 \times 10^{-3}$; Eomes, $P = 1.0 \times 10^{-3}$; SOX, $P = 1.0 \times 10^{-3}$), suggesting that FOXA2 occupancy at these regions may be dependent on multiple TFs. In general, we found that FOXA2 binding at GATA4-stabilized targets did not appreciably increase chromatin accessibility, which indicates that the recruitment of additional chromatin-remodeling machinery had probably not yet occurred (Supplementary Fig. 4d). Instead, cooperative pioneer-TF activity appears to assist in loading to nucleosomal DNA and not nucleosome eviction. Collectively, our data support a model wherein pioneer-factor occupancy is genetically encoded and partially determined by cofactor engagement at specific subsets of target loci.

Transcriptional and epigenetic effects of ectopic FOXA2 binding. To determine the molecular effects of ectopic TF binding, we performed RNA-seq, ATAC–seq and ChIP–seq 48 h after FOXA2 induction. In line with results from previous studies on pioneer factors[7,18], we found only a small number of genes that were immediately responsive to induction (299 genes upregulated and 191 genes downregulated) with only 82 genes directly associated with promoter-proximal FOXA2 binding (±1 kb from the TSS; Supplementary Fig. 5a and Supplementary Table 5). Because of the limited transactivating properties, we focused on chromatin changes after FOXA2 occupancy and found many regions that either acquired de novo or exhibited increased H3K4me1, dimethylated H3 K4 (H3K4me2) and H3K27ac after FOXA2 induction (Supplementary Fig. 5b). De novo H3K4 methylation following FOXA factor occupancy has previously been observed to establish competency at cis-regulatory regions[7,28–30]. Additionally, we found that 1,937 of 4,962 de novo H3K4me2 regions concomitantly gained low enrichment of H3K27ac as well.

Next, we focused on FOXA2-occupied regions within preexisting closed chromatin and measured induced changes to DNA accessibility and histone modification. Occupancy alone appeared to be insufficient to affect global accessibility, because only a fraction (~13%) of BJ^FOXA2 targets demonstrated significant gains in ATAC–seq signal (Fig. 4a,b; $n = 2,092$ of 15,888, $P \leq 2.2 \times 10^{-16}$).
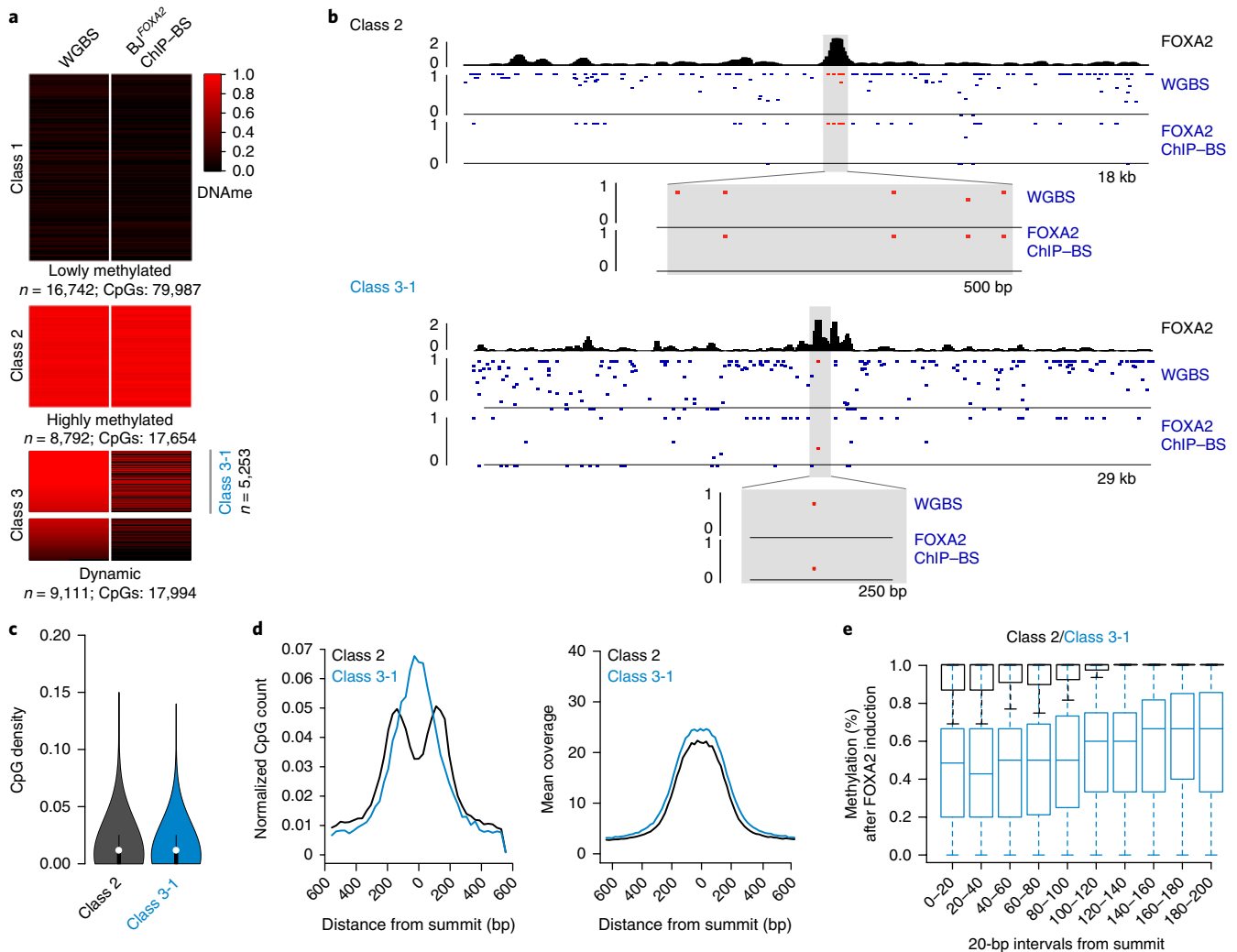
Fig. 5 | Distance-dependent loss of DNAme upon FOXA2 binding. a, Heat map of CpG methylation levels for matched CpGs, comparing BJ WGBS and BJ^{FOXA2} ChIP–BS-seq data. Three main classes of FOXA2 binding responses are displayed. Gray bar in class 3 indicates the subset of dynamic regions with a preinduction methylation level of ≥80% ($n = 5{,}253$ regions further referred to as class 3-1). **b**, Representative genome-browser tracks showing BJ^{FOXA2} FOXA2 enrichment as well as CpG methylation levels prior to and after induction. Top, example of a class 2 region located at chr 12: 54002592–54021127. Bottom, an example of a class 3 region located at chr 18: 32911411–32941267. Zoomed views display CpG locations within the peak. **c**, Violin plot of CpG density of class 2 and class 3 target sites. CpG density was calculated as the number of CpG dinucleotides across 100-bp windows divided by the total number of base pairs. Dots, median values; black lines, interquartile ranges. **d**, Composite plots showing normalized CpG counts of class 2 and class 3 target sites (left). Class 2 was depleted in CpGs toward the center of the peak, whereas class 3 targets were enriched. The mean sequencing coverage between class 2 and class 3 target sites was equivalent (right). **e**, CpG-methylation box plots of 20-bp windows from the peak summit and extended to 200 bp. Methylation measurements were taken from ChIP–BS-seq data. Class 2, black; class 3-1, blue. Boxes indicate interquartile range, and whiskers show maximum and minimum values. Horizontal lines, medians.

