nature genetics

Nucleation of DNA repair factors by FOXA1 links DNA demethylation to transcriptional pioneering

Yu Zhang[1,3], Di Zhang[1,3], Qian Li[1], Jing Liang[1], Luyang Sun[1], Xia Yi[1], Zhe Chen[1], Ruorong Yan[1], Guojia Xie[1], Wanjin Li[1], Shumeng Liu[1], Bosen Xu[1], Lei Li[1], Jianguo Yang[1], Lin He[1] & Yongfeng Shang[1,2]

FOXA1 functions in epigenetic reprogramming and is described as a 'pioneer factor'. However, exactly how FOXA1 achieves these remarkable biological functions is not fully understood. Here we report that FOXA1 associates with DNA repair complexes and is required for genomic targeting of DNA polymerase β (POLB) in human cells. Genome-wide DNA methylomes demonstrate that the FOXA1 DNA repair complex is functionally linked to DNA demethylation in a lineage-specific fashion. Depletion of FOXA1 results in localized reestablishment of methylation in a large portion of FOXA1-bound regions, and the regions with the most consistent hypermethylation exhibit the greatest loss of POLB and are represented by active promoters and enhancers. Consistently, overexpression of FOXA1 commits its binding sites to active DNA demethylation in a POLB-dependent manner. Finally, FOXA1-associated DNA demethylation is tightly coupled with estrogen receptor genomic targeting and estrogen responsiveness. Together, these results link FOXA1-associated DNA demethylation to transcriptional pioneering by FOXA1.

© 2016 Nature America, Inc. All rights reserved.

FOXA1, a member of the forkhead family of winged-helix transcription factors, is implicated in the formation of endodermal tissues and reproductive organogenesis, including that of liver, pancreas, lung, prostate, and mammary gland[1–3]. It has been proposed that the members of the FOXA family occupy distal regulatory enhancers to establish chromatin competency for subsequent recruitment of collaborating transcription factors, instead of promoting immediate transcriptional activation[4–6]. This exceptional property has led to the proposal that FOXA family members act as pioneer transcription factors[4,7]. This paradigm is exemplified by a panel of hepatic nuclear factors and has been extended to transcriptional cooperation between different transcription factors such as FOXA1 and sex hormone receptors, including estrogen and androgen receptors (ER and AR)[8,9]. Consistently, FOXA1 has been shown to act as a reprogramming factor, either facilitating or restricting AR binding to structurally and functionally distinct classes of enhancers in prostate cancer cells[10] and driving the evolution of prostate cancer from an androgen-dependent to an androgen-independent state[11]. In addition, an enhanceosome composed of ER, FOXA1, and GATA3 was found to be sufficient to restore estrogen responsiveness in ER^- breast cancer cells[12].

The exact molecular mechanisms underlying these extraordinary functions of FOXA1 are still not fully understood. The concept that FOXA1 could open a compacted nucleosome when targeted to silent chromatin was based on, at least partially, the structural similarity between the FOXA1 forkhead domain and histone linker H1 (refs. 13–15). Moreover, investigations on lineage-specific occupancy of FOXA1 in discrete genomic regions highlighted the importance of the histone modification marks mono- and dimethylation of histone H3 at lysine 4 (H3K4me1 and H3K4me2) in demarcating FOXA1 regulatory regions[8]. Intriguingly, a genome-wide DNA methylation profile showed that the recruitment of FOXA1 to enhancers is favored in hypomethylated regions[16], and unbiased screening for proteins binding methylated DNA suggested preferential binding of FOXA1 to methylated CpG dinucleotides[17], hinting at a potential role for forkhead proteins in the regulation of DNA methylation dynamics.

Cytosine methylation (resulting in 5-methylcytosine, or 5mC) is a key covalent chemical modification of DNA with the potential to modulate transmissible chromatin architecture and function across the genome[18]. Although DNA methylation was initially believed to confer stable, long-term silencing, it is now clear that this modification is also reversible[19,20]. However, in comparison to the regulation of histone methylation, the mechanisms controlling the plasticity of DNA methylation are more complicated, as active DNA demethylation by direct removal of a methyl group is not likely to occur owing to the thermally unfavorability of breaking the carbon–carbon bond. Instead, evidence from *Arabidopsis thaliana* supports an indirect route in which the DEMETER family of DNA glycosylases recognizes and excises 5mC followed by base-excision repair (BER) to complete active demethylation[21]. Although a counterpart for DEMETER glycosylases in mammals remains elusive, the characterization of ten–eleven translocation (TET) family proteins as 5mC oxidases, which are able to hydroxylate 5mC to 5-hydroxylmethylcytosine and its derivatives

[1]Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education), Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Peking University Health Science Center, Beijing, China. [2]Key Laboratory of Hormones and Development (Ministry of Health), Tianjin Key Laboratory of Medical Epigenetics, Department of Biochemistry and Molecular Biology, School of Basic Medical Sciences, Tianjin Medical University, Tianjin, China. [3]These authors contributed equally to this work. Correspondence should be addressed to Y.S. (yshang@hsc.pku.edu.cn).

Received 12 March; accepted 1 July; published online 8 August 2016; doi:10.1038/ng.3635

5-formylcytosine and 5-carboxylcytosine, provides major insight into active DNA demethylation in mammals[19]. Another alternative pathway of DNA demethylation involves deamination of 5mC to thymine, achieved by activation-induced deaminase (AID) or apolipoprotein B mRNA-editing enzyme, catalytic polypeptide (APOBEC) family proteins[22,23]. Chemically, 5mC editing by either oxidation or deamination generates substrates that are more efficiently modified by DNA glycosylation[24,25]. Regardless of the ways by which cytosine is modified, there is evidence supporting the notion that the DNA repair pathway is an integral component of DNA demethylation machinery[23,26,27]. In effect, DNA glycosylases/lysases recognize damaged or mismatched base pairs and cleave the N-glycosylic bond between the target base and the sugar–phosphate backbone of DNA to produce apurinic or apyrimidinic (AP) sites[28,29]. These sites are then resolved by the BER pathway through successive reactions catalyzed by AP endonuclease, POLB, and DNA ligase III (LIG3) to replenish the AP site[30]. Therefore, active DNA demethylation requires the enlistment of DNA repair factors, especially the BER complex, to replace methylated cytosine.

We sought to explore in this study the mechanistic role of FOXA1 in transcription pioneering and epigenetic reprogramming. We found that the FOXA1 interactome includes a collection of proteins that are involved in the double- and single-strand DNA break repair (DSBR and SSBR) pathways. Genome mapping and functional studies indicate that the FOXA1-associated DNA repair complex is implicated in active DNA demethylation.

RESULTS

FOXA1 is associated with DNA repair factors

To gain molecular insights into the epigenetic reprogramming and pioneering functions of FOXA1, we used epitope-based proteomic screening combined with immunopurification and mass spectrometry to identify proteins that are associated with FOXA1. In these experiments, FOXA1-containing protein complexes were affinity purified from nuclear extracts of HeLa cells stably expressing FLAG-FOXA1 with an antibody to FLAG that was immobilized on agarose beads. The purified protein complexes were resolved by SDS–PAGE and silver stained. The experiment was repeated three times with different washing strengths, allowing capture of the purified proteins with high reproducibility. Subsequent mass spectrometry identified a number of proteins that were associated with FOXA1 (**Fig. 1a** and **Table 1**). These included ATP-dependent DNA helicase 2 subunits 1 and 2 (Ku80 and Ku70), the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), and DNA ligase IV (LIG4), all of which are components of the DSBR complex[31–33], and LIG3, POLB, and X-ray repair cross-complementing protein 1 (XRCC1), all of which are subunits of the SSBR complex[30,34,35], as well as poly(ADP-ribose) polymerase 1 (PARP1), a factor implicated in both the DSBR and SSBR complexes[36]. In addition, topoisomerase IIβ (TOPIIβ) was also identified as a FOXA1-associated protein. The detailed results from mass spectrometry are provided in **Supplementary Table 1**. The presence of DSBR and SSBR complex components in the FOXA1-contaning complex was confirmed by immunoblotting of the iterative eluates from FLAG M2 resin with antibodies against these DNA repair proteins (**Fig. 1b** and **Supplementary Fig. 1**).

Characterization of the endogenous FOXA1 DNA repair complex

We then validated the association of FOXA1 with DSBR and SSBR factors through traditional coimmunoprecipitation experiments performed in human mammary carcinoma MCF-7 cells and prostate carcinoma LNCaP cells. Immunoprecipitation with antibody against FOXA1 followed by immunoblotting with antibody against Ku80, Ku70, PARP1, XRCC1, LIG3, or POLB demonstrated that all these proteins were indeed efficiently coimmunoprecipitated with FOXA1 (**Fig. 2a**). Coimmunoprecipitation with antibody against the SSBR protein XRCC1 showed that XRCC1 was not only coimmunoprecipitated with the SSBR proteins LIG3 and POLB but also was also coimmunoprecipitated with the DSBR proteins Ku80 and Ku70, as well as PARP1 and FOXA1 (**Fig. 2a**). However, although LIG4 and DNA-PKcs were identified in the immunopurification experiments in HeLa cells, these two proteins were not coimmunoprecipitated with FOXA1 or XRCC1 in MCF-7 and LNCaP cells, whereas they efficiently coimmunoprecipitated with Ku80 in these cells (**Fig. 2a**). Whether the discrepancies correspond to the existence of cell-lineage-specific FOXA1-associated complexes *in vivo* or reflect the regulatory (rather than constitutive) nature of LIG4 and DNA-PKcs in the DSBR

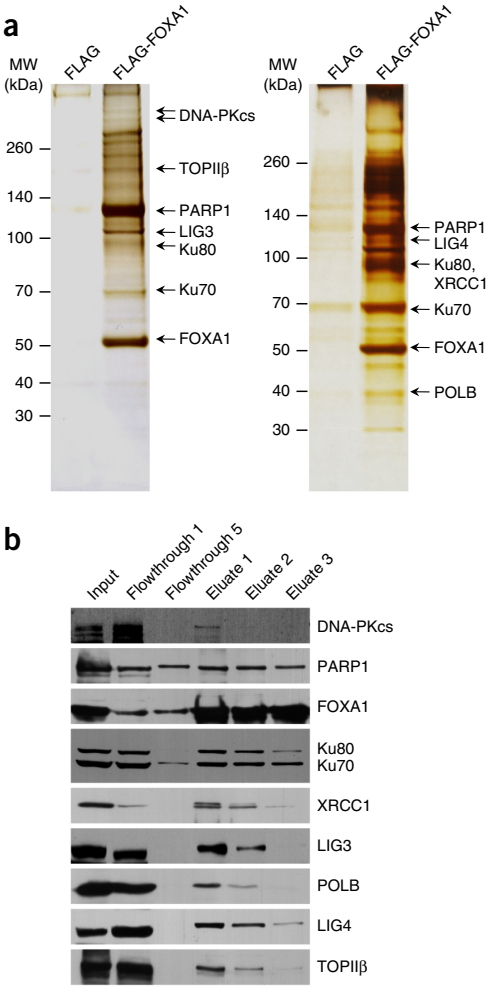


Figure 1 Identification of FOXA1-associated proteins. (**a**) Immunoaffinity purification of FOXA1-containing protein complexes. Cellular extracts from HeLa cells stably expressing FLAG (control) or FLAG-FOXA1 were immunopurified with FLAG M2 resin and eluted with FLAG peptide. Eluted materials were resolved by SDS–PAGE and silver stained. Protein bands were retrieved and analyzed by mass spectrometry. Detailed results from the mass spectrometry analysis are provided in **Supplementary Table 1**. The experiment was repeated three times; results are shown for two experiments. MW, molecular weight. (**b**) Immunoblot analysis of the purified fractions using antibodies against the indicated proteins.

© 2016 Nature America, Inc. All rights reserved.

Table 1 List of FOXA1-associated proteins identified by mass spectrometry

Protein symbol	Hits	Coverage (%) Mass	Coverage (%) Sequence	Approximate MW (kDa)	Activity	Biological process
DNA-PKcs	69	27.3	27.3	300–500	Ser/Thr kinase	DSBR
PARP1	22	38.2	38.1	100–130	Poly(ADP-ribose) polymerase	DSBR, SSBR
TOPIIβ	3	2.6	2.5	90–130	DNA topoisomerase	Topoisomerase
LIG3	5	7.5	7.2	90–130	DNA ligase	SSBR
LIG4	3	6.3	6.4	90–130	DNA ligase	DSBR
Ku80	39	57.7	56.8	80–100	DNA helicase	DSBR
Ku70	27	66.8	66.5	70–90	DNA helicase	DSBR
XRCC1	4	11.8	12.0	70–90	Scaffold of SSBR	SSBR
POLB	3	14.0	13.7	30–50	DNA polymerase	SSBR

complex is currently unknown. Nevertheless, these experiments support our observation that FOXA1 is associated with the components of the DSBR and SSBR complexes.

We next analyzed the co-purification pattern of the endogenous FOXA1 DNA repair complex by size-exclusion chromatography. For this purpose, nuclear extracts from MCF-7 cells were applied to gel filtration columns and separated by fast protein liquid chromatography (FPLC). Immunoblotting analysis of the chromatography fractions showed that FOXA1 under native conditions eluted with an apparent molecular mass much greater than that of the monomeric protein. Notably, the chromatography profile of FOXA1 largely overlapped with those of the DSBR proteins Ku80 and Ku70, the SSBR proteins XRCC1 and LIG3, and PARP1, and partially overlapped with the profile for POLB (**Fig. 2b**). The chromatography profile of DNA-PKcs appeared not to overlap with the profile of FOXA1 in MCF-7 cells.

To further investigate the endogenous FOXA1 DNA repair complex, gel filtration chromatography was performed with MCF-7 nuclear proteins captured by FOXA1-specific antibody as the starting material, allowing a more definitive estimation of the molecular composition of the FOXA1-associated protein complex. Consistent with the results of size-exclusion chromatography of total nuclear extracts, gel filtration chromatography of the purified FOXA1 complex demonstrated that Ku80, Ku70, PARP1, XRCC1, LIG3, and POLB were all associated with FOXA1, whereas LIG4, DNA-PKcs, and TOPIIβ were not identified in the protein complex purified with antibody to FOXA1 (**Fig. 2c**).

As both FOXA1 and DNA repair factors act in the context of chromatin DNA, we then analyzed the binding of the FOXA1 DNA repair

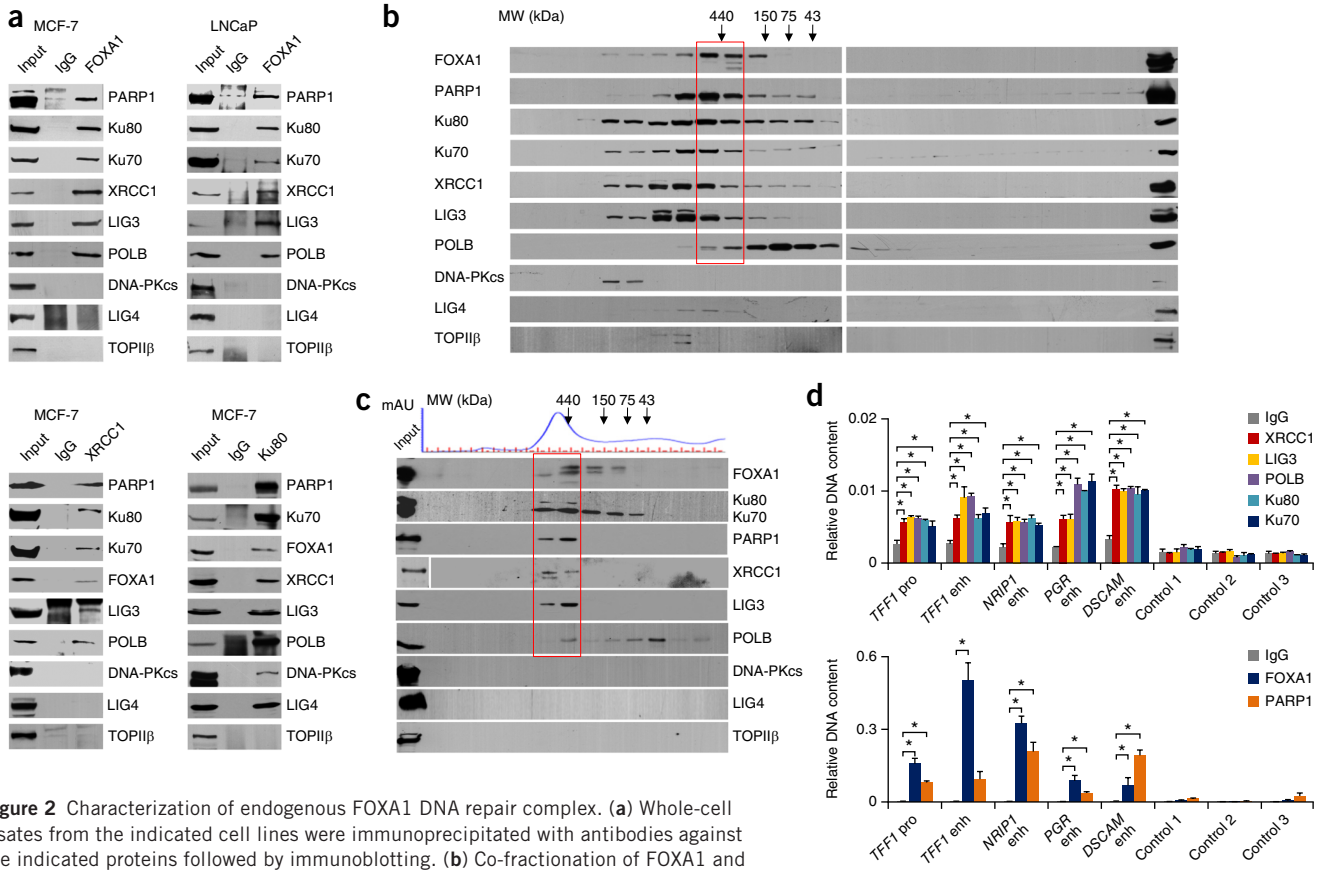


Figure 2 Characterization of endogenous FOXA1 DNA repair complex. (**a**) Whole-cell lysates from the indicated cell lines were immunoprecipitated with antibodies against the indicated proteins followed by immunoblotting. (**b**) Co-fractionation of FOXA1 and DNA repair factors by FPLC. Nuclear extracts from MCF-7 cells were fractionated on Superdex 200 size-exclusion columns. The chromatography elution profiles and immunoblotting analysis of the chromatography fractions are shown. The elution positions of calibration proteins with known molecular masses are indicated at the top. (**c**) Co-fractionation of the endogenous FOXA1 DNA repair complex by FPLC. MCF-7 nuclear proteins captured by antibody specific to FOXA1 were used as the starting material. The homogeneity of the FOXA1 complex and the elution positions of calibration proteins with known molecular masses are indicated at the top. The red boxes in **b** and **c** highlight the overlapping fractions for components of the FOXA1 DNA repair complex. mAU, milli absorption units. (**d**) Co-occupancy of FOXA1-binding sites by components of the FOXA1 DNA repair complex. qChIP assays were performed to determine the binding pattern of individual components of the FOXA1 complex using antibodies against the indicated proteins. Data are displayed as means ± s.d. for triplicate experiments (*$P < 0.01$, one-way ANOVA). pro, promoter; enh, enhancer. Control 1, a region ~5 kb upstream of the *TFF1* transcriptional start site; control 2, a coding region of *GREB1*; control 3, a coding region of *NRIP1*.

© 2016 Nature America, Inc. All rights reserved.