Despite the infrequent increase in DNA accessibility in BJ^{FOXA2}, 61% of unchanged targets (5,144 of 8,443) overlapped with putative gene-regulatory elements that are accessible in at least one other cell type, on the basis of all available ENCODE DNase-hypersensitivity data[31]. However, we detected some low-level increases in ATAC–seq signal even at the target sites that remained inaccessible on the basis of our thresholds (Fig. 4c and Supplementary Fig. 5c). We identified a number of distinguishing features that characterized sites that significantly gained accessibility from those that did not. First, FOXA motifs were more highly enriched and widely distributed across the region (Supplementary Fig. 5d–f). Second, the mean ATAC–seq signal prior to FOXA2 induction was slightly higher (Supplementary Fig. 5g). Third, after FOXA2 induction, we observed enrichment in phased nucleosomes modified by H3K4me1/me2 and H3K27ac surrounding the FOXA2 peak summit (Fig. 4d). Scatter plots of binned

ATAC–seq signal compared with histone-modification enrichment demonstrated a somewhat linear relationship between gain in DNA accessibility and gain in H3K4me1 (along with a weaker enrichment and correlation for H3K4me2 and H3K27ac; Supplementary Fig. 5h). However, the enrichment of activating histone modifications at these regions was quite modest and did not reach enrichment levels similar to those seen for active promoters, for example (Supplementary Fig. 5i). Furthermore, we observed that most of the ATAC–seq signal gained after FOXA2 occupancy was lost within 2 d of factor withdrawal, indicating the transient behavior of this remodeling, in agreement with prior work[3] (Supplementary Fig. 5c,j).

DNA methylation dynamics at FOXA2 targets. We next investigated FOXA2-mediated demethylation in BJ^{FOXA2}, because it has been shown to be associated with loss of methylation at target
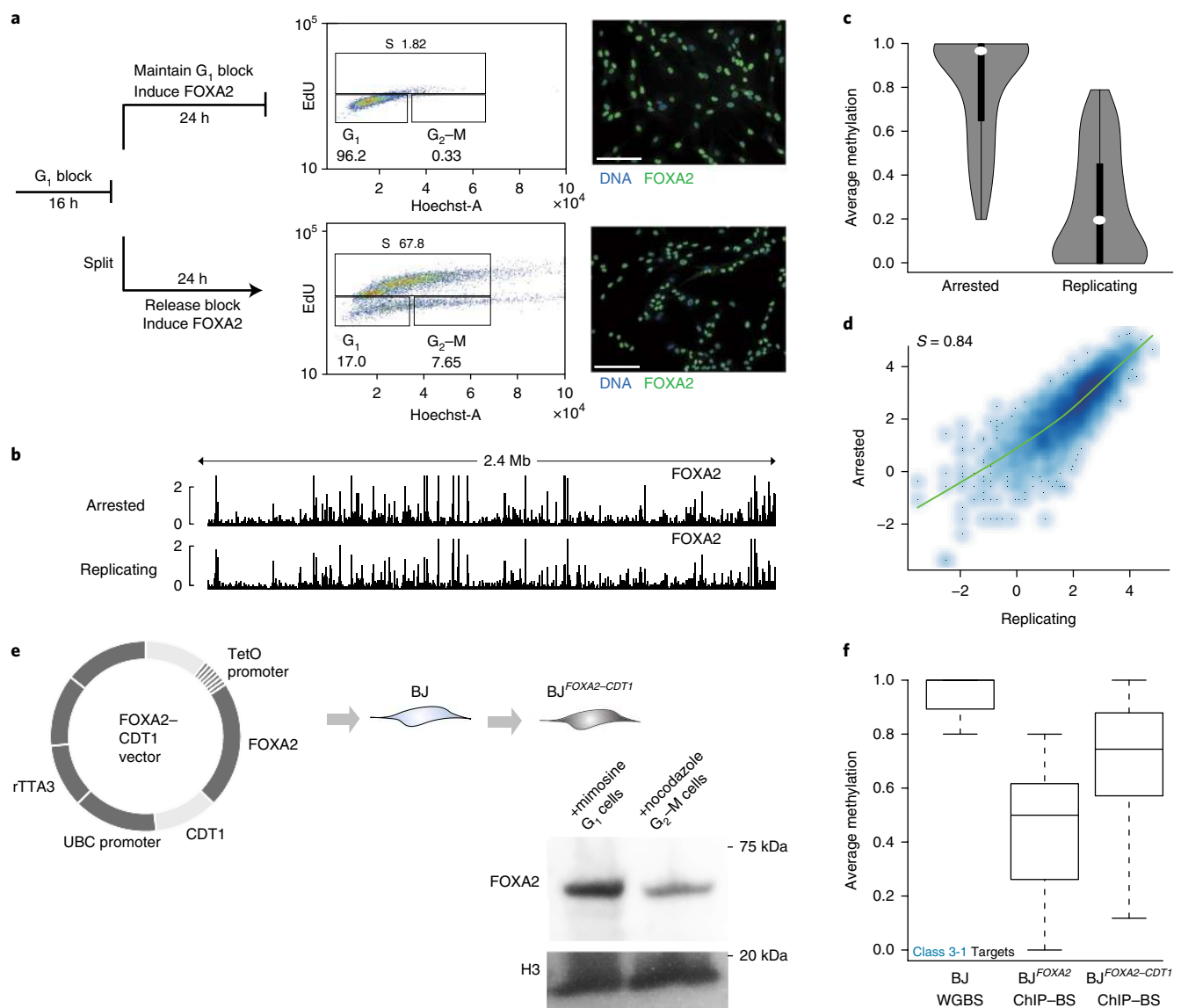
Fig. 6 | FOXA2 can occupy and change DNA accessibility without cell-cycle progression but requires replication to remove DNAme. a, Cells were halted before DNA replication by blocking the BJ^{FOXA2} from progressing through G$_1$ with mimosine, then were either released into normal replicating conditions through cell splitting and washing out the chemical treatment (replicating) or maintained in a halted state through persistent treatment (arrested). FOXA2 was simultaneously induced in both groups for 24 h. Fluorescence-activated cell-sorting analyses using EdU incorporation (488 nm) and Hoechst DNA staining show that halted cells remained in G$_1$ and that cells removed from the chemical block proliferated at a normal rate. Both populations expressed FOXA2, as shown by immunostaining. White scale bar, 345 nm. **b**, Representative genome-browser tracks of a 2.4-Mb window (chr 8: 57156834–59597409) of FOXA2 ChIP–seq data from cells halted in G$_1$ (top) and cells halted and then released back in to normal cycling conditions (bottom). FOXA2 binding and accumulation is visually similar in both experiments. **c**, Violin plots of methylation levels for class 3 targets (4,055 FOXA2 peaks covering 5,386 CpG sites) for cells arrested in G$_1$ compared with those released into replicating conditions. Dynamic regions no longer lost DNAme when cells were arrested in G$_1$. Dots, median values; bars, interquartile ranges. **d**, Scatter plot of ATAC–seq signal (log$_2$ RPKM) after FOXA2 induction in G$_1$-arrested cells compared with cells released back into normal cycling conditions, at regions that become accessible in BJ^{FOXA2} (Fig. 4a). Green line, LOWESS fit curve. Spearman correlation indicates that the induced gain in accessibility between the two samples is highly correlated. **e**, Schematic representation of the generated FOXA2–CDT1 fusion lentiviral construct. Corresponding cropped western blot for FOXA2 in BJ$^{FOXA2-CDT1}$ cells treated with mimosine to enhance the proportion of cells in G$_1$ or with nocadazole to enhance the proportion of cells in G$_2$–M. H3 was used as a loading control. **f**, Methylation box plots of class 3-1 targets in BJ WGBS, BJ^{FOXA2} and BJ$^{FOXA2-CDT1}$ ChIP–BS-seq data, calculated on the basis of regions with at least 10× coverage. Boxes, interquartile ranges; whiskers, maximum and minimum values. Horizontal lines, medians. Outliers were removed.

sites[18,29]. To quantify DNAme levels on fragments that were physically associated with FOXA2, we performed chromatin immunoprecipitation followed by bisulfite sequencing (ChIP–BS-seq)[32] after 4-d induction and compared methylation levels with those from WGBS data from preinduced cells (Supplementary Table 6). We found that FOXA2 occupied three distinct sets of genomic

regions: those in preexisting lowly methylated DNA that remained as such after FOXA2 binding (Fig. 5a; class 1, $n = 16,742$); those that displayed high DNAme levels before and after FOXA2 binding (Fig. 5a,b; class 2, $n = 8,794$); and a unique class of regions displaying a clear loss of DNAme after FOXA2 binding ($\geq 20\%$ change, Fig. 5a,b, dynamic; class 3, $n = 9,111$). Of note, all three classes were

reproducible even when the N- or C-terminal domain of FOXA2 was deleted, indicating that neither domain was responsible for the local demethylation observed at class 3 targets (Supplementary Fig. 6). Because of the unexpected heterogeneity in response to FOXA2 binding, we scrutinized the differences between class 2 and 3 target sites in further detail. First, we confirmed the interaction between DNAme and FOXA2 by using in vitro electromobility shift assays (EMSAs) and found no preference for methylated, hemimethylated or unmethylated DNA (Supplementary Fig. 7a,b). We next selected a stringent subset of class 3 targets with high methylation (≥80%) in uninduced BJ^{FOXA2} cells (Fig. 5a; class 3-1, $n = 5,253$) to be more comparable to the mean methylation of class 2 targets (≥80%). Importantly, FOXA2 enrichment did not correlate with changes in DNAme levels at class 3-1 targets (Supplementary Fig. 7c), and both class 2 and class 3-1 target loci were largely indistinguishable in their genomic location and CpG density (Fig. 5c; mean CpG count 4.2 and 4.8 for class 2 and 3-1, respectively). However, after closer inspection, we found that the class 2 targets were comparatively depleted of CpG dinucleotides toward the peak summit (Fig. 5d). Additionally, the distance from the peak summit to the nearest CpG was significantly greater for class 2 versus class 3-1 targets (average 74 bp and 90 bp, respectively; $P \leq 2.2 \times 10^{-16}$), whereas the average methylation for these nearest CpGs prior to FOXA2 induction was indistinguishable (Supplementary Fig. 7d–f; $P = 0.95$, average 95% methylated in both). Of note, the change in CpG methylation is greatest toward the peak summit (Fig. 5e and Supplementary Fig. 7g), a result that, together with the limited observed chromatin dynamics (Supplementary Fig. 7h,i), suggests that loss of DNAme is unlikely to be a result of recruited histone-modifying enzymes and may be physically linked to FOXA2 occupancy.

Loss of DNAme depends on DNA replication. The transition from a methylated to an unmethylated base may require an active enzymatic removal of the methyl group[33], a passive replication-dependent loss (which would require blocking any maintenance activity after nascent-DNA synthesis) or a combination of both[34,35]. To investigate this mechanism as it occurs for class 3 targets, we used mimosine treatment to reversibly halt BJ^{FOXA2} cells in G$_1$ before DNA replication, then induced FOXA2. We subsequently either continued mimosine treatment or restored cell-cycle progression (as verified by 5-ethynyl-2′-deoxyuridine (EdU) incorporation) for approximately one to two rounds of cell division (Supplementary Fig. 8a). We then collected the cells and performed FOXA2 ChIP–BS-seq (Fig. 6a and Supplementary Tables 7 and 8). Notably, FOXA2 occupied similar genomic regions in both conditions, indicating that it can access these loci even in arrested cells (Fig. 6b and Supplementary Fig. 8b). Quite strikingly, arrested cells displayed no measurable decrease in DNAme levels at FOXA2-occupied regions, despite changes to DNA accessibility (Fig. 6c,d). Together, these results highlight that FOXA2 binding and its effects on DNA accessibility are replication independent, whereas loss of DNAme is replication dependent.