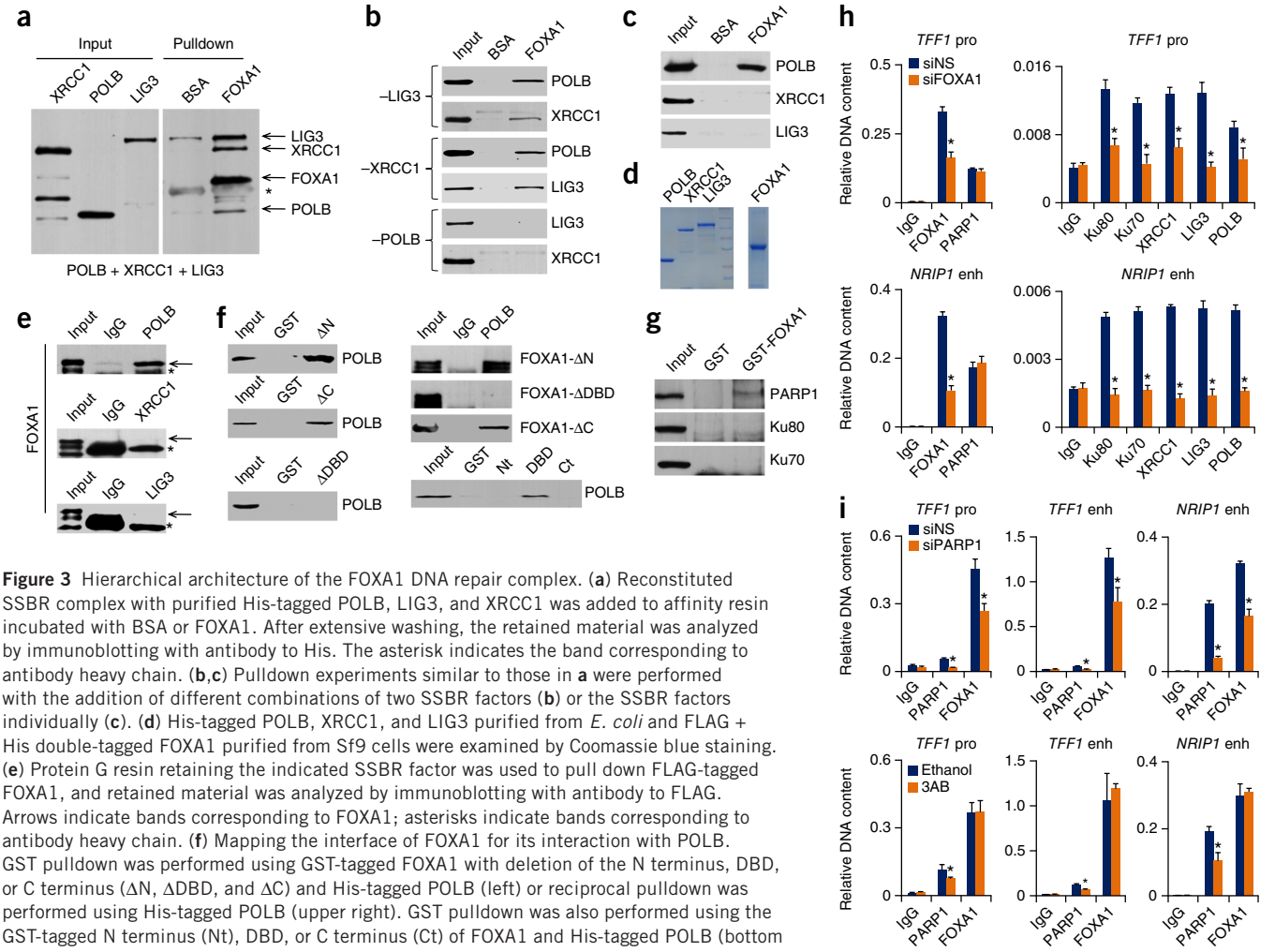


Figure 3 Hierarchical architecture of the FOXA1 DNA repair complex. (**a**) Reconstituted SSBR complex with purified His-tagged POLB, LIG3, and XRCC1 was added to affinity resin incubated with BSA or FOXA1. After extensive washing, the retained material was analyzed by immunoblotting with antibody to His. The asterisk indicates the band corresponding to antibody heavy chain. (**b**,**c**) Pulldown experiments similar to those in **a** were performed with the addition of different combinations of two SSBR factors (**b**) or the SSBR factors individually (**c**). (**d**) His-tagged POLB, XRCC1, and LIG3 purified from *E. coli* and FLAG + His double-tagged FOXA1 purified from Sf9 cells were examined by Coomassie blue staining. (**e**) Protein G resin retaining the indicated SSBR factor was used to pull down FLAG-tagged FOXA1, and retained material was analyzed by immunoblotting with antibody to FLAG. Arrows indicate bands corresponding to FOXA1; asterisks indicate bands corresponding to antibody heavy chain. (**f**) Mapping the interface of FOXA1 for its interaction with POLB. GST pulldown was performed using GST-tagged FOXA1 with deletion of the N terminus, DBD, or C terminus (ΔN, ΔDBD, and ΔC) and His-tagged POLB (left) or reciprocal pulldown was performed using His-tagged POLB (upper right). GST pulldown was also performed using the GST-tagged N terminus (Nt), DBD, or C terminus (Ct) of FOXA1 and His-tagged POLB (bottom right). (**g**) GST pulldown was performed with GST-FOXA1 and DSBR proteins that were transcribed and translated *in vitro*. (**h**) The binding pattern of the indicated components in MCF-7 cells transfected with nonspecific control siRNA (siNS) or siRNA targeting *FOXA1* (siFOXA1) was examined by qChIP. (**i**) PARP1 acts upstream of FOXA1. qChIP was performed to examine the binding of PARP1 and FOXA1 after knockdown of *PARP1* (siPARP1; top) or treatment with the PARP1 inhibitor 3AB (bottom) in MCF-7 cells. Data in **h** and **i** are displayed as means ± s.d. for triplicate experiments (* $P < 0.05$, *t* test). Relative DNA content enriched with specific antibodies was normalized to input DNA.

© 2016 Nature America, Inc. All rights reserved.

complex at the ER target genes *TFF1* (*pS2*), *NRIP1*, *PGR*, and *DSCAM* using a chromatin immunoprecipitation assay coupled with quantitative real-time PCR (qChIP). In support of the finding that FOXA1 and subsets of DNA repair factors were physically associated, FOXA1, Ku80, Ku70, PARP1, XRCC1, LIG3, and POLB were all detected on the regulatory regions of these genes in MCF-7 cells, whereas none of these proteins were detected on control regions (**Fig. 2d**).

Hierarchical architecture of the FOXA1 DNA repair complex

To investigate the molecular details involved in the interaction of FOXA1 with DSBR and SSBR proteins, we performed *in vitro* reconstitution and glutathione *S*-transferase (GST) pulldown experiments with FOXA1 purified from *Spodoptera frugiperda* (Sf9) cells and POLB, LIG3, and XRCC1 purified from *Escherichia coli*. POLB, LIG3, and XRCC1 were first incubated together to form the SSBR complex *in vitro*[34] and were then applied to FLAG-FOXA1 conjugated to FLAG M2 resin. After extensive washing, the resulting materials were analyzed by immunoblotting. The results showed that the SSBR complex was able to bind FOXA1 (**Fig. 3a**). Reconstitution assays in the individual absence of POLB, LIG3, or XRCC1 further showed that POLB was indispensable for association of the SSBR complex with FOXA1 *in vitro* (**Fig. 3b**), suggesting that FOXA1 interacts with this complex through POLB. This notion was strengthened by reciprocal *in vitro* pulldown experiments that showed a direct interaction between FOXA1 and POLB (**Fig. 3c–e**). In addition, GST pulldown assays with GST-tagged FOXA1 mutants in which the N terminus, DNA-binding domain (DBD), or C terminus was deleted (ΔN, ΔDBD, and ΔC, respectively) and bacterially purified POLB indicated that the DBD domain, but not the N terminus or C terminus, was required for the interaction of FOXA1 with POLB (**Fig. 3f** and **Supplementary Fig. 2a**). Consistently, GST pulldown assays with the FOXA1 N terminus, DBD, or C terminus and POLB supported the idea that FOXA1 directly interacts with POLB through its DBD domain (**Fig. 3f**). Although we were unable to perform *in vitro* reconstitution assays for DSBR proteins because of difficulties in purifying them, GST pulldown assays with bacterially purified GST-FOXA1 and PARP1, Ku80, or Ku70 protein that was transcribed and translated *in vitro* showed that FOXA1 could interact directly with PARP1 but not with Ku80 or Ku70 (**Fig. 3g**).

As FOXA1 is the only component in the FOXA1 DNA repair complex with a specific DNA-binding domain, it is reasonable to hypothesize that the chromatin targeting of this complex is dependent on FOXA1. Indeed, binding of Ku80, Ku70, XRCC1, LIG3, and POLB was significantly reduced at FOXA1 target sites upon FOXA1 knockdown by small interfering RNA (siRNA), as determined by qChIP assays

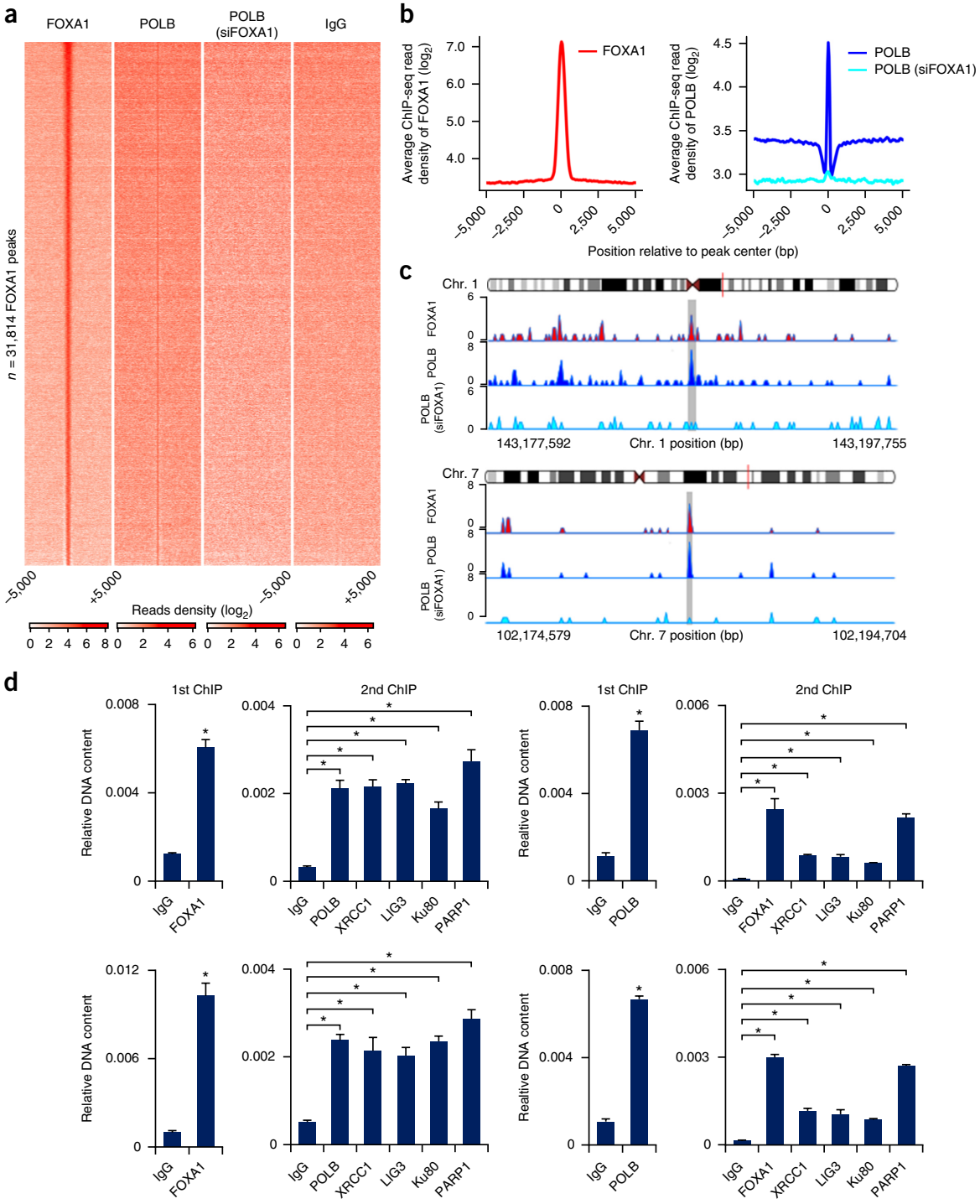


Figure 4 Genomic landscape of the FOXA1 DNA repair complex. (**a**) Heat map showing ChIP-seq read density (log_2 transformed) for the indicated proteins at equivalent genomic regions in 10-kb windows centered on FOXA1-binding sites. The 31,814 regions are sorted by density of FOXA1 signal. The "POLB (siFOXA1)" data set was generated by POLB ChIP-seq in MCF-7 cells transfected with siRNA against *FOXA1*. (**b**) Average ChIP-seq read density profiles (log_2 transformed) for FOXA1 (left) and POLB in the presence or absence of FOXA1 (right). (**c**) ChIP-seq density tracks for two loci with FOXA1-binding sites from chromosomes 1 and 7. The ideograms show the genomic position of the inspected sites. The shaded regions correspond to the called peaks (200 bp) whose sequences served as PCR template for the results in **d**. (**d**) Sequential ChIP (ChIP/re-ChIP) analysis was performed to test the co-binding of subunits of the FOXA1 DNA repair complex at the loci on chromosomes 1 (top) and 7 (bottom) illustrated in **c**. Data are displayed as means ± s.d. for triplicate experiments (*$P < 0.01$, one-way ANOVA).