Mechanistically, we hypothesized that the immediate recruitment of FOXA2 to target regions after DNA replication (S phase) might be sufficient to block maintenance methylation by DNMT1. To explore this possibility, we generated BJ fibroblasts expressing FOXA2 fused to CDT1 (ref. [36]) (BJ$^{FOXA2-CDT1}$) to specifically deplete FOXA2 expression during S phase (Fig. 6e). FOXA2 protein levels were higher in G$_1$ compared with G$_2$–M phases, although the presense of residual protein was still observed, possibly as a result of the supraphysiological expression (Fig. 6e). To ensure similar target-site enrichment of FOXA2 in this new system, we performed ChIP–seq in G$_1$-arrested BJ$^{FOXA2-CDT1}$ cells, in which FOXA2 protein levels were most similar to those in BJ^{FOXA2}, and observed high correlation compared with either arrested or released BJ^{FOXA2} (Supplementary Fig. 8c,d). We then induced FOXA2 for 4 d, performed ChIP–BS-seq in normal cycling conditions, and observed substantially reduced

loss of DNAme at class 3-1 targets (Fig. 6f and Supplementary Fig. 8e,f). Together, our results suggest that FOXA2 occupancy in S phase may be necessary to facilitate targeted loss of DNA methylation.

Discussion

Here, we compiled a set of cis-regulatory elements that are occupied by FOXA2 in endogenously expressing cell types (HepG2, A549 and dEN) and established that ectopic expression in BJ fibroblasts did not recapitulate the high enrichment occupancy at most of these endogenous targets, despite supraphysiological expression. Instead, we observed a minimal overlap between endogenous and ectopic datasets but observed broad low-level enrichment (sampling) for most regions that are occupied by FOXA2 in alternative lineages and cell lines. Sampling may occur as a result of slow chromatin scanning of pioneer factors[37], yet recent evidence from single-molecule-tracking studies suggests that pioneer factors may actually have rapid DNA-residence times that are similar to those of nonpioneer TFs[12]. From these experiments, it is not possible to distinguish whether sampled sites were bound with high frequency/stability in a small number of cells or whether more transient binding occurred at these regions within all cells, because standard ChIP–seq signals are averaged across populations. Future application of competition ChIP experiments or similar approaches might yield further insight.

Nevertheless, sampling appears to be a distinctive characteristic of FOXA2 and GATA4, as OCT4 predominantly exhibited highly enriched cell-type-specific occupancy. This result is consistent with findings from a recent study showing that mouse OCT4 also occupies distinct genomic regions when it is expressed alone or with other reprogramming factors[38]. Notably, the same study has found that, when induced alone, OCT4 primarily occupies regions of preexisting open chromatin[38]. Sampling of alternative target sites may therefore be a defining pioneer-TF quality that factors such as OCT4 acquire within only specific cellular contexts or in combination with additional factors.

We find that cell-type specific binding is most strongly influenced by the presence of additional cofactors, which stabilize FOXA2 at regions that are otherwise only sampled. Modest changes in pioneer-factor occupancy due to cofactor expression have also recently been observed by others[12,13]. The cooperativity that we observed between FOXA2 and GATA4 at this specific subset of target sites appears to be distinct from the dynamic assisted-loading model of TF binding[12,23], because we found little change in DNA accessibility by using ATAC–seq when these regions are cooccupied. Because GATA4 coexpression stabilized FOXA2 binding at a subset of potential sites, the presence of additional factors may be required to establish robust FOXA2 occupancy. It will be interesting to investigate how other TFs (including nonpioneer factors), as well as modulations to cofactor motif sequences at particular loci, affect pioneer-factor occupancy.

The interaction of FOXA2 with repressors may possibly explain the limited gain in DNA accessibility observed at most target sites[7,39,40]. Yet sites that substantially gained accessibility generally also gained modest enrichment of phased and modified histones, potentially indicating recruitment of the ATP-dependent chromatin-remodeling machinery to this subset of target sites. Additionally, FOXA2 may specifically displace linker histone H1 at this subset of regions, which could also increase accessibility[3].

Mechanistic investigations into the loss of methylation observed at some FOXA2 targets uncovered a dependence on DNA replication, because cells arrested in G$_1$ did not dynamically lose DNAme despite occupancy and changes to DNA accessibility. Our study suggests that S-phase binding of FOXA2 may occur rapidly after nascent-strand synthesis and before maintenance methylation. The proximity of CpG dinucleotides to the FOXA2 peak summit is a distinguishing characteristic of targets that become demethylated, supporting a model in which occupancy directly interferes with the DNAme machinery. A recent study has speculated that loss of

DNAme at FOXA1 targets may result from an active demethylation mechanism involving DNA-repair enzymes[41], although that study did not examine DNAme loss in the absence of DNA replication or report the active demethylating enzyme. Therefore, alternative possibilities remain, and more work must be done to clarify the molecular pathway that leads to the removal of this modification. Our data show a clear requirement for DNA replication in the dynamic loss of DNAme, but we cannot rule out the possibility of an enzymatic pathway, such as TET-mediated hydroxymethylation, that either blocks maintenance or marks methylated cytosines for removal following replication. Together, the systematic comparison of endogenous and ectopic TF behaviors reveals relevant mechanistic details and provides more comprehensive understanding of pioneer factors, supporting their rational application toward cellular reprogramming.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41588-017-0034-3.

References

1. Spitz, F. & Furlong, E. E. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* **13**, 613–626 (2012).
2. Cirillo, L. A. et al. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Mol. Cell* **9**, 279–289 (2002).
3. Iwafuchi-Doi, M. et al. The pioneer transcription factor FoxA maintains an accessible nucleosome configuration at enhancers for tissue-specific gene activation. *Mol. Cell* **62**, 79–91 (2016).
4. Iwafuchi-Doi, M. & Zaret, K. S. Pioneer transcription factors in cell reprogramming. *Genes Dev.* **28**, 2679–2692 (2014).
5. Soufi, A., Donahue, G. & Zaret, K. S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* **151**, 994–1004 (2012).
6. Soufi, A. et al. Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* **161**, 555–568 (2015).
7. Zaret, K. S. & Carroll, J. S. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev.* **25**, 2227–2241 (2011).
8. Lupien, M. et al. FoxA1 translates epigenetic signatures into enhancer-driven lineage-specific transcription. *Cell* **132**, 958–970 (2008).
9. Zaret, K. S. & Mango, S. E. Pioneer transcription factors, chromatin dynamics, and cell fate control. *Curr. Opin. Genet. Dev.* **37**, 76–81 (2016).
10. Hurtado, A., Holmes, K. A., Ross-Innes, C. S., Schmidt, D. & Carroll, J. S. FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat. Genet.* **43**, 27–33 (2011).
11. Chen, J. et al. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* **156**, 1274–1285 (2014).
12. Swinstead, E. E. et al. Steroid receptors reprogram FoxA1 occupancy through dynamic chromatin transitions. *Cell* **165**, 593–605 (2016).
13. Liu, Z. & Kraus, W. L. Catalytic-independent functions of PARP-1 determine Sox2 pioneer activity at intractable genomic loci. *Mol. Cell* **65**, 589–603. e9 (2017).
14. Franco, H. L., Nagari, A. & Kraus, W. L. TNFα signaling exposes latent estrogen receptor binding sites to alter the breast cancer cell transcriptome. *Mol. Cell* **58**, 21–34 (2015).
15. Tuteja, G., Jensen, S. T., White, P. & Kaestner, K. H. Cis-regulatory modules in the mammalian liver: composition depends on strength of Foxa2 consensus site. *Nucleic Acids Res.* **36**, 4149–4157 (2008).
16. Li, Z., Schug, J., Tuteja, G., White, P. & Kaestner, K. H. The nucleosome map of the mammalian liver. *Nat. Struct. Mol. Biol.* **18**, 742–746 (2011).
17. Cirillo, L. A. et al. Binding of the winged-helix transcription factor HNF3 to a linker histone site on the nucleosome. *EMBO J.* **17**, 244–254 (1998).
18. Gifford, C. A. et al. Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149–1163 (2013).
19. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
20. Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. Nature **518**, 317–330 (2015).
21. Bailey, T. L. & Machanick, P. Inferring direct DNA binding from ChIP-seq. *Nucleic Acids Res.* **40**, e128 (2012).
22. Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
23. Voss, T. C. et al. Dynamic exchange at regulatory elements during chromatin remodeling underlies assisted loading mechanism. *Cell* **146**, 544–554 (2011).
24. Gualdi, R. et al. Hepatic specification of the gut endoderm in vitro: cell signaling and transcriptional control. *Genes Dev.* **10**, 1670–1682 (1996).
25. Zaret, K. Developmental competence of the gut endoderm: genetic potentiation by GATA and HNF3/fork head proteins. *Dev. Biol.* **209**, 1–10 (1999).
26. Bossard, P. & Zaret, K. S. GATA transcription factors as potentiators of gut endoderm differentiation. *Development* **125**, 4909–4917 (1998).
27. Tsankov, A. M. et al. Transcription factor binding dynamics during human ES cell differentiation. *Nature* **518**, 344–349 (2015).
28. Jozwik, K. M., Chernukhin, I., Serandour, A. A., Nagarajan, S. & Carroll, J. S. FOXA1 directs H3K4 monomethylation at enhancers via recruitment of the methyltransferase MLL3. *Cell Rep.* **17**, 2715–2723 (2016).
29. Sérandour, A. A. et al. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* **21**, 555–565 (2011).
30. Wang, A. et al. Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. *Cell Stem Cell* **16**, 386–399 (2015).
31. Thurman, R. E. et al. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
32. Brinkman, A. B. et al. Sequential ChIP-bisulfite sequencing enables direct genome-scale investigation of chromatin and DNA methylation cross-talk. *Genome Res.* **22**, 1128–1138 (2012).
33. Kohli, R. M. & Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* **502**, 472–479 (2013).
34. Smith, Z. D. & Meissner, A. DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* **14**, 204–220 (2013).
35. Smith, Z. D. & Meissner, A. The simplest explanation: passive DNA demethylation in PGCs. *EMBO J.* **32**, 318–321 (2013).
36. Sakaue-Sawano, A. et al. Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* **132**, 487–498 (2008).
37. Sekiya, T., Muthurajan, U. M., Luger, K., Tulin, A. V. & Zaret, K. S. Nucleosome-binding affinity as a primary determinant of the nuclear mobility of the pioneer transcription factor FoxA. *Genes Dev.* **23**, 804–809 (2009).
38. Chronis, C. et al. Cooperative binding of transcription factors orchestrates reprogramming. *Cell* **168**, 442–459 (2017). e20.
39. Sekiya, T. & Zaret, K. S. Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. *Mol. Cell* **28**, 291–303 (2007).
40. Wang, J.-C. et al. Transducin-like enhancer of split proteins, the human homologs of Drosophila groucho, interact with hepatic nuclear factor 3b. *J. Biol. Chem.* **275**, 18418–18423 (2000).
41. Zhang, Y. et al. Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering. *Nat. Genet.* **48**, 1003–1013 (2016).
42. Stark, R. & Brown, G. DiffBind: differential binding analysis of ChIP–seq peak data. http://bioconductor.org/packages/release/bioc/vignettes/DiffBind/inst/doc/DiffBind.pdf (2011).

Author contributions

J.D. and A.M. conceived and designed the experimental study. J.D. and J.S.C. performed all experiments. S.T. and J.C. performed analysis of next-generation sequencing data. J.D., S.T. and J.C. performed all data interpretation and analysis with guidance from A.M., Z.D.S., and M.J.Z. D.C., C.A.G. and Z.D.S. provided experimental support. Initial data processing and analyses were performed by K.C., R.K., D.R.K. and M.J.Z. with the guidance of A.M. and J.L.R. H.G. and E.K.S. generated WGBS libraries under the guidance of A.G. R.P. assisted with flow cytometry analysis. J.D. and A.M. wrote the manuscript with help from S.T., Z.D.S. and J.C.

Competing interests

The authors declare no competing financial interests.

Additional information

Methods

Cell culture. Clonal FOXA2 doxycycline-inducible cells lines were derived from an immortalized BJ foreskin fibroblast cell line from the ATCC (BJ-5ta; CRL-4001). Cells were cultured in MEM-alpha (Life Technologies 32561-037) with 10% FBS, 1% penicillin–streptomycin, 0.01 mg/mL hygromycin B and 5 ng/mL bFGF. Derived BJ^{FOXA2} lines were grown in the same conditions plus 0.5 µg/mL puromycin.