© 2016 Nature America, Inc. All rights reserved.

(**Fig. 3h**). Intriguingly, however, PARP1 binding was unchanged, suggesting that PARP1 might act upstream of FOXA1. We thus performed qChIP in PARP1-depleted MCF-7 cells and found that knockdown of PARP1 resulted in loss of FOXA1 binding (**Fig. 3i**), whereas it had no effect on FOXA1 expression levels (**Supplementary Figs. 2** and **3**). Moreover, qChIP assays in MCF-7 cells indicated that binding of FOXA1 was not altered by inhibition of the poly(ADP-ribose) polymerase function of PARP1 through addition of the inhibitor 3-aminobenzamide (3AB)[36–38] (**Fig. 3i**), suggesting that the association between FOXA1 and PARP1 is independent of the enzymatic activity of PARP1.

Genomic landscape of the FOXA1 DNA repair complex

Considering the critical function of POLB in the BER pathway and also its direct interaction with FOXA1 in the complex, we analyzed FOXA1 and POLB binding events on a genome-wide scale by ChIP-seq in MCF-7 cells and identified a total of 31,814 binding sites (**Fig. 4a**) for these proteins using MACS software[39]. We observed strong intensity for FOXA1 binding events across all its binding sites, with corresponding global co-enrichment of POLB binding, which had evident but weaker signal (**Fig. 4a**,**b**). To test the possibility that the FOXA1 DNA repair complex has different binding patterns in distinct chromatin environments, we partitioned the epigenome of MCF-7 cells (**Supplementary Figs. 4** and **5**) into seven states by ChromHMM[40] using published profiling data on six representative chromatin marks[41]. Although the FOXA1 complex showed prominent enrichment on active chromatin fractions, there were no discernible binding preferences for either FOXA1 or POLB with respect to the seven ChromHMM chromatin states (**Supplementary Fig. 4**). Notably, with ChIP-seq using antibody to POLB in FOXA1-depleted cells, we demonstrated the dependence of POLB on FOXA1 for its recruitment to chromatin. Although FOXA1 depletion in MCF-7 cells had no effect on the protein levels of DNA repair complex components (**Supplementary Fig. 3**), global enrichment of POLB binding was substantially decreased in these cells in comparison to control cells (**Fig. 4a**,**b**). Sequential ChIP (ChIP/re-ChIP) assays, using antibody to either FOXA1 or POLB in the first immunoprecipitation followed by another pulldown with the other antibody, were used to investigate the binding patterns of these proteins at two binding sites for the FOXA1 complex (**Fig. 4c**). The results demonstrated co-enrichment of these proteins on the same elements but not on control sites (**Fig. 4d** and **Supplementary Fig. 6**).

DNA methylation dynamics following FOXA1 depletion

We next investigated the functional relevance of the physical association of FOXA1 with DNA repair proteins. As stated above, active DNA demethylation requires the participation of DNA damage repair pathways, regardless of the distinctive nature of 5mC conversion[19]. To test whether the FOXA1 DNA repair complex is functionally linked to the regulation of DNA methylation at FOXA1-bound regions, we performed whole-genome bisulfite sequencing (BS-seq)[42] in MCF-7 cells (**Fig. 5**). We obtained methylomes at single-base resolution in two replicates with a combined total of 886 million mapped reads (100-bp paired-end reads). We found that the entire set of FOXA1-binding sites was associated with a prominent low level of DNA methylation (**Fig. 5a** and **Supplementary Fig. 7**). A simple explanation for this observation might be that the FOXA1 binding motif disfavors CpG dinucleotides (**Fig. 6b**). However, lineage-specific analysis showed that DNA methylation levels at FOXA1-binding sites that were either MCF-7 specific or shared by both MCF-7 and hepatocellular HepG2 (ref. 41) cells were much lower than those at sites unique for HepG2 cells (**Fig. 5b** and **Supplementary Fig. 8**), arguing against the simple explanation.

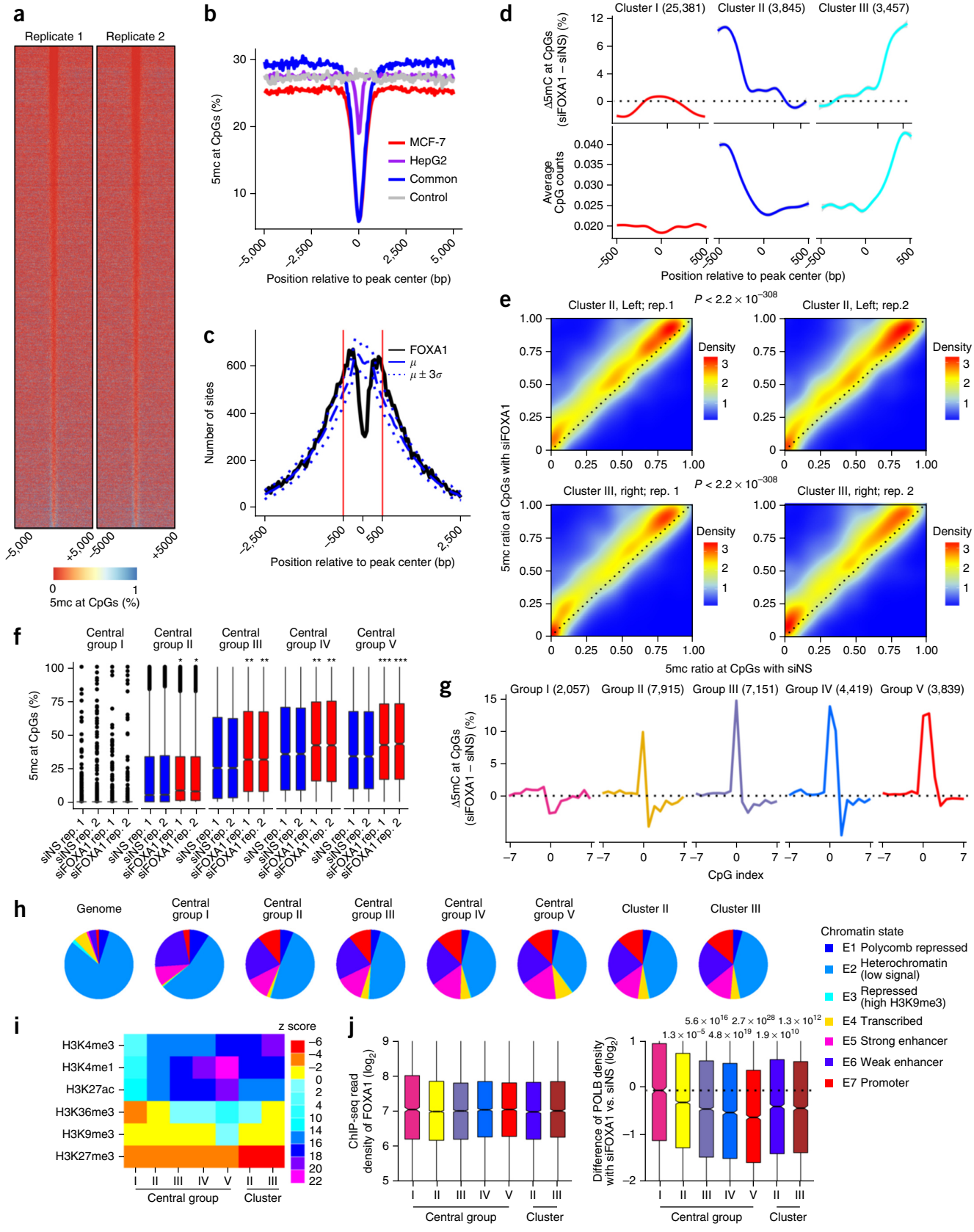
We then performed loss-of-function experiments in which we knocked down the expression of FOXA1 and probed DNA methylome changes using two high-coverage replicates assessed at single-base resolution (868 million mapped reads). Concomitantly, we refined the binding information for the FOXA1 complex to higher resolution by intersecting our data with ChIP–exonuclease results for FOXA1 in the same cell line[43]. We first determined the physical ranges with significant methylation changes after FOXA1 removal, which we identified by comparing the distribution of the most hypermethylated CpG dinucleotide within ±5 kb of a FOXA1 binding summit to the distribution in the background model (**Fig. 5c**). The results indicated that short windows of ±500 bp with respect to peak centers showed significant changes more than 3 s.d. from the mean distribution in null models). Thus, the downstream analysis was performed with methylation data on these 1-kb windows.

To investigate methylation changes for the total of 32,683 refined FOXA1 peaks, we employed a stepdown strategy with an initial coarse classification step using the *k*-means clustering method with *k* set to 3. This analysis indicated that two large sets with 3,845 (cluster II) and 3,457 (cluster III) sites showed asymmetrically increased methylation on either side of the FOXA1 summit, with the increased methylation constrained to a short patch and displaying a striking coincidence with the distribution of CpG dinucleotide content, whereas the remaining 25,381 sites in cluster I showed weak but centralized hypermethylation (**Fig. 5d**). Indeed, pairwise comparisons supported a reproducible and localized pattern of hypermethylation in these clusters ($P < 2.2 \times 10^{-308}$, one-tailed *t* test; **Fig. 5e** and **Supplementary Fig. 9**). For the majority of cluster I sites, because of the lack of precise FOXA1 binding information at single-base resolution and also because of CpG distribution, simple aggregation statistics to quantify CpG methylation signals might conceal the significance of methylation changes in short stretches of CpG dinucleotides as a result of the

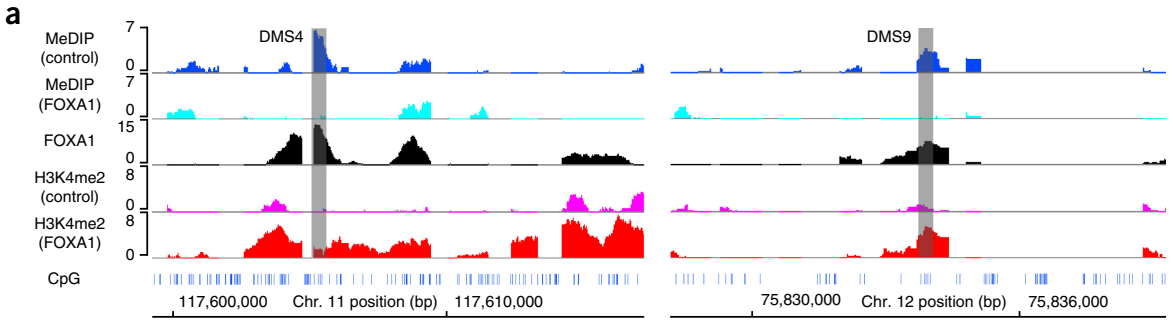
Figure 5 DNA methylation dynamics after FOXA1 depletion. (**a**) Heat map showing 5mC abundance along 10-kb windows centered on FOXA1-binding sites. (**b**) Average 5mC abundance for FOXA1-binding sites unique to MCF-7 or HepG2 cells, or present in both (common), and a set of random genomic intervals (control). (**c**) By considering methylation changes at single-base resolution after FOXA1 depletion, we calculated the distribution of the most hypermethylated CpG dinucleotides within ±5 kb of refined FOXA1 binding summits[43] and the distribution of these CpG sites in the background model. Lines corresponding to mean random density and ±3 s.d. from mean values are shown. μ, mean; σ, s.d. (**d**) Average changes in 5mC abundance with FOXA1 depletion (top) and average number of CpG dinucleotides (bottom) in the three clusters identified by *k*-means clustering. (**e**) Scatterplots showing mean 5mC abundance in the 350-bp window to the left of the FOXA1-binding site for sites in cluster II or the 350-bp window to the right of the FOXA1-binding site in sites from cluster III. rep., replicate. (**f**) The box plots show 5mC abundance within the 200-bp window centered on the LCH midpoint in the five subgroups from cluster I. $^{*}P < 1.0 \times 10^{-5}$, $^{**}P < 1.0 \times 10^{-100}$, $^{***}P < 1.0 \times 10^{-200}$, *t* test. (**g**) Average 5mC abundance for 15 CpG dinucleotides aligned with respect to the center of the LCH. (**h**,**i**) The differential association of epigenetic marks with each subgroup is demonstrated using pie charts of the seven ChromHMM epigenomic states (**Supplementary Fig. 4**) (**h**) or a heat map of *z* scores for enrichment of epigenetic marks against null models (**i**). (**j**) The box plots show FOXA1 binding intensity (left) or changes in POLB binding following FOXA1 depletion (right) for each subgroup. *P* values were calculated by *t* test. For the box plots in **f** and **j**, each box is the interquartile range and the line is the median.