BJ cell-line generation. Cells were infected with pTRIPZ-FOXA2 and pTRIPZ-RFP at an MOI of ~ 1. After infection, cells were selected with puromycin (1 µg/mL) and replated at a high dilution to ensure separation for clonal expansion and isolation. After 2 weeks of growth, individual clones were picked, expanded and screened. Criteria for inclusion in the current study included uniform expression of FOXA2 and minimum basal FOXA2 expression in uninduced controls. Clones were maintained in 0.5 µg/mL puromycin–containing medium after expansion. To induce FOXA2, doxycycline was added at 0.5 µg/mL.

Cloning and constructs. To generate pTRIPZ-FOXA2, pTRIPZ-RFP and pTRIPZ-FOXA2–CDT1 clones, pTRIPZ inducible lentiviral vector (Thermo Scientific) and full-length *FOXA2* were assembled with Gibson Assembly Master Mix (NEB). pTRIPZ empty vector was digested with XhoI and MluI to remove shRNAmir regulatory sequences, and digested ends were blunted. The linearized pTRIPZ backbone was digested with BsiWI to generate two fragments, each with one sticky end. The fragments were gel extracted, purified and ligated with a Quick Ligation Kit (NEB). Primer sequences are listed in Supplementary Table 1.

To generate the HaloTagged-FOXA2 construct, the full-length FOXA2 sequence was ligated into pFN21A (Promega). GATA4-V5 and POU5F1-V5 constructs were obtained from the Broad Institutes Genomics Perturbations platform and are available for purchase through Thermo Fisher.

Protein purification. 293T cells were transfected with pFN21A-FOXA2. Purification was completed with Promega's Halotag Protein Purification System. Briefly, 48 h after transfection, cells were harvested, lysed and gently sonicated four times with a Branson Sonifier at 10% amplitude for 15 s. Samples were diluted 1:3 with protein-purification buffer (1× PBS, 1 mM DTT and 0.0005% NP-40) and centrifuged to remove debris. Halo-Resin was washed in purification buffer, added to lysates and incubated at 4 °C overnight. After incubation, the resin was washed, and FOXA2 protein was cleaved via the addition of TEV protease during an overnight incubation at 4 °C. Purified protein was assessed via Coomassie blue–stained gels and western blotting.

Electrophoretic mobility shift assay. EMSA was performed with a LightShift Chemiluminescent EMSA kit (Pierce). Purified Halo-tagged-FOXA2 protein (3–6 µg) was mixed with duplexed, biotinlyated probes (20 fmol/µL) without competitor DNA. Unlabeled probes (nonbiotinlyated) were added at 10–100× concentrations of biotinlyated probes. Binding reactions were incubated for 20 min at room temperature before being loaded onto a 6% DNA retention gel (Invitrogen). Complexes were transferred to nylon membranes (Invitrogen) and cross-linked via UV radiation in a Stratalinker. Biotinlyated DNA was detected by chemiluminescence.

Chromatin immunoprecipitation. Cells were cross-linked with 1% formaldehyde for 10 min at room temperature and quenched with 125 mM glycine at room temperature. ChIP was performed as previously described[18] by isolating nuclei and shearing DNA to 200–600 bp fragments with a Branson sonicator. Antibody incubation with chromatin was performed overnight. ~ 10 million cells were used per FOXA2 ChIP with 1 µg of antibody/million cells. ~ 1 million cells were used for each histone ChIP. After an overnight incubation, antibody–protein complexes were isolated with Protein G/A beads (Life Technologies), and sequencing libraries were generated. Libraries were generated as previously described[18,42] and were sequenced on a HiSeq 2500 instrument at 11 pM.

Chromatin immunoprecipitation–bisulfite sequencing. To generate bisulfite-converted DNA libraries after ChIP (described above), we used a Nugen Ovation UltraLow Methyl-seq Kit (0335-0336). Bisulfite conversion was performed with an EpiTect Bisulfite kit (Qiagen) with carrier DNA. Libraries were sequenced on a HiSeq 2500 instrument at 8 pM with 35% PhiX spike-in.

Antibodies. ChIPs were performed with antibodies to the following: FOXA2 (R&D, AF2400), H3K4me1 (Millipore, 17-614), H3K4me2 (Active Motif, 39141), H3K27ac (Active Motif, 39133), H3K27me3 (Active Motif, 39155), V5 (MBL, M167-3) and OCT4 (Active Motif, 39811). Immunostaining was performed with antibodies to the following: FOXA2 (R&D, AF2400), V5 (MBL, M167-3) and OCT4 (Active Motif, 39811). Western blots were performed with antibodies to the following: FOXA2 (R&D, AF2400), V5 (MBL, M167-3), H3 (Abcam, ab1791) and OCT4 (Active Motif, 39811). Complete western blots are shown in Supplementary Fig. 9.

Genome-browser tracks. All browser tracks were created in Illustrator by exporting.svg files from Integrated Genome Viewer (IGV). First, data were imported into IGV as normalized tiled data files (tdf) and scaled to the same values (2 for ChIP–seq and 1 for DNAme) unless otherwise specified. Genomic locations displayed are listed in the figure legends.

Whole-genome bisulfite sequencing. 100–200 ng of genomic DNA was fragmented with a Covaris S2 instrument for 6 min according to the following program: duty cycle, 5%; intensity, 10; cycle per burst, 200. The sheared DNA was purified with a DNA Clean and Concentrator kit from Zymo. Bisulfite conversion of DNA was then conducted with an EZ DNA Methylation-Gold kit (Zymo Research), with elution in 15 µl low-TE buffer. To minimize loss during storage, bisulfite-converted DNA was immediately processed for generating WGBS libraries with an Accel-NGS Methyl-Seq DNA library kit (Swift Biosciences). All protocols were carried out according to the manufacturer's recommendations unless specified. The libraries were sequenced with 100-bp paired-end reads on an Illumina HiSeq 2500 or HiSeq 4000 instrument. 10 bp was clipped from the WGBS raw sequencing reads at either end.

ATAC–sequencing. Tagmentation was performed on whole nuclei at 37 °C for 45 min as previously described[21]. DNA was isolated on PCR MinElute columns (Qiagen), and a small amount of the DNA was amplified for 9, 12 and 15 cycles to determine the optimal cycling conditions. The rest of the DNA was then amplified with the chosen number of cycles, and PCR libraries were purified through double-sided Ampure cleanup to remove high-molecular-weight fragments. An 0.55× Ampure volume was added to the reactions, mixed and incubated. Supernatant was removed after magnet separation and cleaned up with a 1× Ampure volume. Libraries were sequenced on a HiSeq 2500 instrument at 8 pM.

RNA sequencing. RNA was isolated with RNeasy columns (Qiagen), and nonstranded libraries were generated with Illumina's standard Tru-Seq kit. Libraries were sequenced on a HiSeq 2500 instrument at 11 pM.

RT–qPCR. cDNA synthesis was performed with 600–2,000 ng of RNA with a RevertAid First Strand cDNA Synthesis Kit (Thermo Scientific) with oligo(dT)18 primer. Quantitative PCR (qPCR) primers were designed with Primer-BLAST (NCBI). Primers were designed to span an exon–exon junction, amplify 70–200 bp of cDNA and amplify all isoforms of a transcript. qPCR was performed with three or four technical replicates with a 1:100 or 1:1,000 dilution of cDNA, Power SYBR Green Master Mix (Applied Biosystems) and 500 nM of forward and reverse primers in a ViiATM 7 Real-Time PCR System (Applied Biosystems). *ACTB* and *HPRT1* were used as endogenous controls. Relative gene expression was calculated with the comparative Ct (ΔΔCt) method in ExpressionSuite Software v1.0.3 (Biosystems).

Western blotting. Nuclear proteins were extracted in standard RIPA buffer supplemented with protease inhibitors (Roche). Equal amounts of extracts were mixed with LDS (Life Technologies) and BME and boiled at 95 °C for 5 min. Samples were loaded onto a NuPage Novex 4–12% Bis-Tris gel (Life Technologies) and electrophoresed for 1 h at 200 V in 1× MES buffer (Life Technologies). Proteins were transferred to PVDF membranes with an iBlot transfer system (Life Technologies). Membranes were blocked in 5% milk/TBST for 1 h at room temperature, then incubated with primary antibodies in 5% milk/TBST overnight at 4 °C. Primary anti-FOXA2 was diluted at 1:5,000, and primary anti-H3 was diluted at 1:10,000. Membranes were washed and incubated in secondary antibodies in TBST at 1:10,000 dilution. Detection was performed with SuperSignal West Dura Chemiluminescent Substrate (Thermo Scientific).

Immunostaining. Cells were fixed in 4% formaldehyde for 15 min at room temperature. After cells were washed, permeabilization and blocking were performed with 4% FBS/0.4% Triton X-100 in PBS for 1 h at room temperature. Primary-antibody staining was performed with 2% FBS/0.2% Triton X-100 in PBS overnight at 4 °C. Secondary staining was performed with fluorophore-conjugated antibodies in PBS for 1 h at room temperature.

Cell-cycle arrest and FACS analysis. Cells were halted in G$_1$ through the addition of 500 mM mimosine (Sigma) treatment overnight. Cell proliferation was determined with a Click-iT Plus EdU Flow Cytometry Assay Kit (Life Technologies). 5 µM EdU was added to the culture medium, and samples were incubated for 18 h. Samples were then fixed, permeabilized and treated with Click-iT EdU reaction cocktail according to the kit instructions. Hoechst and/or Vybrant Dye (Life Technologies) were diluted 1:1,000 to measure DNA content. FACS analysis was performed on a BD LSR II flow cytometry machine.

ChIP–seq analysis. All FOXA2 ChIP–seq datasets from different conditions and cell types were aligned by Bowtie 2 (ref. [43]) to the hg19 human genome reference assembly by using default parameters. Duplicate reads were removed with Picard (http://broadinstitute.github.io/picard/). Genome-browser images were created by converting bam files into tdfs by using IGV tools[44] by normalizing them to 1 million reads. All datasets were subjected to the IDR framework[45] with 0.1 as cutoff, in combination with using MACS2 (ref. [46]) for calling peaks in each

replicate separately. For MACS2 peak calling, we used corresponding whole cell extract (WCE) as a background control and a *P*-value cutoff of 0.01. This initial peak calling using IDR and MACS2 resulted in a set of peaks that were above background for each cell type. As an additional filtering method and to facilitate comparison of peaks from different cell types and conditions, we developed an in-house computational framework to redefine relative peak positions and to standardize peak width. We merged IDR-called peaks from all cell types if they overlapped by at least 20%, while keeping track of the summits of the peaks being merged. This procedure resulted in a master peak set encompassing all FOXA2 datasets. Because several peaks with different peak summits were merged, we devised a simple weighted framework to define new peak summits. To assign a new peak summit, we used the peak height as a measure of weighed distance from the peak center. Using this weighted measure of peak height, we calculated a new peak summit, which was closest to the highest peak that was merged but also represented contributions from smaller peaks in a distance-dependent manner. All peaks were assigned new peak summits by using this formula. To define new peak widths, we extended each peak by 300 bp in both directions from the peak summit to have all peaks of 600 bp. Enrichment of different histone marks at these FOXA2 peaks was calculated with the standard RPKM formula.

Composite plots. Composite plots showing enrichment of different histone marks at FOXA2 peaks were made in the HOMER package[47]. As described in the HOMER documentation, we first created tag directories for each sample or histone mark that we wished to plot around peak regions. Peaks were extended by 2,000 bp in each direction, and tag directories were then used to create a matrix with tag densities at each nucleotide and each individual replicate was normalized for its respective sequencing depth. Matrix files with the tag density at each position within an extended 4,000-bp window were imported into R to create the plots.

Read-density heat maps. Read-density heat maps were created with the EnrichedHeatmap (https://github.com/jokergoo/EnrichedHeatmap/) and ComplexHeatmap (https://github.com/jokergoo/ComplexHeatmap/) packages. We first determined the genome-wide coverage of each sample or histone mark by using coverageBed from the BEDTools package[49]. These coverage files and the peak regions to be plotted were supplied as input to ComplexHeatmap. The heat in each heat map was decided on the basis of the percentile range by capping the maximum at the ninety-ninth percentile to remove outliers.

Differential motif analysis. Differential motif analysis was performed in HOMER[47]. To calculate differential enrichment between two sets of peaks, we used one set as background and then interchanged to calculate for another set. The motifs were scanned in 200-bp regions around the peak center in both directions.