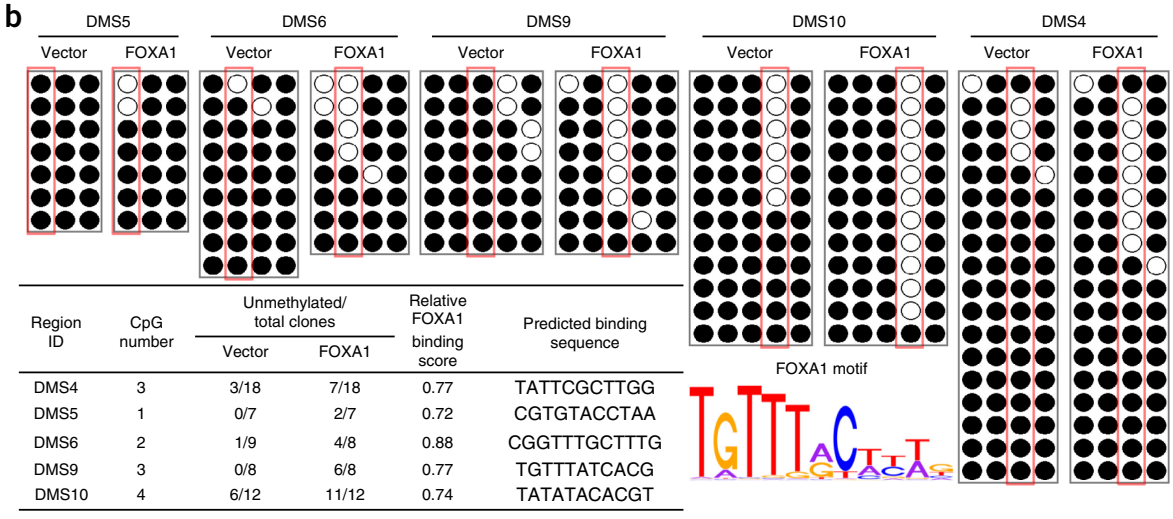
© 2016 Nature America, Inc. All rights reserved.

sequence-specific binding and confined influences of FOXA1. Therefore, we calculated the number of CpG dinucleotides within the longest consecutively hypermethylated (LCH) CpG stretch along each 1-kb window and stratified cluster I sites into five subgroups on the basis of this value (**Supplementary Fig. 10a**), reasoning that sites with longer stretches of consistently hypermethylated CpGs upon FOXA1

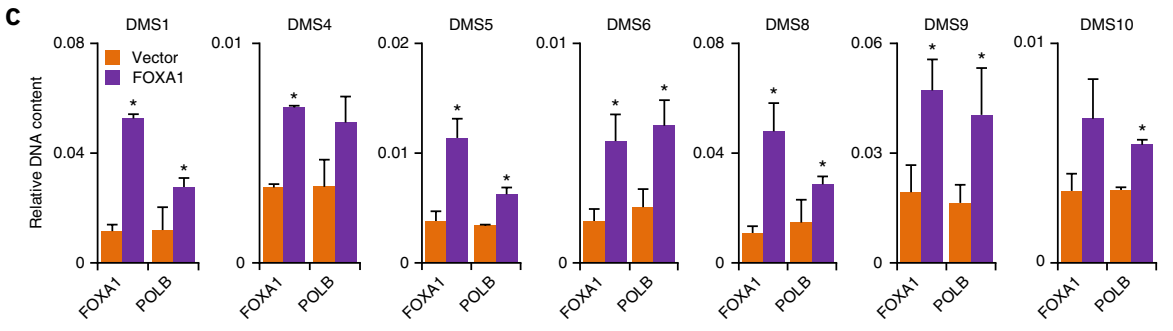


© 2016 Nature America, Inc. All rights reserved.





Region ID	CpG number	Unmethylated/ total clones		Relative FOXA1 binding score	Predicted binding sequence
		Vector	FOXA1		
DMS4	3	3/18	7/18	0.77	TATTCGCTTGG
DMS5	1	0/7	2/7	0.72	CGTGTACCTAA
DMS6	2	1/9	4/8	0.88	CGGTTTGCTTTG
DMS9	3	0/8	6/8	0.77	TGTTTATCACG
DMS10	4	6/12	11/12	0.74	TATATACACGT



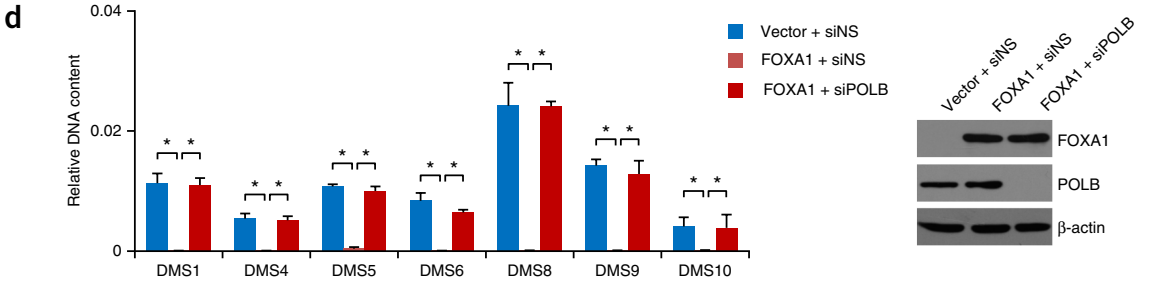


Figure 6 FOXA1 overexpression evokes POLB-dependent demethylation. (**a**) A set of genomic regions was selected on the basis of each region having high basal CpG methylation levels and high affinity for FOXA1 in MDA-MB-231 cells, according to previous chromosome-wide MeDIP on-chip data[16]. A schematic of ChIP on-chip MAT score[55] tracks at two representative sites (shaded boxes) is shown. (**b**) DNA methylation analysis by bisulfite sequencing of cloned PCR products. Top, BSP results for MDA-MB-231 cells infected with empty vector or overexpressing FOXA1 are shown as open or closed circles representing unmethylated and methylated CpGs, respectively. The red boxes highlight the closest CpG dinucleotides to the FOXA1 binding motif. Bottom left, signatures of FOXA1-associated DNA demethylation are listed. Bottom right, the motif used to predict FOXA1-binding sites in demethylated regions was retrieved from the JASPAR database. (**c**) qChIP analysis was performed in MDA-MB-231 cells infected with control lentivirus or lentivirus encoding *FOXA1* cDNA, using antibody against FOXA1 or POLB to detect enrichment at the indicated regions. Data are displayed as means ± s.d. for triplicate experiments (*$P < 0.05$, *t* test comparing the results from two conditions). (**d**) Demethylation was measured by MeDIP-coupled quantitative real-time PCR at the indicated regions. MDA-MB-231 cells infected with lentivirus encoding *FOXA1* cDNA were further challenged by POLB depletion (siPOLB). Left, genomic DNA was prepared and subjected to MeDIP. Data are displayed as means ± s.d. for triplicate experiments (*$P < 0.01$, one-way ANOVA). Right, relative protein expression was examined by immunoblotting.

© 2016 Nature America, Inc. All rights reserved.

depletion would exhibit greater dependency on the FOXA1 DNA repair complex to alter the regional epigenetic environment and transcriptional pioneering. The analysis demonstrated that three large subgroups with 7,151 (central group III), 4,419 (central group IV), and 3,839 (central groups V) sites showed reproducible and significant hypermethylation with two, three, and four or more CpGs in the LCH, respectively ($P < 1.0 \times 10^{-100}$, one-tailed t test; **Fig. 5f**,**g** and **Supplementary Figs. 10** and **11**). Thus, DNA hypermethylation following FOXA1 depletion marked localized features (median difference of ~10–20%) in cluster II and cluster III with flanking hypermethylation patterns and the three subgroups from cluster I with centralized hypermethylation patterns, summing to ~70% of the total FOXA1-binding sites.

Interestingly, by classifying all these groups of sites into interpretable epigenetic domains (**Supplementary Fig. 4**), we found that the central subgroups with longer LCHs contained a greater proportion of sites with an active chromatin state and that this pattern was most evident in central group V, with four or more CpGs in each LCH, whereas sites classified in the subgroups with zero or fewer than two hypermethylated CpGs following FOXA1 depletion were predominately in a repressed state (**Fig. 5h**). This observation was further confirmed by comparing z scores for enrichment of various histone modification marks in the center window of sites from a specified group to the genomic background (**Fig. 5i** and **Supplementary Fig. 12**). Notably, although occupancy of FOXA1 was independent of DNA methylation changes and exhibited similar enrichment across these groups, reduction of POLB binding resulted in a noticeable correlation with DNA hypermethylation upon FOXA1 knockdown (**Fig. 5j**), underscoring the importance of the DNA repair pathway in FOXA1-asscociated hypomethylation.