Epigenetic-state maps. To classify FOXA2-bound regions in different chromatin states, we used a hierarchical classification system. First, all FOXA2 peaks that had ATAC–seq (RPKM >3) enrichment were classified as accessible. Next, peaks that had either H3K27ac or H3K4me1 (RPKM >3) were marked as 'active', and regions with H3K27me3 or H3K9me3 (RPKM >3) were classified as 'repressed'. After classifying all histone modifications, we divided the rest of the regions on the

basis of their DNAme levels. Regions with DNAme levels below 20% were marked as lowly methylated regions (LMR), regions with methylation levels between 20% and 60% were called intermediately methylated regions (IMR), and regions with methylation levels above 60% were termed highly methylated regions (HMR).

ChIP–BS-seq analysis. For the analysis of methylation changes associated with FOXA2 binding, we redefined binding sites to maximize overlap with our ChIP–BS-seq data, because bisulfite conversion on small amounts of input DNA results in material loss. To do so, we combined FOXA2 ChIP–seq data from BJ^{FOXA2} cells 4 d and 10 d after induction. The summit of each peak was determined in MACS, the region 200 bp to either side was then selected, and overlapping regions were merged to generate a list of 113,398 sites. We then intersected these regions with BJ-fibroblast WGBS and 4-d BJ^{FOXA2} ChIP–BS-seq datasets and selected only CpGs covered by at least three reads in both samples, thus yielding a total of 42,086 sites and 135,785 CpGs. We used these same regions to select matched CpGs covered at ≥3× in the ChIP–BS-seq data from the BJ^{FOXA2} mimosine-treated and released samples ($n = 13,494$ sites and 18,429 CpGs). For analysis of BJ$^{FOXA2–CDT1}$, we considered only regions that were 10× covered. We compared individual CpG and mean FOXA2-binding-site methylation, generated CpG counts and coverage plots, and calculated the distances between the summit and the nearest CpG by using custom R scripts. For comparison to the ATAC–seq data (described above), we used HOMER[47] to generate enrichment composite plots for 2 kb to either side of the peaks.

Statistical methods. All *P*-value calculations were done with two-sided Wilcoxon rank-sum tests unless otherwise stated.

Life Sciences Reporting Summary. Further information on experimental design is available in the Life Sciences Reporting Summary.

Data availability. All data have been deposited in the Gene Expression Omnibus database under accession number GSE90456.

References

43. Mikkelsen, T. S. et al. Comparative epigenomic analysis of murine and human adipogenesis. *Cell* **143**, 156–169 (2010).
44. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359 (2012).
45. Robinson, J. T. et al. Integrative genomics viewer. *Nat. Biotechnol.* **29**, 24–26 (2011).
46. Landt, S. G. et al. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Res.* **22**, 1813–1831 (2012).
47. Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
48. 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).
49. Quinlan, A. R. BEDTools: The Swiss-Army Tool for Genome FeatureAnalysis. *Curr. Protoc. Bioinformatics* **47**, 11–12 (2014).

nature research

Corresponding author(s): Alexander Meissner

☐ Initial submission   ☐ Revised version   ☒ Final submission

Life Sciences Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form is intended for publication with all accepted life science papers and provides structure for consistency and transparency in reporting. Every life science submission will use this form; some list items might not apply to an individual manuscript, but all fields must be completed for clarity.

For further information on the points included in this form, see Reporting Life Sciences Research. For further information on Nature Research policies, including our data availability policy, see Authors & Referees and the Editorial Policy Checklist.

▶ Experimental design

1. Sample size

Describe how sample size was determined.

> We did not perform any patient or in-vivo analysis as all data was generated using cell-culture standards. Therefore, we most used three biological replicates most of the data that we collected.

2. Data exclusions

Describe any data exclusions.

> All data are provided. In some analysis we excluded outliers and have indicated this in each panel.

3. Replication

Describe whether the experimental findings were reliably reproduced.

> Yes, they were.

4. Randomization

Describe how samples/organisms/participants were allocated into experimental groups.

> Not applicable

5. Blinding

Describe whether the investigators were blinded to group allocation during data collection and/or analysis.

> Not applicable

Note: all studies involving animals and/or human research participants must disclose whether blinding and randomization were used.

6. Statistical parameters

For all figures and tables that use statistical methods, confirm that the following items are present in relevant figure legends (or in the Methods section if additional space is needed).

n/a	Confirmed	
☐	☒	The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement (animals, litters, cultures, etc.)
☐	☒	A description of how samples were collected, noting whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	A statement indicating how many times each experiment was replicated
☐	☒	The statistical test(s) used and whether they are one- or two-sided (note: only common tests should be described solely by name; more complex techniques should be described in the Methods section)
☐	☒	A description of any assumptions or corrections, such as an adjustment for multiple comparisons
☐	☒	The test results (e.g. P values) given as exact values whenever possible and with confidence intervals noted
☐	☒	A clear description of statistics including central tendency (e.g. median, mean) and variation (e.g. standard deviation, interquartile range)
☐	☒	Clearly defined error bars

See the web collection on statistics for biologists for further resources and guidance.

▶ Software

Policy information about availability of computer code

7. Software

Describe the software used to analyze the data in this study.

> TopHat, Bowtie, Homer, MEME, R packages including ComplexHatmap, EnrichedHeatmap, DESeq

For manuscripts utilizing custom algorithms or software that are central to the paper but not yet described in the published literature, software must be made available to editors and reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). *Nature Methods* guidance for providing algorithms and software for publication provides further information on this topic.

▶ Materials and reagents

Policy information about availability of materials

8. Materials availability

Indicate whether there are restrictions on availability of unique materials or if these materials are only available for distribution by a for-profit company.

> Not applicable

9. Antibodies

Describe the antibodies used and how they were validated for use in the system under study (i.e. assay and species).

> Described in Methods section of the manuscript

10. Eukaryotic cell lines

a. State the source of each eukaryotic cell line used.

> BJ-5ta

b. Describe the method of cell line authentication used.

> Purchased from ATCC

c. Report whether the cell lines were tested for mycoplasma contamination.

> Yes

d. If any of the cell lines used are listed in the database of commonly misidentified cell lines maintained by ICLAC, provide a scientific rationale for their use.

> Not applicable

▶ Animals and human research participants

Policy information about studies involving animals; when reporting animal research, follow the ARRIVE guidelines

11. Description of research animals

Provide details on animals and/or animal-derived materials used in the study.

> Not applicable

Policy information about studies involving human research participants

12. Description of human research participants

Describe the covariate-relevant population characteristics of the human research participants.

> Not applicable