FOXA1 overexpression evokes POLB-dependent demethylation

The maintenance of low methylation levels in cells with high endogenous levels of FOXA1 might be validated with the initialization of DNA demethylation in FOXA1-negative cells through ectopic introduction of FOXA1. To test this possibility, we infected ER$^-$ human mammary carcinoma MDA-MB-231 cells with lentiviruses carrying *FOXA1* cDNA. We selected relevant genomic regions for analysis on the basis of them having high basal CpG methylation levels and high affinity for FOXA1, according to previous chromosome-wide data[16] (**Fig. 6** and **Supplementary Fig. 13**). Using bisulfite sequencing–coupled PCR (BSP), we noted that most of the tested regions experienced DNA demethylation with a localized feature: demethylated sites were confined to regions flanking FOXA1-binding sites without spreading to neighboring sites, and all regions had sparse CpG dinucleotides, differing from the traditional CpG island definition (**Fig. 6a**,**b**). The demethylation was accompanied by physical association of FOXA1 and POLB with the tested regions (**Fig. 6c**). The results were also independently validated using methylated DNA immunoprecipitation (MeDIP) assays (**Fig. 6d**). Moreover, the dependence of FOXA1-associated demethylation on the DNA repair factors nucleated with FOXA1 was further supported by epistasis analysis examining the DNA methylation pattern under overexpression of FOXA1 and knockdown of the key DNA polymerase POLB, which

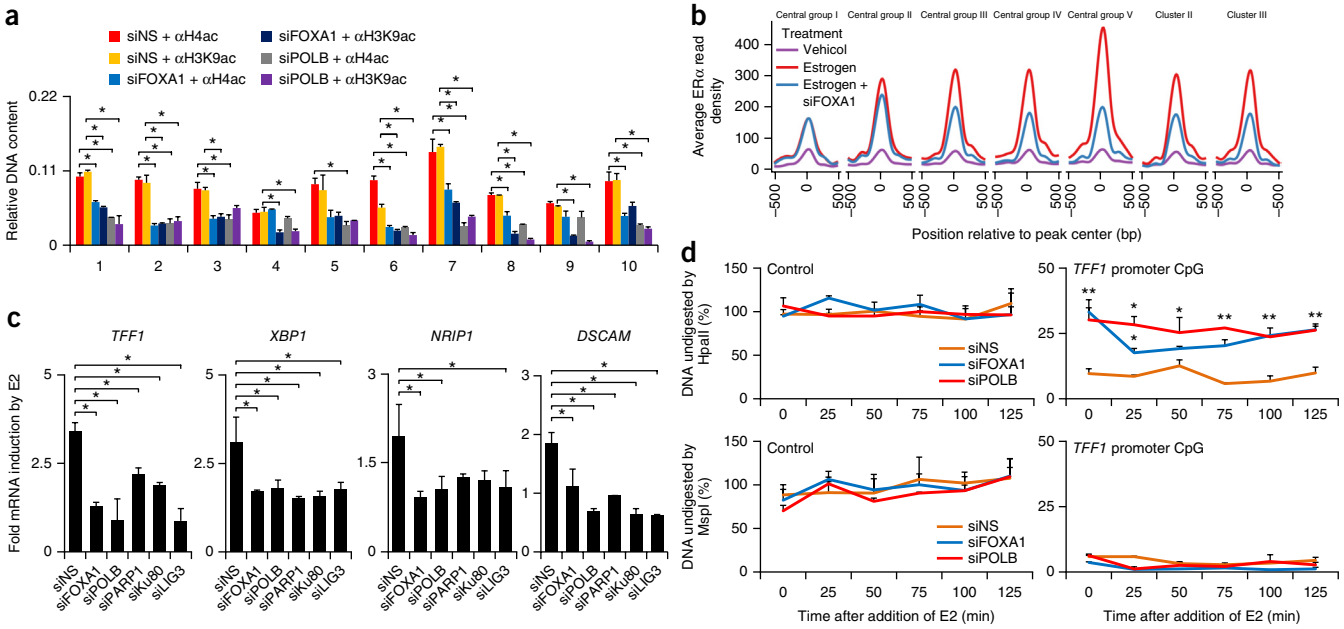


Figure 7 The FOXA1 DNA repair complex in estrogen response. (**a**) qChIP analysis was performed in MCF-7 cells transfected with siNS, siFOXA1, or siPOLB, using antibodies against acetylation of histone H3 at lysine 9 (H3K9ac) and histone H4 pan acetylation (H4ac). Data are displayed as means ± s.d. for triplicate experiments (*$P < 0.01$, one-way ANOVA). (**b**) Impact of DNA methylation on ERα loading. ERα ChIP-seq data[9] from hormone-deprived MCF-7 cells treated with vehicle or estrogen and from MCF-7 cells transfected with siFOXA1 in the presence of estrogen were used. The plot shows the average ChIP-seq read density for ERα in each group with differential hypermethylation patterns (**Fig. 5**). (**c**) MCF-7 cells were transfected with siNS or siRNAs targeting FOXA1 DNA repair factors. The cells were deprived of estrogen for 3 d and then treated with estradiol (E2) for 6 h before measurement of gene expression for *TFF1*, *XBP1*, *NRIP1*, and *DSCAM* by real-time RT–PCR. Data are displayed as means ± s.d. for triplicate experiments (*$P < 0.05$, one-way ANOVA). (**d**) Genomic DNA purified from MCF-7 cells transfected with siNS, siFOXA1, or siPOLB was digested with the methylation-sensitive restriction enzyme HpaII or with the methylation-insensitive isoschizomer MspI. Over a time course of estradiol exposure, the methylation status of a CCGG site in the *TFF1* promoter was measured by PCR with primers amplifying the CpG site. The methylation status of a control region in the *TFF1* promoter without any HpaII sites was also determined. Results were normalized to those for the same region amplified from samples without any restriction enzyme (undigested). Data are displayed as means ± s.d. for triplicate experiments (*$P < 0.05$, t test).

© 2016 Nature America, Inc. All rights reserved.

showed that depletion of POLB severely impaired FOXA1-associated demethylation at all tested regions (**Fig. 6d**).

The FOXA1 DNA repair complex in estrogen response

To link FOXA1-associated DNA hypomethylation to a biologically relevant response, we analyzed histone acetylation patterns at FOXA1-binding sites as an indicator of DNA methylation–mediated repressive chromatin context in FOXA1 complex–depleted MCF-7 cells. qChIP experiments indicated that levels of histone acetylation were significantly reduced at all the tested FOXA1-binding sites upon knockdown of FOXA1 or POLB (**Fig. 7a** and **Supplementary Fig. 14**).

We next tested the dependence of genomic binding of ERα on FOXA1-associated DNA methylation in MCF-7 cells. We analyzed estrogen-induced ERα recruitment and FOXA1-depletion-associated ERα unloading in the seven groups of FOXA1-binding sites showing different hypermethylation patterns (**Fig. 5**) using previous genome-wide data[9]. Interestingly, the central groups with no or single CpG sites in their LCHs had weak ERα engagement after estrogen stimulation and also negligible ERα unloading when FOXA1 was depleted (**Fig. 7b**). Robust FOXA1-dependent ERα loading was detected in central groups with longer LCHs and in the two clusters with flanking hypermethylation patterns, a trend that was maximized in the group with the most extensively affected hypermethylation ranges (**Fig. 7b**). We further annotated the genes associated with different subgroups through GREAT[44] and noted the significant number of biological processes, especially estrogen-related pathways, in the groups with longer LCHs (**Supplementary Tables 2** and **3**).

To directly test the functional relevance of the FOXA1 DNA repair complex in estrogen-triggered transcription, we knocked down the key components of the DSBR and SSBR complexes in MCF-7 cells and measured the expression of the ER target genes *TFF1*, *XBP1*, *NRIP1*, and *DSCAM*. Depletion of any of the tested components severely affected FOXA1-dependent transcription of ER target genes (**Fig. 7c**). We further demonstrated the causal role of FOXA1 and POLB in DNA demethylation using methylation-sensitive restriction enzymes, finding that DNA hypermethylation after FOXA1 or POLB knockdown occurred at a critical CpG site in the *TFF1* promoter (**Fig. 7d**). By using an experimental system in MDA-MB-231 cells[12], we also demonstrated the need for FOXA1 and POLB in DNA demethylation, ERα recruitment, and the 'reprogramming' of estrogen responsiveness (**Supplementary Fig. 15**).

DISCUSSION

The winged-helix transcription factor FOXA1 has been implicated in both epigenetic reprogramming and pioneering competency[7]. However, although the number of proteins with pioneering potential has expanded greatly since the original findings of a panel of hepatic nuclear factors, there is a lack of proportionate mechanistic insight into the enigmatic activity of pioneering competency, especially given the structural diversity of newly identified pioneer factors lacking the histone H1–like winged-helix domain[45–47]. To address this need, we characterized the FOXA1 interactome in HeLa cells by a proteomic approach to identify proteins that are potentially associated with FOXA1. Unexpectedly, we found that FOXA1 is associated with a number of proteins that are components of SSBR and DSBR DNA repair complexes.

We identified seven core components in the FOXA1-containing complex, including Ku80, Ku70, PARP1, POLB, LIG3, and XRCC1. Although this multiple-subunit complex cannot be assigned as a stable stoichiometric complex because of the incomplete overlap of the fractionation profiles of FOXA1 and DNA repair factors, it adheres to the principle of transient regulatory complex–complex interaction networks[48]. This assignment is logical because DNA repair factories would be otherwise restricted and less accessible to detect haphazard damage if stably associated with a sequence-specific DNA-binding factor, and it is also biologically meaningful to allow the dispatch of limited sets of DNA repair proteins to different missions. Indeed, functionally, we found that gain of function of FOXA1 is associated with active removal of DNA methylation, and loss of function of FOXA1 was accompanied by DNA hypermethylation. Mechanistically, we showed that FOXA1-associated DNA demethylation occurs in both a FOXA1-dependent and DNA-repair-factor-dependent manner and exhibits a sequence-specific pattern, differing from TET-catalyzed 5mC oxidation, which seems to occur on general sets of regulatory elements without apparent specification of unique DNA sequences[27,49]. However, it would not be surprising if the FOXA1-nucleated machinery and the TET system intersect at certain levels. This possibility clearly warrants future investigations, especially because we were unable to identify enzymes in the FOXA1 DNA repair complex that are required to catalyze chemical reactions during DNA demethylation. Moreover, clarification of this point also depends on identification of the glycosylase(s) used in the demethylation system, which, unfortunately, generally do not co-purify with repair proteins, as the interactome of thymine DNA glycosylase (TDG) does not contain substantial levels of DNA repair factors[50].

Despite the lack of a unifying mechanistic model to reconcile the findings of active DNA demethylation across different species, a consensus scheme is emerging in which DNA repair factors seem obligatory for active DNA demethylation[19,23,24,26,51]. It is increasingly clear that DNA repair is not restricted to maintenance of genomic integrity in response to environmentally inflicted lesions during replication or recombination but that it is in fact at the heart of an epigenetic regulatory system. It also seems likely that the DNA repair mechanism might have initially been selected as one of the elements involved in DNA demethylation and was subsequently coopted to fulfill the mission of DNA repair during evolution[52]. Along with the general concept of correlation between DNA hypomethylation and transcription factor binding[42,53], linking FOXA1 to DNA repair factors and to DNA demethylation constitutes an attractive model to understand the transcriptional pioneering activity of FOXA1.

METHODS

Methods and any associated references are available in the online version of the paper.

Accession codes. The deep sequencing data reported in this manuscript have been deposited in the Gene Expression Omnibus under accession GSE80808. The mass spectrometry proteomics data have been deposited at the ProteomeXchange via the PRIDE[54] partner repository with data set identifier PXD004271.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

This work was supported by grants (91219201, 81530073, and 81130048 to Y.S. and 81300254 to Y.Z.) from the National Natural Science Foundation of China and a grant (973 Program: 2014CB542004 to J.L.) from the Ministry of Science and Technology of China.

AUTHOR CONTRIBUTIONS

Y.S. and Y.Z. conceived the project. Y.S. and Y.Z. conceived and designed the experiments. Y.Z. and D.Z. performed the experiments. Y.Z. performed computational analysis. Y.S. and Y.Z. wrote the manuscript. Y.S., Y.Z., D.Z.,

© 2016 Nature America, Inc. All rights reserved.

J.L., and Q.L. analyzed the data. Y.Z., D.Z., J.L., Q.L., L.S., X.Y., Z.C., R.Y., G.X., S.L., B.X., W.L., L.L., J.Y., and L.H. performed material preparation.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Augello, M.A., Hickey, T.E. & Knudsen, K.E. FOXA1: master of steroid receptor function in cancer. *EMBO J.* **30**, 3885–3894 (2011).
2. Katoh, M. & Katoh, M. Human FOX gene family (Review). *Int. J. Oncol.* **25**, 1495–1500 (2004).
3. Wijchers, P.J., Burbach, J.P. & Smidt, M.P. In control of biology: of mice, men and Foxes. *Biochem. J.* **397**, 233–246 (2006).
4. Carroll, J.S. *et al.* Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**, 33–43 (2005).
5. 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).
6. Li, Q. *et al.* FOXA1 mediates $p16^{INK4a}$ activation during cellular senescence. *EMBO J.* **32**, 858–873 (2013).
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. 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).
10. Wang, D. *et al.* Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**, 390–394 (2011).
11. Wang, Q. *et al.* Androgen receptor regulates a distinct transcription program in androgen-independent prostate cancer. *Cell* **138**, 245–256 (2009).
12. Kong, S.L., Li, G., Loh, S.L., Sung, W.K. & Liu, E.T. Cellular reprogramming by the conjoint action of ERα, FOXA1, and GATA3 to a ligand-inducible growth state. *Mol. Syst. Biol.* **7**, 526 (2011).
13. 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).
14. Cirillo, L.A. & Zaret, K.S. An early developmental transcription factor complex that is more stable on nucleosome core particles than on free DNA. *Mol. Cell* **4**, 961–969 (1999).
15. Clark, K.L., Halay, E.D., Lai, E. & Burley, S.K. Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* **364**, 412–420 (1993).
16. Sérandour, A.A. *et al.* Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Res.* **21**, 555–565 (2011).
17. Bartke, T. *et al.* Nucleosome-interacting proteins regulated by DNA and histone methylation. *Cell* **143**, 470–484 (2010).
18. Bernstein, B.E., Meissner, A. & Lander, E.S. The mammalian epigenome. *Cell* **128**, 669–681 (2007).
19. Wu, H. & Zhang, Y. Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45–68 (2014).
20. Wu, S.C. & Zhang, Y. Active DNA demethylation: many roads lead to Rome. *Nat. Rev. Mol. Cell Biol.* **11**, 607–620 (2010).
21. Zhu, J.K. Active DNA demethylation mediated by DNA glycosylases. *Annu. Rev. Genet.* **43**, 143–166 (2009).
22. Conticello, S.G. The AID/APOBEC family of nucleic acid mutators. *Genome Biol.* **9**, 229 (2008).
23. Cortellino, S. *et al.* Thymine DNA glycosylase is essential for active DNA demethylation by linked deamination-base excision repair. *Cell* **146**, 67–79 (2011).
24. Rai, K. *et al.* DNA demethylation in zebrafish involves the coupling of a deaminase, a glycosylase, and Gadd45. *Cell* **135**, 1201–1212 (2008).
25. Maiti, A. & Drohat, A.C. Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* **286**, 35334–35338 (2011).
26. Dalton, S.R. & Bellacosa, A. DNA demethylation by TDG. *Epigenomics* **4**, 459–467 (2012).
27. Shen, L. *et al.* Genome-wide analysis reveals TET- and TDG-dependent 5-methylcytosine oxidation dynamics. *Cell* **153**, 692–706 (2013).
28. Lloyd, R.S. The initiation of DNA base excision repair of dipyrimidine photoproducts. *Prog. Nucleic Acid Res. Mol. Biol.* **62**, 155–175 (1999).
29. Schärer, O.D. & Jiricny, J. Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *BioEssays* **23**, 270–281 (2001).
30. Whitehouse, C.J. *et al.* XRCC1 stimulates human polynucleotide kinase activity at damaged DNA termini and accelerates DNA single-strand break repair. *Cell* **104**, 107–117 (2001).
31. Ma, Y. *et al.* A biochemically defined system for mammalian nonhomologous DNA end joining. *Mol. Cell* **16**, 701–713 (2004).
32. Ma, Y., Pannicke, U., Schwarz, K. & Lieber, M.R. Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell* **108**, 781–794 (2002).
33. Burma, S. & Chen, D.J. Role of DNA-PK in the cellular response to DNA double-strand breaks. *DNA Repair (Amst.)* **3**, 909–918 (2004).
34. Kubota, Y. *et al.* Reconstitution of DNA base excision-repair with purified human proteins: interaction between DNA polymerase β and the XRCC1 protein. *EMBO J.* **15**, 6662–6670 (1996).
35. Caldecott, K.W., Aoufouchi, S., Johnson, P. & Shall, S. XRCC1 polypeptide interacts with DNA polymerase β and possibly poly (ADP-ribose) polymerase, and DNA ligase III is a novel molecular 'nick-sensor' *in vitro*. *Nucleic Acids Res.* **24**, 4387–4394 (1996).
36. Kim, M.Y., Zhang, T. & Kraus, W.L. Poly(ADP-ribosyl)ation by PARP-1: 'PAR-laying' NAD^+ into a nuclear signal. *Genes Dev.* **19**, 1951–1967 (2005).
37. Kim, M.Y., Mauro, S., Gévry, N., Lis, J.T. & Kraus, W.L. NAD^+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1. *Cell* **119**, 803–814 (2004).
38. Krishnakumar, R. & Kraus, W.L. PARP-1 regulates chromatin structure and transcription through a KDM5B-dependent pathway. *Mol. Cell* **39**, 736–749 (2010).
39. Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137 (2008).
40. Ernst, J. & Kellis, M. ChromHMM: automating chromatin-state discovery and characterization. *Nat. Methods* **9**, 215–216 (2012).
41. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
42. Stadler, M.B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).
43. Serandour, A.A., Brown, G.D., Cohen, J.D. & Carroll, J.S. Development of an Illumina-based ChIP–exonuclease method provides insight into FoxA1–DNA binding properties. *Genome Biol.* **14**, R147 (2013).
44. McLean, C.Y. *et al.* GREAT improves functional interpretation of *cis*-regulatory regions. *Nat. Biotechnol.* **28**, 495–501 (2010).
45. Menet, J.S., Pescatore, S. & Rosbash, M. CLOCK:BMAL1 is a pioneer-like transcription factor. *Genes Dev.* **28**, 8–13 (2014).
46. Pihlajamaa, P. *et al.* Tissue-specific pioneer factors associate with androgen receptor cistromes and transcription programs. *EMBO J.* **33**, 312–326 (2014).
47. Sherwood, R.I. *et al.* Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat. Biotechnol.* **32**, 171–178 (2014).
48. Malovannaya, A. *et al.* Analysis of the human endogenous coregulator complexome. *Cell* **145**, 787–799 (2011).
49. Yu, M. *et al.* Base-resolution analysis of 5-hydroxymethylcytosine in the mammalian genome. *Cell* **149**, 1368–1380 (2012).
50. Spruijt, C.G. *et al.* Dynamic readers for 5-(hydroxy)methylcytosine and its oxidized derivatives. *Cell* **152**, 1146–1159 (2013).
51. Ooi, S.K. & Bestor, T.H. The colorful history of active DNA demethylation. *Cell* **133**, 1145–1148 (2008).
52. Billaud, M. & Santoro, M. Is co-option a prevailing mechanism during cancer progression? *Cancer Res.* **71**, 6572–6575 (2011).
53. Ziller, M.J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481 (2013).
54. Vizcaíno, J.A. *et al.* 2016 update of the PRIDE database and its related tools. *Nucleic Acids Res.* **44**, D1, D447–D456 (2016).
55. Johnson, W.E. *et al.* Model-based analysis of tiling-arrays for ChIP-chip. *Proc. Natl. Acad. Sci. USA* **103**, 12457–12462 (2006).

© 2016 Nature America, Inc. All rights reserved.

ONLINE METHODS

Antibodies and reagents. Antibodies to FOXA1 (ab23738), XRCC1 (ab1838), POLB (ab3181), LIG3 (ab587), DNA-PKcs (ab1832), LIG4 (ab26039), and TOP2β (ab109524) were purchased from Abcam; antibodies to Ku70 (sc-12729), Ku80 (sc-1485), and PARP1 (sc-56197) were purchased from Santa Cruz Biotechnology; and antibody to FLAG (F3165) was purchased from Sigma. 17-β-estradiol (E2) was purchased from Sigma and used at a 100 nM concentration. Transfection with plasmid DNA was performed with polyethylenimine (Sigma) or Lipofectamine 2000 (Invitrogen) following the manufacturer's instructions, and Lipofectamine RNAiMAX (Invitrogen) was used for transfection with double-stranded siRNA.

Cell lines. The cell lines used were obtained from the American Type Culture Collection. MCF-7 and HeLa cells were cultured in DMEM supplemented with 10% FBS. Cells were maintained in a humidified incubator equilibrated with 5% CO_2 at 37 °C. LNCaP cells were cultured in RPMI-1640 medium supplemented with 10% FBS and maintained in an incubator with 5% CO_2 at 37 °C. All cell lines were tested negative for mycoplasma.

Small interfering RNA. siRNAs were synthesized by Shanghai GenePharma. They were used at a concentration of 20–50 nM. The siRNA targeting sequences are listed in **Supplementary Table 4**.

Immunopurification and mass spectrometry. HeLa cells stably expressing FLAG-FOXA1 were lysed in buffer A (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5% NP-40, 50 mM sodium fluoride, 1 mM DTT, 0.5% sodium deoxycholate, and 1 mM phenylmethylsulfonyl fluoride (PMSF) plus protease inhibitors (Roche)), and lysates were incubated with anti-FLAG M2 beads (Sigma) for 2 h at 4 °C. After washing three times with buffer B (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 0.5% NP-40), FLAG peptide (Sigma) was used to elute the protein complex from the beads following the manufacturer's instructions. The eluted protein complex was then collected and resolved by SDS–PAGE, silver stained, and subjected to liquid chromatography/tandem mass spectrometry for sequencing and data analysis. The peptide sequences identified are shown in **Supplementary Table 1**.

FPLC chromatography. MCF-7 nuclear extracts were prepared, dialyzed against buffer C (20 mM HEPES, pH 8.0, 10% glycerol, 0.1 mM EDTA, and 150 mM NaCl) and then applied to a Superdex 200 column (GE Healthcare) that had been equilibrated with buffer C and calibrated with protein standards. The column was eluted at a flow rate of 0.5 ml/min, and fractions were collected. Alternatively, complex containing endogenous FOXA1 was prepared and subjected to size-exclusion chromatography. In this case, the FOXA1 complex was first enriched by antibody-mediated affinity purification (using antibody to FOXA1; ab23738, Abcam) from MCF-7 nuclear extracts. After three extensive washes with buffer B, the protein complex was competitively eluted with a synthesized FOXA1 peptide (IEPSALEPAYYQGVYSRPVLNTS; MBL), which was the immunogen used to generate the ab23738 antibody to FOXA1. The eluates were then fractionated on a Superdex 200 column.

Immunoprecipitation and immunoblotting. Immunoprecipitation was performed with MCF-7 or LNCaP cellular extract and 2 μg of antibody in a final volume of 500 μl in buffer A. After incubation overnight at 4 °C under constant rotation, 50 μl of 50% protein G beads was added and the mixture was incubated for two more hours. Beads were precipitated by centrifugation (5 min at 500*g*) and were extensively washed five times with buffer A. Washed beads were mixed with 2× SDS–PAGE loading buffer, and samples were analyzed immunoblotting. Immunodetection was performed using enhanced chemiluminescence (ECL System, Amersham Biosciences) according to the manufacturer's instructions.

***In vitro* reconstitution and pulldown assays.** FLAG-tagged FOXA1, Ku70, Ku80, and PARP1 expression plasmids were generated by inserting the corresponding cDNAs into pcDNA3.1(+) vector. Plasmids for GST-FOXA1 and the GST-tagged FOXA1 mutants (ΔN, ΔDBD, ΔC, Nt, DBD, and Ct) were generated by subcloning from the pcDNA3.1(+) vectors encoding FLAG-tagged FOXA1. PET28a vectors encoding LIG3, POLB, and XRCC1 were generously provided by J.L. Parsons (University of Liverpool) and K. Caldecott (University of Sussex). GST fusion proteins and His-tagged proteins were expressed in *E. coli* strain BL21 and purified with Glutathione Sepharose 4B beads (GE Healthcare) and Ni-NTA resin (Qiagen), respectively, according to the manufacturer's instructions. FLAG-tagged Ku70, Ku80, and PARP1 for GST pulldown assays were transcribed and translated *in vitro* (TNT Systems, Promega). For *in vitro* reconstitution and pulldown assays, FLAG, GST, or His fusion proteins were first immobilized on affinity gels and then incubated with the proteins purified or transcribed and translated *in vitro*. After extensive washing, the protein complex was eluted with SDS–PAGE loading buffer and subjected to immunoblotting.

Baculovirus production. For baculovirus expression, cDNAs encoding FOXA1 were inserted into the pFastBac HT A vector (Invitrogen) with a sequence encoding a C-terminal FLAG tag for further affinity purification. The recombinant baculovirus construct was then transfected into Sf9 cells (Invitrogen), and the viruses were generated and amplified according to the manufacturer's protocol. To purify FOXA1 proteins, cells were lysed by sonication and incubated with anti-FLAG M2 beads. After washing three times with buffer A, FLAG peptide was applied to the beads to elute the FOXA1 protein.

ChIP and sequential ChIP. ChIP and sequential ChIP were conducted as previously described[56–58]. Briefly, cells were cross-linked using 1% formaldehyde at room temperature for 10 min, lysed, and sonicated on ice. The sheared chromatin was precleared, incubated with specific antibodies, and washed with low- and high-salt buffers. Then, the DNA was eluted and purified as template for PCR. For sequential ChIP, bead eluates from the first immunoprecipitation were incubated with 10 mM DTT at 37 °C for 30 min, and the resulting samples were diluted 1:50 in dilution buffer (1% Triton X-100, 2 mM EDTA, 150 mM NaCl, and 20 mM Tris-HCl, pH 8.1) followed by immunoprecipitation with the second antibody. The primer sequences for real-time PCR are provided in **Supplementary Table 5**.

Real-time RT–PCR. Total cellular RNA was extracted with TRIzol reagent (Sigma) and used for first-strand cDNA synthesis with the MMLV reverse transcription system (Promega). Quantification of all gene transcripts was performed on an ABI PRISM 7300 instrument (Applied Biosystems) using SYBR Green Master Mix, and RNA levels were normalized to those for *GAPDH* (glyceraldehyde 3-phosphate dehydrogenase). Primers are described in **Supplementary Table 5**.

Lentiviral production and infection. 293T cells were transfected with pLenti6/V5-DEST vector (Invitrogen) encoding FOXA1 and three other helper plasmids (REV, RRE, and VSVG). Forty-eight hours after transfection, the supernatant of transfectants was collected and filtered through a 0.45-μm cellulose acetate filter (Millipore). MDA-MB-231 cells at 20–30% confluence were infected with viruses supplemented with 4 μg/ml polybrene (Millipore).

DNA methylation analysis. Bisulfite treatment and sequencing were performed as described[59] with minor modifications. Briefly, genomic DNA was isolated from MDA-MB-231 cells with SDS treatment and proteinase K digestion followed by phenol-chloroform extraction. One microgram of genomic DNA was first denatured with NaOH for 10 min at 42 °C and then treated with 10 μl of 10 mM hydroquinone and 520 μl of 3 M sodium bisulfite at 50 °C overnight. Modified DNA was purified with a QIAquick PCR Purification kit (Qiagen) and further incubated with 0.3 N NaOH for 5 min. The resulting DNA was purified by phenol-chloroform extraction and dissolved in TE buffer (10 mM Tris-HCl and 1 mM EDTA, pH 8.0). PCR was then performed with the primers described in **Supplementary Table 5**, and the products were cloned into the pEASY-T1 vector (Transgene) for subsequent sequencing. The methylation status of CCGG sites was analyzed through digestion of purified DNA with the methylation-sensitive restriction enzyme HpaII or with the methylation-insensitive isoschizomer MspI for 2 h at 37 °C. Once purified, the digested DNA was subjected to PCR.

© 2016 Nature America, Inc. All rights reserved.

doi:10.1038/ng.3635

Methylated DNA immunoprecipitation. Genomic DNA from MCF-7 or MDA-MB-231 cells was prepared by RNase digestion, proteinase K treatment, and phenol-chloroform extraction. The DNA was sonicated to produce fragments ranging in size from 300 to 1,000 bp, which were then denatured by incubation for 10 min at 95 °C and subjected to immunoprecipitation with 3 μg of monoclonal antibody against 5mC (Eurogentec, BI-MECY-0100). After extensive washes, methylated DNA was eluted and further purified through a QIAquick column. Quantitative real-time PCR was then performed to measure DNA methylation levels using the primers described in **Supplementary Table 5**.

ChIP-seq and whole-genome bisulfite sequencing. Genomic DNA for ChIP-seq was prepared as described[60]. BS-seq was performed as described[42] with few modifications. For construction of the sequencing library, the DNA was fragmented by sonication using a Bioruptor (Diagenode) to a mean size of approximately 250 bp, followed by blunt-ending, polyadenylation, and adaptor ligation (in the case of methylated adaptors to protect from bisulfite conversion), essentially according to the manufacturer's instructions. Ligated DNA was bisulfite converted using the EZ DNA Methylation-Gold kit (Zymo). Fragments corresponding to different insert sizes were excised from the same lane of a 2% TAE agarose gel. Products were purified using the QIAquick Gel Extraction kit (Qiagen) and amplified by PCR. In-depth whole-genome DNA sequencing was performed by the BGI using the Illumina HiSeq 4000 or HiSeq 2500 platform. Sequencing reads were aligned against the GRCh37 human reference genome using Bowtie[61].

Bioinformatic analysis. In-depth whole-genome DNA sequencing was performed by BGI. Peaks from ChIP-seq libraries were called using MACS[39] with default parameters. Cell-lineage-specific FOXA1-binding sites were determined by BEDOPS[62]. To obtain normalized read density for ChIP-seq, we calculated read coverage at each base and then normalized this value by the total number of mappable reads in that library. To refine the FOXA1 binding information, we have intersected our ChIP-seq data with ChIP–exonuclease results[43] and retrieved all short ChIP–exonuclease peaks that overlapped any FOXA1 peaks from our experiment by at least one base. A total of 32,683 ChIP–exonuclease peaks were defined that overlapped with ~85% of the FOXA1 ChIP-seq peaks in our data. For BS-seq, reads were first filtered for adaptor sequences and contamination, and low-quality reads were removed. Then, we mapped the clean reads to the reference genome by BSMAP[63], removed duplicate reads, and merged the mapping results according to each library. Here we calculated the mapping rate and bisulfite conversion rate for each sample. Methylation levels were determined by dividing the number of reads corresponding to each 5mC site by the total number of reads covering that cytosine, which was also equal to the 5mC/C ratio at each reference cytosine[64]. The DNA methylation change at each CpG dinucleotide surrounding FOXA1 binding summits was calculated as the difference in CpG levels between FOXA1-depleted cells and control cells. LCHs were hence defined as the largest set of consecutive CpGs with increased methylation within ±500 bp of a FOXA1 peak summit. We used the GenomicRanges package[65] in R (ref. 66) to handle annotated genomic ranges. We also used R to perform *k*-means clustering and used ggplot2 (ref. 67) to produce heat maps and other statistical plots. ChromHMM was used with default parameters[40]. Enrichment of histone marks in each subgroup was calculated as a *z* score against values from 100 iterations in the background groups with randomly permutated locations serving as null models. We also used some published high-throughput sequencing data from the Gene Expression Omnibus database and the Encyclopedia of DNA Elements (ENCODE)[41] for H3K4me3 (GSM945269), H3K9me3 (GSM945857), H3K27me3 (GSM970218), H3K36me3 (GSM970217), H3K4me1 (ref. 68) (GSM1115994), H3K27ac (GSM945854), and p300 (ref. 69) (GSM1470014); ERα data were obtained under the data series accession GSE25710 (ref. 9).

Statistics. Statistical analysis and *P*-value cutoffs are indicated in the relevant figures.

56. Shang, Y., Hu, X., DiRenzo, J., Lazar, M.A. & Brown, M. Cofactor dynamics and sufficiency in estrogen receptor–regulated transcription. *Cell* **103**, 843–852 (2000).
57. Zhang, H. *et al.* Differential gene regulation by the SRC family of coactivators. *Genes Dev.* **18**, 1753–1765 (2004).
58. Wang, Y. *et al.* LSD1 is a subunit of the NuRD complex and targets the metastasis programs in breast cancer. *Cell* **138**, 660–672 (2009).
59. Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. *Proc. Natl. Acad. Sci. USA* **89**, 1827–1831 (1992).
60. Schmidt, D. *et al.* ChIP-seq: using high-throughput sequencing to discover protein–DNA interactions. *Methods* **48**, 240–248 (2009).
61. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
62. Neph, S. *et al.* BEDOPS: high-performance genomic feature operations. *Bioinformatics* **28**, 1919–1920 (2012).
63. Xi, Y. & Li, W. BSMAP: whole genome bisulfite sequence MAPping program. *BMC Bioinformatics* **10**, 232 (2009).
64. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
65. Lawrence, M. *et al.* Software for computing and annotating genomic ranges. *PLoS Comput. Biol.* **9**, e1003118 (2013).
66. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2014).
67. Wickham, H. *ggplot2: Elegant Graphics for Data Analysis* (Springer-Verlag, 2009).
68. Li, W. *et al.* Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. *Nature* **498**, 516–520 (2013).
69. Liu, Z. *et al.* Enhancer activation requires *trans*-recruitment of a mega transcription factor complex. *Cell* **159**, 358–373 (2014).

© 2016 Nature America, Inc. All rights reserved.